

UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE BIOLOGIA VEGETAL



Caracterização molecular de genes associados a ciliopatias em homens com astenozoospermia por sequenciação de nova geração

Rodrigo Nunes Mateus de Jesus Pinto

Mestrado em Biologia Molecular e Genética

Dissertação orientada por:

Doutor João Gonçalves

(Departamento de Genética Humana – Instituto Nacional de Saúde Dr. Ricardo Jorge)

Professor Doutor Francisco Pina-Martins

(Departamento de Biologia Animal – Faculdade de Ciências da Universidade de Lisboa)

Agradecimentos

O presente projeto resultou de bastante esforço, empenho, trabalho, mas também não teria sido possível sem a contribuição de diversas pessoas. Assim, a todas essas pessoas que, independentemente do impacto, contribuíram nesta fase da minha vida, os meus sinceros agradecimentos.

A todo o Instituto Nacional de Saúde Doutor Ricardo Jorge e, em particular, ao Departamento de Genética Humana, onde tive a oportunidade de desenvolver este projeto. E claro, a todas as pessoas com quem tive contacto durante o tempo que lá passei. Senti-me bem-vindo.

A todos os membros da Unidade de Genética Molecular por me terem acolhido tão bem, e estarem sempre disponíveis para me ensinar e ajudar. Pude inclusive ver e aprender com diferentes pessoas, constatar as diferenças na forma como cada um realiza certos procedimentos laboratoriais, e adaptar a minha forma de trabalhar, desenvolver o meu método. Mas tenho de destacar o apoio que tive do Pedro Rodrigues, pelo tempo que passou comigo para eu aprender as técnicas laboratoriais e sistemas bioinformáticos necessárias para a realização de todo este projeto, e o apoio da Íris Caetano, sempre disponível e uma excelente companhia. Ensinarão-me muito, e sempre com a vossa boa disposição.

Em particular, um enorme agradecimento ao Doutor João Gonçalves pela oportunidade de realizar este projeto, por me ter recebido no seu laboratório, e pela disponibilidade para me ensinar e dar conselhos. Especialmente no início, quando tinha resultados e queria perceber se estavam bem, ou o que poderia ter causado algum erro, ou como prosseguir. Aprendi com quem sabe.

Ao Doutor Luís Vieira e à Unidade de Tecnologia e Inovação, responsáveis pela execução de ensaios experimentais que estiveram na base deste projeto. Para além disso, tenho de agradecer à Doutora Sílvia Duarte e à Doutora Joana Mendonça por terem tido a amabilidade de me permitir acompanhar o seu trabalho e, assim, compreender melhor todos os passos das reações de sequenciação realizadas. Por mais conhecimento teórico que se tenha, trabalho laboratorial faz toda a diferença.

Quero ainda agradecer ao Professor Francisco Pina Martins, da Faculdade de Ciências da Universidade de Lisboa, não só por ter aceite ser meu orientador, mas também pelo seu envolvimento neste projeto. Pude sempre contar com a sua ajuda nesta área na qual não estou tão confortável, e aprendi bastante ao usar este *script* que desenvolveu.

E por fim, não posso deixar de agradecer aos meus amigos mais próximos, à minha família, a quem estive comigo nesta altura. Não apenas nesta etapa, mas em todas pelas quais já passei, e pelo que o futuro me reserva. Tornam os bons momentos ainda melhores, e são fundamentais nos tempos difíceis.

Sumário

As ciliopatias são doenças hereditárias clinicamente e geneticamente heterogêneas, caracterizadas por disfunção ciliar. A patologia discinesia ciliar primária (PCD, *primary ciliary dyskinesia*) resulta de defeitos em cílios móveis, e pode apresentar manifestações como depuração mucociliar defetiva, *situs inversus* e reduzida motilidade de espermatozoides (astenozoospermia). Esta rara ciliopatia, de difícil diagnóstico, está associada a mais de 45 genes. Como tal, a validação da caracterização molecular de múltiplos genes por sequenciação de nova geração (NGS, *next-generation sequencing*) é essencial para o diagnóstico e investigação desta doença. Simultaneamente, procurou-se avaliar a associação entre astenozoospermia e 8 genes tradicionalmente associados a PCD.

Os genes *DNAAF2*, *DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *RSPH9* e *RSPH4A* (223 exões) foram analisados em 48 homens diagnosticados com astenozoospermia, utilizando a plataforma de NGS MiSeq e a metodologia Ampliseq (Illumina). Recorreu-se à sequenciação de Sanger para a avaliação de regiões com cobertura vertical inferior a 20x. A anotação de variantes incluiu as ferramentas VarAFT, Variant Interpreter, HGMD, Alamut Visual-Plus, VarSome, ClinVar, HSF e IGV.

Foram identificadas 57 variantes, em 47 dos 48 doentes, incluindo 9 de significado clínico desconhecido (VUS, *variant of unknown significance*), 2 provavelmente patogénicas e 1 patogénica. Destacam-se as VUS *DNAH11*:c.2909A>G e *DNAH11*:c.9935A>T, possivelmente presentes em *trans* (heterozigotia composta) num homem com oligoastenozoospermia, e *DNAH5*:c.3614C>T e *DNAAF2*:c.2239G>A, uma dupla heterozigotia detetada num doente com oligoastenoteratozoospermia. Adicionalmente, as variantes *DNAH11*:c.8990G>A e *DNAH11*:c.10739G>A, identificadas em 2 homens oligoastenozoospermicos e classificadas como benignas, apresentam classificação oposta na literatura disponível (patogénicas). Contudo, não tendo sido possível estabelecer uma correlação causal direta entre o genótipo e o fenótipo de cada doente, estima-se que os genes analisados não constituam a principal base molecular de astenozoospermia isolada. Ainda assim, serão necessários estudos funcionais e familiares futuros para se poder classificar a patogenicidade destas variantes e compreender melhor o seu impacto na motilidade dos espermatozoides.

Palavras chave: ciliopatias, discinesia ciliar primária, infertilidade masculina, astenozoospermia, sequenciação de nova geração.

Abstract

Ciliopathies are clinically and genetically heterogeneous hereditary diseases, characterized by ciliary dysfunction. Primary ciliary dyskinesia (PCD), in particular, results from defects in motile cilia and, therefore, causes defective muco-ciliary clearance, *situs inversus* and reduced spermatozoa motility (asthenozoospermia). This rare ciliopathy with a challenging diagnosis is currently associated with over 45 genes. As such, validation of next-generation sequencing based panels to identify genetic variants is essential for the diagnosis and research of this disorder. Simultaneously, this study aimed to evaluate the association between asthenozoospermia and a panel of 8 genes related to PCD.

The genes *DNAAF2*, *DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *RSPH9* and *RSPH4A* (223 exons) were analyzed in a group of 48 asthenozoospermic men, using Illumina's MiSeq platform and Ampliseq methodology. Additionally, Sanger sequencing was employed for the evaluation of regions of interest with depth coverage below 20x. Variants annotation involved several bioinformatics tools, including VarAFT, Variant Interpreter, HGMD, Alamut Visual Plus, VarSome, ClinVar, HSF and IGV.

57 variants were identified in 47 out of the 48 patients, including 1 pathogenic, 2 likely pathogenic, and 9 variants of uncertain significance (VUS). Some VUS are of particular interest, namely *DNAH11*:c.2909A>G and *DNAH11*:c.9935A>T, probably present in trans (compound heterozygosity) in an oligoasthenozoospermic man, and *DNAH5*:c.3614C>T and *DNAAF2*:c.2239G>A, a double heterozygosity detected in a patient with oligoasthenoteratozoospermia. Furthermore, the variants *DNAH11*:c.8990G>A and *DNAH11*:c.10739G>A, identified in 2 oligoasthenozoospermic men and described as benign, display contradictory information in the available literature (pathogenic). However, the impossibility of establishing a direct correlation between the genotype and the phenotype of each patient, led us to infer that the 8 genes analyzed are unlikely to constitute the principal molecular base of isolated asthenozoospermia. Even so, future functional and family studies are required in order to properly classify these variants and better understand their impact on spermatozoa motility.

Keywords: ciliopathies, primary ciliary dyskinesia, male infertility, asthenozoospermia, next-generation sequencing.

Índice

Agradecimentos	I
Sumário	II
Abstract	III
Lista de Figuras	VI
Lista de Tabelas	VI
Lista de Abreviaturas, Siglas e Símbolos	VIII
1 Introdução	1
1.1. Ciliopatias e cílios	1
1.1.1. Cílios primários, e características comuns com os cílios móveis.....	1
1.1.2. Cílios móveis	3
1.2. Discinesia ciliar primária.....	4
1.2.1. Etiologia genética da discinesia ciliar primária.....	4
1.3. Infertilidade masculina e astenozoospermia	6
1.4. Objetivos	8
2 Materiais e Métodos	9
2.1. Amostras biológicas.....	9
2.2. Extração de DNA genómico.....	9
2.3. Desenho do painel genético.....	10
2.4. Quantificação de DNA	10
2.5. Sequenciação de nova geração	10
2.5.1. Controlo de qualidade dos ensaios de NGS	11
2.6. Avaliação <i>in silico</i> da cobertura	11
2.7. Reação em cadeia da polimerase, ou PCR	13
2.8. Sequenciação de Sanger.....	15
2.9. Análise <i>in silico</i> de variantes.....	16
2.9.1. Filtragem de variantes.....	16
2.9.2. Anotação e classificação de variantes	17
3 Resultados	18
3.1. Sequenciação de nova geração	18
3.1.1. Quantificação de DNA	18
3.1.2. Controlo de qualidade dos ensaios de NGS	18
3.2. Identificação e sequenciação das regiões não cobertas por NGS.....	19
3.2.1. Avaliação <i>in silico</i> da cobertura.....	19
3.2.2. PCR e sequenciação de Sanger	21

3.3. Anotação e avaliação biopatológica das variantes detetadas	22
4 Discussão	26
Conclusão	29
5 Referências Bibliográficas.....	30
6 Anexos	37

Lista de Figuras

Figura 1.1 A diversidade de manifestações resultantes de disfunção ciliar	1
Figura 1.2 A estrutura ciliar	2
Figura 2.1 Exemplo representativo da diferente visualização de resultados nos <i>software</i> VarAFT e IGV	12
Figura 3.1 Visualização da concentração das amostras PCip32 a PCip41 em gel de agarose	18
Figura 3.2 Representação esquemática das regiões sem cobertura suficiente (20x) no transcrito do gene <i>DNAAF2</i>	20
Figura 3.3 Representação esquemática das regiões sem cobertura suficiente (20x) no transcrito do gene <i>RSPH4A</i>	20
Figura 3.4 Visualização do resultado das amplificações de fragmentos de <i>NME8_Ex5</i> e <i>RSPH9_Ex2</i> , em gel de agarose	21
Figura 3.5 Excertos de um cromatograma exemplificativo dos resultados obtidos após sequenciação de DNA pelo método de Sanger	22
Figura 3.6 Distribuição das 57 variantes em estudo por gene e de acordo com a sua consequência/localização no mesmo	24

Lista de Tabelas

Tabela 1.1 Painel de genes em estudo	5
Tabela 1.2 Exemplos de formas de infertilidade masculina baseadas em alterações do espermograma	7
Tabela 2.1 Doentes que compõem o grupo de 48 homens inférteis, distribuídos de acordo com os ensaios de sequenciação de nova geração em que as amostras foram incluídas, e classificação do respetivo espermograma	9
Tabela 2.2 Iniciadores desenhados para a sequenciação das regiões alvo específicas	14
Tabela 2.3 Reagentes na reação de PCR e respetivas concentrações	15
Tabela 2.4 Programa da PCR	15
Tabela 2.5 Reagentes na reação de sequenciação cíclica de Sanger e respetivas concentrações	16
Tabela 2.6 Ciclo térmico das reações de amplificação	16
Tabela 3.1 Percentagem média, por ensaio, de nucleótidos que atingiram um <i>Q score</i> de 30	19
Tabela 3.2 Regiões alvo sem cobertura suficiente (20x), nos oito genes em estudo, identificadas para posterior amplificação enzimática e sequenciação pelo método de Sanger	20
Tabela 3.3 Regiões sem cobertura suficiente (20x), identificadas pelo <i>script manifest_parser.py</i> , mas não pelo <i>software</i> VarAFT	21

Tabela 3.4 Variantes identificadas na análise aos resultados obtidos após sequenciação de DNA pelo método de Sanger	23
Tabela 3.5 Variantes genéticas de interesse, identificadas no grupo de 48 indivíduos do sexo masculino inférteis	25
Tabela S1 Regiões sem cobertura suficiente (20x), nos oito genes em estudo	37
Tabela S2 Variantes genéticas (n=45) identificadas no grupo de 48 indivíduos do sexo masculino inférteis, classificadas como benignas ou provavelmente benignas	39

Lista de Abreviaturas, Siglas e Símbolos

5' UTR	Região 5' não traduzida (<i>5' untranslated region</i>)
A	Adenina
AB	<i>Applied Biosystems</i>
ACMG-AMP	<i>American College of Medical Genetics and Genomics - Association for Molecular Pathology</i>
Ala	Alanina
Align GVD	<i>Align Grantham variation Grantham deviation</i>
Arg	Arginina
Asn	Asparagina
Asp	Ácido aspártico
ATP	Adenosina trifosfato (<i>adenosine triphosphate</i>)
Atz	Astenozoospermia
B	Benigna
BAM	<i>Binary sequence Alignment/Map</i>
BED	<i>Browser Extensible Data</i>
BWA	<i>Burrows-Wheeler Aligner</i>
C	Citosina
°C	Grau Celsius
CCDC39	<i>Coiled-coil domain 39 molecular ruler complex subunit</i>
CCDC40	<i>Coiled-coil domain 40 molecular ruler complex subunit</i>
cDNA	DNA complementar (<i>complementary DNA</i>)
CFAP43	<i>Cilia and flagella associated protein 43</i>
CFAP44	<i>Cilia and flagella associated protein 44</i>
Cys	Cisteína
ddH₂O	Água bidestilada
ddNTP	Didesoxirribonucleótido trifosfato
DMSO	Dimetilsulfóxido
DNA	Ácido desoxirribonucleico (<i>deoxyribonucleic acid</i>)
DNAAF2	<i>Dynein axonemal assembly factor 2</i>
DNAH1	<i>Dynein axonemal heavy chain 1</i>
DNAH5	<i>Dynein axonemal heavy chain 5</i>
DNAH8	<i>Dynein axonemal heavy chain 8</i>
DNAH9	<i>Dynein axonemal heavy chain 9</i>
DNAH11	<i>Dynein axonemal heavy chain 11</i>
DNAH17	<i>Dynein axonemal heavy chain 17</i>
DNAI1	<i>Dynein axonemal intermediate chain 1</i>
DNAI2	<i>Dynein axonemal intermediate chain 2</i>

dNTP	Desoxirribonucleótido trifosfato
dsDNA	<i>Double-stranded DNA</i>
ESE	<i>Exonic splicing enhancer</i>
esp	Espermatozoides
ESS	<i>Exonic splicing silencer</i>
Ex	Exão
F	<i>Forward</i>
G	Guanina
GATK	<i>Genome Analysis Toolkit</i>
gDNA	DNA genómico (<i>genomic DNA</i>)
Gln	Glutamina
Glu	Ácido glutâmico
Gly	Glicina
GnomAD	<i>Genome Aggregation Database</i>
GRCh37	<i>Genome Reference Consortium Human Build 37</i>
HG19	<i>Human Genome version 19</i>
HGMD	<i>Human Gene Mutation Database</i>
HGVS	<i>Human Genome Variation Society</i>
His	Histidina
HSF	<i>Human Splicing Finder</i>
ICSI	Microinjeção intracitoplasmática de espermatozoides (<i>intracytoplasmic sperm injection</i>)
ID	Identificação da amostra/doente
IDA	Braço de dineína interno (<i>inner dynein arm</i>)
IFT	Transporte intraflagelar (<i>intraflagellar transport</i>)
IGV	<i>Integrative Genomics Viewer</i>
Ile	Isoleucina
IVF	Fertilização <i>in vitro</i> (<i>in vitro fertilization</i>)
LECA	<i>Last eukaryotic common ancestor</i>
Leu	Leucina
Lys	Lisina
MANE	<i>Matched Annotation from NCBI and EMBL-EBI</i>
MaxEntScan	<i>Maximum Entropy Scan</i>
Met	Metionina
mg	Miligrama
min	Minuto
ml	Mililitro
mM	Milimolar

MMAF	Anormalidades morfológicas múltiplas do flagelo (<i>multiple morphological abnormalities of the flagella</i>)
MN	Morfologia normal
MP	Motilidade progressiva
mRNA	RNA mensageiro (<i>messenger RNA</i>)
MTOC	Centro organizador de microtúbulos (<i>microtubule-organizing center</i>)
NCBI	<i>National Center for Biotechnology Information</i>
ng	Nanograma
NGS	Sequenciação de nova geração (<i>next-generation sequencing</i>)
NHLBI GO ESP	<i>National Heart Lung and Blood Institute Grand Opportunity Exome Sequencing Project</i>
NME8	<i>NME/NM23 family member 8</i>
NME/NM	<i>Non-metastatic</i>
NNSPLICE	<i>Neural Network SPLICE</i>
NPHP	<i>Nephronophthisis</i>
OAz	Oligoastenozoospermia
ODA	Braço de dineína externo (<i>outer dynein arm</i>)
OMIM	<i>Online Mendelian Inheritance in Man</i>
OTA	Oligoteratoastenozoospermia
p	Braço curto
P	Probabilidade de uma base ser identificada incorretamente
P	Patogénica
pb	Par de bases
PB	Provavelmente benigna
PCD	Discinesia ciliar primária (<i>primary ciliary dyskinesia</i>)
PCR	Reação em cadeia da polimerase (<i>polymerase chain reaction</i>)
Phe	Fenilalanina
PKD	<i>Polycystic kidney disease</i>
pmol	Picomol
PolyPhen-2	<i>Polymorphism Phenotyping v2</i>
PP	Provavelmente patogénica
Pré-mRNA	mRNA precursor (<i>precursor mRNA</i>)
Pro	Prolina
p/v	Peso por volume
q	Braço longo
Q	<i>Phred quality score</i>
R	<i>Reverse</i>
RNA	Ácido ribonucleico (<i>ribonucleic acid</i>)

RSPH4A	<i>Radial spoke head component 4A</i>
RSPH6A	<i>Radial spoke head 6 homolog A</i>
RSPH9	<i>Radial spoke head component 9</i>
seg	Segundo
Ser	Serina
SIFT	<i>Sorting Intolerant From Tolerant</i>
SNV	<i>Single nucleotide variant</i>
ssDNA	<i>Single-stranded DNA</i>
stdev	Desvio padrão (<i>standard deviation</i>)
T	Timina
TBE	Tris-Borato-EDTA
Te	Temperatura de emparelhamento
TEM	Microscopia eletrônica de transmissão (<i>transmission electron microscopy</i>)
Temp	Temperatura
Ter	Codão de terminação
Thr	Treonina
TOPMed	<i>Trans-Omics for Precision Medicine</i>
Tyr	Tirosina
U	Unidade
Val	Valina
VarAFT	<i>Variant Annotation and Filter Tool</i>
VCF	<i>Variant Call Format</i>
VUS	Variante de significado clínico desconhecido (<i>variant of unknown significance</i>)
WHO	Organização Mundial de Saúde (<i>World Health Organization</i>)
µl	Microlitro

1 | Introdução

1.1. Ciliopatias e cílios

Ciliopatias representam um vasto conjunto de doenças hereditárias, clinicamente e geneticamente heterogêneas, causadas por disfunção ciliar ^{1,2}. Na base destas patologias encontram-se os cílios, complexos organelos ancorados à membrana celular, conservados ao longo da evolução, que divergiram desde o *Last eukaryotic common ancestor* (LECA) por uma enorme diversidade de espécies ^{3,4}. Esta conservação entre espécies filogeneticamente distantes permite que muito do conhecimento que se tem sobre a estrutura e funcionamento destes organelos provenha de organismos como a alga verde unicelular *Chlamydomonas reinhardtii* ³⁻⁵. Contudo, é nos vertebrados que os cílios se encontram mais difundidos, sendo que no ser humano, em particular, estão presentes em quase todos os tipos de células ³. Como tal, defeitos nestes organelos podem causar uma enorme variedade de sintomas distribuídos por todo o organismo (Figura 1.1), o que na prática se reflete tanto em ciliopatias com manifestações isoladas, como em doenças multi-sistémicas com manifestações comuns a outras ciliopatias, o que torna estas patologias particularmente complexas de analisar ¹⁻³.

