

Universidade de Lisboa
Faculdade de Ciências
DEPARTAMENTO DE BIOLOGIA ANIMAL



**ENVOLVIMENTO DA VIA DE SINALIZAÇÃO BDNF/TRKB NA
ETIOLOGIA DO AUTISMO: ANÁLISE GENÉTICA**

Maria Margarida Saial Santos Guiomar Espada

MESTRADO EM BIOLOGIA HUMANA E AMBIENTE

2010

Universidade de Lisboa
Faculdade de Ciências
DEPARTAMENTO DE BIOLOGIA ANIMAL



**ENVOLVIMENTO DA VIA DE SINALIZAÇÃO BDNF/TRKB NA
ETIOLOGIA DO AUTISMO: ANÁLISE GENÉTICA**

Maria Margarida Saial Santos Guiomar Espada

Tese orientada pela Doutora Astrid Moura Vicente (Instituto Nacional de Saúde Dr. Ricardo Jorge e Instituto Gulbenkian de Ciência) e pela Prof.^a Doutora Ana Maria Crespo (Faculdade de Ciências da Universidade de Lisboa)

MESTRADO EM BIOLOGIA HUMANA E AMBIENTE

2010

Resumo

O Autismo é um perturbação do desenvolvimento infantil que se caracteriza por desvios na interacção social e na comunicação verbal e não verbal, assim como, por comportamentos repetidos, estereotipados e ritualizados. Estudos baseados em gémeos e em famílias demonstraram que o autismo apresenta uma forte componente genética. Contudo, pelo facto de ser uma doença complexa, existem várias variantes envolvidas, cada uma conferindo um risco reduzido no fenótipo o que tem dificultado a identificação dos genes de susceptibilidade desta doença.

O objectivo principal deste estudo foi a identificação dos genes de susceptibilidade envolvidos nas vias biológicas activadas pela neurotrofina BDNF (*Brain derived neurotrophic factor*) e pelo seu receptor TrkB (*Neurotrophic tyrosine kinase receptor type 2*). Estudos anteriores encontraram níveis significativamente elevados desta neurotrofina no soro de pacientes com autismo e uma associação entre marcadores do gene que codifica o seu receptor, *NTRK2* e o autismo, colocando a hipótese de que exista uma alteração da via de sinalização BDNF/TrkB nesta patologia. No presente estudo foi explorada esta hipótese, utilizando os resultados de um rastreio genómico de associação levado a cabo pelo consórcio internacional *Autism Genome Project*, onde participaram 1369 famílias, compostas pelo paciente ao qual foi diagnosticado autismo e os respectivos pais, cerca de um terço das quais eram de origem portuguesa (subpopulação PT). Especificamente, investigou-se a associação genética entre o autismo e os genes que codificam factores de transcrição (genes *homeobox*) que se ligam ao promotor do gene *NTRK2* bem como os genes das vias biológicas induzidas pela ligação do BDNF ao seu receptor TrkB (MEK-ERK, PI3K-AKT e PLC γ). Pretendeu-se ainda identificar interacções genéticas entre os genes *homeobox*, *NTRK2* e *BDNF* com impacto na susceptibilidade ao autismo. Em última análise, pretendeu-se analisar o enriquecimento do sinal de associação das vias biológicas supracitadas e identificar outros grupos de genes relevantes na susceptibilidade ao autismo. Em paralelo, foi realizado o rastreio por sequenciação de regiões do gene *NTRK2* anteriormente associadas ao autismo na população portuguesa.

Os resultados revelam novas regiões de susceptibilidade nos genes participantes nas três vias biológicas activadas por BDNF/TrkB, com especial interesse para os genes da via biológica PI3K-AKT e para o gene *NTRK2*. Alterações na expressão do gene *NTRK2* e dos genes participantes na via biológica PI3K-AKT poderão estar envolvidos na susceptibilidade ao autismo. O presente estudo permitiu também identificar três novas vias biológicas que poderão estar envolvidas na susceptibilidade à doença: Via biológica do metabolismo dos compostos isoprenóides, Via biológica do processo de aminoacilação do tRNA (RNA de transferência) na

síntese proteica e a Via de sinalização da m-Calpaína na mobilidade celular. Embora sejam necessários estudos adicionais para esclarecer o papel dos genes e vias identificadas, este estudo constituiu mais um importante passo na compreensão das vias metabólicas que poderão estar alteradas nesta doença, assim como um contributo na detecção de novos factores genéticos de susceptibilidade associados ao autismo.

Palavras-chave: Autismo|Rastreo genómico para associação|BDNF|*NTRK2*| Interacção genética| Análise de vias biológica

Abstract

Autism is a childhood neurodevelopment disorder that is characterized by impairments in social interaction, verbal and non-verbal communication, and stereotyped and ritualistic behaviors. Twin and family studies show that autism has a strong genetic component. However, since it's a complex trait there are several genetic variants involved each one with a small risk to the phenotype, which is a challenge in the identification of the susceptibility genes.

The main goal of this study is to identify susceptibility genes involved in the pathways activated by the neurotrophin BDNF (*Brain derived neurotrophic factor*) and its receptor TrkB (*Neurotrophic tyrosine kinase receptor type 2*). Previous studies found significantly high levels of the neurotrophin BDNF in the platelet-rich plasma of autistic patients and genetic markers of the gene encoding the BDNF receptor, *NTRK2*, were associated with autism. This leads to the hypothesis that alterations in BDNF/TrkB signaling contribute to an increased susceptibility in this disorder. In the present study we explored this hypothesis using the results of a genome-wide association study performed by Autism Genome Project with 1369 families that include the autistic patient and his non-affected parents, of which are one third from Portugal. Specifically, we tested for association between autism and the genes that encode homeodomain transcription factors (homeobox genes) that bind to the promoter region of *NTRK2* gene as well as the genes belonging to the pathways activated by BDNF/TrkB signaling (MEK-ERK, PI3K-AKT and PLC γ). In addition we also pretended to identify genetic interactions between the homeobox, *NTRK2* and *BDNF* genes that contribute to autism susceptibility. Finally, we used a pathway analysis approach with GWAS results to analyze gene sets (of interest) for enrichment of the association signal and to identify new pathways relevant to susceptibility of autism. In parallel, we have conducted a screening test of sequence changes in the *NTRK2* gene regions previously associated with autism, in the Portuguese population.

Results show new susceptibility regions in the genes of the three pathways induced by BDNF/TrkB signaling, specially the genes of the PI3K-AKT pathway and *NTRK2*. Modifications on the expression of the genes *NTRK2* and those from the PI3K-AKT pathway can be involved in the susceptibility of autism. The present study provided also the identification of three new pathways that could be involved in the disease risk: Isoprenoid metabolic pathway, tRNA aminoacylation for protein translation pathway and mCalpain pathway in cell motility. Although future studies will be necessary to clarify the role of the identified genes and pathways, this study was an important step in the knowledge of the metabolic pathways that can be disrupted in the disease, as well as detecting new genetic factors underlying the susceptibility to autism.

Key words: Autism Spectrum Disorder | Genome-wide association study | BDNF | *NTRK2* | Gene-Gene interaction | Pathway analysis

Atribuições

O presente trabalho foi desenvolvido no Departamento de Promoção da Saúde e Doenças Crónicas do Instituto Nacional de Saúde Dr. Ricardo Jorge, no âmbito do projecto PTDC/SAU-GMG/64519/2006 financiado pela Fundação para a Ciência e Tecnologia.

O trabalho envolve uma equipa multidisciplinar, e como tal esta tese resulta da contribuição de vários colaboradores, que em muito contribuíram para a sua realização. A colheita das amostras e a avaliação clínica dos pacientes com autismo foi realizada pela equipa médica do Hospital Pediátrico de Coimbra, chefiada pela Dra. Guiomar Oliveira. O recrutamento e colecção de amostras de sangue de controlos saudáveis foi realizado no Centro Regional de Sangue de Lisboa e Hospital Egas Moniz. A extracção de DNA dos indivíduos usados no estudo foram realizadas por outros colegas de equipa. O rastreio de regiões específicas do gene *NTRK2* através do processo de sequenciação pelo método de Sanger foi realizado e analisado por mim. Os passos que envolvem equipamento específico foram realizados na Unidade Laboratorial Comum do Instituto Nacional de Saúde Dr. Ricardo Jorge.

A análise de associação com o autismo dos genes *homeobox* e dos genes das vias biológicas induzidas pela via de sinalização BDNF/TrkB foi realizada com os genótipos obtidos no âmbito do rastreio genómico do *Autism Genome Project*. Os ensaios de genotipagem foram executados por uma equipa membro do AGP, no laboratório da *University College* em Dublin. A análise estatística genética apresentada (testes de associação e testes de interacção genética) foi feita por mim, assim como, a análise de enriquecimento de associação de vias biológicas, com o apoio do Doutor Tiago Magalhães e de outros membros da equipa de investigação.

Agradecimentos

Na elaboração deste trabalho foram muitas as pessoas que me auxiliaram e quero, por isso, aqui deixar o meu agradecimento às que foram essenciais à sua concretização.

À minha orientadora, Doutora Astrid Moura Vicente, por me ter dado a oportunidade de integrar o seu grupo de Epidemiologia Genética e participar num projecto tão interessante no qual aprendi muito e onde pude desenvolver o meu mestrado. Os ensinamentos que me transmitiu permitiram-me crescer enquanto profissional e enquanto pessoa. Pelos seus conselhos e críticas que serviram sempre, para melhorar o meu empenho e desempenho.

À Prof. Doutora Ana Crespo, minha co-orientadora, professora e coordenadora do mestrado em Biologia Humana e Ambiente, pela sua disponibilidade e atenção sempre que precisei de ajuda.

Um grande obrigado a todos os meus colegas de grupo, Filipa, Catarina, João, Tiago, Helena e Inês, por todo o apoio, ajuda e motivação que me deram durante a realização do meu trabalho para que conseguisse atingir os meus objectivos com a melhor das motivações e também, pelo grande companheirismo e pela boa-disposição (grandes risadas!). Aprendi muito e levo comigo muito boas recordações! Sem vocês esta passagem por Lisboa nunca teria sido a mesma. Um agradecimento muito especial à Filipa, a minha grande companheira de mestrado, para a qual não há palavras suficientes para lhe agradecer. Pela ajuda incansável, apoio, disponibilidade, dedicação e motivação durante os bons e menos bons momentos das nossas pequenas grandes conquistas! Ao Tiago, que tornou possível as tão complexas análises bioinformáticas. Pela sua dedicação, atenção, disponibilidade, motivação e perseverança. Um muito obrigado à Catarina e, mais recentemente, à Inês pelos conhecimentos transmitidos, a experiência, a ajuda, a disponibilidade, a motivação e orientação durante o trabalho. Ao João, um grande companheiro de grupo, pela sua disponibilidade, motivação, ajuda e momentos de risada! À Helena, pela sua ajuda, conhecimento e motivação.

Ao Miguel e às minhas amigas, em especial à Sandra, que me acompanharam durante todo este percurso, por serem a minha força, aturando-me, alegrando-me e motivando-me sempre que precisei.

Aos meus pais e irmão, a gratidão é infinita por tudo o que me têm proporcionado. Obrigada pelo apoio incondicional, imprescindível para que conseguisse conquistar mais esta etapa na minha vida.

Índice Geral

Resumo	i
<i>Abstract</i>	iii
Atribuições	v
Agradecimentos	vi
Índice Geral	vii
Índice de Figuras e de Tabelas	x
Abreviaturas e Símbolos	xii
I. INTRODUÇÃO	1
1. Espectro do Autismo	1
1.1. Perspectiva histórica e definição	1
1.2. Características comportamentais do autismo	2
1.3. Prevalência	3
1.4. Diagnóstico	4
1.5. Etiologia	4
2. Mapeamento genético de doenças complexas	6
2.1. A importância dos marcadores genéticos	6
2.2. Os conceitos de <i>Linkage</i> e Desequilíbrio de <i>Linkage</i>	7
2.3. Estudos de genes candidatos	8
2.3.1. Estudos caso-controlo	9
2.3.2. Estudos baseados em famílias	9
2.4. Rastreio genómico	10
2.5. O consórcio <i>Autism Genome Project</i>	13
3. Análise dos Rastreios genómicos para associação	15
3.1. Pesquisa de variantes raras	15
3.2. Análise de epistasia genética	16
3.3. Análise de vias biológicas	17
4. Factores neurotróficos envolvidos no autismo	17

4.1.	Via de sinalização BDNF/TrkB	18
4.2.	Papel da via de sinalização BDNF/TrkB no autismo	19
II. OBJECTIVOS		22
III. MATERIAL E MÉTODOS		23
1.	Recrutamento e caracterização clínica de pacientes	23
2.	Extracção e quantificação de DNA genómico	23
3.	Sequenciação do gene <i>NTRK2</i>	24
3.1.	Seleccção dos indivíduos para sequenciação	24
3.2.	Optimização das condições de PCR	24
3.3.	Reacção de Sequenciação	26
4.	Genotipagem no rastreio genómico AGP	27
5.	Análise de associação genética	30
5.1.	Análise de associação dos genes <i>homeobox</i> com o autismo	30
5.2.	Análise de associação das vias biológicas activadas por BDNF/TrkB com o autismo	31
6.	Análise de interacção genética	33
7.	Análise de enriquecimento por sinal de associação de vias biológicas	36
IV. RESULTADOS		39
1.	Rastreio das variações genéticas no gene <i>NTRK2</i>	39
1.1.	Variação no intrão 6-7	39
1.1.1.	Caracterização clínica	41
1.1.2.	Análise preditiva	41
1.1.3.	Frequências alélicas e genotípicas	43
1.2.	Variação no intrão 21-22	44
2.	Testes de associação genética	46
2.1.	Genes homeobox <i>DLX1</i> , <i>DLX2</i> , <i>HOXA1</i> , <i>PITX1</i> e <i>EN1</i>	46
2.2.	Genes envolvidos nas vias biológicas activadas por BDNF/TrkB	47
2.2.1.	População AGP	50
2.2.2.	Sub-População PT	50

3. Estudos de interacção genética	53
3.1. Genes <i>NTRK2</i> , <i>DLX1</i> , <i>DLX2</i> , <i>HOXA1</i> , <i>PITX1</i> e <i>EN1</i>	53
3.2. Genes <i>BDNF</i> e <i>NTRK2</i>	54
4. Enriquecimento do sinal de associação de vias biológicas no autismo	55
V. DISCUSSÃO DE RESULTADOS	58
1. Rastreio de variantes no gene <i>NTRK2</i>	58
2. Testes de associação genética	58
2.1. População AGP	58
2.2. Sub-População PT	60
3. Testes de interacção genética	61
4. Análise de enriquecimento de sinal de associação	62
5. Análise estatística: abordagens, vantagens e limitações	63
VI. DISCUSSÃO GERAL E PERSPECTIVAS FUTURAS	66
VII. REFERÊNCIAS BIBLIOGRÁFICAS	69
ANEXOS	78
Anexo A. O gene <i>NTRK2</i>	79
Anexo B. Via biológica das neurotrofinas	80
Anexo C. Protocolo <i>Salting out</i> para extracção de DNA	81
Anexo D. Diagrama do <i>cluster</i> das diferentes populações participantes no GWAS	82
Anexo E. STRING: selecção de genes para análise	83
Anexo F. Estudos de Interacção genética	84

