UNIVERSIDADE DE LISBOA

Faculdade de Medicina

Unravelling longitudinal mitochondrial DNA mutations in Multiple Sclerosis: Association with disease activity and progression

Filipe Carlos Pereira dos Reis Cortes Figueiredo

Orientadores:

Prof. Doutor Friedemann Paul

Prof.ª Doutora Vanessa Alexandra Morais

Tese especialmente elaborada para obtenção do grau de

Doutor em Ciências Biomédicas

Especialidade em Neurociências

2022

Page intentionally left blank

UNIVERSIDADE DE LISBOA

Faculdade de Medicina

Unravelling longitudinal mitochondrial DNA mutations in Multiple Sclerosis: Association with disease activity and progression

Filipe Carlos Pereira dos Reis Cortes Figueiredo

Orientadores:
Prof. Doutor Friedemann Paul
Prof.ª Doutora Vanessa Alexandra Morais

Tese especialmente elaborada para obtenção do grau de
Doutor em Ciências Biomédicas
Especialidade em Neurociências

Júri:

Presidente: Doutor José Augusto Gamito Melo Cristino, Professor Catedrático e Presidente do Conselho Científico da Faculdade de Medicina da Universidade de Lisboa

Vogais:

– Doutora Zofia Chrzanowska-Lightowlers, Full Professor da Newcastle University (United Kingdom);

– Doutora Daniela Galimberti, Professore Associato da Università Degli Studi di Milano (Italy);

– Doutora Ana Maria Ferreira de Sousa Sebastião, Professora Catedrática da Faculdade de Medicina da Universidade de Lisboa;

– Doutor Nuno Luís Barbosa Morais, Professor Associado Convidado da Faculdade de Medicina da Universidade de Lisboa;

– Doutora Vanessa Alexandra dos Santos Morais Epifânio, Professora Associada Convidada da Faculdade de Medicina da Universidade de Lisboa;

– Doutor João Carlos Correia de Sá, Professor Auxiliar Convidado da Faculdade de Medicina da Universidade de Lisboa.

Instituições Financiadoras: Fundação para a Ciência e a Tecnologia (*PD/BD/114122/2015* e *FCT/PTDC/MED-NEU/7976/2020*), Fundação para a Ciência e a Tecnologia e Programa Operacional Regional de Lisboa PORTUGAL 2020 (*PAC-PRECISE LISBOA-01-0145-FEDER-016394*), Merck Germany (restricted research grant), National Multiple Sclerosis Society dos Estados Unidos da América *(PP-1712-29466*)

2022

Page intentionally left blank

A impressão desta tese foi aprovada pelo Conselho Científico da Faculdade de Medicina de Lisboa em reunião de 26 de Abril de 2022.

As opiniões expressas nesta publicação são da exclusiva responsabilidade do seu autor.

Page intentionally left blank

Finally

I'm crossing the threshold

From the ordinary world

To the reveal of my heart

Kieron Menzies, Rick Nowels & Lana Del Rey

Get Free © Universal Music Group, Universal Music Publishing

Page intentionally left blank

TABLE OF CONTENTS

ACKNOWLEDGMENTS XIII

FUNDING XV

MOST RELEVANT ABBREVIATIONS XVII

FIGURES XIX

TABLES XXI

RESUMO XXIII

SUMMARY XXVII

I — INTRODUCTION 1

1.1. The nervous system 3

1.1.1. The central nervous system 4

1.1.2. The peripheral nervous system 8

1.1.3. The cellular constituents of the nervous system 8

1.2. The immune system 12

1.2.1. The innate immune system 15

1.2.2. The adaptive immune system 21

1.3. Neuroimmunology 28

1.4. Multiple Sclerosis 33

1.4.1. Definition and history 33

1.4.2. Epidemiology 34

1.4.3. Clinical course and management 36

1.4.4. Pathology 43

1.4.5. Pathophysiology 47

1.4.6. Etiology 49

1.4.6.1. Environmental risk factors 49

1.4.6.2. Genetic risk factors 52

1.4.6.2.1. Mitochondrial DNA 55
II — OBJECTIVES 71
III — MATERIALS AND METHODS 77
3.1. PrecisionCallerPipeline 79
3.1.1. Sample acquisition/collection 79
3.1.2. DNA sequencing 80
3.1.3. Bioinformatic processing 81
3.1.3.1. PrecisionCallerPipeline 81
3.1.3.2. Torrent Suite™ Software 82
3.1.4. Data analysis 83
3.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort 84
3.2.1. Cohort description 84
3.2.2. Sample processing 85
3.2.2.1. Sample transportation, thawing, and initial processing 85
3.2.2.2. $CD4^+$ T cell enrichment and storage 85
3.2.2.3. Flow cytometry analysis 87
3.2.2.4. DNA extraction, enriched in mtDNA 89
3.2.3. DNA sequencing, bioinformatic processing, and data analysis 90
IV — RESULTS 93
4.1. PrecisionCallerPipeline 95
4.1.1. Pipeline validation with samples from the 1000 Genomes Project 95
4.1.2. Pipeline comparison with an independent set of 50 clinical samples 102
4.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort 104
4.2.1. Cohort description 104
4.2.2. $CD4^+$ T cell enrichment, DNA extraction, and NGS sequencing quality 105
4.2.3. Cross-sectional comparison 108

4.2.4. Longitudinal comparison 119

V — DISCUSSION AND CONCLUSION 125

5.1. PrecisionCallerPipeline 127

5.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort 130

VI — FUTURE PERSPECTIVES 135

VII — REFERENCES 139

VIII — ANNEXES 217

Annex A — PrecisionCallerPipeline I

Annex B — Flow cytometry analysis I

Annex C — mtDNA extraction and NGS sequencing I

Page intentionally left blank

ACKNOWLEDGMENTS

This PhD thesis would not have been possible without the support of my two PhD supervisors. Thus, I would like to thank Friedemann Paul for believing in me from the very start of my journey, ever since I sent him a "cold" e-mail in 2014, when I was still a medical student. His optimism and drive have certainly helped me stay motivated throughout this project and his clinical and incredibly broad expertise in Multiple Sclerosis have been invaluable. As for Vanessa Morais, her scientific passion towards mitochondrial biology is truly inspiring and I was incredibly lucky in being accepted as a PhD student at a particularly complicated stage of my PhD journey. Her rigor, ambition, and scientific reasoning have surely molded me into a better researcher.

In addition, nothing would have happened without: (**I**) our co-authors/collaborators, namely, Hansi Weissensteiner, Susanna Asseyer, Filipa S. Carvalho, Judith Bellmann-Strobl, René Gieß, Sebastian Schönherr, Ana Catarina Fonseca, and José M. Ferro; (**II**) other scientific staff, specifically, Ângela Afonso, Gislaine Castilho de Oliveira, Ana Mafalda Rocha, Rosalie Schmidt, Bettina Vogelreuter, Bibiane Seeger-Schwinge, and Jana Hermann; (**III**) the funders supporting this research, namely, Fundação para a Ciência e Tecnologia, the United States' National Multiple Sclerosis Society, and Merck Germany; and, finally, (**IV**) the subjects who have altruistically donated their samples.

I would also like to thank my Thesis Committee, namely, Ana Espada Sousa, Bruno Silva-Santos, Gabriele De Luca, and Mamede de Carvalho; the External Supervisory Committee of the NeurULisboa PhD Program, namely, Michael Swash, Vincenzo Crunelli, Richard Wise, and the late Fernando Lopes da Silva; the NeurULisboa Program Director Ana Sebastião; and, finally, my fellow NeurULisboa PhD students, who all provided valuable feedback and support throughout my degree.

A special thanks to my one-of-a-kind colleagues at the VMorais Lab, particularly Ana Ramos, Andreia Pereira, Bernardo Antunes, Carlos Ramos, Elvira Leites, Filipa Gonçalves, and Rita Soares, for welcoming me with open arms, a lot of humor, great scientific reasoning, and unique comradery, which I sincerely and deeply cherish. Particularly, I would like to thank Elvira Leites for accompanying me from the very beginning, whom I admire as both a friend and a researcher, and Bernardo Antunes, another colleague turned friend, for the much-needed breaks by the coffee machine and for always lending me his

ear. Also, my time at the Instituto de Medicina Molecular João Lobo Antunes (IMM–JLA) would not have been the same without Eunice Paisana, whom I am lucky to have as a friend, as well as Ana Isabel Afonso, Mariana Ferreira, Marie Bordone, and Nuno Agostinho. Other IMM–JLA members provided valuable scientific input, namely, Ana Catarina Godinho dos Santos (AESousa Lab), Maria Soares (JLacerda Lab), Ana Água-Doce and Filipa Ribeiro (LGraça Lab), João Sabino (SAlmeida Lab), and Diana Macedo (Flow Cytometry Facility).

On a personal matter, I am deeply thankful for all my friends and family, luckily too many to count or enumerate, for their unyielding support, with a special thanks to my mother, Tereza Cortes Figueiredo, and my brother, Miguel Cortes Figueiredo. Additionally, I must acknowledge my 90-year-old grandmother, Maria Beatriz Cortes Figueiredo, who has cultivated my academic interest since I was a small child and who first inspired me to ask (and appreciate!) difficult questions. Furthermore, I would like to dedicate this achievement to my father, whose premature passing prevented him from further pursuing his academic career and from witnessing the powerful impact he has had in my life.

I am also beyond grateful to my husband, Carlos Miguel Oliveira, who has accompanied me throughout the ups and downs of my PhD journey in two countries, three PhD projects, and six years of burnout, symptoms of anxiety and depression, and a mountain of imposter syndrome in both Medicine and Science, with his constant patience, love, and stability. Thank you for being the home I could retreat to whenever the shadows took hold.

Finally, I could not finish this section without acknowledging my therapist, Catarina Janeiro, with whom I started my own self-care journey in 2017, when I was faced with the failure of my first PhD project and without whom I would not be here. A PhD is never easy, and struggling with it or dropping out seems to be, unfortunately, commonplace (Can *et al.*, 2016; Evans *et al.*, 2018; Guthrie *et al.*, 2018; Woolston, 2021). Thus, I would like to acknowledge all those who successfully finished their PhD despite their struggle, while also acknowledging all those who, unfortunately, did not finish, nor survive.

FUNDING

This research was funded by Fundação para a Ciência e Tecnologia (FCT) (*FCT/PTDC/MED-NEU/7976/2020*), and a project cofunded by FEDER (POR Lisboa 2020— Programa Operacional Regional de Lisboa PORTUGAL 2020), and FCT (*PAC-PRECISE LISBOA-01-0145-FEDER-016394*), and the National Multiple Sclerosis Society (NMSS), NMSS Pilot Research Grant (*PP-1712-29466*). My stipend was supported by FCT (*PD/BD/114122/2015*), and by Merck Germany (restricted research grant).

Page intentionally left blank

MOST RELEVANT ABBREVIATIONS

APC — antigen-presenting cell
BBB — blood–brain barrier
BCR — B-cell receptor
bp(s) — base pair(s)
CD — clusters of differentiation
CI — confidence interval
CIS — Clinically Isolated Syndrome
CNS — central nervous system
CSF — cerebrospinal fluid
DAMP — damage-associated molecular pattern
DMD — disease-modifying drug
EAE — experimental autoimmune encephalitis
EBV — Epstein–Barr virus
EDSS — expanded disability status scale
ETC — electron transport chain
FDR — false discovery rate
HC — healthy control
HLA — human leukocyte antigen
IFN — interferon
Ig — immunoglobulin
IL — interleukin
LCR — low complexity region
Max — maximum
MHC — major histocompatibility complex
Min — minimum
MRI — magnetic resonance imaging
MS — Multiple Sclerosis
MSFC — Multiple Sclerosis functional composite
mtDNA — mitochondrial DNA
N — number
nDNA — nuclear DNA
NGS — next-generation sequencing
NK — natural killer
NUMT(s) — nuclear insertion(s) of mitochondrial DNA
OCT — optical coherence tomography
OXPHOS — oxidative phosphorylation
PAMP — pathogen-associated molecular pattern
PBMC — peripheral blood mononuclear cell
PBS+FBS — $PBS^{-/-}$ containing FBS at 2%

PCP — PrecisionCallerPipeline
PNS — peripheral nervous system
PPMS — Primary Progressive Multiple Sclerosis
PRR — pattern recognition receptor
PwMS — patient(s) with MS
qPCR — quantitative polymerase chain reaction
rCRS — revised Cambridge Reference Sequence
RNS — reactive nitrogen species
ROS — reactive oxygen species
RRMS — Relapsing-Remitting Multiple Sclerosis
rRNA — ribosomal RNA
SPMS — Secondary Progressive Multiple Sclerosis
TCR — t-cell receptor
TFAM — mitochondrial transcription factor A
TNF — tumor necrosis factor
Treg — regulatory T
tRNA — transfer RNA
TSS — Ion Torrent Suite™ Software
V1 — patient with Multiple Sclerosis at visit 1; 6 months after disease onset
V2 — patient with Multiple Sclerosis at visit 2; 36 months after disease onset
VCM — variant calling method
VL(s) — variant level(s)
WGS — whole genome sequencing
ΔΨm — mitochondrial membrane potential

FIGURES

Figure 1 — The nervous system: Anatomical subdivision 4
Figure 2 — The central nervous system 5
Figure 3 — The meninges 6
Figure 4 — The cerebrum: Lobe distribution 7
Figure 5 — The spinal cord 8
Figure 6 — The neuron 9
Figure 7 — Types of neurons 10
Figure 8 — Glial cells 12
Figure 9 — The immune-lymphatic system 13
Figure 10 — Hematopoiesis (simplified) 14
Figure 11 — The immunoglobulin 23
Figure 12 — The glymphatic system 32
Figure 13 — Worldwide prevalence of Multiple Sclerosis 35
Figure 14 — Multiple Sclerosis lesions/plaques 44
Figure 15 — Mitochondria and the oxidative phosphorylation chain 56
Figure 16 — Mitochondrial DNA 59
Figure 17 — Mitochondrial DNA replication: Strand-displacement model 60
Figure 18 — Mosaicism in skeletal muscle 62
Figure 19 — Schematic representation of the workflow 96
Figure 20 — Primary analysis: Overall look 97
Figure 21 — Ion Torrent sequencing: Overall look at differences between PCP and TSS 98
Figure 22 — Mixture analysis: Similarity in Grade A/B variants between expected and observed variant levels 100
Figure 23 — Comparison between PCP and TSS in an independent set of samples 103
Figure 24 — Multiple Sclerosis clinical cohort 104
Figure 25 — $CD4^+$ T cell enrichment and mitochondrial DNA next-generation sequencing 106
Figure 26 — Total number of variants: Cross-sectional comparison (subject types and haplogroups) 109
Figure 27 — Total number of variants: Cross-sectional comparison (age and sex) 110
Figure 28 — Total number of variants: Cross-sectional comparison (regions) 111
Figure 29 — Number of deleterious variants: Cross-sectional comparison (subject types and haplogroups) 113
Figure 30 — Cumulative deleterious burden: Cross-sectional comparison (subject types and haplogroups) 114
Figure 31 — Deleterious variants: Cross-sectional comparison (age and sex) 115
Figure 32 — Deleterious variants: Cross-sectional comparison (regions) 116
Figure 33 — tRNA variants: Cross-sectional comparison 118
Figure 34 — Longitudinal changes in PwMS: Intrapersonal changes 120
Figure 35 — Longitudinal changes in PwMS: Proportion of novel variants 121
Figure 36 — Longitudinal changes in PwMS: Variant level of novel variants 122
Figure 37 — Longitudinal changes in PwMS: Variant level change of old variants 123

Page intentionally left blank

TABLES

Table 1 — Immunoglobulin isotypes: Summary 28
Table 2 — Disease modifying drugs for Multiple Sclerosis in the European Union 41
Table 3 — Mitochondrial variants associated with Multiple Sclerosis risk 66
Table 4 — Novel variants (false positives): Summary characteristics in the mutserve variant calling method 102
Table 5 — Magnetic enrichment and DNA extraction: Summary 105
Table 6 — Flow cytometry analysis: Summary 107

Page intentionally left blank

RESUMO

A Esclerose Múltipla (sigla MS em inglês) é uma doença neuroinflamatória e neurodegenerativa crónica com uma etiologia multifatorial indeterminada, que causa desmielinização e perda neuroaxonal no sistema nervoso central (sigla CNS em inglês), conduzindo a uma incapacidade progressiva. Afeta aproximadamente 2,8 milhões de pessoas no Mundo e há um novo diagnóstico de MS a cada cinco minutos. A prevalência de MS tem aumentado nas últimas décadas em todas as regiões do Mundo, afetando principalmente mulheres em idade ativa, com consequências graves para a saúde mental e para a produtividade dos pacientes com MS (sigla PwMS em inglês).

As mitocôndrias estão no cerne da atividade e da progressão da MS: (**I**) no CNS, a disfunção mitocondrial é o principal desencadeador para a perda neuroaxonal e atrofia cerebral; e (**II**) na periferia, o imunometabolismo medeia a autoimunidade do CNS, em que os PwMS apresentam fenótipos mitocondriais inapropriados e/ou insuficiência na fosforilação oxidativa (sigla OXPHOS em inglês), particularmente em linfócitos T $CD4^+$, cujo papel na patogénese da MS é determinante. No entanto, não há atualmente estudos que explorem o genótipo mitocondrial como possível causa da disfunção mitocondrial observada nos linfócitos T $CD4^+$ dos PwMS. Pelo contrário, a maioria dos estudos existentes optou pela análise de amostras de sangue, apesar da especificidade tecidual muito significativa das variantes de ADN mitocondrial (sigla mtDNA em inglês), em que se incluem os linfócitos T.

Globalmente, as abordagens de sequenciação de próxima geração (sigla NGS em inglês) para mtDNA foram prejudicadas pela maior escassez de recursos e por uma menor uniformidade, em comparação com o ADN nuclear (sigla nDNA em inglês). Um exemplo paradigmático é o *kit* de preparação de biblioteca Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, Waltham, Massachusetts, EUA), usado principalmente em ciências forenses, que possibilita uma sequenciação do genoma inteiro (sigla WGS em inglês) de mtDNA com quantidades iniciais muito baixas de ADN. Contudo, o processamento bioinformático inerente ocorre forçosamente no *software* empresarial Ion Torrent Suite™ (sigla TSS em inglês), dando origem a ficheiros BAM alinhados a uma versão pouco ortodoxa da revised Cambridge Reference Sequence (sigla rCRS em inglês). Deste modo, é impedido o processamento bioinformático com outras ferramentas

comumente utilizadas, sendo que há, adicionalmente, um limiar de heteroplasmia elevado de 10%.

De forma a melhor compreender o papel do genótipo mitocondrial na atividade e progressão da MS, pretendi, com este projeto de Doutoramento: (**I**) desenvolver uma nova abordagem para o Precision ID mtDNA Whole Genome Panel, que fosse capaz de produzir ficheiros corretamente alinhados à rCRS, com um limiar de heteroplasmia mais baixo; (**II**) explorar as diferenças entre PwMS recém-diagnosticados e controlos saudáveis (sigla HC em inglês) relativamente ao mtDNA nos linfócitos T $CD4^+$; e (**III**) analisar as alterações basais e longitudinais do mtDNA tendo em conta os vários fenótipos clínicos da MS, permitindo, assim, uma correlação com a atividade e progressão da doença.

Consequentemente, foi desenvolvido um *pipeline* alternativo, personalizável e *open-source*, o PrecisionCallerPipeline (sigla PCP em inglês), com o intuito de processar as amostras sequenciadas no Ion Torrent com o Precision ID mtDNA Whole Genome Panel, dando origem a ficheiros com a rCRS correta. Usando 18 amostras (3 amostras originais e 15 misturas) derivadas do Projeto 1000 Genomes, o PCP obteve métricas de desempenho globalmente melhores relativamente ao TSS, observando-se um desempenho ideal com um limite de heteroplasmia de 2,5%. Como validação adicional, foram analisadas 50 amostras de uma coorte independente de pacientes com acidente vascular cerebral, em que o PCP detetou 98,31% das variantes detetadas pelo TSS, enquanto o TSS detetou apenas 57,92% das variantes detetadas pelo PCP. Além disso, observou-se uma correlação significativa entre os níveis das variantes (sigla VLs em inglês) detetadas por ambas as abordagens.

Curiosamente, nenhum falso positivo foi partilhado entre o PCP e o TSS. Ao analisar os falsos positivos detetados com o PCP, denotou-se, em comparação com os verdadeiros positivos, uma menor *coverage*, um maior rácio da *forward vs. reverse coverage*, um maior número médio de inserções nucleares de mtDNA (sigla NUMTs em inglês), uma menor prevalência de regiões classificadas como *low-complexity* (sigla LCR em inglês) e uma menor distância à extremidade do *amplicon*. Consequentemente, na análise de dados da coorte de MS/Síndrome Clinicamente Isolada (sigla CIS em inglês), foram incluídas todas as variantes presentes em ambas as abordagens (PCP e TSS) com VLs $\geq$ 10% (o limiar do TSS), sendo que variantes com um VL inferior, detetadas apenas com o PCP, foram filtradas de

acordo com os parâmetros característicos de falsos positivos, à exceção da prevalência de LCR, que não pode ser aplicada a variantes individuais.

Deste modo, tendo em conta a coorte de MS/CIS, foi possível efetuar o WGS do mtDNA de 61 PwMS em duas observações clínicas, seis meses (sigla V1 em inglês) e 36 meses (sigla V2 em inglês) após o início da doença, bem como dos respetivos HC, emparelhados de acordo com a idade e o sexo. Esta sequenciação decorreu após um enriquecimento magnético bem-sucedido em linfócitos T $CD4^+$ — rácio médio de 2,59 entre amostras emparelhadas, intervalo de confiança (sigla CI em inglês) de 95% [2,31–2,87]. Apesar deste enriquecimento em linfócitos T $CD4^+$, houve, porém, uma heterogeneidade significativa entre os diferentes tipos de sujeitos, em que os HC demonstraram um número de células superior a PwMS aquando da V1 — diferença média de 0,65 milhões de células, 95% CI [0,16–1,14], bem como uma maior quantidade de ADN extraído — 4,45 ng/µL para as medições com o Synergy HTX, 95% CI [1,06–7,84], e 3,35 ng/µL para as medições com o Qubit®, 95% CI [0,80–5,90].

Globalmente, observou-se um número total de variantes menor nos haplogrupos H e HV, compatível com o haplogrupo H da rCRS. Tendo em conta as variantes previstas *in silico* como provavelmente deletérias para proteínas, os haplogrupos J e T apresentaram um maior número de variantes deletérias e um maior *burden* deletério cumulativo. Curiosamente, os mesmos haplogrupos tinham já sido previamente associados a um risco acrescido de MS. Relativamente às mutações nas regiões de ARN de transferência (sigla tRNA em inglês), não foram detetadas variantes previstas *in silico* como provavelmente ou potencialmente patogénicas, tendo-se observado, porém, um *score* cumulativo de MitoTIP superior nos haplogrupos K e U e inferior nos haplogrupos H e HV.

Contudo, não foram apuradas diferenças estatisticamente significativas no genótipo mitocondrial entre PwMS e HC, bem como entre V1 e V2 relativamente aos PwMS, incluindo na distribuição de haplogrupos. Ao contrário do esperado no estudo piloto previamente efetuado em 20 grupos de três (20 PwMS em V1 e V2, bem como 20 HC emparelhados de acordo com a idade e o sexo), em que se determinou uma discordância na prevalência de variantes deletérias no Complexo I de 35% — 55% para PwMS aquando da V1 (prevalência discordante: 45%) *vs.* 20% para HC (prevalência discordante: 10%), a análise integrada com os restantes 41 grupos de três demonstrou uma discordância na prevalência de 18%. Esta alteração resultou de um aumento de 8% na prevalência de HC,

com um declínio de 9% para os PwMS. Além disso, a análise longitudinal do genoma mitocondrial dos PwMS entre V1 e V2 demonstrou uma estabilidade global e uma ausência de associação com vários subtipos de MS.

Em conclusão, o genótipo mitocondrial não parece determinar um maior risco no desenvolvimento de MS, nem maior gravidade em PwMS. Além disso, é improvável que a disfunção mitocondrial presente nos linfócitos T $CD4^+$ tenha origem no genótipo mitocondrial.

Palavras-chave: ADN mitocondrial; Esclerose Múltipla; linfócitos T $CD4^+$; Precision ID; sequenciação de próxima geração.

SUMMARY

Multiple Sclerosis (MS) is a chronic neuroinflammatory and neurodegenerative disease with a multifactorial undetermined etiology that causes demyelination and neuroaxonal loss in the central nervous system (CNS), leading to disability, particularly in women. Mitochondria are pivotal for MS disease activity and progression: (**I**) in the CNS, mitochondrial dysfunction is the main trigger for neuroaxonal loss and brain atrophy; and (**II**) in the periphery, immunometabolism is critical for CNS autoimmunity, while patients with MS (PwMS) show inappropriate mitochondrial phenotypes and/or oxidative phosphorylation (OXPHOS) insufficiency, particularly in $CD4^+$ T cells. Nonetheless, no studies have explored if mitochondrial genotype could be causing the mitochondrial dysfunction observed in $CD4^+$ T cells from PwMS. In contrast, most studies have opted for whole blood analysis, despite the significant tissue-specificity of mitochondrial DNA (mtDNA) variants, including T cells.

Overall, next-generation sequencing (NGS) approaches to mtDNA have been hindered by fewer resources and less uniformity, in comparison with nuclear DNA (nDNA). A notable example is the Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, Waltham, MA, USA), mostly used in forensic research, which achieves mtDNA whole genome sequencing (WGS) with very low DNA amounts. Despite this, its bioinformatic processing occurs in the enterprise Ion Torrent Suite™ Software (TSS), yielding BAM files aligned to an unorthodox version of the revised Cambridge Reference Sequence (rCRS), thus impeding further bioinformatic processing with third-party tools, with a relatively high heteroplasmy threshold level of 10%.

In order to better understand the role of mitochondrial genotype in MS activity and progression, I aimed to: (**I**) devise a novel approach for the Precision ID mtDNA Whole Genome Panel, which yielded an output in the correct rCRS and a lower heteroplasmy threshold; (**II**) explore differences in mtDNA in $CD4^+$ T cells between newly diagnosed PwMS and healthy controls (HC); and (**III**) analyze baseline and longitudinal mtDNA changes regarding MS clinical phenotypes, enabling a correlation with disease activity and progression.

Accordingly, an alternative customizable and open-source pipeline, the PrecisionCallerPipeline (PCP), was developed for processing samples with the correct rCRS

output after Ion Torrent sequencing with the Precision ID library kit. Using 18 samples (3 original samples and 15 mixtures) derived from the 1000 Genomes Project, PCP achieved overall improved performance metrics in comparison with the proprietary TSS, with optimal performance at a 2.5% heteroplasmy threshold. These findings were further validated with 50 samples from an independent cohort of stroke patients, with PCP finding 98.31% of TSS's variants (TSS found 57.92% of PCP's variants), with a significant correlation between the variant levels (VLs) of variants found with both pipelines.

Regarding the MS/Clinically Isolated Syndrome (CIS) cohort, it was possible to fully sequence the mtDNA of 61 PwMS at two clinical visits, six months (V1) and 36 months (V2) after disease onset, as well as their age- and sex-matched HC, following the successful magnetic enrichment in $CD4^+$ T cells — mean fold change in paired samples of 2.59, 95% confidence interval (CI) [2.31–2.87].

Ultimately, however, there were no statistically significant differences in mitochondrial genotype between PwMS and HC, as well as between clinical visits V1 and V2 within PwMS. Thus, mitochondrial genotype does not seem to determine a higher risk of developing MS, nor a higher disease disability in PwMS. Furthermore, mitochondrial dysfunction in $CD4^+$ T cells is unlikely to derive from mitochondrial genotype.

Keywords: $CD4^+$ T cells; mitochondrial DNA; Multiple Sclerosis; next-generation sequencing; Precision ID.

I — INTRODUCTION

Page intentionally left blank

1.1. The nervous system

The nervous system is responsible for three main functions (Berkowitz, 2016b):

- **Perception** — through afferent pathways, which provide information to the brain and allow it to process the outer world;
- **Cognition** — integrates all perceptual input and plans for the interaction with the external environment;
- **Action** — allows the organism to interact with the external world through efferent pathways, including other organisms.

These three basic functions serve as the basis to the unique set of skills, such as language, creativity, and logic, that allow for "the human condition", separating us from the other primates on Earth (Arendt, 1998; Pääbo, 2014; Waxman, 2020a).

Embryologically, the nervous system arises from the ectoderm at three weeks of development, which further specializes into the neuroectoderm, followed by the neural plate, and, finally, the neural tube and the neural crest (Pozzo-Miller and Theibert, 2020; Mescher, 2021b).

Anatomically, the nervous system can be divided into two main subsystems (**Figure 1**) (Theibert, 2020f; Waxman, 2020a; Mescher, 2021b):

- **Central nervous system (CNS)** — encompasses the brain and the spinal cord, which are surrounded by the meninges and the cerebrospinal fluid (CSF), and further enveloped by bone. It arises from the neural tube;
- **Peripheral nervous system (PNS)** — includes the cranial and spinal nerves, as well as ganglia. It arises from the neural crest.

Functionally, the nervous system can be divided into (Theibert, 2020f; Mescher, 2021b):

- **Somatic** — further subdivided into a sensory division (afferent), which corresponds to information perceived consciously (e.g., vision and hearing),

and a motor division (efferent), which corresponds to motor actions with a conscious or voluntary control (e.g., speech and movement);

- **Autonomic** — further subdivided into a visceral division (afferent), which corresponds to information not perceived consciously (e.g., abdominal organs and the heart), and a motor division (efferent), which corresponds to motor actions without a conscious or voluntary control (e.g., smooth muscle contraction and sweat production);
- **Enteric** — controls the gastrointestinal system independently from the autonomic system, despite the heavy influence of the latter.

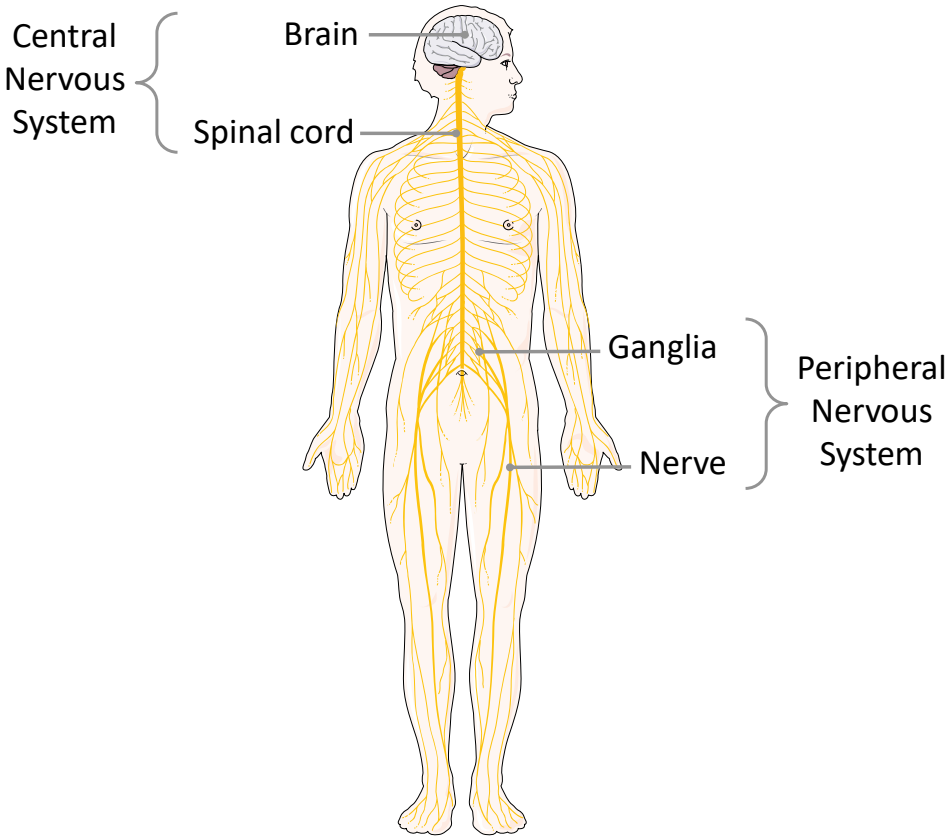


Figure 1 — The nervous system: Anatomical subdivision

Templates adapted from (*SMART*, 2021) (CC BY 3.0).

1.1.1. The central nervous system

The CNS includes both the brain, also called the encephalon, and the spinal cord (**Figure 2**) (Theibert, 2020f; Waxman, 2020a; Mescher, 2021b).

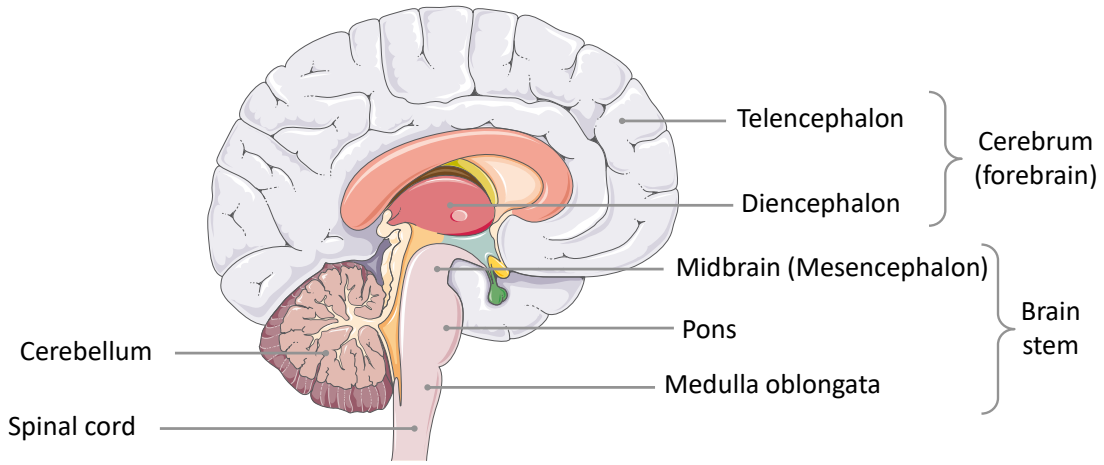


Figure 2 — The central nervous system

Templates adapted from (*SMART*, 2021) (CC BY 3.0).

Both components of the CNS have: (**I**) gray matter, which contains the cellular constituents of the nervous system, namely, neurons, particularly their cell bodies and dendrites, and glial cells; and (**II**) white matter, which predominantly contains the myelinated axons from neurons (Berkowitz, 2016b; Theibert, 2020a, 2020b, 2020f; Waxman, 2020b; Mescher, 2021b). They are also enveloped by the meninges (**Figure 3**), which are, from most superficial to deepest: (**I**) the dura mater, composed of two layers (periosteal and meningeal), which provides structure to the venous sinuses and has several septae that restrict movement of the brain; (**II**) the arachnoid mater, the interface between the dura mater and the pia mater through connective tissue filaments resembling spider webs, which contains the CSF, and, through the arachnoid villi/granulations in the venous sinuses, is also responsible for CSF reabsorption; and (**III**) the pia mater, a delicate layer around the CNS and its blood vessels, which produces CSF, together with the ependymal cells, in the choroid plexuses (Berkowitz, 2016b; Morton, Foreman and Albertine, 2019b; Theibert, 2020f; Mescher, 2021b). Finally, both components of the CNS are further enveloped by bone, namely, the skull for the brain and the vertebral column for the spinal cord (Morton, Foreman and Albertine, 2019a, 2019b; Theibert, 2020f).

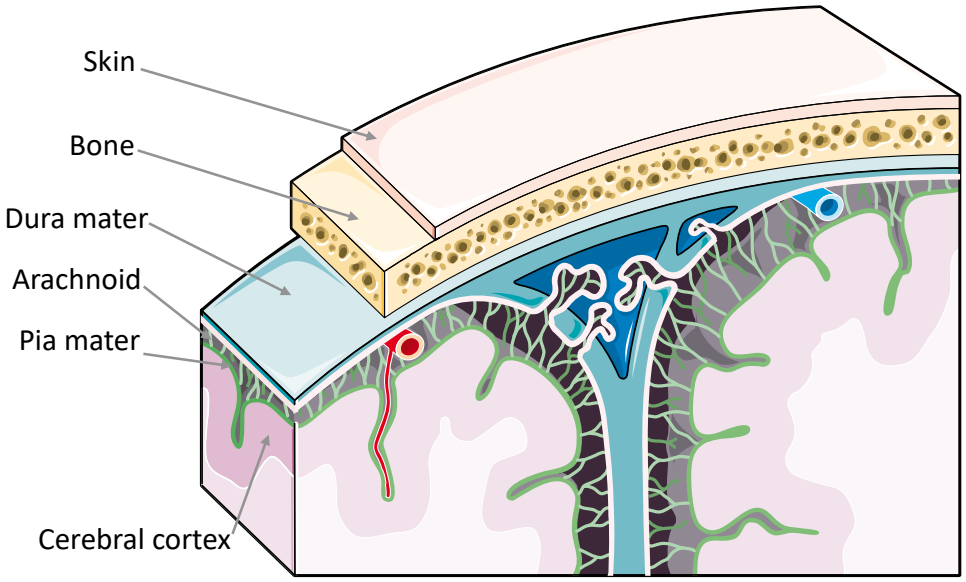


Figure 3 — The meninges

Templates adapted from (*SMART*, 2021) (CC BY 3.0).

The brain is further subdivided into (**Figure 2**) (Morton, Foreman and Albertine, 2019a; Theibert, 2020a, 2020b, 2020f; Waxman, 2020a):

- **Brain stem** — encompasses the midbrain (or mesencephalon), the pons, and the medulla oblongata. It is responsible for vital unconscious functions, such as breathing, swallowing, and coughing;
- **Cerebellum** — is mainly responsible for the coordination of motor function;
- **Cerebrum (forebrain)** — encompasses the telencephalon, which includes the cerebral cortex, cerebral white matter, and subcortical structures, such as the basal ganglia, and the diencephalon, which includes the thalamus, the hypothalamus, the epithalamus, the subthalamus, and the posterior pituitary. It is responsible for most complex processes, such as interpreting stimuli, language, voluntary motor function, and memory. The cerebrum can also be subdivided into two hemispheres (left and right), separated by the longitudinal fissure, where each hemisphere contains four lobes (frontal, parietal, temporal, and occipital), separated by different sulci (**Figure 4**).

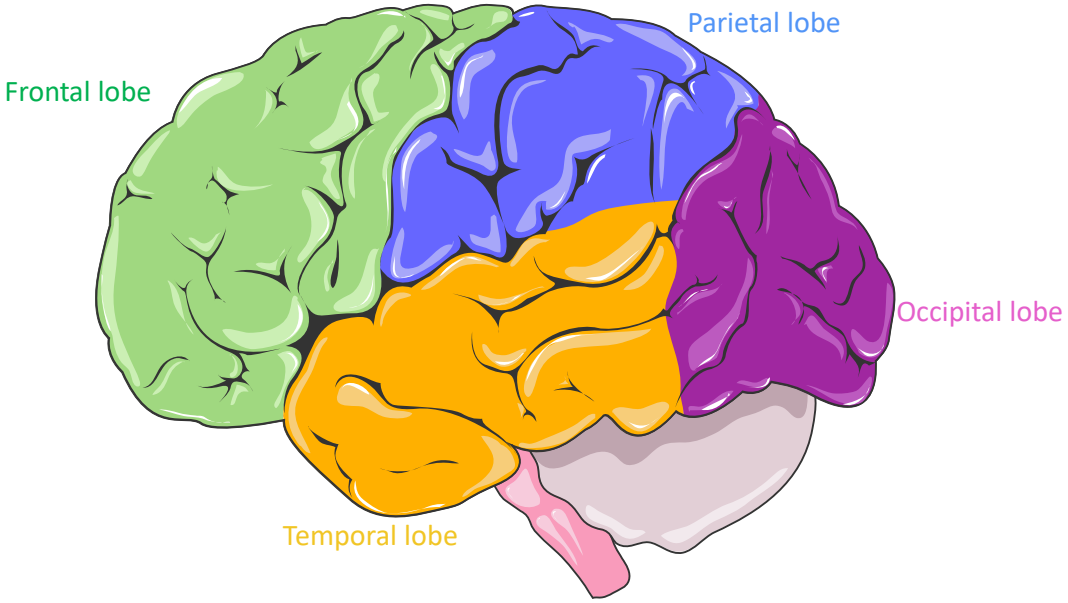


Figure 4 — The cerebrum: Lobe distribution

Templates adapted from (*SMART*, 2021) (CC BY 3.0).

The spinal cord continues the brain stem from the medulla oblongata, until it originates the conus medullaris around the first or second lumbar vertebra (Theibert, 2020b; Waxman, 2020b). Along its track, it gives rise to 31 bilateral spinal nerves, subdivided into five segments: cervical, thoracic, lumbar, sacral, and coccygeal, according to the corresponding region of the vertebral column (Theibert, 2020b). Two main regions are easily distinguishable (**Figure 5**): the peripheral white matter and the central gray matter, around the central canal. The gray matter is subdivided, anatomically and functionally, into two portions: the ventral (anterior) horns, which are responsible for motor function, and the dorsal (posterior) horns, which relay sensory information (Theibert, 2020b, 2020f; Waxman, 2020b). The spinal cord serves as the main communication avenue between the brain and the body (Theibert, 2020b; Waxman, 2020b).

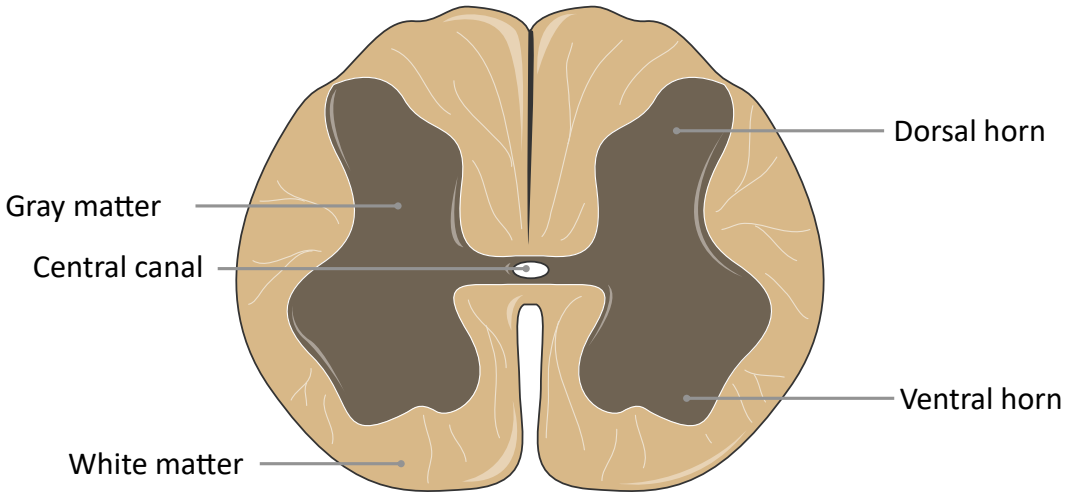


Figure 5 — The spinal cord

Templates adapted from (*SMART*, 2021) (CC BY 3.0).

1.1.2. The peripheral nervous system

The PNS includes cranial and spinal nerves, which are bundles of myelinated or unmyelinated axons of neurons, connected to the CNS, as well as associated ganglia, which are clusters of nerve cells outside the CNS (Berkowitz, 2016b; Theibert, 2020b; Waxman, 2020a; Mescher, 2021b). The PNS establishes the communication between the CNS and the remaining body, relaying both afferent and efferent information to the somatic, autonomic, and enteric subsystems (Theibert, 2020f; Waxman, 2020a).

1.1.3. The cellular constituents of the nervous system

The human brain comprises ~170 billion cells: half are neurons, which are mostly responsible for the complexity of the nervous system through a dense and interconnected network of cells, and half are glial cells, which mainly serve as a support for neuronal function (Azevedo *et al.*, 2009; Theibert, 2020f; Mescher, 2021b).

Morphologically, neurons have three main components (**Figure 6**): (**I**) the soma or cell body, which houses the nucleus and where the majority of molecules of the cytoplasm are synthetized; (**II**) the dendrites, which are multiple short elongated processes covered in dendritic spines aimed at receiving input from other neurons through synapses; and (**III**) the axon, which is a single extension originating from the axon hillock aimed at communicating with other neurons or other cell types, such as muscle and glandular cells, also through synapses (Theibert, 2020f; Mescher, 2021b).

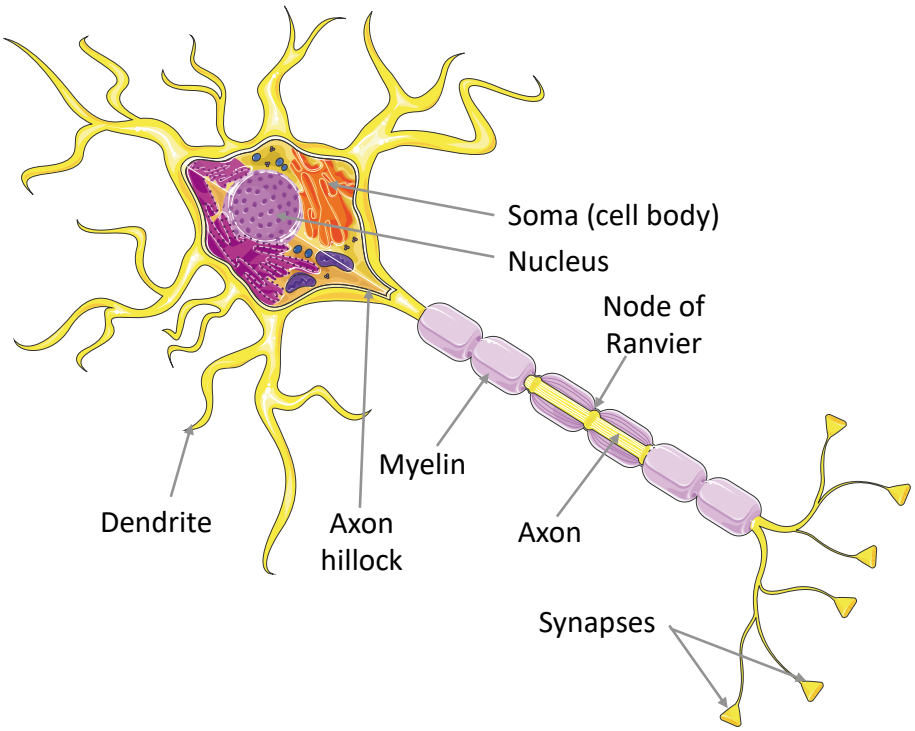


Figure 6 — The neuron

Templates adapted from (*SMART*, 2021) (CC BY 3.0).

Synapses are the main interface between a neuron and a target cell, where an action potential, which is a propagating depolarization of the plasma membrane through voltage-gated Na^+ and K^+ channels after a stimulus above a certain threshold reaches the presynaptic neuron, is transmitted to a postsynaptic cell either electrically, where there is a direct transmission of ions through gap junctions, or, most commonly, chemically, where neurotransmitters, such as acetylcholine, glutamate, γ-aminobutyric acid, dopamine, serotonin, adenosine, and endocannabinoids, among others, act as mediators (Amthor, 2020; Theibert, 2020c, 2020d, 2020e, 2020f; Mescher, 2021b). Most synapses occur between an axon and the soma or dendrites of the target cell — axosomatic and axodendritic synapses, respectively —, although a few synapses occur between two axon terminals — axoaxonic synapses (Theibert, 2020f; Mescher, 2021b). Interestingly, synapse formation (synaptogenesis) and synapse pruning or refinement extend into adulthood (Pozzo-Miller and Theibert, 2020).

Depending on the morphology of the components of each neuron, various types of neurons are usually discernable (**Figure 7**) (Theibert, 2020f; Mescher, 2021b):

- **Unipolar** — neurons have a single process; these cells are rare in mammals and are typical of invertebrates;
- **Pseudo-unipolar** — neurons have a single short process that bifurcates near the soma: one branch extends to the periphery and the other extends to the CNS; these cells are commonly somatosensory neurons;
- **Bipolar** — neurons have one axon and one dendrite; these cells are present in the retina, inner ear, and the olfactory epithelium;
- **Multipolar** — neurons have at least two dendrites and one axon; these are the most common neurons.

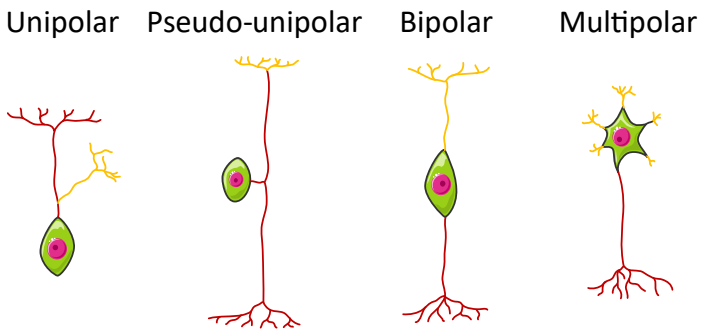


Figure 7 — Types of neurons

Lines in red depict axons, while lines in orange depict dendrites. Templates adapted from (*SMART*, 2021) (CC BY 3.0).

Glial cells, or glia, support neuronal function and are pivotal to the homeostasis of the nervous system (**Figure 8**) (Lago-Baldaia, Fernandes and Ackerman, 2020; Theibert, 2020f; Mescher, 2021b). They can be divided into macroglia, which originate from the ectoderm and include several cell types, and microglia, which derive from the embryonic yolk sac and are a single cell type ubiquitous in the CNS (Ginhoux *et al.*, 2013; Theibert, 2020f).

Major macroglia subdivisions are (Aggarwal, Yurlova and Simons, 2011; Amthor, 2020; Pozzo-Miller and Theibert, 2020; Theibert, 2020f; Mescher, 2021b):

- **Astrocytes** — are the most abundant glial cell type in the CNS and integral in forming the blood–brain barrier (BBB). They provide metabolic support to other cells, particularly neurons, by mediating the absorption and release of metabolites, such as glucose, lactate, and amino acids, as well as neurotransmitters, which makes them crucial to synaptic homeostasis. In the

presence of neuronal death or CNS injury, they are responsible for forming glial scars. Their counterparts in the PNS, namely in ganglia, are satellite cells;

- **Oligodendrocytes** — provide plasma membrane projections that envelop several axons (up to 50) in the CNS in multiple sheaths with very little cytoplasm, collectively named myelin. Myelin serves as an electrical insulator and is mostly composed by lipids (~80%), with a high proportion of long-chain and saturated fatty acids in comparison with other plasma membranes, where proteins, such as proteolipid protein (PLP) and myelin basic protein (MBP), provide structural support. The myelin sheath envelops axons intermittently, leaving bare regions where voltage-gated sodium channels are enriched — nodes of Ranvier — that allow for saltatory action potential conduction, which is 10 times faster than an unmyelinated axon. Their counterparts in the PNS are Schwann cells, which, however, only myelinate a single section of a single axon. Interestingly, as in synaptogenesis and synapse pruning, myelination extends beyond birth until adolescence;
- **Ependymal cells** — line the brain ventricles and the central canal of the spinal cord in the CNS, forming the choroid plexuses, together with the pia mater, which are responsible for CSF production. CSF is synthetized by filtering plasma from blood through the blood–CSF barrier, which is formed by ependymal cells joined by tight junctions in the choroid plexuses. Elsewhere, the blood–CSF barrier is maintained by cells in the arachnoid mater, also joined by tight junctions. In their apical surfaces, ependymal cells have cilia, which facilitate the circulation of CSF;
- **Pericytes** — envelop the nonfenestrated capillary endothelial cells, maintaining their tight junctions. Together with astrocytic end feet, they constitute and maintain the integrity of the BBB. Also, they are able to control blood flow through their contractile properties. In the PNS, they are also responsible for isolating axons from capillaries and establishing the blood–nerve barrier.

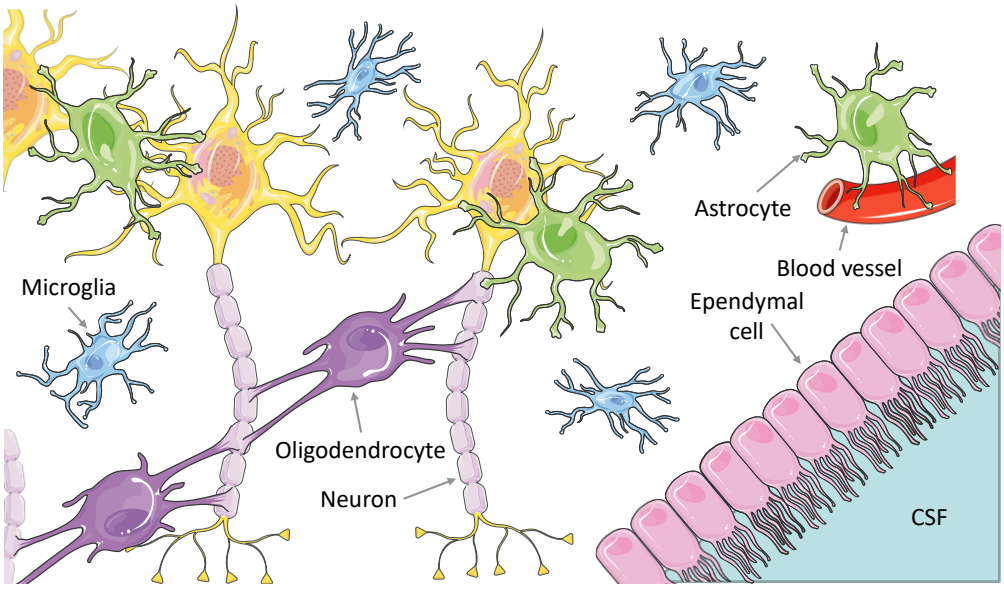


Figure 8 — Glial cells

Templates adapted from (*SMART*, 2021) (CC BY 3.0). Image inspired by (D'Agata *et al.*, 2017). Abbreviations: CSF — cerebrospinal fluid.

Microglia are mainly responsible for the immune surveillance and defense of the CNS, through their phagocytic and antigen-presenting properties, in addition to cytokine production (more details in *1.3. Neuroimmunology*) (Theibert, 2020f; Mescher, 2021b). They are also crucial for brain homeostasis, by removing damaged cells and extracellular debris, and by pruning synapses (Paolicelli Rosa C. *et al.*, 2011; Weinhard *et al.*, 2018; Theibert, 2020f).

1.2. The immune system

The main function of the immune system, through immunity, is to protect the body against foreign and infectious microorganisms that might pose a threat — pathogens —, such as bacteria, viruses, fungi, protozoa, and helminths (Haynes, Soderberg and Fauci, 2018; Mescher, 2018; Levinson *et al.*, 2020g).

Immune cells are ubiquitous in all tissues of the human body and in lymphoid organs, which are connected by the circulation of blood and lymph (Haynes, Soderberg and Fauci, 2018; Mescher, 2018). Histologically, two main tissues are discernable (**Figure 9**) (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Mescher, 2018): (**I**) the primary lymphoid organs, which produce the immune cells, such as the red bone marrow and the thymus; and (**II**) the

secondary lymphoid organs, where immune responses are propagated and refined, such as the lymph nodes, the spleen, and the diffuse lymphoid tissue found in the mucosa of the digestive, respiratory, and urogenital tracts, called mucosa-associated lymphoid tissue (MALT).

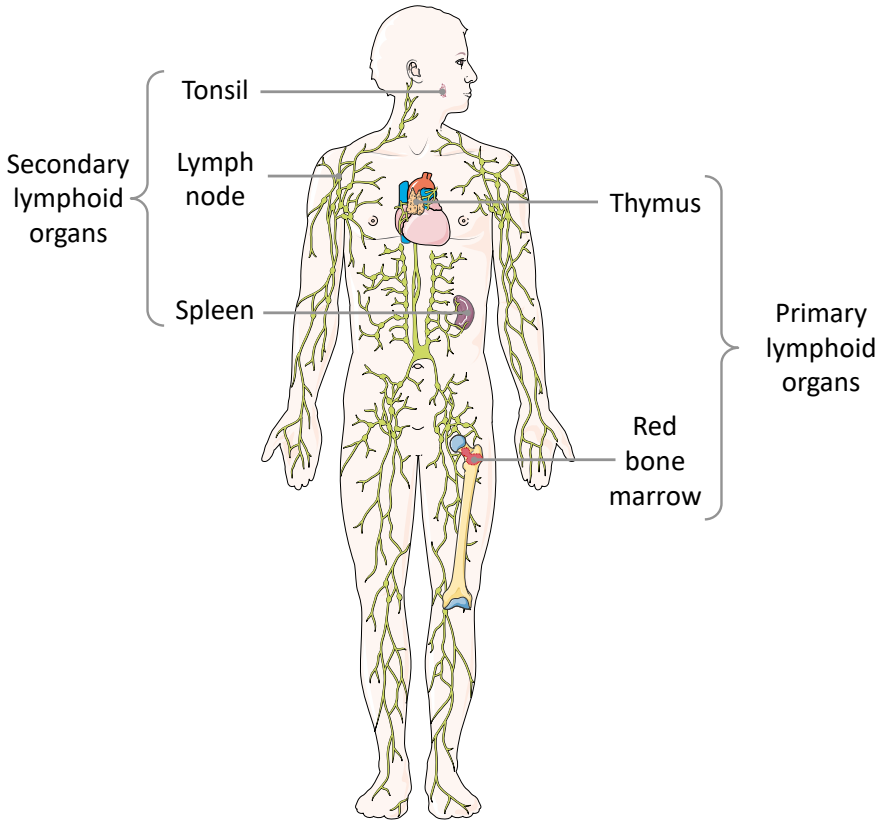


Figure 9 — The immune-lymphatic system

The tonsils are a part of the mucosa-associated lymphoid tissue. Templates adapted from (*SMART*, 2021) (CC BY 3.0). Image inspired by (Mescher, 2018).

Two different lines of defense constitute the immune system (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Mescher, 2018; Levinson *et al.*, 2020g):

- **Innate immune system** — present in both invertebrates and vertebrates, accompanying us from birth; it is nonspecific and serves as the first line of defense against pathogens;
- **Adaptive immune system** — present in vertebrates only, developing more slowly through continuous exposure to pathogens; it exhibits memory and it is specific to certain antigens, which are molecules made of smaller epitopes that are able to trigger the immune system, either from foreign or self agents.

The cellular constituents of the immune system are leukocytes, also called white blood cells, which derive from hematopoietic stem cells in the red bone marrow (Parker *et al.*, 2016; Mescher, 2021a). Embryologically, however, hematopoietic stem cells derive from the yolk sac mesoderm, and, subsequently, the liver and the spleen, before the development of bone marrow in medullary cavities (Mescher, 2021a). The two main hematopoietic lineages are (**Figure 10**) (Parker *et al.*, 2016; Mescher, 2021a): (**I**) myeloid, which gives rise to erythrocytes, platelets, granulocytes, and monocytes; and (**II**) lymphoid, which gives rise to B and T lymphocytes, as well as Natural Killer (NK) cells.

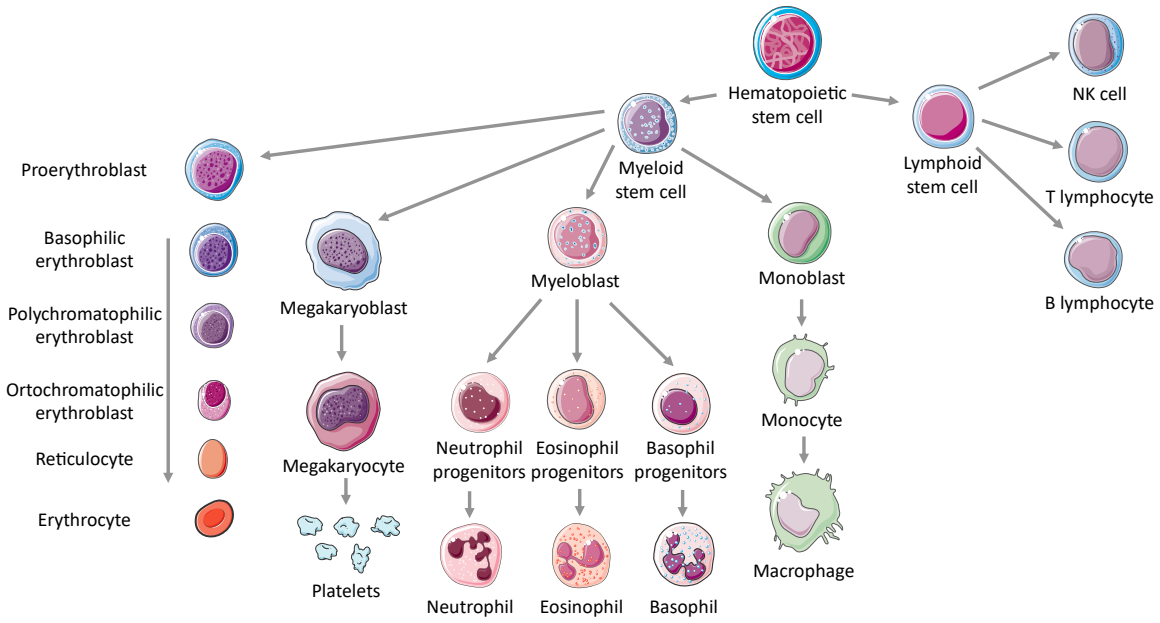


Figure 10 — Hematopoiesis (simplified)

Templates adapted from (*SMART*, 2021) (CC BY 3.0). Image inspired by (Parker *et al.*, 2016; Mescher, 2021a).

Abbreviations: NK — natural killer.

Another important feature of the immune system is the existence of various molecules that are crucial for immunity (Borden *et al.*, 2007; Dinarello, 2007; Peakman, 2012; Clark *et al.*, 2016; Haynes, Soderberg and Fauci, 2018; Mescher, 2018):

- **Clusters of differentiation (CD)** — cell-surface molecules that indicate the lineage and the function of groups of leukocytes. The discovery of human CDs was made possible with the advent of monoclonal antibodies, which target a very specific molecular target;

- **Cytokines** — soluble peptides and glycoproteins with low molecular mass released by a cell that serve as chemical messengers to itself — autocrine — or other cells — paracrine for adjacent cells and endocrine for remote cells, tissues, or organs. They may also exert pleiotropy, which are different effects on different cells, and synergy with other cytokines, where the two cytokines combined exert a more pronounced effect than the expected sum of both. They encompass several different subtypes, such as interleukins (ILs), interferons (IFNs), tumor necrosis factors (TNFs), colony-stimulating factors, and transforming growth factors (TGFs), among others;
- **Chemokines** — cytokines capable of attracting cells through a gradient of concentration, orchestrating immune cell movement. They connect to surface G-coupled protein receptors with seven transmembrane spanning domains. Their nomenclature derives from their cysteine residues, namely, C for a single cysteine, CC for two cysteines close together, CXC for two cysteines separated by another amino acid residue, and CX3C for two cysteines separated by three other amino acid residues.

1.2.1. The innate immune system

The innate immune system accompanies us throughout life with similar intensity and independent from prior contact with pathogens; it is based on recognizing a limited number of generic molecular patterns shared by pathogenic microorganisms. It has a layered approach to immunity, where multiple barriers serve as the first line of defense (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Mescher, 2018; Levinson *et al.*, 2020e, 2020g):

- **Physical barriers** — aimed at limiting the adherence and invasion by microbes, such as the keratinized squamous cells in the epidermis of the skin, sweat production, mucins coating the mucosal surfaces of the gastrointestinal, respiratory, and urogenital tracts, the peristalsis of the gastrointestinal tract, the cilia and the cough reflex in the respiratory tract, and the urine flow in the urogenital tract;

- **Chemical barriers** — aimed at killing or incapacitating microbes, such as hydrochloric acid in the stomach, lysozyme in saliva and tears, defensins in the skin and mucosal surfaces, and surfactant in the respiratory tract;
- **Biological barriers** — aimed at competing with other microorganisms for resources, such as the commensal flora, also called microbiota, present in the skin, the gastrointestinal tract, the nasopharynx, and genitalia.

After breaching the first line of defense, the second line of defense is activated, through the recognition of highly-conserved surface carbohydrates or lipids in microbes that are absent in human cells, called pathogen-associated molecular patterns (PAMPs), by pattern recognition receptors (PRRs) (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Mescher, 2018; Levinson *et al.*, 2020e, 2020g). These receptors can be either extracellular, aimed at pathogens outside the immune cells or within vesicles, i.e., Toll-like receptors (TLRs) and C-type lectin receptors (CLRs), or intracellular, aimed at microbes that invade the plasma membrane, i.e., NOD-like receptors (NLRs) and retinoic acid–inducible gene (RIG)-1-like receptors (RLRs). The recognition of PAMPs through PRRs is critical to set the stage for the subsequent adaptive immune system, since the activation of particular PRRs (extracellular or intracellular) shapes the appropriate cell response for each type of pathogens (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). An important example is the inflammasome, which is a multiprotein cytoplasmatic complex that is assembled after the activation of NLRs by PAMPs or damage-associated molecular patterns (DAMPs), from damaged or mutated host cells, where NLRs, together with other proteins, namely, ASC, caspase protease, and a protein from either the NLR pyrin (NLRP) family or the HIN-200 family, lead to the downstream proteolytic activation, through caspase-1, of the pro-inflammatory IL-1β and IL-18 cytokines (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e; Li and Wu, 2021).

Other soluble molecules are termed pattern-recognition molecules (PRMs), where pentraxins, which include the well-known C-reactive protein, and collectins, which are soluble CLRs and include the mannose-binding lectin (MBL) protein, work together to opsonize bacteria for phagocytosis and activate the complement system (Peakman, 2012; Foo *et al.*, 2015; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e; Li and Wu, 2021). The complement system is a set of circulating proteins mainly synthesized in the

liver, which, once activated, are able to perform: (**I**) cytolysis, through the membrane attack complex; (**II**) chemotaxis, particularly of neutrophils; (**III**) opsonization, promoting phagocytosis of microbes by phagocytes; (**IV**) degranulation of mast cells, leading to increased vascular permeability and bronchospasm; and (**V**) increased antibody production by plasma cells. These proteins may be activated by a first contact with a pathogen, through the MBL protein or exogenous cell surface substances, or by antibody–antigen immune complexes (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020d, 2020e).

Following the recognition of PAMPs through PRRs and PRMs, the innate immune cells are activated, starting with phagocytes, which are responsible for engulfing microbes. The most abundant phagocytes are tissue macrophages (**Figure 10**); in addition to being the first cells that encounter foreign material or injured tissue, they are also ubiquitous in the human body and present before birth (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). They are myeloid cells whose nomenclature and phenotype may change according to the tissue they reside in: Kupffer cells in the liver, alveolar macrophages in the lung, osteoclasts in bone, and microglia in the CNS (more details in *1.1.3. The cellular constituents of the nervous system*). In addition to resident tissue macrophages, circulating monocytes (**Figure 10**), which are myeloid mononuclear $CD14^{+}$ cells, may also differentiate into macrophages upon chemotaxis; together they engulf and kill microbes, release pro-inflammatory cytokines such as IL-1, IL-6, and TNF-α, and are able to fragment pathogens into antigens for antigen presentation to T cells, linking the innate to the adaptive immune system (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e).

Nonetheless, a more robust link between the innate and the adaptive immune systems is provided by dendritic cells, which are mainly responsible for activating naïve T cells and initiating the adaptive immune response (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). Dendritic cells derive from the bone marrow and are mostly present in barrier tissues and in circulation; they encompass four main subtypes: plasmacytoid dendritic cells, conventional or myeloid dendritic cells (which subdivide into two other phenotypes), Langerhans cells, and, in pro-inflammatory conditions, derived from circulating monocytes (Peakman, 2012; Kashem, Haniffa and Kaplan, 2017; Collin and Bigley, 2018; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). Dendritic cells in their immature state undergo constant immunosurveillance until they contact PAMPs or

DAMPs (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). Afterwards they undergo phagocytosis and fully mature into their role as professional antigen-presenting cells (APCs), migrating to secondary lymphoid organs through the lymphatic circulation by expressing the chemokine CCR7, where they interact with naïve T cells (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020c, 2020e).

Antigens are presented in surface major histocompatibility complex (MHC) proteins (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). The MHC proteins are encoded by the human leukocyte antigen (HLA) genes, on the short arm of chromosome 6 (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e, 2020f). For classical class I MHC proteins, expressed in all nucleated cells, each gene *HLA-A*, *HLA-B*, and *HLA-C* encodes a highly polymorphic heavy chain, termed α chain, which associates with a sole β chain, β_2 microglobulin, encoded on chromosome 15, resulting in a molecule with a hypervariable region in its N-terminal region, which is able to present antigens, and a constant region, which binds to the CD8 molecule in $CD8^+$ T cells (Peakman, 2012; Levinson *et al.*, 2020f). Non-classical HLA class I genes, such as *HLA-E*, *HLA-F*, and *HLA-G*, among others, have a more restricted expression and are mainly involved in interactions with NK cells (Peakman, 2012). Regarding class II MHC proteins, which are only present in professional APCs, such as dendritic cells, macrophages and B cells, they are encoded by several *HLA-D* loci with three major subregions, i.e., *DP*, *DQ* and *DR*; each subregion has *A* and *B* genes that form dimeric and highly polymorphic molecules, with a hypervariable region for antigen presentation and a constant region for CD4 binding with $CD4^+$ T cells (Cassell and Schwartz, 1994; Peakman, 2012; Hong *et al.*, 2018; Levinson *et al.*, 2020f; Muntjewerff, Meesters and van den Bogaart, 2020).

Prior to antigen presentation, however, antigens must also be processed (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e). Following the phagocytosis of the pathogens or degraded tissue by APCs, enhanced by prior opsonization and the binding of PRRs/PRMs to PAMPs/DAMPs, the phagosome with the engulfed material fuses with the lysosome of the phagocytic cell, where the pathogen is killed by exposure to proteases, lysozyme, and the respiratory burst, namely, reactive oxygen species (ROS), generated by NADPH oxidase, and reactive nitrogen species (RNS), generated by nitric oxide synthase (Slauch, 2011; Levinson *et al.*, 2020e). Afterwards, the foreign material is digested into smaller peptides that are loaded onto MHC class II

molecules for presentation to $CD4^+$ T cells (exogenous route) or onto MHC class I molecules aimed at $CD8^+$ T cells (cross-presentation) (Peakman, 2012; Levinson *et al.*, 2020e). For all other nucleated cells, MHC class I molecules express a selection of cytosolic peptides of the respective cell; in a healthy state, this yields no response from $CD8^+$ T cells (Peakman, 2012). Regarding other non-peptide antigens, carbohydrates are mostly presented with the MHC class II system (Cobb *et al.*, 2004; Sun *et al.*, 2016), lipid antigens are presented through CD1 molecules (Mori, Lepore and De Libero, 2016; Lepore, Mori and De Libero, 2018; Pereira *et al.*, 2019), and small metabolites are presented through MR1 molecules (Mori, Lepore and De Libero, 2016; Lepore, Mori and De Libero, 2018).

In addition to APCs, other cellular components are vital for the innate immune system (Peakman, 2012; Parker *et al.*, 2016; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020e):

- **Granulocytes** — myeloid effector cells with multilobed nuclei and cytoplasmic granules, they are responsible for amplifying the immune response; three different cell types are discernable:
    - **Neutrophils** — with a nucleus with three to five lobes (**Figure 10**), they are the most abundant immune cell in blood. They are highly phagocytic, especially after opsonization with immunoglobulin (Ig) G antibodies and complement activation, and they are responsible for killing and digesting pathogens, particularly bacteria, through degranulation, i.e., the release of azurophilic granules, mostly intracellularly, which contain myeloperoxidase, lysozyme, defensins, elastase, and cathepsins, among other enzymes, and of specific granules, mostly extracellularly, which contain respiratory burst components, lactoferrin, lysozyme, and collagenase, among other enzymes. Additionally, they release neutrophil extracellular traps (NETs), which are an intertwined mesh of chromatin and neutrophilic granules aimed at killing extracellular pathogens;
    - **Eosinophils** — with a nucleus with two to three lobes and larger granules (**Figure 10**); they are mostly present in tissues, mainly epithelial surfaces, where they interact with IgE antibodies and are mostly responsible for killing parasitic organisms, particularly

tissue-invading nematodes and trematodes. This is achieved through the degranulation and release of major basic protein, which disrupts the cell membrane of the parasites, and pro-inflammatory leukotrienes and peroxidases. Interestingly, their granules also contain anti-inflammatory enzymes, such as histaminase, arylsulfatase, and phospholipase D;

- **Basophils and mast cells** — with a two-lobed nucleus, basophils (**Figure 10**) are found mostly in the blood, while mast cells are found in tissues, particularly the skin and the mucosal surfaces of the respiratory and gastrointestinal tracts. They have granules with histamine, proteoglycans, and proteases that are released upon complement activation or IgE interaction, and produce other pro-inflammatory molecules *de novo*, such as leukotrienes and prostaglandins. They are involved in immediate allergic reactions and anaphylaxis;

- **NK cells** — mononuclear lymphoid cells present in the blood and lymph nodes with a few granules (**Figure 10**); they represent 5–10% of lymphoid cells. These $CD16^{+}CD56^{+}$ cells constantly surveil MHC class I molecules and eliminate abnormal cells through the release of perforins and granzymes, in a mechanism that is nonspecific and absent of memory. They express killer cell immunoglobulin-like activating or inhibitory receptors (KIRs), which mostly interact with class I MHC proteins, and natural cytotoxicity receptors, whose ligands are mostly viral and malignant markers. The death of the target cell depends on the overall balance between inhibitory markers, particularly the expression of healthy "self" molecules, and activating markers, namely, the absence of "self" molecules or the expression of malignant/viral markers. Additionally, NK cells may also perform antibody-dependent cellular cytotoxicity (ADCC), by recognizing target cells coated with IgG through their CD16 receptor. Finally, they also secrete IFN-γ and TNF-α, which have direct cytotoxic effects and activate other components of the innate and adaptive immune systems.

1.2.2. The adaptive immune system

The adaptive immune system is specific to particular antigens, which are recognized by receptors with a unique antigen specificity, whose diversity arises from gene rearrangements taking place throughout life. Following the first contact with an antigen by naïve lymphocytes — immunologic priming —, a response takes around 1–2 weeks to arise; in contrast, a second exposure leads to a quicker (around 3–7 days) and more robust immune response — immunologic memory. The two arms of the adaptive immune system are T lymphocytes, responsible for cellular immunity, and B lymphocytes, responsible for humoral immunity; both derive from the same lymphoid progenitor (**Figure 10**) (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b).

T lymphocytes, also called T cells, migrate from the bone marrow to the thymus (**Figure 9**), where they mature into naïve T cells (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b). In the thymus, they acquire their unique antigen t-cell receptor (TCR), a heterodimer most commonly composed of two chains: α and β, where each chain has a variable region, which interacts with the antigen–MHC complex, and a constant region, which connects the two chains, close to the plasma membrane. Furthermore, the TCR receptor interacts with subunits of the CD3 molecule, which mediates signal transduction of T cell activation through TCR binding (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b). Additionally, only 1% of incoming immature T cells, called thymocytes, leave the thymus as naïve T cells, as they undergo thymic education (Peakman, 2012; Betts *et al.*, 2013; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b):

- **Positive selection** — only cortical thymocytes that successfully bind the MHC self-molecules presented by thymic epithelial cells with low/intermediate affinity are selected;
- **Negative selection** — only thymocytes in the thymic medulla whose CD4 and CD8 molecules are unable to bind to self-antigens presented by APCs survive, with reactive cells undergoing apoptosis;
- **Co-receptor selection** — self-tolerant thymocytes select either CD4 expression, becoming helper T cells, or CD8 expression, becoming cytotoxic T cells.

Most naïve T cells leaving the thymus will be $CD4^+$ (66%) and the vast majority will reside in secondary lymphoid organs, namely lymph nodes and the spleen; only 2% of T cells will circulate in the blood, even though they will still constitute 70–80% of peripheral blood lymphocytes (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b). Although 95% of T cells show a diverse TCR repertoire, through rearrangement of variable (V), diversity (D), and joining (J) gene segments of their variable regions, a few T cells have a restricted TCR repertoire, namely, NK-T cells and mucosal-associated invariant T (MAIT) cells, which use a more limited set of gene segments for their αβ TCRs and are mostly restricted to CD1 molecules and MR1 molecules, respectively, instead of the MHC system, and gamma-delta (γδ) T cells, which have an entirely different TCR with a γδ heterodimer (Mori, Lepore and De Libero, 2016; Haynes, Soderberg and Fauci, 2018; Lepore, Mori and De Libero, 2018; Levinson *et al.*, 2020b).

B lymphocytes, also called B cells, are 3 times less common than T cells, and are mainly responsible for producing antibodies, otherwise known as immunoglobulins, which are crucial in preventing infection by inhibiting pathogens or toxins from reaching their target cells and by promoting the innate immune system in killing or inactivating invading microbes. B cells have their antigen-independent phase maturation in the bone marrow, where they acquire their B-cell receptors (BCRs), also termed surface immunoglobulins, through similar positive and negative selection mechanisms as in T cell development (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a, 2020b).

BCRs are glycoproteins made up of four chains, two identical heavy (H) chains and two identical light (L) chains, selected from the set of four paternal and maternal genes through allelic exclusion (**Figure 11**). Each H and L chain has two main portions: a constant region, closer to the plasma membrane, and a variable region, which binds to antigens and is subjected to VDJ rearrangement (**Figure 11**) (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b). Finally, the immunoglobulin might also be separated into two fragments: the antigen-binding fragment (Fab), which includes the L chains and the most distal portions of the H chains, and the constant or “crystallizable” fragment (Fc), which includes the most proximal portions of the H chains; while the Fab portion interacts with antigens, the Fc portion is crucial for interacting with the plasma membrane of B cells and components of the innate immune system, such as the complement system, APCs,

granulocytes, and NK cells (**Figure 11**) (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b, 2020e).

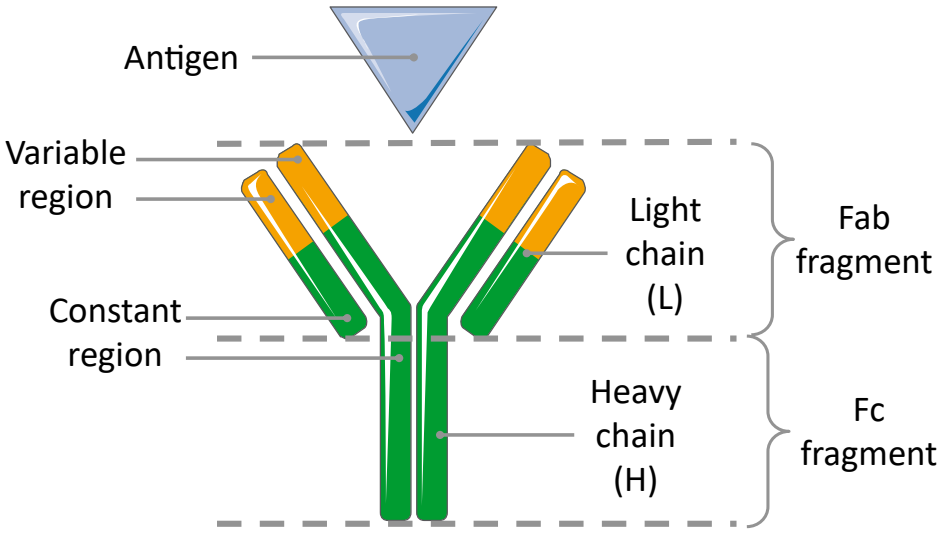


Figure 11 — The immunoglobulin

Templates adapted from (*SMART*, 2021) (CC BY 3.0). Image inspired by (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b). Abbreviations: Fab — antigen binding fragment; Fc — constant or "crystallizable" fragment.

Differences in the constant regions of L and H chains yield different types of chains: L chains are either κ or λ, in a ratio of 2:1, while H chains have five classes, namely, γ, α, μ, ε, and δ, which will determine the function of the antibody (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b). Mature naïve B cells, however, will only have μ and δ H chains as they leave the bone marrow, before undergoing their antigen-dependent phase maturation, which will be described later. As in T cells, most B cells are in secondary lymphoid tissues, mainly lymph nodes and the spleen, and constitute 10–15% of human peripheral blood lymphocytes (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a).

T cells are the primary orchestrators of the adaptive immune response, since they bridge the feedback loop between the adaptive and the innate immune systems, and are capable of influencing virtually all immune cell types (Betts *et al.*, 2013; Levinson *et al.*, 2020c). Mature naïve T cells are activated in secondary lymphoid organs through antigen presentation by either resident APCs or peripheral APCs that have migrated there through the CCR7 chemokine, a processed called priming (Peakman, 2012; Haynes, Soderberg and

Fauci, 2018; Levinson *et al.*, 2020c, 2020e). This activation is mediated through TCR binding with the respective antigen–MHC complex: $CD4^+$ T cells recognize MHC class II molecules and $CD8^+$ T cells recognize MHC class I molecules; this process is further mediated by CD3 molecules, which lead to an influx of calcium, and the subsequent increased expression of IL-2 and the high-affinity IL-2 receptor, through calcineurin (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020c, 2020f). Afterwards, there is a second co-stimulation between the CD28 receptor on T cells and B7 proteins in APCs, termed CD80 for B7-1 and CD86 for B7-2, which are only highly expressed in activated APCs; a previous TCR activation without this co-simulation will lead to T cell anergy rather than activation (Peakman, 2012; Betts *et al.*, 2013; Haynes, Soderberg and Fauci, 2018). Importantly, this co-stimulation may be further balanced by inhibitory surface markers in T cells that compete with the CD28–CD80/CD86 binding, namely, cytotoxic T lymphocyte antigen-4 (CTLA-4) and programed cell death-1 (PD-1) (Betts *et al.*, 2013; Haynes, Soderberg and Fauci, 2018). Finally, T cells undergo clonal proliferation, where IL-2 serves as the main T cell growth factor, giving rise to different progeny cells according to their cytokine environment, albeit reactive to the same antigen (Peakman, 2012; Betts *et al.*, 2013; Haynes, Soderberg and Fauci, 2018):

- **$CD4^+$ cells**:
    - **Effector/helper (Teff or Th) cells** — leave the secondary lymphoid organ and coordinate immune responses in the periphery by enhancing the functions of other cells; depending on the cytokines required for their polarization and the ones they produce; they are further subdivided in:
        - **Th1 cells** — activate macrophages and instruct cytotoxic T cells, aimed at killing bacteria, fungi, and viruses, particularly intracellular microbes; they are polarized by IL-12 and they produce IFN-γ, IL-2 and TNF-α;
        - **Th17 cells** — closely associated with Th1 cells, they are responsible for reinforcing the barrier tissues against bacterial and fungal infections, by stimulating mucosal epithelial cells to increase the production of IL-1, IL-6, and neutrophil-attracting chemokines, as well as the production of antimicrobial

defensins and tight junction proteins; they are polarized by IL-1, IL-6, IL-23, and TGF-β and they produce IL-17, IL-21, and IL-22;

- **Th2 cells** — recruit eosinophils and mast cells, aimed at killing parasites, mainly helminths, and are involved in antibody production, particularly antibody class switching to IgE; they are polarized by IL-4 and they produce IL-4, IL-5, and IL-13.

- **Follicular helper (Tfh) cells** — assist B cells in secondary lymphoid organs, which they find by downregulating the chemokine receptor CCR7 and upregulating the chemokine receptor CXCR5; they produce IL-4 and IL-21;
- **Regulatory (Treg) cells** — crucial in limiting extensive immune responses, which may lead to tissue damage, and in maintaining tolerance of self-antigens and commensal antigens, avoiding autoimmunity; they are both polarized by and produce IL-10 and TGF-β;

- **$CD8^+$ cells** — give rise to cytotoxic/cytolytic T cells (Tc cells), which are crucial for killing virus-infected or malignant cells, which will have altered MHC class I expressions. Tc cells secrete perforins, which form a channel through the membrane of the target cell, and proteases, leading to the direct death of target cells. Additionally, they induce the expression of Fas ligand (FasL) and produce granzyme B, a protease, which lead to the activation of caspases and, ultimately, apoptosis. Finally, they also secrete IFN-γ and TNF-α, which recruit other innate immune cells proficient in killing intracellular pathogens. They require an additional signal to become activated, called licensing, which is transmitted by $CD4^+$ T cells. Through cross-presentation, exogenous antigens may be presented on MHC class I molecules, in parallel with MHC class II presentation. The activation of $CD4^+$ T cells will increase the binding between CD40L on the $CD4^+$ T cell and CD40 on the activated professional APC, which will then lead to an increased expression of co-stimulatory molecules for the APC–Tc cell contact. Also, it will lead to an increased IL-2 production by the activated $CD4^+$ T cell, which will promote the clonal proliferation of the Tc cell.

Additionally, T cell clone proliferation gives rise to memory subsets, which persist through several cell generations in two main pools: effector memory T cells, which reside in nonlymphoid organs and quickly respond to infections with cytokine production and cytotoxicity; and central memory T cells, which reside in lymphoid organs, where they replenish other memory T cells (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020c). Upon a second exposure to a known antigen, a secondary immune response is formed, which is much quicker, more efficient, and more vigorous (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020b, 2020c). This is caused by: (**I**) a greater starting pool of memory cells than naïve cells in the primary immune response; (**II**) a lower threshold of activation, namely, smaller amounts of antigen and less stringent co-stimulation; and (**III**) increased secretion of cytokines by memory cells (Levinson *et al.*, 2020c). However, memory T cells may become depleted in persistent antigen exposure, such as chronic viral infections or malignancies, leading to T cell exhaustion (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020c).

Regarding B cells, in contrast to TCRs, which can only bind MHC–antigen complexes, the BCR can bind to unprocessed native antigens (Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a, 2020b). Antigen binding triggers the antigen-dependent phase maturation for B cells, which begins with the cross-linking and aggregation of multiple BCRs, facilitated by CD19, and, subsequently, the activation of three main pathways: phospholipase C-γ2 (PLC-γ2), phosphatidalyinositol-3-kinase (PI3K), and Bruton tyrosine kinase (Btk) (Woyach, Johnson and Byrd, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a). The second step in B cell activation depends on whether it is T-cell-independent or T-cell-dependent; without this step, B cells will undergo apoptosis or become anergic (Levinson *et al.*, 2020a).

For T-cell-independent activation, the antigen is usually a large multivalent molecule that is able to activate multiple BCRs at once; a second signal for activation will come from either the recognition of PAMPs by PRRs expressed by B cells or from molecules of the complement system that bind to corresponding receptors in B cells. This will lead to short-lived antibody production and mostly IgM secretion, although a few IgG antibodies may also be produced (Levinson *et al.*, 2020a).

For T-cell-dependent activation, B cells will first phagocytize antigens bound to their BCRs and process the resulting peptides for MHC class II presentation as professional APCs;

afterwards, they will come into contact with a Tfh cell, which was previously activated by the same antigen presented by another professional APC, most likely a dendritic cell (Levinson *et al.*, 2020a). Then, through the interaction between the CD40L on Tfh cells and CD40 on B cells and the production of IL-21 by Tfh cell, B cells will undergo: (**I**) clonal proliferation, into either long-lived plasma cells, which secrete more antibodies, or into long-lived memory B cells; (**II**) irreversible H chain class switching, allowing for the production of other H chains, namely, γ, α, and ε, depending on their cytokine environment; and (**III**) somatic hypermutation, which corresponds to nucleotide changes in the variable gene segments of both antibody chains, leading to a competition for antigen binding between slightly different antibodies and, subsequently, Tfh cell communication, and survival, a process termed affinity maturation (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a). Upon a second exposure to the same antigen, this process of affinity maturation will repeat itself, albeit starting with memory B cells, leading to a more rapid response with higher titers of long-lived IgG antibodies, which will also have higher affinity (Levinson *et al.*, 2020a).

H chain class-switching ultimately allows for different antibody isotypes with very different functions (**Table 1**) (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a).

Table 1 — Immunoglobulin isotypes: Summary

Property	IgG	IgA	IgM	IgD	IgE
H chain	γ	α	μ	δ	ε
Main role	Dominant isotype in secondary immune response	Secretory immunoglobulin in mucosal surfaces, it prevents attachment of bacteria and viruses	One of the first surface immunoglobulins in mature naïve B cells and the first isotype to be secreted in a primary immune response	One of the first surface immunoglobulins in mature naïve B cells; additional functions unknown	Defends against parasites, mainly helminths; it is also involved in symptoms of allergy
Binding cells via Fc	Macrophages, neutrophils, NK cells	Lymphocytes	Lymphocytes	-	Mast cells, basophils, B cells
Prevalence in serum for total immunoglobulin (%)	75–85	7–15	5–10	0.2–0.3	0.004–0.019
Structure	Monomer	Monomer or dimer	Pentamer (serum)	Monomer	Monomer
Half-life (days)	21–23	6	5–10	3	2–2.5
Subclasses	G1, G2, G3, G4	A1, A2	-	-	-
Complement activation	++	+	+++	+	-
Opsonization	+	-	-	-	-
Transplacental passage	+	-	-	-	-

Adapted from (Peakman, 2012; Haynes, Soderberg and Fauci, 2018; Levinson *et al.*, 2020a). Abbreviations: Fc — constant or "crystallizable" fragment; H — heavy; Ig — Immunoglobulin; NK — natural killer.

1.3. Neuroimmunology

Neuroimmunology is a fairly recent field of research, which focuses on the crosstalk between the nervous system and the immune system, namely, in the spectrum ranging from physiological processes to pathological conditions ('A neuroimmune odyssey', 2020). Although viewed as very separate entities, both the nervous system and the immune system have arisen in invertebrates, have bidirectional communication, integrate stimuli into memory, and have two main common goals: internal homeostasis and host defense

(Talbot, Foster and Woolf, 2016; Korn and Kallies, 2017; Pavlov and Tracey, 2017; Veiga-Fernandes and Pachnis, 2017; Dantzer, 2018; Morimoto and Nakajima, 2019).

The symbiosis of both systems is further substantiated by the existence of similar molecular patterns between them: neurons, astrocytes, and oligodendrocytes require "immune" molecules, such as cytokines, chemokines, MHC class I molecules, and TLRs, for adequate neural development, synapse plasticity, and higher functions (e.g., memory, cognition, and learning) (Deverman and Patterson, 2009; Marin and Kipnis, 2013; Becher, Spath and Goverman, 2017; Fung, Olson and Hsiao, 2017; Klein, Garber and Howard, 2017; Pavlov and Tracey, 2017; Morimoto and Nakajima, 2019), while immune cells release "neurotransmitters", such as acetylcholine and catecholamines, and express adrenergic receptors, namely, the β2 subtype for innate and adaptive immune cells and the α1 and α2 subtypes for innate immune cells (Pavlov and Tracey, 2017; Dantzer, 2018).

A few notable examples supporting the relevance of neuroimmune interactions in the PNS include (Talbot, Foster and Woolf, 2016; Wohleb *et al.*, 2016; Fung, Olson and Hsiao, 2017; Pavlov and Tracey, 2017; Veiga-Fernandes and Pachnis, 2017; Dantzer, 2018):

- the communication between immune cells and nociceptor neurons, through cytokines, interleukins and neuropeptides, via a feedforward loop once a pathogen is identified, leading to: (**I**) defense mechanisms, such as itchiness, pain, and coughing; (**II**) neurogenic inflammation, characterized by increased vascular leakage as a consequence of neural control of endothelial and vascular smooth muscle cells, which further aids the immune system in fighting the pathogen; and (**III**), in dysfunctional feedforward loops, disorders, such as asthma and allodynia, where non-noxious stimuli are perceived as pain;
- neuroimmune cell units in the intestine, where the autonomous/enteric nervous system is deeply interconnected with Peyer's patches within MALT, leading to: (**I**) homeostasis of the digestive system; (**II**) immune maturation and regulation of the adaptive immune system; (**III**) the sensing of pathogens and molecules from the gut microbiota, the intestinal commensal flora with trillions of microorganisms, which will have a crucial role in the development and functioning of both the immune system and the nervous system, including cognitive functions, behavior, and neurological or psychiatric

diseases, such as Parkinson's disease, Alzheimer's disease, and major depressive disorder;

- the interconnectedness between immune cells and the autonomous nervous system. On the one hand, psychosocial or environmental stress activate the hypothalamic–pituitary–adrenal (HPA) axis and the sympathetic nervous system, which ubiquitously innervates lymphoid organs, leading to glucocorticoid and catecholamine release, which, in turn, are responsible for an influx of myeloid immune cells in circulation, the release of pro-inflammatory citokines, and the retaining of T cells in lymph nodes through β2 adrenergic signaling; this promotes CNS neuroinflammation and, when chronically activated, it may lead to anxiety disorders or major depressive disorder. On the other hand, the parasympathetic nervous system acts as a counterbalance through the inflammatory reflex, where vagal nerve afferents sense cytokines and inflammatory products, communicating with the nucleus tractus solitarius in the brain stem, which then transmits an efferent cholinergic anti-inflammatory pathway. In this pathway, the vagus nerve communicates with the splenic nerve via the celiac ganglion, leading to the release of norepinephrine in the spleen, which is picked up by a subset of splenic T cells able to produce acetylcholine through β2-adrenergic signaling, ultimately leading to the inhibition of TNF-α production by macrophages that express an acetylcholine receptor, namely, the α7 nicotinic acetylcholine receptor. Furthermore, vagus–splenic signaling is also linked to a modulation of humoral immunity, particularly the retention of B cells in the spleen and a decrease in antibody production.

The role of the immune system in the CNS, however, has been a matter of intense debate (Bucchieri *et al.*, 2015; Louveau, Harris and Kipnis, 2015; Mezey and Palkovits, 2015). For decades, the CNS was considered an immune-privileged site (Louveau, Harris and Kipnis, 2015; Engelhardt, Vajkoczy and Weller, 2017; Korn and Kallies, 2017; Prinz and Priller, 2017; Morimoto and Nakajima, 2019), due to: (**I**) the survival of foreign tissue grafts in its parenchyma (Widner and Brundin, 1988); (**II**) the lack of response by the peripheral immune system to CNS parenchymal infections or damage (Stevenson P G *et al.*, 1997); and

(**III**) the very selective barriers around the brain, namely, the BBB, the blood–CSF barrier, and the CSF–brain barrier, which restrict entry of peripheral immune cells (Engelhardt, Vajkoczy and Weller, 2017; Prinz and Priller, 2017).

Despite this, it is now known that: (**I**) the CNS is permeable to cytokines, such as the fever response to pyrogenic cytokines through CNS prostaglandin production (Talbot, Foster and Woolf, 2016; Dantzer, 2018); (**II**) T cells have a critical role in neurogenesis, synaptic plasticity, learning, and memory (Brynskikh *et al.*, 2008; Kipnis, Gadani and Derecki, 2012; Marin and Kipnis, 2013; Radjavi, Smirnov and Kipnis, 2014; Prinz and Priller, 2017; Ribeiro Miguel *et al.*, 2019); (**III**) tissue-resident $CD8^+$ memory T cells are crucial for combating viral parenchymal infections (Stevenson, Bangham and Hawke, 1997; Korn and Kallies, 2017); and, most recently brought to light, (**IV**) dural lymphatic vessels drain CSF and interstitial fluid (ISF) from the CNS, particularly containing immune cells and soluble compounds, to deep cervical lymph nodes (Aspelund *et al.*, 2015; Louveau *et al.*, 2015, 2018; Louveau, Harris and Kipnis, 2015; Absinta *et al.*, 2017).

However, this latest description of CSF and ISF drainage to deep cervical lymph nodes through meningeal lymphatic vessels, named the glymphatic system (**Figure 12**) (Aspelund *et al.*, 2015; Louveau *et al.*, 2015; Absinta *et al.*, 2017; Engelhardt, Vajkoczy and Weller, 2017; Prinz and Priller, 2017), was not as novel as it seemed (Bucchieri *et al.*, 2015; Mezey and Palkovits, 2015), since CSF drainage to deep cervical lymph nodes was first described in 1869 (Schwalbe, 1869) and dural lymphatic vessels were first described in 1787 (Mascagni and Sanctius, 1787).

In the CNS, microglia are the resident immune cells, representing 5–20% of total brain cells and 80% of brain immune cells and maintaining their CNS population independently from peripheral immune cells through self-renewal (Fung, Olson and Hsiao, 2017; Prinz and Priller, 2017; Dantzer, 2018; Li and Barres, 2018; Morimoto and Nakajima, 2019). During the development of the CNS, they are directly involved in a myriad of processes, such as (Wohleb *et al.*, 2016; Dantzer, 2018; Li and Barres, 2018; Morimoto and Nakajima, 2019): (**I**) angiogenesis and vascular anastomosis modulation; (**II**) neurocircuitry refinement, by eliminating synapses in an activity-dependent manner using elements of the complement system; (**III**) regulation of neurogenesis, through supporting neuronal survival and inducing the cell death of neurons and neuronal precursors; and (**IV**) the laminar positioning of neurons. Postnatally, as mentioned previously (more details in *1.1.3. The cellular*

constituents of the nervous system), they are further involved in synapse maturation and pruning, as well as promoting the survival and differentiation of oligodendrocyte progenitor cells (OPCs) (Paolicelli Rosa C. *et al.*, 2011; Li and Barres, 2018; Weinhard *et al.*, 2018; Morimoto and Nakajima, 2019; Theibert, 2020f). In adults, they phagocytize myelin, while maintaining the pool of OPCs, they regulate neurogenesis through the phagocytosis of neuronal precursors, and they monitor neuronal activity, by sensing fluctuations in neurotransmitters and ATP (Wohleb *et al.*, 2016; Li and Barres, 2018).

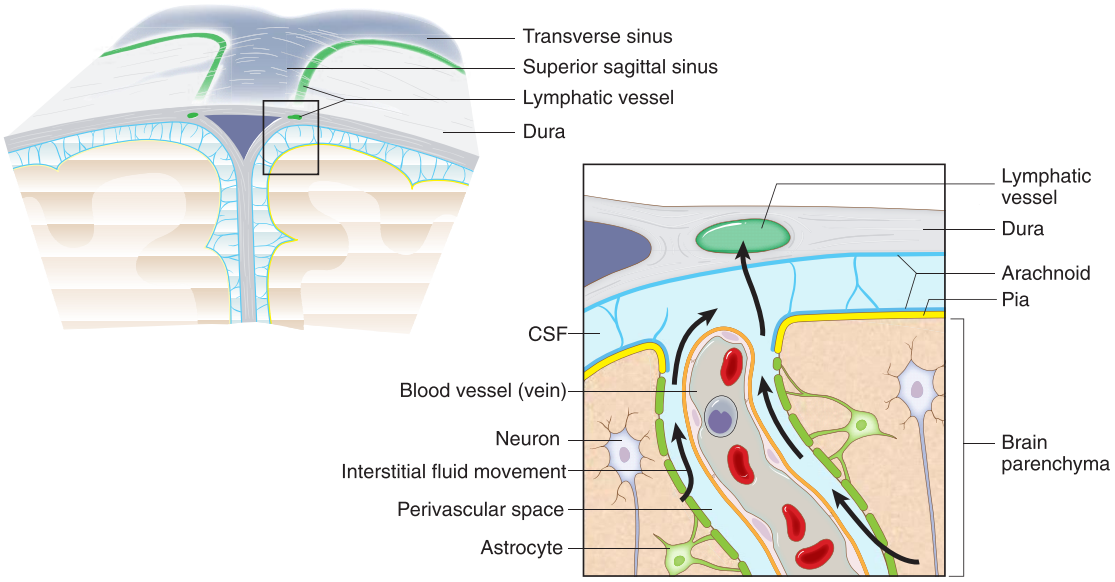


Figure 12 — The glymphatic system

Unedited image extracted from (Louveau *et al.*, 2015). Copyright © 2015 Macmillan Publishers Limited. Abbreviations: CSF — cerebrospinal fluid.

In addition to microglia, astrocytes are also pivotal modulators of neuroinflammation, due to their role in detecting PAMPs and DAMPs, presenting antigens to adaptive immune cells through MHC class II molecules, producing cytokines and chemokines, and, ultimately, integrating the neuroinflammatory environment and responding with either pro-inflammatory or anti-inflammatory signaling pathways (Dong and Benveniste, 2001; Farina, Aloisi and Meinl, 2007; Colombo and Farina, 2016; Fung, Olson and Hsiao, 2017; Klein, Garber and Howard, 2017; O'Reilly and Tom, 2020).

Other immune cells, namely, granulocytes, monocytes/macrophages, dendritic cells, T cells, B cells, and NK cells, are also present in the brain, although almost exclusively in the

meninges and choroid plexus (Korin *et al.*, 2017; Van Hove *et al.*, 2019). The meninges (more details in *1.1.1. The central nervous system*) integrate both resident and peripheral cells of the innate and adaptive immune systems and serve as a conduit for neuroimmune communication and CNS immune surveillance (Rua and McGavern, 2018; Alves de Lima, Rustenhoven and Kipnis, 2020). Upon CNS injury or neuroinflammation, microglia and meningeal macrophages will release pro-inflammatory cytokines, e.g., IL-1β, TNF-α, and IFN-γ, and chemokines, e.g., CCL2, CCL5, CXCL9–11, and CXCL12, altering the usually restrictive barriers around the CNS, which will start expressing cell adhesion molecules, e.g., VCAM-1 and ICAM-1, and lead to an infiltration of immune cells in the CNS parenchyma (Engelhardt and Ransohoff, 2012; Louveau, Harris and Kipnis, 2015; Schläger *et al.*, 2016; Becher, Spath and Goverman, 2017; Engelhardt, Vajkoczy and Weller, 2017; Klein, Garber and Howard, 2017; Korn and Kallies, 2017; Prinz and Priller, 2017; Li and Barres, 2018; Rua and McGavern, 2018; Alves de Lima, Rustenhoven and Kipnis, 2020; O'Reilly and Tom, 2020).

Although peripheral immune cells are vital in supporting the resident CNS cells in killing pathogens, phagocytizing debris, and tissue remodeling, the tissue damage secondary to pro-inflammatory cytokines, increased ROS production, and extensive phagocytosis might be detrimental, leading to demyelination, extensive scar formation, and neuroaxonal loss (Yamasaki *et al.*, 2014; O'Reilly and Tom, 2020). Ultimately, this disruption might lead to the development of neuroinflammatory disorders, which account for a significant proportion of disability in people within working age, attributable to a progressive decline in neurologic function (Meador, 2020).

1.4. Multiple Sclerosis

1.4.1. Definition and history

Multiple Sclerosis (MS) is chronic neuroinflammatory and neurodegenerative disease, characterized by an autoimmune demyelination secondary to the infiltration of immune cells in the CNS, further leading to reactive scarring — gliosis —, oligodendrocyte death, and neuroaxonal loss (Dendrou, Fugger and Friese, 2015; Berkowitz, 2016a; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019; Meador, 2020).

MS was first described in a series of lectures in May 1868 as S*clérose en plaques* by physician and scientist Jean Martin Charcot working at the Salpêtrière hospital in Paris, who was able to integrate previous scientific observations with his own into a mysterious and heterogeneous disease, which he presented in full, including its anatomo-pathology, symptomatology, and, finally, etiology and treatment (Jean-Martin Charcot, 1875; Zalc, 2018; Ropper *et al.*, 2019).

In the past 150 years, MS has been through a revolution in disease diagnosis and management, which has contributed to partially altering the "extremely severe" prognosis in the age of Charcot (Jean-Martin Charcot, 1875; Zalc, 2018) by reducing relapses, hospitalizations, and deaths, albeit significant disability still remains in many patients (Dendrou, Fugger and Friese, 2015; Kobelt *et al.*, 2017; Reich, Lucchinetti and Calabresi, 2018; Wallin *et al.*, 2019).

1.4.2. Epidemiology

Worldwide, approximately 2.8 million people are living with MS and someone in the world is diagnosed with MS every five minutes (Multiple Sclerosis International Federation, 2020; Walton *et al.*, 2020). MS prevalence has increased in the past decades in all world regions, regardless of the significant regional differences in prevalence, ranging from 4.79 per 100,000 people in the Western Pacific to 142.81 per 100,000 people in Europe (**Figure 13**) (James *et al.*, 2018; Wallin *et al.*, 2019; Multiple Sclerosis International Federation, 2020; Walton *et al.*, 2020). Similar regional differences are also present within Europe, with higher latitude countries having an increased incidence and prevalence (Kingwell *et al.*, 2013), an aspect that is also present in other regions (Simpson *et al.*, 2019). Regarding temporal changes in MS incidence, however, studies have shown conflicting results (Koch-Henriksen *et al.*, 2018; Rotstein *et al.*, 2018; Hawkes *et al.*, 2020; Iljicsov *et al.*, 2020; Hosseinzadeh *et al.*, 2021).

Globally, women are twice as likely to have MS than men, a pattern that remains in all world regions (James *et al.*, 2018; Wallin *et al.*, 2019; Multiple Sclerosis International Federation, 2020; Walton *et al.*, 2020). MS is usually diagnosed in young adults, with a mean age at diagnosis of 32 years, although a symptomatic prodromal phase usually takes place a few years earlier (Kister *et al.*, 2013; Kobelt *et al.*, 2017; Disanto *et al.*, 2018; Multiple Sclerosis International Federation, 2020). In addition, although pediatric cases are

much rarer (approximately 1.5% of global MS cases), prevalence has risen substantially in recent years (Multiple Sclerosis International Federation, 2020).

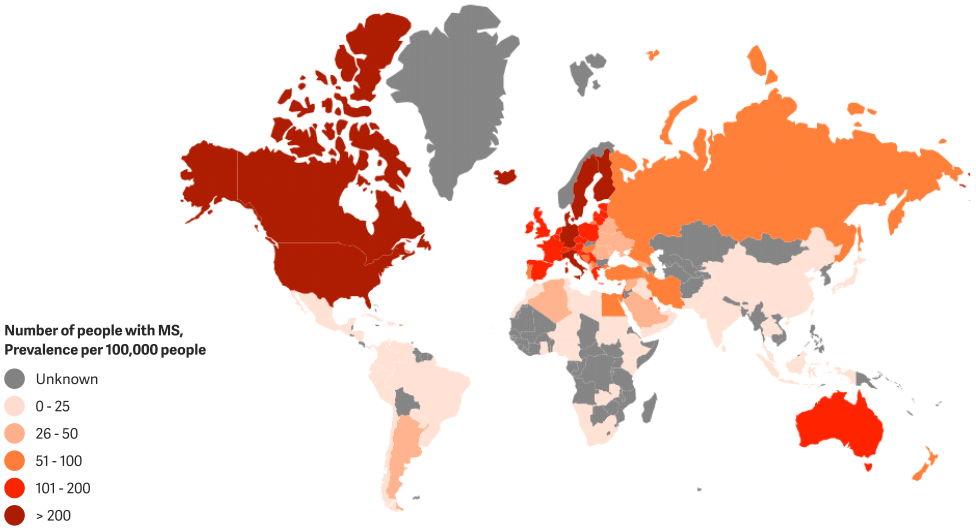


Figure 13 — Worldwide prevalence of Multiple Sclerosis

Unedited image extracted from (Multiple Sclerosis International Federation, 2020). Copyright © Multiple Sclerosis International Federation (MSIF).

Accordingly, MS has a significant impact in patients with MS (PwMS), leading to increased rates of psychological distress (Patti and Vila, 2014; Kobelt *et al.*, 2017), unemployment due to disability (Patti and Vila, 2014; Kobelt *et al.*, 2017), and higher suicide rates (Shen *et al.*, 2019; Erlangsen *et al.*, 2020). A recent study (Cortese *et al.*, 2022) found that female PwMS show an accelerated age-related decline in physical health by 15–30 years, in comparison with healthy controls (HC).

Ultimately, this further leads to significant societal burden as disease progresses (Kobelt *et al.*, 2017), driven in part by rising costs in disease management (Hartung *et al.*, 2015). For instance, in Germany, where 252,000 PwMS live (Multiple Sclerosis International Federation, 2020), MS affected productivity at work for 80% of PwMS and the mean cost of a relapse was €2468 (Flachenecker *et al.*, 2017). In parallel, in Portugal, where 5,787 PwMS live (Multiple Sclerosis International Federation, 2020), productivity was affected in 72% of PwMS and the mean cost of a relapse was €2931 (Sá *et al.*, 2017).

1.4.3. Clinical course and management

Since MS may affect any region of the CNS, it may lead to a very heterogeneous presentation of clinical symptoms, including issues in mobility, dexterity, the sensory system, balance, sphincter function (bladder and bowel), vision, and cognition (Kister *et al.*, 2013; Patti and Vila, 2014; Dendrou, Fugger and Friese, 2015; Berkowitz, 2016a; Reich, Lucchinetti and Calabresi, 2018; Thompson *et al.*, 2018; Ropper *et al.*, 2019; Meador, 2020). Diagnosis is made more difficult by the possibility of having a single focal neurological deficit as a debut event, before a dissemination throughout the CNS takes place (Thompson *et al.*, 2018; Ropper *et al.*, 2019). Overall, fatigue is the most prevalent and most burdensome symptom amongst patients (Kister *et al.*, 2013; Patti and Vila, 2014; Kobelt *et al.*, 2017; Penner and Paul, 2017; Meador, 2020), with a clinically challenging management, which requires an interdisciplinary and individualized therapeutic strategy aimed at addressing comorbid, sleep, and depressive disorders, as well as medication side effects (Veauthier *et al.*, 2016).

The majority of initial lesions affects (Brownlee *et al.*, 2017; Ropper *et al.*, 2019): (**I**) the optic nerve, giving rise to optic neuritis, where there is a partial or total loss of visual acuity, usually unilaterally and with a quick and full recovery; (**II**) the spinal cord, giving rise to acute myelitis with a short vertical extent, where there is usually an asymmetrical and incomplete paraparesis or paraplegia with ascending paresthesia, among other clinical signs depending on the lesion's level and extent; and (**III**) the brain stem, giving rise to a myriad of symptoms, such as diplopia, vertigo, facial pain or numbness, and dysarthria. Other classic signs of MS are (Berkowitz, 2016a; Ropper *et al.*, 2019):

- **Uhthoff's phenomenon** — the recurrence or emergence of neurologic symptoms with heat, mainly driven by the altered nerve conduction of demyelinated axons;
- **Lhermitte's sign** — an electrical sensation down the spine upon forward flexion of the neck, present in any cervical myelopathy;
- **Internuclear ophthalmoplegia** — which is a paresis of the medial rectus on attempted lateral gaze with a coarse nystagmus in the abducting eye, due to disruption of the medial longitudinal fasciculi; in MS, it is usually bilateral.

Patients with a suspicious clinical presentation are usually assessed with paraclinical investigations, namely brain and spinal cord magnetic resonance imaging (MRI), which is able to identify either symptomatic or asymptomatic lesions suggestive of MS, and have their CSF collected for further analysis, where they may show an increased protein content, particularly in gamma immunoglobulins not present in serum — oligoclonal bands (Dendrou, Fugger and Friese, 2015; Berkowitz, 2016a; Brownlee *et al.*, 2017; Reich, Lucchinetti and Calabresi, 2018; Thompson *et al.*, 2018; Ropper *et al.*, 2019; Meador, 2020; Wattjes *et al.*, 2021).

Typical MRI lesions in PwMS mostly occur in the white matter and are usually located in particular regions, namely, in the brain, in periventricular, cortical/juxtacortical, and infratentorial regions, and in the spinal cord (Geraldes *et al.*, 2018; Thompson *et al.*, 2018). MRI lesions will usually be multifocal, well delineated, oval or linear, and radially oriented, with an hyperintensity on T2-weighted images, particularly on T2 fluid-attenuated inversion recovery (T2-FLAIR) images (Berkowitz, 2016a; Ropper *et al.*, 2019; Meador, 2020; Wattjes *et al.*, 2021). Acute lesions will tend to demonstrate edema and an abnormal enhancement following the administration of gadolinium in a C-shaped partial or open ring, due to the disruption in the BBB, while chronic lesions might be observed as T1 “black holes”, with T1 hypointensity having been shown as being inversely proportional to the degree of remyelination (Barkhof *et al.*, 2003; Berkowitz, 2016a; Ropper *et al.*, 2019; Meador, 2020).

Other paraclinical investigations include optical coherence tomography (OCT), which measures neuroaxonal loss in the retina, particularly, the retinal nerve fiber layer (RNFL) and the ganglion cell layer (GCL) (Berkowitz, 2016a; Brownlee *et al.*, 2017; Reich, Lucchinetti and Calabresi, 2018; Thompson *et al.*, 2018; Ropper *et al.*, 2019; Lambe, Saidha and Bermel, 2020), and evoked potentials in visual, somatosensory, or auditory pathways (Berkowitz, 2016a; Brownlee *et al.*, 2017; Reich, Lucchinetti and Calabresi, 2018; Thompson *et al.*, 2018; Ropper *et al.*, 2019; Meador, 2020).

A patient with a clinical presentation suggestive of MS, namely, an acute or subacute monophasic clinical episode with patient-reported symptoms and objective findings that reflect a focal or multifocal inflammatory demyelinating event in the CNS, with a duration of at least 24 h, irrespective of recovery, and in the absence of fever or infection, requires

two main thresholds for diagnosis, defined by the 2017 revisions of the McDonald criteria (Thompson *et al.*, 2018):

- **Dissemination in space (DIS)** — either the occurrence of a previous clinical episode in another area of the CNS, or the presence of MRI lesions in two distinct anatomical locations within the most typically affected in the CNS (periventricular, cortical/juxtacortical, and infratentorial regions in the brain, and in the spinal cord);
- **Dissemination in time (DIT)** — either a reoccurrence of a clinical episode in a different CNS site, the simultaneous presence of gadolinium-enhancing and non-enhancing lesions at any time, or the presence of a new T2-hyperintense or gadolinium-enhancing lesion on a follow-up MRI, in comparison with a baseline scan. Alternatively, CSF-specific oligoclonal bands may replace DIT.

Patients fulfilling the criteria above will be diagnosed as PwMS; conversely, patients who do not fulfil the criteria for DIS and DIT, and patients with incidental and asymptomatic MRI lesions, ruling out other diagnoses or explanations, will be diagnosed with a Clinically Isolated Syndrome (CIS) and a Radiologically Isolated Syndrome (RIS), respectively (Berkowitz, 2016a; Thompson *et al.*, 2018; Ropper *et al.*, 2019; Meador, 2020). Conversion from CIS to MS eventually takes place in the majority (60–70%) of CIS patients, where MRI disease burden or pattern of lesions and the presence of CSF oligoclonal bands are relevant risk factors (Fisniku *et al.*, 2008; Kuhle *et al.*, 2015; H. Zhang *et al.*, 2019; Chung *et al.*, 2020; Kolčava *et al.*, 2020).

Most PwMS (85–90%) are diagnosed with Relapsing-Remitting MS (RRMS), where neurological symptoms remit and recur, in a cyclical manner, while a minority of the patients (10–15%) have Primary Progressive MS (PPMS), where disability steadily increases without recurring (Dendrou, Fugger and Friese, 2015; Berkowitz, 2016a; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019; Meador, 2020; Multiple Sclerosis International Federation, 2020). Typically, a relapse occurs every 13–15 months without a discernable precipitating factor (Ropper *et al.*, 2019; Meador, 2020). Rarely, patients present with a rapidly progressive and highly debilitating form with large acute lesions with mass effects on surrounding structures that require a biopsy — tumefactive MS or Marburg variant (Berkowitz, 2016a; Ropper *et al.*, 2019; Meador, 2020).

A variable proportion of patients with RRMS (20–60%) (Tremlett, Yinshan Zhao and Devonshire, 2008; University of California *et al.*, 2016; Chung *et al.*, 2020) will also develop a similar progressive pattern of disease — Secondary Progressive MS (SPMS) — after a few decades, where increased relapse rate or incomplete recovery from the first event (Langer-Gould *et al.*, 2006; Scalfari *et al.*, 2014; Alroughani *et al.*, 2015; Novotna *et al.*, 2015; Fambiatos *et al.*, 2020), older age at onset (Scalfari *et al.*, 2014; Alroughani *et al.*, 2015; Fambiatos *et al.*, 2020), male sex (Scalfari *et al.*, 2014; Alroughani *et al.*, 2015), and disease duration (Scalfari *et al.*, 2014; Fambiatos *et al.*, 2020) are predictive of higher disability, while immunomodulatory treatment with disease-modifying drugs (DMDs) is protective of disability (Bergamaschi *et al.*, 2016; Fambiatos *et al.*, 2020).

Disability is usually measured clinically with two scales (Cohen *et al.*, 2012): (**I**) the revised expanded disability status scale (EDSS) (Kurtzke, 1983), with a range of 0–10, which is based on a structured neurological examination that qualitatively assesses functional system scores (visual, brain stem, pyramidal, cerebellar, sensory, bowel and bladder, cerebral), mobility, and independence in daily living; and (**II**) the more recent Multiple Sclerosis functional composite (MSFC) (Cutter *et al.*, 1999), which quantitively assesses ambulation, upper-extremity function, and cognition, normalized to the study population through Z scores.

Regarding disease management, treatment of acute exacerbations is managed through corticosteroids, either intravenously or orally, namely, methylprednisolone (most commonly) or prednisone; these appear to hasten the recovery from the clinical event, but do not affect the likelihood of developing MS from CIS, nor prevent disability (Berkowitz, 2016a; Ropper *et al.*, 2019; Meador, 2020). For PwMS whose disability is resistant to treatment with corticosteroids, apheresis techniques, i.e., plasmapheresis/plasma exchange or immunoadsorption, are used as an alternative, where pathogenic molecules in circulation, such as autoantibodies, immune complexes, and inflammatory mediators, are removed through extracorporeal plasma purification (Dorst *et al.*, 2019; Rolfes *et al.*, 2019; Marrodan *et al.*, 2021).

For long-term management, there are, as of January 6 2022, 16 DMDs marketed in the European Union (EU) specifically for treating MS (**Table 2**) (Human Medicines Evaluation Division, European Medicines Agency, 2021; European Medicines Agency, 2022), with the aim of reducing relapses and impeding disability (Berkowitz, 2016a; Trojano

et al., 2017; Montalban *et al.*, 2018; Ropper *et al.*, 2019; Tintore, Vidal-Jordana and Sastre-Garriga, 2019; Meador, 2020). The number of DMDs has skyrocketed in recent years: 81% of the currently active DMDs in the EU were authorized in the past 16 years (Human Medicines Evaluation Division, European Medicines Agency, 2021; European Medicines Agency, 2022). The prescription of DMDs is done in specialized centers based on diverse factors, such as disease subset (CIS, RRMS, PPMS, or SPMS), clinical and/or radiological disease activity, comorbidities, family planning, response to current and previous DMDs, and drug safety profile (Montalban *et al.*, 2018). Generally, PwMS have started treatment with less efficacious drugs (**Table 2**), switching to a more efficacious drug if their disease activity worsened; however, current clinical practice is shifting towards treatment of PwMS with highly efficacious drugs from the start (Spelman *et al.*, 2021; Sue Hughes, 2021).

Table 2 — Disease modifying drugs for Multiple Sclerosis in the European Union

Drug	Comm. name(s)	Mode of administration	Mechanism of action	Most common side effects	EMA date of authorization	Efficacy (Atlas of MS)
Interferon beta-1b	Betaferon, Extavia	Injectable	Diverse immunomodulatory effects	Flu-like symptoms and reactions at the site of injection	30/11/1995 (Betaferon) and 20/05/2008 (Extavia)	Moderate
Interferon beta-1a	Avonex, Rebif			Headache, flu-like symptoms, fever, chills, sweating, neutropenia, lymphopenia, leucopenia, thrombocytopenia, anemia, reactions at the site of injection, and increases in liver enzymes.	13/03/1997 (Avonex) and 03/05/1998 (Rebif)	Moderate
Peginterferon beta-1a	Plegridy			Headache, muscle pain, joint pain, flu-like symptoms, fever, chills, asthenia, and erythema, reactions at the site of injection	17/07/2014	Moderate
Glatiramer acetate	Copaxone (among others)	Injectable	Diverse immunomodulatory effects	Anxiety, nausea, rashes, difficulty breathing, vasodilation, back pain, thoracic pain, reactions at the site of injection	NA (decentralized process); FDA: 20/12/1996	Moderate
Natalizumab	Tysabri	IV infusion	Monoclonal antibody targetting the α4β1 integrin, used by lymphocytes to enter the brain and spinal cord	Urinary tract infections, nasopharyngitis, headache, dizziness, nausea, joint pain, asthenia. Risk of progressive multifocal leukoencephalopathy due to John Cunningham (JC) virus infection	27/06/2006	High

Table 2 — Disease modifying drugs for Multiple Sclerosis in the European Union (continued)

Drug	Comm. name(s)	Mode of administration	Mechanism of action	Most common side effects	EMA date of authorization	Efficacy (Atlas of MS)
Fingolimod	Gilenya	Oral	Blocks the action of sphingosine-1-phosphate receptors on lymphocytes, stopping lymphocytes from exiting the lymph nodes	Flu-like symptoms, sinusitis, headache, cough, diarrhea, back pain, and raised liver enzyme levels	17/03/2011	Good
Siponimod	Mayzent			Headache and hypertension	13/01/2020	Good
Ozanimod	Zeposia			Nasopharyngitis and increased levels of liver enzymes	20/05/2020	NA
Ponesimod	Ponvory				19/05/2021	NA
Teriflunomide	Aubagio	Oral	Blocks dihydroorotate dehydrogenase; interferes with the production of DNA by lymphocytes	Headache, diarrhea, increased liver enzymes, nausea, and alopecia	26/08/2013	Moderate
Alemtuzumab	Lemtrada	IV infusion	Monoclonal antibody against CD52, which is found on mature lymphocytes	Rash, headache, fever, and respiratory tract infections	12/09/2013	High
Dimethyl fumarate	Tecfidera	Oral	Activates Nrf2 protein, regulating oxidative stress by increased production of antioxidants	Flushing and gastrointestinal problems (diarrhea, nausea, and abdominal pain)	30/01/2014	Good
Diroximel fumarate	Vumerity				15/11/2021	NA
Cladribine	Mavenclad	Oral	Purine analogue; interferes with the production of DNA by lymphocytes	Lymphopenia, infections with herpes zoster virus, rashes, alopecia, and neutropenia	22/08/2017	Good

Table 2 — Disease modifying drugs for Multiple Sclerosis in the European Union (continued)

Drug	Comm. name(s)	Mode of administration	Mechanism of action	Most common side effects	EMA date of authorization	Efficacy (Atlas of MS)
Ocrelizumab	Ocrevus	IV infusion	Monoclonal antibody targeting CD20 on lymphocytes, predominantly B cells	Infusion reactions (pruritus, rash, and difficulty breathing) and infections	08/01/2018	High
Ofatumumab	Kesimpta	Injectable		Upper respiratory tract infections, urinary tract infections, reactions at the site of injection, and injection-related reactions (fever, headache, myalgias, chills, and asthenia)	26/03/2021	NA

Data was compiled primarily from official EMA's data (European Medicines Agency, 2022), the 3rd edition of the Atlas of MS — part 2 (Multiple Sclerosis International Federation, 2021), and other book chapters on MS (Berkowitz, 2016a; Ropper *et al.*, 2019; Meador, 2020), with third-party additions for glatiramer acetate (INFARMED, 2020; United States Food & Drug Administration, 2022), which did not undergo centralized approval by EMA (Human Medicines Evaluation Division, European Medicines Agency, 2021). Abbreviations: CD — cluster of differentiation; comm. — commercial; EMA — European Medicines Agency; IV — intravenous; NA — not available.

1.4.4. Pathology

Macroscopically, and without sectioning the CNS of PwMS, no significant differences from healthy CNS are discernable (Ropper *et al.*, 2019). Upon sectioning the CNS, however, multiple, scattered, well delineated, and confluent lesions/plaques, ranging from less than a millimeter to a few centimeters, are discernable in both white and gray matter of the encephalon and the spinal cord (**Figure 14**), not extending beyond the root entry zones of the cranial and spinal nerves, thus excluding the PNS. These lesions correspond to regions of the CNS with a loss of myelin sheaths and oligodendrocytes (more details in *1.1.3. The cellular constituents of the nervous system*), mostly sparing neurons and axons, particularly in early lesions, and in the vicinity of postcapillary venules and veins or adjacent to the pia mater (more details in *1.1.1. The central nervous system*) (Dendrou, Fugger and Friese, 2015; Lassmann, 2018; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019).

Active lesions initially show a proximal disruption of the BBB by the infiltration of lymphocytes in the perivascular space, namely $CD4^+$ T cells, B cells, and plasma cells, with a parenchymal infiltration of $CD8^+$ T cells, particularly with a resident memory phenotype (Dendrou, Fugger and Friese, 2015; Lassmann, 2018; Ropper *et al.*, 2019; Absinta, Lassmann and Trapp, 2020). At the center of the active demyelinating lesion with oligodendrocyte loss, macrophages undergo extensive myelin phagocytosis, accompanied by astrocytic activation, relatively sparing axons and neurons, while activated microglia are at the lesion's edges (Dendrou, Fugger and Friese, 2015; Lassmann, 2018; Ropper *et al.*, 2019).

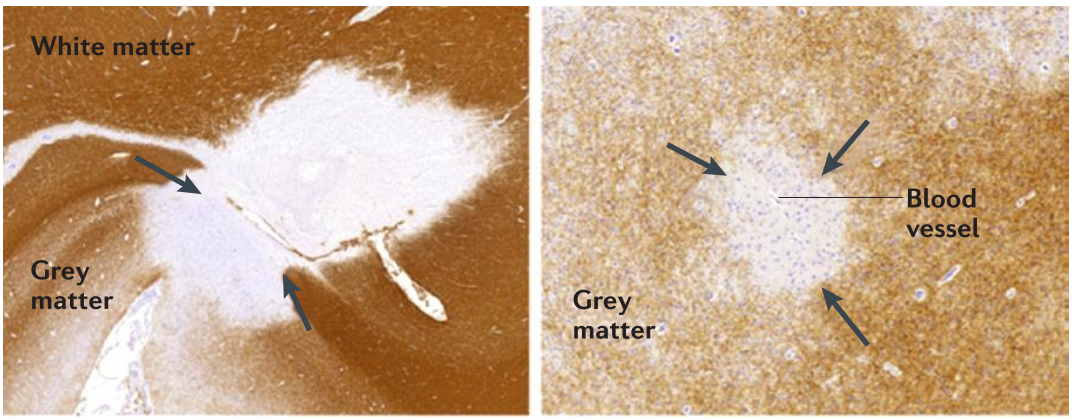


Figure 14 — Multiple Sclerosis lesions/plaques

Arrows indicate the positions of lesions. Image extracted and adapted from (Calabrese *et al.*, 2015). Copyright © 2015 Macmillan Publishers Limited.

As lesions evolve into an inactive state, there is a decrease in lymphocyte and macrophage infiltration, with the center of the lesions presenting demyelinated axons surrounded by astrocytic gliosis, while microglia and OPCs remain at the lesion's edge (Lassmann, 2018; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019). In addition to being arrested at the plaque's edge in PwMS (Reich, Lucchinetti and Calabresi, 2018), the number of OPCs is lower in PwMS in comparison with HC, further contributing to the depletion of fully mature and stable oligodendrocytes (Jäkel *et al.*, 2019). Nonetheless, partial remyelination takes place occasionally, yielding uniformly thin myelin sheaths enveloping the undamaged axons, most abundantly in the cortex, termed "shadow plaques" (Dendrou, Fugger and Friese, 2015; Lassmann, 2018; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019).

Additionally, a few lesions will be chronically active, also called “slowly expanding” or “smoldering”, where, despite their inactive lesion center (described above), the edge of the lesion will continue to expand due to activated microglia/macrophages with myelin debris and iron [iron brain homeostasis is mostly regulated by oligodendrocytes (Mukherjee *et al.*, 2020)], leading to continued demyelination and axonal injury (Lassmann, 2018; Calvi *et al.*, 2020; Absinta *et al.*, 2021). While microglia might play a beneficial role in active lesions (Yamasaki *et al.*, 2014), it appears that they are crucial for the deleterious expansion of chronically active lesions (Masuda *et al.*, 2019; Jäckle *et al.*, 2020; Absinta *et al.*, 2021).

Regarding the remaining CNS, namely, normal-appearing white and gray matter, (NAWM and NAGM, respectively), chronic and diffuse changes arise from either the peripheral inflamed meningeal compartments or from focal lesions, through trans-synaptic axonal degeneration, where lymphocytic infiltration, microglial activation, axonal injury, and astrocytic gliosis are discernable (Dziedzic *et al.*, 2010; Gabilondo *et al.*, 2014; Lassmann, 2018; Absinta *et al.*, 2021).

All types of lesions are observable in all PwMS; nonetheless, PwMS in earlier stages or with RRMS will mostly have active lesions with a variable degree of subpial cortical demyelination (Lucchinetti *et al.*, 2011; Lassmann, 2018; Absinta, Lassmann and Trapp, 2020), while PwMS in progressive stages will have fewer invading immune cells, mostly chronic active lesions, an expansion of the diffuse changes in NAWM and NAGM, and extensive subpial cortical demyelination associated with meningeal inflammation — compartmentalized inflammation (Dendrou, Fugger and Friese, 2015; Lassmann, 2018; Absinta, Lassmann and Trapp, 2020). Some PwMS with SPMS will also have follicle-like tertiary lymphoid structures in the meningeal compartment, resembling germinal centers (more details in *1.2.2. The adaptive immune system*), associated with subpial cortical lesions (Serafini *et al.*, 2004; Magliozzi *et al.*, 2007; Howell *et al.*, 2011; Lassmann, 2018).

As MS develops, particularly in progressive stages, direct axonal injury induced by $CD8^+$ T cell cytotoxicity, macrophages, and microglia, among other pro-inflammatory cytokines, will eventually lead to neuronal loss, primarily through mitochondrial dysfunction (Dendrou, Fugger and Friese, 2015; Lassmann and van Horssen, 2016; Lassmann, 2018; Correale, Marrodan and Ysrraelit, 2019). Mitochondrial dysfunction is primarily induced by increased oxidative stress, secondary to both the release of ROS and RNS produced during the oxidative burst of activated microglia and macrophages (more

details in *1.2.1. The innate immune system*), and the extracellular release of divalent iron, which, through the Fenton and Haber-Weiss reactions, gives rise to highly reactive hydroxyl radicals (Vladimirova *et al.*, 1998; Haider *et al.*, 2011; Fischer *et al.*, 2012; Lassmann and van Horssen, 2016; Andravizou *et al.*, 2019). This then leads to a decrease in peroxisome proliferator-activated receptor (PPAR)-gamma coactivator 1-alpha (PGC-1α), a master regulator of metabolism and mitochondrial function (more details in *1.4.6.2.1. Mitochondrial DNA*), including antioxidant mechanisms, while ROS and RNS directly damage mitochondrial DNA (mtDNA), culminating in reduced mitochondrial function, particularly oxidative phosphorylation (Mahad *et al.*, 2009; Campbell *et al.*, 2011, 2012; Witte *et al.*, 2013; Lassmann and van Horssen, 2016; Campbell and Mahad, 2018; Campbell, Licht-Mayer and Mahad, 2019).