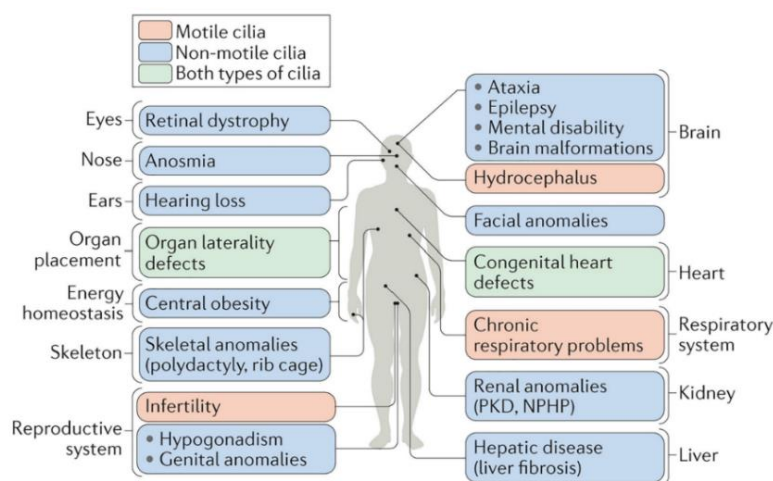


Figura 1.1 A diversidade de manifestações resultantes de disfunção ciliar. Figura representativa dos diferentes órgãos e tecidos afetados por ciliopatias, as principais manifestações, e a classe de cílios correspondente. PKD *polycystic kidney disease*, NPHP *nephronophthisis*. Retirado de Reiter e Leroux, 2017 ⁶.

Os cílios podem ser divididos em dois grupos: cílios primários, por vezes identificados como imóveis ou sensoriais, e cílios móveis, geradores de movimento de fluidos ou da motilidade de espermatozoides ⁶. Para uma melhor compreensão das suas diferenças, e da sua relevância para este projeto, é necessário considerar as funções e a ultraestrutura destes organelos.

1.1.1. Cílios primários, e características comuns com os cílios móveis

Relativamente à sua estrutura, cílios primários e cílios móveis apresentam diversas semelhanças (Figura 1.2.e). Na superfície apical celular encontra-se ancorado o corpo basal, uma estrutura de 9 tripletos de microtúbulos de γ -tubulina, derivada de um centríolo da célula mãe, e conectada à membrana ciliar por fibras de transição (também conhecidas como apêndices distais) e apêndices subdistais ^{7,8}. Logo após o corpo basal encontra-se a zona de transição, que inclui estruturas *Y-shaped* conectadas à membrana ciliar de modo a separar o compartimento ciliar, permitindo ao cílio estabelecer a composição proteica

necessária para as suas funções ⁸. É nesta zona de transição que a partir do corpo basal se desenvolve o axonema, uma estrutura central composta por 9 pares de α e β -tubulina (Figura 1.2.g), à qual está associado o transporte intraflagelar (IFT, *intraflagellar transport*) ⁹. Este complexo mecanismo de transporte bidirecional é responsável pelo movimento de proteínas e outros compostos da célula para a ponta do cílio e vice-versa, através de partículas de IFT, ou *IFT-trains* (i.e., complexos multiproteicos associados aos microtúbulos do axonema e deslocados ao longo destes por proteínas motoras kinesina-2, para movimento anterógrado, e dineína 2, para movimento retrógrado) ^{5,9}.

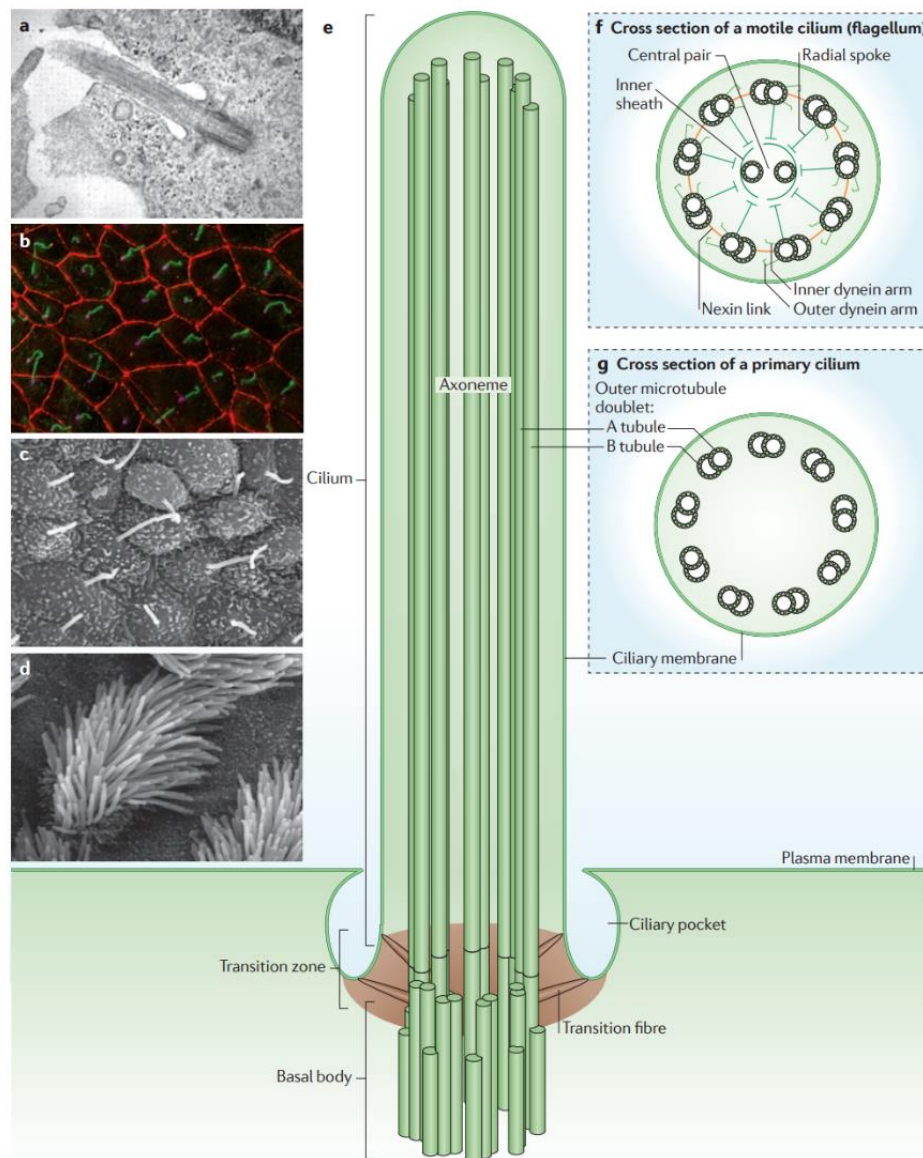


Figura 1.2 A estrutura ciliar. (a) Microscopia eletrônica de transmissão de um cílio primário de células do epitélio pigmentar da retina. (b) Visualização por imunofluorescência de cílios primários de células do ducto coletor medular interno. (c,d) Microscopia eletrônica de varrimento de cílios nodais e cílios móveis da traqueia de ratos, respetivamente. (e) Estrutura esquemática de um cílio primário. (f,g) Representação transversal da estrutura de um cílio móvel e de um cílio primário, respetivamente. Retirado de Ishikawa e Marshall, 2011 ⁹.

O corpo basal, um dos componentes dos cílios já enunciado, deriva do centríolo-mãe, o qual, juntamente com o centríolo-filho associado, faz parte do centrossoma, um organelo regulador da divisão celular e definido como o principal centro organizador de microtúbulos (MTOC, *microtubule-organizing center*) ⁷.

Assim, para que ocorra a divisão celular é necessário que a célula reabsorva o cílio previamente à mitose, e só após a citocinese se volta a formar um novo cílio a partir do centrossoma presente em cada célula, num processo denominado ciliogénese⁹⁻¹¹. Neste processo, o centríolo-mãe adquire apêndices distais e subdistais que lhe irão permitir ancorar-se à membrana celular, e é a partir de microtúbulos do já denominado corpo basal que se inicia a formação do axonema^{9,10}. A posterior extensão desta estrutura depende da importação de proteínas citosólicas transportadas até à ponta do cílio pelo mecanismo IFT, que mesmo após a extensão do axonema mantém um contínuo *turnover* de tubulina^{9,10}. Células multiciliadas (Figura 1.2.d) requerem, contudo, uma multiplicação inicial para formação de até 300 centríolos a partir do centrossoma inicial¹².

Cílios primários estão presentes em quase todas as células quiescentes humanas, um cílio por célula (Figura 1.2.b). São responsáveis pela perceção de estímulos extracelulares através de recetores transmembranares e pela coordenação da resposta celular, ao estimularem ou reprimirem as vias de sinalização correspondentes, sendo exemplos as vias de sinalização Hedgehog e Wnt, com impacto ao nível da proliferação, diferenciação e manutenção celular, entre outros¹³. Consequentemente, qualquer alteração que prejudique as funções sensoriais e de sinalização destes organelos difundidos por todo o corpo humano poderá dar origem a uma impressionante diversidade de manifestações nos mais diversos órgãos e tecidos, conforme ilustrado na Figura 1.1.

1.1.2. Cílios móveis

Embora sejam estruturalmente semelhantes aos cílios primários, cílios móveis apresentam algumas estruturas adicionais ao nível do axonema (Figura 1.2.f), responsáveis pela motilidade que caracteriza estes organelos. Para além dos nove pares de microtúbulos que conferem a conformação 9+0 aos primeiros, os cílios móveis incluem um par central de microtúbulos (conformação 9+2), raios radiais entre este e os restantes pares, e ligações elásticas, denominadas de complexos regulatórios de nexina-dineína, a conectar pares de microtúbulos adjacentes^{14,15}. Atualmente, considera-se que estas três estruturas estejam envolvidas na regulação do padrão do movimento ciliar, ao coordenarem a atividade dos braços de dineína¹⁶⁻¹⁸. Os braços de dineína são proteínas motoras, associados aos nove pares de microtúbulos, responsáveis pela regulação do batimento do cílio no caso dos braços de dineína internos (IDA, ou *inner dynein arm*), e capazes de gerar a força necessária para este movimento com gasto de adenosina trifosfato (ATP, *adenosine triphosphate*) no caso dos braços de dineína externos (ODA, ou *outer dynein arm*)^{16,19}.

Os cílios móveis podem atingir dezenas ou até centenas de organelos por célula (Figura 1.2.d), e tendem a estar presentes num grupo mais restrito de células e tecidos, nomeadamente: i) no epitélio respiratório, onde promovem a depuração mucociliar (i.e., eliminação de agentes nocivos inalados e presentes nas vias respiratórias através do fluxo de muco presente nas mesmas); ii) nas células do epêndima que revestem os ventrículos cerebrais, nos quais circula o líquido cefalorraquidiano; iii) no epitélio das trompas de Falópio, para ajudar no movimento do óvulo do ovário para o útero; iv) em células nodais da gástrula, durante o desenvolvimento embrionário, que apresentam um único cílio móvel (9+0) por célula, desprovido do par central de microtúbulos e de raios radiais, o que lhes confere um movimento rotatório, ainda assim capaz de criar um fluxo do fluido extracelular (fluxo nodal) implicado na determinação do eixo corporal direita-esquerda; e v) em espermatozoides, que possuem o flagelo, um cílio móvel especializado no qual o axonema (9+2) se encontra envolto por estruturas acessórias (bainha mitocondrial, bainha fibrosa, e fibras densas externas, que contribuem para, por exemplo, a rigidez do organelo e o fornecimento de energia), e que difere na sua função, ao conferir motilidade à célula²⁰⁻²⁴.

Defeitos em cílios móveis podem afetar vários órgãos (Figura 1.1), mas estão associados essencialmente a uma ciliopatia, a discinesia ciliar primária.

1.2. Discinesia ciliar primária

Discinesia ciliar primária (PCD, *primary ciliary dyskinesia*) é uma ciliopatia resultante de defeitos em cílios móveis, associada primariamente a incapacidade de remoção de muco nas vias respiratórias ²⁵. Entre as manifestações clínicas inclui-se infecções respiratórias recorrentes, com tendência a desenvolver bronquiectasias, rinite e sinusite crônicas, hidrocefalia, infertilidade feminina, infertilidade masculina, *situs inversus* (presente em cerca de 50% dos casos de PCD, nos quais a doença passa a ser denominada por síndrome de Kartagener), entre outras ^{25,26}.

Esta doença rara e pouco conhecida, cuja prevalência é difícil de determinar, tem uma incidência estimada de 1/15 000 a 1/30 000 nascimentos, mas que poderá estar subestimada devido ao seu difícil diagnóstico ²⁷. Por norma PCD expressa-se desde o nascimento, podendo estar associada a insuficiência respiratória neonatal sem explicação, bem como tosse produtiva crônica e congestão nasal crônica com início nos primeiros meses de vida, para além da existência de *situs inversus totalis* ou outras deficiências de lateralidade ²⁸. Contudo, a falta de conhecimento em relação à doença por parte de pediatras, a sobreposição de sintomas com outras doenças, a ausência de um teste diagnóstico único, e a consequente necessidade de um conjunto de testes especializados e nem sempre disponíveis – como a medição do óxido nítrico nasal, cujos valores diminuem com a doença, a microscopia eletrônica de transmissão (TEM, *transmission electron microscopy*), para identificação de defeitos na ultraestrutura ciliar, a microscopia de vídeo de alta velocidade, que permite avaliar a frequência e padrão do movimento ciliar, e testes genéticos –, dificultam o atempado diagnóstico desta patologia, essencial para um melhor acompanhamento, prevenção de danos pulmonares graves e irreversíveis, e fundamentalmente melhor qualidade de vida ²⁹⁻³¹.

Atualmente não existem terapias capazes de recuperar a função ciliar em indivíduos afetados por PCD. Como tal, cada caso deve ser monitorizado e gerido consoante as manifestações presentes, por norma com medidas para melhorar a eliminação do muco e prevenir e tratar infecções nas vias respiratórias superiores e inferiores ^{25,32}. Contudo, com exceção do antibiótico azitromicina, associado a uma redução significativa de exacerbações respiratórias em pacientes com PCD, a ausência de ensaios clínicos obriga a que a abordagem definida para estes doentes seja baseada no tratamento de doenças com sintomas semelhantes, como a fibrose quística ^{32,33}. De futuro, o tratamento de PCD poderá passar por edição génica para recuperação da motilidade ciliar, tal como demonstrado por um estudo promissor, embora inicial, de Lai *et al.* (2016), em células epiteliais respiratórias e com o gene *DNAH11* ³⁴.

1.2.1. Etiologia genética da discinesia ciliar primária

Discinesia ciliar primária é uma patologia de grande heterogeneidade genética que apresenta, por norma, transmissão autossómica recessiva e, mais raramente, transmissão ligada ao cromossoma X recessiva ou transmissão autossómica dominante ^{25,35,36}. Conhecem-se atualmente mais de 45 genes associados à doença, mas ainda assim não é identificada a causa genética responsável em aproximadamente um terço dos casos, um valor que poderá decrescer nos próximos anos com a identificação de mais genes e variantes patogénicas associados a PCD ^{25,27}. Neste projeto serão estudados 8 genes, *DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *DNAAF2*, *RSPH4A* e *RSPH9* (Tabela 1.1), todos com transmissão autossómica recessiva.

Tabela 1.1 Painel de genes em estudo.

Gene	OMIM	Localização cromossômica	Estrutura ciliar afetada
<i>DNAH5</i> (dynein axonemal heavy chain 5)	603335	5p15.2	ODA
<i>DNAH11</i> (dynein axonemal heavy chain 11)	603339	7p15.3	ODA
<i>DNAI1</i> (dynein axonemal intermediate chain 1)	604366	9p13.3	ODA
<i>DNAI2</i> (dynein axonemal intermediate chain 2)	605483	17q25.1	ODA
<i>NME8</i> (NME/NM23 family member 8)	607421	7p14.1	ODA
<i>DNAAF2</i> (dynein axonemal assembly factor 2)	612517	14q21.3	ODA e IDA
<i>RSPH9</i> (radial spoke head component 9)	612648	6p21.1	Raio radial e par central
<i>RSPH4A</i> (radial spoke head component 4A)	612647	6q22.1	Raio radial e par central

NME/NM non-metastatic, OMIM Online Mendelian Inheritance in Man, p braço curto, q braço longo, ODA braço de dineína externo, IDA braço de dineína interno

1.2.1.1. DNAH5 e DNAH11, DNAI1 e DNAI2, e NME8, componentes dos braços de dineína externos

A ausência de braços de dineína, ou o encurtamento destas estruturas, representam a principal causa de discinesia ciliar primária ²⁵. Os braços de dineína externos, em particular, são indispensáveis para o correto funcionamento dos cílios móveis devido à propulsão que conferem a cada batimento do organelo e, como tal, a patologia PCD tende a ser causada por defeitos em proteínas da sua estrutura, proteínas citoplasmáticas necessárias para a sua formação (e de IDA), e proteínas presentes no complexo de ancoragem de ODA ao microtúbulo ³⁷. Estruturalmente, cada ODA é constituído por duas proteínas motoras, denominadas de cadeias pesadas, e por duas cadeias intermediárias (DNAI1 e DNAI2) e diversas cadeias leves, que se considera estarem envolvidas na regulação das cadeias pesadas e serem necessárias para a sua estabilidade ^{38,39}. Estudos em células epiteliais respiratórias humanas identificaram a existência de dois subtipos de ODA: ODA tipo 1, presente na região proximal do cílio e composto pelas proteínas motoras DNAH5 e DNAH11, e ODA tipo 2, na região distal, com DNAH5 e DNAH9 (*dynein axonemal heavy chain 9*) ^{40,41}. Variantes patogénicas nos genes em estudo *DNAH5*, *DNAH11*, *DNAI1* (alguns dos genes mais frequentemente associados à doença) e *DNAI2* são responsáveis pelo encurtamento ou até ausência de ODA com impacto na motilidade do cílio e, como tal, estão associados a problemas respiratórios, *situs inversus* e outros sintomas característicos de PCD ^{25,37,42}. Contudo, estudos recentes apontam para a existência de um terceiro tipo de ODA, composto pelas cadeias pesadas DNAH8 (*dynein axonemal heavy chain 8*) e DNAH17 (*dynein axonemal heavy chain 17*), e presente nos testículos ao invés de ODA tipo 1 e ODA tipo 2 ^{43,44}. Estas conclusões baseiam-se, por um lado, na elevada expressão de *DNAH8* e *DNAH17* nos testículos e na muito baixa expressão de *DNAH5* e *DNAH11* (e *DNAH9*, na maior parte da espermatogénese) nos mesmos, e o oposto ao nível de células respiratórias, e por outro, na identificação de variantes patogénicas nos genes *DNAH17* e *DNAH8* em indivíduos masculinos, diagnosticados com infertilidade isolada, caracterizada por perda de ODA em células germinativas

masculinas, mas não nos cílios de células respiratórias dos pulmões ⁴³⁻⁴⁵. Ainda assim, existem casos reportados de PCD com infertilidade masculina associados aos genes *DNAH5* e *DNAH11* ^{46,47}.

Similarmente aos genes mencionados no parágrafo anterior, *NME8* codifica uma proteína constituinte dos braços de dineína externos e, como tal, variantes patogênicas neste gene são responsáveis pelo encurtamento ou ausência desta estrutura, identificado por TEM, e consequentemente responsáveis por diversos sintomas associados a PCD, nomeadamente ao nível das vias respiratórias e na localização dos órgãos ⁴⁸. Existe, contudo, uma significativa falta de informação sobre este gene e a sua relação com a doença devido à quase total ausência de casos diagnosticados, não tendo ainda sido descritos casos de infertilidade masculina associada ao gene *NME8*, embora este tenha sido inicialmente identificado e caracterizado em células germinativas testiculares humanas ⁴⁹.