Índice de Figuras e de Tabelas

Figuras

Figura 1.	Modelo de sub-fenótipos no autismo, proposto por Beglinger and Smith.	2
Figura 2.	Princípios do teste TDT.	10
Figura 3.	Esquema representativo da classificação de variantes genéticas segundo a sua frequência alélica de risco e <i>effect size</i> .	12
Figura 4.	Resultado do Rastreo Genómico para <i>Linkage</i> realizado pelo consórcio AGP, para 1181 famílias <i>multiplex</i> .	14
Figura 5.	Resultados do rastreo genómico para <i>Linkage</i> replicado na população AGP, em 1031 famílias <i>multiplex</i> .	14
Figura 6.	Esquema representativo das três vias biológicas activadas pela ligação da neurotrofina BDNF ao seu receptor de alta afinidade, TrkB.	19
Figura 7.	Etapas do procedimento experimental de Infinium Human 1M-single Infinium Beadchip da Illumina.	30
Figura 8.	Os diferentes passos envolvidos no método GMDR.	36
Figura 9.	Esquema representativo do <i>software</i> GenGen.	39
Figura 10.	Sequência nucleotídica dos exões 6/7 e respectivas regiões intrónicas do gene <i>NTRK2</i> .	41
Figura 11.	Resultado da análise da sequenciação dos exões 6/7 no programa <i>Staden</i> .	41
Figura 12.	Cromatograma dos resultados de sequenciação dos exões 6/7.	41
Figura 13.	Esquema representativo dos resultados dos programas de análise preditiva no intrão 6/7.	43
Figura 14.	Sequência nucleotídica do exão 22 e respectiva região intrónica do gene <i>NTRK2</i> .	45
Figura 15.	Resultado da sequenciação do exão 22.	45
Figura 16.	Resultado do cromatograma de sequenciação do exão 22.	46
Figura 17.	Padrão de <i>splicing</i> alternativo do gene <i>NTRK2</i> .	80
Figura 18.	Estrutura de três proteínas : TRkB, TrkB-T1 e TRkB-T-Shc.	80
Figura 19.	Esquema representativo da via biológica das neurotrofinas (adaptado de KEGG).	81
Figura 20.	Diagrama de <i>clustering</i> dos resultados de <i>Principal Component Analysis</i> (PCA) para as diferentes populações que constituem a amostra AGP.	83
Figura 21.	Resultados da selecção de genes através do programa STRING.	84
Figura 22.	Modelo gráfico multifactorial de <i>NTRK2</i> e <i>DLX1</i> .	85
Figura 23.	Modelo gráfico multifactorial de <i>NTRK2</i> e <i>PITX1</i> .	85
Figura 24.	Modelo gráfico multifactorial de <i>NTRK2</i> e <i>EN1</i> .	85
Figura 25.	Modelo gráfico multifactorial de <i>NTRK2</i> e <i>HOXA1</i> .	86
Figura 26.	Modelo gráfico multifactorial de <i>NTRK2</i> e <i>DLX2</i> .	86
Figura 27.	Modelo gráfico multifactorial de <i>NTRK2</i> e <i>BDNF</i> .	86

Tabelas

Tabela 1.	Haplótipos de risco do gene <i>NTRK2</i> associados ao autismo por Correia <i>et al.</i> (2010).	25
Tabela 2.	Sequências e temperaturas de hibridação dos <i>primer forward</i> e <i>reverse</i> utilizados para os exões e respectivo tamanho do produto de PCR amplificado.	26
Tabela 3.	Mistura de reacção de PCR, para os cinco exões do gene <i>NTRK2</i> .	26
Tabela 4.	Condições de reacção de PCR para os cinco exões do gene <i>NTRK2</i> .	27
Tabela 5.	Mistura de reacção de sequenciação.	27
Tabela 6.	Condições da reacção de sequenciação.	28
Tabela 7.	Características dos marcadores genéticos presentes <i>Infinium DNA Analysis Human 1M Beadchip</i> , da Illumina.	29
Tabela 8.	Tabela indicativa do número de SNP genotipados para os genes <i>HOXA1</i> , <i>PITX1</i> , <i>EN1</i> , <i>DLX1</i> e <i>DLX2</i>	32
Tabela 9.	Lista de 27 genes pertencentes às três vias de sinalização activadas pela ligação BDNF/TrkB e a respectiva molécula codificada por cada gene.	33
Tabela 10.	Lista dos resultados do programa de interacção proteica STRING para valores de confiança superiores a 95%.	34
Tabela 11.	Número de marcadores genéticos para o par de genes utilizado em cada ensaio de interacção genética	37
Tabela 12.	Alterações de DNA encontradas nas regiões sequenciadas no gene <i>NTRK2</i> .	40
Tabela 13.	Caracterização clínica referente aos três indivíduos com o a alteração c.288-15T>C.	42
Tabela 14.	Frequências genotípicas absolutas e frequências alélicas para o polimorfismo c.288-15T>C. Resultados do teste de χ^2 em 100 indivíduos das populações de casos e controlo	44
Tabela 15.	Caracterização clínica dos quatro indivíduos com o polimorfismo no intrão 21-21, rs3739807.	46
Tabela 16.	Frequências genotípicas absolutas de rs3739807 em 100 indivíduos com autismo, na população portuguesa.	47
Tabela 17.	Resultados significativos do teste de associação genética TDT, na população AGP, para os genes <i>EN1</i> , <i>PITX1</i> , <i>HOXA1</i> , <i>DLX1</i> e <i>DLX2</i> .	48
Tabela 18.	Resultados significativos do teste de associação genética TDT na população PT para os genes <i>EN1</i> , <i>PITX1</i> , <i>HOXA1</i> , <i>DLX1</i> e <i>DLX2</i> .	48
Tabela 19.	Resultado significativos do teste de associação genética TDT, na população AGP, para os genes envolvidos nas vias biológicas activadas por BDNF/TrkB.	49
Tabela 20.	Resultado significativos (p -value $\leq 0,05$) do teste de associação genética TDT, na população PT, para os genes envolvidos nas vias biológicas activadas por BDNF/TrkB.	53
Tabela 21.	Resultado de PGMDR para a interacção genética entre o gene <i>NTRK2</i> e os genes <i>DLX1</i> , <i>DLX2</i> , <i>HOXA1</i> , <i>PITX1</i> e <i>EN1</i> .	55
Tabela 22.	Resultados de PGMDR para a interacção entre os marcadores dos genes <i>NTRK2</i> e <i>BDNF</i> .	56
Tabela 23.	Resultado significativos do enriquecimento por sinal de associação dos genes de 1655 vias biológicas/ grupo de genes de interesse, em 1369 famílias da população AGP, através do <i>software</i> GenGen.	56
Tabela 24.	Resultado da análise de "Leave One Out" na via biológica mCalpaína.	58

Abreviaturas e Símbolos

3'UTR	Região não traduzida 3'UTR (<i>Three prime Untranslated Region</i>)
5'UTR	Região não traduzida 5'UTR (<i>Five prime Untranslated region</i>)
AARS	<i>Aminoacyl tRNA synthetase</i>
ADHD	Distúrbio de Déficit de Atenção com Hiperatividade (<i>Attention Deficit Hyperactivity Disorder</i>)
ADI-R	Entrevista de Diagnóstico de Autismo Revista (<i>Autism Diagnostic Interview-Revised</i>)
ADOS	Diagnóstico de Observação de Autismo e Distúrbios do Espectro do Autismo (<i>Autism Diagnostic Observation Schedule</i>)
AGP	<i>Autism Genome Project</i>
AKT	Proteína cinase serina/Treonina Akt (<i>Protein kinase Akt</i>)
ASD	Distúrbios do Espectro do Autismo (<i>Autism Spectrum Disorders</i>)
BDNF	<i>Brain-derived neurotrophic factor</i>
BDNF	Gene codificante de <i>Brain-derived neurotrophic factor</i>
bp	Par de bases (<i>base pair</i>)
BP	<i>Branch point</i>
cM	Centimorgan
CNV	<i>Copy number variation</i>
CVC	<i>Crossvalidation Consistency</i>
DLX1	Gene codificante de <i>Distal-less homeobox 1</i>
DLX2	Gene codificante de <i>Distal-less homeobox 2</i>
DNA	Ácido desoxirribonucleico (Deoxyribonucleic acid)
Downstream	A jusante
Dz	Gêmeos dizigóticos
EDTA	Ácido etilenodiamino tetra-acético
EEG	Electroencefalograma
EN1	Gene codificante de <i>Engrailed homeobox 1</i>
ENSEMBL	<i>Ensembl genome Project</i>
ERK	<i>Extracellular signal-regulated kinases</i>
ES	Valor de Enriquecimento do sinal de associação (<i>Enrichment score</i>)
ESE	<i>Exonic Splicing Enhancer</i>
ESS	<i>Exonic Splicing Silencer</i>
FDR	<i>False Discovery Rate</i>
FRS2	<i>Fibroblast growth factor receptor substrate 2</i>
FWER	<i>Familywise Error Rate</i>
GAB1	Gene codificante de <i>GRB2-associated binding protein 1</i>
GMDR	<i>Generalized Multifactor Dimensionality Reduction</i>
GWAS	Rastreo genómico para associação (<i>Genome wide association studies</i>)
hnRNP	<i>Heterogeneous nuclear ribonucleoproteins</i>
HOMF	<i>Homeodomain Transcription Factors</i>
HOXA1	Gene codificante de <i>homeobox A1</i>
HOXF	<i>Factors with moderate activity to homeodomain consensus sequence</i>
HuR	<i>ELAV-like protein 1</i>
HWE	Equilíbrio de Hardy-Weinberg (<i>Hardy Weinberg Equilibrium</i>)
ISE	<i>Intronic Splicing Enhancer</i>
ISS	<i>Intronic Splicing Silencer</i>
Kb	Quilo bases (<i>Kilo base pair</i>)

KEGG	<i>Kyoto Encyclopedia of Genes and Genomes</i>
LD	Desequilíbrio de <i>Linkage</i> (<i>Linkage disequilibrium</i>)
MAF	Frequência do alelo menos comum (<i>Minor allele frequency</i>)
MAPK	<i>Mitogen-activated protein (MAP) kinases</i>
MDR	<i>Multifactor Dimensionality Reduction</i>
MEK	<i>Mitogen-activated protein kinase</i>
MLE	<i>Maximum-Likelihood Estimates</i>
mRNA	RNA mensageiro (<i>Messenger RNA</i>)
Mz	Gêmeos monozigóticos
NCBI	<i>National Center for Biotechnology Information</i>
NES	Valor de enriquecimento normalizado (<i>Normalized Enrichment Score</i>)
NT	Não transmitido (<i>Untransmitted</i>)
<i>NTRK2</i>	Gene codificante de <i>neurotrophic tyrosine kinase receptor type 2</i>
OR	<i>Odds ratio</i>
PCR	<i>Polymerase Chain Reaction</i>
PGMDR	<i>Pedigree Generalized Multifactor Dimensionality Reduction</i>
PI3K	Fosfoinositol 3-quinase (<i>Phosphoinositide 3-kinases</i>)
<i>PIK3R1</i>	Gene codificante de <i>PI3-kinase regulatory subunit alpha²</i>
<i>PIK3R3</i>	Gene codificante de <i>Phosphatidylinositol 3-kinase regulatory subunit gamma</i>
<i>PIK3R5</i>	Gene codificante de <i>Phosphoinositide 3-kinase regulatory subunit 5</i>
<i>PIK3CA</i>	Gene codificante de <i>Phosphoinositide-3-kinase, catalytic, alpha polypeptide</i>
<i>PIK3CD</i>	Gene codificante de <i>Phosphoinositide-3-kinase, catalytic, delta polypeptide</i>
<i>PIK3CG</i>	Gene codificante de <i>Phosphatidylinositol-4,5-bisphosphate 3-kinase 110 catalytic subunit gamma</i>
<i>PITX1</i>	Gene codificante de <i>Paired-like homeodomain 1</i>
PLCG	<i>Phospholipase C gamma</i>
PPT	<i>Polypyrimidine tract site</i>
PT	Portuguesa (população)
PTB	<i>Polypyrimidine tract-binding protein</i>
RNA	Ácido Ribonucleico (<i>Ribonucleic Acid</i>)
<i>SHC3</i>	Gene codificante de <i>SHC transforming protein 3</i>
SNC	Sistema Nervoso Central
SNP	Polimorfismo de uma única base (<i>Single Nucleotide Polymorphism</i>)
<i>SOS</i>	Gene codificante de <i>Son of sevenless</i>
SRp20	<i>Serine/arginine-rich protein p20</i>
T	Transmitido (<i>Transmitted</i>)
TBA	<i>Testing Balanced Accuracy</i>
TBE	Tris base, ácido bórico e EDTA
TDI	<i>Transmission Disequilibrium Test</i>
TrkB	<i>Neurotrophic tyrosine kinase receptor type 2</i>
U	Unidade (<i>Unit</i>)
<i>Upstream</i>	A montante
WISC	Wechsler Intelligence Scale for Children
WGS	Sequenciação do genoma (<i>Whole-genome Sequencing</i>)
χ^2	Chi-quadrado

I. INTRODUÇÃO

1. ESPECTRO DO AUTISMO

1.1. Perspectiva histórica e definição

O autismo foi pela primeira vez descrito por Leo Kanner, em 1943, para definir uma condição observada num grupo de 11 crianças, na sua maioria rapazes, com um atraso no desenvolvimento e isolamento social, o qual denominou de "early infantile autism" ou "inborn autist disturbances of affective contact" (Kanner, 1943). Um ano mais tarde e sem conhecimento da descrição de Kanner, Hans Asperger descreveu quatro pacientes com comportamentos similares e definiu a condição como "Autistic psychopatology of childhood" (Asperger, 1944). O reconhecimento de Autismo como uma doença distinta foi feito com base na definição proposta por Michael Rutter, em 1978 (Rutter, 1978). Várias décadas passaram após a primeira descrição até que o autismo fosse reconhecido como uma doença de etiologia complexa, na qual os factores genéticos têm um papel relevante (Wing, 1997; Folstein and Rosen-Sheidley, 2001).

O termo "Perturbações do Espectro do Autismo" (no presente estudo referidas como *Autism Spectrum Disorder*, ASD) refere-se ao conjunto de perturbações do comportamento e desenvolvimento neuronal, com etiologia bastante heterogénea e, em parte desconhecida, caracterizados por desvios em três categorias comportamentais: na interacção social; na comunicação verbal e não verbal; e em comportamentos repetitivos e estereotipados (Wing, 1997; Folstein and Rosen-Sheidley, 2001). O espectro do autismo estende-se para além de padrões definidos caracterizando-se por uma combinação de áreas comportamentais, como ilustrado na figura 1. Num dos extremos do espectro encontram-se indivíduos com baixo potencial cognitivo caracterizados por deficiências na comunicação verbal e comportamentos repetitivos acompanhados de atraso mental. No extremo oposto encontram-se os indivíduos com elevado potencial cognitivo, como os indivíduos com Síndrome de Asperger, caracterizados por défices na interacção social, com comportamentos estereotipados e repetitivos, mas sem atraso no desenvolvimento da linguagem (Berglinger and Smith, 2001).

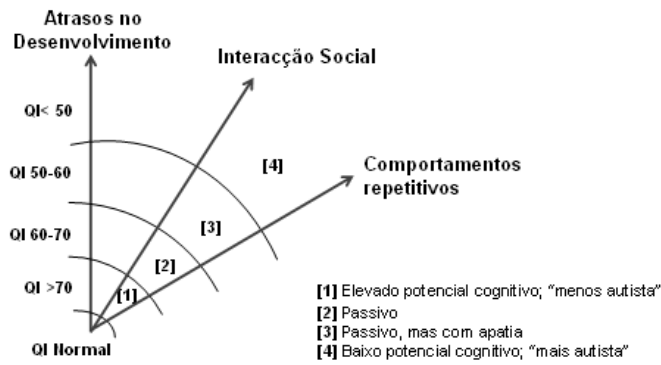


Figura 1. Modelo de sub-fenótipos no autismo, proposto por Beglinger and Smith (2001) (adaptado).

Dentro do grupo das ASD, os indivíduos cujo fenótipo clínico não se encaixa nos padrões restritos de perturbação autista pertencem ao grupo de Perturbações Globais de Desenvolvimentos Não Especificadas (*Pervasive development disorders not otherwise specified*, PDD-NOS). As ASD estão integradas no conjunto clínico de Perturbações Globais de Desenvolvimento, PGD (adaptado de *Pervasive development disorders*, PDD) que se definem como o conjunto de distúrbios clínicos com início precoce que afectam simultaneamente múltiplas áreas básicas do comportamento e desenvolvimento, nos quais se incluem também o Síndrome de Rett e o Distúrbio Desintegrativo de Segunda Infância (*Childhood Disintegrative disorder*, CDD) (Wing, 1997; Oliveira *et al.*, 2007).