Due to demyelination, which increases axonal energy demands, this reduction in ATP production will lead to the failure of the Na^+/K^+ pump and a redistribution of sodium channels throughout the demyelinated axons rather than in the nodes of Ranvier; as intra-axonal Na^+ levels rise, Na^+ ions will be replaced by Ca^{2+} through the Na^+/Ca^{2+} exchanger, which further exacerbates axonal degeneration, through the activation of proteases, phospholipases, and calpains (Craner *et al.*, 2004; Young *et al.*, 2008; Bouafia *et al.*, 2014; Lassmann and van Horssen, 2016; Campbell and Mahad, 2018; Andravizou *et al.*, 2019; Campbell, Licht-Mayer and Mahad, 2019; Correale, Marrodan and Ysrraelit, 2019).

This will further culminate in neuroaxonal loss, leading to brain and retinal atrophy, which might even be detectable in patients with early MS and CIS (Chard *et al.*, 2002; Oberwahrenbrock *et al.*, 2013; Andravizou *et al.*, 2019; Lin *et al.*, 2021). The importance of mitochondrial dysfunction in disease progression has also been demonstrated in various animal models of MS (Ziabreva *et al.*, 2010; Sadeghian *et al.*, 2016; Smorodchenko *et al.*, 2017; Buonvicino *et al.*, 2019; Ng *et al.*, 2019; Varhaug *et al.*, 2020; Martínez-Pinilla *et al.*, 2021). Additionally, in accordance with clinical treatment of PwMS with dimethyl or diroximel fumarate (**Table 2**), which increase the transcription factor Nrf2, involved in combating oxidative stress (Mills *et al.*, 2018; European Medicines Agency, 2022), experimental attempts to rescue the oxidative stress and/or mitochondrial dysfunction in a diverse range of animal models ameliorate disease activity (Mao *et al.*, 2013; Wang *et al.*, 2016; Warne *et al.*, 2016; Liessem-Schmitz *et al.*, 2018; Bando and Geisler, 2019; Fetisova *et al.*, 2019; Largani *et al.*, 2019; Martín-Montañez *et al.*, 2019).

1.4.5. Pathophysiology

Our current understanding of MS's pathogenesis comes mainly from animal models, particularly the almost 80-year-old experimental autoimmune encephalitis (EAE) model (Rivers, Sprunt and Berry, 1933; Rivers and Schwentker, 1935; Kabat, Wolf and Bezer, 1947), where animals are immunized, actively or passively, with immune adjuvants and emulsified CNS antigens, particularly myelin components, such as MBP and PLP (more details in *1.1.3. The cellular constituents of the nervous system*) (Constantinescu *et al.*, 2011; Dendrou, Fugger and Friese, 2015; Ropper *et al.*, 2019).

In this model, dormant autoreactive myelin-specific $CD4^+$ T cells are activated either in the periphery, in secondary lymphoid organs, or by CNS antigens in circulation following damage (Miller *et al.*, 1997; Dendrou, Fugger and Friese, 2015; Korn and Kallies, 2017; Prinz and Priller, 2017). Activated $CD4^+$ T cells will then differentiate into pro-inflammatory Th1 and Th17 subsets (more details in *1.2.2. The adaptive immune system*), which are able to disrupt and cross the BBB, leading to the invasion of CNS by peripheral immune cells (more details in *1.3. Neuroimmunology*) (Bartholomäus *et al.*, 2009; Dendrou, Fugger and Friese, 2015; Schläger *et al.*, 2016; Korn and Kallies, 2017; Prinz and Priller, 2017; Kaskow and Baecher-Allan, 2018). The interaction between the peripheral immune system and the resident microglia and astrocytes will eventually give rise to neuroaxonal loss, mostly mediated by mitochondrial dysfunction and oxidative stress (more details in *1.4.4. Pathology*) (Dendrou, Fugger and Friese, 2015; Lassmann and van Horssen, 2016; Campbell and Mahad, 2018; Campbell, Licht-Mayer and Mahad, 2019; Correale, Marrodan and Ysrraelit, 2019).

In the EAE model, this autoimmune response is further exacerbated by an ineffective suppressive activity from Treg cells (Kohm *et al.*, 2002; McGeachy, Stephens and Anderton, 2005; Lowther *et al.*, 2013). Accordingly, Treg cells from PwMS have shown: (**I**) a decrease in circulation or in the CNS (Venken, Hellings, Broekmans, *et al.*, 2008; Venken, Hellings, Thewissen, *et al.*, 2008; Fritzsching *et al.*, 2011; Rodi *et al.*, 2016); (**II**) a defective suppressive activity (Viglietta *et al.*, 2004; Haas *et al.*, 2005; Huan *et al.*, 2005; Feger *et al.*, 2007; Venken, Hellings, Broekmans, *et al.*, 2008; Venken, Hellings, Thewissen, *et al.*, 2008; Fletcher *et al.*, 2009; Tapia-Maltos *et al.*, 2021); (**III**) a resistance from effector T cells mediated by IL-6 (Schneider Anya *et al.*, 2013; Bhela *et al.*, 2015); and (**IV**) a complete

distortion of the usual Treg profile into a pro-inflammatory Th1-like phenotype, mediated by IL-12 (Dominguez-Villar, Baecher-Allan and Hafler, 2011; Kitz *et al.*, 2016).

Contrary to EAE, however, the view on MS pathogenesis has slowly been shifting away from a sole T-cell mediated disease into a multicellular pathophysiology, where B cells and monocytes are increasingly viewed as key contributors (Dendrou, Fugger and Friese, 2015; Severa *et al.*, 2019; Bar-Or and Li, 2021). Interestingly, various researchers have indeed suggested that MS might be mainly driven by B cells, propelled by the relevance of oligoclonal bands, the presence of tertiary lymphoid structures in PwMS with SPMS, and the efficacy of anti-CD20 DMDs (**Table 2**) (Wekerle, 2017; Baker *et al.*, 2018, 2020; Li, Patterson and Bar-Or, 2018; Sabatino, Pröbstel and Zamvil, 2019; Comi *et al.*, 2021; Holloman *et al.*, 2021).

In fact, B cells are naturally found in meningeal compartments (Korin *et al.*, 2017; Schafflick *et al.*, 2021), and their functions extend beyond antigen-presentation and antibody synthesis (more details in *1.2.2. The adaptive immune system*) into the production of both pro-inflammatory and anti-inflammatory cytokines that may influence the innate and the adaptive immune systems (Wekerle, 2017; Baker *et al.*, 2018; Li, Patterson and Bar-Or, 2018; Sabatino, Pröbstel and Zamvil, 2019; Holloman *et al.*, 2021). Furthermore, recent clinical studies have found that intrathecal IgM synthesis is an independent biomarker of disease activity and severity in RRMS (Oechtering *et al.*, 2021) and that the presence of $CD20^+$ cells in brain slices of PwMS, interpreted as an indicator of B cells, correlated with worse clinical outcomes and CNS-compartmentalized inflammation (Moccia *et al.*, 2022).

Nonetheless, oligoclonal bands are not directed towards CNS autoantigens in particular (Brändle *et al.*, 2016), antibodies directed towards myelin components, such as myelin oligodendrocyte glycoprotein (MOG), are quite rare amongst PwMS (Jarius *et al.*, 2018; Spadaro *et al.*, 2018), and therapies specifically targeting antibody production in PwMS have either had no effect on (Silk and Nantz, 2018) or exacerbated disease activity (Kappos *et al.*, 2014). Moreover, in experimental models, in contrast to autoreactive T cells, which are able to resume CNS autoimmunity on their own (Ben-Nun, Wekerle and Cohen, 1981), autoreactive B cells/antibodies require the collaboration of $CD4^+$ T cells to trigger disease (Pierson, Stromnes and Goverman, 2014; Parker Harp *et al.*, 2018; Spadaro *et al.*, 2018). Finally, growing evidence has shown that CD20 is also expressed in T cells, which show pro-inflammatory properties; these $CD20^+$ T cells are increased in PwMS and directly

linked to disease progression, while also appearing to play a role in MS white matter lesions and in the therapeutic response to anti-CD20 drugs (Palanichamy *et al.*, 2014; Schuh *et al.*, 2016; Gingele *et al.*, 2019; Sabatino *et al.*, 2019; von Essen *et al.*, 2019; Boldrini *et al.*, 2021; Fernández-Velasco *et al.*, 2021; Hsiao *et al.*, 2021; Meinl and Hohlfeld, 2021; Quendt *et al.*, 2021). Most likely, B cells, particularly memory B cells, are relevant modulators of $CD4^+$ T cells through cytokine production and antigen presentation (Jelcic *et al.*, 2018; Li, Patterson and Bar-Or, 2018; Sabatino, Pröbstel and Zamvil, 2019; Baker *et al.*, 2020).

Thus, currently, the most commonly accepted hypothesis for the autoimmune response seen in MS is based on a dysfunctional autoreactive $CD4^+$ T cell compartment (Ropper *et al.*, 2019; Gerdes *et al.*, 2020; Kunkl *et al.*, 2020; Arneth, 2021; Cruciani *et al.*, 2021), whose dysfunction is further substantiated by unbiased high-dimensional single-cell mass cytometry in samples from PwMS (Böttcher *et al.*, 2019; Galli *et al.*, 2019), that might be exacerbated by both (**I**) a lack/disruption of suppressive functions by Treg cells (Viglietta *et al.*, 2004; Haas *et al.*, 2005; Huan *et al.*, 2005; Feger *et al.*, 2007; Venken, Hellings, Broekmans, *et al.*, 2008; Venken, Hellings, Thewissen, *et al.*, 2008; Fletcher *et al.*, 2009; Dominguez-Villar, Baecher-Allan and Hafler, 2011; Fritzsching *et al.*, 2011; Schneider Anya *et al.*, 2013; Bhela *et al.*, 2015; Kitz *et al.*, 2016; Rodi *et al.*, 2016; Tapia-Maltos *et al.*, 2021) and (**II**) pro-inflammatory functions of B cells and monocyte-derived cells (Jelcic *et al.*, 2018; Li, Patterson and Bar-Or, 2018; Sabatino, Pröbstel and Zamvil, 2019; Baker *et al.*, 2020).

1.4.6. Etiology

The exact etiology behind the autoimmune response in MS remains unknown, although it appears to be multifactorial, with both environmental risk factors and genetic predisposition playing a role in disease triggering, most likely several years before clinical onset (Dendrou, Fugger and Friese, 2015; Berkowitz, 2016a; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019; Meador, 2020).

1.4.6.1. Environmental risk factors

The most reliable environmental risk factors associated with an increased risk of MS, according to a recent umbrella review of meta-analyses and Mendelian randomization studies (Belbasis *et al.*, 2020), where genetic variants are used as instrumental variables for modifiable risk factors (Davies, Holmes and Davey Smith, 2018), are: (**I**) history of infectious

mononucleosis and/or evidence of previous Epstein–Barr virus (EBV) infection, documented through anti-Epstein–Barr virus nuclear antigen (EBNA) IgG seropositivity or anti-viral capsid antigen (VCA) IgG seropositivity; (**II**) cigarette smoking; (**III**) high body mass index (BMI), particularly during childhood or adolescence; and (**IV**) low serum vitamin D levels.

EBV is a double-stranded DNA virus and a member of the Herpesviridae family that infects 90–95% of all human beings, mostly asymptomatically, showing a particular predilection to B cells, usually persisting in the pool of memory B cells (Young, Yap and Murray, 2016; Ryan, 2022). Primary infection is due to direct exposure to contaminated saliva, with the virus infecting host epithelial cells, through integrins, and B cells, through CD21, which is also the receptor for the C3d component of the complement system, with MHC class II as a co-receptor (more details in *1.2.1. The innate immune system*). As B cells become infected *in vivo*, they undergo polyclonal activation with benign proliferation, with only a small fraction producing EBV. Interestingly, *in vitro*, they become immortalized and proliferate indefinitely. In parallel, there is also a clonal expansion of $CD8^+$ T cells, leading to a relative reduction of $CD4^+$ T cells (Cohen, 2018; Ryan, 2022).

Primary infections during childhood are usually asymptomatic, while 50% of infected adolescents or young adults will have a symptomatic infection termed infectious mononucleosis, characterized by fever, malaise, sore throat, lymphadenopathy, atypical lymphocytosis, and splenomegaly. The vast majority of infectious mononucleosis cases (90%) are attributable to EBV and are self-limited, with EBV infection subsiding due to cellular rather than humoral immunity, particularly derived from $CD8^+$ T cells and NK cells. In children, EBV infection might lead to several CNS diseases, namely, meningitis, encephalitis, Guillain-Barré syndrome, acute transverse myelitis, and peripheral neuritis. Chronically, particularly in immunocompromised patients, EBV might also lead to several neoplastic diseases, namely, gastric carcinoma (9% of tumors), Burkitt's lymphoma, particularly in Africa, several other lymphomas, and anaplastic nasopharyngeal carcinoma, particularly in southern China. As a marker of disease, IgG against VCA or against EBNA persist for life (Cohen, 2018; Ryan, 2022).

In MS, reported infectious mononucleosis and humoral evidence of past infection are consistently linked to increased risk of disease (Handel *et al.*, 2010; Almohmeed *et al.*, 2013; Lossius *et al.*, 2014; Bjornevik *et al.*, 2022), although data on disease progression is

conflicting (Gieß *et al.*, 2017; Jakimovski *et al.*, 2020). Transcriptomic data has demonstrated a link between latent EBV infection in brain samples of patients with progressive MS (Veroni *et al.*, 2018) and, peripherally, EBV reactivation of B cells is linked to relapses in female PwMS (Irizar *et al.*, 2014). Interestingly, a recent study has shown that the EBV transcription factor EBNA1 shows high-affinity molecular mimicry to the CNS protein GlialCAM, with PwMS showing increased IgG reactivity against EBNA1 and GlialCAM in comparison with HC, while EBNA1 immunization exacerbated disease activity in EAE mice (Lanz *et al.*, 2022). This is further consistent with an additional observational study, which demonstrated that hospital-diagnosed infections during adolescence (but not during childhood), particularly infectious mononucleosis and CNS infectious, are strongly linked to MS risk (Xu, Smith, *et al.*, 2021), suggesting that immune dysregulation during this crucial stage of development might trigger CNS autoimmune disease.

Cigarette smoking releases more than 7000 compounds that provide an irritant and pro-inflammatory trigger that leads to direct cellular damage, release of DAMPs, and activation of the innate immune system. This activation of the innate immune system mediates an oxidative burst from phagocytes, which, together with the direct exposure to free radicals from cigarette smoking, might lead to the presentation of unusual or altered self-antigens or the antigen presentation of external smoke components to T cells. In turn, this might lead to an autoimmune response to CNS antigens through cross-reactivity (Rosso and Chitnis, 2020; Ringh *et al.*, 2021). Smoking shows a dose-response relationship with MS risk (Poorolajal *et al.*, 2017) and has been associated with an increased relapse activity in PwMS (Petersen *et al.*, 2018), as well as worse long-term cognitive function and neuronal integrity (Cortese *et al.*, 2020). Recently, however, a large Mendelian randomization study has failed to corroborate previous observational studies (Mitchell *et al.*, 2020).

Obesity, which corresponds to the accumulation of abnormal or excessive fat, is directly responsible for chronic inflammation due to a dysfunctional adipose tissue (Ellulu *et al.*, 2017). This will lead to an increased release of pro-inflammatory cytokines, higher Th1 responses, and a decrease in the number of Treg cells (Hedström, Olsson and Alfredsson, 2015). Obesity during childhood and adolescence has been linked to a dose-response risk of MS (Wesnes *et al.*, 2015; Xu, Hiyoshi, *et al.*, 2021) and, in female PwMS, a younger age at symptom onset (Kavak *et al.*, 2015).

Vitamin D is mostly generated by the skin upon exposure to ultraviolet radiation from sunlight, following metabolization in the liver and kidney into the active form of 1-α,25-dihydroxyvitamin D3, also known as calcitriol. It is viewed as the main driver behind the latitude effect seen in MS (more details in *1.4.2. Epidemiology*) and it exerts immunomodulatory functions through type I IFN-family cytokines, particularly in monocytes, which appear to be protective of MS (Koduah, Paul and Dörr, 2017; Ostkamp *et al.*, 2021). Consequently, low vitamin D levels and low sun exposure have been associated with both an increased risk of MS (Bjørnevik *et al.*, 2014; Duan *et al.*, 2014; Cortese *et al.*, 2015) and higher disability in PwMS (Cortese *et al.*, 2020; Ostkamp *et al.*, 2021). Surprisingly, this protective effect is not the same for all individuals (Rito *et al.*, 2016) and might be related to a specific polymorphism in the melanocortin 1 receptor (MC1R), which is responsible for melanin synthesis upon ultraviolet radiation, associated with both higher sensitivity to sun exposure and higher disease disability (Ostkamp *et al.*, 2021). Surprisingly, a recent study has also noted a synergistic effect between past exposure to EBV and low sun exposure (Hedström *et al.*, 2021).

Finally, all aforementioned well-established environmental risk factors for MS are, remarkably, additional regulators of mitochondrial function (Fernández-Sánchez *et al.*, 2011; Anand and Tikoo, 2013; Consiglio *et al.*, 2014; de Mello *et al.*, 2018; Ricca *et al.*, 2018; Dikalov *et al.*, 2019; Malińska *et al.*, 2019; Wang *et al.*, 2019; Ashcroft *et al.*, 2020), which has proved to be central in driving disease progression (more details in *1.4.4. Pathology*). Nonetheless, the search for other risk or protective determinants, such as the microbiota (Berer *et al.*, 2011), breastfeeding (Ragnedda *et al.*, 2015), physical activity (Wesnes *et al.*, 2018), and level of education (Bjørnevik *et al.*, 2016), is ongoing.

1.4.6.2. Genetic risk factors

A genetic basis for MS has been suspected for several decades (Miller, 1968; Ebers *et al.*, 1986, 2000; Ebers, Sadovnick and Risch, 1995; Robertson *et al.*, 1996; Willer *et al.*, 2003; Dyment *et al.*, 2006; Patsopoulos, 2018), mainly driven by an increased familial recurrence risk of MS, particularly amongst monozygotic twins who share the same nuclear DNA (nDNA), with heritability estimates in large-scale studies or meta-analyses ranging from 0.48–0.64, contrasting with the estimated environmental impact of 0.01–0.29 (Ristori *et al.*, 2006; O'Gorman *et al.*, 2013; Westerlind *et al.*, 2014; Fagnani *et al.*, 2015).

Nonetheless, the exact genetic determinants of MS only started to be unveiled with the advent of DNA sequencing. The sequencing of the ~3 billion base pairs (bps) that convey the set of genetic instructions behind the existence of our species (Alberts *et al.*, 2015; Jameson and Kopp, 2018) was finally accomplished in 2004 (International Human Genome Sequencing Consortium, 2004), having already been subjected to several revisions until its latest version (Schneider *et al.*, 2017). DNA sequencing has undergone a complete revolution in the past decades, moving away from genotyping through Southern blots, associated with restriction enzymes, and quantitative polymerase chain reaction (qPCR), to more high-throughput methods, such as genome wide analysis studies (GWAS) through DNA microarrays, Sanger sequencing, and, most recently, massively parallel sequencing (MPS) or next-generation sequencing (NGS) (Cortes-Figueiredo, Morais and Pinheiro, 2021).

NGS differs from the previous technologies by bioinformatically aligning numerous sequences to a consensus sequence. Most commonly, it starts with DNA fragmentation and adapter ligation, which constitute the library preparation, moving on to parallel sequencing of each DNA fragment, identified by its exclusive adapter, after cluster amplification (Cortes-Figueiredo, Morais and Pinheiro, 2021). The most common approaches are (Quail *et al.*, 2012; Alberts *et al.*, 2015; Adams and Eng, 2018):

- **Illumina sequencing** — based on sequencing-by-synthesis on a microarray. It uses nucleotides labelled with different fluorochromes that block DNA elongation. The tags, however, may be removed enzymatically, allowing for the addition of a further tagged nucleotide. Thus, the DNA sequence is determined by consecutively reading the different fluorescent colors;
- **Ion torrent sequencing** — based on semiconductor sequencing on DNA-covered beads dispersed throughout multiple wells. Each cluster covers a single bead, and the DNA sequence is determined by detecting changes in pH, through a voltage-sensitive semiconductor chip. A different nucleotide is added to the well sequentially until a pH change is registered, corresponding to the release of a proton as the correct nucleotide is incorporated into the DNA molecule. Thus, the DNA sequence is determined by sequentially registering successful nucleotides in changing the pH in the well.

NGS has given rise to multiple areas of research, such as transcriptomics (RNA-seq) and epigenomics (methylation sequencing), additionally driven by a consistent decrease in its costs, particularly since 2008 (Metzker, 2010; Lowe *et al.*, 2017; Adams and Eng, 2018; Barros-Silva *et al.*, 2018; Gulley, 2020; Wetterstrand, 2021). Currently, the focus has also gradually shifted from bulk DNA or RNA sequencing, to single-cell genomics (Lodato *et al.*, 2018; Minussi *et al.*, 2021), transcriptomics (Enge *et al.*, 2017; Han *et al.*, 2020), or even both at once (Bian *et al.*, 2018; Ono *et al.*, 2021).

Moreover, both GWAS and NGS research areas have originated collective efforts to share genetic data: (**I**) the International HapMap Project collects, organizes, and shares haplotypes, which are a collection of single nucleotide polymorphisms (SNPs) usually inherited together (International HapMap Consortium, 2003), while the 1000 Genomes Project aims to collect and share a diverse set of whole genomes (1000 Genomes Project Consortium *et al.*, 2015).

Specific genetic risk factors for MS were first discovered 50 years ago (Jersild, Svejgaard and Fog, 1972; NAITO *et al.*, 1972), namely, in several HLA loci (more details in *1.2.1. The innate immune system*) (Gourraud *et al.*, 2012), while non-HLA SNPs were only discovered in the first GWAS by the International Multiple Sclerosis Genetics Consortium (IMSGC), in 2007 (International Multiple Sclerosis Genetics Consortium *et al.*, 2007). The latest IMSGC study has identified 233 genome-wide variants significant to MS susceptibility: 32 HLA-related, 200 autosomal non-HLA, and one in chromosome X, based on genetic data from 47,429 PwMS and 68,374 HC, and explaining up to 48% of disease heritability (International Multiple Sclerosis Genetics Consortium *et al.*, 2019).

The identified loci implicate cells and tissues from both the innate and adaptive immune systems, as well as microglia within the CNS, as relevant contributors to MS genetic risk, which further indicates the relevance of the immune system in triggering the disease (International Multiple Sclerosis Genetics Consortium *et al.*, 2019). This was in line with previous IMSGC studies (International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2, 2011; International Multiple Sclerosis Genetics Consortium, 2013b, 2013a, 2018) and further validated by pathway analysis and protein interaction networks (International Multiple Sclerosis Genetics Consortium, 2019), along with a novel deep learning analysis focused on regulatory SNPs (Arloth *et al.*, 2020). A recent study has also validated these findings for Hispanic and African American

populations in the United States for 76% and 68% of non-HLA variants, respectively (Beecham *et al.*, 2020). In contrast to the clear genetic risk of MS development, however, research on disease progression has yielded conflicting results (George *et al.*, 2016; Isobe *et al.*, 2016; Pan *et al.*, 2016).

*The HLA-DRB1*15:01* allele is the most well-established genotype associated with an increased risk of MS (Gourraud *et al.*, 2012; Patsopoulos *et al.*, 2013; International Multiple Sclerosis Genetics Consortium, 2015), located in a HLA class II locus, which is directly involved in antigen presentation to $CD4^+$ T cells (more details in *1.2.1. The innate immune system*). It is carried by 28–33% of Northern Caucasian PwMS, contrasting with 9–15% of HC (Gourraud *et al.*, 2012), and its presence has been linked to multiple known MS disease mechanisms: (**I**) a synergistic effect with infectious mononucleosis in driving MS risk (Jacobs *et al.*, 2020); (**II**) a higher probability of having CSF oligoclonal bands, as well as higher levels of CSF IgG (Goris *et al.*, 2015); (**III**) a decrease in peripheral naïve $CD8^+$ T cells in children at age 6, thereby increasing their $CD4^+/CD8^+$ ratio (Mol *et al.*, 2021); and (**IV**), experimentally, increased antigen-presentation and autoreactivity regarding self-peptides (Wang *et al.*, 2020), with memory B cells serving as important mediators in antigen presentation to $CD4^+$ T cells, which will subsequently differentiate into CNS-homing Th1 cells (Jelcic *et al.*, 2018).

Interestingly, while HLA class II loci have been associated with an increased risk in MS development, HLA class I loci, which encode for MHC proteins involved in antigen presentation to $CD8^+$ T cells, have a protective effect (Patsopoulos *et al.*, 2013; International Multiple Sclerosis Genetics Consortium, 2015).

1.4.6.2.1. Mitochondrial DNA

Mitochondria (**Figure 15**) are thought to derive from a symbiotic relationship between a primitive anaerobic predator cell and an aerobic bacterial cell 1.5 billion years ago, originating eukaryotic cells (Alberts *et al.*, 2015), as popularized by Lynn Margulis in 1967 (Sagan, 1967; Gray, 2017). With time, the previously independent aerobic organisms transferred a majority of their genome to the nDNA, originating nuclear insertions of mtDNA (NUMTs) (Lopez *et al.*, 1994; Alberts *et al.*, 2015). Incidentally, NUMTs have been linked to disease due to a disruption in nDNA (Li *et al.*, 2021).

Thus, the assembly of the electron transport chain (ETC), which is responsible for the oxidative phosphorylation (OXPHOS) behind ATP production involves a delicate and

complex orchestra of several nucleus–mitochondria molecular interactions. The ETC (**Figure 15**) is comprised of four multimeric protein complexes embedded in the inner mitochondrial membrane (IMM): (**I**) complex I or NADH-ubiquinone oxidoreductase; (**II**) complex II or succinate dehydrogenase; (**III**) complex III or cytochrome bc_1 oxidoreductase; and (**IV**) complex IV or cytochrome *c* oxidase. These protein complexes are further joined by two mobile electron carriers, namely, the membrane-embedded ubiquinone and the soluble cytochrome *c*. Apart from complex II, all complexes are partially encoded by mtDNA. Intriguingly, recent studies have uncovered the assembly of complexes I, III, and IV into supercomplexes, whose functional relevance is still being unveiled (Signes and Fernandez-Vizarra, 2018; Vercellino and Sazanov, 2021).

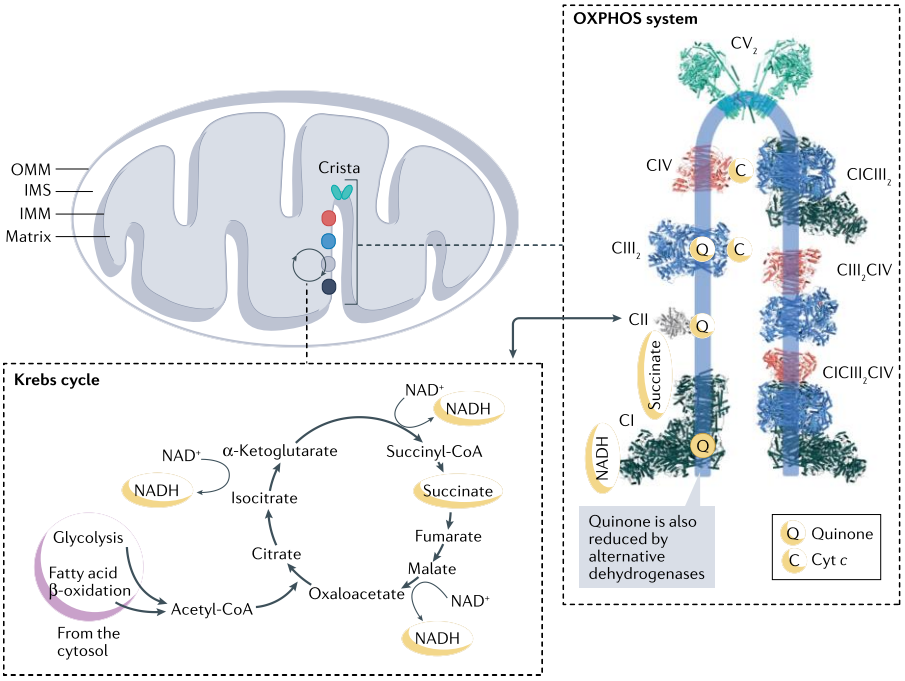


Figure 15 — Mitochondria and the oxidative phosphorylation chain

Image extracted and adapted from (Vercellino and Sazanov, 2021). Copyright © Springer Nature Limited 2021.

Abbreviations: CI — complex I; CII — complex II; CIII — complex III; CIV — complex IV; CV — complex V; Cyt c — cytochrome c; IMM — inner mitochondrial membrane; IMS — intermembrane space; OMM — outer mitochondrial membrane; OXPHOS — oxidative phosphorylation.

The goal of the ETC is to sequentially transfer electrons from NADH and succinate, mainly arising from the Krebs cycle (**Figure 15**), with the contribution of ubiquinone and cytochrome *c*, as well as cofactors, such as iron–sulfur clusters and heme groups (Signes and Fernandez-Vizarra, 2018; Vercellino and Sazanov, 2021). This will allow complexes I, III,

and IV to pump protons into the intermembrane space (IMS), generating an electrochemical proton gradient across the IMM and, thus, a mitochondrial membrane potential (ΔΨm) (Signes and Fernandez-Vizarra, 2018; Zorova *et al.*, 2018; Vercellino and Sazanov, 2021). Finally, the electrons will reduce molecular oxygen, which, together with protons, produces water, while the proton gradient is used by complex V, or ATP synthase, to phosphorylate ADP to ATP (Signes and Fernandez-Vizarra, 2018; Vercellino and Sazanov, 2021).

OXPHOS is not 100% efficient and it is, thus, associated with electron leak, which will lead to the production of harmful ROS, namely, superoxide and hydrogen peroxide, through the interaction of electrons with oxygen (Jastroch *et al.*, 2010; Zhao *et al.*, 2019). Furthermore, ROS might interact with nitric oxide (NO), giving rise to RNS (Valerio and Nisoli, 2015), which might also mediate cellular damage (more details in *1.2.1. The innate immune system* and *1.4.4. Pathology*). Contrariwise, increased ROS due to electron leak is counterbalanced by proton leak, which increases exponentially with increasing ΔΨm, where protons bypass complex V into the mitochondrial matrix in a process most importantly, albeit not exclusively, mediated by uncoupling proteins (UCPs) (Divakaruni and Brand, 2011; Zhao *et al.*, 2019). Finally, increased oxidative stress is further counterbalanced by several enzymes and small non-enzymatic molecules that constitute the cell's antioxidant defense system (Fransen *et al.*, 2012; Snezhkina *et al.*, 2019).

Nonetheless, the role of mitochondria extends much more beyond ATP production and oxidative stress. Mitochondria are highly dynamic organelles, whose functioning correlates to the mitochondrial network that is continuously established through either fission, mediated by the proteins mitochondrial fission 1 (Fis1), mitochondrial fission factor (Mff), and dynamin-related protein 1 (Drp1), or fusion, mediated by proteins mitofusins 1 and 2 (Mfn1 and Mfn2), as well as optic atrophy 1 (OPA1) (Leites and Morais, 2018; Tilokani *et al.*, 2018). Recently, it has been demonstrated that quantitative 3D mapping of this mitochondrial network is significantly altered upon pathogenic mtDNA mutations (Vincent *et al.*, 2019).

Mitochondria are also able to eliminate defective mitochondria through mitophagy, which is particularly, although not exclusively, mediated by an interaction between the proteins PTEN induced putative kinase 1 (PINK1) and the ubiquitin E3 ligase Parkin, which, through ubiquitination of the outer mitochondrial membrane (OMM), are able to signal

mitochondria for elimination through the autophagy machinery (Leites and Morais, 2018; Yao *et al.*, 2021). While mutations in the *PINK1* gene and PINK1 dysfunction have been consistently linked to Parkinson's disease (Morais and De Strooper, 2010; Gonçalves and Morais, 2021; Leites and Morais, 2021), PwMS show increased levels of PINK1 and Parkin in their CSF and serum (Patergnani *et al.*, 2018; Castellazzi *et al.*, 2019; Hassanpour *et al.*, 2020; Joodi khanghah *et al.*, 2020; Cossu *et al.*, 2021).

Another putative alternative to deal with defective mitochondria has been reported recently, consisting of intercellular exchange of mitochondria in extracellular vesicles (Hayakawa *et al.*, 2016; Hough *et al.*, 2018). A recent study in the EAE model found that providing mice with mitochondria-containing extracellular vesicles induced significant amelioration of clinical deficits (Peruzzotti-Jametti *et al.*, 2021).

Lastly, mitochondria are crucial for intracellular Ca^{2+} homeostasis, which is intimately linked with cell death, either from necrosis or apoptosis (Rizzuto *et al.*, 2012). Also, mitochondria themselves further mediate the intrinsic apoptosis pathway, by releasing cytochrome *c* to the cytosol, depending on regulation from the Bcl2 family of proteins (Alberts *et al.*, 2015).

Regarding mtDNA, it was first reported in 1963, by two researchers from Sweden (M. M. NASS and NASS, 1963; S. NASS and NASS, 1963); 18 years later, the human mtDNA sequence was published, using Sanger sequencing (Anderson *et al.*, 1981). Human mtDNA (**Figure 16**) consists of double-stranded circular DNA with 16,569 bps, encoding for 13 proteins of the ETC (**Figure 15**), two ribosomal RNA (rRNA) components, and 22 transfer RNAs (tRNAs) (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015; Silva-Pinheiro and Minczuk, 2021). Additionally, various other mitochondrial-derived peptides (MDPs), which are polypeptides encoded in distinct open reading frames within the mitochondrial genome, have also been described, whose biological relevance is still being determined (Hashimoto *et al.*, 2001; Lee, Yen and Cohen, 2013; Lee *et al.*, 2015; Cobb *et al.*, 2016; Okada *et al.*, 2017).

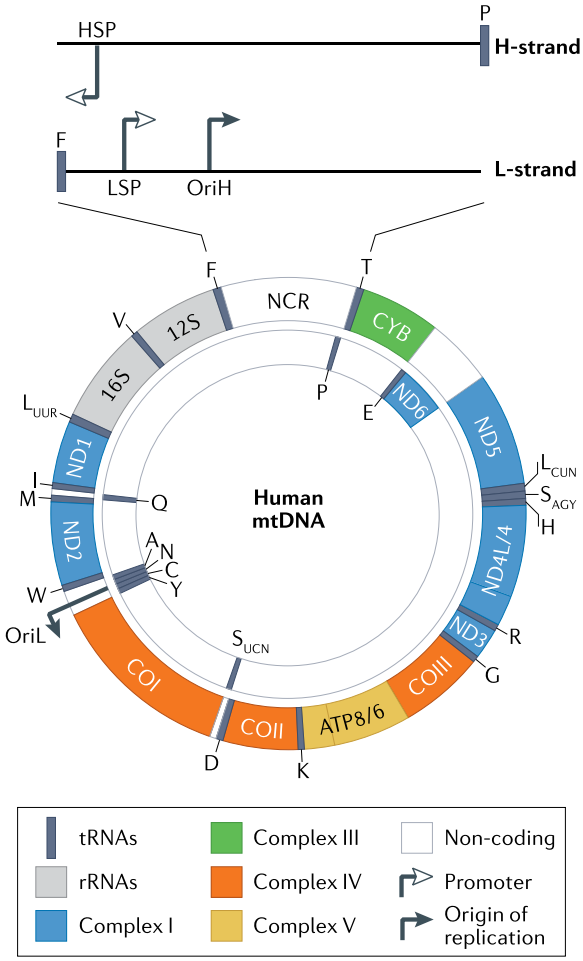


Figure 16 — Mitochondrial DNA

Unedited image extracted from (Silva-Pinheiro and Minczuk, 2021). Copyright © Springer Nature Limited 2021.

Abbreviations: mtDNA — mitochondrial DNA; rRNAs — ribosomal RNAs; tRNAs — transfer RNAs.

Although the full molecular pathway is still being studied, replication of mtDNA is performed independently of nuclear replication by several nuclear-encoded proteins: (**I**) DNA polymerase γ (POL_γ), for nucleotide synthesis; (**II**) RNA polymerase POLRMT, for primer synthesis; (**III**) helicase TWINKLE, for unwinding the double-stranded DNA; and (**IV**) mitochondrial single-stranded binding protein (mtSSB), for protection of the nascent strand against the formation of secondary structures or undesired primer synthesis. Currently, the most widely accepted model for mtDNA replication is the strand-displacement model (**Figure 17**), where DNA synthesis is asynchronous for the strands being replicated, named according to their buoyancy in a cesium chloride density gradient (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015; Falkenberg, 2018; Sanchez-Contreras *et al.*, 2021):

- **Heavy (H) strand** — rich in guanine, begins its replication at OriH after primer synthesis by POLRMT (**Figure 16**). The OriH is located in the largest non-coding region (NCR) in mtDNA, which is also called the D-loop region, due to the displacement loop that is formed between the nascent H strand and the existing H strand. POL_{γ} will continue the synthesis of the new H strand aided by mtSSB and TWINKLE until it reaches OriL;
- **Light (L) strand** — rich in cytosine, begins its replication at OriL (**Figure 16**), after primer synthesis by POLRMT, encouraged by a stem-loop structure at OriL (**Figure 17**). POL_{γ} will, then, initiate synthesis of the nascent L strand as simultaneous H strand synthesis carries on.

Ultimately, two full-length, double-stranded mtDNA molecules are formed (**Figure 17**) (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015; Falkenberg, 2018; Sanchez-Contreras *et al.*, 2021).

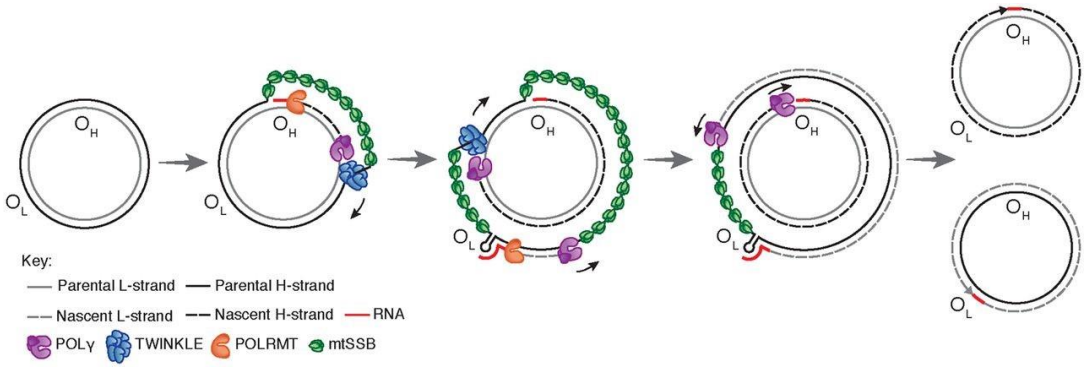


Figure 17 — Mitochondrial DNA replication: Strand-displacement model

Unedited image from (Falkenberg, 2018) (CC BY-NC 4.0).

Transcription of mtDNA begins with the aforementioned POLRMT being recruited by nuclear-encoded mitochondrial transcription factor A (TFAM) and mitochondrial transcription factor B2 (TFB2M) at the two promoter sites: HSP for the H strand and LSP for the L strand (**Figure 16**) (Taylor and Turnbull, 2005; Asin-Cayuela and Gustafsson, 2007; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015; D'Souza and Minczuk, 2018). Previously, two promoter sites were suggested for the H strand (Taylor and Turnbull, 2005; Asin-

Cayuela and Gustafsson, 2007; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015), but recent evidence has shown that a single promoter site is present (D'Souza and Minczuk, 2018).

After transcription is initiated, POLRMT will require the aid of the nuclear-encoded mitochondrial transcription elongation factor (TEFM) for the elongation stage, producing long polycystronic transcripts, which will later be processed into functional messenger RNAs (mRNAs), rRNAs, and tRNAs, by various nuclear-encoded proteins. Afterwards, mitochondrial translation will require further nuclear-encoded proteins, both for the assembly of mitoribosomal proteins, which will incorporate the two mitochondrial rRNAs, and for the synthesis of the mitochondrial polypeptides, where a deviation of the standard genetic code is observed (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; D'Souza and Minczuk, 2018; Kummer and Ban, 2021).

Interestingly, transcription from LSP will yield primers for H strand replication, linking transcription and replication (Asin-Cayuela and Gustafsson, 2007; Stewart and Chinnery, 2015). This relationship is further supported by the organization of mtDNA itself, since mtDNA is packaged into nucleoids in the mitochondrial matrix by its replication and transcription machinery, and in the proximity of mitoribosomes (Gilkerson *et al.*, 2013; Falkenberg, 2018). The interaction between TFAM and the Lon protease, which are both core nucleoid proteins, is particularly relevant. On the one hand, Lon will degrade unbound or phosphorylated TFAM, thereby decreasing mtDNA transcription (Lu *et al.*, 2013; Quirós *et al.*, 2014). On the other hand, mitochondrial biogenesis will be driven by the expression of TFAM, as well as of several other nuclear-encoded mitochondrial genes, which will be increased by PGC-1α, through binding to nuclear respiration factors 1 and 2 (NRF-1 and NRF-2), among other transcription factors (Ventura-Clapier, Garnier and Veksler, 2008; Cardanho-Ramos and Morais, 2021).

In comparison with nDNA, mtDNA shows polyploidy, where multiple mtDNA molecules coexist, allowing for the existence of multiple genotypes within the same cell, a concept termed heteroplasmy, in contrast to homoplasmy, where all mtDNA molecules are mutated (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015; Silva-Pinheiro and Minczuk, 2021). Heteroplasmy is ubiquitous in humans and it involves tissue mosaicism (**Figure 18**), where different cells of the same tissue, including postmitotic tissues, may have remarkably different mitochondrial genotypes and, thus, phenotypes (Calloway *et al.*, 2000; He *et al.*, 2010; Sosa *et al.*, 2012; Payne *et al.*, 2013; Ye *et al.*, 2014a;

Li *et al.*, 2015; Naue *et al.*, 2015; Hahn and Zuryn, 2019; J. Xu *et al.*, 2019; Ludwig *et al.*, 2019). For instance, individual muscle fibres are affected by mtDNA mutations independently from each other (Roos *et al.*, 2021) and only specific tissues may be affected at all in the same person (Peverelli *et al.*, 2014). Furthermore, mitochondrial DNA levels will also differ significantly between different tissues, overall decreasing with age (Kelly *et al.*, 2012; Ding *et al.*, 2015; Wachsmuth *et al.*, 2016; Chen *et al.*, 2020).

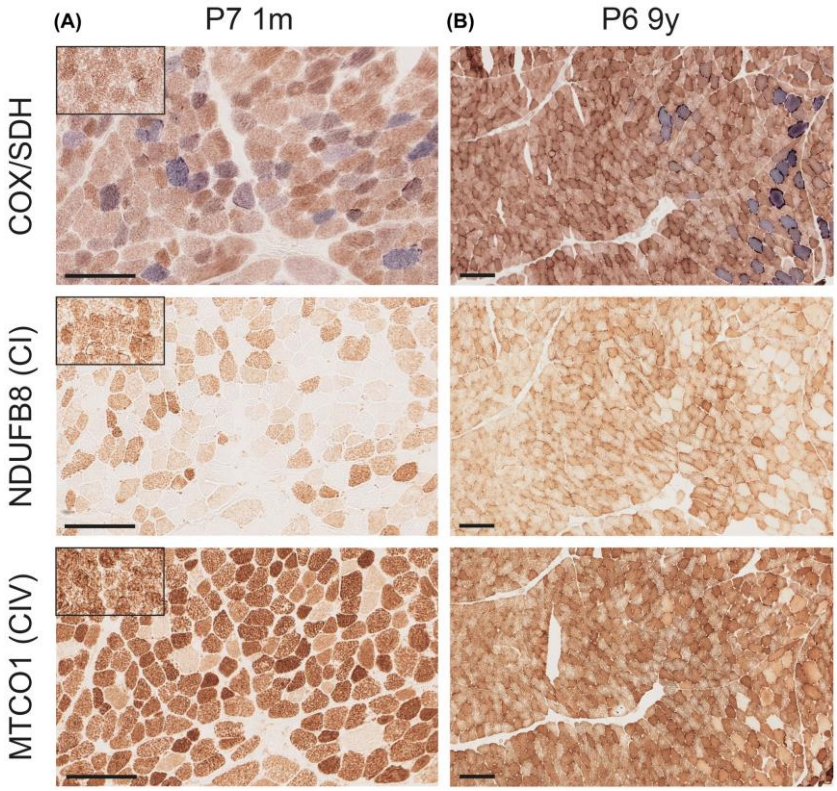


Figure 18 — Mosaicism in skeletal muscle

Serial sections of skeletal muscle of affected individuals with reversible infantile respiratory chain deficiency in the same family (subjects P7 and P6). Fibres lacking complex III (COX) activity are stained blue, while fibres with both complex III an II (SDH), which is solely nuclear-encoded, are stained brown. Immunohistochemistry was also used to visualize NDUFB8 (complex I) and MTCO1 (complex IV). (**A**) P7 at 1 month of age. Insets show staining of an age-matched control; (**B**) P6 at nine years of age. Scale bars measure 100 µm. Unedited image from (Roos *et al.*, 2021) (CC BY-NC 3.0).

This diverse mutational pattern arises from both inherited and somatic mtDNA mutations. On the one hand, inherited mtDNA mutations are currently considered to derive from the maternal progenitor (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; Stewart and Chinnery, 2015; Silva-Pinheiro and Minczuk, 2021), despite a recent report of biparental inheritance in humans (Luo *et al.*, 2018), which was later disputed by several studies (Rius

et al., 2019; Wei *et al.*, 2020; Bai *et al.*, 2021). Most likely, the autosomal transmission of NUMTs was behind the apparent biparental pattern in mtDNA (Wei *et al.*, 2020; Bai *et al.*, 2021). Similar to nDNA and haplotypes (more details in *1.4.6.2. Genetic risk factors)*, inherited mutational patterns might be classified into phylogenetic clusters — haplogroups —, which are mostly used in population genetics and forensics (van Oven, 2015; Weissensteiner, Pacher, *et al.*, 2016; Dür, Huber and Parson, 2021).

Interestingly, there might be a dramatic shift in variant levels (VLs) for the same mutations between mother and child, due to the germline bottleneck, where primordial germ cells undergo rapid replication of mtDNA during oocyte maturation from a reduced pool of available mtDNA molecules, separating into daughter cells with different mutation loads. Nonetheless, mitochondrial segregation appears to be less random than initially thought, with influences from mito–nuclear interactions and maternal age, and possibly extending to very low-level heteroplasmic variants (Guo, C.-I. Li, *et al.*, 2013; Floros *et al.*, 2018; Latorre-Pellicer *et al.*, 2019; Wei *et al.*, 2019; Zaidi *et al.*, 2019).

In addition to inherited mtDNA mutations, there are also somatic mtDNA mutations, most commonly mtDNA deletions for post-mitotic cells and point mutations for mitotic cells (Cortopassi and Arnheim, 1990; Simonetti *et al.*, 1992; Nekhaeva *et al.*, 2002; Taylor *et al.*, 2003; Bender *et al.*, 2006; Campbell *et al.*, 2011, 2012; Nido *et al.*, 2018; Lawless *et al.*, 2020). The main driver behind these mutations, with point mutation rates ~300-fold higher than nDNA (Khrapko *et al.*, 1997), is still unknown, albeit strongly correlated with age (Cortopassi and Arnheim, 1990; Wei, 1992; Michikawa Yuichi *et al.*, 1999; Calloway *et al.*, 2000; Wang *et al.*, 2001; Williams *et al.*, 2013; Greaves *et al.*, 2014; Li *et al.*, 2015; Naue *et al.*, 2015; Kang *et al.*, 2016). Previously, the absence of protective histones, exposure to ROS and RNS, and more rudimentary repair systems were presented as likely mechanisms for mutagenesis (Harman, 1972; Taylor and Turnbull, 2005; Tuppen *et al.*, 2010), but these mechanisms have been called into question by recent experimental data (Alexeyev *et al.*, 2013; Kennedy *et al.*, 2013; Chocron, Munkácsy and Pickering, 2019). Currently, point mutations seem to arise with mtDNA replication (Zheng *et al.*, 2006; Payne *et al.*, 2013; Pantic *et al.*, 2021; Sanchez-Contreras *et al.*, 2021), while the mechanisms behind mtDNA deletions are less well understood (Lawless *et al.*, 2020).

Mitochondrial diseases were first described in 1962 (Luft *et al.*, 1962), while the first pathogenic mtDNA mutations were discovered in 1988 (Holt, Harding and Morgan-Hughes,

1988; Wallace *et al.*, 1988). Mitochondrial diseases are very heterogeneous diseases affecting multiple organ systems and with a large variety of clinical symptoms and timing of disease onset, most commonly affecting the CNS and neuromuscular systems, since these tissues have a high energy demand, mostly dependent on OXPHOS (Taylor and Turnbull, 2005; Tuppen *et al.*, 2010; Chung *et al.*, 2021; Ng *et al.*, 2021). Adult-onset mitochondrial diseases affect 1 in 4,300 adults, where mtDNA mutations are ~7 times more prevalent than nDNA in triggering disease (Gorman *et al.*, 2015). The level of heteroplasmy of the pathogenic mutation plays a significant role in triggering disease, where a high mutational burden in the affected tissue must usually surpass a specific threshold, typically 60–80% (Stewart and Chinnery, 2015). Yet, peripheral mutational burden has also been linked to higher disease severity in the CNS for patients with mitochondrial diseases harboring the m.3243A>G mutation, such as mitochondrial encephalopathy lactic acidosis and stroke-like episodes (MELAS) and mitochondrial inherited deafness and diabetes (MIDD) (Haast *et al.*, 2018).

Similar to nDNA, NGS has dramatically changed research on mtDNA-derived diseases (Yuan *et al.*, 2020; Cai *et al.*, 2021; Poole *et al.*, 2021) and mtDNA forensics (Canale, Parson and Holland, 2022; Massilani *et al.*, 2022), additionally moving towards single-cell genomics (Lareau *et al.*, 2021; Penter *et al.*, 2021). However, the peculiar nature of mtDNA has hindered the development of specific data analysis guidelines, since the large majority of datasets, guidelines, and bioinformatic pipelines for variant discovery analysis in NGS are primarily, and sometimes exclusively, focused on nDNA (Van der Auwera *et al.*, 2013; 1000 Genomes Project Consortium *et al.*, 2015), with mtDNA NGS analysis lagging years behind (*gatk4-mitochondria-pipeline*, 2020).

Out of the existing bioinformatic tools for mtDNA, with quality control assessment, variant calling, and haplogroup assignment (Lee *et al.*, 2008; Fan and Yao, 2011; Zhidkov *et al.*, 2011; Guo, J. Li, *et al.*, 2013; Lott *et al.*, 2013; Vianello *et al.*, 2013; Yang *et al.*, 2013; Calabrese *et al.*, 2014; Falk *et al.*, 2015; Navarro-Gomez *et al.*, 2015; Vellarikkal *et al.*, 2015; Weissensteiner, Forer, *et al.*, 2016; Weissensteiner, Pacher, *et al.*, 2016; Ishiya and Ueda, 2017; Rueda and Torkamani, 2017; Preste *et al.*, 2019), very few incorporate user accessibility, integrative data analysis, and regular updates. A suitable option is mtDNA-Server (Weissensteiner, Forer, *et al.*, 2016), which allows for FASTQ and BAM files as input and, considering the revised Cambridge Reference Sequence (rCRS) (Andrews *et al.*, 1999)

as a reference, identifies heteroplasmic and homoplasmic variants, assigns a haplogroup with HaploGrep 2 (Weissensteiner, Pacher, *et al.*, 2016), based on PhyloTree Build 17 (van Oven, 2015) with an updated algorithm (Dür, Huber and Parson, 2021), and performs a contamination check and coverage analysis.

Recently, a novel approach to mtDNA NGS through whole genome sequencing (WGS) has been developed with the Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, Waltham, MA, USA) (Chaitanya *et al.*, 2015). This library preparation kit, with 162 amplicons that target the whole mtDNA, is mostly used for forensic samples, achieving reliable results in often degraded samples (Strobl *et al.*, 2018; Wai, Barash and Gunn, 2018; Yao *et al.*, 2018; Cuenca *et al.*, 2020) and with low DNA input (usually 0.1 ng of genomic DNA, but 6.25 pg (Pereira, Longobardi and Børsting, 2018) and, very recently, 0.6 pg (Faccinetto *et al.*, 2021) have been reported). Despite its most common use in forensic sciences, it has also been used in a range of other applications, from worldwide lineage studies (Strobl *et al.*, 2019), to rare mtDNA differences in monozygotic twins (Ming *et al.*, 2019).

A disadvantage of this approach, however, is its use of a modified mtDNA reference with 16,649 bps, instead of the conventional 16,569 bps. Since the last amplicon corresponds to the last 28 bps and the first 80 bps in the widely used rCRS, the Precision ID mtDNA reference duplicates the first 80 bps of the rCRS at the end of the reference. Thus, most of its bioinformatic processing must be done opaquely in the enterprise software, Ion Torrent Suite™ Software (TSS), with the possible addition of other company-owned software (Woerner *et al.*, 2018; Roth *et al.*, 2019; Cihlar, Amory, *et al.*, 2020; Faccinetto *et al.*, 2021). In addition to producing an unorthodox reference (Parson *et al.*, 2014), which impedes further processing in open-source bioinformatic tools, sensitivity analyses have set its heteroplasmy threshold level to 10% (Cho *et al.*, 2018; Churchill *et al.*, 2018; Pereira, Longobardi and Børsting, 2018).

Regarding mtDNA mutations in MS, a few researchers initially focused on the rare Harding's disease, where patients with mtDNA mutations associated with a known mitochondrial disease, Leber's hereditary optic neuropathy (LHON), had an MS-like disease course, particularly women with the m.11778G>A mutation (Harding *et al.*, 1992; Joshi and Kermode, 2019; Uittenbogaard *et al.*, 2019). However, further analysis has yielded no

association between typical MS and Harding's disease (Pfeffer *et al.*, 2013; Andalib *et al.*, 2017).

In parallel, considering typical MS cohorts, data on mtDNA mutations has been conflicting, with a few studies finding no significant differences between HC and PwMS (Chalmers *et al.*, 1996; Kalman and Alder, 1998; Ohlenbusch, Wilichowski and Hanefeld, 1998a, 1998b; Wilichowski, Ohlenbusch and Hanefeld, 1998; Otaegui *et al.*, 2004; Fernández *et al.*, 2008; Andalib *et al.*, 2015b, 2015a, 2019; Souren *et al.*, 2016; Pienaar *et al.*, 2021), and other studies finding several associations with the disease (Kalman, Lublin and Alder, 1995, 1996; Mayr-Wohlfart *et al.*, 1996; Kalman *et al.*, 1999; Houshmand *et al.*, 2005; Vogler *et al.*, 2005; Hassani-Kumleh *et al.*, 2006; Mihailova *et al.*, 2007; Vyshkina *et al.*, 2008; Venkateswaran *et al.*, 2011; Poursadegh Zonouzi *et al.*, 2014; Kozin *et al.*, 2018; M. Kozin *et al.*, 2020; M. S. Kozin *et al.*, 2020). Data from larger cohorts and a systematic review (>1,000 cases and controls) (Yu *et al.*, 2008; Hudson *et al.*, 2014; Andalib *et al.*, 2016; Yonova-Doing *et al.*, 2021) has demonstrated that a few mutations are associated with higher MS risk, while others seem to be protective (**Table 3**).

Table 3 — Mitochondrial variants associated with Multiple Sclerosis risk

Position	Mutation	Region	*N* PwMS	*N* HC	OR (95% CI)	Ref
4126	4216C	*MT-ND1*	297/1403	236/1436	1.38 (1.13–1.67)	(Andalib *et al.*, 2016)
5657	5657G	*MT-TN*	211/8809	59/4902	2.07 (1.56–2.74)	(Hudson *et al.*, 2014)
6261	6261A	*MT-CO1*	132/8809	118/4902	0.63 (0.49–0.80)	(Hudson *et al.*, 2014)
6777	6777C	*MT-CO1*	370/8809	265/4902	0.78 (0.66–0.91)	(Hudson *et al.*, 2014)
7559	7559G	*MT-TD*	18/1149	1289/278030	2.06 (1.59–2.67)	(Yonova-Doing *et al.*, 2021)
10035	10035C	*MT-TG*	229/8809	162/4902	0.78 (0.63–0.95)	(Hudson *et al.*, 2014)
10398	10398G	*MT-ND3*	297/1149	58567/278030	1.15 (1.07–1.23)	(Yonova-Doing *et al.*, 2021)
10399	10399A	*MT-ND3*	1823/8809	1127/4902	0.87 (0.80–0.95)	(Hudson *et al.*, 2014)
11486	11486C	*MT-ND4*	150/8809	127/4902	0.67 (0.53–0.85)	(Hudson *et al.*, 2014)
13117	13117G	*MT-ND5*	17/1149	1550/278030	1.65 (1.30–2.11)	(Yonova-Doing *et al.*, 2021)
13708	13708A	*MT-ND5*	157/1092	89/1082	1.71 (1.29–2.27)	(Yu *et al.*, 2008)

Only data from cohorts with >1,000 participants was included. Abbreviations: CI — confidence interval; HC — healthy controls; *N* — number; OR — odds ratio; PwMS — patients with Multiple Sclerosis; Ref — reference.

Additionally, large cohort studies have also established that a few haplogroups are associated with an increased MS risk, namely, J (Yu *et al.*, 2008; Tranah *et al.*, 2015), JT (Tranah *et al.*, 2015), T (Tranah *et al.*, 2015), and U (Ban *et al.*, 2008). The J haplogroup was further associated with an increased risk of PPMS (Tranah *et al.*, 2015). However, no association with haplogroup was found in the most recent and largest study (Yonova-Doing *et al.*, 2021).

In accordance with the pivotal role of mitochondrial dysfunction in the CNS for driving disease progression in MS (more details in *1.4.4. Pathology*), immune cells of PwMS in the periphery have also demonstrated significant mitochondrial dysfunction. Peripheral blood mononuclear cells (PBMCs) from PwMS, which include lymphocytes, monocytes, and dendritic cells (**Figure 10**) (Kleiveland, 2015; Alfatni *et al.*, 2020), exhibit increased ROS production (Gonzalo *et al.*, 2019; De Rasmo *et al.*, 2020), a significant decrease in protein content of the ETC complexes I, III, and V (Gonzalo *et al.*, 2019), and decreased antioxidant capacity (Gonzalo *et al.*, 2019). Focusing on lymphocytes from PwMS, they further show a decreased ΔΨm (Armon-Omer *et al.*, 2020). Finally, $CD4^+$ T cells from PwMS have impaired OXPHOS (De Riccardis *et al.*, 2015; La Rocca *et al.*, 2017), particularly driven by complex I and IV (De Riccardis *et al.*, 2015), with PPMS patients having a more significant disruption in their metabolism, in comparison with SPMS patients (De Biasi *et al.*, 2019), which is a similar pattern to the one associated with aging (Bektas *et al.*, 2019; Severin *et al.*, 2021). Interestingly, while glycolytic activity in $CD4^+$ T cells has yielded conflicting results (De Riccardis *et al.*, 2015; La Rocca *et al.*, 2017), serum or extracellular lactate is consistently elevated in PwMS (De Riccardis *et al.*, 2015; De Biasi *et al.*, 2019; Gonzalo *et al.*, 2019; Esmael *et al.*, 2021) and correlates with disease activity (Esmael *et al.*, 2021).

This is particularly relevant when we consider the impact of metabolism in $CD4^+$ T cells. Following a compatible engagement between a TCR and an antigenic peptide bound to an MHC class II molecule from an APC (more details in *1.2.1. The innate immune system* and *1.2.2. The adaptive immune system*), a naïve $CD4^+$ T cell can differentiate into multiple subsets according to the cytokine environment (more details in *1.2.2. The adaptive immune system*), including Th17 cells, which are pro-inflammatory and express RORγt/c, RORα, and STAT3, and Treg cells, which are anti-inflammatory and express FOXP3. These peripheral $CD4^+$-derived Treg cells are called adaptive or induced Treg (iTreg) cells to differentiate them from thymus-derived “natural” Treg (nTreg) cells. Even though both Treg subtypes

have immunosuppressive functions, nTreg cells are particularly more susceptible to their environment and are, thus, able to switch to an inflammatory response when exposed to pro-inflammatory environments, including a Th17-like phenotype in the presence of IL-6 (Zhu, Yamane and Paul, 2010; Kleinewietfeld and Hafler, 2013; Lin *et al.*, 2013).

External cues are, however, not the sole factors that influence the phenotype of immune cells, as cellular metabolism has recently arisen as an important immunomodulator. While Treg cells are usually dependent on lipid oxidation, the citric acid cycle, and subsequent mitochondrial OXPHOS, effector T cells rely on cytosolic glycolysis, with the conversion of glucose to lactate, amino acid metabolism, and fatty acid synthesis (Windt and Pearce, 2012; Buck, O'Sullivan and Pearce, 2015; Freitag *et al.*, 2016; O'Neill, Kishton and Rathmell, 2016).

On the one hand, when naïve $CD4^+$ cells are activated, they experience a shift from a quiescent anabolic state to a catabolic glycolytic state — the Warburg effect. This switch is thought to be a result of the activation of the mammalian target of rapamycin (mTOR) that encompasses two complexes: mTORC1 and mTORC2. Following the activation of PI3K by TCR–CD3/CD28 stimulation or growth factors as IL-2, 3-phosphoinositide–dependent protein kinase-1 (PDPK1) and Akt are recruited. Next, PDPK1 and mTORC2 phosphorylate Akt, which activates mTORC1. This leads to a promotion of cytosolic glycolysis over OXPHOS and the subsequent differentiation, growth, and function of effector T cells (Michalek *et al.*, 2011; Windt and Pearce, 2012; Buck, O'Sullivan and Pearce, 2015; Freitag *et al.*, 2016).

On the other hand, Treg cells have an increased AMP kinase (AMPK) phosphorylation and activation, which suppresses mTOR activity and promotes catabolism, particularly lipid oxidation. Moreover, nuclear FOXP3 alone is capable of increasing mitochondrial respiration and lipid oxidation capacity (Michalek *et al.*, 2011; Windt and Pearce, 2012; Buck, O'Sullivan and Pearce, 2015; Freitag *et al.*, 2016; Howie *et al.*, 2017).

Indeed, manipulation of the metabolic substrates allows for an increased plasticity between Th17 and Tregs phenotypes: inhibition of cytosolic glycolysis (Shi *et al.*, 2011) and of fatty acid synthesis (Berod *et al.*, 2014) decreases the number of Th17 cells and promotes Treg generation.

Experimental MS models have further demonstrated that immunometabolism is critical for MS pathogenesis and constitutes an important putative target for treatment: CNS-homing $CD4^+$ and $CD8^+$ T cells are highly dependent on cytosolic glycolysis (Seki *et al.*,

2017), Treg cells show an increased oxidative stress (Alissafi *et al.*, 2020), and Th17 cells have higher levels of cytosolic glycolysis (D. Zhang *et al.*, 2019) and polyamine metabolism (Wagner *et al.*, 2021), while the promotion of OXPHOS ameliorates disease activity (Luu *et al.*, 2019).

Remarkably, current DMDs (more details in *1.4.3. Clinical course and management*) also have an impact on the metabolism of $CD4^+$ T cells from PwMS: (**I**) interferon-beta leads to a dose-response decrease of intracellular ATP production and of $\Delta\Psi_m$ (Haghikia *et al.*, 2015), despite an increase in both cytosolic glycolysis and OXPHOS (La Rocca *et al.*, 2017); (**II**) glatiramer acetate also leads to a decrease in $\Delta\Psi_m$, albeit with an increase in OXPHOS through rescuing complex I and IV activities, a decrease in cytosolic glycolysis, a decrease in lactate production, and an increase in antioxidant capacity (De Riccardis *et al.*, 2016); and (**III**) dimethyl fumarate has been shown to downregulate cytosolic glycolysis (Diebold *et al.*, 2018; Kornberg *et al.*, 2018), while enhancing oxidative stress by decreasing antioxidative capacity (Diebold *et al.*, 2018; Liebmann *et al.*, 2021).

It should be noted, however, that there are mechanisms that are shared between different cellular subtypes and that a black-and-white separation is not always possible. For example, in iTreg cells glycolysis is necessary for the expression of FOXP3 exon 2 splicing variants (De Rosa *et al.*, 2015), *de novo* lipogenesis is necessary for Treg cell function (Galgani *et al.*, 2016), OXPHOS is necessary for naïve $CD4^+$ activation (Chang *et al.*, 2013) and Th1 cell induction (Kolev *et al.*, 2015), and AMPK can be used by effector T cells to induce a glutamine-dependent mitochondrial metabolism when there is a lack of glucose (Blagih *et al.*, 2015). Accordingly, OXPHOS has been shown to be pivotal in driving Th17 activity in EAE models (Shin *et al.*, 2020), while the immunomodulatory effects of teriflunomide and dimethyl fumarate in PwMS appear to also derive from an inhibition of OXPHOS, in addition to the inhibition of cytosolic glycolysis (Klotz *et al.*, 2019; Liebmann *et al.*, 2021).

Interestingly, despite the evidence of mtDNA polymorphisms modulating both metabolism and immunity (Uittenbogaard *et al.*, 2019; Beadnell *et al.*, 2020; Schilf *et al.*, 2021), and the tissue-specificity of mtDNA variants, including T cells (Yao, Kajigaya and Young, 2015; Walker *et al.*, 2020; Rausser *et al.*, 2021; Tang *et al.*, 2021), no studies have addressed if the OXPHOS dysfunction of $CD4^+$ T cells from PwMS might be linked to mitochondrial genotype.

Page intentionally left blank

II — OBJECTIVES

Page intentionally left blank

MS is a chronic neuroinflammatory and neurodegenerative disease that causes demyelination and neuroaxonal loss in the CNS, leading to disability, particularly in women (Dendrou, Fugger and Friese, 2015; Berkowitz, 2016a; Reich, Lucchinetti and Calabresi, 2018; Ropper *et al.*, 2019; Meador, 2020). Mitochondria are pivotal for MS disease activity and progression:

I. in the CNS, mitochondrial dysfunction is the main trigger for neuroaxonal loss and brain atrophy (Vladimirova *et al.*, 1998; Craner *et al.*, 2004; Young *et al.*, 2008; Mahad *et al.*, 2009; Ziabreva *et al.*, 2010; Campbell *et al.*, 2011, 2012; Haider *et al.*, 2011; Fischer *et al.*, 2012; Mao *et al.*, 2013; Witte *et al.*, 2013; Bouafia *et al.*, 2014; Dendrou, Fugger and Friese, 2015; Lassmann and van Horssen, 2016; Sadeghian *et al.*, 2016; Wang *et al.*, 2016; Warne *et al.*, 2016; Smorodchenko *et al.*, 2017; Campbell and Mahad, 2018; Lassmann, 2018; Liessem-Schmitz *et al.*, 2018; Mills *et al.*, 2018; Andravizou *et al.*, 2019; Bando and Geisler, 2019; Buonvicino *et al.*, 2019; Campbell, Licht-Mayer and Mahad, 2019; Correale, Marrodan and Ysrraelit, 2019; Fetisova *et al.*, 2019; Largani *et al.*, 2019; Martín-Montañez *et al.*, 2019; Ng *et al.*, 2019; Varhaug *et al.*, 2020; Martínez-Pinilla *et al.*, 2021);
II. in the periphery, immunometabolism is critical for CNS autoimmunity, while PwMS show inappropriate mitochondrial phenotypes and/or OXPHOS insufficiency, particularly in $CD4^+$ T cells (De Riccardis *et al.*, 2015, 2016; Haghikia *et al.*, 2015; La Rocca *et al.*, 2017; Seki *et al.*, 2017; Diebold *et al.*, 2018; Kornberg *et al.*, 2018; De Biasi *et al.*, 2019; D. Zhang *et al.*, 2019; Gonzalo *et al.*, 2019; Klotz *et al.*, 2019; Luu *et al.*, 2019; Alissafi *et al.*, 2020; Armon-Omer *et al.*, 2020; De Rasmo *et al.*, 2020; Shin *et al.*, 2020; Esmael *et al.*, 2021; Liebmann *et al.*, 2021; Wagner *et al.*, 2021).

However, no studies have explored if mitochondrial genotype could be behind the mitochondrial dysfunction in $CD4^+$ T cells from PwMS. Instead, most studies have opted for whole blood analysis (Kalman, Lublin and Alder, 1995, 1996; Chalmers *et al.*, 1996; Mayr-Wohlfart *et al.*, 1996; Kalman and Alder, 1998; Ohlenbusch, Wilichowski and Hanefeld, 1998a, 1998b; Wilichowski, Ohlenbusch and Hanefeld, 1998; Kalman *et al.*, 1999; Otaegui *et al.*, 2004; Houshmand *et al.*, 2005; Vogler *et al.*, 2005; Hassani-Kumleh *et al.*, 2006;

Mihailova *et al.*, 2007; Ban *et al.*, 2008; Fernández *et al.*, 2008; Vyshkina *et al.*, 2008; Yu *et al.*, 2008; Venkateswaran *et al.*, 2011; Hudson *et al.*, 2014; Poursadegh Zonouzi *et al.*, 2014; Andalib *et al.*, 2015b, 2015a, 2016, 2019; Tranah *et al.*, 2015; Souren *et al.*, 2016; Kozin *et al.*, 2018; M. Kozin *et al.*, 2020; M. S. Kozin *et al.*, 2020; Pienaar *et al.*, 2021; Yonova-Doing *et al.*, 2021), despite the significant tissue-specificity of mtDNA variants (Calloway *et al.*, 2000; He *et al.*, 2010; Li *et al.*, 2015; Naue *et al.*, 2015; Hahn and Zuryn, 2019; J. Xu *et al.*, 2019; Ludwig *et al.*, 2019), including T cells (Yao, Kajigaya and Young, 2015; Walker *et al.*, 2020; Rausser *et al.*, 2021; Tang *et al.*, 2021).

Nevertheless, NGS approaches to mtDNA have been hindered by fewer resources, in comparison with nDNA (Lee *et al.*, 2008; Fan and Yao, 2011; Zhidkov *et al.*, 2011; Guo, J. Li, *et al.*, 2013; Lott *et al.*, 2013; Van der Auwera *et al.*, 2013; Vianello *et al.*, 2013; Yang *et al.*, 2013; Calabrese *et al.*, 2014; 1000 Genomes Project Consortium *et al.*, 2015; Falk *et al.*, 2015; Navarro-Gomez *et al.*, 2015; Vellarikkal *et al.*, 2015; Weissensteiner, Forer, *et al.*, 2016; Weissensteiner, Pacher, *et al.*, 2016; Ishiya and Ueda, 2017; Rueda and Torkamani, 2017; Preste *et al.*, 2019; *gatk4-mitochondria-pipeline*, 2020). A suitable example is the Precision ID mtDNA Whole Genome Panel, mostly used in forensic research, which achieves mtDNA WGS with very low DNA amounts (Chaitanya *et al.*, 2015; Pereira, Longobardi and Børsting, 2018; Strobl *et al.*, 2018, 2019; Wai, Barash and Gunn, 2018; Yao *et al.*, 2018; Ming *et al.*, 2019; Cuenca *et al.*, 2020; Faccinetto *et al.*, 2021). In contrast, it uses a modified rCRS (Andrews *et al.*, 1999) as a reference (Woerner *et al.*, 2018), thus impeding bioinformatic processing with third-party tools (Roth *et al.*, 2019; Cihlar, Amory, *et al.*, 2020; Faccinetto *et al.*, 2021). Also, it has a relatively high heteroplasmy threshold level (Cho *et al.*, 2018; Churchill *et al.*, 2018; Pereira, Longobardi and Børsting, 2018).

In order to better understand the role of mitochondrial genotype in MS activity and progression, I aimed to provide insight regarding the following questions:

1. Is there an approach for the Precision ID mtDNA Whole Genome Panel, which yields an output in the correct rCRS and a lower heteroplasmy threshold?
2. Is mtDNA different in $CD4^+$ T cells between newly diagnosed PwMS and HC?
3. Does the analysis of baseline and longitudinal mtDNA changes regarding MS clinical phenotypes enable a correlation with disease activity and progression?

Ultimately, I hoped to determine if mtDNA had a significant predictive value for disease activity and progression, opening the possibility of identifying novel prognostic factors and the future development of diagnostic approaches.

Page intentionally left blank

III — MATERIALS AND METHODS

Page intentionally left blank

3.1. PrecisionCallerPipeline

3.1.1. Sample acquisition/collection

In order to validate the bioinformatic pipeline, termed PrecisionCallerPipeline (PCP), which has already been published (Cortes-Figueiredo *et al.*, 2021) and is available in *Annex A — Page LXIII*, three cell-line samples from the Coriell Institute for Medical Research (Camden, NJ, USA) were acquired, i.e., HG00256, HG01626, and HG01757. These samples have been sequenced within the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2015) on Illumina HiSeq (Illumina, Inc., San Diego, CA, USA), with three different sequencing runs, mostly targeted at nDNA: (**I**) exome sequencing — exome; (**II**) low-coverage sequencing — lowCov; and (**III**) high coverage sequencing — highCov.

With the aim of evaluating the sensitivity, specificity, precision, and F_1 score (harmonic mean of precision and sensitivity) (Weissensteiner, Forer, *et al.*, 2016; Fazzini *et al.*, 2021) with different variant detection thresholds, I performed three different mixtures at five different mixture levels (1, 2, 5, 10, and 25%):

- **Mixture HT** — Minor component: HG00256 (haplogroup H5b2, short identifier H) + Major component: HG01626 (haplogroup T2a1b1a1b, short identifier T);
- **Mixture TU** — Minor component: HG01626 (haplogroup T2a1b1a1b, short identifier T) + Major component: HG01757 (haplogroup U4a, short identifier U);
- **Mixture UH** — Minor component: HG01757 (haplogroup U4a, short identifier U) + Major component: HG00256 (haplogroup H5b2, short identifier H).

DNA concentration measured prior to shipping ranged from 307 to 331 ng/µL. Thus, DNA mixtures were volume-based (**Figure 19A**).

As an independent set of samples, I analyzed 50 samples derived from a prospective stroke cohort of patients at the Hospital de Santa Maria, Centro Hospitalar Universitário Lisboa Norte. Blood samples were collected within 72 h of hospital admission and all cases were reviewed and confirmed by trained neurologists. Inclusion criteria were: (**I**) ischemic stroke; (**II**) age ≥ 18 years; (**III**) blood samples collected up to 72 h after symptom onset. The exclusion criteria were: (**I**) active cancer diagnosis; (**II**) previous cerebral revascularization surgeries; (**III**) modified Rankin score (Bamford *et al.*, 1989) ≥ 5. The approval of the institutional review board (IRB) was conceded by the Comissão de Ética do

Centro Académico de Medicina de Lisboa (reference 435/16, approved on December 14, 2016), informed consents were given by every subject, and the study followed the standards of the Declaration of Helsinki.

DNA extraction of these samples was almost exclusively performed by Filipa Carvalho, PhD and Bernardo Antunes, MSc with the QIAamp® DNA Blood Midi Kit (QIAGEN GmbH, Hilden, Germany) after PBMC isolation, according to the manufacturer's instructions.

Namely, heparinized blood taken from patients was mixed with 1:1 volume of $PBS^{+/+}$ and 5 mL of Ficoll, followed by a centrifugation at 805 x *g* for 20 min (acceleration 4, break 0). Afterwards, the buffy coat containing the PBMCs was collected and mixed with 200 µL of QIAGEN Protease, and, if necessary, $PBS^{-/-}$ would be added to bring the total volume up to 2 mL. Subsequently, 2.4 mL of Buffer AL were added, followed by an incubation at 70° C for 10 min in a pre-heated water bath. Then, 2 mL of ethanol 100% were added and mixed in. Each half of the mixture was transferred to the QIAamp Midi column, followed by two centrifugations at 1811 x *g* for 3 min (half of the mixture in each centrifugation). Afterwards, 2 mL of Buffer AW1 were added to the column and centrifuged at 3220 x *g* for 1 min, followed by the addition of 2 mL of Buffer AW2, and another centrifugation at 3220 x *g* for 30 min.

After that, the QIAamp Midi column was transferred to a 15 mL centrifuge tube, incubated in dry heat at 70° C for 10 min, and 300 µL of DNAse/RNAse-free water at room temperature (RT) were added onto the membrane's column with an incubation of 5 min. Then, after a centrifugation at 3220 x *g* for 2 min, the eluate was reloaded on the QIAamp Midi column, followed by a 5 min incubation at RT, and centrifuged at 3220 x *g* for 2 min.

Afterwards, DNA was collected, and its concentration was measured through spectral analysis at 1 nm increments from 230 nm to 290 nm with the Synergy HTX Multi-mode Reader, in combination with its Take3 microvolume plate (BioTek Instruments, Winooski, VT, USA). The 260/280 nm absorbance ratio was used for quality control, with the aim of a 1.8 value (Nicklas and Buel, 2003; Peter Brescia, 2021). Finally, DNA was stored at -20° C.

3.1.2. DNA sequencing

Samples were sequenced without a prior long-range polymerase chain reaction (LR-PCR) with the Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo

Fisher Scientific, Waltham, MA, USA), in combination with the Ion Torrent™ Ion S5™ (Thermo Fisher Scientific, Waltham, MA, USA), in accordance with the manufacturer's instructions. Briefly, DNA was quantified with a Qubit® 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and samples were diluted to 0.0067 ng/µL for an input of 0.1 ng of genomic DNA in 15 µL. Libraries were prepared using the Ion Chef™ automated protocol, and samples were then run on 530™ chips with the Ion Torrent™ Ion S5™ at IPATIMUP — Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Porto, Portugal).

Samples with a coverage uniformity < 85% and/or a mean coverage < 1500 reads were resequenced.

3.1.3. Bioinformatic processing

3.1.3.1. PrecisionCallerPipeline

The PCP pipeline automatically takes the FASTQ files from the sequencing facility and outputs fully aligned BAM files mapped to the commonly used reference sequence, rCRS (Andrews *et al.*, 1999). The workflow is based on Snakemake (Köster and Rahmann, 2012), which uses: (**I**) Awk, for SAM file editing (Aho, Kernighan and Weinberger, 1979); (**II**) BEDTools, for BAM to FASTQ conversion (Quinlan and Hall, 2010); (**III**) BWA-MEM, for read alignment (Li, 2013); (**IV**) Pycision, for amplicon delimitation and selection (Woerner *et al.*, 2018); (**V**) SAMtools for BAM conversion, sorting, indexing, and merging (H. Li *et al.*, 2009); and (**VI**) Trimmomatic for read quality control and trimming (Bolger, Lohse and Usadel, 2014) (**Figure 19B**). Removal of NUMTs was tested before and after final BAM merging with RtN! (Woerner *et al.*, 2020) (**Figure 19C**).

Samples were processed in a Linux-based system. For simplicity, a predetermined file structure and the necessary files (except for the files from external software) may be downloaded from https://github.com/filcfig/PCP.git. A separate Snakefile (Snakefile_no RtN) is also provided to run the samples without the removal of NUMTs.

Read quality analysis was performed with FastQC (Andrews *et al.*, 2010) and MultiQC (Ewels *et al.*, 2016). For read quality control, I opted to crop reads at 160 bps, taking into account: (**I**) the visual inspection of Phred score patterns per read bp (**Figure 21D**); (**II**) the length range of 73–137 bps for the known "callable" sections of each amplicon (Woerner

et al., 2018), also considering the company-reported amplicon average length of 163 bps, as the exact coordinates of the amplicon themselves are unknown.

Variant calling was performed with freebayes v1.3.5 (Garrison and Marth, 2012), mutserve v2, the command line interface and successor of the mtDNA-Server pipeline (Weissensteiner, Forer, *et al.*, 2016; Weissensteiner *et al.*, 2021), and VarScan 2 v2.3.7 (Koboldt *et al.*, 2012). Initially, 21 different heteroplasmy thresholds were used, ranging from 0.4% to 10.0% (**Figure 19C** and **Figure S1**). After optimization, the maximum F_1 score was achieved with a heteroplasmy threshold of 2.5% and with NUMT removal before the final BAM merge (**Figure 19B**, **Table S2**, and **Table S3**); this processing was then used in all variant calling methods (VCMs). For the Illumina sequencing runs (exome, lowCov, and highCov), a threshold of 0.4% was maintained as a reference. Very similar parameters as (Fazzini *et al.*, 2021) were adopted, with the exception of a base quality score of 20 for all VCMs, as well as, for freebayes, where I used *"--ploidy 1 --pooled-continuous"*. Similar to (Fazzini *et al.*, 2021), variants in positions 302–315 (position 310 was blacklisted in the analysis), 523–524, and 3104–3110 were excluded. Only single nucleotide substitutions were considered, and variants below the established threshold were filtered.

Haplogroup calling was carried out through HaploGrep v2.4.0 (Weissensteiner, Pacher, *et al.*, 2016), and a contamination check was done with Haplocheck v1.3.3 (Weissensteiner *et al.*, 2021), based on the output from mutserve.

3.1.3.2. Torrent Suite™ Software

For the current gold standard, data from each run was processed using the Ion Torrent™ specific pipeline software, TSS, using the reference sequence PrecisionID_mtDNA_rCRS, and target regions PrecisionID_mtDNA_WG_targets with the plugins CoverageAnalysis and VariantCaller. FASTQ and BAM files were generated using the plugin FileExporter. The software versions ranged from v5.8, v5.10, and v5.12, according to the date of each run.

I compiled the VCF files arising from the sequencing runs and corrected all positions > 16,569 to the first 80 bps in the rCRS. If there were equal variants with different coverages and VLs, particularly in the first 80 bps, the mean coverage and mean VL were calculated per mutation. Only single nucleotide substitutions were considered.

After exporting the corrected variants in a VCF format, where GT 1/0 was defined for VL ≥ 90%, samples were run through HaploGrep and Haplocheck, similar to PCP.

3.1.4. Data analysis

Data analysis was performed with R version 4.1.1 (R Core Team, 2021) in RStudio (RStudio Team, 2021) with the packages extrafont (Chang, 2014), infer (Bray *et al.*, 2021), magick (Ooms, 2021), patchwork (Pedersen, 2020), readxl (Wickham and Bryan, 2019), remotes (Hester *et al.*, 2021), scales (Wickham and Seidel, 2020), svglite (Wickham *et al.*, 2021), and tidyverse (Wickham *et al.*, 2019), as well as Excel 2016 (Microsoft Corporation, Redmond, WA, USA).

Analyses were performed in parallel for PCP's output and TSS's output. For the primary analysis, variants from the resequenced unmixed samples were compared to the ones identified in Illumina for the exome, lowCov, and highCov sequencing runs, and classified according to their reliability:

1. Grade A variants: homoplasmic variants (mean VL ≥ 95%) found in both highCov and lowCov, regardless of exome;
2. Grade B variants: heteroplasmic variants (mean VL ≥ 0.4% and ≤ 95%) found in both highCov and lowCov, regardless of exome, or found in highCov plus exome, or lowCov plus exome;
3. Grade C variants: found in a single sequencing run;
4. Novel variants: found in the Ion Torrent runs only.

For the mixture analysis, variants from the 15 mixtures were classified according to the interaction between both VL and primary variant classifications (explained previously) in both components (minor and major). I began with a manual curation of all 512 theoretical combinations, which gave rise to 128 possible scenarios, grouped in 47 micro classes, seven meso classes, and five macro classes (**Table S1**). In this case, due to the possibility of shared variants — variants in the same position in two different samples — having two different classifications, I assumed that lower grade variants would prevail in combinations. Hence, a shared variant with both Grade A and Grade C variant classifications would receive a Grade C classification.

3.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort

3.2.1. Cohort description

The study cohort was based on the Berlin CIS-Cohort (Charite University, Berlin, Germany, 2011) at the Charité — Universitätsmedizin Berlin, Berlin, Germany, which includes individuals with age ≥ 18 years with either a CIS diagnosis within six months or a RRMS diagnosis within two years from symptom onset (diagnoses were updated according to the 2017 revisions of the McDonald criteria (Thompson *et al.*, 2018)); PwMS with other diagnoses were excluded.

I had access to ready-available clinical data, collected by the team at Charité. To address PwMS' disease activity and progression, a follow-up period of three years was considered, where the following variables were assessed: number of relapses and relapse proximity; EDSS (Kurtzke, 1983) and MSFC (Cutter *et al.*, 1999) scores; T2 hyperintense lesions, T1 gadolinium-enhancing lesions — MRI; and GCL and RNFL thickness — OCT. Additionally, blood samples were collected at these clinical visits and, subsequently, PBMCs were isolated.

Recruitment of HC without a family history of MS was finalized on May 20, 2019, where HC were matched to PwMS, in a 1:1 ratio, according to sex and a maximum age difference of five years.

IRB approval had already been obtained, informed consents were given by every subject, and the study followed the standards of the Declaration of Helsinki (World Medical Association, 2013).

Following the longitudinal data availability described above, 61 PwMS met the inclusion criteria and were included in this study cohort. In addition, since mtDNA WGS in $CD4^+$ T cells from PwMS is unreported in the literature (more details in *1.4.6.2.1. Mitochondrial DNA*), a pilot study with 20 triplets (20 PwMS at V1&V2 + 20 age- and sex-matched HC) was performed to determine the appropriateness of the sample size through the Dupont method (Dupont, 1988) (more details in *4.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort*).

3.2.2. Sample processing

3.2.2.1. Sample transportation, thawing, and initial processing

Samples were stored at -80° C and sample processing was performed at the same time for each triplet (PwMS at V1&V2 + HC), to minimize differences within processing. Two PBMC aliquots (~3 mL in total) per subject were used.

Processing began by gently thawing the cells in a pre-warmed water bath at 37° C until they were nearly thawed (~3 min). Then, the cryopreservation tubes were carried to the hood, swiped with 70% ethanol, and the samples were pipetted into a separate 15 mL conical tube and mixed with 9 mL of pre-warmed (at 37° C) $PBS^{-/-}$ containing FBS at 2% (PBS+FBS). After a centrifugation at 300 x *g* for 5 min at 4° C, supernatant was removed until ~100 µL. Cells were resuspended by flicking the tube and by pipetting after adding 7.5 mL of pre-warmed (at 37° C) PBS+FBS. Cells were passed through a Falcon® 70 µm cell strainer (#352350, Corning Inc., NY, USA), and 10 µL were taken for cell counting, while the remaining samples were kept at 4° C. Cell counting was performed manually with 0.4% Trypan blue (#T8154, Sigma-Aldrich®, Merck KGaA, Darmstadt, Germany), in a 1:1 ratio, with the aid of a Neubauer chamber to calculate cell density.

An aliquot of 375,000 cells suspended in 375 µL of PBS+FBS was taken for flow cytometry analysis if the total number of cells was > 10.5 million cells. If feasible, a particular attempt was made to take an aliquot per subject type on each day, where the sample with the highest number of cells was selected. This aliquot would then be stored at 4° C until further processing (more details in *3.2.2.3. Flow cytometry analysis*).

Finally, samples were further centrifuged at 300 x *g* for 5 min at 4° C, and the supernatant was discarded. Then, 100 µL of PBS+FBS at RT were added, and cells were resuspended by pipetting and flicking the tube. Cells were then transferred to a 5 mL round bottom tube.

3.2.2.2. $CD4^+$ T cell enrichment and storage

For testing, two different $CD4^+$ T cell enrichment protocols were performed: (**I**) a positive selection, by first depleting monocytes with the MojoSort™ Human CD14 Selection Kit (#480025, BioLegend, San Diego, CA, USA), followed by a capture of labelled cells with the MojoSort™ Human CD4 Nanobeads (#480014, BioLegend, San Diego, CA, USA), as suggested by the manufacturer (Miller, 2016); and (**II**) a negative selection with the

MojoSort™ Human CD4 T Cell Isolation Kit (#480010, BioLegend, San Diego, CA, USA). After comparing the two options from the same parent sample, the MojoSort™ Human CD4 T Cell Isolation Kit yielded more events in flow cytometry and a larger percentage of viable $CD4^+$ T cells **(Table S27**, **Figure S7**, and **Figure S8).** Thus, the negative selection protocol was chosen for the processing of the samples.

The protocol I followed was adapted from the manufacturer's instructions with slight adaptions arising from a protocol already tested and validated by the Ana Espada de Sousa Lab at Instituto de Medicina Molecular João Lobo Antunes (Godinho-Santos *et al.*, 2020). Namely, 5 µL of the Biotin-Antibody Cocktail (Biotin anti-CD8a, CD11b, CD16, CD14, CD19, CD20, CD36, CD56, TCR-γδ, CD123, CD235ab) were added to the 100 µL resulting from the previous protocol (more details in *3.2.2.1. Sample transportation, thawing, and initial processing*), mixed, and incubated for 15 min on ice, on a shaker. Then, after resuspending the Streptavidin Nanobeads (vortex, maximum speed, five touches), 5 µL of the beads were added and mixed, followed by a 15 min incubation on ice with slow agitation.

Afterwards, 3 mL of RT PBS+FBS were added and mixed by pipetting. Samples were then placed on MojoSort™ Magnets (#480019, BioLegend, San Diego, CA, USA)/EasySep™ Magnets (#18000, STEMCELL Technologies, Vancouver, BC, Canada) for 5 min. After the incubation, the cells of interest were poured into a 15 mL conical tube. The previous process was then repeated twice more, for a total of three times.

Afterwards, these 15 mL tubes were centrifuged at 300 x *g* for 5 min at 4° C, and, after supernatant removal, cells were resuspended in 1.5 mL of PBS+FBS and transferred to 2 mL conical tubes. Subsequently, 10 µL were taken for counting, while cells were kept at 4° C. The labelled portions of the samples, not enriched for $CD4^+$ T cells, were collected into 1.5 mL conical tubes with 1 mL of PBS+FBS and kept at 4° C.

If the number of cells was > 2 million, particularly if an aliquot for flow cytometry analysis was taken in the beginning (more details in *3.2.2.1. Sample transportation, thawing, and initial processing*), 375,000 cells suspended in 375 µL of PBS+FBS were taken as an "output" aliquot and kept at 4° C until further processing (more details *in 3.2.2.3. Flow cytometry analysis*).

Subsequently, all conical tubes were centrifuged at 300 x *g* for 5 min at 4° C, supernatant was removed, and a freshly-made freezing solution at 4° C (80% PBS+FBS, 10% FBS, 10% DMSO) was added to each tube: 1 mL to the 1.5 mL tubes (without $CD4^+$ T cells)

and 1.5 mL to the 2 mL tubes (with $CD4^+$ T cells). Finally, samples were placed in a Nalgene® Mr. Frosty (Sigma-Aldrich®, Merck KGaA, Darmstadt, Germany) for a minimum of 4 h at -80° C until they were transported to long-term storage in an ultrafreezer, kept at the same temperature.

3.2.2.3. Flow cytometry analysis

I performed flow cytometry analysis to ensure the desired $CD4^+$ T cell enrichment. To interrogate the main cellular populations in the samples, different antibodies were used:

- anti-CD14, aimed at identifying monocytes, with the antibody CD14-eFluor450 (#48014941, clone 61D3, eBioscience™, Thermo Fisher Scientific, Waltham, MA, USA) at 1.6%;
- anti-CD19, aimed at identifying B cells, with the antibody CD19-Alexa Fluor 647 (#302222, clone HIB19, BioLegend, San Diego, CA, USA) at 2%;
- anti-CD3, aimed at identifying $CD4^+$ T cells, together with the anti-CD4 antibody described below, with the antibody CD3-PerCP-Cy5.5 (#45003741, clone OKT3, eBioscience™, Thermo Fisher Scientific, Waltham, MA, USA) at 1%;
- anti-CD4, aimed at identifying $CD4^+$ T cells, together with the anti-CD3 antibody described above, with the antibody CD4-BV711 (#317439, clone OKT4, BioLegend, San Diego, CA, USA) at 1%;
- anti-CD56, aimed at identifying NK cells, with the antibody CD56-PE-Cy7 (#25056741, clone CMSSB/NCAM, eBioscience™, Thermo Fisher Scientific, Waltham, MA, USA) at 0.4%.

The antibodies above were chosen according to their spectral profiles. Flow cytometry was performed in a BD LSRFortessa™ X-20 Cell Analyzer (BD Biosciences, Franklin Lakes, NJ, USA), after voltage and compensation parameters were appropriately set.

To infer the exact number of cells needed for the flow cytometry analysis and the optimal strategy of how cells should be fixed (before or after surface staining), several approaches were tested (**Figure S9**, **Table S28**, and **Figure S10**). Ultimately, fixing the cells prior to staining and a minimum of 300,000 cells was determined as optimal. Thus, in order

to account for cell loss in sample processing, 375,000 cells were used for flow cytometry analysis (more details in *3.2.2.1. Sample transportation, thawing, and initial processing* and *3.2.2.2. CD4+ T cell enrichment and storage*).

Initially, cells were frozen prior to fixation and stored in the same freezing solution described previously (more details in *3.2.2.2. CD4+ T cell enrichment and storage*). On the day of flow cytometry analysis, cells were quickly thawed, washed, fixed, and stained. Namely, cells were thawed in a pre-warmed water bath set at 37° C, centrifuged at 300 x *g* for 5 min at 4° C, and the supernatant was discarded. Then, cells were resuspended in 1 mL of pre-warmed (at 37° C) PBS+FBS and centrifuged at 300 x *g* for 5 min at 4° C, followed by supernatant removal. Finally, 100 µL of 2% paraformaldehyde (PFA) were added, followed by a 30 min incubation on ice. Afterwards, 1 mL of $PBS^{-/-}$ was added, followed by a centrifugation at 1000 x *g* for 5 min at 4° C, and supernatant removal until ~25 µL. Then, 25 µL of the antibody mixture was added, mixed by vortexing at a low speed, followed by an incubation protected from light for 30 min at 4° C. After the incubation, 1 mL of $PBS^{-/-}$ was added, followed by a centrifugation at 1000 x *g* for 5 min at 4° C, and the removal of the supernatant. Finally, cells were resuspended in 500 µL of $PBS^{-/-}$ and transferred to 5 mL round bottom tubes for flow cytometry analysis.

However, since cell loss was high and the number of recorded events in the flow cytometer was highly variable (**Table S29**), I switched to fixing the cells immediately after finishing the magnetic enrichment. Thus, the aliquots for flow cytometry taken previously (more details in *3.2.2.1. Sample transportation, thawing, and initial processing* and *3.2.2.2. CD4+ T cell enrichment and storage*) were centrifuged at 300 x *g* for 5 min at 4° C, followed by supernatant removal. Subsequently, cells were fixed with 2% PFA, washed with $PBS^{-/-}$, and left at a volume of 25 µL at 4° C, as described above. The aforementioned staining protocol was performed on the day of the flow cytometry analysis. Samples were left unanalyzed for a maximum of five days for freshly fixed samples and 11 days for previously frozen samples. Out of a total of 132 suitable aliquots analyzed through flow cytometry, only 23 samples were previously frozen (**Table S29**).

Flow cytometry acquisition was performed until 30,000 events or 10 min had passed (whichever occurred first). Rate was kept at “low” unless < 10,500 events were acquired at 3.5 min; if so, the rate was increased to “medium”.

Regarding the gating strategy (**Figure S11** and **Figure S12**), FSC-A *vs.* FSC-H followed by SSC-A *vs.* SSC-H, were used to select single cells, followed by a gating on $CD3^+$ cells, then $CD4^+$ cells, and, finally, on $CD19^-$ and $CD56^-$ cells. CD14 was excluded from the gating strategy, as $CD3^+CD4^+CD14^+$ cells likely represent doublets of $CD4^+$ T cells associated with monocytes and not a distinct cell lineage (Burel *et al.*, 2019). Also, I did not take into account cell viability in the flow cytometry analysis, since mtDNA would be extracted from the samples without previous isolation of live cells and no functional assays would take place.

Four different batch gating strategies with FlowJo™ v10 (BD Biosciences, Franklin Lakes, NJ, USA) were adopted: (**I**) cells frozen prior to fixation, processed in 2018 (**Figure S11A–B**); (**II**) cells freshly fixed, processed in 2018 (**Figure S11C–D**); (**III**) cells freshly fixed, processed in 2019, before a scheduled maintenance on the BD LSRFortessa™ X-20 Cell Analyzer on August 12, 2019, where a blue laser was substituted and the flow cell was optimized (**Figure S12A–B**); and (**IV**) cells freshly fixed, processed in 2019, after the scheduled maintenance (**Figure S12C–D**).

3.2.2.4. DNA extraction, enriched in mtDNA

For DNA extraction enriched in mtDNA, I used a very similar protocol to the one for the stroke clinical cohort (more details in *3.1.1. Sample acquisition/collection*), without the need to collect the buffy coat. Thus, samples were thawed with a Thermomixer® (Eppendorf, Hamburg, Germany) set at 37° C until cells were nearly thawed. Then, the cryopreservation tubes were carried to the hood, swiped with 70% ethanol, and 1/3 of the volume (500 µL) was taken from each sample and transferred to a new 15 mL conical tube. The remaining volume was kept at 4° C.

Subsequently, the protocol followed the one described previously (more details in *3.1.1. Sample acquisition/collection*), after the addition of 200 µL of QIAGEN Protease, 1.5 mL of $PBS^{-/-}$, and 2.4 mL of Buffer AL. After DNA was quantified and its quality checked (more details in *3.1.1. Sample acquisition/collection*), DNA was stored at -20° C. The remaining sample volume was placed in a Nalgene® Mr. Frosty (Sigma-Aldrich®, Merck KGaA, Darmstadt, Germany) for a minimum of 4 h at -80° C until they were transported to long-term storage in an ultrafreezer, kept at the same temperature.

As in other protocols, sample processing was performed at the same time for each triplet (PwMS at V1&V2 + HC), to minimize differences within processing. In 2018, for the pilot study (more details in *3.2.1. Cohort description*), samples were processed in ascending number of cells after $CD4^+$ T cell enrichment, while in 2019 samples were processed in the same order as their magnetic enrichment.

The enrichment of mtDNA was confirmed through a qPCR on other samples not included in the cohort, which, however, underwent the same magnetic enrichment and DNA extraction protocols (**Table S31**, **Figure S13**, and **Figure S14**). For that, two sets of primers were used: (**I**) aimed at GAPDH (Uhm *et al.*, 2012; Liu, Tsao and Hou, 2015) — 5'-AGGGCTGCTTTTAACTCTGGT-3' (forward) and 5'-CCCCACTTGATTTTGGAGGGA-3' (reverse) —; and (**II**) aimed at mtDNA (Siddiqui *et al.*, 2012; Jansen *et al.*, 2015) — 5'-CATGCAAGCATCCCCGTTCC-3' (forward) and 5'-CTGTTTCCCGTGGGGGTGTG-3' (reverse). Each sample was run in triplicate, with a no template control (NTC) for each primer set. Each non-NTC well consisted of 0.7 µL of a forward primer (10 µM), 0.7 µL of a reverse primer (10 µM), 5 µL of 10 ng of DNA resuspended in DNAse/RNase-free water, and 3.6 µL of iTaq™ Universal SYBR® Green Supermix (#1725124, Bio-Rad Laboratories, Inc., Hercules, CA, USA), in an Applied Biosystems™ MicroAmp™ Optical 384-Well Reaction Plate with Barcode (#LS4309849, Thermo Fisher Scientific, Waltham, MA, USA) with a compatible optical adhesive film. Finally, plates were centrifuged at 900 x *g* for 5 min at RT and analyzed on an Applied Biosystems® ViiA™ 7 Real-Time PCR System (Thermo Fisher Scientific, Waltham, MA, USA) for 40 cycles, with a denaturing temperature of 95° C, and an annealing temperature of 58° C.

3.2.3. DNA sequencing, bioinformatic processing, and data analysis

NGS mtDNA sequencing and data analysis followed the protocols described previously (more details in *3.1.2. DNA sequencing* and *3.1.4. Data analysis*), with the addition of the package rstatix (Kassambara, 2021).

Regarding bioinformatic processing, the PCP pipeline was used (Cortes-Figueiredo *et al.*, 2021) (more details in *3.1.3. Bioinformatic processing*) with mutserve v2 (Weissensteiner, Forer, *et al.*, 2016; Weissensteiner *et al.*, 2021) for variant calling, HaploGrep v2.4.0 (Weissensteiner, Pacher, *et al.*, 2016) for haplogroup calling, and Haplocheck v1.3.3 (Weissensteiner *et al.*, 2021) for a contamination check.

To account for false positives, in accordance with (Cortes-Figueiredo *et al.*, 2021) (more details in *4.1. PrecisionCallerPipeline*), variants with a VL $\geq$ 10% were only accepted if they were found with both PCP and TSS, while variants with a VL $<$ 10% only present in PCP were only accepted if: (**I**) the normalized coverage, coverage ratio, mean value of reported NUMTs, and the distance to the amplicon's edge was not lower, higher, higher, or lower, respectively (more details in **Table 4**), than their corresponding outliers for variants found in both PCP and TSS, calculated through the Tukey's fences method (Tukey, 1977); and (**II**) variants were previously reported in either the HelixMTdb database (Bolze *et al.*, 2020), the ClinVar database (until October 2021) (Landrum *et al.*, 2018) through MitImpact 3D 3.0.7 (Castellana *et al.*, 2021), or in the MITOMAP database (Lott *et al.*, 2013). Variants only present in TSS were excluded.

Page intentionally left blank

Part of the research data presented herein has already been published (Cortes-Figueiredo et al., 2021).

IV — RESULTS

Page intentionally left blank

4.1. PrecisionCallerPipeline

4.1.1. Pipeline validation with samples from the 1000 Genomes Project

Three DNA samples were acquired, which were previously sequenced within the 1000 Genomes Project (1000 Genomes Project Consortium *et al.*, 2015) on Illumina HiSeq (Illumina, Inc., San Diego, CA, USA), with three different sequencing runs (more details in *3.1.1. Sample acquisition/collection*): (**I**) exome sequencing — exome (mean coverage: 420.59 reads/bp); (**II**) low-coverage sequencing — lowCov (mean coverage 3863.37 reads/bp); and (**III**) high-coverage sequencing — highCov (mean coverage: 15079.28 reads/bp). Samples were then processed and mixed at five different mixtures levels, ranging from 1 to 25% (**Figure 19A**), and sequenced on Ion Torrent™ Ion S5™ (Thermo Fisher Scientific, Waltham, MA, USA).

The original sequenced samples and their mixtures then underwent bioinformatic analysis with the PCP pipeline (**Figure 19B**). For that, I aimed to optimize the protocol for the heteroplasmy threshold and the removal of NUMTs. Correspondingly, the samples were run with three different NUMT removal approaches and 21 thresholds, ranging from 0.4% to 10% (**Figure 19C**), through the mutserve variant caller, an offline and command-line version of mtDNA-Server (Weissensteiner, Forer, *et al.*, 2016; Weissensteiner *et al.*, 2021). After data analysis and variant classification for primary and mixture samples (**Table S1**), I determined the performance metrics (**Figure S1**) considering both Grade A variants, which were homoplasmic in the three independent sequencing runs, and Grade B variants, which were heteroplasmic in the same three sequencing runs (more details in *3.1.4. Data analysis*). The primary outcome was the highest possible F_1 score.

NUMT removal immediately after the first alignment and before merging the BAM files (**Figure 19B**) was superior to the other NUMT removal approaches (**Table S2**), while the 2.5% heteroplasmy threshold with NUMT removal before merging yielded the highest mean F_1 score between the four different datasets — Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B — (**Table S3**).

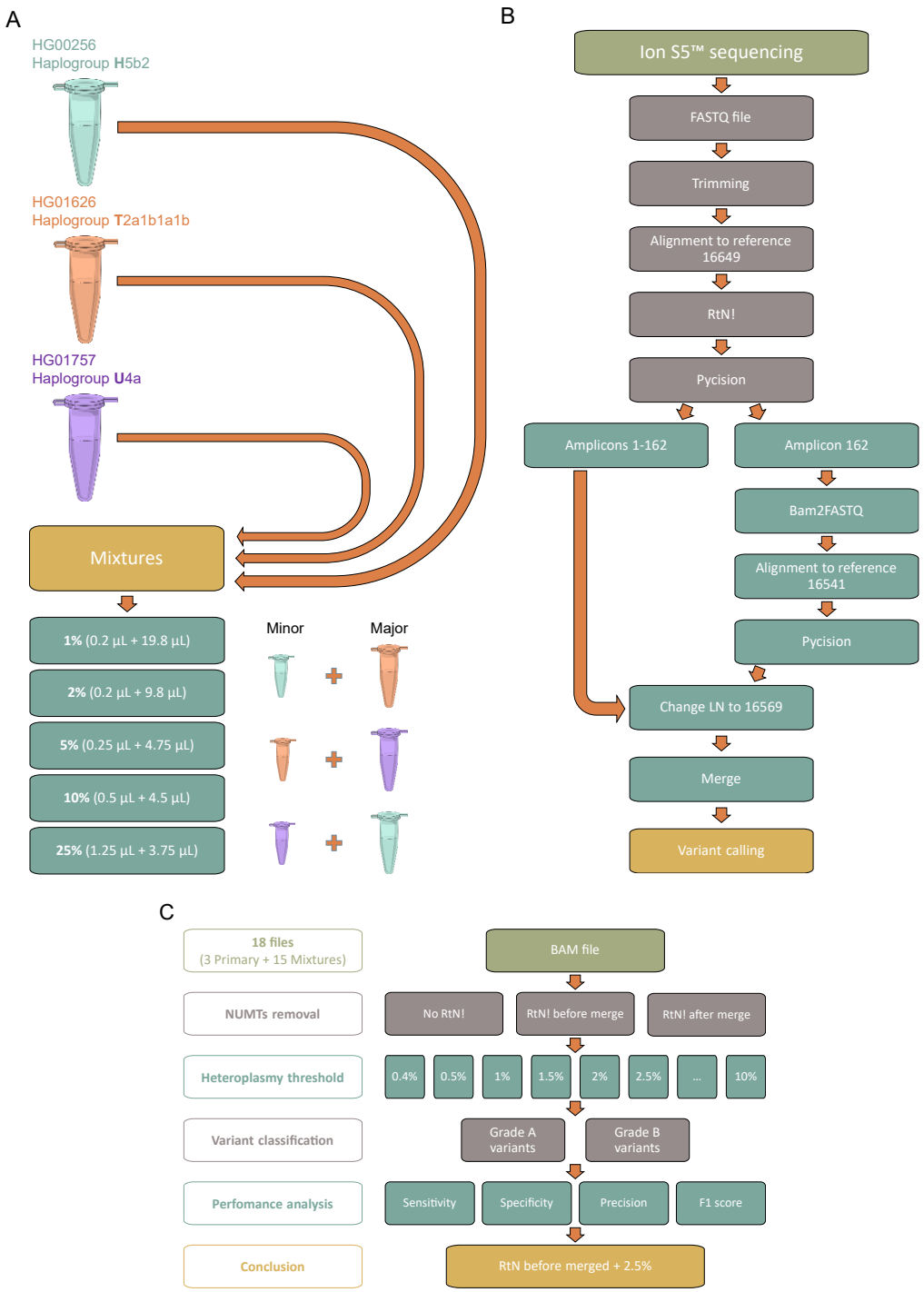


Figure 19 — Schematic representation of the workflow

(**A**) Wet lab approach to performing the 15 mixtures from the Coriell Institute for Medical Research's samples, which have been analyzed within the 1000 Genomes Project — templates adapted from (*SMART*, 2021) (CC BY 3.0); (**B**) Visual representation of the PrecisionCallerPipeline (PCP); (**C**) Visual representation of the optimization approach to NUMT removal and heteroplasmy threshold. Abbreviations: NUMTs — nuclear insertions of mitochondrial DNA.

Correspondingly, all BAM files were then run through the mutserve variant caller with a 2.5% threshold and NUMT removal before merging (**Figure 19B**). For simplicity, samples

are denoted by their simplified haplogroup identifier. Thus, HG00256 as H, HG01626 as T, and HG01757 as U. Similarly, names for mixtures follow an analogous pattern with the minor component being followed by its absolute mixture level and the major component. Therefore, *U0.01H* is sample HG01757 (U) at 1% mixed with sample HG00256 (H) at 99%.

Genome coverage and mappability, defined as the ability to read a bp above a certain coverage threshold, were adequate in comparison to other sequencing runs in the primary analysis. Haplogroup assignment was also homogeneous for both PCP and the output from TSS (**Table S4** and **Figure 20**).

Interestingly, PCP had an increased number of mutations in comparison with TSS (**Figure 20B**), despite a reduction in the mean coverage in PCP (**Figure 20A**). To better understand this phenomenon, I considered all samples analyzed on the Ion Torrent (**Table S5** and **Figure 21**).

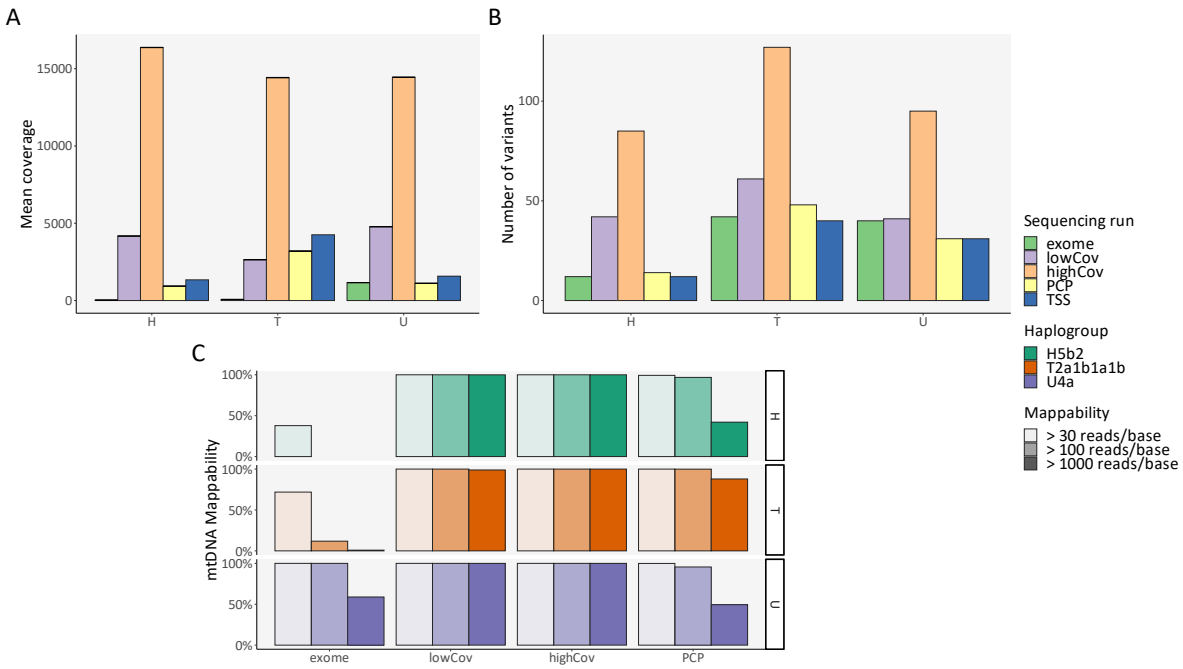


Figure 20 — Primary analysis: Overall look

Variant threshold was set at 0.4% for exome, lowCov, and highCov, and 2.5% for PCP. (**A**) Mean coverage per sample and per sequencing run/pipeline; (**B**) Number of variants per sequencing run/pipeline; (**C**) Mappability and haplogroup assignment per sequencing run/pipeline (TSS is not shown since it does not output raw data). Error bars denote standard error of the mean (SEM). Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

In PCP, coverage and mappability were uniform, haplogroup assignment was successful for mixtures up to 10%, and sample contamination was correctly detected, with

5% as the lowest detection limit (**Table S5** and **Figure S2**). TSS correctly identified haplogroups up to 10% as well, albeit with errors in the contamination detection for unmixed samples U and T (lower than TSS's own limit of ~10%), and a higher contamination detection limit of 10% (**Table S5**).

In comparison with TSS, samples processed with PCP have a reduced total number of sequences (**Table S5** and **Figure 21A**) — mean difference of 16.94%, 95% CI [15.98–17.89%], an adjusted *p*-value of 1.43 x 10^{-17}, and a significantly lower mean coverage (**Table S5** and **Figure 21B**) — mean difference of 25.83%, 95% [24.97–26.70%], an adjusted *p*-value of 3.87 x 10^{-21}; paired t-tests adjusted with false discovery rate (FDR).

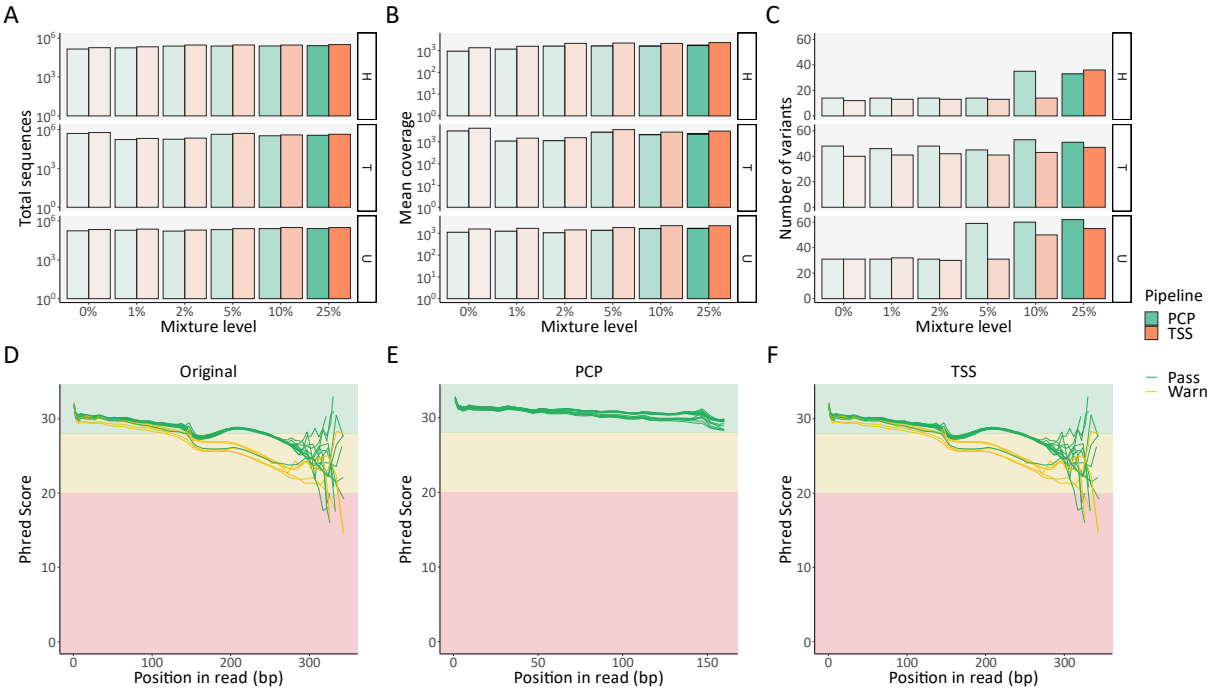


Figure 21 — Ion Torrent sequencing: Overall look at differences between PCP and TSS

A 0% mixture level denotes unmixed samples. (**A**) Total number of sequences per pipeline and major sample contributor; (**B**) Mean coverage per pipeline and major sample contributor; (**C**) Number of variants per pipeline and major sample contributor; (**D–F**) MultiQC's Phred score per base pair in the unprocessed FASTQ files, PCP's BAM to FASTQ files, and TSS's BAM to FASTQ files, respectively. Abbreviations: bp — base pair; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

This arises from the protocol for read trimming, NUMT removal, and quality control, since, in comparison to the unaltered FASTQ files, the BAM files from TSS have the exact same number of reads and Phred score patterns (**Figure 21D** and **Figure 21F**).

In comparison with PCP, these results indicate that TSS achieves a higher coverage at the expense of a lack of read selection from the initial FASTQ files. Despite having significantly reduced coverage and fewer reads, more variants are called with PCP than with TSS (**Table S5** and **Figure 21C**) — mean difference of 5.83, 95% CI [1.95–9.72], an adjusted *p*-value of 5.60×10^{-3}; paired t-tests with FDR.

To further dissect this difference in the number of variants, I first looked at the variants in the primary analysis (**Figure S3A**, and **Table S6–Table S9**).

When comparing variants found with the Ion Torrent sequencing method to the ones previously found on the other three Illumina runs, we observe that: (**I**) PCP significantly increases the proportion of correctly found Grade A/B variants (true positives) in contrast to lost Grade A/B variants (false negatives), while no Grade A/B lost variant with PCP was picked up by TSS (**Table S7**); (**II**) TSS significantly overestimates the VL in found Grade A/B variants in comparison with PCP (**Table S8**); (**III**) despite PCP picking up more variants than TSS, the difference in the proportion of "novel" variants (false positives) was not statistically significant between the two pipelines (**Table S9**).

We then looked at the variants in the mixture analysis (**Figure 22**, **Figure S3B–H**, **Table S10–Table S16**, **Figure S4**, **Table S17–Table S24**). Samples with a macro classification of "Mixed status" (more details in **Table S1**) were excluded because of the impossibility of determining their origin.

Similar to the primary analysis, Grade A/B variants were best retrieved with PCP, in comparison with TSS, with a single variant found by TSS and lost with PCP (**Table S11**). Moreover, true positive variants below 11% are completely absent in TSS, while PCP correctly picks up variants until its threshold of 2.5% (**Figure 22** and **Table S10**).

Regarding the difference between the expected and the observed VL, I defined two different approaches:

- **External comparison** — the difference to the other Illumina sequencing methods (exome, lowCov, and highCov);
- **Internal comparison** — the difference within Ion Torrent, taking the primary sequencing as a reference.

Consequently, I also considered two different relative differences, where I divided each difference by its expected level in order to observe a percent variance per VL (**Figure**

22A, **Figure S3C–H**). When performing a paired comparison regarding differences for Grade A/B variants (**Table S12**), I observed no significant differences between TSS and PCP for raw VL differences taking found variants into consideration. However, considering lost variants, TSS performed worse in the external comparison (**Table S12**). Regarding relative differences (**Table S13**), no statistically significant differences between PCP and TSS were found.

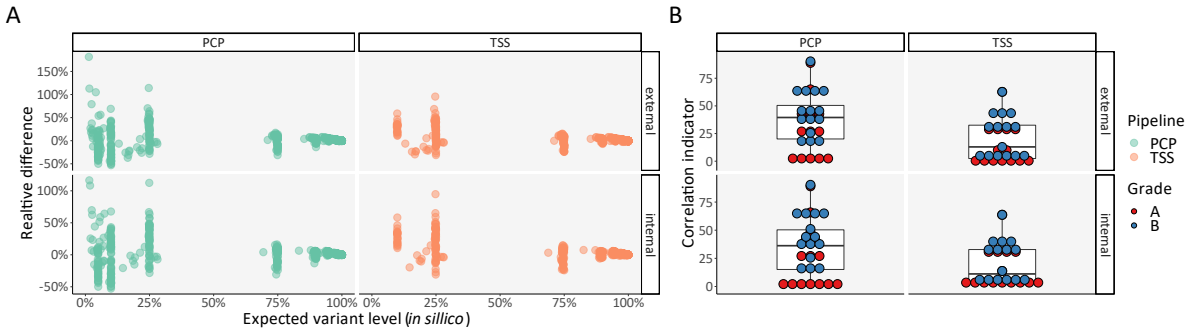


Figure 22 — Mixture analysis: Similarity in Grade A/B variants between expected and observed variant levels

(**A**) Relative difference between expected and observed variant level, per pipeline, and per internal and external comparison; (**B**) Correlation indicator for internal and external comparisons, per pipeline, and per variant grade. Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

In order to assess the similarity between the expected and the observed VLs (internal and external), I also performed linear models per mixture, considering Grade A/B variants for each pipeline, for each comparison (internal and external), and for each grade variant classification (A or B), as they might have varied between them. After extracting the adjusted R^2 and the adjusted *p*-value with FDR, I created an indicator for correlation strength determined by the log10 of the ratio between the adjusted R^2 and the adjusted *p*-value. Thus, a greater value would indicate a stronger correlation between both variables. PCP showed a significantly increased correlation indicator in both comparisons (internal and external) for Grade B variants, while no significant differences for Grade A variants were found (**Figure 22B** and **Table S14**).

Since PCP appeared to have increased sensitivity in comparison with TSS, I then looked at the prevalence of novel variants (false positives). Interestingly, no differences

were found (**Table S15**), albeit with TSS showing novel variants at a higher VL, namely, novel variants already present in the primary analysis (**Table S16**).

Considering the location of these novel variants in both analyses (primary and mixture), PCP had one mutation in 10/18 samples (4318T), one mutation in 8/18 samples (8649C), while the remaining four were isolated (2463G, 5752G, 6698G, 8152A). TSS had two mutations shared in 10/18 samples (310C, 10958C), one shared in 4/18 samples (14777C), one shared in 3/18 samples (8249C), and one shared in 2/18 samples (318C).

Looking into the different variant classifications (found, lost, novel, and primary novel variants), several conclusions can be made: (**I**) PCP showed an increased rate (adjusted to the size of each region) of found variants throughout the genome, while maintaining a lower rate of lost variants (**Figure S4A–B**, **Table S17**, and **Table S18**); (**II**) novel variants (primary and mixture) showed the lowest normalized coverage (normalized for the mean coverage of each sample in **Table S4** and **Table S5**) for both PCP and TSS, while the highest strand bias, calculated as a coverage ratio (forward vs. reverse), was observed in mixture novel variants for PCP, and in primary novel variants for TSS (**Figure S4C–E**, **Table S19**, and **Table S20**); (**III**) since false variants might arise from NUMTs, I calculated the mean value of reported NUMTs based on two different databases (Weissensteiner, Forer, *et al.*, 2016; Smart *et al.*, 2019) for each position — the highest rate of NUMTs was observed in primary novel variants for PCP, and in found variants for TSS (**Figure S4F** and **Table S21**); (**IV**) regarding the proportion of variants in positions classified as a low-complexity region (LCR) (Weissensteiner, Forer, *et al.*, 2016), lost variants showed the highest proportion of LCR variants for PCP, while the same was true in primary novel variants for TSS (**Figure S4G** and **Table S22**); (**V**) finally, I tested the hypothesis that the distance to the "callable" extremity of each amplicon might be significantly different, depending on each variant class. The lowest distance was observed in primary and mixture novel variants for PCP, while the same was true for primary novel variants alone for TSS, although less consistently (**Figure S4H** and **Table S23**). **Table 4** offers a summary for novel variants (false positives).

Ultimately, I performed a performance analysis for four different datasets: Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B, with separate paired statistical tests for each dataset (**Figure S5** and **Table S24**). PCP achieved a higher sensitivity and F_1 score in both the primary and mixture analyses for Grade B variants, without performing worse than TSS in the other indicators or in the other datasets (**Table S24**).

In order to further validate these findings, I performed the same analyses with two other variant callers (freebayes (Garrison and Marth, 2012) and varscan 2 (Koboldt *et al.*, 2012)), which yielded overall overlapping results (**Table S6–Table S24**).

Table 4 — Novel variants (false positives): Summary characteristics in the mutserve variant calling method

Variable	PCP		TSS	
	Primary novel variants	Mixture novel variants	Primary novel variants	Mixture novel variants
Normalized coverage	↓	↓↓	↓↓↓	↓↓
Coverage ratio	-	↑↑↑	↑↑↑	↑
Number of NUMTs	↑↑↑	↓	-	↓↓
LCR prevalence	↓↓↓	↓↓↓	↑↑↑	-
Distance to amplicon edge	↓↓↓	↓↓	↓	-

Using found variants (true positives) as a reference: "-" denotes nonsignificant changes; "↑" or "↓" denote significant changes in the 25–50% range; "↑↑" or "↓↓" denote significant changes in the 51–75% range; and "↑↑↑" or "↓↓↓" denote significant changes > 75% (more details in **Table S19–Table S23**). Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; NUMTs — nuclear insertions of mitochondrial DNA; LCR — low-complexity region.

4.1.2. Pipeline comparison with an independent set of 50 clinical samples

As a way of demonstrating the added value of the PCP pipeline, I selected 50 independent samples from a prospective study in patients with ischemic stroke (more details in *3.1.1. Sample acquisition/collection)* and analyzed them in parallel with TSS and PCP.

In PCP, genome coverage followed a similar pattern in all samples (**Figure 23A**), with an overall uniform mappability (**Table S25**). The genome coverage was also in accordance with what is typical in Ion Torrent™ sequencing with the Precision ID mtDNA Whole Genome Panel (Cihlar, Amory, *et al.*, 2020). The haplogroup was identical for both pipelines in 45/50 samples (**Table S25**). In PCP, three samples were flagged for contamination in contrast to TSS, which had 11 samples, 10/11 were well below its ~10% threshold (**Table S25**). In PCP, the number of variants varied extremely, from 12 to 503, while in TSS, they ranged from 10 to 80 (**Table S25**). Thus, I eliminated Samples #21, #37, and #44 because of

their contamination status, and Samples #2 and #3 because of their extremely high number of variants in PCP, from the subsequent analysis.

Consistent with what I described previously, while mean coverage was significantly decreased in PCP (**Figure 23B**) — mean difference of 30.87%, 95% CI [29.17–32.58%], an adjusted *p*-value of 3.01 x 10^{-34}, a higher number of variants with the PCP pipeline was still observed (**Figure 23D**) — mean difference of 15.11, 95% [10.07–20.16], an adjusted *p*-value of 2.97 x 10^{-7}; paired t-tests adjusted with FDR.

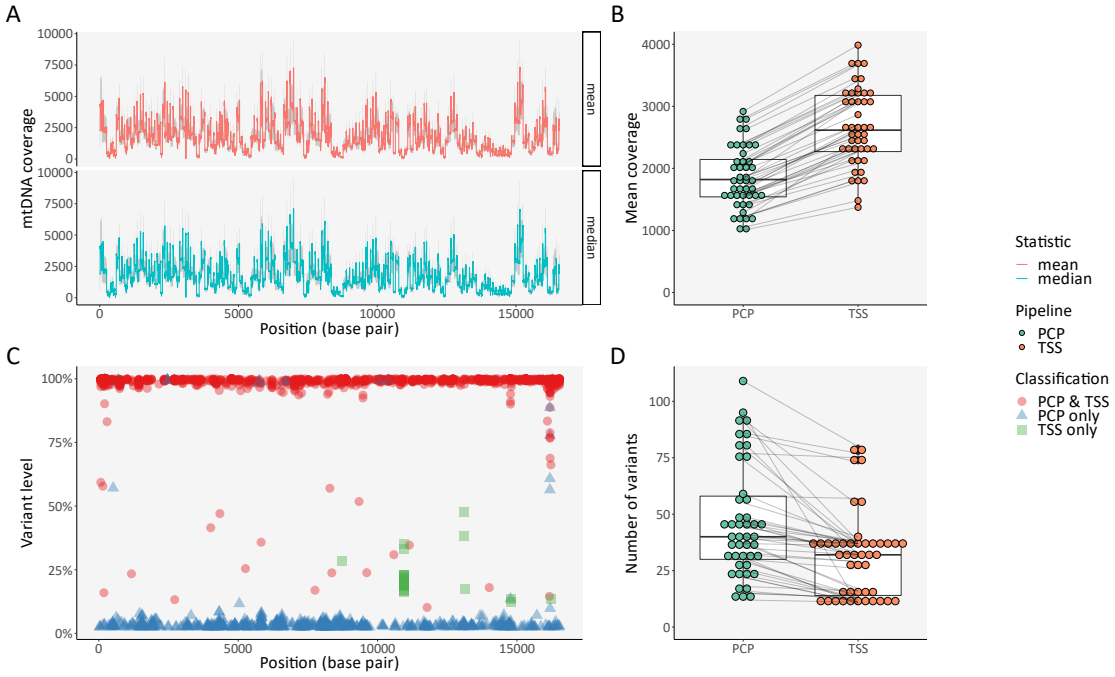


Figure 23 — Comparison between PCP and TSS in an independent set of samples

Data was obtained from an ischemic stroke prospective cohort (after filtering, *N* = 45). (**A**) Coverage per base pair with PCP data, per statistic (mean and median) — the grey area represents the standard deviation of each statistic per base pair; (**B**) Mean coverage per sample and per pipeline — lines connect the same sample; (**C**) Distribution of variants per base pair and per variant classification — for variants found in PCP and TSS, mean variant level was plotted; (**D**) Mean number of variants per sample and per pipeline — lines connect the same sample. Abbreviations: *N* — number; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

This difference mostly derives from an increase in variants below TSS's heteroplasmy threshold (**Figure 23C**): variants only present in PCP have a mean VL of 7.45%, while variants only present in TSS have a mean VL of 22.5%. Interestingly, PCP finds 98.31% of TSS's variants, while TSS only finds 57.92% of PCP's variants. Although not exclusively, when we consider discordant variants (variants not present in both pipelines) with more than one

occurrence (**Table S26**), we observe similar patterns as false positives in the previous analyses (**Figure S4C–H** and **Table S19–Table S23**). In PCP, for example, we encounter the same variants seen in the previous analysis: 8649C (present in 25/45 samples) and 4318T (present in 20/45 samples). These two samples have a mean VL <5%, a low mean normalized coverage, and a small distance to the amplicon's "callable" edge. Similarly, in TSS, we observe variants that were also previously found: 10958C (present in 17/45 samples) and 14777C (present in 2/45 samples). These two variants have a low mean normalized coverage, without other discernible features.

Regardless of the discordant variants, in the variants found in both pipelines, there is a statistically significant correlation between the VL with TSS, and with PCP (**Figure S6**) — adjusted R^2 of 0.97 and a *p*-value < 2.2×10^{-16}; linear regression model.

4.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort

4.2.1. Cohort description

The cohort consists of 61 PwMS with clinical visits at six months (V1) and 36 months (V2) after disease onset, with an average age of 33 years, where females account for 66% of the cohort. HC were matched to PwMS according to sex and a maximum age difference of 5 years. Most patients presented with RRMS at both clinical visits (**Figure 24**).

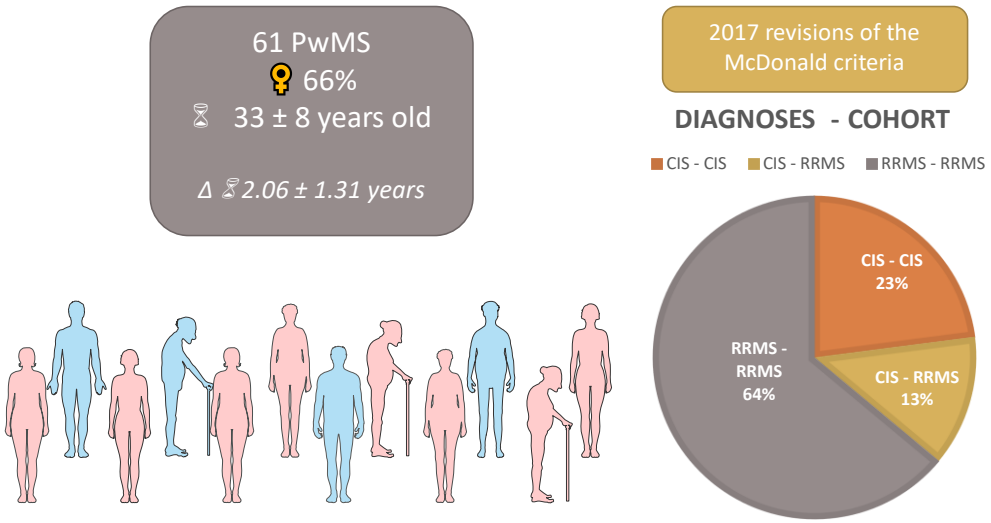


Figure 24 — Multiple Sclerosis clinical cohort

Females account for 66% of subjects and the average age is 33 ± 8 years. The average age difference between PwMS at V1 and HC is 2.06 ± 1.31 years — templates adapted from (*SMART*, 2021) (CC BY 3.0). Abbreviations: CIS — Clinically Isolated Syndrome; PwMS — patients with Multiple Sclerosis; RRMS — Relapsing-Remitting Multiple Sclerosis.