1.2.1.2. DNAAF2, um fator de montagem de dineína no citoplasma

O gene *DNAAF2*, juntamente com vários outros fatores de montagem de dineína, é responsável pela formação de ODA e de IDA, ainda no citoplasma, antes de serem transportados pelo mecanismo IFT até ao local da sua função ⁵⁰⁻⁵². Relativamente a este processo de formação, sabe-se que *DNAAF2* interage com a proteína *DNAI2* de ODA, e que quando a primeira se encontra defetiva não ocorre estabilização e adição das cadeias pesadas, exceto em ODA tipo 1 que aparenta não ser afetado ^{50,53}. Este encurtamento ou ausência dos braços de dineína devido a variantes patogênicas em *DNAAF2* é responsável por cílios móveis sem motilidade e, naturalmente, encontra-se associado aos diversos sintomas de PCD previamente descritos, incluindo infertilidade masculina ^{50,53}.

1.2.1.3. RSPH4A e RSPH9, constituintes dos raios radiais

Embora os braços de dineína sejam essenciais para o batimento ciliar, não são o único fator que, se afetado, pode causar PCD. Os raios radiais são estruturas de ligação entre o par central de microtúbulos e os restantes nove pares, que assistem na regulação do movimento dos cílios móveis, como indicado anteriormente (secção 1.1.2. Cílios móveis) ¹⁷. Variantes patogênicas em diversos genes desta estrutura, incluindo *RSPH4A* e *RSPH9*, afetam a integridade dos raios radiais e do complexo do par central e alteram o padrão de movimento do cílio para um movimento circular anormal o que, consequentemente, causa complicações respiratórias, infertilidade e outros sintomas previamente identificados de PCD, com exceção da disrupção do eixo corporal direita-esquerda devido à natural ausência de raios radiais em células nodais ⁵⁴⁻⁵⁶. Contudo, *RSPH4A* poderá não ser necessária para a produção de espermatozoides viáveis, já que estudos apontam para uma muito baixa expressão do gene *RSPH4A* nos testículos, possivelmente compensado pelo gene parálogo *RSPH6A* (*radial spoke head 6 homolog A*), expresso essencialmente nestes órgãos ^{45,57}. Ainda assim, Wang *et al.* (2022) identificaram duas variantes patogênicas no gene *RSPH4A* (heterozigotia composta) como a causa de PCD e infertilidade (astenoteratozoospermia) de um indivíduo do sexo masculino ⁵⁸.

1.3. Infertilidade masculina e astenozoospermia

Segundo a Organização Mundial de Saúde (WHO, *World Health Organization*) a infertilidade é um problema de saúde pública definido como a incapacidade de engravidar após um período de 12 meses de relações sexuais regulares desprotegidas ⁵⁹. Estima-se que possa afetar aproximadamente 1 em cada 6 pessoas em idade reprodutiva a nível mundial em algum momento da sua vida, ou seja, 17,5% da

população adulta, embora a dificuldade na avaliação desta prevalência leve a que tais estudos sejam pouco frequentes e apresentem uma considerável variação de resultados entre si ⁵⁹. Relativamente à causa da incapacidade de engravidar destes casais inférteis, verifica-se uma divisão semelhante entre os casos motivados por fatores femininos e os que derivam de fatores masculinos, ambos responsáveis por aproximadamente um terço dos mesmos, enquanto que o último terço se encontra dividido similarmente entre casos resultantes da combinação de fatores masculinos e femininos e casos de infertilidade sem explicação ⁶⁰.

Os casos de infertilidade masculina, em particular, podem derivar de diversos fatores com impacto na produção dos espermatozoides, dever-se a características anormais na qualidade e função destes, ou resultar da obstrução do trato reprodutor ou de outras disrupções da função ejacutória ^{61,62}. São exemplos varicocele, trauma testicular, infeções urogenitais, tratamentos médicos como a quimioterapia, e alterações hormonais, e ainda condições genéticas como a síndrome de Klinefelter, microdeleções no cromossoma Y e fibrose quística ^{61,62}. Fatores ambientais e fatores do estilo de vida também têm impacto na fertilidade masculina, sendo exemplos o tabaco, álcool, drogas, obesidade, exposição a toxinas, e a própria idade do indivíduo ^{61,62}. Por norma, estes fatores causam alterações perceptíveis num espermograma (Tabela 1.2), o principal exame para determinação da contribuição de fatores masculinos para a infertilidade do casal, que consiste na análise de diversos parâmetros do ejaculado incluindo o seu volume, viscosidade e pH, bem como a concentração, motilidade e morfologia dos espermatozoides ⁶³. Se forem detetados parâmetros anormais deverá ser necessária uma avaliação mais exaustiva para identificar a causa da alteração e, conseqüentemente, a sua correção, ou caso tal não seja possível, para se perceber se o casal poderá conceber com recurso a tecnologias de reprodução assistida, como a fertilização *in vitro* (IVF, *in vitro fertilization*) por microinjeção intracitoplasmática de espermatozoides (ICSI, *intracytoplasmic sperm injection*) – particularmente utilizada em casos de infertilidade masculina causados, por exemplo, por problemas na mobilidade dos espermatozoides ou baixa concentração dos mesmos –, ou se será necessário recorrer a espermatozoides de dador ou a adoção ^{64,65}. Adicionalmente, identificar e compreender a causa responsável pela infertilidade masculina pode ser de extrema importância para a saúde do indivíduo, caso se verifique que o fator é prejudicial para a sua saúde e até potencialmente fatal (e.g., cancro testicular), e para a saúde da descendência, nos casos de comorbidades genéticas hereditárias, para que se possa proporcionar um correto aconselhamento ao casal ⁶⁴. Contudo, estima-se que em cerca de 30% a 40% dos casos de infertilidade masculina com espermograma anormal não seja identificado um fator associado à variação (infertilidade masculina idiopática) ⁶⁶. Como tal, tendo em conta que existem milhares de genes envolvidos na espermatogénese, e que causas genéticas são responsáveis por aproximadamente 15 a 17% dos casos de infertilidade masculina, é possível que a identificação de novas causas genéticas venha a contribuir para o diagnóstico de uma percentagem significativa destes casos de infertilidade masculina idiopática ^{60,67}.

Tabela 1.2 Exemplos de formas de infertilidade masculina baseadas em alterações do espermograma. Com exceção da aspermia e azoospermia, a presença de uma alteração espermática num determinado indivíduo não exclui a presença simultânea de outra alteração. O diagnóstico é feito de acordo com valores de referência e, quando nenhum dos parâmetros se encontra alterado, a amostra é classificada como normozoospermica ⁶⁸.

Característica espermática	Fenótipo	Valor de referência
Aspermia	Ausência completa de ejaculado	-
Astenozoospermia	Espermatozoides com defeitos na motilidade	< 40% esp móveis, < 32% esp com MP
Azoospermia	Ausência de espermatozoides no ejaculado	0 esp

Necrozoospermia	Baixa percentagem de espermatozoides vivos	< 52% esp vivos
Oligozoospermia	Baixa concentração de espermatozoides	< 15 000 000 esp/ml
Teratozoospermia	Espermatozoides com morfologia anormal	< 4% esp com MN

esp espermatozoides, *MP* motilidade progressiva, *MN* morfologia normal

Tal como referido nas secções 1.1.2. Cílios móveis e 1.2. Discinesia ciliar primária, a patologia discinesia ciliar primária pode causar infertilidade masculina devido à função que os vários genes envolvidos na doença têm no correto funcionamento do flagelo, uma estrutura crucial na motilidade dos espermatozoides. Como tal, a patologia tende a estar associada a uma característica espermática específica, a astenozoospermia. Esta forma de infertilidade masculina é caracterizada por baixa motilidade dos espermatozoides, sendo diagnosticada quando a análise do esperma revela que menos de 40% dos espermatozoides possuem motilidade, ou menos de 32% possuem motilidade progressiva, de acordo com a WHO ^{68,69}. Contudo, tendo em conta que PCD está hoje associada a mais de 45 genes, um número que até poderá continuar a aumentar, torna-se evidente a necessidade de uma metodologia que permita a análise simultânea de múltiplos genes e doentes por ensaio, o que a sequenciação de nova geração (NGS, *next-generation sequencing*) é capaz de cumprir com custos e tempo despendido aceitáveis ⁷⁰. Não obstante, a sua implementação requer uma melhor compreensão dos genes a analisar, e é precisamente isso que se pretende atingir com este projeto: avaliar a associação entre os genes *DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *DNAAF2*, *RSPH9* e *RSPH4A* e a astenozoospermia.

1.4. Objetivos

Tendo por base a informação previamente descrita, este projeto foi desenvolvido essencialmente com dois propósitos. Pretendeu-se, numa vertente mais técnica, implementar e validar experimentalmente a caracterização molecular de um painel de genes em estudo (*DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *DNAAF2*, *RSPH4A* e *RSPH9*) por sequenciação de nova geração. Adicionalmente, procurou-se progredir no conhecimento científico em relação à infertilidade masculina no contexto da discinesia ciliar primária, ao determinar se alterações nestes oito genes poderiam estar associadas a astenozoospermia, sem que os doentes manifestassem outras características fenotípicas mais graves associadas às ciliopatias. Para tal, foram estabelecidos três objetivos principais:

- I. Sequenciação de nova geração do painel de genes em estudo, num conjunto de amostras de homens inférteis astenozoospermicos, para identificação das variantes presentes;
- II. Assegurar a sequenciação e análise de todas as regiões alvo, através da avaliação *in silico* da cobertura das regiões sequenciadas na fase anterior e da utilização de sequenciação de Sanger para pesquisa de variantes em regiões com cobertura inferior a 20x;
- III. Anotação e avaliação biopatológica das variantes identificadas nas duas fases anteriores, com previsão de consequências funcionais e estruturais em espermatozoides, para estabelecimento de correlações genótipo-fenótipo.

2 | Materiais e Métodos

2.1. Amostras biológicas

O presente estudo compreendeu um total de 164 amostras, obtidas de utentes com origem geográfica em Portugal. Destas, 116 são provenientes de indivíduos que não apresentam a patologia em estudo, pelo que constituem o grupo controlo. Os restantes 48 utentes, todos do sexo masculino, foram diagnosticados com infertilidade por possuírem astenozoospermia isolada, ou seja, não sindrómica, frequentemente associada a outras alterações do espermograma, tal como indicado na Tabela 2.1. Todas as amostras de DNA (ácido desoxirribonucleico, do inglês *deoxyribonucleic acid*) referidas fazem parte da coleção de amostras de DNA do Departamento de Genética Humana do Instituto Nacional de Saúde Doutor Ricardo Jorge, as quais foram obtidas após consentimento informado dos respetivos utentes no âmbito da realização de testes genéticos e de cedência das mesmas para investigação biomédica, tendo ainda sido codificadas e irreversivelmente anonimizadas.

Tabela 2.1 Doentes que compõem o grupo de 48 homens inférteis, distribuídos de acordo com os ensaios de sequenciação de nova geração em que as amostras foram incluídas, e classificação do respetivo espermograma. Os primeiros 3 ensaios de NGS (sequenciação de nova geração, do inglês *next-generation sequencing*) incluíram apenas amostras controlo. ID - identificação da amostra/doente; OTA - oligoteratoastenozoospermia; OAz - oligoastenozoospermia; Atz - astenozoospermia.

4º ensaio NGS		5º ensaio NGS		6º ensaio NGS	
ID	Espermograma	ID	Espermograma	ID	Espermograma
PCip1	OTA	PCip15	OTA	PCip30	OTA
PCip2	OTA	PCip16	OTA	PCip31	OTA
PCip3	OAz	PCip17	OTA	PCip32	OTA
PCip4	OAz	PCip18	OTA	PCip33	OTA
PCip5	OTA	PCip19	OAz	PCip34	OTA
PCip6	OTA	PCip20	OAz	PCip35	Atz
PCip7	OAz	PCip21	OTA	PCip36	OTA
PCip8	OAz	PCip22	Atz	PCip37	OTA
PCip9	OTA	PCip23	OTA	PCip38	OTA
PCip10	OAz	PCip24	OTA	PCip39	OAz
PCip11	OAz	PCip25	OAz	PCip40	OTA
PCip12	OTA	PCip26	OTA	PCip41	OAz
PCip13	OTA	PCip27	Atz	PCip42	OTA
PCip14	OTA	PCip28	OTA	PCip43	OTA
		PCip29	OTA	PCip44	OTA
				PCip45	OAz
				PCip46	OTA
				PCip47	OTA
				PCip48	OTA

2.2. Extração de DNA genómico

O DNA genómico, ou gDNA (*genomic DNA*), foi extraído a partir de células nucleadas do sangue periférico, utilizando o equipamento MagNA Pure LC 2.0 (Roche). Neste equipamento é utilizado um conjunto de reagentes específico, o kit *MagNA Pure LC DNA Isolation Kit - Large Volume* (Roche), e

executa todo o processo de purificação de ácidos nucleicos de forma robotizada, obtendo-se no final do processo de extração uma solução tamponada que contém o gDNA dissolvido.

2.3. Desenho do painel genético

O painel genético em estudo é composto pelos genes *DNAAF2*, *DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *RSPH4A* e *RSPH9*, associados a discinesia ciliar primária. Este painel customizado foi criado com recurso ao *software* DesignStudio, da Illumina, o qual procede à escolha/desenho dos amplicões específicos que compreendem os 223 exões codificantes e respetivas regiões intrónicas proximais dos referidos genes. Estes amplicões definem, assim, as regiões genómicas alvo analisadas por NGS. O genoma de referência humano usado corresponde ao HG19 (*Human Genome version 19*), equivalente ao genoma de referência GRCh37 (*Genome Reference Consortium Human Build 37*).

2.4. Quantificação de DNA

De modo a obter a concentração de DNA genómico requerida para a reação de NGS, 4 ng/μl, de acordo com as especificações da Illumina (protocolo *AmpliSeq for Illumina On-Demand, Custom, and Community Panels Reference Guide*, documento # 1000000036408 v09), procedeu-se à quantificação do gDNA de cada amostra no fluorímetro Qubit 3.0 (Invitrogen). Para maior precisão, o protocolo estabelecido compreendeu duas quantificações. A primeira quantificação, realizada com o kit *Qubit dsDNA Broad Range* (2 a 1000 ng) de acordo com as instruções do fabricante, permitiu diluir cada amostra de gDNA original para uma concentração de aproximadamente 10 ng/μl. De seguida procedeu-se à segunda quantificação, desta vez da solução de 10 ng/μl referida, com o kit *Qubit dsDNA High Sensitivity* (0 a 100 ng) de acordo com as instruções do fabricante, para a preparação das soluções de DNA com concentração desejada (4 ng/μl).

Para confirmação da concentração das amostras quantificadas foi realizada uma análise qualitativa e semiquantitativa do DNA através de eletroforese em gel de agarose (SeaKem) a 1% (p/v), preparado com tampão TBE (Tris-Borato-EDTA, 1x concentrado) e suplementado com brometo de etídio (10 mg/ml). DNA de concentração conhecida (10 ng/μl) do fago lambda foi utilizado para comparação com as amostras em estudo, permitindo inferir aproximadamente a concentração destas últimas e avaliar o grau de integridade do gDNA. Ambas as amostras de DNA foram suplementadas com azul de bromofenol.

2.5. Sequenciação de nova geração

Os ensaios de sequenciação de nova geração para identificação da sequência de nucleótidos das regiões genómicas alvo foram realizados na Unidade de Tecnologia e Inovação do Instituto Nacional de Saúde Doutor Ricardo Jorge, pela metodologia Ampliseq. Primeiro, procedeu-se à preparação de bibliotecas de gDNA, de acordo com o protocolo *AmpliSeq for Illumina On-Demand, Custom, and Community Panels Reference Guide* (documento # 1000000036408 v09), da Illumina, num processo que requer a fragmentação do gDNA e posterior amplificação dos fragmentos/amplicões com adição de regiões de ligação dos iniciadores de sequenciação (tagmentação), dos índices para identificação de cada amostra, e de adaptadores específicos que permitem a ligação a oligonucleótidos complementares imobilizados na *flow cell*. A desnaturação e a diluição do *pool* das bibliotecas foram feitas como indicado no guia *MiSeq System Denature and Dilute Libraries Guide* (documento # 15039740 v10), protocolo C: *AmpliSeq for Illumina Panels Normalization Method*.

A segunda e terceira etapas de NGS, denominadas geração de *clusters* e sequenciação, decorreram no equipamento MiSeq, da Illumina, com recurso ao kit *MiSeq Reagent Kit v2 (500-cycles)*. Na geração de *clusters* os fragmentos de ssDNA (*single-stranded DNA*) são depositados na *flow cell*, na qual decorrem múltiplos ciclos de amplificação em ponte (i.e., dobra do amplicão, ligação do adaptador livre ao oligonucleótido complementar da *flow cell*, extensão da cadeia molde, e desnaturação) com formação de agregados de clones de todos os fragmentos. Por fim, a sequenciação paralela destes fragmentos ocorre através da síntese da cadeia complementar, com a adição de um nucleótido terminal associado a um fluoróforo específico de uma base, seguido de excitação por um laser para identificação da base incorporada. Após clivagem do terminador e do fluoróforo, o ciclo é repetido para o nucleótido seguinte.

Terminada a sequenciação, a informação é disponibilizada em ficheiros de texto em formato FASTQ, que agregam a sequência de nucleótidos de cada leitura e os *quality scores* correspondentes. Para o mapeamento destas sequências contra o genoma de referência utilizou-se o software BWA ⁷¹ (*Burrows-Wheeler Aligner*) v0.7.13, com produção de ficheiros BAM (*Binary sequence Alignment/Map*). A identificação de variantes nestes ficheiros foi feita utilizando a ferramenta informática GATK ⁷² (*Genome Analysis Toolkit*) v4.3.0.0, com geração de ficheiros VCF (*Variant Call Format*). Este software faz parte da aplicação DNA Amplicon v2.1.1, da Illumina.

2.5.1. Controlo de qualidade dos ensaios de NGS

Todos os ensaios de NGS incluíram uma mesma amostra controlo que permitiu demonstrar a repetibilidade e reprodutibilidade do procedimento entre ensaios. Adicionalmente, após desnaturação e diluição das bibliotecas de ssDNA é adicionada uma biblioteca controlo PhiX, derivada do pequeno genoma já sequenciado do fago PhiX, permitindo maior confiança nos resultados das restantes etapas do ensaio.

Para controlo da qualidade de cada ensaio de NGS utilizou-se ainda o coeficiente de qualidade Q30, fornecido pela aplicação DNA Amplicon. *Phred quality score*, ou *Q score*, é calculado de acordo com a equação $Q = -10\log_{10}P$, e permite avaliar o desempenho do procedimento ao estimar a probabilidade de uma base ser identificada incorretamente (*P*). Um *Q score* de 30 indica que a probabilidade de determinada base ser incorretamente identificada é de 1 em 1000, ou seja, é identificada com uma precisão de 99,9%, que se reflete numa probabilidade mínima de se encontrar um falso positivo devido a erros de sequenciação.

2.6. Avaliação *in silico* da cobertura

A fim de se validar, amostra a amostra, os resultados dos ensaios de NGS, procedeu-se à avaliação *in silico* da cobertura das regiões alvo. Ou seja, para cada nucleótido entre a 25ª posição a montante e a 25ª a jusante de cada exão, bem como para os 50 nucleótidos a 5' e 3' do primeiro e último exão de cada gene, respetivamente, foi necessário analisar o número de leituras obtidas na reação de NGS, e determinar se cada nucleótido apresenta uma cobertura adequada de pelo menos 20 leituras (20x). Esta informação foi previamente incorporada nos ficheiros de alinhamento de sequências BAM (secção 2.5 Sequenciação de nova geração), e para a sua análise recorreu-se a duas metodologias diferentes: 1) o software VarAFT ⁷³ (*Variant Annotation and Filter Tool*, <https://varaft.eu/>) v2.17-2 e IGV ⁷⁴ (*Integrative Genomics Viewer*, <https://software.broadinstitute.org/software/igv/>) v2.13.2, e 2) o script *manifest_parser.py* (<https://gitlab.com/StuntsPT/misc.bam.scripts>, commit 903d16a1), que depende da biblioteca *pysam* ⁷⁵.

O *software* de identificação e anotação de variantes VarAFT foi utilizado, neste projeto, exclusivamente para a avaliação da cobertura das regiões genômicas acima descritas, através da sua função de análise de cobertura, por ser capaz de calcular a proporção de nucleótidos de uma dada região sequenciada que apresente a cobertura mínima definida de 20 leituras (Figura 2.1.A). Este *software* requer o uso de um ficheiro BED (*Browser Extensible Data*), no qual estão indicadas as coordenadas de cada região genômica de interesse, e que o próprio programa concebe de acordo com os genes e outras informações indicadas pelo utilizador. As regiões sinalizadas foram analisadas ao nível do nucleótido através do *software* IGV, que revela informação detalhada quanto à cobertura de cada base e existência de variantes (Figura 2.1.B), e como tal foi bastante útil na confirmação de resultados específicos.