1.2. Características comportamentais do autismo

As características do autismo, descritas por Kanner, incluem desvios na interacção social recíproca, na comunicação verbal e não verbal, bem como interesses invulgares e comportamentos estereotipados (Lord *et al.*, 2000). O autismo manifesta-se durante a infância, frequentemente identificado entre os 18 e os 30 meses de idade, onde é notório a ausência ou atraso do desenvolvimento da linguagem e de interacção social. As rotinas repetitivas aparecem nos primeiros 36 meses e, até aos 5 anos, os comportamentos repetitivos e isolamento social atingem o seu extremo (Wing, 1997).

As crianças com autismo têm um desvio muito acentuado nos comportamentos comunicativos e de interacção social, tais como ausência de expressão facial, postura corporal e contacto visual inadequado à situação social (Folstein and Rosen-sheidley, 2001). Mostram ausência de sentimentos espontâneos e reciprocidade emocional com outros indivíduos, mantendo-se afastados e indiferentes aos outros. As crianças não sabem como brincar, manipulando os brinquedos sem a noção aparente do que eles representam e não apresentam jogo simbólico (Rapin, 1997). Na adolescência estas características podem ser substituídas por depressão ou isolamento. As crianças têm pouca capacidade de concentração, embora por vezes tenham

momentos de atenção bastante prolongados durante as actividades por elas iniciadas (Rapin, 1997).

No domínio verbal, um terço dos indivíduos não desenvolve a linguagem, expressando apenas palavras isoladas, não conseguindo articular um discurso. As características mais comuns são o atraso no desenvolvimento da linguagem, a qual apresenta anomalias características que incluem a imitação das expressões bem como a repetição de palavras ou frases de forma estereotipada e monotónica. A deficiência na pragmática é a anomalia da linguagem típica do autismo. A pragmática refere-se aos aspectos sociais da comunicação e inclui a coordenação do contacto visual e fala, demonstração de interesse pela opinião dos outros ou saber quando parar de falar. Os indivíduos com autismo usam a linguagem para exprimir necessidades ou dar informações mas raramente para trocar informações ou socializar (Wing, 1997; Folstein and Rosen-sheidley, 2001).

O comportamento repetitivo consiste em movimentos estranhos dos membros do corpo como bater ou entrelaçar as mãos, designadas por estereotipias motoras, no fascínio por estímulos sensoriais simples como luzes brilhantes e na resistência à alteração às suas rotinas no que diz respeito a horas ou caminhos percorridos (Wing, 1997). Comportamentos de auto-agressão podem ser uma forma severa das estereotipias (Rapin, 1997). Também incluídos nesta categoria estão os interesses ou preocupações exageradas por determinados tópicos (Folstein and Rosen-Sheidley, 2001). Outras características incluem uma sensibilidade elevada a certos sons, estímulos tácteis ou dor, distúrbios do sono, postura e modo de andar singular (Wing, 1997).

1.3. Prevalência

A doença manifesta-se nos três primeiros anos de vida, atingindo quatro vezes mais os indivíduos do sexo masculino e ocorrendo com uma prevalência de cerca de uma em cada mil crianças na população portuguesa (Folstein and Rosen-Sheidley, 2001; Oliveira *et al.*, 2007). Fombonne (2009) estima que a prevalência de todas as PGD combinadas seja de, aproximadamente, 60-70 indivíduos por cada 10 000, tornando-se uma das mais frequentes perturbações do desenvolvimento infantil, que tem vindo a aumentar nas últimas décadas. A prevalência de ASD nos estudos mais recentes é de 20 indivíduos por cada 10 000, dos quais um terço são indivíduos com Síndrome de Asperger, com uma prevalência de 6 indivíduos por cada 10 000 (Fombonne, 2009).

1.4. Diagnóstico

Não existem marcadores biológicos para o autismo. O diagnóstico assenta na avaliação comportamental por profissionais experientes. A maioria dos indivíduos afectados revelam deficiências no desenvolvimento infantil, que permanecem ocultas até ao segundo ano de vida, altura em que surgem desvios e atrasos na construção de frases e interacção social (Wing, 1997).

O diagnóstico é feito com base na investigação detalhada do desenvolvimento da criança e historial familiar, com incidência nas três áreas de sintomatologia com recurso à Entrevista de Diagnóstico de Autismo Revista (*Autism Diagnostic Interview-Revised*, ADI-R) e à Escala de Diagnóstico de Observação de Autismo e Perturbações do Espectro do Autismo (*Autism Diagnostic Observation Schedule*, ADOS) (Lord *et al.*, 1994; Oliveira *et al.*, 2007). Os quocientes de inteligência e desenvolvimento são determinados através da Escala II de Ruth Griffiths de desenvolvimento mental (*Ruth Griffiths Mental Scale II*) ou a escala de inteligência de Wechsler para crianças (*Wechsler Intelligence Scale for Children*, WISC) (Griffiths, 1984; Marques, 1970). A investigação de possíveis causas, distúrbios físicos ou psicológicos é realizada através de electroencefalograma (EEG), análises cromossómicas, bioquímicas e genéticas (Wing, 1997).

1.5. Etiologia

O autismo é uma doença complexa, cuja etiologia é desconhecida na maioria dos casos. A heterogeneidade de fenótipos resulta provavelmente da interacção entre múltiplos genes, vias biológicas e factores ambientais, ainda não definidos, e que contribuem no seu conjunto para o risco da doença. Em 90% dos casos, a causa de autismo não é conhecida e denomina-se por autismo idiopático. Nos restantes 10%, a causa está associada a uma condição médica conhecida, como o Síndrome de X-Frágil (1% a 5% dos pacientes com ASD), Síndrome de Rett (0,5% dos pacientes com ASD), Esclerose tuberosa (1% dos pacientes com ASD), Síndrome de Down (1% dos pacientes com ASD), Neurofibromatose (0,2% a 1,4% dos pacientes com ASD) e anomalias cromossómicas, denominando-se por autismo sindrómico (Folstein and Rosen-Sheidley, 2001; Abrahams and Geschwind, 2008; Kelleher and Bear, 2008). Vários distúrbios metabólicos como a doença ou disfunção mitocondrial e a fenilcetonúria também se encontram associados ao autismo (Lombard, 1998; Muhle *et al.*, 2004). No estudo realizado em Portugal 20,3% dos 69 pacientes com autismo apresentavam hiperlactacidemia, um marcador bioquímico de disfunção mitocondrial. Estudos funcionais mitocondriais confirmaram a doença mitocondrial em cinco dos pacientes participantes (Oliveira *et al.*, 2005).

Várias anomalias cromossómicas, envolvendo quase todos os cromossomas, foram descritas em indivíduos com autismo, tais como, aneuploidias, deleções, duplicações, translocações e

inversões (Folstein and Rosen-sheidley, 2001). As duplicações nas regiões 15q11-q13, deleções de 2q37, 22q11, 22q13.3 e Xp22.3 têm sido recorrentemente descritas em estudos de citogenética, sendo a região 15q11-q13 a anomalia citogenética mais comum, ocorrendo em 1% a 3% dos pacientes com autismo. Esta região inclui a região crítica para os síndromes de Prader-Willi e Angelman e está descrita na literatura como estando sujeita ao processo de *imprinting* genómico (Gillberg, 1998; Castermans *et al.*, 2004). Um número considerável de anomalias cromossómicas têm sido encontradas na região 7q, no entanto, estas anomalias estão localizadas por todo o braço longo do cromossoma, o que dificulta a identificação de genes de susceptibilidade (Scherer *et al.*, 2003). As metodologias baseadas em *arrays*, tais como CGC (*Comparative Genomic Hybridization*) ou o método de FISH (*Fluorescent in situ hybridization*) permitem identificar deleções ou duplicações em determinadas regiões do genoma. Actualmente, a detecção de *copy number variation* (CNV) é a técnica frequentemente utilizada para detectar micro variações estruturais nos cromossomas (Abrahams and Geschwind, 2008). A sua potencialidade poderá ser uma mais-valia para o diagnóstico clínico.

O autismo é uma doença multifactorial com elevada heritabilidade, na ordem dos 90% (Rutter, 2005). A heritabilidade de um fenótipo é definida como a proporção da variância total que pode ser explicada por factores genéticos (Hirschhorn and Daly, 2005; Manolio *et al.*, 2009). No entanto, a importância da componente genética só se tornou clara na década de 1980, pois até aqui, a causa era atribuída a deficiências na educação dos pais que teriam lacunas a nível social como, por exemplo, a hipótese das "mães-frigorífico" ou a experiências traumáticas precoces (Eisenberg, 1957; Bettelheim, 1967). Para determinar se a doença é hereditária são realizados estudos de famílias, nos quais é calculado o risco de recorrência (λ_R) de um familiar de probando(s) afectado(s) comparativamente ao risco da população em geral (Ott, 1999). No autismo, o risco de recorrência nos familiares é aumentado entre 2% a 8% em relação à população em geral (Muhle, 2004). É 75 a 150 vezes mais prevalente na população afectada comparativamente com outras doenças multifactoriais (Rutter, 2005). Várias características como défices de interacção social, dificuldades de comunicação (aspectos pragmáticos de linguagem) e preferência por rotinas são encontradas mais frequentemente em pais de indivíduos afectados em relação à população em geral (Folstein and Rosen-Sheidley, 2001). Contudo, o risco elevado nos familiares não significa que a doença tenha uma componente hereditária causada por variação genética. Desta forma, o estudo de famílias não permite compreender se a doença é uma consequência do efeito de factores genéticos ou ambientais. Para esclarecer o contributo de ambos são realizados estudos de gémeos. Os gémeos monozigóticos (Mz) são geneticamente idênticos e partilham ambientes semelhantes ao contrário dos gémeos dizigóticos (Dz) que apenas partilham metade do património genético por descendência. O objectivo é comparar as taxas de concordância entre os dois tipos de gémeos. No autismo, a concordância nos gémeos Mz varia entre 60% a 90% e nos gémeos Dz entre de

2% a 10% o que evidencia o contributo dos factores genéticos na etiologia da doença (Folstein and Rosen-Sheidley, 2001; Muhle *et al.*, 2004; Abrahams and Geschwind, 2008).

O autismo pode ser o resultado do pequeno efeito de múltiplos genes e o seu funcionamento em conjunto nas vias biológicas de desenvolvimento cerebral serão a possível causa da variabilidade fenotípica característica da doença (Levitt and Campbell, 2009).

2. MAPEAMENTO GENÉTICO DE DOENÇAS COMPLEXAS

Existem múltiplas estratégias para identificar genes de susceptibilidade e mecanismos biológicos numa doença complexa. Os testes de associação, um dos métodos mais utilizados, procuram identificar variantes genéticas que estão associadas a uma patologia, comparando os indivíduos afectados com uma população controlo. No decorrer do capítulo, será feita uma descrição do método, apresentando as vantagens e limitações.

2.1. A importância dos marcadores genéticos

A presença de marcadores genéticos no genoma humano possibilitou o mapeamento genómico. Os marcadores genéticos são pequenas sequências de nucleótidos no DNA com localização definida e são polimórficos, isto é, num determinado *locus* existem duas ou mais variantes com frequências significativas na população. São exemplos de marcadores genéticos os minisatélites, os microsátélites, os VNTR (*Variable Number Tandem Repeat*), os STR (*Short Tandem Repeat*), entre outros. Entre os marcadores genéticos mais usados actualmente estão os polimorfismos de uma única base ou SNP (*Single Nucleotide Polymorphism*). São a variante genética mais comum que, por definição, ocorre quando existe uma alteração num único nucleótido da sequência de DNA. Estão descritos cerca de 11 milhões de SNP no genoma humano distribuídos uniformemente e, como tal, constituem importantes biomarcadores no rastreio genómico em doenças complexas (Kruglyak and Nickerson, 2001; Frazer *et al.*, 2009). O grande impulso na utilização de rastreios genómicos teve lugar com o início do projecto *International Haplotype Map*, designado por HapMap (The International HapMap Consortium, 2003). Este projecto internacional pretende catalogar os marcadores genéticos que ocorrem no genoma humano e, mais relevante, determinar as estruturas de desequilíbrio de *linkage* (*Linkage disequilibrium*, LD) entre os marcadores nas diferentes populações. Foi possível determinar blocos de haplótipos de SNP que se encontram em LD, identificando os denominados *tag* SNP para cada bloco. Um *tag* SNP é um SNP representativo de uma região genómica com elevado LD que permite identificar a variabilidade genética sem ser necessário genotipar todos os SNP da mesma região. A utilização destes SNP informativos tornou-se

altamente eficaz com as novas tecnologias de genotipagem. Contudo, os SNP por si só não podem explicar a diversidade genética humana, nem a susceptibilidade a doenças complexas, embora seja inquestionável a sua importância como ferramenta de análise (Kruglyak and Nickerson, 2001; Seng and Seng, 2008; Frazer *et al.*, 2009).

2.2. Os conceitos de *Linkage* e Desequilíbrio de *Linkage*

O termo *linkage* refere-se à proximidade física de *loci* ao longo do cromossoma. Dois *loci* estão em *linkage* se estiverem suficientemente próximos para que os seus alelos não sejam separados durante o processo de *crossing-over*, pelo que alelos num pequeno segmento cromossômico tendem a ser transmitidos em bloco, ou seja são cosegregados. A análise de *linkage* baseia-se no estudo da cosegregação de marcadores polimórficos de DNA com uma doença dentro de famílias com indivíduos afectados em várias gerações (Borecki and Suarez, 2001). Estes estudos identificam e localizam regiões de *linkage* com grandes dimensões, na ordem das megabases, que podem incluir várias dezenas de genes. Neste método, a probabilidade de *linkage* é estimada através do valor de LOD score (*Logarithm of an odds ratio*), que é o logaritmo da probabilidade de *linkage*, ou seja, a razão entre a probabilidade de *linkage* para uma dada fracção de recombinação e a probabilidade esperada assumindo que não existe *linkage* (hipótese nula). O valor de LOD superior ou igual a 3 é considerado como estatisticamente significativo. Significa que a hipótese de *linkage* é 10^3 vezes mais provável que a hipótese de que dois *loci* não estão em *linkage* (Ott, 1999; Borecki and Suarez, 2001).

Os estudos de *linkage* pretendem identificar *loci* que cosegregam com a doença dentro de famílias, enquanto os estudos de associação procuram identificar os alelos de marcadores genéticos associados ao fenótipo, na população. A associação entre um marcador e o fenótipo de interesse pode ocorrer se se verificarem uma de duas circunstâncias: quando a variante funcional (causal) é marcador genético ou quando o marcador genético está em desequilíbrio de *linkage* (*Linkage disequilibrium*, LD) com a variante funcional (causal) (Borecki and Suarez, 2001). Para se compreender o conceito de LD é necessário definir o conceito de Equilíbrio de Hardy-Weinberg (*Hardy-Weinberg Equilibrium*, HWE). Considere-se uma população infinita, com cruzamentos ao acaso e ausência de selecção, mutação e migração, onde as frequências genotípicas se mantêm constantes e dependem apenas das frequências alélicas. Então, diz-se que esta população está em HWE. No caso de dois ou mais *loci*, existe equilíbrio de *linkage* quando todas as combinações possíveis de alelos nos dois *loci* ocorrem tão frequentemente quanto seria de esperar pela associação aleatória das frequências alélicas individuais. Os desvios à ocorrência de alelos ao acaso são referidos como associação alélica ou LD. Considere-se uma mutação T, num determinado gene, flanqueada por outros marcadores genéticos. Esta mutação encontra-se associada aos alelos "vizinhos" sendo apenas observável quando estes o são. Assim, os alelos dos marcadores "vizinhos" de uma variante genética causal da doença,

quando a mutação T que lhe deu origem foi introduzida na população, podem permanecer na sua proximidade através de várias gerações, isto é permanecem em LD. Se T é o alelo mutante no *locus* da doença existirão diferenças nos valores de frequências alélicas desses marcadores entre indivíduos afectados e não afectados. Esta análise é eficaz quando a variante de susceptibilidade à doença provém de uma única mutação que lhe deu origem e os marcadores estão muito próximos do *locus* da doença, sendo o sinal de LD superior em mutações mais recentes. Uma medida de LD entre dois *loci* é D' (*Deviation from LD*) que varia entre 0 e 1. Quando dois *loci* se encontram completamente separados, D' é 0; se dois loci forem idênticos, ou seja, o *locus* do marcador é o *locus* causal, neste caso o D' é igual a 1 (Ott, 1999; Borecki and Suarez, 2001; Chapman and Thompson, 2001; Nordborg and Tavaré, 2002).