As mentioned previously (more details in *3.2.1. Cohort description*), a pilot study with 20 triplets (20 PwMS at V1&V2 + 20 age- and sex-matched HC) was performed to determine the appropriateness of the sample size through the Dupont method (Dupont, 1988). Following WGS data analysis of this subset, a 35% discordance in prevalence was detected for deleterious variants in Complex I — 55% for PwMS at V1 (discordant prevalence: 45%) *vs.* 20% for HC (discordant prevalence: 10%). Hence, the sample size was adequate according to this endpoint.

4.2.2. $CD4^+$ T cell enrichment, DNA extraction, and NGS sequencing quality

Initially, a magnetic enrichment in $CD4^+$ T cells was performed (more details in *3.2.2. Sample processing*), where a wide range of the number of cells and cell mortality was obtained, for all subject types (**Table 5** and **Table S30**). In comparison with HC, PwMS at V1 showed a lower number of cells, in millions — mean difference of 0.65, 95% CI [0.16–1.14], an adjusted *p*-value of 3.66×10^{-2}; paired t-test adjusted with FDR. In comparison with the V2 visit, cells from PwMS at V1 had lower mortality — mean difference of 2.77%, 95% CI [0.56–4.97%], an adjusted *p*-value of 3.67×10^{-2}; paired t-test adjusted with FDR.

Table 5 — Magnetic enrichment and DNA extraction: Summary

Subject type	Number of cells after enrichment (millions)	Mortality	DNA (Synergy HTX) (ng/µL)	260/280 ratio	DNA (Qubit®) (ng/µL)
HC	3.03 [0.37–8.92]	15.15% [4.34–70.86%]	20.04 [1.45–65.24]	1.87 [1.19–2.62]	13.78 [1.91–56.00]
V1	2.38 [0.57–6.40]	14.97% [5.57–33.41%]	15.59 [2.93–45.33]	1.89 [1.19–3.37]	10.43 [2.54–37.00]
V2	2.48 [0.10–8.10]	17.73% [7.43–56.25%]	15.94 [3.48–53.46]	1.93 [1.15–5.20]	10.54 [0.70–27.20]
All	**2.63 [0.10–8.92]**	**15.95% [4.34–70.86%]**	**17.19 [1.45–65.24]**	**1.90 [1.15–5.20]**	**11.58 [0.70–56.00]**

Numbers show the mean for each value followed by the minimum and maximum values between brackets.

Abbreviations: HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Nonetheless, following the magnetic enrichment, a significant increase in $CD4^+$ T cells ($CD3^+$ $CD4^+$ $CD56^-$ $CD19^-$) was achieved (**Figure 25A, Table 6,** and **Table S29**) — mean fold change in paired samples of 2.59, 95% CI [2.31–2.87], a *p*-value of 1.66×10^{-15}; paired t-test. Considering both paired and unpaired samples, the mean prevalence of $CD4^+$ T cells

raised from 29.68% (range: 14.27–53.84%) to 67.38% (range: 32.78–91.39%), without any influence from subject type (**Table 6**).

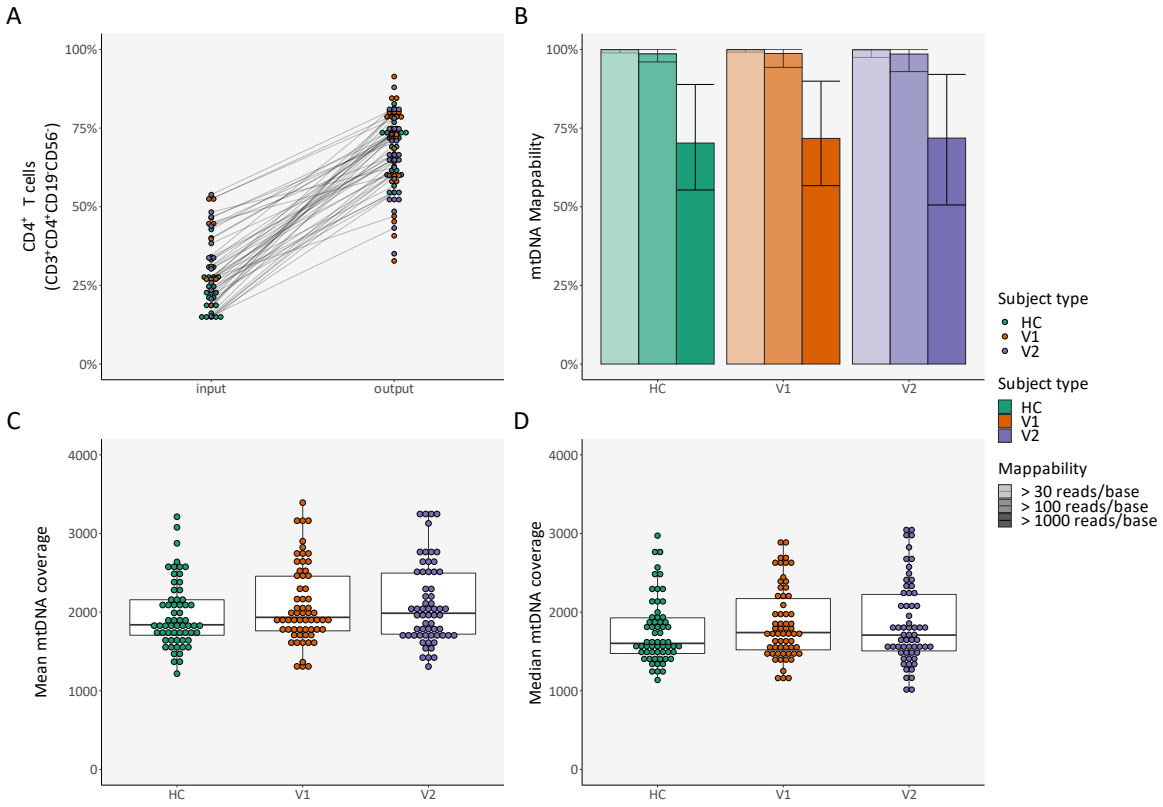


Figure 25 — CD4⁺ T cell enrichment and mitochondrial DNA next-generation sequencing

(**A**) Percentage of CD4⁺ T cells (CD3⁺ CD4⁺ CD56⁻ CD19⁻) before and after magnetic enrichment, as assessed with flow cytometry — lines connect the same sample; (**B**) Mean mappability, per subject type and per definition — error bars denote minimum and maximum values; (**C**) Mean mtDNA coverage, per subject type; (**D**) Median mtDNA coverage, per subject type. Abbreviations: HC — healthy control; mtDNA — mitochondrial DNA; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Relevant exceptions to the gating strategy in 2019 (**Figure S12**) are highlighted in **Table S29**, namely:

- On May 23, 2019, the MojoSort™ Human CD14 Selection Kit was used by mistake instead of the MojoSort™ Human CD4 T Cell Isolation Kit. Cells were kept overnight at 4° C and underwent the proper CD4⁺ T cell enrichment protocol the next day. Even though CD4⁺ T cells increased afterwards, CD4 signal was significantly reduced in these samples. Thus, these samples were excluded from the flow cytometry analysis;

- On July 25, 2019, one of the seven analyzed samples had an inexplicably low CD4 signal. This sample was excluded from the flow cytometry analysis;
- On August 23, 2019, the samples that I analyzed fit better with the batch gating strategy aimed at cells processed prior to the scheduled maintenance on August 12, 2019. I used the previous gating strategy and these samples were included in the flow cytometry analysis.

Table 6 — Flow cytometry analysis: Summary

Subject type	Input or output?	$CD4^+$ T cells ($CD3^+CD4^+CD19^-CD56^-$)	*N*
HC	input	23.79%	*25*
HC	output	67.99%	*30*
V1	input	37.10%	*12*
V1	output	68.71%	*28*
V2	input	34.51%	*12*
V2	output	65.15%	*25*
All	**input**	**29.68%**	***49***
All	**output**	**67.38%**	***83***

Abbreviations: HC — healthy control; *N* — number; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Subsequently, following the extraction of DNA enriched in mtDNA (more details in *3.2.2.4. DNA extraction, enriched in mtDNA*), variable amounts of DNA were achieved (**Table 5** and **Table S30**). In comparison with PwMS at V1, HC had higher yields of extracted DNA, in ng/µL — mean difference of 4.45 for the measurement with Synergy HTX, 95% CI [1.06–7.84], an adjusted *p*-value of 3.66×10^{-2}, and mean difference of 3.35 for the measurement with Qubit®, 95% CI [0.80–5.90], an adjusted *p*-value of 3.66×10^{-2}; paired t-tests adjusted with FDR.

There is a significant correlation between the two DNA measurements (**Figure S15A**) — adjusted R^2 of 0.88 and a *p*-value < 2.2×10^{-16}, linear regression model; and between the number of cells used for DNA extraction and DNA concentration (**Figure S15B**) — adjusted R^2 of 0.77 and a *p*-value < 2.2×10^{-16}, linear regression model. However, DNA concentration had no effect on mtDNA coverage (**Figure S15C–D**), with most variability appearing to arise from each sequencing run (**Figure S15E–F**). Regarding differences between different subject types (**Figure 25B–D** and **Figure S15E–F**), HC's significantly lower mean and median

mtDNA coverages (mean differences: 122.98 reads/bp and 102.62 reads/bp, respectively) did not remain after adjusting for multiple comparisons — initial *p*-values 1.07 x 10^{-2} and 2.44 x 10^{-2}; adjusted *p*-values 1.92 x 10^{-1} and 2.20 x 10^{-1}, respectively; paired t-tests adjusted with FDR.

Notable issues with DNA extraction were (samples highlighted in **Table S30**):

- On July 19, 2018, although no obvious procedural errors took place, four out of 12 samples had a lower volume than the expected 300 µL;
- On June 5, 2019, three out of 30 samples were used in full, as they had lower cell numbers than the minimum previously registered in 2018, namely 70,000 cells. Thus, all 1.5 mL of the sample (instead of 500 µL) were used and only 500 µL of $PBS^{-/-}$ (instead of 1.5 mL) were added;
- On July 14, 2019, one sample out of 30 was used in full, as it had a lower cell number than the minimum previously registered, as mentioned previously;
- On July 14, 2019, one sample fell to the floor and the tube broke in the midst of the DNA extraction; 500 µL more were taken from the original sample and the protocol was resumed;
- On August 26, 2019, three samples out of 33 were used in full, as they had lower cell numbers than the minimum previously registered, as mentioned previously.

Taking into account all samples analyzed through NGS sequencing (**Table S32**), no contamination was detected and all haplogroups corresponded to European lineages. PwMS at both visits had the same haplogroup with two minor exceptions (Triplets #5 and #48), albeit without changing their simplified haplogroup (**Table S32**).

4.2.3. Cross-sectional comparison

No significant differences in the total number of variants were found between different subject types, i.e., between HC and PwMS at V1 and between V1 and V2 for PwMS (**Figure 26** and **Table S32**) — paired t-tests adjusted with FDR. Variant distribution was also bimodal in all subject types (**Figure 26A**), likely arising from each sample's haplogroup (**Figure 26B–C**), since the mean number of variants in HC and PwMS at V1 is significantly

different in haplogroups with at least three samples — *p*-value 3.34 x 10^{-16}; Kruskal-Wallis test. Haplogroups H and HV had significantly fewer variants (**Table S33**), which is consistent with rCRS's own H haplogroup (Bandelt *et al.*, 2014). Nonetheless, haplogroup distribution is independent from subject type (HC and PwMS at V1) — Fisher's exact test with Monte Carlo simulations.

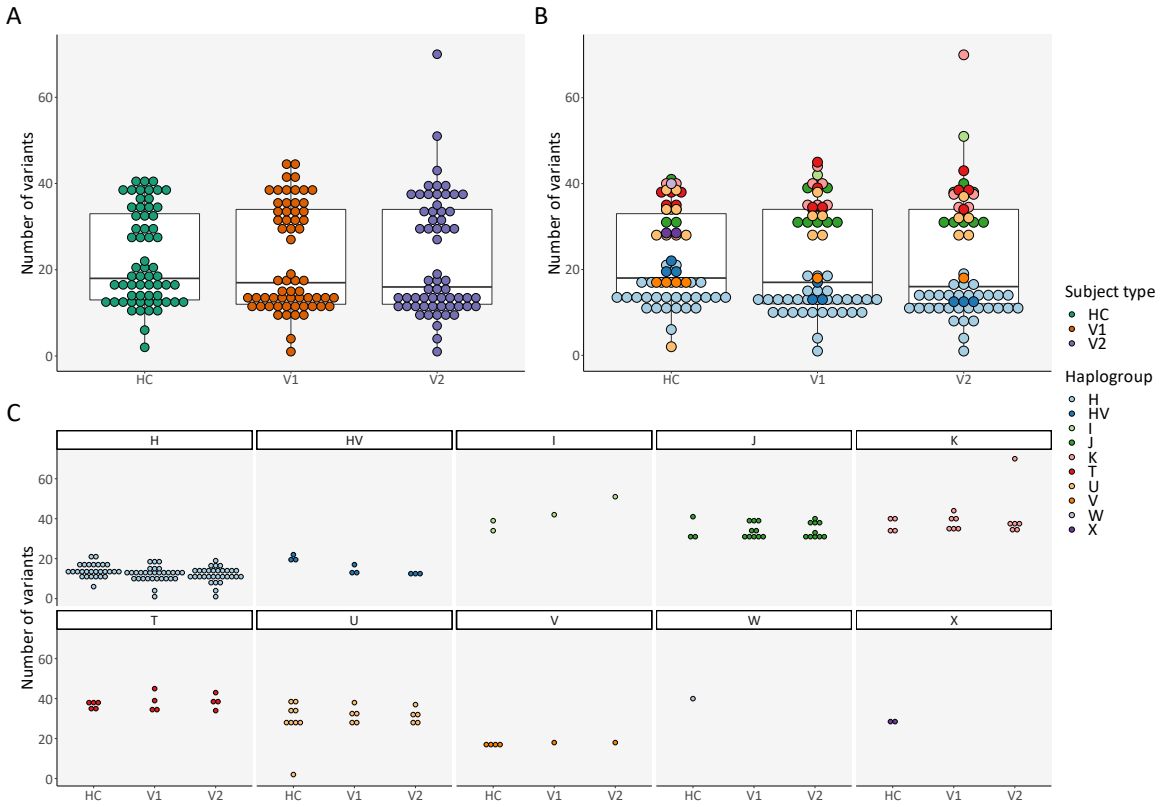


Figure 26 — Total number of variants: Cross-sectional comparison (subject types and haplogroups)

(**A**) Number of variants, per subject type; (**B**) Number of variants, per subject type and per haplogroup; (**C**) Expansion of **Figure 26B** for better visualization. Abbreviations: HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

When comparing HC and PwMS at V1, we observe that the total number of variants did not vary according to age (**Figure 27A**) — linear regression model. The difference in the number of mutations between HC and PwMS at V1 was also not influenced by sex (**Figure 27B–C**) — two sample t-test (**Figure 27B**), χ^2 test of independence (**Figure 27C**), and two-proportions z-test (**Figure 27C**).

When looking at the differences between PwMS at V1 and HC in various mtDNA regions, *MT-ND3* showed the highest discordancy in prevalence for PwMS at V1 and *MT-RNR2* for HC (**Figure 28A**), while the region with the highest discordant adjusted mutational rate, which is the sum of all VLs in a specific region divided by its region size, was *MT-TA* for PwMS at V1 and *MT-TT* for HC (**Figure 28B**). Only the discordant mutational rate between PwMS at V1 and HC regarding *MT-CO2* was significant prior to correcting for multiple comparisons — initial *p*-value 2.09×10^{-2}; adjusted *p*-value 6.69×10^{-1}; paired t-test adjusted with FDR. However, there were no statistically significant differences for discordant prevalence (**Figure 28A**) — McNemar's tests adjusted with FDR; nor for discordant mutational rates (**Figure 28B**) — paired t-tests adjusted with FDR.

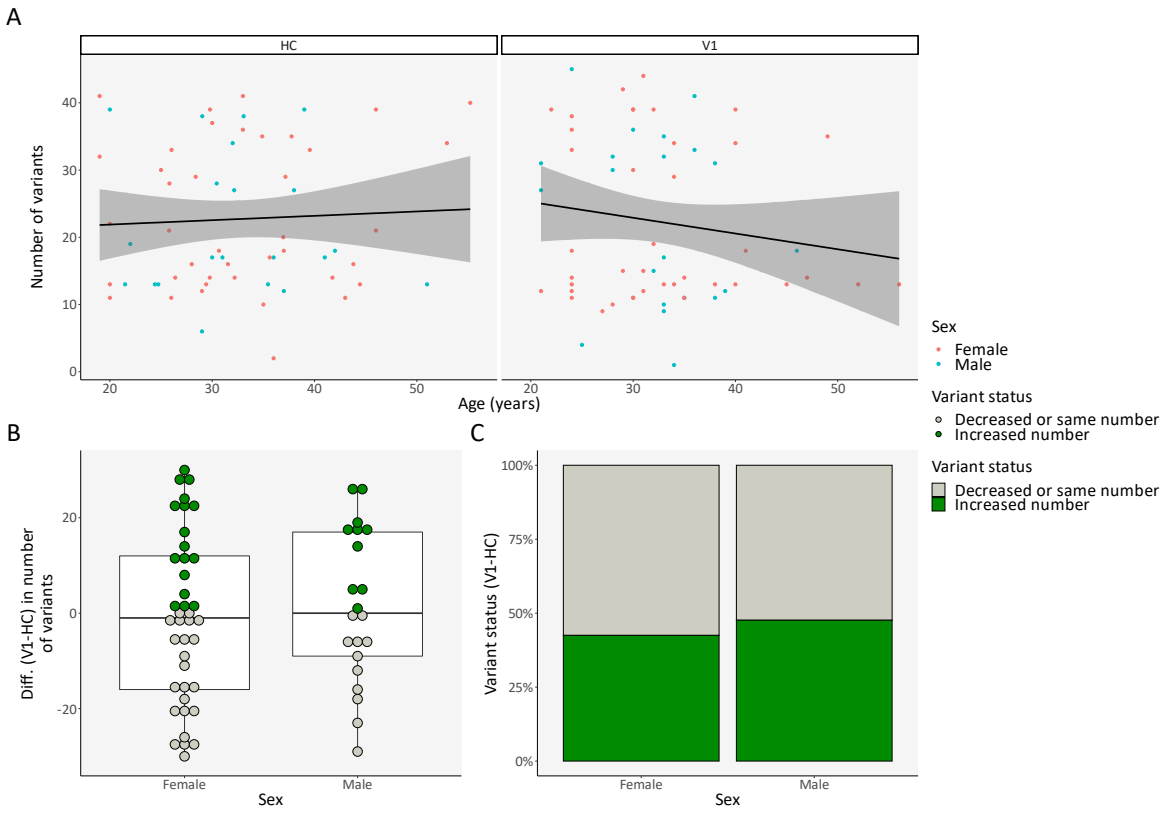


Figure 27 — Total number of variants: Cross-sectional comparison (age and sex)

(**A**) Correlation between number of variants and age, per subject type and sex; (**B**) Difference in number of variants per triplet (PwMS at V1-HC), per sex; (**C**) Proportion of triplets with an increased or equal/decreased number of variants, per sex. Abbreviations: Diff. — difference; HC — healthy control; PwMS — patient with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1.

Regarding macro regions in mtDNA, *Other*, which refers to positions in rCRS with an overlap between the two strands or left unannotated, was the most commonly discordant

affected region for PwMS at V1 and tRNA for HC (**Figure 28C**), while Complex III was the region with the highest discordant adjusted mutational rate for PwMS at V1 and *Other* for HC (**Figure 28D**). Nonetheless, there were no statistically significant differences for discordant prevalence (**Figure 28C**) — McNemar's tests adjusted with FDR; nor for discordant mutational rates (**Figure 28D**) — paired t-tests adjusted with FDR.

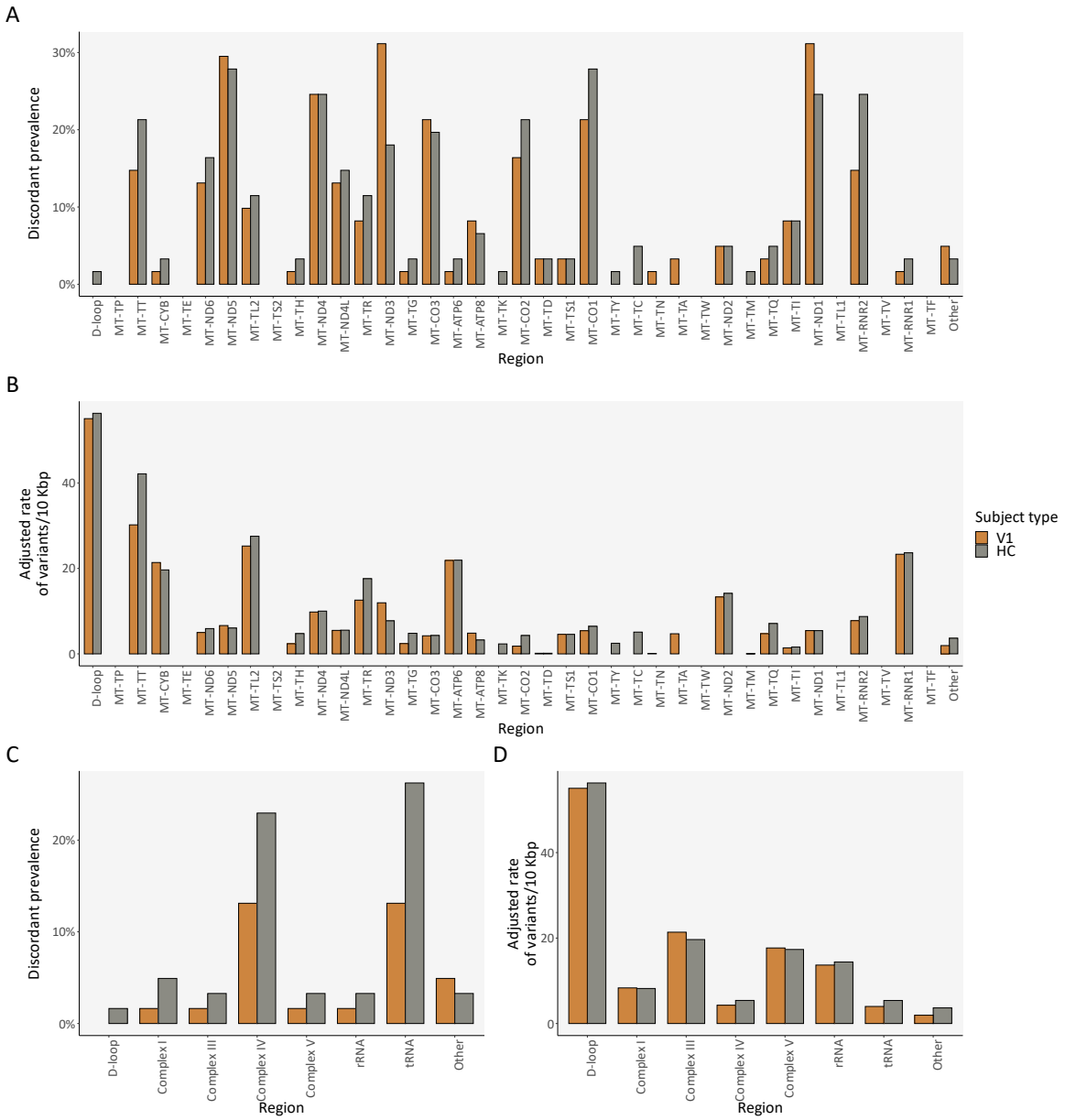


Figure 28 — Total number of variants: Cross-sectional comparison (regions)

(**A,C**) Discordant prevalence of variants for each mtDNA region/locus and for each macro mtDNA region, per subject type, respectively; (**B,D**) Relative variant burden for each mtDNA region/locus and for each macro mtDNA region, per subject type, respectively. Region *Other* refers to positions in rCRS with an overlap between the two strands or unannotated. Abbreviations: bp — base pair; HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1.

Afterwards, I further took all variants into account, to see if individual variants differed between HC and PwMS at V1, including variants previously associated with MS risk (more details in *1.4.6.2.1. Mitochondrial DNA*). The variant with the highest discordant prevalence in PwMS was m.12612A>G in the *MT-ND5* gene, which is a common non-pathogenic polymorphism associated with the J haplogroup, while, for HC, it was the variant m.8251G>A in the *MT-CO2* gene, which is also a common non-pathogenic polymorphism, albeit associated with the I haplogroup (**Table S34**) (Butler, 2012; Lott *et al.*, 2013). However, significant differences were not found — McNemar's tests adjusted with FDR.

Interestingly, a single PwMS had a pathogenic variant, namely, m.11778G>A in the *MT-ND4* gene (**Table S34**), which is associated with LHON and neuropathy, ataxia, and retinitis pigmentosa (NARP) (Chung *et al.*, 2021), albeit at low VLs: 5.8% and 7.3% for V1 and V2, respectively; well below the VLs usually required for pathogenesis (more details in *1.4.6.2.1. Mitochondrial DNA*).

Subsequently, I took into account variants predicted *in silico* to be likely deleterious for proteins, defined by a mean value > 0.5 from two independent scores: (**I**) MutPred (B. Li *et al.*, 2009; Pereira *et al.*, 2011; Piotrowska-Nowak *et al.*, 2018); and (**II**) APOGEE (Castellana *et al.*, 2017) (**Figure 29** and **Figure 30**).

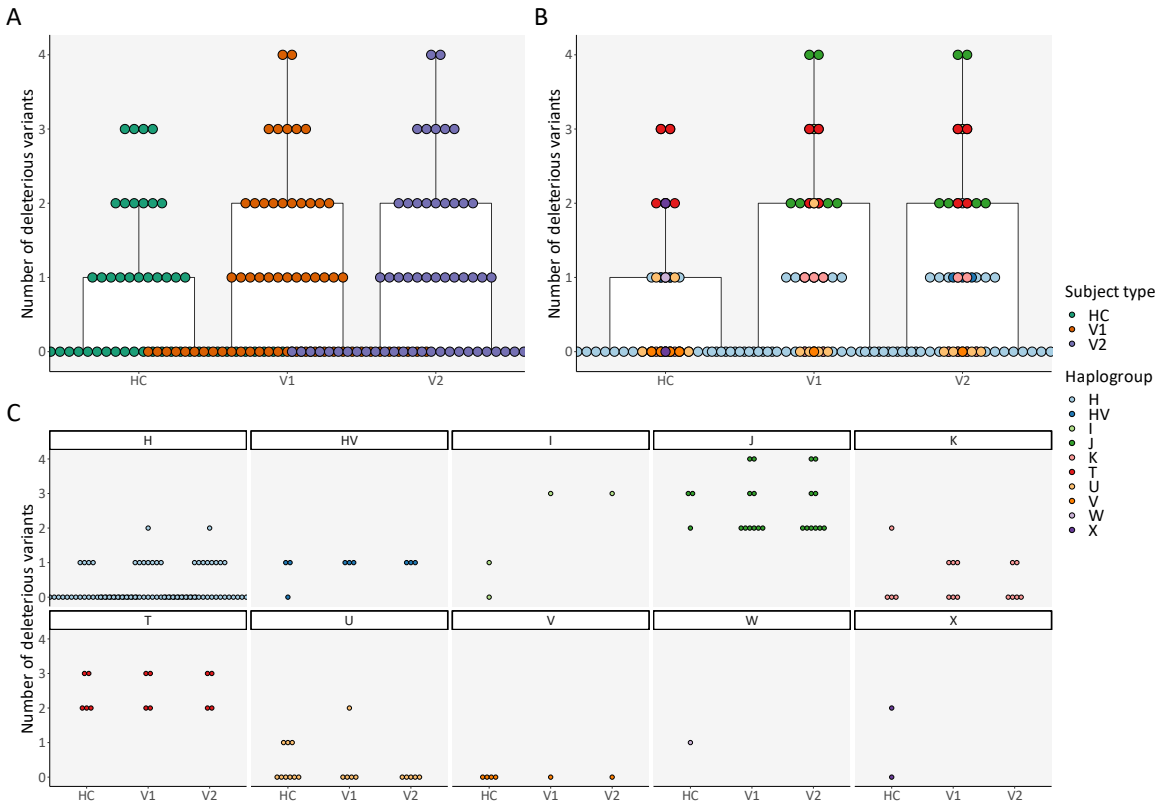


Figure 29 — Number of deleterious variants: Cross-sectional comparison (subject types and haplogroups)

(**A**) Number of deleterious variants, per subject type; (**B**) Number of deleterious variants, per subject type and haplogroup; (**C**) Expansion of **Figure 29B** for better visualization. Abbreviations: HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Nevertheless, there were no significant differences between different subject types for number of deleterious variants (**Figure 29A**), nor for cumulative deleterious burden (**Figure 30A**), which corresponds to the total sum of a variant's VL multiplied by its deleterious score per sample — paired t-tests adjusted with FDR.

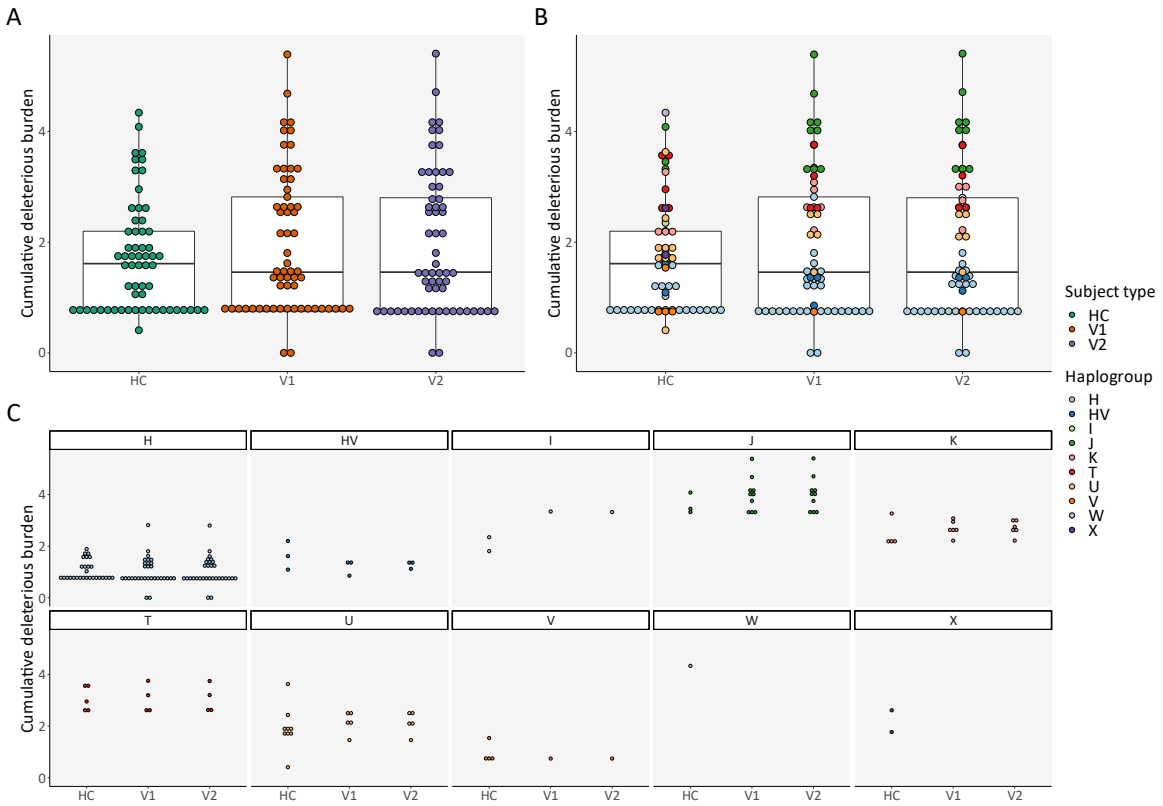


Figure 30 — Cumulative deleterious burden: Cross-sectional comparison (subject types and haplogroups)

(**A**) Cumulative deleterious burden, per subject type; (**B**) Cumulative deleterious burden, per subject type and haplogroup; (**C**) Expansion of **Figure 30B** for better visualization. Abbreviations: HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

We observe a similar haplogroup-specific effect for both the number of deleterious mutations (**Figure 29B–C**) and cumulative deleterious burden (**Figure 30B–C**) — *p*-values 4.61×10^{-13} and 1.38×10^{-15}, respectively; Kruskal-Wallis test. Haplogroups J and T had significantly more deleterious variants (**Table S35**) and cumulative deleterious burden (**Table S36**).

Considering the number of deleterious variants, when comparing HC and PwMS at V1, we note that they did not vary according to age (**Figure 31A**) — Kendall rank correlation test. The difference in the number of deleterious mutations between HC and PwMS at V1 was also not influenced by sex (**Figure 31B–C**) — two sample t-test (**Figure 31B**), χ^2 test of independence (**Figure 31C**), and two-proportions z-test (**Figure 31C**). The same was true for cumulative deleterious burden — linear regression model for age (**Figure 31D**) and two sample t-test for sex (**Figure 31E**).

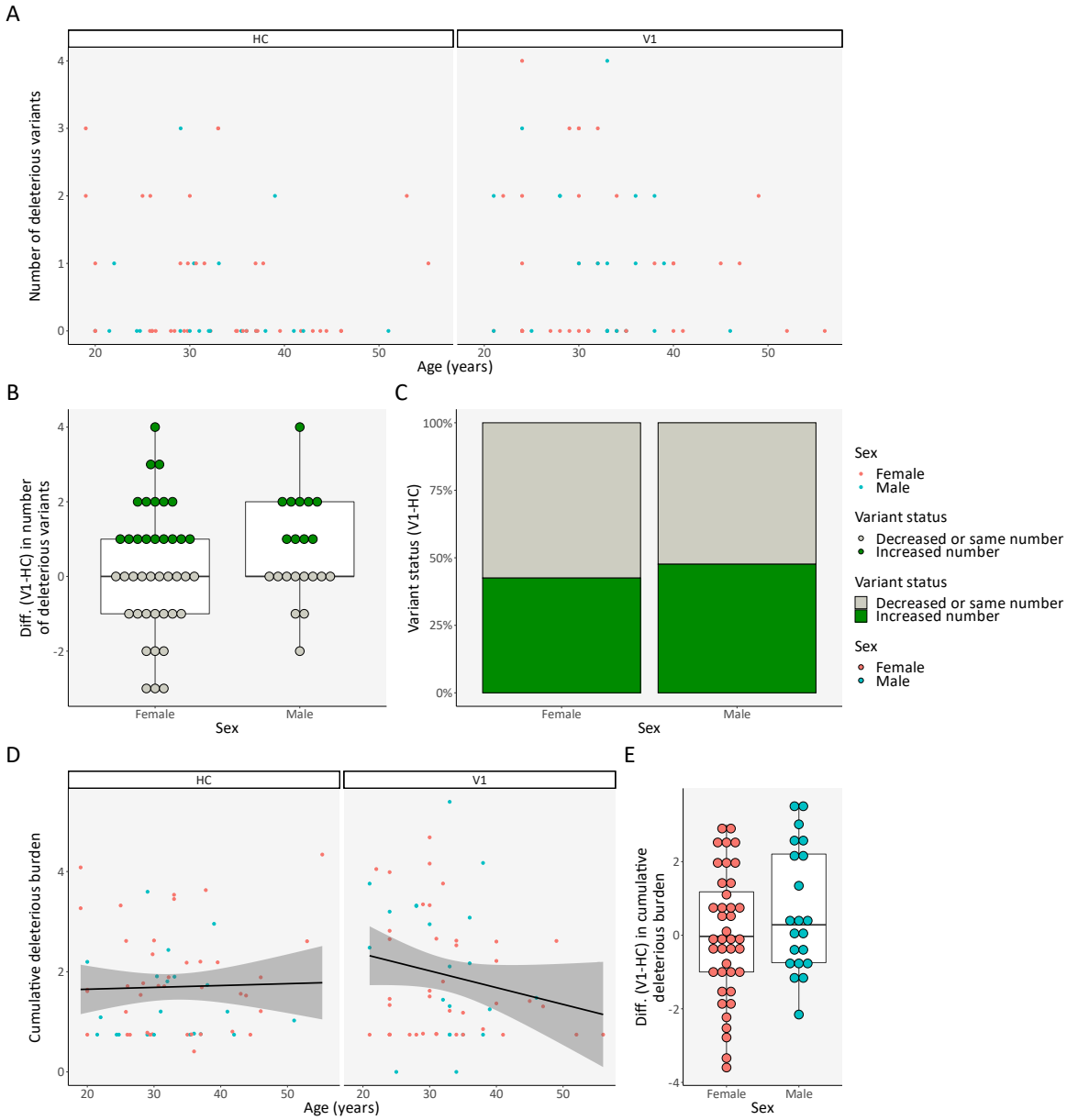


Figure 31 — Deleterious variants: Cross-sectional comparison (age and sex)

(**A**) Correlation between number of deleterious variants and age, per subject type and sex; (**B**) Difference in number of deleterious variants per triplet (PwMS at V1-HC), per sex; (**C**) Proportion of triplets with an increased or equal/decreased number of deleterious variants, per sex; (**D**) Correlation between cumulative deleterious burden and age, per subject type and sex; (**E**) Difference in cumulative deleterious burden per triplet (PwMS at V1-HC), per sex. Abbreviations: Diff. — difference; HC — healthy control; PwMS — patient with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1.

Regarding differences between PwMS at V1 and HC in various mtDNA regions, *MT-ND5* showed the highest discordancy in prevalence for PwMS at V1 and *MT-ATP6* for HC (**Figure 32A**).

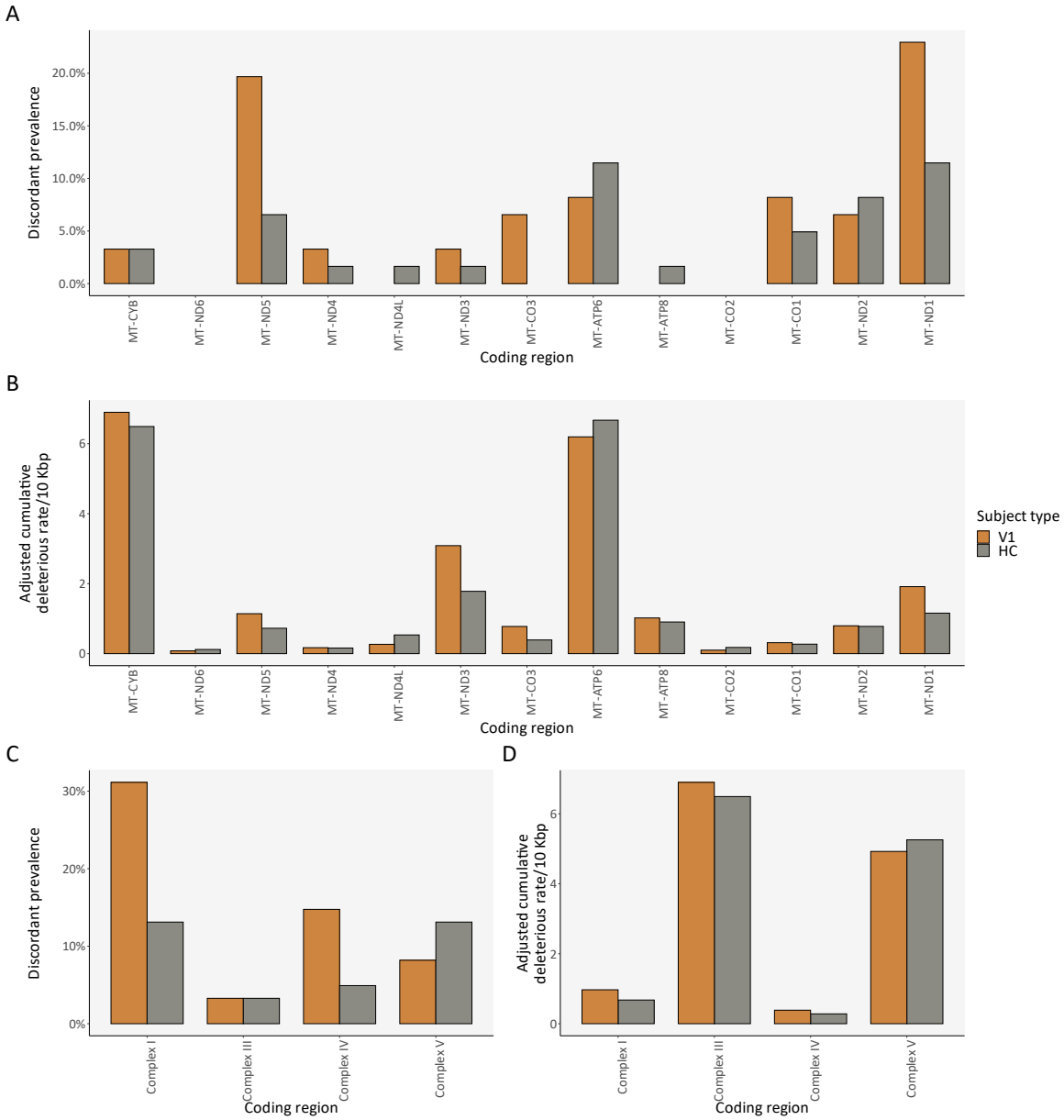


Figure 32 — Deleterious variants: Cross-sectional comparison (regions)

(**A,C**) Discordant prevalence of deleterious variants for each mtDNA coding region/locus and for each macro mtDNA coding region, per subject type, respectively; (**B,D**) Relative cumulative deleterious burden for each mtDNA coding region/locus and for each macro mtDNA coding region, per subject type, respectively. Abbreviations: bp — base pair; HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1.

In parallel, the region with the highest discordant adjusted cumulative deleterious rate, which is the cumulative deleterious rate for a specific region relative to its region size, was *MT-ND3* for PwMS at V1 and *MT-ATP6* for HC (**Figure 32B**). Regardless, no statistically significant differences were found for discordant prevalence (**Figure 32A**) — McNemar's

tests adjusted with FDR; nor for discordant cumulative deleterious rates (**Figure 32B**) — paired t-tests adjusted with FDR.

Regarding macro regions in mtDNA, Complex I was the most commonly discordant affected region for PwMS at V1 and Complex V for HC (**Figure 32C**), while Complex III was the region with the highest adjusted cumulative deleterious rate for PwMS at V1 and Complex V for HC (**Figure 32D**). Interestingly, in contrast to the pilot study for sample size appraisal (more details in *4.2.1. Cohort description*), the discordancy in prevalence for Complex I diminished with the increased sample size. Consequently, there were no statistically significant differences for discordant prevalence (**Figure 32C**) — McNemar's tests adjusted with FDR; nor for cumulative deleterious rate (**Figure 32D**) — paired t-tests adjusted with FDR.

Finally, I looked into tRNA mutations predicted to be pathogenic according to the MitoTIP score (Sonney *et al.*, 2017): no variants were found (**Figure 33A**). The cumulative MitoTIP score, which corresponds to the total sum of a variant's VL multiplied by its MitoTIP score per sample, was similar in all subject types (**Figure 33B–C**) — paired t-tests adjusted with FDR.

Nonetheless, as observed previously, haplogroup had a significant influence in the cumulative MitoTIP score (**Figure 33B–C**) — *p*-value 3.59×10^{-13}; Kruskal-Wallis test, with haplogroups K and U having higher cumulative MitoTIP scores, whereas haplogroups H and HV had lower scores (**Table S37**). No influence was found regarding age (**Figure 33D**) — linear regression model; or sex (**Figure 33E**) — two sample t-test.

The region with the highest discordant adjusted cumulative MitoTIP rate, which is the cumulative MitoTIP rate for a specific region relative to its region size, was *MT-TA* for PwMS at V1 and *MT-TT* for HC (**Figure 33F**). However, no statistically significant differences were found — paired t-tests adjusted with FDR.

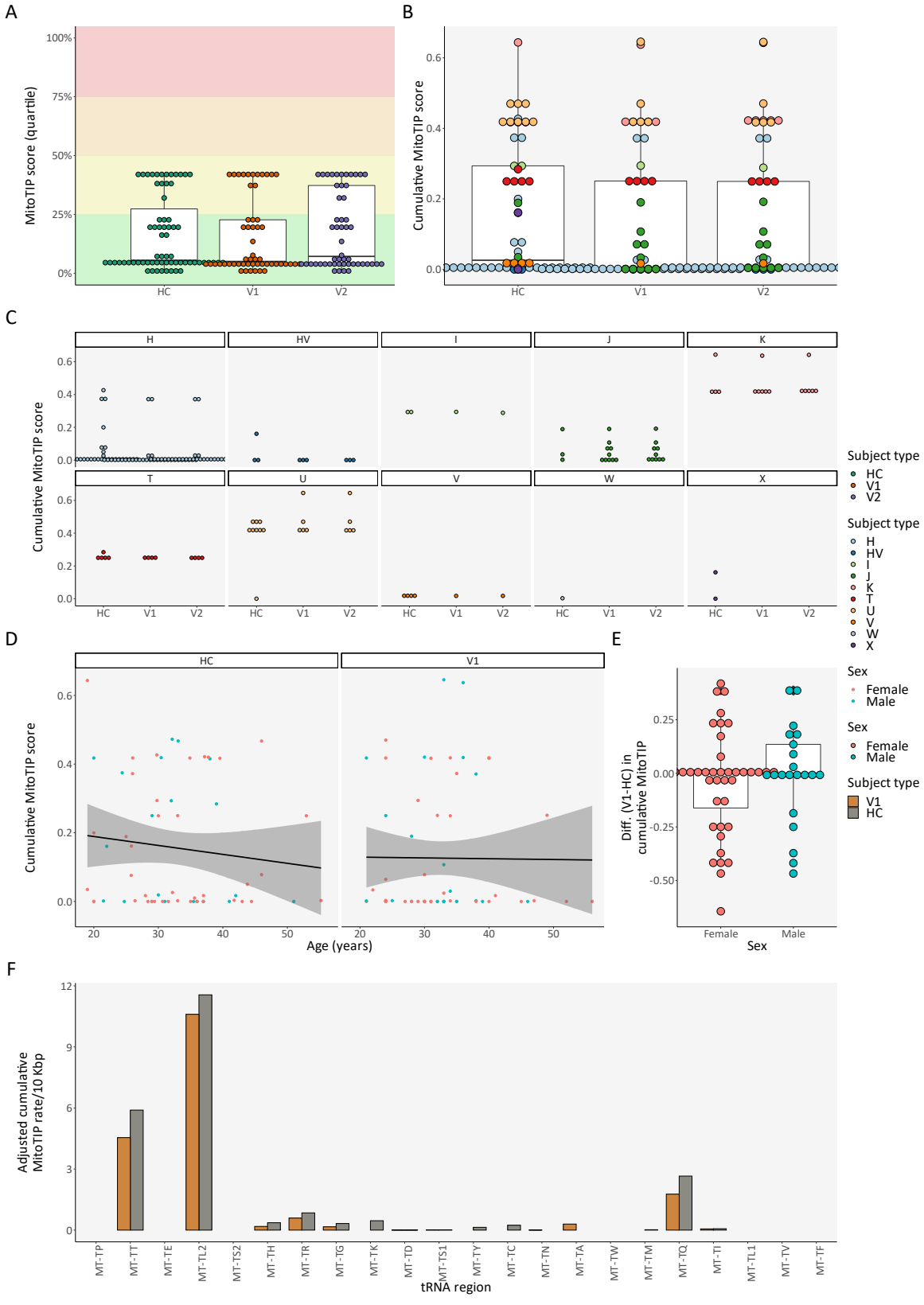


Figure 33 — tRNA variants: Cross-sectional comparison

(**A**) MitoTIP score (in quartiles) of all variants, per subject type — the green background indicates the range for “likely benign” variants, yellow for “possibly benign” variants, orange for “possibly pathogenic” variants, and red for “likely pathogenic” variants; (**B**) Cumulative MitoTIP score, per subject type and haplogroup; (**C**) Expansion of **Figure 33B** for

better visualization; (**D**) Correlation between cumulative MitoTIP score and age, per subject type and sex; (**E**) Difference in cumulative MitoTIP score per triplet (PwMS at V1-HC), per sex; (**F**) Relative cumulative MitoTIP score for each tRNA region/locus, per subject type. Abbreviations: bp — base pair; Diff. — difference; HC — healthy control; PwMS — patient with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1.

4.2.4. Longitudinal comparison

As mentioned previously, there were no significant differences between PwMS at V1 and the same patient at V2 concerning: total number of variants (**Figure 26A**), number of deleterious variants (**Figure 29A**), cumulative deleterious burden (**Figure 30A**), and cumulative MitoTIP score (**Figure 33B**) — paired t-tests adjusted with FDR.

The mean proportion of mutations present in a single visit — novel variants — were 4.01% (range: 0.00–53.33%) for V1 and 3.38% (range: 0.00–37.14%) for V2 (**Figure 34A**), with mean VLs of 8.39% (range: 2.50–100.00%) for V1 and 4.66% (range: 2.50–19.60%) for V2 (**Figure 34B**); no differences were found between different visits — paired t-tests.

Subsequently, I considered the difference in VLs for the previously found variants between the second and the first visits — old variants. The mean VL change was 0.06% in all variants (**Figure 34C**) and 0.09% per subject (**Figure 34D**) (range: -10.90–50.30%). Both VL changes were statistically similar to zero — one-sample t-test.

When I took into account all novel and old variants, novel variants had a significantly lower VL than old variants (**Figure 34E**) and a lower coverage ratio (**Figure 34G**) — adjusted *p*-values 1.80×10^{-150} and 1.09×10^{-6}; two-sample t-tests adjusted with FDR. There were no differences for other quality control variables (**Figure 34F** and **Figure 34H–I**) (more details in *3.2.3. DNA sequencing, bioinformatic processing, and data analysis*), nor for protein deleterious score and MitoTIP score (**Figure 34J–K**) — two-sample t-tests adjusted with FDR.

Since each PwMS had a very similar mutational pattern between the two visits, I then moved on from an intrapersonal comparison to an interpersonal comparison, with a focus on exploring the origin of these novel variants and the VL change in old variants in the full PwMS cohort. For that purpose, I considered a PwMS as a single data point, where I performed the mean of the results from both sequencing runs.

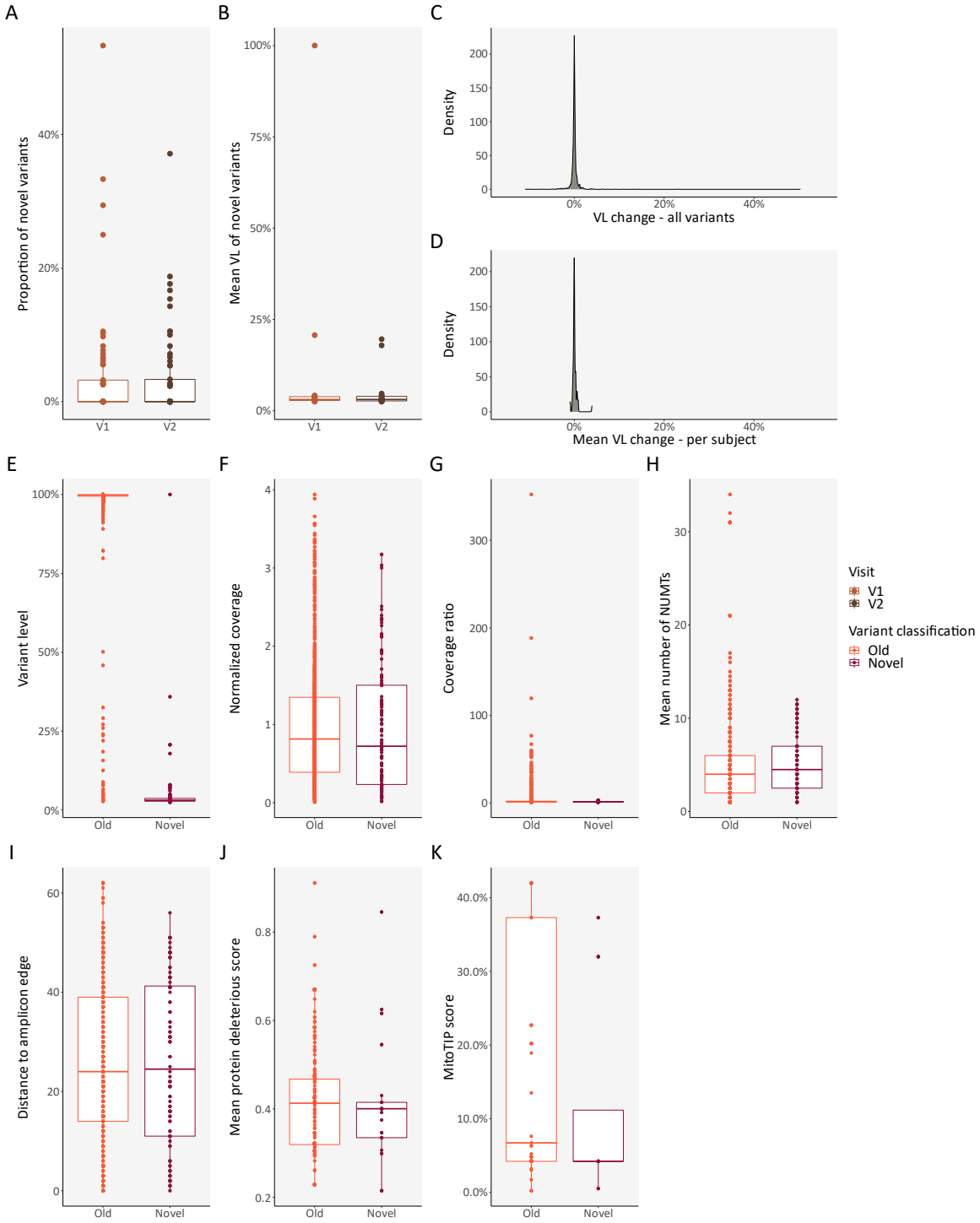


Figure 34 — Longitudinal changes in PwMS: Intrapersonal changes

(**A**) Proportion of novel variants, per subject and clinical visit; (**B**) Mean variant level of novel variants, per subject and clinical visit; (**C–D**) Density plots of variant level change in old variants, for all mutations and per subject, respectively; (**E–D**) Comparison between novel and old variants for variant level, normalized coverage, coverage ratio, mean number of NUMTs, distance to amplicon edge, mean protein deleterious score, and MitoTIP score, respectively. Abbreviations: NUMTs — nuclear insertions of mitochondrial DNA; PwMS — patients with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2; VL — variant level.

Excluding patients without novel variants, haplogroup showed a significant effect, albeit close to non-significance, in the mean proportion of novel variants (excluding haplogroups with less than three paired samples) (**Figure 35A**) — *p*-value 4.99 x 10^{-2}; Kruskal-Wallis test. However, subsequent pairwise comparisons yielded no statistically significant results — Dunn's test with FDR. No influence was found regarding age (**Figure 35B**) — linear regression model; or sex (**Figure 35C**) — two sample t-test.

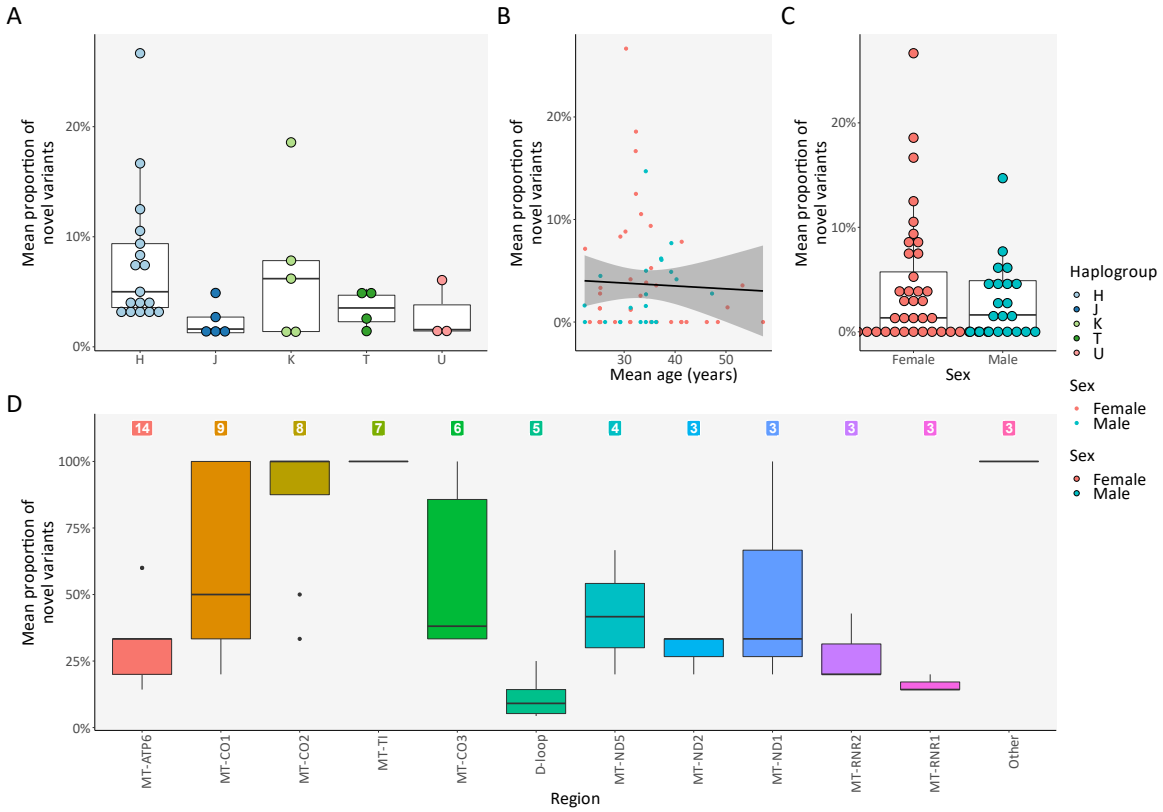


Figure 35 — Longitudinal changes in PwMS: Proportion of novel variants

(**A**) Mean proportion of novel variants per PwMS (both V1 and V2), per haplogroup with at least three samples and excluding samples without novel variants; (**B**) Correlation between mean proportion of novel variants per PwMS (both V1 and V2) and age; (**C**) Mean proportion of novel variants per PwMS (both V1 and V2), per sex; (**D**) Mean proportion of novel variants per PwMS (both V1 and V2), per region with at least three samples — colored labels indicate the number of samples for each region. Abbreviations: PwMS — patient(s) with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Interestingly, different regions have a significant effect on the proportion of novel variants (excluding regions with mutations from less than three paired samples) (**Figure 35D**) — *p*-value 5.90 x 10^{-6}; Kruskal-Wallis test, where *MT-TI* had the highest prevalence and D-loop the lowest (**Table S38**).

Regarding the mean VL of these novel variants, as well as the mean VL change in old variants, there was no effect from haplogroup (**Figure 36A** and **Figure 37A**), age (**Figure 36B** and **Figure 37B**), sex (**Figure 36C** and **Figure 37C**), nor region (**Figure 36D** and **Figure 37D**) — Kruskal-Wallis test, linear regression model, two sample t-test, and Kruskal-Wallis test, respectively.

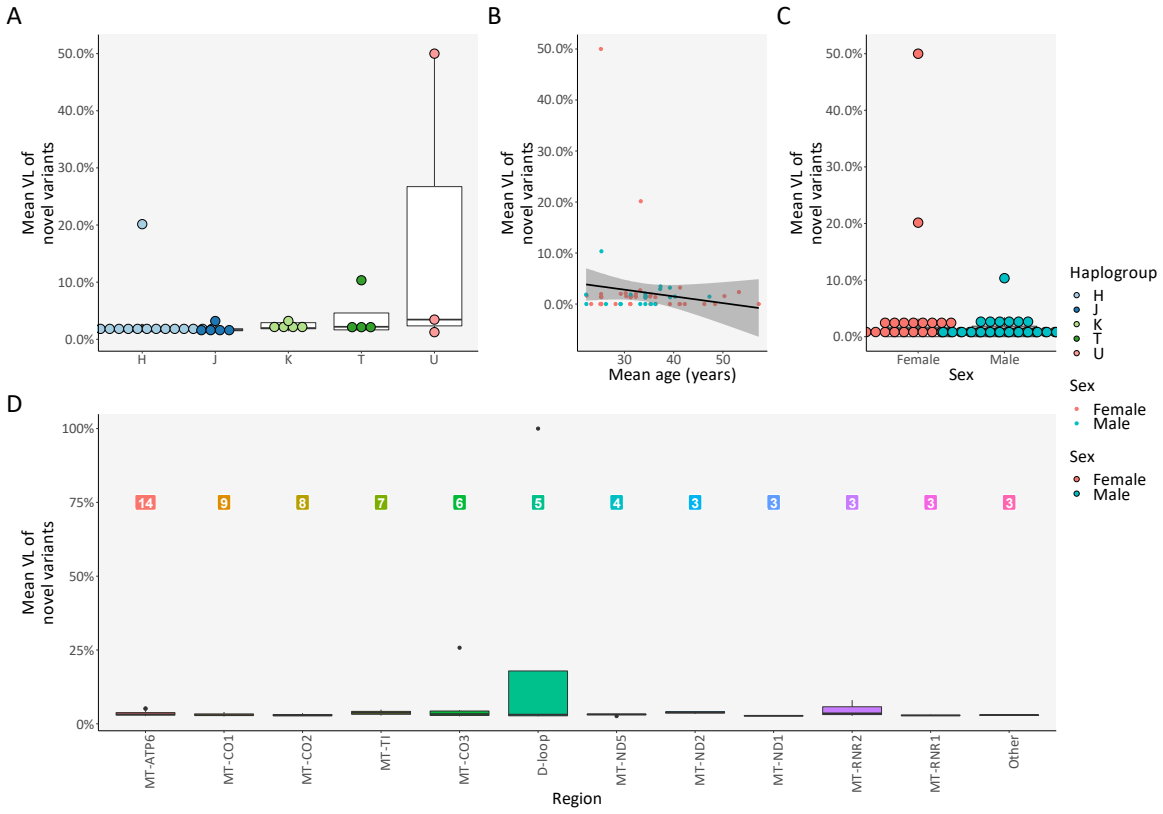


Figure 36 — Longitudinal changes in PwMS: Variant level of novel variants

(**A**) Mean variant level of novel variants per PwMS (both V1 and V2), per haplogroup with at least three samples and excluding samples without novel variants; (**B**) Correlation between mean variant level of novel variants per PwMS (both V1 and V2) and age; (**C**) Mean variant level of novel variants per PwMS (both V1 and V2), per sex; (**D**) Mean variant level of novel variants per PwMS (both V1 and V2), per region with at least three samples — colored labels indicate the number of samples for each region. Abbreviations: PwMS — patient(s) with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2; VL — variant level.

Regarding the clinical variables in the cohort, PwMS under MS treatment increased from 26.23% to 47.54% (**Table S39**). However, change in MS medication was independent from diagnosis change (**Figure S16**) — Fisher's exact test with Monte Carlo simulations.

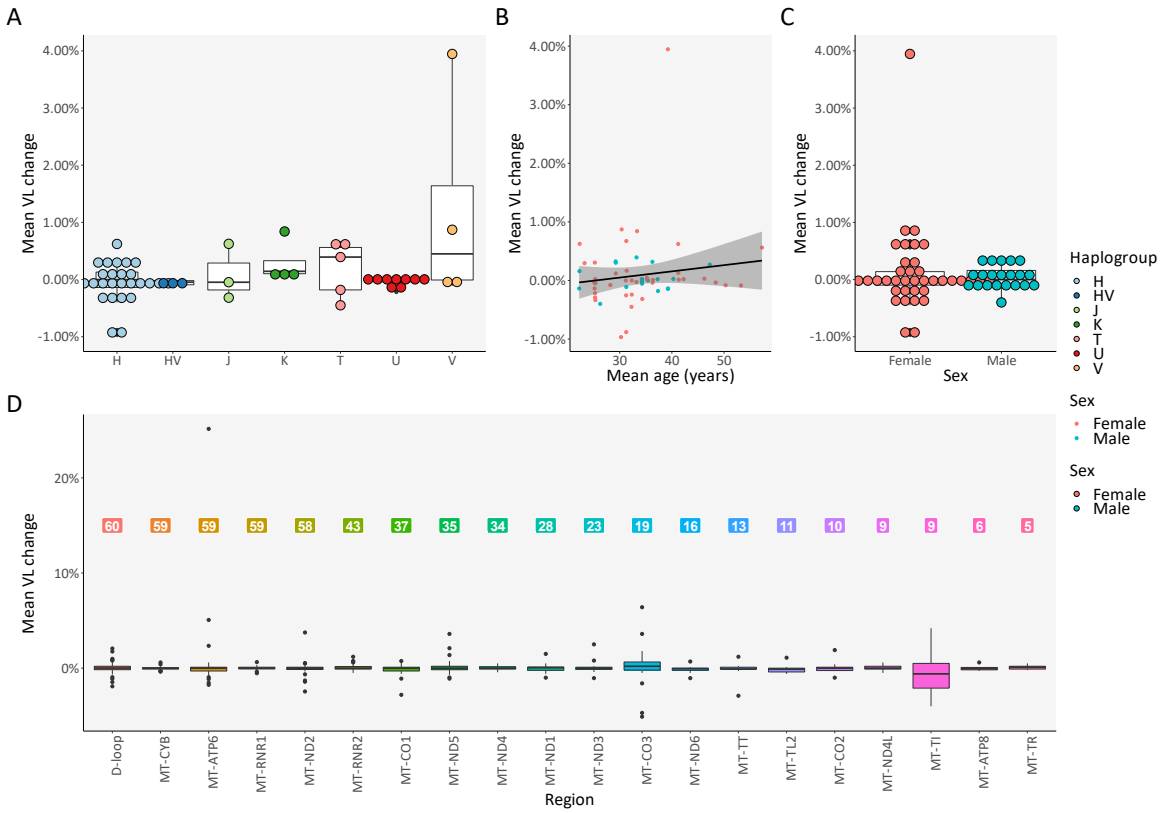


Figure 37 — Longitudinal changes in PwMS: Variant level change of old variants

(**A**) Mean variant level change of old variants per PwMS (both V1 and V2), per haplogroup with at least three samples; (**B**) Correlation between mean variant level change of old variants per PwMS (both V1 and V2) and age; (**C**) Mean variant level change of old variants per PwMS (both V1 and V2), per sex; (**D**) Mean variant level change of old variants per PwMS (both V1 and V2), per region with at least three samples — colored labels indicate the number of samples for each region. Abbreviations: PwMS — patient(s) with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2; VL — variant level.

Taking full advantage of the clinical data in the cohort and hoping to identify the variables most strongly associated with clinical disability, I performed batch Kruskal-Wallis tests for each available clinical variable with diagnosis change as a grouping variable (**Figure S17** and **Figure S18**). Two different approaches were tested: (**I**) the mean between V1 and V2; and (**II**) the change from V1 to V2, by subtracting V2 and V1. After multiple comparison correction with FDR, only the mean values of MRI data, namely, number and volume of T2 hyperintense lesions, were associated with diagnostic change (**Table S40**). In comparison with the V2-V1 difference, mean values had consistently lower adjusted *p*-values — *p*-value 1.76×10^{-2}; paired t test.

Similarly, I performed batch Kruskal-Wallis tests with diagnosis change as a grouping variable for each NGS variable explained previously, namely, number of variants, number

of deleterious variants, cumulative protein deleterious burden, cumulative MitoTIP, proportion of novel variants, mean VL of novel variants, and mean VL change (**Figure S19**); no statistically significant differences were found (**Table S41**). Additionally, haplogroup was independent from diagnosis change (**Figure S20**) — Fisher's exact test with Monte Carlo simulations.

Since I had further longitudinal clinical data from a subset of the cohort (55.74%, *N*=34), encompassing an additional clinical visit six years after cohort entry (V3), and given the likely additional spectrum of clinical disability, I repeated the Kruskal-Wallis batch analysis grouped on the diagnosis evolution (V1, V2, and V3) and the mean value of the clinical variables described above. However, clinical variables did not differ significantly per diagnosis evolution (**Figure S21** and **Table S42**). The same was true for the chosen NGS variables (**Figure S22** and **Table S43**). Finally, haplogroup was also independent from diagnosis evolution (**Figure S23**) — Fisher's exact test with Monte Carlo simulations.

V — DISCUSSION AND CONCLUSION

Page intentionally left blank

5.1. PrecisionCallerPipeline

In accordance with the goal I had previously set (more details in *II — OBJECTIVES*), I was able to develop a fully customizable pipeline — PCP — for the Precision ID mtDNA Whole Genome Panel, which produced BAMs accurately mapped to the rCRS and that has revisited the heteroplasmy threshold of the current gold standard method — TSS.

On the one hand, despite arising from the same set of samples and sequencing runs, the output from PCP and TSS was sufficiently different to yield significant changes between the two pipelines. In comparison with TSS, PCP accomplished better performance metrics in previously sequenced samples, namely, higher sensitivity and F_1 scores without a decrease in specificity and precision. This was mostly due to the additional detection of heteroplasmic variants with a VL above 2.5%, but below TSS's ~10% threshold. Interestingly, this increase in the number of correct variants and the improved performance was achieved despite PCP having lower coverage and fewer sequences. Moreover, the same pattern of a higher number of variants at a lower threshold, notwithstanding a lower mean coverage and fewer sequences, was also observed in an independent set of clinical samples.

On the other hand, mtDNA differences between the same sample are not unheard of, as DNA polymerases, amplification protocols, sequencing runs, and variant callers are frequent sources of disparity in genomics (Cihlar, Amory, *et al.*, 2020; Fazzini *et al.*, 2021). Discerning sequencing errors/false positives from true heteroplasmic variants is a challenging task, usually achieved through post-sequencing curation, looking for signs of poor amplification, strand bias, mutations in LCR, and the presence of NUMTs (Ring *et al.*, 2018; Smart *et al.*, 2019; Cihlar, Strobl, *et al.*, 2020; Wei *et al.*, 2020; Woerner *et al.*, 2020; Marshall and Parson, 2021).

Although sometimes ignored (Yonova-Doing *et al.*, 2021), heteroplasmic variants are ubiquitous (Sosa *et al.*, 2012; Payne *et al.*, 2013; Ye *et al.*, 2014a; Li *et al.*, 2015; Naue *et al.*, 2015) and show significant tissue-specificity (Calloway *et al.*, 2000; He *et al.*, 2010; Li *et al.*, 2015; Naue *et al.*, 2015; Hahn and Zuryn, 2019; J. Xu *et al.*, 2019; Ludwig *et al.*, 2019). Interestingly, low-level heteroplasmic variants (VL <10%) have shown matrilineal inheritance (Guo, C.-I. Li, *et al.*, 2013; Zaidi *et al.*, 2019) with increased relevance in aging (Wachsmuth *et al.*, 2016), and in clinical settings, particularly cancer (Fendt *et al.*, 2020; Yuan *et al.*, 2020). Nonetheless, the reliability of these low-level variants has been a matter

of debate (Just, Irwin and Parson, 2014, 2015; Ye *et al.*, 2014b) and, thus, current guidelines/workflows suggest a heteroplasmy threshold of 10%, based on the limitations of Sanger sequencing and electrophoresis technology (Parson *et al.*, 2014; Bris *et al.*, 2018; Brandhagen, Just and Irwin, 2020; Cihlar, Amory, *et al.*, 2020; Laricchia *et al.*, 2021; Poole *et al.*, 2021).

False positives in both PCP and TSS showed a decreased normalized coverage, which indicates poor amplification in those regions, and were mostly transitions, which is in accordance with the literature (Bolze *et al.*, 2020; Laricchia *et al.*, 2021). However, the remaining putative mechanisms of error were very different between the two pipelines, and even showed differences within the same pipeline, depending on where the novel variants were found (unmixed vs. mixed samples). Correspondingly, in the analysis of the 1000 Genomes Project's samples, no false positives were shared.

In parallel, discordant variants in the stroke cohort showed most of the variants previously flagged as false positives, albeit with many more variants only present in PCP at a low level, which had not been flagged before. This different mutational pattern from the samples sequenced within the 1000 Genomes Project might arise from the tissue-specificity of heteroplasmic variants, since blood, which was used for DNA extraction in the stroke cohort, has a lower level of heteroplasmy (Li *et al.*, 2015; Naue *et al.*, 2015) and a very diverse mixture of mtDNA content (Rausser *et al.*, 2021), as it encompasses multiple cell subtypes.

The PCP pipeline offers multiple advantages in comparison with TSS. Firstly, it is in line with the Findable, Accessible, Interoperable, Reusable (FAIR) principles (Wilkinson *et al.*, 2016), increasingly important in the field of genomics (Chervova *et al.*, 2019). Thus, this open-access approach makes it easier to implement validated and uniform bioinformatic protocols in genomics (Van der Auwera *et al.*, 2013; Jennings *et al.*, 2017; Marshall *et al.*, 2020). Secondly, it outputs a BAM file with the correct rCRS reference, thus allowing for integration into other NGS workflows and greatly facilitating variant annotation. This is particularly important when dealing with high-throughput datasets, where manual curation and inspection are not feasible nor efficient, thus favoring computational approaches to DNA variant analysis (Wong *et al.*, 2021). Finally, it is fully customizable. In a world of limited resources (personnel, time, samples, funding), this workflow is easily adaptable, depending on the focus of each research project and tissue mutational pattern.

On the one hand, if one wishes to prioritize sample diversity in favor of replicates, variants might be called with PCP and filtered if not present in TSS, since PCP captures ~98% of all TSS variants with a very high VL correlation, and TSS discordant variants are likely to be false positives, in accordance with the previous analysis. On the other hand, if one prefers to sacrifice sample diversity in favor of VCM performance, the protocol for NUMT removal and heteroplasmy threshold optimization in unmixed and mixed samples yields an optimal workflow specific to that set of samples. Furthermore, the two approaches might be combined, or even expanded, into FASTQ quality control, alignment, or further variant annotation, with the addition of hypervariable segments (Amorim, Fernandes and Taveira, 2019; Syndercombe Court, 2021), poly C-stretches (Zhao *et al.*, 2010), which are particularly difficult in Ion Torrrent™ sequencing (Sturk-Andreaggi *et al.*, 2020), and the presence of homoplasmic or heteroplasmic variants at the HelixMTdb (Bolze *et al.*, 2020) (**Table S26** provides an example).

In summary, future clinical research will benefit from using this open-source bioinformatic processing since it keeps the advantages of the Precision ID library kit, particularly its low DNA input, while circumventing the limitations of TSS, namely, its modified reference sequence, proprietary nature, and 10% heteroplasmy threshold. For separating true variants from false variants, there is a range of options that can be used in combination or not, depending on the samples and research focus: (**I**) a dedicated mixture optimization protocol; (**II**) variant filtering based on normalized coverage, strand bias, the presence of NUMTs, among other parameters; (**III**) filtering the output from PCP with TSS.

Nonetheless, the research protocol for PCP also has a few limitations. Firstly, a limited set of samples were used for the primary and mixture analysis, and they are not representative of the range of possible haplogroup combinations. Secondly, samples from the stroke cohort were not sequenced in duplicate nor mixed, which did not allow for a replication of the performance metrics. Finally, insertions and deletions were not taken into account, and, thus, I am unable to provide any recommendations for the analysis of those variants with the PCP pipeline.

Overall, I have developed the first open-source alternative to the enterprise software, Ion Torrent Suite™ Software (TSS), for the Precision ID mtDNA Whole Genome Panel, with improved performance metrics, and with an output in the correct rCRS reference. Since the majority of existing mtDNA bioinformatic tools (Lee *et al.*, 2008; Fan and Yao, 2011; Zhidkov

et al., 2011; Guo, J. Li, *et al.*, 2013; Lott *et al.*, 2013; Vianello *et al.*, 2013; Yang *et al.*, 2013; Calabrese *et al.*, 2014; Falk *et al.*, 2015; Navarro-Gomez *et al.*, 2015; Vellarikkal *et al.*, 2015; Weissensteiner, Forer, *et al.*, 2016; Weissensteiner, Pacher, *et al.*, 2016; Ishiya and Ueda, 2017; Rueda and Torkamani, 2017; Preste *et al.*, 2019) are only compatible with the correct rCRS format, and the few tools that perform variant calling in samples sequenced with the Precision ID library kit (Woerner *et al.*, 2018; Roth *et al.*, 2019; Cihlar, Amory, *et al.*, 2020; Faccinetto *et al.*, 2021) keep TSS's modified reference sequence, PCP is currently the sole option that bridges this bioinformatic gap, allowing for the universal variant calling of samples sequenced with the aforementioned library.

As mentioned previously, the herein presented pipeline, PCP, is available in the GitHub repository, https://github.com/filcfig/PCP and has already been published (Cortes-Figueiredo *et al.*, 2021). Additionally, the data generated with the samples previously sequenced within the 1000 Genomes Project are openly available in Zenodo (European Organization For Nuclear Research and OpenAIRE, 2013) at doi:10.5281/zenodo.5524539. Access to the data analyzed within the stroke cohort is restricted due to ethical and privacy concerns.

5.2. Multiple Sclerosis and Clinically Isolated Syndrome cohort

In accordance with the previously set aims (more details in *II — OBJECTIVES*), I was able to successfully perform WGS from $CD4^+$ T cells in a matched cohort of PwMS, with two different approaches: cross-sectional and longitudinal. However, there were no significant differences between PwMS and HC, and between V1 and V2 within PwMS.

Importantly, the MS/CIS cohort herein described is representative of the overall population of PwMS in Germany, namely, the mean age at onset of 33 years (Ohle *et al.*, 2021), the female predominance (Engelhard *et al.*, 2021; Ohle *et al.*, 2021), the distribution of the diagnoses, with a vast majority of RRMS (Engelhard *et al.*, 2021; Ohle *et al.*, 2021), and the evolution of DMDs, namely, the increase in the proportion of patients treated with dimethyl fumarate with time (Engelhard *et al.*, 2021).

According to the flow cytometry analysis performed in a subset of magnetically enriched samples, $CD4^+$ T cell enrichment was successful with the MojoSort™ Human CD4 T Cell Isolation Kit, which was expected (Fonseca Valter R. *et al.*, 2017; Godinho-Santos *et al.*, 2020). The Precision ID mtDNA Whole Genome Panel, in combination with the PCP

pipeline (Cortes-Figueiredo *et al.*, 2021), proved particularly valuable, as it was able to obtain mtDNA WGS from all samples without signs of contamination, despite a wide range of the number of cells after processing the samples, and, consequently, of DNA yield, which is consistent with previous reports (Pereira, Longobardi and Børsting, 2018; Strobl *et al.*, 2018; Wai, Barash and Gunn, 2018; Yao *et al.*, 2018; Cuenca *et al.*, 2020; Faccinetto *et al.*, 2021).

Overall, in the CIS/MS cohort, the increased number of mutations in the D-loop and the presence of a few pathogenic mutations at low VLs, such as the single m.11778G>A variant mentioned previously, are consistent with data from human mtDNA analysis (Sosa *et al.*, 2012; Ye *et al.*, 2014a; Li *et al.*, 2015). Also, the vast majority of tRNA variants are non-pathogenic (Sonney *et al.*, 2017), which might explain the absence of pathogenic tRNA variants in this cohort. In contrast to previous studies that have clearly demonstrated an increase in the number of mtDNA variants with aging (Calloway *et al.*, 2000; Li *et al.*, 2015; Naue *et al.*, 2015), there was a lack of effect from age on the number of mtDNA variants in the CIS/MS cohort. This might derive from the more limited age range of the MS/CIS cohort, varying from 19 to 56 years, while in other reports (Calloway *et al.*, 2000; Li *et al.*, 2015; Naue *et al.*, 2015) ages ranged from <20 to >80 years.

Curiously, amongst the 21 most prevalent mutations in PwMS, 5 had already been associated with MS, ranging from case reports to large epidemiological studies (Kalman, Lublin and Alder, 1996; Vyshkina *et al.*, 2008; Yu *et al.*, 2008; Andalib *et al.*, 2016; Yonova-Doing *et al.*, 2021). Furthermore, remarkably, the high number of deleterious variants in complex I and IV matches the pattern of decreased activity in these complexes in $CD4^+$ T cells of PwMS in previous studies (De Riccardis *et al.*, 2015, 2016). Moreover, the higher number of deleterious variants and cumulative deleterious burden from haplogroups J and T is consistent with previous reports of increased MS risk for these haplogroups (Yu *et al.*, 2008; Tranah *et al.*, 2015).

Nevertheless, mitochondrial genotype did not differ significantly between PwMS and HC, nor between V1 and V2 within PwMS. This is consistent with the two previous WGS studies for mtDNA in PwMS: (**I**) a 2016 study that analyzed mtDNA from blood and buccal swabs in 49 monozygotic twin pairs clinically discordant for MS, which did not find any significant differences between subjects (Souren *et al.*, 2016); and (**II**) a 2021 study that analyzed mtDNA from post-mortem brain tissue samples of 217 PwMS, namely, from snap

frozen tissue blocks from the cerebellum, with the aim of correlating deleterious cumulative burden from heteroplasmic variants with disease disability, which also did not find any significant differences (Pienaar *et al.*, 2021).

The current project was limited by its sample size, which was apparently insufficient for the previously set endpoint, namely, the discordance in prevalence for deleterious variants in Complex I. In contrast to the 35% difference registered in the pilot study with 20 triplets in 2018, the final difference was 18%, due to an increase in prevalence of 8% for HC, while in PwMS it decreased by 9%. Currently, more than 100 PwMS + HC pairs would be needed to detect a significant difference, with the final value in mind (Dupont, 1988). As a comparison, a recently published longitudinal study successfully addressing the role of mtDNA variants in Huntington's disease analyzed samples from 1,549 patients and 182 HC, where 169 samples were processed for the longitudinal section (Y. Wang *et al.*, 2021). Furthermore, variant calling was performed until a VL of 0.5% (for blood samples) and follow-up was longer (5–9 years).

The sample size might also be behind the lack of a statistically significant difference for most clinical variables per diagnosis change, apart from the MRI variables, which are usually the strongest predictors of MS activity and progression (Fisniku *et al.*, 2008; Kuhle *et al.*, 2015; H. Zhang *et al.*, 2019; Chung *et al.*, 2020; Kolčava *et al.*, 2020). Disease progression usually correlates with EDSS and MSFC (Cohen *et al.*, 2012), number of relapses and proximity to relapse (Langer-Gould *et al.*, 2006; Scalfari *et al.*, 2014; Alroughani *et al.*, 2015; Novotna *et al.*, 2015; Fambiatos *et al.*, 2020), and OCT data (Zimmermann *et al.*, 2018; Lambe, Saidha and Bermel, 2020; Bsteh *et al.*, 2021; Giedraitiene *et al.*, 2021).

Despite these limitations, however, this project has provided a very thorough analysis of mtDNA WGS in PwMS, and its sole longitudinal analysis, in addition to being the only project to link mitochondrial dysfunction in $CD4^+$ T cells directly to mitochondrial genotype. Moreover, longitudinal studies are rare in mtDNA NGS (Ortiz *et al.*, 2011; Arbeithuber *et al.*, 2020; Ma *et al.*, 2021; Penter *et al.*, 2021; Y. Wang *et al.*, 2021), and most studies dedicated to aging have performed cross-sectional comparisons alone (Cortopassi and Arnheim, 1990; Wei, 1992; Michikawa Yuichi *et al.*, 1999; Calloway *et al.*, 2000; Wang *et al.*, 2001; Williams *et al.*, 2013; Greaves *et al.*, 2014; Li *et al.*, 2015; Naue *et al.*, 2015; Kang *et al.*, 2016).

Overall, mitochondrial genotype does not seem to determine a higher risk of developing MS, nor a higher disease disability in PwMS. Furthermore, mitochondrial dysfunction in $CD4^+$ T cells is unlikely to derive from mitochondrial genotype.

Page intentionally left blank

VI — FUTURE PERSPECTIVES

Page intentionally left blank

In general, this project was successful in: (**I**) creating an open-source alternative approach to the existing Precision ID mtDNA Whole Genome Panel, with an output in the correct rCRS and a lower heteroplasmy threshold; (**II**) determining the cross-sectional profile of $CD4^+$ T cell mtDNA genomes of newly diagnosed PwMS and age- and sex-matched HC; and (**III**) assessing the longitudinal $CD4^+$ T cell mtDNA changes in PwMS, and correlate them with disease activity and progression.

In the end, however, there were no statistically significant differences between PwMS and HC, as well as between clinical visits V1 and V2 within PwMS, thereby pointing to a lack of association between mitochondrial genotype and phenotype in MS.

Nonetheless, since various studies have demonstrated that the peripheral immune system of PwMS, particularly $CD4^+$ T cells, shows mitochondrial dysfunction and/or loss of OXPHOS function (De Riccardis *et al.*, 2015, 2016; Haghikia *et al.*, 2015; La Rocca *et al.*, 2017; De Biasi *et al.*, 2019; D. Zhang *et al.*, 2019; Gonzalo *et al.*, 2019; Alissafi *et al.*, 2020; Armon-Omer *et al.*, 2020; De Rasmo *et al.*, 2020; Shin *et al.*, 2020; Esmael *et al.*, 2021; Wagner *et al.*, 2021), further studies should aim to address the interaction between mtDNA and nDNA (McManus *et al.*, 2019), as well as the pathways responsible for the transcriptional, post-transcriptional, and protein lifecycle regulation of ETC assembly and function (Sirey and Ponting, 2016; Boczonadi, Ricci and Horvath, 2018; D'Souza and Minczuk, 2018; Signes and Fernandez-Vizarra, 2018; Yambire *et al.*, 2019; Kummer and Ban, 2021; Vercellino and Sazanov, 2021).

Regarding the future of mtDNA NGS in general, the shift towards single-cell sequencing (Lareau *et al.*, 2021) is essential, since it circumvents the allelic tissue-specificity (Calloway *et al.*, 2000; He *et al.*, 2010; Li *et al.*, 2015; Naue *et al.*, 2015; Hahn and Zuryn, 2019; J. Xu *et al.*, 2019; Ludwig *et al.*, 2019). Incidentally, a recent study has demonstrated that somatic mutations in nDNA also show significant tissue-specificity (Moore *et al.*, 2021). Thus, this is not an issue reserved to mtDNA alone.

Finally, Genetic Medicine has lately undergone a revolution from a more passive and descriptive approach to a highly active and intervening one (Jameson and Kopp, 2018; Lindstrand *et al.*, 2019; Stranneheim *et al.*, 2021; Gorzynski *et al.*, 2022). Currently, genetic treatments have begun to bring hope to once doomed diseases (L. Xu *et al.*, 2019; Cehajic-Kapetanovic *et al.*, 2020; Gillmore *et al.*, 2021; Sahel *et al.*, 2021; Temperton, 2021), with mitochondrial diseases in the horizon (Bacman *et al.*, 2018; Gammage *et al.*, 2018; Silva-

Pinheiro and Minczuk, 2021). Thus, I believe it is safe to say: the future is genomics (Denny and Collins, 2021; Karlsson Linnér *et al.*, 2021; Q. Wang *et al.*, 2021).

VII — REFERENCES

Page intentionally left blank

1000 Genomes Project Consortium, Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A. and Abecasis, G.R. (2015) 'A global reference for human genetic variation', *Nature*, 526(7571), pp. 68–74. doi:10.1038/nature15393.

'A neuroimmune odyssey' (2020) *Nature Reviews Immunology*, 20(4), pp. 203–203. doi:10.1038/s41577-020-0293-6.

Absinta, M., Ha, S.-K., Nair, G., Sati, P., Luciano, N.J., Palisoc, M., Louveau, A., Zaghloul, K.A., Pittaluga, S., Kipnis, J. and Reich, D.S. (2017) 'Human and nonhuman primate meninges harbor lymphatic vessels that can be visualized noninvasively by MRI', *eLife*. Edited by H. Johansen-Berg, 6, p. e29738. doi:10.7554/eLife.29738.

Absinta, M., Lassmann, H. and Trapp, B.D. (2020) 'Mechanisms underlying progression in multiple sclerosis', *Current Opinion in Neurology*, 33(3), pp. 277–285. doi:10.1097/WCO.0000000000000818.

Absinta, M., Maric, D., Gharagozloo, M., Garton, T., Smith, M.D., Jin, J., Fitzgerald, K.C., Song, A., Liu, P., Lin, J.-P., Wu, T., Johnson, K.R., McGavern, D.B., Schafer, D.P., Calabresi, P.A. and Reich, D.S. (2021) 'A lymphocyte–microglia–astrocyte axis in chronic active multiple sclerosis', *Nature*, 597(7878), pp. 709–714. doi:10.1038/s41586-021-03892-7.

Adams, D.R. and Eng, C.M. (2018) 'Next-Generation Sequencing to Diagnose Suspected Genetic Disorders', *The New England Journal of Medicine*, 379(14), pp. 1353–1362. doi:10.1056/NEJMra1711801.

Aggarwal, S., Yurlova, L. and Simons, M. (2011) 'Central nervous system myelin: structure, synthesis and assembly', *Trends in Cell Biology*, 21(10), pp. 585–593. doi:10.1016/j.tcb.2011.06.004.

Aho, A.V., Kernighan, B.W. and Weinberger, P.J. (1979) 'Awk — a pattern scanning and processing language', *Software: Practice and Experience*, 9(4), pp. 267–279. doi:10.1002/spe.4380090403.

Alberts, B., Johnson, A., Lewis, J., Morgan, D., Raff, M., Roberts, K. and Walter, P. (2015) *Molecular Biology of the Cell*. Sixth edition. New York, NY: Garland Science, Taylor and Francis Group.

Alexeyev, M., Shokolenko, I., Wilson, G. and LeDoux, S. (2013) 'The Maintenance of Mitochondrial DNA Integrity—Critical Analysis and Update', *Cold Spring Harbor Perspectives in Biology*, 5(5). doi:10.1101/cshperspect.a012641.

Alfatni, A., Riou, M., Charles, A.-L., Meyer, A., Barnig, C., Andres, E., Lejay, A., Talha, S. and Geny, B. (2020) 'Peripheral Blood Mononuclear Cells and Platelets Mitochondrial Dysfunction, Oxidative Stress, and Circulating mtDNA in Cardiovascular Diseases', *Journal of Clinical Medicine*, 9(2), p. E311. doi:10.3390/jcm9020311.

Alissafi, T., Kalafati, L., Lazari, M., Filia, A., Kloukina, I., Manifava, M., Lim, J.-H., Alexaki, V.I., Ktistakis, N.T., Doskas, T., Garinis, G.A., Chavakis, T., Boumpas, D.T. and Verginis, P. (2020)

'Mitochondrial Oxidative Damage Underlies Regulatory T Cell Defects in Autoimmunity', *Cell Metabolism*, 32(4), pp. 591-604.e7. doi:10.1016/j.cmet.2020.07.001.

Almohmeed, Y.H., Avenell, A., Aucott, L. and Vickers, M.A. (2013) 'Systematic Review and Meta-Analysis of the Sero-Epidemiological Association between Epstein Barr Virus and Multiple Sclerosis', *PLoS ONE*. Edited by G. Khan, 8(4), p. e61110. doi:10.1371/journal.pone.0061110.

Alroughani, R.A., Akhtar, S., Ahmed, S.F. and Al-Hashel, J.Y. (2015) 'Clinical predictors of disease progression in multiple sclerosis patients with relapsing onset in a nation-wide cohort', *International Journal of Neuroscience*, 125(11), pp. 831–837. doi:10.3109/00207454.2014.976641.

Alves de Lima, K., Rustenhoven, J. and Kipnis, J. (2020) 'Meningeal Immunity and Its Function in Maintenance of the Central Nervous System in Health and Disease', *Annual Review of Immunology*, 38(1), pp. 597–620. doi:10.1146/annurev-immunol-102319-103410.

Amorim, A., Fernandes, T. and Taveira, N. (2019) 'Mitochondrial DNA in human identification: a review', *PeerJ*, 7, p. e7314. doi:10.7717/peerj.7314.

Amthor, F.R. (2020) 'Resting Membrane Potential & Action Potential', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242634 (Accessed: 22 December 2021).

Anand, S.K. and Tikoo, S.K. (2013) 'Viruses as Modulators of Mitochondrial Functions', *Advances in Virology*, 2013, pp. 1–17. doi:10.1155/2013/738794.

Andalib, S., Emamhadi, M., Yousefzadeh-Chabok, S., Salari, A., Sigaroudi, A.E. and Vafaee, M.S. (2016) 'MtDNA T4216C variation in multiple sclerosis: a systematic review and meta-analysis', *Acta Neurologica Belgica*, 116(4), pp. 439–443. doi:10.1007/s13760-016-0675-5.

Andalib, S., Talebi, M., Sakhinia, E., Farhoudi, M., Sadeghi-Bazargani, H. and Gjedde, A. (2015a) 'Lack of association between mitochondrial DNA G15257A and G15812A variations and multiple sclerosis', *Journal of the Neurological Sciences*, 356(1–2), pp. 102–106. doi:10.1016/j.jns.2015.06.022.

Andalib, S., Talebi, M., Sakhinia, E., Farhoudi, M., Sadeghi-Bazargani, H. and Gjedde, A. (2015b) 'Mitochondrial DNA T4216C and A4917G variations in multiple sclerosis', *Journal of the Neurological Sciences*, 356(1–2), pp. 55–60. doi:10.1016/j.jns.2015.04.050.

Andalib, S., Talebi, M., Sakhinia, E., Farhoudi, M., Sadeghi-Bazargani, H., Masoudian, N., Michel, T.M., Vafaee, M.S. and Gjedde, A. (2019) 'Mitochondrial DNA G15927A and G15928A variations in patients with multiple sclerosis', *Multiple Sclerosis and Related Disorders*, 27, pp. 9–12. doi:10.1016/j.msard.2018.09.004.

Andalib, S., Talebi, M., Sakhinia, E., Farhoudi, M., Sadeghi-Bazargani, H., Masoudian, N., Vafaee, M.S. and Gjedde, A. (2017) 'No evidence of association between optic neuritis and secondary LHON mtDNA mutations in patients with multiple sclerosis', *Mitochondrion*, 36, pp. 182–185. doi:10.1016/j.mito.2017.08.005.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn, M.H.L., Coulson, A.R., Drouin, J., Eperon, I.C., Nierlich, D.P., Roe, B.A., Sanger, F., Schreier, P.H., Smith, A.J.H., Staden, R. and Young, I.G. (1981) 'Sequence and organization of the human mitochondrial genome', *Nature*, 290(5806), pp. 457–465. doi:10.1038/290457a0.

Andravizou, A., Dardiotis, E., Artemiadis, A., Sokratous, M., Siokas, V., Tsouris, Z., Aloizou, A.-M., Nikolaidis, I., Bakirtzis, C., Tsivgoulis, G., Deretzi, G., Grigoriadis, N., Bogdanos, D.P. and Hadjigeorgiou, G.M. (2019) 'Brain atrophy in multiple sclerosis: mechanisms, clinical relevance and treatment options', *Autoimmunity Highlights*, 10(1), p. 7. doi:10.1186/s13317-019-0117-5.

Andrews, R.M., Kubacka, I., Chinnery, P.F., Lightowlers, R.N., Turnbull, D.M. and Howell, N. (1999) 'Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA', *Nature Genetics*, 23(2), pp. 147–147. doi:10.1038/13779.

Andrews, S., Krueger, F., Segonds-Pichon, A., Biggins, L., Krueger, C. and Wingett, S. (2010) *FastQC: a quality control tool for high throughput sequence data*. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (Accessed: 18 August 2021).

Arbeithuber, B., Hester, J., Cremona, M.A., Stoler, N., Zaidi, A., Higgins, B., Anthony, K., Chiaromonte, F., Diaz, F.J. and Makova, K.D. (2020) 'Age-related accumulation of de novo mitochondrial mutations in mammalian oocytes and somatic tissues', *PLOS Biology*. Edited by L.D. Hurst, 18(7), p. e3000745. doi:10.1371/journal.pbio.3000745.

Arendt, H. (1998) *The human condition*. 2nd ed. Chicago: University of Chicago Press.

Arloth, J., Eraslan, G., Andlauer, T.F.M., Martins, J., Iurato, S., Kühnel, B., Waldenberger, M., Frank, J., Gold, R., Hemmer, B., Luessi, F., Nischwitz, S., Paul, F., Wiendl, H., Gieger, C., Heilmann-Heimbach, S., Kacprowski, T., Laudes, M., Meitinger, T., Peters, A., Rawal, R., Strauch, K., Lucae, S., Müller-Myhsok, B., Rietschel, M., Theis, F.J., Binder, E.B. and Mueller, N.S. (2020) 'DeepWAS: Multivariate genotype-phenotype associations by directly integrating regulatory information using deep learning', *PLOS Computational Biology*. Edited by T.M. Przytycka, 16(2), p. e1007616. doi:10.1371/journal.pcbi.1007616.

Armon-Omer, A., Neuman, H., Sharabi-Nov, A. and Shahien, R. (2020) 'Mitochondrial activity is impaired in lymphocytes of MS patients in correlation with disease severity', *Multiple Sclerosis and Related Disorders*, 41, p. 102025. doi:10.1016/j.msard.2020.102025.

Arneth, B. (2021) 'Contributions of T cells in multiple sclerosis: what do we currently know?', *Journal of Neurology*, 268(12), pp. 4587–4593. doi:10.1007/s00415-020-10275-x.

Ashcroft, S.P., Bass, J.J., Kazi, A.A., Atherton, P.J. and Philp, A. (2020) 'The vitamin D receptor regulates mitochondrial function in C2C12 myoblasts', *American Journal of Physiology-Cell Physiology*, 318(3), pp. C536–C541. doi:10.1152/ajpcell.00568.2019.

Asin-Cayuela, J. and Gustafsson, C.M. (2007) 'Mitochondrial transcription and its regulation in mammalian cells', *Trends in Biochemical Sciences*, 32(3), pp. 111–117. doi:10.1016/j.tibs.2007.01.003.

Aspelund, A., Antila, S., Proulx, S.T., Karlsen, T.V., Karaman, S., Detmar, M., Wiig, H. and Alitalo, K. (2015) 'A dural lymphatic vascular system that drains brain interstitial fluid and macromolecules', *Journal of Experimental Medicine*, 212(7), pp. 991–999. doi:10.1084/jem.20142290.

Azevedo, F.A.C., Carvalho, L.R.B., Grinberg, L.T., Farfel, J.M., Ferretti, R.E.L., Leite, R.E.P., Jacob Filho, W., Lent, R. and Herculano-Houzel, S. (2009) 'Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain', *The Journal of Comparative Neurology*, 513(5), pp. 532–541. doi:10.1002/cne.21974.

Bacman, S.R., Kauppila, J.H.K., Pereira, C.V., Nissanka, N., Miranda, M., Pinto, M., Williams, S.L., Larsson, N.-G., Stewart, J.B. and Moraes, C.T. (2018) 'MitoTALEN reduces mutant mtDNA load and restores tRNAAla levels in a mouse model of heteroplasmic mtDNA mutation', *Nature Medicine*, 24(11), pp. 1696–1700. doi:10.1038/s41591-018-0166-8.

Bai, R., Cui, H., Devaney, J.M., Allis, K.M., Balog, A.M., Liu, X., Schnur, R.E., Shapiro, F.L., Brautbar, A., Estrada-Veras, J.I., Hochstetler, L., McConkie-Rosell, A., McDonald, M.T., Solomon, B.D., Hofherr, S., Richard, G. and Suchy, S.F. (2021) 'Interference of nuclear mitochondrial DNA segments in mitochondrial DNA testing resembles biparental transmission of mitochondrial DNA in humans', *Genetics in Medicine*, 23(8), pp. 1514–1521. doi:10.1038/s41436-021-01166-1.

Baker, D., Pryce, G., Amor, S., Giovannoni, G. and Schmierer, K. (2018) 'Learning from other autoimmunities to understand targeting of B cells to control multiple sclerosis', *Brain*, 141(10), pp. 2834–2847. doi:10.1093/brain/awy239.

Baker, D., Pryce, G., James, L.K., Schmierer, K. and Giovannoni, G. (2020) 'Failed B cell survival factor trials support the importance of memory B cells in multiple sclerosis', *European Journal of Neurology*, 27(2), pp. 221–228. doi:10.1111/ene.14105.

Bamford, J.M., Sandercock, P.A., Warlow, C.P. and Slattery, J. (1989) 'Interobserver agreement for the assessment of handicap in stroke patients', *Stroke*, 20(6), p. 828. doi:10.1161/01.str.20.6.828.

Ban, M., Elson, J., Walton, A., Turnbull, D., Compston, A., Chinnery, P. and Sawcer, S. (2008) 'Investigation of the role of mitochondrial DNA in multiple sclerosis susceptibility', *PloS One*, 3(8), p. e2891. doi:10.1371/journal.pone.0002891.

Bandelt, H.-J., Kloss-Brandstätter, A., Richards, M.B., Yao, Y.-G. and Logan, I. (2014) 'The case for the continuing use of the revised Cambridge Reference Sequence (rCRS) and the standardization of notation in human mitochondrial DNA studies', *Journal of Human Genetics*, 59(2), pp. 66–77. doi:10.1038/jhg.2013.120.

Bando, Y. and Geisler, J.G. (2019) 'Disease modifying mitochondrial uncouplers, MP101, and a slow release ProDrug, MP201, in models of Multiple Sclerosis', *Neurochemistry International*, 131, p. 104561. doi:10.1016/j.neuint.2019.104561.

Barkhof, F., Brück, W., De Groot, C.J.A., Bergers, E., Hulshof, S., Geurts, J., Polman, C.H. and van der Valk, P. (2003) 'Remyelinated Lesions in Multiple Sclerosis: Magnetic Resonance Image Appearance', *Archives of Neurology*, 60(8), pp. 1073–1081. doi:10.1001/archneur.60.8.1073.

Bar-Or, A. and Li, R. (2021) 'Cellular immunology of relapsing multiple sclerosis: interactions, checks, and balances', *The Lancet Neurology*, 20(6), pp. 470–483. doi:10.1016/S1474-4422(21)00063-6.

Barros-Silva, D., Marques, C.J., Henrique, R. and Jerónimo, C. (2018) 'Profiling DNA Methylation Based on Next-Generation Sequencing Approaches: New Insights and Clinical Applications', *Genes*, 9(9). doi:10.3390/genes9090429.

Bartholomäus, I., Kawakami, N., Odoardi, F., Schläger, C., Miljkovic, D., Ellwart, J.W., Klinkert, W.E.F., Flügel-Koch, C., Issekutz, T.B., Wekerle, H. and Flügel, A. (2009) 'Effector T cell interactions with meningeal vascular structures in nascent autoimmune CNS lesions', *Nature*, 462(7269), pp. 94–98. doi:10.1038/nature08478.

Beadnell, T.C., Fain, C., Vivian, C.J., King, J.C.G., Hastings, R., Markiewicz, M.A. and Welch, D.R. (2020) 'Mitochondrial genetics cooperate with nuclear genetics to selectively alter immune cell development/trafficking', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1866(5), p. 165648. doi:10.1016/j.bbadis.2019.165648.

Becher, B., Spath, S. and Goverman, J. (2017) 'Cytokine networks in neuroinflammation', *Nature Reviews Immunology*, 17(1), pp. 49–59. doi:10.1038/nri.2016.123.

Beecham, A., Amezcua, L., Chinea, A., Manrique, C., Rubi, C., Isobe, N., Lund, B., Santaniello, A., Beecham, G., Burchard, E., Comabella, M., Patsopoulos, N., Fitzgerald, K., Calabresi, P., De Jager, P., Conti, D., Delgado, S., Oksenberg, J. and McCauley, J. (2020) 'The genetic diversity of multiple sclerosis risk among Hispanic and African American populations living in the United States', *Multiple Sclerosis Journal*, 26(11), pp. 1329–1339. doi:10.1177/1352458519863764.

Bektas, A., Schurman, S.H., Gonzalez-Freire, M., Dunn, C.A., Singh, A.K., Macian, F., Cuervo, A.M., Sen, R. and Ferrucci, L. (2019) 'Age-associated changes in human CD4+ T cells point to mitochondrial dysfunction consequent to impaired autophagy', *Aging*, 11(21), pp. 9234–9263. doi:10.18632/aging.102438.

Belbasis, L., Bellou, V., Evangelou, E. and Tzoulaki, I. (2020) 'Environmental factors and risk of multiple sclerosis: Findings from meta-analyses and Mendelian randomization studies', *Multiple Sclerosis Journal*, 26(4), pp. 397–404. doi:10.1177/1352458519872664.

Bender, A., Krishnan, K.J., Morris, C.M., Taylor, G.A., Reeve, A.K., Perry, R.H., Jaros, E., Hersheson, J.S., Betts, J., Klopstock, T., Taylor, R.W. and Turnbull, D.M. (2006) 'High levels

of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease', *Nature Genetics*, 38(5), pp. 515–517. doi:10.1038/ng1769.

Ben-Nun, A., Wekerle, H. and Cohen, I.R. (1981) 'The rapid isolation of clonable antigen-specific T lymphocyte lines capable of mediating autoimmune encephalomyelitis', *European Journal of Immunology*, 11(3), pp. 195–199. doi:10.1002/eji.1830110307.

Berer, K., Mues, M., Koutrolos, M., Rasbi, Z.A., Boziki, M., Johner, C., Wekerle, H. and Krishnamoorthy, G. (2011) 'Commensal microbiota and myelin autoantigen cooperate to trigger autoimmune demyelination', *Nature*, 479(7374), pp. 538–541. doi:10.1038/nature10554.

Bergamaschi, R., Quaglini, S., Tavazzi, E., Amato, M.P., Paolicelli, D., Zipoli, V., Romani, A., Tortorella, C., Portaccio, E., D'Onghia, M., Garberi, F., Bargiggia, V. and Trojano, M. (2016) 'Immunomodulatory therapies delay disease progression in multiple sclerosis', *Multiple Sclerosis Journal*, 22(13), pp. 1732–1740. doi:10.1177/1352458512445941.

Berkowitz, A.L. (2016a) 'Demyelinating Diseases of the Central Nervous System', in *Clinical Neurology and Neuroanatomy: A Localization-Based Approach*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1160205018 (Accessed: 4 January 2022).

Berkowitz, A.L. (2016b) 'Overview of the Anatomy of the Nervous System', in *Clinical Neurology and Neuroanatomy: A Localization-Based Approach*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1160203432 (Accessed: 16 December 2021).

Berod, L., Friedrich, C., Nandan, A., Freitag, J., Hagemann, S., Harmrolfs, K., Sandouk, A., Hesse, C., Castro, C.N., Bähre, H., Tschirner, S.K., Gorinski, N., Gohmert, M., Mayer, C.T., Huehn, J., Ponimaskin, E., Abraham, W.-R., Müller, R., Lochner, M. and Sparwasser, T. (2014) 'De novo fatty acid synthesis controls the fate between regulatory T and T helper 17 cells', *Nature Medicine*, 20(11), pp. 1327–1333. doi:10.1038/nm.3704.

Betts, J.G., Young, K.A., Wise, J.A., Johnson, E., Poe, B., Kruse, D.H., Korol, O., Johnson, J.E., Womble, M. and DeSaix, P. (2013) 'The Adaptive Immune Response: T lymphocytes and Their Functional Types', in *Anatomy and Physiology*. Houston, Texas: OpenStax. Available at: https://openstax.org/books/anatomy-and-physiology/pages/21-3-the-adaptive-immune-response-t-lymphocytes-and-their-functional-types (Accessed: 28 December 2021).

Bhela, S., Kempsell, C., Manohar, M., Dominguez-Villar, M., Griffin, R., Bhatt, P., Kivisakk-Webb, P., Fuhlbrigge, R., Kupper, T., Weiner, H. and Baecher-Allan, C. (2015) 'Nonapoptotic and Extracellular Activity of Granzyme B Mediates Resistance to Regulatory T Cell (Treg) Suppression by HLA-DR^{-}CD25^{hi}CD127^{lo} Tregs in Multiple Sclerosis and in Response to IL-6', *The Journal of Immunology*, 194(5), p. 2180. doi:10.4049/jimmunol.1303257.

Bian, S., Hou, Y., Zhou, X., Li, X., Yong, J., Wang, Y., Wang, W., Yan, J., Hu, B., Guo, H., Wang, J., Gao, S., Mao, Y., Dong, J., Zhu, P., Xiu, D., Yan, L., Wen, L., Qiao, J., Tang, F. and Fu, W.

(2018) 'Single-cell multiomics sequencing and analyses of human colorectal cancer', *Science*, 362(6418), pp. 1060–1063. doi:10.1126/science.aao3791.

Bjornevik, K., Cortese, M., Healy, B.C., Kuhle, J., Mina, M.J., Leng, Y., Elledge, S.J., Niebuhr, D.W., Scher, A.I., Munger, K.L. and Ascherio, A. (2022) 'Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis', *Science*, p. science.abj8222. doi:10.1126/science.abj8222.

Bjørnevik, K., Riise, T., Casetta, I., Drulovic, J., Granieri, E., Holmøy, T., Kampman, M.T., Landtblom, A.-M., Lauer, K., Lossius, A., Magalhaes, S., Myhr, K.-M., Pekmezovic, T., Wesnes, K., Wolfson, C. and Pugliatti, M. (2014) 'Sun exposure and multiple sclerosis risk in Norway and Italy: The EnvIMS study', *Multiple Sclerosis Journal*, 20(8), pp. 1042–1049. doi:10.1177/1352458513513968.

Bjørnevik, K., Riise, T., Cortese, M., Holmøy, T., Kampman, M.T., Magalhaes, S., Myhr, K.-M., Wolfson, C. and Pugliatti, M. (2016) 'Level of education and multiple sclerosis risk after adjustment for known risk factors: The EnvIMS study', *Multiple Sclerosis Journal*, 22(1), pp. 104–111. doi:10.1177/1352458515579444.

Blagih, J., Coulombe, F., Vincent, E.E., Dupuy, F., Galicia-Vázquez, G., Yurchenko, E., Raissi, T.C., van der Windt, G.J.W., Viollet, B., Pearce, E.L., Pelletier, J., Piccirillo, C.A., Krawczyk, C.M., Divangahi, M. and Jones, R.G. (2015) 'The Energy Sensor AMPK Regulates T Cell Metabolic Adaptation and Effector Responses In Vivo', *Immunity*, 42(1), pp. 41–54. doi:10.1016/j.immuni.2014.12.030.

Boczonadi, V., Ricci, G. and Horvath, R. (2018) 'Mitochondrial DNA transcription and translation: clinical syndromes', *Essays in Biochemistry*. Edited by C. Garone and M. Minczuk, 62(3), pp. 321–340. doi:10.1042/EBC20170103.

Boldrini, V.O., Quintiliano, R.P.S., Silva, L.S., Damasceno, A., Santos, L.M.B. and Farias, A.S. (2021) 'Cytotoxic profile of CD3+CD20+ T cells in progressive multiple sclerosis', *Multiple Sclerosis and Related Disorders*, 52. doi:10.1016/j.msard.2021.103013.

Bolger, A.M., Lohse, M. and Usadel, B. (2014) 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30(15), pp. 2114–2120. doi:10.1093/bioinformatics/btu170.

Bolze, A., Mendez, F., White, S., Tanudjaja, F., Isaksson, M., Jiang, R., Rossi, A.D., Cirulli, E.T., Rashkin, M., Metcalf, W.J., Grzymski, J.J., Lee, W., Lu, J.T. and Washington, N.L. (2020) 'A catalog of homoplasmic and heteroplasmic mitochondrial DNA variants in humans', *bioRxiv*, p. 798264. doi:10.1101/798264.

Borden, E.C., Sen, G.C., Uze, G., Silverman, R.H., Ransohoff, R.M., Foster, G.R. and Stark, G.R. (2007) 'Interferons at age 50: past, current and future impact on biomedicine', *Nature Reviews Drug Discovery*, 6(12), pp. 975–990. doi:10.1038/nrd2422.

Böttcher, C., Fernández-Zapata, C., Schlickeiser, S., Kunkel, D., Schulz, A.R., Mei, H.E., Weidinger, C., Gieß, R.M., Asseyer, S., Siegmund, B., Paul, F., Ruprecht, K. and Priller, J. (2019) 'Multi-parameter immune profiling of peripheral blood mononuclear cells by

multiplexed single-cell mass cytometry in patients with early multiple sclerosis', *Scientific Reports*, 9(1), p. 19471. doi:10.1038/s41598-019-55852-x.

Bouafia, A., Golmard, J.-L., Thuries, V., Sazdovitch, V., Hauw, J.J., Fontaine, B. and Seilhean, D. (2014) 'Axonal expression of sodium channels and neuropathology of the plaques in multiple sclerosis', *Neuropathology and Applied Neurobiology*, 40(5), pp. 579–590. doi:10.1111/nan.12059.

Brandhagen, M.D., Just, R.S. and Irwin, J.A. (2020) 'Validation of NGS for mitochondrial DNA casework at the FBI Laboratory', *Forensic Science International. Genetics*, 44, p. 102151. doi:10.1016/j.fsigen.2019.102151.

Brändle, S.M., Obermeier, B., Senel, M., Bruder, J., Mentele, R., Khademi, M., Olsson, T., Tumani, H., Kristoferitsch, W., Lottspeich, F., Wekerle, H., Hohlfeld, R. and Dornmair, K. (2016) 'Distinct oligoclonal band antibodies in multiple sclerosis recognize ubiquitous self-proteins', *Proceedings of the National Academy of Sciences*, 113(28), p. 7864. doi:10.1073/pnas.1522730113.

Bray, A., Ismay, C., Chasnovski, E., Baumer, B. and Cetinkaya-Rundel, M. (2021) *infer: Tidy Statistical Inference*. Available at: https://CRAN.R-project.org/package=infer (Accessed: 2 February 2022).

Bris, C., Goudenege, D., Desquiret-Dumas, V., Charif, M., Colin, E., Bonneau, D., Amati-Bonneau, P., Lenaers, G., Reynier, P. and Procaccio, V. (2018) 'Bioinformatics Tools and Databases to Assess the Pathogenicity of Mitochondrial DNA Variants in the Field of Next Generation Sequencing', *Frontiers in Genetics*, 9, p. 632. doi:10.3389/fgene.2018.00632.

Brownlee, W.J., Hardy, T.A., Fazekas, F. and Miller, D.H. (2017) 'Diagnosis of multiple sclerosis: progress and challenges', *The Lancet*, 389(10076), pp. 1336–1346. doi:10.1016/S0140-6736(16)30959-X.

Brynskikh, A., Warren, T., Zhu, J. and Kipnis, J. (2008) 'Adaptive immunity affects learning behavior in mice', *Brain, Behavior, and Immunity*, 22(6), pp. 861–869. doi:10.1016/j.bbi.2007.12.008.

Bsteh, G., Berek, K., Hegen, H., Altmann, P., Wurth, S., Auer, M., Zinganell, A., Di Pauli, F., Rommer, P., Leutmezer, F., Deisenhammer, F. and Berger, T. (2021) 'Macular ganglion cell–inner plexiform layer thinning as a biomarker of disability progression in relapsing multiple sclerosis', *Multiple Sclerosis Journal*, 27(5), pp. 684–694. doi:10.1177/1352458520935724.

Bucchieri, F., Farina, F., Zummo, G. and Cappello, F. (2015) 'Lymphatic vessels of the dura mater: a new discovery?', *Journal of Anatomy*, 227(5), pp. 702–703. doi:10.1111/joa.12381.

Buck, M.D., O'Sullivan, D. and Pearce, E.L. (2015) 'T cell metabolism drives immunity', *Journal of Experimental Medicine*, 212(9), pp. 1345–1360. doi:10.1084/jem.20151159.

Buonvicino, D., Ranieri, G., Pratesi, S., Guasti, D. and Chiarugi, A. (2019) 'Neuroimmunological characterization of a mouse model of primary progressive

experimental autoimmune encephalomyelitis and effects of immunosuppressive or neuroprotective strategies on disease evolution', *Experimental Neurology*, 322, p. 113065. doi:10.1016/j.expneurol.2019.113065.

Burel, J.G., Pomaznoy, M., Lindestam Arlehamn, C.S., Weiskopf, D., da Silva Antunes, R., Jung, Y., Babor, M., Schulten, V., Seumois, G., Greenbaum, J.A., Premawansa, S., Premawansa, G., Wijewickrama, A., Vidanagama, D., Gunasena, B., Tippalagama, R., deSilva, A.D., Gilman, R.H., Saito, M., Taplitz, R., Ley, K., Vijayanand, P., Sette, A. and Peters, B. (2019) 'Circulating T cell-monocyte complexes are markers of immune perturbations', *eLife*, 8, p. e46045. doi:10.7554/eLife.46045.

Butler, J.M. (2012) 'Chapter 14 - Mitochondrial DNA Analysis', in Butler, J.M. (ed.) *Advanced Topics in Forensic DNA Typing: Methodology*. San Diego: Academic Press, pp. 405–456. doi:10.1016/B978-0-12-374513-2.00014-2.

Cai, N., Gomez-Duran, A., Yonova-Doing, E., Kundu, K., Burgess, A.I., Golder, Z.J., Calabrese, C., Bonder, M.J., Camacho, M., Lawson, R.A., Li, L., Williams-Gray, C.H., Di Angelantonio, E., Roberts, D.J., Watkins, N.A., Ouwehand, W.H., Butterworth, A.S., Stewart, I.D., Pietzner, M., Wareham, N.J., Langenberg, C., Danesh, J., Walter, K., Rothwell, P.M., Howson, J.M.M., Stegle, O., Chinnery, P.F., Soranzo, N., and ICICLE-PD Study Group (2021) 'Mitochondrial DNA variants modulate N-formylmethionine, proteostasis and risk of late-onset human diseases', *Nature Medicine*, 27(9), pp. 1564–1575. doi:10.1038/s41591-021-01441-3.

Calabrese, C., Simone, D., Diroma, M.A., Santorsola, M., Guttà, C., Gasparre, G., Picardi, E., Pesole, G. and Attimonelli, M. (2014) 'MToolBox: a highly automated pipeline for heteroplasmy annotation and prioritization analysis of human mitochondrial variants in high-throughput sequencing', *Bioinformatics (Oxford, England)*, 30(21), pp. 3115–3117. doi:10.1093/bioinformatics/btu483.

Calabrese, M., Magliozzi, R., Ciccarelli, O., Geurts, J.J.G., Reynolds, R. and Martin, R. (2015) 'Exploring the origins of grey matter damage in multiple sclerosis', *Nature Reviews Neuroscience*, 16(3), pp. 147–158. doi:10.1038/nrn3900.

Calloway, C.D., Reynolds, R.L., Herrin, G.L. and Anderson, W.W. (2000) 'The Frequency of Heteroplasmy in the HVII Region of mtDNA Differs across Tissue Types and Increases with Age', *The American Journal of Human Genetics*, 66(4), pp. 1384–1397. doi:10.1086/302844.

Calvi, A., Haider, L., Prados, F., Tur, C., Chard, D. and Barkhof, F. (2020) 'In vivo imaging of chronic active lesions in multiple sclerosis', *Multiple Sclerosis Journal*, p. 1352458520958589. doi:10.1177/1352458520958589.

Campbell, G., Licht-Mayer, S. and Mahad, D. (2019) 'Targeting mitochondria to protect axons in progressive MS', *Neuroscience Letters*, 710, p. 134258. doi:10.1016/j.neulet.2019.05.012.

Campbell, G. and Mahad, D. (2018) 'Neurodegeneration in Progressive Multiple Sclerosis', *Cold Spring Harbor Perspectives in Medicine*, 8(10), p. a028985. doi:10.1101/cshperspect.a028985.

Campbell, G.R., Kraytsberg, Y., Krishnan, K.J., Ohno, N., Ziabreva, I., Reeve, A., Trapp, B.D., Newcombe, J., Reynolds, R., Lassmann, H., Khrapko, K., Turnbull, D.M. and Mahad, D.J. (2012) 'Clonally expanded mitochondrial DNA deletions within the choroid plexus in multiple sclerosis', *Acta Neuropathologica*, 124(2), pp. 209–220. doi:10.1007/s00401-012-1001-9.

Campbell, G.R., Ziabreva, I., Reeve, A.K., Krishnan, K.J., Reynolds, R., Howell, O., Lassmann, H., Turnbull, D.M. and Mahad, D.J. (2011) 'Mitochondrial DNA deletions and neurodegeneration in multiple sclerosis', *Annals of Neurology*, 69(3), pp. 481–492. doi:10.1002/ana.22109.

Can, E., Richter, F., Valchanova, R. and Dewey, M. (2016) 'Supervisors' perspective on medical thesis projects and dropout rates: survey among thesis supervisors at a large German university hospital', *BMJ open*, 6(10), p. e012726. doi:10.1136/bmjopen-2016-012726.

Canale, L.C., Parson, W. and Holland, M.M. (2022) 'The time is now for ubiquitous forensic MTMPS analysis', *WIREs Forensic Science*, 4(1). doi:10.1002/wfs2.1431.

Cardanho-Ramos, C. and Morais, V.A. (2021) 'Mitochondrial Biogenesis in Neurons: How and Where', *International Journal of Molecular Sciences*, 22(23), p. 13059. doi:10.3390/ijms222313059.

Cassell, D.J. and Schwartz, R.H. (1994) 'A quantitative analysis of antigen-presenting cell function: activated B cells stimulate naive CD4 T cells but are inferior to dendritic cells in providing costimulation.', *Journal of Experimental Medicine*, 180(5), pp. 1829–1840. doi:10.1084/jem.180.5.1829.

Castellana, S., Biagini, T., Petrizzelli, F., Parca, L., Panzironi, N., Caputo, V., Vescovi, A.L., Carella, M. and Mazza, T. (2021) 'MitImpact 3: modeling the residue interaction network of the Respiratory Chain subunits', *Nucleic Acids Research*, 49(D1), pp. D1282–D1288. doi:10.1093/nar/gkaa1032.

Castellana, S., Fusilli, C., Mazzoccoli, G., Biagini, T., Capocefalo, D., Carella, M., Vescovi, A.L. and Mazza, T. (2017) 'High-confidence assessment of functional impact of human mitochondrial non-synonymous genome variations by APOGEE', *PLoS computational biology*, 13(6), p. e1005628. doi:10.1371/journal.pcbi.1005628.

Castellazzi, M., Patergnani, S., Donadio, M., Giorgi, C., Bonora, M., Fainardi, E., Casetta, I., Granieri, E., Pugliatti, M. and Pinton, P. (2019) 'Correlation between auto/mitophagic processes and magnetic resonance imaging activity in multiple sclerosis patients', *Journal of Neuroinflammation*, 16(1), p. 131. doi:10.1186/s12974-019-1526-0.

Cehajic-Kapetanovic, J., Xue, K., Martinez-Fernandez de la Camara, C., Nanda, A., Davies, A., Wood, L.J., Salvetti, A.P., Fischer, M.D., Aylward, J.W., Barnard, A.R., Jolly, J.K., Luo, E., Lujan, B.J., Ong, T., Girach, A., Black, G.C.M., Gregori, N.Z., Davis, J.L., Rosa, P.R., Lotery, A.J., Lam, B.L., Stanga, P.E. and MacLaren, R.E. (2020) 'Initial results from a first-in-human gene therapy trial on X-linked retinitis pigmentosa caused by mutations in RPGR', *Nature Medicine*, 26(3), pp. 354–359. doi:10.1038/s41591-020-0763-1.

Chaitanya, L., Ralf, A., van Oven, M., Kupiec, T., Chang, J., Lagacé, R. and Kayser, M. (2015) 'Simultaneous Whole Mitochondrial Genome Sequencing with Short Overlapping Amplicons Suitable for Degraded DNA Using the Ion Torrent Personal Genome Machine', *Human Mutation*, 36(12), pp. 1236–1247. doi:10.1002/humu.22905.

Chalmers, R.M., Harding, A.E., Robertson, N. and Compston, D.A.S. (1996) 'Sequence of mitochondrial DNA in patients with multiple sclerosis', *Annals of Neurology*, 40(2), pp. 239–243. doi:10.1002/ana.410400216.

Chang, C.-H., Curtis, J.D., Maggi, L.B., Faubert, B., Villarino, A.V., O'Sullivan, D., Huang, S.C.-C., van der Windt, G.J.W., Blagih, J., Qiu, J., Weber, J.D., Pearce, E.J., Jones, R.G. and Pearce, E.L. (2013) 'Posttranscriptional Control of T Cell Effector Function by Aerobic Glycolysis', *Cell*, 153(6), pp. 1239–1251. doi:10.1016/j.cell.2013.05.016.

Chang, W. (2014) *extrafont: Tools for using fonts*. Available at: https://CRAN.R-project.org/package=extrafont (Accessed: 2 February 2022).

Chard, D.T., Griffin, C.M., Parker, G.J.M., Kapoor, R., Thompson, A.J. and Miller, D.H. (2002) 'Brain atrophy in clinically early relapsing–remitting multiple sclerosis', *Brain*, 125(2), pp. 327–337. doi:10.1093/brain/awf025.

Charite University, Berlin, Germany (2011) *Clinically Isolated Syndrome and Newly Diagnosed Multiple Sclerosis: Diagnostic, Prognostic and Therapy - Response Markers - a Prospective Observational Study (Berlin CIS-COHORT)*. Clinical trial registration NCT01371071. clinicaltrials.gov. Available at: https://clinicaltrials.gov/ct2/show/NCT01371071 (Accessed: 23 November 2021).

Chen, J., Zheng, Q., Peiffer, L.B., Hicks, J.L., Haffner, M.C., Rosenberg, A.Z., Levi, M., Wang, X.X., Ozbek, B., Baena-Del Valle, J., Yegnasubramanian, S. and De Marzo, A.M. (2020) 'An in Situ Atlas of Mitochondrial DNA in Mammalian Tissues Reveals High Content in Stem and Proliferative Compartments', *The American Journal of Pathology*, 190(7), pp. 1565–1579. doi:10.1016/j.ajpath.2020.03.018.

Chervova, O., Conde, L., Guerra-Assunção, J.A., Moghul, I., Webster, A.P., Berner, A., Larose Cadieux, E., Tian, Y., Voloshin, V., Jesus, T.F., Hamoudi, R., Herrero, J. and Beck, S. (2019) 'The Personal Genome Project-UK, an open access resource of human multi-omics data', *Scientific Data*, 6(1), p. 257. doi:10.1038/s41597-019-0205-4.

Cho, S., Kim, M.Y., Lee, J.H. and Lee, S.D. (2018) 'Assessment of mitochondrial DNA heteroplasmy detected on commercial panel using MPS system with artificial mixture samples', *International Journal of Legal Medicine*, 132(4), pp. 1049–1056. doi:10.1007/s00414-017-1755-7.

Chocron, E.S., Munkácsy, E. and Pickering, A.M. (2019) 'Cause or casualty: The role of mitochondrial DNA in aging and age-associated disease', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1865(2), pp. 285–297. doi:10.1016/j.bbadis.2018.09.035.

Chung, C.-Y., Valdebenito, G.E., Chacko, A.R. and Duchen, M.R. (2021) 'Rewiring cell signalling pathways in pathogenic mtDNA mutations', *Trends in Cell Biology*, p. S0962892421002075. doi:10.1016/j.tcb.2021.10.005.

Chung, K.K., Altmann, D., Barkhof, F., Miszkiel, K., Brex, P.A., O'Riordan, J., Ebner, M., Prados, F., Cardoso, M.J., Vercauteren, T., Ourselin, S., Thompson, A., Ciccarelli, O. and Chard, D.T. (2020) 'A 30-Year Clinical and Magnetic Resonance Imaging Observational Study of Multiple Sclerosis and Clinically Isolated Syndromes', *Annals of Neurology*, 87(1), pp. 63–74. doi:10.1002/ana.25637.

Churchill, J.D., Stoljarova, M., King, J.L. and Budowle, B. (2018) 'Massively parallel sequencing-enabled mixture analysis of mitochondrial DNA samples', *International Journal of Legal Medicine*, 132(5), pp. 1263–1272. doi:10.1007/s00414-018-1799-3.

Cihlar, J.C., Amory, C., Lagacé, R., Roth, C., Parson, W. and Budowle, B. (2020) 'Developmental Validation of a MPS Workflow with a PCR-Based Short Amplicon Whole Mitochondrial Genome Panel', *Genes*, 11(11), p. E1345. doi:10.3390/genes11111345.

Cihlar, J.C., Strobl, C., Lagacé, R., Muenzler, M., Parson, W. and Budowle, B. (2020) 'Distinguishing mitochondrial DNA and NUMT sequences amplified with the precision ID mtDNA whole genome panel', *Mitochondrion*, 55, pp. 122–133. doi:10.1016/j.mito.2020.09.001.

Clark, G., Stockinger, H., Balderas, R., van Zelm, M.C., Zola, H., Hart, D. and Engel, P. (2016) 'Nomenclature of CD molecules from the Tenth Human Leucocyte Differentiation Antigen Workshop', *Clinical & Translational Immunology*, 5(1), p. e57. doi:10.1038/cti.2015.38.

Cobb, B.A., Wang, Q., Tzianabos, A.O. and Kasper, D.L. (2004) 'Polysaccharide processing and presentation by the MHCII pathway', *Cell*, 117(5), pp. 677–687. doi:10.1016/j.cell.2004.05.001.

Cobb, L.J., Lee, C., Xiao, J., Yen, K., Wong, R.G., Nakamura, H.K., Mehta, H.H., Gao, Q., Ashur, C., Huffman, D.M., Wan, J., Muzumdar, R., Barzilai, N. and Cohen, P. (2016) 'Naturally occurring mitochondrial-derived peptides are age-dependent regulators of apoptosis, insulin sensitivity, and inflammatory markers', *Aging*, 8(4), pp. 796–809. doi:10.18632/aging.100943.

Cohen, J.A., Reingold, S.C., Polman, C.H. and Wolinsky, J.S. (2012) 'Disability outcome measures in multiple sclerosis clinical trials: current status and future prospects', *The Lancet Neurology*, 11(5), pp. 467–476. doi:10.1016/S1474-4422(12)70059-5.

Cohen, J.I. (2018) 'Epstein-Barr Virus Infections, Including Infectious Mononucleosis', in Jameson, J.L., Fauci, A.S., Kasper, D.L., Hauser, S.L., Longo, D.L., and Loscalzo, J. (eds) *Harrison's Principles of Internal Medicine, 20e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1160014147 (Accessed: 10 January 2022).

Collin, M. and Bigley, V. (2018) 'Human dendritic cell subsets: an update', *Immunology*, 154(1), pp. 3–20. doi:10.1111/imm.12888.

Colombo, E. and Farina, C. (2016) 'Astrocytes: Key Regulators of Neuroinflammation', *Trends in Immunology*, 37(9), pp. 608–620. doi:10.1016/j.it.2016.06.006.