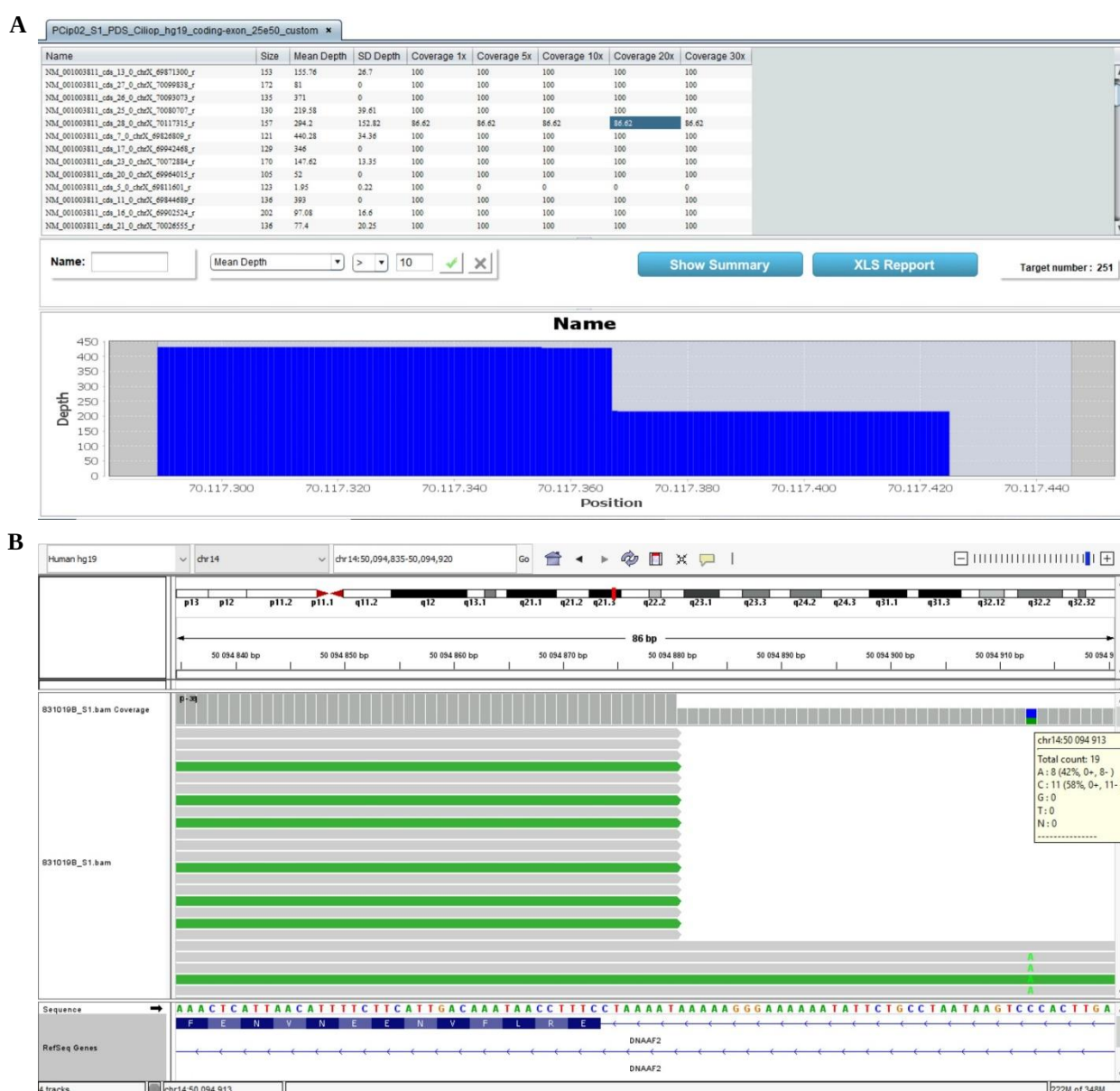


Figura 2.1 Exemplo representativo da diferente visualização de resultados nos *software* VarAFT e IGV. Enquanto que a plataforma VarAFT (A) permite uma análise direta e intuitiva de que exões sequenciados não apresentam a cobertura desejada (20x) ao longo de todo o fragmento, é necessário utilizar o *software* IGV (B) por permitir a identificação de que nucleótidos não atingem a cobertura necessária, e simultaneamente indicar as suas coordenadas.

O *script manifest_parser.py*, escrito na linguagem de programação *Python*, foi desenvolvido com o intuito de determinar e indicar que nucleótidos, dentro das regiões alvo, apresentam uma cobertura inferior à desejada. Para tal, introduz-se o seguinte comando:

```
python3 manifest_parser.py "Manifest_file.txt" "min_cov(int)"  
"flank_region(int)" files.bam
```

O *script* requer um *manifest file* (i.e., um ficheiro de texto com a identificação e as coordenadas de cada exão em estudo); o menor número de leituras aceite para cada nucleótido ("min_cov(int)" – "20" neste projeto); o número de nucleótidos a analisar nas regiões flanqueantes ("flank_region(int)", "25" pb); e ainda o ficheiro BAM da amostra em estudo, ou vários se assim for pretendido. Após analisar no(s) ficheiro(s) BAM os valores de cobertura para as regiões indicadas no *manifest file* e regiões flanqueantes, o *script* indica as coordenadas do primeiro e último nucleótido de cada sequência de nucleótidos que não atinjam a cobertura requerida.

Adicionalmente, de modo a avaliar a possível existência de variantes de interesse em regiões não compreendidas pelos amplicões definidos, fez-se uma pesquisa de variantes descritas na base de dados HGMD ⁷⁶ (*Human Gene Mutation Database*, <https://www.hgmd.cf.ac.uk/ac/index.php>), e na literatura disponível. Toda a componente bioinformática descrita nesta secção foi necessária para detetar que regiões dos genes em estudo iriam requerer confirmação por sequenciação cíclica usando o método de Sanger.

2.7. Reação em cadeia da polimerase, ou PCR

A amplificação enzimática do DNA *in vitro* por reação em cadeia da polimerase (PCR, *polymerase chain reaction*) foi realizada de acordo com os procedimentos padrão, com preparação de uma mistura de reação composta pelos reagentes indicados na Tabela 2.3, e de acordo com as condições de desnaturação do DNA, emparelhamento dos iniciadores à sequência complementar no DNA genómico, e extensão de cada cadeia complementar a partir da extremidade 3'OH de cada iniciador, indicados na Tabela 2.4, tendo sido usado o termociclador Tgradient, da Biometra. Todos os pares de iniciadores, apresentados na Tabela 2.2, foram desenhados com recurso ao *software* Primer-BLAST do *National Center for Biotechnology Information* (NCBI, <https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi>) e sintetizados pela Invitrogen.

Tabela 2.2 Iniciadores desenhados para a sequenciação das regiões alvo específicas. A designação de cada iniciador indica o gene, o número do respetivo exão (cada par de iniciadores compreende a região onde a cobertura mínima pretendida não foi atingida na NGS), e a direção do iniciador. As duas temperaturas de emparelhamento indicadas referem-se, respetivamente, à reação de PCR e à reação de sequenciação cíclica de Sanger, sendo indicado o iniciador utilizado no segundo procedimento. Estas temperaturas foram definidas após otimização de ambos os procedimentos para cada par de iniciadores utilizado. Caso o par *DNAH11_Ex33-34F1* e *DNAH11_Ex34R1* não permitisse amplificar corretamente a sequência alvo, devido ao significativo tamanho do fragmento, seria utilizado o par de iniciadores *DNAH11_Ex33-34F2* e *DNAH11_Ex34R1*. Não tendo sido necessário, não se determinou as temperaturas de emparelhamento deste último par.

Designação	Sequência (5'-3')	Tamanho esperado do fragmento amplificado (pb)	Te (°C)
<i>DNAAF2_Ex1F</i>	TCCTCCCGAATGTTAGTGGTG	460	64 / 62 ^F
<i>DNAAF2_Ex2R</i>	GCAGACATTACAAAGCAGCG		
<i>DNAH5_Ex25F</i>	ACCTCTTGCGGTTTGAAGTAG	483	58 / 58 ^R
<i>DNAH5_Ex25R</i>	TGGGAATTCCTAGGAAGTTTCAC		
<i>DNAH5_Ex63F</i>	TGCCTCCAGTTTGATTTGTGG	582	58 / 58 ^R
<i>DNAH5_Ex63R</i>	GCATCTGCTACTGCTCTACCT		
<i>DNAH11_Ex1F</i>	AAGTAGCAGCAGGTGGGAGAC	581	65 / 62 ^F
<i>DNAH11_Ex1R</i>	GAATCCAAAGGGGAGCGACTG		
<i>DNAH11_Ex10F</i>	CTGTGAACGGAAATTGCAGAC	544	58 / 58 ^R
<i>DNAH11_Ex10R</i>	ACCATCTCCATGAGGAAACCA		
<i>DNAH11_Ex18F</i>	GAAATGGAGCTGGATGTTTCAG	714	58 / 58 ^F
<i>DNAH11_Ex18R</i>	GTCATAAAACACACCGAGAGTG		
<i>DNAH11_Ex33-34F1</i>	CACAGCTCTTAGGAAATGAGATC	785	58 / 58 ^R
<i>DNAH11_Ex34R1</i>	ATGAGACAGGGAGAGGTAGC		
<i>DNAH11_Ex33-34F2</i>	CTCATATTAGACTGTGATGTTTGTG	571	—
<i>DNAH11_Ex35F</i>	TTGCAGTGTTTGCATTTGTC	338	58 / 58 ^R
<i>DNAH11_Ex35R</i>	TGTCTATAAAAGGAGCCCAAGG		
<i>DNAH11_Ex45F</i>	AAGTGAACACCACCGTTACAA	600	58 / 58 ^R
<i>DNAH11_Ex45R</i>	TTCCCAGAAGCTCAGCAGAT		
<i>DNAH11_Ex55F</i>	GTCACAGTTGGCTGCACTCC	571	63 / 62 ^F
<i>DNAH11_Ex55R</i>	ATCCTGGGCTTCCAGTGAGTC		
<i>DNAI2_Ex6F</i>	GGAAAGACTGGAGGGGCTTC	616	63 / 62 ^R
<i>DNAI2_Ex6R</i>	GCAGCAGGATTGGCTCTGAC		
<i>NME8_Ex5F</i>	TGAGAGTAGGGTAGAAGGGAG	427	63 / 62 ^F
<i>NME8_Ex5R</i>	TTGCACCATTACACACCTGG		
<i>RSPH4A_Ex3F1</i>	GAAAAAGACAAGGGAAGCAAGC	570	64 / 62 ^F
<i>RSPH4A_Ex3R1</i>	GGTGGGTAGCTTATGATGGGA		
<i>RSPH4A_Ex3F2</i>	TGCAATGAACCAGGAAGACC	587	60 / 58 ^R
<i>RSPH4A_Ex3R2</i>	AGTCTGTTTGGAGATCCCTGG		
<i>RSPH9_Ex2F</i>	GGGCTTCTAACAGTCTCACG	402	58 / 58 ^F
<i>RSPH9_Ex2R</i>	GTTGCCATAATTCTATTAGGCTGG		

Ex exão, F forward, R reverse, pb par de bases, Te temperatura de emparelhamento

Tabela 2.3 Reagentes na reação de PCR e respectivas concentrações. DMSO foi adicionado apenas para amplificação com o par de iniciadores *DNAH11_Ex1*.

Reagentes da PCR
ddH ₂ O
Tampão de PCR, AB (10x)
dNTP (25 mM)
Iniciador <i>forward</i> (10 pmol/μl)
Iniciador <i>reverse</i> (10 pmol/μl)
DMSO
<i>AmpliTaq DNA polymerase</i> , AB (5 U/μl)
DNA (~ 50 ng/μl)

ddH₂O água bidestilada, AB Applied Biosystems, dNTP desoxirribonucleótido trifosfato, DMSO dimetilsulfóxido

Tabela 2.4 Programa da PCR. As temperaturas de emparelhamento encontram-se na Tabela 2.2.

Ciclos	Temp (°C)	Tempo
1	94	5 min
35	94	45 seg
	Te	1 min
	72	1 min
1	72	5 min
	10	-

Temp temperatura, *Te* temperatura de emparelhamento, *seg* segundo, *min* minuto

A avaliação da qualidade e tamanho dos produtos da reação de PCR foi realizada por eletroforese horizontal em gel de agarose (SeaKem) a 2% (p/v), com tampão TBE concentrado 1x, suplementado com brometo de etídio (10 mg/ml). As amostras foram comparadas com o marcador de peso molecular *100 bp plus DNA Ladder* (Bioron), ambos suplementados com azul de bromofenol. Após eletroforese procedeu-se à visualização do gel no transiluminador de luz ultravioleta Uvitec Cambridge FireReader Max (Uvitec).

2.8. Sequenciação de Sanger

Para a determinação da sequência de nucleótidos de cada região em análise que não atingiu uma cobertura de 20x no ensaio de NGS, recorreu-se à sequenciação de DNA pelo método de Sanger, uma técnica de referência baseada na sequenciação cíclica do DNA e eletroforese capilar. Contudo, primeiramente, os fragmentos amplificados na reação de PCR foram purificados com recurso à mistura enzimática *Illustra ExoStar 1-Step* (GE Healthcare), constituída por fosfatases alcalinas e exonucleases I. A incubação (37°C, 15 minutos) e a posterior inativação dos enzimas (80°C, 15 minutos) decorreram no termociclador Biometra Trio.

De seguida procedeu-se à reação de sequenciação cíclica, através da mistura de reação composta pelos reagentes indicados na Tabela 2.5, de acordo com as condições apresentadas na Tabela 2.6, no mesmo termociclador Biometra Trio. Esta etapa é caracterizada pela desnaturação de dsDNA (*double-stranded DNA*), emparelhamento de iniciadores, e alongamento da cadeia complementar mediado pela DNA polimerase, num processo simultâneo para as inúmeras cadeias complementares em extensão e que termina com a incorporação de um ddNTP (didesoxirribonucleótido trifosfato), um nucleótido terminal que, ao contrário dos dNTP's (desoxirribonucleótido trifosfato) em solução, não apresenta grupo hidroxilo na posição 3', mas ao qual está acoplado um fluorocromo específico da base.

Por fim, a eletroforese capilar do produto da reação anterior decorreu na Unidade de Tecnologia e Inovação do Instituto Nacional de Saúde Doutor Ricardo Jorge, com recurso ao sequenciador *3500 genetic analyzer* da Applied Biosystems. Nesta etapa ocorre a separação dos fragmentos de DNA de acordo com o seu tamanho, e posterior excitação por laser dos fluorocromos das diversas cadeias de tamanhos diferentes, para leitura e determinação dos nucleótidos que compõem a sequência original.

Para a visualização dos resultados e confirmação da qualidade dos mesmos foi utilizada a aplicação de visualização de cromatogramas Chromas (<http://technelysium.com.au/wp/chromas/>) v2.6.6, da Technelysium.

Tabela 2.5 Reagentes na reação de sequenciação cíclica de Sanger e respectivas concentrações.

Mistura reacional
ddH ₂ O
<i>Big Dye Terminator v1.1</i> , AB (1:20)
Tampão de sequenciação (5x)
Iniciador (1 pmol/μl)
Produto de PCR (~ 0,5 ng/μl)

ddH₂O água bidestilada, AB Applied Biosystems

Tabela 2.6 Ciclo térmico das reações de amplificação. As temperaturas de emparelhamento encontram-se na Tabela 2.2.

Ciclos	Temp (°C)	Tempo
1	96	30 seg
	96	10 seg
25	Te	5 seg
	60	4 min
	10	-

Temp temperatura, *Te* temperatura de emparelhamento, *seg* segundo, *min* minuto

2.9. Análise *in silico* de variantes

A análise de dados dos ensaios de NGS, previamente iniciada com a transformação dos dados (secção 2.5. Sequenciação de nova geração) e análise de coberturas (secção 2.6. Avaliação *in silico* da cobertura), terminou com a anotação de variantes e identificação do seu significado clínico. Para tal, foram utilizados os ficheiros VCF previamente produzidos (secção 2.5. Sequenciação de nova geração), e a aplicação Variant Interpreter v2.16.1.300 (Illumina), que utiliza estes ficheiros para revelar SNV (*single nucleotide variants*) e pequenas deleções ou inserções presentes em cada amostra, juntamente com informações sobre as mesmas, como a sua localização no gene, a alteração ao nível da sequência codificante e da proteína, e o número e proporção de leituras que apresentam a variante.

Paralelamente, para a identificação de variantes presentes nas regiões sequenciadas pelo método de Sanger, comparou-se cada sequência de nucleótidos obtida com a sequência de referência - MANE (*Matched Annotation from NCBI and EMBL-EBI*). Para tal, foi utilizado o *software* Nucleotide BLAST ⁷⁷ do NCBI (https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome).

2.9.1. Filtragem de variantes

O *pool* de variantes resultante, por ser muito extenso e incluir um significativo número de variantes sem relevância para o projeto dado que correspondem maioritariamente a polimorfismos não patogénicos e frequentes na população, foi filtrado de acordo com os seguintes parâmetros: 1) frequência populacional, não tendo sido consideradas as variantes que apresentassem frequências superiores a 2% para toda a população nas bases de dados GnomAD ⁷⁸ (*Genome Aggregation Database*, <https://gnomad.broadinstitute.org/>) v2.1.1, 1000 Genomes Project ⁷⁹ (<https://www.internationalgenome.org/>), TOPMed ⁸⁰ (*Trans-Omics for Precision Medicine*, <https://topmed.nhlbi.nih.gov/>), e NHLBI GO ESP (*National Heart Lung and Blood Institute "Grand Opportunity" Exome Sequencing Project*, <https://esp.gs.washington.edu/drupal/>); 2) localização no genoma, tendo sido desprezadas as variantes que não se encontravam nas regiões genómicas em estudo previamente definidas; e 3) presença de homopolímero, ou seja, variantes localizadas em sequências nucleotídicas repetitivas e consequentemente propensas a erros de leitura

durante a sequenciação por dificuldade na interpretação destas sequências foram excluídas. Nesta fase utilizou-se simultaneamente o Alamut Visual Plus (<https://www.sophiagenetics.com/platform/alamut-visual-plus/>) v1.6.1, um *software* especializado na interpretação de variantes que permite a sua visualização na sequência nucleotídica, providencia informação de diversas bases de dados, e apresenta preditores do efeito de variantes *missense* e de alterações que possam ter impacto ao nível do *splicing*.

Deste modo, todas as variantes identificadas pela metodologia NGS ou pela reação de sequenciação cíclica de Sanger, com frequências populacionais inferiores a 2%, presentes nas regiões codificantes e regiões não codificantes proximais em estudo, e que não integravam homopolímeros, foram analisadas, interpretadas e classificadas através de preditores de patogenicidade *in silico*, bases de dados *online*, e *software* polivalentes.

2.9.2. Anotação e classificação de variantes

As variantes *missense* foram analisadas com auxílio de quatro programas, Align GVGD ⁸¹ (*Align Grantham variation Grantham deviation*, <http://agvgd.hci.utah.edu/about.php>), PolyPhen-2 ⁸² (*Polymorphism Phenotyping v2*, <http://genetics.bwh.harvard.edu/pph2/>), SIFT ⁸³ (*Sorting Intolerant From Tolerant*, <https://sift.bii.a-star.edu.sg/>) v6.2.0, e MutationTaster2 ⁸⁴ (<https://www.mutationtaster.org/>). Estes programas, incorporados no *software* Alamut Visual Plus, avaliam o impacto da substituição de um aminoácido por outro na função da proteína, após consideração de fatores como as propriedades físico-químicas dos dois aminoácidos, a conservação evolutiva dos mesmos entre espécies, e a estrutura e domínios funcionais da proteína.