2.3. Estudos de genes candidatos

Genes candidatos são genes envolvidos em vias biológicas patofisiologicamente relevantes ou localizados em regiões cromossómicas de susceptibilidade e que foram anteriormente associados à doença (Ott, 1999). Na última década foram realizados muitos estudos de genes candidatos para se testarem potenciais regiões de susceptibilidade ao autismo, no entanto os resultados têm-se revelado inconclusivos. Muitos genes têm sido testados como genes candidatos no autismo idiopático. O pequeno tamanho da amostra e os distintos *backgrounds* genéticos dos grupos étnicos poderão explicar em parte a ausência de replicação de resultados de associação entre os vários estudos. Mesmo quando a evidência de associação é fraca há alguns genes que se mostram como fortes candidatos na susceptibilidade à doença. Por exemplo, baseado na observação de uma ligação entre a hiperserotonémia e o autismo, há diferentes genes envolvidos no sistema serotoninérgico que aparecem fortemente associados nos pacientes com hiperserotonémia (Coutinho *et al.*, 2004; Klauck, 2006). Com especial relevância para o presente estudo, Correia *et al.*, (2010) testou a associação de determinados genes candidatos envolvidos na determinação dos níveis de BDNF (*Brain derived neurotrophic factor*) na patofisiologia do autismo. Foram encontrados resultados significativos, na população portuguesa, entre os níveis de BDNF (no soro) e o autismo e descrita pela primeira vez a associação entre o gene *NTRK2* (*Neurotrophic Tyrosine Kinase Receptor type 2*) e o autismo.

O estudo de genes candidatos pode ser realizado através de estudos de associação genética com diferentes marcadores genéticos, tais como microsatélites e SNP. A utilização dos microsatélites é vantajosa no sentido em que os marcadores são mais polimórficos, ou seja, são mais informativos. No entanto, actualmente utilizam-se os SNP por serem mais frequentes ao longo de todo o genoma. Existe associação entre um marcador e um fenótipo quando estes ocorrem juntos no mesmo indivíduo mais frequentemente que o observado (Ott, 1999).

2.3.1. Estudos caso-controlo

Estes estudos populacionais baseiam-se na comparação de frequências alélicas ou genotípicas entre indivíduos afectados, não relacionados entre si, e controlos. Comparam a frequência de alelos observados nos pacientes e a frequência dos alelos observados nos controlos. Neste sentido, um alelo diz-se associado com a doença em estudo se ocorrerem desvios nos valores das frequências alélicas dos casos comparativamente aos controlos. A análise estatística envolve um teste estatístico com distribuição de Chi-quadrado (χ^2). Se a frequência de um marcador for significativamente mais elevado nos casos que nos controlos podemos inferir que está na vizinhança de uma variante causal (Ott, 1999; Schork *et al.*, 2001; Gordon and Finch, 2005).

A principal limitação deste tipo de estudo é a estratificação da população, isto é, os pacientes podem ser provenientes de uma população e os controlos de outra. A associação detectada pode ser relativa à origem genética e não estar relacionada com a doença. Assim, existirão diferenças nas frequências alélicas nas duas populações o que poderá comprometer o poder do teste de associação, dado que poderão surgir falsos positivos (erros tipo I) (Schork *et al.*, 2001; Gordon and Finch, 2005).

2.3.2. Estudos baseados em famílias

Os estudos baseados em famílias surgiram com o intuito de minimizar os problemas de estratificação populacional que podem ocorrer nos estudos caso-controlo e utilizam como controlo interno os alelos dos pais que não foram transmitidos para a descendência afectada. O método mais usado, denominado por *Transmission Disequilibrium Test* (TDT) e desenvolvido por Spielman *et al.* (1993), testa se o alelo de um marcador é transmitido por um progenitor heterozigótico para um probando afectado um número de vezes significativamente mais elevado do que o número de vezes que não é transmitido (Dekker and van Duijn, 2003; Gordon and Finch, 2005) (figura 2). Para que o teste seja informativo os progenitores têm de ser heterozigóticos para o marcador em estudo.

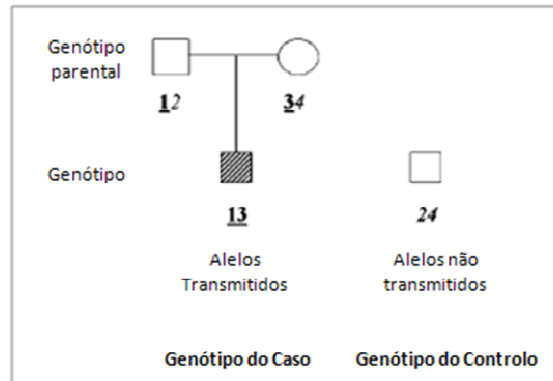


Figura 2. Princípios do teste TDT. Os alelos não transmitidos do indivíduo afectado (quadrado branco) são usados para construir os genótipos dos controlos internos. Um alelo está associado a uma determinada doença se for mais frequentemente transmitido ao indivíduo afectado comparativamente ao controlo interno (adaptado de Decker and van Duin, 2003).

O TDT testa a hipótese nula de que qualquer alelo de um marcador em análise será transmitido ao acaso (isto é, 50% das vezes) pelos pais heterocigóticos para a descendência afectada, isto é, não existe *linkage* e/ou associação entre a doença e o marcador. A hipótese alternativa é de que o alelo será transmitido mais frequentemente que o esperado, ou seja, existe *linkage* e associação entre a doença e o *locus* do marcador. O método estatístico é um teste de McNemar com uma distribuição de chi-quadrado, cujo resultado significativo indica que o marcador testado está na vizinhança de um *locus* de susceptibilidade à doença (Borecki and Suarez, 2001; Gordon and Finch, 2005).

2.4. Rastreamento genómico

Na ausência de bons genes candidatos e de pouco conhecimento sobre a etiologia do autismo, a estratégia de análise mais apropriada envolve todo o genoma através dos rastreios genómicos. A primeira estratégia de análise utilizada nas doenças complexas foi o rastreamento genómico para detecção de *linkage*, que permite identificar regiões dos cromossomas ou uma vasta região genómica que contenha vários genes associados à doença. Espera-se que os marcadores genéticos nos indivíduos afectados pela doença, dentro da mesma família sejam partilhados mais frequentemente do que seria expectável se a segregação fosse independente. Esta é a indicação de que nessa determinada região cromossómica existem genes que causam ou aumentam o risco de susceptibilidade à doença (Thomson, 2001). Desde o primeiro rastreamento genómico de *linkage*, em 1998 (*International Molecular Genetic Study of Autism Consortium*) foram publicados vários estudos com populações maioritariamente provenientes da Europa e Estados Unidos da América. As regiões de interesse situam-se nos cromossomas 1p, 2q, 5q, 7q, 15q, 16p, 17q, 19p e Xq (Klauck, 2006; Yang and Gill, 2007). Actualmente, esta metodologia foi abandonada e só se usa quando existem dados referentes a várias gerações afectadas dentro

de famílias. A sua substituta denomina-se por rastreio genómico para associação (*Genome-wide association studies*, GWAS), método este que utiliza milhões de marcadores genéticos e requer amostras populacionais de grande dimensão.

Para compreender de que forma as variantes genéticas contribuem para a diversidade fenotípica de uma doença, foram iniciados vários estudos de larga escala que identificaram associações estatísticas em muitos *loci* do genoma humano em doenças complexas. Os contínuos avanços nas tecnologias de genotipagem e o aparecimento de melhores algoritmos de análise contribuíram para que os GWAS se tornassem possíveis. Estes são estudos que exploram as relações entre as variantes genéticas comuns (SNPs) ao longo de todo o genoma e a predisposição à doença, com o objectivo de mapear regiões de susceptibilidade. A interpretação dos resultados tem permitido identificar novos *loci* de susceptibilidade, novos mecanismos biológicos e compreender mais sobre a complexidade das relações entre a componente genética e a heterogeneidade fenotípica. Muitos dos desafios que se colocam nos próximos anos será a identificação de sinais de associação e, talvez mais relevante, definir os mecanismos moleculares que influenciam o risco de susceptibilidade à doença e/ou a expressão fenotípica (McCarthy *et al.*, 2008). Os GWAS são mais um gerador de hipóteses do que um método que testa uma hipótese (Seng and Seng, 2008). É importante reconhecer que não é esperado encontrar todas as variantes causais, mas sim encontrar padrões de variantes causais comuns na população de casos em relação aos controlos (Manolio *et al.*, 2009). É indiscutível o papel destes estudos nas doenças complexas, no seu grande contributo na identificação de novas variantes de susceptibilidade, nos esforços feitos para o recrutamento de maiores amostras populacionais, identificação de regiões de LD e *tag* SNP no genoma, no desenvolvimento de novas e melhores plataformas tecnológicas e novas metodologias de análise bioinformática (Hirschhorn and Daly, 2005; McCarthy *et al.*, 2008; Frazer *et al.*, 2009; Manolio *et al.*, 2009). Comparativamente aos estudos de *linkage* e de genes candidatos, os GWAS têm um maior poder e eficácia no rastreio do genoma humano, utilizando centenas de milhares de SNP em milhares de indivíduos, tornando-se uma das ferramentas de investigação da arquitectura genética nas doenças complexas.

A hipótese por detrás deste método é a de "*Common Disease-Common Variant*" (CDCV) (Pritchard and Cox, 2002). Considera que as múltiplas variantes causais comuns, existentes tanto na população de casos como de controlos, estão na base de doenças comuns em associação com factores ambientais. O GWAS é realizado com variantes genéticas comuns, isto é, a frequência do alelo menos comum (*Minor Allele Frequency*, MAF) é superior a 5%. Dos 11 milhões de SNPs conhecidos no genoma humano, apenas 20% destes são detectáveis nas plataformas de genotipagem de alta densidade e não é possível detectar as variantes raras e novas variantes genéticas, que se caracterizam por valores de MAF entre 0,1% e 3% e inferior a 0,1%, respectivamente (Frazer *et al.*, 2009) (figura 3).

Frequentemente estas variantes comuns têm baixo efeito de risco na doença (*Effect size*) (Figura 3). Este parâmetro é medido por *Odds ratio* (OR) que representa a probabilidade de risco da doença quando se verifica um determinado genótipo específico no indivíduo. Os valores de OR obtidos no GWAS (para doenças complexas) variam aproximadamente entre 1 e 1,5, o que significa que a presença de determinado genótipo aumenta em 1,5 o risco de ter a doença. Comparativamente, nas variantes raras os valores de OR sobem para 3 até 5. (Bodmer and Bonilla, 2008; Manolio *et al.*, 2009). O tamanho da amostra populacional é também um factor limitante. Os GWAS com amostras populacionais entre 2000 e 5000 indivíduos têm poder estatístico para detectarem variantes comuns. O aumento da amostra permite aumentar o seu poder, embora os estudos feitos indiquem que mesmo uma amostra com 100 000 indivíduos, se encontrarão factores genéticos que apenas podem explicar 10-15% da variancia genética de uma doença (Frazer *et al.*, 2009).

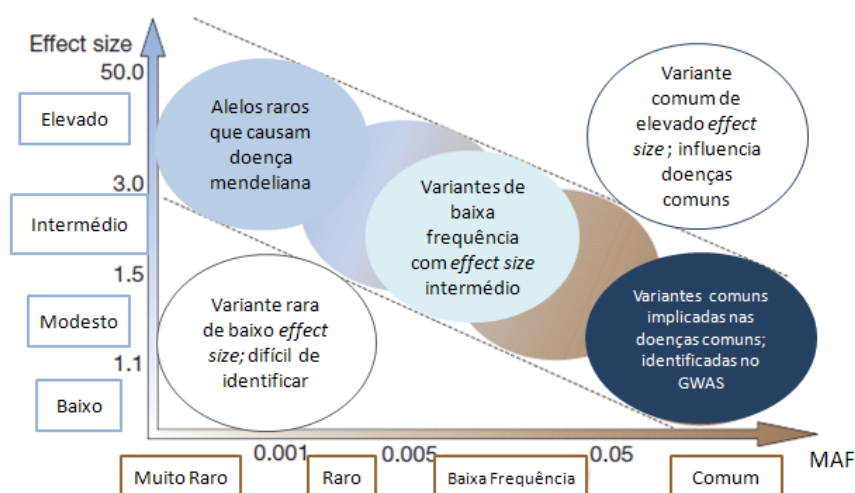


Figura 3. Esquema representativo da classificação de variantes genéticas segundo a frequência alélica do alelo menos comum (MAF) e *effect size* (OR) (adaptado de Manolio *et al.*, 2009).

Os GWAS realizados revelaram resultados novos e promissores no conhecimento genómico de algumas doenças complexas (Dermitzakis and Clark, 2009). Segundo a recente base de dados de GWAS, pertencente ao *National Human Genome Research Institute*, os resultados dos diversos estudos revelam que mais de 300 sinais de associação foram replicados para mais de 70 doenças complexas ^[1]. A título de exemplo, na doença de Crohn há mais de 30 *loci* associados (Barrett *et al.*, 2008) e 20 *loci* de susceptibilidade na Diabetes Tipo 2 (Zeggini *et al.*, 2008). Na esquizofrenia e doença bipolar os resultados dos GWAS mostram três genes significativamente associados ao risco das doenças: *ZNF804A* (associado à esquizofrenia) e *CACNA1A* e *ANKK3*, ambos associados à doença bipolar (Moskvina *et al.*, 2009; O' Donovan *et al.*, 2009). Na doença de Alzheimer, e após oito GWAS publicados na literatura, foram revelados 20 *loci* de susceptibilidade à doença, de entre os quais três já se encontram replicados noutras

populações: 14q32.13, 14q31.2 e 6q24.1 (Bertram and Tanzi, 2009) e identificaram os genes *GAB2*, *GALP*, *PGBD* e *TNK1*. O primeiro GWAS no autismo realizou-se em 2006 (Lauritsen *et al.*, 2006) numa amostra de 12 indivíduos afectados e 44 controlos (numa população isolada da micronésia). Os microsatélites usados no estudo estavam distribuídos por 14 regiões cromossómicas e foram identificadas 18 *loci* de susceptibilidade (Yang and Gill, 2007). Desde então, foram realizados outros GWAS que permitiram identificar dois *loci* de susceptibilidade à doença, as regiões 5p15.2 (Weiss *et al.*, 2009) e 5p14.1 (Ma *et al.*, 2009; Wang *et al.*, 2009). Muitos dos rastreio genómicos para associação vêm corroborar os resultados dos picos de *linkage*, nas doenças complexas.

2.5. O consórcio *Autism Genome Project*

Sendo o autismo uma doença complexa, a identificação dos seus factores de risco requer uma grande amostra populacional bem caracterizada e uma cooperação internacional elevada. Em 2003, investigadores de mais de 50 instituições espalhados pelos Estados Unidos da América, Europa (Portugal, França, Irlanda, Suécia e Reino Unido) e Canadá uniram esforços no sentido de reunir recursos e competências a fim de progredir na compreensão do autismo, formando o consórcio *Autism Genome Project* (AGP) ^[2] cujo principal objectivo é identificar e caracterizar os genes do autismo. Na primeira fase do projecto, finalizada em 2007, o AGP reuniu o maior biobanco mundial e levou a cabo uma análise genómica de *linkage* para o mapeamento de genes de susceptibilidade para o autismo, que incluiu famílias *multiplex* (famílias com dois ou mais indivíduos afectados) e *simplex* (famílias com apenas um indivíduo afectado). A segunda fase, ainda a decorrer, teve início em 2007 e pretende rastrear todo o genoma para identificar variantes genéticas comuns e raras, assim como, variações estruturais (especificamente, *Copy Number Variations*, CNV) associadas ao autismo utilizando novas tecnologias de genotipagem e sequenciação. O primeiro estadio da segunda fase já terminou e do qual resultaram duas publicações (Anney *et al.*, 2010; Pinto *et al.*, 2010). Os resultados obtidos irão gerar hipóteses sobre os genes envolvidos e permitir a identificação das variantes genéticas e alterações em vias biológicas relevantes que, eventualmente, conduzem ao aumento de susceptibilidade no autismo, assim como, o desenvolvimento de tratamentos.