Comi, G., Bar-Or, A., Lassmann, H., Uccelli, A., Hartung, H.-P., Montalban, X., Sørensen, P.S., Hohlfeld, R., Hauser, S.L., and Expert Panel of the 27th Annual Meeting of the European Charcot Foundation (2021) 'Role of B Cells in Multiple Sclerosis and Related Disorders', *Annals of Neurology*, 89(1), pp. 13–23. doi:10.1002/ana.25927.

Consiglio, M., Destefanis, M., Morena, D., Foglizzo, V., Forneris, M., Pescarmona, G. and Silvagno, F. (2014) 'The Vitamin D Receptor Inhibits the Respiratory Chain, Contributing to the Metabolic Switch that Is Essential for Cancer Cell Proliferation', *PLoS ONE*. Edited by M. Makishima, 9(12), p. e115816. doi:10.1371/journal.pone.0115816.

Constantinescu, C.S., Farooqi, N., O'Brien, K. and Gran, B. (2011) 'Experimental autoimmune encephalomyelitis (EAE) as a model for multiple sclerosis (MS)', *British Journal of Pharmacology*, 164(4), pp. 1079–1106. doi:10.1111/j.1476-5381.2011.01302.x.

Correale, J., Marrodan, M. and Ysrraelit, M. (2019) 'Mechanisms of Neurodegeneration and Axonal Dysfunction in Progressive Multiple Sclerosis', *Biomedicines*, 7(1), p. 14. doi:10.3390/biomedicines7010014.

Cortese, M., Bjornevik, K., Chitnis, T., Ascherio, A. and Munger, K.L. (2022) 'Aging with multiple sclerosis: A longitudinal study of physical function, mental health, and memory in two cohorts of US women', *Multiple Sclerosis Journal*, 28(1), pp. 121–131. doi:10.1177/13524585211007739.

Cortese, M., Munger, K.L., Martínez-Lapiscina, E.H., Barro, C., Edan, G., Freedman, M.S., Hartung, H.-P., Montalbán, X., Foley, F.W., Penner, I.K., Hemmer, B., Fox, E.J., Schippling, S., Wicklein, E.-M., Kappos, L., Kuhle, J., Ascherio, A., and on behalf of the BENEFIT Study Group (2020) 'Vitamin D, smoking, EBV, and long-term cognitive performance in MS: 11-year follow-up of BENEFIT', *Neurology*, 94(18), pp. e1950–e1960. doi:10.1212/WNL.0000000000009371.

Cortese, M., Riise, T., Bjørnevik, K., Holmøy, T., Kampman, M.T., Magalhaes, S., Pugliatti, M., Wolfson, C. and Myhr, K.-M. (2015) 'Timing of use of cod liver oil, a vitamin D source, and multiple sclerosis risk: The EnvIMS study', *Multiple Sclerosis Journal*, 21(14), pp. 1856–1864. doi:10.1177/1352458515578770.

Cortes-Figueiredo, F., Carvalho, F.S., Fonseca, A.C., Paul, F., Ferro, J.M., Schönherr, S., Weissensteiner, H. and Morais, V.A. (2021) 'From Forensics to Clinical Research: Expanding the Variant Calling Pipeline for the Precision ID mtDNA Whole Genome Panel', *International Journal of Molecular Sciences*, 22(21), p. 12031. doi:10.3390/ijms222112031.

Cortes-Figueiredo, F., Morais, V.A. and Pinheiro, H. (2021) 'From Bedside to Bench: Methods in Precision Medicine', in Fonseca, A.C. and Ferro, J.M. (eds) *Precision Medicine in Stroke*. Cham: Springer International Publishing, pp. 289–307. doi:10.1007/978-3-030-70761-3_12.

Cortopassi, G.A. and Arnheim, N. (1990) 'Detection of a specific mitochondrial DNA deletion in tissues of older humans', *Nucleic Acids Research*, 18(23), pp. 6927–6933. doi:10.1093/nar/18.23.6927.

Cossu, D., Yokoyama, K., Sechi, L.A. and Hattori, N. (2021) 'Potential of PINK1 and PARKIN Proteins as Biomarkers for Active Multiple Sclerosis: A Japanese Cohort Study', *Frontiers in Immunology*, 12, p. 3172. doi:10.3389/fimmu.2021.681386.

Craner, M.J., Newcombe, J., Black, J.A., Hartle, C., Cuzner, M.L. and Waxman, S.G. (2004) 'Molecular changes in neurons in multiple sclerosis: Altered axonal expression of $Na_v1.2$ and $Na_v1.6$ sodium channels and $Na^+/Ca^{2}+$ exchanger', *Proceedings of the National Academy of Sciences of the United States of America*, 101(21), p. 8168. doi:10.1073/pnas.0402765101.

Cruciani, C., Puthenparampil, M., Tomas-Ojer, P., Jelcic, Ivan, Docampo, M.J., Planas, R., Manogaran, P., Opfer, R., Wicki, C., Reindl, M., Jelcic, Ilijas, Lutterotti, A., Martin, R. and Sospedra, M. (2021) 'T-Cell Specificity Influences Disease Heterogeneity in Multiple Sclerosis', *Neurology - Neuroimmunology Neuroinflammation*, 8(6), p. e1075. doi:10.1212/NXI.0000000000001075.

Cuenca, D., Battaglia, J., Halsing, M. and Sheehan, S. (2020) 'Mitochondrial Sequencing of Missing Persons DNA Casework by Implementing Thermo Fisher's Precision ID mtDNA Whole Genome Assay', *Genes*, 11(11), p. E1303. doi:10.3390/genes11111303.

Cutter, G.R., Baier, M.L., Rudick, R.A., Cookfair, D.L., Fischer, J.S., Petkau, J., Syndulko, K., Weinshenker, B.G., Antel, J.P., Confavreux, C., Ellison, G.W., Lublin, F., Miller, A.E., Rao, S.M., Reingold, S., Thompson, A. and Willoughby, E. (1999) 'Development of a multiple sclerosis functional composite as a clinical trial outcome measure', *Brain: A Journal of Neurology*, 122 (Pt 5), pp. 871–882. doi:10.1093/brain/122.5.871.

D'Agata, F., Ruffinatti, F.A., Boschi, S., Stura, I., Rainero, I., Abollino, O., Cavalli, R. and Guiot, C. (2017) 'Magnetic Nanoparticles in the Central Nervous System: Targeting Principles, Applications and Safety Issues', *Molecules (Basel, Switzerland)*, 23(1), p. E9. doi:10.3390/molecules23010009.

Dantzer, R. (2018) 'Neuroimmune Interactions: From the Brain to the Immune System and Vice Versa', *Physiological Reviews*, 98(1), pp. 477–504. doi:10.1152/physrev.00039.2016.

Davies, N.M., Holmes, M.V. and Davey Smith, G. (2018) 'Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians', *BMJ*, 362, p. k601. doi:10.1136/bmj.k601.

De Biasi, S., Simone, A.M., Bianchini, E., Lo Tartaro, D., Pecorini, S., Nasi, M., Patergnani, S., Carnevale, G., Gibellini, L., Ferraro, D., Vitetta, F., Pinton, P., Sola, P., Cossarizza, A. and Pinti, M. (2019) 'Mitochondrial functionality and metabolism in T cells from progressive multiple sclerosis patients', *European Journal of Immunology*, 49(12), pp. 2204–2221. doi:10.1002/eji.201948223.

De Rasmo, D., Ferretta, A., Russo, S., Ruggieri, M., Lasorella, P., Paolicelli, D., Trojano, M. and Signorile, A. (2020) 'PBMC of Multiple Sclerosis Patients Show Deregulation of OPA1 Processing Associated with Increased ROS and PHB2 Protein Levels', *Biomedicines*, 8(4), p. 85. doi:10.3390/biomedicines8040085.

De Riccardis, L., Ferramosca, A., Danieli, A., Trianni, G., Zara, V., De Robertis, F. and Maffia, M. (2016) 'Metabolic response to glatiramer acetate therapy in multiple sclerosis patients', *BBA Clinical*, 6, pp. 131–137. doi:10.1016/j.bbacli.2016.10.004.

De Riccardis, L., Rizzello, A., Ferramosca, A., Urso, E., De Robertis, F., Danieli, A., Giudetti, A.M., Trianni, G., Zara, V. and Maffia, M. (2015) 'Bioenergetics profile of CD4 + T cells in relapsing remitting multiple sclerosis subjects', *Journal of Biotechnology*, 202, pp. 31–39. doi:10.1016/j.jbiotec.2015.02.015.

De Rosa, V., Galgani, M., Porcellini, A., Colamatteo, A., Santopaolo, M., Zuchegna, C., Romano, A., De Simone, S., Procaccini, C., La Rocca, C., Carrieri, P.B., Maniscalco, G.T., Salvetti, M., Buscarinu, M.C., Franzese, A., Mozzillo, E., La Cava, A. and Matarese, G. (2015) 'Glycolysis controls the induction of human regulatory T cells by modulating the expression of FOXP3 exon 2 splicing variants', *Nature Immunology*, 16(11), pp. 1174–1184. doi:10.1038/ni.3269.

Dendrou, C.A., Fugger, L. and Friese, M.A. (2015) 'Immunopathology of multiple sclerosis', *Nature Reviews Immunology*, 15(9), pp. 545–558. doi:10.1038/nri3871.

Denny, J.C. and Collins, F.S. (2021) 'Precision medicine in 2030—seven ways to transform healthcare', *Cell*, 184(6), pp. 1415–1419. doi:10.1016/j.cell.2021.01.015.

Deverman, B.E. and Patterson, P.H. (2009) 'Cytokines and CNS Development', *Neuron*, 64(1), pp. 61–78. doi:10.1016/j.neuron.2009.09.002.

Diebold, M., Sievers, C., Bantug, G., Sanderson, N., Kappos, L., Kuhle, J., Lindberg, R.L.P. and Derfuss, T. (2018) 'Dimethyl fumarate influences innate and adaptive immunity in multiple sclerosis', *Journal of Autoimmunity*, 86, pp. 39–50. doi:10.1016/j.jaut.2017.09.009.

Dikalov, S., Itani, H., Richmond, B., Arslanbaeva, L., Vergeade, A., Rahman, S.M.J., Boutaud, O., Blackwell, T., Massion, P.P., Harrison, D.G. and Dikalova, A. (2019) 'Tobacco smoking induces cardiovascular mitochondrial oxidative stress, promotes endothelial dysfunction, and enhances hypertension', *American Journal of Physiology-Heart and Circulatory Physiology*, 316(3), pp. H639–H646. doi:10.1152/ajpheart.00595.2018.

Dinarello, C.A. (2007) 'Historical insights into cytokines', *European Journal of Immunology*, 37 Suppl 1, pp. S34-45. doi:10.1002/eji.200737772.

Ding, J., Sidore, C., Butler, T.J., Wing, M.K., Qian, Y., Meirelles, O., Busonero, F., Tsoi, L.C., Maschio, A., Angius, A., Kang, H.M., Nagaraja, R., Cucca, F., Abecasis, G.R. and Schlessinger, D. (2015) 'Assessing Mitochondrial DNA Variation and Copy Number in Lymphocytes of ~2,000 Sardinians Using Tailored Sequencing Analysis Tools', *PLOS Genetics*. Edited by G. Gibson, 11(7), p. e1005306. doi:10.1371/journal.pgen.1005306.

Disanto, G., Zecca, C., MacLachlan, S., Sacco, R., Handunnetthi, L., Meier, U.C., Simpson, A., McDonald, L., Rossi, A., Benkert, P., Kuhle, J., Ramagopalan, S.V. and Gobbi, C. (2018) 'Prodromal symptoms of multiple sclerosis in primary care', *Annals of Neurology*, 83(6), pp. 1162–1173. doi:10.1002/ana.25247.

Divakaruni, A.S. and Brand, M.D. (2011) 'The regulation and physiology of mitochondrial proton leak', *Physiology (Bethesda, Md.)*, 26(3), pp. 192–205. doi:10.1152/physiol.00046.2010.

Dominguez-Villar, M., Baecher-Allan, C.M. and Hafler, D.A. (2011) 'Identification of T helper type 1–like, Foxp3+ regulatory T cells in human autoimmune disease', *Nature Medicine*, 17(6), pp. 673–675. doi:10.1038/nm.2389.

Dong, Y. and Benveniste, E.N. (2001) 'Immune function of astrocytes', *Glia*, 36(2), pp. 180–190. doi:10.1002/glia.1107.

Dorst, J., Fangerau, T., Taranu, D., Eichele, P., Dreyhaupt, J., Michels, S., Schuster, J., Ludolph, A.C., Senel, M. and Tumani, H. (2019) 'Safety and efficacy of immunoadsorption versus plasma exchange in steroid-refractory relapse of multiple sclerosis and clinically isolated syndrome: A randomised, parallel-group, controlled trial', *EClinicalMedicine*, 16, pp. 98–106. doi:10.1016/j.eclinm.2019.10.017.

D'Souza, A.R. and Minczuk, M. (2018) 'Mitochondrial transcription and translation: overview', *Essays in Biochemistry*. Edited by C. Garone and M. Minczuk, 62(3), pp. 309–320. doi:10.1042/EBC20170102.

Duan, S., Lv, Z., Fan, X., Wang, L., Han, F., Wang, H. and Bi, S. (2014) 'Vitamin D status and the risk of multiple sclerosis: A systematic review and meta-analysis', *Neuroscience Letters*, 570, pp. 108–113. doi:10.1016/j.neulet.2014.04.021.

Dupont, W.D. (1988) 'Power calculations for matched case-control studies', *Biometrics*, 44(4), pp. 1157–1168.

Dür, A., Huber, N. and Parson, W. (2021) 'Fine-Tuning Phylogenetic Alignment and Haplogrouping of mtDNA Sequences', *International Journal of Molecular Sciences*, 22(11), p. 5747. doi:10.3390/ijms22115747.

Dyment, D.A., Yee, I.M.L., Ebers, G.C. and Sadovnick, A.D. (2006) 'Multiple sclerosis in stepsiblings: recurrence risk and ascertainment', *Journal of Neurology, Neurosurgery & Psychiatry*, 77(2), p. 258. doi:10.1136/jnnp.2005.063008.

Dziedzic, T., Metz, I., Dallenga, T., König, F.B., Müller, S., Stadelmann, C. and Brück, W. (2010) 'Wallerian Degeneration: A Major Component of Early Axonal Pathology in Multiple Sclerosis', *Brain Pathology*, 20(5), pp. 976–985. doi:10.1111/j.1750-3639.2010.00401.x.

Ebers, G.C., Bulman, D.E., Sadovnick, A.D., Paty, D.W., Warren, S., Hader, W., Murray, T.J., Seland, T.P., Duquette, P., Grey, T., Nelson, R., Nicolle, M. and Brunet, D. (1986) 'A Population-Based Study of Multiple Sclerosis in Twins', *New England Journal of Medicine*, 315(26), pp. 1638–1642. doi:10.1056/NEJM198612253152603.

Ebers, G.C., Sadovnick, A.D. and Risch, N.J. (1995) 'A genetic basis for familial aggregation in multiple sclerosis', *Nature*, 377(6545), pp. 150–151. doi:10.1038/377150a0.

Ebers, G.C., Yee, I.M.L., Sadovnick, A.D. and Duquette, P. (2000) 'Conjugal multiple sclerosis: Population-based prevalence and recurrence risks in offspring', *Annals of Neurology*, 48(6), pp. 927–931. doi:10.1002/1531-8249(200012)48:6<927::AID-ANA14>3.0.CO;2-F.

Ellulu, M.S., Patimah, I., Khaza'ai, H., Rahmat, A. and Abed, Y. (2017) 'Obesity and inflammation: the linking mechanism and the complications', *Archives of Medical Science*, 4, pp. 851–863. doi:10.5114/aoms.2016.58928.

Enge, M., Arda, H.E., Mignardi, M., Beausang, J., Bottino, R., Kim, S.K. and Quake, S.R. (2017) 'Single-Cell Analysis of Human Pancreas Reveals Transcriptional Signatures of Aging and Somatic Mutation Patterns', *Cell*, 171(2), pp. 321-330.e14. doi:10.1016/j.cell.2017.09.004.

Engelhard, J., Oleske, D.M., Schmitting, S., Wells, K.E., Talapala, S. and Barbato, L.M. (2021) 'Multiple sclerosis by phenotype in Germany', *Multiple Sclerosis and Related Disorders*, p. 103326. doi:10.1016/j.msard.2021.103326.

Engelhardt, B. and Ransohoff, R.M. (2012) 'Capture, crawl, cross: the T cell code to breach the blood–brain barriers', *Trends in Immunology*, 33(12), pp. 579–589. doi:10.1016/j.it.2012.07.004.

Engelhardt, B., Vajkoczy, P. and Weller, R.O. (2017) 'The movers and shapers in immune privilege of the CNS', *Nature Immunology*, 18(2), pp. 123–131. doi:10.1038/ni.3666.

Erlangsen, A., Stenager, Egon, Conwell, Y., Andersen, P.K., Hawton, K., Benros, M.E., Nordentoft, M. and Stenager, Elsebeth (2020) 'Association Between Neurological Disorders and Death by Suicide in Denmark', *JAMA*, 323(5), p. 444. doi:10.1001/jama.2019.21834.

Esmael, A., Talaat, M., Egila, H. and Eltoukhy, K. (2021) 'Mitochondrial dysfunction and serum lactate as a biomarker for the progression and disability in MS and its correlation with the radiological findings', *Neurological Research*, 43(7), pp. 582–590. doi:10.1080/01616412.2021.1893567.

von Essen, M.R., Ammitzbøll, C., Hansen, R.H., Petersen, E.R.S., McWilliam, O., Marquart, H.V., Damm, P. and Sellebjerg, F. (2019) 'Proinflammatory CD20+ T cells in the pathogenesis of multiple sclerosis', *Brain*, 142(1), pp. 120–132. doi:10.1093/brain/awy301.

European Medicines Agency (2022) *Medicines | European Medicines Agency*, *European Medicines Agency*. Available at: https://www.ema.europa.eu/en/medicines/field_ema_web_categories%253Aname_field/Human/search_api_aggregation_ema_therapeutic_area_name/Multiple%20Sclerosis/search_api_aggregation_ema_therapeutic_area_name/Multiple%20Sclerosis%2C%20Relapsing-Remitting (Accessed: 6 January 2022).

European Organization For Nuclear Research and OpenAIRE (2013) 'Zenodo: Research. Shared.' doi:10.25495/7GXK-RD71.

Evans, T.M., Bira, L., Gastelum, J.B., Weiss, L.T. and Vanderford, N.L. (2018) 'Evidence for a mental health crisis in graduate education', *Nature Biotechnology*, 36(3), pp. 282–284. doi:10.1038/nbt.4089.

Ewels, P., Magnusson, M., Lundin, S. and Käller, M. (2016) 'MultiQC: summarize analysis results for multiple tools and samples in a single report', *Bioinformatics (Oxford, England)*, 32(19), pp. 3047–3048. doi:10.1093/bioinformatics/btw354.

Faccinetto, C., Sabbatini, D., Serventi, P., Rigato, M., Salvoro, C., Casamassima, G., Margiotta, G., De Fanti, S., Sarno, S., Staiti, N., Luiselli, D., Marino, A. and Vazza, G. (2021) 'Internal validation and improvement of mitochondrial genome sequencing using the Precision ID mtDNA Whole Genome Panel', *International Journal of Legal Medicine*, 135(6), pp. 2295–2306. doi:10.1007/s00414-021-02686-w.

Fagnani, C., Neale, M.C., Nisticò, L., Stazi, M.A., Ricigliano, V.A., Buscarinu, M.C., Salvetti, M. and Ristori, G. (2015) 'Twin studies in multiple sclerosis: A meta-estimation of heritability and environmentality', *Multiple Sclerosis Journal*, 21(11), pp. 1404–1413. doi:10.1177/1352458514564492.

Falk, M.J., Shen, L., Gonzalez, M., Leipzig, J., Lott, M.T., Stassen, A.P.M., Diroma, M.A., Navarro-Gomez, D., Yeske, P., Bai, R., Boles, R.G., Brilhante, V., Ralph, D., DaRe, J.T., Shelton, R., Terry, S.F., Zhang, Z., Copeland, W.C., van Oven, M., Prokisch, H., Wallace, D.C., Attimonelli, M., Krotoski, D., Zuchner, S., Gai, X., MSeqDR Consortium Participants, MSeqDR Consortium participants: Sherri Bale, Jirair Bedoyan, Doron Behar, Penelope Bonnen, Lisa Brooks, Claudia Calabrese, Sarah Calvo, Patrick Chinnery, John Christodoulou, Deanna Church, Rosanna Clima, Bruce H. Cohen, Richard G. Cotton, IFM de Coo, Olga Derbenevoa, Johan T. den Dunnen, David Dimmock, Gregory Enns, Giuseppe Gasparre, Amy Goldstein, Iris Gonzalez, Katrina Gwinn, Sihoun Hahn, Richard H. Haas, Hakon Hakonarson, Michio Hirano, Douglas Kerr, Dong Li, Maria Lvova, Finley Macrae, Donna Maglott, Elizabeth McCormick, Grant Mitchell, Vamsi K. Mootha, Yasushi Okazaki, Aurora Pujol, Melissa Parisi, Juan Carlos Perin, Eric A. Pierce, Vincent Procaccio, Shamima Rahman, Honey Reddi, Heidi Rehm, Erin Riggs, Richard Rodenburg, Yaffa Rubinstein, Russell Saneto, Mariangela Santorsola, Curt Scharfe, and Claire Sheldon, Eric A. Shoubridge, Domenico Simone, Bert Smeets, Jan A. Smeitink, Christine Stanley, Anu Suomalainen, Mark Tarnopolsky, Isabelle Thiffault, David R. Thorburn, Johan Van Hove, Lynne Wolfe, and Lee-Jun Wong (2015) 'Mitochondrial Disease Sequence Data Resource (MSeqDR): a global grass-roots consortium to facilitate deposition, curation, annotation, and integrated analysis of genomic data for the mitochondrial disease clinical and research communities', *Molecular Genetics and Metabolism*, 114(3), pp. 388–396. doi:10.1016/j.ymgme.2014.11.016.

Falkenberg, M. (2018) 'Mitochondrial DNA replication in mammalian cells: overview of the pathway', *Essays in Biochemistry*, 62(3), pp. 287–296. doi:10.1042/EBC20170100.

Fambiatos, A., Jokubaitis, V., Horakova, D., Kubala Havrdova, E., Trojano, M., Prat, A., Girard, M., Duquette, P., Lugaresi, A., Izquierdo, G., Grand'Maison, F., Grammond, P., Sola, P., Ferraro, D., Alroughani, R., Terzi, M., Hupperts, R., Boz, C., Lechner-Scott, J., Pucci, E., Bergamaschi, R., Van Pesch, V., Ozakbas, S., Granella, F., Turkoglu, R., Iuliano, G., Spitaleri,

D., McCombe, P., Solaro, C., Slee, M., Ampapa, R., Soysal, A., Petersen, T., Sanchez-Menoyo, J.L., Verheul, F., Prevost, J., Sidhom, Y., Van Wijmeersch, B., Vucic, S., Cristiano, E., Saladino, M.L., Deri, N., Barnett, M., Olascoaga, J., Moore, F., Skibina, O., Gray, O., Fragoso, Y., Yamout, B., Shaw, C., Singhal, B., Shuey, N., Hodgkinson, S., Altintas, A., Al-Harbi, T., Csepany, T., Taylor, B., Hughes, J., Jun, J.-K., van der Walt, A., Spelman, T., Butzkueven, H. and Kalincik, T. (2020) 'Risk of secondary progressive multiple sclerosis: A longitudinal study', *Multiple Sclerosis Journal*, 26(1), pp. 79–90. doi:10.1177/1352458519868990.

Fan, L. and Yao, Y.-G. (2011) 'MitoTool: a web server for the analysis and retrieval of human mitochondrial DNA sequence variations', *Mitochondrion*, 11(2), pp. 351–356. doi:10.1016/j.mito.2010.09.013.

Farina, C., Aloisi, F. and Meinl, E. (2007) 'Astrocytes are active players in cerebral innate immunity', *Trends in Immunology*, 28(3), pp. 138–145. doi:10.1016/j.it.2007.01.005.

Fazzini, F., Fendt, L., Schönherr, S., Forer, L., Schöpf, B., Streiter, G., Losso, J.L., Kloss-Brandstätter, A., Kronenberg, F. and Weissensteiner, H. (2021) 'Analyzing Low-Level mtDNA Heteroplasmy—Pitfalls and Challenges from Bench to Benchmarking', *International Journal of Molecular Sciences*, 22(2), p. 935. doi:10.3390/ijms22020935.

Feger, U., Luther, C., Poeschel, S., Melms, A., Tolosa, E. and Wiendl, H. (2007) 'Increased frequency of CD4+ CD25+ regulatory T cells in the cerebrospinal fluid but not in the blood of multiple sclerosis patients', *Clinical and Experimental Immunology*, 147(3), pp. 412–418. doi:10.1111/j.1365-2249.2006.03271.x.

Fendt, L., Fazzini, F., Weissensteiner, H., Bruckmoser, E., Schönherr, S., Schäfer, G., Losso, J.L., Streiter, G.A., Lamina, C., Rasse, M., Klocker, H., Kofler, B., Kloss-Brandstätter, A., Huck, C.W., Kronenberg, F. and Laimer, J. (2020) 'Profiling of Mitochondrial DNA Heteroplasmy in a Prospective Oral Squamous Cell Carcinoma Study', *Cancers*, 12(7), p. E1933. doi:10.3390/cancers12071933.

Fernández, O., Fernández, V., Martinez-Cabrera, V., Mayorga, C., Alonso, A., León, A., Arnal, C., Hens, M., Luque, G., de Ramón, E., Caballero, A. and Leyva, L. (2008) 'Multiple sclerosis in Gypsies from southern Spain: prevalence, mitochondrial DNA haplogroups and HLA class II association', *Tissue Antigens*, 71(5), pp. 426–433. doi:10.1111/j.1399-0039.2008.01016.x.

Fernández-Sánchez, A., Madrigal-Santillán, E., Bautista, M., Esquivel-Soto, J., Morales-González, Á., Esquivel-Chirino, C., Durante-Montiel, I., Sánchez-Rivera, G., Valadez-Vega, C. and Morales-González, J.A. (2011) 'Inflammation, Oxidative Stress, and Obesity', *International Journal of Molecular Sciences*, 12(5), pp. 3117–3132. doi:10.3390/ijms12053117.

Fernández-Velasco, J.I., Kuhle, J., Monreal, E., Meca-Lallana, V., Meca-Lallana, J., Izquierdo, G., Gascón-Giménez, F., Sainz de la Maza, S., Walo-Delgado, P.E., Maceski, A., Rodríguez-Martín, E., Roldán, E., Villarrubia, N., Saiz, A., Blanco, Y., Sánchez, P., Carreón-Guarnizo, E., Aladro, Y., Brieva, L., Íñiguez, C., González-Suárez, I., Rodríguez de Antonio, L.A., Masjuan, J., Costa-Frossard, L. and Villar, L.M. (2021) 'Effect of Ocrelizumab in Blood Leukocytes of Patients With Primary Progressive MS', *Neurology - Neuroimmunology Neuroinflammation*, 8(2), p. e940. doi:10.1212/NXI.0000000000000940.

Fetisova, E.K., Muntyan, M.S., Lyamzaev, K.G. and Chernyak, B.V. (2019) 'Therapeutic Effect of the Mitochondria-Targeted Antioxidant SkQ1 on the Culture Model of Multiple Sclerosis', *Oxidative Medicine and Cellular Longevity*, 2019, p. 2082561. doi:10.1155/2019/2082561.

Fischer, M.T., Sharma, R., Lim, J.L., Haider, L., Frischer, J.M., Drexhage, J., Mahad, D., Bradl, M., van Horssen, J. and Lassmann, H. (2012) 'NADPH oxidase expression in active multiple sclerosis lesions in relation to oxidative tissue damage and mitochondrial injury', *Brain*, 135(3), pp. 886–899. doi:10.1093/brain/aws012.

Fisniku, L.K., Brex, P.A., Altmann, D.R., Miszkiel, K.A., Benton, C.E., Lanyon, R., Thompson, A.J. and Miller, D.H. (2008) 'Disability and T2 MRI lesions: a 20-year follow-up of patients with relapse onset of multiple sclerosis', *Brain*, 131(3), pp. 808–817. doi:10.1093/brain/awm329.

Flachenecker, P., Kobelt, G., Berg, J., Capsa, D. and Gannedahl, M. (2017) 'New insights into the burden and costs of multiple sclerosis in Europe: Results for Germany', *Multiple Sclerosis Journal*, 23(2_suppl), pp. 78–90. doi:10.1177/1352458517708141.

Fletcher, J.M., Lonergan, R., Costelloe, L., Kinsella, K., Moran, B., O'Farrelly, C., Tubridy, N. and Mills, K.H.G. (2009) 'CD39^{+}Foxp3^{+} Regulatory T Cells Suppress Pathogenic Th17 Cells and Are Impaired in Multiple Sclerosis', *The Journal of Immunology*, 183(11), p. 7602. doi:10.4049/jimmunol.0901881.

Floros, V.I., Pyle, A., Dietmann, S., Wei, W., Tang, W.C.W., Irie, N., Payne, B., Capalbo, A., Noli, L., Coxhead, J., Hudson, G., Crosier, M., Strahl, H., Khalaf, Y., Saitou, M., Ilic, D., Surani, M.A. and Chinnery, P.F. (2018) 'Segregation of mitochondrial DNA heteroplasmy through a developmental genetic bottleneck in human embryos', *Nature Cell Biology*, 20(2), pp. 144–151. doi:10.1038/s41556-017-0017-8.

Fonseca Valter R., Agua-Doce Ana, Maceiras Ana Raquel, Pierson Wim, Ribeiro Filipa, Romão Vasco C., Pires Ana Rita, da Silva Susana Lopes, Fonseca João Eurico, Sousa Ana E., Linterman Michelle A., and Graca Luis (2017) 'Human blood Tfr cells are indicators of ongoing humoral activity not fully licensed with suppressive function', *Science Immunology*, 2(14), p. eaan1487. doi:10.1126/sciimmunol.aan1487.

Foo, S.-S., Reading, P.C., Jaillon, S., Mantovani, A. and Mahalingam, S. (2015) 'Pentraxins and Collectins: Friend or Foe during Pathogen Invasion?', *Trends in Microbiology*, 23(12), pp. 799–811. doi:10.1016/j.tim.2015.09.006.

Fransen, M., Nordgren, M., Wang, B. and Apanasets, O. (2012) 'Role of peroxisomes in ROS/RNS-metabolism: Implications for human disease', *Metabolic Functions and Biogenesis of Peroxisomes in Health and Disease*, 1822(9), pp. 1363–1373. doi:10.1016/j.bbadis.2011.12.001.

Freitag, J., Berod, L., Kamradt, T. and Sparwasser, T. (2016) 'Immunometabolism and autoimmunity', *Immunology & Cell Biology*, 94(10), pp. 925–934. doi:10.1038/icb.2016.77.

Fritzsching, B., Haas, J., König, F., Kunz, P., Fritzsching, E., Pöschl, J., Krammer, P.H., Brück, W., Suri-Payer, E. and Wildemann, B. (2011) 'Intracerebral Human Regulatory T Cells: Analysis of CD4+CD25+FOXP3+ T Cells in Brain Lesions and Cerebrospinal Fluid of Multiple Sclerosis Patients', *PLOS ONE*, 6(3), p. e17988. doi:10.1371/journal.pone.0017988.

Fung, T.C., Olson, C.A. and Hsiao, E.Y. (2017) 'Interactions between the microbiota, immune and nervous systems in health and disease', *Nature Neuroscience*, 20(2), pp. 145–155. doi:10.1038/nn.4476.

Gabilondo, I., Martínez-Lapiscina, E.H., Martínez-Heras, E., Fraga-Pumar, E., Llufriu, S., Ortiz, S., Bullich, S., Sepulveda, M., Falcon, C., Berenguer, J., Saiz, A., Sanchez-Dalmau, B. and Villoslada, P. (2014) 'Trans-synaptic axonal degeneration in the visual pathway in multiple sclerosis', *Annals of Neurology*, 75(1), pp. 98–107. doi:10.1002/ana.24030.

Galgani, M., De Rosa, V., La Cava, A. and Matarese, G. (2016) 'Role of Metabolism in the Immunobiology of Regulatory T Cells', *The Journal of Immunology*, 197(7), pp. 2567–2575. doi:10.4049/jimmunol.1600242.

Galli, E., Hartmann, F.J., Schreiner, B., Ingelfinger, F., Arvaniti, E., Diebold, M., Mrdjen, D., van der Meer, F., Krieg, C., Nimer, F.A., Sanderson, N., Stadelmann, C., Khademi, M., Piehl, F., Claassen, M., Derfuss, T., Olsson, T. and Becher, B. (2019) 'GM-CSF and CXCR4 define a T helper cell signature in multiple sclerosis', *Nature Medicine*, 25(8), pp. 1290–1300. doi:10.1038/s41591-019-0521-4.

Gammage, P.A., Viscomi, C., Simard, M.-L., Costa, A.S.H., Gaude, E., Powell, C.A., Van Haute, L., McCann, B.J., Rebelo-Guiomar, P., Cerutti, R., Zhang, L., Rebar, E.J., Zeviani, M., Frezza, C., Stewart, J.B. and Minczuk, M. (2018) 'Genome editing in mitochondria corrects a pathogenic mtDNA mutation in vivo', *Nature Medicine*, 24(11), pp. 1691–1695. doi:10.1038/s41591-018-0165-9.

Garrison, E. and Marth, G. (2012) 'Haplotype-based variant detection from short-read sequencing', *arXiv:1207.3907 [q-bio]* [Preprint].

gatk4-mitochondria-pipeline (2020). Available at: https://github.com/gatk-workflows/gatk4-mitochondria-pipeline (Accessed: 18 August 2021).

George, M.F., Briggs, F.B.S., Shao, X., Gianfrancesco, M.A., Kockum, I., Harbo, H.F., Celius, E.G., Bos, S.D., Hedström, A., Shen, L., Bernstein, A., Alfredsson, L., Hillert, J., Olsson, T., Patsopoulos, N.A., De Jager, P.L., Oturai, A.B., Søndergaard, H.B., Sellebjerg, F., Sorensen, P.S., Gomez, R., Caillier, S.J., Cree, B.A.C., Oksenberg, J.R., Hauser, S.L., D'Alfonso, S., Leone, M.A., Boneschi, F.M., Sorosina, M., van der Mei, I., Taylor, B.V., Zhou, Y., Schaefer, C. and Barcellos, L.F. (2016) 'Multiple sclerosis risk loci and disease severity in 7,125 individuals from 10 studies', *Neurology Genetics*, 2(4), p. e87. doi:10.1212/NXG.0000000000000087.

Geraldes, R., Ciccarelli, O., Barkhof, F., De Stefano, N., Enzinger, C., Filippi, M., Hofer, M., Paul, F., Preziosa, P., Rovira, A., DeLuca, G.C., Kappos, L., Yousry, T., Fazekas, F., Frederiksen, J., Gasperini, C., Sastre-Garriga, J., Evangelou, N., Palace, J., and on behalf of the MAGNIMS study group (2018) 'The current role of MRI in differentiating multiple sclerosis from its

imaging mimics', *Nature Reviews Neurology*, 14(4), pp. 199–213. doi:10.1038/nrneurol.2018.14.

Gerdes, L.A., Janoschka, C., Eveslage, M., Mannig, B., Wirth, T., Schulte-Mecklenbeck, A., Lauks, S., Glau, L., Gross, C.C., Tolosa, E., Flierl-Hecht, A., Ertl-Wagner, B., Barkhof, F., Meuth, S.G., Kümpfel, T., Wiendl, H., Hohlfeld, R. and Klotz, L. (2020) 'Immune signatures of prodromal multiple sclerosis in monozygotic twins', *Proceedings of the National Academy of Sciences*, 117(35), pp. 21546–21556. doi:10.1073/pnas.2003339117.

Giedraitiene, N., Drukteiniene, E., Kizlaitiene, R., Cimbalas, A., Asoklis, R. and Kaubrys, G. (2021) 'Cognitive Decline in Multiple Sclerosis Is Related to the Progression of Retinal Atrophy and Presence of Oligoclonal Bands: A 5-Year Follow-Up Study', *Frontiers in Neurology*, 12, p. 678735. doi:10.3389/fneur.2021.678735.

Gieß, R.M., Pfuhl, C., Behrens, J.R., Rasche, L., Freitag, E., Khalighy, N., Otto, C., Wuerfel, J., Brandt, A.U., Hofmann, J., Eberspächer, B., Bellmann-Strobl, J., Paul, F. and Ruprecht, K. (2017) 'Epstein-Barr virus antibodies in serum and DNA load in saliva are not associated with radiological or clinical disease activity in patients with early multiple sclerosis', *PLOS ONE*. Edited by G. Khan, 12(4), p. e0175279. doi:10.1371/journal.pone.0175279.

Gilkerson, R., Bravo, L., Garcia, I., Gaytan, N., Herrera, A., Maldonado, A. and Quintanilla, B. (2013) 'The mitochondrial nucleoid: integrating mitochondrial DNA into cellular homeostasis', *Cold Spring Harbor Perspectives in Biology*, 5(5), p. a011080. doi:10.1101/cshperspect.a011080.

Gillmore, J.D., Gane, E., Taubel, J., Kao, J., Fontana, M., Maitland, M.L., Seitzer, J., O'Connell, D., Walsh, K.R., Wood, K., Phillips, J., Xu, Y., Amaral, A., Boyd, A.P., Cehelsky, J.E., McKee, M.D., Schiermeier, A., Harari, O., Murphy, A., Kyratsous, C.A., Zambrowicz, B., Soltys, R., Gutstein, D.E., Leonard, J., Sepp-Lorenzino, L. and Lebwohl, D. (2021) 'CRISPR-Cas9 In Vivo Gene Editing for Transthyretin Amyloidosis', *New England Journal of Medicine*, 385(6), pp. 493–502. doi:10.1056/NEJMoa2107454.

Gingele, S., Jacobus, T.L., Konen, F.F., Hümmert, M.W., Sühs, K.-W., Schwenkenbecher, P., Ahlbrecht, J., Möhn, N., Müschen, L.H., Bönig, L., Alvermann, S., Schmidt, R.E., Stangel, M., Jacobs, R. and Skripuletz, T. (2019) 'Ocrelizumab Depletes CD20+ T Cells in Multiple Sclerosis Patients', *Cells*, 8(1). doi:10.3390/cells8010012.

Ginhoux, F., Lim, S., Hoeffel, G., Low, D. and Huber, T. (2013) 'Origin and differentiation of microglia', *Frontiers in Cellular Neuroscience*, 7, p. 45. doi:10.3389/fncel.2013.00045.

Godinho-Santos, A., Foxall, R.B., Antão, A.V., Tavares, B., Ferreira, T., Serra-Caetano, A., Matoso, P. and Sousa, A.E. (2020) 'Follicular Helper T Cells Are Major Human Immunodeficiency Virus-2 Reservoirs and Support Productive Infection', *The Journal of Infectious Diseases*, 221(1), pp. 122–126. doi:10.1093/infdis/jiz431.

Gonçalves, F.B. and Morais, V.A. (2021) 'PINK1: A Bridge between Mitochondria and Parkinson's Disease', *Life*, 11(5), p. 371. doi:10.3390/life11050371.

Gonzalo, H., Nogueras, L., Gil-Sánchez, A., Hervás, J.V., Valcheva, P., González-Mingot, C., Martin-Gari, M., Canudes, M., Peralta, S., Solana, M.J., Pamplona, R., Portero-Otin, M., Boada, J., Serrano, J.C.E. and Brieva, L. (2019) 'Impairment of Mitochondrial Redox Status in Peripheral Lymphocytes of Multiple Sclerosis Patients', *Frontiers in Neuroscience*, 13, p. 938. doi:10.3389/fnins.2019.00938.

Goris, A., Pauwels, I., Gustavsen, M.W., van Son, B., Hilven, K., Bos, S.D., Celius, E.G., Berg-Hansen, P., Aarseth, J., Myhr, K.-M., D'Alfonso, S., Barizzone, N., Leone, M.A., Martinelli Boneschi, F., Sorosina, M., Liberatore, G., Kockum, I., Olsson, T., Hillert, J., Alfredsson, L., Bedri, S.K., Hemmer, B., Buck, D., Berthele, A., Knier, B., Biberacher, V., van Pesch, V., Sindic, C., Bang Oturai, A., Søndergaard, H.B., Sellebjerg, F., Jensen, P.E.H., Comabella, M., Montalban, X., Pérez-Boza, J., Malhotra, S., Lechner-Scott, J., Broadley, S., Slee, M., Taylor, B., Kermode, A.G., Gourraud, P.-A., International Multiple Sclerosis Genetics Consortium, Sawcer, S.J., Andreassen, B.K., Dubois, B. and Harbo, H.F. (2015) 'Genetic variants are major determinants of CSF antibody levels in multiple sclerosis', *Brain*, 138(3), pp. 632–643. doi:10.1093/brain/awu405.

Gorman, G.S., Schaefer, A.M., Ng, Y., Gomez, N., Blakely, E.L., Alston, C.L., Feeney, C., Horvath, R., Yu-Wai-Man, P., Chinnery, P.F., Taylor, R.W., Turnbull, D.M. and McFarland, R. (2015) 'Prevalence of nuclear and mitochondrial DNA mutations related to adult mitochondrial disease', *Annals of Neurology*, 77(5), pp. 753–759. doi:10.1002/ana.24362.

Gorzynski, J.E., Goenka, S.D., Shafin, K., Jensen, T.D., Fisk, D.G., Grove, M.E., Spiteri, E., Pesout, T., Monlong, J., Baid, G., Bernstein, J.A., Ceresnak, S., Chang, P.-C., Christle, J.W., Chubb, H., Dalton, K.P., Dunn, K., Garalde, D.R., Guillory, J., Knowles, J.W., Kolesnikov, A., Ma, M., Moscarello, T., Nattestad, M., Perez, M., Ruzhnikov, M.R.Z., Samadi, M., Setia, A., Wright, C., Wusthoff, C.J., Xiong, K., Zhu, T., Jain, M., Sedlazeck, F.J., Carroll, A., Paten, B. and Ashley, E.A. (2022) 'Ultrarapid Nanopore Genome Sequencing in a Critical Care Setting', *New England Journal of Medicine*, p. NEJMc2112090. doi:10.1056/NEJMc2112090.

Gourraud, P.-A., Harbo, H.F., Hauser, S.L. and Baranzini, S.E. (2012) 'The genetics of multiple sclerosis: an up-to-date review', *Immunological Reviews*, 248(1), pp. 87–103. doi:10.1111/j.1600-065X.2012.01134.x.

Gray, M.W. (2017) 'Lynn Margulis and the endosymbiont hypothesis: 50 years later', *Molecular Biology of the Cell*, 28(10), pp. 1285–1287. doi:10.1091/mbc.E16-07-0509.

Greaves, L.C., Nooteboom, M., Elson, J.L., Tuppen, H.A.L., Taylor, G.A., Commane, D.M., Arasaradnam, R.P., Khrapko, K., Taylor, R.W., Kirkwood, T.B.L., Mathers, J.C. and Turnbull, D.M. (2014) 'Clonal expansion of early to mid-life mitochondrial DNA point mutations drives mitochondrial dysfunction during human ageing', *PLoS genetics*, 10(9), p. e1004620. doi:10.1371/journal.pgen.1004620.

Gulley, M.L. (2020) 'Clinical Practice: Molecular Pathology', in Reisner, H.M. (ed.) *Pathology: A Modern Case Study*. 2nd edn. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1173768182 (Accessed: 11 January 2022).

Guo, Y., Li, C.-I., Sheng, Q., Winther, J.F., Cai, Q., Boice, J.D. and Shyr, Y. (2013) 'Very low-level heteroplasmy mtDNA variations are inherited in humans', *Journal of Genetics and Genomics = Yi Chuan Xue Bao*, 40(12), pp. 607–615. doi:10.1016/j.jgg.2013.10.003.

Guo, Y., Li, J., Li, C.-I., Shyr, Y. and Samuels, D.C. (2013) 'MitoSeek: extracting mitochondria information and performing high-throughput mitochondria sequencing analysis', *Bioinformatics (Oxford, England)*, 29(9), pp. 1210–1211. doi:10.1093/bioinformatics/btt118.

Guthrie, S., Lichten, C.A., Van Belle, J., Ball, S., Knack, A. and Hofman, J. (2018) 'Understanding mental health in the research environment: A Rapid Evidence Assessment', *Rand Health Quarterly*, 7(3), p. 2.

Haas, J., Hug, A., Viehöver, A., Fritzsching, B., Falk, C.S., Filser, A., Vetter, T., Milkova, L., Korporal, M., Fritz, B., Storch-Hagenlocher, B., Krammer, P.H., Suri-Payer, E. and Wildemann, B. (2005) 'Reduced suppressive effect of CD4+CD25high regulatory T cells on the T cell immune response against myelin oligodendrocyte glycoprotein in patients with multiple sclerosis', *European Journal of Immunology*, 35(11), pp. 3343–3352. doi:10.1002/eji.200526065.

Haast, R.A.M., Ivanov, D., IJsselstein, R.J.T., Sallevelt, S.C.E.H., Jansen, J.F.A., Smeets, H.J.M., de Coo, I.F.M., Formisano, E. and Uludağ, K. (2018) 'Anatomic & metabolic brain markers of the m.3243A>G mutation: A multi-parametric 7T MRI study', *NeuroImage: Clinical*, 18, pp. 231–244. doi:10.1016/j.nicl.2018.01.017.

Haghikia, A., Faissner, S., Pappas, D., Pula, B., Akkad, D.A., Arning, L., Ruhrmann, S., Duscha, A., Gold, R., Baranzini, S.E., Malhotra, S., Montalban, X., Comabella, M. and Chan, A. (2015) 'Interferon-beta affects mitochondrial activity in CD4 $^{+}$ lymphocytes: Implications for mechanism of action in multiple sclerosis', *Multiple Sclerosis Journal*, 21(10), pp. 1262–1270. doi:10.1177/1352458514561909.

Hahn, A. and Zuryn, S. (2019) 'The Cellular Mitochondrial Genome Landscape in Disease', *Trends in Cell Biology*, 29(3), pp. 227–240. doi:10.1016/j.tcb.2018.11.004.

Haider, L., Fischer, M.T., Frischer, J.M., Bauer, J., Hoftberger, R., Botond, G., Esterbauer, H., Binder, C.J., Witztum, J.L. and Lassmann, H. (2011) 'Oxidative damage in multiple sclerosis lesions', *Brain*, 134(7), pp. 1914–1924. doi:10.1093/brain/awr128.

Han, X., Zhou, Z., Fei, L., Sun, H., Wang, R., Chen, Y., Chen, H., Wang, J., Tang, H., Ge, W., Zhou, Y., Ye, F., Jiang, M., Wu, J., Xiao, Y., Jia, X., Zhang, T., Ma, X., Zhang, Q., Bai, X., Lai, S., Yu, C., Zhu, L., Lin, R., Gao, Y., Wang, M., Wu, Y., Zhang, J., Zhan, R., Zhu, S., Hu, H., Wang, C., Chen, M., Huang, H., Liang, T., Chen, J., Wang, W., Zhang, D. and Guo, G. (2020) 'Construction of a human cell landscape at single-cell level', *Nature*, 581(7808), pp. 303–309. doi:10.1038/s41586-020-2157-4.

Handel, A.E., Williamson, A.J., Disanto, G., Handunnetthi, L., Giovannoni, G. and Ramagopalan, S.V. (2010) 'An Updated Meta-Analysis of Risk of Multiple Sclerosis following Infectious Mononucleosis', *PLoS ONE*. Edited by S. Jacobson, 5(9), p. e12496. doi:10.1371/journal.pone.0012496.

Harding, A.E., Sweeney, M.G., Miller, D.H., Mumford, C.J., Kellar-Wood, H., Menard, D., McDonald, W.I. and Compston, D.A. (1992) 'Occurrence of a multiple sclerosis-like illness in women who have a Leber's hereditary optic neuropathy mitochondrial DNA mutation', *Brain: A Journal of Neurology*, 115 (Pt 4), pp. 979–989. doi:10.1093/brain/115.4.979.

Harman, D. (1972) 'The biologic clock: the mitochondria?', *Journal of the American Geriatrics Society*, 20(4), pp. 145–147. doi:10.1111/j.1532-5415.1972.tb00787.x.

Hartung, D.M., Bourdette, D.N., Ahmed, S.M. and Whitham, R.H. (2015) 'The cost of multiple sclerosis drugs in the US and the pharmaceutical industry', *Neurology*, 84(21), p. 2185. doi:10.1212/WNL.0000000000001608.

Hashimoto, Y., Ito, Y., Niikura, T., Shao, Z., Hata, M., Oyama, F. and Nishimoto, I. (2001) 'Mechanisms of Neuroprotection by a Novel Rescue Factor Humanin from Swedish Mutant Amyloid Precursor Protein', *Biochemical and Biophysical Research Communications*, 283(2), pp. 460–468. doi:10.1006/bbrc.2001.4765.

Hassani-Kumleh, H., Houshmand, M., Panahi, M.S.S., Riazi, G.H., Sanati, M.H., Gharagozli, K. and Ghabaee, M. (2006) 'Mitochondrial D-loop variation in Persian multiple sclerosis patients: K and A haplogroups as a risk factor!!', *Cellular and Molecular Neurobiology*, 26(2), pp. 119–125. doi:10.1007/s10571-006-9026-z.

Hassanpour, M., Cheraghi, O., Laghusi, D., Nouri, M. and Panahi, Y. (2020) 'The relationship between ANT1 and NFL with autophagy and mitophagy markers in patients with multiple sclerosis', *Journal of Clinical Neuroscience*, 78, pp. 307–312. doi:10.1016/j.jocn.2020.04.122.

Hawkes, C.H., Giovannoni, G., Lechner-Scott, J. and Levy, M. (2020) 'Is the incidence of multiple sclerosis really increasing?', *Multiple Sclerosis and Related Disorders*, 45, p. 102527. doi:10.1016/j.msard.2020.102527.

Hayakawa, K., Esposito, E., Wang, X., Terasaki, Y., Liu, Y., Xing, C., Ji, X. and Lo, E.H. (2016) 'Transfer of mitochondria from astrocytes to neurons after stroke', *Nature*, 535(7613), pp. 551–555. doi:10.1038/nature18928.

Haynes, B.F., Soderberg, K.A. and Fauci, A.S. (2018) 'Introduction to the Immune System', in Jameson, J.L., Fauci, A.S., Kasper, D.L., Hauser, S.L., Longo, D.L., and Loscalzo, J. (eds) *Harrison's Principles of Internal Medicine, 20e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1156520587 (Accessed: 24 December 2021).

He, Y., Wu, J., Dressman, D.C., Iacobuzio-Donahue, C., Markowitz, S.D., Velculescu, V.E., Diaz Jr, L.A., Kinzler, K.W., Vogelstein, B. and Papadopoulos, N. (2010) 'Heteroplasmic mitochondrial DNA mutations in normal and tumour cells', *Nature*, 464(7288), pp. 610–614. doi:10.1038/nature08802.

Hedström, A.K., Huang, J., Brenner, N., Butt, J., Kockum, I., Waterboer, T., Olsson, T. and Alfredsson, L. (2021) 'Low sun exposure acts synergistically with high Epstein−Barr nuclear

antigen 1 (EBNA-1) antibody levels in multiple sclerosis etiology', *European Journal of Neurology*, 28(12), pp. 4146–4152. doi:10.1111/ene.15082.

Hedström, A.K., Olsson, T. and Alfredsson, L. (2015) 'The Role of Environment and Lifestyle in Determining the Risk of Multiple Sclerosis', in La Flamme, A.C. and Orian, J.M. (eds) *Emerging and Evolving Topics in Multiple Sclerosis Pathogenesis and Treatments*. Cham: Springer International Publishing (Current Topics in Behavioral Neurosciences), pp. 87–104. doi:10.1007/7854_2015_372.

Hester, J., Csárdi, G., Wickham, H., Chang, W., Morgan, M. and Tenenbaum, D. (2021) *remotes: R Package Installation from Remote Repositories, Including 'GitHub'*. Available at: https://CRAN.R-project.org/package=remotes (Accessed: 2 February 2022).

Holloman, J.P., Axtell, R.C., Monson, N.L. and Wu, G.F. (2021) 'The Role of B Cells in Primary Progressive Multiple Sclerosis', *Frontiers in Neurology*, 12, p. 877. doi:10.3389/fneur.2021.680581.

Holt, I.J., Harding, A.E. and Morgan-Hughes, J.A. (1988) 'Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies', *Nature*, 331(6158), pp. 717–719. doi:10.1038/331717a0.

Hong, S., Zhang, Zhimin, Liu, H., Tian, M., Zhu, X., Zhang, Zhuqiang, Wang, W., Zhou, X., Zhang, F., Ge, Q., Zhu, B., Tang, H., Hua, Z. and Hou, B. (2018) 'B Cells Are the Dominant Antigen-Presenting Cells that Activate Naive CD4+ T Cells upon Immunization with a Virus-Derived Nanoparticle Antigen', *Immunity*, 49(4), pp. 695-708.e4. doi:10.1016/j.immuni.2018.08.012.

Hosseinzadeh, A., Sedighi, B., Kermanchi, J., Heidari, M. and Haghdoost, A.A. (2021) 'Is the sharp increasing trend of multiple sclerosis incidence real in Iran?', *BMC Neurology*, 21(1), p. 7. doi:10.1186/s12883-020-02031-5.

Hough, K.P., Trevor, J.L., Strenkowski, J.G., Wang, Y., Chacko, B.K., Tousif, S., Chanda, D., Steele, C., Antony, V.B., Dokland, T., Ouyang, X., Zhang, J., Duncan, S.R., Thannickal, V.J., Darley-Usmar, V.M. and Deshane, J.S. (2018) 'Exosomal transfer of mitochondria from airway myeloid-derived regulatory cells to T cells', *Redox Biology*, 18, pp. 54–64. doi:10.1016/j.redox.2018.06.009.

Houshmand, M., Sanati, M.H., Babrzadeh, F., Ardalan, A., Teimori, M., Vakilian, M., Akuchekian, M., Farhud, D. and Lotfi, J. (2005) 'Population screening for association of mitochondrial haplogroups BM, J, K and M with multiple sclerosis: interrelation between haplogroup J and MS in Persian patients', *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 11(6), pp. 728–730. doi:10.1191/1352458505ms1228sr.

Howell, O.W., Reeves, C.A., Nicholas, R., Carassiti, D., Radotra, B., Gentleman, S.M., Serafini, B., Aloisi, F., Roncaroli, F., Magliozzi, R. and Reynolds, R. (2011) 'Meningeal inflammation is widespread and linked to cortical pathology in multiple sclerosis', *Brain*, 134(9), pp. 2755–2771. doi:10.1093/brain/awr182.

Howie, D., Cobbold, S.P., Adams, E., Ten Bokum, A., Necula, A.S., Zhang, W., Huang, H., Roberts, D.J., Thomas, B., Hester, S.S., Vaux, D.J., Betz, A.G. and Waldmann, H. (2017) 'Foxp3 drives oxidative phosphorylation and protection from lipotoxicity', *JCI Insight*, 2(3). doi:10.1172/jci.insight.89160.

Hsiao, C.-C., Fransen, N.L., van den Bosch, A.M.R., Brandwijk, K.I.M., Huitinga, I., Hamann, J. and Smolders, J. (2021) 'White matter lesions in multiple sclerosis are enriched for CD20dim CD8+ tissue-resident memory T cells', *European Journal of Immunology*, 51(2), pp. 483–486. doi:10.1002/eji.202048665.

Huan, J., Culbertson, N., Spencer, L., Bartholomew, R., Burrows, G.G., Chou, Y.K., Bourdette, D., Ziegler, S.F., Offner, H. and Vandenbark, A.A. (2005) 'Decreased FOXP3 levels in multiple sclerosis patients', *Journal of Neuroscience Research*, 81(1), pp. 45–52. doi:10.1002/jnr.20522.

Hudson, G., Gomez-Duran, A., Wilson, I.J. and Chinnery, P.F. (2014) 'Recent mitochondrial DNA mutations increase the risk of developing common late-onset human diseases', *PLoS genetics*, 10(5), p. e1004369. doi:10.1371/journal.pgen.1004369.

Human Medicines Evaluation Division, European Medicines Agency (2021) *List of nationally authorised medicinal products. Active substance: Glatiramer. Procedure no.: PSUSA/00001529/202011. EMA/466570/2021*, *European Medicines Agency*. Available at: https://www.ema.europa.eu/en/documents/psusa/glatiramer-list-nationally-authorised-medicinal-products-psusa-00001529-202011_en.pdf (Accessed: 6 January 2022).

Iljicsov, A., Milanovich, D., Ajtay, A., Oberfrank, F., Bálint, M., Dobi, B., Bereczki, D. and Simó, M. (2020) 'Incidence and prevalence of multiple sclerosis in Hungary based on record linkage of nationwide multiple healthcare administrative data', *PLOS ONE*, 15(7), p. e0236432. doi:10.1371/journal.pone.0236432.

INFARMED (2020) *RESUMO DAS CARACTERÍSTICAS DO MEDICAMENTO - Copaxone 20 mg/ml Solução Injetável, seringa pré-cheia*. Available at: https://extranet.infarmed.pt/INFOMED-fo/pesquisa-avancada.xhtml (Accessed: 6 January 2022).

International HapMap Consortium (2003) 'The International HapMap Project', *Nature*, 426(6968), pp. 789–796. doi:10.1038/nature02168.

International Human Genome Sequencing Consortium (2004) 'Finishing the euchromatic sequence of the human genome', *Nature*, 431(7011), pp. 931–945. doi:10.1038/nature03001.

International Multiple Sclerosis Genetics Consortium (2013a) 'Analysis of immune-related loci identifies 48 new susceptibility variants for multiple sclerosis', *Nature Genetics*, 45(11), pp. 1353–1360. doi:10.1038/ng.2770.

International Multiple Sclerosis Genetics Consortium (2013b) 'Network-based multiple sclerosis pathway analysis with GWAS data from 15,000 cases and 30,000 controls', *American Journal of Human Genetics*, 92(6), pp. 854–865. doi:10.1016/j.ajhg.2013.04.019.

International Multiple Sclerosis Genetics Consortium (2015) 'Class II HLA interactions modulate genetic risk for multiple sclerosis', *Nature Genetics*, 47(10), pp. 1107–1113. doi:10.1038/ng.3395.

International Multiple Sclerosis Genetics Consortium (2018) 'Low-Frequency and Rare-Coding Variation Contributes to Multiple Sclerosis Risk', *Cell*, 175(6), pp. 1679-1687.e7. doi:10.1016/j.cell.2018.09.049.

International Multiple Sclerosis Genetics Consortium (2019) 'A systems biology approach uncovers cell-specific gene regulatory effects of genetic associations in multiple sclerosis', *Nature Communications*, 10(1), p. 2236. doi:10.1038/s41467-019-09773-y.

International Multiple Sclerosis Genetics Consortium & The Wellcome Trust Case Control Consortium 2 (2011) 'Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis', *Nature*, 476(7359), pp. 214–219. doi:10.1038/nature10251.

International Multiple Sclerosis Genetics Consortium, Hafler, D.A., Compston, A., Sawcer, S., Lander, E.S., Daly, M.J., De Jager, P.L., de Bakker, P.I.W., Gabriel, S.B., Mirel, D.B., Ivinson, A.J., Pericak-Vance, M.A., Gregory, S.G., Rioux, J.D., McCauley, J.L., Haines, J.L., Barcellos, L.F., Cree, B., Oksenberg, J.R. and Hauser, S.L. (2007) 'Risk alleles for multiple sclerosis identified by a genomewide study', *The New England Journal of Medicine*, 357(9), pp. 851–862. doi:10.1056/NEJMoa073493.

International Multiple Sclerosis Genetics Consortium, Patsopoulos, N.A., Baranzini, S.E., Santaniello, A., Shoostari, P., Cotsapas, C., Wong, G., Beecham, A.H., James, T., Replogle, J., Vlachos, I.S., McCabe, C., Pers, T.H., Brandes, A., White, C., Keenan, B., Cimpean, M., Winn, P., Panteliadis, I.-P., Robbins, A., Andlauer, T.F.M., Zarzycki, O., Dubois, B., Goris, A., Søndergaard, H.B., Sellebjerg, F., Sorensen, P.S., Ullum, H., Thørner, L.W., Saarela, J., Cournu-Rebeix, I., Damotte, V., Fontaine, B., Guillot-Noel, L., Lathrop, M., Vukusic, S., Berthele, A., Pongratz, V., Buck, D., Gasperi, C., Graetz, C., Grummel, V., Hemmer, B., Hoshi, M., Knier, B., Korn, T., Lill, C.M., Luessi, F., Mühlau, M., Zipp, F., Dardiotis, E., Agliardi, C., Amoroso, A., Barizzone, N., Benedetti, M.D., Bernardinelli, L., Cavalla, P., Clarelli, F., Comi, G., Cusi, D., Esposito, F., Ferrè, L., Galimberti, D., Guaschino, C., Leone, M.A., Martinelli, V., Moiola, L., Salvetti, M., Sorosina, M., Vecchio, D., Zauli, A., Santoro, S., Mancini, N., Zuccalà, M., Mescheriakova, J., van Duijn, C., Bos, S.D., Celius, E.G., Spurkland, A., Comabella, M., Montalban, X., Alfredsson, L., Bomfim, I.L., Gomez-Cabrero, D., Hillert, J., Jagodic, M., Lindén, M., Piehl, F., Jelčić, I., Martin, R., Sospedra, M., Baker, A., Ban, M., Hawkins, C., Hysi, P., Kalra, S., Karpe, F., Khadake, J., Lachance, G., Molyneux, P., Neville, M., Thorpe, J., Bradshaw, E., Caillier, S.J., Calabresi, P., Cree, B.A.C., Cross, A., Davis, M., de Bakker, P.W.I., Delgado, S., Dembele, M., Edwards, K., Fitzgerald, K., Frohlich, I.Y., Gourraud, P.-A., Haines, J.L., Hakonarson, H., Kimbrough, D., Isobe, N., Konidari, I., Lathi, E., Lee, M.H., Li, T., An, D., Zimmer, A., Madireddy, L., Manrique, C.P., Mitrovic, M., Olah, M., Patrick, E., Pericak-Vance, M.A., Piccio, L., Schaefer, C., Weiner, H., Lage, K., ANZgene, IIBDGC, WTCCC2, Compston, A., Hafler, D., Harbo, H.F., Hauser, S.L., Stewart, G., D'Alfonso, S., Hadjigeorgiou, G., Taylor, B., Barcellos, L.F., Booth, D., Hintzen, R., Kockum, I., Martinelli-Boneschi, F., McCauley, J.L., Oksenberg, J.R., Oturai, A., Sawcer, S., Ivinson, A.J., Olsson, T. and De Jager,

P.L. (2019) 'Multiple sclerosis genomic map implicates peripheral immune cells and microglia in susceptibility', *Science*, 365(6460), p. eaav7188. doi:10.1126/science.aav7188.

Irizar, H., Muñoz-Culla, M., Sepúlveda, L., Sáenz-Cuesta, M., Prada, Á., Castillo-Triviño, T., Zamora-López, G., de Munain, A.L., Olascoaga, J. and Otaegui, D. (2014) 'Transcriptomic Profile Reveals Gender-Specific Molecular Mechanisms Driving Multiple Sclerosis Progression', *PLoS ONE*. Edited by F.J. Esteban, 9(2), p. e90482. doi:10.1371/journal.pone.0090482.

Ishiya, K. and Ueda, S. (2017) 'MitoSuite: a graphical tool for human mitochondrial genome profiling in massive parallel sequencing', *PeerJ*, 5, p. e3406. doi:10.7717/peerj.3406.

Isobe, N., Keshavan, A., Gourraud, P.-A., Zhu, A.H., Datta, E., Schlaeger, R., Caillier, S.J., Santaniello, A., Lizée, A., Himmelstein, D.S., Baranzini, S.E., Hollenbach, J., Cree, B.A.C., Hauser, S.L., Oksenberg, J.R. and Henry, R.G. (2016) 'Association of HLA Genetic Risk Burden With Disease Phenotypes in Multiple Sclerosis', *JAMA Neurology*, 73(7), p. 795. doi:10.1001/jamaneurol.2016.0980.

Jäckle, K., Zeis, T., Schaeren-Wiemers, N., Junker, A., van der Meer, F., Kramann, N., Stadelmann, C. and Brück, W. (2020) 'Molecular signature of slowly expanding lesions in progressive multiple sclerosis', *Brain*, 143(7), pp. 2073–2088. doi:10.1093/brain/awaa158.

Jacobs, B.M., Giovannoni, G., Cuzick, J. and Dobson, R. (2020) 'Systematic review and meta-analysis of the association between Epstein–Barr virus, multiple sclerosis and other risk factors', *Multiple Sclerosis Journal*, 26(11), pp. 1281–1297. doi:10.1177/1352458520907901.

Jäkel, S., Agirre, E., Mendanha Falcão, A., van Bruggen, D., Lee, K.W., Knuesel, I., Malhotra, D., ffrench-Constant, C., Williams, A. and Castelo-Branco, G. (2019) 'Altered human oligodendrocyte heterogeneity in multiple sclerosis', *Nature*, 566(7745), pp. 543–547. doi:10.1038/s41586-019-0903-2.

Jakimovski, D., Ramanathan, M., Weinstock-Guttman, B., Bergsland, N., Ramasamay, D.P., Carl, E., Dwyer, M.G. and Zivadinov, R. (2020) 'Higher EBV response is associated with more severe gray matter and lesion pathology in relapsing multiple sclerosis patients: A case-controlled magnetization transfer ratio study', *Multiple Sclerosis Journal*, 26(3), pp. 322–332. doi:10.1177/1352458519828667.

James, S.L., Abate, D., Abate, K.H., Abay, S.M., Abbafati, C., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdela, J., Abdelalim, A., Abdollahpour, I., Abdulkader, R.S., Abebe, Z., Abera, S.F., Abil, O.Z., Abraha, H.N., Abu-Raddad, L.J., Abu-Rmeileh, N.M.E., Accrombessi, M.M.K., Acharya, D., Acharya, P., Ackerman, I.N., Adamu, A.A., Adebayo, O.M., Adekanmbi, V., Adetokunboh, O.O., Adib, M.G., Adsuar, J.C., Afanvi, K.A., Afarideh, M., Afshin, A., Agarwal, G., Agesa, K.M., Aggarwal, R., Aghayan, S.A., Agrawal, S., Ahmadi, A., Ahmadi, M., Ahmadieh, H., Ahmed, M.B., Aichour, A.N., Aichour, I., Aichour, M.T.E., Akinyemiju, T., Akseer, N., Al-Aly, Z., Al-Eyadhy, A., Al-Mekhlafi, H.M., Al-Raddadi, R.M., Alahdab, F., Alam, K., Alam, T., Alashi, A., Alavian, S.M., Alene, K.A., Alijanzadeh, M., Alizadeh-Navaei, R., Aljunid, S.M., Alkerwi, A., Alla, F., Allebeck, P., Alouani, M.M.L., Altirkawi, K., Alvis-Guzman, N., Amare, A.T., Aminde, L.N., Ammar, W., Amoako, Y.A., Anber, N.H., Andrei, C.L.,

Androudi, S., Animut, M.D., Anjomshoa, M., Ansha, M.G., Antonio, C.A.T., Anwari, P., Arabloo, J., Arauz, A., Aremu, O., Ariani, F., Armoon, B., Ärnlöv, J., Arora, A., Artaman, A., Aryal, K.K., Asayesh, H., Asghar, R.J., Ataro, Z., Atre, S.R., Ausloos, M., Avila-Burgos, L., Avokpaho, E.F.G.A., Awasthi, A., Quintanilla, B.P.A., Ayer, R., Azzopardi, P.S., Babazadeh, A., Badali, H., Badawi, A., Bali, A.G., Ballesteros, K.E., Ballew, S.H., Banach, M., Banoub, J.A.M., Banstola, A., Barac, A., Barboza, M.A., Barker-Collo, S.L., Bärnighausen, T.W., Barrero, L.H., Baune, B.T., Bazargan-Hejazi, S., Bedi, N., Beghi, E., Behzadifar, Masoud, Behzadifar, Meysam, Béjot, Y., Belachew, A.B., Belay, Y.A., Bell, M.L., Bello, A.K., Bensenor, I.M., Bernabe, E., Bernstein, R.S., Beuran, M., Beyranvand, T., Bhala, N., Bhattarai, S., Bhaumik, S., Bhutta, Z.A., Biadgo, B., Bijani, A., Bikbov, B., Bilano, V., Bililign, N., Sayeed, M.S.B., Bisanzio, D., Blacker, B.F., Blyth, F.M., Bou-Orm, I.R., Boufous, S., Bourne, R., Brady, O.J., Brainin, M., Brant, L.C., Brazinova, A., Breitborde, N.J.K., Brenner, H., Briant, P.S., Briggs, A.M., Briko, A.N., Britton, G., Brugha, T., Buchbinder, R., Busse, R., Butt, Z.A., Cahuana-Hurtado, L., Cano, J., Cárdenas, R., Carrero, J.J., Carter, A., Carvalho, F., Castañeda-Orjuela, C.A., Rivas, J.C., Castro, F., Catalá-López, F., Cercy, K.M., Cerin, E., Chaiah, Y., Chang, A.R., Chang, H.-Y., Chang, J.-C., Charlson, F.J., Chattopadhyay, A., Chattu, V.K., Chaturvedi, P., Chiang, P.P.-C., Chin, K.L., Chitheer, A., Choi, J.-Y.J., Chowdhury, R., Christensen, H., Christopher, D.J., Cicuttini, F.M., Ciobanu, L.G., Cirillo, M., Claro, R.M., Collado-Mateo, D., Cooper, C., Coresh, J., Cortesi, P.A., Cortinovis, M., Costa, M., Cousin, E., Criqui, M.H., Cromwell, E.A., Cross, M., Crump, J.A., Dadi, A.F., Dandona, L., Dandona, R., Dargan, P.I., Daryani, A., Gupta, R.D., Neves, J.D., Dasa, T.T., Davey, G., Davis, A.C., Davitoiu, D.V., Courten, B.D., Hoz, F.P.D.L., Leo, D.D., Neve, J.-W.D., Degefa, M.G., Degenhardt, L., Deiparine, S., Dellavalle, R.P., Demoz, G.T., Deribe, K., Dervenis, N., Jarlais, D.C.D., Dessie, G.A., Dey, S., Dharmaratne, S.D., Dinberu, M.T., Dirac, M.A., Djalalinia, S., Doan, L., Dokova, K., Doku, D.T., Dorsey, E.R., Doyle, K.E., Driscoll, T.R., Dubey, M., Dubljanin, E., Duken, E.E., Duncan, B.B., Duraes, A.R., Ebrahimi, H., Ebrahimpour, S., Echko, M.M., Edvardsson, D., Effiong, A., Ehrlich, J.R., Bcheraoui, C.E., Zaki, M.E.S., El-Khatib, Z., Elkout, H., Elyazar, I.R.F., Enayati, A., Endries, A.Y., Er, B., Erskine, H.E., Eshrati, B., Eskandarieh, S., Esteghamati, A., Esteghamati, S., Fakhim, H., Omrani, V.F., Faramarzi, M., Fareed, M., Farhadi, F., Farid, T.A., Farinha, C.S.E. sá, Farioli, A., Faro, A., Farvid, M.S., Farzadfar, F., Feigin, V.L., Fentahun, N., Fereshtehnejad, S.-M., Fernandes, E., Fernandes, J.C., Ferrari, A.J., Feyissa, G.T., Filip, I., Fischer, F., Fitzmaurice, C., Foigt, N.A., Foreman, K.J., Fox, J., Frank, T.D., Fukumoto, T., Fullman, N., Fürst, T., Furtado, J.M., Futran, N.D., Gall, S., Ganji, M., Gankpe, F.G., Garcia-Basteiro, A.L., Gardner, W.M., Gebre, A.K., Gebremedhin, A.T., Gebremichael, T.G., Gelano, T.F., Geleijnse, J.M., Genova-Maleras, R., Geramo, Y.C.D., Gething, P.W., Gezae, K.E., Ghadiri, K., Falavarjani, K.G., Ghasemi-Kasman, M., Ghimire, M., Ghosh, R., Ghoshal, A.G., Giampaoli, S., Gill, P.S., Gill, T.K., Ginawi, I.A., Giussani, G., Gnedovskaya, E.V., Goldberg, E.M., Goli, S., Gómez-Dantés, H., Gona, P.N., Gopalani, S.V., Gorman, T.M., Goulart, A.C., Goulart, B.N.G., Grada, A., Grams, M.E., Grosso, G., Gugnani, H.C., Guo, Y., Gupta, P.C., Gupta, Rahul, Gupta, Rajeev, Gupta, T., Gyawali, B., Haagsma, J.A., Hachinski, V., Hafezi-Nejad, N., Bidgoli, H.H., Hagos, T.B., Hailu, G.B., Haj-Mirzaian, Arvin, Haj-Mirzaian, Arya, Hamadeh, R.R., Hamidi, S., Handal, A.J., Hankey, G.J., Hao, Y., Harb, H.L., Harikrishnan, S., Haro, J.M., Hasan, M., Hassankhani, H., Hassen, H.Y., Havmoeller, R., Hawley, C.N., Hay, R.J., Hay, S.I., Hedayatizadeh-Omran, A., Heibati, B., Hendrie, D., Henok, A., Herteliu, C., Heydarpour, S., Hibstu, D.T., Hoang, H.T., Hoek, H.W., Hoffman, H.J., Hole, M.K., Rad, E.H., Hoogar, P., Hosgood, H.D., Hosseini, S.M., Hosseinzadeh, M., Hostiuc, M., Hostiuc, S., Hotez, P.J., Hoy, D.G., Hsairi, M., Htet, A.S., Hu, G., Huang, J.J., Huynh, C.K., Iburg, K.M.,

Ikeda, C.T., Ileanu, B., Ilesanmi, O.S., Iqbal, U., Irvani, S.S.N., Irvine, C.M.S., Islam, S.M.S., Islami, F., Jacobsen, K.H., Jahangiry, L., Jahanmehr, N., Jain, S.K., Jakovljevic, M., Javanbakht, M., Jayatilleke, A.U., Jeemon, P., Jha, R.P., Jha, V., Ji, J.S., Johnson, C.O., Jonas, J.B., Jozwiak, J.J., Jungari, S.B., Jürisson, M., Kabir, Z., Kadel, R., Kahsay, A., Kalani, R., Kanchan, T., Karami, M., Matin, B.K., Karch, A., Karema, C., Karimi, N., Karimi, S.M., Kasaeian, A., Kassa, D.H., Kassa, G.M., Kassa, T.D., Kassebaum, N.J., Katikireddi, S.V., Kawakami, N., Karyani, A.K., Keighobadi, M.M., Keiyoro, P.N., Kemmer, L., Kemp, G.R., Kengne, A.P., Keren, A., Khader, Y.S., Khafaei, B., Khafaie, M.A., Khajavi, A., Khalil, I.A., Khan, E.A., Khan, M.S., Khan, M.A., Khang, Y.-H., Khazaei, M., Khoja, A.T., Khosravi, A., Khosravi, M.H., Kiadaliri, A.A., Kiirithio, D.N., Kim, C.-I., Kim, D., Kim, P., Kim, Y.-E., Kim, Y.J., Kimokoti, R.W., Kinfu, Y., Kisa, A., Kissimova-Skarbek, K., Kivimäki, M., Knudsen, A.K.S., Kocarnik, J.M., Kochhar, S., Kokubo, Y., Kolola, T., Kopec, J.A., Kosen, S., Kotsakis, G.A., Koul, P.A., Koyanagi, A., Kravchenko, M.A., Krishan, K., Krohn, K.J., Defo, B.K., Bicer, B.K., Kumar, G.A., Kumar, M., Kyu, H.H., Lad, D.P., Lad, S.D., Lafranconi, A., Lalloo, R., Lallukka, T., Lami, F.H., Lansingh, V.C., Latifi, A., Lau, K.M.-M., Lazarus, J.V., Leasher, J.L., Ledesma, J.R., Lee, P.H., Leigh, J., Leung, J., Levi, M., Lewycka, S., Li, S., Li, Y., Liao, Y., Liben, M.L., Lim, L.-L., Lim, S.S., Liu, S., Lodha, R., Looker, K.J., Lopez, A.D., Lorkowski, S., Lotufo, P.A., Low, N., Lozano, R., Lucas, T.C.D., Lucchesi, L.R., Lunevicius, R., Lyons, R.A., Ma, S., Macarayan, E.R.K., Mackay, M.T., Madotto, F., Razek, H.M.A.E., Razek, M.M.A.E., Maghavani, D.P., Mahotra, N.B., Mai, H.T., Majdan, M., Majdzadeh, R., Majeed, A., Malekzadeh, R., Malta, D.C., Mamun, A.A., Manda, A.-L., Manguerra, H., Manhertz, T., Mansournia, M.A., Mantovani, L.G., Mapoma, C.C., Maravilla, J.C., Marcenes, W., Marks, A., Martins-Melo, F.R., Martopullo, I., März, W., Marzan, M.B., Mashamba-Thompson, T.P., Massenburg, B.B., Mathur, M.R., Matsushita, K., Maulik, P.K., Mazidi, M., McAlinden, C., McGrath, J.J., McKee, M., Mehndiratta, M.M., Mehrotra, R., Mehta, K.M., Mehta, V., Mejia-Rodriguez, F., Mekonen, T., Melese, A., Melku, M., Meltzer, M., Memiah, P.T.N., Memish, Z.A., Mendoza, W., Mengistu, D.T., Mengistu, G., Mensah, G.A., Mereta, S.T., Meretoja, A., Meretoja, T.J., Mestrovic, T., Mezerji, N.M.G., Miazgowski, B., Miazgowski, T., Millear, A.I., Miller, T.R., Miltz, B., Mini, G.K., Mirarefin, M., Mirrakhimov, E.M., Misganaw, A.T., Mitchell, P.B., Mitiku, H., Moazen, B., Mohajer, B., Mohammad, K.A., Mohammadifard, N., Mohammadnia-Afrouzi, M., Mohammed, M.A., Mohammed, S., Mohebi, F., Moitra, M., Mokdad, A.H., Molokhia, M., Monasta, L., Moodley, Y., Moosazadeh, M., Moradi, G., Moradi-Lakeh, M., Moradinazar, M., Moraga, P., Morawska, L., Velásquez, I.M., Morgado-Da-Costa, J., Morrison, S.D., Moschos, M.M., Mountjoy-Venning, W.C., Mousavi, S.M., Mruts, K.B., Muche, A.A., Muchie, K.F., Mueller, U.O., Muhammed, O.S., Mukhopadhyay, S., Muller, K., Mumford, J.E., Murhekar, M., Musa, J., Musa, K.I., Mustafa, G., Nabhan, A.F., Nagata, C., Naghavi, M., Naheed, A., Nahvijou, A., Naik, G., Naik, N., Najafi, F., Naldi, L., Nam, H.S., Nangia, V., Nansseu, J.R., Nascimento, B.R., Natarajan, G., Neamati, N., Negoi, I., Negoi, R.I., Neupane, S., Newton, C.R.J., Ngunjiri, J.W., Nguyen, A.Q., Nguyen, Ha Thu, Nguyen, H.L.T., Nguyen, Huong Thanh, Nguyen, L.H., Nguyen, M., Nguyen, N.B., Nguyen, S.H., Nichols, E., Ningrum, D.N.A., Nixon, M.R., Nolutshungu, N., Nomura, S., Norheim, O.F., Noroozi, M., Norrving, B., Noubiap, J.J., Nouri, H.R., Shiadeh, M.N., Nowroozi, M.R., Nsoesie, E.O., Nyasulu, P.S., Odell, C.M., Ofori-Asenso, R., Ogbo, F.A., Oh, I.-H., Oladimeji, O., Olagunju, A.T., Olagunju, T.O., Olivares, P.R., Olsen, H.E., Olusanya, B.O., Ong, K.L., Ong, S.K., Oren, E., Ortiz, A., Ota, E., Otstavnov, S.S., Øverland, S., Owolabi, M.O., A, M.P., Pacella, R., Pakpour, A.H., Pana, A., Panda-Jonas, S., Parisi, A., Park, E.-K., Parry, C.D.H., Patel, S., Pati, S., Patil, S.T., Patle, A., Patton, G.C., Paturi, V.R., Paulson, K.R., Pearce, N., Pereira, D.M., Perico, N., Pesudovs, K., Pham, H.Q., Phillips,

M.R., Pigott, D.M., Pillay, J.D., Piradov, M.A., Pirsaheb, M., Pishgar, F., Plana-Ripoll, O., Plass, D., Polinder, S., Popova, S., Postma, M.J., Pourshams, A., Poustchi, H., Prabhakaran, D., Prakash, S., Prakash, V., Purcell, C.A., Purwar, M.B., Qorbani, M., Quistberg, D.A., Radfar, A., Rafay, A., Rafiei, A., Rahim, F., Rahimi, K., Rahimi-Movaghar, A., Rahimi-Movaghar, V., Rahman, M., Rahman, M.H. ur, Rahman, M.A., Rahman, S.U., Rai, R.K., Rajati, F., Ram, U., Ranjan, P., Ranta, A., Rao, P.C., Rawaf, D.L., Rawaf, S., Reddy, K.S., Reiner, R.C., Reinig, N., Reitsma, M.B., Remuzzi, G., Renzaho, A.M.N., Resnikoff, S., Rezaei, S., Rezai, M.S., Ribeiro, A.L.P., Roberts, N.L.S., Robinson, S.R., Roever, L., Ronfani, L., Roshandel, G., Rostami, A., Roth, G.A., Roy, A., Rubagotti, E., Sachdev, P.S., Sadat, N., Saddik, B., Sadeghi, E., Moghaddam, S.S., Safari, H., Safari, Y., Safari-Faramani, R., Safdarian, M., Safi, S., Safiri, S., Sagar, R., Sahebkar, A., Sahraian, M.A., Sajadi, H.S., Salam, N., Salama, J.S., Salamati, P., Saleem, K., Saleem, Z., Salimi, Y., Salomon, J.A., Salvi, S.S., Salz, I., Samy, A.M., Sanabria, J., Sang, Y., Santomauro, D.F., Santos, I.S., Santos, J.V., Milicevic, M.M.S., Jose, B.P.S., Sardana, M., Sarker, A.R., Sarrafzadegan, N., Sartorius, B., Sarvi, S., Sathian, B., Satpathy, M., Sawant, A.R., Sawhney, M., Saxena, S., Saylan, M., Schaeffner, E., Schmidt, M.I., Schneider, I.J.C., Schöttker, B., Schwebel, D.C., Schwendicke, F., Scott, J.G., Sekerija, M., Sepanlou, S.G., Serván-Mori, E., Seyedmousavi, S., Shabaninejad, H., Shafieesabet, A., Shahbazi, M., Shaheen, A.A., Shaikh, M.A., Shams-Beyranvand, M., Shamsi, M., Shamsizadeh, M., Sharafi, H., Sharafi, K., Sharif, M., Sharif-Alhoseini, M., Sharma, M., Sharma, R., She, J., Sheikh, A., Shi, P., Shibuya, K., Shigematsu, M., Shiri, R., Shirkoohi, R., Shishani, K., Shiue, I., Shokraneh, F., Shoman, H., Shrime, M.G., Si, S., Siabani, S., Siddiqi, T.J., Sigfusdottir, I.D., Sigurvinsdottir, R., Silva, J.P., Silveira, D.G.A., Singam, N.S.V., Singh, J.A., Singh, N.P., Singh, V., Sinha, D.N., Skiadaresi, E., Slepak, E.L.N., Sliwa, K., Smith, D.L., Smith, M., Filho, A.M.S., Sobaih, B.H., Sobhani, S., Sobngwi, E., Soneji, S.S., Soofi, M., Soosaraei, M., Sorensen, R.J.D., Soriano, J.B., Soyiri, I.N., Sposato, L.A., Sreeramareddy, C.T., Srinivasan, V., Stanaway, J.D., Stein, D.J., Steiner, C., Steiner, T.J., Stokes, M.A., Stovner, L.J., Subart, M.L., Sudaryanto, A., Sufiyan, M.B., Sunguya, B.F., Sur, P.J., Sutradhar, I., Sykes, B.L., Sylte, D.O., Tabarés-Seisdedos, R., Tadakamadla, S.K., Tadesse, B.T., Tandon, N., Tassew, S.G., Tavakkoli, M., Taveira, N., Taylor, H.R., Tehrani-Banihashemi, A., Tekalign, T.G., Tekelemedhin, S.W., Tekle, M.G., Temesgen, H., Temsah, M.-H., Temsah, O., Terkawi, A.S., Teweldemedhin, M., Thankappan, K.R., Thomas, N., Tilahun, B., To, Q.G., Tonelli, M., Topor-Madry, R., Topouzis, F., Torre, A.E., Tortajada-Girbés, M., Touvier, M., Tovani-Palone, M.R., Towbin, J.A., Tran, B.X., Tran, K.B., Troeger, C.E., Truelsen, T.C., Tsilimbaris, M.K., Tsoi, D., Car, L.T., Tuzcu, E.M., Ukwaja, K.N., Ullah, I., Undurraga, E.A., Unutzer, J., Updike, R.L., Usman, M.S., Uthman, O.A., Vaduganathan, M., Vaezi, A., Valdez, P.R., Varughese, S., Vasankari, T.J., Venketasubramanian, N., Villafaina, S., Violante, F.S., Vladimirov, S.K., Vlassov, V., Vollset, S.E., Vosoughi, K., Vujcic, I.S., Wagnew, F.S., Waheed, Y., Waller, S.G., Wang, Y., Wang, Y.-P., Weiderpass, E., Weintraub, R.G., Weiss, D.J., Weldegebreal, F., Weldegwergs, K.G., Werdecker, A., West, T.E., Whiteford, H.A., Widecka, J., Wijeratne, T., Wilner, L.B., Wilson, S., Winkler, A.S., Wiyeh, A.B., Wiysonge, C.S., Wolfe, C.D.A., Woolf, A.D., Wu, S., Wu, Y.-C., Wyper, G.M.A., Xavier, D., Xu, G., Yadgir, S., Yadollahpour, A., Jabbari, S.H.Y., Yamada, T., Yan, L.L., Yano, Y., Yaseri, M., Yasin, Y.J., Yeshaneh, A., Yimer, E.M., Yip, P., Yisma, E., Yonemoto, N., Yoon, S.-J., Yotebieng, M., Younis, M.Z., Yousefifard, M., Yu, C., Zadnik, V., Zaidi, Z., Zaman, S.B., Zamani, M., Zare, Z., Zeleke, A.J., Zenebe, Z.M., Zhang, K., Zhao, Z., Zhou, M., Zodpey, S., Zucker, I., Vos, T. and Murray, C.J.L. (2018) 'Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden

of Disease Study 2017', *The Lancet*, 392(10159), pp. 1789–1858. doi:10.1016/S0140-6736(18)32279-7.

Jameson, J.L. and Kopp, P. (2018) 'Principles of Human Genetics', in Jameson, J.L., Fauci, A.S., Kasper, D.L., Hauser, S.L., Longo, D.L., and Loscalzo, J. (eds) *Harrison's Principles of Internal Medicine*. 20th edn. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1160636532 (Accessed: 11 January 2022).

Jansen, R.J., Fonseca-Williams, S., Bamlet, W.R., Ayala-Peña, S., Oberg, A.L., Petersen, G.M. and Torres-Ramos, C.A. (2015) 'Detection of DNA damage in peripheral blood mononuclear cells from pancreatic cancer patients', *Molecular Carcinogenesis*, 54(10), pp. 1220–1226. doi:10.1002/mc.22189.

Jarius, S., Ruprecht, K., Stellmann, J.P., Huss, A., Ayzenberg, I., Willing, A., Trebst, C., Pawlitzki, M., Abdelhak, A., Grüter, T., Leypoldt, F., Haas, J., Kleiter, I., Tumani, H., Fechner, K., Reindl, M., Paul, F. and Wildemann, B. (2018) 'MOG-IgG in primary and secondary chronic progressive multiple sclerosis: a multicenter study of 200 patients and review of the literature', *Journal of Neuroinflammation*, 15(1), p. 88. doi:10.1186/s12974-018-1108-6.

Jastroch, M., Divakaruni, A.S., Mookerjee, S., Treberg, J.R. and Brand, M.D. (2010) 'Mitochondrial proton and electron leaks', *Essays in Biochemistry*, 47, pp. 53–67. doi:10.1042/bse0470053.

Jean-Martin Charcot (1875) 'Sixième Leçon – Huitième Leçon', in Adrien Delahaye and Désiré-Magloire Bourneville (eds) *Leçons sur les maladies du système nerveux faites à la Salpêtrière*. 2nd edn. Paris, pp. 189–271. Available at: http://catalogue.bnf.fr/ark:/12148/cb30222614z (Accessed: 6 January 2022).

Jelcic, Ivan, Al Nimer, F., Wang, J., Lentsch, V., Planas, R., Jelcic, Ilijas, Madjovski, A., Ruhrmann, S., Faigle, W., Frauenknecht, K., Pinilla, C., Santos, R., Hammer, C., Ortiz, Y., Opitz, L., Grönlund, H., Rogler, G., Boyman, O., Reynolds, R., Lutterotti, A., Khademi, M., Olsson, T., Piehl, F., Sospedra, M. and Martin, R. (2018) 'Memory B Cells Activate Brain-Homing, Autoreactive CD4+ T Cells in Multiple Sclerosis', *Cell*, 175(1), pp. 85-100.e23. doi:10.1016/j.cell.2018.08.011.

Jennings, L.J., Arcila, M.E., Corless, C., Kamel-Reid, S., Lubin, I.M., Pfeifer, J., Temple-Smolkin, R.L., Voelkerding, K.V. and Nikiforova, M.N. (2017) 'Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists', *The Journal of molecular diagnostics: JMD*, 19(3), pp. 341–365. doi:10.1016/j.jmoldx.2017.01.011.

Jersild, C., Svejgaard, A. and Fog, T. (1972) 'HL-A antigens and multiple sclerosis', *Lancet (London, England)*, 1(7762), pp. 1240–1241. doi:10.1016/s0140-6736(72)90962-2.

Joodi khanghah, O., Nourazarian, A., Khaki-Khatibi, F., Nikanfar, M., Laghousi, D., Vatankhah, A.M. and Moharami, S. (2020) 'Evaluation of the Diagnostic and Predictive

Value of Serum Levels of ANT1, ATG5, and Parkin in Multiple Sclerosis', *Clinical Neurology and Neurosurgery*, 197, p. 106197. doi:10.1016/j.clineuro.2020.106197.

Joshi, S. and Kermode, A.G. (2019) 'Harding's disease: an important MS mimic', *BMJ Case Reports*, 12(3), p. e228337. doi:10.1136/bcr-2018-228337.

Just, R.S., Irwin, J.A. and Parson, W. (2014) 'Questioning the prevalence and reliability of human mitochondrial DNA heteroplasmy from massively parallel sequencing data', *Proceedings of the National Academy of Sciences of the United States of America*, 111(43), pp. E4546-4547. doi:10.1073/pnas.1413478111.

Just, R.S., Irwin, J.A. and Parson, W. (2015) 'Mitochondrial DNA heteroplasmy in the emerging field of massively parallel sequencing', *Forensic Science International. Genetics*, 18, pp. 131–139. doi:10.1016/j.fsigen.2015.05.003.

Kabat, E.A., Wolf, A. and Bezer, A.E. (1947) 'THE RAPID PRODUCTION OF ACUTE DISSEMINATED ENCEPHALOMYELITIS IN RHESUS MONKEYS BY INJECTION OF HETEROLOGOUS AND HOMOLOGOUS BRAIN TISSUE WITH ADJUVANTS', *The Journal of Experimental Medicine*, 85(1), pp. 117–130. doi:10.1084/jem.85.1.117.

Kalman, B. and Alder, H. (1998) 'Is the mitochondrial DNA involved in determining susceptibility to multiple sclerosis?', *Acta Neurologica Scandinavica*, 98(4), pp. 232–237. doi:10.1111/j.1600-0404.1998.tb07301.x.

Kalman, B., Li, S., Chatterjee, D., O'Connor, J., Voehl, M.R., Brown, M.D. and Alder, H. (1999) 'Large scale screening of the mitochondrial DNA reveals no pathogenic mutations but a haplotype associated with multiple sclerosis in Caucasians', *Acta Neurologica Scandinavica*, 99(1), pp. 16–25. doi:10.1111/j.1600-0404.1999.tb00653.x.

Kalman, B., Lublin, F.D. and Alder, H. (1995) 'Mitochondrial DNA mutations in multiple sclerosis', *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 1(1), pp. 32–36. doi:10.1177/135245859500100106.

Kalman, B., Lublin, F.D. and Alder, H. (1996) 'Characterization of the mitochondrial DNA in patients with multiple sclerosis', *Journal of the Neurological Sciences*, 140(1–2), pp. 75–84. doi:10.1016/0022-510x(96)00112-8.

Kang, E., Wang, X., Tippner-Hedges, R., Ma, H., Folmes, C.D.L., Gutierrez, N.M., Lee, Y., Van Dyken, C., Ahmed, R., Li, Y., Koski, A., Hayama, T., Luo, S., Harding, C.O., Amato, P., Jensen, J., Battaglia, D., Lee, D., Wu, D., Terzic, A., Wolf, D.P., Huang, T. and Mitalipov, S. (2016) 'Age-Related Accumulation of Somatic Mitochondrial DNA Mutations in Adult-Derived Human iPSCs', *Cell Stem Cell*, 18(5), pp. 625–636. doi:10.1016/j.stem.2016.02.005.

Kappos, L., Hartung, H.-P., Freedman, M.S., Boyko, A., Radü, E.W., Mikol, D.D., Lamarine, M., Hyvert, Y., Freudensprung, U., Plitz, T. and Beek, J. van (2014) 'Atacicept in multiple sclerosis (ATAMS): a randomised, placebo-controlled, double-blind, phase 2 trial', *The Lancet Neurology*, 13(4), pp. 353–363. doi:10.1016/S1474-4422(14)70028-6.

Karlsson Linnér, R., Mallard, T.T., Barr, P.B., Sanchez-Roige, S., Madole, J.W., Driver, M.N., Poore, H.E., de Vlaming, R., Grotzinger, A.D., Tielbeek, J.J., Johnson, E.C., Liu, M., Rosenthal, S.B., Ideker, T., Zhou, H., Kember, R.L., Pasman, J.A., Verweij, K.J.H., Liu, D.J., Vrieze, S., Porjesz, B., Hesselbrock, V., Foroud, T.M., Agrawal, A., Edenberg, H.J., Nurnberger Jr, J.I., Liu, Y., Kuperman, S., Kramer, J., Meyer, J.L., Kamarajan, C., Pandey, A.K., Bierut, L., Rice, J., Bucholz, K.K., Schuckit, M.A., Tischfield, J., Brooks, A., Hart, R.P., Almasy, L., Dick, D.M., Salvatore, J.E., Goate, A., Kapoor, M., Slesinger, P., Scott, D.M., Bauer, L., Wetherill, L., Xuei, X., Lai, D., O'Connor, S.J., Plawecki, M.H., Lourens, S., Acion, L., Chan, G., Chorlian, D.B., Zhang, J., Kinreich, S., Pandey, G., Chao, M.J., Anokhin, A.P., McCutcheon, V.V., Saccone, S., Aliev, F., Barr, P.B., Chin, H., Parsian, A., Kranzler, H.R., Gelernter, J., Harris, K.M., Tucker-Drob, E.M., Waldman, I.D., Palmer, A.A., Harden, K.P., Koellinger, P.D., Dick, D.M., and COGA Collaborators (2021) 'Multivariate analysis of 1.5 million people identifies genetic associations with traits related to self-regulation and addiction', *Nature Neuroscience*, 24(10), pp. 1367–1376. doi:10.1038/s41593-021-00908-3.

Kashem, S.W., Haniffa, M. and Kaplan, D.H. (2017) 'Antigen-Presenting Cells in the Skin', *Annual Review of Immunology*, 35(1), pp. 469–499. doi:10.1146/annurev-immunol-051116-052215.

Kaskow, B.J. and Baecher-Allan, C. (2018) 'Effector T Cells in Multiple Sclerosis', *Cold Spring Harbor Perspectives in Medicine*, 8(4), p. a029025. doi:10.1101/cshperspect.a029025.

Kassambara, A. (2021) *rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. Available at: https://CRAN.R-project.org/package=rstatix (Accessed: 2 February 2022).

Kavak, K.S., Teter, B.E., Hagemeier, J., Zakalik, K. and Weinstock-Guttman, B. (2015) 'Higher weight in adolescence and young adulthood is associated with an earlier age at multiple sclerosis onset', *Multiple Sclerosis Journal*, 21(7), pp. 858–865. doi:10.1177/1352458514555787.

Kelly, R.D.W., Mahmud, A., McKenzie, M., Trounce, I.A. and St John, J.C. (2012) 'Mitochondrial DNA copy number is regulated in a tissue specific manner by DNA methylation of the nuclear-encoded DNA polymerase gamma A', *Nucleic Acids Research*, 40(20), pp. 10124–10138. doi:10.1093/nar/gks770.

Kennedy, S.R., Salk, J.J., Schmitt, M.W. and Loeb, L.A. (2013) 'Ultra-Sensitive Sequencing Reveals an Age-Related Increase in Somatic Mitochondrial Mutations That Are Inconsistent with Oxidative Damage', *PLOS Genetics*, 9(9), p. e1003794. doi:10.1371/journal.pgen.1003794.

Khrapko, K., Coller, H.A., André, P.C., Li, X.-C., Hanekamp, J.S. and Thilly, W.G. (1997) 'Mitochondrial mutational spectra in human cells and tissues', *Proceedings of the National Academy of Sciences*, 94(25), p. 13798. doi:10.1073/pnas.94.25.13798.

Kingwell, E., Marriott, J.J., Jetté, N., Pringsheim, T., Makhani, N., Morrow, S.A., Fisk, J.D., Evans, C., Béland, S.G., Kulaga, S., Dykeman, J., Wolfson, C., Koch, M.W. and Marrie, R.A. (2013) 'Incidence and prevalence of multiple sclerosis in Europe: a systematic review', *BMC Neurology*, 13(1), p. 128. doi:10.1186/1471-2377-13-128.

Kipnis, J., Gadani, S. and Derecki, N.C. (2012) 'Pro-cognitive properties of T cells', *Nature Reviews Immunology*, 12(9), pp. 663–669. doi:10.1038/nri3280.

Kister, I., Bacon, T.E., Chamot, E., Salter, A.R., Cutter, G.R., Kalina, J.T. and Herbert, J. (2013) 'Natural History of Multiple Sclerosis Symptoms', *International Journal of MS Care*, 15(3), pp. 146–156. doi:10.7224/1537-2073.2012-053.

Kitz, A., de Marcken, M., Gautron, A.-S., Mitrovic, M., Hafler, D.A. and Dominguez-Villar, M. (2016) 'AKT isoforms modulate Th1-like Treg generation and function in human autoimmune disease', *EMBO reports*, 17(8), pp. 1169–1183. doi:10.15252/embr.201541905.

Klein, R.S., Garber, C. and Howard, N. (2017) 'Infectious immunity in the central nervous system and brain function', *Nature Immunology*, 18(2), pp. 132–141. doi:10.1038/ni.3656.

Kleinewietfeld, M. and Hafler, D.A. (2013) 'The plasticity of human Treg and Th17 cells and its role in autoimmunity', *Seminars in Immunology*, 25(4), pp. 305–312. doi:10.1016/j.smim.2013.10.009.

Kleiveland, C.R. (2015) 'Peripheral Blood Mononuclear Cells', in Verhoeckx, K., Cotter, P., López-Expósito, I., Kleiveland, C., Lea, T., Mackie, A., Requena, T., Swiatecka, D., and Wichers, H. (eds) *The Impact of Food Bioactives on Health: in vitro and ex vivo models*. Cham (CH): Springer. Available at: http://www.ncbi.nlm.nih.gov/books/NBK500157/ (Accessed: 16 January 2022).

Klotz, L., Eschborn, M., Lindner, M., Liebmann, M., Herold, M., Janoschka, C., Torres Garrido, B., Schulte-Mecklenbeck, A., Gross, C.C., Breuer, J., Hundehege, P., Posevitz, V., Pignolet, B., Nebel, G., Glander, S., Freise, N., Austermann, J., Wirth, T., Campbell, G.R., Schneider-Hohendorf, T., Eveslage, M., Brassat, D., Schwab, N., Loser, K., Roth, J., Busch, K.B., Stoll, M., Mahad, D.J., Meuth, S.G., Turner, T., Bar-Or, A. and Wiendl, H. (2019) 'Teriflunomide treatment for multiple sclerosis modulates T cell mitochondrial respiration with affinity-dependent effects', *Science Translational Medicine*, 11(490), p. eaao5563. doi:10.1126/scitranslmed.aao5563.

Kobelt, G., Thompson, A., Berg, J., Gannedahl, M. and Eriksson, J. (2017) 'New insights into the burden and costs of multiple sclerosis in Europe', *Multiple Sclerosis Journal*, 23(8), pp. 1123–1136. doi:10.1177/1352458517694432.

Koboldt, D.C., Zhang, Q., Larson, D.E., Shen, D., McLellan, M.D., Lin, L., Miller, C.A., Mardis, E.R., Ding, L. and Wilson, R.K. (2012) 'VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing', *Genome Research*, 22(3), pp. 568–576. doi:10.1101/gr.129684.111.

Koch-Henriksen, N., Thygesen, L.C., Stenager, E., Laursen, B. and Magyari, M. (2018) 'Incidence of MS has increased markedly over six decades in Denmark particularly with late onset and in women', *Neurology*, 90(22), p. e1954. doi:10.1212/WNL.0000000000005612.

Koduah, P., Paul, F. and Dörr, J.-M. (2017) 'Vitamin D in the prevention, prediction and treatment of neurodegenerative and neuroinflammatory diseases', *The EPMA journal*, 8(4), pp. 313–325. doi:10.1007/s13167-017-0120-8.

Kohm, A.P., Carpentier, P.A., Anger, H.A. and Miller, S.D. (2002) 'Cutting Edge: $CD4^+CD25^+$ Regulatory T Cells Suppress Antigen-Specific Autoreactive Immune Responses and Central Nervous System Inflammation During Active Experimental Autoimmune Encephalomyelitis', *The Journal of Immunology*, 169(9), p. 4712. doi:10.4049/jimmunol.169.9.4712.

Kolčava, J., Kočica, J., Hulová, M., Dušek, L., Horáková, M., Keřkovský, M., Stulík, J., Dostál, M., Kuhn, M., Vlčková, E., Bednařík, J. and Benešová, Y. (2020) 'Conversion of clinically isolated syndrome to multiple sclerosis: a prospective study', *Multiple Sclerosis and Related Disorders*, 44. doi:10.1016/j.msard.2020.102262.

Kolev, M., Dimeloe, S., Le Friec, G., Navarini, A., Arbore, G., Povoleri, G.A., Fischer, M., Belle, R., Loeliger, J., Develioglu, L., Bantug, G.R., Watson, J., Couzi, L., Afzali, B., Lavender, P., Hess, C. and Kemper, C. (2015) 'Complement Regulates Nutrient Influx and Metabolic Reprogramming during Th1 Cell Responses', *Immunity*, 42(6), pp. 1033–1047. doi:10.1016/j.immuni.2015.05.024.

Korin, B., Ben-Shaanan, T.L., Schiller, M., Dubovik, T., Azulay-Debby, H., Boshnak, N.T., Koren, T. and Rolls, A. (2017) 'High-dimensional, single-cell characterization of the brain's immune compartment', *Nature Neuroscience*, 20(9), pp. 1300–1309. doi:10.1038/nn.4610.

Korn, T. and Kallies, A. (2017) 'T cell responses in the central nervous system', *Nature Reviews Immunology*, 17(3), pp. 179–194. doi:10.1038/nri.2016.144.

Kornberg, M.D., Bhargava, P., Kim, P.M., Putluri, V., Snowman, A.M., Putluri, N., Calabresi, P.A. and Snyder, S.H. (2018) 'Dimethyl fumarate targets GAPDH and aerobic glycolysis to modulate immunity', *Science*, 360(6387), pp. 449–453. doi:10.1126/science.aan4665.

Köster, J. and Rahmann, S. (2012) 'Snakemake--a scalable bioinformatics workflow engine', *Bioinformatics (Oxford, England)*, 28(19), pp. 2520–2522. doi:10.1093/bioinformatics/bts480.

Kozin, M., Kulakova, O., Kiselev, I., Baulina, N., Boyko, A. and Favorova, O. (2020) 'Mitonuclear interactions influence multiple sclerosis risk', *Gene*, 758, p. 144962. doi:10.1016/j.gene.2020.144962.

Kozin, M.S., Kulakova, O.G., Kiselev, I.S., Balanovsky, O.P., Boyko, A.N. and Favorova, O.O. (2018) 'Variants of Mitochondrial Genome and Risk of Multiple Sclerosis Development in Russians', *Acta Naturae*, 10(4), pp. 79–86.

Kozin, M.S., Kulakova, O.G., Kiselev, I.S., Boyko, A.N. and Favorova, O.O. (2020) '[Variability of the Mitochondrial Genome and Development of the Primary Progressing form of Multiple Sclerosis]', *Molekuliarnaia Biologiia*, 54(4), pp. 596–602. doi:10.31857/S0026898420040084.

Kuhle, J., Disanto, G., Dobson, R., Adiutori, R., Bianchi, L., Topping, J., Bestwick, J., Meier, U.-C., Marta, M., Costa, G.D., Runia, T., Evdoshenko, E., Lazareva, N., Thouvenot, E., Iaffaldano, P., Direnzo, V., Khademi, M., Piehl, F., Comabella, M., Sombekke, M., Killestein, J., Hegen, H., Rauch, S., D'Alfonso, S., Alvarez-Cermeño, J., Kleinová, P., Horáková, D., Roesler, R., Lauda, F., Llufriu, S., Avsar, T., Uygunoglu, U., Altintas, A., Saip, S., Menge, T., Rajda, C., Bergamaschi, R., Moll, N., Khalil, M., Marignier, R., Dujmovic, I., Larsson, H., Malmestrom, C., Scarpini, E., Fenoglio, C., Wergeland, S., Laroni, A., Annibali, V., Romano, S., Martínez, A., Carra, A., Salvetti, M., Uccelli, A., Torkildsen, Ø., Myhr, K., Galimberti, D., Rejdak, K., Lycke, J., Frederiksen, J., Drulovic, J., Confavreux, C., Brassat, D., Enzinger, C., Fuchs, S., Bosca, I., Pelletier, J., Picard, C., Colombo, E., Franciotta, D., Derfuss, T., Lindberg, R., Yaldizli, Ö., Vécsei, L., Kieseier, B., Hartung, H., Villoslada, P., Siva, A., Saiz, A., Tumani, H., Havrdová, E., Villar, L., Leone, M., Barizzone, N., Deisenhammer, F., Teunissen, C., Montalban, X., Tintoré, M., Olsson, T., Trojano, M., Lehmann, S., Castelnovo, G., Lapin, S., Hintzen, R., Kappos, L., Furlan, R., Martinelli, V., Comi, G., Ramagopalan, S. and Giovannoni, G. (2015) 'Conversion from clinically isolated syndrome to multiple sclerosis: A large multicentre study', *Multiple Sclerosis Journal*, 21(8), pp. 1013–1024. doi:10.1177/1352458514568827.

Kummer, E. and Ban, N. (2021) 'Mechanisms and regulation of protein synthesis in mitochondria', *Nature Reviews Molecular Cell Biology*, 22(5), pp. 307–325. doi:10.1038/s41580-021-00332-2.

Kunkl, M., Frascolla, S., Amormino, C., Volpe, E. and Tuosto, L. (2020) 'T Helper Cells: The Modulators of Inflammation in Multiple Sclerosis', *Cells*, 9(2). doi:10.3390/cells9020482.

Kurtzke, J.F. (1983) 'Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)', *Neurology*, 33(11), pp. 1444–1452. doi:10.1212/wnl.33.11.1444.

La Rocca, C., Carbone, F., De Rosa, V., Colamatteo, A., Galgani, M., Perna, F., Lanzillo, R., Brescia Morra, V., Orefice, G., Cerillo, I., Florio, C., Maniscalco, G.T., Salvetti, M., Centonze, D., Uccelli, A., Longobardi, S., Visconti, A. and Matarese, G. (2017) 'Immunometabolic profiling of T cells from patients with relapsing-remitting multiple sclerosis reveals an impairment in glycolysis and mitochondrial respiration', *Metabolism*, 77, pp. 39–46. doi:10.1016/j.metabol.2017.08.011.

Lago-Baldaia, I., Fernandes, V.M. and Ackerman, S.D. (2020) 'More Than Mortar: Glia as Architects of Nervous System Development and Disease', *Frontiers in Cell and Developmental Biology*, 8, p. 1527. doi:10.3389/fcell.2020.611269.

Lambe, J., Saidha, S. and Bermel, R.A. (2020) 'Optical coherence tomography and multiple sclerosis: Update on clinical application and role in clinical trials', *Multiple Sclerosis Journal*, 26(6), pp. 624–639. doi:10.1177/1352458519872751.

Landrum, M.J., Lee, J.M., Benson, M., Brown, G.R., Chao, C., Chitipiralla, S., Gu, B., Hart, J., Hoffman, D., Jang, W., Karapetyan, K., Katz, K., Liu, C., Maddipatla, Z., Malheiro, A., McDaniel, K., Ovetsky, M., Riley, G., Zhou, G., Holmes, J.B., Kattman, B.L. and Maglott, D.R. (2018) 'ClinVar: improving access to variant interpretations and supporting evidence', *Nucleic Acids Research*, 46(D1), pp. D1062–D1067. doi:10.1093/nar/gkx1153.

Langer-Gould, A., Popat, R.A., Huang, S.M., Cobb, K., Fontoura, P., Gould, M.K. and Nelson, L.M. (2006) 'Clinical and Demographic Predictors of Long-term Disability in Patients With Relapsing-Remitting Multiple Sclerosis: A Systematic Review', *Archives of Neurology*, 63(12), pp. 1686–1691. doi:10.1001/archneur.63.12.1686.

Lanz, T.V., Brewer, R.C., Ho, P.P., Moon, J.-S., Jude, K.M., Fernandez, D., Fernandes, R.A., Gomez, A.M., Nadj, G.-S., Bartley, C.M., Schubert, R.D., Hawes, I.A., Vazquez, S.E., Iyer, M., Zuchero, J.B., Teegen, B., Dunn, J.E., Lock, C.B., Kipp, L.B., Cotham, V.C., Ueberheide, B.M., Aftab, B.T., Anderson, M.S., DeRisi, J.L., Wilson, M.R., Bashford-Rogers, R.J.M., Platten, M., Garcia, K.C., Steinman, L. and Robinson, W.H. (2022) 'Clonally Expanded B Cells in Multiple Sclerosis Bind EBV EBNA1 and GlialCAM', *Nature* [Preprint]. doi:10.1038/s41586-022-04432-7.

Lareau, C.A., Ludwig, L.S., Muus, C., Gohil, S.H., Zhao, T., Chiang, Z., Pelka, K., Verboon, J.M., Luo, W., Christian, E., Rosebrock, D., Getz, G., Boland, G.M., Chen, F., Buenrostro, J.D., Hacohen, N., Wu, C.J., Aryee, M.J., Regev, A. and Sankaran, V.G. (2021) 'Massively parallel single-cell mitochondrial DNA genotyping and chromatin profiling', *Nature Biotechnology*, 39(4), pp. 451–461. doi:10.1038/s41587-020-0645-6.

Largani, S.H.H., Borhani-Haghighi, M., Pasbakhsh, P., Mahabadi, V.P., Nekoonam, S., Shiri, E., Kashani, I.R. and Zendehdel, A. (2019) 'Oligoprotective effect of metformin through the AMPK-dependent on restoration of mitochondrial hemostasis in the cuprizone-induced multiple sclerosis model', *Journal of Molecular Histology*, 50(3), pp. 263–271. doi:10.1007/s10735-019-09824-0.

Laricchia, K.M., Lake, N.J., Watts, N.A., Shand, M., Haessly, A., Gauthier, L., Benjamin, D., Banks, E., Soto, J., Garimella, K., Emery, J., Genome Aggregation Database Consortium, Rehm, H.L., MacArthur, D.G., Tiao, G., Lek, M., Mootha, V.K. and Calvo, S.E. (2021) 'Mitochondrial DNA variation across 56,434 individuals in gnomAD', *bioRxiv*, p. 2021.07.23.453510. doi:10.1101/2021.07.23.453510.

Lassmann, H. (2018) 'Multiple Sclerosis Pathology', *Cold Spring Harbor Perspectives in Medicine*, 8(3), p. a028936. doi:10.1101/cshperspect.a028936.

Lassmann, H. and van Horssen, J. (2016) 'Oxidative stress and its impact on neurons and glia in multiple sclerosis lesions', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1862(3), pp. 506–510. doi:10.1016/j.bbadis.2015.09.018.

Latorre-Pellicer, A., Lechuga-Vieco, A.V., Johnston, I.G., Hämäläinen, R.H., Pellico, J., Justo-Méndez, R., Fernández-Toro, J.M., Clavería, C., Guaras, A., Sierra, R., Llop, J., Torres, M., Criado, L.M., Suomalainen, A., Jones, N.S., Ruíz-Cabello, J. and Enríquez, J.A. (2019) 'Regulation of Mother-to-Offspring Transmission of mtDNA Heteroplasmy', *Cell Metabolism*, 30(6), pp. 1120-1130.e5. doi:10.1016/j.cmet.2019.09.007.

Lawless, C., Greaves, L., Reeve, A.K., Turnbull, D.M. and Vincent, A.E. (2020) 'The rise and rise of mitochondrial DNA mutations', *Open Biology*, 10(5), p. 200061. doi:10.1098/rsob.200061.

Lee, C., Yen, K. and Cohen, P. (2013) 'Humanin: a harbinger of mitochondrial-derived peptides?', *Trends in Endocrinology & Metabolism*, 24(5), pp. 222–228. doi:10.1016/j.tem.2013.01.005.

Lee, C., Zeng, J., Drew, B.G., Sallam, T., Martin-Montalvo, A., Wan, J., Kim, S.-J., Mehta, H., Hevener, A.L., de Cabo, R. and Cohen, P. (2015) 'The Mitochondrial-Derived Peptide MOTS-c Promotes Metabolic Homeostasis and Reduces Obesity and Insulin Resistance', *Cell Metabolism*, 21(3), pp. 443–454. doi:10.1016/j.cmet.2015.02.009.

Lee, H.Y., Song, I., Ha, E., Cho, S.-B., Yang, W.I. and Shin, K.-J. (2008) 'mtDNAmanager: a Web-based tool for the management and quality analysis of mitochondrial DNA control-region sequences', *BMC bioinformatics*, 9, p. 483. doi:10.1186/1471-2105-9-483.

Leites, E.P. and Morais, V.A. (2018) 'Mitochondrial quality control pathways: PINK1 acts as a gatekeeper', *Mitochondrial Dynamics*, 500(1), pp. 45–50. doi:10.1016/j.bbrc.2017.06.096.

Leites, E.P. and Morais, V.A. (2021) 'The PINK1-Mediated Crosstalk between Neural Cells and the Underlying Link to Parkinson's Disease', *Cells*, 10(6), p. 1395. doi:10.3390/cells10061395.

Lepore, M., Mori, L. and De Libero, G. (2018) 'The Conventional Nature of Non-MHC-Restricted T Cells', *Frontiers in Immunology*, 9, p. 1365. doi:10.3389/fimmu.2018.01365.

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020a) 'Adaptive Immunity: B Cells & Antibodies', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1171926486 (Accessed: 26 December 2021).

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020b) 'Adaptive Immunity: Lymphocyte Antigen Receptors', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1171926331 (Accessed: 26 December 2021).

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020c) 'Adaptive Immunity: T-Cell–Mediated Immunity', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1171926408 (Accessed: 26 December 2021).

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020d) 'Complement', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1171926571 (Accessed: 26 December 2021).

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020e) 'Innate Immunity', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1176200364 (Accessed: 26 December 2021).

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020f) 'Major Histocompatibility Complex & Transplantation', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1171926542 (Accessed: 26 December 2021).

Levinson, W., Chin-Hong, P., Joyce, E.A., Nussbaum, J. and Schwartz, B. (2020g) 'Overview of Immunity', in *Review of Medical Microbiology & Immunology: A Guide to Clinical Infectious Diseases, 16e*. New York, NY: McGraw Hill.

Li, B., Krishnan, V.G., Mort, M.E., Xin, F., Kamati, K.K., Cooper, D.N., Mooney, S.D. and Radivojac, P. (2009) 'Automated inference of molecular mechanisms of disease from amino acid substitutions', *Bioinformatics (Oxford, England)*, 25(21), pp. 2744–2750. doi:10.1093/bioinformatics/btp528.

Li, D. and Wu, M. (2021) 'Pattern recognition receptors in health and diseases', *Signal Transduction and Targeted Therapy*, 6(1), p. 291. doi:10.1038/s41392-021-00687-0.

Li, H. (2013) 'Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM', *arXiv:1303.3997 [q-bio]* [Preprint].

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and 1000 Genome Project Data Processing Subgroup (2009) 'The Sequence Alignment/Map format and SAMtools', *Bioinformatics (Oxford, England)*, 25(16), pp. 2078–2079. doi:10.1093/bioinformatics/btp352.

Li, M., Schröder, R., Ni, S., Madea, B. and Stoneking, M. (2015) 'Extensive tissue-related and allele-related mtDNA heteroplasmy suggests positive selection for somatic mutations', *Proceedings of the National Academy of Sciences of the United States of America*, 112(8), pp. 2491–2496. doi:10.1073/pnas.1419651112.

Li, Q. and Barres, B.A. (2018) 'Microglia and macrophages in brain homeostasis and disease', *Nature Reviews Immunology*, 18(4), pp. 225–242. doi:10.1038/nri.2017.125.

Li, R., Patterson, K.R. and Bar-Or, A. (2018) 'Reassessing B cell contributions in multiple sclerosis', *Nature Immunology*, 19(7), pp. 696–707. doi:10.1038/s41590-018-0135-x.

Li, X., Xu, D., Cheng, B., Zhou, Y., Chen, Z. and Wang, Y. (2021) 'Mitochondrial DNA insert into CD40 ligand gene-associated X-linked hyper-IgM syndrome', *Molecular Genetics & Genomic Medicine*, 9(5). doi:10.1002/mgg3.1646.

Liebmann, M., Korn, L., Janoschka, C., Albrecht, S., Lauks, S., Herrmann, A.M., Schulte-Mecklenbeck, A., Schwab, N., Schneider-Hohendorf, T., Eveslage, M., Wildemann, B.,

Luessi, F., Schmidt, S., Diebold, M., Bittner, S., Gross, C.C., Kovac, S., Zipp, F., Derfuss, T., Kuhlmann, T., König, S., Meuth, S.G., Wiendl, H. and Klotz, L. (2021) 'Dimethyl fumarate treatment restrains the antioxidative capacity of T cells to control autoimmunity', *Brain*, 144(10), pp. 3126–3141. doi:10.1093/brain/awab307.

Liessem-Schmitz, A., Teske, N., Scheld, M., Nyamoya, S., Zendedel, A., Beyer, C., Clarner, T. and Fragoulis, A. (2018) 'Nrf2 Signaling in Sodium Azide-Treated Oligodendrocytes Restores Mitochondrial Functions', *Journal of Molecular Neuroscience*, 66(2), pp. 229–237. doi:10.1007/s12031-018-1159-2.

Lin, T.-Y., Vitkova, V., Asseyer, S., Martorell Serra, I., Motamedi, S., Chien, C., Ditzhaus, M., Papadopoulou, A., Benkert, P., Kuhle, J., Bellmann-Strobl, J., Ruprecht, K., Paul, F., Brandt, A.U. and Zimmermann, H.G. (2021) 'Increased Serum Neurofilament Light and Thin Ganglion Cell-Inner Plexiform Layer Are Additive Risk Factors for Disease Activity in Early Multiple Sclerosis', *Neurology(R) Neuroimmunology & Neuroinflammation*, 8(5), p. e1051. doi:10.1212/NXI.0000000000001051.

Lin, X., Chen, M., Liu, Y., Guo, Z., He, X., Brand, D. and Zheng, S.G. (2013) 'Advances in distinguishing natural from induced Foxp3(+) regulatory T cells', *International Journal of Clinical and Experimental Pathology*, 6(2), pp. 116–123.

Lindstrand, A., Eisfeldt, J., Pettersson, M., Carvalho, C.M.B., Kvarnung, M., Grigelioniene, G., Anderlid, B.-M., Bjerin, O., Gustavsson, P., Hammarsjö, A., Georgii-Hemming, P., Iwarsson, E., Johansson-Soller, M., Lagerstedt-Robinson, K., Lieden, A., Magnusson, M., Martin, M., Malmgren, H., Nordenskjöld, M., Norling, A., Sahlin, E., Stranneheim, H., Tham, E., Wincent, J., Ygberg, S., Wedell, A., Wirta, V., Nordgren, A., Lundin, J. and Nilsson, D. (2019) 'From cytogenetics to cytogenomics: whole-genome sequencing as a first-line test comprehensively captures the diverse spectrum of disease-causing genetic variation underlying intellectual disability', *Genome Medicine*, 11(1), p. 68. doi:10.1186/s13073-019-0675-1.

Liu, J.-F., Tsao, Y.-T. and Hou, C.-H. (2015) 'Amphiregulin enhances intercellular adhesion molecule-1 expression and promotes tumor metastasis in human osteosarcoma', *Oncotarget*, 6(38), pp. 40880–40895. doi:10.18632/oncotarget.5679.

Lodato, M.A., Rodin, R.E., Bohrson, C.L., Coulter, M.E., Barton, A.R., Kwon, M., Sherman, M.A., Vitzthum, C.M., Luquette, L.J., Yandava, C.N., Yang, P., Chittenden, T.W., Hatem, N.E., Ryu, S.C., Woodworth, M.B., Park, P.J. and Walsh, C.A. (2018) 'Aging and neurodegeneration are associated with increased mutations in single human neurons', *Science*, 359(6375), pp. 555–559. doi:10.1126/science.aao4426.

Lopez, J.V., Yuhki, N., Masuda, R., Modi, W. and O'Brien, S.J. (1994) 'Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat', *Journal of Molecular Evolution*, 39(2), pp. 174–190. doi:10.1007/BF00163806.

Lossius, A., Riise, T., Pugliatti, M., Bjørnevik, K., Casetta, I., Drulovic, J., Granieri, E., Kampman, M.T., Landtblom, A.-M., Lauer, K., Magalhaes, S., Myhr, K.-M., Pekmezovic, T., Wesnes, K., Wolfson, C. and Holmøy, T. (2014) 'Season of infectious mononucleosis and risk

of multiple sclerosis at different latitudes; the EnvIMS Study', *Multiple Sclerosis Journal*, 20(6), pp. 669–674. doi:10.1177/1352458513505693.

Lott, M.T., Leipzig, J.N., Derbeneva, O., Xie, H.M., Chalkia, D., Sarmady, M., Procaccio, V. and Wallace, D.C. (2013) 'mtDNA Variation and Analysis Using Mitomap and Mitomaster', *Current Protocols in Bioinformatics*, 44, p. 1.23.1-26. doi:10.1002/0471250953.bi0123s44.

Louveau, A., Harris, T.H. and Kipnis, J. (2015) 'Revisiting the Mechanisms of CNS Immune Privilege', *Special Issue: Neuroimmunology*, 36(10), pp. 569–577. doi:10.1016/j.it.2015.08.006.

Louveau, A., Herz, J., Alme, M.N., Salvador, A.F., Dong, M.Q., Viar, K.E., Herod, S.G., Knopp, J., Setliff, J.C., Lupi, A.L., Da Mesquita, S., Frost, E.L., Gaultier, A., Harris, T.H., Cao, R., Hu, S., Lukens, J.R., Smirnov, I., Overall, C.C., Oliver, G. and Kipnis, J. (2018) 'CNS lymphatic drainage and neuroinflammation are regulated by meningeal lymphatic vasculature', *Nature Neuroscience*, 21(10), pp. 1380–1391. doi:10.1038/s41593-018-0227-9.

Louveau, A., Smirnov, I., Keyes, T.J., Eccles, J.D., Rouhani, S.J., Peske, J.D., Derecki, N.C., Castle, D., Mandell, J.W., Lee, K.S., Harris, T.H. and Kipnis, J. (2015) 'Structural and functional features of central nervous system lymphatic vessels', *Nature*, 523(7560), pp. 337–341. doi:10.1038/nature14432.

Lowe, R., Shirley, N., Bleackley, M., Dolan, S. and Shafee, T. (2017) 'Transcriptomics technologies', *PLoS computational biology*, 13(5), p. e1005457. doi:10.1371/journal.pcbi.1005457.

Lowther, D.E., Chong, D.L., Ascough, S., Ettorre, A., Ingram, R.J., Boyton, R.J. and Altmann, D.M. (2013) 'Th1 not Th17 cells drive spontaneous MS-like disease despite a functional regulatory T cell response', *Acta Neuropathologica*, 126(4), pp. 501–515. doi:10.1007/s00401-013-1159-9.

Lu, B., Lee, J., Nie, X., Li, M., Morozov, Y.I., Venkatesh, S., Bogenhagen, D.F., Temiakov, D. and Suzuki, C.K. (2013) 'Phosphorylation of Human TFAM in Mitochondria Impairs DNA Binding and Promotes Degradation by the AAA+ Lon Protease', *Molecular Cell*, 49(1), pp. 121–132. doi:10.1016/j.molcel.2012.10.023.

Lucchinetti, C.F., Popescu, B.F.G., Bunyan, R.F., Moll, N.M., Roemer, S.F., Lassmann, H., Brück, W., Parisi, J.E., Scheithauer, B.W., Giannini, C., Weigand, S.D., Mandrekar, J. and Ransohoff, R.M. (2011) 'Inflammatory Cortical Demyelination in Early Multiple Sclerosis', *New England Journal of Medicine*, 365(23), pp. 2188–2197. doi:10.1056/NEJMoa1100648.

Ludwig, L.S., Lareau, C.A., Ulirsch, J.C., Christian, E., Muus, C., Li, L.H., Pelka, K., Ge, W., Oren, Y., Brack, A., Law, T., Rodman, C., Chen, J.H., Boland, G.M., Hacohen, N., Rozenblatt-Rosen, O., Aryee, M.J., Buenrostro, J.D., Regev, A. and Sankaran, V.G. (2019) 'Lineage Tracing in Humans Enabled by Mitochondrial Mutations and Single-Cell Genomics', *Cell*, 176(6), pp. 1325-1339.e22. doi:10.1016/j.cell.2019.01.022.

Luft, R., Ikkos, D., Palmieri, G., Ernster, L. and Afzelius, B. (1962) 'A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial

respiratory control: a correlated clinical, biochemical, and morphological study', *The Journal of Clinical Investigation*, 41, pp. 1776–1804. doi:10.1172/JCI104637.

Luo, S., Valencia, C.A., Zhang, J., Lee, N.-C., Slone, J., Gui, B., Wang, X., Li, Z., Dell, S., Brown, J., Chen, S.M., Chien, Y.-H., Hwu, W.-L., Fan, P.-C., Wong, L.-J., Atwal, P.S. and Huang, T. (2018) 'Biparental Inheritance of Mitochondrial DNA in Humans', *Proceedings of the National Academy of Sciences*, 115(51), p. 13039. doi:10.1073/pnas.1810946115.

Luu, M., Pautz, S., Kohl, V., Singh, R., Romero, R., Lucas, S., Hofmann, J., Raifer, H., Vachharajani, N., Carrascosa, L.C., Lamp, B., Nist, A., Stiewe, T., Shaul, Y., Adhikary, T., Zaiss, M.M., Lauth, M., Steinhoff, U. and Visekruna, A. (2019) 'The short-chain fatty acid pentanoate suppresses autoimmunity by modulating the metabolic-epigenetic crosstalk in lymphocytes', *Nature Communications*, 10(1), p. 760. doi:10.1038/s41467-019-08711-2.

Ma, H., Van Dyken, C., Darby, H., Mikhalchenko, A., Marti-Gutierrez, N., Koski, A., Liang, D., Li, Y., Tippner-Hedges, R., Kang, E., Lee, Y., Sidener, H., Ramsey, C., Hodge, T., Amato, P. and Mitalipov, S. (2021) 'Germline transmission of donor, maternal and paternal mtDNA in primates', *Human Reproduction*, 36(2), pp. 493–505. doi:10.1093/humrep/deaa308.

Magliozzi, R., Howell, O., Vora, A., Serafini, B., Nicholas, R., Puopolo, M., Reynolds, R. and Aloisi, F. (2007) 'Meningeal B-cell follicles in secondary progressive multiple sclerosis associate with early onset of disease and severe cortical pathology', *Brain*, 130(4), pp. 1089–1104. doi:10.1093/brain/awm038.

Mahad, D.J., Ziabreva, I., Campbell, G., Lax, N., White, K., Hanson, P.S., Lassmann, H. and Turnbull, D.M. (2009) 'Mitochondrial changes within axons in multiple sclerosis', *Brain*, 132(5), pp. 1161–1174. doi:10.1093/brain/awp046.

Malińska, D., Więckowski, M.R., Michalska, B., Drabik, K., Prill, M., Patalas-Krawczyk, P., Walczak, J., Szymański, J., Mathis, C., Van der Toorn, M., Luettich, K., Hoeng, J., Peitsch, M.C., Duszyński, J. and Szczepanowska, J. (2019) 'Mitochondria as a possible target for nicotine action', *Journal of Bioenergetics and Biomembranes*, 51(4), pp. 259–276. doi:10.1007/s10863-019-09800-z.

Mao, P., Manczak, M., Shirendeb, U.P. and Reddy, P.H. (2013) 'MitoQ, a mitochondria-targeted antioxidant, delays disease progression and alleviates pathogenesis in an experimental autoimmune encephalomyelitis mouse model of multiple sclerosis', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1832(12), pp. 2322–2331. doi:10.1016/j.bbadis.2013.09.005.

Marin, I. and Kipnis, J. (2013) 'Learning and memory … and the immune system', *Learning & Memory*, 20(10), pp. 601–606. doi:10.1101/lm.028357.112.

Marrodan, M., Crema, S., Rubstein, A., Alessandro, L., Fernandez, J., Correale, J. and Ysrraelit, M.C. (2021) 'Therapeutic plasma exchange in MS refractory relapses: Long-term outcome', *Multiple Sclerosis and Related Disorders*, 55, p. 103168. doi:10.1016/j.msard.2021.103168.

Marshall, C. and Parson, W. (2021) 'Interpreting NUMTs in forensic genetics: Seeing the forest for the trees', *Forensic Science International. Genetics*, 53, p. 102497. doi:10.1016/j.fsigen.2021.102497.

Marshall, C.R., Chowdhury, S., Taft, R.J., Lebo, M.S., Buchan, J.G., Harrison, S.M., Rowsey, R., Klee, E.W., Liu, P., Worthey, E.A., Jobanputra, V., Dimmock, D., Kearney, H.M., Bick, D., Kulkarni, S., Taylor, S.L., Belmont, J.W., Stavropoulos, D.J., Lennon, N.J., and Medical Genome Initiative (2020) 'Best practices for the analytical validation of clinical whole-genome sequencing intended for the diagnosis of germline disease', *NPJ genomic medicine*, 5, p. 47. doi:10.1038/s41525-020-00154-9.