A análise de variantes intrónicas, e das variantes sinónimas e *missense* que possam afetar a correta maturação do pré-mRNA (ácido ribonucleico mensageiro precursor, do inglês *precursor messenger ribonucleic acid*) em mRNA que caracteriza o *splicing*, foi realizada com recurso ao *software* HSF ⁸⁵ (*Human Splicing Finder*, <https://www.genomnis.com/access-hsf>) v4.3.2 e a quatro algoritmos incorporados no programa Alamut Visual Plus: SpliceSiteFinder-like ⁸⁶, MaxEntScan ⁸⁷ (*Maximum Entropy Scan*), NNSPLICE 0.9 ⁸⁸ (*Neural Network SPLICE*, https://www.fruitfly.org/seq_tools/splice.html) e GeneSplicer ⁸⁹. Estes preditores foram desenhados para identificar locais aceitadores e locais dadores de *splicing*, pontos de ramificação (*branch points*), *exonic splicing enhancers* (ESE) e *exonic splicing silencers* (ESS) presentes em proximidade à variante. Adicionalmente, permitem prever as consequências que modificações ou criação de sinais de *splicing* possam ter na excisão de intrões e, fundamentalmente, no transcrito.

Por fim, teve-se ainda em consideração toda a informação e avaliações expressas na aplicação Variant Interpreter (Illumina) e na base de dados HGMD (secção 2.6. Avaliação *in silico* da cobertura), bem como nas plataformas NCBI ClinVar ⁹⁰ (<https://www.ncbi.nlm.nih.gov/clinvar/>), que integra inúmeras variantes genómicas e respetivas interpretações de patogenicidade submetidas por entidades externas/laboratórios, e Varsome ⁹¹ (<https://varsome.com/>) v11.3.3, que agrega informação de mais de 140 fontes com o objetivo de analisar o impacto de variantes genéticas e partilhar conhecimento sobre estas. Com base em toda a informação adquirida, as variantes em estudo foram avaliadas de acordo com as diretrizes recomendadas pelo *American College of Medical Genetics and Genomics* e pela *Association for Molecular Pathology* (ACMG-AMP guidelines), tendo-se classificado cada variante de acordo com a sua patogenicidade em: benigna, provavelmente benigna, de patogenicidade desconhecida, provavelmente patogénica, ou patogénica ⁹².

3 | Resultados

3.1. Sequenciação de nova geração

Como indicado anteriormente (secção 1.3. Infertilidade masculina e astenozoospermia), é expectável que uma percentagem significativa dos casos de infertilidade masculina idiopática resulte de alterações genéticas patogénicas. A patologia discinesia ciliar primária, por ser uma doença rara e pouco estudada, de difícil diagnóstico e de incidência subestimada, mas que pode causar infertilidade masculina, poderá ser responsável por uma parte destes casos de infertilidade sem causa aparente. Assim, o primeiro objetivo deste projeto consistiu na sequenciação do DNA genómico de 48 indivíduos do sexo masculino inférteis (características espermáticas indicadas na Tabela 2.1) e 116 amostras controlo, para o conjunto de genes associados a PCD *DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *DNAAF2*, *RSPH4A* e *RSPH9*.

3.1.1. Quantificação de DNA

Antes de cada ensaio de NGS, foi necessário quantificar e diluir as respetivas amostras de gDNA à concentração de 4 ng/μl, ideal para a reação de sequenciação. Cada amostra diluída foi visualizada e avaliada por eletroforese em gel de agarose, como exemplificado na Figura 3.1 para as amostras PCip32 a PCip41.

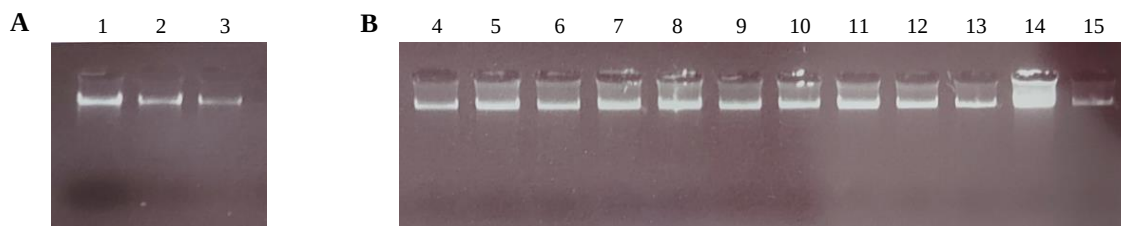


Figura 3.1 Visualização da concentração das amostras PCip32 a PCip41 em gel de agarose. Exemplo representativo da análise do DNA quantificado e diluído antes de cada reação de NGS, relativa ao terceiro ensaio de NGS para amostras de indivíduos inférteis. A) DNA de concentração conhecida (10 ng/μl) do fago lambda, em três poços com as massas 100, 50 e 25 ng (poços 1, 2 e 3, respetivamente), para comparação com o DNA das amostras em estudo. B) DNA quantificado e diluído das amostras PCip32 a PCip37 (poços 4 a 9), PCip38 a PCip41 (poços 12 a 15), e de duas amostras quantificadas, mas posteriormente eliminadas, devido a problemas associados à sequenciação (poços 10 e 11).

A comparação da intensidade da banda de cada amostra em estudo com as três bandas relativas ao fago lambda permitiu determinar, em cada ensaio, que amostras diluídas não apresentavam a concentração desejada. Neste caso, o poço 14 (amostra PCip40) não apresentou o resultado esperado e, como tal, esta amostra requereu uma segunda visualização em gel de agarose para confirmação. Inversamente, todas as amostras que não evidenciaram problemas foram posteriormente sequenciadas.

3.1.2. Controlo de qualidade dos ensaios de NGS

Após cada ensaio de NGS, recorreu-se ao *Phred quality score*, ou *Q score*, para a avaliação inicial do desempenho da reação de sequenciação. Na Tabela 3.1 encontram-se as médias, por ensaio, da percentagem de nucleótidos que apresentaram um *Q score* de 30 (Q30) ou superior.

Tabela 3.1 Percentagem média, por ensaio, de nucleótidos que atingiram um Q score de 30. Um *quality score* de 30 (Q30) indica que a probabilidade de uma base ser identificada incorretamente é de 1/1000.

Ensaio NGS	1	2	3	4	5	6
Média (stdev)	93,75% (0,64)	92,08% (0,69)	95,06% (0,85)	91,79% (0,92)	95,59% (0,77)	92,43% (0,86)

NGS sequenciação de nova geração, *stdev* desvio padrão

Todos os ensaios apresentaram percentagens médias acima de 90%, e ao nível das amostras verificou-se que entre 88,44% (Ct095, ensaio 4) e 96,48% (PCip18, ensaio 5) das bases atingiram um *Q score* de 30. Estes elevados valores comprovam a qualidade dos seis ensaios de NGS e, como tal, pôde-se prosseguir para a análise da cobertura das regiões sequenciadas.

3.2. Identificação e sequenciação das regiões não cobertas por NGS

O uso de plataformas de NGS possibilitou a sequenciação simultânea das regiões codificantes e regiões intrónicas flanqueantes para o painel de genes estipulado (Tabela 1.1). Ainda assim, é necessário ter em conta a possibilidade de existirem regiões de interesse sem a cobertura mínima (20 leituras, ou 20x). Como tal, para se assegurar que todas as regiões alvo foram devidamente sequenciadas, procedeu-se à análise *in silico* complementar da cobertura de todos os nucleótidos nas regiões em estudo, e sequenciou-se pelo método de Sanger quaisquer regiões com cobertura inferior a 20x.

3.2.1. Avaliação *in silico* da cobertura

Entre a informação disponibilizada pela aplicação DNA Amplicon após cada ensaio de NGS, encontram-se dados relativos à cobertura dos diversos amplicões definidos. Nestes, é possível destacar os elevados valores médios de cobertura de cada amostra, tendo-se obtido uma cobertura média de aproximadamente 548,7x entre as 164 amostras sequenciadas (controlo e doentes), variando entre aproximadamente 156,6x para a amostra Ct109 e 1441,9x para a amostra Ct027.

Estes valores encontram-se bastante acima do limite de confiança de 20x, mas, sendo médias e não a cobertura de cada nucleótido, não permitem identificar que sequências não atingem a cobertura desejada. Para isso, recorreu-se ao *software* VarAFT, com apoio do *software* IGV, e simultaneamente ao *script* *manifest_parser.py*, para comparação entre as duas metodologias. Esta análise bioinformática permitiu identificar um conjunto significativo de regiões sem a cobertura desejada, identificadas na Tabela S1. Tendo em conta o objetivo expresso de assegurar a sequenciação e análise de todas as regiões alvo nos 8 genes em estudo, foram desenhados iniciadores específicos que compreendessem as sequências com cobertura insuficiente. Estas regiões encontram-se indicadas na Tabela 3.2, e as sequências dos iniciadores utilizados constam na Tabela 2.2. Para melhor visualização e compreensão, como exemplo ilustrativo, procedeu-se à representação das regiões sem cobertura suficiente nos genes *DNAAF2* e *RSPH4A* (Figura 3.2 e Figura 3.3, respetivamente).

Tabela 3.2 Regiões alvo sem cobertura suficiente (20x), nos oito genes em estudo, identificadas para posterior amplificação enzimática e sequenciação pelo método de Sanger. Todas as regiões alvo do gene *DNAI1* foram devidamente sequenciadas nos ensaios de NGS, para as amostras pertencentes ao grupo de homens inférteis.

Amostra	Gene	Exão	Coordenadas
Todas	<i>DNAAF2</i>	1	50 101 694 a 50 101 917
Todas	<i>DNAH5</i>	25	13 867 858 a 13 867 888
Todas		63	13 753 317 a 13 753 361
6 amostras Ct, 9 PCip	<i>DNAH11</i>	1	21 582 814 a 21 583 120
13 amostras Ct, 12 PCip		10	21 627 657 a 21 627 844
98 amostras Ct, 38 PCip		18	21 640 089 a 21 641 067
Todas		34	21 726 975 a 21 727 170
9 amostras Ct, 9 PCip		35	21 730 358 a 21 730 524
8 amostras Ct, 10 PCip		45	21 765 404 a 21 765 550
9 amostras Ct, 4 PCip		55	21 805 021 a 21 805 232
8 amostras Ct, 11 PCip	<i>DNAI2</i>	6	72 287 275 a 72 287 297
8 amostras Ct, PCip10 e 29	<i>NME8</i>	5	37 901 642 a 37 901 771
107 amostras Ct, 43 PCip	<i>RSPH4A</i>	3	116 948 971 a 116 949 267
Todas	<i>RSPH9</i>	2	43 618 087 a 43 618 302

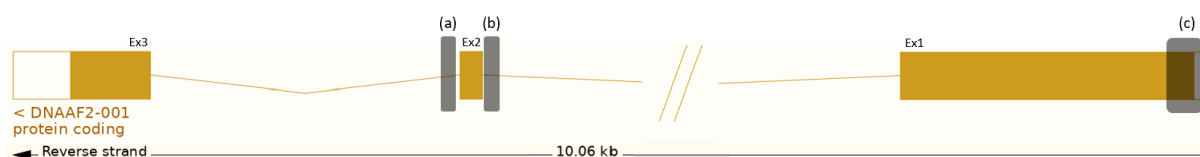


Figura 3.2 Representação esquemática das regiões sem cobertura suficiente (20x) no transcrito do gene *DNAAF2*. Estão assinalados os seus três exões, bem como as várias regiões não cobertas identificadas: (a) e (b) regiões sem cobertura suficiente na amostra Ct084; (c) região sem cobertura suficiente em todas as amostras, para a qual foram desenhados iniciadores. Ex – exão.



Figura 3.3 Representação esquemática das regiões sem cobertura suficiente (20x) no transcrito do gene *RSPH4A*. Estão assinalados os seus seis exões, bem como as várias regiões não cobertas identificadas: (a) região sem cobertura suficiente na amostra Ct084; (b) região sem cobertura suficiente em 107 amostras Ct e em 43 amostras PCip, para a qual foram desenhados iniciadores; (c) região sem cobertura suficiente em todas as amostras. Ex - exão.

É de salientar que as regiões sem a cobertura desejada identificadas na Tabela 3.2 e na Tabela S1 e representadas na Figura 3.2 e na Figura 3.3 podem variar ligeiramente entre amostras. Como tal, as coordenadas indicadas nestas tabelas são relativas à sequência de nucleótidos mais extensa sem cobertura adequada para cada região identificada.

Adicionalmente, o uso de duas abordagens (i.e., o uso do *software* VarAFT e IGV e do *script* *manifest_parser.py*) provou ter sido uma escolha acertada. Ambas as metodologias permitiram identificar com bastante precisão as regiões sem cobertura suficiente, mas foram detetadas algumas discrepâncias, expressas na Tabela 3.3.

Tabela 3.3 Regiões sem cobertura suficiente (20x) identificadas pelo *script manifest_parser.py*, mas não pelo software VarAFT. O programa IGV, por revelar o número de leituras de cada nucleótido, foi utilizado para se confirmar se as regiões sinalizadas tinham sido identificadas corretamente.

Amostra	Gene	Exão	Coordenadas	IGV
40 amostras Ct, 14 PCip	<i>DNAH5</i>	18	13 886 245	Confirmada
20 amostras Ct, 15 PCip		18	13 886 245 a 13 886 246	Confirmada
6 amostras Ct, PCip43		18	13 886 245 a 13 886 247	Confirmada
Ct111		18	13 886 245 a 13 886 248	Confirmada
16 amostras Ct, 9 PCip		25	13 867 858 a 13 867 888; 13 868 104	Confirmada
Ct066, PCip20, 48 e 50		43	13 814 977	Confirmada
PCip48	<i>DNAH11</i>	55	21 805 228 a 21 805 232	Confirmada
14 amostras Ct, 9 PCip		55	21 805 229 a 21 805 230	Confirmada
13 amostras Ct, 6 PCip		55	21 805 229 a 21 805 232	Confirmada
8 amostras Ct, PCip12	<i>NME8</i>	1	37 889 839 a 37 889 842	Confirmada

O *software* VarAFT não identificou as regiões indicadas (Tabela 3.3). Este erro deverá estar relacionado com o facto de estas sequências de poucos nucleótidos fazerem parte de regiões homopoliméricas, embora não seja claro o porquê de estas estruturas afetarem os resultados do *software* VarAFT, mas não do *script* utilizado. Inversamente, o *script manifest_parser.py* demonstrou excelentes resultados, não tendo havido nenhuma região sem a cobertura mínima que o VarAFT tenha sido capaz de identificar e o *script* não.

3.2.2. PCR e sequenciação de Sanger

Antes de se poder sequenciar as regiões indicadas na Tabela 3.2 pelo método de Sanger, e identificar quaisquer variantes presentes, procedeu-se à amplificação enzimática do gDNA por PCR. A avaliação dos resultados de cada reação de PCR foi realizada por eletroforese em gel de agarose, como exemplificado na Figura 3.4, relativa à otimização inicial da temperatura de emparelhamento dos pares de iniciadores *NME8_Ex5* e *RSPH9_Ex2*.

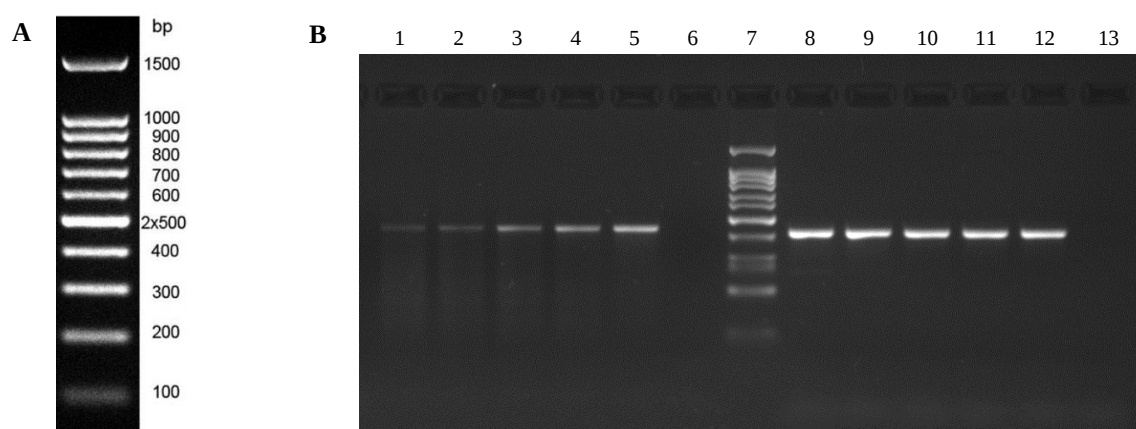


Figura 3.4 Visualização do resultado das amplificações de fragmentos de *NME8_Ex5* e *RSPH9_Ex2*, em gel de agarose. Exemplo representativo da análise dos produtos da reação de PCR, para otimização da temperatura de emparelhamento aplicada a cada par de iniciadores nesta mesma reação. A) Marcador de peso molecular 100 bp plus DNA Ladder, da Bioron. B) Poços 1 a 5: DNA amplificado de um fragmento do quinto exão do gene *NME8*, com um tamanho esperado de 427 pb. Poços 8 a 12:

DNA amplificado de um fragmento do segundo exão do gene *RSPH9*, com um tamanho esperado de 402 pb. Poços 6 e 13: os controlos negativos para *NME8_Ex5* e *RSPH9_Ex2*, respetivamente; Poço 7: o marcador de peso molecular. Temperaturas aplicadas: 55°C para os poços 1 e 8; 57°C para 2 e 9; 58,3°C para 3 e 10, e para os controlos negativos 6 e 13; 59,7°C para 4 e 11; 62,2°C para 5 e 12. Inversamente à amplificação de DNA das amostras de indivíduos inférteis, na otimização da temperatura de emparelhamento foi sempre utilizado DNA de uma única amostra controlo. Amplificação com recurso à *AmpliTaq DNA polymerase*, da Applied Biosystems.

As temperaturas de emparelhamento definidas para cada um dos catorze pares de iniciadores encontram-se na Tabela 2.2. A amplificação das regiões identificadas para as respetivas amostras de indivíduos inférteis e a purificação dos fragmentos amplificados antecederam a sequenciação destas mesmas regiões, pelo método de sequenciação cíclica de Sanger. Para a visualização dos resultados obtidos recorreu-se à aplicação Chromas, como exemplificado na Figura 3.5.

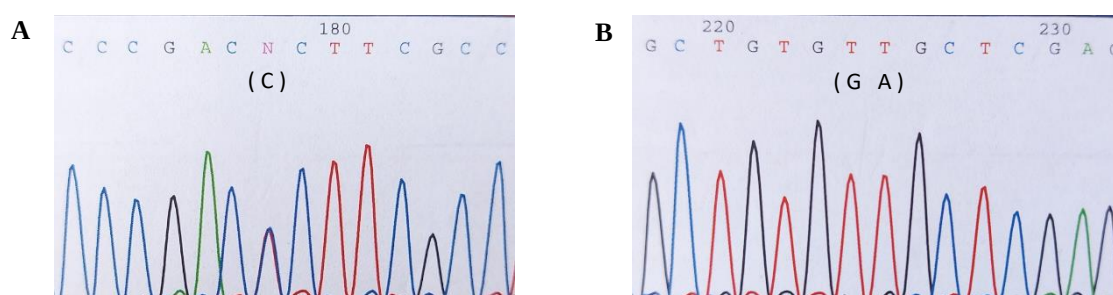


Figura 3.5 Excertos de um cromatograma exemplificativo dos resultados obtidos após sequenciação de DNA pelo método de Sanger. Na visualização dos resultados foi utilizada a aplicação Chromas, da Technelysium. Posteriormente (secção 3.3. Anotação e avaliação biopatológica das variantes detetadas), a comparação dos resultados obtidos com a sequência de referência MANE do gene *DNAH11* (exão 1) permitiu a identificação das variantes c.54C>T (em A) e c.100_101delinsTT (B).

3.3. Anotação e avaliação biopatológica das variantes detetadas

Terminada a sequenciação das regiões de interesse no painel de genes em estudo, pôde-se prosseguir com a análise das variantes presentes. Para tal, primeiro foram fornecidos os ficheiros VCF de cada um dos 6 ensaios de NGS à aplicação Variant Interpreter, que os utilizou para revelar um total de 365 variantes nas 164 amostras (incluindo grupo controlo e doentes).

Paralelamente, procedeu-se à comparação entre as regiões sequenciadas pelo método de Sanger e a sequência de referência MANE, com o objetivo de identificar todas as variantes presentes nos fragmentos amplificados. No total, foram identificadas 26 variantes em 6 dos 8 genes em estudo (Tabela 3.4), sendo as exceções os genes *RSPH4A*, por não se ter identificado nenhuma variante no exão sequenciado pelo método de Sanger, e o gene *DNAI1*, corretamente sequenciada por NGS para todas as amostras de indivíduos inférteis nas regiões de interesse.