O programa *The Autism Simplex Collection* (TASC), independente do AGP, surgiu com a finalidade de armazenar o maior número de amostras biológicas possível. Toda a informação clínica e DNA dos indivíduos afectados e respectivos familiares participantes neste estudo está depositada neste biobanco. A avaliação de diagnóstico clínico é rigorosamente caracterizada de acordo com os critérios supracitados (ver secção I.1.4.). O AGP, assim como o programa TASC, são apoiados e financiados pela *Autism Speaks*, uma organização internacional fundada por familiares de pessoas com autismo que acreditam que, através da investigação a nível mundial, será possível chegar a um tratamento para esta patologia.

Para além da participação portuguesa no consórcio, o grupo de investigação tem o objectivo de reunir um elevado número de amostras da população portuguesa de indivíduos afectados e respectivos familiares provenientes de Portugal continental e Açores e que actualmente integram o Biobanco Português de Autismo. Mais de 500 famílias já pertencem a este biobanco que está sediado no Instituto Nacional de Saúde Dr. Ricardo Jorge. A informação biológica e respectiva caracterização clínica do paciente é realizada no Hospital Pediátrico de Coimbra (ver secção I.1.4.) e armazenada numa base de dados (BcGENE), anónima e de acesso restrito, no Instituto Gulbenkian de Ciência.

Em 2007, no âmbito do consórcio AGP, foi realizado um rastreio genómico de *linkage*, em 1181 famílias *multiplex* (com pelo menos dois indivíduos afectados). Os resultados revelam um pico de *linkage* na região cromossómica 11p12-p13 (Autism Genome Project Consortium, 2007) (figura 4).

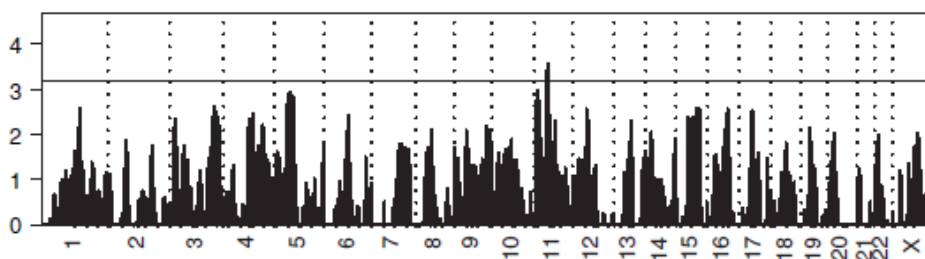


Figura 4. Resultado do rastreio genómico para *Linkage* realizado pelo consórcio AGP, para 1181 famílias *multiplex*. A linha vertical a ponteados separa os cromossomas. A linha horizontal corresponde ao LOD, valor a partir do qual o resultado da análise é considerado significativo ($LOD > 3,18$) (adaptado de Autism Genome Project Consortium, 2007).

Em 2009, um rastreio genómico para *linkage* identificou novo *loci* de susceptibilidade no autismo em 1031 famílias, na região cromossómica 20p13 (Weiss *et al.*, 2009). A replicação de resultados foi realizada na população do AGP (figura 5). A seguinte figura mostra os resultados:

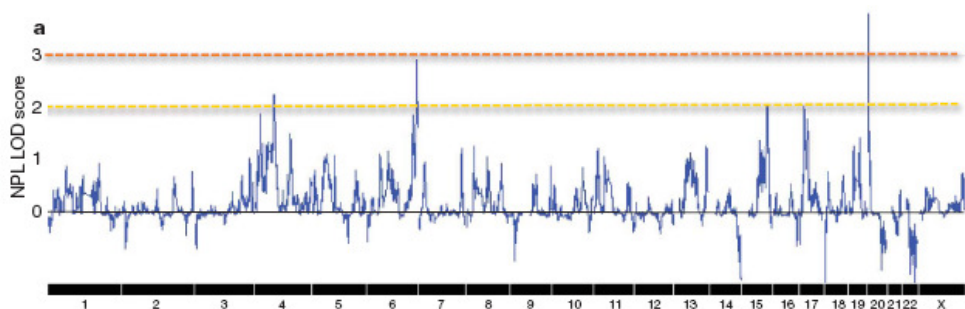


Figura 5. Resultados do rastreio genómico para *linkage*, em 1031 famílias *multiplex*, replicados na população AGP. Os números no eixo das abcissas correspondem aos cromossomas. A linha horizontal de referência (cor-de-laranja) corresponde ao valor de LOD considerado estatisticamente significativo ($LOD > 3$) (adaptado de Weiss *et al.*, 2009).

No âmbito do último rastreio genómico para associação realizado pelo AGP, utilizando o *BeadChip* da Illumina com 1 milhão de SNPs, em 1369 famílias, foi identificada uma multiplicidade de variantes genéticas raras nos pacientes com autismo (Pinto *et al.*, 2010). Este estudo detectou a existência de inúmeros CNVs (*de novo* e herdados dos progenitores) presentes mais frequentemente em indivíduos com autismo. O mesmo estudo sugeriu novos genes implicados na doença: *SHANK2*, *SYNGAP1*, *DLGAP2* e *locus DDX53-PTCHD1*. Os resultados deste estudo evidenciam a importância das variantes genéticas raras na genética do autismo. O mesmo rastreio genómico realizado pela AGP permitiu a identificação de variantes genéticas comuns utilizando como abordagem os estudos de associação. Os resultados desta análise identificaram uma forte associação com o autismo do SNP rs4141463 ($P < 5 \times 10^{-8}$), localizado no gene *MACROD2* (20p12.1) (Anney *et al.*, 2010).

3. ANÁLISE DOS RASTREIOS GENÓMICOS PARA ASSOCIAÇÃO

Os rastreios genómicos para associação publicados na literatura representam uma considerável quantidade de dados sobre diversas doenças complexas. É inegável o contributo que deram na compreensão da arquitectura genética e mostraram a importância da utilização de variantes não codificantes na etiologia destas doenças. Tal como os estudos de *linkage* já o tinham demonstrado, estas doenças não podem ser apenas explicadas por um número limitado de variantes comuns e de *effect size* moderado. Muita da análise dos resultados ainda permanece por fazer a fim de compreender de que forma as vias biológicas, genes e as suas variações contribuem para a heterogeneidade do fenótipo da doença. Enquanto não se compreender a arquitectura genética por detrás das doenças complexas, não se poderão fazer predições sobre o risco individual a uma determinada doença. No autismo, os factores genéticos contribuem cerca de 90% na etiologia da doença, no entanto, os resultados dos GWAS apenas conseguem explicar uma pequena parte da heritabilidade da doença. Desta forma, procuram-se os factores causais que explicam a ausência de heritabilidade. Estes factores incluem a identificação de variantes raras e estruturais de *effect size* moderado a elevado, a identificação de interacções entre diferentes *loci* e a identificação de vias biológicas que poderão estar implicadas no risco de susceptibilidade à doença (Maher, 2008; Manolio *et al.*, 2009; Dermitzakis and Clark, 2009).

3.1. Pesquisa de variantes raras

As variantes raras (MAF entre 0,1% e 3%) não estão em LD com variantes comuns e, consequentemente, não são detectadas no GWAS (Frazer *et al.*, 2009). No entanto, no caso dos CNVs alguns são detectados nos GWAS. Estes estudos demonstraram a importância das características destas variantes e revelaram a necessidade de desenvolver metodologia capaz

de as detectar (Bodmer and Bonilla, 2008). Evolutivamente, as variantes raras são mais recentes em relação às variantes comuns, que têm uma frequência elevada na população, o que sugere que as variantes raras são específicas de determinadas populações, originadas por efeitos fundadores, em gerações recentes. Neste sentido, diferentes populações poderão ter diferentes variantes causais na origem da doença, para explicar a heterogeneidade fenotípica da mesma. As variantes raras poderão ter maior *effect size*, em relação às variantes comuns, e associadas a determinados fenótipos explicar uma determinada característica da doença (Frazer *et al.*, 2009; Levitt and Campbell, 2009). No futuro próximo, a tecnologia para detectar variantes raras é a sequenciação total do genoma humano (*Whole Genome Sequencing*, WGS) (Li and Leal, 2009). Estas novas técnicas de sequenciação irão gerar milhões de sequências promovendo o conhecimento sobre determinadas regiões do genoma ainda desconhecidas.

3.2. Análise de epistasia genética

Epistasia define-se como a interacção entre dois ou mais genes, nos quais o efeito de um alelo particular num fenótipo depende de um outro alelo noutra *locus* diferente e o seu efeito fenotípico combinado desvia-se da aditividade. Por aditividade entende-se a soma dos efeitos de um *loci* no fenótipo (Phillips, 2008; Cordell, 2009). Os GWAS utilizam uma estratégia de análise *locus a locus*, na qual cada variante é testada individualmente para associação com um fenótipo específico. Contudo, uma das razões que explica a ausência de associação entre as variantes genéticas e a doença poderá estar relacionado com a existência de interacções entre *loci* (Phillips, 2008). Pelo facto de estarmos perante uma doença complexa, se se analisarem os genes isoladamente poderemos menosprezar as suas potenciais interacções e respectivo contributo na explicação do fenótipo. Torna-se pertinente a detecção de interacções estatísticas entre *loci* com o intuito de compreender as vias biológicas envolvidas nas doenças complexas (Moore, 2003). Vários métodos computacionais foram desenvolvidos nos últimos anos para detectar quais os genes que interagem. Pela incapacidade dos tradicionais métodos de regressão de lidar com modelos não lineares e com elevados dados amostrais com diversas variáveis surgiram os outros métodos, eficientes computacionalmente e capazes de lidar com amostras grandes (Moore, 2003; Cordell, 2009).

Alguns estudos têm utilizado a abordagem de interacção entre múltiplos *loci*. Neste sentido, espera-se que os resultados dos GWAS sejam ainda mais potenciadores deste tipo de análise. No autismo já foram reveladas evidências de epistasia estatística entre *SLC6A4* e *ITGB3* (Weiss *et al.*, 2006; Coutinho *et al.*, 2007) dois genes que codificam moléculas participantes nas vias do sistema serotoninérgico e que estão envolvidas na etiologia da doença.

3.3. Análise de vias biológicas

As vias biológicas representam uma larga variedade de processos biológicos que incluem funções celulares, processos metabólicos, biosíntese, processamento da informação genética, sinalização celular, respostas do sistema imunitário, desenvolvimento embrionário. Num contexto celular, uma via biológica representa um conjunto de acções entre moléculas que participam numa determinada função celular. Os genes codificadores das moléculas que participam numa mesma tarefa celular são agrupados numa mesma via biológica, que reflecte o processo celular. Estes genes formam conjunto de genes (*gene set*) e são frequentemente colocados em diagramas ilustrativos das vias biológicas que indicam a ordem pela qual os genes actuam e interagem dentro da via biológica. No entanto, estas listas de genes podem ser comuns a diferentes vias biológicas, o que lhes confere um carácter ambíguo e dificulta a sua análise (Cantor *et al.*, 2010). A análise de vias biológicas é a mais recente estratégia na análise de GWAS que poderá ter um contributo bastante significativo na compreensão da etiologia de doenças complexas, que não será possível numa análise *locus a locus*. A análise de vias biológicas poderá conduzir a novas descobertas de vias biológicas, gerar novas hipóteses, introduzir novas questões de índole biológica e funcional na análise de uma doença complexa.

Levitt and Campbell (2009) referem várias vias biológicas associadas à patofisiologia do autismo, como por exemplo, disfunções em vias biológicas responsáveis pelo desenvolvimento e formação de sinapses, o que inclui variações estruturais em proteínas que participam na sinapse, como neuroliginas e neuroxinas. Correia *et al.* (2010) evidenciam que alterações na via de sinalização BDNF/TrkB contribuem para a susceptibilidade ao autismo e sugere que as cascatas de sinalização por estes activadas (as vias de sinalização de MAPK- MEK/ERK e PI3K-AKT e PLC γ) poderão também estar alteradas (Levitt and Campbell, 2009). Algumas variações genéticas nestas vias de sinalização já estão associadas ao autismo e a algumas patologias como Neurofibromatose, Macrocefalia, Esclerose tuberosa (via de sinalização PI3K), encontradas em indivíduos com o autismo síndromico (Levitt and Campbell, 2009).

4. FACTORES NEUROTRÓFICOS ENVOLVIDOS NO AUTISMO

Vários estudos têm sido realizados com o intuito de elucidar os principais mecanismos neurofisiológicos subjacentes aos défices cognitivos associados ao autismo. A sua heterogeneidade fenotípica coloca a hipótese de diferentes mecanismos genéticos influenciarem o desenvolvimento do circuito neuronal, a diferentes níveis. A vulnerabilidade genética reside na perturbação de processos celulares, devido à alteração de proteínas codificadas por genes em neurocircuitos específicos que poderão ser influenciados pela interacção entre factores

genéticos e ambientais. Estudos de genética, em humanos e animais, sugerem que a alteração de vias metabólicas e processos biológicos chave no desenvolvimento e formação sináptica (sinaptogénese) representam um dos possíveis caminhos na compreensão da etiologia do autismo (Levitt and Campbell, 2009).

4.1. Via de sinalização BDNF/TrkB

O BDNF (*Brain derived neurotrophic factor*) é um membro da família das neurotrofinas com elevada expressão no hipocampo e córtex cerebral da criança e jovem adulto. Esta família desempenha importantes funções na formação do sistema nervoso central e periférico, onde controlam a função sináptica e plasticidade, a sobrevivência, morfologia e diferenciação neuronal e estão envolvidos (em especial o BDNF) nos processos de aprendizagem e memória (Yamada and Nabeshima, 2003; Reichardt, 2006). Participam na regulação de mecanismos celulares, crescimento dos axónios e dendrites e expressão de proteínas. A família é constituída por NGF (*Nerve Growth Factor*), BDNF, NT-3 (*Neurotrophin-3*) e NT-4 (*Neurotrophin-4*). Os efeitos do BDNF são mediados pela sua ligação ao receptor de alta afinidade, TrkB (*tropomyosin-related kinase B*, tirosina cinase) que é membro da família dos receptores de tirosina cinase (RTK), em conjunto com TrkA e TrkC (Bibel, 1999). O receptor completo (*full-length*) é predominantemente expresso no cérebro, embora existam várias isoformas (cerca de 36 isoformas conhecidas), resultado de mecanismos de *splicing* alternativo que diferem nos domínios intracelulares e nas regiões onde são expressas (Stoilov *et al.*, 2002; Luberg *et al.*, 2010) (ver anexo A, figura 17). As formas truncadas mais comuns, TrkB-T1 e TrkB-t-Shc, não têm domínio intracelular (tirosina cinase) não permitindo a sua ligação correcta ao citoplasma e, possivelmente, não permitindo a sinalização de outras vias biológicas. A variante TrkB-T1 é expressa em vários tecidos, incluindo o tecido neuronal enquanto a variante TrkB-T-Shc apenas é expressa no cérebro (Stoilov *et al.*, 2002). O domínio extracelular do receptor contém duas regiões ricas em cisteína, separadas por um domínio rico em leucina, seguidas pelo domínio transmembranar e intracelular, domínio de cinase de tirosina (ver anexo A, figura 18).