Martínez-Pinilla, E., Rubio-Sardón, N., Villar-Conde, S., Navarro, G., del Valle, E., Tolivia, J., Franco, R. and Navarro, A. (2021) 'Cuprizone-Induced Neurotoxicity in Human Neural Cell Lines Is Mediated by a Reversible Mitochondrial Dysfunction: Relevance for Demyelination Models', *Brain Sciences*, 11(2), p. 272. doi:10.3390/brainsci11020272.

Martín-Montañez, E., Pavia, J., Valverde, N., Boraldi, F., Lara, E., Oliver, B., Hurtado-Guerrero, I., Fernandez, O. and Garcia-Fernandez, M. (2019) 'The S1P mimetic fingolimod phosphate regulates mitochondrial oxidative stress in neuronal cells', *Free Radical Biology and Medicine*, 137, pp. 116–130. doi:10.1016/j.freeradbiomed.2019.04.022.

Mascagni, P. and Sanctius, C. (1787) *Vasorum lymphaticorum corporis humani historia et ichnographia*. Ex typographia Pazzini Carli. Available at: https://anatomia.library.utoronto.ca/islandora/object/anatomia%3ARBAI052 (Accessed: 2 January 2022).

Massilani, D., Morley, M.W., Mentzer, S.M., Aldeias, V., Vernot, B., Miller, C., Stahlschmidt, M., Kozlikin, M.B., Shunkov, M.V., Derevianko, A.P., Conard, N.J., Wurz, S., Henshilwood, C.S., Vasquez, J., Essel, E., Nagel, S., Richter, J., Nickel, B., Roberts, R.G., Pääbo, S., Slon, V., Goldberg, P. and Meyer, M. (2022) 'Microstratigraphic preservation of ancient faunal and hominin DNA in Pleistocene cave sediments', *Proceedings of the National Academy of Sciences*, 119(1), p. e2113666118. doi:10.1073/pnas.2113666118.

Masuda, T., Sankowski, R., Staszewski, O., Böttcher, C., Amann, L., Sagar, Scheiwe, C., Nessler, S., Kunz, P., van Loo, G., Coenen, V.A., Reinacher, P.C., Michel, A., Sure, U., Gold, R., Grün, D., Priller, J., Stadelmann, C. and Prinz, M. (2019) 'Spatial and temporal heterogeneity of mouse and human microglia at single-cell resolution', *Nature*, 566(7744), pp. 388–392. doi:10.1038/s41586-019-0924-x.

Mayr-Wohlfart, U., Paulus, C., Henneberg, A. and Rödel, G. (1996) 'Mitochondrial DNA mutations in multiple sclerosis patients with severe optic involvement', *Acta Neurologica Scandinavica*, 94(3), pp. 167–171. doi:10.1111/j.1600-0404.1996.tb07048.x.

McGeachy, M.J., Stephens, L.A. and Anderton, S.M. (2005) 'Natural Recovery and Protection from Autoimmune Encephalomyelitis: Contribution of $CD4^+CD25^+$ Regulatory Cells within the Central Nervous System', *The Journal of Immunology*, 175(5), p. 3025. doi:10.4049/jimmunol.175.5.3025.

McManus, M.J., Picard, M., Chen, H.-W., De Haas, H.J., Potluri, P., Leipzig, J., Towheed, A., Angelin, A., Sengupta, P., Morrow, R.M., Kauffman, B.A., Vermulst, M., Narula, J. and Wallace, D.C. (2019) 'Mitochondrial DNA Variation Dictates Expressivity and Progression of Nuclear DNA Mutations Causing Cardiomyopathy', *Cell Metabolism*, 29(1), pp. 78-90.e5. doi:10.1016/j.cmet.2018.08.002.

Meador, W. (2020) 'Neuroimmunology & Neuroinflammatory Disorders', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174246890 (Accessed: 3 January 2022).

Meinl, E. and Hohlfeld, R. (2021) 'CD20+ T Cells as Pathogenic Players and Therapeutic Targets in MS', *Annals of Neurology*, 90(5), pp. 722–724. doi:10.1002/ana.26232.

de Mello, A.H., Costa, A.B., Engel, J.D.G. and Rezin, G.T. (2018) 'Mitochondrial dysfunction in obesity', *Life Sciences*, 192, pp. 26–32. doi:10.1016/j.lfs.2017.11.019.

Mescher, A.L. (2018) 'The Immune System & Lymphoid Organs', in *Junqueira's Basic Histology: Text and Atlas, 15e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1160662533 (Accessed: 24 December 2021).

Mescher, A.L. (2021a) 'Hemopoiesis', in *Junqueira's Basic Histology Text and Atlas, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1180413252 (Accessed: 24 December 2021).

Mescher, A.L. (2021b) 'Nerve Tissue & the Nervous System', in *Junqueira's Basic Histology Text and Atlas, 16e*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1180412769 (Accessed: 16 December 2021).

Metzker, M.L. (2010) 'Sequencing technologies - the next generation', *Nature Reviews. Genetics*, 11(1), pp. 31–46. doi:10.1038/nrg2626.

Mezey, É. and Palkovits, M. (2015) 'Forgotten findings of brain lymphatics', *Nature*, 524(7566), pp. 415–415. doi:10.1038/524415b.

Michalek, R.D., Gerriets, V.A., Jacobs, S.R., Macintyre, A.N., MacIver, N.J., Mason, E.F., Sullivan, S.A., Nichols, A.G. and Rathmell, J.C. (2011) 'Cutting Edge: Distinct Glycolytic and Lipid Oxidative Metabolic Programs Are Essential for Effector and Regulatory CD4 $^{+}$ T Cell Subsets', *The Journal of Immunology*, 186(6), pp. 3299–3303. doi:10.4049/jimmunol.1003613.

Michikawa Yuichi, Mazzucchelli Franca, Bresolin Nereo, Scarlato Guglielmo, and Attardi Giuseppe (1999) 'Aging-Dependent Large Accumulation of Point Mutations in the Human mtDNA Control Region for Replication', *Science*, 286(5440), pp. 774–779. doi:10.1126/science.286.5440.774.

Mihailova, S.M., Ivanova, M.I., Quin, L.M. and Naumova, E.J. (2007) 'Mitochondrial DNA variants in Bulgarian patients affected by multiple sclerosis', *European Journal of Neurology*, 14(1), pp. 44–47. doi:10.1111/j.1468-1331.2006.01541.x.

Miller, H. (1968) 'Multiple Sclerosis: Epidemiological Aspects of Multiple Sclerosis', *Proceedings of the Royal Society of Medicine*, 61(9), pp. 937–940. doi:10.1177/003591576806100934.

Miller, K. (2016) 'MojoSort™ Human CD4 Nanobeads No Wash Protocol v1'. doi:10.17504/protocols.io.e3ibgke.

Miller, S.D., Vanderlugt, C.L., Begolka, W.S., Pao, W., Yauch, R.L., Neville, K.L., Katz-Levy, Y., Carrizosa, A. and Kim, B.S. (1997) 'Persistent infection with Theiler's virus leads to CNS autoimmunity via epitope spreading', *Nature Medicine*, 3(10), pp. 1133–1136. doi:10.1038/nm1097-1133.

Mills, E.A., Ogrodnik, M.A., Plave, A. and Mao-Draayer, Y. (2018) 'Emerging Understanding of the Mechanism of Action for Dimethyl Fumarate in the Treatment of Multiple Sclerosis', *Frontiers in Neurology*, 9, p. 5. doi:10.3389/fneur.2018.00005.

Ming, T., Wang, M., Zheng, M., Zhou, Y., Hou, Y. and Wang, Z. (2019) 'Exploring of rare differences in mtGenomes between MZ twins using massively parallel sequencing', *Forensic Science International: Genetics Supplement Series*, 7(1), pp. 70–72. doi:10.1016/j.fsigss.2019.09.028.

Minussi, D.C., Nicholson, M.D., Ye, H., Davis, A., Wang, K., Baker, T., Tarabichi, M., Sei, E., Du, H., Rabbani, M., Peng, C., Hu, M., Bai, S., Lin, Y., Schalck, A., Multani, A., Ma, J., McDonald, T.O., Casasent, A., Barrera, A., Chen, H., Lim, B., Arun, B., Meric-Bernstam, F., Van Loo, P., Michor, F. and Navin, N.E. (2021) 'Breast tumours maintain a reservoir of subclonal diversity during expansion', *Nature*, 592(7853), pp. 302–308. doi:10.1038/s41586-021-03357-x.

Mitchell, R.E., Bates, K., Wootton, R.E., Harroud, A., Richards, J.B., Davey Smith, G. and Munafò, M.R. (2020) 'Little evidence for an effect of smoking on multiple sclerosis risk: A Mendelian Randomization study', *PLOS Biology*, 18(11), p. e3000973. doi:10.1371/journal.pbio.3000973.

Moccia, M., Haider, L., Eshaghi, A., van de Pavert, S.H.P., Brescia Morra, V., Patel, A., Wheeler-Kingshott, C.A.M., Barkhof, F. and Ciccarelli, O. (2022) 'B Cells in the CNS at Postmortem Are Associated With Worse Outcome and Cell Types in Multiple Sclerosis', *Neurology - Neuroimmunology Neuroinflammation*, 9(1), p. e1108. doi:10.1212/NXI.0000000000001108.

Mol, C.L., Looman, K.I.M., Luijn, M.M., Kreft, K.L., Jansen, P.R., Zelm, M.C., Smolders, J.J.F.M., White, T.J.H., Moll, H.A. and Neuteboom, R.F. (2021) 'T cell composition and polygenic multiple sclerosis risk: A population-based study in children', *European Journal of Neurology*, 28(11), pp. 3731–3741. doi:10.1111/ene.15019.

Montalban, X., Gold, R., Thompson, A.J., Otero-Romero, S., Amato, M.P., Chandraratna, D., Clanet, M., Comi, G., Derfuss, T., Fazekas, F., Hartung, H.P., Havrdova, E., Hemmer, B., Kappos, L., Liblau, R., Lubetzki, C., Marcus, E., Miller, D.H., Olsson, T., Pilling, S., Selmaj, K., Siva, A., Sorensen, P.S., Sormani, M.P., Thalheim, C., Wiendl, H. and Zipp, F. (2018) 'ECTRIMS/EAN Guideline on the pharmacological treatment of people with multiple sclerosis', *Multiple Sclerosis Journal*, 24(2), pp. 96–120. doi:10.1177/1352458517751049.

Moore, L., Cagan, A., Coorens, T.H.H., Neville, M.D.C., Sanghvi, R., Sanders, M.A., Oliver, T.R.W., Leongamornlert, D., Ellis, P., Noorani, A., Mitchell, T.J., Butler, T.M., Hooks, Y., Warren, A.Y., Jorgensen, M., Dawson, K.J., Menzies, A., O'Neill, L., Latimer, C., Teng, M., van Boxtel, R., Iacobuzio-Donahue, C.A., Martincorena, I., Heer, R., Campbell, P.J., Fitzgerald, R.C., Stratton, M.R. and Rahbari, R. (2021) 'The mutational landscape of human somatic and germline cells', *Nature*, 597(7876), pp. 381–386. doi:10.1038/s41586-021-03822-7.

Morais, V.A. and De Strooper, B. (2010) 'Mitochondria dysfunction and neurodegenerative disorders: cause or consequence', *Journal of Alzheimer's disease: JAD*, 20 Suppl 2, pp. S255-263. doi:10.3233/JAD-2010-100345.

Mori, L., Lepore, M. and De Libero, G. (2016) 'The Immunology of CD1- and MR1-Restricted T Cells', *Annual Review of Immunology*, 34(1), pp. 479–510. doi:10.1146/annurev-immunol-032414-112008.

Morimoto, K. and Nakajima, K. (2019) 'Role of the Immune System in the Development of the Central Nervous System', *Frontiers in Neuroscience*, 13, p. 916. doi:10.3389/fnins.2019.00916.

Morton, D.A., Foreman, K.B. and Albertine, K.H. (2019a) 'Brain', in *The Big Picture: Gross Anatomy, 2e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1158277411 (Accessed: 16 December 2021).

Morton, D.A., Foreman, K.B. and Albertine, K.H. (2019b) 'Scalp, Skull, and Meninges', in *The Big Picture: Gross Anatomy, 2e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1158277369 (Accessed: 17 December 2021).

Mukherjee, C., Kling, T., Russo, B., Miebach, K., Kess, E., Schifferer, M., Pedro, L.D., Weikert, U., Fard, M.K., Kannaiyan, N., Rossner, M., Aicher, M.-L., Goebbels, S., Nave, K.-A., Krämer-Albers, E.-M., Schneider, A. and Simons, M. (2020) 'Oligodendrocytes Provide Antioxidant Defense Function for Neurons by Secreting Ferritin Heavy Chain', *Cell Metabolism*, 32(2), pp. 259-272.e10. doi:10.1016/j.cmet.2020.05.019.

Multiple Sclerosis International Federation (2020) 'Atlas of MS – 3rd Edition. Part 1: Mapping multiple sclerosis around the world. Key epidemiology findings'. Available at: https://www.atlasofms.org/ (Accessed: 4 January 2022).

Multiple Sclerosis International Federation (2021) 'Atlas of MS – 3rd Edition. Part 2: Clinical management of multiple sclerosis around the world. Key findings about the diagnosis and

clinical management of MS'. Available at: https://www.atlasofms.org/ (Accessed: 4 January 2022).

Muntjewerff, E.M., Meesters, L.D. and van den Bogaart, G. (2020) 'Antigen Cross-Presentation by Macrophages', *Frontiers in Immunology*, 11, p. 1276. doi:10.3389/fimmu.2020.01276.

NAITO, S., NAMEROW, N., MICKEY, M.R. and TERASAKI, P.I. (1972) 'Multiple Sclerosis: Association with HL—A3', *Tissue Antigens*, 2(1), pp. 1–4. doi:10.1111/j.1399-0039.1972.tb00111.x.

NASS, M.M. and NASS, S. (1963) 'INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS. I. FIXATION AND ELECTRON STAINING REACTIONS', *The Journal of cell biology*, 19(3), pp. 593–611. doi:10.1083/jcb.19.3.593.

NASS, S. and NASS, M.M. (1963) 'INTRAMITOCHONDRIAL FIBERS WITH DNA CHARACTERISTICS. II. ENZYMATIC AND OTHER HYDROLYTIC TREATMENTS', *The Journal of cell biology*, 19(3), pp. 613–629. doi:10.1083/jcb.19.3.613.

Naue, J., Hörer, S., Sänger, T., Strobl, C., Hatzer-Grubwieser, P., Parson, W. and Lutz-Bonengel, S. (2015) 'Evidence for frequent and tissue-specific sequence heteroplasmy in human mitochondrial DNA', *Mitochondrion*, 20, pp. 82–94. doi:10.1016/j.mito.2014.12.002.

Navarro-Gomez, D., Leipzig, J., Shen, L., Lott, M., Stassen, A.P.M., Wallace, D.C., Wiggs, J.L., Falk, M.J., van Oven, M. and Gai, X. (2015) 'Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier', *Bioinformatics (Oxford, England)*, 31(8), pp. 1310–1312. doi:10.1093/bioinformatics/btu825.

Nekhaeva, E., Bodyak, N.D., Kraytsberg, Y., McGrath, S.B., Van Orsouw, N.J., Pluzhnikov, A., Wei, J.Y., Vijg, J. and Khrapko, K. (2002) 'Clonally expanded mtDNA point mutations are abundant in individual cells of human tissues', *Proceedings of the National Academy of Sciences*, 99(8), pp. 5521–5526. doi:10.1073/pnas.072670199.

Ng, X., Sadeghian, M., Heales, S. and Hargreaves, I.P. (2019) 'Assessment of Mitochondrial Dysfunction in Experimental Autoimmune Encephalomyelitis (EAE) Models of Multiple Sclerosis', *International Journal of Molecular Sciences*, 20(20), p. 4975. doi:10.3390/ijms20204975.

Ng, Y.S., Bindoff, L.A., Gorman, G.S., Klopstock, T., Kornblum, C., Mancuso, M., McFarland, R., Sue, C.M., Suomalainen, A., Taylor, R.W., Thorburn, D.R. and Turnbull, D.M. (2021) 'Mitochondrial disease in adults: recent advances and future promise', *The Lancet. Neurology*, 20(7), pp. 573–584. doi:10.1016/S1474-4422(21)00098-3.

Nicklas, J.A. and Buel, E. (2003) 'Quantification of DNA in forensic samples', *Analytical and Bioanalytical Chemistry*, 376(8), pp. 1160–1167. doi:10.1007/s00216-003-1924-z.

Nido, G.S., Dölle, C., Flønes, I., Tuppen, H.A., Alves, G., Tysnes, O.-B., Haugarvoll, K. and Tzoulis, C. (2018) 'Ultradeep mapping of neuronal mitochondrial deletions in Parkinson's

disease', *Neurobiology of Aging*, 63, pp. 120–127. doi:10.1016/j.neurobiolaging.2017.10.024.

Novotna, M., Paz Soldán, M.M., Abou Zeid, N., Kale, N., Tutuncu, M., Crusan, D.J., Atkinson, E.J., Siva, A., Keegan, B.M., Pirko, I., Pittock, S.J., Lucchinetti, C.F., Noseworthy, J.H., Weinshenker, B.G., Rodriguez, M. and Kantarci, O.H. (2015) 'Poor early relapse recovery affects onset of progressive disease course in multiple sclerosis', *Neurology*, 85(8), p. 722. doi:10.1212/WNL.0000000000001856.

Oberwahrenbrock, T., Ringelstein, M., Jentschke, S., Deuschle, K., Klumbies, K., Bellmann-Strobl, J., Harmel, J., Ruprecht, K., Schippling, S., Hartung, H.-P., Aktas, O., Brandt, A.U. and Paul, F. (2013) 'Retinal ganglion cell and inner plexiform layer thinning in clinically isolated syndrome', *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 19(14), pp. 1887–1895. doi:10.1177/1352458513489757.

Oechtering, J., Schaedelin, S., Benkert, P., Müller, S., Achtnichts, L., Vehoff, J., Disanto, G., Findling, O., Fischer-Barnicol, B., Orleth, A., Chan, A., Pot, C., Barakovic, M., Rahmanzadeh, R., Galbusera, R., Heijnen, I., Lalive, P.H., Wuerfel, J., Subramaniam, S., Aeschbacher, S., Conen, D., Naegelin, Y., Maceski, A., Meier, S., Berger, K., Wiendl, H., Lincke, T., Lieb, J., Yaldizli, Ö., Sinnecker, T., Derfuss, T., Regeniter, A., Zecca, C., Gobbi, C., Kappos, L., Granziera, C., Leppert, D., Kuhle, J., and and for the Swiss Multiple Sclerosis Cohort Study (2021) 'Intrathecal Immunoglobulin M Synthesis is an Independent Biomarker for Higher Disease Activity and Severity in Multiple Sclerosis', *Annals of Neurology*, 90(3), pp. 477–489. doi:10.1002/ana.26137.

O'Gorman, C., Lin, R., Stankovich, J. and Broadley, S.A. (2013) 'Modelling Genetic Susceptibility to Multiple Sclerosis with Family Data', *Neuroepidemiology*, 40(1), pp. 1–12. doi:10.1159/000341902.

Ohle, L.-M., Ellenberger, D., Flachenecker, P., Friede, T., Haas, J., Hellwig, K., Parciak, T., Warnke, C., Paul, F., Zettl, U.K. and Stahmann, A. (2021) 'Chances and challenges of a long-term data repository in multiple sclerosis: 20th birthday of the German MS registry', *Scientific Reports*, 11(1), p. 13340. doi:10.1038/s41598-021-92722-x.

Ohlenbusch, A., Wilichowski, E. and Hanefeld, F. (1998a) 'Characterization of the mitochondrial genome in childhood multiple sclerosis. I. Optic neuritis and LHON mutations', *Neuropediatrics*, 29(4), pp. 175–179. doi:10.1055/s-2007-973557.

Ohlenbusch, A., Wilichowski, E. and Hanefeld, F. (1998b) 'Characterization of the mitochondrial genome in childhood multiple sclerosis. III. Multiple sclerosis without optic neuritis and the non-LHON-associated genes', *Neuropediatrics*, 29(6), pp. 313–319. doi:10.1055/s-2007-973584.

Okada, A.K., Teranishi, K., Lobo, F., Isas, J.M., Xiao, J., Yen, K., Cohen, P. and Langen, R. (2017) 'The Mitochondrial-Derived Peptides, HumaninS14G and Small Humanin-like Peptide 2, Exhibit Chaperone-like Activity', *Scientific Reports*, 7(1), p. 7802. doi:10.1038/s41598-017-08372-5.

O'Neill, L.A.J., Kishton, R.J. and Rathmell, J. (2016) 'A guide to immunometabolism for immunologists', *Nature Reviews Immunology*, 16(9), pp. 553–565. doi:10.1038/nri.2016.70.

Ono, H., Arai, Y., Furukawa, E., Narushima, D., Matsuura, T., Nakamura, H., Shiokawa, D., Nagai, M., Imai, T., Mimori, K., Okamoto, K., Hippo, Y., Shibata, T. and Kato, M. (2021) 'Single-cell DNA and RNA sequencing reveals the dynamics of intra-tumor heterogeneity in a colorectal cancer model', *BMC Biology*, 19(1), p. 207. doi:10.1186/s12915-021-01147-5.

Ooms, J. (2021) *magick: Advanced Graphics and Image-Processing in R*. Available at: https://CRAN.R-project.org/package=magick (Accessed: 2 February 2022).

O'Reilly, M.L. and Tom, V.J. (2020) 'Neuroimmune System as a Driving Force for Plasticity Following CNS Injury', *Frontiers in Cellular Neuroscience*, 14, p. 187. doi:10.3389/fncel.2020.00187.

Ortiz, M., Poloni, E.S., Furrer, H., Kovari, H., Martinez, R., Arnedo, M., Elzi, L., Bernasconi, E., Vernazza, P., Hirschel, B., Cavassini, M., Ledergerber, B., Günthard, H.F., Telenti, A., Tarr, P.E., and the Swiss HIV Cohort Study (2011) 'No Longitudinal Mitochondrial DNA Sequence Changes in HIV-infected Individuals With and Without Lipoatrophy', *The Journal of Infectious Diseases*, 203(5), pp. 620–624. doi:10.1093/infdis/jiq106.

Ostkamp, P., Salmen, A., Pignolet, B., Görlich, D., Andlauer, T.F.M., Schulte-Mecklenbeck, A., Gonzalez-Escamilla, G., Bucciarelli, F., Gennero, I., Breuer, J., Antony, G., Schneider-Hohendorf, T., Mykicki, N., Bayas, A., Then Bergh, F., Bittner, S., Hartung, H.-P., Friese, M.A., Linker, R.A., Luessi, F., Lehmann-Horn, K., Mühlau, M., Paul, F., Stangel, M., Tackenberg, B., Tumani, H., Warnke, C., Weber, F., Wildemann, B., Zettl, U.K., Ziemann, U., Müller-Myhsok, B., Kümpfel, T., Klotz, L., Meuth, S.G., Zipp, F., Hemmer, B., Hohlfeld, R., Brassat, D., Gold, R., Gross, C.C., Lukas, C., Groppa, S., Loser, K., Wiendl, H., Schwab, N., and on behalf of the German Competence Network Multiple Sclerosis (KKNMS) and the BIONAT Network (2021) 'Sunlight exposure exerts immunomodulatory effects to reduce multiple sclerosis severity', *Proceedings of the National Academy of Sciences*, 118(1), p. e2018457118. doi:10.1073/pnas.2018457118.

Otaegui, D., Sáenz, A., Martínez-Zabaleta, M., Villoslada, P., Fernández-Manchola, I., de Arcaya, A. álvarez, Emparanza, J.I. and de Munain, A.L. (2004) 'Mitochondrial haplogroups in Basque multiple sclerosis patients', *Multiple Sclerosis Journal*, 10(5), pp. 532–535. doi:10.1191/1352458504ms1069oa.

van Oven, M. (2015) 'PhyloTree Build 17: Growing the human mitochondrial DNA tree', *Forensic Science International: Genetics Supplement Series*, 5, pp. e392–e394. doi:10.1016/j.fsigss.2015.09.155.

Pääbo, S. (2014) 'The Human Condition—A Molecular Approach', *Cell*, 157(1), pp. 216–226. doi:10.1016/j.cell.2013.12.036.

Palanichamy, A., Jahn, S., Nickles, D., Derstine, M., Abounasr, A., Hauser, S.L., Baranzini, S.E., Leppert, D. and von Büdingen, H.-C. (2014) 'Rituximab Efficiently Depletes Increased

CD20-Expressing T Cells in Multiple Sclerosis Patients', *The Journal of Immunology*, 193(2), p. 580. doi:10.4049/jimmunol.1400118.

Pan, G., Simpson, S., van der Mei, I., Charlesworth, J.C., Lucas, R., Ponsonby, A.-L., Zhou, Y., Wu, F. and Taylor, B.V. (2016) 'Role of genetic susceptibility variants in predicting clinical course in multiple sclerosis: a cohort study', *Journal of Neurology, Neurosurgery & Psychiatry*, 87(11), pp. 1204–1211. doi:10.1136/jnnp-2016-313722.

Pantic, B., Ives, D., Mennuni, M., Perez-Rodriguez, D., Fernandez-Pelayo, U., Lopez de Arbina, A., Muñoz-Oreja, M., Villar-Fernandez, M., Dang, T.J., Vergani, L., Johnston, I.G., Pitceathly, R.D.S., McFarland, R., Hanna, M.G., Taylor, R.W., Holt, I.J. and Spinazzola, A. (2021) '2-Deoxy-D-glucose couples mitochondrial DNA replication with mitochondrial fitness and promotes the selection of wild-type over mutant mitochondrial DNA', *Nature Communications*, 12(1), p. 6997. doi:10.1038/s41467-021-26829-0.

Paolicelli Rosa C., Bolasco Giulia, Pagani Francesca, Maggi Laura, Scianni Maria, Panzanelli Patrizia, Giustetto Maurizio, Ferreira Tiago Alves, Guiducci Eva, Dumas Laura, Ragozzino Davide, and Gross Cornelius T. (2011) 'Synaptic Pruning by Microglia Is Necessary for Normal Brain Development', *Science*, 333(6048), pp. 1456–1458. doi:10.1126/science.1202529.

Parker Harp, C.R., Archambault, A.S., Sim, J., Shlomchik, M.J., Russell, J.H. and Wu, G.F. (2018) 'B cells are capable of independently eliciting rapid reactivation of encephalitogenic CD4 T cells in a murine model of multiple sclerosis', *PLOS ONE*, 13(6), p. e0199694. doi:10.1371/journal.pone.0199694.

Parker, N., Schneegurt, M., Tu, A.-H.T., Lister, P. and Forster, B.M. (2016) 'Cellular Defenses', in *Microbiology*. Houston, Texas: OpenStax. Available at: https://openstax.org/books/microbiology/pages/17-3-cellular-defenses (Accessed: 24 December 2021).

Parson, W., Gusmão, L., Hares, D.R., Irwin, J.A., Mayr, W.R., Morling, N., Pokorak, E., Prinz, M., Salas, A., Schneider, P.M., Parsons, T.J., and DNA Commission of the International Society for Forensic Genetics (2014) 'DNA Commission of the International Society for Forensic Genetics: revised and extended guidelines for mitochondrial DNA typing', *Forensic Science International. Genetics*, 13, pp. 134–142. doi:10.1016/j.fsigen.2014.07.010.

Patergnani, S., Castellazzi, M., Bonora, M., Marchi, S., Casetta, I., Pugliatti, M., Giorgi, C., Granieri, E. and Pinton, P. (2018) 'Autophagy and mitophagy elements are increased in body fluids of multiple sclerosis-affected individuals', *Journal of Neurology, Neurosurgery & Psychiatry*, 89(4), p. 439. doi:10.1136/jnnp-2017-316234.

Patsopoulos, N.A. (2018) 'Genetics of Multiple Sclerosis: An Overview and New Directions', *Cold Spring Harbor Perspectives in Medicine*, 8(7), p. a028951. doi:10.1101/cshperspect.a028951.

Patsopoulos, N.A., Barcellos, L.F., Hintzen, R.Q., Schaefer, C., van Duijn, C.M., Noble, J.A., Raj, T., IMSGC, ANZgene, Gourraud, P.-A., Stranger, B.E., Oksenberg, J., Olsson, T., Taylor, B.V., Sawcer, S., Hafler, D.A., Carrington, M., De Jager, P.L. and de Bakker, P.I.W. (2013)

'Fine-Mapping the Genetic Association of the Major Histocompatibility Complex in Multiple Sclerosis: HLA and Non-HLA Effects', *PLOS Genetics*, 9(11), p. e1003926. doi:10.1371/journal.pgen.1003926.

Patti, F. and Vila, C. (2014) 'Symptoms, Prevalence and Impact of Multiple Sclerosis in Younger Patients: A Multinational Survey', *Neuroepidemiology*, 42(4), pp. 211–218. doi:10.1159/000360423.

Pavlov, V.A. and Tracey, K.J. (2017) 'Neural regulation of immunity: molecular mechanisms and clinical translation', *Nature Neuroscience*, 20(2), pp. 156–166. doi:10.1038/nn.4477.

Payne, B.A.I., Wilson, I.J., Yu-Wai-Man, P., Coxhead, J., Deehan, D., Horvath, R., Taylor, R.W., Samuels, D.C., Santibanez-Koref, M. and Chinnery, P.F. (2013) 'Universal heteroplasmy of human mitochondrial DNA', *Human Molecular Genetics*, 22(2), pp. 384–390. doi:10.1093/hmg/dds435.

Peakman, M. (2012) 'The immune system and disease', in Kumar, P.J. and Clark, M. (eds) *Kumar & Clark's Clinical Medicine*. 8. ed. Edinburgh: Saunders, Elsevier, pp. 49–72.

Pedersen, T.L. (2020) *patchwork: The Composer of Plots*. Available at: https://CRAN.R-project.org/package=patchwork (Accessed: 2 February 2022).

Penner, I.-K. and Paul, F. (2017) 'Fatigue as a symptom or comorbidity of neurological diseases', *Nature Reviews Neurology*, 13(11), pp. 662–675. doi:10.1038/nrneurol.2017.117.

Penter, L., Gohil, S.H., Lareau, C., Ludwig, L.S., Parry, E.M., Huang, T., Li, S., Zhang, W., Livitz, D., Leshchiner, I., Parida, L., Getz, G., Rassenti, L.Z., Kipps, T.J., Brown, J.R., Davids, M.S., Neuberg, D.S., Livak, K.J., Sankaran, V.G. and Wu, C.J. (2021) 'Longitudinal Single-Cell Dynamics of Chromatin Accessibility and Mitochondrial Mutations in Chronic Lymphocytic Leukemia Mirror Disease History', *Cancer Discovery* [Preprint]. doi:10.1158/2159-8290.CD-21-0276.

Pereira, C.S., Pérez-Cabezas, B., Ribeiro, H., Maia, M.L., Cardoso, M.T., Dias, A.F., Azevedo, O., Ferreira, M.F., Garcia, P., Rodrigues, E., Castro-Chaves, P., Martins, E., Aguiar, P., Pineda, M., Amraoui, Y., Fecarotta, S., Leão-Teles, E., Deng, S., Savage, P.B. and Macedo, M.F. (2019) 'Lipid Antigen Presentation by CD1b and CD1d in Lysosomal Storage Disease Patients', *Frontiers in Immunology*, 10, p. 1264. doi:10.3389/fimmu.2019.01264.

Pereira, L., Soares, P., Radivojac, P., Li, B. and Samuels, D.C. (2011) 'Comparing phylogeny and the predicted pathogenicity of protein variations reveals equal purifying selection across the global human mtDNA diversity', *American Journal of Human Genetics*, 88(4), pp. 433–439. doi:10.1016/j.ajhg.2011.03.006.

Pereira, V., Longobardi, A. and Børsting, C. (2018) 'Sequencing of mitochondrial genomes using the Precision ID mtDNA Whole Genome Panel', *Electrophoresis*, 39(21), pp. 2766–2775. doi:10.1002/elps.201800088.

Peruzzotti-Jametti, L., Bernstock, J.D., Willis, C.M., Manferrari, G., Rogall, R., Fernandez-Vizarra, E., Williamson, J.C., Braga, A., van den Bosch, A., Leonardi, T., Krzak, G., Kittel, Á., Benincá, C., Vicario, N., Tan, S., Bastos, C., Bicci, I., Iraci, N., Smith, J.A., Peacock, B., Muller, K.H., Lehner, P.J., Buzas, E.I., Faria, N., Zeviani, M., Frezza, C., Brisson, A., Matheson, N.J., Viscomi, C. and Pluchino, S. (2021) 'Neural stem cells traffic functional mitochondria via extracellular vesicles', *PLOS Biology*. Edited by K.G. Storey, 19(4), p. e3001166. doi:10.1371/journal.pbio.3001166.

Peter Brescia (2021) 'Application Note: Microvolume Purity Assessment of Nucleic Acids Using A260/A280 Ratio and Spectral Scanning.' Available at: https://www.agilent.com/cs/library/applications/A260A280-spectral-scanning-5994-2538EN-agilent.pdf (Accessed: 16 January 2022).

Petersen, E.R., Oturai, A.B., Koch-Henriksen, N., Magyari, M., Sørensen, P.S., Sellebjerg, F. and Søndergaard, H.B. (2018) 'Smoking affects the interferon beta treatment response in multiple sclerosis', *Neurology*, 90(7), p. e593. doi:10.1212/WNL.0000000000004949.

Peverelli, L., Gold, C.A., Naini, A.B., Tanji, K., Akman, H.O., Hirano, M. and Dimauro, S. (2014) 'Mitochondrial myopathy with dystrophic features due to a novel mutation in the *MTTM* gene: Mitochondrial Dystrophy', *Muscle & Nerve*, 50(2), pp. 292–295. doi:10.1002/mus.24262.

Pfeffer, G., Burke, A., Yu-Wai-Man, P., Compston, D.A.S. and Chinnery, P.F. (2013) 'Clinical features of MS associated with Leber hereditary optic neuropathy mtDNA mutations', *Neurology*, 81(24), pp. 2073–2081. doi:10.1212/01.wnl.0000437308.22603.43.

Pienaar, I.S., Mohammed, R., Courtley, R., Gledson, M.R., Reynolds, R., Nicholas, R. and Elson, J.L. (2021) 'Investigation of the correlation between mildly deleterious mtDNA Variations and the clinical progression of multiple sclerosis', *Multiple Sclerosis and Related Disorders*, 53, p. 103055. doi:10.1016/j.msard.2021.103055.

Pierson, E.R., Stromnes, I.M. and Goverman, J.M. (2014) 'B Cells Promote Induction of Experimental Autoimmune Encephalomyelitis by Facilitating Reactivation of T Cells in the Central Nervous System', *The Journal of Immunology*, 192(3), p. 929. doi:10.4049/jimmunol.1302171.

Piotrowska-Nowak, A., Elson, J.L., Sobczyk-Kopciol, A., Piwonska, A., Puch-Walczak, A., Drygas, W., Ploski, R., Bartnik, E. and Tonska, K. (2018) 'New mtDNA Association Model, MutPred Variant Load, Suggests Individuals With Multiple Mildly Deleterious mtDNA Variants Are More Likely to Suffer From Atherosclerosis', *Frontiers in Genetics*, 9, p. 702. doi:10.3389/fgene.2018.00702.

Poole, O.V., Pizzamiglio, C., Murphy, D., Falabella, M., Macken, W.L., Bugiardini, E., Woodward, C.E., Labrum, R., Efthymiou, S., Salpietro, V., Chelban, V., Kaiyrzhanov, R., Maroofian, R., SYNaPS Study Group, Amato, A.A., Gregory, A., Hayflick, S.J., Queen Square Genomics, Jonvik, H., Wood, N., Houlden, H., Vandrovcova, J., Hanna, M.G., Pittman, A. and Pitceathly, R.D.S. (2021) 'Mitochondrial DNA Analysis from Exome Sequencing Data Improves Diagnostic Yield in Neurological Diseases', *Annals of Neurology*, 89(6), pp. 1240–1247. doi:10.1002/ana.26063.

Poorolajal, J., Bahrami, M., Karami, M. and Hooshmand, E. (2017) 'Effect of smoking on multiple sclerosis: a meta-analysis', *Journal of Public Health*, 39(2), pp. 312–320. doi:10.1093/pubmed/fdw030.

Poursadegh Zonouzi, A., Ghorbian, S., Abkar, M., Poursadegh Zonouzi, A.A. and Azadi, A. (2014) 'Mitochondrial complex I gene variations; as a potential genetic risk factor in pathogenesis of multiple sclerosis', *Journal of the Neurological Sciences*, 345(1–2), pp. 220–223. doi:10.1016/j.jns.2014.07.051.

Pozzo-Miller, L. and Theibert, A.B. (2020) 'Development of the Nervous System', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242196 (Accessed: 16 December 2021).

Preste, R., Vitale, O., Clima, R., Gasparre, G. and Attimonelli, M. (2019) 'HmtVar: a new resource for human mitochondrial variations and pathogenicity data', *Nucleic Acids Research*, 47(D1), pp. D1202–D1210. doi:10.1093/nar/gky1024.

Prinz, M. and Priller, J. (2017) 'The role of peripheral immune cells in the CNS in steady state and disease', *Nature Neuroscience*, 20(2), pp. 136–144. doi:10.1038/nn.4475.

Quail, M.A., Smith, M., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P. and Gu, Y. (2012) 'A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers', *BMC genomics*, 13, p. 341. doi:10.1186/1471-2164-13-341.

Quendt, C., Ochs, J., Häusser-Kinzel, S., Häusler, D. and Weber, M.S. (2021) 'Proinflammatory CD20+ T Cells are Differentially Affected by Multiple Sclerosis Therapeutics', *Annals of Neurology*, 90(5), pp. 834–839. doi:10.1002/ana.26216.

Quinlan, A.R. and Hall, I.M. (2010) 'BEDTools: a flexible suite of utilities for comparing genomic features', *Bioinformatics*, 26(6), pp. 841–842. doi:10.1093/bioinformatics/btq033.

Quirós, P.M., Español, Y., Acín-Pérez, R., Rodríguez, F., Bárcena, C., Watanabe, K., Calvo, E., Loureiro, M., Fernández-García, M.S., Fueyo, A., Vázquez, J., Enríquez, J.A. and López-Otín, C. (2014) 'ATP-Dependent Lon Protease Controls Tumor Bioenergetics by Reprogramming Mitochondrial Activity', *Cell Reports*, 8(2), pp. 542–556. doi:10.1016/j.celrep.2014.06.018.

R Core Team (2021) *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. Available at: https://www.R-project.org/ (Accessed: 2 February 2022).

Radjavi, A., Smirnov, I. and Kipnis, J. (2014) 'Brain antigen-reactive CD4+ T cells are sufficient to support learning behavior in mice with limited T cell repertoire', *Brain, Behavior, and Immunity*, 35, pp. 58–63. doi:10.1016/j.bbi.2013.08.013.

Ragnedda, G., Leoni, S., Parpinel, M., Casetta, I., Riise, T., Myhr, K.-M., Wolfson, C. and Pugliatti, M. (2015) 'Reduced duration of breastfeeding is associated with a higher risk of

multiple sclerosis in both Italian and Norwegian adult males: the EnvIMS study', *Journal of Neurology*, 262(5), pp. 1271–1277. doi:10.1007/s00415-015-7704-9.

Rausser, S., Trumpff, C., McGill, M.A., Junker, A., Wang, W., Ho, S., Mitchell, A., Karan, K.R., Monk, C., Segerstrom, S.C., Reed, R.G. and Picard, M. (2021) 'Mitochondrial phenotypes in purified human immune cell subtypes and cell mixtures', *bioRxiv*, p. 2020.10.16.342923. doi:10.1101/2020.10.16.342923.

Reich, D.S., Lucchinetti, C.F. and Calabresi, P.A. (2018) 'Multiple Sclerosis', *New England Journal of Medicine*, 378(2), pp. 169–180. doi:10.1056/NEJMra1401483.

Ribeiro Miguel, Brigas Helena C., Temido-Ferreira Mariana, Pousinha Paula A., Regen Tommy, Santa Cátia, Coelho Joana E., Marques-Morgado Inês, Valente Cláudia A., Omenetti Sara, Stockinger Brigitta, Waisman Ari, Manadas Bruno, Lopes Luísa V., Silva-Santos Bruno, and Ribot Julie C. (2019) 'Meningeal γδ T cell–derived IL-17 controls synaptic plasticity and short-term memory', *Science Immunology*, 4(40), p. eaay5199. doi:10.1126/sciimmunol.aay5199.

Ricca, C., Aillon, A., Bergandi, L., Alotto, D., Castagnoli, C. and Silvagno, F. (2018) 'Vitamin D Receptor Is Necessary for Mitochondrial Function and Cell Health', *International Journal of Molecular Sciences*, 19(6), p. 1672. doi:10.3390/ijms19061672.

Ring, J.D., Sturk-Andreaggi, K., Alyse Peck, M. and Marshall, C. (2018) 'Bioinformatic removal of NUMT-associated variants in mitotiling next-generation sequencing data from whole blood samples', *Electrophoresis*, 39(21), pp. 2785–2797. doi:10.1002/elps.201800135.

Ringh, M.V., Hagemann-Jensen, M., Needhamsen, M., Kullberg, S., Wahlström, J., Grunewald, J., Brynedal, B., Jagodic, M., Ekström, T.J., Öckinger, J. and Kular, L. (2021) 'Methylome and transcriptome signature of bronchoalveolar cells from multiple sclerosis patients in relation to smoking', *Multiple Sclerosis Journal*, 27(7), pp. 1014–1026. doi:10.1177/1352458520943768.

Ristori, G., Cannoni, S., Stazi, M.A., Vanacore, N., Cotichini, R., Alfò, M., Pugliatti, M., Sotgiu, S., Solaro, C., Bomprezzi, R., Di Giovanni, S., Figà Talamanca, L., Nisticò, L., Fagnani, C., Neale, M.C., Cascino, I., Giorgi, G., Battaglia, M.A., Buttinelli, C., Tosi, R., Salvetti, M., and Italian Study Group on Multiple Sclerosis in Twins (2006) 'Multiple sclerosis in twins from continental Italy and Sardinia: A nationwide study', *Annals of Neurology*, 59(1), pp. 27–34. doi:10.1002/ana.20683.

Rito, Y., Flores, J., Fernández Aguilar, Á., Escalante Membrillo, C., Gutiérrez Lanz, E., Barboza, M.A., Rivas Alonso, V., Treviño Frenk, I. and Corona Vázquez, T. (2016) 'Vitamin D in multiple sclerosis patients: Not the same risk for everybody', *Multiple Sclerosis Journal*, 22(1), pp. 126–127. doi:10.1177/1352458515576986.

Rius, R., Cowley, M.J., Riley, L., Puttick, C., Thorburn, D.R. and Christodoulou, J. (2019) 'Biparental inheritance of mitochondrial DNA in humans is not a common phenomenon', *Genetics in Medicine*, 21(12), pp. 2823–2826. doi:10.1038/s41436-019-0568-0.

Rivers, T.M. and Schwentker, F.F. (1935) 'ENCEPHALOMYELITIS ACCOMPANIED BY MYELIN DESTRUCTION EXPERIMENTALLY PRODUCED IN MONKEYS', *Journal of Experimental Medicine*, 61(5), pp. 689–702. doi:10.1084/jem.61.5.689.

Rivers, T.M., Sprunt, D.H. and Berry, G.P. (1933) 'OBSERVATIONS ON ATTEMPTS TO PRODUCE ACUTE DISSEMINATED ENCEPHALOMYELITIS IN MONKEYS', *Journal of Experimental Medicine*, 58(1), pp. 39–53. doi:10.1084/jem.58.1.39.

Rizzuto, R., De Stefani, D., Raffaello, A. and Mammucari, C. (2012) 'Mitochondria as sensors and regulators of calcium signalling', *Nature Reviews Molecular Cell Biology*, 13(9), pp. 566–578. doi:10.1038/nrm3412.

Robertson, N.P., Fraser, M., Deans, J., Clayton, D., Walker, N. and Compston, D.A.S. (1996) 'Age–adjusted recurrence risks for relatives of patients with multiple sclerosis', *Brain*, 119(2), pp. 449–455. doi:10.1093/brain/119.2.449.

Rodi, M., Dimisianos, N., De Lastic, A.-L., Sakellaraki, P., Deraos, G., Matsoukas, J., Papathanasopoulos, P. and Mouzaki, A. (2016) 'Regulatory Cell Populations in Relapsing-Remitting Multiple Sclerosis (RRMS) Patients: Effect of Disease Activity and Treatment Regimens', *International Journal of Molecular Sciences*, 17(9). doi:10.3390/ijms17091398.

Rolfes, L., Pfeuffer, S., Ruck, T., Melzer, N., Pawlitzki, M., Heming, M., Brand, M., Wiendl, H. and Meuth, S.G. (2019) 'Therapeutic Apheresis in Acute Relapsing Multiple Sclerosis: Current Evidence and Unmet Needs-A Systematic Review', *Journal of Clinical Medicine*, 8(10), p. E1623. doi:10.3390/jcm8101623.

Roos, S., Hedberg-Oldfors, C., Visuttijai, K., Stein, M., Kollberg, G., Elíasdóttir, Ó., Lindberg, C., Darin, N. and Oldfors, A. (2021) 'Expression pattern of mitochondrial respiratory chain enzymes in skeletal muscle of patients with mitochondrial myopathy associated with the homoplasmic m.14674T>C variant', *Brain Pathology* [Preprint]. doi:10.1111/bpa.13038.

Ropper, A.H., Samuels, M.A., Klein, J.P. and Prasad, S. (2019) 'Multiple Sclerosis and Other Inflammatory Demyelinating Diseases', in *Adams and Victor's Principles of Neurology, 11e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1162596445 (Accessed: 4 January 2022).

Rosso, M. and Chitnis, T. (2020) 'Association Between Cigarette Smoking and Multiple Sclerosis: A Review', *JAMA Neurology*, 77(2), pp. 245–253. doi:10.1001/jamaneurol.2019.4271.

Roth, C., Parson, W., Strobl, C., Lagacé, R. and Short, M. (2019) 'MVC: an integrated mitochondrial variant caller for forensics', *Australian Journal of Forensic Sciences*, 51(sup1), pp. S52–S55. doi:10.1080/00450618.2019.1569150.

Rotstein, D.L., Chen, H., Wilton, A.S., Kwong, J.C., Marrie, R.A., Gozdyra, P., Krysko, K.M., Kopp, A., Copes, R. and Tu, K. (2018) 'Temporal trends in multiple sclerosis prevalence and incidence in a large population', *Neurology*, 90(16), p. e1435. doi:10.1212/WNL.0000000000005331.

RStudio Team (2021) *RStudio: Integrated Development for R*. Boston, MA, USA: RStudio, PBC. Available at: https://www.rstudio.com/ (Accessed: 2 February 2022).

Rua, R. and McGavern, D.B. (2018) 'Advances in Meningeal Immunity', *Trends in Molecular Medicine*, 24(6), pp. 542–559. doi:10.1016/j.molmed.2018.04.003.

Rueda, M. and Torkamani, A. (2017) 'SG-ADVISER mtDNA: a web server for mitochondrial DNA annotation with data from 200 samples of a healthy aging cohort', *BMC bioinformatics*, 18(1), p. 373. doi:10.1186/s12859-017-1778-6.

Ryan, K.J. (2022) 'Herpesviruses', in *Sherris & Ryan's Medical Microbiology, 8e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1185629987 (Accessed: 10 January 2022).

Sá, M.J., Kobelt, G., Berg, J., Capsa, D., Dalén, J., and The European Multiple Sclerosis Platform (2017) 'New insights into the burden and costs of multiple sclerosis in Europe: Results for Portugal', *Multiple Sclerosis Journal*, 23(2_suppl), pp. 143–154. doi:10.1177/1352458517708667.

Sabatino, J.J., Pröbstel, A.-K. and Zamvil, S.S. (2019) 'B cells in autoimmune and neurodegenerative central nervous system diseases', *Nature Reviews Neuroscience*, 20(12), pp. 728–745. doi:10.1038/s41583-019-0233-2.

Sabatino, J.J., Wilson, M.R., Calabresi, P.A., Hauser, S.L., Schneck, J.P. and Zamvil, S.S. (2019) 'Anti-CD20 therapy depletes activated myelin-specific $CD8^+$ T cells in multiple sclerosis', *Proceedings of the National Academy of Sciences*, 116(51), p. 25800. doi:10.1073/pnas.1915309116.

Sadeghian, M., Mastrolia, V., Rezaei Haddad, A., Mosley, A., Mullali, G., Schiza, D., Sajic, M., Hargreaves, I., Heales, S., Duchen, M.R. and Smith, K.J. (2016) 'Mitochondrial dysfunction is an important cause of neurological deficits in an inflammatory model of multiple sclerosis', *Scientific Reports*, 6(1), p. 33249. doi:10.1038/srep33249.

Sagan, L. (1967) 'On the origin of mitosing cells', *Journal of Theoretical Biology*, 14(3), pp. 255–274. doi:10.1016/0022-5193(67)90079-3.

Sahel, J.-A., Boulanger-Scemama, E., Pagot, C., Arleo, A., Galluppi, F., Martel, J.N., Esposti, S.D., Delaux, A., de Saint Aubert, J.-B., de Montleau, C., Gutman, E., Audo, I., Duebel, J., Picaud, S., Dalkara, D., Blouin, L., Taiel, M. and Roska, B. (2021) 'Partial recovery of visual function in a blind patient after optogenetic therapy', *Nature Medicine*, 27(7), pp. 1223–1229. doi:10.1038/s41591-021-01351-4.

Sanchez-Contreras, M., Sweetwyne, M.T., Kohrn, B.F., Tsantilas, K.A., Hipp, M.J., Schmidt, E.K., Fredrickson, J., Whitson, J.A., Campbell, M.D., Rabinovitch, P.S., Marcinek, D.J. and Kennedy, S.R. (2021) 'A replication-linked mutational gradient drives somatic mutation accumulation and influences germline polymorphisms and genome composition in mitochondrial DNA', *Nucleic Acids Research*, 49(19), pp. 11103–11118. doi:10.1093/nar/gkab901.

Scalfari, A., Neuhaus, A., Daumer, M., Muraro, P.A. and Ebers, G.C. (2014) 'Onset of secondary progressive phase and long-term evolution of multiple sclerosis', *Journal of Neurology, Neurosurgery & Psychiatry*, 85(1), p. 67. doi:10.1136/jnnp-2012-304333.

Schafflick, D., Wolbert, J., Heming, M., Thomas, C., Hartlehnert, M., Börsch, A.-L., Ricci, A., Martín-Salamanca, S., Li, X., Lu, I.-N., Pawlak, M., Minnerup, J., Strecker, J.-K., Seidenbecher, T., Meuth, S.G., Hidalgo, A., Liesz, A., Wiendl, H. and Meyer zu Horste, G. (2021) 'Single-cell profiling of CNS border compartment leukocytes reveals that B cells and their progenitors reside in non-diseased meninges', *Nature Neuroscience*, 24(9), pp. 1225–1234. doi:10.1038/s41593-021-00880-y.

Schilf, P., Künstner, A., Olbrich, M., Waschina, S., Fuchs, B., Galuska, C.E., Braun, A., Neuschütz, K., Seutter, M., Bieber, K., Hellberg, L., Sina, C., Laskay, T., Rupp, J., Ludwig, R.J., Zillikens, D., Busch, H., Sadik, C.D., Hirose, M. and Ibrahim, S.M. (2021) 'A Mitochondrial Polymorphism Alters Immune Cell Metabolism and Protects Mice from Skin Inflammation', *International Journal of Molecular Sciences*, 22(3), p. 1006. doi:10.3390/ijms22031006.

Schläger, C., Körner, H., Krueger, M., Vidoli, S., Haberl, M., Mielke, D., Brylla, E., Issekutz, T., Cabañas, C., Nelson, P.J., Ziemssen, T., Rohde, V., Bechmann, I., Lodygin, D., Odoardi, F. and Flügel, A. (2016) 'Effector T-cell trafficking between the leptomeninges and the cerebrospinal fluid', *Nature*, 530(7590), pp. 349–353. doi:10.1038/nature16939.

Schneider Anya, Long Sarah Alice, Cerosaletti Karen, Ni Chester T., Samuels Peter, Kita Mariko, and Buckner Jane Hoyt (2013) 'In Active Relapsing-Remitting Multiple Sclerosis, Effector T Cell Resistance to Adaptive Tregs Involves IL-6–Mediated Signaling', *Science Translational Medicine*, 5(170), pp. 170ra15-170ra15. doi:10.1126/scitranslmed.3004970.

Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., Fulton, R.S., Kremitzki, M., Magrini, V., Markovic, C., McGrath, S., Steinberg, K.M., Auger, K., Chow, W., Collins, J., Harden, G., Hubbard, T., Pelan, S., Simpson, J.T., Threadgold, G., Torrance, J., Wood, J.M., Clarke, L., Koren, S., Boitano, M., Peluso, P., Li, H., Chin, C.-S., Phillippy, A.M., Durbin, R., Wilson, R.K., Flicek, P., Eichler, E.E. and Church, D.M. (2017) 'Evaluation of GRCh38 and de novo haploid genome assemblies demonstrates the enduring quality of the reference assembly', *Genome Research*, 27(5), pp. 849–864. doi:10.1101/gr.213611.116.

Schuh, E., Berer, K., Mulazzani, M., Feil, K., Meinl, I., Lahm, H., Krane, M., Lange, R., Pfannes, K., Subklewe, M., Gürkov, R., Bradl, M., Hohlfeld, R., Kümpfel, T., Meinl, E. and Krumbholz, M. (2016) 'Features of Human $CD3^{+}CD20^{+}$ T Cells', *The Journal of Immunology*, 197(4), p. 1111. doi:10.4049/jimmunol.1600089.

Schwalbe, G. (1869) 'Der Arachnoidalraum, ein Lymphraum und sein Zusammenhang mit dem Perichorioidalraum', *Zentralblatt für medizinische Wissenschaft*, (7), pp. 465–467.

Seki, S.M., Stevenson, M., Rosen, A.M., Arandjelovic, S., Gemta, L., Bullock, T.N.J. and Gaultier, A. (2017) 'Lineage-Specific Metabolic Properties and Vulnerabilities of T Cells in the Demyelinating Central Nervous System', *The Journal of Immunology*, 198(12), pp. 4607–4617. doi:10.4049/jimmunol.1600825.

Serafini, B., Rosicarelli, B., Magliozzi, R., Stigliano, E. and Aloisi, F. (2004) 'Detection of Ectopic B-cell Follicles with Germinal Centers in the Meninges of Patients with Secondary Progressive Multiple Sclerosis', *Brain Pathology*, 14(2), pp. 164–174. doi:10.1111/j.1750-3639.2004.tb00049.x.

Severa, M., Rizzo, F., Srinivasan, S., Di Dario, M., Giacomini, E., Buscarinu, M.C., Cruciani, M., Etna, M.P., Sandini, S., Mechelli, R., Farina, A., Trivedi, P., Hertzog, P.J., Salvetti, M., Farina, C. and Coccia, E.M. (2019) 'A cell type-specific transcriptomic approach to map B cell and monocyte type I interferon-linked pathogenic signatures in Multiple Sclerosis', *Journal of Autoimmunity*, 101, pp. 1–16. doi:10.1016/j.jaut.2019.04.006.

Severin, Y., Hale, B.D., Mena, J., Goslings, D., Frey, B.M. and Snijder, B. (2021) 'Multiplexed high-throughput immune cell imaging reveals molecular health-associated phenotypes', *bioRxiv*, p. 2021.12.03.471105. doi:10.1101/2021.12.03.471105.

Shen, Q., Lu, H., Xie, D., Wang, H., Zhao, Q. and Xu, Y. (2019) 'Association between suicide and multiple sclerosis: An updated meta-analysis', *Multiple Sclerosis and Related Disorders*, 34, pp. 83–90. doi:10.1016/j.msard.2019.06.012.

Shi, L.Z., Wang, R., Huang, G., Vogel, P., Neale, G., Green, D.R. and Chi, H. (2011) 'HIF1α–dependent glycolytic pathway orchestrates a metabolic checkpoint for the differentiation of TH17 and Treg cells', *Journal of Experimental Medicine*, 208(7), pp. 1367–1376. doi:10.1084/jem.20110278.

Shin, B., Benavides, G.A., Geng, J., Koralov, S.B., Hu, H., Darley-Usmar, V.M. and Harrington, L.E. (2020) 'Mitochondrial Oxidative Phosphorylation Regulates the Fate Decision between Pathogenic Th17 and Regulatory T Cells', *Cell Reports*, 30(6), pp. 1898-1909.e4. doi:10.1016/j.celrep.2020.01.022.

Siddiqui, A., Rivera-Sánchez, S., Castro, M. del R., Acevedo-Torres, K., Rane, A., Torres-Ramos, C.A., Nicholls, D.G., Andersen, J.K. and Ayala-Torres, S. (2012) 'Mitochondrial DNA damage is associated with reduced mitochondrial bioenergetics in Huntington's disease', *Free Radical Biology & Medicine*, 53(7), pp. 1478–1488. doi:10.1016/j.freeradbiomed.2012.06.008.

Signes, A. and Fernandez-Vizarra, E. (2018) 'Assembly of mammalian oxidative phosphorylation complexes I–V and supercomplexes', *Essays in Biochemistry*. Edited by C. Garone and M. Minczuk, 62(3), pp. 255–270. doi:10.1042/EBC20170098.

Silk, M. and Nantz, E. (2018) 'Efficacy and Safety of Tabalumab in Patients with Relapsing-Remitting Multiple Sclerosis: A Randomized, Double-Blind, Placebo-Controlled Study (P3.397)', *Neurology*, 90(15 Supplement), p. P3.397.

Silva-Pinheiro, P. and Minczuk, M. (2021) 'The potential of mitochondrial genome engineering', *Nature Reviews Genetics* [Preprint]. doi:10.1038/s41576-021-00432-x.

Simonetti, S., Chen, X., DiMauro, S. and Schon, E.A. (1992) 'Accumulation of deletions in human mitochondrial DNA during normal aging: analysis by quantitative PCR', *Biochimica*

et Biophysica Acta (BBA) - Molecular Basis of Disease, 1180(2), pp. 113–122. doi:10.1016/0925-4439(92)90059-V.

Simpson, S., Wang, W., Otahal, P., Blizzard, L., van der Mei, I.A.F. and Taylor, B.V. (2019) 'Latitude continues to be significantly associated with the prevalence of multiple sclerosis: an updated meta-analysis', *Journal of Neurology, Neurosurgery & Psychiatry*, 90(11), p. 1193. doi:10.1136/jnnp-2018-320189.

Sirey, T.M. and Ponting, C.P. (2016) 'Insights into the post-transcriptional regulation of the mitochondrial electron transport chain', *Biochemical Society Transactions*, 44(5), pp. 1491–1498. doi:10.1042/BST20160100.

Slauch, J.M. (2011) 'How does the oxidative burst of macrophages kill bacteria? Still an open question', *Molecular Microbiology*, 80(3), pp. 580–583. doi:10.1111/j.1365-2958.2011.07612.x.

SMART (2021) *Servier Medical Art*. Available at: https://smart.servier.com/ (Accessed: 17 December 2021).

Smart, U., Budowle, B., Ambers, A., Soares Moura-Neto, R., Silva, R. and Woerner, A.E. (2019) 'A novel phylogenetic approach for de novo discovery of putative nuclear mitochondrial (pNumt) haplotypes', *Forensic Science International. Genetics*, 43, p. 102146. doi:10.1016/j.fsigen.2019.102146.

Smorodchenko, A., Schneider, S., Rupprecht, A., Hilse, K., Sasgary, S., Zeitz, U., Erben, R.G. and Pohl, E.E. (2017) 'UCP2 up-regulation within the course of autoimmune encephalomyelitis correlates with T-lymphocyte activation', *Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease*, 1863(4), pp. 1002–1012. doi:10.1016/j.bbadis.2017.01.019.

Snezhkina, A.V., Kudryavtseva, A.V., Kardymon, O.L., Savvateeva, M.V., Melnikova, N.V., Krasnov, G.S. and Dmitriev, A.A. (2019) 'ROS Generation and Antioxidant Defense Systems in Normal and Malignant Cells', *Oxidative Medicine and Cellular Longevity*. Edited by M. Isaguliants, 2019, p. 6175804. doi:10.1155/2019/6175804.

Sonney, S., Leipzig, J., Lott, M.T., Zhang, S., Procaccio, V., Wallace, D.C. and Sondheimer, N. (2017) 'Predicting the pathogenicity of novel variants in mitochondrial tRNA with MitoTIP', *PLOS Computational Biology*, 13(12), p. e1005867. doi:10.1371/journal.pcbi.1005867.

Sosa, M.X., Sivakumar, I.K.A., Maragh, S., Veeramachaneni, V., Hariharan, R., Parulekar, M., Fredrikson, K.M., Harkins, T.T., Lin, J., Feldman, A.B., Tata, P., Ehret, G.B. and Chakravarti, A. (2012) 'Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency', *PLoS computational biology*, 8(10), p. e1002737. doi:10.1371/journal.pcbi.1002737.

Souren, N.Y.P., Gerdes, L.A., Kümpfel, T., Lutsik, P., Klopstock, T., Hohlfeld, R. and Walter, J. (2016) 'Mitochondrial DNA Variation and Heteroplasmy in Monozygotic Twins Clinically Discordant for Multiple Sclerosis', *Human Mutation*, 37(8), pp. 765–775. doi:10.1002/humu.23003.

Spadaro, M., Winklmeier, S., Beltrán, E., Macrini, C., Höftberger, R., Schuh, E., Thaler, F.S., Gerdes, L.A., Laurent, S., Gerhards, R., Brändle, S., Dornmair, K., Breithaupt, C., Krumbholz, M., Moser, M., Krishnamoorthy, G., Kamp, F., Jenne, D., Hohlfeld, R., Kümpfel, T., Lassmann, H., Kawakami, N. and Meinl, E. (2018) 'Pathogenicity of human antibodies against myelin oligodendrocyte glycoprotein', *Annals of Neurology*, 84(2), pp. 315–328. doi:10.1002/ana.25291.

Spelman, T., Magyari, M., Piehl, F., Svenningsson, A., Rasmussen, P.V., Kant, M., Sellebjerg, F., Joensen, H., Hillert, J. and Lycke, J. (2021) 'Treatment Escalation vs Immediate Initiation of Highly Effective Treatment for Patients With Relapsing-Remitting Multiple Sclerosis: Data From 2 Different National Strategies', *JAMA Neurology*, 78(10), pp. 1197–1204. doi:10.1001/jamaneurol.2021.2738.

Stevenson P G, Hawke S, Sloan D J, and Bangham C R (1997) 'The immunogenicity of intracerebral virus infection depends on anatomical site', *Journal of Virology*, 71(1), pp. 145–151. doi:10.1128/jvi.71.1.145-151.1997.

Stevenson, P.G., Bangham, C.R.M. and Hawke, S. (1997) 'Recruitment, activation and proliferation of CD8+ memory T cells in an immunoprivileged site', *European Journal of Immunology*, 27(12), pp. 3259–3268. doi:10.1002/eji.1830271225.

Stewart, J.B. and Chinnery, P.F. (2015) 'The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease', *Nature Reviews Genetics*, 16(9), pp. 530–542. doi:10.1038/nrg3966.

Stranneheim, H., Lagerstedt-Robinson, K., Magnusson, M., Kvarnung, M., Nilsson, D., Lesko, N., Engvall, M., Anderlid, B.-M., Arnell, H., Johansson, C.B., Barbaro, M., Björck, E., Bruhn, H., Eisfeldt, J., Freyer, C., Grigelioniene, G., Gustavsson, P., Hammarsjö, A., Hellström-Pigg, M., Iwarsson, E., Jemt, A., Laaksonen, M., Enoksson, S.L., Malmgren, H., Naess, K., Nordenskjöld, M., Oscarson, M., Pettersson, M., Rasi, C., Rosenbaum, A., Sahlin, E., Sardh, E., Stödberg, T., Tesi, B., Tham, E., Thonberg, H., Töhönen, V., von Döbeln, U., Vassiliou, D., Vonlanthen, S., Wikström, A.-C., Wincent, J., Winqvist, O., Wredenberg, A., Ygberg, S., Zetterström, R.H., Marits, P., Soller, M.J., Nordgren, A., Wirta, V., Lindstrand, A. and Wedell, A. (2021) 'Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients', *Genome Medicine*, 13(1), p. 40. doi:10.1186/s13073-021-00855-5.

Strobl, C., Churchill Cihlar, J., Lagacé, R., Wootton, S., Roth, C., Huber, N., Schnaller, L., Zimmermann, B., Huber, G., Lay Hong, S., Moura-Neto, R., Silva, R., Alshamali, F., Souto, L., Anslinger, K., Egyed, B., Jankova-Ajanovska, R., Casas-Vargas, A., Usaquén, W., Silva, D., Barletta-Carrillo, C., Tineo, D.H., Vullo, C., Würzner, R., Xavier, C., Gusmão, L., Niederstätter, H., Bodner, M., Budowle, B. and Parson, W. (2019) 'Evaluation of mitogenome sequence concordance, heteroplasmy detection, and haplogrouping in a worldwide lineage study using the Precision ID mtDNA Whole Genome Panel', *Forensic Science International. Genetics*, 42, pp. 244–251. doi:10.1016/j.fsigen.2019.07.013.

Strobl, C., Eduardoff, M., Bus, M.M., Allen, M. and Parson, W. (2018) 'Evaluation of the precision ID whole MtDNA genome panel for forensic analyses', *Forensic Science International. Genetics*, 35, pp. 21–25. doi:10.1016/j.fsigen.2018.03.013.

Sturk-Andreaggi, K., Parson, W., Allen, M. and Marshall, C. (2020) 'Impact of the sequencing method on the detection and interpretation of mitochondrial DNA length heteroplasmy', *Forensic Science International. Genetics*, 44, p. 102205. doi:10.1016/j.fsigen.2019.102205.

Sue Hughes (2021) *Updated MS Guidelines Advocate Earlier, More Aggressive Treatment*, *Medscape*. Available at: http://www.medscape.com/viewarticle/961161 (Accessed: 5 January 2022).

Sun, L., Middleton, D.R., Wantuch, P.L., Ozdilek, A. and Avci, F.Y. (2016) 'Carbohydrates as T-cell antigens with implications in health and disease', *Glycobiology*, 26(10), pp. 1029–1040. doi:10.1093/glycob/cww062.

Syndercombe Court, D. (2021) 'Mitochondrial DNA in forensic use', *Emerging Topics in Life Sciences*. Edited by N.N. Daeid and L. Hackman, 5(3), pp. 415–426. doi:10.1042/ETLS20210204.

Talbot, S., Foster, S.L. and Woolf, C.J. (2016) 'Neuroimmunity: Physiology and Pathology', *Annual Review of Immunology*, 34(1), pp. 421–447. doi:10.1146/annurev-immunol-041015-055340.

Tang, Z., Lu, Z., Chen, B., Zhang, W., Chang, H.Y., Hu, Z. and Xu, J. (2021) 'A genetic bottleneck of mitochondrial DNA during human lymphocyte development', *bioRxiv*, p. 2021.09.28.462268. doi:10.1101/2021.09.28.462268.

Tapia-Maltos, M., Treviño-Frenk, I., García-González, H., Rosetti, M., Barriga-Maldonado, V., Morales-Ramírez, F., López-Hernández, D., Rosetti, F. and Crispín, J. (2021) 'Identification of regulatory T cell molecules associated with severity of multiple sclerosis', *Multiple Sclerosis Journal*, 27(11), pp. 1695–1705. doi:10.1177/1352458520977045.

Taylor, R.W., Barron, M.J., Borthwick, G.M., Gospel, A., Chinnery, P.F., Samuels, D.C., Taylor, G.A., Plusa, S.M., Needham, S.J., Greaves, L.C., Kirkwood, T.B.L. and Turnbull, D.M. (2003) 'Mitochondrial DNA mutations in human colonic crypt stem cells', *Journal of Clinical Investigation*, 112(9), pp. 1351–1360. doi:10.1172/JCI19435.

Taylor, R.W. and Turnbull, D.M. (2005) 'Mitochondrial DNA mutations in human disease', *Nature Reviews. Genetics*, 6(5), pp. 389–402. doi:10.1038/nrg1606.

Temperton, J. (2021) 'A dying child, a mother's love and the drug that changed medicine', *Wired UK*, 26 March. Available at: https://www.wired.co.uk/article/milasen-aso-gene-therapy (Accessed: 17 January 2022).

Theibert, A.B. (2020a) 'Functional Anatomy of the Nervous System I: Cerebrum & Subcortical Structures', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242299 (Accessed: 16 December 2021).

Theibert, A.B. (2020b) 'Functional Anatomy of the Nervous System II: The Brainstem, Cerebellum, Spinal Cord, Peripheral Nervous System, & Supporting Systems', in Amthor,

F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242441 (Accessed: 16 December 2021).

Theibert, A.B. (2020c) 'Mechanisms of Synaptic Transmission', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242697 (Accessed: 22 December 2021).

Theibert, A.B. (2020d) 'Neurotransmitter Systems I: Acetylcholine & the Amino Acids', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242774 (Accessed: 22 December 2021).

Theibert, A.B. (2020e) 'Neurotransmitter Systems II: Monoamines, Purines, Neuropeptides, & Unconventional Neurotransmitters', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242855 (Accessed: 22 December 2021).

Theibert, A.B. (2020f) 'Organization & Cells of the Nervous System', in Amthor, F.R., Theibert, A.B., Standaert, D.G., and Roberson, E.D. (eds) *Essentials of Modern Neuroscience*. New York, NY: McGraw Hill. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1174242128 (Accessed: 16 December 2021).

Thompson, A.J., Banwell, B.L., Barkhof, F., Carroll, W.M., Coetzee, T., Comi, G., Correale, J., Fazekas, F., Filippi, M., Freedman, M.S., Fujihara, K., Galetta, S.L., Hartung, H.P., Kappos, L., Lublin, F.D., Marrie, R.A., Miller, A.E., Miller, D.H., Montalban, X., Mowry, E.M., Sorensen, P.S., Tintoré, M., Traboulsee, A.L., Trojano, M., Uitdehaag, B.M.J., Vukusic, S., Waubant, E., Weinshenker, B.G., Reingold, S.C. and Cohen, J.A. (2018) 'Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria', *The Lancet. Neurology*, 17(2), pp. 162–173. doi:10.1016/S1474-4422(17)30470-2.

Tilokani, L., Nagashima, S., Paupe, V. and Prudent, J. (2018) 'Mitochondrial dynamics: overview of molecular mechanisms', *Essays in Biochemistry*. Edited by C. Garone and M. Minczuk, 62(3), pp. 341–360. doi:10.1042/EBC20170104.

Tintore, M., Vidal-Jordana, A. and Sastre-Garriga, J. (2019) 'Treatment of multiple sclerosis — success from bench to bedside', *Nature Reviews Neurology*, 15(1), pp. 53–58. doi:10.1038/s41582-018-0082-z.

Tranah, G.J., Santaniello, A., Caillier, S.J., D'Alfonso, S., Martinelli Boneschi, F., Hauser, S.L. and Oksenberg, J.R. (2015) 'Mitochondrial DNA sequence variation in multiple sclerosis', *Neurology*, 85(4), pp. 325–330. doi:10.1212/WNL.0000000000001744.

Tremlett, H., Yinshan Zhao and Devonshire, V. (2008) 'Natural history of secondary-progressive multiple sclerosis', *Multiple Sclerosis Journal*, 14(3), pp. 314–324. doi:10.1177/1352458507084264.

Trojano, M., Tintore, M., Montalban, X., Hillert, J., Kalincik, T., Iaffaldano, P., Spelman, T., Sormani, M.P. and Butzkueven, H. (2017) 'Treatment decisions in multiple sclerosis — insights from real-world observational studies', *Nature Reviews Neurology*, 13(2), pp. 105–118. doi:10.1038/nrneurol.2016.188.

Tukey, J.W. (1977) *Exploratory data analysis*. Reading, Mass: Addison-Wesley Pub. Co (Addison-Wesley series in behavioral science).

Tuppen, H.A.L., Blakely, E.L., Turnbull, D.M. and Taylor, R.W. (2010) 'Mitochondrial DNA mutations and human disease', *Biochimica Et Biophysica Acta*, 1797(2), pp. 113–128. doi:10.1016/j.bbabio.2009.09.005.

Uhm, T.G., Lee, S.K., Kim, B.S., Kang, J.H., Park, C.S., Rhim, T.Y., Chang, H.S., Kim, D.J. and Chung, I.Y. (2012) 'CpG methylation at GATA elements in the regulatory region of CCR3 positively correlates with CCR3 transcription', *Experimental & Molecular Medicine*, 44(4), pp. 268–280. doi:10.3858/emm.2012.44.4.022.

Uittenbogaard, M., Brantner, C.A., Fang, Z., Wong, L.-J., Gropman, A. and Chiaramello, A. (2019) 'The m.11778 A > G variant associated with the coexistence of Leber's hereditary optic neuropathy and multiple sclerosis-like illness dysregulates the metabolic interplay between mitochondrial oxidative phosphorylation and glycolysis', *Mitochondrion*, 46, pp. 187–194. doi:10.1016/j.mito.2018.06.001.

United States Food & Drug Administration (2022) *New Drug Application (NDA): 020622. Company: TEVA PHARMS USA*, *Drugs@FDA: FDA-Approved Drugs*. Available at: https://www.accessdata.fda.gov/scripts/cder/daf/index.cfm?event=BasicSearch.process (Accessed: 6 January 2022).

University of California, S.F.M.-E.T., Cree, B.A.C., Gourraud, P.-A., Oksenberg, J.R., Bevan, C., Crabtree-Hartman, E., Gelfand, J.M., Goodin, D.S., Graves, J., Green, A.J., Mowry, E., Okuda, D.T., Pelletier, D., von Büdingen, H.-C., Zamvil, S.S., Agrawal, A., Caillier, S., Ciocca, C., Gomez, R., Kanner, R., Lincoln, R., Lizee, A., Qualley, P., Santaniello, A., Suleiman, L., Bucci, M., Panara, V., Papinutto, N., Stern, W.A., Zhu, A.H., Cutter, G.R., Baranzini, S., Henry, R.G. and Hauser, S.L. (2016) 'Long-term evolution of multiple sclerosis disability in the treatment era', *Annals of Neurology*, 80(4), pp. 499–510. doi:10.1002/ana.24747.

Valerio, A. and Nisoli, E. (2015) 'Nitric oxide, interorganelle communication, and energy flow: a novel route to slow aging', *Frontiers in Cell and Developmental Biology*, 3, p. 6. doi:10.3389/fcell.2015.00006.

Van der Auwera, G.A., Carneiro, M.O., Hartl, C., Poplin, R., Del Angel, G., Levy-Moonshine, A., Jordan, T., Shakir, K., Roazen, D., Thibault, J., Banks, E., Garimella, K.V., Altshuler, D., Gabriel, S. and DePristo, M.A. (2013) 'From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline', *Current Protocols in Bioinformatics*, 43, p. 11.10.1-33. doi:10.1002/0471250953.bi1110s43.

Van Hove, H., Martens, L., Scheyltjens, I., De Vlaminck, K., Pombo Antunes, A.R., De Prijck, S., Vandamme, N., De Schepper, S., Van Isterdael, G., Scott, C.L., Aerts, J., Berx, G., Boeckxstaens, G.E., Vandenbroucke, R.E., Vereecke, L., Moechars, D., Guilliams, M., Van Ginderachter, J.A., Saeys, Y. and Movahedi, K. (2019) 'A single-cell atlas of mouse brain macrophages reveals unique transcriptional identities shaped by ontogeny and tissue environment', *Nature Neuroscience*, 22(6), pp. 1021–1035. doi:10.1038/s41593-019-0393-4.

Varhaug, K.N., Kråkenes, T., Alme, M.N., Vedeler, C.A. and Bindoff, L.A. (2020) 'Mitochondrial complex IV is lost in neurons in the cuprizone mouse model', *Mitochondrion*, 50, pp. 58–62. doi:10.1016/j.mito.2019.09.003.

Veauthier, C., Hasselmann, H., Gold, S.M. and Paul, F. (2016) 'The Berlin Treatment Algorithm: recommendations for tailored innovative therapeutic strategies for multiple sclerosis-related fatigue', *The EPMA journal*, 7, p. 25. doi:10.1186/s13167-016-0073-3.

Veiga-Fernandes, H. and Pachnis, V. (2017) 'Neuroimmune regulation during intestinal development and homeostasis', *Nature Immunology*, 18(2), pp. 116–122. doi:10.1038/ni.3634.

Vellarikkal, S.K., Dhiman, H., Joshi, K., Hasija, Y., Sivasubbu, S. and Scaria, V. (2015) 'mit-o-matic: a comprehensive computational pipeline for clinical evaluation of mitochondrial variations from next-generation sequencing datasets', *Human Mutation*, 36(4), pp. 419–424. doi:10.1002/humu.22767.

Venkateswaran, S., Zheng, K., Sacchetti, M., Gagne, D., Arnold, D.L., Sadovnick, A.D., Scherer, S.W., Banwell, B., Bar-Or, A., Simon, D.K., and Canadian Pediatric Demyelinating Disease Network (2011) 'Mitochondrial DNA haplogroups and mutations in children with acquired central demyelination', *Neurology*, 76(9), pp. 774–780. doi:10.1212/WNL.0b013e31820ee1bb.

Venken, K., Hellings, N., Broekmans, T., Hensen, K., Rummens, J.-L. and Stinissen, P. (2008) 'Natural Naive $CD4^{+}CD25^{+}CD127^{low}$ Regulatory T Cell (Treg) Development and Function Are Disturbed in Multiple Sclerosis Patients: Recovery of Memory Treg Homeostasis during Disease Progression', *The Journal of Immunology*, 180(9), p. 6411. doi:10.4049/jimmunol.180.9.6411.

Venken, K., Hellings, N., Thewissen, M., Somers, V., Hensen, K., Rummens, J.-L., Medaer, R., Hupperts, R. and Stinissen, P. (2008) 'Compromised CD4+ CD25high regulatory T-cell function in patients with relapsing-remitting multiple sclerosis is correlated with a reduced frequency of FOXP3-positive cells and reduced FOXP3 expression at the single-cell level', *Immunology*, 123(1), pp. 79–89. doi:10.1111/j.1365-2567.2007.02690.x.

Ventura-Clapier, R., Garnier, A. and Veksler, V. (2008) 'Transcriptional control of mitochondrial biogenesis: the central role of PGC-1', *Cardiovascular Research*, 79(2), pp. 208–217. doi:10.1093/cvr/cvn098.

Vercellino, I. and Sazanov, L.A. (2021) 'The assembly, regulation and function of the mitochondrial respiratory chain', *Nature Reviews Molecular Cell Biology* [Preprint]. doi:10.1038/s41580-021-00415-0.

Veroni, C., Serafini, B., Rosicarelli, B., Fagnani, C. and Aloisi, F. (2018) 'Transcriptional profile and Epstein-Barr virus infection status of laser-cut immune infiltrates from the brain of patients with progressive multiple sclerosis', *Journal of Neuroinflammation*, 15(1), p. 18. doi:10.1186/s12974-017-1049-5.

Vianello, D., Sevini, F., Castellani, G., Lomartire, L., Capri, M. and Franceschi, C. (2013) 'HAPLOFIND: a new method for high-throughput mtDNA haplogroup assignment', *Human Mutation*, 34(9), pp. 1189–1194. doi:10.1002/humu.22356.

Viglietta, V., Baecher-Allan, C., Weiner, H.L. and Hafler, D.A. (2004) 'Loss of Functional Suppression by CD4+CD25+ Regulatory T Cells in Patients with Multiple Sclerosis', *Journal of Experimental Medicine*, 199(7), pp. 971–979. doi:10.1084/jem.20031579.

Vincent, A.E., White, K., Davey, T., Philips, J., Ogden, R.T., Lawless, C., Warren, C., Hall, M.G., Ng, Y.S., Falkous, G., Holden, T., Deehan, D., Taylor, R.W., Turnbull, D.M. and Picard, M. (2019) 'Quantitative 3D Mapping of the Human Skeletal Muscle Mitochondrial Network', *Cell Reports*, 26(4), pp. 996-1009.e4. doi:10.1016/j.celrep.2019.01.010.

Vladimirova, O., O'Connor, J., Cahill, A., Alder, H., Butunoi, C. and Kalman, B. (1998) 'Oxidative damage to DNA in plaques of MS brains', *Multiple Sclerosis Journal*, 4(5), pp. 413–418. doi:10.1177/135245859800400503.

Vogler, S., Goedde, R., Miterski, B., Gold, R., Kroner, A., Koczan, D., Zettl, U.-K., Rieckmann, P., Epplen, J.T. and Ibrahim, S.M. (2005) 'Association of a common polymorphism in the promoter of UCP2 with susceptibility to multiple sclerosis', *Journal of Molecular Medicine (Berlin, Germany)*, 83(10), pp. 806–811. doi:10.1007/s00109-005-0661-5.

Vyshkina, T., Sylvester, A., Sadiq, S., Bonilla, E., Canter, J.A., Perl, A. and Kalman, B. (2008) 'Association of common mitochondrial DNA variants with multiple sclerosis and systemic lupus erythematosus', *Clinical Immunology (Orlando, Fla.)*, 129(1), pp. 31–35. doi:10.1016/j.clim.2008.07.011.

Wachsmuth, M., Hübner, A., Li, M., Madea, B. and Stoneking, M. (2016) 'Age-Related and Heteroplasmy-Related Variation in Human mtDNA Copy Number', *PLoS genetics*, 12(3), p. e1005939. doi:10.1371/journal.pgen.1005939.

Wagner, A., Wang, C., Fessler, J., DeTomaso, D., Avila-Pacheco, J., Kaminski, J., Zaghouani, S., Christian, E., Thakore, P., Schellhaass, B., Akama-Garren, E., Pierce, K., Singh, V., Ron-Harel, N., Douglas, V.P., Bod, L., Schnell, A., Puleston, D., Sobel, R.A., Haigis, M., Pearce, E.L., Soleimani, M., Clish, C., Regev, A., Kuchroo, V.K. and Yosef, N. (2021) 'Metabolic modeling of single Th17 cells reveals regulators of autoimmunity', *Cell*, 184(16), pp. 4168-4185.e21. doi:10.1016/j.cell.2021.05.045.

Wai, K.T., Barash, M. and Gunn, P. (2018) 'Performance of the Early Access AmpliSeq™ Mitochondrial Panel with degraded DNA samples using the Ion Torrent™ platform', *Electrophoresis*, 39(21), pp. 2776–2784. doi:10.1002/elps.201700371.

Walker, M.A., Lareau, C.A., Ludwig, L.S., Karaa, A., Sankaran, V.G., Regev, A. and Mootha, V.K. (2020) 'Purifying Selection against Pathogenic Mitochondrial DNA in Human T Cells', *New England Journal of Medicine*, 383(16), pp. 1556–1563. doi:10.1056/NEJMoa2001265.

Wallace, D.C., Singh, G., Lott, M.T., Hodge, J.A., Schurr, T.G., Lezza, A.M., Elsas, L.J. and Nikoskelainen, E.K. (1988) 'Mitochondrial DNA mutation associated with Leber's hereditary optic neuropathy', *Science (New York, N.Y.)*, 242(4884), pp. 1427–1430. doi:10.1126/science.3201231.

Wallin, M.T., Culpepper, W.J., Nichols, E., Bhutta, Z.A., Gebrehiwot, T.T., Hay, S.I., Khalil, I.A., Krohn, K.J., Liang, X., Naghavi, M., Mokdad, A.H., Nixon, M.R., Reiner, R.C., Sartorius, B., Smith, M., Topor-Madry, R., Werdecker, A., Vos, T., Feigin, V.L. and Murray, C.J.L. (2019) 'Global, regional, and national burden of multiple sclerosis 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016', *The Lancet Neurology*, 18(3), pp. 269–285. doi:10.1016/S1474-4422(18)30443-5.

Walton, C., King, R., Rechtman, L., Kaye, W., Leray, E., Marrie, R.A., Robertson, N., La Rocca, N., Uitdehaag, B., van der Mei, I., Wallin, M., Helme, A., Angood Napier, C., Rijke, N. and Baneke, P. (2020) 'Rising prevalence of multiple sclerosis worldwide: Insights from the Atlas of MS, third edition', *Multiple Sclerosis Journal*, 26(14), pp. 1816–1821. doi:10.1177/1352458520970841.

Wang, J., Jelcic, I., Mühlenbruch, L., Haunerdinger, V., Toussaint, N.C., Zhao, Y., Cruciani, C., Faigle, W., Naghavian, R., Foege, M., Binder, T.M.C., Eiermann, T., Opitz, L., Fuentes-Font, L., Reynolds, R., Kwok, W.W., Nguyen, J.T., Lee, J.-H., Lutterotti, A., Münz, C., Rammensee, H.-G., Hauri-Hohl, M., Sospedra, M., Stevanovic, S. and Martin, R. (2020) 'HLA-DR15 Molecules Jointly Shape an Autoreactive T Cell Repertoire in Multiple Sclerosis', *Cell*, 183(5), pp. 1264-1281.e20. doi:10.1016/j.cell.2020.09.054.

Wang, J., Zhao, C., Kong, P., Sun, H., Sun, Z., Bian, G., Sun, Y. and Guo, L. (2016) 'Treatment with NAD+ inhibited experimental autoimmune encephalomyelitis by activating AMPK/SIRT1 signaling pathway and modulating Th1/Th17 immune responses in mice', *International Immunopharmacology*, 39, pp. 287–294. doi:10.1016/j.intimp.2016.07.036.

Wang, L.W., Shen, H., Nobre, L., Ersing, I., Paulo, J.A., Trudeau, S., Wang, Z., Smith, N.A., Ma, Y., Reinstadler, B., Nomburg, J., Sommermann, T., Cahir-McFarland, E., Gygi, S.P., Mootha, V.K., Weekes, M.P. and Gewurz, B.E. (2019) 'Epstein-Barr-Virus-Induced One-Carbon Metabolism Drives B Cell Transformation', *Cell Metabolism*, 30(3), pp. 539-555.e11. doi:10.1016/j.cmet.2019.06.003.

Wang, Q., Dhindsa, R.S., Carss, K., Harper, A.R., Nag, A., Tachmazidou, I., Vitsios, D., Deevi, S.V.V., Mackay, A., Muthas, D., Hühn, M., Monkley, S., Olsson, H., Angermann, B.R., Artzi, R., Barrett, C., Belvisi, M., Bohlooly-Y, M., Burren, O., Buvall, L., Challis, B., Cameron-Christie, S., Cohen, S., Davis, A., Danielson, R.F., Dougherty, B., Georgi, B., Ghazoui, Z., Hansen, P.B.L., Hu, F., Jeznach, M., Jiang, X., Kumar, C., Lai, Z., Lassi, G., Lewis, S.H., Linghu,

B., Lythgow, K., Maccallum, P., Martins, C., Matakidou, A., Michaëlsson, E., Moosmang, S., O'Dell, S., Ohne, Y., Okae, J., O'Neill, A., Paul, D.S., Reznichenko, A., Snowden, M.A., Walentinsson, A., Zeron, J., Pangalos, M.N., Wasilewski, S., Smith, K.R., March, R., Platt, A., Haefliger, C., Petrovski, S., and AstraZeneca Genomics Initiative (2021) 'Rare variant contribution to human disease in 281,104 UK Biobank exomes', *Nature*, 597(7877), pp. 527–532. doi:10.1038/s41586-021-03855-y.

Wang, Y., Guo, X., Ye, K., Orth, M. and Gu, Z. (2021) 'Accelerated expansion of pathogenic mitochondrial DNA heteroplasmies in Huntington's disease', *Proceedings of the National Academy of Sciences*, 118(30), p. e2014610118. doi:10.1073/pnas.2014610118.

Wang, Y., Michikawa, Y., Mallidis, C., Bai, Y., Woodhouse, L., Yarasheski, K.E., Miller, C.A., Askanas, V., Engel, W.K., Bhasin, S. and Attardi, G. (2001) 'Muscle-specific mutations accumulate with aging in critical human mtDNA control sites for replication', *Proceedings of the National Academy of Sciences of the United States of America*, 98(7), pp. 4022–4027. doi:10.1073/pnas.061013598.

Warne, J., Pryce, G., Hill, J.M., Shi, X., Lennerås, F., Puentes, F., Kip, M., Hilditch, L., Walker, P., Simone, M.I., Chan, A.W.Edit., Towers, G.J., Coker, A.R., Duchen, M.R., Szabadkai, G., Baker, D. and Selwood, D.L. (2016) 'Selective Inhibition of the Mitochondrial Permeability Transition Pore Protects against Neurodegeneration in Experimental Multiple Sclerosis', *Journal of Biological Chemistry*, 291(9), pp. 4356–4373. doi:10.1074/jbc.M115.700385.

Wattjes, M.P., Ciccarelli, O., Reich, D.S., Banwell, B., de Stefano, N., Enzinger, C., Fazekas, F., Filippi, M., Frederiksen, J., Gasperini, C., Hacohen, Y., Kappos, L., Li, D.K.B., Mankad, K., Montalban, X., Newsome, S.D., Oh, J., Palace, J., Rocca, M.A., Sastre-Garriga, J., Tintoré, M., Traboulsee, A., Vrenken, H., Yousry, T., Barkhof, F., Rovira, À., Magnetic Resonance Imaging in Multiple Sclerosis study group, Consortium of Multiple Sclerosis Centres, and North American Imaging in Multiple Sclerosis Cooperative MRI guidelines working group (2021) '2021 MAGNIMS-CMSC-NAIMS consensus recommendations on the use of MRI in patients with multiple sclerosis', *The Lancet. Neurology*, 20(8), pp. 653–670. doi:10.1016/S1474-4422(21)00095-8.

Waxman, S.G. (2020a) 'Fundamentals of the Nervous System', in *Clinical Neuroanatomy, 29e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1171519650 (Accessed: 16 December 2021).

Waxman, S.G. (2020b) 'The Spinal Cord', in *Clinical Neuroanatomy, 29e*. New York, NY: McGraw-Hill Education. Available at: accessmedicine.mhmedical.com/content.aspx?aid=1186188964 (Accessed: 16 December 2021).

Wei, W., Pagnamenta, A.T., Gleadall, N., Sanchis-Juan, A., Stephens, J., Broxholme, J., Tuna, S., Odhams, C.A., Genomics England Research Consortium, NIHR BioResource, Fratter, C., Turro, E., Caulfield, M.J., Taylor, J.C., Rahman, S. and Chinnery, P.F. (2020) 'Nuclear-mitochondrial DNA segments resemble paternally inherited mitochondrial DNA in humans', *Nature Communications*, 11(1), p. 1740. doi:10.1038/s41467-020-15336-3.

Wei, W., Tuna, S., Keogh, M.J., Smith, K.R., Aitman, T.J., Beales, P.L., Bennett, D.L., Gale, D.P., Bitner-Glindzicz, M.A.K., Black, G.C., Brennan, P., Elliott, P., Flinter, F.A., Floto, R.A., Houlden, H., Irving, M., Koziell, A., Maher, E.R., Markus, H.S., Morrell, N.W., Newman, W.G., Roberts, I., Sayer, J.A., Smith, K.G.C., Taylor, J.C., Watkins, H., Webster, A.R., Wilkie, A.O.M., Williamson, C., NIHR BioResource–Rare Diseases, 100,000 Genomes Project–Rare Diseases Pilot, Ashford, S., Penkett, C.J., Stirrups, K.E., Rendon, A., Ouwehand, W.H., Bradley, J.R., Raymond, F.L., Caulfield, M., Turro, E., Chinnery, P.F., Huissoon, A.P., Crisp-Hihn, A., Shaw, A.C., Mead, A.J., Levine, A.P., Thrasher, A.J., Bierzynska, A., Hassan, A., Kumar, A., Sanchis-Juan, A., Richter, A., Lawrie, A., Frary, A.J., Nemeth, A.H., Olschewski, A., Themistocleous, A.C., Browning, A.C., Mumford, A.D., Schaefer, A.M., Marshall, A., Wilkie, A.O.M., Peacock, A., Harper, A.R., Webster, A.R., Rice, A.S.C., Pyle, A., Koziell, A., Drazyk, A.M., Kelly, A.M., Wagner, A., Attwood, A., De Soyza, A., Vandersteen, A.M., Moore, A.T., Vonk Noordegraaf, A., Rao, A., Herwadkar, A., Houweling, A., Sen, A., Rendon, A., Worth, A., Girerd, B., Madan, B., Wilson, B.T., Diz, C.B., Treacy, C., Brewer, C., Campbell, C., Millar, C., Roughley, C., Titterton, C., Williamson, C., Compton, C.J., Danesino, C., Thys, C., Hadinnapola, C., Deshpande, C., Toh, C.-H., Van Geet, C., Babbs, C., Woods, C.G., Penkett, C.J., Watt, C., Harris, C., Lentaigne, C., Palles, C., Searle, C., Pilkington, C., Church, C., French, C.E., Samarghitean, C., Layton, D.M., Evans, D.G., Smedley, D., Greene, D., Hart, D., Gale, D.P., Kiely, D.G., Gosal, D., Allsup, D.J., Bennett, D.L., Montani, D., Parry, D., Thomas, D., Ruddy, D.M., Whitehorn, D., Grozeva, D., Bockenhauer, D., Kumararatne, D., Josifova, D., Maher, E.R., Wong, E.K.S., Dewhurst, E.F., Louka, E., Colby, E., Ormondroyd, E., Thomas, E., Swietlik, E., Staples, E., Matthews, E., Woodward, E., Turro, E., Haque, E., Raymond, F.L., Hu, F., Lalloo, F., Soubrier, F., Cheng, F., Flinter, F.A., Kovacs, G., Arno, G., Hudson, G., Sayer, G., Carr-White, G., Coghlan, G., Evans, G., Black, G.C., Hayman, G., Cook, H.T., Alachkar, H., Allen, H.L., Kazkaz, H., Stark, H., Marschall, H.-U., Bogaard, H., Maxwell, H., Baxendale, H.E., Hanson, H.L., Gall, H., Houlden, H., Longhurst, H., Fassihi, H., Olschewski, H., Ghofrani, H.-A., Markus, H.S., Watkins, H., Tomlinson, I.P., Simeoni, I., Roberts, I., Edgar, J.D.M., Gibbs, J.S.R., Thaventhiran, J.E., Fox, J., Ware, J.S., Whitworth, J., Collins, J., Suntharalingam, J., Jolley, J., Martin, J., O'Sullivan, J., Taylor, J.C., Chambers, J., Maimaris, J., Clayton-Smith, J., Pepke-Zaba, J., Graham, J., Sayer, J.A., Westwood, J.-P., Burn, J., Davis, J., Wharton, J., Taylor, J., Hoffman, J., Stephens, J., Adlard, J., von Ziegenweidt, J., Wessels, J., Ong, K.R., Edwards, K., Downes, K., Gibson, K., Talks, K., Thomson, K., Peerlinck, K., Smith, K.R., Yates, K., Stirrups, K.E., Freson, K., Snape, K., Gomez, K., Sibson, K., Muir, K.W., Poole, K.E.S., Smith, K.G.C., Carss, K., Marchbank, K.J., Gilmour, K.C., Harkness, K., Abulhoul, L., Scelsi, L., Robert, L., Lorenzo, L.E., Izatt, L., Side, L., Wedderburn, L., Howard, L.S., Greenhalgh, L., Chitre, M., Kurian, M.A., Humbert, M., Tischkowitz, M., Bitner-Glindzicz, M.A.K., Estiu, M.C., Erwood, M., Scully, M., Caulfield, M., Gurnell, M., McCarthy, M.I., Toshner, M., Bleda, M., Vazquez-Lopez, M., Wilkins, M.R., Mathias, M., Brown, M., Sims, M.C., Hall, M., Daniels, M.J., Buckland, M.S., Traylor, M., Haimel, M., Cleary, M., Dattani, M., Eyries, M., Chan, M.M.Y., Ekani, M.N., Irving, M., Laffan, M.A., Gattens, M., Browning, M.J., Newnham, M., Simpson, M., Wright, M., Michaelides, M., Lambert, M., Saleem, M., Thomas, M.J., Mozere, M., Ahmed, M., Brod, N.C., Kingston, N., Shah, N., Jurkute, N., Cooper, N., Morrell, N.W., Curry, N.S., Burrows, N., Koelling, N., Roy, N.B., Shamardina, O., Spasic-Boskovic, O., Sadeghi-Alavijeh, O., Gresele, P., Chinnery, P.F., Yong, P.F.K., Yu-Wai-Man, P., Lyons, P.A., Aurora, P., Brennan, P., Corris, P., McAlinden, P., Rayner-Matthews, P.J., Gordins, P., Elliott, P., Dixon, P.H., Kelleher, P., Collins, P.W., Syrris, P., Beales, P.L., Ancliff, P., Griffiths, P., Twiss, P., Yu, P., Waisfisz, Q., Floto, R.A., Tait, R.C., Buchan, R.J., Linger, R., Kazmi, R., Sargur, R.B.,

Favier, R., Tan, R.Y.Y., Antrobus, R., Quinton, R., Scott, R., Trembath, R., Horvath, R., Sarkany, R.N., Ross-Russell, R., MacKenzie Ross, R.V., Condliffe, R., James, R., Hague, R., Mapeta, R., Armstrong, R., Casey, R., Noorani, S., Tuna, S., Johnson, S.A., Malka, S., Obaji, S., Boyce, S., Goddard, S., Rose, S.J., Westbury, S.K., Mangles, S., Mehta, S., Hackett, S., Moledina, S., Rahman, S., Mohammed, S.N., Banka, S., Holden, S., Pearce, S., Satchell, S., Staines, S., Savic, S., Patel, S., Douzgou, S., Grigoriadou, S., Papadia, S., Ashford, S., Schulman, S., Park, S.-M., Deevi, S.V.V., Gräf, S., Abbs, S., Wort, S.J., Jolles, S., Marks, S., Okoli, S., Cook, S., Meacham, S., Walker, S.M., Shapiro, S.E., Sivapalaratnam, S., Kuijpers, T.W., Bariana, T., Bakchoul, T., Everington, T., Renton, T., Bueser, T., Dent, T., Aitman, T.J., Biss, T., Dutt, T., Fowler, T., Vale, T., Lester, T., Cole, T.R.P., Ganesan, V., Sewell, W.A.C., Wei, W., Erber, W.N., Seeger, W., Kelsall, W., Ouwehand, W.H., Egner, W., Newman, W.G., Hague, W.M. and Wood, Y. (2019) 'Germline selection shapes human mitochondrial DNA diversity', *Science*, 364(6442), p. eaau6520. doi:10.1126/science.aau6520.

Wei, Y.H. (1992) 'Mitochondrial DNA alterations as ageing-associated molecular events', *Mutation Research*, 275(3–6), pp. 145–155. doi:10.1016/0921-8734(92)90019-l.

Weinhard, L., di Bartolomei, G., Bolasco, G., Machado, P., Schieber, N.L., Neniskyte, U., Exiga, M., Vadisiute, A., Raggioli, A., Schertel, A., Schwab, Y. and Gross, C.T. (2018) 'Microglia remodel synapses by presynaptic trogocytosis and spine head filopodia induction', *Nature Communications*, 9(1), p. 1228. doi:10.1038/s41467-018-03566-5.

Weissensteiner, H., Forer, L., Fendt, L., Kheirkhah, A., Salas, A., Kronenberg, F. and Schoenherr, S. (2021) 'Contamination detection in sequencing studies using the mitochondrial phylogeny', *Genome Research* [Preprint]. doi:10.1101/gr.256545.119.

Weissensteiner, H., Forer, L., Fuchsberger, C., Schöpf, B., Kloss-Brandstätter, A., Specht, G., Kronenberg, F. and Schönherr, S. (2016) 'mtDNA-Server: next-generation sequencing data analysis of human mitochondrial DNA in the cloud', *Nucleic Acids Research*, 44(W1), pp. W64-69. doi:10.1093/nar/gkw247.

Weissensteiner, H., Pacher, D., Kloss-Brandstätter, A., Forer, L., Specht, G., Bandelt, H.-J., Kronenberg, F., Salas, A. and Schönherr, S. (2016) 'HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing', *Nucleic Acids Research*, 44(W1), pp. W58-63. doi:10.1093/nar/gkw233.

Wekerle, H. (2017) 'B cells in multiple sclerosis', *Autoimmunity*, 50(1), pp. 57–60. doi:10.1080/08916934.2017.1281914.

Wesnes, K., Myhr, K.-M., Riise, T., Cortese, M., Pugliatti, M., Boström, I., Landtblom, A.-M., Wolfson, C. and Bjørnevik, K. (2018) 'Physical activity is associated with a decreased multiple sclerosis risk: The EnvIMS study', *Multiple Sclerosis Journal*, 24(2), pp. 150–157. doi:10.1177/1352458517694088.

Wesnes, K., Riise, T., Casetta, I., Drulovic, J., Granieri, E., Holmøy, T., Kampman, M.T., Landtblom, A.-M., Lauer, K., Lossius, A., Magalhaes, S., Pekmezovic, T., Bjørnevik, K., Wolfson, C., Pugliatti, M. and Myhr, K.-M. (2015) 'Body size and the risk of multiple sclerosis in Norway and Italy: The EnvIMS study', *Multiple Sclerosis Journal*, 21(4), pp. 388–395. doi:10.1177/1352458514546785.

Westerlind, H., Ramanujam, R., Uvehag, D., Kuja-Halkola, R., Boman, M., Bottai, M., Lichtenstein, P. and Hillert, J. (2014) 'Modest familial risks for multiple sclerosis: a registry-based study of the population of Sweden', *Brain*, 137(3), pp. 770–778. doi:10.1093/brain/awt356.

Wetterstrand, K.A. (2021) *DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP)*. Available at: https://www.genome.gov/sequencingcostsdata (Accessed: 1 January 2022).

Wickham, H., Averick, M., Bryan, J., Chang, W., McGowan, L., François, R., Grolemund, G., Hayes, A., Henry, L., Hester, J., Kuhn, M., Pedersen, T., Miller, E., Bache, S., Müller, K., Ooms, J., Robinson, D., Seidel, D., Spinu, V., Takahashi, K., Vaughan, D., Wilke, C., Woo, K. and Yutani, H. (2019) 'Welcome to the Tidyverse', *Journal of Open Source Software*, 4(43), p. 1686. doi:10.21105/joss.01686.

Wickham, H. and Bryan, J. (2019) *readxl: Read Excel Files*. Available at: https://CRAN.R-project.org/package=readxl (Accessed: 4 November 2019).

Wickham, H., Henry, L., Pedersen, T.L., Luciani, T.J., Decorde, M. and Lise, V. (2021) *svglite: An 'SVG' Graphics Device*. Available at: https://CRAN.R-project.org/package=svglite (Accessed: 2 February 2022).

Wickham, H. and Seidel, D. (2020) *scales: Scale Functions for Visualization*. Available at: https://CRAN.R-project.org/package=scales (Accessed: 2 February 2022).

Widner, H. and Brundin, P. (1988) 'Immunological aspects of grafting in the mammalian central nervous system. A review and speculative synthesis', *Brain Research Reviews*, 13(3), pp. 287–324. doi:10.1016/0165-0173(88)90010-0.

Wilichowski, E., Ohlenbusch, A. and Hanefeld, F. (1998) 'Characterization of the mitochondrial genome in childhood multiple sclerosis. II. Multiple sclerosis without optic neuritis and LHON-associated genes', *Neuropediatrics*, 29(6), pp. 307–312. doi:10.1055/s-2007-973583.

Wilkinson, M.D., Dumontier, M., Aalbersberg, I.J.J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L.B., Bourne, P.E., Bouwman, J., Brookes, A.J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C.T., Finkers, R., Gonzalez-Beltran, A., Gray, A.J.G., Groth, P., Goble, C., Grethe, J.S., Heringa, J., 't Hoen, P.A.C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S.J., Martone, M.E., Mons, A., Packer, A.L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M.A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J. and Mons, B. (2016) 'The FAIR Guiding Principles for scientific data management and stewardship', *Scientific Data*, 3, p. 160018. doi:10.1038/sdata.2016.18.

Willer, C.J., Dyment, D.A., Risch, N.J., Sadovnick, A.D., Ebers, G.C., and The Canadian Collaborative Study Group (2003) 'Twin concordance and sibling recurrence rates in multiple sclerosis', *Proceedings of the National Academy of Sciences*, 100(22), pp. 12877–12882. doi:10.1073/pnas.1932604100.

Williams, S.L., Mash, D.C., Züchner, S. and Moraes, C.T. (2013) 'Somatic mtDNA mutation spectra in the aging human putamen', *PLoS genetics*, 9(12), p. e1003990. doi:10.1371/journal.pgen.1003990.

Windt, G.J.W. and Pearce, E.L. (2012) 'Metabolic switching and fuel choice during T-cell differentiation and memory development', *Immunological Reviews*, 249(1), pp. 27–42. doi:10.1111/j.1600-065X.2012.01150.x.

Witte, M.E., Nijland, P.G., Drexhage, J.A.R., Gerritsen, W., Geerts, D., van het Hof, B., Reijerkerk, A., de Vries, H.E., van der Valk, P. and van Horssen, J. (2013) 'Reduced expression of PGC-1α partly underlies mitochondrial changes and correlates with neuronal loss in multiple sclerosis cortex', *Acta Neuropathologica*, 125(2), pp. 231–243. doi:10.1007/s00401-012-1052-y.

Woerner, A.E., Ambers, A., Wendt, F.R., King, J.L., Moura-Neto, R.S., Silva, R. and Budowle, B. (2018) 'Evaluation of the precision ID mtDNA whole genome panel on two massively parallel sequencing systems', *Forensic Science International. Genetics*, 36, pp. 213–224. doi:10.1016/j.fsigen.2018.07.015.

Woerner, A.E., Cihlar, J.C., Smart, U. and Budowle, B. (2020) 'Numt identification and removal with RtN!', *Bioinformatics (Oxford, England)*, 36(20), pp. 5115–5116. doi:10.1093/bioinformatics/btaa642.

Wohleb, E.S., Franklin, T., Iwata, M. and Duman, R.S. (2016) 'Integrating neuroimmune systems in the neurobiology of depression', *Nature Reviews Neuroscience*, 17(8), pp. 497–511. doi:10.1038/nrn.2016.69.

Wong, A.K., Sealfon, R.S.G., Theesfeld, C.L. and Troyanskaya, O.G. (2021) 'Decoding disease: from genomes to networks to phenotypes', *Nature Reviews. Genetics* [Preprint]. doi:10.1038/s41576-021-00389-x.

Woolston, C. (2021) 'How burnout and imposter syndrome blight scientific careers', *Nature*, 599(7886), pp. 703–705. doi:10.1038/d41586-021-03042-z.

World Medical Association (2013) 'World Medical Association Declaration of Helsinki: ethical principles for medical research involving human subjects', *JAMA*, 310(20), pp. 2191–2194. doi:10.1001/jama.2013.281053.

Woyach, J.A., Johnson, A.J. and Byrd, J.C. (2012) 'The B-cell receptor signaling pathway as a therapeutic target in CLL', *Blood*, 120(6), pp. 1175–1184. doi:10.1182/blood-2012-02-362624.

Xu, J., Nuno, K., Litzenburger, U.M., Qi, Y., Corces, M.R., Majeti, R. and Chang, H.Y. (2019) 'Single-cell lineage tracing by endogenous mutations enriched in transposase accessible mitochondrial DNA', *eLife*, 8, p. e45105. doi:10.7554/eLife.45105.

Xu, L., Wang, J., Liu, Y., Xie, L., Su, B., Mou, D., Wang, L., Liu, T., Wang, X., Zhang, B., Zhao, L., Hu, L., Ning, H., Zhang, Y., Deng, K., Liu, L., Lu, X., Zhang, T., Xu, J., Li, C., Wu, H., Deng, H. and Chen, H. (2019) 'CRISPR-Edited Stem Cells in a Patient with HIV and Acute Lymphocytic

Leukemia', *New England Journal of Medicine*, 381(13), pp. 1240–1247. doi:10.1056/NEJMoa1817426.

Xu, Y., Hiyoshi, A., Brand, J.S., Smith, K.A., Bahmanyar, S., Alfredsson, L., Olsson, T. and Montgomery, S. (2021) 'Higher body mass index at ages 16 to 20 years is associated with increased risk of a multiple sclerosis diagnosis in subsequent adulthood among men', *Multiple Sclerosis Journal*, 27(1), pp. 147–150. doi:10.1177/1352458520928061.

Xu, Y., Smith, K.A., Hiyoshi, A., Piehl, F., Olsson, T. and Montgomery, S. (2021) 'Hospital-diagnosed infections before age 20 and risk of a subsequent multiple sclerosis diagnosis', *Brain*, 144(8), pp. 2390–2400. doi:10.1093/brain/awab100.

Yamasaki, R., Lu, H., Butovsky, O., Ohno, N., Rietsch, A.M., Cialic, R., Wu, P.M., Doykan, C.E., Lin, J., Cotleur, A.C., Kidd, G., Zorlu, M.M., Sun, N., Hu, W., Liu, L., Lee, J.-C., Taylor, S.E., Uehlein, L., Dixon, D., Gu, J., Floruta, C.M., Zhu, M., Charo, I.F., Weiner, H.L. and Ransohoff, R.M. (2014) 'Differential roles of microglia and monocytes in the inflamed central nervous system', *Journal of Experimental Medicine*, 211(8), pp. 1533–1549. doi:10.1084/jem.20132477.

Yambire, K.F., Fernandez-Mosquera, L., Steinfeld, R., Mühle, C., Ikonen, E., Milosevic, I. and Raimundo, N. (2019) 'Mitochondrial biogenesis is transcriptionally repressed in lysosomal lipid storage diseases', *eLife*, 8, p. e39598. doi:10.7554/eLife.39598.

Yang, I.S., Lee, H.Y., Yang, W.I. and Shin, K.-J. (2013) 'mtDNAprofiler: a Web application for the nomenclature and comparison of human mitochondrial DNA sequences', *Journal of Forensic Sciences*, 58(4), pp. 972–980. doi:10.1111/1556-4029.12139.

Yao, L., Xu, Z., Zhao, H., Tu, Z., Liu, Z., Li, W., Hu, L. and Wan, L. (2018) 'Concordance of mitochondrial DNA sequencing methods on bloodstains using Ion PGM™', *Legal Medicine (Tokyo, Japan)*, 32, pp. 27–30. doi:10.1016/j.legalmed.2018.02.005.

Yao, R.-Q., Ren, C., Xia, Z.-F. and Yao, Y.-M. (2021) 'Organelle-specific autophagy in inflammatory diseases: a potential therapeutic target underlying the quality control of multiple organelles', *Autophagy*, 17(2), pp. 385–401. doi:10.1080/15548627.2020.1725377.

Yao, Y.-G., Kajigaya, S. and Young, N.S. (2015) 'Mitochondrial DNA mutations in single human blood cells', *Mutation Research*, 779, pp. 68–77. doi:10.1016/j.mrfmmm.2015.06.009.

Ye, K., Lu, J., Ma, F., Keinan, A. and Gu, Z. (2014a) 'Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals', *Proceedings of the National Academy of Sciences of the United States of America*, 111(29), pp. 10654–10659. doi:10.1073/pnas.1403521111.

Ye, K., Lu, J., Ma, F., Keinan, A. and Gu, Z. (2014b) 'Reply to Just et al.: Mitochondrial DNA heteroplasmy could be reliably detected with massively parallel sequencing technologies', *Proceedings of the National Academy of Sciences of the United States of America*, 111(43), pp. E4548-4550. doi:10.1073/pnas.1415171111.

Yonova-Doing, E., Calabrese, C., Gomez-Duran, A., Schon, K., Wei, W., Karthikeyan, S., Chinnery, P.F. and Howson, J.M.M. (2021) 'An atlas of mitochondrial DNA genotype-phenotype associations in the UK Biobank', *Nature Genetics*, 53(7), pp. 982–993. doi:10.1038/s41588-021-00868-1.

Young, E.A., Fowler, C.D., Kidd, G.J., Chang, A., Rudick, R., Fisher, E. and Trapp, B.D. (2008) 'Imaging correlates of decreased axonal Na+/K+ ATPase in chronic multiple sclerosis lesions', *Annals of Neurology*, 63(4), pp. 428–435. doi:10.1002/ana.21381.

Young, L.S., Yap, L.F. and Murray, P.G. (2016) 'Epstein–Barr virus: more than 50 years old and still providing surprises', *Nature Reviews Cancer*, 16(12), pp. 789–802. doi:10.1038/nrc.2016.92.

Yu, X., Koczan, D., Sulonen, A.-M., Akkad, D.A., Kroner, A., Comabella, M., Costa, G., Corongiu, D., Goertsches, R., Camina-Tato, M., Thiesen, H.-J., Nyland, H.I., Mørk, S.J., Montalban, X., Rieckmann, P., Marrosu, M.G., Myhr, K.-M., Epplen, J.T., Saarela, J. and Ibrahim, S.M. (2008) 'mtDNA nt13708A Variant Increases the Risk of Multiple Sclerosis', *PLOS ONE*, 3(2), p. e1530. doi:10.1371/journal.pone.0001530.

Yuan, Y., Ju, Y.S., Kim, Y., Li, J., Wang, Y., Yoon, C.J., Yang, Y., Martincorena, I., Creighton, C.J., Weinstein, J.N., Xu, Y., Han, L., Kim, H.-L., Nakagawa, H., Park, K., Campbell, P.J., Liang, H., and PCAWG Consortium (2020) 'Comprehensive molecular characterization of mitochondrial genomes in human cancers', *Nature Genetics*, 52(3), pp. 342–352. doi:10.1038/s41588-019-0557-x.

Zaidi, A.A., Wilton, P.R., Su, M.S.-W., Paul, I.M., Arbeithuber, B., Anthony, K., Nekrutenko, A., Nielsen, R. and Makova, K.D. (2019) 'Bottleneck and selection in the germline and maternal age influence transmission of mitochondrial DNA in human pedigrees', *Proceedings of the National Academy of Sciences*, 116(50), pp. 25172–25178. doi:10.1073/pnas.1906331116.

Zalc, B. (2018) 'One hundred and fifty years ago Charcot reported multiple sclerosis as a new neurological disease', *Brain*, 141(12), pp. 3482–3488. doi:10.1093/brain/awy287.

Zhang, D., Jin, W., Wu, R., Li, J., Park, S.-A., Tu, E., Zanvit, P., Xu, J., Liu, O., Cain, A. and Chen, W. (2019) 'High Glucose Intake Exacerbates Autoimmunity through Reactive-Oxygen-Species-Mediated TGF-β Cytokine Activation', *Immunity*, 51(4), pp. 671-681.e5. doi:10.1016/j.immuni.2019.08.001.

Zhang, H., Alberts, E., Pongratz, V., Mühlau, M., Zimmer, C., Wiestler, B. and Eichinger, P. (2019) 'Predicting conversion from clinically isolated syndrome to multiple sclerosis–An imaging-based machine learning approach', *NeuroImage: Clinical*, 21, p. 101593. doi:10.1016/j.nicl.2018.11.003.

Zhao, H., Shen, J., Medico, L., Platek, M. and Ambrosone, C.B. (2010) 'Length heteroplasmies in human mitochondrial DNA control regions and breast cancer risk', *International Journal of Molecular Epidemiology and Genetics*, 1(3), pp. 184–192.

Zhao, R., Jiang, S., Zhang, L. and Yu, Z. (2019) 'Mitochondrial electron transport chain, ROS generation and uncoupling (Review)', *International Journal of Molecular Medicine*, 44(1), pp. 3–15. doi:10.3892/ijmm.2019.4188.

Zheng, W., Khrapko, K., Coller, H.A., Thilly, W.G. and Copeland, W.C. (2006) 'Origins of human mitochondrial point mutations as DNA polymerase γ-mediated errors', *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*, 599(1), pp. 11–20. doi:10.1016/j.mrfmmm.2005.12.012.

Zhidkov, I., Nagar, T., Mishmar, D. and Rubin, E. (2011) 'MitoBamAnnotator: A web-based tool for detecting and annotating heteroplasmy in human mitochondrial DNA sequences', *Mitochondrion*, 11(6), pp. 924–928. doi:10.1016/j.mito.2011.08.005.

Zhu, J., Yamane, H. and Paul, W.E. (2010) 'Differentiation of Effector CD4 T Cell Populations', *Annual Review of Immunology*, 28(1), pp. 445–489. doi:10.1146/annurev-immunol-030409-101212.

Ziabreva, I., Campbell, G., Rist, J., Zambonin, J., Rorbach, J., Wydro, M.M., Lassmann, H., Franklin, R.J.M. and Mahad, D. (2010) 'Injury and differentiation following inhibition of mitochondrial respiratory chain complex IV in rat oligodendrocytes', *Glia*, 58(15), pp. 1827–1837. doi:10.1002/glia.21052.

Zimmermann, H.G., Knier, B., Oberwahrenbrock, T., Behrens, J., Pfuhl, C., Aly, L., Kaminski, M., Hoshi, M.-M., Specovius, S., Giess, R.M., Scheel, M., Mühlau, M., Bellmann-Strobl, J., Ruprecht, K., Hemmer, B., Korn, T., Paul, F. and Brandt, A.U. (2018) 'Association of Retinal Ganglion Cell Layer Thickness With Future Disease Activity in Patients With Clinically Isolated Syndrome', *JAMA Neurology*, 75(9), p. 1071. doi:10.1001/jamaneurol.2018.1011.

Zorova, L.D., Popkov, V.A., Plotnikov, E.Y., Silachev, D.N., Pevzner, I.B., Jankauskas, S.S., Babenko, V.A., Zorov, S.D., Balakireva, A.V., Juhaszova, M., Sollott, S.J. and Zorov, D.B. (2018) 'Mitochondrial membrane potential', *Analytical Biochemistry*, 552, pp. 50–59. doi:10.1016/j.ab.2017.07.009.

VIII — ANNEXES

Page intentionally left blank

Annex A — PrecisionCallerPipeline

Page intentionally left blank

Table S1 — Mixture analysis: Variant classification methodology

Minor component: primary analysis VL	Minor component: primary analysis variant class	Major component: primary analysis VL	Major component: primary analysis variant class	Mixture: VL	Mixture: micro	Mixture: meso	Mixture: macro
0	-	> 0	Grade A variant found	> 0	Major Grade A variant found (Ion Torrent and mixture)	Found variants	Found variants
0	-	> 0	Grade B variant found	> 0	Major Grade B variant found (Ion Torrent and mixture)	Found variants	Found variants
0	-	> 0	Grade C variant found	> 0	Major Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade A variant found	0	-	> 0	Minor Grade A variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade B variant found	0	-	> 0	Minor Grade B variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade C variant found	0	-	> 0	Minor Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade A variant found	> 0	Grade A variant found	> 0	Shared Grade A variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade A variant found	> 0	Grade B variant found	> 0	Shared Grade B variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade B variant found	> 0	Grade A variant found	> 0	Shared Grade B variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade B variant found	> 0	Grade B variant found	> 0	Shared Grade B variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade A variant found	> 0	Grade C variant found	> 0	Shared Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade B variant found	> 0	Grade C variant found	> 0	Shared Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade C variant found	> 0	Grade A variant found	> 0	Shared Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade C variant found	> 0	Grade B variant found	> 0	Shared Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
> 0	Grade C variant found	> 0	Grade C variant found	> 0	Shared Grade C variant found (Ion Torrent and mixture)	Found variants	Found variants
0	-	0	Grade A variant lost	> 0	Major Grade A variant resurfaced (mixture)	Mixture found variants	Found variants
0	-	0	Grade B variant lost	> 0	Major Grade B variant resurfaced (mixture)	Mixture found variants	Found variants
0	-	0	Grade C variant lost	> 0	Major Grade C variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade A variant lost	0	-	> 0	Minor Grade A variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade B variant lost	0	-	> 0	Minor Grade B variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade C variant lost	0	-	> 0	Minor Grade C variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade A variant lost	0	Grade A variant lost	> 0	Shared Grade A variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade A variant lost	0	Grade B variant lost	> 0	Shared Grade B variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade B variant lost	0	Grade A variant lost	> 0	Shared Grade B variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade B variant lost	0	Grade B variant lost	> 0	Shared Grade B variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade A variant lost	0	Grade C variant lost	> 0	Shared Grade C variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade B variant lost	0	Grade C variant lost	> 0	Shared Grade C variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade C variant lost	0	Grade A variant lost	> 0	Shared Grade C variant resurfaced (mixture)	Mixture found variants	Found variants

Table S1 — Mixture analysis: Variant classification methodology (continued)

Minor component: primary analysis VL	Minor component: primary analysis variant class	Major component: primary analysis VL	Major component: primary analysis variant class	Mixture: VL	Mixture: micro	Mixture: meso	Mixture: macro
0	Grade C variant lost	0	Grade B variant lost	> 0	Shared Grade C variant resurfaced (mixture)	Mixture found variants	Found variants
0	Grade C variant lost	0	Grade C variant lost	> 0	Shared Grade C variant resurfaced (mixture)	Mixture found variants	Found variants
0	-	> 0	Grade A variant found	0	Major Grade A variant lost (mixture)	Mixture lost variants	Lost variants
0	-	> 0	Grade B variant found	0	Major Grade B variant lost (mixture)	Mixture lost variants	Lost variants
0	-	> 0	Grade C variant found	0	Major Grade C variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade A variant found	0	-	0	Minor Grade A variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade B variant found	0	-	0	Minor Grade B variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade C variant found	0	-	0	Minor Grade C variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade A variant found	> 0	Grade A variant found	0	Shared Grade A variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade A variant found	> 0	Grade B variant found	0	Shared Grade B variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade B variant found	> 0	Grade A variant found	0	Shared Grade B variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade B variant found	> 0	Grade B variant found	0	Shared Grade B variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade A variant found	> 0	Grade C variant found	0	Shared Grade C variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade B variant found	> 0	Grade C variant found	0	Shared Grade C variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade C variant found	> 0	Grade A variant found	0	Shared Grade C variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade C variant found	> 0	Grade B variant found	0	Shared Grade C variant lost (mixture)	Mixture lost variants	Lost variants
> 0	Grade C variant found	> 0	Grade C variant found	0	Shared Grade C variant lost (mixture)	Mixture lost variants	Lost variants
0	-	0	Grade A variant lost	0	Major Grade A variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	-	0	Grade B variant lost	0	Major Grade B variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	-	0	Grade C variant lost	0	Major Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade A variant lost	0	-	0	Minor Grade A variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade B variant lost	0	-	0	Minor Grade B variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade C variant lost	0	-	0	Minor Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade A variant lost	0	Grade A variant lost	0	Shared Grade A variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade A variant lost	0	Grade B variant lost	0	Shared Grade B variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade B variant lost	0	Grade A variant lost	0	Shared Grade B variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade B variant lost	0	Grade B variant lost	0	Shared Grade B variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade A variant lost	0	Grade C variant lost	0	Shared Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants

Table S1 — Mixture analysis: Variant classification methodology (continued)

Minor component: primary analysis VL	Minor component: primary analysis variant class	Major component: primary analysis VL	Major component: primary analysis variant class	Mixture: VL	Mixture: micro	Mixture: meso	Mixture: macro
0	Grade B variant lost	0	Grade C variant lost	0	Shared Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade C variant lost	0	Grade A variant lost	0	Shared Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade C variant lost	0	Grade B variant lost	0	Shared Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	Grade C variant lost	0	Grade C variant lost	0	Shared Grade C variant lost (Ion Torrent and mixture)	Old lost variants	Lost variants
0	-	> 0	Novel variant	> 0	Major novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Novel variant	0	-	> 0	Minor novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Grade A variant found	> 0	Novel variant	> 0	Shared novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Grade B variant found	> 0	Novel variant	> 0	Shared novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Grade C variant found	> 0	Novel variant	> 0	Shared novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Novel variant	> 0	Grade A variant found	> 0	Shared novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Novel variant	> 0	Grade B variant found	> 0	Shared novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Novel variant	> 0	Grade C variant found	> 0	Shared novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
> 0	Novel variant	> 0	Novel variant	> 0	Shared true novel variant found (Ion Torrent and mixture)	Found variants	Novel variants
0	-	0	-	> 0	Novel variant: mixture	Novel variants	Novel variants
0	-	> 0	Novel variant	0	Major novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Novel variant	0	-	0	Minor novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Grade A variant found	> 0	Novel variant	0	Shared novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Grade B variant found	> 0	Novel variant	0	Shared novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Grade C variant found	> 0	Novel variant	0	Shared novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Novel variant	> 0	Grade A variant found	0	Shared novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Novel variant	> 0	Grade B variant found	0	Shared novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Novel variant	> 0	Grade C variant found	0	Shared novel variant lost (mixture)	Mixture lost variants	Primary novel variants
> 0	Novel variant	> 0	Novel variant	0	Shared true novel variant lost (mixture)	Mixture lost variants	Primary novel variants
0	Grade A variant lost	> 0	Grade A variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade A variant lost	> 0	Grade B variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade A variant lost	> 0	Grade C variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade A variant lost	> 0	Novel variant	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade B variant lost	> 0	Grade A variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status

Table S1 — Mixture analysis: Variant classification methodology (continued)

Minor component: primary analysis VL	Minor component: primary analysis variant class	Major component: primary analysis VL	Major component: primary analysis variant class	Mixture: VL	Mixture: micro	Mixture: meso	Mixture: macro
0	Grade B variant lost	> 0	Grade B variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade B variant lost	> 0	Grade C variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade B variant lost	> 0	Novel variant	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade C variant lost	> 0	Grade A variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade C variant lost	> 0	Grade B variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade C variant lost	> 0	Grade C variant found	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade C variant lost	> 0	Novel variant	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade A variant found	0	Grade A variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade A variant found	0	Grade B variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade A variant found	0	Grade C variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade B variant found	0	Grade A variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade B variant found	0	Grade B variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade B variant found	0	Grade C variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade C variant found	0	Grade A variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade C variant found	0	Grade B variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Grade C variant found	0	Grade C variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Novel variant	0	Grade A variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Novel variant	0	Grade B variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
> 0	Novel variant	0	Grade C variant lost	> 0	Shared variant found: mixed classification	Shared mixed found variants	Mixed status
0	Grade A variant lost	> 0	Grade A variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade A variant lost	> 0	Grade B variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade A variant lost	> 0	Grade C variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade A variant lost	> 0	Novel variant	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade B variant lost	> 0	Grade A variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade B variant lost	> 0	Grade B variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade B variant lost	> 0	Grade C variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade B variant lost	> 0	Novel variant	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade C variant lost	> 0	Grade A variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status

Table S1 — Mixture analysis: Variant classification methodology (continued)

Minor component: primary analysis VL	Minor component: primary analysis variant class	Major component: primary analysis VL	Major component: primary analysis variant class	Mixture: VL	Mixture: micro	Mixture: meso	Mixture: macro
0	Grade C variant lost	> 0	Grade B variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade C variant lost	> 0	Grade C variant found	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	Grade C variant lost	> 0	Novel variant	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade A variant found	0	Grade A variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade A variant found	0	Grade B variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade A variant found	0	Grade C variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade B variant found	0	Grade A variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade B variant found	0	Grade B variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade B variant found	0	Grade C variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade C variant found	0	Grade A variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade C variant found	0	Grade B variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Grade C variant found	0	Grade C variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Novel variant	0	Grade A variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Novel variant	0	Grade B variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
> 0	Novel variant	0	Grade C variant lost	0	Shared variant lost: mixed classification	Shared mixed lost variants	Mixed status
0	-	0	-	0	Not possible	Not possible	Not possible

Abbreviations: VL — variant level.

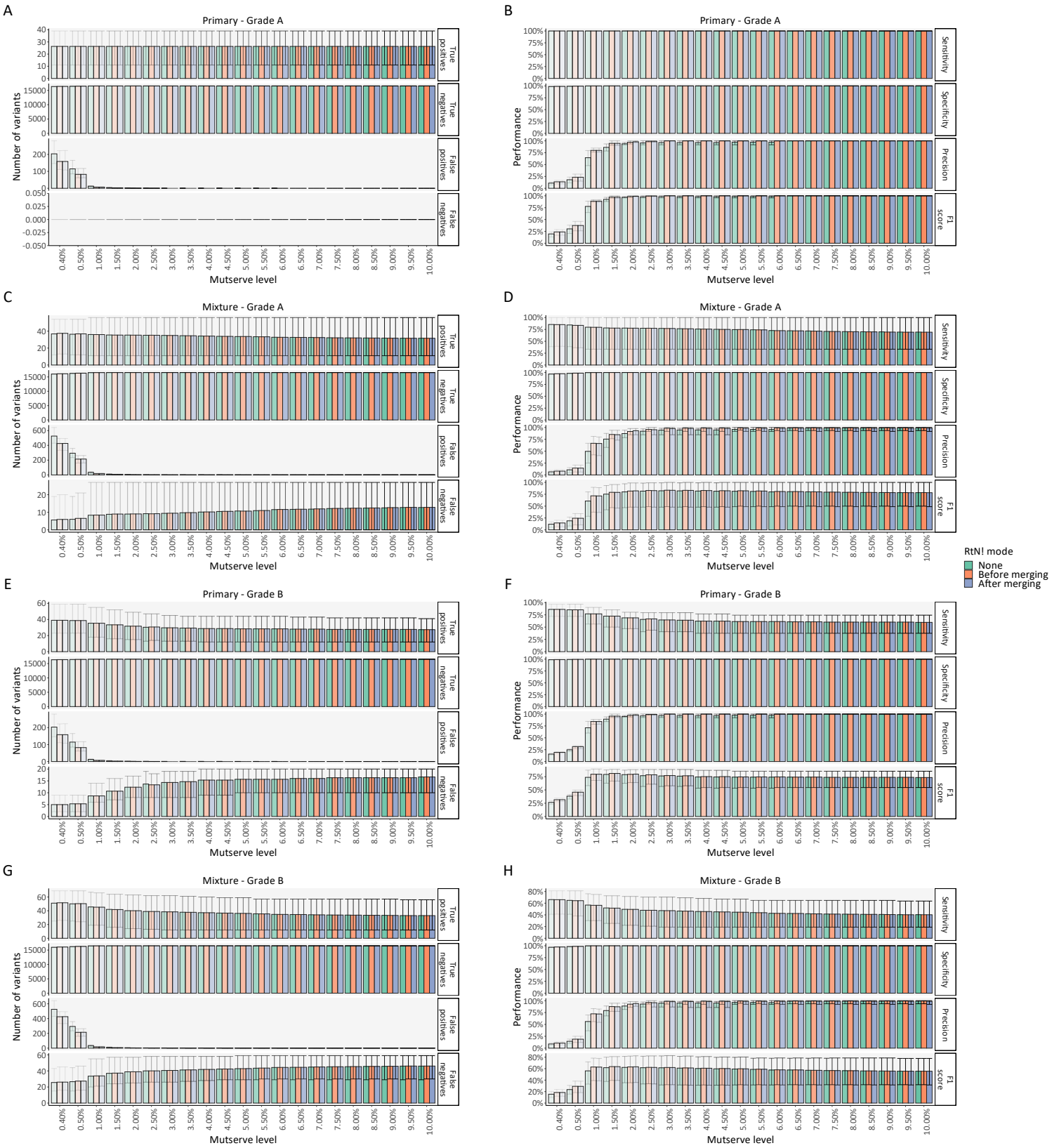


Figure S1 — Optimization of the bioinformatic processing

(**A, C, E, G**) Performance metric variant classification according to the approach to NUMT removal and heteroplasmy threshold for Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B datasets, respectively; (**B, D, F, H**) Performance metrics according to the approach to NUMT removal and heteroplasmy threshold for Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B datasets, respectively. Error bars denote minimum and maximum values. Abbreviations: NUMT — nuclear insertions of mitochondrial DNA.