Tabela 3.4 Variantes identificadas na análise aos resultados obtidos após sequenciação de DNA pelo método de Sanger. As variantes foram descritas de acordo com as recomendações da HGVS. Os valores de frequência alélica apresentados provêm da base de dados populacional GnomAD. As variantes c.-57dup e c.100_101delinsTT (*DNAH11*) não se encontram representadas nesta base de dados.

Gene	Exão	Variante	Frequência alélica (%)	Região coberta por NGS?
<i>DNAAF2</i>	1	c.-79C>T	1,822	Não
	1	c.186G>C	64,204	Sim
<i>DNAH5</i>	25	c.3835-3del	41,341	Sim
	25	c.3987A>G	1,066	Sim
	25	c.4053+32C>T	10,953	Não
	63	c.10556-72del	0,125	Não
<i>DNAH11</i>	1	c.-57dup	-	Sim
	1	c.54C>T	10,141	Não
	1	c.100_101delinsTT	-	Não
	10	c.1848+90A>G	16,915	Não
	18	c.3630G>A	77,563	Sim
	18	c.3648+33A>G	12,616	Sim
	18	c.3648+46C>A	0,026	Sim
	33	c.5622-7C>A	0,016	Sim
	34	c.5779-33G>A	0,129	Sim
	34	c.5924+41A>G	24,109	Não
	35	c.6041+18T>C	0,420	Não
	45	c.7290C>T	44,913	Não
	45	c.7335G>A	19,113	Não
	45	c.7440+34T>C	85,141	Sim
	45	c.7440+63C>A	21,569	Sim
	55	c.9102+8G>A	1,218	Não
	55	c.9102+20G>A	10,126	Não
<i>DNAI2</i>	6	c.724+158G>A	30,662	Sim
<i>NME8</i>	5	c.271-49G>A	88,879	Sim
<i>RSPH9</i>	2	c.228-86C>T	19,517	Não

Entre as 26 variantes identificadas por sequenciação de Sanger, 13 pertencem a regiões sem a cobertura mínima requerida. Destas, 7 variantes encontram-se dentro dos limites das regiões genômicas em estudo anteriormente estabelecidas. Adicionalmente, 16 variantes apresentam frequências populacionais elevadas, superiores a 2%, pelo que correspondem a polimorfismos populacionais e, presumivelmente, não são patogênicas. Estes dois parâmetros (frequência populacional e localização genômica), juntamente com a eliminação de variantes associadas a homopolímeros, foram utilizados na seleção das variantes identificadas pela reação de sequenciação cíclica de Sanger, bem como na seleção das variantes obtidas por NGS. Este processo resultou, assim, num total de 57 variantes em estudo (Figura 3.6).

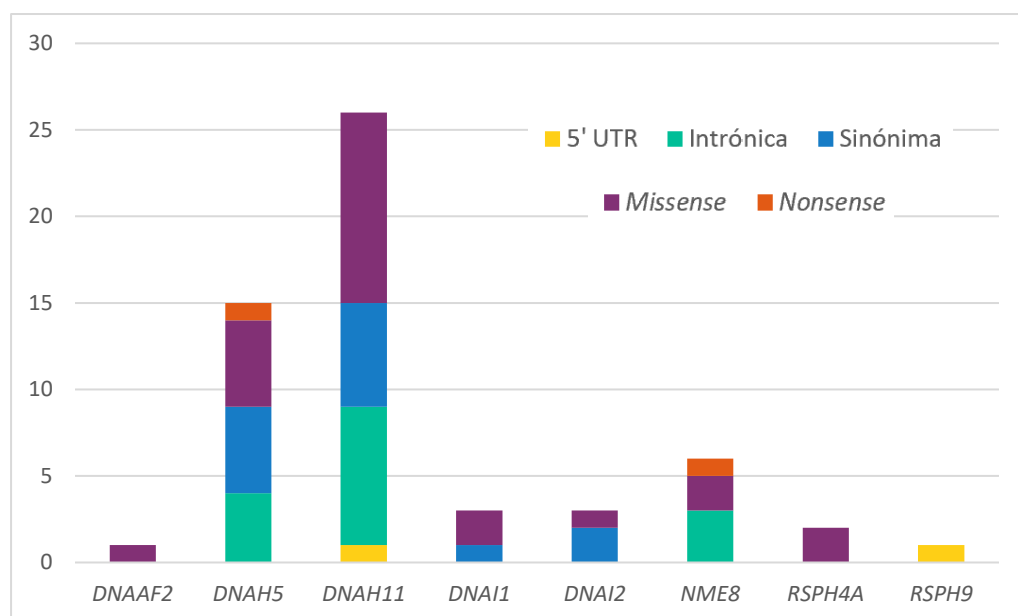


Figura 3.6 Distribuição das 57 variantes em estudo por gene e de acordo com a sua consequência/localização no mesmo.

A distribuição das 57 variantes em estudo de acordo com os seus genes e consequência/localização no gene permitiu observar que a grande maioria das variantes (53 em 57) se encontra em genes que codificam proteínas presentes nos braços de dineína, predominantemente nos genes *DNAH5* (15) e *DNAH11* (26 variantes). Tal não surpreendeu, visto que, não só 5 dos 8 genes em estudo codificam proteínas estruturais dos braços de dineína, como os genes *DNAH5* (79 exões) e *DNAH11* (82 exões) são significativamente maiores que os restantes (62 exões em conjunto, num total de 223 em estudo). Por outro lado, verificou-se uma predominância de variantes *missense* (24), seguidas pelas variantes intrónicas (15) e sinónimas (14). Inversamente, apenas foram identificadas 2 variantes localizadas na região correspondente à região 5' não traduzida (5' UTR, do inglês *5' untranslated region*) do mRNA e 2 variantes *nonsense*. A predominância de variantes em regiões codificantes também era esperada, tendo em consideração as características do estudo (compreender os exões, os 25 nucleótidos a montante e a jusante de cada exão de cada gene, e os 50 nucleótidos a 5' do primeiro exão e a 3' do último).

Estas variantes foram analisadas e classificadas quanto à sua patogenicidade em benigna (B), provavelmente benigna (PB), significado clínico desconhecido (VUS, *variant of unknown significance*), provavelmente patogénica (PP) ou patogénica (P), de acordo com as diretrizes da ACMG-AMP. Para tal, foram tidos em consideração dados disponibilizados pela aplicação Variant Interpreter, por plataformas e bases de dados *online*, e por preditores de patogenicidade *in silico*, como explicado anteriormente (secção 2.9.2. Anotação e classificação de variantes). As 45 variantes B ou PB encontram-se na Tabela S2. De igual modo, as 12 variantes VUS, PP ou P, e uma variante B e outra PB com informação conflituosa, constam na Tabela 3.5.

A interpretação das variantes genéticas de interesse levou à classificação de 21 variantes como B, e de 24 como PB. Essencialmente, entre os critérios que pesaram nesta classificação, verificou-se frequentemente: preditores bioinformáticos considerarem a variante como benigna ou outra classificação específica do preditor com significado semelhante, a variante ter sido reportada anteriormente como benigna por múltiplas entidades/investigadores, frequências alélicas superiores ao esperado para uma doença genética rara, e a variante ser identificada em homozigotia em adultos e idosos saudáveis. Em certos casos, porém, a falta de informação disponível a suportar a classificação aparente ou a existência de informação não totalmente concordante levou a que algumas variantes tenham sido classificadas como PB.

Adicionalmente, as variantes *DNAH11*:c.100_101delinsTT e *NME8*:c.1478_1479delinsCT, que não se encontravam descritas na base de dados GnomAD, não foram eliminadas durante o processo de seleção de variantes dado não serem conhecidas as suas frequências populacionais. Contudo, a anotação das variantes obtidas neste estudo permitiu perceber que as duas variantes mencionadas são muito frequentes, tanto em indivíduos inférteis (presentes em 37 e 29 das 48 amostras de doentes, respetivamente) como nas amostras controlo (80 e 62 das 116 amostras) e, como tal, foram classificadas como benignas.

Tabela 3.5 Variantes genéticas de interesse, identificadas no grupo de 48 indivíduos do sexo masculino inférteis. As variantes foram descritas de acordo com as recomendações da HGVS ao nível do cDNA e da proteína. A classificação de cada variante teve por base informação disponibilizada pela aplicação Variant Interpreter, por plataformas e bases de dados *online*, e por preditores de patogenicidade *in silico*, como explicado na secção 2.9.2. Anotação e classificação de variantes, e foi realizada de acordo com as diretrizes da ACMG-AMP. * Também presentes na Tabela S2.

Amostra PCip	Gene	HGVS cDNA	HGVS proteína	Tipo de variante	Frequência GnomAD (%)	Classificação
11 e 41	<i>DNAH11</i>	c.10739G>A	p.(Arg3580His)	<i>Missense</i>	0,88611	B*
	<i>DNAH11</i>	c.8990G>A	p.(Arg2997Gln)	<i>Missense</i>	1,1237	PB*
03	<i>DNAH11</i>	c.2909A>G	p.(Tyr970Cys)	<i>Missense</i>	0,00393	VUS
	<i>DNAH11</i>	c.9935A>T	p.(Asp3312Val)	<i>Missense</i>	0,35355	VUS
13	<i>NME8</i>	c.1169G>C	p.(Arg390Thr)	<i>Missense</i>	0,01524	VUS
22	<i>NME8</i>	c.455-5dup	p.(?)	Intrónica	0,00369	VUS
26	<i>DNAH11</i>	c.1084G>A	p.(Ala362Thr)	<i>Missense</i>	0,00143	VUS
27	<i>DNAH5</i>	c.12465A>G	p.(Gln4155=)	Sinónima	0,11394	VUS
34	<i>DNAH5</i>	c.6748G>A	p.(Glu2250Lys)	<i>Missense</i>	0,00990	VUS
47	<i>DNAH5</i>	c.3614C>T	p.(Ala1205Val)	<i>Missense</i>	0,00080	VUS
	<i>DNAAF2</i>	c.2239G>A	p.(Glu747Lys)	<i>Missense</i>	0,00066	VUS
01	<i>NME8</i>	c.1277T>A	p.(Leu426Ter)	<i>Nonsense</i>	0,00837	PP
21	<i>DNAI1</i>	c.400G>C	p.(Gly134Arg)	<i>Missense</i>	0,04035	PP
04	<i>DNAH5</i>	c.13486C>T	p.(Arg4496Ter)	<i>Nonsense</i>	0,00602	P

Ala alanina, *Arg* arginina, *Asp* ácido aspártico, *Cys* cisteína, *Gln* glutamina, *Glu* ácido glutâmico, *Gly* glicina, *His* histidina, *Leu* leucina, *Lys* lisina, *Thr* treonina, *Tyr* tirosina, *Val* valina, *Ter* codão de terminação, *B* benigna, *PB* provavelmente benigna, *VUS* variant of unknown significance, *PP* provavelmente patogénica, *P* patogénica.

Relativamente às 12 variantes de interesse, 9 foram classificadas como VUS. Neste conjunto encontram-se algumas variantes com frequências alélicas particularmente baixas, para as quais a falta de informação e de critérios para classificação leva a que seja difícil prever a sua patogenicidade. *DNAH5*:c.12465A>G, contudo, difere das restantes por apresentar simultaneamente indícios de que poderá ser benigna, nomeadamente a frequência alélica em África (1,226%) acima do esperado, e indícios de que poderá ser patogénica, por poder ter impacto significativo no *splicing* segundo o preditor HSF e vários algoritmos do programa Alamut Visual Plus, os quais sugerem a criação de um novo sítio aceitador de *splicing*.

Apenas 3 variantes foram classificadas como provavelmente patogénicas ou patogénicas: as variantes *nonsense* *NME8*:c.1277T>A e *DNAH5*:c.13486C>T, e a variante *missense* *DNAI1*:c.400G>C que, de acordo com preditores *in silico*, deverá afetar o local aceitador de *splicing* adjacente. Tal classificação deveu-se ao impacto que se prevê que tenham na estrutura das respetivas proteínas, com perda de função das mesmas. Consequentemente, poderão estar relacionadas com a patologia discinesia ciliar primária.

O conjunto de variantes de interesse inclui, ainda, *DNAH11*:c.10739G>A e *DNAH11*:c.8990G>A. Estas duas variantes classificadas, respetivamente, como B e PB segundo os critérios definidos para este projeto, encontram-se descritas como patogénicas na literatura disponível⁹³.

4 | Discussão

Atualmente, a infertilidade masculina continua a ser, em muitos casos, uma condição de difícil diagnóstico devido à enorme diversidade de fatores genéticos envolvidos na gametogênese e ao reduzido número de causas genéticas identificadas no âmbito da investigação molecular^{60,67}. Como tal, para se poder identificar os fatores associados aos casos de infertilidade masculina idiopática, é necessário compreender melhor as múltiplas doenças genéticas raras associadas ao sintoma em causa. Nas últimas duas décadas foram identificadas variantes responsáveis pela patologia discinesia ciliar primária em mais de 45 genes, mas a compreensão de como estes genes podem causar infertilidade masculina é ainda escassa^{25,27}. O presente projeto incidiu predominantemente na sequenciação de oito genes associados a PCD (*DNAH5*, *DNAH11*, *DNAI1*, *DNAI2*, *NME8*, *DNAAF2*, *RSPH4A* e *RSPH9*), e na identificação e caracterização das variantes presentes. No total, foram analisadas 57 variantes em 47 homens inférteis, 14 das quais possivelmente relacionadas com a patologia e o fenótipo em estudo (Tabela 3.5). Assim, por último, resta proceder à avaliação de que forma cada variante se correlaciona com a patologia em causa, de acordo com a informação recolhida quanto à própria variante, ao gene e à classificação do espermograma do respetivo doente.

Nesta avaliação é necessário ter-se em conta que os 8 genes em estudo apresentam transmissão autossómica recessiva. Visto que as 14 variantes de interesse foram identificadas em heterozigotia, nenhuma poderá ser, por si só, a causa dos sintomas identificados. Nenhuma das 12 variantes VUS/PP/P foi identificada em 2 ou mais doentes, e 8 destes apresentavam apenas 1 variante de interesse. Neste último conjunto incluem-se: as variantes *missense* *DNAH5*:c.6748G>A, *DNAH11*:c.1084G>A e *NME8*:c.1169G>C, de patogenicidade desconhecida devido a falta de dados; as variantes com possível impacto no *splicing* *NME8*:c.455-5dup (VUS, por potencial quebra do local aceitador de *splicing* e falta de informação), *DNAH5*:c.12465A>G (VUS, por possível novo local aceitador de *splicing* e informação contraditória) e *DNAI1*:c.400G>C (PP, com prevista quebra do local aceitador de *splicing* e perda de função da proteína); e ainda as duas variantes *nonsense* *NME8*:c.1277T>A (PP) e *DNAH5*:c.13486C>T (P), já descritas na literatura, que originam um codão de terminação da tradução prematuro e, presumivelmente, uma proteína truncada, com perda de função da mesma^{48,94,95}. Como indicado, devido a todas estas variantes se encontrarem em heterozigotia e não ter sido identificada mais nenhuma alteração que possa explicar a patologia nas respetivas 8 amostras, não é possível considerá-las como a causa da característica espermatogénica diagnosticada.

As variantes *DNAH11*:c.8990G>A e *DNAH11*:c.10739G>A, identificadas aparentemente em heterozigotia composta nas amostras PCip11 e PCip41 e classificadas, respetivamente, como provavelmente benigna e como benigna, encontram-se entre as 14 variantes de interesse (Tabela 3.5) devido a informação contraditória na literatura disponível em relação às classificações definidas neste projeto. Boon *et al.* (2014) identificou este par de variantes em dois doentes com PCD, diagnosticados com *situs solitus* (i.e., posicionamento normal de órgãos no tórax e abdómen) e ultraestrutura ciliar anormal, e classificou-as como provavelmente patogénicas somente por não terem sido reportadas na literatura disponível até à altura e por terem sido as únicas variantes identificadas nos dois indivíduos estudados⁹³. Este artigo não representa, assim, razão suficiente para se considerar a classificação atribuída às duas variantes em causa e as apontar como responsáveis pela oligoastenozoospermia diagnosticada nos doentes PCip11 e PCip41, tendo-se inclusivamente identificado as mesmas variantes num indivíduo saudável, não infértil, do grupo controlo (Ct061, homozigótico para *DNAH11*:c.8990G>A e heterozigótico para *DNAH11*:c.10739G>A), e a variante *DNAH11*:c.8990G>A em homozigotia no indivíduo Ct063, também não infértil.

Foram ainda detetadas 4 variantes *missense* raras de particular interesse, todas presentes em genes codificantes para proteínas associadas aos braços de dineína externos. As variantes *DNAH11*:c.2909A>G e *DNAH11*:c.9935A>T foram identificadas na amostra PCip03 (doente com oligoastenozoospermia) e, como tal, representam um possível caso de heterozigotia composta. Ao nível dos aminoácidos, verifica-se uma alteração de tirosina para cisteína e de ácido aspártico para valina, respetivamente. Ambos os aminoácidos constitutivos, da sequência de referência, são altamente conservados entre espécies, e é prevista uma diferença físico-química alta entre estes e os aminoácidos resultantes da sequência alterada. Observou-se ainda que parte dos preditores de patogenicidade de variantes *missense* classificam estas duas variantes como deletérias, ou equivalente, nomeadamente o *SIFT* para ambas. Em sentido inverso, as variantes *DNAH5*:c.3614C>T e *DNAAF2*:c.2239G>A, presentes na amostra PCip47 (doente com oligoastenoteratozoospermia) em dupla heterozigotia, são classificadas como benignas, ou equivalente, pela maioria dos preditores, incluindo o MutationTaster para ambas. A primeira variante resulta na substituição de alanina, pouco conservada, por valina, e a segunda leva à troca de ácido glutâmico, moderadamente conservado, por lisina, sendo previstas diferenças físico-químicas baixas entre os dois aminoácidos em ambos os casos. Assim, aparenta ser mais provável que o par de variantes heterozigóticas *DNAH11*:c.2909A>G e *DNAH11*:c.9935A>T possa afetar a estrutura e função da proteína negativamente (variantes hipomórficas), e possivelmente perturbar a motilidade do espermatozoide, do que as variantes *DNAH5*:c.3614C>T e *DNAAF2*:c.2239G>A. Contudo, devido à escassa informação disponível para estas variantes, e à consequente ausência de critérios para a sua classificação de acordo com as diretrizes da ACMG-AMP, não é possível estabelecer uma correlação causal segura entre os genótipos dos doentes com as variantes *missense* em causa e os respetivos fenótipos, pelo que o mais adequado é considerar as variantes de patogenicidade desconhecida.

Embora os resultados obtidos não permitam tirar conclusões definitivas quanto à causa dos vários casos de infertilidade masculina em análise, existem ainda algumas possibilidades que devem ser consideradas.

Após uma extensa análise de variantes nos genes *DNAI1*, *DNAH5* e *DNAH11* em homens com astenozoospermia não sindrómica, e estudos familiares associados, Zucarello *et al.* (2008) manifestou a hipótese de certas variantes heterozigóticas em genes associados a ciliopatias poderem afetar de forma menos grave a atividade ciliar e causar infertilidade masculina como um sintoma isolado, devido a haploinsuficiência ou efeito dominante negativo, por exemplo ⁹⁶. Contudo, no presente projeto, a análise das amostras controlo permitiu identificar a variante VUS *DNAH11*:c.9935A>T em 4 indivíduos não inférteis (Ct025, 049, 060 e 086), a variante VUS *NME8*:c.1169G>C noutros dois (Ct013 e Ct077) e, particularmente, a variante provavelmente patogénica *DNAI1*:c.400G>C num último indivíduo não infértil (Ct084), todas em heterozigotia.