Os domínios citoplasmáticos dos receptores TrkB contêm várias tirosinas que são substratos para a fosforilação. Quando fosforilados estes resíduos formam locais específicos de ligação para enzimas e *scaffolds*, que são intermediários na activação das cascatas de sinalização (Reichardt, 2006). As neurotrofinas provocam a dimerização do receptor e o aumento na autofosforilação, levando à activação dos domínios intracelulares de três cascatas de sinalização intracelulares: a activação mediada por sch/Frs2 e Ras/Raf da via de MAPK que envolve MEK e ERK; a activação mediada por shc/Frs2 e GAB1 da via PI3K e, a jusante a via de Akt; e, por último, a única via de sinalização activada directamente pelo receptor, PLC γ , que produz IP3 (*Inositol triphosphate*) e DAG (*diacylglycerol*) aumentando os níveis de cálcio intracelular (Reichardt, 2006; Gokce *et al.*, 2009) (ver figura 6 e anexo B). As vias de sinalização de MAPK e

PI3K têm um papel fundamental na síntese, tradução e/ou transporte de proteínas com funções sinápticas e a via de sinalização PLC γ regula o cálcio intracelular que conduz o processo de transcrição através de AMP cíclico e proteína cinase C (Yoshii and Constantine-Paton, 2010).

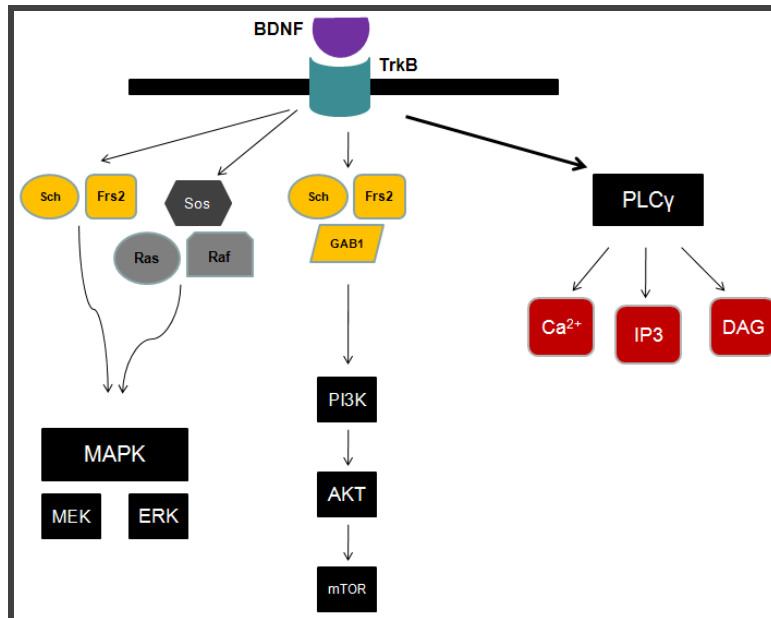


Figura 6. Esquema representativo das três vias biológicas activadas pela ligação da neurotrofina BDNF ao seu receptor de alta afinidade, TrkB. Nos rectângulos pretos estão representados as 3 vias activadas: MAPK/MEK-ERK, PI3K-AKT e PLC γ .

4.2. Papel da via de sinalização BDNF/TrkB no autismo

A via de sinalização BDNF/TrkB está envolvida na transcrição e tradução de proteínas durante várias fases do desenvolvimento sináptico, plasticidade sináptica, sobrevivência neuronal e migração celular (Yoshii and Constantine-Paton, 2010), através das três vias biológicas que activa. Pelas funções que desempenham, os genes que codificam as moléculas participantes nestas três vias biológicas são potenciais candidatos na etiologia das ASD.

A título de exemplo, a via biológica PI3K desempenha funções na regulação do ciclo celular, sobrevivência, diferenciação e proliferação celular e AKT participa na regulação de proteínas apoptóticas e factores de transcrição. Estão descritas na literatura mutações nos genes *TSC1*, *TSC2*, *NF1* e *PTEN* associadas ao autismo síndromico (Esclerose tuberosa, neurofibromatose e macrocefalia), que participam na via biológica PI3K-AKT (Rodger and Theibert, 2002; Leslie and Downes, 2002; Haapasalo *et al.*, 2002; Levitt and Campbell, 2009; Bourgeron, 2009).

Alterações na neurotrofina BDNF têm sido associadas à etiologia de várias doenças neuropsiquiátricas como a esquizofrenia, a depressão e doença bipolar (Pardo and Eberhart, 2007; Dmitrzak-Weglarz *et al.*, 2008). Foram descritos em diversos estudos, e também na

população portuguesa, níveis significativamente mais elevados de BDNF no plasma de crianças com autismo comparativamente à população controlo (Nelson *et al.*, 2001; Correia *et al.*, 2010). Em 2007, o consórcio AGP realizou um rastreio genómico de *linkage* e encontrou um pico de *linkage* na região cromossómica 11p12-13, onde está localizado o gene *BDNF* (AGP, 2007). No mesmo ano um outro estudo (Nishimura *et al.*, 2007) reporta o sinal de associação entre um haplótipo de SNP pertencentes ao gene *BDNF* e o autismo.

O receptor TrkB, codificado por *NTRK2*, medeia processos de crescimento, diferenciação e apoptose celular no SNC e periférico. Algumas variantes genéticas de *NTRK2* estão associadas com o autismo, a depressão, a doença bipolar, a doença de Alzheimer e a doença obsessiva-compulsiva (Alonso *et al.*, 2008; Chen *et al.*, 2008; Dmitrzak-Weglaz *et al.*, 2008; Smith *et al.*, 2009; Kohli *et al.*, 2010). Na população portuguesa, Correia *et al.* (2010) encontraram evidência de associação com o autismo entre variantes genéticas do gene *NTRK2* e o autismo, num SNP localizado na região promotora (5') e os restantes nas regiões intrónicas 5-6-7-8 e 22-23. O estudo verifica que a presença do alelo T no SNP da região promotora de *NTRK2* (SNP associado à doença) altera o local de ligação de factores de transcrição das famílias HOMF (*Homeodomain Transcription Factors*) e HOXF (*Factors with moderate activity to homeodomain consensus sequence*). Estes factores de transcrição são codificados por genes *homeobox* (com aproximadamente 180bp) que codificam domínios proteicos ou homeodomínios e, quando expressos, ligam-se a sequências específicas de DNA. Foi encontrada associação entre o gene *EN1* (*Engrailed homeobox 1*) e a doença de Parkinson (Haubenberger *et al.*, 2009) e associação entre os marcadores dos genes *DLX1* (*Distal-less homeobox 1*), *DLX2* (*Distal-less homeobox 2*) (todos pertencentes à família HOMF), *PITX1* (*Paired-like homeodomain transcription factor 1*) e *HOXA1* (*Homeobox protein Hox-A1*) (pertencentes à família HOXF) e o autismo (Conciatori *et al.*, 2000; Ingram *et al.*, 2000; Philippi *et al.*, 2007; Liu *et al.*, 2009), embora existam estudos com resultados contraditórios (Devlin *et al.*, 2002; Li *et al.*, 2002; Romano *et al.*, 2003; Gallagher *et al.*, 2004).

Desta forma, coloca-se a hipótese de que alterações ao nível da ligação BDNF/TrkB e/ou ao nível das três vias biológicas MAPK/MEK/ERK, PI3k-AKT e PLC γ poderão levar à disrupção dos mecanismos de desenvolvimento sináptico e, desta forma, contribuirão como factores de susceptibilidade na etiologia do autismo (Levitt and Campbell, 2009; Yoshii and Constantine-Paton, 2010).

II. OBJECTIVOS

Correia *et al.* (2010) descreveram pela primeira vez uma associação do gene *NTRK2* com o autismo e a alteração dos níveis de BDNF (no soro) nesta doença, na população portuguesa. Sugerem a desregulação na via de sinalização que envolve a neurotrofina e o seu receptor (BDNF/TrkB) e a possível alteração de uma ou várias das três vias de sinalização por si activadas. A hipótese de que a alteração da via de sinalização BDNF/TrkB está implicada no autismo é a ideia unificadora e ponto de partida para este estudo. O objectivo principal do presente estudo é a identificação dos genes de susceptibilidade para o autismo envolvidos nas vias biológicas activadas por BDNF /TrkB.

Os objectivos específicos deste estudo são:

1. Identificação de variantes genéticas do gene *NTRK2* nas regiões anteriormente associadas ao autismo.
2. A análise da associação genética dos genes reguladores da via de sinalização BDNF/TrkB com o autismo, com ênfase nos genes das famílias de factores de transcrição HOMF e HOXF e no gene *NTRK2*.
3. A análise da associação genética dos genes envolvidos nas vias biológicas induzidas por BDNF/TrkB com o autismo.
4. A análise da interacção genética entre os genes *NTRK2*, *DLX1*, *DLX2*, *EN1*, *HOXA1*, *PITX1*, assim como entre os genes *NTRK2* e *BDNF* que contribuem na susceptibilidade ao autismo.
5. A análise do enriquecimento do sinal de associação *in silico* das vias biológicas de interesse e de outras vias biológicas implicadas na susceptibilidade ao autismo.

A estratégia metodológica do presente estudo dividiu-se nas seguintes etapas:

- a. Sequenciação das diferentes regiões de susceptibilidade do gene *NTRK2*, numa amostra populacional portuguesa.
- b. Com base nos resultados do GWAS obtidos a partir do consórcio AGP realizaram-se:

- Testes de associação genética baseados em famílias nos genes envolvidos nas vias biológicas induzidas pela ligação BDNF/TrkB.
- Testes de interação genética entre os gene *homeobox* e *NTRK2*, assim como entre *BDNF* e *NTRK2*.
- Análise de enriquecimento do sinal de associação *in silico* das vias biológicas.

III. MATERIAIS E MÉTODOS

1. RECRUTAMENTO E CARACTERIZAÇÃO CLÍNICA DE PACIENTES

O rastreio de variações genéticas foi realizado numa sub-amostra populacional portuguesa de 100 pacientes com autismo (idade entre 2 e 18 anos). Estes indivíduos e os pais não afectados pertencem à amostra populacional portuguesa contida num biobanco português do autismo, sediado no Instituto Nacional de Saúde Dr. Ricardo Jorge. As amostras são provenientes de Portugal Continental e Açores e foram recrutadas no Hospital Pediátrico de Coimbra. As famílias participantes no estudo, maioritariamente caucasianas, foram diagnosticadas e caracterizadas de acordo com os seguintes instrumentos de avaliação clínica para o autismo: a entrevista de diagnóstico de autismo revista (*Autism Diagnostic Interview-Revised*, ADI-R) (Lord *et al.*, 1994) e a escala de diagnóstico de observação de autismo e distúrbios do espectro do autismo, (*Autism Diagnostic Observation Schedule*, ADOS) (Lord *et al.*, 2000). Os quocientes de inteligência e desenvolvimento foram determinados através da Escala II de Ruth Griffiths de desenvolvimento mental (*Ruth Griffiths Mental Scale II*) ou da escala de inteligência de Wechsler para crianças, WISC (*Wechsler Intelligence Scale for Children*) (Griffiths, 1984; Marques, 1970). Foram excluídos indivíduos com esclerose tuberosa, síndrome do X-Frágil, anomalias cromossómicas, distúrbios metabólicos e rubela congénita. A população controlo foi recrutada no Centro Regional de Sangue de Lisboa, maioritariamente caucasiana e tem idade média de 40 anos.

O GWAS foi realizado pelo consórcio AGP que engloba indivíduos provenientes da América do Norte (Canadá e E.U.A.) e Europa (Portugal, França, Reino Unido, Irlanda). Este consórcio mundial reuniu 1369 famílias com ASD e os respectivos familiares não afectados. O diagnóstico da amostra populacional portuguesa foi realizado pela equipa do Hospital Pediátrico de Coimbra, segundo os critérios supracitados. A amostra portuguesa estudada no AGP inclui um total de 229 famílias, compostos pelo paciente com autismo e os respectivos pais não afectados, provenientes de Portugal Continental e Açores.

2. EXTRACÇÃO E QUANTIFICAÇÃO DE DNA GENÓMICO

A extracção de DNA dos indivíduos que constituem as duas amostras populacionais foi realizada a partir de amostras de sangue manuseadas em câmara de fluxo laminar. A extracção foi realizada por um processo de *salting-out*, segundo o protocolo adaptado do método descrito

por Lahiri e Nurnberger (1991). A composição das soluções preparadas, assim como, o protocolo de extracção encontram-se no anexo C. Após a extracção de DNA, a sua concentração foi estimada através do uso do espectrofotómetro *Nanodrop* ND-1000. Este aparelho detecta e quantifica ácidos nucleicos com elevada precisão necessitando apenas 1µl de volume da amostra, sem recurso a diluições, cuvetes ou capilares. É possível aferir a qualidade do DNA, através dos rácios de absorvância 260/230 nm e 260/280 nm, determinados pelo espectrofotómetro. As soluções *stock* de DNA são armazenadas a 4°C assim como as soluções trabalho, preparadas posteriormente, com uma concentração final de 25 ng/µl e mantidas a 4°C.

3. SEQUENCIAÇÃO DO GENE *NTRK2*

O processo pelo qual se determina a ordem das bases nucleotídicas ao longo da cadeia de DNA denomina-se por sequenciação, sendo aqui utilizado, o método de reacção em cadeia denominado de método de Sanger. Foram sequenciados quatro produtos de *Polymerase Chain Reaction* (PCR) referentes a cinco exões do gene *NTRK2*, com o objectivo de identificar a existência de polimorfismos nas regiões exónicas e respectivas fronteiras intrónicas.

3.1. Selecção de indivíduos para sequenciação

Correia *et al.* (2010) detectou forte sinal associação de dois haplótipos do gene *NTRK2* no autismo. Os haplótipos estão representados na tabela 1:

Tabela 1. Haplótipos de risco do gene *NTRK2* associados ao autismo por Correia *et al.*, 2010.

Haplótipo	<i>P-value</i>	<i>Over-transmitted</i>
[1] rs893584_rs1187352	0,0003	CA
[2] rs4271046_rs4578034	0,0059	GC

O haplótipo [1] encontra-se na região dos exões 6, 7 e 8 (correspondente ao domínio de ligação extracelular), estando o haplótipo [2] na região dos exões 22 e 23 (correspondente à região catalítica, domínio de cinase de tirosina). Foram seleccionados os indivíduos homocigóticos para cada haplótipo de risco, respectivamente CA CA e GC GC, e os indivíduos heterocigóticos TG CA, TA CA, CG CA e GC AC, GC GT, respectivamente, num total de cem indivíduos para cada exão.

3.2. Optimização das condições de PCR

O PCR é um método de amplificação enzimática exponencial de DNA, *in vitro*, que resulta na obtenção de grandes quantidades de DNA. A amplificação ocorre em ciclos repetidos de três

passos consecutivos denominados por desnaturação, hibridação (ou *annealing*) e extensão. Os reagentes necessários para a reacção incluem uma DNA polimerase (Bioline), quatro trifosfatos de desoxirribonucleótidos em concentrações equimolares (Promega), DNA molde (neste caso, genómico), cloreto de magnésio (Bioline) e solução tampão (Bioline). A técnica requer também duas pequenas sequências de oligonucleótidos de iniciação (*primers*) que são específicos na região de interesse. Os *primers* utilizados foram desenhados com auxílio do programa *on line* Primer3 (disponível em <http://frodo.wi.mit.edu/primer3>). A seguinte tabela mostra as características dos *primers* (Eurofins MWG) utilizados para as reacções de PCR dos cinco exões. De notar que os exões 6 e 7 por serem muito pequenos e próximos foram amplificados num único fragmento de PCR, incluindo o intrão entre estes (tabela 2).

Tabela 2. Sequências e temperaturas de hibridação dos *primers forward* e *reverse* utilizados para os exões e respectivo tamanho do produto de PCR amplificado.

Exão	Primers	Tm	Tamanho produto PCR amplificado
6/7	5' – GTCAGTGCATTCACTCTCT -3' (F)	58°C	368bp
	5' – GAAGGAAAATAAACAGAGGG – 3' (R)	58°C	
8	5' – TGAAAGAGAGAGAGATCTGGATG- 3' (F)	58,9°C	356bp
	5' – TGAGAGACTGAAACTTGAGAAATCA – 3' (R)	58,1°C	
22	5' – AGTCAGGCAGCATCTTTAGC -3' (F)	63,7°C	432bp
	5' – CCTCAGCAAACAATACAGCA – 3' (R)	63,4°C	
23	5' – GCAAATAAGGAAAGCAAACAG – 3' (F)	62,3°C	365bp
	5' – CCCATAAGATCGTGTCTGAC – 3' (R)	63,4°C	

Legenda: bp- pares de bases nucleotídicas; Tm – Temperatura de hibridação

O procedimento de PCR inicia-se com a desnaturação, a 95°C, de 37,5ng de DNA juntamente com 2µl água bidestilada. De seguida, a 4°C para evitar a renaturação, adicionou-se a mistura de reacção, perfazendo um volume final de reacção de 25µl. Os volumes (em µl) e as respectivas concentrações dos componentes da reacção estão na tabela 3.