Table S2 — Optimizing the approach to NUMT removal. Paired t-test with FDR

Performance metric	Grade	RtN! before merged - no RtN!			RtN! before merged - RtN! after merged		
		Estimate [95% CI]	*p*-value	*p*-value (adj)	Estimate [95% CI]	*p*-value	*p*-value (adj)
F_1 score	Grade A	1.63% [1.32–1.93%]	3.14E-22	8.36E-22	0.12% [0.07–0.16%]	2.40E-07	1.92E-06
	Grade B	1.03% [0.81–1.25%]	5.52E-18	1.10E-17	0.05% [0.03–0.07%]	5.36E-05	7.14E-05
Sensitivity	Grade A	-0.23% [-0.33–-0.13%]	7.33E-06	7.33E-06	0.01% [-0.01–0.02%]	3.18E-01	3.64E-01
	Grade B	-0.16% [-0.23–-0.10%]	1.40E-06	1.60E-06	0.00% [-0.02–0.01%]	6.38E-01	6.38E-01
Specificity	Grade A	0.06% [0.04–0.08%]	8.33E-11	1.11E-10	0.00% [0.00–0.00%]	5.96E-06	1.52E-05
	Grade B	0.06% [0.04–0.08%]	8.34E-11	1.11E-10	0.00% [0.00–0.00%]	5.95E-06	1.52E-05
Precision	Grade A	3.32% [2.82–3.82%]	1.73E-31	6.91E-31	0.50% [0.28–0.71%]	9.51E-06	1.52E-05
	Grade B	3.12% [2.66–3.58%]	2.94E-32	2.35E-31	0.46% [0.26–0.65%]	7.61E-06	1.52E-05

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval.

Table S3 — Optimizing the heteroplasmy threshold in mutserve, in datasets processed with RtN! before merging, with the F_1 score

Mutserve level	Primary — Grade A	Mixture — Grade A	Primary — Grade B	Mixture — Grade B	Mean
2.50%	99.58%	82.94%	78.56%	62.20%	80.82%
2.00%	98.47%	81.76%	79.41%	63.00%	80.66%
3.00%	100.00%	83.49%	77.08%	62.01%	80.65%
3.50%	100.00%	83.28%	76.72%	61.27%	80.32%
1.50%	97.26%	79.10%	81.08%	63.52%	80.24%
4.00%	100.00%	82.93%	75.10%	60.75%	79.70%
4.50%	100.00%	82.50%	75.10%	60.20%	79.45%
5.00%	100.00%	82.41%	74.55%	59.82%	79.20%
5.50%	100.00%	81.94%	74.55%	59.02%	78.88%
6.00%	100.00%	80.90%	74.55%	57.95%	78.35%
6.50%	100.00%	80.69%	74.18%	57.76%	78.16%
7.00%	100.00%	80.23%	74.18%	57.14%	77.89%
7.50%	100.00%	79.63%	73.80%	56.69%	77.53%
8.00%	100.00%	79.25%	73.80%	56.35%	77.35%
8.50%	100.00%	78.97%	73.80%	56.08%	77.21%
9.00%	100.00%	78.63%	73.80%	55.82%	77.06%
9.50%	100.00%	78.51%	73.80%	55.58%	76.97%
10.00%	100.00%	78.51%	73.41%	55.58%	76.87%
1.00%	88.78%	71.70%	79.61%	62.83%	75.73%
0.50%	37.41%	24.51%	45.78%	29.20%	34.23%
0.40%	23.85%	14.53%	31.54%	18.47%	22.10%

Table S4 — Primary analysis: Overall look

Sample	Sequencing run		Mean coverage	Mappability >30 reads/base	Mappability >100 reads/base	Mappability >1000 reads/base	Number of variants	Haplogroup	Haplogroup quality (%)
HG00256 (*H*)	exome		29.10	37.84%	0.00%	0.00%	12	H5b2	98.98%
	lowCov		4172.55	100.00%	100.00%	99.95%	42	H5b2	98.98%
	highCov		16370.49	100.00%	100.00%	99.99%	85	H5b2	95.80%
	Ion Torrent	PCP	932.28	99.29%	96.80%	42.07%	14	H5b2	98.98%
		TSS	1342.00	-	-	-	12	H5b2	98.98%
HG01626 (*T*)	exome		72.86	71.85%	11.86%	0.86%	42	T2a1b1a1b	97.07%
	lowCov		2641.51	100.00%	99.99%	98.91%	61	T2a1b1a1b	97.60%
	highCov		14422.54	100.00%	100.00%	99.99%	127	T2a1b1a1b	97.60%
	Ion Torrent	PCP	3196.55	99.98%	99.91%	87.91%	48	T2a1b1a1b	97.60%
		TSS	4256.00	-	-	-	40	T2a1b1a1b	94.94%
HG01757 (*U*)	exome		1159.80	99.99%	99.99%	58.89%	40	U4a	95.92%
	lowCov		4776.03	100.00%	100.00%	99.99%	41	U4a	95.92%
	highCov		14444.80	100.00%	100.00%	99.99%	95	U4a	94.33%
	Ion Torrent	PCP	1115.99	99.90%	95.59%	49.45%	31	U4a	95.92%
		TSS	1577.00	-	-	-	31	U4a	95.92%

Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

Table S5 — Ion Torrent sequencing: Overall look on differences between PCP and TSS

Sample	PCP							TSS						
	Haplogroup	Haplogroup quality (%)	Contamination	Contamination level	Number of sequences	Mean coverage	Number of variants	Haplogroup	Haplogroup quality (%)	Contamination	Contamination level	Number of sequences	Mean coverage	Number of variants
H	H5b2	98.98%	No	-	144178	932.28	14	H5b2	98.98%	No	-	184905	1342.00	12
U0.01H			No	-	177989	1163.89	14			No	-	214373	1567.00	13
U0.02H			No	-	245198	1596.93	14			No	-	292422	2131.00	13
U0.05H			No	-	245786	1636.51	14			No	-	297105	2206.00	13
U0.10H		96.53%	Yes	7.50%	248124	1612.49	35		96.53%	No	-	291568	2122.00	14
U0.25H	H5'36	98.57%	Yes	31.30%	265832	1730.69	33	H+195	100.00%	Yes	29.00%	318735	2325.00	36
T	T2a1b1a1b	97.60%	No	-	494101	3196.55	48	T2a1b1a1b	94.94%	Yes	0.80%	586187	4256.00	40
H0.01T			No	-	168591	1098.88	46			No	-	205371	1494.00	41
H0.02T			No	-	177652	1144.59	48			No	-	217075	1569.00	42
H0.05T			No	-	429697	2802.03	45			No	-	502798	3672.00	41
H0.10T			Yes	6.30%	329892	2158.23	53			No	-	384980	2825.00	43
H0.25T	H	100.00%	Yes	18.80%	360527	2347.62	51	H	100.00%	Yes	18.90%	428989	3141.00	47
U	U4a	95.92%	No	-	172011	1115.99	31	U4a	95.92%	Yes	0.60%	216643	1577.00	31
T0.01U			No	-	191343	1250.31	31			No	-	231015	1687.00	32
T0.02U		94.69%	No	-	163090	1062.47	31			No	-	196335	1432.00	30
T0.05U		93.99%	Yes	4.40%	211826	1373.27	59			No	-	253305	1835.00	31
T0.10U	R	94.91%	Yes	13.40%	255253	1656.45	60	R	100.00%	Yes	13.40%	308232	2240.00	50
T0.25U	R	98.75%	Yes	36.60%	258009	1676.07	62	R		Yes	36.60%	307354	2242.00	55

Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

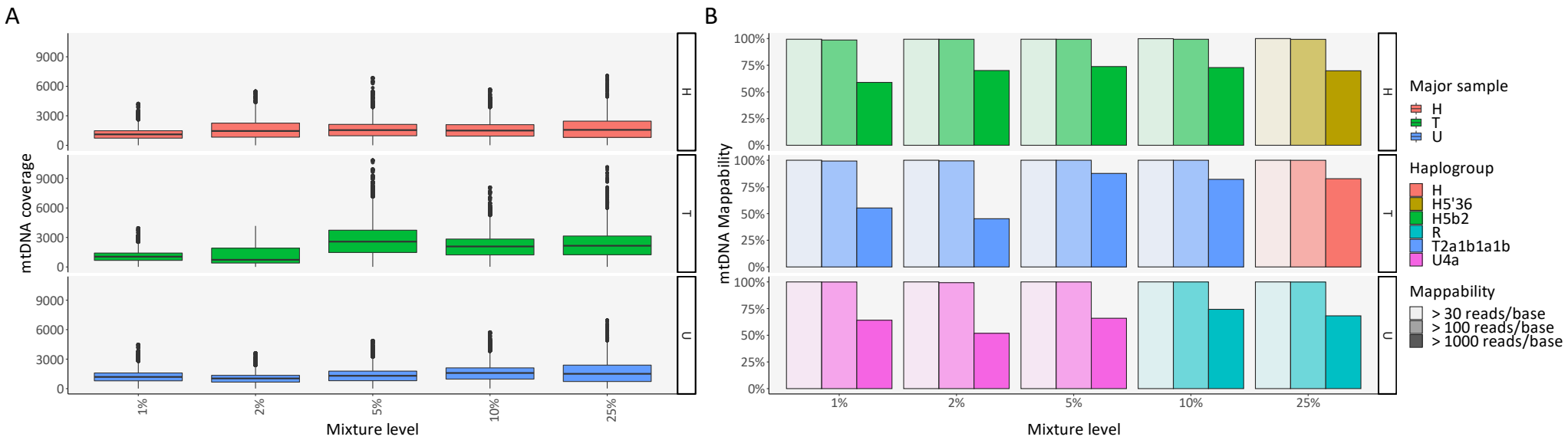


Figure S2 — Mixture analysis: Overall look with PCP, in mutserve

(**A**) Coverage per sample, (**B**) Mappability and haplogroup assignment per sample. Abbreviations: mtDNA — mitochondrial DNA; PCP — PrecisionCallerPipeline.

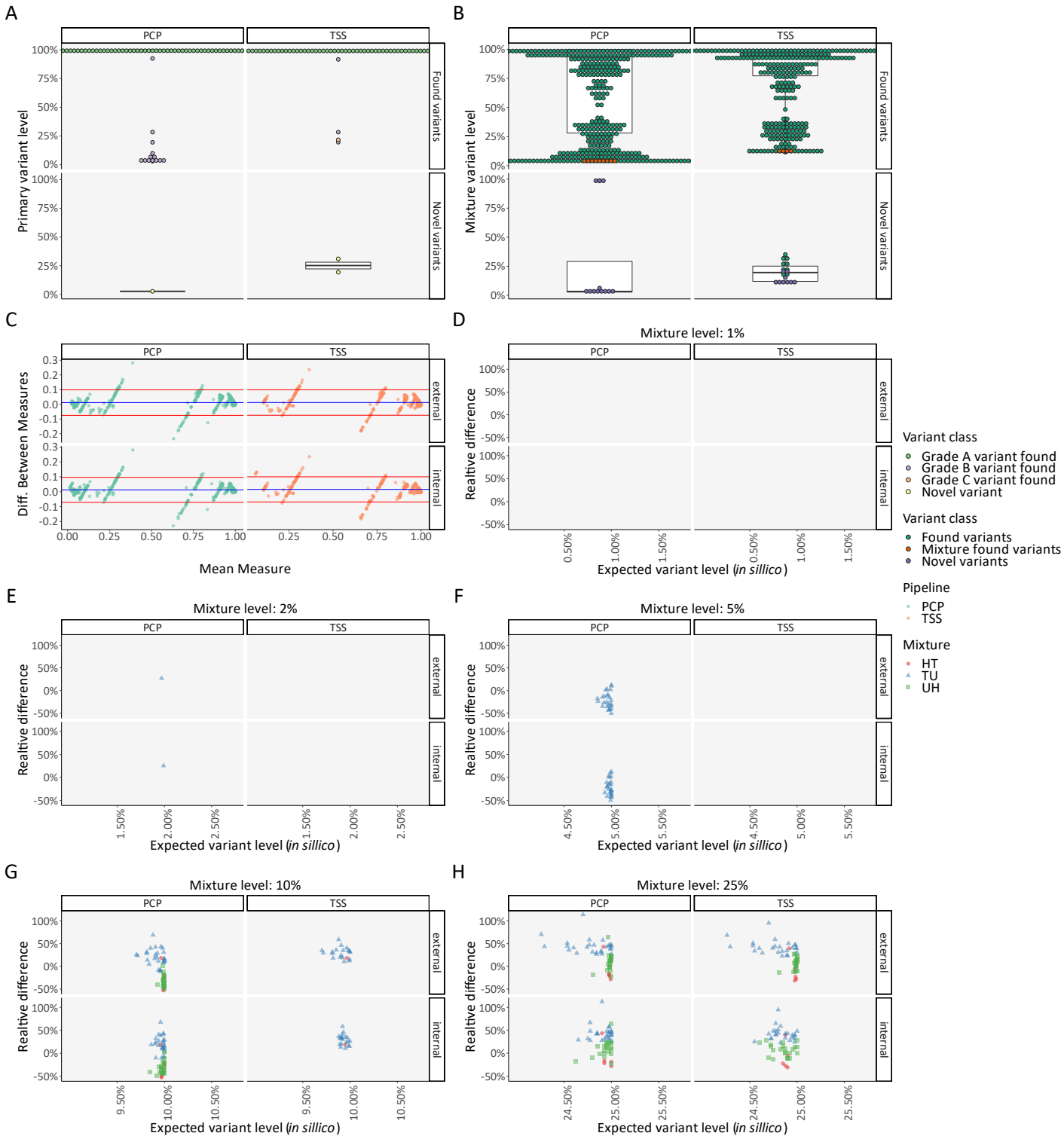


Figure S3 — Distribution of variant levels with mutserve, per pipeline

(**A**) Primary analysis — distribution of variant levels per variant classification, per pipeline, (**B**) Mixture analysis — distribution of variant levels per macro variant classification, per pipeline, (**C**) Bland–Altman plot for internal and external differences, per pipeline; (**D-H**) close-up of Figure 22A with the relative difference between expected and observed variant level for each internal and external comparison, per pipeline and per mixture, depending on mixture level: 1%, 2%, 5%, 10%, and 25%, respectively. Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

Table S6 — Primary analysis: Distribution of variants, depending on the VCM chosen

VCM	Variant description		PCP				TSS			
	Broad classification	Detailed classification	*N*	Mean observed VL [Min–Max]	Mean primary VL [Min–Max]	Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Mean primary VL [Min–Max]	Mean absolute difference in VLs [Min–Max]
freebayes	Grade A variant found	including exome	*43*	99.72% [98.56–100.00%]	99.56% [96.99–99.99%]	0.48% [0.00–2.71%]	*44*	99.41% [98.60–100.00%]	99.57% [96.99–99.99%]	0.68% [0.03–2.64%]
	Grade B variant found	including exome	*5*	26.10% [3.52–92.55%]	28.59% [2.57–94.75%]	2.88% [0.95–7.42%]	*2*	55.65% [19.60–91.70%]	58.56% [22.37–94.75%]	2.91% [2.77–3.05%]
		excluding exome	*7*	7.72% [2.67–28.39%]	6.86% [1.12–28.31%]	0.87% [0.03–2.06%]	*1*	28.30%	28.31%	0.01%
		from highCov and exome	*3*	99.94% [99.83–100.00%]	99.97% [99.94–99.99%]	0.08% [0.02–0.16%]	*3*	99.47% [99.30–99.60%]	99.97% [99.94–99.99%]	0.51% [0.34–0.69%]
		from lowCov and exome	*21*	99.74% [98.46–100.00%]	99.30% [95.83–100.00%]	0.75% [0.00–3.88%]	*21*	99.50% [98.80–100.00%]	99.30% [95.83–100.00%]	0.87% [0.00–4.17%]
	Grade C variant found	from exome	*4*	99.78% [99.56–100.00%]	100.00% [100.00–100.00%]	0.22% [0.00–0.44%]	*5*	99.62% [99.40–99.90%]	99.96% [99.79–100.00%]	0.38% [0.11–0.60%]
	Grade A variant lost	including exome	*1*	-	99.99%	-	-	-	-	-
	Grade B variant lost	including exome	*5*	-	0.83% [0.54–1.27%]	-	*8*	-	3.75% [0.54–17.37%]	-
		excluding exome	*32*	-	1.19% [0.43–5.75%]	-	*38*	-	1.52% [0.43–6.13%]	-
		from lowCov and exome	*7*	-	2.03% [0.44–4.59%]	-	*7*	-	2.03% [0.44–4.59%]	-
	Grade C variant lost	from highCov	*210*	-	0.76% [0.40–2.52%]	-	*210*	-	0.76% [0.40–2.52%]	-
		from lowCov	*19*	-	0.88% [0.40–2.35%]	-	*19*	-	0.88% [0.40–2.35%]	-
		from exome	*10*	-	10.54% [0.43–99.79%]	-	*9*	-	0.63% [0.43–1.04%]	-
	Novel variant	only present in IonTorrent	*3*	4.57% [2.52–8.44%]	-	-	*7*	66.87% [19.30–99.90%]	-	-

Table S6 — Primary analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description		PCP				TSS			
	Broad classification	Detailed classification	*N*	Mean observed VL [Min–Max]	Mean primary VL [Min–Max]	Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Mean primary VL [Min–Max]	Mean absolute difference in VLs [Min–Max]
mutserve	Grade A variant found	including exome	*79*	99.73% [98.50–100.00%]	99.61% [97.03–100.00%]	0.44% [0.00–2.73%]	*77*	99.43% [98.20–100.00%]	99.60% [97.03–100.00%]	0.65% [0.00–2.77%]
	Grade B variant found	including exome	*5*	26.06% [3.50–92.50%]	28.59% [2.57–94.77%]	2.91% [0.93–7.47%]	*2*	55.65% [19.60–91.70%]	58.57% [22.37–94.77%]	2.92% [2.77–3.07%]
		excluding exome	*8*	7.07% [2.50–28.40%]	6.36% [1.10–28.35%]	0.79% [0.05–2.10%]	*1*	28.30%	28.35%	0.05%
	Grade C variant found	from highCov	-	-	-	-	*1*	21.70%	0.40%	21.30%
	Grade A variant lost	including exome	-	-	-	-	*2*	-	100.00% [100.00–100.00%]	-
	Grade B variant lost	including exome	*4*	-	0.88% [0.50–1.27%]	-	*7*	-	4.19% [0.50–17.37%]	-
		excluding exome	*34*	-	1.10% [0.40–5.75%]	-	*41*	-	1.46% [0.40–6.20%]	-
		from lowCov and exome	*2*	-	0.48% [0.45–0.50%]	-	*2*	-	0.48% [0.45–0.50%]	-
	Grade C variant lost	from highCov	*177*	-	0.82% [0.40–5.80%]	-	*176*	-	0.83% [0.40–5.80%]	-
		from lowCov	*12*	-	0.93% [0.40–3.80%]	-	*12*	-	0.93% [0.40–3.80%]	-
		from exome	*4*	-	0.70% [0.40–1.00%]	-	*4*	-	0.70% [0.40–1.00%]	-
	Novel variant	only present in IonTorrent	*1*	2.70%	-	-	*2*	25.15% [19.30–31.00%]	-	-
varscan	Grade A variant found	including exome	*74*	99.75% [98.46–100.00%]	99.62% [97.07–100.00%]	0.44% [0.00–2.78%]	*77*	99.43% [98.20–100.00%]	99.62% [97.07–100.00%]	0.66% [0.00–2.78%]
	Grade B variant found	including exome	*5*	26.10% [3.52–92.55%]	28.61% [2.58–94.80%]	2.89% [0.94–7.43%]	*2*	55.65% [19.60–91.70%]	58.59% [22.38–94.80%]	2.94% [2.78–3.10%]
		excluding exome	*7*	7.73% [2.67–28.41%]	6.88% [1.11–28.38%]	0.87% [0.03–2.08%]	*1*	28.30%	28.38%	0.08%

Table S6 — Primary analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description		PCP				TSS			
	Broad classification	Detailed classification	*N*	Mean observed VL [Min–Max]	Mean primary VL [Min–Max]	Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Mean primary VL [Min–Max]	Mean absolute difference in VLs [Min–Max]
	Grade A variant lost	including exome	*5*	-	99.79% [99.29–99.94%]	-	*2*	-	99.99% [99.98–100.00%]	-
	Grade B variant lost	including exome	*2*	-	0.97% [0.68–1.27%]	-	*5*	-	5.56% [0.68–17.38%]	-
		excluding exome	*31*	-	1.24% [0.46–5.75%]	-	*37*	-	1.57% [0.46–6.20%]	-
	Grade C variant lost	from highCov	*66*	-	0.72% [0.41–2.71%]	-	*66*	-	0.72% [0.41–2.71%]	-
		from lowCov	*7*	-	0.52% [0.40–0.90%]	-	*7*	-	0.52% [0.40–0.90%]	-
		from exome	*2*	-	0.94% [0.84–1.04%]	-	*2*	-	0.94% [0.84–1.04%]	-
	Novel variant	Only present in IonTorrent	*1*	2.75%	-	-	*3*	24.00% [19.30–31.00%]	-	-

Abbreviations: PCP — PrecisionCallerPipeline; Max — maximum; Min — minimum; *N* — number; TSS — Ion Torrent Suite™ Software; VCM — variant calling method; VL — variant level.

Table S7 — Primary analysis: Distribution of Grade A/B variants, according to the VCM chosen. McNemar's test with FDR

VCM	Variants found with PCP and TSS	Variants found with PCP & lost with TSS	Variants lost with PCP & found with TSS	Variants lost with PCP and TSS	*p*-value	*p*-value (adj)
freebayes	70	9	1	44	2.69E-02	4.03E-02
mutserve	80	12	0	40	1.50E-03	4.49E-03
varscan	75	11	5	33	2.11E-01	2.11E-01

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S8 — Primary analysis: Paired difference between the observed VL in Ion Torrent and the expected VL in Grade A/B variants with other sequencing methods, according to the VCM chosen. Bootstrap with FDR

VCM	Mean TSS-PCP paired difference [95% CI]	*p*-value	*p*-value (adj)
freebayes	0.33% [0.12–0.50%]	0.00E+00	0.00E+00
mutserve	0.30% [0.13–0.41%]	0.00E+00	0.00E+00
varscan	0.32% [0.14–0.40%]	0.00E+00	0.00E+00

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method; VL — variant level.

Table S9 — Primary analysis: Paired difference between the proportion of "Novel variants", depending on the VCM chosen. Bootstrap with FDR

VCM	Mean TSS-PCP paired difference [95% CI]	*p*-value	*p*-value (adj)
freebayes	1.05% [0.00–2.44%]	7.47E-02	2.24E-01
mutserve	0.38% [-0.78–1.92%]	5.29E-01	5.29E-01
varscan	1.04% [0.00–3.12%]	5.05E-01	5.29E-01

Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
freebayes	Found variants	Found variants	Major grade A	*160*	92.54% [56.88–100.00%]	3.54% [0.06–18.12%]	3.81% [0.02–18.08%]	*165*	92.21% [56.80–100.00%]	3.46% [0.11–18.20%]	3.66% [0.07–18.16%]
			Major grade B	*99*	64.69% [2.60–100.00%]	2.87% [0.10–22.99%]	3.38% [0.02–23.53%]	*80*	83.39% [11.80–100.00%]	3.19% [0.14–17.10%]	3.42% [0.02–17.33%]
			Major grade C	*18*	91.91% [60.41–99.84%]	3.51% [0.57–14.59%]	3.36% [0.24–14.59%]	*15*	90.23% [57.60–99.60%]	3.68% [0.50–17.17%]	3.43% [0.20–17.40%]
			Minor grade A	*78*	17.64% [2.99–52.77%]	4.50% [0.03–27.86%]	4.55% [0.01–28.07%]	*47*	25.41% [11.00–48.30%]	5.60% [0.00–23.50%]	5.64% [0.01–23.60%]
			Minor grade B	*38*	14.75% [2.82–39.27%]	4.03% [0.15–14.29%]	4.04% [0.11–14.40%]	*19*	24.80% [11.90–37.50%]	6.17% [1.99–12.70%]	6.18% [2.05–12.62%]
			Minor grade C	*6*	16.18% [5.04–35.66%]	5.48% [1.86–10.77%]	5.47% [1.82–10.66%]	*5*	23.26% [11.80–34.80%]	4.61% [1.85–9.92%]	4.63% [1.80–9.80%]
			Shared grade A	*30*	99.82% [99.15–100.00%]	0.10% [0.00–0.36%]	0.15% [0.00–0.78%]	*30*	99.36% [98.60–100.00%]	0.28% [0.00–1.30%]	0.60% [0.02–1.33%]
			Shared grade B	*75*	99.93% [99.62–100.00%]	0.08% [0.00–0.40%]	0.08% [0.00–0.35%]	*75*	99.68% [98.60–100.00%]	0.18% [0.00–0.69%]	0.31% [0.00–1.32%]
			Shared grade C	*5*	99.70% [99.64–99.75%]	0.04% [0.00–0.08%]	3.51% [2.79–3.80%]	*10*	99.89% [99.60–100.00%]	0.06% [0.00–0.29%]	1.93% [0.06–4.08%]

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
		Mixture found variants	Major grade B	*9*	4.08% [2.58–5.68%]	-	0.44% [0.02–1.67%]	*4*	12.60% [11.90–13.20%]	-	3.12% [0.83–4.00%]
			Shared grade C	*13*	14.36% [3.95–99.80%]	-	5.77% [0.01–16.43%]	-	-	-	-
	Lost variants	Mixture lost variants	Major grade B	*26*	-	19.63% [2.00–97.48%]	19.09% [0.84–98.91%]	-	-	-	-
			Major grade C	*2*	-	82.37% [74.88–89.86%]	82.50% [75.00–90.00%]	-	-	-	-
			Minor grade A	*82*	-	2.26% [0.99–5.00%]	2.26% [0.97–5.00%]	*113*	-	3.75% [0.99–10.00%]	3.76% [0.97–10.00%]
			Minor grade B	*92*	-	1.93% [0.03–24.93%]	1.91% [0.01–24.98%]	*61*	-	3.61% [0.20–10.00%]	3.64% [0.22–10.00%]
			Minor grade C	*9*	-	2.66% [1.00–5.00%]	2.67% [1.00–5.00%]	*15*	-	4.12% [0.99–9.99%]	4.13% [1.00–10.00%]
		Old lost variants	Major grade A	*5*	-	-	91.39% [74.99–98.99%]	-	-	-	-
			Major grade B	*186*	-	-	0.89% [0.32–2.67%]	*236*	-	-	1.47% [0.32–16.50%]
			Major grade C	*480*	-	-	0.57% [0.30–1.81%]	*480*	-	-	0.57% [0.30–1.81%]
			Minor grade B	*190*	-	-	0.10% [0.00–1.44%]	*235*	-	-	0.16% [0.00–4.34%]
			Minor grade C	*490*	-	-	0.14% [0.00–24.95%]	*485*	-	-	0.05% [0.00–0.46%]
			Shared grade B	*5*	-	-	0.65% [0.56–0.83%]	*5*	-	-	0.65% [0.56–0.83%]

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
			Shared grade C	*722*	-	-	1.47% [0.41–99.84%]	*730*	-	-	0.92% [0.41–4.57%]
	Novel variants	Found variants	Major novel	*2*	6.61% [6.45–6.77%]	0.68% [0.12–1.25%]	-	*20*	74.87% [22.30–99.80%]	3.94% [0.10–13.49%]	-
			Minor novel	*3*	3.57% [2.76–4.44%]	2.50% [2.33–2.65%]	-	*8*	24.80% [16.80–31.00%]	13.55% [5.90–24.07%]	-
			Shared novel	-	-	-	-	*10*	99.87% [99.60–100.00%]	0.25% [0.00–0.40%]	-
			Shared true novel	-	-	-	-	*4*	31.65% [28.00–35.10%]	12.10% [8.10–15.78%]	-
		Novel variants	Novel variant: mixture	*24*	4.85% [2.53–13.97%]	-	-	*9*	13.92% [10.60–21.10%]	-	-
	Primary novel variants	Mixture lost variants	Major novel	*13*	-	3.71% [1.89–8.36%]	-	*5*	-	28.33% [23.25–30.69%]	-
			Minor novel	*12*	-	0.22% [0.03–0.69%]	-	*17*	-	3.48% [0.19–9.99%]	-
			Shared true novel	-	-	-	-	*1*	-	19.35%	-
mutserve	Found variants	Found variants	Major grade A	*270*	92.51% [51.50–100.00%]	3.51% [0.00–22.90%]	3.73% [0.00–23.42%]	*260*	92.10% [56.80–100.00%]	3.40% [0.10–18.20%]	3.55% [0.00–18.15%]
			Major grade B	*39*	20.82% [2.60–94.30%]	1.73% [0.16–6.32%]	2.09% [0.29–6.10%]	*15*	43.18% [11.80–94.20%]	2.26% [0.14–6.29%]	2.25% [0.08–4.98%]

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
			Major grade C	-	-	-	-	*5*	28.48% [22.30–34.10%]	8.65% [1.03–13.49%]	28.11% [21.91–33.72%]
			Minor grade A	*133*	16.95% [2.50–52.90%]	4.38% [0.02–28.00%]	4.41% [0.06–28.20%]	*73*	25.31% [11.00–48.30%]	5.73% [0.00–23.50%]	5.75% [0.02–23.60%]
			Minor grade B	*8*	7.54% [3.30–26.00%]	1.90% [0.32–3.95%]	1.72% [0.14–4.18%]	*1*	26.30%	3.38%	2.61%
			Shared grade A	*125*	99.89% [99.20–100.00%]	0.09% [0.00–0.40%]	0.20% [0.00–2.41%]	*125*	99.64% [98.60–100.00%]	0.20% [0.00–1.30%]	0.45% [0.00–2.71%]
		Mixture found variants	Major grade B	*9*	4.08% [2.60–5.70%]	-	0.44% [0.03–1.69%]	*4*	12.60% [11.90–13.20%]	-	3.12% [0.83–3.99%]
	Lost variants	Mixture lost variants	Major grade B	*26*	-	2.77% [1.88–3.56%]	2.00% [0.83–2.72%]	-	-	-	-
			Minor grade A	*137*	-	2.44% [0.98–24.80%]	2.44% [0.97–24.98%]	*187*	-	3.86% [0.98–10.00%]	3.87% [0.97–10.00%]
			Minor grade B	*57*	-	0.48% [0.03–4.63%]	0.48% [0.01–4.74%]	*14*	-	2.65% [0.20–9.17%]	2.78% [0.22–9.48%]
		Old lost variants	Major grade A	-	-	-	-	*10*	-	-	91.40% [75.00–99.00%]
			Major grade B	*186*	-	-	0.81% [0.30–2.70%]	*241*	-	-	1.44% [0.30–16.50%]
			Major grade C	*430*	-	-	0.72% [0.30–5.74%]	*425*	-	-	0.73% [0.30–5.74%]
			Minor grade A	-	-	-	-	*10*	-	-	8.60% [1.00–25.00%]

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
			Minor grade B	*195*	-	-	0.09% [0.00–1.44%]	*245*	-	-	0.16% [0.00–4.34%]
			Minor grade C	*430*	-	-	0.06% [0.00–0.95%]	*425*	-	-	0.06% [0.00–0.95%]
			Shared grade C	*540*	-	-	0.86% [0.40–2.69%]	*540*	-	-	0.86% [0.40–2.69%]
	Novel variants	Found variants	Minor novel	-	-	-	-	*5*	21.24% [16.80–26.00%]	18.07% [11.98–24.07%]	-
			Shared novel	-	-	-	-	*4*	31.65% [28.00–35.10%]	12.10% [8.10–15.78%]	-
		Novel variants	Novel variant: mixture	*12*	27.26% [2.50–99.20%]	-	-	*9*	13.92% [10.60–21.10%]	-	-
	Primary novel variants	Mixture lost variants	Major novel	*5*	-	2.47% [2.02–2.67%]	-	*5*	-	28.33% [23.25–30.69%]	-
			Minor novel	*5*	-	0.23% [0.03–0.68%]	-	*5*	-	1.15% [0.19–3.10%]	-
			Shared novel	-	-	-	-	*1*	-	19.35%	-
varscan	Found variants	Found variants	Major grade A	*250*	92.67% [56.88–100.00%]	3.42% [0.06–18.12%]	3.64% [0.01–18.08%]	*260*	92.10% [56.80–100.00%]	3.40% [0.10–18.20%]	3.53% [0.01–18.17%]
			Major grade B	*39*	20.82% [2.60–94.34%]	1.72% [0.10–6.28%]	2.09% [0.31–6.15%]	*15*	43.18% [11.80–94.20%]	2.26% [0.14–6.29%]	2.26% [0.10–4.99%]
			Minor grade A	*123*	17.24% [2.82–52.77%]	4.44% [0.03–27.86%]	4.48% [0.01–28.07%]	*73*	25.31% [11.00–48.30%]	5.73% [0.00–23.50%]	5.74% [0.01–23.60%]

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
			Minor grade B	*8*	7.54% [3.27–26.01%]	1.90% [0.30–3.95%]	1.72% [0.12–4.18%]	*1*	26.30%	3.38%	2.60%
			Shared grade A	*120*	99.92% [99.15–100.00%]	0.09% [0.00–0.40%]	0.19% [0.00–2.75%]	*125*	99.64% [98.60–100.00%]	0.20% [0.00–1.30%]	0.45% [0.00–2.73%]
		Mixture found variants	Major grade A	*3*	98.06% [94.77–100.00%]	-	4.07% [1.71–5.41%]	-	-	-	-
			Major grade B	*9*	4.08% [2.58–5.68%]	-	0.44% [0.02–1.67%]	*4*	12.60% [11.90–13.20%]	-	3.13% [0.84–4.01%]
			Minor grade A	*2*	14.83% [8.84–20.83%]	-	2.65% [1.16–4.15%]	-	-	-	-
			Shared grade A	*5*	100.00% [100.00–100.00%]	-	0.12% [0.07–0.22%]	-	-	-	-
	Lost variants	Mixture lost variants	Major grade B	*21*	-	2.88% [2.00–3.58%]	1.89% [0.83–2.71%]	-	-	-	-
			Minor grade A	*127*	-	2.39% [0.98–24.93%]	2.39% [0.97–24.90%]	*187*	-	3.86% [0.98–10.00%]	3.87% [0.97–10.00%]
			Minor grade B	*52*	-	0.51% [0.03–4.63%]	0.51% [0.01–4.74%]	*14*	-	2.65% [0.20–9.17%]	2.78% [0.22–9.48%]
		Old lost variants	Major grade A	*17*	-	-	90.67% [74.47–98.92%]	*10*	-	-	91.39% [74.99–99.00%]
			Major grade B	*156*	-	-	0.94% [0.34–2.70%]	*206*	-	-	1.60% [0.34–16.52%]
			Major grade C	*205*	-	-	0.55% [0.30–2.36%]	*205*	-	-	0.55% [0.30–2.36%]

Table S10 — Mixture analysis: Distribution of variants, depending on the VCM chosen (continued)

VCM	Variant description			PCP				TSS			
	Macro	Meso	Micro	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]	*N*	Mean observed VL [Min–Max]	Within platform: Mean absolute difference in VLs [Min–Max]	Other platforms: Mean absolute difference in VLs [Min–Max]
			Minor grade A	*18*	-	-	7.60% [1.00–24.98%]	*10*	-	-	8.60% [1.00–25.00%]
			Minor grade B	*165*	-	-	0.10% [0.00–1.44%]	*210*	-	-	0.18% [0.00–4.35%]
			Minor grade C	*205*	-	-	0.05% [0.00–0.39%]	*205*	-	-	0.05% [0.00–0.39%]
			Shared grade C	*170*	-	-	0.84% [0.42–2.70%]	*170*	-	-	0.84% [0.42–2.70%]
	Novel variants	Found variants	Major novel	-	-	-	-	*5*	28.48% [22.30–34.10%]	8.65% [1.03–13.49%]	-
			Minor novel	-	-	-	-	*5*	21.24% [16.80–26.00%]	18.07% [11.98–24.07%]	-
			Shared true novel	-	-	-	-	*4*	31.65% [28.00–35.10%]	12.10% [8.10–15.78%]	-
		Novel variants	Novel variant: mixture	*10*	3.39% [2.53–6.25%]	-	-	*9*	13.92% [10.60–21.10%]	-	-
	Primary novel variants	Mixture lost variants	Major novel	*5*	-	2.51% [2.06–2.72%]	-	*5*	-	28.33% [23.25–30.69%]	-
			Minor novel	*5*	-	0.24% [0.03–0.69%]	-	*5*	-	1.15% [0.19–3.10%]	-
			Shared true novel	-	-	-	-	*1*	-	19.35%	-

Abbreviations: PCP — PrecisionCallerPipeline; Max — maximum; Min — minimum; *N* — number; TSS — Ion Torrent Suite™ Software; VCM — variant calling method; VL — variant level.

Table S11 — Mixture analysis: Distribution of Grade A/B variants, according to the VCM chosen. McNemar's test with FDR

VCM	Variants found with PCP and TSS	Variants found with PCP & lost with TSS	Variants lost with PCP & found with TSS	Variants lost with PCP and TSS	*p*-value	*p*-value (adj)
freebayes	409	80	11	570	1.02E-12	1.02E-12
mutserve	477	107	1	600	5.32E-24	1.60E-23
varscan	457	102	21	535	5.46E-13	8.19E-13

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S12 — Mixture analysis: Paired difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
Found variants	freebayes	internal (within plaftorm)	mean(PCP - expected VL)	1.20% [0.39–2.02%]	6.89E-03	1.13E-02
			mean(TSS - expected VL)	1.55% [0.87–2.22%]	2.27E-04	8.16E-04
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.35% [-0.15–0.84%]	1.55E-01	1.99E-01
		external (other plaftorms)	mean(PCP - expected VL)	1.07% [0.22–1.93%]	1.77E-02	2.65E-02
			mean(TSS - expected VL)	1.14% [0.38–1.90%]	6.08E-03	1.11E-02
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.07% [-0.37–0.51%]	7.42E-01	7.86E-01

Table S12 — Mixture analysis: Paired difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR (continued)

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
	mutserve	internal (within plaftorm)	mean(PCP - expected VL)	1.15% [0.38–1.92%]	6.56E-03	1.12E-02
			mean(TSS - expected VL)	1.46% [0.82–2.10%]	2.26E-04	8.16E-04
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.31% [-0.17–0.79%]	1.88E-01	2.33E-01
		external (other plaftorms)	mean(PCP - expected VL)	1.02% [0.21–1.82%]	1.68E-02	2.63E-02
			mean(TSS - expected VL)	1.07% [0.37–1.77%]	5.35E-03	1.11E-02
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.06% [-0.38–0.50%]	7.75E-01	7.97E-01
	varscan	internal (within plaftorm)	mean(PCP - expected VL)	2.46% [1.42–3.50%]	1.75E-04	8.16E-04
			mean(TSS - expected VL)	1.46% [0.82–2.10%]	2.26E-04	8.16E-04
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	-1.00% [-1.93–-0.06%]	3.81E-02	5.08E-02
		external (other plaftorms)	mean(PCP - expected VL)	1.03% [0.20–1.87%]	1.90E-02	2.66E-02

Table S12 — Mixture analysis: Paired difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR (continued)

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
			mean(TSS - expected VL)	1.05% [0.35–1.76%]	6.19E-03	1.11E-02
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.02% [-0.46–0.50%]	9.31E-01	9.31E-01
Lost variants	freebayes	internal (within plaftorm)	mean(PCP - expected VL)	-1.45% [-2.14–-0.75%]	5.27E-04	1.46E-03
			mean(TSS - expected VL)	-0.92% [-1.52–-0.31%]	5.84E-03	1.11E-02
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.53% [-0.34–1.40%]	2.12E-01	2.46E-01
		external (other plaftorms)	mean(PCP - expected VL)	-2.59% [-3.11–-2.06%]	4.71E-08	1.70E-06
			mean(TSS - expected VL)	-1.55% [-2.11–-0.99%]	3.77E-05	2.26E-04
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	1.04% [0.20–1.88%]	1.92E-02	2.66E-02
	mutserve	internal (within plaftorm)	mean(PCP - expected VL)	-0.66% [-0.96–-0.36%]	3.68E-04	1.20E-03
			mean(TSS - expected VL)	-1.00% [-1.63–-0.37%]	4.31E-03	9.69E-03

Table S12 — Mixture analysis: Paired difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR (continued)

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	-0.34% [-1.00–0.31%]	2.84E-01	3.19E-01
		external (other plaftorms)	mean(PCP - expected VL)	-0.92% [-1.24–-0.61%]	2.22E-05	1.75E-04
			mean(TSS - expected VL)	-3.08% [-4.16–-2.01%]	2.44E-05	1.75E-04
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	-2.16% [-3.36–-0.96%]	1.71E-03	4.40E-03
	varscan	internal (within plaftorm)	mean(PCP - expected VL)	-0.65% [-0.96–-0.34%]	5.01E-04	1.46E-03
			mean(TSS - expected VL)	-1.11% [-1.81–-0.41%]	4.29E-03	9.69E-03
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	-0.46% [-1.21–0.28%]	2.06E-01	2.46E-01
		external (other plaftorms)	mean(PCP - expected VL)	-4.10% [-5.38–-2.81%]	8.09E-06	1.46E-04
			mean(TSS - expected VL)	-3.39% [-4.56–-2.21%]	2.35E-05	1.75E-04
			mean[mean(TSS - expected VL) - mean(PCP - expected VL)]	0.71% [-1.38–2.79%]	4.79E-01	5.22E-01

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method; VL — variant level.

Table S13 — Mixture analysis: Paired relative difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
Found variants	freebayes	internal (within plaftorm)	mean[(PCP - expected VL)/expected VL]	1.82% [-3.26–6.90%]	4.55E-01	5.08E-01
			mean[(TSS - expected VL)/expected VL]	3.47% [0.94–6.00%]	1.07E-02	6.40E-02
			mean{mean[(PCP - expected VL)/expected VL] - mean[(TSS - expected VL)/expected VL]}	1.65% [-2.05–5.36%]	3.54E-01	5.08E-01
		external (other plaftorms)	mean[(PCP - expected VL)/expected VL]	1.81% [-3.19–6.81%]	4.51E-01	5.08E-01
			mean[(TSS - expected VL)/expected VL]	2.95% [0.29–5.61%]	3.24E-02	9.71E-02
			mean{mean[(PCP - expected VL)/expected VL] - mean[(TSS - expected VL)/expected VL]}	1.14% [-2.40–4.68%]	5.01E-01	5.08E-01

Table S13 — Mixture analysis: Paired relative difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR (continued)

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
	mutserve	internal (within plaftorm)	mean[(PCP - expected VL)/expected VL]	1.62% [-3.19–6.43%]	4.82E-01	5.08E-01
			mean[(TSS - expected VL)/expected VL]	3.29% [1.01–5.57%]	7.85E-03	6.40E-02
			mean{mean[(PCP - expected VL)/expected VL] - mean[(TSS - expected VL)/expected VL]}	1.67% [-1.97–5.31%]	3.41E-01	5.08E-01
		external (other plaftorms)	mean[(PCP - expected VL)/expected VL]	1.50% [-3.23–6.23%]	5.07E-01	5.08E-01
			mean[(TSS - expected VL)/expected VL]	2.78% [0.40–5.16%]	2.52E-02	9.48E-02
			mean{mean[(PCP - expected VL)/expected VL] - mean[(TSS - expected VL)/expected VL]}	1.28% [-2.21–4.77%]	4.46E-01	5.08E-01
	varscan	internal (within plaftorm)	mean[(PCP - expected VL)/expected VL]	1.74% [-3.33–6.80%]	4.74E-01	5.08E-01
			mean[(TSS - expected VL)/expected VL]	3.29% [1.01–5.57%]	7.85E-03	6.40E-02
			mean{mean[(PCP - expected VL)/expected VL] - mean[(TSS - expected VL)/expected VL]}	1.55% [-2.31–5.41%]	4.03E-01	5.08E-01
		external (other plaftorms)	mean[(PCP - expected VL)/expected VL]	1.61% [-3.29–6.50%]	4.93E-01	5.08E-01
			mean[(TSS - expected VL)/expected VL]	2.75% [0.37–5.13%]	2.63E-02	9.48E-02

Table S13 — Mixture analysis: Paired relative difference between the observed VL and the expected VL (internal and external) in Grade A/B variants, according to the VCM chosen. Paired t-test with FDR (continued)

Macro variant description	VCM	Comparison	Calculation	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
			mean{mean[(PCP - expected VL)/expected VL] - mean[(TSS - expected VL)/expected VL]}	1.15% [-2.48–4.77%]	5.08E-01	5.08E-01

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method; VL — variant level.

Table S14 — Mixture analysis: Paired difference between correlation indicators (internal and external) in Grade A/B variants, depending on the VCM chosen. Paired t-test with FDR

VCM	Variant grade	Comparison	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
freebayes	A	PCP internal - PCP external	0.59 [0.07–1.11]	2.79E-02	7.44E-02
		TSS internal - TSS external	0.28 [-0.05–0.61]	9.31E-02	1.60E-01
		TSS internal - PCP internal	-5.77 [-13.12–1.58]	1.15E-01	1.72E-01
		TSS external - PCP external	-5.45 [-12.63–1.73]	1.26E-01	1.74E-01
	B	PCP internal - PCP external	0.77 [-0.26–1.79]	1.30E-01	1.74E-01
		TSS internal - TSS external	0.11 [-1.16–1.39]	8.50E-01	8.50E-01

Table S14 — Mixture analysis: Paired difference between correlation indicators (internal and external) in Grade A/B variants, depending on the VCM chosen. Paired t-test with FDR (continued)

VCM	Variant grade	Comparison	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
		TSS internal - PCP internal	-16.23 [-25.82–-6.64]	2.73E-03	9.37E-03
		TSS external - PCP external	-15.57 [-25.24–-5.91]	3.86E-03	1.16E-02
mutserve	A	PCP internal - PCP external	0.34 [-0.04–0.72]	7.70E-02	1.51E-01
		TSS internal - TSS external	1.37 [0.69–2.06]	7.48E-04	7.84E-03
		TSS internal - PCP internal	-10.65 [-22.82–1.52]	8.16E-02	1.51E-01
		TSS external - PCP external	-11.68 [-23.90–0.53]	5.94E-02	1.30E-01
	B	PCP internal - PCP external	0.38 [-0.64–1.40]	4.35E-01	5.22E-01
		TSS internal - TSS external	0.25 [-1.02–1.52]	6.81E-01	7.78E-01
		TSS internal - PCP internal	-20.98 [-32.15–-9.81]	1.25E-03	7.84E-03
		TSS external - PCP external	-20.84 [-32.31–-9.38]	1.61E-03	7.84E-03
varscan	A	PCP internal - PCP external	-10.82 [-20.92–-0.72]	3.76E-02	9.02E-02
		TSS internal - TSS external	1.29 [0.60–1.98]	1.32E-03	7.84E-03
		TSS internal - PCP internal	3.22 [-4.92–11.35]	4.11E-01	5.19E-01

Table S14 — Mixture analysis: Paired difference between correlation indicators (internal and external) in Grade A/B variants, depending on the VCM chosen. Paired t-test with FDR (continued)

VCM	Variant grade	Comparison	Mean paired difference [95% CI]	*p*-value	*p*-value (adj)
		TSS external - PCP external	-8.89 [-20.02–2.23]	1.09E-01	1.72E-01
	B	PCP internal - PCP external	-20.89 [-32.54–-9.23]	1.79E-03	7.84E-03
		TSS internal - TSS external	0.16 [-1.08–1.40]	7.85E-01	8.19E-01
		TSS internal - PCP internal	1.66 [-8.28–11.60]	7.25E-01	7.91E-01
		TSS external - PCP external	-19.38 [-30.33–-8.44]	1.96E-03	7.84E-03

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S15 — Mixture analysis: Paired difference between the proportion of "Novel variants", depending on the VCM chosen. Paired t-test with FDR

VCM	Mean TSS-PCP paired difference [95% CI]	*p*-value	*p*-value (adj)
freebayes	-0.52% [-0.81–-0.24%]	1.58E-03	4.74E-03
mutserve	-0.12% [-0.41–0.17%]	4.02E-01	6.03E-01
varscan	-0.07% [-0.47–0.32%]	6.94E-01	6.94E-01

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S16 — Mixture analysis: Paired difference between TSS and PCP VL for "Novel variants", depending on the VCM chosen. Paired t-test with FDR

VCM	Comparison	Mean TSS-PCP paired difference [95% CI]	*p*-value	*p*-value (adj)
freebayes	Found novel	68.58% [59.44–77.73%]	2.02E-10	1.21E-09
	New novel	3.85% [-0.68–8.38%]	8.89E-02	1.07E-01
mutserve	Found novel	26.69% [21.65–31.73%]	4.78E-06	9.57E-06
	New novel	-7.03% [-29.49–15.43%]	5.01E-01	5.01E-01
varscan	Found novel	27.38% [24.09–30.66%]	4.31E-10	1.29E-09
	New novel	7.51% [2.36–12.67%]	8.77E-03	1.31E-02

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method; VL — variant level.

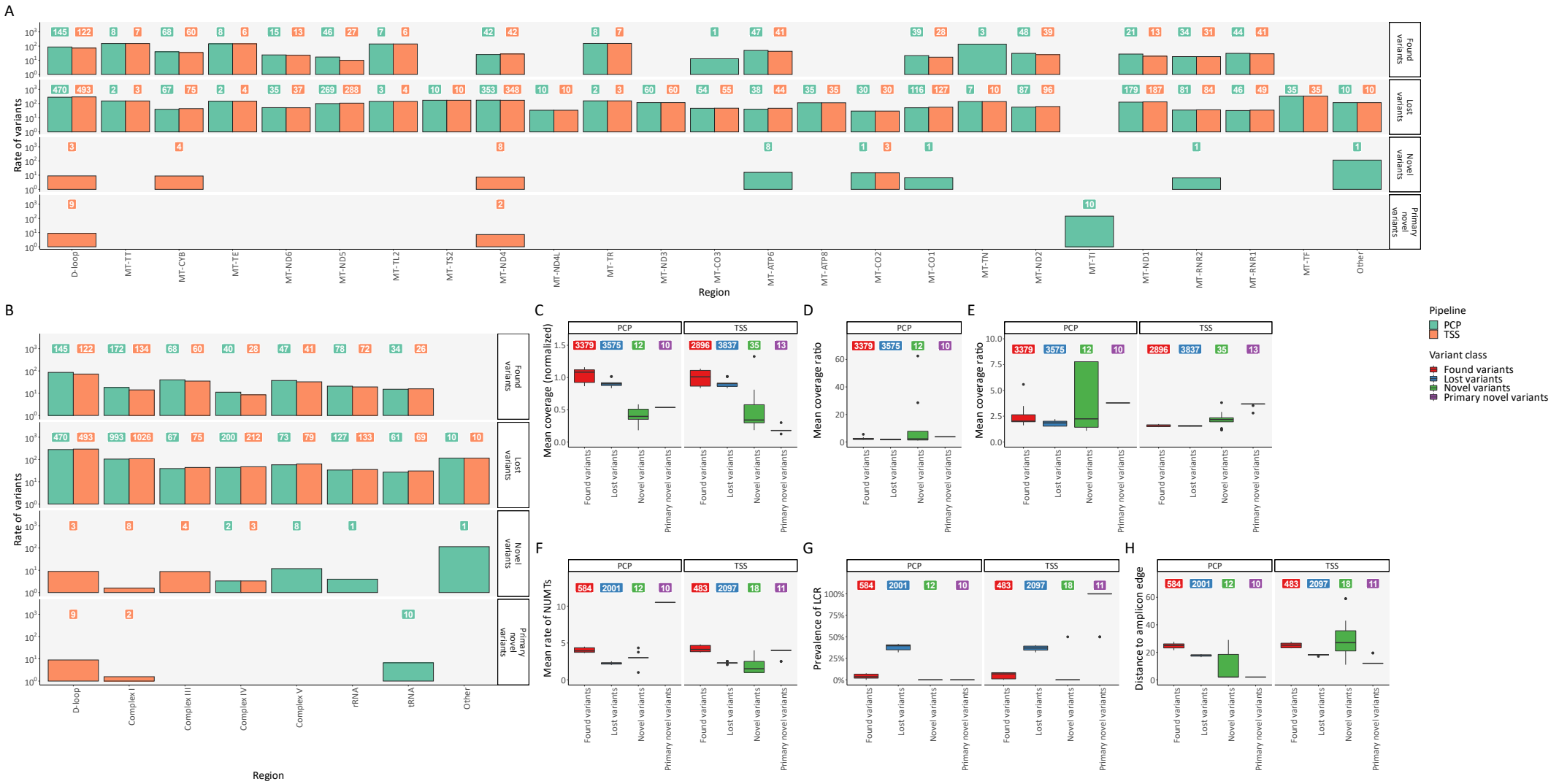


Figure S4 — Differences between different variant classifications, per pipeline

(**A**) Relative variant burden of different mtDNA regions, per pipeline and per variant classification; (**B**) Relative variant burden of different macro mtDNA regions, per pipeline and per variant classification; (**C**) Mean normalized coverage per sample, per pipeline and per variant classification; (**D**) Mean coverage ratio per sample in PCP, per variant classification; (**E**) Mean coverage ratio per sample, per pipeline and per variant classification, with a close-up of Figure S4D regarding PCP; (**F**) Mean rate of NUMTs per sample, per pipeline and per variant classification; (**G**) Mean prevalence of LCR variants per sample, per pipeline and per variant classification; (**H**) Mean distance to “callable” amplicon edge per sample, per pipeline and per variant classification. Labels above the columns/boxplots represent the number of variants for each

x-value. Region *Other* refers to positions in rCRS with an overlap between the two strands or unannotated. For the mean normalized coverage and for the mean coverage ratio, I considered all variants from both the primary and the mixture analysis, including the output from the Illumina sequencing; otherwise, it would not be possible to have coverage and its ratio for false negatives. Abbreviations: LCR — low-complexity region; mtDNA — mitochondrial DNA; NUMTs — nuclear insertions of mitochondrial DNA; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

Table S17 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. All regions

VCM	Pipeline	Macro variant description	Macro region	Region	Relative burden
freebayes	PCP	Found variants	Complex I	MT-ND1	27.46
freebayes	PCP	Found variants	Complex I	MT-ND2	30.71
freebayes	PCP	Found variants	Complex I	MT-ND4	28.13
freebayes	PCP	Found variants	Complex I	MT-ND5	11.04
freebayes	PCP	Found variants	Complex I	MT-ND6	23.81
freebayes	PCP	Found variants	Complex III	MT-CYB	39.73
freebayes	PCP	Found variants	Complex IV	MT-CO1	21.08
freebayes	PCP	Found variants	Complex IV	MT-CO3	12.77
freebayes	PCP	Found variants	Complex V	MT-ATP6	33.60
freebayes	PCP	Found variants	D-loop	D-loop	66.55
freebayes	PCP	Found variants	rRNA	MT-RNR1	30.75
freebayes	PCP	Found variants	rRNA	MT-RNR2	17.66
freebayes	PCP	Found variants	tRNA	MT-TE	144.93
freebayes	PCP	Found variants	tRNA	MT-TL2	140.85
freebayes	PCP	Found variants	tRNA	MT-TN	136.99
freebayes	PCP	Found variants	tRNA	MT-TR	153.85
freebayes	PCP	Found variants	tRNA	MT-TT	151.52
freebayes	PCP	Lost variants	Complex I	MT-ND1	205.02
freebayes	PCP	Lost variants	Complex I	MT-ND2	46.07
freebayes	PCP	Lost variants	Complex I	MT-ND3	144.51
freebayes	PCP	Lost variants	Complex I	MT-ND4	199.85
freebayes	PCP	Lost variants	Complex I	MT-ND4L	33.67
freebayes	PCP	Lost variants	Complex I	MT-ND5	134.29
freebayes	PCP	Lost variants	Complex I	MT-ND6	36.63
freebayes	PCP	Lost variants	Complex III	MT-CYB	53.75
freebayes	PCP	Lost variants	Complex IV	MT-CO1	41.50
freebayes	PCP	Lost variants	Complex IV	MT-CO2	19.49
freebayes	PCP	Lost variants	Complex IV	MT-CO3	58.75
freebayes	PCP	Lost variants	Complex V	MT-ATP6	66.14
freebayes	PCP	Lost variants	Complex V	MT-ATP8	177.13
freebayes	PCP	Lost variants	D-loop	D-loop	251.34
freebayes	PCP	Lost variants	rRNA	MT-RNR1	39.13
freebayes	PCP	Lost variants	rRNA	MT-RNR2	41.53
freebayes	PCP	Lost variants	tRNA	MT-TE	144.93
freebayes	PCP	Lost variants	tRNA	MT-TF	422.54
freebayes	PCP	Lost variants	tRNA	MT-TL1	133.33
freebayes	PCP	Lost variants	tRNA	MT-TL2	140.85
freebayes	PCP	Lost variants	tRNA	MT-TN	136.99
freebayes	PCP	Lost variants	tRNA	MT-TR	153.85
freebayes	PCP	Lost variants	tRNA	MT-TT	151.52
freebayes	PCP	Lost variants	tRNA	MT-TW	147.06
freebayes	PCP	Novel variants	Complex I	MT-ND4	7.29
freebayes	PCP	Novel variants	Complex I	MT-ND5	5.52
freebayes	PCP	Novel variants	Complex IV	MT-CO2	14.62
freebayes	PCP	Novel variants	Complex V	MT-ATP6	17.06

Table S17 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. All regions (continued)

VCM	Pipeline	Macro variant description	Macro region	Region	Relative burden
freebayes	PCP	Novel variants	tRNA	MT-TE	144.93
freebayes	PCP	Primary novel variants	Complex V	MT-ATP6	15.75
freebayes	PCP	Primary novel variants	tRNA	MT-TE	144.93
freebayes	PCP	Primary novel variants	tRNA	MT-TI	144.93
freebayes	TSS	Found variants	Complex I	MT-ND1	19.43
freebayes	TSS	Found variants	Complex I	MT-ND2	24.95
freebayes	TSS	Found variants	Complex I	MT-ND4	24.53
freebayes	TSS	Found variants	Complex I	MT-ND5	9.93
freebayes	TSS	Found variants	Complex I	MT-ND6	22.51
freebayes	TSS	Found variants	Complex III	MT-CYB	35.06
freebayes	TSS	Found variants	Complex IV	MT-CO1	16.51
freebayes	TSS	Found variants	Complex V	MT-ATP6	35.79
freebayes	TSS	Found variants	D-loop	D-loop	65.36
freebayes	TSS	Found variants	rRNA	MT-RNR1	28.65
freebayes	TSS	Found variants	rRNA	MT-RNR2	18.10
freebayes	TSS	Found variants	tRNA	MT-TE	144.93
freebayes	TSS	Found variants	tRNA	MT-TL2	140.85
freebayes	TSS	Found variants	tRNA	MT-TR	153.85
freebayes	TSS	Found variants	tRNA	MT-TT	151.52
freebayes	TSS	Lost variants	Complex I	MT-ND1	210.60
freebayes	TSS	Lost variants	Complex I	MT-ND2	51.82
freebayes	TSS	Lost variants	Complex I	MT-ND3	144.51
freebayes	TSS	Lost variants	Complex I	MT-ND4	208.12
freebayes	TSS	Lost variants	Complex I	MT-ND4L	33.67
freebayes	TSS	Lost variants	Complex I	MT-ND5	135.39
freebayes	TSS	Lost variants	Complex I	MT-ND6	36.73
freebayes	TSS	Lost variants	Complex III	MT-CYB	58.43
freebayes	TSS	Lost variants	Complex IV	MT-CO1	46.26
freebayes	TSS	Lost variants	Complex IV	MT-CO2	19.49
freebayes	TSS	Lost variants	Complex IV	MT-CO3	59.60
freebayes	TSS	Lost variants	Complex V	MT-ATP6	62.99
freebayes	TSS	Lost variants	Complex V	MT-ATP8	177.13
freebayes	TSS	Lost variants	D-loop	D-loop	252.53
freebayes	TSS	Lost variants	rRNA	MT-RNR1	41.23
freebayes	TSS	Lost variants	rRNA	MT-RNR2	42.39
freebayes	TSS	Lost variants	tRNA	MT-TE	144.93
freebayes	TSS	Lost variants	tRNA	MT-TF	422.54
freebayes	TSS	Lost variants	tRNA	MT-TL1	133.33
freebayes	TSS	Lost variants	tRNA	MT-TL2	140.85
freebayes	TSS	Lost variants	tRNA	MT-TN	136.99
freebayes	TSS	Lost variants	tRNA	MT-TR	153.85
freebayes	TSS	Lost variants	tRNA	MT-TT	151.52
freebayes	TSS	Lost variants	tRNA	MT-TW	147.06
freebayes	TSS	Novel variants	Complex I	MT-ND4	7.29
freebayes	TSS	Novel variants	Complex III	MT-CYB	8.76
freebayes	TSS	Novel variants	Complex IV	MT-CO2	14.62

Table S17 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. All regions (continued)

VCM	Pipeline	Macro variant description	Macro region	Region	Relative burden
freebayes	TSS	Novel variants	Complex V	MT-ATP6	25.20
freebayes	TSS	Novel variants	D-loop	D-loop	19.10
freebayes	TSS	Primary novel variants	Complex I	MT-ND4	7.29
freebayes	TSS	Primary novel variants	Complex V	MT-ATP6	15.75
freebayes	TSS	Primary novel variants	D-loop	D-loop	16.84
mutserve	PCP	Found variants	Complex I	MT-ND1	27.46
mutserve	PCP	Found variants	Complex I	MT-ND2	30.71
mutserve	PCP	Found variants	Complex I	MT-ND4	25.53
mutserve	PCP	Found variants	Complex I	MT-ND5	16.92
mutserve	PCP	Found variants	Complex I	MT-ND6	23.81
mutserve	PCP	Found variants	Complex III	MT-CYB	39.73
mutserve	PCP	Found variants	Complex IV	MT-CO1	21.08
mutserve	PCP	Found variants	Complex IV	MT-CO3	12.77
mutserve	PCP	Found variants	Complex V	MT-ATP6	49.34
mutserve	PCP	Found variants	D-loop	D-loop	86.16
mutserve	PCP	Found variants	rRNA	MT-RNR1	30.75
mutserve	PCP	Found variants	rRNA	MT-RNR2	18.20
mutserve	PCP	Found variants	tRNA	MT-TE	144.93
mutserve	PCP	Found variants	tRNA	MT-TL2	140.85
mutserve	PCP	Found variants	tRNA	MT-TN	136.99
mutserve	PCP	Found variants	tRNA	MT-TR	153.85
mutserve	PCP	Found variants	tRNA	MT-TT	151.52
mutserve	PCP	Lost variants	Complex I	MT-ND1	124.83
mutserve	PCP	Lost variants	Complex I	MT-ND2	55.66
mutserve	PCP	Lost variants	Complex I	MT-ND3	115.61
mutserve	PCP	Lost variants	Complex I	MT-ND4	171.65
mutserve	PCP	Lost variants	Complex I	MT-ND4L	33.67
mutserve	PCP	Lost variants	Complex I	MT-ND5	98.97
mutserve	PCP	Lost variants	Complex I	MT-ND6	51.28
mutserve	PCP	Lost variants	Complex III	MT-CYB	39.15
mutserve	PCP	Lost variants	Complex IV	MT-CO1	50.15
mutserve	PCP	Lost variants	Complex IV	MT-CO2	29.24
mutserve	PCP	Lost variants	Complex IV	MT-CO3	45.98
mutserve	PCP	Lost variants	Complex V	MT-ATP6	39.90
mutserve	PCP	Lost variants	Complex V	MT-ATP8	112.72
mutserve	PCP	Lost variants	D-loop	D-loop	279.26
mutserve	PCP	Lost variants	Other	Other	114.94
mutserve	PCP	Lost variants	rRNA	MT-RNR1	32.15
mutserve	PCP	Lost variants	rRNA	MT-RNR2	34.68
mutserve	PCP	Lost variants	tRNA	MT-TE	144.93
mutserve	PCP	Lost variants	tRNA	MT-TF	328.64
mutserve	PCP	Lost variants	tRNA	MT-TL2	140.85
mutserve	PCP	Lost variants	tRNA	MT-TN	136.99
mutserve	PCP	Lost variants	tRNA	MT-TR	153.85
mutserve	PCP	Lost variants	tRNA	MT-TS2	169.49
mutserve	PCP	Lost variants	tRNA	MT-TT	151.52

Table S17 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. All regions (continued)

VCM	Pipeline	Macro variant description	Macro region	Region	Relative burden
mutserve	PCP	Novel variants	Complex IV	MT-CO1	6.49
mutserve	PCP	Novel variants	Complex IV	MT-CO2	14.62
mutserve	PCP	Novel variants	Complex V	MT-ATP6	15.75
mutserve	PCP	Novel variants	Other	Other	114.94
mutserve	PCP	Novel variants	rRNA	MT-RNR2	6.42
mutserve	PCP	Primary novel variants	tRNA	MT-TI	144.93
mutserve	TSS	Found variants	Complex I	MT-ND1	19.43
mutserve	TSS	Found variants	Complex I	MT-ND2	24.95
mutserve	TSS	Found variants	Complex I	MT-ND4	27.85
mutserve	TSS	Found variants	Complex I	MT-ND5	9.93
mutserve	TSS	Found variants	Complex I	MT-ND6	22.51
mutserve	TSS	Found variants	Complex III	MT-CYB	35.06
mutserve	TSS	Found variants	Complex IV	MT-CO1	16.51
mutserve	TSS	Found variants	Complex V	MT-ATP6	43.04
mutserve	TSS	Found variants	D-loop	D-loop	72.49
mutserve	TSS	Found variants	rRNA	MT-RNR1	28.65
mutserve	TSS	Found variants	rRNA	MT-RNR2	18.10
mutserve	TSS	Found variants	tRNA	MT-TE	144.93
mutserve	TSS	Found variants	tRNA	MT-TL2	140.85
mutserve	TSS	Found variants	tRNA	MT-TR	153.85
mutserve	TSS	Found variants	tRNA	MT-TT	151.52
mutserve	TSS	Lost variants	Complex I	MT-ND1	130.40
mutserve	TSS	Lost variants	Complex I	MT-ND2	61.42
mutserve	TSS	Lost variants	Complex I	MT-ND3	115.61
mutserve	TSS	Lost variants	Complex I	MT-ND4	169.22
mutserve	TSS	Lost variants	Complex I	MT-ND4L	33.67
mutserve	TSS	Lost variants	Complex I	MT-ND5	105.96
mutserve	TSS	Lost variants	Complex I	MT-ND6	50.34
mutserve	TSS	Lost variants	Complex III	MT-CYB	43.82
mutserve	TSS	Lost variants	Complex IV	MT-CO1	54.91
mutserve	TSS	Lost variants	Complex IV	MT-CO2	29.24
mutserve	TSS	Lost variants	Complex IV	MT-CO3	46.83
mutserve	TSS	Lost variants	Complex V	MT-ATP6	46.19
mutserve	TSS	Lost variants	Complex V	MT-ATP8	112.72
mutserve	TSS	Lost variants	D-loop	D-loop	292.93
mutserve	TSS	Lost variants	Other	Other	114.94
mutserve	TSS	Lost variants	rRNA	MT-RNR1	34.24
mutserve	TSS	Lost variants	rRNA	MT-RNR2	35.97
mutserve	TSS	Lost variants	tRNA	MT-TE	144.93
mutserve	TSS	Lost variants	tRNA	MT-TF	328.64
mutserve	TSS	Lost variants	tRNA	MT-TL2	140.85
mutserve	TSS	Lost variants	tRNA	MT-TN	136.99
mutserve	TSS	Lost variants	tRNA	MT-TR	153.85
mutserve	TSS	Lost variants	tRNA	MT-TS2	169.49
mutserve	TSS	Lost variants	tRNA	MT-TT	151.52
mutserve	TSS	Novel variants	Complex I	MT-ND4	7.29

Table S17 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. All regions (continued)

VCM	Pipeline	Macro variant description	Macro region	Region	Relative burden
mutserve	TSS	Novel variants	Complex III	MT-CYB	8.76
mutserve	TSS	Novel variants	Complex IV	MT-CO2	14.62
mutserve	TSS	Novel variants	D-loop	D-loop	8.91
mutserve	TSS	Primary novel variants	Complex I	MT-ND4	7.29
mutserve	TSS	Primary novel variants	D-loop	D-loop	8.91
varscan	PCP	Found variants	Complex I	MT-ND1	27.46
varscan	PCP	Found variants	Complex I	MT-ND2	30.71
varscan	PCP	Found variants	Complex I	MT-ND4	25.53
varscan	PCP	Found variants	Complex I	MT-ND5	16.92
varscan	PCP	Found variants	Complex I	MT-ND6	23.81
varscan	PCP	Found variants	Complex III	MT-CYB	37.98
varscan	PCP	Found variants	Complex IV	MT-CO1	21.08
varscan	PCP	Found variants	Complex IV	MT-CO3	12.77
varscan	PCP	Found variants	Complex V	MT-ATP6	48.29
varscan	PCP	Found variants	D-loop	D-loop	74.87
varscan	PCP	Found variants	rRNA	MT-RNR1	30.75
varscan	PCP	Found variants	rRNA	MT-RNR2	17.66
varscan	PCP	Found variants	tRNA	MT-TE	144.93
varscan	PCP	Found variants	tRNA	MT-TL2	140.85
varscan	PCP	Found variants	tRNA	MT-TN	136.99
varscan	PCP	Found variants	tRNA	MT-TR	153.85
varscan	PCP	Found variants	tRNA	MT-TT	151.52
varscan	PCP	Lost variants	Complex I	MT-ND1	51.60
varscan	PCP	Lost variants	Complex I	MT-ND2	36.47
varscan	PCP	Lost variants	Complex I	MT-ND3	77.07
varscan	PCP	Lost variants	Complex I	MT-ND4	71.97
varscan	PCP	Lost variants	Complex I	MT-ND4L	33.67
varscan	PCP	Lost variants	Complex I	MT-ND5	45.62
varscan	PCP	Lost variants	Complex I	MT-ND6	36.63
varscan	PCP	Lost variants	Complex III	MT-CYB	26.29
varscan	PCP	Lost variants	Complex IV	MT-CO1	32.86
varscan	PCP	Lost variants	Complex IV	MT-CO2	19.49
varscan	PCP	Lost variants	Complex IV	MT-CO3	33.21
varscan	PCP	Lost variants	Complex V	MT-ATP6	40.94
varscan	PCP	Lost variants	Complex V	MT-ATP8	96.62
varscan	PCP	Lost variants	D-loop	D-loop	168.75
varscan	PCP	Lost variants	rRNA	MT-RNR1	18.17
varscan	PCP	Lost variants	rRNA	MT-RNR2	35.11
varscan	PCP	Lost variants	tRNA	MT-TE	144.93
varscan	PCP	Lost variants	tRNA	MT-TL2	140.85
varscan	PCP	Lost variants	tRNA	MT-TN	136.99
varscan	PCP	Lost variants	tRNA	MT-TR	153.85
varscan	PCP	Lost variants	tRNA	MT-TT	151.52
varscan	PCP	Novel variants	Complex IV	MT-CO2	14.62
varscan	PCP	Novel variants	Complex V	MT-ATP6	15.75
varscan	PCP	Novel variants	Other	Other	114.94

Table S17 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. All regions (continued)

VCM	Pipeline	Macro variant description	Macro region	Region	Relative burden
varscan	PCP	Primary novel variants	tRNA	MT-TI	144.93
varscan	TSS	Found variants	Complex I	MT-ND1	19.43
varscan	TSS	Found variants	Complex I	MT-ND2	24.95
varscan	TSS	Found variants	Complex I	MT-ND4	24.53
varscan	TSS	Found variants	Complex I	MT-ND5	9.93
varscan	TSS	Found variants	Complex I	MT-ND6	22.51
varscan	TSS	Found variants	Complex III	MT-CYB	35.06
varscan	TSS	Found variants	Complex IV	MT-CO1	16.51
varscan	TSS	Found variants	Complex V	MT-ATP6	43.04
varscan	TSS	Found variants	D-loop	D-loop	72.49
varscan	TSS	Found variants	rRNA	MT-RNR1	28.65
varscan	TSS	Found variants	rRNA	MT-RNR2	18.10
varscan	TSS	Found variants	tRNA	MT-TE	144.93
varscan	TSS	Found variants	tRNA	MT-TL2	140.85
varscan	TSS	Found variants	tRNA	MT-TR	153.85
varscan	TSS	Found variants	tRNA	MT-TT	151.52
varscan	TSS	Lost variants	Complex I	MT-ND1	57.18
varscan	TSS	Lost variants	Complex I	MT-ND2	42.23
varscan	TSS	Lost variants	Complex I	MT-ND3	77.07
varscan	TSS	Lost variants	Complex I	MT-ND4	74.40
varscan	TSS	Lost variants	Complex I	MT-ND4L	33.67
varscan	TSS	Lost variants	Complex I	MT-ND5	52.61
varscan	TSS	Lost variants	Complex I	MT-ND6	36.73
varscan	TSS	Lost variants	Complex III	MT-CYB	29.21
varscan	TSS	Lost variants	Complex IV	MT-CO1	37.61
varscan	TSS	Lost variants	Complex IV	MT-CO2	19.49
varscan	TSS	Lost variants	Complex IV	MT-CO3	34.06
varscan	TSS	Lost variants	Complex V	MT-ATP6	46.19
varscan	TSS	Lost variants	Complex V	MT-ATP8	96.62
varscan	TSS	Lost variants	D-loop	D-loop	171.12
varscan	TSS	Lost variants	rRNA	MT-RNR1	20.27
varscan	TSS	Lost variants	rRNA	MT-RNR2	35.97
varscan	TSS	Lost variants	tRNA	MT-TE	144.93
varscan	TSS	Lost variants	tRNA	MT-TL2	140.85
varscan	TSS	Lost variants	tRNA	MT-TN	136.99
varscan	TSS	Lost variants	tRNA	MT-TR	153.85
varscan	TSS	Lost variants	tRNA	MT-TT	151.52
varscan	TSS	Novel variants	Complex I	MT-ND4	7.29
varscan	TSS	Novel variants	Complex III	MT-CYB	8.76
varscan	TSS	Novel variants	Complex IV	MT-CO2	14.62
varscan	TSS	Novel variants	D-loop	D-loop	8.91
varscan	TSS	Primary novel variants	Complex I	MT-ND4	7.29
varscan	TSS	Primary novel variants	D-loop	D-loop	8.91

Abbreviations: mtDNA — mitochondrial DNA; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S18 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. Macro regions only

VCM	Pipeline	Macro variant description	Macro region	Relative burden
freebayes	PCP	Found variants	Complex I	17.64
freebayes	PCP	Found variants	Complex III	39.73
freebayes	PCP	Found variants	Complex IV	11.08
freebayes	PCP	Found variants	Complex V	25.34
freebayes	PCP	Found variants	D-loop	66.55
freebayes	PCP	Found variants	rRNA	20.44
freebayes	PCP	Found variants	tRNA	15.04
freebayes	PCP	Lost variants	Complex I	131.46
freebayes	PCP	Lost variants	Complex III	53.75
freebayes	PCP	Lost variants	Complex IV	40.99
freebayes	PCP	Lost variants	Complex V	93.43
freebayes	PCP	Lost variants	D-loop	251.34
freebayes	PCP	Lost variants	rRNA	40.62
freebayes	PCP	Lost variants	tRNA	29.20
freebayes	PCP	Novel variants	Complex I	2.00
freebayes	PCP	Novel variants	Complex IV	3.32
freebayes	PCP	Novel variants	Complex V	12.87
freebayes	PCP	Novel variants	tRNA	6.64
freebayes	PCP	Primary novel variants	Complex V	11.88
freebayes	PCP	Primary novel variants	tRNA	9.01
freebayes	TSS	Found variants	Complex I	13.55
freebayes	TSS	Found variants	Complex III	35.06
freebayes	TSS	Found variants	Complex IV	8.46
freebayes	TSS	Found variants	Complex V	26.99
freebayes	TSS	Found variants	D-loop	65.36
freebayes	TSS	Found variants	rRNA	19.12
freebayes	TSS	Found variants	tRNA	15.68
freebayes	TSS	Lost variants	Complex I	135.56
freebayes	TSS	Lost variants	Complex III	58.43
freebayes	TSS	Lost variants	Complex IV	43.65
freebayes	TSS	Lost variants	Complex V	91.05
freebayes	TSS	Lost variants	D-loop	252.53
freebayes	TSS	Lost variants	rRNA	41.95
freebayes	TSS	Lost variants	tRNA	32.74
freebayes	TSS	Novel variants	Complex I	1.58
freebayes	TSS	Novel variants	Complex III	8.76
freebayes	TSS	Novel variants	Complex IV	3.32
freebayes	TSS	Novel variants	Complex V	19.00
freebayes	TSS	Novel variants	D-loop	19.10
freebayes	TSS	Primary novel variants	Complex I	1.58
freebayes	TSS	Primary novel variants	Complex V	11.88
freebayes	TSS	Primary novel variants	D-loop	16.84
mutserve	PCP	Found variants	Complex I	18.06
mutserve	PCP	Found variants	Complex III	39.73
mutserve	PCP	Found variants	Complex IV	11.08

Table S18 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. Macro regions only (continued)

VCM	Pipeline	Macro variant description	Macro region	Relative burden
mutserve	PCP	Found variants	Complex V	37.21
mutserve	PCP	Found variants	D-loop	86.16
mutserve	PCP	Found variants	rRNA	20.71
mutserve	PCP	Found variants	tRNA	15.04
mutserve	PCP	Lost variants	Complex I	104.27
mutserve	PCP	Lost variants	Complex III	39.15
mutserve	PCP	Lost variants	Complex IV	44.31
mutserve	PCP	Lost variants	Complex V	57.80
mutserve	PCP	Lost variants	D-loop	279.26
mutserve	PCP	Lost variants	Other	114.94
mutserve	PCP	Lost variants	rRNA	33.72
mutserve	PCP	Lost variants	tRNA	26.99
mutserve	PCP	Novel variants	Complex IV	3.32
mutserve	PCP	Novel variants	Complex V	11.88
mutserve	PCP	Novel variants	Other	114.94
mutserve	PCP	Novel variants	rRNA	3.98
mutserve	PCP	Primary novel variants	tRNA	6.64
mutserve	TSS	Found variants	Complex I	14.07
mutserve	TSS	Found variants	Complex III	35.06
mutserve	TSS	Found variants	Complex IV	8.46
mutserve	TSS	Found variants	Complex V	32.46
mutserve	TSS	Found variants	D-loop	72.49
mutserve	TSS	Found variants	rRNA	19.12
mutserve	TSS	Found variants	tRNA	15.68
mutserve	TSS	Lost variants	Complex I	107.73
mutserve	TSS	Lost variants	Complex III	43.82
mutserve	TSS	Lost variants	Complex IV	46.97
mutserve	TSS	Lost variants	Complex V	62.55
mutserve	TSS	Lost variants	D-loop	292.93
mutserve	TSS	Lost variants	Other	114.94
mutserve	TSS	Lost variants	rRNA	35.31
mutserve	TSS	Lost variants	tRNA	30.52
mutserve	TSS	Novel variants	Complex I	1.58
mutserve	TSS	Novel variants	Complex III	8.76
mutserve	TSS	Novel variants	Complex IV	3.32
mutserve	TSS	Novel variants	D-loop	8.91
mutserve	TSS	Primary novel variants	Complex I	1.58
mutserve	TSS	Primary novel variants	D-loop	8.91
varscan	PCP	Found variants	Complex I	18.06
varscan	PCP	Found variants	Complex III	37.98
varscan	PCP	Found variants	Complex IV	11.08
varscan	PCP	Found variants	Complex V	36.42
varscan	PCP	Found variants	D-loop	74.87
varscan	PCP	Found variants	rRNA	20.44
varscan	PCP	Found variants	tRNA	14.60
varscan	PCP	Lost variants	Complex I	50.19

Table S18 — Mixture analysis: Relative burden of different macro variants in different mtDNA regions, depending on the VCM chosen. Macro regions only (continued)

VCM	Pipeline	Macro variant description	Macro region	Relative burden
varscan	PCP	Lost variants	Complex III	26.29
varscan	PCP	Lost variants	Complex IV	29.91
varscan	PCP	Lost variants	Complex V	54.63
varscan	PCP	Lost variants	D-loop	168.75
varscan	PCP	Lost variants	rRNA	28.67
varscan	PCP	Lost variants	tRNA	8.68
varscan	PCP	Novel variants	Complex IV	3.32
varscan	PCP	Novel variants	Complex V	11.88
varscan	PCP	Novel variants	Other	114.94
varscan	PCP	Primary novel variants	tRNA	6.64
varscan	TSS	Found variants	Complex I	13.55
varscan	TSS	Found variants	Complex III	35.06
varscan	TSS	Found variants	Complex IV	8.46
varscan	TSS	Found variants	Complex V	32.46
varscan	TSS	Found variants	D-loop	72.49
varscan	TSS	Found variants	rRNA	19.12
varscan	TSS	Found variants	tRNA	15.68
varscan	TSS	Lost variants	Complex I	54.71
varscan	TSS	Lost variants	Complex III	29.21
varscan	TSS	Lost variants	Complex IV	32.57
varscan	TSS	Lost variants	Complex V	58.59
varscan	TSS	Lost variants	D-loop	171.12
varscan	TSS	Lost variants	rRNA	30.00
varscan	TSS	Lost variants	tRNA	11.38
varscan	TSS	Novel variants	Complex I	1.58
varscan	TSS	Novel variants	Complex III	8.76
varscan	TSS	Novel variants	Complex IV	3.32
varscan	TSS	Novel variants	D-loop	8.91
varscan	TSS	Primary novel variants	Complex I	1.58
varscan	TSS	Primary novel variants	D-loop	8.91

Abbreviations: mtDNA — mitochondrial DNA; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S19 — Mixture analysis: Paired comparison of mean normalized coverage per sample in different macro variants, depending on the VCM chosen. Pairwise t-test with FDR

VCM	Pipeline	Comparison	Means	*p*-value
freebayes	PCP	Lost variants - Found variants	0.89 ~ 1.06	2.69E-05
freebayes	PCP	Lost variants - Novel variants	0.89 ~ 0.22	8.55E-25
freebayes	PCP	Novel variants - Found variants	0.22 ~ 1.06	6.67E-29
freebayes	PCP	Primary novel variants - Found variants	0.34 ~ 1.06	7.20E-26
freebayes	PCP	Primary novel variants - Lost variants	0.34 ~ 0.89	6.18E-21
freebayes	PCP	Primary novel variants - Novel variants	0.34 ~ 0.22	1.54E-03
freebayes	TSS	Lost variants - Found variants	0.90 ~ 1.02	3.69E-02
freebayes	TSS	Lost variants - Novel variants	0.90 ~ 0.52	3.30E-08
freebayes	TSS	Novel variants - Found variants	0.52 ~ 1.02	3.13E-11
freebayes	TSS	Primary novel variants - Found variants	0.36 ~ 1.02	7.54E-13
freebayes	TSS	Primary novel variants - Lost variants	0.36 ~ 0.90	1.54E-10
freebayes	TSS	Primary novel variants - Novel variants	0.36 ~ 0.52	2.06E-02
mutserve	PCP	Lost variants - Found variants	0.91 ~ 1.04	1.06E-04
mutserve	PCP	Lost variants - Novel variants	0.91 ~ 0.41	3.92E-18
mutserve	PCP	Novel variants - Found variants	0.41 ~ 1.04	1.64E-21
mutserve	PCP	Primary novel variants - Found variants	0.54 ~ 1.04	1.72E-18
mutserve	PCP	Primary novel variants - Lost variants	0.54 ~ 0.91	3.59E-14
mutserve	PCP	Primary novel variants - Novel variants	0.54 ~ 0.41	1.36E-03
mutserve	TSS	Lost variants - Found variants	0.91 ~ 1.00	1.70E-01
mutserve	TSS	Lost variants - Novel variants	0.91 ~ 0.49	1.43E-07
mutserve	TSS	Novel variants - Found variants	0.49 ~ 1.00	1.68E-09
mutserve	TSS	Primary novel variants - Found variants	0.19 ~ 1.00	1.15E-13
mutserve	TSS	Primary novel variants - Lost variants	0.19 ~ 0.91	2.89E-12
mutserve	TSS	Primary novel variants - Novel variants	0.19 ~ 0.49	4.23E-04
varscan	PCP	Lost variants - Found variants	0.98 ~ 1.06	7.37E-03
varscan	PCP	Lost variants - Novel variants	0.98 ~ 0.37	1.89E-22
varscan	PCP	Novel variants - Found variants	0.37 ~ 1.06	2.73E-24
varscan	PCP	Primary novel variants - Found variants	0.54 ~ 1.06	1.19E-20
varscan	PCP	Primary novel variants - Lost variants	0.54 ~ 0.98	4.27E-18
varscan	PCP	Primary novel variants - Novel variants	0.54 ~ 0.37	4.21E-05
varscan	TSS	Lost variants - Found variants	0.99 ~ 1.01	6.26E-01
varscan	TSS	Lost variants - Novel variants	0.99 ~ 0.31	1.30E-21
varscan	TSS	Novel variants - Found variants	0.31 ~ 1.01	4.95E-22
varscan	TSS	Primary novel variants - Found variants	0.17 ~ 1.01	2.62E-22
varscan	TSS	Primary novel variants - Lost variants	0.17 ~ 0.99	3.75E-22
varscan	TSS	Primary novel variants - Novel variants	0.17 ~ 0.31	5.58E-03

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S20 — Mixture analysis: Paired comparison of mean coverage ratios per sample in different macro variants, depending on the VCM chosen. Pairwise t-test with FDR

VCM	Pipeline	Comparison	Means	*p*-value
freebayes	PCP	Lost variants - Found variants	17.17 ~ 4.09	1.50E-16
freebayes	PCP	Lost variants - Novel variants	17.17 ~ 4.95	5.33E-14
freebayes	PCP	Novel variants - Found variants	4.95 ~ 4.09	4.46E-01
freebayes	PCP	Primary novel variants - Found variants	1.04 ~ 4.09	1.13E-02
freebayes	PCP	Primary novel variants - Lost variants	1.04 ~ 17.17	8.00E-18
freebayes	PCP	Primary novel variants - Novel variants	1.04 ~ 4.95	4.02E-03
freebayes	TSS	Lost variants - Found variants	16.10 ~ 1.59	1.76E-34
freebayes	TSS	Lost variants - Novel variants	16.10 ~ 2.28	9.12E-34
freebayes	TSS	Novel variants - Found variants	2.28 ~ 1.59	1.26E-01
freebayes	TSS	Primary novel variants - Found variants	3.39 ~ 1.59	1.52E-03
freebayes	TSS	Primary novel variants - Lost variants	3.39 ~ 16.10	2.64E-29
freebayes	TSS	Primary novel variants - Novel variants	3.39 ~ 2.28	4.41E-02
mutserve	PCP	Lost variants - Found variants	1.78 ~ 2.46	8.35E-01
mutserve	PCP	Lost variants - Novel variants	1.78 ~ 12.10	3.82E-02
mutserve	PCP	Novel variants - Found variants	12.10 ~ 2.46	3.82E-02
mutserve	PCP	Primary novel variants - Found variants	3.78 ~ 2.46	8.35E-01
mutserve	PCP	Primary novel variants - Lost variants	3.78 ~ 1.78	8.35E-01
mutserve	PCP	Primary novel variants - Novel variants	3.78 ~ 12.10	9.18E-02
mutserve	TSS	Lost variants - Found variants	1.55 ~ 1.58	8.69E-01
mutserve	TSS	Lost variants - Novel variants	1.55 ~ 2.18	1.26E-04
mutserve	TSS	Novel variants - Found variants	2.18 ~ 1.58	1.70E-04
mutserve	TSS	Primary novel variants - Found variants	3.57 ~ 1.58	5.98E-16
mutserve	TSS	Primary novel variants - Lost variants	3.57 ~ 1.55	5.98E-16
mutserve	TSS	Primary novel variants - Novel variants	3.57 ~ 2.18	1.41E-10
varscan	PCP	Lost variants - Found variants	1.45 ~ 1.50	9.73E-01
varscan	PCP	Lost variants - Novel variants	1.45 ~ 5.33	5.88E-02
varscan	PCP	Novel variants - Found variants	5.33 ~ 1.50	5.88E-02
varscan	PCP	Primary novel variants - Found variants	3.81 ~ 1.50	2.08E-01
varscan	PCP	Primary novel variants - Lost variants	3.81 ~ 1.45	2.08E-01
varscan	PCP	Primary novel variants - Novel variants	3.81 ~ 5.33	4.59E-01
varscan	TSS	Lost variants - Found variants	1.45 ~ 1.57	2.91E-01
varscan	TSS	Lost variants - Novel variants	1.45 ~ 2.35	4.94E-10
varscan	TSS	Novel variants - Found variants	2.35 ~ 1.57	1.83E-08
varscan	TSS	Primary novel variants - Found variants	3.62 ~ 1.57	4.64E-20
varscan	TSS	Primary novel variants - Lost variants	3.62 ~ 1.45	8.02E-21
varscan	TSS	Primary novel variants - Novel variants	3.62 ~ 2.35	1.78E-12

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S21 — Mixture analysis: Paired comparison of NUMT rates per sample in different macro variants, depending on the VCM chosen. Pairwise t-test with FDR

VCM	Pipeline	Comparison	Means	*p*-value
freebayes	PCP	Lost variants - Found variants	2.27 ~ 3.82	2.12E-04
freebayes	PCP	Novel variants - Found variants	1.87 ~ 3.82	1.07E-05
freebayes	PCP	Novel variants - Lost variants	1.87 ~ 2.27	3.16E-01
freebayes	PCP	Primary novel variants - Found variants	6.32 ~ 3.82	8.70E-08
freebayes	PCP	Primary novel variants - Lost variants	6.32 ~ 2.27	6.49E-14
freebayes	PCP	Primary novel variants - Novel variants	6.32 ~ 1.87	7.60E-15
freebayes	TSS	Lost variants - Found variants	2.25 ~ 4.19	1.38E-02
freebayes	TSS	Novel variants - Found variants	3.91 ~ 4.19	6.64E-01
freebayes	TSS	Novel variants - Lost variants	3.91 ~ 2.25	2.92E-02
freebayes	TSS	Primary novel variants - Found variants	5.14 ~ 4.19	2.59E-01
freebayes	TSS	Primary novel variants - Lost variants	5.14 ~ 2.25	2.20E-03
freebayes	TSS	Primary novel variants - Novel variants	5.14 ~ 3.91	1.64E-01
mutserve	PCP	Lost variants - Found variants	2.25 ~ 4.02	8.02E-15
mutserve	PCP	Novel variants - Found variants	3.01 ~ 4.02	1.13E-06
mutserve	PCP	Novel variants - Lost variants	3.01 ~ 2.25	9.41E-05
mutserve	PCP	Primary novel variants - Found variants	10.50 ~ 4.02	2.43E-35
mutserve	PCP	Primary novel variants - Lost variants	10.50 ~ 2.25	1.77E-39
mutserve	PCP	Primary novel variants - Novel variants	10.50 ~ 3.01	1.14E-35
mutserve	TSS	Lost variants - Found variants	2.30 ~ 4.21	6.02E-10
mutserve	TSS	Novel variants - Found variants	1.97 ~ 4.21	3.54E-11
mutserve	TSS	Novel variants - Lost variants	1.97 ~ 2.30	1.93E-01
mutserve	TSS	Primary novel variants - Found variants	3.67 ~ 4.21	6.39E-02
mutserve	TSS	Primary novel variants - Lost variants	3.67 ~ 2.30	1.34E-05
mutserve	TSS	Primary novel variants - Novel variants	3.67 ~ 1.97	5.27E-07
varscan	PCP	Lost variants - Found variants	2.45 ~ 4.09	1.50E-19
varscan	PCP	Novel variants - Found variants	2.97 ~ 4.09	8.06E-12
varscan	PCP	Novel variants - Lost variants	2.97 ~ 2.45	9.69E-05
varscan	PCP	Primary novel variants - Found variants	10.50 ~ 4.09	2.19E-42
varscan	PCP	Primary novel variants - Lost variants	10.50 ~ 2.45	2.60E-46
varscan	PCP	Primary novel variants - Novel variants	10.50 ~ 2.97	4.91E-43
varscan	TSS	Lost variants - Found variants	2.49 ~ 4.24	7.52E-13
varscan	TSS	Novel variants - Found variants	1.57 ~ 4.24	3.79E-19
varscan	TSS	Novel variants - Lost variants	1.57 ~ 2.49	7.02E-06
varscan	TSS	Primary novel variants - Found variants	3.67 ~ 4.24	7.80E-03
varscan	TSS	Primary novel variants - Lost variants	3.67 ~ 2.49	1.06E-06
varscan	TSS	Primary novel variants - Novel variants	3.67 ~ 1.57	3.94E-13

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate; NUMT — nuclear insertion of mitochondrial DNA; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S22 — Mixture analysis: Paired comparison of mean prevalence of LCR variants per sample in different macro variants, depending on the VCM chosen. Pairwise t-test with FDR

VCM	Pipeline	Comparison	Means	*p*-value
freebayes	PCP	Lost variants - Found variants	38.19% ~ 7.38%	5.85E-35
freebayes	PCP	Novel variants - Found variants	0.00% ~ 7.38%	4.38E-09
freebayes	PCP	Novel variants - Lost variants	0.00% ~ 38.19%	3.42E-39
freebayes	PCP	Primary novel variants - Found variants	0.00% ~ 7.38%	4.38E-09
freebayes	PCP	Primary novel variants - Lost variants	0.00% ~ 38.19%	3.42E-39
freebayes	PCP	Primary novel variants - Novel variants	0.00% ~ 0.00%	1.00E+00
freebayes	TSS	Lost variants - Found variants	37.67% ~ 5.23%	1.14E-11
freebayes	TSS	Novel variants - Found variants	1.67% ~ 5.23%	3.34E-01
freebayes	TSS	Novel variants - Lost variants	1.67% ~ 37.67%	7.97E-13
freebayes	TSS	Primary novel variants - Found variants	45.37% ~ 5.23%	1.69E-12
freebayes	TSS	Primary novel variants - Lost variants	45.37% ~ 37.67%	8.87E-02
freebayes	TSS	Primary novel variants - Novel variants	45.37% ~ 1.67%	2.93E-13
mutserve	PCP	Lost variants - Found variants	38.18% ~ 4.11%	4.77E-35
mutserve	PCP	Novel variants - Found variants	0.00% ~ 4.11%	3.71E-04
mutserve	PCP	Novel variants - Lost variants	0.00% ~ 38.18%	1.17E-34
mutserve	PCP	Primary novel variants - Found variants	0.00% ~ 4.11%	3.11E-04
mutserve	PCP	Primary novel variants - Lost variants	0.00% ~ 38.18%	4.77E-35
mutserve	PCP	Primary novel variants - Novel variants	0.00% ~ 0.00%	1.00E+00
mutserve	TSS	Lost variants - Found variants	36.41% ~ 4.85%	2.32E-09
mutserve	TSS	Novel variants - Found variants	3.85% ~ 4.85%	8.21E-01
mutserve	TSS	Novel variants - Lost variants	3.85% ~ 36.41%	2.32E-09
mutserve	TSS	Primary novel variants - Found variants	88.89% ~ 4.85%	2.67E-21
mutserve	TSS	Primary novel variants - Lost variants	88.89% ~ 36.41%	5.52E-14
mutserve	TSS	Primary novel variants - Novel variants	88.89% ~ 3.85%	2.67E-21
varscan	PCP	Lost variants - Found variants	31.02% ~ 1.38%	2.25E-32
varscan	PCP	Novel variants - Found variants	0.00% ~ 1.38%	2.29E-01
varscan	PCP	Novel variants - Lost variants	0.00% ~ 31.02%	5.17E-31
varscan	PCP	Primary novel variants - Found variants	0.00% ~ 1.38%	2.29E-01
varscan	PCP	Primary novel variants - Lost variants	0.00% ~ 31.02%	1.94E-31
varscan	PCP	Primary novel variants - Novel variants	0.00% ~ 0.00%	1.00E+00
varscan	TSS	Lost variants - Found variants	27.98% ~ 4.85%	1.43E-06
varscan	TSS	Novel variants - Found variants	3.33% ~ 4.85%	7.18E-01
varscan	TSS	Novel variants - Lost variants	3.33% ~ 27.98%	4.92E-07
varscan	TSS	Primary novel variants - Found variants	88.89% ~ 4.85%	2.31E-22
varscan	TSS	Primary novel variants - Lost variants	88.89% ~ 27.98%	7.76E-17
varscan	TSS	Primary novel variants - Novel variants	88.89% ~ 3.33%	2.13E-22

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate; LCR — low-complexity region; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S23 — Mixture analysis: Paired comparison of mean distance to the amplicon's edge per sample in different macro variants, depending on the VCM chosen. Pairwise t-test with FDR

VCM	Pipeline	Comparison	Means	*p*-value
freebayes	PCP	Lost variants - Found variants	19.00 ~ 25.21	6.10E-02
freebayes	PCP	Novel variants - Found variants	25.64 ~ 25.21	8.91E-01
freebayes	PCP	Novel variants - Lost variants	25.64 ~ 19.00	6.05E-02
freebayes	PCP	Primary novel variants - Found variants	9.79 ~ 25.21	2.99E-05
freebayes	PCP	Primary novel variants - Lost variants	9.79 ~ 19.00	1.04E-02
freebayes	PCP	Primary novel variants - Novel variants	9.79 ~ 25.64	2.99E-05
freebayes	TSS	Lost variants - Found variants	19.31 ~ 24.88	2.79E-03
freebayes	TSS	Novel variants - Found variants	30.42 ~ 24.88	2.79E-03
freebayes	TSS	Novel variants - Lost variants	30.42 ~ 19.31	1.53E-07
freebayes	TSS	Primary novel variants - Found variants	24.03 ~ 24.88	6.63E-01
freebayes	TSS	Primary novel variants - Lost variants	24.03 ~ 19.31	2.26E-02
freebayes	TSS	Primary novel variants - Novel variants	24.03 ~ 30.42	2.79E-03
mutserve	PCP	Lost variants - Found variants	17.78 ~ 24.72	6.41E-04
mutserve	PCP	Novel variants - Found variants	9.61 ~ 24.72	1.79E-08
mutserve	PCP	Novel variants - Lost variants	9.61 ~ 17.78	6.33E-04
mutserve	PCP	Primary novel variants - Found variants	2.00 ~ 24.72	1.78E-13
mutserve	PCP	Primary novel variants - Lost variants	2.00 ~ 17.78	4.05E-09
mutserve	PCP	Primary novel variants - Novel variants	2.00 ~ 9.61	2.16E-03
mutserve	TSS	Lost variants - Found variants	18.10 ~ 25.09	2.89E-02
mutserve	TSS	Novel variants - Found variants	30.09 ~ 25.09	1.22E-01
mutserve	TSS	Novel variants - Lost variants	30.09 ~ 18.10	6.46E-04
mutserve	TSS	Primary novel variants - Found variants	13.67 ~ 25.09	2.52E-03
mutserve	TSS	Primary novel variants - Lost variants	13.67 ~ 18.10	1.90E-01
mutserve	TSS	Primary novel variants - Novel variants	13.67 ~ 30.09	9.79E-05
varscan	PCP	Lost variants - Found variants	19.25 ~ 25.59	1.74E-03
varscan	PCP	Novel variants - Found variants	7.61 ~ 25.59	2.90E-10
varscan	PCP	Novel variants - Lost variants	7.61 ~ 19.25	3.65E-06
varscan	PCP	Primary novel variants - Found variants	2.00 ~ 25.59	6.01E-14
varscan	PCP	Primary novel variants - Lost variants	2.00 ~ 19.25	2.90E-10
varscan	PCP	Primary novel variants - Novel variants	2.00 ~ 7.61	2.13E-02
varscan	TSS	Lost variants - Found variants	19.80 ~ 25.09	4.51E-02
varscan	TSS	Novel variants - Found variants	30.94 ~ 25.09	4.08E-02
varscan	TSS	Novel variants - Lost variants	30.94 ~ 19.80	2.14E-04
varscan	TSS	Primary novel variants - Found variants	13.67 ~ 25.09	6.86E-04
varscan	TSS	Primary novel variants - Lost variants	13.67 ~ 19.80	4.51E-02
varscan	TSS	Primary novel variants - Novel variants	13.67 ~ 30.94	2.55E-06

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.



Figure S5 — Performance analysis in PCP and TSS

(**A, C, E, G**) Performance metric variant classification per pipeline for Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B datasets, respectively; (**B, D, F, H**) Performance metrics per pipeline for Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B datasets, respectively. Error bars denote minimum and maximum values. Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

Table S24 — Performance analysis: Paired difference between different performance indicators for Grade A/B variants, depending on the VCM chosen. Bootstrap with FDR for the primary analysis and paired t-test with FDR for the mixture analysis

VCM	Dataset	Grade	Performance indicator	Mean TSS-PCP paired difference [95% CI]	*p*-value	*p*-value (adj)
freebayes	primary	A	Sensitivity	1.45% [0.00–4.35%]	5.15E-01	5.80E-01
			Specificity	-0.01% [-0.02–0.00%]	7.63E-02	1.57E-01
			Precision	-5.20% [-12.39–0.00%]	7.29E-02	1.57E-01
			F_1 score	-2.23% [-6.96–0.26%]	5.13E-01	5.80E-01
		B	Sensitivity	-5.82% [-8.62–-3.70%]	0.00E+00	0.00E+00
			Specificity	-0.01% [-0.02–0.00%]	7.28E-02	1.57E-01
			Precision	-4.39% [-10.22–0.00%]	7.85E-02	1.57E-01
			F_1 score	-5.99% [-7.07–-4.05%]	0.00E+00	0.00E+00
	mixture	A	Sensitivity	-7.64% [-21.53–6.25%]	2.58E-01	2.58E-01
			Specificity	-0.01% [-0.01–0.00%]	3.89E-03	2.39E-02
			Precision	-9.39% [-17.58–-1.19%]	2.77E-02	5.02E-02
			F_1 score	-7.81% [-19.01–3.39%]	1.57E-01	1.98E-01
		B	Sensitivity	-6.39% [-11.86–-0.93%]	2.49E-02	4.98E-02
			Specificity	-0.01% [-0.01–0.00%]	3.90E-03	2.39E-02
			Precision	-7.12% [-11.30–-2.95%]	2.58E-03	2.39E-02
			F_1 score	-7.06% [-12.83–-1.29%]	2.00E-02	4.81E-02
mutserve	primary	A	Sensitivity	-1.71% [-5.13–0.00%]	5.21E-01	5.80E-01
			Specificity	0.00% [-0.01–0.01%]	7.37E-01	7.40E-01
			Precision	-1.32% [-6.45–2.50%]	5.13E-01	5.80E-01
			F_1 score	-1.57% [-3.33–0.00%]	7.44E-02	1.57E-01
		B	Sensitivity	-8.16% [-13.11–-5.13%]	0.00E+00	0.00E+00
			Specificity	0.00% [-0.01–0.01%]	7.40E-01	7.40E-01
			Precision	-1.46% [-6.45–2.08%]	5.13E-01	5.80E-01
			F_1 score	-6.76% [-8.24–-5.71%]	0.00E+00	0.00E+00

Table S24 — Performance analysis: Paired difference between different performance indicators for Grade A/B variants, depending on the VCM chosen. Bootstrap with FDR for the primary analysis and paired t-test with FDR for the mixture analysis (continued)

VCM	Dataset	Grade	Performance indicator	Mean TSS-PCP paired difference [95% CI]	*p*-value	*p*-value (adj)
varscan	mixture	A	Sensitivity	-10.45% [-20.93–0.04%]	5.07E-02	7.16E-02
			Specificity	0.00% [-0.01–0.00%]	2.50E-01	2.58E-01
			Precision	-4.25% [-8.10–-0.40%]	3.29E-02	5.02E-02
			F_1 score	-7.44% [-15.11–0.23%]	5.62E-02	7.50E-02
		B	Sensitivity	-8.84% [-14.66–-3.02%]	5.71E-03	2.39E-02
			Specificity	0.00% [-0.01–0.00%]	2.50E-01	2.58E-01
			Precision	-4.27% [-7.84–-0.69%]	2.28E-02	4.98E-02
			F_1 score	-8.55% [-14.33–-2.76%]	6.85E-03	2.39E-02
	primary	A	Sensitivity	7.50% [-2.56–18.18%]	7.09E-02	1.57E-01
			Specificity	0.00% [-0.01–0.00%]	5.25E-01	5.80E-01
			Precision	-2.17% [-6.45–0.00%]	7.28E-02	1.57E-01
			F_1 score	2.97% [-1.33–10.00%]	5.32E-01	5.80E-01
		B	Sensitivity	-2.81% [-11.86–3.45%]	5.16E-01	5.80E-01
			Specificity	0.00% [-0.01–0.00%]	5.17E-01	5.80E-01
			Precision	-2.27% [-6.45–0.00%]	7.43E-02	1.57E-01
			F_1 score	-2.38% [-8.00–3.54%]	2.98E-01	5.49E-01
	mixture	A	Sensitivity	-6.40% [-16.75–3.96%]	2.06E-01	2.36E-01
			Specificity	-0.01% [-0.01–0.00%]	1.04E-02	2.78E-02
			Precision	-5.03% [-8.45–-1.61%]	6.98E-03	2.39E-02
			F_1 score	-4.87% [-12.74–3.00%]	2.05E-01	2.36E-01
		B	Sensitivity	-7.02% [-13.23–-0.80%]	2.96E-02	5.02E-02
			Specificity	-0.01% [-0.01–0.00%]	1.04E-02	2.78E-02
			Precision	-5.08% [-8.27–-1.89%]	4.16E-03	2.39E-02
			F_1 score	-6.73% [-12.86–-0.61%]	3.34E-02	5.02E-02

Significant *p*-values are highlighted. Abbreviations: adj — adjusted; CI — confidence interval; FDR — false discovery rate; PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software; VCM — variant calling method.

Table S25 — Ion Torrent sequencing of an independent set of 50 clinical samples: Overall look on differences between PCP and TSS

ID	PCP								TSS					
	Mappability >30 reads/base	Mappability >100 reads/base	Haplogroup	Haplogroup quality (%)	Contamination	Contamination level	Mean coverage	Number of variants	Haplogroup	Haplogroup quality (%)	Contamination	Contamination level	Mean coverage	Number of variants
1	99.95%	98.05%	H5a1c1a	96.27%	No	-	2417.09	18	H5a1c1a	96.27%	No	-	3282.00	16
2	99.88%	95.78%	H1	81.75%	No	-	1310.67	503	H1	90.08%	Yes	0.30%	1895.00	14
3	99.93%	96.92%	H	86.08%	No	-	1833.92	355	H7a1	92.25%	No	-	2717.00	14
4	99.54%	96.82%	H7	93.82%	No	-	1852.59	95	H7	93.82%	No	-	2524.00	10
5	99.05%	98.60%	V10b	93.92%	No	-	1820.96	37	V10b	93.92%	No	-	2479.00	17
6	99.95%	99.71%	H1+152	83.47%	No	-	1215.88	26	H1+152	83.47%	No	-	1823.00	14
7	99.90%	98.02%	T1a2*1	94.78%	No	-	2073.41	79	T1a2*1	95.32%	No	-	2866.00	37
8	99.82%	95.96%	L1b1a16	94.75%	No	-	1289.37	109	L1b1a16	94.65%	Yes	0.60%	2271.00	80
9	99.98%	99.95%	L3e1a2	97.54%	No	-	1393.70	55	L3e1a2	98.63%	No	-	1891.00	40
10	99.99%	99.94%	H80	96.54%	No	-	1575.96	15	H80	96.54%	No	-	2080.00	13
11	99.30%	97.91%	J1c2	94.04%	No	-	1048.60	74	J1c2	94.04%	No	-	1480.00	36
12	99.92%	98.94%	H1ak1	90.53%	No	-	1219.96	29	H1ak1	90.53%	No	-	1830.00	14
13	99.95%	98.79%	K1a4a1	94.44%	No	-	1216.21	41	K1a4a1	94.44%	No	-	1771.00	36
14	99.41%	98.58%	X2b+226	89.97%	No	-	1577.57	82	X2b	90.33%	No	-	2355.00	32
15	99.96%	99.60%	H1a	97.22%	No	-	1546.07	23	H1a	97.22%	No	-	2279.00	13
16	99.95%	99.20%	L1b1a12a	98.96%	No	-	1529.49	86	L1b1a12a	97.62%	No	-	2271.00	77
17	99.95%	99.81%	I2'3	95.13%	No	-	1687.41	35	I2'3	95.13%	Yes	0.40%	2659.00	32
18	99.23%	98.65%	L3e4a	98.57%	No	-	2125.65	32	L3e4a	98.57%	No	-	3039.00	29
19	99.91%	98.57%	L3b2a	95.89%	No	-	2367.50	59	L3b2a	95.89%	No	-	3411.00	38
20	99.96%	99.28%	I1a1	96.20%	No	-	2061.17	46	I1a1	96.22%	No	-	3108.00	37
21	99.89%	99.29%	H	82.33%	Yes	12.60%	2110.02	39	R0	88.61%	Yes	12.40%	3085.00	38
22	99.93%	99.23%	H1c	95.40%	No	-	1703.82	12	H1c	95.40%	Yes	0.70%	2616.00	11
23	99.90%	98.84%	L2a1c6	97.35%	No	-	1541.49	58	L2a1c6	97.35%	No	-	2324.00	57
24	99.65%	98.94%	J1b1a1	99.80%	No	-	2058.41	49	J1b1a1	98.20%	No	-	3192.00	37
25	99.68%	98.89%	H1e1a	100.00%	No	-	1636.08	16	H1e1a	100.00%	No	-	2627.00	12
26	99.95%	99.22%	I1a1	96.80%	No	-	1394.89	40	I1a1	95.01%	No	-	2168.00	37
27	99.96%	99.30%	H1	91.45%	No	-	1849.73	15	H1	91.45%	No	-	2705.00	13

Table S25 — Ion Torrent sequencing of an independent set of 50 clinical samples: Overall look on differences between PCP and TSS (continued)

ID	PCP								TSS					
	Mappability >30 reads/base	Mappability >100 reads/base	Haplogroup	Haplogroup quality (%)	Contamination	Contamination level	Mean coverage	Number of variants	Haplogroup	Haplogroup quality (%)	Contamination	Contamination level	Mean coverage	Number of variants
28	98.95%	94.82%	H1+16189	100.00%	No	-	1158.02	31	H1+16189	100.00%	Yes	0.80%	2497.00	10
29	99.94%	98.86%	U6a2c	92.83%	No	-	1562.96	40	U6a2c	91.44%	No	-	2360.00	32
30	99.96%	99.57%	H1	93.15%	No	-	1752.64	22	H1	93.15%	No	-	2584.00	11
31	99.97%	99.86%	U5b3	100.00%	No	-	2143.96	30	U5b3	100.00%	No	-	3176.00	26
32	99.97%	99.90%	H18	100.00%	No	-	2053.76	22	H18	100.00%	Yes	1.30%	3077.00	11
33	99.96%	99.89%	H1ba	89.65%	No	-	1702.02	30	H1	92.86%	No	-	2399.00	11
34	99.95%	99.51%	L2a1b1a	98.12%	No	-	1822.10	92	L2a1b1a	98.45%	No	-	3088.00	54
35	99.93%	97.59%	K1a+195	89.99%	No	-	2380.44	44	K1a3a	93.79%	No	-	3221.00	38
36	99.98%	99.89%	H5a4a1	91.99%	No	-	2239.32	38	H5a4a1	91.20%	No	-	3082.00	16
37	99.98%	99.89%	H1r	100.00%	Yes	3.30%	2871.44	34	H1r	100.00%	No	-	3906.00	10
38	99.97%	99.87%	U5a1a1	96.65%	No	-	2610.27	33	U5a1a1	96.65%	No	-	3479.00	29
39	99.96%	99.82%	J2b1a	96.74%	No	-	1864.20	45	J2b1a	96.74%	No	-	2513.00	33
40	99.31%	98.58%	L2b2a	98.39%	No	-	2917.67	85	L2b2a	99.33%	Yes	0.50%	3986.00	73
41	99.97%	99.20%	T1a5a	90.88%	No	-	1966.95	39	T1a5a	92.30%	No	-	2700.00	37
42	99.37%	98.18%	T2b3+151	94.86%	No	-	2410.06	47	T2b3+151	94.86%	Yes	0.40%	3254.00	38
43	99.98%	99.29%	K2a6	95.59%	No	-	2339.70	45	K2a6	98.22%	No	-	3212.00	32
44	99.99%	99.97%	H11a	94.57%	Yes	7.00%	2760.81	78	H11a	94.57%	No	-	3595.00	21
45	99.99%	99.98%	T1a1+@152	94.41%	No	-	2808.41	48	T1a1+@152	94.41%	Yes	0.60%	3717.00	37
46	99.99%	99.98%	J1c1b	96.39%	No	-	2777.32	47	J1c1b	96.39%	No	-	3683.00	31
47	99.94%	96.71%	H31	96.46%	No	-	1436.10	25	H31	96.46%	No	-	1983.00	13
48	99.97%	99.87%	W5	94.82%	No	-	2669.32	91	W5	94.82%	Yes	1.60%	3665.00	38
49	99.90%	97.68%	L2b1a	95.62%	No	-	1600.22	77	L2b1a	95.90%	No	-	2170.00	75
50	99.89%	97.37%	L3b1a1*	96.58%	No	-	1003.83	37	L3b1a1*	97.04%	No	-	1373.00	36

Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

Table S26 — Ion Torrent sequencing of an independent set of 50 clinical samples: Discordant variants between PCP and TSS

Mutation	Classification	*N*	Position blacklisted?	Mean VL	Mean normalized coverage	Coverage ratio	NUMT classification	Mean number of NUMTs	LCR	Distance to amplicon edge	HVS?	Poly-C?	*N* as homoplasmic variant in Helixmtdb	*N* as heteroplasmic variant in Helixmtdb
8649C	PCP only	*25*		3.96%	0.30	1.88	pNUMT	3.00	No	2				
5821A	PCP only	*21*		4.29%	2.15	1.18	NUMT (both)	16.00	No	22			*339*	*25*
4318T	PCP only	*20*		4.20%	0.47	5.01	NUMT (both)	10.50	No	2			*7*	*0*
5840T	PCP only	*20*		4.23%	2.18	1.10	NUMT (both)	21.00	No	18			*17*	*3*
10958C	TSS only	*17*		21.66%	0.17	2.87	pNUMT	1.00	No	27				
6569A	PCP only	*16*		4.21%	0.32	4.63	NUMT (both)	11.50	No	48				
1556T	PCP only	*13*		4.12%	1.56	1.20	NUMT (both)	16.00	No	7			*1*	*2*
8152A	PCP only	*13*		3.42%	0.76	19.95	NUMT (both)	4.50	No	35			*298*	*33*
3172A	PCP only	*7*		3.20%	0.90	3.06	pNUMT	1.00	No	10			*30*	*0*
4253A	PCP only	*6*		2.73%	0.55	7.87	NUMT (both)	1.00	No	37				
7957T	PCP only	*6*		3.55%	0.57	3.21	NUMT (both)	1.50	No	37			*8*	*1*
7959A	PCP only	*6*		3.23%	0.25	7.89	pNUMT	1.00	No	35				
8140T	PCP only	*6*		3.05%	1.32	1.27	NUMT (both)	2.00	No	23			*12*	*0*
8943T	PCP only	*6*		2.93%	0.71	1.56	NUMT (both)	5.50	No	51			*41*	*5*
9434G	PCP only	*6*		3.05%	0.84	1.70	NUMT (both)	3.50	No	34			*1*	*1*
14715C	PCP only	*6*		3.43%	0.39	21.15	NUMT (both)	4.50	No	1				
4456T	PCP only	*5*		3.26%	0.86	1.44	NUMT (both)	8.50	No	31			*12*	*6*
6542T	PCP only	*5*		4.10%	0.45	1.56	NUMT (both)	9.50	No	21			*23*	*0*
7146G	PCP only	*5*		2.74%	1.09	1.86	NUMT (both)	7.50	No	33			*1174*	*282*
8065A	PCP only	*5*		3.88%	2.29	1.12	NUMT (both)	7.00	No	41			*330*	*16*
11677A	PCP only	*5*		2.86%	0.31	4.69	NUMT (both)	2.50	No	11			*4*	*0*
14429A	PCP only	*5*		3.24%	0.26	10.45	NUMT (both)	4.50	No	5				*1*
204C	PCP only	*4*		51.45%	0.89	1.65	NUMT (both)	1.50	No	43	HVS-II		*11099*	*696*
498A	PCP only	*4*		3.55%	0.15	11.53	NUMT (both)	1.00	No	18	HVS-III			
1905A	PCP only	*4*		3.45%	0.63	5.63	NUMT (both)	2.50	No	35				
2463G	PCP only	*4*		99.48%	0.68	129.66	NUMT (both)	1.00	No	29				*1*
3913T	PCP only	*4*		2.95%	0.45	7.22	NUMT (mtDNA-server)	1.00	No	18				

Table S26 — Ion Torrent sequencing of an independent set of 50 clinical samples: Discordant variants between PCP and TSS (continued)

Mutation	Classification	*N*	Position blacklisted?	Mean VL	Mean normalized coverage	Coverage ratio	NUMT classification	Mean number of NUMTs	LCR	Distance to amplicon edge	HVS?	Poly-C?	*N* as homoplasmic variant in Helixmtdb	*N* as heteroplasmic variant in Helixmtdb
4048A	PCP only	*4*		2.98%	0.66	1.38	NUMT (both)	1.00	No	44			*677*	*45*
4312T	PCP only	*4*		3.08%	1.21	1.17	NUMT (both)	14.00	No	8			*419*	*7*
5320T	PCP only	*4*		3.63%	0.32	1.85	NUMT (both)	2.00	No	17			*13*	*0*
5752G	PCP only	*4*		98.83%	0.42	54.48	NUMT (mtDNA-server)	2.00	No	36				
6410T	PCP only	*4*		3.90%	0.81	1.53	NUMT (both)	15.50	No	3			*34*	*0*
7521A	PCP only	*4*		3.35%	0.94	3.63	NUMT (both)	9.00	No	31			*2209*	*1190*
8021G	PCP only	*4*		4.03%	2.25	1.28	NUMT (both)	6.00	No	16			*20*	*1*
8152T	PCP only	*4*		3.68%	0.60	12.32	NUMT (both)	4.50	No	35			*3*	*0*
8167C	PCP only	*4*		3.15%	1.39	1.24	NUMT (both)	3.00	No	50			*200*	*15*
8856T	PCP only	*4*		3.13%	0.32	1.46	NUMT (both)	7.00	No	25			*4*	*0*
544A	PCP only	*3*		2.97%	0.13	14.95	NUMT (both)	1.00	No	1	HVS-III			*2*
1520C	PCP only	*3*		3.20%	1.66	1.33	NUMT (both)	6.00	No	43			*54*	*2*
1536G	PCP only	*3*		4.33%	1.64	1.18	NUMT (both)	6.00	No	27			*160*	*4*
1883A	PCP only	*3*		3.23%	1.07	1.19	NUMT (both)	13.50	No	13				
1888A	PCP only	*3*		3.97%	1.10	1.18	NUMT (both)	13.00	No	18			*18596*	*59*
1977C	PCP only	*3*		3.27%	1.07	1.15	NUMT (both)	13.50	No	0			*89*	*5*
3897A	PCP only	*3*		3.20%	0.35	12.89	NUMT (mtDNA-server)	14.00	No	2			*3*	*2*
4141A	PCP only	*3*		2.90%	0.28	8.47	NUMT (both)	3.50	No	29				
4883A	PCP only	*3*		3.70%	0.13	2.55	NUMT (both)	7.50	No	48				
5235A	PCP only	*3*		3.87%	0.10	1.28	pNUMT	3.00	No	7				
5306A	PCP only	*3*		3.03%	0.24	1.38	NUMT (both)	2.00	No	31				
5491A	PCP only	*3*		3.17%	0.50	7.15	pNUMT	4.00	No	26				
5756T	PCP only	*3*		3.17%	0.58	2.64	NUMT (mtDNA-server)	12.00	No	32				
6366A	PCP only	*3*		4.13%	0.89	1.91	NUMT (both)	4.50	No	47			*176*	*12*
6383A	PCP only	*3*		4.13%	0.88	1.85	NUMT (both)	6.00	No	30			*239*	*11*
7232T	PCP only	*3*		3.70%	2.18	1.21	NUMT (both)	10.00	No	1			*3*	*2*
7316A	PCP only	*3*		3.27%	1.70	1.29	NUMT (both)	6.00	No	21			*25*	*8*
8503C	PCP only	*3*		4.00%	0.23	5.14	NUMT (both)	3.00	No	45			*237*	*7*

Table S26 — Ion Torrent sequencing of an independent set of 50 clinical samples: Discordant variants between PCP and TSS (continued)

Mutation	Classification	*N*	Position blacklisted?	Mean VL	Mean normalized coverage	Coverage ratio	NUMT classification	Mean number of NUMTs	LCR	Distance to amplicon edge	HVS?	Poly-C?	*N* as homoplasmic variant in Helixmtdb	*N* as heteroplasmic variant in Helixmtdb
8701G	PCP only	*3*		4.37%	0.08	4.10	NUMT (both)	2.50	No	43			*17154*	*47*
8718G	PCP only	*3*		3.97%	0.08	2.60	NUMT (both)	3.00	No	51			*189*	*7*
9060A	PCP only	*3*		3.27%	0.60	1.09	NUMT (both)	4.00	No	49				
9075T	PCP only	*3*		3.73%	0.47	1.61	NUMT (both)	4.00	No	36			*13*	*0*
9629G	PCP only	*3*		2.83%	1.45	1.08	NUMT (both)	3.00	Yes	11			*49*	*5*
10635T	PCP only	*3*		3.13%	0.23	1.59	NUMT (mtDNA-server)	1.00	No	19				
11657T	PCP only	*3*		3.43%	0.71	1.80	pNUMT	1.00	No	31				
12684A	PCP only	*3*		3.43%	1.46	1.36	NUMT (both)	10.00	No	22			*155*	*40*
12971A	PCP only	*3*		2.60%	0.71	3.84	pNUMT	1.00	No	2				
13062G	PCP only	*3*		3.63%	0.45	17.24	NUMT (both)	8.00	No	12			*11*	*44*
13095C	PCP only	*3*		3.73%	1.04	2.58	NUMT (both)	2.50	No	11			*290*	*50*
13105G	PCP only	*3*		3.63%	1.08	2.48	NUMT (both)	11.00	No	21			*4882*	*106*
14159A	PCP only	*3*		3.43%	0.14	9.43	NUMT (both)	1.50	No	7				
14248A	PCP only	*3*		3.93%	0.13	6.47	NUMT (both)	5.50	No	27				
14531A	PCP only	*3*		4.77%	0.07	4.17	pNUMT	1.00	No	6				
15518A	PCP only	*3*		3.23%	0.17	1.71	NUMT (mtDNA-server)	3.00	No	52			*1*	*0*
16183C	PCP only	*3*		75.77%	0.11	7.72	pNUMT	5.00	Yes	38	HVS-I			
34T	PCP only	*2*		2.60%	1.31	1.25	NUMT (both)	11.00	No	19				
203A	PCP only	*2*		99.20%	0.43	5.45	NUMT (both)	1.50	Yes	44	HVS-II		*1201*	*21*
709A	PCP only	*2*		98.90%	1.13	1.43	NUMT (both)	11.00	No	0			*28750*	*168*
710C	PCP only	*2*		99.35%	1.04	1.42	NUMT (both)	16.50	No	1			*998*	*5*
1120T	PCP only	*2*		2.85%	0.64	3.13	NUMT (both)	10.50	No	2			*67*	*3*
1462A	PCP only	*2*		6.65%	0.62	1.04	NUMT (both)	14.50	No	1			*584*	*12*
1476A	PCP only	*2*		2.55%	1.57	1.29	NUMT (both)	7.50	No	2				
1488A	PCP only	*2*		2.70%	1.42	1.51	NUMT (both)	7.50	No	14				
1518T	PCP only	*2*		2.80%	1.60	1.23	NUMT (both)	7.50	No	44			*18*	*1*
1534A	PCP only	*2*		2.85%	1.36	1.22	pNUMT	11.00	No	29				
1944T	PCP only	*2*		3.65%	1.11	1.20	NUMT (both)	2.50	No	33			*1*	*0*

Table S26 — Ion Torrent sequencing of an independent set of 50 clinical samples: Discordant variants between PCP and TSS (continued)

Mutation	Classification	*N*	Position blacklisted?	Mean VL	Mean normalized coverage	Coverage ratio	NUMT classification	Mean number of NUMTs	LCR	Distance to amplicon edge	HVS?	Poly-C?	*N* as homoplasmic variant in Helixmtdb	*N* as heteroplasmic variant in Helixmtdb
2219A	PCP only	*2*		2.75%	0.21	2.00	NUMT (mtDNA-server)	3.00	No	20			*1*	*0*
2772A	PCP only	*2*		2.60%	0.87	1.33	NUMT (mtDNA-server)	2.00	No	0			*5*	*0*
3914T	PCP only	*2*		2.75%	0.71	1.55	Non-NUMT		No	19				
4463A	PCP only	*2*		3.10%	0.64	1.11	NUMT (both)	5.00	No	24				
4479A	PCP only	*2*		3.50%	0.57	2.50	NUMT (both)	2.00	No	8				
4856C	PCP only	*2*		2.85%	0.38	2.57	NUMT (both)	2.00	No	21			*204*	*6*
5111T	PCP only	*2*		2.90%	0.78	1.68	NUMT (both)	1.50	No	16			*19*	*0*
5147A	PCP only	*2*		3.05%	0.27	2.27	NUMT (both)	4.50	No	9			*8213*	*664*
5580C	PCP only	*2*		2.60%	1.06	1.13	NUMT (both)	4.50	No	6			*383*	*29*
6173A	PCP only	*2*		2.70%	0.83	4.27	NUMT (both)	2.50	No	4				
6381T	PCP only	*2*		3.00%	0.29	6.50	NUMT (both)	3.00	No	32				
6424A	PCP only	*2*		3.80%	0.30	8.86	NUMT (both)	3.00	No	0				
6641C	PCP only	*2*		2.60%	2.09	1.23	NUMT (both)	8.00	No	4			*42*	*15*
7256T	PCP only	*2*		3.35%	2.08	1.12	NUMT (both)	11.00	No	23			*4195*	*151*
7705C	PCP only	*2*		3.50%	1.21	1.29	NUMT (both)	3.50	No	48			*1158*	*58*
8276A	PCP only	*2*		2.70%	0.53	20.71	NUMT (both)	2.00	No	48			*10*	*1*
8455T	PCP only	*2*		3.70%	0.23	7.46	NUMT (both)	2.00	No	15			*109*	*2*
8545A	PCP only	*2*		3.65%	0.26	13.28	NUMT (both)	4.00	No	3			*138*	*16*
8677C	PCP only	*2*		3.20%	0.09	4.50	NUMT (both)	4.50	No	19				
9055T	PCP only	*2*		2.90%	0.45	2.06	pNUMT	2.00	No	44				*1*
9478A	PCP only	*2*		4.40%	1.04	7.46	NUMT (both)	3.50	No	1				
9573A	PCP only	*2*		2.90%	0.20	2.16	pNUMT	1.00	No	5				
10196A	PCP only	*2*		3.10%	0.61	21.24	NUMT (mtDNA-server)	1.00	No	34				
10385G	PCP only	*2*		99.30%	0.84	95.78	NUMT (mtDNA-server)	1.00	No	8			*4*	*1*
11063A	PCP only	*2*		2.85%	0.46	2.23	pNUMT	1.00	No	35				
11093T	PCP only	*2*		3.40%	0.24	1.07	pNUMT	1.00	No	5				
11237A	PCP only	*2*		2.85%	0.18	2.39	pNUMT	3.00	No	52				
11708G	PCP only	*2*		2.70%	1.21	1.74	NUMT (both)	2.00	No	9			*6*	*3*

Table S26 — Ion Torrent sequencing of an independent set of 50 clinical samples: Discordant variants between PCP and TSS (continued)

Mutation	Classification	*N*	Position blacklisted?	Mean VL	Mean normalized coverage	Coverage ratio	NUMT classification	Mean number of NUMTs	LCR	Distance to amplicon edge	HVS?	Poly-C?	*N* as homoplasmic variant in Helixmtdb	*N* as heteroplasmic variant in Helixmtdb
11767T	PCP only	*2*		2.60%	1.19	1.55	NUMT (both)	5.50	No	4			*9*	*1*
12705T	PCP only	*2*		3.75%	1.49	1.43	NUMT (both)	6.50	Yes	43			*35323*	*91*
12838T	PCP only	*2*		3.10%	1.29	1.20	Non-NUMT		No	30				
13359A	PCP only	*2*		3.05%	0.76	1.21	NUMT (both)	14.50	No	33			*346*	*26*
13368A	PCP only	*2*		3.40%	0.74	1.15	NUMT (both)	10.00	No	42			*17444*	*42*
13386C	PCP only	*2*		3.10%	0.77	1.16	NUMT (both)	13.50	No	56			*81*	*4*
13440T	PCP only	*2*		3.05%	0.74	1.16	NUMT (both)	4.50	No	2			*44*	*0*
13766A	PCP only	*2*		4.85%	0.06	1.44	NUMT (both)	1.00	No	12			*5*	*0*
14493A	PCP only	*2*		3.70%	0.20	18.96	NUMT (both)	4.50	No	44				
14771T	PCP only	*2*		4.10%	0.13	13.45	NUMT (both)	1.50	No	55			*8*	*0*
14777C	TSS only	*2*		12.85%	0.28	2.15	NUMT (both)	1.50	No	59				*1*
16131C	PCP only	*2*		2.70%	2.11	1.15	pNUMT	2.00	No	0	HVS-I		*60*	*18*
16380A	PCP only	*2*		2.90%	0.32	12.35	pNUMT	6.00	No	39				

Abbreviations: HVS — hypervariable segments; LCR — low-complexity region; *N* — number; NUMT(s) — nuclear insertion(s) of mitochondrial DNA; PCP — PrecisionCallerPipeline; Poly-C — poly C-stretches; TSS — Ion Torrent Suite™ Software; VL — variant level.