Em sentido inverso, estudos recentes observaram baixos valores de expressão de *DNAH5*, *DNAH11* e *RSPH4A* nos testículos, e concluíram que estas proteínas poderão não ser essenciais para a fertilidade masculina ^{43,45}. Estes resultados, explicados anteriormente nas secções 1.2.1.1. (*DNAH5* e *DNAH11*, *DNAI1* e *DNAI2*, e *NME8*, componentes dos braços de dineína externos) e 1.2.1.3. (*RSPH4A* e *RSPH9*, constituintes dos raios radiais), sugerem que uma parte das variantes de interesse identificadas no presente projeto poderão não estar associadas ao sintoma em estudo, incluindo o possível caso de heterozigotia composta (amostra PCip03) caracterizado pela substituição de dois aminoácidos altamente conservados. Contudo, a descoberta de um segundo mecanismo através do qual PCD causa infertilidade masculina, associado aos ductos eferentes, poderá explicar estes e outros casos resultantes de variantes patogénicas nos genes *DNAH5*, *DNAH11* e *RSPH4A*. Os ductos eferentes são túbulos integrantes do sistema reprodutor masculino, responsáveis pelo transporte dos espermatozoides dos testículos (local de

produção) para os epidídimos (onde são armazenados), e que desempenham um papel importante no aumento da concentração destas células através da reabsorção de fluido seminífero luminal por células não ciliadas do epitélio^{97,98}. Adicionalmente, o epitélio destes túbulos inclui células multiciliadas que, através do batimento ciliar, mantém os espermatozoides em suspensão, o que promove a reabsorção do fluido e previne a agregação de espermatozoides e consequente bloqueio dos ductos⁹⁹. Em 2021, Aprea *et al.* constatou que defeitos na motilidade dos cílios de dutos eferentes devido à deficiência de *DNAH5* também causam a agregação de espermatozoides e bloqueio destes túbulos¹⁰⁰. Assim, comprovou a existência deste segundo mecanismo através do qual PCD pode causar infertilidade masculina, não dependente de alterações no flagelo de espermatozoides. É ainda importante constatar que estes casos estão invariavelmente associados a oligozoospermia (i.e., reduzido número de espermatozoides no ejaculado), mas a análise à qualidade do esperma não identificou alterações ao nível da motilidade e morfologia dos espermatozoides¹⁰⁰. Deste modo, este último estudo reforça a hipótese de que os ODA do flagelo de espermatozoides apresentam, de fato, composição diferente de outros cílios móveis, e sugere que as variantes identificadas no presente projeto nos genes *DNAH5* (incluindo a variante patogénica *DNAH5:c.13486C>T*) e *DNAH11* (incluindo o possível caso de heterozigotia composta de *DNAH11:c.2909A>G* e *DNAH11:c.9935A>T*) poderão não estar diretamente associadas ao sintoma em estudo, dado que a astenozoospermia é comum aos 48 doentes inférteis analisados. Ainda assim, não se pode desprezar a possibilidade de existir um outro fator que explique o diagnóstico de astenozoospermia nestes e noutros casos, nomeadamente a presença de uma variante patogénica num gene associado a PCD não pertencente ao painel definido. Em sentido inverso, a identificação de uma variante patogénica nos genes *DNAI1*, *NME8* e *DNAAF2*, entre outros, responsável por alterações na ultraestrutura do flagelo identificadas por TEM, não se traduzirá necessariamente na alteração da morfologia do flagelo do espermatozoide (teratozoospermia) após visualização por microscopia ótica¹⁰¹. Deste modo, não é possível excluir que um possível efeito dominante negativo, a dupla heterozigotia e a heterozigotia composta possam perturbar a presença de espermatozoides no ejaculado dos homens em causa.

Como indicado no parágrafo anterior, não se pode deixar de considerar a possibilidade de que não foram obtidos resultados conclusivos pelo fato de a ou as variantes responsáveis por cada caso de infertilidade diagnosticada poderem estar presentes em genes não estudados no presente projeto. Atualmente, a patologia discinesia ciliar primária está associada a mais de 45 genes, o que representa aproximadamente dois terços dos casos diagnosticados com a doença^{25,27}. Entre os vários genes que não foram incluídos no painel de genes em estudo encontram-se fatores de montagem de dineína no citoplasma e constituintes dos raios radiais, bem como componentes do complexo regulatório de nexina-dineína, do complexo de ancoragem de ODA e do par central do axonema, entre outros³⁷. Em particular, destaca-se *CCDC39* (*coiled-coil domain 39 molecular ruler complex subunit*) e *CCDC40* (*coiled-coil domain 40 molecular ruler complex subunit*), dois genes frequentemente identificados como causa de PCD com infertilidade masculina, em casos caracterizados por desorganização microtubular e defeitos em IDA e nos complexos regulatórios de nexina-dineína^{25,46,102,103}. Adicionalmente, os genes *DNAH8* e *DNAH17*, apresentados anteriormente devido à elevada expressão nos testículos contrastante com *DNAH5* e *DNAH11* (secção 1.2.1.1. *DNAH5* e *DNAH11*, *DNAI1* e *DNAI2*, e *NME8*, componentes dos braços de dineína externos), não devem ser desvalorizados uma vez que variantes patogénicas nestes genes causam infertilidade masculina isolada, caracterizada por defeitos na motilidade dos espermatozoides e na morfologia dos flagelos, que não são identificados em cílios móveis de outros grupos de células^{43,44}. Este fenótipo específico de espermatozoides é denominado de MMAF (anormalidades morfológicas múltiplas do flagelo, do inglês *multiple morphological abnormalities of the flagella*), e é comum a vários genes do axonema não associados a PCD, incluindo *DNAH1* (*dynein axonemal heavy chain 1*), que codifica uma cadeia pesada de IDA e é fundamental na organização do axonema, e *CFAP43* (*cilia and flagella associated protein 43*) e *CFAP44* (*cilia and flagella associated protein 44*), dois genes

envolvidos na organização e função do axonema do flagelo ^{104–107}. Atualmente, a informação sobre MMAF, os genes associados e a relação com PCD é escassa, mas seria certamente interessante ver que resultados seriam obtidos com o estudo destes genes nos doentes analisados, por MMAF incidir predominantemente em doentes com astenozoospermia não síndrômica.

Conclusão

Numa análise global, o procedimento proposto foi concluído com sucesso, tendo-se implementado e validado a aplicação de NGS para a caracterização molecular de um painel de 8 genes relevantes na patologia discinesia ciliar primária, num conjunto significativo de 48 doentes inférteis diagnosticados com astenozoospermia não síndrômica, em combinação com sequenciação de Sanger.

No total foram identificadas 57 variantes, 14 das quais potencialmente associadas à característica fenotípica em estudo. Dado que não foi possível estabelecer correlações genótipo/fenótipo definitivas de acordo com o padrão de hereditariedade dos genes em causa, os resultados obtidos permitem inferir, à luz dos conhecimentos atuais, que os genes analisados não devem constituir a principal base molecular da astenozoospermia quando presente em homens sem outras manifestações fenotípicas de PCD.

No entanto, é relevante destacar que só através da realização de estudos funcionais e estudos familiares futuros se poderá classificar devidamente estas variantes quanto à sua patogenicidade e compreender o seu impacto na motilidade dos cílios. Ainda assim, os resultados obtidos neste projeto não deixam de contribuir para o conhecimento científico em torno desta doença rara e da sua etiologia, e estudos similares permitirão continuar a desenvolver a compreensão e a capacidade de diagnóstico desta complexa doença.

5 | Referências Bibliográficas

- 1 Waters AM, Beales PL. Ciliopathies: An expanding disease spectrum. *Pediatric Nephrology* 2011; **26**: 1039–1056. doi:10.1007/s00467-010-1731-7
- 2 Oud MM, Lamers IJ, Arts HH. Ciliopathies: Genetics in Pediatric Medicine. *J Pediatr Genet* 2016; **6**: 18–29. doi:10.1055/s-0036-1593841
- 3 Fliegauf M, Benzing T, Omran H. When cilia go bad: Cilia defects and ciliopathies. *Nat Rev Mol Cell Biol* 2007; **8**: 880–893. doi:10.1038/nrm2278
- 4 Mitchell DR. Evolution of Cilia. *Cold Spring Harb Perspect Biol* 2017; **9**. doi:10.1101/cshperspect.a028290
- 5 Kozminski KG, Johnson KA, Forscher P, Rosenbaum JL. A motility in the eukaryotic flagellum unrelated to flagellar beating. *Proceedings of the National Academy of Sciences USA* 1993; **90**: 5519–5523. doi:10.1073/pnas.90.12.5519
- 6 Reiter JF, Leroux MR. Genes and molecular pathways underpinning ciliopathies. *Nat Rev Mol Cell Biol* 2017; **18**: 533–547. doi:10.1038/nrm.2017.60
- 7 Carvalho-Santos Z, Azimzadeh J, Pereira-Leal JB, Bettencourt-Dias M. Evolution: Tracing the origins of centrioles, cilia, and flagella. *Journal of Cell Biology* 2011; **194**: 165–175. doi:10.1083/jcb.201011152
- 8 Reiter JF, Blacque OE, Leroux MR. The base of the cilium: roles for transition fibres and the transition zone in ciliary formation, maintenance and compartmentalization. *EMBO Rep* 2012; **13**: 608–618. doi:10.1038/embor.2012.73
- 9 Ishikawa H, Marshall WF. Ciliogenesis: building the cell’s antenna. *Nat Rev Mol Cell Biol* 2011; **12**: 222–234. doi:10.1038/nrm3085
- 10 Avasthi P, Marshall WF. Stages of ciliogenesis and regulation of ciliary length. *Differentiation* 2012; **83**. doi:10.1016/j.diff.2011.11.015.
- 11 Sorokin SP. Reconstructions of centriole formation and ciliogenesis in mammalian lungs. *J Cell Sci* 1968; **3**: 207. doi:10.1242/jcs.3.2.207
- 12 Spassky N, Meunier A. The development and functions of multiciliated epithelia. *Nat Rev Mol Cell Biol* 2017; **18**: 423–436. doi:10.1038/nrm.2017.21
- 13 Wheway G, Nazlamova L, Hancock JT. Signaling through the primary cilium. *Front Cell Dev Biol* 2018; **6**. doi:10.3389/fcell.2018.00008
- 14 Oda T, Yanagisawa H, Yagi T, Kikkawa M. Mechanosignaling between central apparatus and radial spokes controls axonemal dynein activity. *Journal of Cell Biology* 2014; **204**: 807–819. doi:10.1083/jcb.201312014
- 15 Heuser T, Raytchev M, Krell J, Porter ME, Nicastro D. The dynein regulatory complex is the nexin link and a major regulatory node in cilia and flagella. *Journal of Cell Biology* 2009; **187**: 921–933. doi:10.1083/jcb.200908067
- 16 Lin J, Nicastro D. Asymmetric distribution and spatial switching of dynein activity generates ciliary motility. *Science* 2018; **360**. doi:10.1126/science.aar1968.

- 17 Grossman-Haham I *et al.* Structure of the radial spoke head and insights into its role in mechanoregulation of ciliary beating. *Nat Struct Mol Biol* 2021; **28**: 20–28. doi:10.1038/s41594-020-00519-9
- 18 Gui L *et al.* Scaffold subunits support associated subunit assembly in the Chlamydomonas ciliary nexin-dynein regulatory complex. *Proc Natl Acad Sci U S A* 2019; **116**: 23152–23162. doi:10.1073/pnas.1910960116
- 19 Antony D, Brunner HG, Schmidts M. Ciliary dyneins and dynein related ciliopathies. *Cells* 2021; **10**. doi:10.3390/cells10081885. doi:10.3390/cells10081885
- 20 Bustamante-Marin XM, Ostrowski LE. Cilia and mucociliary clearance. *Cold Spring Harb Perspect Biol* 2017; **9**. doi:10.1101/cshperspect.a028241
- 21 Ibañez-Tallon I *et al.* Dysfunction of axonemal dynein heavy chain Mdnah5 inhibits ependymal flow and reveals a novel mechanism for hydrocephalus formation. *Hum Mol Genet* 2004; **13**: 2133–2141. doi:10.1093/hmg/ddh219
- 22 Lyons RA, Saridogan E, Djahanbakhch O. The reproductive significance of human Fallopian tube cilia. *Hum Reprod Update* 2006; **12**: 363–372. doi:10.1093/humupd/dml012
- 23 Nonaka S *et al.* Randomization of Left-Right Asymmetry due to Loss of Nodal Cilia Generating Leftward Flow of Extraembryonic Fluid in Mice Lacking KIF3B Motor Protein. *Cell* 1998; **95**: 829–837. doi:10.1016/s0092-8674(00)81705-5
- 24 Lehti MS, Sironen A. Formation and function of sperm tail structures in association with sperm motility defects. *Biol Reprod* 2017; **97**: 522–536. doi:10.1093/biolre/iox096
- 25 Zariwala MA, Knowles MR, Leigh M. Primary Ciliary Dyskinesia. In: Adam MP *et al.* (eds). *GeneReviews*. University of Washington, Seattle, 2019.
- 26 Goutaki M *et al.* Clinical manifestations in primary ciliary dyskinesia: systematic review and meta-analysis. *European Respiratory Journal* 2016; **48**: 1081–1095. doi:10.1183/13993003.00736-2016
- 27 Knowles M, Zariwala M. Primary ciliary dyskinesia. Orphanet. 2019. <https://www.orpha.net/en/disease/detail/244?name=Primary%20ciliary%20diskinesia&mode=name> (accessed 18 Feb2024).
- 28 Leigh MW *et al.* Clinical features and associated likelihood of primary ciliary dyskinesia in children and adolescents. *Ann Am Thorac Soc* 2016; **13**: 1305–1313. doi:10.1513/AnnalsATS.201511-748OC
- 29 Lucas JS *et al.* European Respiratory Society guidelines for the diagnosis of primary ciliary dyskinesia. *European Respiratory Journal* 2017; **49**. doi:10.1183/13993003.01090-2016
- 30 Shapiro AJ *et al.* Diagnosis of primary ciliary dyskinesia: An official American thoracic society clinical practice guideline. *Am J Respir Crit Care Med* 2018; **197**: e24–e39. doi:10.1164/rccm.201805-0819ST
- 31 Lucas JS *et al.* Clinical care of children with primary ciliary dyskinesia. *Expert Rev Respir Med* 2017; **11**: 779–790. doi:10.1080/17476348.2017.1360770
- 32 Paff T, Omran H, Nielsen KG, Haarman EG. Current and future treatments in primary ciliary dyskinesia. *Int J Mol Sci* 2021; **22**. doi:10.3390/ijms22189834

- 33 Kobbernagel HE *et al.* Efficacy and safety of azithromycin maintenance therapy in primary ciliary dyskinesia (BESTCILIA): a multicentre, double-blind, randomised, placebo-controlled phase 3 trial. *Lancet Respir Med* 2020; **8**: 493–505. doi:10.1016/S2213-2600(20)30058-8
- 34 Lai M *et al.* Gene editing of DNAH11 restores normal cilia motility in primary ciliary dyskinesia. *J Med Genet* 2016; **53**: 242–249. doi:10.1136/jmedgenet-2015-103539
- 35 Paff T *et al.* Mutations in PIH1D3 Cause X-Linked Primary Ciliary Dyskinesia with Outer and Inner Dynein Arm Defects. *Am J Hum Genet* 2017; **100**: 160–168. doi:10.1016/j.ajhg.2016.11.019
- 36 Shapiro AJ *et al.* Autosomal dominant variants in FOXJ1 causing primary ciliary dyskinesia in two patients with obstructive hydrocephalus. *Mol Genet Genomic Med* 2021; **9**. doi:10.1002/mgg3.1726
- 37 Wallmeier J *et al.* Motile ciliopathies. *Nat Rev Dis Primers* 2020; **6**: 77. doi:10.1038/s41572-020-0209-6
- 38 Fabczak H, Osinka A. Role of the novel Hsp90 co-chaperones in dynein arms' preassembly. *Int J Mol Sci* 2019; **20**: 6174. doi:10.3390/ijms20246174
- 39 Oda T, Abe T, Yanagisawa H, Kikkawa M. Structure and function of outer dynein arm intermediate and light chain complex. *Mol Biol Cell* 2016; **27**: 1051–1059. doi:10.1091/mbc.E15-10-0723
- 40 Fliegauf M *et al.* Mislocalization of DNAH5 and DNAH9 in respiratory cells from patients with primary ciliary dyskinesia. *Am J Respir Crit Care Med* 2005; **171**: 1343–1349. doi:10.1164/rccm.200411-1583OC
- 41 Dougherty GW *et al.* DNAH11 localization in the proximal region of respiratory cilia defines distinct outer dynein arm complexes. *Am J Respir Cell Mol Biol* 2016; **55**: 213–224. doi:10.1165/rcmb.2015-0353OC
- 42 Shoemark A *et al.* Primary ciliary dyskinesia with normal ultrastructure: Three-dimensional tomography detects absence of DNAH11. *European Respiratory Journal* 2018; **51**. doi:10.1183/13993003.01809-2017
- 43 Whitfield M *et al.* Mutations in DNAH17, Encoding a Sperm-Specific Axonemal Outer Dynein Arm Heavy Chain, Cause Isolated Male Infertility Due to Asthenozoospermia. *Am J Hum Genet* 2019; **105**: 198–212. doi:10.1016/j.ajhg.2019.04.015
- 44 Liu C *et al.* Bi-allelic DNAH8 Variants Lead to Multiple Morphological Abnormalities of the Sperm Flagella and Primary Male Infertility. *Am J Hum Genet* 2020; **107**: 330–341. doi:10.1016/j.ajhg.2020.06.004
- 45 Fagerberg L *et al.* Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & Cellular Proteomics* 2014; **13**: 397–406. doi:10.1074/mcp.M113.035600
- 46 Vanaken GJ *et al.* Infertility in an adult cohort with primary ciliary dyskinesia: Phenotype-gene association. *European Respiratory Journal* 2017; **50**. doi:10.1183/13993003.00314-2017
- 47 Zhu D *et al.* Association of DNAH11 gene polymorphisms with asthenozoospermia in Northeast Chinese patients. *Biosci Rep* 2019; **39**. doi:10.1042/BSR20181450

- 48 Duriez B *et al.* A common variant in combination with a nonsense mutation in a member of the thioredoxin family causes primary ciliary dyskinesia. *Proceedings of the National Academy of Sciences* 2007; **104**: 3336–3341. doi:10.1073/pnas.0611405104
- 49 Sadek CM, Damdimopoulos AE, Pelto-Huikko M, Gustafsson JÅ, Spyrou G, Miranda-Vizuete A. Sptrx-2, a fusion protein composed of one thioredoxin and three tandemly repeated NDP-kinase domains is expressed in human testis germ cells. *Genes to Cells* 2001; **6**: 1077–1090. doi:10.1046/j.1365-2443.2001.00484.x
- 50 Omran H *et al.* Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins. *Nature* 2008; **456**: 611–616. doi:10.1038/nature07471
- 51 Fowkes ME, Mitchell DR. The Role of Preassembled Cytoplasmic Complexes in Assembly of Flagellar Dynein Subunits. *Mol Biol Cell* 1998; **9**: 2337–2347. doi:10.1091/mbc.9.9.2337
- 52 Ahmed NT, Gao C, Lucker BF, Cole DG, Mitchell DR. ODA16 aids axonemal outer row dynein assembly through an interaction with the intraflagellar transport machinery. *Journal of Cell Biology* 2008; **183**: 313–322. doi:10.1083/jcb.200802025
- 53 Aprea I *et al.* Defects in the cytoplasmic assembly of axonemal dynein arms cause morphological abnormalities and dysmotility in sperm cells leading to male infertility. *PLoS Genet* 2021; **17**. doi:10.1371/journal.pgen.1009306
- 54 Castleman VH *et al.* Mutations in radial spoke head protein genes RSPH9 and RSPH4A cause primary ciliary dyskinesia with central-microtubular-pair abnormalities. *Am J Hum Genet* 2008; **84**: 197–209. doi:10.1016/j.ajhg.2009.01.011
- 55 Frommer A *et al.* Immunofluorescence analysis and diagnosis of primary ciliary dyskinesia with radial spoke defects. *Am J Respir Cell Mol Biol* 2015; **53**: 563–573. doi:10.1165/rcmb.2014-0483OC
- 56 Aprea I *et al.* Pathogenic gene variants in CCDC39, CCDC40, RSPH1, RSPH9, HYDIN, and SPEF2 cause defects of sperm flagella composition and male infertility. *Front Genet* 2023; **14**. doi:10.3389/fgene.2023.1117821
- 57 Abbasi F *et al.* RSPH6A is required for sperm flagellum formation and male fertility in mice. *J Cell Sci* 2018; **131**. doi:10.1242/jcs.221648
- 58 Wang L *et al.* Novel RSPH4A Variants Associated With Primary Ciliary Dyskinesia–Related Infertility in Three Chinese Families. *Front Genet* 2022; **13**. doi:10.3389/fgene.2022.922287
- 59 Infertility prevalence estimates, 1990–2021. Geneva: World Health Organization. 2023.
- 60 Pandruvada S *et al.* Lack of trusted diagnostic tools for undetermined male infertility. *J Assist Reprod Genet* 2021; **38**: 265–276. doi:10.1007/s10815-020-02037-5
- 61 Infertility. Geneva: World Health Organization. 2024. <https://www.who.int/news-room/fact-sheets/detail/infertility> (accessed 20 Jun 2024).
- 62 Agarwal A *et al.* Male infertility. *The Lancet* 2021; **397**: 319–333. doi:10.1016/S0140-6736(20)32667-2
- 63 WHO laboratory manual for the examination and processing of human semen, sixth edition. Geneva: World Health Organization. 2021.