Tabela 3. Mistura de reacção de PCR, para os cinco exões do gene *NTRK2* (todos os valores na tabela são referentes a volumes em µl).

	Exão 6/7	Exão 8	Exão 22	Exão 23
Água Bidestilada	15,55	15,125	13,125	15,875
Buffer 10x	2,5	2,5	2,5	2,5
MgCl ₂ [50mM]	0,75 [1,5mM]	1,25 [2,5mM]	1,25 [2,5mM]	0,5 [1mM]
dNTPs [2mM]	0,5 [0,04mM]	0,5 [0,04mM]	2,5 [0,2mM]	0,5 [0,04mM]
Primers [10mM]	1+1 [0,4mM]	1+1 [0,4mM]	1+1 [0,4mM]	1+1 [0,4mM]
Taq	0,125 (0,625U)	0,125 (0,625U)	0,125 (0,625U)	0,125 (0,625U)

As condições usadas na reacção de PCR estão descritas na tabela 4.

Tabela 4. Condições de reacção de PCR para os cinco exões do gene *NTRK2*.

Etapas do ciclo	Tempo	T°C	Exão6/7	Exão8	Exão22	Exão23	
Desnaturação inicial	5'	95°C					
Desnaturação	30"	95°C					30 ciclos
<i>Annealing</i>	30"		57°C	59°C	59°C	57°C	
Extensão	30"	72°C					
Extensão final	10'	72°C					

A confirmação dos produtos de PCR amplificados foi realizada por electroforese em gel de agarose, uma técnica de separação de moléculas que envolve a migração das mesmas, mediante a aplicação de uma diferença de potencial. As moléculas são separadas de acordo com o seu tamanho, sendo que as de menor massa migram mais rapidamente no gel. A concentração do gel de agarose é dependente do tamanho das moléculas, tendo sido utilizado géis de 1,5% (1,5g de agarose para cada 100ml de TBE). O agente intercalante usado foi *Sybr safe* (*Invitrogen*).

3.3. Reacção de sequenciação

Após a reacção de PCR, um volume de 2,5µl dos produtos amplificados foi purificado, a fim de retirar nucleótidos e *primers* em excesso, com recurso a uma combinação de duas enzimas: uma exonuclease e uma fosfatase alcalina (*Fermentas*). As enzimas são activadas a 37°C durante 15 minutos e em seguida colocadas a 80°C, 15 minutos, para a sua inactivação. Seguidamente, preparou-se a mistura de reacção, para cada um dos *primers*, com um volume final de 10µl. Cada reacção contém 1µl de produto de PCR purificado e os restantes componentes estão na tabela 5:

Tabela 5. Mistura de reacção de sequenciação, com volume final de 10µl.

Componentes	Volume (µl)
Big Dye Terminator v1.1 (Applied Biosystems)	2
<i>Primer Forward</i> ou <i>Reverse</i> (Bonsai)	1
Água bidestilada	6
Produto de PCR purificado	1µl

Para cada indivíduo, foram sequenciadas as cadeias *forward* e *reverse*, num total de 100 indivíduos com autismo, para cada exão. Após juntar-se a mistura de reacção e o produto de PCR purificado os tubos foram colocados no termociclador, segundo as condições descritas na seguinte tabela:

Tabela 6. Condições da reacção de sequenciação.

Etapas do ciclo	Tempo	T°C	
Desnaturação inicial	1'	96°C	
Desnaturação	10"	96°C	
<i>Annealing</i>	5"	50°C / 59°C (*)	25 ciclos
Extensão	4'	60°C	

(*)Temperatura de *annealing* para o exão 22.

Após a reacção de sequenciação, as amostras seguiram para o serviço de sequenciação, onde foi realizada a análise das amostras no sequenciador automático ABI 3130XL da Applied Biosystems.

A análise das sequências foi efectuada com recurso ao programa informático *Staden Package* (Staden, 1996 – <http://www.mrc-lmb.cam.ac.uk:80/pubseq>). O programa emparelha as sequências de DNA de cadeia simples obtidas, comparando-as simultaneamente com a sequência padrão do exão (retirada da base de dados genética ENSEMBL). Dentro do pacote existe o programa Pregap que avalia a qualidade das sequências e formata-as compativelmente com o programa de análise Gap4. As suas funcionalidades permitem visualizar os cromatogramas de sequenciação, realçar as diferenças entre sequências e remover os extremos das sequências nos quais a ordem dos nucleótidos não é correctamente identificada pelo sequenciador. Nos indivíduos onde foram encontradas alterações repetiu-se a reacção de sequenciação para confirmação.

4. GENOTIPAGEM NO RASTREIO GENÓMICO AGP

O processo de genotipagem de DNA foi realizado através da plataforma Illumina para mais de um milhão de SNP "Infinium Human 1M-single Infinium BeadChip" que permite uma cobertura muito densa do genoma humano (Tabela 7). A selecção dos marcadores é feita pelo fabricante a partir das bases de dados RefSeq (NCBI) e ENSEMBL e do projecto HapMap. Incluem *tag* SNP, SNP sinónimos e não sinónimos localizados a menos de 10kb dos genes referenciados na RefSeq, assim como SNP em regiões de *Copy Number Variations* (CNV) e outros marcadores para preencher os espaços entre os anteriores, permitindo uma análise mais completa, uniforme e abrangente do genoma. Foram ainda incluídos marcadores localizados nos genes da região do Complexo de Histocompatibilidade Maior e marcadores localizados nos 333 genes denominados ADME (*Absorption, Distribution, Metabolism and Excretion*), isto é, genes anteriormente associados em estudos de farmacogenómica. A distância de separação entre marcadores situa-se entre os 1,7Kb e os 2,7 Kb para assegurar a cobertura de todo o genoma.

A descrição detalhada do tipo e respectivo número de marcadores genéticos presentes no *chip* está na tabela 7.

Tabela 7. Características dos marcadores genéticos presentes no *Infinium Human 1M-single Infinium BeadChip* da Illumina (adaptado de Infinium® DNA Analysis Beadchips).

Conteúdo	Número
Todos os marcadores	1072820
SNP dos genes referenciados¹	565718
SNP não sinónimos²	23288
SNP em regiões de CNV previamente descritos³	206665
Sondas e SNP em regiões candidatas para CNV	~52000
SNP no genoma mitocondrial	163
SNP nos cromossomas X/Y/PAR <i>region</i>⁴	40097/2283/686
SNP localizados a menos de 10kb de um gene ADME	15468
SNP na região do Complexo de Histocompatibilidade Maior	10073

¹ Marcadores localizados a menos de 10Kb dos genes descritos na base de dados RefSeq (NCBI).

² Informação dos marcadores retirada das bases de dados RefSeq e ENSEMBL.

³ Regiões de CNV descritas na base de dados DGV (*Database of Genomic Variants*).

⁴ PAR *region* - região de homologia entre os cromossomas X e Y.

Para cada *BeadChip* são genotipados os marcadores de 1 indivíduo, sendo necessário 750 ng de DNA genómico por amostra. De forma sucinta, o método inicia-se com a amplificação e incubação do DNA genómico isotermicamente (ver figura 8, passos 1 e 2). O produto de amplificação é fragmentado através de um processo enzimático (passo 3). De seguida, o precipitado de DNA é ressuspensão e inicia-se a preparação e colocação da amostra no *BeadChip* (passos 4 e 5). O passo 6 é de hibridação: cada tipo de sonda corresponde a cada alelo de um determinado *locus*. A especificidade alélica é conferida através de uma reacção enzimática e é marcado com fluorescência (passo 7). A intensidade da fluorescência emitida pelas sondas é detectada pelo *Illumina BeadArray Reader* e analisada no software *Bead Studio* da Illumina (passos 8 e 9).

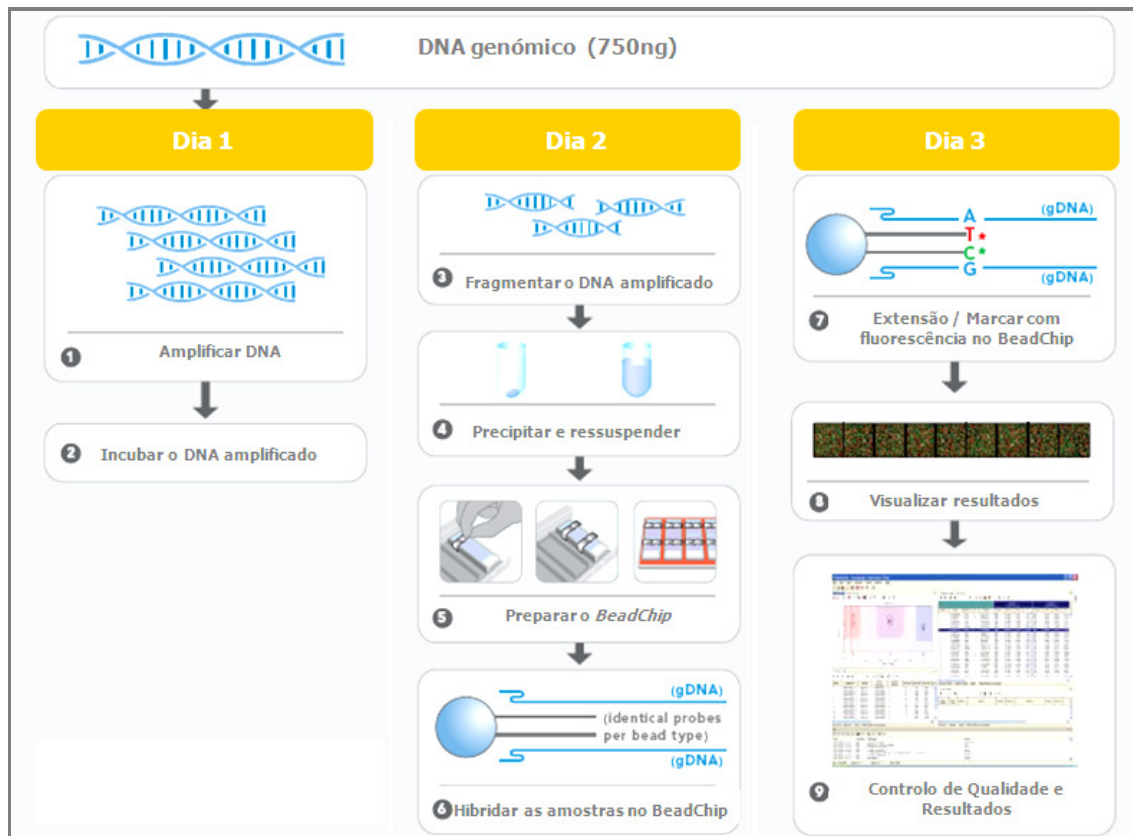


Figura 7. Etapas do procedimento experimental de *Infinium Human 1M-single Infinium Beadchip*, da plataforma Illumina (adaptado de *Infinium® DNA Analysis Beadchips*).

As amostras de DNA foram preparadas e enviadas para uma instituição, pertencente ao consórcio AGP, que centraliza todas as genotipagens. Os dados de genotipagem do AGP são utilizados em todas as análises estatísticas realizadas neste trabalho. Um total de 842 348 marcadores passaram os testes de controlo de qualidade com uma *call rate* superior a 98%, MAF superior a 0,05 e desvios ao HWE superiores a 0,05. Um total de 1369 famílias AGP compostas por pelo menos um indivíduo ao qual foi diagnosticado ASD e os respectivos pais passaram os testes de controlo de qualidade, das quais 229 famílias correspondem a famílias portuguesas (sub-população PT).

O SNP rs1187321 (A/T) está localizado na região 5' do gene *NTRK2*. Correia *et al.* (2010) encontrou associação entre o marcador e o autismo, na população portuguesa. A importância da sua análise deve-se ao facto deste SNP alterar o local de ligação de factores de transcrição com homeodomínio na presença do alelo T e por isso é uma parte de interesse neste estudo. O SNP foi anteriormente genotipado através da tecnologia de Sequenom MassARRAY (Sequenom) (Correia *et al.*, 2010).

5. ANÁLISE DE ASSOCIAÇÃO GENÉTICA

Para aferir o contributo individual de cada marcador genético no risco da doença realizaram-se testes de associação baseados em famílias utilizando os dados do GWAS do AGP, numa população de 1369 famílias (população AGP). Foram também realizados testes de associação genética na sub-população portuguesa (sub-população PT) composta por 229 famílias, pelo facto de ser distinta geneticamente das restantes populações que integraram o GWAS (ver anexo D).

O método para calcular o sinal de associação genética usado denomina-se *Transmission Disequilibrium Test* (TDT) e utiliza os genótipos não transmitidos como controlo interno. O TDT é um método de análise não paramétrico, na medida em que não requer a especificação de um modelo de transmissão para a doença. O TDT analisa a frequência com que um alelo é transmitido dos pais heterozigóticos para a descendência afectada. Se os alelos não estiverem associados com a doença, a probabilidade de transmissão mendeliana é igual, isto é, 50% para cada alelo. A hipótese nula de não existir associação entre o marcador e a doença é testada através do teste estatístico de McNemar $(b - c)^2 / (b + c)$, que segue uma distribuição de qui-quadrado, em que b e c representam o número de vezes que um alelo é transmitido ou não é transmitido, respectivamente. O teste estatístico testa a validade da hipótese nula, ou seja, existem ou não diferenças significativas entre o número de vezes que um alelo é ou não transmitido (Spielman *et al.*, 2003)

Para realizar o teste TDT foi utilizada uma ferramenta de análise denominada PLINK (versão 1.07), disponível *online*, em <http://pngu.mgh.harvard.edu/purcell/plink/> (Purcell *et al.*, 2007). Esta ferramenta de análise permite utilizar os dados genotípicos e fenotípicos resultantes de rastreio genómicos para associação e realizar sobre estes uma bateria de testes, entre os quais o TDT.

5.1. Análise de associação dos genes *homeobox* com o autismo

Correia *et al.* (2010) mostraram que o SNP rs1187321 (A/T), na região regulatória do gene *NTRK2*, está localizado num domínio de ligação de factores de transcrição com homeodomínio e encontraram evidência preditiva de alteração na ligação de factores de transcrição de duas famílias, quando o alelo T está presente (Correia *et al.*, 2010). As famílias acima referidas são: a família HOXF que inclui os factores de transcrição *HOXA1* e *PITX1* e a família HOMF que inclui os factores de transcrição *EN1*, *DLX1*, *DLX2*. A fim de aferir o contributo individual destes genes *homeobox* no autismo, realizou-se o teste de TDT para os respectivos genes na população AGP

e isoladamente na população PT. A tabela 8 indica o número de marcadores genotipados para cada um dos genes utilizados no teste.

Tabela 8. Tabela indicativa do número de SNP genotipados para os genes *HOXA1*, *PITX1*, *EN1*, *DLX1* e *DLX2*.

Gene	Número de SNP genotipados
<i>PITX1</i>	58
<i>HOXA1</i>	26
<i>EN1</i>	148
<i>DLX1</i>	8
<i>DLX2</i>	61

5.2. Análise de associação dos genes das vias biológicas activadas por BDNF/TrkB com o autismo

Os resultados encontrados por Correia *et al.* (2010) sugerem que alterações na via de sinalização BDNF/TrkB poderão estar associadas ao autismo. Estas alterações poderão ocorrer ao nível da ligação da neurotrofina ao seu receptor e, conseqüentemente, afectar as moléculas participantes nas três principais vias biológicas activadas. Pelo facto de serem vias biológicas com um elevado número de genes utilizou-se um critério de selecção para restringir o número de genes. A selecção dos genes decorreu em duas etapas diferentes: 1) identificação dos genes da via de sinalização de referência das neurotrofinas (Anexo B); 2) selecção dos genes pelo elevado nível de confiança de interacção proteica putativa. A via biológica de referência está disponível *online* na base de dados KEGG (*Kyoto Encyclopedia of Genes and Genomes*) (apresentada no anexo B). A tabela 9 é o resultado da primeira etapa, ou seja, a lista de genes seleccionados e as respectivas moléculas participantes na via biológica das neurotrofinas acima referida e descrita na secção I.4.1..