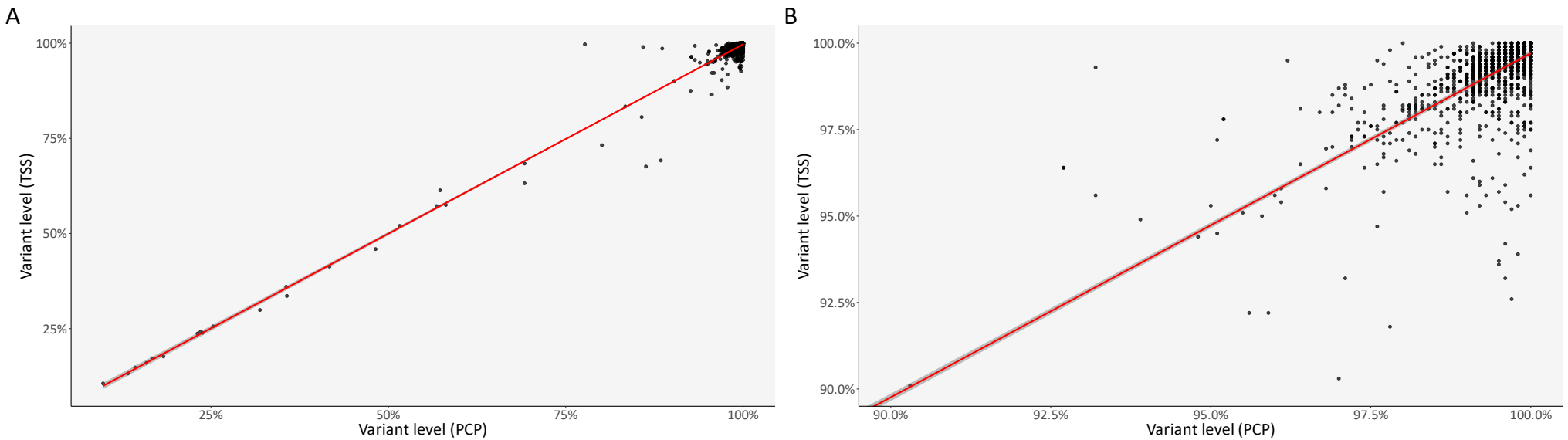


Figure S6 — Ion Torrent sequencing of an independent set of 50 clinical samples: Linear regression model between the variant level in PCP and the variant level in TSS, for variants found in both PCP and TSS

(A) Linear regression model between the variant level in PCP and the variant level in TSS for all variants found in both PCP and TSS, (**B**) Close-up of Figure S6A, focusing on homoplasmic variants. Abbreviations: PCP — PrecisionCallerPipeline; TSS — Ion Torrent Suite™ Software.

International Journal of *Molecular Sciences*

MDPI

Article

From Forensics to Clinical Research: Expanding the Variant Calling Pipeline for the Precision ID mtDNA Whole Genome Panel

Filipe Cortes-Figueiredo [1,2], Filipa S. Carvalho [1], Ana Catarina Fonseca [3,4], Friedemann Paul [2,5], José M. Ferro [3,4], Sebastian Schönherr [6], Hansi Weissensteiner [6,*] and Vanessa A. Morais [1,*]

1 VMorais Lab—Mitochondria Biology & Neurodegeneration, Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisbon, Portugal; filipe.figueiredo@medicina.ulisboa.pt (F.C.-F.); filcarvalho.uc@gmail.com (F.S.C.)
2 NeuroCure Clinical Research Center, Charité—Universitätsmedizin Berlin, 10117 Berlin, Germany; friedemann.paul@charite.de
3 José Ferro Lab—Clinical Research in Non-communicable Neurological Diseases, Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisbon, Portugal; acfonseca@medicina.ulisboa.pt (A.C.F.); jmferro@medicina.ulisboa.pt (J.M.F.)
4 Serviço de Neurologia, Hospital de Santa Maria, Centro Hospitalar Universitário Lisboa Norte, 1649-035 Lisbon, Portugal
5 Experimental and Clinical Research Center, Charité—Universitätsmedizin Berlin and Max Delbrück Center for Molecular Medicine, 13125 Berlin, Germany
6 Institute of Genetic Epidemiology, Department of Genetics and Pharmacology, Medical University of Innsbruck, 6020 Innsbruck, Austria; sebastian.schoenherr@i-med.ac.at
* Correspondence: hansi.weissensteiner@i-med.ac.at (H.W.); vmorais@medicina.ulisboa.pt (V.A.M.); Tel.: +43-0512-9003-70560 (H.W.); +351-217-999-573 (V.A.M.)

Abstract: Despite a multitude of methods for the sample preparation, sequencing, and data analysis of mitochondrial DNA (mtDNA), the demand for innovation remains, particularly in comparison with nuclear DNA (nDNA) research. The Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, USA) is an innovative library preparation kit suitable for degraded samples and low DNA input. However, its bioinformatic processing occurs in the enterprise Ion Torrent Suite™ Software (TSS), yielding BAM files aligned to an unorthodox version of the revised Cambridge Reference Sequence (rCRS), with a heteroplasmy threshold level of 10%. Here, we present an alternative customizable pipeline, the PrecisionCallerPipeline (PCP), for processing samples with the correct rCRS output after Ion Torrent sequencing with the Precision ID library kit. Using 18 samples (3 original samples and 15 mixtures) derived from the 1000 Genomes Project, we achieved overall improved performance metrics in comparison with the proprietary TSS, with optimal performance at a 2.5% heteroplasmy threshold. We further validated our findings with 50 samples from an ongoing independent cohort of stroke patients, with PCP finding 98.31% of TSS's variants (TSS found 57.92% of PCP's variants), with a significant correlation between the variant levels of variants found with both pipelines.

Keywords: mitochondrial DNA; next-generation sequencing; massively parallel sequencing; whole genome sequencing; Precision ID; Thermo Fisher Scientific; variant calling; mixture; performance metrics

check for updates

Citation: Cortes-Figueiredo, F.; Carvalho, F.S.; Fonseca, A.C.; Paul, F.; Ferro, J.M.; Schönherr, S.; Weissensteiner, H.; Morais, V.A. From Forensics to Clinical Research: Expanding the Variant Calling Pipeline for the Precision ID mtDNA Whole Genome Panel. *Int. J. Mol. Sci.* **2021**, *22*, 12031. https://doi.org/10.3390/ijms222112031

Academic Editor: Andrey V. Kuznetsov

Received: 7 October 2021
Accepted: 29 October 2021
Published: 6 November 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mitochondria are the primary source of cellular ATP, while also prominently contributing to cell survival, differentiation, and apoptosis. They contain their own double-stranded circular DNA (mtDNA), with 16,569 base pairs (bps) in humans, responsible for encoding 22 transfer RNAs, two ribosomal RNAs, and 13 essential proteins of the oxidative phosphorylation (OXPHOS) chain. In comparison with nuclear DNA (nDNA), mtDNA has a

higher copy number per cell (polyploidy), replicating independently of the nuclear genome, which allows for the existence of multiple genotypes within the same cell (heteroplasmy). Additionally, mtDNA shows a higher mutation rate due to the absence of protective histones, exposure to reactive oxygen and nitrogen species, and more rudimentary repair systems [1,2].

The peculiar nature of mtDNA has hindered the development of specific data analysis guidelines, since the large majority of datasets, guidelines, and bioinformatic pipelines for variant discovery analysis in next-generation sequencing (NGS)/massively parallel sequencing (MPS) are primarily, and sometimes exclusively, focused on nDNA [3,4], with mtDNA NGS analysis lagging years behind [5]. Out of the existing bioinformatic tools for mtDNA, with quality control assessment, variant calling, and haplogroup assignment [6–21], very few incorporate user accessibility, integrative data analysis, and regular updates. A suitable option is mtDNA-Server [17], which allows for FASTQ and BAM files as input and, considering the revised Cambridge Reference Sequence (rCRS) [22] as a reference, identifies heteroplasmic and homoplasmic variants, assigns a haplogroup with HaploGrep 2 [18], based on PhyloTree Build 17 [23] with an updated algorithm [24], and performs a contamination check and coverage analysis.

Recently, a novel approach to mtDNA NGS through whole genome sequencing (WGS) has been developed with the Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, USA) [25]. This library preparation kit, with 162 amplicons that target the whole mtDNA, is mostly used for forensic samples, achieving reliable results in often degraded samples [26–29] and with low DNA input (usually 0.1 ng of genomic DNA, but 6.25 pg [30] and, very recently, 0.6 pg [31] have been reported). Despite its most common use in forensic sciences, it has also been used in a range of other applications, from worldwide lineage studies [32], to rare mtDNA differences in monozygotic twins [33].

A disadvantage of this approach, however, is its use of a modified mtDNA reference with 16,649 bps, instead of the conventional 16,569 bps. Since the last amplicon corresponds to the last 28 bps and the first 80 bps in the widely used rCRS, the Precision ID mtDNA reference duplicates the first 80 bps of the rCRS at the end of the reference. Thus, most of its bioinformatic processing must be done opaquely in the enterprise software, Ion Torrent Suite™ Software (TSS), with the possible addition of other company-owned software [31,34–36]. In addition to producing an unorthodox reference [37], which impedes further processing in open-source bioinformatic tools, sensitivity analyses have set its heteroplasmy threshold level to 10% [30,38,39].

In this study, we aimed to establish an alternative fully customizable pipeline—PrecisionCallerPipeline (PCP)—for the Precision ID mtDNA Whole Genome Panel, which (**I**) Produced BAMs accurately mapped to the rCRS, allowing complete integration with existing open-source mtDNA pipelines; and (**II**) Revisited the TSS's heteroplasmy threshold of 10%. Our approach yielded improved performance metrics in comparison with TSS, additionally enabling the detection of heteroplasmic variants at the 2.5% threshold.

On the basis of this extensive validation process, we compared our pipeline to TSS by investigating 50 clinical samples from an ongoing stroke cohort, where we further validated our approach.

2. Results

2.1. Pipeline Validation with Samples from the 1000 Genomes Project

We acquired three DNA samples previously sequenced within the 1000 Genomes Project [4] on Illumina HiSeq (Illumina, Inc., USA), with three different sequencing runs (more details in Section 4.1. Sample acquisition/collection): (**I**) Exome sequencing—exome (mean coverage: 420.59 reads/base pair); (**II**) Low-coverage sequencing—lowCov (mean coverage 3863.37 reads/base pair); and (**III**) High-coverage sequencing—highCov (mean coverage: 15079.28 reads/base pair). Samples were then processed and mixed at five different mixtures levels, ranging from 1 to 25% (Figure 1A), and sequenced on Ion Torrent™ Ion S5™ (Thermo Fisher Scientific, USA). The original sequenced samples and their mixtures

then underwent bioinformatic analysis with our pipeline, the PrecisionCallerPipeline (PCP) (Figure 1B). For that, we aimed to optimize the protocol for the heteroplasmy threshold and the removal of nuclear insertions of mitochondrial DNA (NUMTs). Correspondingly, we ran the samples with three different NUMT removal approaches and 21 thresholds, ranging from 0.4% to 10% (Figure 1C), through the mutserve variant caller, an offline and command-line version of mtDNA-Server [17,40]. After data analysis and variant classification for primary and mixture samples (Table S1), we determined the performance metrics (Figure S1) considering both Grade A variants, which were homoplasmic in the three independent sequencing runs, and Grade B variants, which were heteroplasmic in the same three sequencing runs (more details in Section 4.4. Data analysis). Our primary outcome was the highest possible F_1 score.

NUMT removal immediately after the first alignment and before merging the BAM files (Figure 1B) was superior to the other NUMT removal approaches (Table S2), while the 2.5% heteroplasmy threshold with NUMT removal before merging yielded the highest mean F_1 score between the four different datasets—Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B—(Table S3).

Correspondingly, all BAM files were then run through the mutserve variant caller with a 2.5% threshold and NUMT removal before merging (Figure 1B). For simplicity, we denote samples by their simplified haplogroup identifier. Thus, HG00256 as H, HG01626 as T, and HG01757 as U. Similarly, names for mixtures follow an analogous pattern with the minor component being followed by its absolute mixture level and the major component. Therefore, *U0.01H* is sample HG01757 (U) at 1% mixed with sample HG00256 (H) at 99%.

Genome coverage and mappability, defined as the ability to read a base pair (bp) above a certain coverage threshold, were adequate in comparison to other sequencing runs in the primary analysis. Haplogroup assignment was also homogeneous for both PCP and the output from Ion Torrent Suite™ Software (TSS) (Table S4 and Figure 2).

Interestingly, PCP had an increased number of mutations in comparison with TSS (Figure 2B), despite a reduction in the mean coverage in PCP (Figure 2A). To better understand this phenomenon, we considered all samples analyzed on the Ion Torrent (Table S5 and Figure 3).

In PCP, coverage and mappability were uniform, haplogroup assignment was successful for mixtures up to 10%, and sample contamination was correctly detected, with 5% as the lowest detection limit (Table S5 and Figure S2). TSS correctly identified haplogroups up to 10% as well, albeit with errors in the contamination detection for unmixed samples U and T (lower than TSS's own limit of ~10%), and a higher contamination detection limit of 10% (Table S5).

In comparison with TSS, samples processed with PCP have a reduced total number of sequences (Table S5 and Figure 3A)—mean difference of 16.94%, 95% CI [15.98%–17.89%], an adjusted p-value of 1.43×10^{-17}, and a significantly lower mean coverage (Table S5 and Figure 3B)—mean difference of 25.83%, 95% [24.97%–26.70%], an adjusted p-value of 3.87×10^{-21}; paired t-tests adjusted with false discovery rate (FDR). This arises from our read trimming, NUMT removal, and quality control protocol, since, in comparison to the unaltered FASTQ files, the BAM files from TSS have the exact same number of reads and Phred score patterns (Figures 3D and 3F).

In comparison with PCP, these results indicate that TSS achieves a higher coverage at the expense of a lack of read selection from the initial FASTQ files. Despite having significantly reduced coverage and fewer reads, more variants are called with PCP than with TSS (Table S5 and Figure 3C)—mean difference of 5.83, 95% CI [1.95–9.72], an adjusted p-value of 5.60×10^{-3}; paired t-tests with FDR.

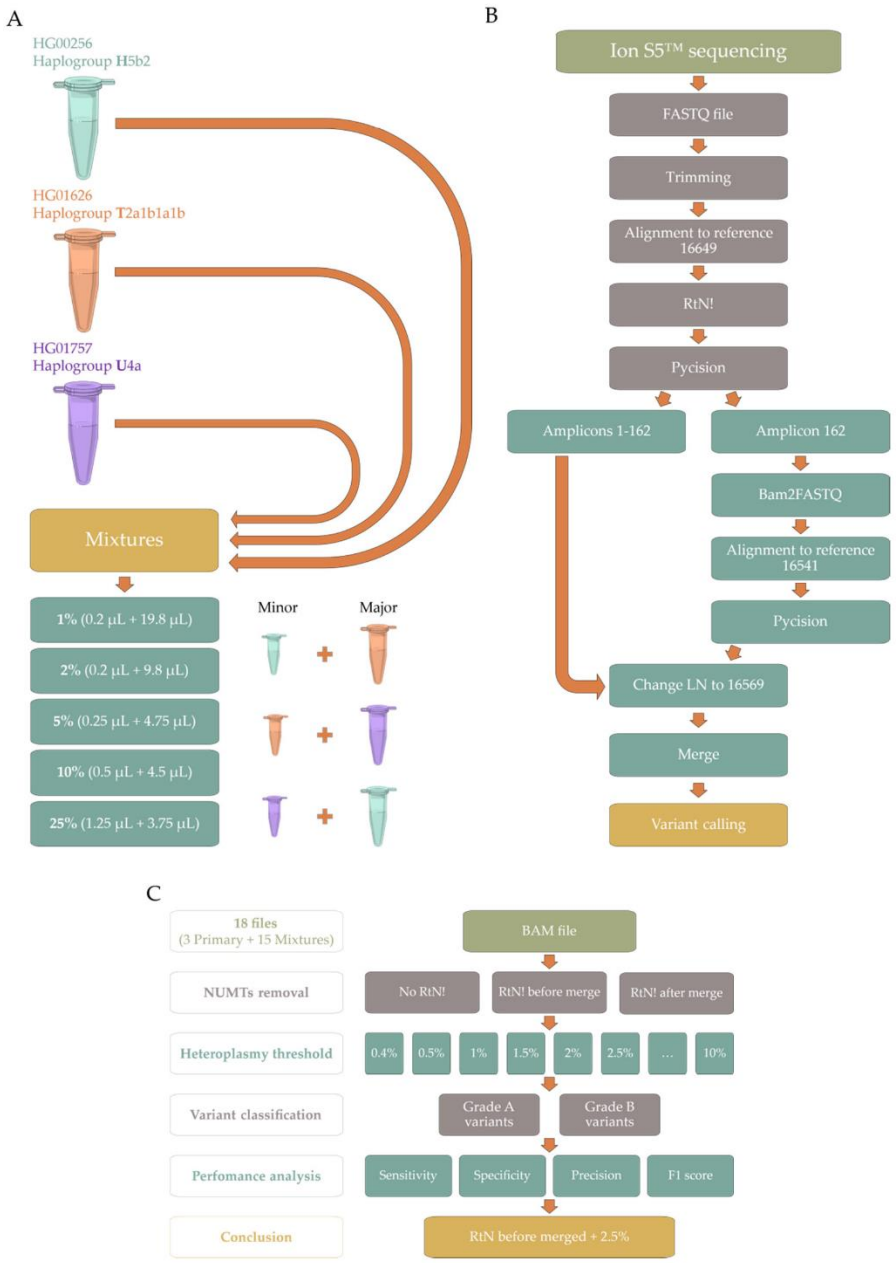


Figure 1. Schematic representation of our workflow. (**A**) Wet lab approach to performing the 15 mixtures from the Coriell Institute for Medical Research's samples, which have been analyzed within the 1000 Genomes Project—templates from Servier Medical Art (CC BY 3.0) were adapted, and are freely available at https://smart.servier.com (accessed on 1 November 2021); (**B**) Visual representation of the PrecisionCallerPipeline (PCP); (**C**) Visual representation of the optimization approach to NUMT removal and heteroplasmy threshold. Abbreviations: NUMTs—nuclear insertions of mitochondrial DNA.

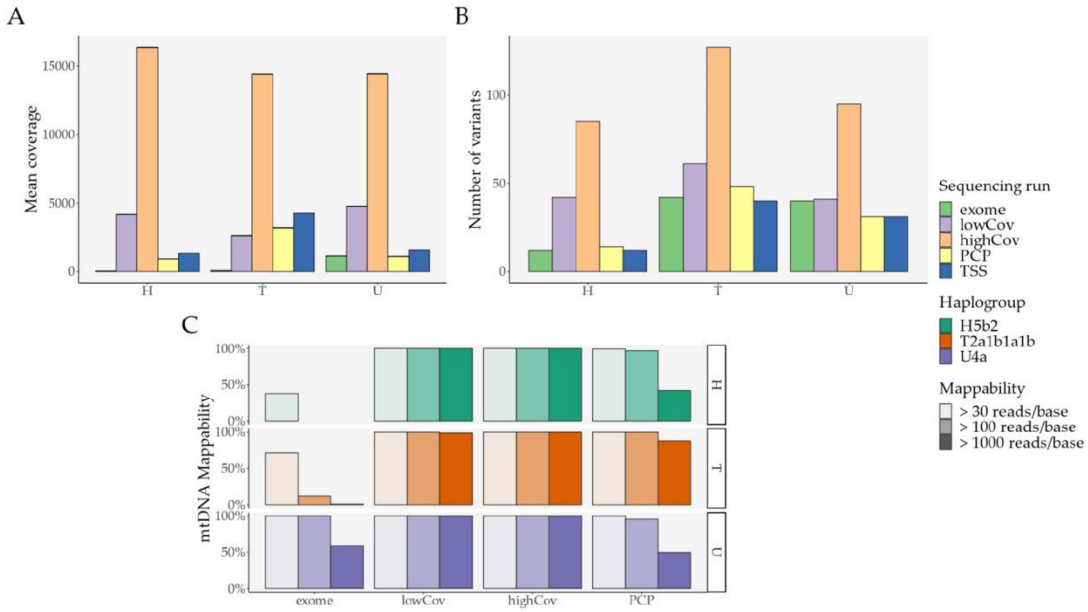


Figure 2. Primary analysis: Overall look; variant threshold was set at 0.4% for exome, lowCov, and highCov, and 2.5% for PCP. (**A**) Mean coverage per sample and per sequencing run/pipeline; (**B**) Number of variants per sequencing run/pipeline; (**C**) Mappability and haplogroup assignment per sequencing run/pipeline (TSS is not shown since it does not output raw data). Error bars denote standard error of the mean (SEM). Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software.

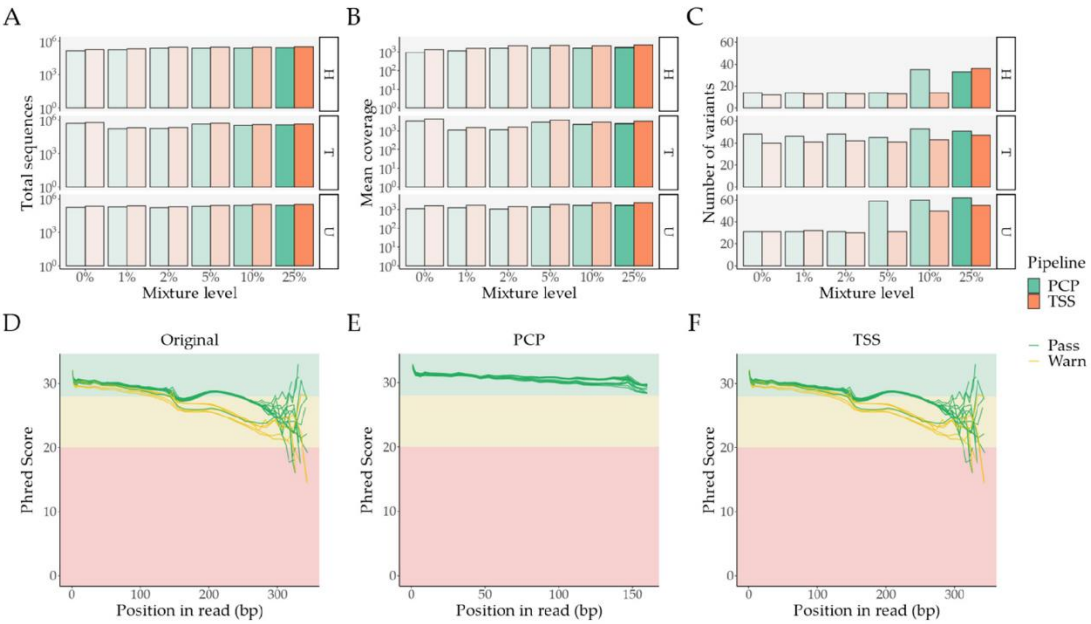


Figure 3. Ion Torrent sequencing: Overall look at differences between PCP and TSS; a 0% mixture level denotes unmixed samples. (**A**) Total number of sequences per pipeline and major sample contributor; (**B**) Mean coverage per pipeline and major sample contributor; (**C**) Number of variants per pipeline and major sample contributor; (**D–F**) MultiQC's Phred score per base pair in the unprocessed FASTQ files, PCP's BAM to FASTQ files, and TSS's BAM to FASTQ files, respectively. Abbreviations: bp—base pair; PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software.

To further dissect this difference in the number of variants, we first looked at the variants in the primary analysis (Table A1, Figure S3A, and Tables S6–S8).

When comparing variants found with the Ion Torrent sequencing method to the ones previously found on the other three Illumina runs, we observe that: (**I**) PCP significantly increases the proportion of correctly found Grade A/B variants (true positives) in contrast to lost Grade A/B variants (false negatives), while no Grade A/B lost variant with PCP was picked up by TSS (Table S6); (**II**) TSS significantly overestimates the variant level (VL) in found Grade A/B variants in comparison with PCP (Table S7); (**III**) Despite PCP picking up more variants than TSS, the difference in the proportion of "novel" variants (false positives) was not statistically significant between the two pipelines (Table S8).

We then looked at the variants in the mixture analysis (Figure 4, Table A2, Figure S3B–H, Tables S9–S14, Figure S4, Tables S15–S22). Samples with a macro classification of "Mixed status" (more details in Table S1) were excluded because of the impossibility of determining their origin.

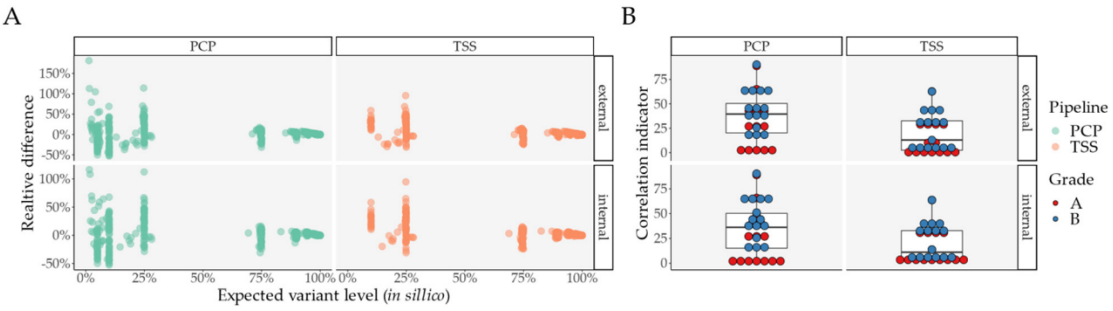


Figure 4. Mixture analysis: similarity in Grade A/B variants between expected and observed variant levels, for internal and external comparisons, per pipeline. (**A**) Relative difference between expected and observed variant level, per pipeline, and per internal and external comparison; (**B**) Correlation indicator for internal and external comparisons, per pipeline, and per variant grade. Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software.

Similar to the primary analysis, Grade A/B variants were best retrieved with PCP, in comparison with TSS, with a single variant found by TSS and lost with PCP (Table S9). Moreover, true positive variants below 11% are completely absent in TSS, while PCP correctly picks up variants until its threshold of 2.5% (Figure 4 and Table A2).

Regarding the difference between the expected and the observed VL, we defined two different approaches:

- External comparison—the difference to the other Illumina sequencing methods (exome, lowCov, and highCov);
- Internal comparison—the difference within Ion Torrent, taking the primary sequencing as a reference.

Consequently, we also considered two different relative differences, where we divided each difference by its expected level in order to observe a percent variance per VL (Figure 4A, Figure S3C–H). When performing a paired comparison regarding differences for Grade A/B variants (Table S10), we observed no significant differences between TSS and PCP for raw VL differences taking found variants into consideration. However, when we considered lost variants, TSS performed worse in the external comparison (Table S10). Regarding relative differences (Table S11), we found no statistically significant differences between PCP and TSS.

In order to assess the similarity between the expected and the observed VLs (internal and external), we also performed linear models per mixture, considering Grade A/B variants for each pipeline, for each comparison (internal and external), and for each grade variant classification (A or B), as they might have varied between them. After extracting the adjusted R^2 and the adjusted *p*-value with FDR, we created an indicator for correlation strength determined by the log10 of the ratio between the adjusted R2 and the adjusted

p-value. Thus, a greater value would indicate a stronger correlation between both variables. PCP showed a significantly increased correlation indicator in both comparisons (internal and external) for Grade B variants, while no significant differences for Grade A variants were found (Figure 4B and Table S12).

Since PCP appeared to have increased sensitivity in comparison with TSS, we then looked at the prevalence of novel variants (false positives). Interestingly, no differences were found (Table S13), albeit with TSS showing novel variants at a higher VL, namely, novel variants already present in the primary analysis (Table S14).

When we looked at the location of these novel variants in both analyses (primary and mixture), PCP had one mutation in 10/18 samples (4318T), one mutation in 8/18 samples (8649C), while the remaining four were isolated (2463G, 5752G, 6698G, 8152A). TSS had two mutations shared in 10/18 samples (310C, 10958C), one shared in 4/18 samples (14777C), one shared in 3/18 samples (8249C), and one shared in 2/18 samples (318C).

Looking into the different variant classifications (found, lost, novel, and primary novel variants), we noticed that: (**I**) PCP showed an increased rate (adjusted to the size of each region) of found variants throughout the genome, while maintaining a lower rate of lost variants (Figure S4A–B and Tables S15 and S16); (**II**) Novel variants (primary and mixture) showed the lowest normalized coverage (normalized for the mean coverage of each sample in Tables S4–S5) for both PCP and TSS, while the highest strand bias, calculated as a coverage ratio (forward vs. reverse), was observed in mixture novel variants for PCP, and in primary novel variants for TSS (Figure S4C–E and Tables S17 and S18); (**III**) Since false variants might arise from NUMTs, we calculated the mean value of reported NUMTs based on two different databases [17,41] for each position—the highest rate of NUMTs was observed in primary novel variants for PCP, and in found variants for TSS (Figure S4F and Table S19); (**IV**) Regarding the proportion of variants in positions classified as a low-complexity region (LCR) [17], lost variants showed the highest proportion of LCR variants for PCP, while the same was true in primary novel variants for TSS (Figure S4G and Table S20); (**IV**) Finally, we tested the hypothesis that the distance to the "callable" extremity of each amplicon might be significantly different, depending on each variant class. The lowest distance was observed in primary and mixture novel variants for PCP, while the same was true for primary novel variants alone for TSS, although less consistently (Figure S4H and Table S21). Table 1 offers a summary for novel variants (false positives).

Table 1. Novel variants (false positives): Summary characteristics in the mutserve variant calling method (VCM).

Variable	PCP		TSS	
	Primary Novel Variants	**Mixture Novel Variants**	**Primary Novel Variants**	**Mixture Novel Variants**
Normalized coverage	↓	↓↓	↓↓↓	↓↓
Coverage ratio	-	↑↑↑	↑↑↑	↑
Number of NUMTs	↑↑↑	↓	-	↓↓
LCR prevalence	↓↓↓	↓↓↓	↑↑↑	-
Distance to amplicon edge	↓↓↓	↓↓	↓	-

Using found variants (true positives) as a reference: "-" denotes nonsignificant changes; "↑" or "↓" denote significant changes in the 25–50% range; "↑↑" or "↓↓" denote significant changes in the 51–75% range; and "↑↑↑" or "↓↓↓" denote significant changes > 75% (more details in Tables S17–S21). Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software; NUMTs—nuclear insertions of mitochondrial DNA; LCR—low-complexity region.

Ultimately, we performed a performance analysis for four different datasets: Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B, with separate paired statistical tests for each dataset (Figure A1 and Table S22). PCP achieved a higher sensitivity and F_1 score in both the primary and mixture analyses for Grade B variants, without performing worse than TSS in the other indicators or in the other datasets (Table S22).

In order to further validate our findings, we performed the same analyses with two other variant callers (freebayes [42] and varscan 2 [43]), which yielded overall overlapping results (Tables S6–S24).

2.2. Pipeline Comparison with An Independent Set of 50 Clinical Samples

As a way of demonstrating the added value of our pipeline, we selected 50 independent samples from a prospective study in patients with ischemic stroke (more details in Section 4.1. Sample acquisition/collection) and analyzed them in parallel with TSS and PCP.

In PCP, genome coverage followed a similar pattern in all samples (Figure 5A), with an overall uniform mappability (Table S25). The genome coverage was also in accordance with what is typical in Ion Torrent™ sequencing with the Precision ID mtDNA Whole Genome Panel [35]. The haplogroup was identical for both pipelines in 45/50 samples (Table S25). In PCP, 3 samples were flagged for contamination in contrast to TSS, which had 11 samples, 10/11 were well below its ~10% threshold (Table S25). In PCP, the number of variants varied extremely, from 12 to 503, while in TSS, they ranged from 10 to 80 (Table S25). Thus, we eliminated Samples #21, #37, and #44 because of their contamination status, and Samples #2 and #3 because of their extremely high number of variants in PCP, from our subsequent analysis.

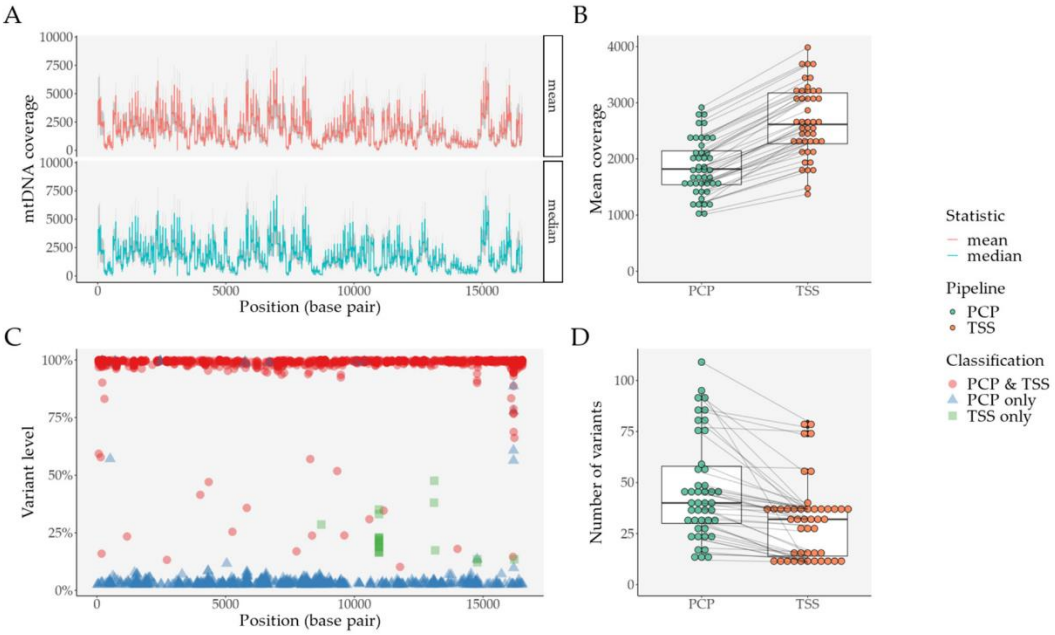


Figure 5. Comparison between PCP and TSS in an independent set of samples (after filtering, N = 45) derived from clinical practice (an ischemic stroke prospective cohort). (**A**) Coverage per base pair with PCP data, per statistic (mean and median)—the grey area represents the standard deviation of each statistic per base pair; (**B**) Mean coverage per sample and per pipeline—lines connect the same sample; (**C**) Distribution of variants per base pair and per variant classification—for variants found in PCP and TSS, mean variant level was plotted; (**D**) Mean number of variants per sample and per pipeline—lines connect the same sample. Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software.

Consistent with what we described previously, while mean coverage was significantly decreased in PCP (Figure 5B)—mean difference of 30.87%, 95% CI [29.17%–32.58%], an adjusted *p*-value of 3.01×10^{-34}, we still observed a higher number of variants with our pipeline (Figure 5D)—mean difference of 15.11%, 95% [10.07–20.16], an adjusted *p*-value of 2.97×10^{-7}; paired t-tests adjusted with FDR.

This difference mostly derives from an increase in variants below TSS's heteroplasmy threshold (Figure 5C): variants only present in PCP have a mean VL of 7.45%, while variants only present in TSS have a mean VL of 22.5%. Interestingly, PCP finds 98.31% of TSS's variants, while TSS only finds 57.92% of PCP's variants. Although not exclusively, when we consider discordant variants (variants not present in both pipelines) with more than one occurrence (Table S26), we observe similar patterns as false positives in the previous analyses (Figure S4C–H and Tables S17–21). In PCP, for example, we encounter the same variants seen in the previous analysis: 8649C (present in 25/45 samples) and 4318T (present in 20/45 samples). These two samples have a mean VL <5%, a low mean normalized coverage, and a small distance to the amplicon's "callable" edge. Similarly, in TSS, we observe variants that were also previously found: 10958C (present in 17/45 samples) and 14777C (present in 2/45 samples). These two variants have a low mean normalized coverage, without other discernible features.

Regardless of the discordant variants, in the variants found in both pipelines, there is a statistically significant correlation between the VL with TSS, and with PCP (Figure S5)—adjusted R^2 of 0.97 and a *p*-value < 2.2×10^{-16}; linear regression model.

3. Discussion and Conclusions

In this study, we have developed a fully customizable pipeline—PCP—for the Precision ID mtDNA Whole Genome Panel, which produced BAMs accurately mapped to the rCRS and that has revisited the heteroplasmy threshold of the current gold standard method—TSS.

On the one hand, despite arising from the same set of samples and sequencing runs, the output from PCP and TSS was sufficiently different to yield significant changes between the two pipelines. In comparison with TSS, our method accomplished better performance metrics in previously sequenced samples, namely, higher sensitivity and F_1 scores without a decrease in specificity and precision. This was mostly due to the additional detection of heteroplasmic variants with a VL above 2.5%, but below TSS's ~10% threshold. Interestingly, this increase in the number of correct variants and the improved performance was achieved despite PCP having lower coverage and fewer sequences. Moreover, the same pattern of a higher number of variants at a lower threshold, notwithstanding a lower mean coverage and fewer sequences, was also observed in an independent set of clinical samples.

On the other hand, mtDNA differences between the same sample are not unheard of, as DNA polymerases, amplification protocols, sequencing runs, and variant callers are frequent sources of disparity in genomics [35,44]. Discerning sequencing errors/false positives from true heteroplasmic variants is a challenging task, usually achieved through post-sequencing curation, looking for signs of poor amplification, strand bias, mutations in LCR, and the presence of NUMTs [41,45–49].

Although sometimes ignored [50], heteroplasmic variants are ubiquitous [51–54] and show significant tissue-specificity [53,54]. Interestingly, low-level heteroplasmic variants (VL <10%) have shown matrilineal inheritance [55,56] with increased relevance in aging [57], and in clinical settings, particularly cancer [58,59]. Nonetheless, the reliability of these low-level variants has been a matter of debate [60–62] and, thus, current guidelines/workflows suggest a heteroplasmy threshold of 10%, based on the limitations of Sanger sequencing and electrophoresis technology [35,37,63–66].

False positives in both PCP and TSS showed a decreased normalized coverage, which indicates poor amplification in those regions, and were mostly transitions, which is in accordance with the literature [66,67]. However, the remaining putative mechanisms of error were very different between the two pipelines, and even showed differences within

the same pipeline, depending on where the novel variants were found (unmixed vs. mixed samples). Correspondingly, in the analysis of the 1000 Genomes Project's samples, no false positives were shared.

In parallel, discordant variants in the stroke cohort showed most of the variants previously flagged as false positives, albeit with many more variants only present in PCP at a low level, which had not been flagged before. This different mutational pattern from the samples sequenced within the 1000 Genomes Project might arise from the tissue-specificity of heteroplasmic variants, since blood, from where we extracted the DNA in our stroke cohort, has a lower level of heteroplasmy [53,54] and a very diverse mixture of mtDNA content [68], as it encompasses multiple cell subtypes.

Our pipeline offers multiple advantages in comparison with TSS. Firstly, it is in line with the Findable, Accessible, Interoperable, Reusable (FAIR) principles [69], increasingly important in the field of genomics [70]. Thus, our open-access approach makes it easier to implement validated and uniform bioinformatic protocols in genomics [3,71,72]. Secondly, it outputs a BAM file with the correct rCRS reference, thus allowing for integration into other NGS workflows and greatly facilitating variant annotation. This is particularly important when dealing with high-throughput datasets, where manual curation and inspection are not feasible nor efficient, thus favoring computational approaches to DNA variant analysis [73]. Finally, it is fully customizable. In a world of limited resources (personnel, time, samples, funding), our workflow is easily adaptable, depending on the focus of each research project and tissue mutational pattern. On the one hand, if one wishes to prioritize sample diversity in favor of replicates, variants might be called with PCP and filtered if not present in TSS, since PCP captures ~98% of all TSS variants with a very high VL correlation, and TSS discordant variants are likely to be false positives, in accordance with our analysis. On the other hand, if one prefers to sacrifice sample diversity in favor of VCM performance, the protocol for NUMT removal and heteroplasmy threshold optimization in unmixed and mixed samples yields an optimal workflow specific to that set of samples. Furthermore, the two approaches might be combined, or even expanded, into FASTQ quality control, alignment, or further variant annotation, with the addition of hypervariable segments [74], poly C-stretches [75], which are particularly difficult in Ion Torrrent™ sequencing [76], and the presence of homoplasmic or heteroplasmic variants at the HelixMTdb [67] (see Table S26 for an example). In summary, future clinical research will benefit from using our open-source bioinformatic processing since it keeps the advantages of the Precision ID library kit, particularly its low DNA input, while circumventing the limitations of TSS, namely, its modified reference sequence, proprietary nature, and 10% heteroplasmy threshold. For separating true variants from false variants, we also present a range of options that can be used in combination or not, depending on the samples and research focus: (**I**) A dedicated mixture optimization protocol; (**II**) Variant filtering based on normalized coverage, strand bias, the presence of NUMTs, among other parameters; (**III**) Filtering the output from PCP with TSS.

Nonetheless, our study also has a few limitations. Firstly, we used a limited set of samples for the primary and mixture analysis, which are not representative of the range of possible haplogroup combinations. Secondly, samples from our stroke cohort were not sequenced in duplicate nor mixed, which did not allow for a replication of the performance metrics we did previously. Finally, we did not consider insertions nor deletions and, thus, we are unable to provide any recommendations for the analysis of those variants with our pipeline.

Overall, we have developed the first open-source alternative to the enterprise software, Ion Torrent Suite™ Software (TSS), for the Precision ID mtDNA Whole Genome Panel, with improved performance metrics, and with an output in the correct rCRS reference. Since the majority of existing mtDNA bioinformatic tools [6–21] are only compatible with the correct rCRS format, and the few tools that perform variant calling in samples sequenced with the Precision ID library kit [31,34–36] keep TSS's modified reference sequence, PCP

is currently the sole option that bridges this bioinformatic gap, allowing for the universal variant calling of samples sequenced with the aforementioned library.

The herein presented pipeline, PCP, is available in the GitHub repository, https://github.com/filcfig/PCP (accessed on 1 November 2021). Additionally, we provide the generated data for validation (see Data Availability Statement).

4. Materials and Methods

4.1. Sample Acquisition/Collection

In order to validate our pipeline—PrecisionCallerPipeline *(PCP)*, we acquired three cell-line samples from the Coriell Institute for Medical Research (Camden, NJ, USA), i.e., HG00256, HG01626, and HG01757. These samples have been sequenced within the 1000 Genomes Project [4] on Illumina HiSeq (Illumina, Inc., San Diego, CA, USA), with three different sequencing runs, mostly targeted at nuclear DNA: (**I**) Exome sequencing—exome; (**II**) Low-coverage sequencing—lowCov; and (**III**) High coverage sequencing—highCov.

With the aim of evaluating the sensitivity, specificity, precision, and F_1 score (harmonic mean of precision and sensitivity) [17,44] with different variant detection thresholds, we performed three different mixtures at five different mixture levels (1, 2, 5, 10, and 25%):

- Mixture HT—Minor component: HG00256 (haplogroup H5b2, short identifier H) + Major component: HG01626 (haplogroup T2a1b1a1b, short identifier T);
- Mixture TU—Minor component: HG01626 (haplogroup T2a1b1a1b, short identifier T) + Major component: HG01757 (haplogroup U4a, short identifier U);
- Mixture UH—Minor component: HG01757 (haplogroup U4a, short identifier U) + Major component: HG00256 (haplogroup H5b2, short identifier H).

DNA concentration measured prior to shipping ranged from 307 to 331 ng/µL. Thus, DNA mixtures were volume-based (Figure 1A).

As an independent set of samples, we analyzed 50 samples derived from a prospective stroke cohort of patients at the Hospital de Santa Maria, Centro Hospitalar Universitário Lisboa Norte. Blood samples were collected within 72 h of hospital admission and all cases were reviewed and confirmed by trained neurologists. Inclusion criteria were: (**I**) Ischemic stroke; (**II**) Age $\geq$ 18 years old; (**III**) Blood samples collected up to 72h after symptom onset. The exclusion criteria were: (**I**) Active cancer diagnosis; (**II**) Previous cerebral revascularization surgeries; (**III**) Modified Rankin score [77] $\geq$ 5. The approval of the institutional review board (IRB) was conceded by the Comissão de Ética do Centro Académico de Medicina de Lisboa (reference 435/16, approved on 14 December 2016), informed consents were given by every subject, and the study followed the standards of the Declaration of Helsinki. DNA extraction was performed after PBMC isolation with the QIAamp® DNA Blood Midi Kit (QIAGEN GmbH, Hilden, Germany), according to the manufacturer's instructions.

4.2. DNA Sequencing

Samples were sequenced without prior long-range PCR (LR-PCR) with the Applied Biosystems™ Precision ID mtDNA Whole Genome Panel (Thermo Fisher Scientific, Waltham, MA, USA), in conjunction with the Ion Torrent™ Ion S5™ (Thermo Fisher Scientific, Waltham, MA, USA), in accordance with the manufacturer's instructions. Briefly, DNA was quantified with a Qubit® 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA) and samples were diluted to 0.0067 ng/µL for an input of 0.1 ng of genomic DNA in 15 µL. Libraries were prepared using the Ion Chef™ automated protocol, and samples were then run on 530™ chips with the Ion Torrent™ Ion S5™ at Ipatimup—Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Porto, Portugal).

4.3. Bioinformatic Processing

4.3.1. PrecisionCallerPipeline (PCP)

Our PCP pipeline automatically takes the FASTQ files from the sequencing facility and outputs fully aligned BAM files mapped to the commonly used reference sequence, rCRS [22]. We use a workflow based on Snakemake [78] that uses: (**I**) Awk, for SAM file editing [79]; (**II**) BEDTools, for BAM to FASTQ conversion [80]; (**III**) BWA-MEM, for read alignment [81]; (**IV**) Pycision, for amplicon delimitation and selection [34]; (**V**) SAMtools for BAM conversion, sorting, indexing, and merging [82]; and (**VI**) Trimmomatic for read quality control and trimming [83] (Figure 1B). Removal of NUMTs was tested before and after final BAM merging with RtN! [47] (Figure 1C).

Samples were processed in a Linux-based system. For simplicity, a predetermined file structure and the necessary files (except for the files from external software) may be downloaded from https://github.com/filcfig/PCP.git (accessed on 1 November 2021). A separate Snakefile (Snakefile_no RtN) is also provided to run the samples without the removal of NUMTs.

Read quality analysis was performed with FastQC [84] and MultiQC [85]. For read quality control, we opted to crop reads at 160 base pairs (bps), taking into account: (**I**) The visual inspection of Phred score patterns per read bp (Figure 3D); (**II**) The length range of 73–137 bps for the known "callable" sections of each amplicon [34], also considering the company-reported amplicon average length of 163 bps, as the exact coordinates of the amplicon themselves are unknown.

Variant calling was performed with freebayes v1.3.5 [42], mutserve v2, the command line interface and successor of the mtDNA-Server pipeline [17,40], and VarScan 2 v2.3.7 [43]. Initially, we used 21 different heteroplasmy thresholds, ranging from 0.4% to 10.0% (Figure 1C and Figure S1). After optimization, the maximum F_1 score was achieved with a heteroplasmy threshold of 2.5% and with NUMT removal before the final BAM merge (Figure 1B, Tables S2 and S3); this processing was then used in all variant calling methods (VCMs). For the Illumina sequencing runs (exome, lowCov, and highCov), a threshold of 0.4% was maintained as a reference. We used very similar parameters as [44], with the exception of a base quality score of 20 for all VCMs, as well as, for freebayes, where we used *"–ploidy 1 –pooled-continuous"*. Similar to [44], variants in positions 302–315 (position 310 was blacklisted in our analysis), 523–524, and 3104–3110 were excluded. Only single nucleotide substitutions were considered, and variants below the established threshold were filtered.

Haplogroup calling was carried out through HaploGrep v2.4.0 [18], and a contamination check was done with Haplocheck v1.3.3 [40], based on the output from mutserve.

4.3.2. Ion Torrent Suite™ Software (TSS)

For the current gold standard, data from each run was processed using the Ion Torrent™ specific pipeline software, Ion Torrent Suite™ Software (TSS), using the reference sequence PrecisionID_mtDNA_rCRS, and target regions PrecisionID_mtDNA_WG_targets with the plugins CoverageAnalysis and VariantCaller. FASTQ and BAM files were generated using the plugin FileExporter. The software versions ranged from v5.8, v5.10, and v5.12, according to the date of each run.

We compiled the VCF files arising from the sequencing runs and corrected all positions > 16,569 to the first 80 bps in the rCRS. When we observed equal variants with different coverages and variant levels (VLs), particularly in the first 80 bps, we calculated the mean coverage and VL per mutation. Only single nucleotide substitutions were considered.

After exporting the corrected variants in a VCF format, where GT 1/0 was defined for VL $\geq$ 90%, we ran the samples through HaploGrep and Haplocheck, similar to PCP.

4.4. Data Analysis

Data analysis was performed with R version 4.1.1 [86] in RStudio [87] with the packages extrafont [88], infer [89], magick [90], patchwork [91], readxl [92], remotes [93],

scales [94], svglite [95], and tidyverse [96], as well as Excel 2016 (Microsoft Corporation, Redmond, WA, USA).

Analyses were performed in parallel for PCP's output and TSS's output. For the primary analysis, variants from the resequenced unmixed samples were compared to the ones identified in Illumina for the exome, lowCov, and highCov sequencing runs, and classified according to their reliability:

1. Grade A variants: homoplasmic variants (mean variant level $\geq$ 95%) found in both highCov and lowCov, regardless of exome;
2. Grade B variants: heteroplasmic variants (mean variant level $\geq$ 0.4% and $\leq$ 95%) found in both highCov and lowCov, regardless of exome, or found in highCov plus exome, or lowCov plus exome;
3. Grade C variants: found in a single sequencing run;
4. Novel variants: found in the Ion Torrent runs only.

For the mixture analysis, variants from the 15 mixtures were classified according to the interaction between both VL and primary variant classifications (explained previously) in both components (minor and major). We began with a manual curation of all 512 theoretical combinations, which gave rise to 128 possible scenarios, grouped in 47 micro classes, 7 meso classes, and 5 macro classes (Table S1). In this case, due to the possibility of shared variants—variants in the same position in two different samples—having two different classifications, we assumed that lower grade variants would prevail in combinations. Hence, a shared variant with both Grade A and Grade C variant classifications would receive a Grade C classification.

Supplementary Materials: The following are available online at https://www.mdpi.com/article/10.3390/ijms222112031/s1.

Author Contributions: Conceptualization, F.C.-F., H.W. and V.A.M.; formal analysis, F.C.-F., H.W., and V.A.M.; funding acquisition, F.P. and V.A.M.; investigation, F.C.-F., F.S.C. and A.C.F.; methodology, F.C.-F., S.S., H.W. and V.A.M.; resources, A.C.F., J.M.F. and V.A.M.; software, F.C.-F., S.S. and H.W.; supervision, H.W. and V.A.M.; visualization, F.C.-F. and H.W.; writing—original draft preparation, F.C.-F., H.W. and V.A.M.; writing—review and editing, F.C.-F., F.S.C., A.C.F., F.P., J.M.F., S.S., H.W. and V.A.M.. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Fundação para a Ciência e Tecnologia (FCT) (FCT/PTDC/MED-NEU/7976/2020), and a project cofunded by FEDER (POR Lisboa 2020—Programa Operacional Regional de Lisboa PORTUGAL 2020), and FCT (PAC-PRECISE LISBOA-01-0145-FEDER-016394), and the National Multiple Sclerosis Society (NMSS), NMSS Pilot Research Grant (PP-1712-29466). F.C.F.'s stipend was supported by FCT (PD/BD/114122/2015), and by Merck Germany (restricted research grant). V.A.M. is an iFCT researcher (IF/01693/2014; IMM/CT/27-2020).

Institutional Review Board Statement: Sample collection within the 1000 Genomes Project was performed with an informed consent by every subject, which explicitly outlined the creation of cell lines from the patients' samples. The distribution of genetic material was carried out by the Coriell Institute. Regarding the independent set of samples from stroke patients, the study was conducted according to the guidelines of the Declaration of Helsinki, and it was approved by the Ethics Committee of the Centro Académico de Medicina de Lisboa (reference 435/16, approved on December 14, 2016).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study, namely, the data generated with the samples previously sequenced within the 1000 Genomes Project, are openly available in Zenodo [97] at doi:10.5281/zenodo.5524539. Data analyzed within the stroke cohort are available upon request from the corresponding author; the data are not publicly available because of ethical and privacy restrictions.

Acknowledgments: We would like to thank the staff at Biobanco-iMM CAML for training and sample handling, namely, Ângela Afonso, nurse Gislaine Castilho de Oliveira for sample collection

and transportation; and the staff at Ipatimup—Instituto de Patologia e Imunologia Molecular da Universidade do Porto (Porto, Portugal) for DNA sequencing, namely, Ana Mafalda Rocha. Finally, we would like to thank all the subjects who have altruistically donated their samples for this research project.

Conflicts of Interest: F.P. is a member of the Novartis OCTIMS study steering committee and the MedImmune/Viela Bio steering committee; has received speaker honoraria/travel grants from Bayer, Novartis, Biogen Idec, Teva, Sanofi-Aventis/Genzyme, Merck Serono, Alexion, Chugai, MedImmune, Shire, Roche, Actelion, and Celgene and; is an Academic Editor of PLoS ONE, and an Associate Editor of *Neurology: Neuroimmunology & Neuroinflammation*; has provided consultancies for SanofiGenzyme, BiogenIdec, MedImmune, Shire, and Alexion; and has received research support from Bayer, Novartis, Biogen Idec, Teva, Sanofi-Aventis/Genzyme, Alexion, Merck Serono, the German Research Council (DFG Exc 257), Werth Stiftung of the City of Cologne, the German Ministry of Education and Research (BMBF Competence Network Multiple Sclerosis), Arthur Arnstein Stiftung Berlin, the EU FP7 Framework Program (combims.eu), the Arthur Arnstein Foundation Berlin, the Guthy Jackson Charitable Foundation, and the National Multiple Sclerosis Society of the USA. The remaining authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Table A1. Primary analysis: Distribution of variants with the mutserve variant calling method (VCM).

Variant Description		PCP				TSS			
Broad Classification	**Detailed Classification**	***N***	**Mean Observed VL**	**Mean Primary VL**	**Mean Absolute Difference in VLs**	***N***	**Mean Observed VL**	**Mean Primary VL**	**Mean Absolute Difference in VLs**
			[Min–Max]	**[Min–Max]**	**[Min–Max]**		**[Min–Max]**	**[Min–Max]**	**[Min–Max]**
Grade A variant found	including exome	*79*	99.73% [98.50%–100.00%]	99.61% [97.03%–100.00%]	0.44% [0.00%–2.73%]	*77*	99.43% [98.20%–100.00%]	99.60% [97.03%–100.00%]	0.65% [0.00%–2.77%]
Grade B variant found	including exome	*5*	26.06% [3.50%–92.50%]	28.59% [2.57%–94.77%]	2.91% [0.93%–7.47%]	*2*	55.65% [19.60%–91.70%]	58.57% [22.37%–94.77%]	2.92% [2.77%–3.07%]
	excluding exome	*8*	7.07% [2.50%–28.40%]	6.36% [1.10%–28.35%]	0.79% [0.05%–2.10%]	*1*	28.30%	28.35%	0.05%
Grade C variant found	from highCov	-	-	-	-	*1*	21.70%	0.40%	21.30%
Grade A variant lost	including exome	-	-	-	-	*2*	-	100.00% [100.00%–100.00%]	-
Grade B variant lost	including exome	*4*	-	0.88% [0.50%–1.27%]	-	*7*	-	4.19% [0.50%–17.37%]	-
	excluding exome	*34*	-	1.10% [0.40%–5.75%]	-	*41*	-	1.46% [0.40%–6.20%]	-
	from lowCov and exome	*2*	-	0.48% [0.45%–0.50%]	-	*2*	-	0.48% [0.45%–0.50%]	-
Grade C variant lost	from highCov	*177*	-	0.82% [0.40%–5.80%]	-	*176*	-	0.83% [0.40%–5.80%]	-
	from lowCov	*12*	-	0.93% [0.40%–3.80%]	-	*12*	-	0.93% [0.40%–3.80%]	-
	from exome	*4*	-	0.70% [0.40%–1.00%]	-	*4*	-	0.70% [0.40%–1.00%]	-
Novel variant	Only present in Ion Torrent	*1*	2.70%	-	-	*2*	25.15% [19.30%–31.00%]	-	-

Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software; N—number; VL—variant level; Min—minimum; Max—maximum.

Table A2. Mixture analysis: Distribution of variants with the mutserve VCM.

Variant Description				PCP				TSS		
					Within Platform	Other Platforms			Within Platform	Other Platforms
Macro	**Meso**	**Micro**	***N***	**Mean Observed VL**	**Mean Absolute Difference in VLs**	**Mean Absolute Difference in VLs**	***N***	**Mean Observed VL**	**Mean Absolute Difference in VLs**	**Mean Absolute Difference in VLs**
				[Min–Max]	**[Min–Max]**	**[Min–Max]**		**[Min–Max]**	**[Min–Max]**	**[Min–Max]**
Found variants	Found variants	Major Grade A	*270*	92.51% [51.50%–100.00%]	3.51% [0.00%–22.90%]	3.73% [0.00%–23.42%]	*260*	92.10% [56.80%–100.00%]	3.40% [0.10%–18.20%]	3.55% [0.00%–18.15%]
		Major Grade B	*39*	20.82% [2.60%–94.30%]	1.73% [0.16%–6.32%]	2.09% [0.29%–6.10%]	*15*	43.18% [11.80%–94.20%]	2.26% [0.14%–6.29%]	2.25% [0.08%–4.98%]
		Major Grade C	-	-	-	-	*5*	28.48% [22.30%–34.10%]	8.65% [1.03%–13.49%]	28.11% [21.91%–33.72%]
		Minor Grade A	*133*	16.95% [2.50%–52.90%]	4.38% [0.02%–28.00%]	4.41% [0.06%–28.20%]	*73*	25.31% [11.00%–48.30%]	5.73% [0.00%–23.50%]	5.75% [0.02%–23.60%]
		Minor Grade B	*8*	7.54% [3.30%–26.00%]	1.90% [0.32%–3.95%]	1.72% [0.14%–4.18%]	*1*	26.30%	3.38%	2.61%
	Mixture found variants	Shared Grade A	*125*	99.89% [99.20%–100.00%]	0.09% [0.00%–0.40%]	0.20% [0.00%–2.41%]	*125*	99.64% [98.60%–100.00%]	0.20% [0.00%–1.30%]	0.45% [0.00%–2.71%]
		Major Grade B	*9*	4.08% [2.60%–5.70%]	-	0.44% [0.03%–1.69%]	*4*	12.60% [11.90%–13.20%]	-	3.12% [0.83%–3.99%]
Lost variants	Mixture lost variants	Major Grade B	*26*	-	2.77% [1.88%–3.56%]	2.00% [0.83%–2.72%]	-	-	-	-
		Minor Grade A	*137*	-	2.44% [0.98%–24.80%]	2.44% [0.97%–24.98%]	*187*	-	3.86% [0.98%–10.00%]	3.87% [0.97%–10.00%]
		Minor Grade B	*57*	-	0.48% [0.03%–4.63%]	0.48% [0.01%–4.74%]	*14*	-	2.65% [0.20%–9.17%]	2.78% [0.22%–9.48%]
	Old lost variants	Major Grade A	-	-	-	-	*10*	-	-	91.40% [75.00%–99.00%]
		Major Grade B	*186*	-	-	0.81% [0.30%–2.70%]	*241*	-	-	1.44% [0.30%–16.50%]
		Major Grade C	*430*	-	-	0.72% [0.30%–5.74%]	*425*	-	-	0.73% [0.30%–5.74%]
		Minor Grade A	-	-	-	-	*10*	-	-	8.60% [1.00%–25.00%]

Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software; N—number; VL—variant level; Min—minimum; Max—maximum.

Table A2. *Cont.*

Variant Description				PCP				TSS		
					Within Platform	Other Platforms			Within Platform	Other Platforms
Macro	**Meso**	**Micro**	***N***	**Mean Observed VL**	**Mean Absolute Difference in VLs**	**Mean Absolute Difference in VLs**	***N***	**Mean Observed VL**	**Mean Absolute Difference in VLs**	**Mean Absolute Difference in VLs**
				[Min–Max]	**[Min–Max]**	**[Min–Max]**		**[Min–Max]**	**[Min–Max]**	**[Min–Max]**
		Minor Grade B	*195*	-	-	0.09% [0.00%–1.44%]	*245*	-	-	0.16% [0.00%–4.34%]
		Minor Grade C	*430*	-	-	0.06% [0.00%–0.95%]	*425*	-	-	0.06% [0.00%–0.95%]
		Shared Grade C	*540*	-	-	0.86% [0.40%–2.69%]	*540*	-	-	0.86% [0.40%–2.69%]
Novel variants	Found variants	Minor novel	-	-	-	-	*5*	21.24% [16.80%–26.00%]	18.07% [11.98%–24.07%]	-
		Shared novel	-	-	-	-	*4*	31.65% [28.00%–35.10%]	12.10% [8.10%–15.78%]	-
	Novel variants	Novel variant: mixture	*12*	27.26% [2.50%–99.20%]	-	-	*9*	13.92% [10.60%–21.10%]	-	-
Primary novel variants	Mixture lost variants	Major novel	*5*	-	2.47% [2.02%–2.67%]	-	*5*	-	28.33% [23.25%–30.69%]	-
		Minor novel	*5*	-	0.23% [0.03%–0.68%]	-	*5*	-	1.15% [0.19%–3.10%]	-
		Shared novel	-	-	-	-	*1*	-	19.35%	-

Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software; N—number; VL—variant level; Min—minimum; Max—maximum.

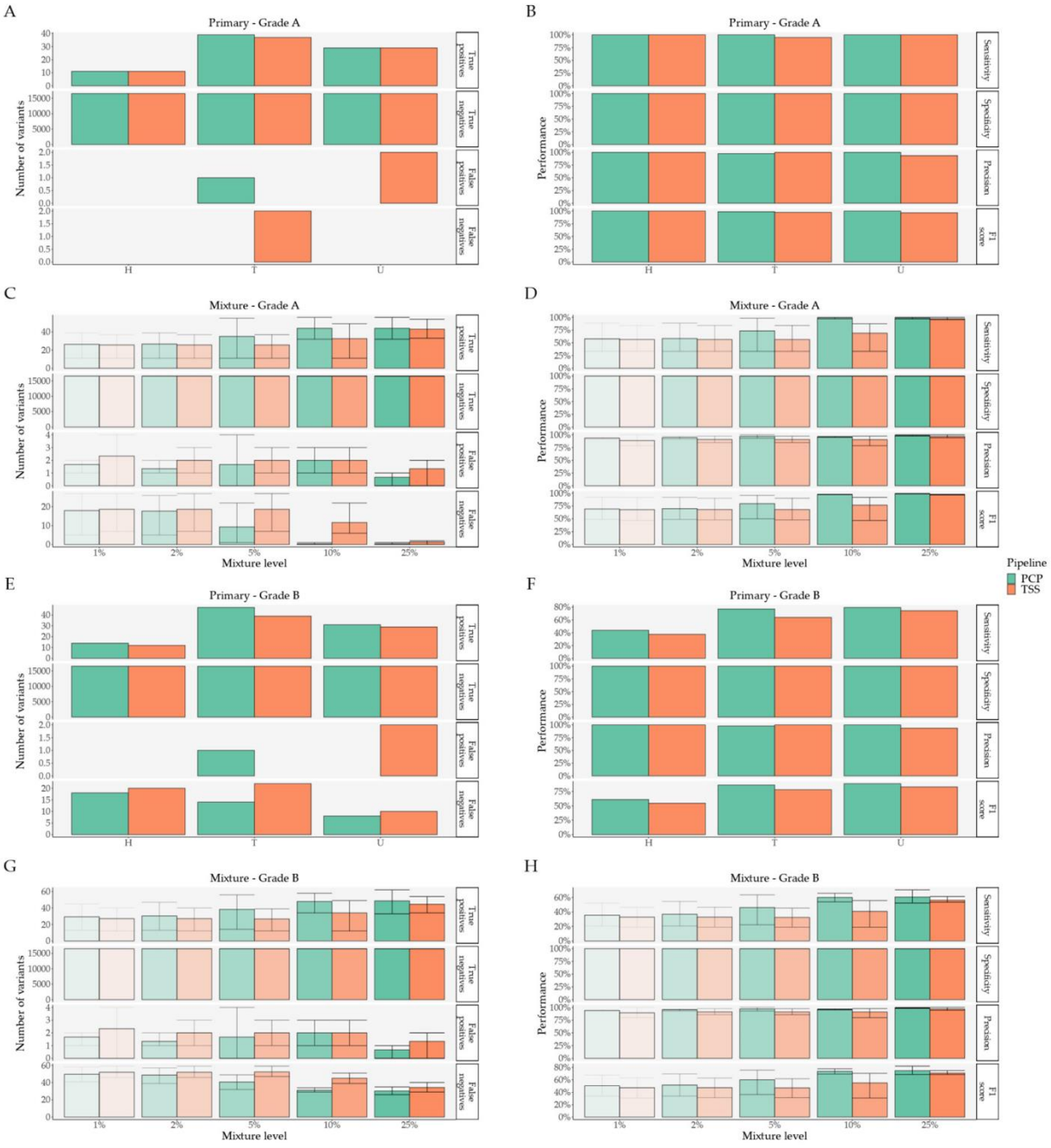


Figure A1. Performance analysis in PCP and TSS. (**A**,**C**,**E**,**G**) Performance metric variant classification per pipeline for Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B datasets, respectively; (**B**,**D**,**F**,**H**) Performance metrics per pipeline for Primary Grade A, Mixture Grade A, Primary Grade B, and Mixture Grade B datasets, respectively. Error bars denote minimum and maximum values. Abbreviations: PCP—PrecisionCallerPipeline; TSS—Ion Torrent Suite™ Software.

References

1. Taylor, R.W.; Turnbull, D.M. Mitochondrial DNA Mutations in Human Disease. *Nat. Rev. Genet.* **2005**, *6*, 389–402. [CrossRef]
2. Tuppen, H.A.L.; Blakely, E.L.; Turnbull, D.M.; Taylor, R.W. Mitochondrial DNA Mutations and Human Disease. *Biochim. Biophys. Acta* **2010**, *1797*, 113–128. [CrossRef]
3. Van der Auwera, G.A.; Carneiro, M.O.; Hartl, C.; Poplin, R.; Del Angel, G.; Levy-Moonshine, A.; Jordan, T.; Shakir, K.; Roazen, D.; Thibault, J.; et al. From FastQ Data to High Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Curr. Protoc. Bioinform.* **2013**, *43*, 11.10.1–11.10.33. [CrossRef]
4. 1000 Genomes Project Consortium; Auton, A.; Brooks, L.D.; Durbin, R.M.; Garrison, E.P.; Kang, H.M.; Korbel, J.O.; Marchini, J.L.; McCarthy, S.; McVean, G.A.; et al. A Global Reference for Human Genetic Variation. *Nature* **2015**, *526*, 68–74. [CrossRef]
5. Gatk4-Mitochondria-Pipeline. Available online: https://github.com/gatk-workflows/gatk4-mitochondria-pipeline (accessed on 18 August 2021).
6. Lee, H.Y.; Song, I.; Ha, E.; Cho, S.-B.; Yang, W.I.; Shin, K.-J. MtDNAmanager: A Web-Based Tool for the Management and Quality Analysis of Mitochondrial DNA Control-Region Sequences. *BMC Bioinform.* **2008**, *9*, 483. [CrossRef] [PubMed]
7. Fan, L.; Yao, Y.-G. MitoTool: A Web Server for the Analysis and Retrieval of Human Mitochondrial DNA Sequence Variations. *Mitochondrion* **2011**, *11*, 351–356. [CrossRef]
8. Zhidkov, I.; Nagar, T.; Mishmar, D.; Rubin, E. MitoBamAnnotator: A Web-Based Tool for Detecting and Annotating Heteroplasmy in Human Mitochondrial DNA Sequences. *Mitochondrion* **2011**, *11*, 924–928. [CrossRef]
9. Guo, Y.; Li, J.; Li, C.-I.; Shyr, Y.; Samuels, D.C. MitoSeek: Extracting Mitochondria Information and Performing High-Throughput Mitochondria Sequencing Analysis. *Bioinformatics* **2013**, *29*, 1210–1211. [CrossRef] [PubMed]
10. Vianello, D.; Sevini, F.; Castellani, G.; Lomartire, L.; Capri, M.; Franceschi, C. HAPLOFIND: A New Method for High-Throughput MtDNA Haplogroup Assignment. *Hum. Mutat.* **2013**, *34*, 1189–1194. [CrossRef] [PubMed]
11. Yang, I.S.; Lee, H.Y.; Yang, W.I.; Shin, K.-J. MtDNAprofiler: A Web Application for the Nomenclature and Comparison of Human Mitochondrial DNA Sequences. *J. Forensic Sci.* **2013**, *58*, 972–980. [CrossRef] [PubMed]
12. Lott, M.T.; Leipzig, J.N.; Derbeneva, O.; Xie, H.M.; Chalkia, D.; Sarmady, M.; Procaccio, V.; Wallace, D.C. MtDNA Variation and Analysis Using Mitomap and Mitomaster. *Curr. Protoc. Bioinform.* **2013**, *44*, 1.23.1–1.23.26. [CrossRef]
13. Calabrese, C.; Simone, D.; Diroma, M.A.; Santorsola, M.; Guttà, C.; Gasparre, G.; Picardi, E.; Pesole, G.; Attimonelli, M. MToolBox: A Highly Automated Pipeline for Heteroplasmy Annotation and Prioritization Analysis of Human Mitochondrial Variants in High-Throughput Sequencing. *Bioinformatics* **2014**, *30*, 3115–3117. [CrossRef]
14. Navarro-Gomez, D.; Leipzig, J.; Shen, L.; Lott, M.; Stassen, A.P.M.; Wallace, D.C.; Wiggs, J.L.; Falk, M.J.; van Oven, M.; Gai, X. Phy-Mer: A Novel Alignment-Free and Reference-Independent Mitochondrial Haplogroup Classifier. *Bioinformatics* **2015**, *31*, 1310–1312. [CrossRef] [PubMed]
15. Falk, M.J.; Shen, L.; Gonzalez, M.; Leipzig, J.; Lott, M.T.; Stassen, A.P.M.; Diroma, M.A.; Navarro-Gomez, D.; Yeske, P.; Bai, R.; et al. Mitochondrial Disease Sequence Data Resource (MSeqDR): A Global Grass-Roots Consortium to Facilitate Deposition, Curation, Annotation, and Integrated Analysis of Genomic Data for the Mitochondrial Disease Clinical and Research Communities. *Mol. Genet. Metab.* **2015**, *114*, 388–396. [CrossRef] [PubMed]
16. Vellarikkal, S.K.; Dhiman, H.; Joshi, K.; Hasija, Y.; Sivasubbu, S.; Scaria, V. Mit-o-Matic: A Comprehensive Computational Pipeline for Clinical Evaluation of Mitochondrial Variations from next-Generation Sequencing Datasets. *Hum. Mutat.* **2015**, *36*, 419–424. [CrossRef] [PubMed]
17. Weissensteiner, H.; Forer, L.; Fuchsberger, C.; Schöpf, B.; Kloss-Brandstätter, A.; Specht, G.; Kronenberg, F.; Schönherr, S. MtDNA-Server: Next-Generation Sequencing Data Analysis of Human Mitochondrial DNA in the Cloud. *Nucleic Acids Res.* **2016**, *44*, W64–W69. [CrossRef]
18. Weissensteiner, H.; Pacher, D.; Kloss-Brandstätter, A.; Forer, L.; Specht, G.; Bandelt, H.-J.; Kronenberg, F.; Salas, A.; Schönherr, S. HaploGrep 2: Mitochondrial Haplogroup Classification in the Era of High-Throughput Sequencing. *Nucleic Acids Res.* **2016**, *44*, W58–W63. [CrossRef]
19. Ishiya, K.; Ueda, S. MitoSuite: A Graphical Tool for Human Mitochondrial Genome Profiling in Massive Parallel Sequencing. *PeerJ* **2017**, *5*, e3406. [CrossRef]
20. Rueda, M.; Torkamani, A. SG-ADVISER MtDNA: A Web Server for Mitochondrial DNA Annotation with Data from 200 Samples of a Healthy Aging Cohort. *BMC Bioinform.* **2017**, *18*, 373. [CrossRef]
21. Preste, R.; Vitale, O.; Clima, R.; Gasparre, G.; Attimonelli, M. HmtVar: A New Resource for Human Mitochondrial Variations and Pathogenicity Data. *Nucleic Acids Res.* **2019**, *47*, D1202–D1210. [CrossRef]
22. Andrews, R.M.; Kubacka, I.; Chinnery, P.F.; Lightowlers, R.N.; Turnbull, D.M.; Howell, N. Reanalysis and Revision of the Cambridge Reference Sequence for Human Mitochondrial DNA. *Nat. Genet.* **1999**, *23*, 147. [CrossRef]
23. Van Oven, M. PhyloTree Build 17: Growing the Human Mitochondrial DNA Tree. *Forensic Sci. Int. Genet. Suppl. Ser.* **2015**, *5*, e392–e394. [CrossRef]
24. Dür, A.; Huber, N.; Parson, W. Fine-Tuning Phylogenetic Alignment and Haplogrouping of MtDNA Sequences. *Int. J. Mol. Sci.* **2021**, *22*, 5747. [CrossRef]
25. Chaitanya, L.; Ralf, A.; van Oven, M.; Kupiec, T.; Chang, J.; Lagacé, R.; Kayser, M. Simultaneous Whole Mitochondrial Genome Sequencing with Short Overlapping Amplicons Suitable for Degraded DNA Using the Ion Torrent Personal Genome Machine. *Hum. Mutat.* **2015**, *36*, 1236–1247. [CrossRef] [PubMed]

26. Wai, K.T.; Barash, M.; Gunn, P. Performance of the Early Access AmpliSeq™ Mitochondrial Panel with Degraded DNA Samples Using the Ion Torrent™ Platform. *Electrophoresis* **2018**, *39*, 2776–2784. [CrossRef] [PubMed]
27. Yao, L.; Xu, Z.; Zhao, H.; Tu, Z.; Liu, Z.; Li, W.; Hu, L.; Wan, L. Concordance of Mitochondrial DNA Sequencing Methods on Bloodstains Using Ion PGM™. *Leg. Med.* **2018**, *32*, 27–30. [CrossRef]
28. Strobl, C.; Eduardoff, M.; Bus, M.M.; Allen, M.; Parson, W. Evaluation of the Precision ID Whole MtDNA Genome Panel for Forensic Analyses. *Forensic. Sci. Int. Genet.* **2018**, *35*, 21–25. [CrossRef]
29. Cuenca, D.; Battaglia, J.; Halsing, M.; Sheehan, S. Mitochondrial Sequencing of Missing Persons DNA Casework by Implementing Thermo Fisher's Precision ID MtDNA Whole Genome Assay. *Genes* **2020**, *11*, 1303. [CrossRef] [PubMed]
30. Pereira, V.; Longobardi, A.; Børsting, C. Sequencing of Mitochondrial Genomes Using the Precision ID MtDNA Whole Genome Panel. *Electrophoresis* **2018**, *39*, 2766–2775. [CrossRef]
31. Faccinetto, C.; Sabbatini, D.; Serventi, P.; Rigato, M.; Salvoro, C.; Casamassima, G.; Margiotta, G.; De Fanti, S.; Sarno, S.; Staiti, N.; et al. Internal Validation and Improvement of Mitochondrial Genome Sequencing Using the Precision ID MtDNA Whole Genome Panel. *Int. J. Legal Med.* **2021**, *135*, 2295–2306. [CrossRef]
32. Strobl, C.; Churchill Cihlar, J.; Lagacé, R.; Wootton, S.; Roth, C.; Huber, N.; Schnaller, L.; Zimmermann, B.; Huber, G.; Lay Hong, S.; et al. Evaluation of Mitogenome Sequence Concordance, Heteroplasmy Detection, and Haplogrouping in a Worldwide Lineage Study Using the Precision ID MtDNA Whole Genome Panel. *Forensic Sci. Int. Genet.* **2019**, *42*, 244–251. [CrossRef] [PubMed]
33. Ming, T.; Wang, M.; Zheng, M.; Zhou, Y.; Hou, Y.; Wang, Z. Exploring of Rare Differences in MtGenomes between MZ Twins Using Massively Parallel Sequencing. *Forensic Sci. Int. Genet. Suppl. Ser.* **2019**, *7*, 70–72. [CrossRef]
34. Woerner, A.E.; Ambers, A.; Wendt, F.R.; King, J.L.; Moura-Neto, R.S.; Silva, R.; Budowle, B. Evaluation of the Precision ID MtDNA Whole Genome Panel on Two Massively Parallel Sequencing Systems. *Forensic Sci. Int. Genet.* **2018**, *36*, 213–224. [CrossRef] [PubMed]
35. Cihlar, J.C.; Amory, C.; Lagacé, R.; Roth, C.; Parson, W.; Budowle, B. Developmental Validation of a MPS Workflow with a PCR-Based Short Amplicon Whole Mitochondrial Genome Panel. *Genes* **2020**, *11*, 1345. [CrossRef]
36. Roth, C.; Parson, W.; Strobl, C.; Lagacé, R.; Short, M. MVC: An Integrated Mitochondrial Variant Caller for Forensics. *Aust. J. Forensic Sci.* **2019**, *51*, S52–S55. [CrossRef]
37. Parson, W.; Gusmão, L.; Hares, D.R.; Irwin, J.A.; Mayr, W.R.; Morling, N.; Pokorak, E.; Prinz, M.; Salas, A.; Schneider, P.M.; et al. DNA Commission of the International Society for Forensic Genetics: Revised and Extended Guidelines for Mitochondrial DNA Typing. *Forensic Sci. Int. Genet.* **2014**, *13*, 134–142. [CrossRef]
38. Cho, S.; Kim, M.Y.; Lee, J.H.; Lee, S.D. Assessment of Mitochondrial DNA Heteroplasmy Detected on Commercial Panel Using MPS System with Artificial Mixture Samples. *Int. J. Legal Med.* **2018**, *132*, 1049–1056. [CrossRef]
39. Churchill, J.D.; Stoljarova, M.; King, J.L.; Budowle, B. Massively Parallel Sequencing-Enabled Mixture Analysis of Mitochondrial DNA Samples. *Int. J. Legal Med.* **2018**, *132*, 1263–1272. [CrossRef]
40. Weissensteiner, H.; Forer, L.; Fendt, L.; Kheirkhah, A.; Salas, A.; Kronenberg, F.; Schoenherr, S. Contamination Detection in Sequencing Studies Using the Mitochondrial Phylogeny. *Genome Res.* **2021**, *31*, 309–316. [CrossRef]
41. Smart, U.; Budowle, B.; Ambers, A.; Soares Moura-Neto, R.; Silva, R.; Woerner, A.E. A Novel Phylogenetic Approach for de Novo Discovery of Putative Nuclear Mitochondrial (PNumt) Haplotypes. *Forensic Sci. Int. Genet.* **2019**, *43*, 102146. [CrossRef]
42. Garrison, E.; Marth, G. Haplotype-Based Variant Detection from Short-Read Sequencing. *arXiv* **2012**, arXiv:1207.3907.
43. Koboldt, D.C.; Zhang, Q.; Larson, D.E.; Shen, D.; McLellan, M.D.; Lin, L.; Miller, C.A.; Mardis, E.R.; Ding, L.; Wilson, R.K. VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing. *Genome Res.* **2012**, *22*, 568–576. [CrossRef]
44. Fazzini, F.; Fendt, L.; Schönherr, S.; Forer, L.; Schöpf, B.; Streiter, G.; Losso, J.L.; Kloss-Brandstätter, A.; Kronenberg, F.; Weissensteiner, H. Analyzing Low-Level MtDNA Heteroplasmy—Pitfalls and Challenges from Bench to Benchmarking. *Int. J. Mol. Sci.* **2021**, *22*, 935. [CrossRef]
45. Ring, J.D.; Sturk-Andreaggi, K.; Alyse Peck, M.; Marshall, C. Bioinformatic Removal of NUMT-Associated Variants in Mitotiling next-Generation Sequencing Data from Whole Blood Samples. *Electrophoresis* **2018**, *39*, 2785–2797. [CrossRef]
46. Genomics England Research Consortium; NIHR BioResource; Wei, W.; Pagnamenta, A.T.; Gleadall, N.; Sanchis-Juan, A.; Stephens, J.; Broxholme, J.; Tuna, S.; Odhams, C.A.; et al. Nuclear-Mitochondrial DNA Segments Resemble Paternally Inherited Mitochondrial DNA in Humans. *Nat. Commun.* **2020**, *11*, 1740. [CrossRef] [PubMed]
47. Woerner, A.E.; Cihlar, J.C.; Smart, U.; Budowle, B. Numt Identification and Removal with RtN! *Bioinformatics* **2020**, *36*, 5115–5116. [CrossRef] [PubMed]
48. Cihlar, J.C.; Strobl, C.; Lagacé, R.; Muenzler, M.; Parson, W.; Budowle, B. Distinguishing Mitochondrial DNA and NUMT Sequences Amplified with the Precision ID MtDNA Whole Genome Panel. *Mitochondrion* **2020**, *55*, 122–133. [CrossRef] [PubMed]
49. Marshall, C.; Parson, W. Interpreting NUMTs in Forensic Genetics: Seeing the Forest for the Trees. *Forensic Sci. Int. Genet.* **2021**, *53*, 102497. [CrossRef]
50. Yonova-Doing, E.; Calabrese, C.; Gomez-Duran, A.; Schon, K.; Wei, W.; Karthikeyan, S.; Chinnery, P.F.; Howson, J.M.M. An Atlas of Mitochondrial DNA Genotype-Phenotype Associations in the UK Biobank. *Nat. Genet.* **2021**, *53*, 982–993. [CrossRef]
51. Sosa, M.X.; Sivakumar, I.K.A.; Maragh, S.; Veeramachaneni, V.; Hariharan, R.; Parulekar, M.; Fredrikson, K.M.; Harkins, T.T.; Lin, J.; Feldman, A.B.; et al. Next-Generation Sequencing of Human Mitochondrial Reference Genomes Uncovers High Heteroplasmy Frequency. *PLoS Comput. Biol.* **2012**, *8*, e1002737. [CrossRef]

52. Ye, K.; Lu, J.; Ma, F.; Keinan, A.; Gu, Z. Extensive Pathogenicity of Mitochondrial Heteroplasmy in Healthy Human Individuals. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 10654–10659. [CrossRef]
53. Li, M.; Schröder, R.; Ni, S.; Madea, B.; Stoneking, M. Extensive Tissue-Related and Allele-Related MtDNA Heteroplasmy Suggests Positive Selection for Somatic Mutations. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 2491–2496. [CrossRef]
54. Naue, J.; Hörer, S.; Sänger, T.; Strobl, C.; Hatzer-Grubwieser, P.; Parson, W.; Lutz-Bonengel, S. Evidence for Frequent and Tissue-Specific Sequence Heteroplasmy in Human Mitochondrial DNA. *Mitochondrion* **2015**, *20*, 82–94. [CrossRef]
55. Guo, Y.; Li, C.-I.; Sheng, Q.; Winther, J.F.; Cai, Q.; Boice, J.D.; Shyr, Y. Very Low-Level Heteroplasmy MtDNA Variations Are Inherited in Humans. *J. Genet. Genom.* **2013**, *40*, 607–615. [CrossRef] [PubMed]
56. Zaidi, A.A.; Wilton, P.R.; Su, M.S.-W.; Paul, I.M.; Arbeithuber, B.; Anthony, K.; Nekrutenko, A.; Nielsen, R.; Makova, K.D. Bottleneck and Selection in the Germline and Maternal Age Influence Transmission of Mitochondrial DNA in Human Pedigrees. *Proc. Natl. Acad. Sci. USA* **2019**, *116*, 25172–25178. [CrossRef] [PubMed]
57. Wachsmuth, M.; Hübner, A.; Li, M.; Madea, B.; Stoneking, M. Age-Related and Heteroplasmy-Related Variation in Human MtDNA Copy Number. *PLoS Genet.* **2016**, *12*, e1005939. [CrossRef] [PubMed]
58. Yuan, Y.; Ju, Y.S.; Kim, Y.; Li, J.; Wang, Y.; Yoon, C.J.; Yang, Y.; Martincorena, I.; Creighton, C.J.; Weinstein, J.N.; et al. Comprehensive Molecular Characterization of Mitochondrial Genomes in Human Cancers. *Nat. Genet.* **2020**, *52*, 342–352. [CrossRef] [PubMed]
59. Fendt, L.; Fazzini, F.; Weissensteiner, H.; Bruckmoser, E.; Schönherr, S.; Schäfer, G.; Losso, J.L.; Streiter, G.A.; Lamina, C.; Rasse, M.; et al. Profiling of Mitochondrial DNA Heteroplasmy in a Prospective Oral Squamous Cell Carcinoma Study. *Cancers* **2020**, *12*, 1933. [CrossRef]
60. Just, R.S.; Irwin, J.A.; Parson, W. Questioning the Prevalence and Reliability of Human Mitochondrial DNA Heteroplasmy from Massively Parallel Sequencing Data. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 4546–4547. [CrossRef]
61. Ye, K.; Lu, J.; Ma, F.; Keinan, A.; Gu, Z. Reply to Just et al. Mitochondrial DNA Heteroplasmy Could Be Reliably Detected with Massively Parallel Sequencing Technologies. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, E4548–E4550. [CrossRef] [PubMed]
62. Just, R.S.; Irwin, J.A.; Parson, W. Mitochondrial DNA Heteroplasmy in the Emerging Field of Massively Parallel Sequencing. *Forensic Sci. Int. Genet.* **2015**, *18*, 131–139. [CrossRef]
63. Bris, C.; Goudenege, D.; Desquiret-Dumas, V.; Charif, M.; Colin, E.; Bonneau, D.; Amati-Bonneau, P.; Lenaers, G.; Reynier, P.; Procaccio, V. Bioinformatics Tools and Databases to Assess the Pathogenicity of Mitochondrial DNA Variants in the Field of Next Generation Sequencing. *Front. Genet.* **2018**, *9*, 632. [CrossRef] [PubMed]
64. Brandhagen, M.D.; Just, R.S.; Irwin, J.A. Validation of NGS for Mitochondrial DNA Casework at the FBI Laboratory. *Forensic Sci. Int. Genet.* **2020**, *44*, 102151. [CrossRef] [PubMed]
65. Poole, O.V.; Pizzamiglio, C.; Murphy, D.; Falabella, M.; Macken, W.L.; Bugiardini, E.; Woodward, C.E.; Labrum, R.; Efthymiou, S.; Salpietro, V.; et al. Mitochondrial DNA Analysis from Exome Sequencing Data Improves Diagnostic Yield in Neurological Diseases. *Ann. Neurol.* **2021**, *89*, 1240–1247. [CrossRef]
66. Laricchia, K.M.; Lake, N.J.; Watts, N.A.; Shand, M.; Haessly, A.; Gauthier, L.; Benjamin, D.; Banks, E.; Soto, J.; Garimella, K.; et al. Mitochondrial DNA Variation across 56,434 Individuals in GnomAD. *bioRxiv* **2021**. bioRxiv:2021.07.23.453510. [CrossRef]
67. Bolze, A.; Mendez, F.; White, S.; Tanudjaja, F.; Isaksson, M.; Jiang, R.; Rossi, A.D.; Cirulli, E.T.; Rashkin, M.; Metcalf, W.J.; et al. A Catalog of Homoplasmic and Heteroplasmic Mitochondrial DNA Variants in Humans. *bioRxiv* **2020**. bioRxiv:798264. [CrossRef]
68. Rausser, S.; Trumpff, C.; McGill, M.A.; Junker, A.; Wang, W.; Ho, S.; Mitchell, A.; Karan, K.R.; Monk, C.; Segerstrom, S.C.; et al. Mitochondrial Phenotypes in Purified Human Immune Cell Subtypes and Cell Mixtures. *bioRxiv* **2021**. bioRxiv:2020.10.16.342923. [CrossRef]
69. Wilkinson, M.D.; Dumontier, M.; Aalbersberg, I.J.J.; Appleton, G.; Axton, M.; Baak, A.; Blomberg, N.; Boiten, J.-W.; da Silva Santos, L.B.; Bourne, P.E.; et al. The FAIR Guiding Principles for Scientific Data Management and Stewardship. *Sci. Data* **2016**, *3*, 160018. [CrossRef]
70. Chervova, O.; Conde, L.; Guerra-Assunção, J.A.; Moghul, I.; Webster, A.P.; Berner, A.; Larose Cadieux, E.; Tian, Y.; Voloshin, V.; Jesus, T.F.; et al. The Personal Genome Project-UK, an Open Access Resource of Human Multi-Omics Data. *Sci. Data* **2019**, *6*, 257. [CrossRef]
71. Jennings, L.J.; Arcila, M.E.; Corless, C.; Kamel-Reid, S.; Lubin, I.M.; Pfeifer, J.; Temple-Smolkin, R.L.; Voelkerding, K.V.; Nikiforova, M.N. Guidelines for Validation of Next-Generation Sequencing-Based Oncology Panels: A Joint Consensus Recommendation of the Association for Molecular Pathology and College of American Pathologists. *J. Mol. Diagn.* **2017**, *19*, 341–365. [CrossRef] [PubMed]
72. Marshall, C.R.; Chowdhury, S.; Taft, R.J.; Lebo, M.S.; Buchan, J.G.; Harrison, S.M.; Rowsey, R.; Klee, E.W.; Liu, P.; Worthey, E.A.; et al. Best Practices for the Analytical Validation of Clinical Whole-Genome Sequencing Intended for the Diagnosis of Germline Disease. *NPJ Genom. Med.* **2020**, *5*, 47. [CrossRef]
73. Wong, A.K.; Sealfon, R.S.G.; Theesfeld, C.L.; Troyanskaya, O.G. Decoding Disease: From Genomes to Networks to Phenotypes. *Nat. Rev. Genet.* **2021**. [CrossRef]
74. Amorim, A.; Fernandes, T.; Taveira, N. Mitochondrial DNA in Human Identification: A Review. *PeerJ* **2019**, *7*, e7314. [CrossRef] [PubMed]
75. Zhao, H.; Shen, J.; Medico, L.; Platek, M.; Ambrosone, C.B. Length Heteroplasmies in Human Mitochondrial DNA Control Regions and Breast Cancer Risk. *Int. J. Mol. Epidemiol. Genet.* **2010**, *1*, 184–192. [PubMed]

76. Sturk-Andreaggi, K.; Parson, W.; Allen, M.; Marshall, C. Impact of the Sequencing Method on the Detection and Interpretation of Mitochondrial DNA Length Heteroplasmy. *Forensic Sci. Int. Genet.* **2020**, *44*, 102205. [CrossRef]
77. Bamford, J.M.; Sandercock, P.A.; Warlow, C.P.; Slattery, J. Interobserver Agreement for the Assessment of Handicap in Stroke Patients. *Stroke* **1989**, *20*, 828. [CrossRef] [PubMed]
78. Köster, J.; Rahmann, S. Snakemake—A Scalable Bioinformatics Workflow Engine. *Bioinformatics* **2012**, *28*, 2520–2522. [CrossRef] [PubMed]
79. Aho, A.V.; Kernighan, B.W.; Weinberger, P.J. Awk—A Pattern Scanning and Processing Language. *Softw. Pract. Exper.* **1979**, *9*, 267–279. [CrossRef]
80. Quinlan, A.R.; Hall, I.M. BEDTools: A Flexible Suite of Utilities for Comparing Genomic Features. *Bioinformatics* **2010**, *26*, 841–842. [CrossRef]
81. Li, H. Aligning Sequence Reads, Clone Sequences and Assembly Contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997.
82. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map Format and SAMtools. *Bioinformatics* **2009**, *25*, 2078–2079. [CrossRef] [PubMed]
83. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A Flexible Trimmer for Illumina Sequence Data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef] [PubMed]
84. Andrews, S.; Krueger, F.; Segonds-Pichon, A.; Biggins, L.; Krueger, C.; Wingett, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 18 August 2021).
85. Ewels, P.; Magnusson, M.; Lundin, S.; Käller, M. MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report. *Bioinformatics* **2016**, *32*, 3047–3048. [CrossRef] [PubMed]
86. R Core Team. *R: A Language and Environment for Statistical Computing*; R Foundation for Statistical Computing: Vienna, Austria, 2021.
87. RStudio Team. *RStudio: Integrated Development Environment for R*; RStudio, PBC: Boston, MA, USA, 2021.
88. Chang, W. *Extrafont: Tools for Using Fonts*; 2014.
89. Bray, A.; Ismay, C.; Chasnovski, E.; Baume, B.; Cetinkaya-Rundel, M. *Infer: Tidy Statistical Inference*; 2021.
90. Ooms, J. *Magick: Advanced Graphics and Image-Processing in R*; 2021.
91. Pedersen, T.L. *Patchwork: The Composer of Plots*; 2020.
92. Wickham, H.; Bryan, J. *Readxl: Read Excel Files*; 2019.
93. Hester, J.; Csárdi, G.; Wickham, H.; Chang, W.; Morgan, M.; Tenenbaum, D. *Remotes: R Package Installation from Remote Repositories, Including "GitHub"*; 2021.
94. Wickham, H.; Seidel, D. *Scales: Scale Functions for Visualization*; 2020.
95. Wickham, H.; Henry, L.; Pedersen, T.L.; Luciani, T.J.; Decorde, M.; Lise, V. *Svglite: An "SVG" Graphics Device*; 2021.
96. Wickham, H.; Averick, M.; Bryan, J.; Chang, W.; McGowan, L.; François, R.; Grolemund, G.; Hayes, A.; Henry, L.; Hester, J.; et al. Welcome to the Tidyverse. *JOSS* **2019**, *4*, 1686. [CrossRef]
97. European Organization for Nuclear Research. *OpenAIRE Zenodo: Research. Shared*; 2013. [CrossRef]

Annex B — Flow cytometry analysis

Page intentionally left blank

As explained previously (more details in *3.2.2.2. CD4+ T cell enrichment and storage*), two different approaches to enrich the PBMC samples in $CD4^+$ T cells were tried out: (**I**) a positive selection, by first depleting monocytes with the MojoSort™ Human CD14 Selection Kit (#480025, BioLegend, San Diego, CA, USA), followed by a capture of labelled cells with the MojoSort™ Human CD4 Nanobeads (#480014, BioLegend, San Diego, CA, USA), as suggested by the manufacturer (Miller, 2016); and (**II**) a negative selection with the MojoSort™ Human CD4 T Cell Isolation Kit (#480010, BioLegend, San Diego, CA, USA). **Table S27** shows the result of the trial described above.

Table S27 — $CD4^+$ T cell enrichment trial: Summary

	Manual cell count with Neubauer chamber			Flow cytometry analysis				
Source	Number of cells (millions)	Viable cells (%)	From input (%)	Number of events	Single cells (%)	Viable cells (%)	Viable T cells ($CD3^+$) (%)	Viable $CD4^+$ T cells ($CD3^+CD4^+CD14^-CD19^-$) (%)
Input	6,650,000	85.71%	-	10000	81.46%	87.74%	46.31%	24.26%
Output selection +	453,333	94.12%	15.49%	2012	52.39%	90.23%	42.17%	16.82%
Output selection -	720,000	93.06%	24.61%	20000	88.94%	94.63%	83.77%	62.85%

In addition to the antibodies listed previously (more details in *3.2.2.3. Flow cytometry analysis*), I used:

- anti-CD123, aimed at identifying dendritic cells, with the antibody CD123-FITC (#11123942, clone 6H6, eBioscience™, Thermo Fisher Scientific, Waltham, MA, USA) at 1.6%;
- Fixable Viability Dye eFluor™ 506, aimed at identifying dead cells (#650866, eBioscience™, Thermo Fisher Scientific, Waltham, MA, USA) at 0.1%.

Thus, together with the protocol mentioned previously (more details in *3.2.2.3. Flow cytometry analysis*), cells were stained prior to fixation with the Fixable Viability Dye eFluor™ 506 for 30 min at 4° C, protected from light. After analyzing the data (**Table S27**), according to the temporary gating strategy (**Figure S7**), as well as considering the main contaminants in the enrichment protocols (**Figure S8**), I decided to: (**I**) use the MojoSort™ Human CD4 T Cell Isolation Kit for $CD4^+$ T cell enrichment; and (**II**) eliminate CD123 from the staining protocol, as it was the least relevant contaminant.

To test the optimal strategy regarding the number of cells for each flow cytometry aliquot and the staining and fixation protocol, I tried eight different cell amounts in duplicate (**Figure S9**): (**I**) with staining before fixation; and (**II**) with staining after fixation. After analyzing the data (**Table S28**), according to the

temporary gating strategy (**Figure S10**), I decided to take 375,000 cells for flow cytometry and performing the staining after the fixation (more details in *3.2.2.3. Flow cytometry analysis*).

Table S29 lists the samples analyzed with flow cytometry. As mentioned previously (more details in *3.2.2.3. Flow cytometry analysis*), viability and CD14 were removed from the final gating strategy, made in batches (**Figure S11** and **Figure S12**).

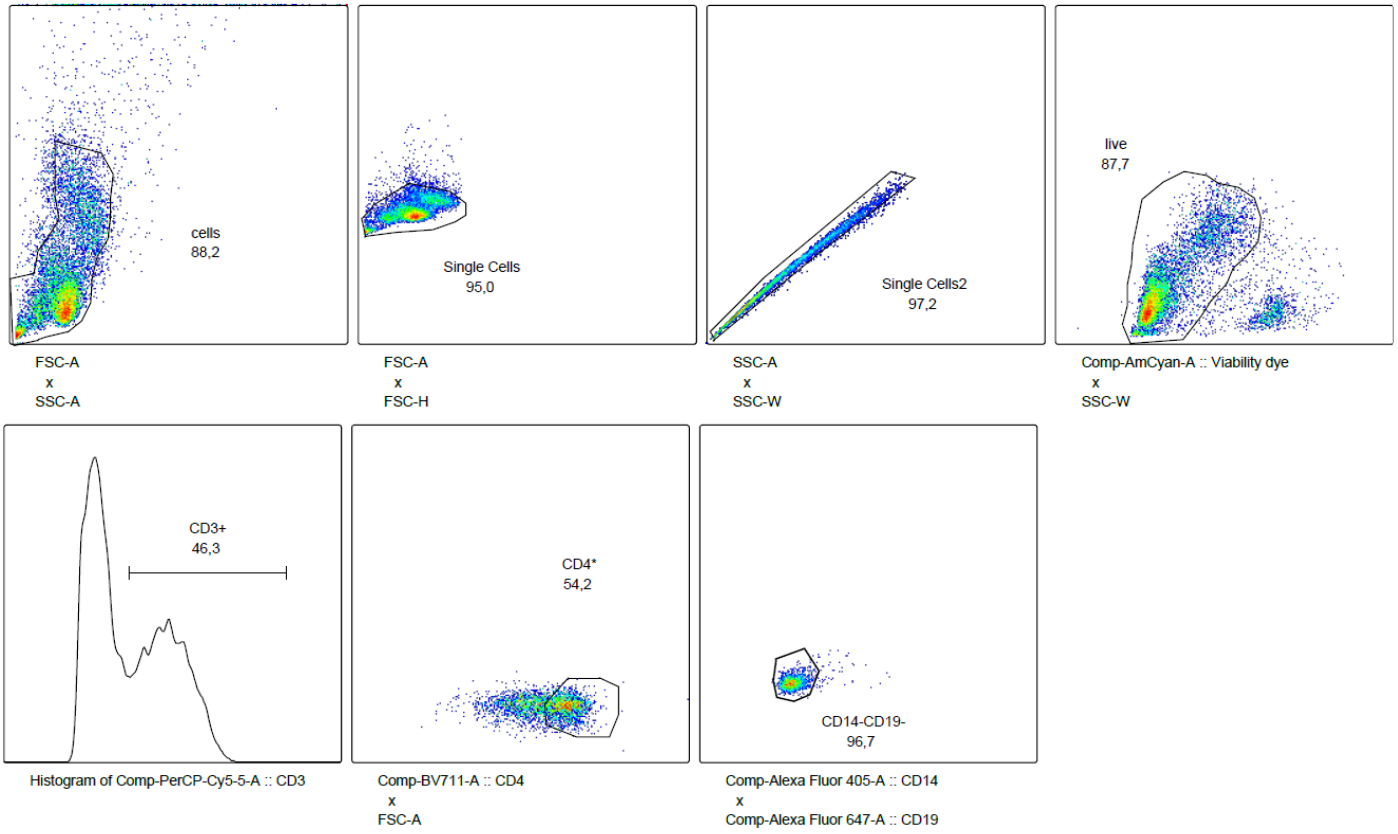


Figure S7 — Temporary gating strategy in the CD4⁺ T cell enrichment trial

The sample represented is the parent sample (input) from where CD4⁺ T cell enrichment occurred. The order is: FSC-A *vs*. SSC-A, FSC-A *vs*. FSC-H, SSC-A *vs*. SSC-W, Fixable Viability Dye eFluor™ 506 *vs.* SSC-W, CD3-PerCP-Cy5.5 (histogram), CD4-BV711 *vs.* FSC-A, and CD14-eFluor450 *vs.* CD19-Alexa Fluor 647. The gaiting is aimed at identifying viable CD4⁺ T cells ($CD3^+CD4^+CD14^-CD19^-$).

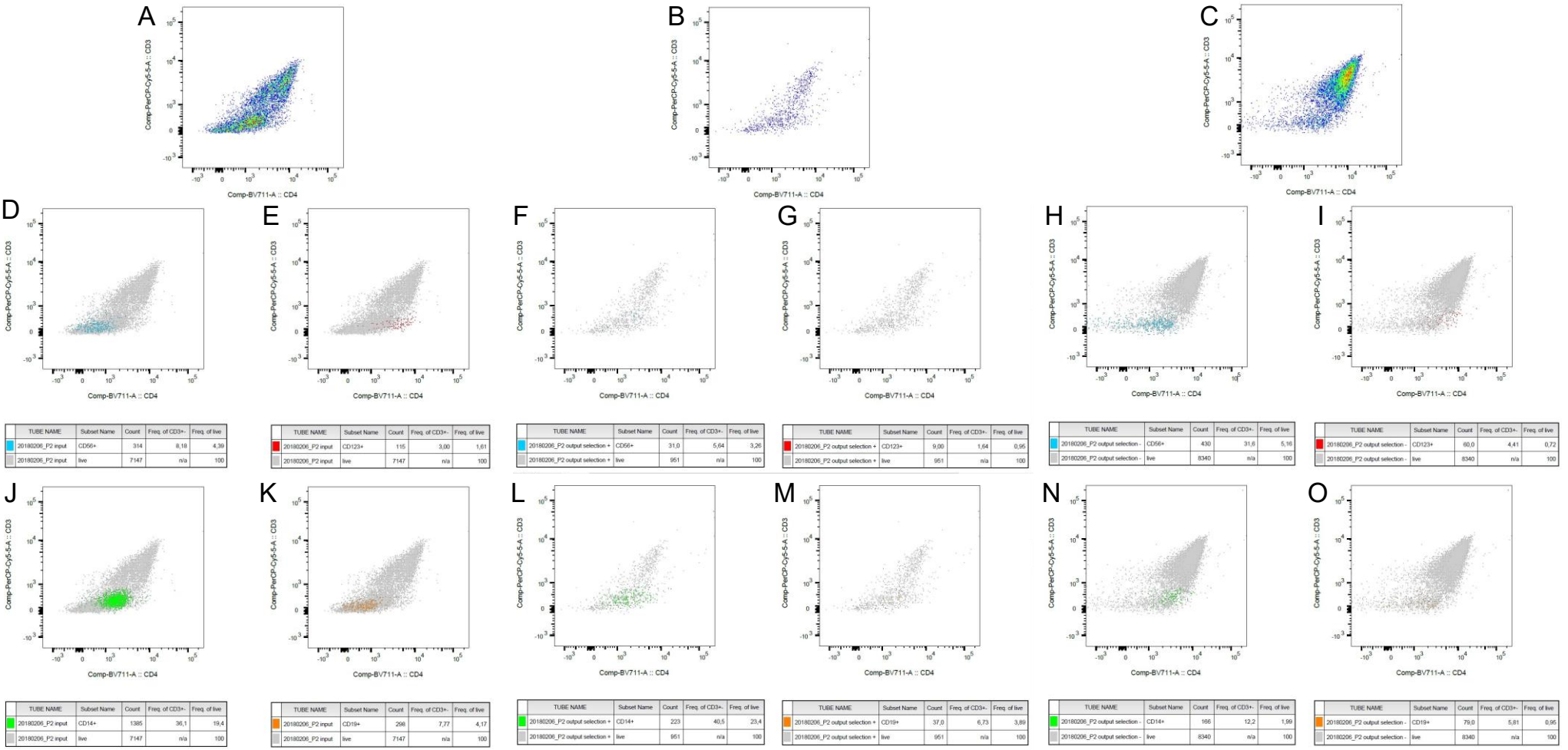


Figure S8 — Analysis of the contaminants in the CD4⁺ T cell enrichment trial

(**A, B, C**) Viable CD3⁺ cells (gating strategy in **Figure S7**) plotted with CD4-BV711 *vs.* CD3-PerCP-Cy5.5 from the parent sample, the positive selection, and the negative selection, respectively; (**D, F, H**) CD3⁺CD56⁺ cells from the parent sample, the positive selection, and the negative selection, respectively; (**E, G, I**) CD3⁺CD123⁺ cells from the parent sample, the positive selection, and the negative selection, respectively; (**J, L, N**) CD3⁺CD14⁺ cells from the parent sample, the positive selection, and the negative selection, respectively; (**K, M, O**) CD3⁺CD19⁺ cells from the parent sample, the positive selection, and the negative selection, respectively.

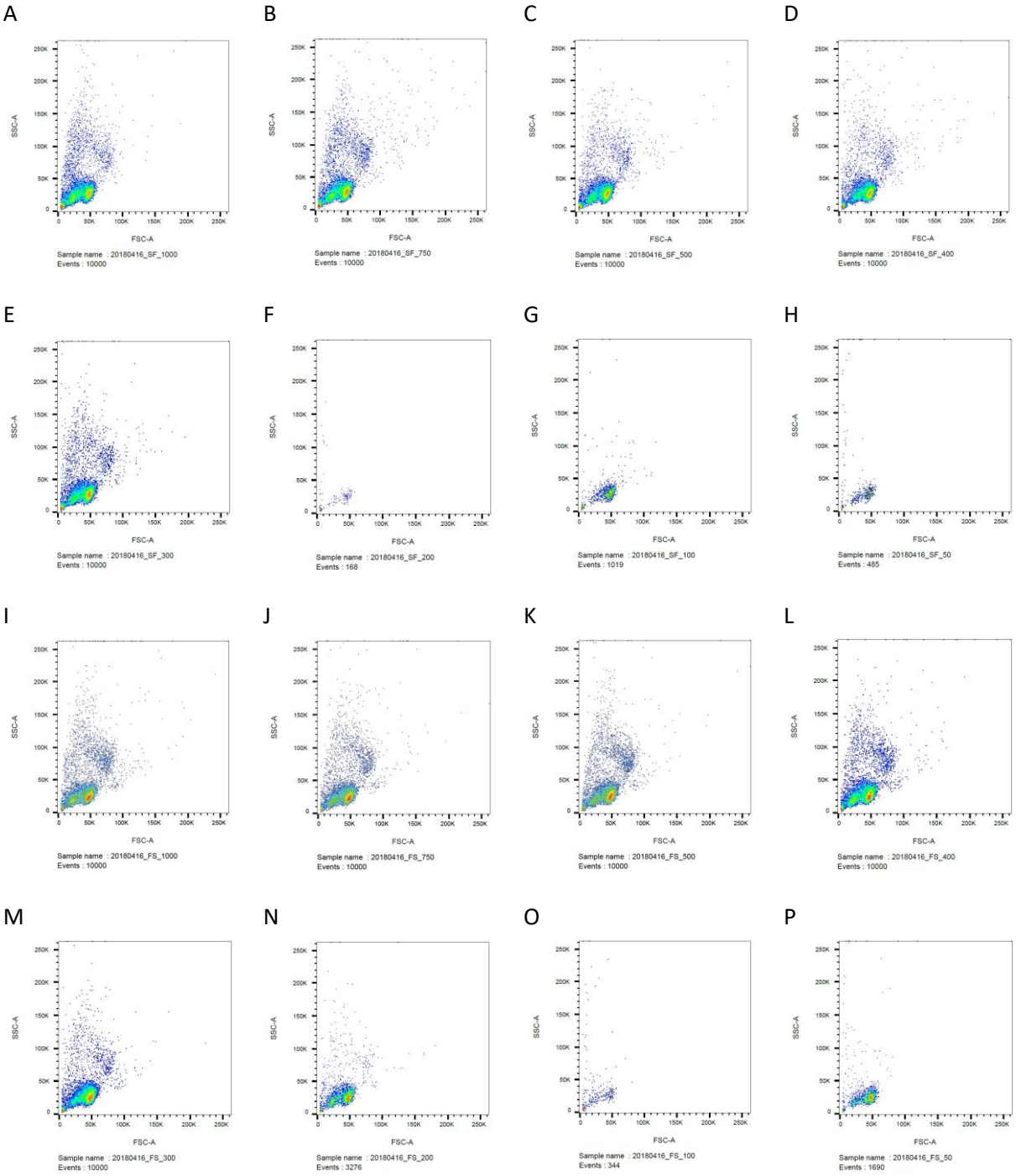


Figure S9 — Staining and fixation protocol optimization, with multiple cell amounts

(**A–H**) Samples analyzed with flow cytometry, with staining before fixation, with 1,000,000 cells, 750,000 cells, 500,000 cells, 400,000 cells, 300,000 cells, 200,000 cells, 100,000 cells, and 50,000 cells, respectively; (**I–P**) Samples analyzed with flow cytometry, with fixation before staining, with 1,000,000 cells, 750,000 cells, 500,000 cells, 400,000 cells, 300,000 cells, 200,000 cells, 100,000 cells, and 50,000 cells, respectively.

Table S28 — Staining and fixation protocol optimization: Summary

Method	Number of cells (x1000)	Number of events	Single cells (%)	$CD4^+$ T cells ($CD3^+CD4^+CD56^-CD19^-$) (%)
Stain-Fix	All	6459	96.42%	27.83%
Stain-Fix	≥300	10000	96.35%	27.10%
Stain-Fix	<300	557.33	96.54%	29.07%
Stain-Fix	50	485	95.05%	35.14%
Stain-Fix	100	1019	98.14%	28.60%
Stain-Fix	200	168	96.43%	23.46%
Stain-Fix	300	10000	97.17%	25.48%
Stain-Fix	400	10000	96.93%	26.21%
Stain-Fix	500	10000	97.05%	29.35%
Stain-Fix	750	10000	94.29%	28.35%
Stain-Fix	1000	10000	96.30%	26.09%
Fix-Stain	All	6914	96.95%	29.25%
Fix-Stain	≥300	10000	96.78%	31.49%
Fix-Stain	<300	1770	97.24%	25.52%
Fix-Stain	50	1690	98.22%	28.98%
Fix-Stain	100	344	95.35%	13.72%
Fix-Stain	200	3276	98.14%	33.87%
Fix-Stain	300	10000	98.27%	28.56%
Fix-Stain	400	10000	96.67%	32.99%
Fix-Stain	500	10000	96.68%	31.32%
Fix-Stain	750	10000	96.80%	33.73%
Fix-Stain	1000	10000	95.46%	30.85%

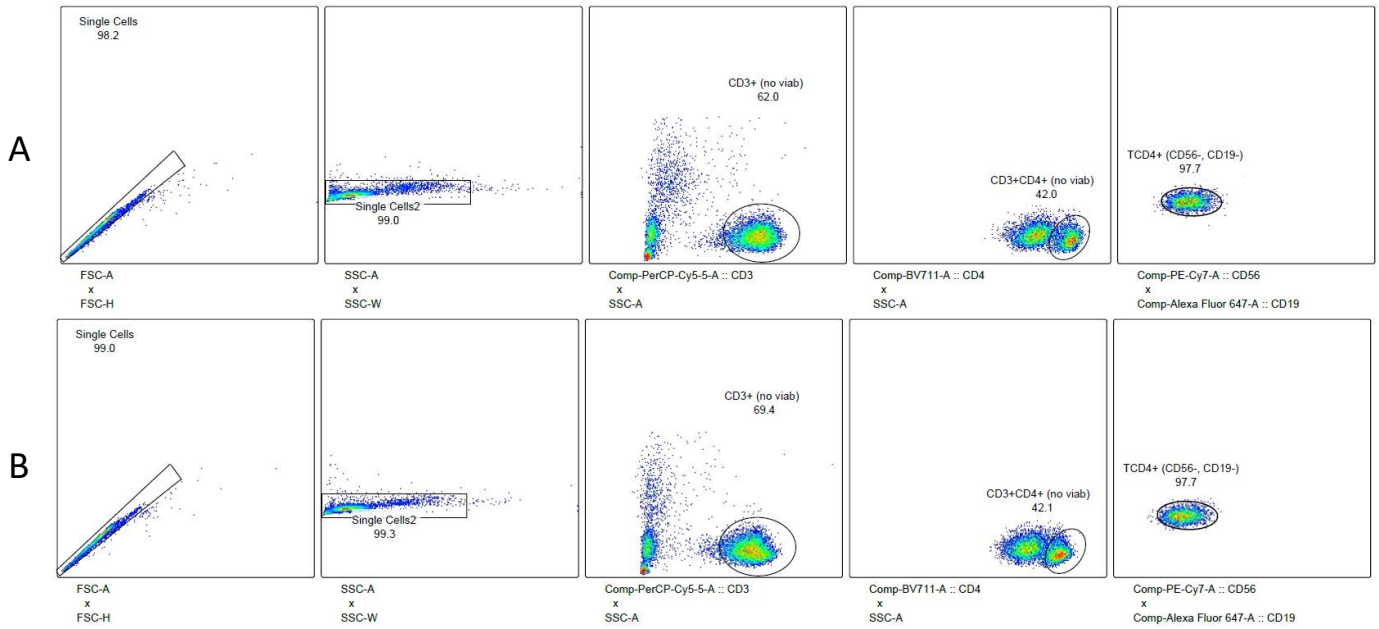


Figure S10 — Staining and fixation protocol optimization: Temporary gating strategy

The order is: FSC-A *vs*. FSC-H, SSC-A *vs*. SSC-W, CD3-PerCP-Cy5.5 *vs.* SSC-A, CD4-BV711 *vs.* SSC-A, and CD56-PE-Cy7 *vs.* CD19-Alexa Fluor 647. The gaiting is aimed at identifying $CD4^+$ T cells ($CD3^+CD4^+CD56^-CD19^-$). (**A**) Gating strategy for sample with 300,000 cells that was stained before fixation; (**B**) Gating strategy for sample with 300,000 cells that was fixed before staining.

Table S29 — Samples analyzed with flow cytometry

Subject Type	ID	Separation date	Analysis date	Input or output?	Frozen?	Number of single cells	Single cells (%)	$CD4^+$ T cells ($CD3^+CD4^+CD19^-CD56^-$)
HC	HC_2	21/04/2018	02/05/2018	input	Yes	7710	96.58%	31.21%
HC	HC_2	21/04/2018	02/05/2018	output	Yes	3806	97.89%	73.33%
V1	V1_1	21/04/2018	02/05/2018	output	Yes	964	92.78%	40.77%
V1	V1_2	21/04/2018	02/05/2018	input	Yes	1815	95.58%	34.05%
V1	V1_2	21/04/2018	02/05/2018	output	Yes	2631	96.48%	64.31%
V2	V2_1	21/04/2018	02/05/2018	output	Yes	8615	98.61%	48.52%
HC	HC_3	22/04/2018	02/05/2018	input	Yes	5280	98.40%	24.41%
HC	HC_3	22/04/2018	02/05/2018	output	Yes	5389	94.86%	54.85%
V1	V1_4	22/04/2018	02/05/2018	input	Yes	2074	94.62%	27.15%
V1	V1_4	22/04/2018	02/05/2018	output	Yes	5895	99.48%	79.56%
V2	V2_3	22/04/2018	02/05/2018	input	Yes	880	83.41%	15.23%
V2	V2_3	22/04/2018	02/05/2018	output	Yes	4145	97.76%	43.23%
HC	HC_5	29/04/2018	10/05/2018	input	Yes	28650	95.50%	22.02%
HC	HC_5	29/04/2018	10/05/2018	output	Yes	27878	92.93%	54.23%
V1	V1_5	29/04/2018	10/05/2018	input	Yes	29135	97.12%	27.43%
V1	V1_5	29/04/2018	10/05/2018	output	Yes	27296	90.99%	47.00%
V2	V2_5	29/04/2018	10/05/2018	input	Yes	28857	96.19%	30.25%
V2	V2_5	29/04/2018	10/05/2018	output	Yes	28329	94.43%	65.44%
HC	HC_8	30/04/2018	10/05/2018	input	Yes	29278	97.59%	23.37%
HC	HC_8	30/04/2018	10/05/2018	output	Yes	27533	91.78%	60.54%
V1	V1_8	30/04/2018	10/05/2018	output	Yes	29203	97.34%	72.52%
V2	V2_8	30/04/2018	10/05/2018	input	Yes	29205	97.35%	25.84%
V2	V2_8	30/04/2018	10/05/2018	output	Yes	27330	91.10%	54.53%
HC	HC_9	05/05/2018	09/05/2018	input		28338	94.46%	40.09%
HC	HC_9	05/05/2018	09/05/2018	output		28675	95.57%	72.92%
V1	V1_10	05/05/2018	09/05/2018	input		29151	96.31%	52.55%
V1	V1_10	05/05/2018	09/05/2018	output		28842	95.95%	77.90%
V2	V2_10	05/05/2018	09/05/2018	input		29178	96.39%	46.75%
V2	V2_10	05/05/2018	09/05/2018	output		28957	96.52%	70.50%
HC	HC_11	06/05/2018	09/05/2018	input		28486	94.95%	42.79%
HC	HC_11	06/05/2018	09/05/2018	output		28187	93.96%	74.96%
V1	V1_11	06/05/2018	09/05/2018	input		28441	94.80%	46.46%
V1	V1_11	06/05/2018	09/05/2018	output		28095	93.65%	72.96%
V2	V2_11	06/05/2018	09/05/2018	input		28780	95.93%	53.84%
V2	V2_11	06/05/2018	09/05/2018	output		29009	96.70%	81.09%
HC	HC_13	11/05/2018	16/05/2018	input		28813	96.04%	27.50%
HC	HC_13	11/05/2018	16/05/2018	output		28516	95.05%	56.63%
V1	V1_13	11/05/2018	16/05/2018	input		27929	93.10%	45.06%
V1	V1_13	11/05/2018	16/05/2018	output		28671	95.57%	60.23%
V2	V2_13	11/05/2018	16/05/2018	input		28936	96.45%	48.29%
V2	V2_13	11/05/2018	16/05/2018	output		29501	98.34%	62.55%
HC	HC_16	12/05/2018	16/05/2018	input		20151	98.24%	30.44%
HC	HC_16	12/05/2018	16/05/2018	output		28319	94.40%	61.22%

Table S29 — Samples analyzed with flow cytometry (continued)

Subject Type	ID	Separation date	Analysis date	Input or output?	Frozen?	Number of single cells	Single cells (%)	$CD4^+$ T cells ($CD3^+CD4^+CD19^-CD56^-$)
V1	V1_16	12/05/2018	16/05/2018	input		28731	95.77%	39.76%
V1	V1_16	12/05/2018	16/05/2018	output		29105	97.02%	72.97%
V2	V2_15	12/05/2018	16/05/2018	input		29315	97.72%	38.35%
V2	V2_15	12/05/2018	16/05/2018	output		28750	95.83%	64.70%
V1	V1_17	14/05/2018	19/05/2018	input		29160	97.20%	52.43%
V1	V1_17	14/05/2018	19/05/2018	output		28894	96.21%	79.26%
V2	V2_18	14/05/2018	19/05/2018	input		29685	98.95%	43.85%
V2	V2_18	14/05/2018	19/05/2018	output		29595	98.65%	66.31%
HC	HC_20	15/05/2018	19/05/2018	input		28901	96.34%	28.02%
HC	HC_20	15/05/2018	19/05/2018	output		27965	93.22%	61.93%
V1	V1_20	15/05/2018	19/05/2018	input		29396	97.99%	44.18%
V1	V1_20	15/05/2018	19/05/2018	output		29072	96.91%	71.38%
V2	V2_19	15/05/2018	19/05/2018	input		29210	97.37%	33.09%
V2	V2_19	15/05/2018	19/05/2018	output		28344	94.48%	66.30%
HC	HC_21	24/05/2019	27/05/2019	AfterCD4		25757	97.23%	70.61%
HC	HC_21	24/05/2019	27/05/2019	AfterCD14		32958	95.08%	3.23%
HC	HC_21	24/05/2019	27/05/2019	BeforeCD14		48928	95.04%	1.28%
HC	HC_57	24/05/2019	27/05/2019	AfterCD4		30278	98.17%	52.58%
HC	HC_57	24/05/2019	27/05/2019	AfterCD14		29016	95.77%	0.26%
HC	HC_57	24/05/2019	27/05/2019	BeforeCD14		22499	95.72%	0.48%
V1	V1_21	24/05/2019	27/05/2019	AfterCD4		9198	95.42%	54.40%
V1	V1_21	24/05/2019	27/05/2019	AfterCD14		24624	93.99%	2.27%
V2	V2_21	24/05/2019	27/05/2019	AfterCD14		32566	95.81%	1.16%
HC	HC_38	29/05/2019	02/06/2019	input		50769	94.23%	21.78%
HC	HC_38	29/05/2019	02/06/2019	output		31727	97.03%	73.64%
V1	V1_45	30/05/2019	02/06/2019	output		30923	98.35%	91.39%
V2	V2_45	30/05/2019	02/06/2019	input		55499	93.69%	34.42%
V2	V2_45	30/05/2019	02/06/2019	output		32563	96.11%	81.68%
HC	HC_35	31/05/2019	02/06/2019	output		18786	96.34%	71.20%
V1	V1_50	31/05/2019	02/06/2019	output		33237	91.62%	59.46%
V2	V2_35	31/05/2019	02/06/2019	output		8887	92.34%	35.06%
HC	HC_44	01/06/2019	02/06/2019	input		56248	95.44%	15.15%
HC	HC_44	01/06/2019	02/06/2019	output		34832	92.00%	54.31%
V1	V1_44	01/06/2019	02/06/2019	output		37810	91.05%	57.32%
V2	V2_44	01/06/2019	02/06/2019	output		38185	88.14%	51.84%
HC	HC_55	16/06/2019	17/06/2019	input		49031	94.84%	24.91%
HC	HC_55	16/06/2019	17/06/2019	output		32833	93.50%	74.18%
V1	V1_55	16/06/2019	17/06/2019	output		14810	82.85%	32.78%
V2	V2_47	16/06/2019	17/06/2019	output		19463	96.36%	52.10%
HC	HC_22	19/06/2019	21/06/2019	output		31154	95.79%	78.74%
HC	HC_58	19/06/2019	21/06/2019	input		62526	93.74%	15.66%
HC	HC_58	19/06/2019	21/06/2019	output		35522	93.78%	64.43%
V1	V1_22	19/06/2019	21/06/2019	output		30529	98.47%	84.92%
V2	V2_58	19/06/2019	21/06/2019	output		37227	91.16%	52.74%
HC	HC_24	20/06/2019	21/06/2019	output		31725	97.43%	82.71%

Table S29 — Samples analyzed with flow cytometry (continued)

Subject Type	ID	Separation date	Analysis date	Input or output?	Frozen?	Number of single cells	Single cells (%)	$CD4^+$ T cells ($CD3^+CD4^+CD19^-CD56^-$)
HC	HC_61	20/06/2019	21/06/2019	input		74544	96.13%	18.34%
HC	HC_61	20/06/2019	21/06/2019	output		32200	96.25%	79.86%
V1	V1_61	20/06/2019	21/06/2019	output		32298	96.06%	80.43%
V2	V2_24	20/06/2019	21/06/2019	output		34559	93.56%	66.69%
HC	HC_59	26/06/2019	28/06/2019	input		58174	94.81%	15.33%
HC	HC_59	26/06/2019	28/06/2019	output		34640	90.20%	60.82%
V1	V1_39	26/06/2019	28/06/2019	output		25663	92.68%	74.52%
V2	V2_59	26/06/2019	28/06/2019	input		67924	96.45%	20.71%
V2	V2_59	26/06/2019	28/06/2019	output		32369	95.82%	78.11%
HC	HC_23	27/06/2019	28/06/2019	input		70070	97.45%	14.27%
HC	HC_23	27/06/2019	28/06/2019	output		37649	89.69%	59.52%
V1	V1_42	27/06/2019	28/06/2019	output		32743	93.61%	79.21%
V2	V2_42	27/06/2019	28/06/2019	input		62336	96.37%	23.52%
V2	V2_42	27/06/2019	28/06/2019	output		34279	94.66%	74.26%
HC	HC_40	02/07/2019	04/07/2019	input		74232	94.16%	14.97%
HC	HC_40	02/07/2019	04/07/2019	output		35794	90.91%	62.81%
V1	V1_40	02/07/2019	04/07/2019	output		33408	93.38%	79.15%
HC	HC_51	03/07/2019	04/07/2019	output		35464	92.86%	68.45%
HC	HC_52	03/07/2019	04/07/2019	input		69810	96.93%	16.18%
HC	HC_52	03/07/2019	04/07/2019	output		33785	95.47%	72.98%
V1	V1_51	03/07/2019	04/07/2019	output		37565	90.56%	58.81%
V1	V1_52	03/07/2019	04/07/2019	input		81832	95.85%	18.61%
V1	V1_52	03/07/2019	04/07/2019	output		33064	95.45%	80.65%
V2	V2_51	03/07/2019	04/07/2019	output		34099	91.09%	71.45%
V2	V2_52	03/07/2019	04/07/2019	output		31346	97.01%	87.97%
HC	HC_29	17/07/2019	19/07/2019	input		55974	94.10%	22.61%
HC	HC_29	17/07/2019	19/07/2019	output		33272	95.16%	72.09%
V2	V2_29	17/07/2019	19/07/2019	output		31799	96.03%	80.23%
HC	HC_28	18/07/2019	19/07/2019	input		55953	93.98%	19.24%
HC	HC_28	18/07/2019	19/07/2019	output		33869	95.32%	69.72%
V1	V1_28	18/07/2019	19/07/2019	output		51012	89.98%	45.26%
V2	V2_41	18/07/2019	19/07/2019	output		33883	92.51%	74.87%
HC	HC_53	20/07/2019	25/07/2019	input		48227	94.13%	33.17%
HC	HC_53	20/07/2019	25/07/2019	output		31770	96.35%	72.55%
V1	V1_53	20/07/2019	25/07/2019	output		34281	93.27%	34.56%
HC	HC_60	21/07/2019	25/07/2019	input		50347	92.08%	20.42%
HC	HC_60	21/07/2019	25/07/2019	output		34385	89.20%	60.26%
V1	V1_60	21/07/2019	25/07/2019	output		34905	92.43%	60.41%
V2	V2_60	21/07/2019	25/07/2019	output		33070	95.94%	58.88%
HC	HC_31	15/08/2019	19/08/2019	input		38484	97.95%	27.26%
HC	HC_31	15/08/2019	19/08/2019	output		32284	98.60%	65.57%
V2	V2_31	15/08/2019	19/08/2019	output		10155	98.33%	64.30%
HC	HC_26	16/08/2019	19/08/2019	input		33807	94.12%	27.61%
HC	HC_26	16/08/2019	19/08/2019	output		31716	95.38%	74.78%
V1	V1_26	16/08/2019	19/08/2019	output		30452	98.47%	84.11%

Table S29 — Samples analyzed with flow cytometry (continued)

Subject Type	ID	Separation date	Analysis date	Input or output?	Frozen?	Number of single cells	Single cells (%)	$CD4^+$ T cells ($CD3^+CD4^+CD19^-CD56^-$)
HC	HC_54	17/08/2019	19/08/2019	input		39576	98.84%	18.07%
HC	HC_54	17/08/2019	19/08/2019	output		30936	96.60%	76.87%
V1	V1_32	17/08/2019	19/08/2019	input		39072	96.79%	26.54%
V1	V1_32	17/08/2019	19/08/2019	output		31878	95.89%	68.53%
V1	V1_27	20/08/2019	23/08/2019	input		40459	95.74%	31.01%
V1	V1_27	20/08/2019	23/08/2019	output		32569	89.84%	66.77%
V2	V2_27	20/08/2019	23/08/2019	output		30849	97.33%	75.29%
HC	HC_56	21/08/2019	23/08/2019	output		31742	90.42%	73.52%
V1	V1_56	21/08/2019	23/08/2019	output		30571	98.04%	81.31%

Abbreviations: HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

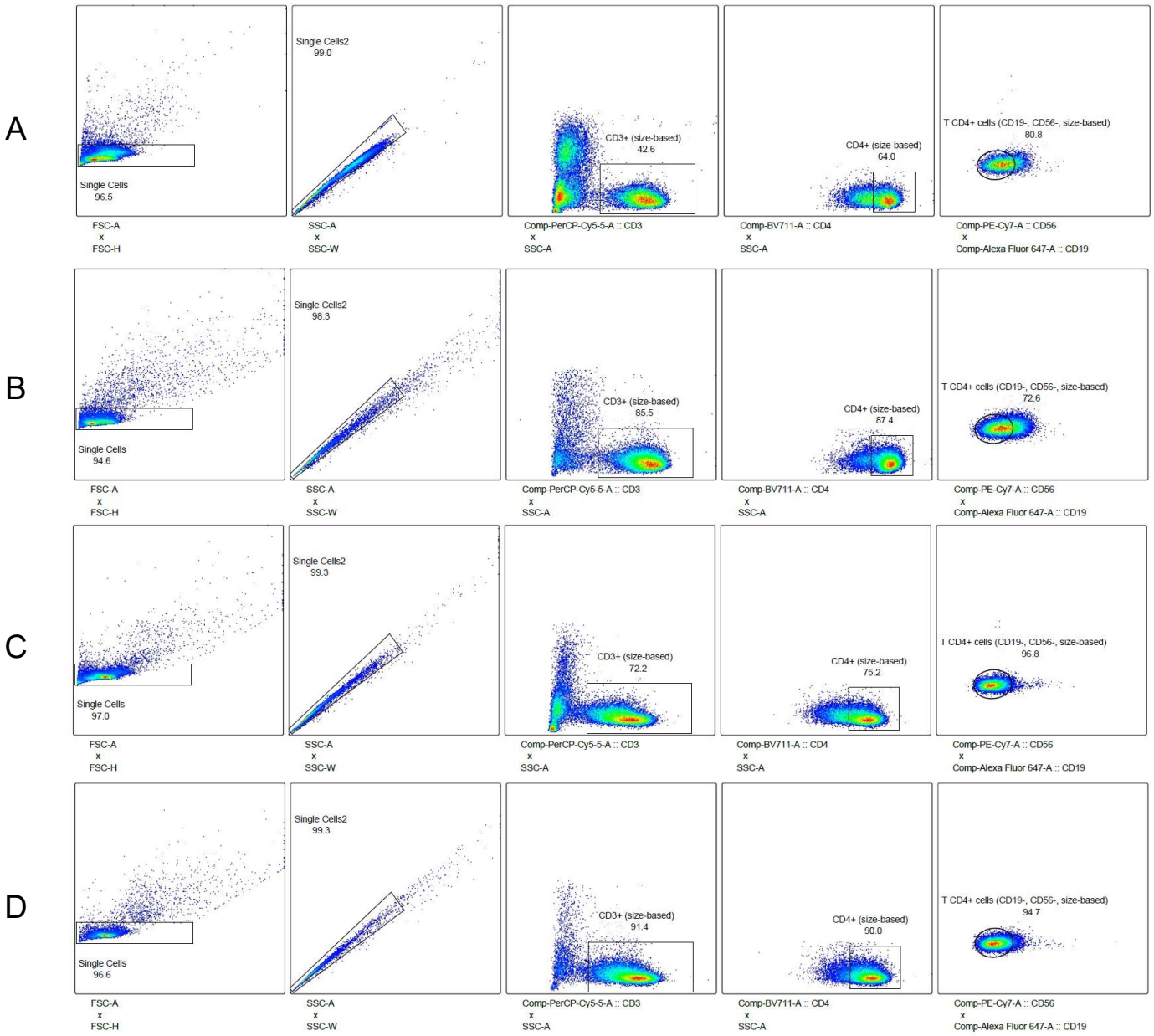


Figure S11 — Gating strategy in 2018 (Pilot study)

The order is: FSC-A *vs*. FSC-H, SSC-A *vs*. SSC-W, CD3-PerCP-Cy5.5 *vs.* SSC-A, CD4-BV711 *vs.* SSC-A, and CD56-PE-Cy7 *vs.* CD19-Alexa Fluor 647. The gaiting is aimed at identifying CD4⁺ T cells ($CD3^+CD4^+CD56^-CD19^-$). (**A, B**) show a previously frozen sample from a healthy control before and after undergoing CD4⁺ T cell enrichment, respectively; (**C, D**) show a freshly-fixed sample from a patient at V1 before and after undergoing CD4⁺ T cell enrichment, respectively.