- 64 Schlegel PN *et al.* Diagnosis and Treatment of Infertility in Men: AUA/ASRM Guideline Part I. *Journal of Urology* 2021; **205**: 36–43. doi:10.1097/JU.0000000000001521
- 65 Schlegel PN *et al.* Diagnosis and Treatment of Infertility in Men: AUA/ASRM Guideline Part II. *Journal of Urology* 2021; **205**: 44–51. doi:10.1097/JU.0000000000001520
- 66 Jungwirth A, Diemer T, Kopa Z, Krausz C, Minhas S, Tournaye H. EAU Guidelines on Male Infertility. European Association of Urology. 2019.
- 67 Joseph S, Mahale SD. Male Infertility Knowledgebase: Decoding the genetic and disease landscape. *Database* 2021; **2021**: 1–14. doi:10.1093/database/baab049
- 68 Cooper TG *et al.* World Health Organization reference values for human semen characteristics. *Hum Reprod Update* 2009; **16**: 231–245. doi:10.1093/humupd/dmp048
- 69 Moro AM *et al.* What is asthenozoospermia? – definition, causes and treatment. inviTRA. 2022. <https://www.invitro.com/en/asthenozoospermia/> (accessed 18 Feb2024).
- 70 Pabinger S *et al.* A survey of tools for variant analysis of next-generation genome sequencing data. *Brief Bioinform* 2014; **15**: 256–278. doi:10.1093/bib/bbs086
- 71 Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint* 2013. doi:10.48550/arXiv.1303.3997
- 72 Depristo MA *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; **43**: 491–498. doi:10.1038/ng.806
- 73 Desvignes JP *et al.* VarAFT: A variant annotation and filtration system for human next generation sequencing data. *Nucleic Acids Res* 2018; **46**: W545–W553. doi:10.1093/nar/gky471
- 74 Robinson JT *et al.* Integrative genomics viewer. *Nat Biotechnol* 2011; **29**: 24–26. doi:10.1038/nbt.1754
- 75 Li H *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079. doi:10.1093/bioinformatics/btp352
- 76 Stenson PD, Mort M, Ball E V., Shaw K, Phillips AD, Cooper DN. The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine. *Hum Genet* 2014; **133**: 1–9. doi:10.1007/s00439-013-1358-4
- 77 Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol* 1990; **215**: 403–410. doi:10.1016/S0022-2836(05)80360-2
- 78 Karczewski KJ *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature* 2020; **581**: 434–443. doi:10.1038/s41586-020-2308-7
- 79 Auton A *et al.* A global reference for human genetic variation. *Nature* 2015; **526**: 68–74. doi:10.1038/nature15393
- 80 Taliun D *et al.* Sequencing of 53,831 diverse genomes from the NHLBI TOPMed Program. *Nature* 2021; **590**: 290–299. doi:10.1038/s41586-021-03205-y

- 81 Tavgitian S V. *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet* 2006; **43**: 295–305. doi:10.1136/jmg.2005.033878
- 82 Adzhubei IA *et al.* A method and server for predicting damaging missense mutations. *Nat Methods* 2010; **7**: 248–249. doi:10.1038/nmeth0410-248
- 83 Ng PC, Henikoff S. Predicting deleterious amino acid substitutions. *Genome Res* 2001; **11**: 863–874. doi:10.1101/gr.176601
- 84 Schwarz JM, Cooper DN, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods* 2014; **11**: 361–362. doi:10.1038/nmeth.2890
- 85 Desmet FO, Hamroun D, Lalande M, Collod-B  roud G, Claustres M, B  roud C. Human Splicing Finder: An online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009; **37**. doi:10.1093/nar/gkp215
- 86 Shapiro MB, Senapathy P. RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acids Res* 1987; **15**: 7155–7174. doi:10.1093/nar/15.17.7155
- 87 Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology* 2004; **11**: 377–394. doi:10.1089/1066527041410418
- 88 Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. *Journal of Computational Biology* 1997; **4**: 311–323. doi:10.1089/cmb.1997.4.311
- 89 Pertea M. GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Res* 2001; **29**: 1185–1190. doi:10.1093/nar/29.5.1185
- 90 Landrum MJ *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res* 2014; **42**: D980–D985. doi:10.1093/nar/gkt1113
- 91 Kopanos C *et al.* VarSome: the human genomic variant search engine. *Bioinformatics* 2019; **35**: 1978–1980. doi:10.1093/bioinformatics/bty897
- 92 Richards S *et al.* Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine* 2015; **17**: 405–424. doi:10.1038/gim.2015.30
- 93 Boon M *et al.* Primary ciliary dyskinesia: critical evaluation of clinical symptoms and diagnosis in patients with normal and abnormal ultrastructure. *Orphanet J Rare Dis* 2014; **9**: 11. doi:10.1186/1750-1172-9-11
- 94 Hornef N *et al.* DNAH5 mutations are a common cause of primary ciliary dyskinesia with outer dynein arm defects. *Am J Respir Crit Care Med* 2006; **174**: 120–126. doi:10.1164/rccm.200601-084OC
- 95 Davis SD *et al.* Primary ciliary dyskinesia: longitudinal study of lung disease by ultrastructure defect and genotype. *Am J Respir Crit Care Med* 2019; **199**: 190–198. doi:10.1164/rccm.201803-0548OC

- 96 Zuccarello D *et al.* Mutations in dynein genes in patients affected by isolated non-syndromic asthenozoospermia. *Human Reproduction* 2008; **23**: 1957–1962. doi:10.1093/humrep/den193
- 97 Clulow J, Hansen L, Jones R. In vivo microperfusion of the ductuli efferentes testis of the rat: flow dependence of fluid reabsorption. *Exp Physiol* 1996; **81**: 633–644. doi:10.1113/expphysiol.1996.sp003964
- 98 Hess RA. Small tubules, surprising discoveries: from efferent ductules in the turkey to the discovery that estrogen receptor alpha is essential for fertility in the male. *Anim Reprod* 2015; **12**: 7–23.
- 99 Yuan S *et al.* Motile cilia of the male reproductive system require miR-34/miR-449 for development and function to generate luminal turbulence. *Proceedings of the National Academy of Sciences* 2019; **116**: 3584–3593. doi:10.1073/pnas.1817018116
- 100 Aprea I *et al.* Motility of efferent duct cilia aids passage of sperm cells through the male reproductive system. *Mol Hum Reprod* 2021; **27**. doi:10.1093/molehr/gaab009
- 101 Coutton C, Escoffier J, Martinez G, Arnoult C, Ray PF. Teratozoospermia: spotlight on the main genetic actors in the human. *Hum Reprod Update* 2015; **21**: 455–485. doi:10.1093/humupd/dmv020
- 102 Merveille A-C *et al.* CCDC39 is required for assembly of inner dynein arms and the dynein regulatory complex and for normal ciliary motility in humans and dogs. *Nat Genet* 2011; **43**: 72–78. doi:10.1038/ng.726
- 103 Becker-Heck A *et al.* The coiled-coil domain containing protein CCDC40 is essential for motile cilia function and left-right axis formation. *Nat Genet* 2011; **43**: 79–84. doi:10.1038/ng.727
- 104 Wang W-L, Tu C-F, Tan Y-Q. Insight on multiple morphological abnormalities of sperm flagella in male infertility: what is new? *Asian J Androl* 2020; **22**: 236. doi:10.4103/aja.aja_53_19
- 105 Ben Khelifa M *et al.* Mutations in DNAH1, which encodes an inner arm heavy chain dynein, lead to male infertility from multiple morphological abnormalities of the sperm flagella. *The American Journal of Human Genetics* 2014; **94**: 95–104. doi:10.1016/j.ajhg.2013.11.017
- 106 Tang S *et al.* Biallelic mutations in CFAP43 and CFAP44 cause male infertility with multiple morphological abnormalities of the sperm flagella. *The American Journal of Human Genetics* 2017; **100**: 854–864. doi:10.1016/j.ajhg.2017.04.012
- 107 Coutton C *et al.* Mutations in CFAP43 and CFAP44 cause male infertility and flagellum defects in Trypanosoma and human. *Nat Commun* 2018; **9**: 686. doi:10.1038/s41467-017-02792-7

6 | Anexos

Este projeto foi apresentado em forma de poster no congresso da “26ª Reunião Anual da Sociedade Portuguesa de Genética Humana”, que se realizou nos dias 17, 18 e 19 de novembro de 2022, em Coimbra.

Tabela S1 Regiões sem cobertura suficiente (20x), nos oito genes em estudo.

Amostra	Gene	Exão	Coordenadas	Sequenciar?
Todas	<i>DNAAF2</i>	1	50 101 694 a 50 101 917	Sim
Ct084		2	50 094 705 a 50 094 706; 50 094 881 a 50 094 898	Não
Ct106, 109 e 111, PCip29	<i>DNAH5</i>	12	13 911 470 a 13 911 504	Não
Ct057		14	13 901 484 a 13 901 518	Não
Ct084, 109 e 111		17	13 891 124 a 13 891 167	Não
69 amostras Ct, 28 amostras PCip		18	13 886 197 a 13 886 263	Não
Ct084, PCip30, 35 e 36		19	13 885 165 a 13 885 362	Não
Ct057		22	13 876 945 a 13 876 951	Não
Ct109		24	13 870 893 a 13 870 903	Não
Todas		25	13 867 858 a 13 867 888	Sim
Ct084		33	13 841 835 a 13 842 038	Não
Ct014 e 084		37	13 830 110 a 13 830 347	Não
Todas		41	13 820 632 a 13 820 633	Não
Ct066 e 084, PCip16, 35 e 37		43	13 814 689 a 13 814 980	Não
Ct048		48	13 794 158 a 13 794 192	Não
Ct109		59	13 766 060 a 13 766 085; 13 766 281 a 13 766 313	Não
Todas		63	13 753 317 a 13 753 361	Sim
Ct084	<i>DNAH11</i>	66	13 737 336 a 13 737 390; 13 737 582 a 13 737 629	Não
Ct109		77	13 701 529 a 13 701 570	Não
6 amostras Ct, 9 amostras PCip		1	21 582 814 a 21 583 120	Sim
Ct048		2	21 584 599 a 21 584 792	Não
Ct084		6	21 603 779 a 21 603 812; 21 603 991 a 21 604 040	Não
13 amostras Ct, 12 amostras PCip	<i>DNAH11</i>	10	21 627 657 a 21 627 844	Sim
Ct084		13	21 630 511 a 21 630 558	Não
98 amostras Ct, 38 amostras PCip		18	21 640 089 a 21 641 067	Sim
Ct048 e 109, PCip23 e 29		27	21 677 343 a 21 677 345	Não

Ct057		28	21 678 696 a 21 678 708	Não
Todas		34	21 726 975 a 21 727 170	Sim
9 amostras Ct, 9 amostras PCip		35	21 730 358 a 21 730 524	Sim
Ct111		38	21 744 027 a 21 744 039; 21 744 215 a 21 744 271	Não
Ct106		40	21 747 292 a 21 747 295; 21 747 466 a 21 747 478	Não
Ct084		41	21 750 146 a 21 750 164; 21 750 333 a 21 750 346	Não
8 amostras Ct, 10 amostras PCip		45	21 765 404 a 21 765 550	Sim
Ct057		51	21 784 463 a 21 784 478; 21 784 667 a 21 784 706	Não
9 amostras Ct, 4 amostras PCip		55	21 805 021 a 21 805 232	Sim
8 amostras Ct, 8 amostras PCip		67	21 892 060 a 21 892 064	Não
Ct084, 109 e 111		77	21 932 038 a 21 932 103; 21 932 115 a 21 932 129	Não
Ct109	<i>DNAI1</i>	11	34 500 842 a 34 500 062	Não
Todas		20	34 520 790 a 34 520 805	Não
8 amostras Ct, 11 amostras PCip	<i>DNAI2</i>	6	72 287 275 a 72 287 297	Sim
4 amostras Ct, PCip30, 31 e 36		12	72 308 117 a 72 308 318	Não
Todas		13	72 310 394 a 72 310 408	Não
8 amostras Ct, PCip10 e 29	<i>NME8</i>	5	37 901 642 a 37 901 771	Sim
Ct084		13	37 928 024 a 37 928 055	Não
Ct111		14	37 934 043 a 37 934 072	Não
Ct084	<i>RSPH4A</i>	1	116 937 815 a 116 937 877	Não
107 amostras Ct, 43 amostras PCip		3	116 948 971 a 116 949 267	Sim
Todas		6	116 953 630 a 116 953 654	Não
Ct057 e 109, PCip30	<i>RSPH9</i>	1	43 612 954 a 43 613 087	Não
Todas		2	43 618 087 a 43 618 302	Sim
Ct057, 084, 109 e 111		3	43 623 274 a 43 623 314	Não
Ct057 e 109		4	43 624 289 a 43 624 308; 43 624 475 a 43 624 486	Não

Tabela S2 Variantes genéticas (n=45) identificadas no grupo de 48 indivíduos do sexo masculino inférteis, classificadas como benignas ou provavelmente benignas. As variantes foram descritas de acordo com as recomendações da HGVS ao nível do cDNA e da proteína. A classificação de cada variante teve por base informação disponibilizada pela aplicação Variant Interpreter, por plataformas e bases de dados *online*, e por preditores de patogenicidade *in silico*, como explicado na secção 2.9.2. Anotação e classificação de variantes, e foi realizada de acordo com as diretrizes da ACMG-AMP. A variante c.100_101delinsTT (*DNAH11*) foi identificada nas amostras PCip01, 03, 06, 07, 08, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 25, 26, 27, 29, 30, 31, 33, 34, 35, 37, 38, 39, 41, 43, 44, 45, 46, 48, e a variante c.1478_1479delinsCT (*NME8*) nas amostras PCip02, 03, 05, 06, 07, 10, 11, 12, 14, 15, 16, 17, 18, 21, 23, 24, 26, 31, 32, 33, 34, 35, 36, 37, 38, 42, 43, 44, 46.

Amostra PCip	Gene	HGVS cDNA	HGVS proteína	Tipo de variante	Frequência GnomAD (%)	Classificação
Várias (37)	<i>DNAH11</i>	c.100_101delinsTT	p.(Glu34Leu)	Missense	-	B
Várias (29)	<i>NME8</i>	c.1478_1479delinsCT	p.(Ile493Thr)	Missense	-	B
02, 24, 25, 31, 43	<i>RSPH4A</i>	c.731G>A	p.(Arg244His)	Missense	1,9328	B
08	<i>DNAI2</i>	c.1462C>T	p.(Leu488Phe)	Missense	0,2649	B
09	<i>DNAH11</i>	c.9102+8G>A	p.(?)	Intrónica	1,2176	B
11 e 41	<i>DNAH11</i>	c.10739G>A	p.(Arg3580His)	Missense	0,88611	B
13	<i>DNAH11</i>	c.13263G>C	p.(Pro4421=)	Sinónima	1,1387	B
16	<i>DNAH5</i>	c.2052+22C>A	p.(?)	Intrónica	0,76954	B
20 e 32	<i>DNAH5</i>	c.1647C>G	p.(Asn549Lys)	Missense	1,008	B
21	<i>DNAH11</i>	c.6041+18T>C	p.(?)	Intrónica	0,41968	B
22	<i>DNAH11</i>	c.9102+21_9102+22dup	p.(?)	Intrónica	0,28792	B
22	<i>DNAH11</i>	c.9102+23A>T	p.(?)	Intrónica	1,5147	B
22	<i>DNAH11</i>	c.10027-10T>C	p.(?)	Intrónica	0,15542	B
28	<i>DNAH11</i>	c.1535T>A	p.(Met512Lys)	Missense	0,16259	B
28	<i>DNAH11</i>	c.9098T>C	p.(Ile3033Thr)	Missense	0,51869	B
35	<i>DNAH5</i>	c.299G>C	p.(Gly100Ala)	Missense	0,09093	B
36	<i>DNAH5</i>	c.3987A>G	p.(Lys1329=)	Sinónima	1,0664	B
38	<i>DNAH5</i>	c.2053-23A>C	p.(?)	Intrónica	0,88959	B
38	<i>DNAH5</i>	c.2253C>A	p.(Asn751Lys)	Missense	0,90332	B
39	<i>NME8</i>	c.528+11T>C	p.(?)	Intrónica	0,47138	B
43 e 46	<i>RSPH4A</i>	c.1489G>A	p.(Val497Ile)	Missense	1,2565	B
02	<i>DNAI2</i>	c.1516C>A	p.(Arg506=)	Sinónima	0,04306	PB
09	<i>DNAH11</i>	c.7646-25G>A	p.(?)	Intrónica	0,32875	PB
10	<i>DNAH5</i>	c.5703T>C	p.(Asp1901=)	Sinónima	0,00120	PB
10	<i>DNAH11</i>	c.939C>T	p.(Ser313=)	Sinónima	0,48209	PB
11	<i>DNAH11</i>	c.6274-5T>C	p.(?)	Intrónica	0,03659	PB
11 e 41	<i>DNAH11</i>	c.8990G>A	p.(Arg2997Gln)	Missense	1,1237	PB
17	<i>DNAH11</i>	c.4282A>G	p.(Thr1428Ala)	Missense	0,33316	PB
17	<i>DNAH11</i>	c.10417C>T	p.(Leu3473=)	Sinónima	-	PB
19 e 24	<i>DNAH5</i>	c.624C>T	p.(Asn208=)	Sinónima	0,25068	PB
21	<i>DNAH11</i>	c.4920C>T	p.(Leu1640=)	Sinónima	0,00161	PB
22	<i>DNAH11</i>	c.11200A>C	p.(Lys3734Gln)	Missense	0,24498	PB
23	<i>DNAH11</i>	c.5622-7C>A	p.(?)	Intrónica	0,01551	PB
25	<i>DNAH11</i>	c.-27C>A	p.(?)	5' UTR	0,85189	PB
27	<i>DNAH5</i>	c.7230+24A>G	p.(?)	Intrónica	0,01670	PB
28	<i>DNAH11</i>	c.5270C>T	p.(Ala1757Val)	Missense	0,12049	PB
41	<i>DNAH5</i>	c.4597-22T>C	p.(?)	Intrónica	0,02426	PB
42	<i>RSPH9</i>	c.-37A>C	p.(?)	5' UTR	0,00567	PB
44	<i>DNAI1</i>	c.462C>T	p.(Pro154=)	Sinónima	-	PB
45	<i>DNAH5</i>	c.1335C>T	p.(Cys445=)	Sinónima	0,01067	PB
47	<i>DNAH11</i>	c.7059T>C	p.(Tyr2353=)	Sinónima	0,00361	PB
47	<i>DNAH11</i>	c.11469A>C	p.(Leu3823=)	Sinónima	0,00614	PB
47	<i>DNAI2</i>	c.759G>A	p.(Ala253=)	Sinónima	0,13761	PB

47	<i>NME8</i>	c.994+19A>G	p.(?)	Intrónica	0,27776	PB
48	<i>DNAI1</i>	c.1889C>T	p.(Ala630Val)	Missense	0,05449	PB

Ala alanina, *Arg* arginina, *Asn* asparagina, *Asp* ácido aspártico, *Cys* cisteína, *Gln* glutamina, *Glu* ácido glutâmico, *Gly* glicina, *His* histidina, *Ile* isoleucina, *Leu* leucina, *Lys* lisina, *Met* metionina, *Phe* fenilalanina, *Pro* prolina, *Ser* serina, *Thr* treonina, *Tyr* tirosina, *Val* valina, *B* benigna, *PB* provavelmente benigna.