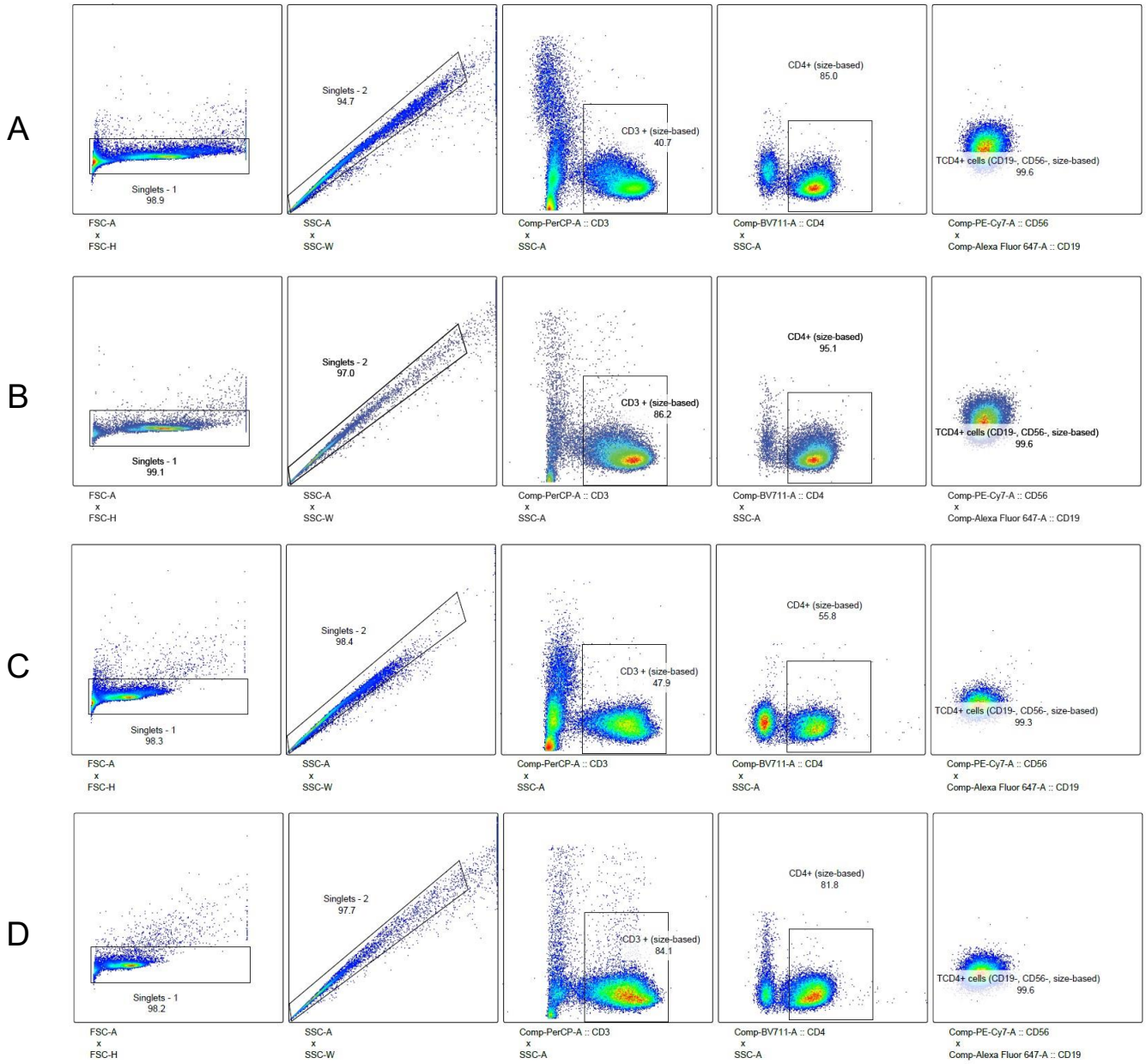


Figure S12 — Gating strategy in 2019

The order is: FSC-A *vs*. FSC-H, SSC-A *vs*. SSC-W, CD3-PerCP-Cy5.5 *vs.* SSC-A, CD4-BV711 *vs.* SSC-A, and CD56-PE-Cy7 *vs.* CD19-Alexa Fluor 647. The gaiting is aimed at identifying $CD4^+$ T cells ($CD3^+CD4^+CD56^-CD19^-$). (**A, B**) show a sample analyzed before the scheduled maintenance of the cell analyzer, on August 12, 2019, from a patient at V2 before and after undergoing $CD4^+$ T cell enrichment, respectively; (**C, D**) show a sample analyzed after the scheduled maintenance of the cell analyzer, from a patient at V1 before and after undergoing $CD4^+$ T cell enrichment, respectively.

Annex C — mtDNA extraction and NGS sequencing

Page intentionally left blank

Table S30 — DNA extraction from clinical samples after CD4⁺ T cell enrichment

Subject Type	ID	Separation date	DNA extraction date	Number of cells after enrichment (millions)	Mortality	Flow cytometry analysis?	DNA (Synergy HTX) (ng/µL)	260/280 ratio	DNA (Qubit®) (ng/µL)
HC	HC_1	21/04/2018	06/07/2018	2.63	48.05%	No	20.12	1.96	13.00
V1	V1_1	21/04/2018	06/07/2018	4.11	8.62%	Yes, output only	38.40	1.86	21.00
V2	V2_1	21/04/2018	06/07/2018	2.38	10.73%	Yes, output only	6.89	2.38	9.64
HC	HC_2	21/04/2018	06/07/2018	4.92	8.71%	Yes	38.83	1.92	26.00
V1	V1_2	21/04/2018	06/07/2018	3.94	6.50%	Yes	23.06	1.93	14.80
V2	V2_2	21/04/2018	06/07/2018	2.62	11.18%	No	15.83	2.01	13.00
HC	HC_3	22/04/2018	19/07/2018	4.42	15.27%	Yes	24.32	1.91	9.22
V1	V1_3	22/04/2018	19/07/2018	3.83	12.47%	No	20.16	1.94	16.10
V2	V2_3	22/04/2018	19/07/2018	6.09	11.61%	Yes	28.31	2.18	26.80
HC	HC_4	22/04/2018	14/06/2018	1.27	11.67%	No	5.25	1.86	5.40
V1	V1_4	22/04/2018	14/06/2018	1.93	6.78%	Yes	7.30	2.02	4.80
V2	V2_4	22/04/2018	14/06/2018	5.40	9.76%	No	26.78	1.95	19.70
HC	HC_5	29/04/2018	11/07/2018	2.69	10.70%	Yes	24.29	1.66	9.20
V1	V1_5	29/04/2018	11/07/2018	6.40	9.53%	Yes	45.33	1.83	37.00
V2	V2_5	29/04/2018	11/07/2018	5.29	13.49%	Yes	33.65	2.10	17.10
HC	HC_6	29/04/2018	14/06/2018	4.62	13.58%	No	45.79	2.28	29.80
V1	V1_6	29/04/2018	14/06/2018	1.80	12.65%	No	6.30	3.05	6.10
V2	V2_6	29/04/2018	14/06/2018	2.13	22.36%	No	14.70	2.34	10.60
HC	HC_7	30/04/2018	29/06/2018	4.00	11.67%	No	26.96	2.11	18.20
V1	V1_7	30/04/2018	29/06/2018	1.81	11.63%	No	10.87	1.19	4.80
V2	V2_7	30/04/2018	29/06/2018	3.08	12.46%	No	15.93	1.97	11.30
HC	HC_8	30/04/2018	06/07/2018	8.27	15.29%	Yes	65.24	1.91	56.00
V1	V1_8	30/04/2018	06/07/2018	2.30	16.08%	Yes, output only	2.93	1.36	5.50
V2	V2_8	30/04/2018	06/07/2018	3.50	19.52%	Yes	34.04	2.00	19.70
HC	HC_9	05/05/2018	11/07/2018	2.66	12.56%	Yes	17.53	1.55	16.30
V1	V1_9	05/05/2018	11/07/2018	3.07	17.78%	No	16.22	2.06	14.10
V2	V2_9	05/05/2018	11/07/2018	4.01	15.96%	No	25.25	1.94	22.40
HC	HC_10	05/05/2018	19/07/2018	4.34	16.96%	No	26.07	2.08	16.90
V1	V1_10	05/05/2018	19/07/2018	4.78	11.08%	Yes	20.79	2.43	23.60
V2	V2_10	05/05/2018	19/07/2018	8.10	8.63%	Yes	53.46	1.92	27.20
HC	HC_11	06/05/2018	19/07/2018	2.78	13.57%	Yes	9.67	2.19	8.10
V1	V1_11	06/05/2018	19/07/2018	5.27	10.21%	Yes	27.86	1.84	19.30
V2	V2_11	06/05/2018	19/07/2018	7.38	9.60%	Yes	26.71	1.99	20.80
HC	HC_12	06/05/2018	29/06/2018	1.97	14.04%	No	6.66	2.02	6.20
V1	V1_12	06/05/2018	29/06/2018	5.59	11.12%	No	28.59	1.78	19.30
V2	V2_12	06/05/2018	29/06/2018	4.42	8.94%	No	32.19	1.31	26.20
HC	HC_13	11/05/2018	11/07/2018	5.43	11.70%	Yes	39.35	2.04	27.00
V1	V1_13	11/05/2018	11/07/2018	2.64	12.98%	Yes	12.47	1.96	9.48
V2	V2_13	11/05/2018	11/07/2018	3.88	8.35%	Yes	17.34	1.91	13.90
HC	HC_14	11/05/2018	29/06/2018	1.94	40.28%	No	8.86	1.51	6.50
V1	V1_14	11/05/2018	29/06/2018	2.36	17.60%	No	11.13	1.94	7.96
V2	V2_14	11/05/2018	29/06/2018	4.13	12.67%	No	24.46	1.98	17.40

Table S30 — DNA extraction from clinical samples after CD4⁺ T cell enrichment (continued)

Subject Type	ID	Separation date	DNA extraction date	Number of cells after enrichment (millions)	Mortality	Flow cytometry analysis?	DNA (Synergy HTX) (ng/µL)	260/280 ratio	DNA (Qubit®) (ng/µL)
HC	HC_15	12/05/2018	29/06/2018	4.57	10.73%	No	20.80	1.97	17.50
V1	V1_15	12/05/2018	29/06/2018	2.08	12.78%	No	11.49	1.94	9.60
V2	V2_15	12/05/2018	29/06/2018	2.75	9.90%	Yes	11.57	1.82	8.40
HC	HC_16	12/05/2018	11/07/2018	3.77	29.30%	Yes	11.70	2.06	11.50
V1	V1_16	12/05/2018	11/07/2018	2.59	14.16%	Yes	9.56	1.69	5.90
V2	V2_16	12/05/2018	11/07/2018	4.54	21.02%	No	24.41	1.98	17.30
HC	HC_17	14/05/2018	28/05/2018	1.81	11.75%	No	7.36	1.76	8.00
V1	V1_17	14/05/2018	28/05/2018	3.19	7.75%	Yes	13.16	1.90	10.80
V2	V2_17	14/05/2018	28/05/2018	1.01	14.46%	No	3.48	1.66	3.70
HC	HC_18	14/05/2018	28/05/2018	0.70	60.08%	No	1.45	1.19	2.70
V1	V1_18	14/05/2018	28/05/2018	2.39	15.39%	No	9.87	1.76	7.80
V2	V2_18	14/05/2018	28/05/2018	5.54	13.74%	Yes	33.12	2.17	19.90
HC	HC_19	15/05/2018	19/07/2018	4.39	15.90%	No	20.63	1.90	15.20
V1	V1_19	15/05/2018	19/07/2018	3.56	11.32%	No	16.58	1.75	12.20
V2	V2_19	15/05/2018	19/07/2018	4.17	12.50%	Yes	18.32	1.93	13.80
HC	HC_20	15/05/2018	06/07/2018	4.65	16.55%	Yes	20.96	2.16	19.00
V1	V1_20	15/05/2018	06/07/2018	3.83	9.08%	Yes	19.14	1.80	13.90
V2	V2_20	15/05/2018	06/07/2018	2.67	21.30%	No	11.55	2.21	10.30
HC	HC_21	24/05/2019	05/06/2019	2.20	14.65%	Yes	15.18	2.39	9.62
V1	V1_21	24/05/2019	05/06/2019	1.75	30.09%	Yes	6.60	2.78	5.34
V2	V2_21	24/05/2019	05/06/2019	1.22	28.13%	Yes, input only	11.12	1.99	6.10
HC	HC_57	24/05/2019	05/06/2019	1.70	14.49%	Yes	12.66	1.99	6.40
V1	V1_57	24/05/2019	05/06/2019	1.68	22.56%	No	13.77	3.37	6.90
V2	V2_57	24/05/2019	05/06/2019	1.57	18.99%	No	10.08	2.31	5.38
HC	HC_38	29/05/2019	05/06/2019	2.12	5.36%	Yes	12.55	1.97	7.80
V1	V1_38	29/05/2019	05/06/2019	0.86	9.27%	No	11.98	1.67	4.44
V2	V2_38	29/05/2019	05/06/2019	1.07	7.99%	No	10.22	2.07	2.12
HC	HC_30	29/05/2019	05/06/2019	1.70	12.42%	No	10.50	2.07	6.78
V1	V1_30	29/05/2019	05/06/2019	1.77	23.77%	No	8.81	1.69	5.00
V2	V2_30	29/05/2019	05/06/2019	0.10	56.25%	No	4.26	2.24	0.70
HC	HC_37	30/05/2019	05/06/2019	1.54	18.75%	No	7.37	1.58	5.20
V1	V1_37	30/05/2019	05/06/2019	1.75	19.29%	No	16.70	1.81	8.36
V2	V2_37	30/05/2019	05/06/2019	2.37	9.54%	No	16.86	1.90	7.58
HC	HC_45	30/05/2019	05/06/2019	1.51	22.38%	No	11.88	2.01	7.54
V1	V1_45	30/05/2019	05/06/2019	2.26	5.57%	Yes, output only	17.20	1.83	12.20
V2	V2_45	30/05/2019	05/06/2019	3.26	14.26%	Yes	25.09	2.00	16.70
HC	HC_50	31/05/2019	05/06/2019	0.37	70.86%	No	6.36	1.82	1.91
V1	V1_50	31/05/2019	05/06/2019	2.29	15.04%	Yes, output only	17.54	2.03	11.00
V2	V2_50	31/05/2019	05/06/2019	1.25	12.10%	No	7.32	1.89	4.22
HC	HC_35	31/05/2019	05/06/2019	1.78	14.05%	Yes, output only	11.54	1.97	6.52
V1	V1_35	31/05/2019	05/06/2019	1.33	13.33%	No	7.88	1.86	4.70
V2	V2_35	31/05/2019	05/06/2019	2.17	33.86%	Yes, output only	12.18	2.09	5.94
HC	HC_46	01/06/2019	05/06/2019	2.71	8.97%	No	13.52	1.83	9.60
V1	V1_46	01/06/2019	05/06/2019	1.73	23.08%	No	10.71	1.86	6.28

Table S30 — DNA extraction from clinical samples after CD4⁺ T cell enrichment (continued)

Subject Type	ID	Separation date	DNA extraction date	Number of cells after enrichment (millions)	Mortality	Flow cytometry analysis?	DNA (Synergy HTX) (ng/µL)	260/280 ratio	DNA (Qubit®) (ng/µL)
V2	V2_46	01/06/2019	05/06/2019	0.59	32.71%	No	8.30	1.91	4.72
HC	HC_44	01/06/2019	05/06/2019	2.20	10.72%	Yes	18.04	2.01	10.90
V1	V1_44	01/06/2019	05/06/2019	1.65	14.21%	Yes, output only	11.38	1.84	8.34
V2	V2_44	01/06/2019	05/06/2019	1.85	22.78%	Yes, output only	13.67	2.21	9.90
HC	HC_47	16/06/2019	14/07/2019	2.55	8.50%	No	23.12	1.95	16.10
V1	V1_47	16/06/2019	14/07/2019	2.04	23.17%	No	15.60	2.63	7.80
V2	V2_47	16/06/2019	14/07/2019	2.04	22.69%	Yes, output only	11.35	1.94	6.90
HC	HC_55	16/06/2019	14/07/2019	2.23	4.34%	Yes	17.82	1.98	10.90
V1	V1_55	16/06/2019	14/07/2019	1.75	18.90%	Yes, output only	12.15	1.99	7.90
V2	V2_55	16/06/2019	14/07/2019	1.36	20.56%	No	9.31	1.79	8.10
HC	HC_22	19/06/2019	14/07/2019	3.64	7.64%	Yes, output only	17.40	1.98	14.80
V1	V1_22	19/06/2019	14/07/2019	1.75	14.50%	Yes, output only	15.86	1.80	6.60
V2	V2_22	19/06/2019	14/07/2019	1.01	28.21%	No	6.58	2.34	4.50
HC	HC_58	19/06/2019	14/07/2019	2.75	5.00%	Yes	20.14	1.95	13.00
V1	V1_58	19/06/2019	14/07/2019	0.57	17.66%	No	26.26	1.89	16.90
V2	V2_58	19/06/2019	14/07/2019	2.09	25.35%	Yes, output only	13.42	1.90	7.96
HC	HC_61	20/06/2019	14/07/2019	2.90	8.05%	Yes	13.83	2.62	7.84
V1	V1_61	20/06/2019	14/07/2019	1.69	17.91%	Yes, output only	14.03	2.23	7.82
V2	V2_61	20/06/2019	14/07/2019	1.92	8.40%	No	14.34	2.03	10.40
HC	HC_24	20/06/2019	14/07/2019	8.92	4.42%	Yes, output only	53.27	2.11	35.40
V1	V1_24	20/06/2019	14/07/2019	2.06	33.41%	No	9.72	1.97	6.40
V2	V2_24	20/06/2019	14/07/2019	1.78	18.54%	Yes, output only	10.10	1.74	6.20
HC	HC_59	26/06/2019	14/07/2019	4.30	7.99%	Yes	24.79	1.93	17.70
V1	V1_59	26/06/2019	14/07/2019	2.17	12.85%	No	11.23	1.89	8.30
V2	V2_59	26/06/2019	14/07/2019	3.18	9.92%	Yes	16.52	1.91	12.30
HC	HC_39	26/06/2019	14/07/2019	3.26	5.12%	No	19.04	1.94	13.90
V1	V1_39	26/06/2019	14/07/2019	1.88	18.36%	Yes, output only	9.09	1.94	6.10
V2	V2_39	26/06/2019	14/07/2019	3.16	11.40%	No	25.08	2.33	13.10
HC	HC_42	27/06/2019	14/07/2019	4.56	7.16%	No	35.64	1.74	24.80
V1	V1_42	27/06/2019	14/07/2019	2.31	9.20%	Yes, output only	20.34	1.87	10.60
V2	V2_42	27/06/2019	14/07/2019	2.71	14.03%	Yes	17.69	1.96	12.00
HC	HC_23	27/06/2019	14/07/2019	2.85	8.70%	Yes	24.91	1.96	17.10
V1	V1_23	27/06/2019	14/07/2019	1.73	12.42%	No	13.64	2.08	8.24
V2	V2_23	27/06/2019	14/07/2019	2.30	11.66%	No	24.65	1.91	16.50
HC	HC_40	02/07/2019	23/07/2019	3.39	26.22%	Yes	24.42	1.76	18.00
V1	V1_40	02/07/2019	23/07/2019	3.01	5.96%	Yes, output only	20.75	1.75	15.50
V2	V2_40	02/07/2019	23/07/2019	1.95	18.27%	No	9.57	1.71	7.42
HC	HC_51	03/07/2019	23/07/2019	3.94	11.57%	Yes, output only	23.99	1.93	19.30
V1	V1_51	03/07/2019	23/07/2019	1.65	20.00%	Yes, output only	11.74	1.91	7.06
V2	V2_51	03/07/2019	23/07/2019	2.12	14.73%	Yes, output only	10.68	1.80	9.08
HC	HC_52	03/07/2019	23/07/2019	1.68	10.18%	Yes	6.67	1.98	5.32
V1	V1_52	03/07/2019	23/07/2019	4.03	9.62%	Yes	26.70	1.91	19.00
V2	V2_52	03/07/2019	23/07/2019	1.95	7.43%	Yes, output only	10.71	2.15	5.60
HC	HC_49	03/07/2019	23/07/2019	2.30	13.51%	No	14.37	1.61	7.66

Table S30 — DNA extraction from clinical samples after CD4⁺ T cell enrichment (continued)

Subject Type	ID	Separation date	DNA extraction date	Number of cells after enrichment (millions)	Mortality	Flow cytometry analysis?	DNA (Synergy HTX) (ng/µL)	260/280 ratio	DNA (Qubit®) (ng/µL)
V1	V1_49	03/07/2019	23/07/2019	2.24	8.97%	No	15.27	1.93	9.58
V2	V2_49	03/07/2019	23/07/2019	1.68	16.63%	No	10.62	1.85	7.66
HC	HC_34	17/07/2019	23/07/2019	1.77	23.99%	No	11.59	1.73	8.64
V1	V1_34	17/07/2019	23/07/2019	1.44	9.69%	No	11.03	1.77	8.80
V2	V2_34	17/07/2019	23/07/2019	1.12	16.38%	No	10.71	1.80	5.22
HC	HC_29	17/07/2019	24/07/2019	3.09	12.20%	Yes	20.51	1.89	13.10
V1	V1_29	17/07/2019	24/07/2019	1.75	30.36%	No	16.37	1.94	12.00
V2	V2_29	17/07/2019	24/07/2019	1.60	13.95%	Yes, output only	7.57	2.32	6.12
HC	HC_41	18/07/2019	24/07/2019	2.57	8.68%	No	19.95	1.81	14.60
V1	V1_41	18/07/2019	24/07/2019	2.20	14.68%	No	15.91	1.80	10.40
V2	V2_41	18/07/2019	24/07/2019	3.24	19.28%	Yes, output only	15.86	1.81	10.60
HC	HC_28	18/07/2019	24/07/2019	2.52	8.97%	Yes	12.55	1.84	10.40
V1	V1_28	18/07/2019	24/07/2019	1.92	9.93%	Yes, output only	16.30	1.81	13.80
V2	V2_28	18/07/2019	24/07/2019	1.31	21.77%	No	9.95	1.46	5.10
HC	HC_48	20/07/2019	24/07/2019	3.86	5.20%	No	25.23	1.83	16.20
V1	V1_48	20/07/2019	24/07/2019	2.09	15.38%	No	10.62	1.77	7.70
V2	V2_48	20/07/2019	24/07/2019	1.20	22.99%	No	13.44	1.77	8.32
HC	HC_53	20/07/2019	24/07/2019	3.77	20.66%	Yes	25.47	1.84	17.40
V1	V1_53	20/07/2019	24/07/2019	2.50	15.13%	Yes, output only	13.93	1.99	10.10
V2	V2_53	20/07/2019	24/07/2019	1.20	11.93%	No	4.91	5.20	4.24
HC	HC_25	21/07/2019	26/08/2019	2.12	7.86%	No	18.61	1.65	12.60
V1	V1_25	21/07/2019	26/08/2019	1.77	11.39%	No	14.99	1.69	7.92
V2	V2_25	21/07/2019	26/08/2019	0.33	18.75%	No	11.50	1.56	7.18
HC	HC_60	21/07/2019	26/08/2019	3.10	6.31%	Yes	27.14	1.71	18.80
V1	V1_60	21/07/2019	26/08/2019	3.50	12.42%	Yes, output only	24.05	1.77	16.20
V2	V2_60	21/07/2019	26/08/2019	1.67	19.09%	Yes, output only	13.11	1.47	8.90
HC	HC_31	15/08/2019	26/08/2019	5.72	7.71%	Yes	41.07	1.82	28.20
V1	V1_31	15/08/2019	26/08/2019	1.84	8.05%	No	17.67	1.47	10.20
V2	V2_31	15/08/2019	26/08/2019	1.64	22.81%	Yes, output only	7.01	1.28	3.20
HC	HC_43	15/08/2019	26/08/2019	2.63	7.92%	No	18.82	1.65	12.10
V1	V1_43	15/08/2019	26/08/2019	1.69	23.02%	No	12.40	1.52	7.04
V2	V2_43	15/08/2019	26/08/2019	0.64	43.65%	No	11.74	1.15	7.84
HC	HC_26	16/08/2019	26/08/2019	2.78	7.65%	Yes	23.59	1.62	15.60
V1	V1_26	16/08/2019	26/08/2019	1.58	10.01%	Yes, output only	14.90	1.50	8.68
V2	V2_26	16/08/2019	26/08/2019	1.22	26.24%	No	9.13	1.51	5.34
HC	HC_36	16/08/2019	26/08/2019	1.49	16.99%	No	11.14	1.50	6.36
V1	V1_36	16/08/2019	26/08/2019	0.82	26.92%	No	11.77	1.75	2.54
V2	V2_36	16/08/2019	26/08/2019	0.53	15.79%	No	25.70	1.66	16.40
HC	HC_54	17/08/2019	26/08/2019	2.49	19.20%	Yes	15.78	1.58	8.00
V1	V1_54	17/08/2019	26/08/2019	0.93	27.33%	No	8.08	1.36	4.26
V2	V2_54	17/08/2019	26/08/2019	0.74	27.21%	No	8.13	1.30	3.12
HC	HC_32	17/08/2019	26/08/2019	3.19	11.45%	No	24.14	1.72	17.20
V1	V1_32	17/08/2019	26/08/2019	3.08	20.46%	Yes	20.36	1.61	15.20
V2	V2_32	17/08/2019	26/08/2019	0.75	18.45%	No	9.64	1.39	5.58

Table S30 — DNA extraction from clinical samples after CD4⁺ T cell enrichment (continued)

Subject Type	ID	Separation date	DNA extraction date	Number of cells after enrichment (millions)	Mortality	Flow cytometry analysis?	DNA (Synergy HTX) (ng/µL)	260/280 ratio	DNA (Qubit®) (ng/µL)
HC	HC_33	20/08/2019	26/08/2019	1.16	17.73%	No	11.71	1.59	5.84
V1	V1_33	20/08/2019	26/08/2019	0.89	16.36%	No	11.49	1.61	4.78
V2	V2_33	20/08/2019	26/08/2019	1.58	19.39%	No	13.98	1.52	7.48
HC	HC_27	20/08/2019	26/08/2019	1.68	24.50%	No	18.20	1.54	10.60
V1	V1_27	20/08/2019	26/08/2019	2.29	14.13%	Yes	23.33	1.76	16.00
V2	V2_27	20/08/2019	26/08/2019	3.02	14.01%	Yes, output only	21.56	1.68	13.00
HC	HC_56	21/08/2019	26/08/2019	2.14	7.77%	Yes, output only	15.95	1.67	9.00
V1	V1_56	21/08/2019	26/08/2019	1.71	13.11%	Yes, output only	11.44	1.54	6.14
V2	V2_56	21/08/2019	26/08/2019	1.60	17.46%	No	10.73	1.39	5.10

Abbreviations: HC — healthy control; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Table S31 — qPCR on a set of trial samples: mtDNA *vs.* nuclear DNA

Target	Sample	Ct SD (internal)	Mean Ct
mtDNA	NTC	0.912	32.604
mtDNA	FCF 1 — P1 CD4+	0.108	14.696
mtDNA	FCF 2 — P1 CD4-	0.115	14.881
mtDNA	FCF 3 — P2 CD4-	0.260	15.045
GAPDH	NTC		36.986
GAPDH	FCF 1 — P1 CD4+	0.105	17.027
GAPDH	FCF 2 — P1 CD4-	0.184	17.152
GAPDH	FCF 3 — P2 CD4-	0.445	18.171

Abbreviations: Ct — threshold cycle; mtDNA — mitochondrial DNA; NTC — no template control; qPCR — quantitative polymerase chain reaction; SD — standard deviation.

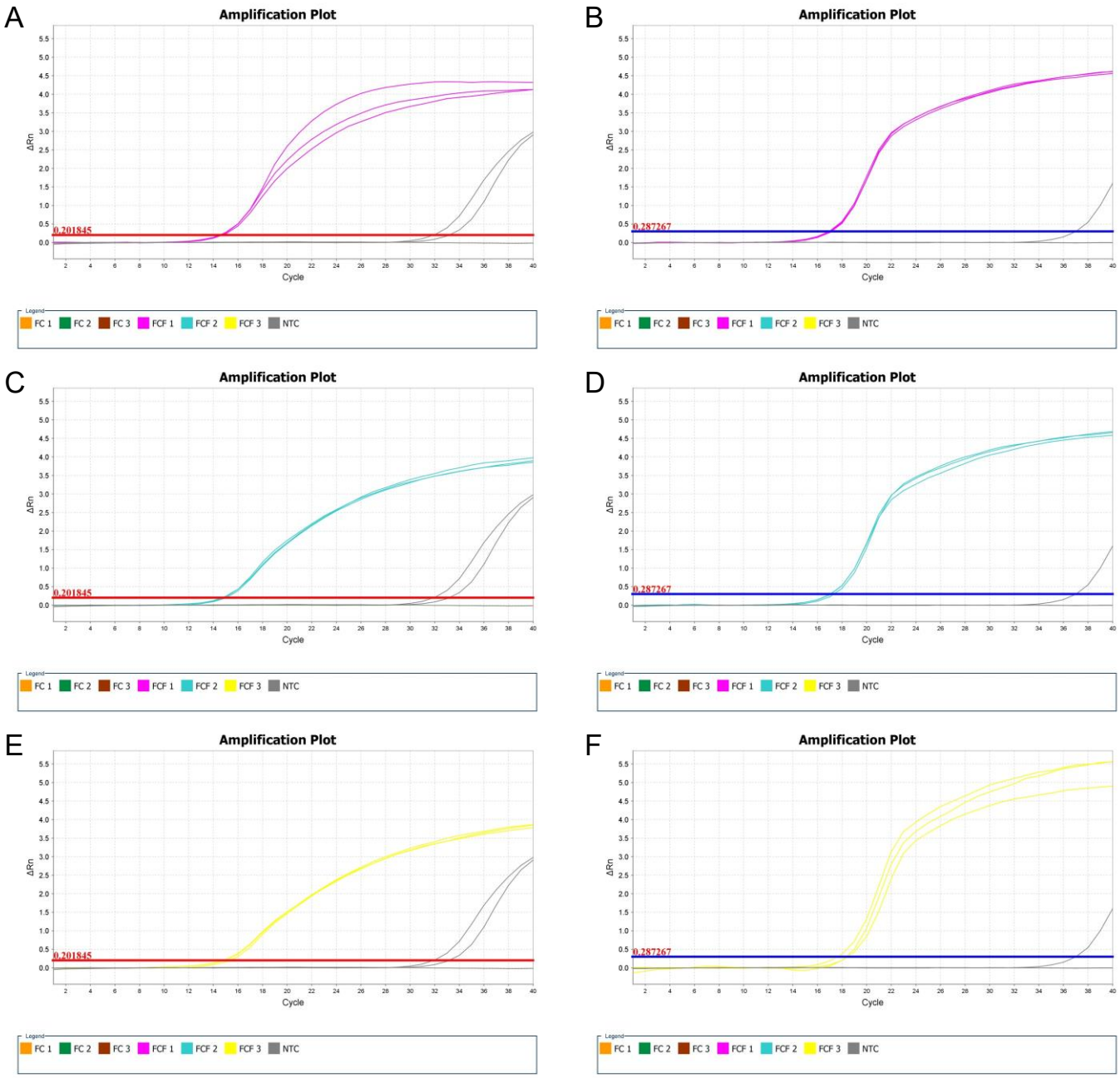


Figure S13 — Amplification plots of a qPCR from DNA extracted from trial samples

(**A, B**) Sample FCF 1 — P1 CD4+'s amplification plots of the mtDNA wells and the GAPDH wells, respectively; (**C, D**) Sample FCF 2 — P1 CD4-'s amplification plots of the mtDNA wells and the GAPDH wells, respectively; (**E, F**) Sample FCF 3 — P2 CD4-'s amplification plots of the mtDNA wells and the GAPDH wells, respectively. Grey depicts NTC wells from each set of primers. Abbreviations: mtDNA — mitochondrial DNA; NTC — no template control; qPCR — quantitative polymerase chain reaction.

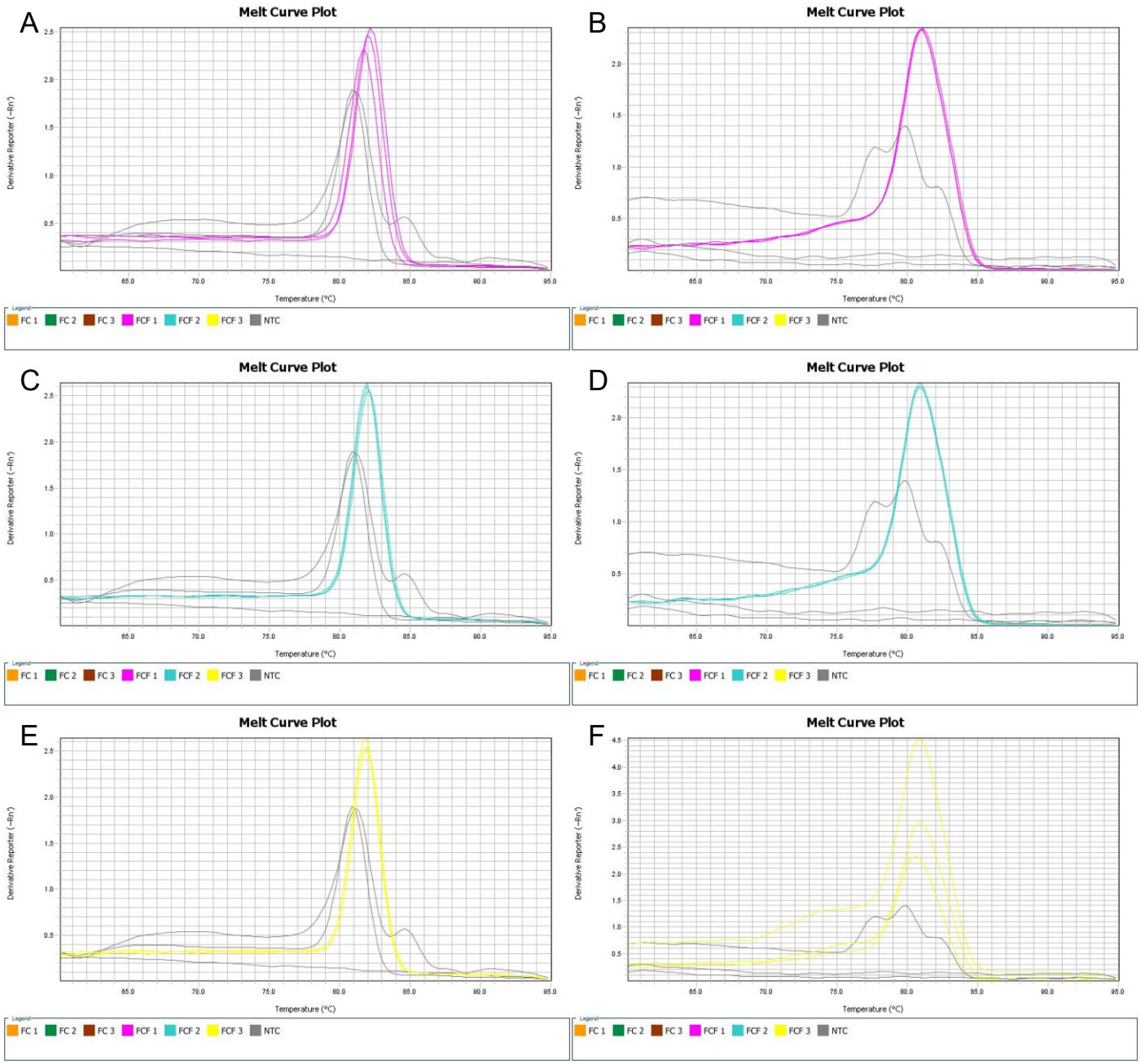


Figure S14 — Melt curve plots of a qPCR from DNA extracted from trial samples

(**A, B**) Sample FCF 1 — P1 CD4+'s melt curve plots of the mtDNA wells and the GAPDH wells, respectively; (**C, D**) Sample FCF 2 — P1 CD4-'s melt curve plots of the mtDNA wells and the GAPDH wells, respectively; (**E, F**) Sample FCF 3 — P2 CD4-'s melt curve plots of the mtDNA wells and the GAPDH wells, respectively. Grey depicts NTC wells from each set of primers. Abbreviations: mtDNA — mitochondrial DNA; NTC — no template control; qPCR — quantitative polymerase chain reaction.

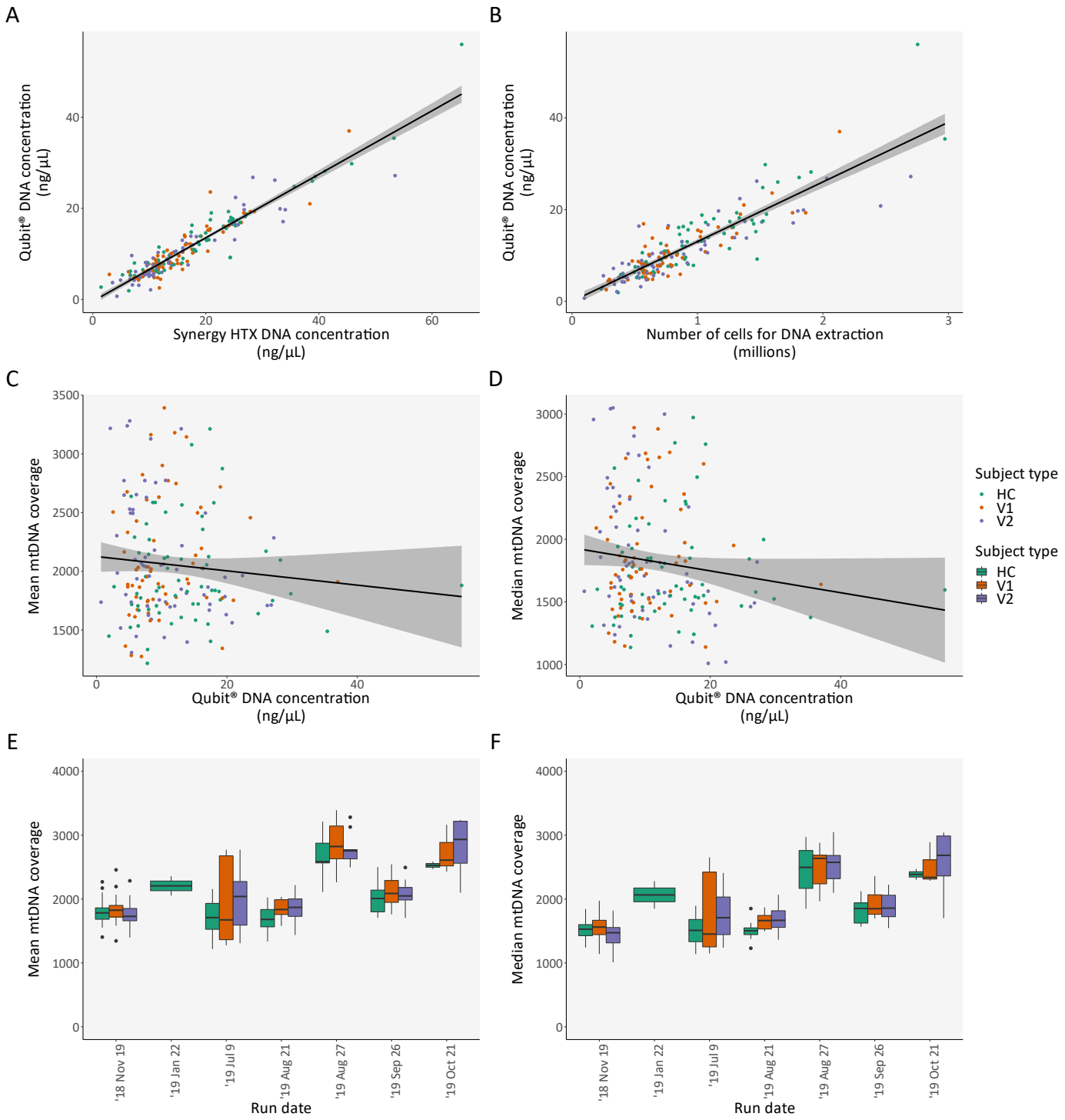


Figure S15 — NGS sequencing coverage analysis

(**A**) Correlation between two DNA measurements: one with the Synergy HTX Multi-mode Reader (BioTek Instruments, Winooski, VT, USA), and the other with a Qubit® 3.0 fluorometer (Thermo Fisher Scientific, Waltham, MA, USA); (**B**) Correlation between the number of cells used for DNA extraction and DNA concentration with a Qubit® 3.0 fluorometer; (**C,D**) Correlation between DNA concentration with a Qubit® 3.0 fluorometer and mean and median mtDNA coverage, respectively; (**E,F**) Mean and median mtDNA coverage, per subject type and per run date, respectively. Abbreviations: HC — healthy control; mtDNA — mitochondrial DNA; NGS — next-generation sequencing; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Table S32 — NGS sequencing in the CIS/MS cohort: Overall look

ID	Mappability >30 reads/base	Mappability >100 reads/base	Mean coverage	Haplogroup	Haplogroup (simplified)	Haplogroup quality (%)	Contamination?	Haplogroup origin	Number of variants
HC_1	99.95%	98.96%	2103.88	U4a1a	U	97.29%	No	Europe	33
V1_1	99.94%	98.78%	1753.64	H6a1a4	H	97.56%	No	Europe	18
V2_1	99.94%	98.48%	1757.07	H6a1a4	H	97.56%	No	Europe	17
HC_2	99.94%	99.20%	2171.30	H5a2	H	100.00%	No	Europe	14
V1_2	99.97%	99.39%	2068.36	I1a1	I	93.90%	No	Europe	42
V2_2	99.96%	98.29%	1848.08	I1a1	I	90.83%	No	Europe	51
HC_3	99.94%	97.91%	1833.79	H6a1a	H	94.18%	-	Europe	16
V1_3	99.92%	98.59%	2195.22	J1c2c2	J	95.46%	No	Europe	39
V2_3	99.91%	96.56%	1713.43	J1c2c2	J	95.46%	No	Europe	38
HC_4	99.92%	98.96%	1718.53	H	H	75.99%	No	Europe	21
V1_4	99.95%	98.70%	1888.68	T2b1	T	99.61%	No	Europe	35
V2_4	99.90%	97.43%	1949.34	T2b1	T	99.61%	No	Europe	34
HC_5	99.92%	96.96%	1706.79	H1b2	H	92.33%	No	Europe	14
V1_5	99.65%	97.54%	1910.67	H1a	H	92.20%	No	Europe	14
V2_5	99.66%	98.71%	1865.97	H1a1	H	94.96%	No	Europe	15
HC_6	98.90%	96.84%	1809.44	U5a2b1a	U	92.35%	No	Europe	29
V1_6	99.94%	97.88%	1782.18	H1au	H	89.76%	No	Europe	13
V2_6	99.92%	96.64%	1739.10	H1au	H	93.56%	No	Europe	16
HC_7	99.92%	96.60%	1837.93	H7a1	H	92.82%	No	Europe	13
V1_7	99.89%	97.02%	1867.96	J1c4	J	97.79%	No	Europe	31
V2_7	99.93%	97.21%	1697.30	J1c4	J	97.79%	No	Europe	30
HC_8	99.91%	98.35%	1880.49	H5a1j	H	100.00%	No	Europe	13
V1_8	99.94%	97.29%	1879.02	H2a2a1	H	50.00%	No	Europe	4
V2_8	99.93%	97.07%	1785.10	H2a2a1	H	50.00%	No	Europe	4
HC_9	99.97%	99.93%	2357.52	H1ap	H	92.83%	No	Europe	13
V1_9	99.92%	99.00%	1897.20	J1c3	J	96.67%	No	Europe	31
V2_9	99.70%	97.06%	1963.22	J1c3	J	95.34%	No	Europe	30
HC_10	99.98%	99.81%	2055.22	H1au	H	89.76%	No	Europe	17
V1_10	99.92%	98.56%	2456.93	T2f1a1	T	91.19%	No	Europe	39
V2_10	99.95%	98.15%	2285.53	T2f1a1	T	91.19%	No	Europe	39
HC_11	99.94%	98.80%	2270.24	H13a2b2a	H	96.74%	No	Europe	16
V1_11	99.93%	98.68%	1772.31	H15a1	H	98.69%	No	Europe	14
V2_11	99.95%	97.77%	1562.48	H15a1	H	98.69%	No	Europe	14
HC_12	99.93%	98.66%	1810.32	H5c1	H	94.58%	No	Europe	14
V1_12	99.93%	97.54%	1344.66	V*1	V	93.25%	No	Europe	18
V2_12	99.93%	97.49%	1709.44	V*1	V	93.25%	No	Europe	18
HC_13	99.92%	97.50%	1751.79	W1e1a	W	97.13%	No	Europe	40
V1_13	99.92%	99.12%	1894.20	H1z1	H	100.00%	-	Europe	13
V2_13	99.90%	97.05%	1397.82	H1z1	H	100.00%	No	Europe	14
HC_14	99.82%	98.77%	1668.46	K1a+150	K	94.66%	No	Europe	35
V1_14	99.46%	98.69%	1643.22	H1+152	H	81.34%	No	Europe	19
V2_14	99.89%	99.02%	1719.68	H1+152	H	81.34%	No	Europe	19

Table S32 — NGS sequencing in the CIS/MS cohort: Overall look (continued)

ID	Mappability >30 reads/base	Mappability >100 reads/base	Mean coverage	Haplogroup	Haplogroup (simplified)	Haplogroup quality (%)	Contamination?	Haplogroup origin	Number of variants
HC_15	99.92%	96.48%	1404.58	H1b	H	93.28%	No	Europe	13
V1_15	99.92%	97.04%	1582.60	J1c2t	J	100.00%	No	Europe	30
V2_15	99.90%	98.18%	1591.83	J1c2t	J	100.00%	No	Europe	30
HC_16	99.89%	97.47%	1553.89	V11	V	97.25%	No	Europe	17
V1_16	99.92%	98.18%	1613.20	K1b2a1	K	98.43%	No	Europe	36
V2_16	97.46%	92.96%	1669.34	K1b2a1	K	98.43%	No	Europe	35
HC_17	99.92%	96.95%	1675.36	U5b2b1a1	U	95.66%	No	Europe	35
V1_17	99.90%	96.93%	1722.89	U5b2a1a1	U	96.05%	No	Europe	29
V2_17	99.91%	96.55%	1517.45	U5b2a1a1	U	95.51%	No	Europe	29
HC_18	99.94%	98.25%	1869.41	X2b+226	X	96.12%	No	Europe	28
V1_18	99.87%	97.49%	1706.42	J1b1a1b	J	95.45%	No	Europe	39
V2_18	99.60%	97.09%	1631.77	J1b1a1b	J	96.69%	No	Europe	39
HC_19	99.16%	96.80%	1584.42	K1a1a	K	97.99%	No	Europe	33
V1_19	99.93%	99.11%	1780.61	K1c1f	K	95.47%	No	Europe	34
V2_19	99.89%	98.56%	1813.83	K1c1f	K	97.10%	No	Europe	34
HC_20	99.97%	98.27%	1757.74	I1a1c	I	97.35%	No	Europe	39
V1_20	99.95%	98.73%	1933.71	H16b	H	100.00%	No	Europe	9
V2_20	99.95%	98.34%	2061.30	H16b	H	100.00%	No	Europe	9
HC_21	99.95%	97.38%	1646.22	J1c1b1	J	98.58%	No	Europe	32
V1_21	99.92%	95.67%	1284.49	H28a	H	91.40%	No	Europe	12
V2_21	99.91%	95.57%	1446.05	H28a	H	91.40%	No	Europe	14
HC_22	99.96%	98.49%	1722.85	K1a1b2a1a	K	94.18%	No	Europe	41
V1_22	99.95%	97.89%	1865.40	H1c	H	90.83%	-	Europe	13
V2_22	99.95%	97.78%	1801.82	H1c	H	90.83%	No	Europe	13
HC_23	99.95%	98.80%	1550.76	HV0e	HV	90.08%	No	Europe	22
V1_23	99.94%	97.15%	1580.13	H	H	84.45%	No	Europe	11
V2_23	99.95%	97.92%	1799.29	H	H	84.45%	No	Europe	11
HC_24	99.96%	96.56%	1489.69	H1c1	H	94.85%	No	Europe	13
V1_24	99.94%	97.74%	2033.71	K1a+150	K	88.85%	No	Europe	36
V2_24	99.95%	99.02%	1936.47	K1a+150	K	88.85%	No	Europe	37
HC_25	99.98%	99.90%	1922.72	H3	H	87.84%	No	Europe	11
V1_25	99.97%	99.91%	2038.19	H3ae	H	87.65%	No	Europe	12
V2_25	99.98%	99.88%	2049.16	H3ae	H	84.17%	No	Europe	12
HC_26	99.98%	99.87%	1722.38	J1c2t	J	98.51%	No	Europe	30
V1_26	99.99%	99.80%	1762.09	U2e1b1	U	96.62%	No	Europe	38
V2_26	99.99%	99.92%	1737.76	U2e1b1	U	95.31%	No	Europe	37
HC_27	99.99%	99.94%	2503.62	H	H	75.99%	No	Europe	21
V1_27	99.99%	99.94%	2544.61	J2b1	J	94.21%	No	Europe	33
V2_27	99.99%	99.96%	3213.83	J2b1	J	94.21%	No	Europe	33
HC_28	99.97%	99.91%	2111.60	H5a1	H	95.32%	No	Europe	11
V1_28	99.98%	99.93%	3144.49	H1c	H	96.20%	No	Europe	10
V2_28	99.98%	99.94%	3281.08	H1c	H	96.20%	No	Europe	12
HC_29	99.98%	99.29%	2565.71	V1a	V	95.13%	No	Europe	16
V1_29	99.98%	99.95%	3180.15	H2a1	H	100.00%	No	Europe	15
V2_29	99.99%	99.96%	2653.38	H2a1	H	100.00%	No	Europe	7

Table S32 — NGS sequencing in the CIS/MS cohort: Overall look (continued)

ID	Mappability >30 reads/base	Mappability >100 reads/base	Mean coverage	Haplogroup	Haplogroup (simplified)	Haplogroup quality (%)	Contamination?	Haplogroup origin	Number of variants
HC_30	99.63%	98.03%	1773.66	X2c1a	X	95.54%	No	Europe	29
V1_30	99.92%	97.73%	1628.40	H7	H	91.58%	No	Europe	11
V2_30	99.95%	97.31%	1737.71	H7	H	91.58%	No	Europe	11
HC_31	99.99%	99.90%	2095.59	H24a	H	87.49%	No	Europe	12
V1_31	99.97%	99.86%	1869.81	H3*1	H	88.86%	No	Europe	11
V2_31	99.98%	99.87%	1985.97	H3*1	H	88.86%	No	Europe	12
HC_32	99.99%	99.93%	2125.28	H5u1	H	100.00%	No	Europe	13
V1_32	99.98%	99.90%	2136.68	J1c2	J	97.95%	No	Europe	30
V2_32	99.98%	99.93%	2494.11	J1c2	J	97.95%	No	Europe	30
HC_33	99.98%	99.93%	2291.16	T2e1	T	94.25%	No	Europe	37
V1_33	99.67%	99.61%	2332.25	J1b1a1+146	J	98.06%	No	Europe	39
V2_33	99.66%	99.02%	2119.40	J1b1a1+146	J	98.06%	No	Europe	40
HC_34	99.97%	99.93%	2586.54	H36	H	96.12%	No	Europe	18
V1_34	99.98%	99.81%	2261.72	H23	H	99.26%	No	Europe	12
V2_34	99.98%	99.93%	2496.00	H23	H	99.26%	No	Europe	9
HC_35	99.95%	97.41%	2157.32	H2a1	H	78.36%	No	Europe	14
V1_35	99.93%	99.17%	2677.27	K1a4c1	K	92.62%	No	Europe	44
V2_35	99.98%	99.89%	2100.89	K1a4c1	K	94.21%	No	Europe	70
HC_36	99.98%	99.23%	1706.13	T2b3+151	T	96.27%	No	Europe	36
V1_36	99.98%	99.94%	2504.74	H10e	H	100.00%	No	Europe	15
V2_36	99.98%	99.93%	2025.57	H10e	H	100.00%	No	Europe	10
HC_37	99.93%	97.21%	1523.01	J2a1a1a2	J	100.00%	No	Europe	41
V1_37	99.37%	97.63%	2020.24	H	H	88.79%	No	Europe	13
V2_37	99.35%	97.06%	2039.61	H	H	94.05%	No	Europe	12
HC_38	99.90%	96.02%	1217.10	H24	H	93.19%	No	Europe	10
V1_38	99.90%	95.08%	1363.63	T2b28	T	100.00%	No	Europe	34
V2_38	99.99%	99.96%	3217.46	T2b28	T	100.00%	No	Europe	38
HC_39	99.97%	98.77%	1828.62	U5a2a	U	51.90%	No	Europe	2
V1_39	99.96%	99.48%	2013.47	H1c+152	H	90.31%	No	Europe	14
V2_39	99.95%	99.27%	2218.68	H1c+152	H	90.31%	No	Europe	13
HC_40	99.98%	99.95%	2583.79	HV11	HV	90.91%	No	Europe	20
V1_40	99.97%	99.92%	2498.29	H5a1	H	100.00%	No	Europe	11
V2_40	99.98%	99.92%	2773.15	H5a1	H	100.00%	No	Europe	11
HC_41	99.98%	99.95%	3078.20	V11	V	97.84%	No	Europe	18
V1_41	99.98%	99.94%	3391.94	HV6	HV	95.99%	No	Europe	13
V2_41	99.98%	99.90%	2774.15	HV6	HV	95.99%	No	Europe	13
HC_42	99.94%	98.22%	1639.94	H11a	H	96.96%	No	Europe	11
V1_42	99.92%	99.23%	1806.80	K1a4d	K	98.84%	No	Europe	39
V2_42	99.94%	98.15%	1705.32	K1a4d	K	98.47%	No	Europe	37
HC_43	99.98%	99.90%	1792.53	H1b2	H	100.00%	No	Europe	13
V1_43	99.98%	99.87%	1921.85	H1b*	H	100.00%	No	Europe	13
V2_43	99.98%	99.87%	1703.96	H1b*	H	100.00%	No	Europe	13
HC_44	99.95%	99.64%	2123.54	U2e2a1a	U	95.91%	No	Europe	39
V1_44	99.99%	99.96%	3163.42	HV+16311	HV	91.58%	No	Europe	13
V2_44	99.94%	98.80%	2052.74	HV+16311	HV	91.58%	No	Europe	13

Table S32 — NGS sequencing in the CIS/MS cohort: Overall look (continued)

ID	Mappability >30 reads/base	Mappability >100 reads/base	Mean coverage	Haplogroup	Haplogroup (simplified)	Haplogroup quality (%)	Contamination?	Haplogroup origin	Number of variants
HC_45	99.94%	96.21%	1955.70	T2b4	T	94.76%	No	Europe	34
V1_45	99.96%	97.62%	2747.78	H1b	H	90.51%	No	Europe	13
V2_45	99.96%	98.87%	2496.57	H1b	H	92.51%	No	Europe	13
HC_46	99.93%	97.97%	1862.67	K1a4a1b2	K	99.29%	No	Europe	39
V1_46	99.93%	99.16%	1673.86	U5a1c1	U	100.00%	No	Europe	27
V2_46	99.99%	99.97%	3238.84	U5a1c1	U	100.00%	No	Europe	27
HC_47	99.95%	98.85%	1869.42	HV1a1a	HV	95.91%	No	Europe	19
V1_47	99.96%	98.20%	1955.61	T2b34	T	100.00%	No	Europe	45
V2_47	99.97%	97.41%	2068.20	T2b34	T	100.00%	No	Europe	43
HC_48	99.99%	99.95%	2468.72	H2a	H	87.12%	No	Europe	6
V1_48	99.98%	99.95%	2609.59	J1c2e1	J	98.74%	No	Europe	32
V2_48	99.96%	99.93%	3128.08	J1c2	J	97.79%	No	Europe	32
HC_49	99.98%	99.94%	2404.45	T2d1b1	T	95.07%	No	Europe	38
V1_49	99.98%	99.92%	2631.74	H1c2	H	90.67%	No	Europe	15
V2_49	99.99%	99.95%	2629.15	H1c2	H	90.67%	No	Europe	15
HC_50	99.95%	96.73%	1448.76	U5b3b2	U	98.54%	No	Europe	28
V1_50	99.95%	99.22%	2772.15	H1	H	90.86%	No	Europe	10
V2_50	99.96%	98.81%	2772.45	H1	H	90.86%	No	Europe	10
HC_51	99.98%	99.54%	2874.89	H1a1c	H	90.95%	No	Europe	17
V1_51	99.97%	99.93%	2823.02	HV	HV	89.69%	No	Europe	17
V2_51	99.98%	99.94%	2755.36	HV	HV	89.69%	No	Europe	12
HC_52	99.99%	99.96%	2638.26	I3a	I	97.06%	No	Europe	34
V1_52	99.98%	99.93%	2719.08	J2b1a	J	94.43%	No	Europe	35
V2_52	99.98%	99.95%	2525.17	J2b1a	J	94.43%	No	Europe	37
HC_53	99.98%	99.94%	3212.76	U5a2b1c	U	94.68%	No	Europe	27
V1_53	99.99%	99.95%	2901.51	U5a1a1+152	U	93.79%	No	Europe	32
V2_53	99.98%	99.93%	2649.38	U5a1a1+152	U	93.79%	No	Europe	31
HC_54	99.69%	99.63%	2145.05	U2d1	U	91.43%	No	Europe	38
V1_54	99.98%	99.92%	2165.20	H56	H	100.00%	No	Europe	9
V2_54	99.98%	99.93%	2303.96	H56	H	100.00%	No	Europe	10
HC_55	99.96%	98.70%	2025.92	H1a1	H	100.00%	No	Europe	17
V1_55	99.96%	97.88%	1752.22	H2a2a1	H	50.00%	No	Europe	1
V2_55	99.97%	99.22%	1959.85	H2a2a1	H	50.00%	No	Europe	1
HC_56	99.99%	99.96%	2586.49	H+152	H	96.01%	No	Europe	12
V1_56	99.98%	99.95%	2428.34	H24a	H	100.00%	No	Europe	11
V2_56	99.99%	99.95%	2530.23	H24a	H	100.00%	No	Europe	11
HC_57	99.90%	96.83%	1535.55	U4a	U	96.38%	No	Europe	27
V1_57	99.12%	94.30%	1275.44	K1a1b1f	K	98.74%	No	Europe	41
V2_57	99.14%	94.87%	1307.83	K1a1b1f	K	97.25%	No	Europe	38
HC_58	99.95%	98.44%	1607.76	T2b	T	96.02%	No	Europe	39
V1_58	99.93%	99.05%	1776.84	U5b2a2b	U	98.50%	No	Europe	33
V2_58	99.96%	97.56%	1626.77	U5b2a2b	U	98.50%	No	Europe	33
HC_59	99.96%	98.26%	1838.69	H1a	H	86.75%	No	Europe	17
V1_59	99.95%	99.52%	2000.74	H5a1	H	100.00%	No	Europe	11
V2_59	99.96%	99.49%	2015.76	H5a1	H	100.00%	No	Europe	13

Table S32 — NGS sequencing in the CIS/MS cohort: Overall look (continued)

ID	Mappability >30 reads/base	Mappability >100 reads/base	Mean coverage	Haplogroup	Haplogroup (simplified)	Haplogroup quality (%)	Contamination?	Haplogroup origin	Number of variants
HC_60	99.97%	99.86%	1828.78	V7a	V	96.86%	No	Europe	18
V1_60	99.97%	99.91%	2025.34	H85	H	95.09%	No	Europe	12
V2_60	99.98%	99.92%	2180.39	H85	H	95.09%	No	Europe	11
HC_61	99.95%	97.28%	1334.85	H1a3a	H	99.06%	No	Europe	13
V1_61	99.95%	98.04%	1743.89	H1a3a3	H	97.38%	No	Europe	18
V2_61	99.93%	97.24%	1436.93	H1a3a3	H	97.38%	No	Europe	17

Abbreviations: CIS — Clinically Isolated Syndrome; HC — healthy control; MS — Multiple Sclerosis; NGS — next-generation sequencing; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Table S33 — Mean number of mutations per haplogroup (simplified). Dunn test, after a Kruskal-Wallis test, with FDR

Haplogroup (simplified)	Mean number of mutations	Adjusted *p*-value
I ~ J	38.33 ~ 34.00	6.11E-01
I ~ K	38.33 ~ 37.80	9.30E-01
I ~ T	38.33 ~ 37.44	9.30E-01
I ~ U	38.33 ~ 29.79	3.48E-01
I ~ V	38.33 ~ 17.40	1.48E-01
K ~ T	37.80 ~ 37.44	9.30E-01
K ~ U	37.80 ~ 29.79	1.84E-01
K ~ V	37.80 ~ 17.40	6.26E-02
T ~ U	37.44 ~ 29.79	2.17E-01
T ~ V	37.44 ~ 17.40	7.92E-02
J ~ K	34.00 ~ 37.80	5.14E-01
J ~ T	34.00 ~ 37.44	5.97E-01
J ~ U	34.00 ~ 29.79	4.94E-01
J ~ V	34.00 ~ 17.40	1.56E-01
U ~ V	29.79 ~ 17.40	3.48E-01
HV ~ I	17.33 ~ 38.33	8.90E-02
HV ~ J	17.33 ~ 34.00	7.92E-02
HV ~ K	17.33 ~ 37.80	2.57E-02
HV ~ T	17.33 ~ 37.44	3.51E-02
HV ~ U	17.33 ~ 29.79	2.17E-01
HV ~ V	17.33 ~ 17.40	9.20E-01
H ~ HV	13.05 ~ 17.33	2.18E-01
H ~ I	13.05 ~ 38.33	1.87E-03
H ~ J	13.05 ~ 34.00	2.03E-07
H ~ K	13.05 ~ 37.80	4.28E-08
H ~ T	13.05 ~ 37.44	2.03E-07
H ~ U	13.05 ~ 29.79	2.84E-05
H ~ V	13.05 ~ 17.40	1.84E-01

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate.

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
12612G	3	0	47	11	6.14E-02	1	18.03%	4.92%	13.11%	-
3010A	7	5	34	15	1.36E-01	1	24.59%	11.48%	13.11%	-
14798C	4	2	44	11	1.21E-01	1	18.03%	6.56%	11.48%	MS (risk) (Kalman, Lublin and Alder, 1996)
16069T	3	0	48	10	9.61E-02	1	16.39%	4.92%	11.48%	-
295T	3	0	48	10	9.61E-02	1	16.39%	4.92%	11.48%	-
489C	3	0	48	10	9.61E-02	1	16.39%	4.92%	11.48%	-
10398G	9	2	35	15	3.07E-01	1	24.59%	14.75%	9.84%	MS (risk) (Yonova-Doing *et al.*, 2021)
11251G	7	1	40	13	2.64E-01	1	21.31%	11.48%	9.84%	-
15452A	7	1	40	13	2.64E-01	1	21.31%	11.48%	9.84%	-
16126C	7	2	39	13	2.64E-01	1	21.31%	11.48%	9.84%	-
4216C	7	1	40	13	2.64E-01	1	21.31%	11.48%	9.84%	MS (risk) (Andalib *et al.*, 2016)
462T	2	0	51	8	1.14E-01	1	13.11%	3.28%	9.84%	-
13708A	4	1	47	9	2.67E-01	1	14.75%	6.56%	8.20%	MS (risk) (Yu *et al.*, 2008)
146C	3	0	51	7	3.43E-01	1	11.48%	4.92%	6.56%	-
185A	2	1	52	6	2.89E-01	1	9.84%	3.28%	6.56%	-
8021G	0	0	57	4	1.34E-01	1	6.56%	0.00%	6.56%	-
8065A	0	0	57	4	1.34E-01	1	6.56%	0.00%	6.56%	-
16311C	6	3	43	9	6.06E-01	1	14.75%	9.84%	4.92%	-
228A	3	0	52	6	5.05E-01	1	9.84%	4.92%	4.92%	-
477C	1	0	56	4	3.71E-01	1	6.56%	1.64%	4.92%	-
9055A	3	1	51	6	5.05E-01	1	9.84%	4.92%	4.92%	MS (risk) (Vyshkina *et al.*, 2008)
10172A	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
10237C	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
10550G	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-
11299C	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-
1189C	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
12850G	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
13879C	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
13934T	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
14167T	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-
15812A	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
16193T	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
16222T	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
16224C	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-
16296T	2	0	55	4	6.83E-01	1	6.56%	3.28%	3.28%	-
16390A	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
16399G	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
16519C	14	26	5	16	8.55E-01	1	26.23%	22.95%	3.28%	-
188G	2	0	55	4	6.83E-01	1	6.56%	3.28%	3.28%	-
2158C	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
242T	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
3480G	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-
4732G	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
5460A	1	0	57	3	6.17E-01	1	4.92%	1.64%	3.28%	-
5580C	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
5633T	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
8269A	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
8557A	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
8701G	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
8715C	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
8718G	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
9629G	0	0	59	2	4.80E-01	1	3.28%	0.00%	3.28%	-
9698C	3	1	52	5	7.24E-01	1	8.20%	4.92%	3.28%	-
10169T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
10211T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10232G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10394T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10410C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10632C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10680A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10698T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11050C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11071T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11176A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11191T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11204C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11410C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11428T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11470G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11485C	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	-
11563G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11617C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11642A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
11719A	15	10	20	16	1.00E+00	1	26.23%	24.59%	1.64%	-
11778A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	NARP, LHON (Chung *et al.*, 2021)
11788T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
12083G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
12738G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
12771A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
12813T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
12814A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
12858T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
13064C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13134G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13145A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13191C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13281C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13590A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13637G	0	1	59	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13710G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13781C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13782A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13802T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13827G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13854T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
13933G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14016A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14097T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14153C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14218C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14370G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14386C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14470A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14766T	15	10	20	16	1.00E+00	1	26.23%	24.59%	1.64%	-
14893C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14953T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
14956C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15028A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15067C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15172A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
15218G	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	-
15257A	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	MS (risk) (Kalman, Lublin and Alder, 1995; Mayr-Wohlfart *et al.*, 1996)
15262C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15388C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15401G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15440C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15449C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15511C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15784C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15906G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15914G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
15930A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16093C	3	0	54	4	1.00E+00	1	6.56%	4.92%	1.64%	-
16111T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16145A	2	0	56	3	1.00E+00	1	4.92%	3.28%	1.64%	-
16147T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16180G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16188T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16217C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16221T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16239T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16245T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16246T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16249C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16256T	1	1	57	2	1.00E+00	1	3.28%	1.64%	1.64%	-
16259T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16260T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16262T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
16265G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16266T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16271C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16287T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16291T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16294A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16325C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16360T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16366T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16368C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
16398A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
1721T	0	1	59	1	1.00E+00	1	1.64%	0.00%	1.64%	-
186A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
199C	4	0	52	5	1.00E+00	1	8.20%	6.56%	1.64%	-
2028A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
214G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
2405A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
263G	2	56	0	3	1.00E+00	1	4.92%	3.28%	1.64%	-
2757G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
2857C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3212T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
327T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3335C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3337A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3337C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3396C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
340T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3507T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
372C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3796G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
3912G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
4502C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
4634C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
4755C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
4823C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
508G	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	-
5264T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5277C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5302C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5351G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5426C	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	-
5437T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5495C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
55C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5664G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
57C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
5913A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6023A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6119T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6253C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6266G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6302G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6323G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6465A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6489A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6528T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
6569A	2	0	56	3	1.00E+00	1	4.92%	3.28%	1.64%	-
6755A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6776C	0	1	59	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6852A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
6872G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7051C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7094C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7109T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
723C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
735G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7374G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7424G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7476T	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	-
7598A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7849T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
7852A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8179G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8378G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8461T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8464T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8468T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8558A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8598C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8666G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8793C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8842C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8843C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8850G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
8865A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8881C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8911C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
8961G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
896G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9010A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9075T	2	0	56	3	1.00E+00	1	4.92%	3.28%	1.64%	-
9093G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9123T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9129T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9145A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9163A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9185C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9254G	1	0	58	2	1.00E+00	1	3.28%	1.64%	1.64%	-
9355G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9456G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9527T	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9540C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9591A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9632G	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9861C	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
988A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
9948A	0	0	60	1	1.00E+00	1	1.64%	0.00%	1.64%	-
10685A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
10876G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
10915C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
11253C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
11377A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
114T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
12007A	2	0	57	2	1.00E+00	1	3.28%	3.28%	0.00%	-
12441C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
13020C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
13095C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
13105G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
13260C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
13404C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
13617C	2	2	55	2	1.00E+00	1	3.28%	3.28%	0.00%	-
13734C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
14182C	1	1	58	1	1.00E+00	1	1.64%	1.64%	0.00%	-
1438G	3	55	0	3	1.00E+00	1	4.92%	4.92%	0.00%	-
14793G	1	1	58	1	1.00E+00	1	1.64%	1.64%	0.00%	-
14978G	0	1	60	0	-		0.00%	0.00%	0.00%	-
15043A	2	0	57	2	1.00E+00	1	3.28%	3.28%	0.00%	-
15355A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
15497A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
15607G	4	0	53	4	1.00E+00	1	6.56%	6.56%	0.00%	-
15619T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
15833T	2	0	57	2	1.00E+00	1	3.28%	3.28%	0.00%	-
15907G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
15924G	3	0	55	3	1.00E+00	1	4.92%	4.92%	0.00%	-
15947G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16129A	3	0	55	3	1.00E+00	1	4.92%	4.92%	0.00%	-
16129C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16148T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16172C	4	0	53	4	1.00E+00	1	6.56%	6.56%	0.00%	-
16247G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
16320T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16324C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16354T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16362C	3	1	54	3	1.00E+00	1	4.92%	4.92%	0.00%	-
16482G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
16T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
1700C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
1811G	5	2	49	5	1.00E+00	1	8.20%	8.20%	0.00%	-
217C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
319C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
3333T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
3720G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
3915A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
3990T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
41T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
4318T	5	9	42	5	1.00E+00	1	8.20%	8.20%	0.00%	-
4727G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
4769G	3	55	0	3	1.00E+00	1	4.92%	4.92%	0.00%	-
4793G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
497T	4	0	53	4	1.00E+00	1	6.56%	6.56%	0.00%	-
5390G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
6045T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
6152C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
6734A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
7521A	2	0	57	2	1.00E+00	1	3.28%	3.28%	0.00%	-
7650T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
7768G	1	1	58	1	1.00E+00	1	1.64%	1.64%	0.00%	-
8140T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
8277C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
8616T	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
8706G	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
8943T	2	1	56	2	1.00E+00	1	3.28%	3.28%	0.00%	-
9060A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
9196A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
922A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
930A	3	0	55	3	1.00E+00	1	4.92%	4.92%	0.00%	-
9380A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
9434G	2	0	57	2	1.00E+00	1	3.28%	3.28%	0.00%	-
9477A	2	2	55	2	1.00E+00	1	3.28%	3.28%	0.00%	-
951A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
9530C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
9938C	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
9947A	1	0	59	1	1.00E+00	1	1.64%	1.64%	0.00%	-
10034C	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
10197A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10220G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10238C	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
10248C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10336C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10352G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10454C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10595C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10609C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10646A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10750G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10966C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
10993A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11053G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11260C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11467G	7	5	43	6	1.00E+00	1	9.84%	11.48%	-1.64%	-
11653G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11654G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11662C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11674T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11812G	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	-
11827C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11840T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11893G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11899C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11944C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11947G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
11978A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12127A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12172G	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
12308G	7	5	43	6	1.00E+00	1	9.84%	11.48%	-1.64%	MS (risk) (M. S. Kozin *et al.*, 2020)
12372A	7	5	43	6	1.00E+00	1	9.84%	11.48%	-1.64%	-
12373G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12414C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1243C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12501A	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
12510T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12557T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12618A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12634G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
12684A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12855G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
12937G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13056T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13174C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13326C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13368A	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	-
13449A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13488C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13575T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13630G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13722G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13740C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13759A	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
13762G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
13780G	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
13815A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14133G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14233G	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	-
14258A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14527G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14560A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1462A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14770A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14872T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
14905A	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	-
14926G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
151T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
15326G	2	58	0	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
15380G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15454C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15713G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15721C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15758G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15884C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15903G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
15928A	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	MS (risk) (Mayr-Wohlfart *et al.*, 1996)
16051G	2	1	57	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
16067T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16070G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16092C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16108T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16114A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16134T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16153A	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
16183C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16218T	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
16219G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16231C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16234T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16240G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16255A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16258G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16263C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16278T	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
16293G	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
16295T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16342C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16356C	3	1	55	2	1.00E+00	1	3.28%	4.92%	-1.64%	-
16361A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16391A	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
16400T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16463G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
16465T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1711T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1737G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1850C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1888A	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	-
189G	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
1900G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
1901T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
194T	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
2065G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
2075C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
215G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
2259T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
225A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
227G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
2294G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
2352C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
235G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
239C	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
250C	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
2765G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
2783G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3094A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3394C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3447G	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
3505G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3525A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3552C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3561G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3753C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3777C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3849A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
3861G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4025T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4122G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4227G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4336C	3	0	56	2	1.00E+00	1	3.28%	4.92%	-1.64%	-
4456T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4490T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4529T	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
4553C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4639C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
464G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
471C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4736C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
482C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4859C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
4917G	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	MS (risk) (Kalman, Lublin and Alder, 1995; Mayr-Wohlfart *et al.*, 1996; Vyshkina *et al.*, 2008)
4924A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
5046A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
5408G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
5471A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
5477T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
5747G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
5772A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
5839T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6040G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6249A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6260A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6278C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6406C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6446A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6620C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
6951A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
709A	6	1	49	5	1.00E+00	1	8.20%	9.84%	-1.64%	-
7146G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7184G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7226A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7256T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7269A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7278C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7295G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7316A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7340A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
73G	17	13	15	16	1.00E+00	1	26.23%	27.87%	-1.64%	-
7444A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7468T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
750G	2	58	0	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
7691C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7705C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
770T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7729G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7789A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7853A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
7864T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8014T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8098G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8167C	2	0	58	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
8242C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8322C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8393T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8410T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8448C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8473C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8477G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8578T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8639C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8659G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8697A	5	0	52	4	1.00E+00	1	6.56%	8.20%	-1.64%	-
8705C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8840T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8860G	2	58	0	1	1.00E+00	1	1.64%	3.28%	-1.64%	-
8869G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8887G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8938G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
8982G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
8994A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9150G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9151G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9157A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9300A	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9336C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9371T	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
93G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9617G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
961C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
961G	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
9800C	1	0	60	0	1.00E+00	1	0.00%	1.64%	-1.64%	-
10463C	6	0	51	4	7.52E-01	1	6.56%	9.84%	-3.28%	-
10499G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
11332T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
11620G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
11914A	4	0	55	2	6.83E-01	1	3.28%	6.56%	-3.28%	-
13966G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	MS (risk) (Kalman, Lublin and Alder, 1996)
14470C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
15098G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
15244G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
152C	9	2	43	7	8.03E-01	1	11.48%	14.75%	-3.28%	-
153G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
15693C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
15773A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
15927A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	MS (risk) (Mayr-Wohlfart *et al.*, 1996)
16080G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
16086C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
16162G	3	1	56	1	6.17E-01	1	1.64%	4.92%	-3.28%	-
16184T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
16207G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
16209C	4	0	55	2	6.83E-01	1	3.28%	6.56%	-3.28%	-
16261T	3	0	57	1	6.17E-01	1	1.64%	4.92%	-3.28%	-
16270T	3	2	55	1	6.17E-01	1	1.64%	4.92%	-3.28%	-
16292T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
16294T	6	0	51	4	7.52E-01	1	6.56%	9.84%	-3.28%	-
16355T	3	0	57	1	6.17E-01	1	1.64%	4.92%	-3.28%	-
207A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
2706G	18	14	13	16	8.64E-01	1	26.23%	29.51%	-3.28%	-
3184T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
3197C	4	2	53	2	6.83E-01	1	3.28%	6.56%	-3.28%	-
3357A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
449C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
4646C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
499A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
513A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
5147A	5	1	52	3	7.24E-01	1	4.92%	8.20%	-3.28%	-
5773A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
5999C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
6047G	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
6221C	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
6365C	3	0	57	1	6.17E-01	1	1.64%	4.92%	-3.28%	-
6371T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
6542T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
7028T	18	14	13	16	8.64E-01	1	26.23%	29.51%	-3.28%	-

Table S34 — Mutation distribution between HC and PwMS at V1. McNemar's test with FDR (continued)

Mutation	HC mutated & V1 non-mutated	HC & V1 mutated	HC & V1 non-mutated	V1 mutated & HC non-mutated	*p*-value	*p*-value (adjusted)	V1 mutation (discordant)	HC mutation (discordant)	V1-HC	Literature
72C	3	0	57	1	6.17E-01	1	1.64%	4.92%	-3.28%	-
8818T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
9548A	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
9908T	2	0	59	0	4.80E-01	1	0.00%	3.28%	-3.28%	-
14620T	3	0	58	0	2.48E-01	1	0.00%	4.92%	-4.92%	-
150T	6	2	50	3	5.05E-01	1	4.92%	9.84%	-4.92%	-
15904T	4	0	56	1	3.71E-01	1	1.64%	6.56%	-4.92%	-
16192T	5	0	54	2	4.50E-01	1	3.28%	8.20%	-4.92%	-
16298C	5	0	54	2	4.50E-01	1	3.28%	8.20%	-4.92%	-
16526A	4	0	56	1	3.71E-01	1	1.64%	6.56%	-4.92%	-
204C	3	0	58	0	2.48E-01	1	0.00%	4.92%	-4.92%	-
4580A	4	0	56	1	3.71E-01	1	1.64%	6.56%	-4.92%	-
16223T	5	0	55	1	2.21E-01	1	1.64%	8.20%	-6.56%	-
16304C	9	0	47	5	4.23E-01	1	8.20%	14.75%	-6.56%	-
195C	9	0	47	5	4.23E-01	1	8.20%	14.75%	-6.56%	-
200G	4	0	57	0	1.34E-01	1	0.00%	6.56%	-6.56%	-
456T	6	0	53	2	2.89E-01	1	3.28%	9.84%	-6.56%	-
5320T	5	0	55	1	2.21E-01	1	1.64%	8.20%	-6.56%	-
12705T	6	0	54	1	1.31E-01	1	1.64%	9.84%	-8.20%	-
16189C	9	1	47	4	2.67E-01	1	6.56%	14.75%	-8.20%	-
1719A	6	0	54	1	1.31E-01	1	1.64%	9.84%	-8.20%	-
6366A	7	0	52	2	1.82E-01	1	3.28%	11.48%	-8.20%	-
6383A	7	0	52	2	1.82E-01	1	3.28%	11.48%	-8.20%	-
8251A	7	0	53	1	7.71E-02	1	1.64%	11.48%	-9.84%	-

Abbreviations: FDR — false discovery rate; HC — healthy control; PwMS — patient with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1.

Table S35 — Mean number of deleterious mutations per haplogroup (simplified). Dunn test, after a Kruskal-Wallis test, with FDR

Haplogroups	Mean number of deleterious mutations	Adjusted *p*-value
J ~ K	2.62 ~ 0.50	2.62E-04
J ~ T	2.62 ~ 2.44	9.37E-01
J ~ U	2.62 ~ 0.36	1.11E-05
J ~ V	2.62 ~ 0.00	5.63E-05
T ~ U	2.44 ~ 0.36	9.65E-05
T ~ V	2.44 ~ 0.00	1.55E-04
I ~ J	1.33 ~ 2.62	2.08E-01
I ~ K	1.33 ~ 0.50	3.89E-01
I ~ T	1.33 ~ 2.44	2.44E-01
I ~ U	1.33 ~ 0.36	2.65E-01
I ~ V	1.33 ~ 0.00	1.39E-01
HV ~ I	0.83 ~ 1.33	9.37E-01
HV ~ J	0.83 ~ 2.62	7.91E-02
HV ~ K	0.83 ~ 0.50	3.47E-01
HV ~ T	0.83 ~ 2.44	1.09E-01
HV ~ U	0.83 ~ 0.36	1.99E-01
HV ~ V	0.83 ~ 0.00	9.29E-02
K ~ T	0.50 ~ 2.44	1.20E-03
K ~ U	0.50 ~ 0.36	6.93E-01
K ~ V	0.50 ~ 0.00	3.26E-01
U ~ V	0.36 ~ 0.00	4.36E-01
H ~ HV	0.22 ~ 0.83	7.91E-02
H ~ I	0.22 ~ 1.33	1.63E-01
H ~ J	0.22 ~ 2.62	9.77E-10
H ~ K	0.22 ~ 0.50	3.89E-01
H ~ T	0.22 ~ 2.44	3.46E-07
H ~ U	0.22 ~ 0.36	6.64E-01
H ~ V	0.22 ~ 0.00	5.75E-01

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate.

Table S36 — Mean cumulative deleterious burden per haplogroup (simplified). Dunn test, after a Kruskal-Wallis test, with FDR

Haplogroups	Mean cumulative deleterious burden	Adjusted *p*-value
J ~ K	3.92 ~ 2.60	2.47E-01
J ~ T	3.92 ~ 3.06	4.99E-01
J ~ U	3.92 ~ 2.00	1.32E-02
J ~ V	3.92 ~ 0.90	9.92E-06
T ~ U	3.06 ~ 2.00	1.42E-01
T ~ V	3.06 ~ 0.90	3.74E-04
K ~ T	2.60 ~ 3.06	6.13E-01
K ~ U	2.60 ~ 2.00	3.23E-01
K ~ V	2.60 ~ 0.90	1.50E-03
I ~ J	2.50 ~ 3.92	3.95E-01
I ~ K	2.50 ~ 2.60	9.23E-01
I ~ T	2.50 ~ 3.06	6.49E-01
I ~ U	2.50 ~ 2.00	5.56E-01
I ~ V	2.50 ~ 0.90	2.37E-02
U ~ V	2.00 ~ 0.90	1.37E-02
HV ~ I	1.41 ~ 2.50	3.00E-01
HV ~ J	1.41 ~ 3.92	5.22E-03
HV ~ K	1.41 ~ 2.60	1.25E-01
HV ~ T	1.41 ~ 3.06	4.35E-02
HV ~ U	1.41 ~ 2.00	4.32E-01
HV ~ V	1.41 ~ 0.90	1.73E-01
H ~ HV	1.01 ~ 1.41	2.61E-01
H ~ I	1.01 ~ 2.50	2.76E-02
H ~ J	1.01 ~ 3.92	5.12E-11
H ~ K	1.01 ~ 2.60	3.87E-05
H ~ T	1.01 ~ 3.06	5.48E-06
H ~ U	1.01 ~ 2.00	1.50E-03
H ~ V	1.01 ~ 0.90	4.63E-01

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate.

Table S37 — Mean cumulative MitoTIP score per haplogroup (simplified). Dunn test, after a Kruskal-Wallis test, with FDR

Haplogroups	Mean cumulative MitoTIP score	Adjusted *p*-value
K ~ T	0.46 ~ 0.25	2.54E-01
K ~ U	0.46 ~ 0.42	8.95E-01
K ~ V	0.46 ~ 0.02	4.45E-02
U ~ V	0.42 ~ 0.02	4.45E-02
I ~ J	0.29 ~ 0.05	2.26E-01
I ~ K	0.29 ~ 0.46	6.06E-01
I ~ T	0.29 ~ 0.25	8.17E-01
I ~ U	0.29 ~ 0.42	6.28E-01
I ~ V	0.29 ~ 0.02	3.20E-01
T ~ U	0.25 ~ 0.42	2.54E-01
T ~ V	0.25 ~ 0.02	3.17E-01
J ~ K	0.05 ~ 0.46	2.49E-03
J ~ T	0.05 ~ 0.25	1.30E-01
J ~ U	0.05 ~ 0.42	1.52E-03
J ~ V	0.05 ~ 0.02	8.17E-01
H ~ HV	0.04 ~ 0.03	6.65E-01
H ~ I	0.04 ~ 0.29	3.72E-02
H ~ J	0.04 ~ 0.05	2.14E-01
H ~ K	0.04 ~ 0.46	2.20E-07
H ~ T	0.04 ~ 0.25	1.44E-03
H ~ U	0.04 ~ 0.42	7.37E-09
H ~ V	0.04 ~ 0.02	2.54E-01
HV ~ I	0.03 ~ 0.29	4.27E-02
HV ~ J	0.03 ~ 0.05	2.42E-01
HV ~ K	0.03 ~ 0.46	1.99E-04
HV ~ T	0.03 ~ 0.25	1.25E-02
HV ~ U	0.03 ~ 0.42	1.47E-04
HV ~ V	0.03 ~ 0.02	2.42E-01

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate.

Table S38 — Mean proportion of novel variants per region. Dunn test, after a Kruskal-Wallis test, with FDR

Regions	Mean proportion of novel variants	Adjusted *p*-value
MT-TI ~ MT-ND1	1.00 ~ 0.51	1.98E-01
MT-TI ~ MT-RNR2	1.00 ~ 0.28	4.36E-02
MT-TI ~ MT-RNR1	1.00 ~ 0.16	3.93E-03
MT-TI ~ Other	1.00 ~ 1.00	1.00E+00
MT-CO2 ~ MT-CO1	0.85 ~ 0.63	3.93E-01
MT-CO2 ~ MT-ND2	0.85 ~ 0.29	1.16E-01
MT-CO2 ~ MT-TI	0.85 ~ 1.00	6.71E-01
MT-CO2 ~ MT-ND1	0.85 ~ 0.51	2.98E-01
MT-CO2 ~ MT-RNR2	0.85 ~ 0.28	9.25E-02
MT-CO2 ~ MT-RNR1	0.85 ~ 0.16	9.64E-03
MT-CO2 ~ Other	0.85 ~ 1.00	7.52E-01
MT-CO1 ~ MT-ND2	0.63 ~ 0.29	3.26E-01
MT-CO1 ~ MT-TI	0.63 ~ 1.00	2.26E-01
MT-CO1 ~ MT-ND1	0.63 ~ 0.51	6.57E-01
MT-CO1 ~ MT-RNR2	0.63 ~ 0.28	2.58E-01
MT-CO1 ~ MT-RNR1	0.63 ~ 0.16	5.56E-02
MT-CO1 ~ Other	0.63 ~ 1.00	3.42E-01
MT-CO3 ~ MT-ATP6	0.57 ~ 0.31	2.16E-01
MT-CO3 ~ MT-CO2	0.57 ~ 0.85	3.93E-01
MT-CO3 ~ MT-CO1	0.57 ~ 0.63	9.54E-01
MT-CO3 ~ MT-ND2	0.57 ~ 0.29	3.71E-01
MT-CO3 ~ MT-TI	0.57 ~ 1.00	2.33E-01
MT-CO3 ~ MT-ND1	0.57 ~ 0.51	7.04E-01
MT-CO3 ~ MT-RNR2	0.57 ~ 0.28	3.26E-01
MT-CO3 ~ MT-RNR1	0.57 ~ 0.16	9.07E-02
MT-CO3 ~ Other	0.57 ~ 1.00	3.42E-01
MT-ND1 ~ MT-RNR2	0.51 ~ 0.28	6.08E-01
MT-ND1 ~ MT-RNR1	0.51 ~ 0.16	2.38E-01
MT-ND1 ~ Other	0.51 ~ 1.00	2.57E-01
MT-ND5 ~ MT-CO3	0.42 ~ 0.57	6.57E-01
MT-ND5 ~ MT-ATP6	0.42 ~ 0.31	5.85E-01
MT-ND5 ~ MT-CO2	0.42 ~ 0.85	2.26E-01
MT-ND5 ~ MT-CO1	0.42 ~ 0.63	5.85E-01
MT-ND5 ~ MT-ND2	0.42 ~ 0.29	6.71E-01
MT-ND5 ~ MT-TI	0.42 ~ 1.00	1.22E-01
MT-ND5 ~ MT-ND1	0.42 ~ 0.51	9.73E-01
MT-ND5 ~ MT-RNR2	0.42 ~ 0.28	6.13E-01
MT-ND5 ~ MT-RNR1	0.42 ~ 0.16	2.33E-01
MT-ND5 ~ Other	0.42 ~ 1.00	2.19E-01
MT-ATP6 ~ MT-CO2	0.31 ~ 0.85	9.64E-03
MT-ATP6 ~ MT-CO1	0.31 ~ 0.63	1.16E-01
MT-ATP6 ~ MT-ND2	0.31 ~ 0.29	1.00E+00
MT-ATP6 ~ MT-TI	0.31 ~ 1.00	3.93E-03

Table S38 — Mean proportion of novel variants per region. Dunn test, after a Kruskal-Wallis test, with FDR (continued)

Regions	Mean proportion of novel variants	Adjusted *p*-value
MT-ATP6 ~ MT-ND1	0.31 ~ 0.51	5.85E-01
MT-ATP6 ~ MT-RNR2	0.31 ~ 0.28	9.47E-01
MT-ATP6 ~ MT-RNR1	0.31 ~ 0.16	3.42E-01
MT-ATP6 ~ Other	0.31 ~ 1.00	3.92E-02
MT-ND2 ~ MT-TI	0.29 ~ 1.00	5.79E-02
MT-ND2 ~ MT-ND1	0.29 ~ 0.51	6.62E-01
MT-ND2 ~ MT-RNR2	0.29 ~ 0.28	9.54E-01
MT-ND2 ~ MT-RNR1	0.29 ~ 0.16	4.70E-01
MT-ND2 ~ Other	0.29 ~ 1.00	1.16E-01
MT-RNR2 ~ MT-RNR1	0.28 ~ 0.16	5.54E-01
MT-RNR2 ~ Other	0.28 ~ 1.00	9.64E-02
MT-RNR1 ~ Other	0.16 ~ 1.00	1.44E-02
D-loop ~ MT-ND5	0.12 ~ 0.42	1.35E-01
D-loop ~ MT-CO3	0.12 ~ 0.57	2.52E-02
D-loop ~ MT-ATP6	0.12 ~ 0.31	2.12E-01
D-loop ~ MT-CO2	0.12 ~ 0.85	1.19E-03
D-loop ~ MT-CO1	0.12 ~ 0.63	9.64E-03
D-loop ~ MT-ND2	0.12 ~ 0.29	3.42E-01
D-loop ~ MT-TI	0.12 ~ 1.00	4.25E-04
D-loop ~ MT-ND1	0.12 ~ 0.51	1.59E-01
D-loop ~ MT-RNR2	0.12 ~ 0.28	3.93E-01
D-loop ~ MT-RNR1	0.12 ~ 0.16	9.47E-01
D-loop ~ Other	0.12 ~ 1.00	3.93E-03

Significant *p*-values are highlighted. Abbreviations: FDR — false discovery rate.

Table S39 — PwMS' medications: Summary

Medication type	Drug	*N* at V1	*N* at V2
Multiple Sclerosis	None	45	32
Multiple Sclerosis	Glatiramer acetate	9	11
Multiple Sclerosis	Interferon beta	5	8
Multiple Sclerosis	Fumarate	2	8
Multiple Sclerosis	Fingolimod	-	2
CNS and Neuromuscular	Biperiden	1	-
CNS and Neuromuscular	Lorazepam	1	-
CNS and Neuromuscular	Ziprasidone	1	1
CNS and Neuromuscular	Zopiclone	1	-
CNS and Neuromuscular	Baclofen	-	1
CNS and Neuromuscular	Citalopram	-	1
CNS and Neuromuscular	Mirtazapine	-	1
CNS and Neuromuscular	Sertraline	-	1
Cardiovascular	Ramipril	2	1
Cardiovascular	Irbesartan + Hydrochlorothiazide	1	1
Cardiovascular	Losartan	1	1
Cardiovascular	Metoprolol	1	2
Cardiovascular	Acetylsalicylic acid	-	2
Cardiovascular	Valsartan	-	2
Cardiovascular	Nebivolol	-	1
Endocrine	Levothyroxine	10	11
Endocrine	Oral contraceptive	7	9
Endocrine	Triiodothyronine	-	1
Supplemental	Vitamin D	16	24
Supplemental	Vitamin B12	3	3
Supplemental	Selenium	2	1
Supplemental	Vitamin B9	2	-
Supplemental	Iron	1	1
Supplemental	Isoflavone	1	1
Supplemental	Lipoic acid	1	-
Supplemental	Omega-3 fatty acids	1	-
Supplemental	Vitamin B complex	1	1
Supplemental	Vitamin B6 + Vitamin B12	1	-
Supplemental	Vitamin D + Calcium	1	-
Supplemental	Vitamin E	1	-
Supplemental	Vitamin K	1	-
Supplemental	Zinc	1	-
Supplemental	Magnesium	-	2
Other	Ibuprofen	3	3
Other	Desloratadine	1	-
Other	Hydroxychloroquine	1	1
Other	Intrauterine contraceptive device	1	1
Other	Pantoprazole	1	1
Other	Rupatadine	1	-

Table S39 — PwMS' medications: Summary (continued)

Medication type	Drug	*N* at V1	*N* at V2
Other	Topical corticosteroid	1	2
Other	Doxycycline	-	1
Other	Formoterol	-	1
Other	Ivermectin (topical)	-	1
Other	Paracetamol	-	1

Abbreviations: CNS — central nervous system; *N* — number; PwMS — patients with Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

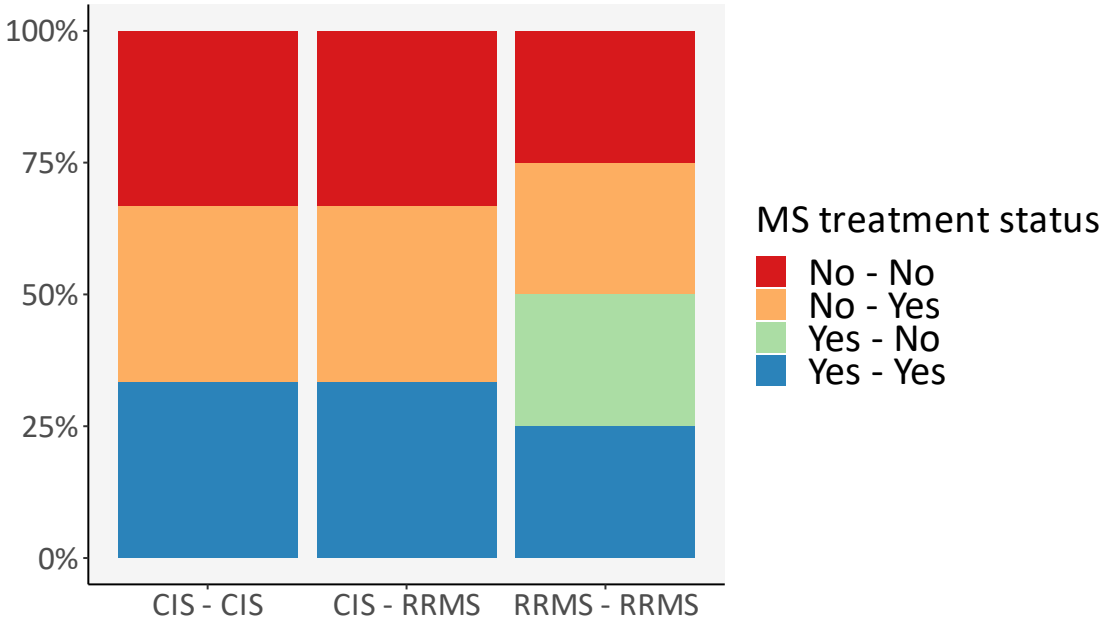


Figure S16 — MS treatment status between V1 and V2, per diagnosis change

Abbreviations: CIS — Clinically Isolated Syndrome; MS — Multiple Sclerosis; RRMS — Relapsing-Remitting Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

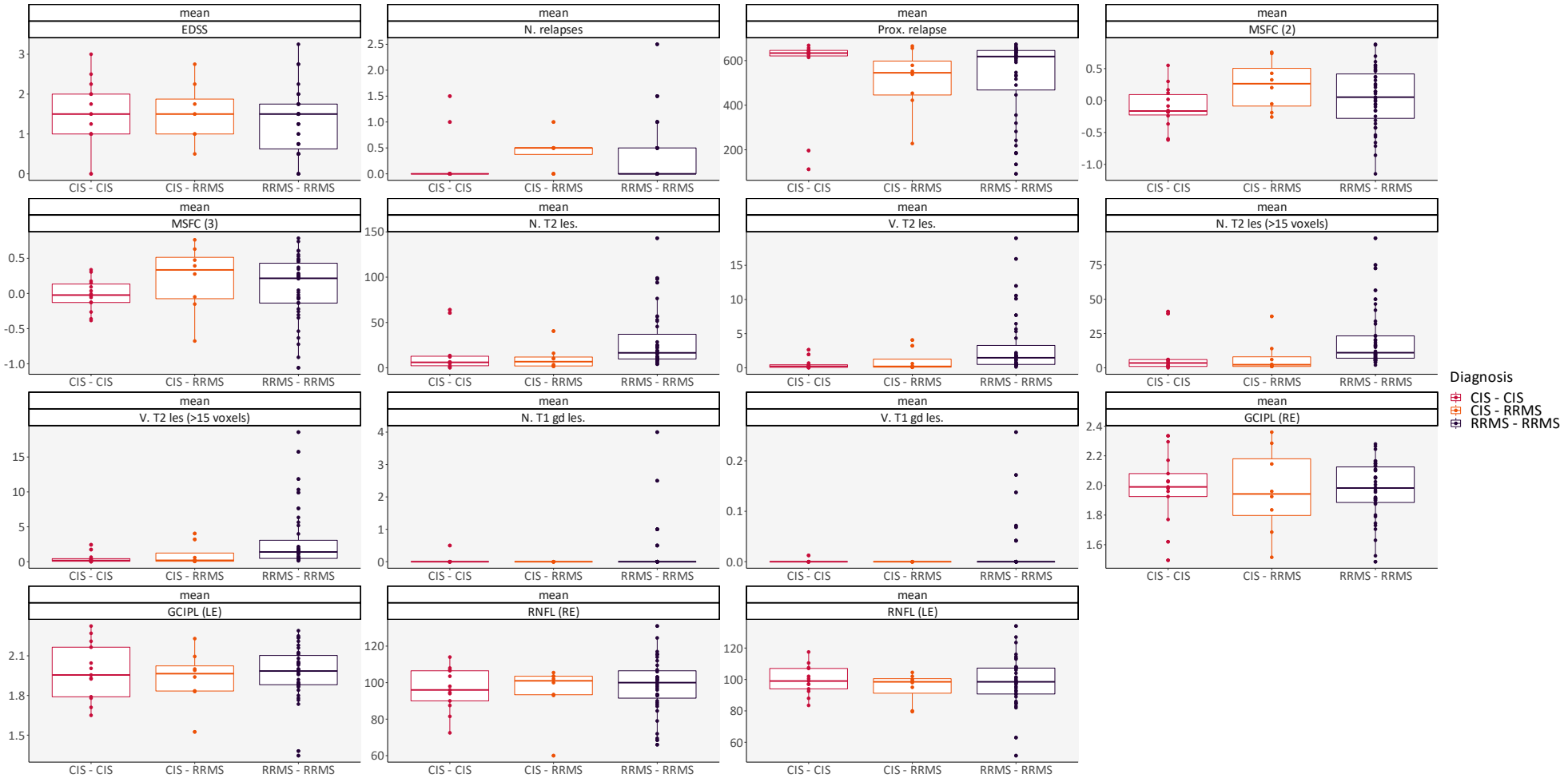


Figure S17 — Clinical variables and diagnosis change: Mean values

Distribution according to diagnosis change, from V1 to V2, regarding the mean value between V1 and V2 for EDSS, number of relapses, proximity to relapse, MSFC without PASAT, MSFC with PASAT, number of T2 hyperintense lesions, volume of T2 hyperintense lesions in mL, number of T2 hyperintense lesions (larger than 15 voxels), volume of T2 hyperintense lesions (larger than 15 voxels) in mL, number of gadolinium-enhancing T1 lesions, volume of gadolinium-enhancing T1 lesions, GCIPL volume in the right eye in mm^3, GCIPL volume in the left eye in mm^3, RNFL thickness in the right eye in µm, RNFL thickness in the left eye in µm. Abbreviations: CIS — Clinically Isolated Syndrome; EDSS — expanded disability status scale; GCIPL — ganglion cell-inner plexiform layer; gd — gadolinium; LE — left eye; les. — lesions; MSFC — Multiple Sclerosis functional composite; N. — number; PASAT — Paced Auditory Serial Addition Test; Prox. — proximity; RE — right eye; RNFL — retinal nerve fiber layer; RRMS — Relapsing-Remitting Multiple Sclerosis; V. — volume; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

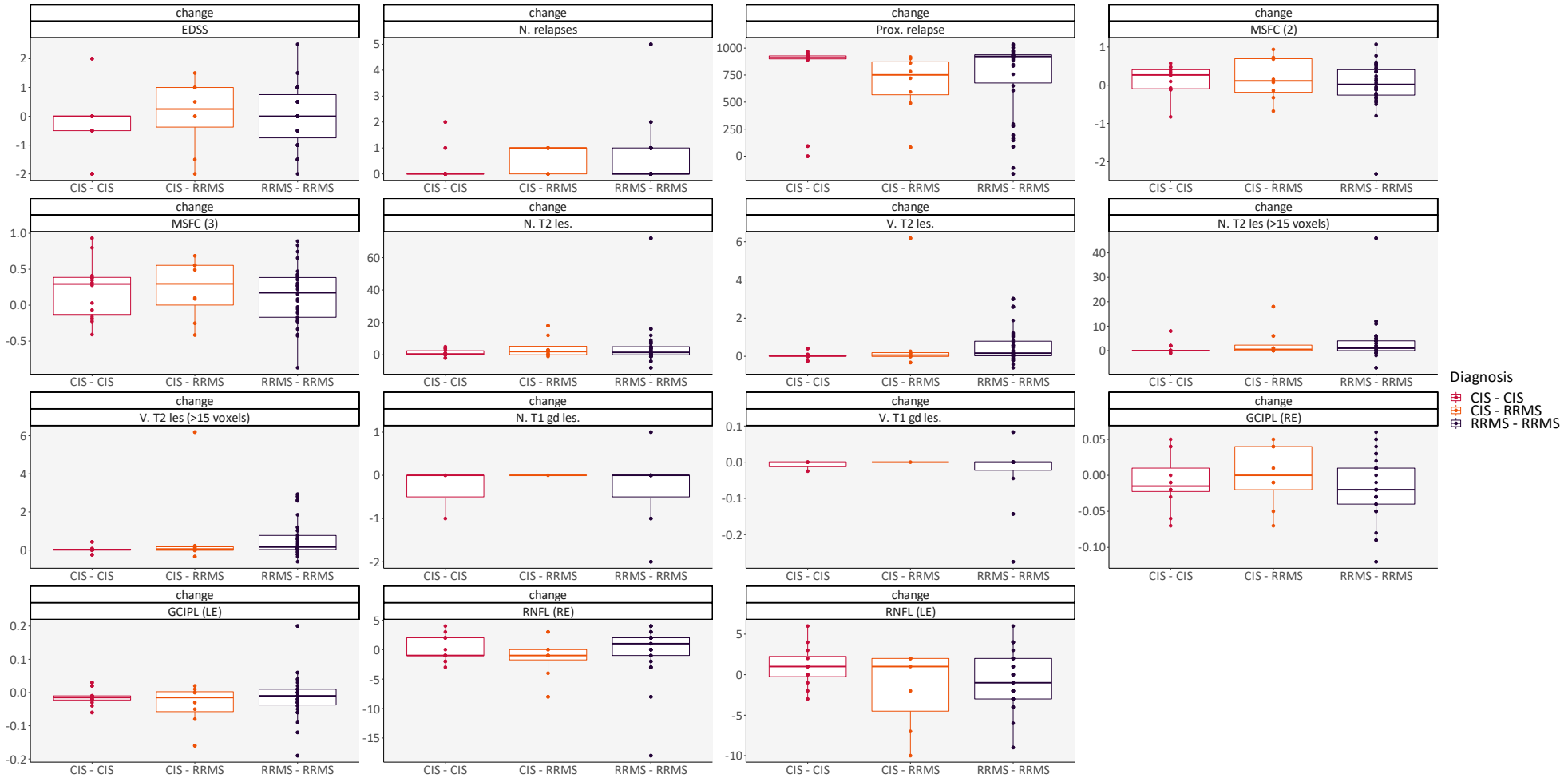


Figure S18 — Clinical variables and diagnosis change: Difference

Distribution according to diagnosis change, from V1 to V2, regarding the difference between V2 and V1 for EDSS, number of relapses, proximity to relapse, MSFC without PASAT, MSFC with PASAT, number of T2 hyperintense lesions, volume of T2 hyperintense lesions in mL, number of T2 hyperintense lesions (larger than 15 voxels), volume of T2 hyperintense lesions (larger than 15 voxels) in mL, number of gadolinium-enhancing T1 lesions, volume of gadolinium-enhancing T1 lesions, GCIPL volume in the right eye in mm^3, GCIPL volume in the left eye in mm^3, RNFL thickness in the right eye in µm, RNFL thickness in the left eye in µm. Abbreviations: CIS — Clinically Isolated Syndrome; EDSS — expanded disability status scale; GCIPL — ganglion cell-inner plexiform layer; gd — gadolinium; LE — left eye; les. — lesions; MSFC — Multiple Sclerosis functional composite; N. — number; PASAT — Paced Auditory Serial Addition Test; Prox. — proximity; RE — right eye; RNFL — retinal nerve fiber layer; RRMS — Relapsing-Remitting Multiple Sclerosis; V. — volume; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2.

Table S40 — Clinical variables and diagnosis change: Mean values *vs.* Differences. Batch Kruskal-Wallis tests, with FDR

Comparison	Mean adjusted *p*-value	Difference adjusted *p*-value
EDSS ~ Diagnosis	6.84E-01	8.26E-01
GCIPL (LE) ~ Diagnosis	9.52E-01	9.12E-01
GCIPL (RE) ~ Diagnosis	9.52E-01	8.26E-01
MSFC (2) ~ Diagnosis	4.61E-01	8.26E-01
MSFC (3) ~ Diagnosis	4.72E-01	8.82E-01
N. relapses ~ Diagnosis	2.76E-01	4.44E-01
N. T1 gd les. ~ Diagnosis	4.72E-01	9.12E-01
N. T2 les (>15 voxels) ~ Diagnosis	1.26E-03	4.44E-01
N. T2 les. ~ Diagnosis	1.20E-02	8.26E-01
Prox. relapse ~ Diagnosis	6.30E-01	4.44E-01
RNFL (LE) ~ Diagnosis	8.96E-01	7.11E-01
RNFL (RE) ~ Diagnosis	9.52E-01	7.11E-01
V. T1 gd les. ~ Diagnosis	4.72E-01	9.12E-01
V. T2 les (>15 voxels) ~ Diagnosis	1.26E-03	4.05E-01
V. T2 les. ~ Diagnosis	1.26E-03	4.05E-01

Significant *p*-values are highlighted. Abbreviations: EDSS — expanded disability status scale; FDR — false discovery rate; GCIPL — ganglion cell-inner plexiform layer; gd — gadolinium; LE — left eye; les. — lesions; MSFC — Multiple Sclerosis functional composite; N. — number; Prox. — proximity; RE — right eye; RNFL — retinal nerve fiber layer; V. — volume.

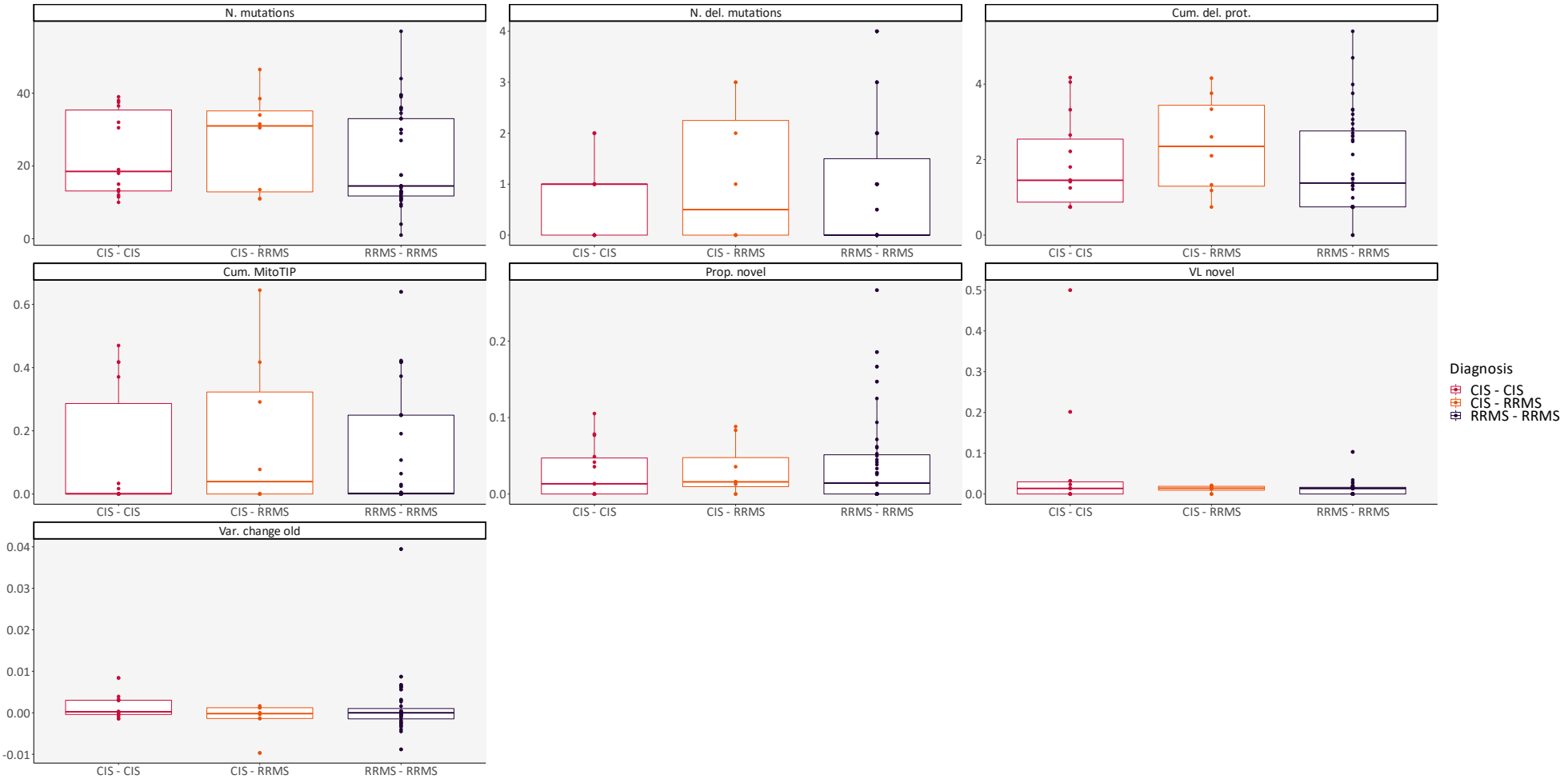


Figure S19 — NGS variables and diagnosis change

Distribution according to diagnosis change, from V1 to V2, regarding number of variants, number of deleterious variants, cumulative protein deleterious burden, cumulative MitoTIP, proportion of novel variants, mean VL of novel variants, and mean variant level change. Abbreviations: CIS — Clinically Isolated Syndrome; Cum. del. prot. — cumulative protein deleterious burden; Cum. MitoTIP — cumulative MitoTIP; del. — deleterious; N. — number; NGS — next-generation sequencing; Prop. — proportion; RRMS — Relapsing-Remitting Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2; Var. — variant; VL — variant level.

Table S41 — NGS variables and diagnosis change. Batch Kruskal-Wallis tests, with FDR

Comparison	Adjusted *p*-value
Cum. del. prot. ~ Diagnosis	9.24E-01
Cum. MitoTIP ~ Diagnosis	9.24E-01
N. del. mutations ~ Diagnosis	9.24E-01
N. mutations ~ Diagnosis	9.24E-01
Prop. novel ~ Diagnosis	9.24E-01
Var. change old ~ Diagnosis	9.24E-01
VL novel ~ Diagnosis	9.24E-01

Abbreviations: Cum. del. prot. — cumulative protein deleterious burden; Cum. MitoTIP — cumulative MitoTIP; del. — deleterious; FDR — false discovery rate; N. — number; NGS — next-generation sequencing; Prop. — proportion; Var. — variant; VL — variant level.

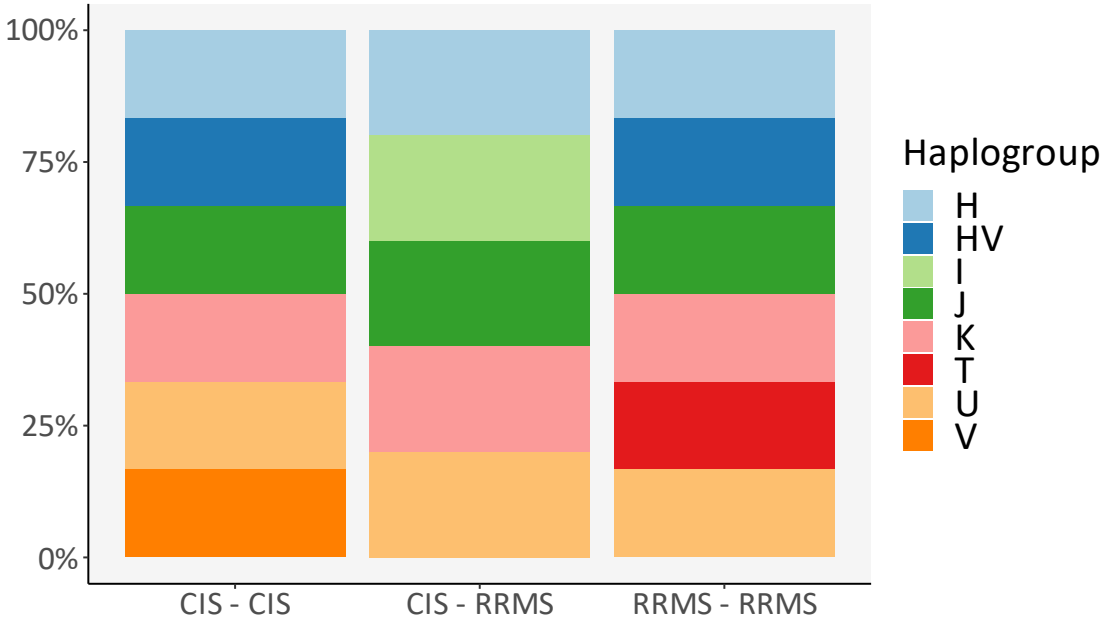


Figure S20 — Haplogroup distribution per diagnosis change

Abbreviations: CIS — Clinically Isolated Syndrome; RRMS — Relapsing-Remitting Multiple Sclerosis.

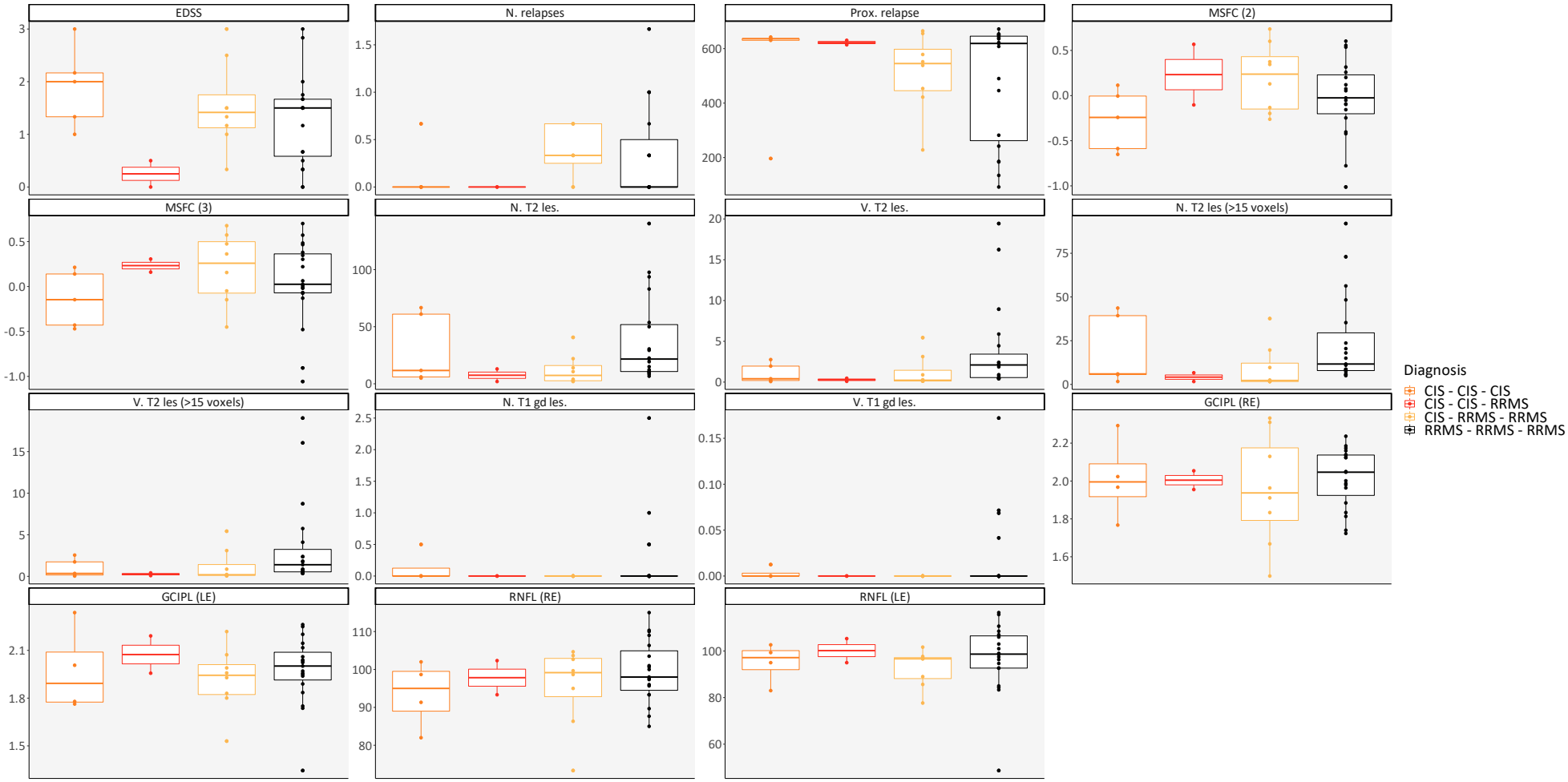


Figure S21— Clinical variables and diagnosis evolution

Distribution according to diagnosis evolution, from V1 to V3, regarding the mean value between V1, V2, and V3 for EDSS, number of relapses, proximity to relapse, MSFC without PASAT, MSFC with PASAT, number of T2 hyperintense lesions, volume of T2 hyperintense lesions in mL, number of T2 hyperintense lesions (larger than 15 voxels), volume of T2 hyperintense lesions (larger than 15 voxels) in mL, number of gadolinium-enhancing T1 lesions, volume of gadolinium-enhancing T1 lesions, GCIPL volume in the right eye in mm3, GCIPL volume in the left eye in mm3, RNFL thickness in the right eye in µm, RNFL thickness in the left eye in µm. Abbreviations: CIS — Clinically Isolated Syndrome; EDSS — expanded disability status scale; GCIPL — ganglion cell-inner plexiform layer; gd — gadolinium; LE — left eye; les. — lesions; MSFC — Multiple Sclerosis functional composite; N. — number; PASAT — Paced Auditory Serial Addition Test; Prox. — proximity; RE — right eye; RNFL — retinal nerve fiber layer; RRMS — Relapsing-Remitting Multiple Sclerosis; V. — volume; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2; V3 — patient with Multiple Sclerosis at visit 3.

Table S42 — Clinical variables and diagnosis evolution: Batch Kruskal-Wallis tests, with FDR

Comparison	Adjusted *p*-value
EDSS ~ Diagnosis	4.73E-01
GCIPL (LE) ~ Diagnosis	9.16E-01
GCIPL (RE) ~ Diagnosis	9.48E-01
MSFC (2) ~ Diagnosis	5.71E-01
MSFC (3) ~ Diagnosis	7.48E-01
N. relapses ~ Diagnosis	5.71E-01
N. T1 gd les. ~ Diagnosis	7.90E-01
N. T2 les (>15 voxels) ~ Diagnosis	3.47E-01
N. T2 les. ~ Diagnosis	3.47E-01
Prox. relapse ~ Diagnosis	9.48E-01
RNFL (LE) ~ Diagnosis	8.59E-01
RNFL (RE) ~ Diagnosis	9.16E-01
V. T1 gd les. ~ Diagnosis	7.90E-01
V. T2 les (>15 voxels) ~ Diagnosis	3.47E-01
V. T2 les. ~ Diagnosis	3.47E-01

Abbreviations: EDSS — expanded disability status scale; FDR — false discovery rate; GCIPL — ganglion cell-inner plexiform layer; gd — gadolinium; LE — left eye; les. — lesions; MSFC — Multiple Sclerosis functional composite; N. — number; Prox. — proximity; RE — right eye; RNFL — retinal nerve fiber layer; V. — volume.

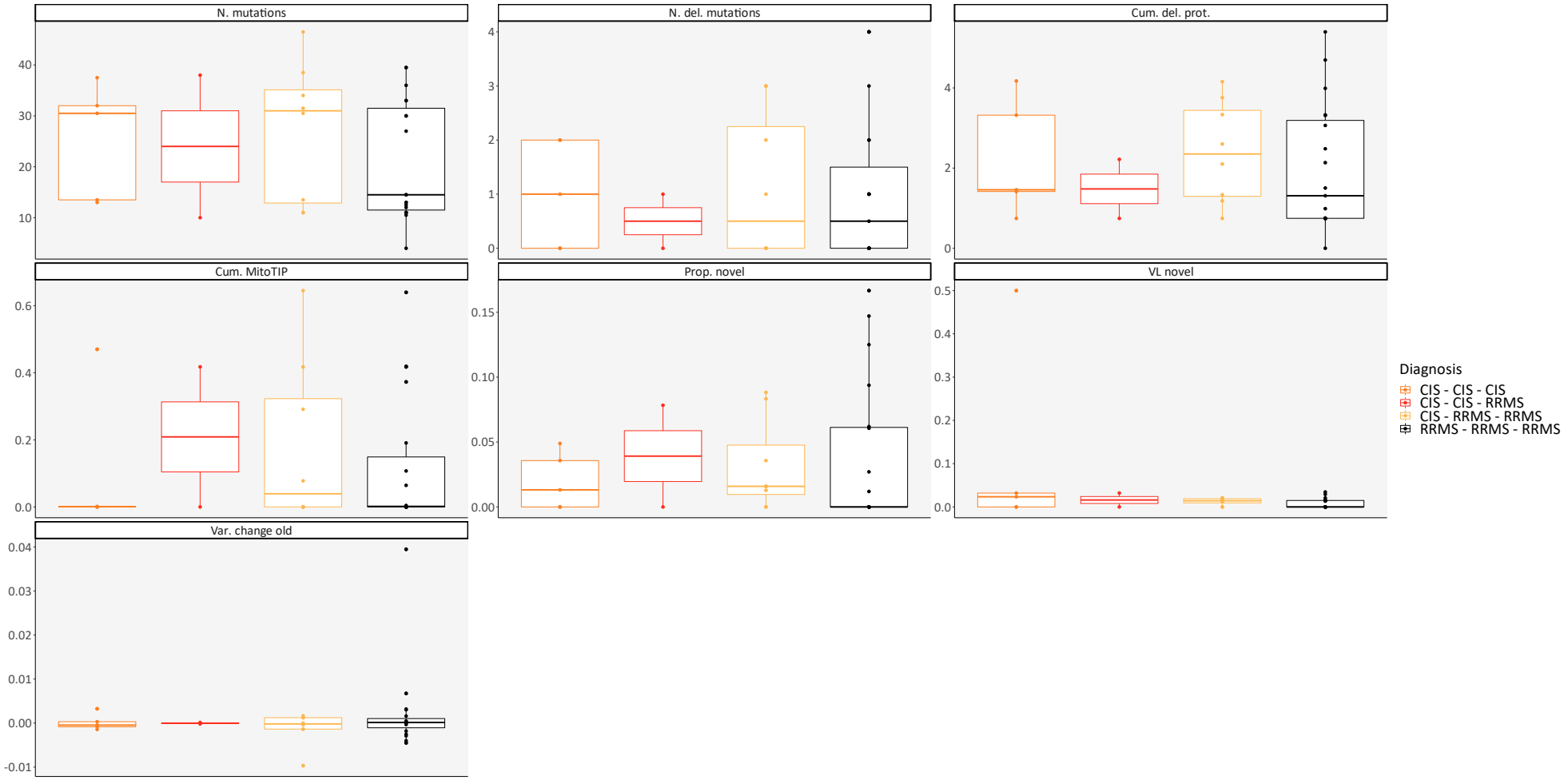


Figure S22 — NGS variables and diagnosis evolution

Distribution according to diagnosis evolution, from V1 to V3, regarding number of variants, number of deleterious variants, cumulative protein deleterious burden, cumulative MitoTIP, proportion of novel variants, mean VL of novel variants, and mean variant level change. Abbreviations: CIS — Clinically Isolated Syndrome; Cum. del. prot. — cumulative protein deleterious burden; Cum. MitoTIP — cumulative MitoTIP; del. — deleterious; N. — number; NGS — next-generation sequencing; Prop. — proportion; RRMS — Relapsing-Remitting Multiple Sclerosis; V1 — patient with Multiple Sclerosis at visit 1; V2 — patient with Multiple Sclerosis at visit 2; V3 — patient with Multiple Sclerosis at visit 3; Var. — variant; VL — variant level.

Table S43 — NGS variables and diagnosis evolution. Batch Kruskal-Wallis tests, with FDR

Comparison	Adjusted *p*-value
Cum. del. prot. ~ Diagnosis	9.70E-01
Cum. MitoTIP ~ Diagnosis	9.70E-01
N. del. mutations ~ Diagnosis	9.70E-01
N. mutations ~ Diagnosis	9.70E-01
Prop. novel ~ Diagnosis	9.70E-01
Var. change old ~ Diagnosis	9.70E-01
VL novel ~ Diagnosis	9.70E-01

Abbreviations: Cum. del. prot. — cumulative protein deleterious burden; Cum. MitoTIP — cumulative MitoTIP; del. — deleterious; FDR — false discovery rate; N. — number; NGS — next-generation sequencing; Prop. — proportion; Var. — variant; VL — variant level.

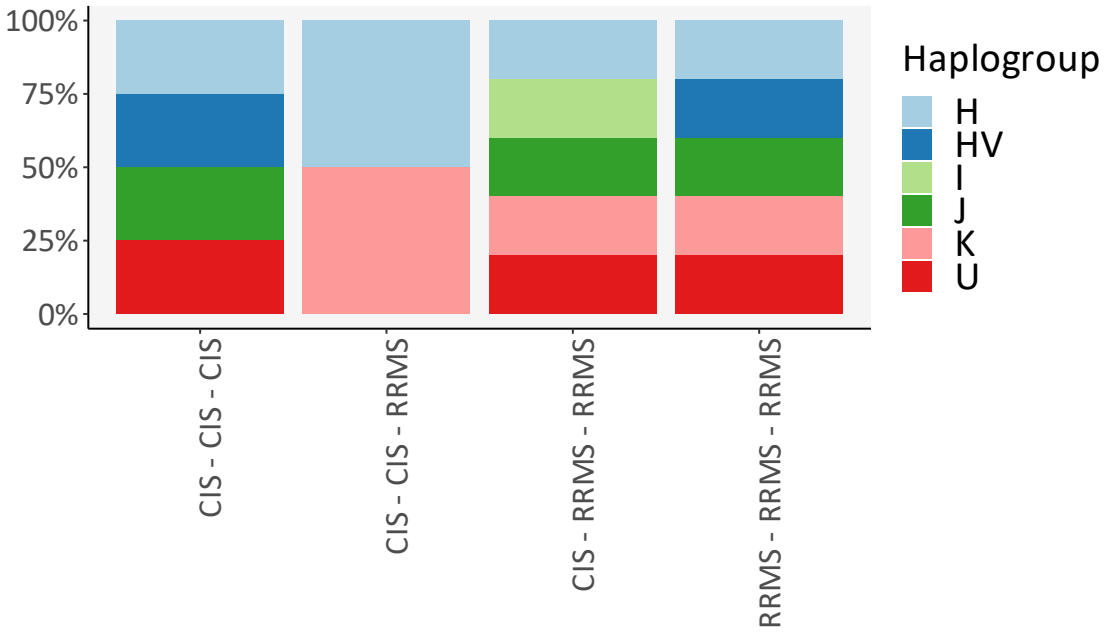


Figure S23 — Haplogroup distribution per diagnosis evolution

Abbreviations: CIS — Clinically Isolated Syndrome; RRMS — Relapsing-Remitting Multiple Sclerosis.