Tabela 9. Lista de 27 genes pertencentes às três vias de sinalização activadas pela ligação BDNF/TrkB e a respectiva molécula codificada por cada gene.

Gene	Molécula codificada
<i>BDNF</i>	Bdnf
<i>NTRK2</i>	TrkB
<i>SHC1</i>	Shc1
<i>SHC3</i>	Shc3
<i>GRB2</i>	Grb2
<i>SOS1</i>	Sos1
<i>SOS2</i>	Sos2
<i>FRS2</i>	Frs2
<i>GAB1</i>	Gab 1
<i>HRAS</i>	Ras
<i>RAF1 ou cRAF</i>	Raf
<i>MAP2K1 ou MEK1</i>	Mek1
<i>MAP2K2 ou MEK2</i>	Mek2
<i>MAPK3 ou ERK1</i>	Erk1
<i>PIK3R1</i>	PI3K-p85 α
<i>PIK3R2</i>	PI3K-p85 β
<i>PIK3R3</i>	PI3K-p55 γ
<i>PIK3R5</i>	PI3K-p101
<i>PIK3CA</i>	PI3K-p110 α
<i>PIK3CB</i>	PI3K-p110 β
<i>PIK3CD</i>	PI3K-p110 δ
<i>PIK3CG</i>	PI3K-p110 γ
<i>AKT1</i>	Akt1
<i>AKT2</i>	Akt2
<i>AKT3</i>	Akt3
<i>PLCG1 e PLCG2</i>	PLC γ

Na segunda etapa foram escolhidos, de entre todos os genes, os que apresentavam valores de confiança de interacção proteica superiores a 95% (ver anexo E), de acordo com o programa *online* STRING (v.8.2) (Jensen *et al.*, 2009) ^[4]. Este programa prediz, com base na literatura científica, em estudos genómicos e de expressão génica, quais as interacções proteicas e a cada associa um valor de confiança estatística. Dos vinte e seis genes introduzidos no STRING (tabela 9) vinte e cinco apresentam valores de confiança de interacção superiores a 95% (Tabela 10). Foi excluído o gene *AKT3*. O diagrama dos resultados do programa STRING para níveis de confiança superiores a 95% encontra-se disponível no anexo E.

Testou-se a associação de cada gene com o autismo através de um teste de TDT na população AGP e também na sub-população PT. A lista de genes e o número de marcadores genéticos genotipado para cada gene está representada na tabela 10.

Tabela 10. Lista dos resultados do programa de interacção proteica STRING para valores de confiança superiores a 95%. Os 26 genes escolhidos pertencem às três vias de sinalização activadas pela ligação BDNF/TrkB. A coluna da direita representa o número de SNPs genotipados para cada gene.

Gene	Nº de SNP
<i>BDNF</i>	81
<i>NTRK2</i>	219
<i>SHC1</i>	17
<i>SHC3</i>	73
<i>GRB2</i>	66
<i>SOS1</i>	51
<i>SOS2</i>	32
<i>GAB1</i>	60
<i>FRS2</i>	47
<i>HRAS</i>	4
<i>RAF1 ou cRAF</i>	40
<i>MAP2K1 ou MEK1</i>	27
<i>MAP2K2 ou MEK2</i>	18
<i>MAPK3 ou ERK1</i>	11
<i>PIK3R1</i>	343
<i>PIK3R2</i>	34
<i>PIK3R3</i>	25
<i>PIK3R5</i>	35
<i>PIK3CA</i>	54
<i>PIK3CB</i>	44
<i>PIK3CD</i>	45
<i>PIK3CG</i>	84
<i>AKT1</i>	26
<i>AKT2</i>	38
<i>PLCG1</i>	23
<i>PLCG2</i>	142

6. ANÁLISE DE INTERACÇÃO GENÉTICA

Após testar-se a associação genética dos genes *NTRK2*, *BDNF* e genes *homeobox* procurou-se evidência de interacção genética entre os seus marcadores. Pretendeu-se avaliar o efeito sobre o fenótipo de diferentes combinações dos marcadores genéticos destes genes. Para realizar a

análise estatística foi utilizado o programa PGMDR. Segue-se uma breve descrição da metodologia.

Ritchie e colegas (2001) introduziram o método MDR (*Multifactor Dimensionality Reduction*) com poder para detectar interacções genéticas entre dois ou mais *loci*, permitindo identificar combinações de polimorfismos que poderão estar associados ao risco de uma determinada doença. É um método que não assume nenhum modelo de hereditariedade e é aplicável a estudos de caso-controlo e estudos baseados em famílias, no último caso com utilização de uma extensão do algoritmo original.

Em 2007, foi introduzido um método baseado no MDR denominado por GMDR (*Generalized Multifactor Dimensionality Reduction*) (Lou *et al.*, 2007). Este método apresenta vantagens sobre MDR porque permite ajustar para covariáveis discretas e quantitativas e é aplicável a fenótipos dicotómicos e contínuos. Os métodos utilizam a mesma estratégia de redução dos dados: as possíveis células são classificadas por um conjunto de factores e agrupadas em dois grupos distintos, o que permite reduzir a dimensão (de multidimensionais para unidimensionais) e, deste modo, identificar de entre todas as potenciais combinações, as combinações específicas que mostram o sinal de associação mais forte com o fenótipo. A diferença entre os métodos baseia-se na substituição do rácio caso:controlo em cada célula, por um valor estatístico para classificar as células em elevado e baixo risco e estimar o erro de predição. O restante procedimento mantém-se igual a MDR.

Segue-se uma descrição do método. A figura 8 representa os diferentes passos envolvidos no método de GMDR. No primeiro passo os dados de genotipagem para todos os indivíduos são divididos ao acaso em dez partes iguais. Uma subdivisão é usada como o conjunto de dados de teste (*testing set*) e as restantes subdivisões (9/10) são conjuntos de dados de treino (*training set*). O modelo é desenvolvido em 9/10 dos dados e testado no 1/10 restante (passos 2-5 para o *training set* e passo 6 para o *testing set*). Para todos os indivíduos, foi calculado o valor estatístico através do método *Maximum-Likelihood Estimates* (MLE). A hipótese nula assume que não há efeitos dos putativos genótipos ou das suas interacções, isto é, os seus valores estatísticos serão os mesmos para as diferentes combinações de *locus* (genótipos). No passo 3, o valor de teste (cumulativo) é calculado para cada célula. Se o valor médio estatístico for superior a 0 (zero) a célula é classificada de "elevado risco"; se este valor não exceder 0 (zero) então a célula é classificada de "baixo risco" (passo 4). Todas as potenciais combinações são avaliadas pela sua capacidade de classificar os indivíduos como caso ou controlo no *training set*, isto é, o programa prediz se um determinado genótipo está mais presente nos casos ou controlos (passo 5). É escolhido o modelo que apresenta o menor erro de classificação. No passo seguinte, o *testing set* é usado para estimar o erro de predição (*Testing Balanced Accuracy*, TBA) do(s) melhor(es) modelo(s) seleccionado(s) anteriormente através de um

módulo de permutações denominado *Cross Validation* (CV). O valor do erro de predição do modelo é um valor médio das 10 partes. A *Cross Validation Consistency* (CVC) é a medida do número de vezes que uma determinada combinação é a melhor combinação ao longo das 10 divisões dos dados. Os melhores modelos apresentam o menor valor de significância (*p-value*) do erro de predição (TBA) e/ou o máximo de CVC.

Os modelos são verdadeiros positivos se são capazes de gerar conjuntos de dados independentes ao qual corresponderá um valor de TBA superior a 0,5 ou 50%. Se a contribuição dos factores for nula então TBA terá valores de 0,50. Como estamos perante doenças complexas e considerando que vários genes com pequeno efeito (*effect size*) podem contribuir para a doença, estes limites para os valores de predição são biologicamente aceitáveis na literatura. O limite máximo deste parâmetro, ou seja TBA de 100%, significa que há uma contribuição muito elevada dos factores genéticos e/ou fenotípicos em estudo na etiologia da doença.

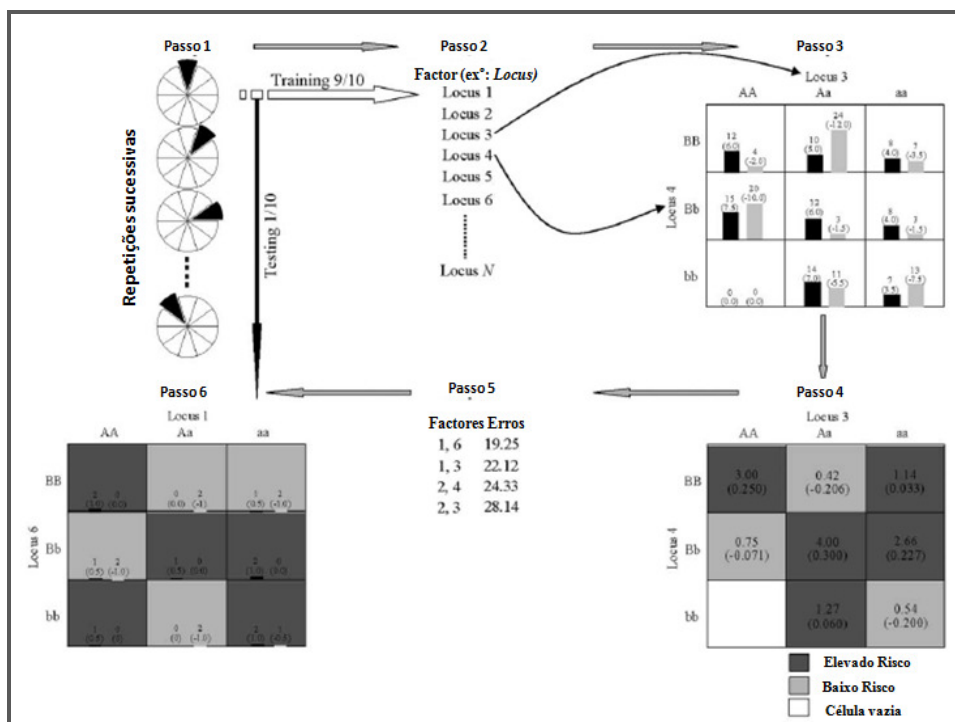


Figura 8. Os diferentes passos envolvidos no método GMDR. Os seis passos estão descritos detalhadamente no texto. No passo 3, dentro de cada célula, as barras correspondem às distribuições hipotéticas de casos (barra esquerda) e controlos (barra direita). Os números por cima das barras são o número de casos e controlos e os números entre parêntesis são o somatório dos valores estatísticos. Nos passos 4 e 6, os números sem parêntesis correspondem aos rácios caso:controlo e os números entre parêntesis são os valores estatísticos médios. As células de elevado risco (*High risk*) são marcadas com cinzento-escuro, enquanto as células de baixo risco (*Low risk*) estão assinaladas a cinzento claro (Lou *et al.*, 2007).

Em 2008, Lou *et al*, propôs o método PGMDR (*Pedigree Generalized Multifactor Dimensionality Reduction*), que permite utilizar indivíduos relacionados entre si, ou seja, dados de estudos baseados em famílias. O teste utilizado para análise de interação entre *loci* foi PGMDR (disponível em <http://www.healthsystem.virginia.edu/internet/addiction-genomics>). O programa foi utilizado em 229 famílias e o único factor fenotípico testado foi a variável *Affection* (ou seja, indivíduo é caso ou controlo). O programa realiza-se em duas etapas distintas. Na primeira etapa ocorre a transformação dos dados de famílias para dados de caso e controlo e geraram-se os ficheiros intermédios contendo os alelos não transmitidos, que serão utilizados como os genótipos dos controlos "virtuais". Os ficheiros contendo os dados genotípicos relativos aos marcadores de cada gene são obtidos através do *PLINK* (Purcell *et al*, 2007). No fim desta primeira etapa, os três ficheiros *output* criados por PedGMDR (*ped.txt*; *FTND.txt*; *pi.txt*) contêm os genótipos dos indivíduos que representam os casos (os genótipos transmitidos) e os recém obtidos genótipos dos "controlos". Com os dados genotípicos caso-controlo (segunda etapa) realizou-se a análise de interação entre as possíveis combinações de todos os marcadores, com o software GMDR. Foram seleccionadas combinações de marcadores dois a dois até três a três, para cada par de genes (*NTRK2 vs.BDNF*, *NTRK2 vs.DLX1*, *NTRK2 vs.DLX2*, *NTRK2 vs.EN1*, *NTRK2 vs.PITX1* e *NTRK2 vs.HOXA1*). A precisão de predição foi calculada através do modelo de *10fold cross-validation* (CV). O melhor modelo de interação é avaliado mediante três parâmetros: TBA superior a 0,50, valor de significância estatística ($p < 0,05$) e valor de CVC de 10. O número de marcadores genéticos utilizado está representado na tabela 11.

Tabela 11. Número de marcadores genéticos para o par de genes utilizado em cada ensaio de interação genética.

Ensaio	Nº de SNP no teste
<i>BDNFxNTRK2</i>	323
<i>PITX1xNTRK2</i>	277
<i>HOXA1 xNTRK2</i>	246
<i>EN1 xNTRK2</i>	369
<i>DLX1 xNTRK2</i>	227
<i>DLX2 xNTRK2</i>	280

7. ANÁLISE DE ENRIQUECIMENTO POR SINAL DE ASSOCIAÇÃO DE VIAS BIOLÓGICAS

A análise de vias biológicas (*Pathway analysis*) é uma nova estratégia de análise para potenciar os resultados provenientes dos GWAS e encontrar vias biológicas que poderão estar implicadas na susceptibilidade às doenças complexas, como o autismo.

A análise bioinformática foi realizada com os dados do GWAS realizado pelo AGP. Todos os indivíduos foram genotipados utilizando a plataforma *Infinium Human 1M-single Infinium Beadchip* e a metodologia de genotipagem utilizada está descrita detalhadamente na secção III.4. Os genótipos foram obtidos usando o software *Illumina BeadArray*, que utiliza o HapMap como referência. Foram excluídos os indivíduos com heterozigotia elevada e género ambíguo. Os marcadores com MAF inferior a 0,05 e desvios significativos ($p < 0,05$) ao HWE foram excluídos. Na análise foram utilizados 684731 marcadores, em 1369 famílias AGP. Foram analisadas 1655 grupos de genes/vias biológicas que incluem: as bases de dados de vias biológicas disponíveis *online* KEGG, BioCarta e Gene Ontology (GO) e três grupos de genes de interesse (indicados na tabela 9) correspondentes às três vias biológicas a jusante da ligação BDNF/trkB (denominados de "PI3K-AKTsignaling", "MAPKsignaling" e "PLCGsignaling"). Foram excluídas as vias biológicas com menos de 5 genes e mais de 200 genes para evitar a repetição de genes nos diferentes conjuntos e erros de *multiple testing*.

O método adaptado por Wang (2007) tem como base um algoritmo denominado *Gene Set Enrichment Analysis* (GSEA) desenvolvido para a análise de estudos de expressão (Subramanian *et al.*, 2005). Wang aplicou esta ideia à análise de dados resultantes de GWAS (Wang *et al.*, 2007) e desenvolveu o *software* *GenGen Package (Genetic Genomics Analysis of Complex Data)* (versão Agosto de 2009, disponível em www.openbioinformatics.org/gengen) utilizado neste estudo (Figura 9). Este método ordena todos os genes pelos seus sinais significativos de associação (TDT) e testa se um determinado grupo de genes ou via biológica está mais associado no topo da lista de genes do que seria esperado.

O *software* tem dois programas distintos: `calculate_association.pl` e `calculate_gsea.pl`. O primeiro programa executa um teste de TDT para todos os SNPs em análise e simultaneamente calcula os *p-values* de permutação dos dados de TDT, permutando os alelos transmitidos e não transmitidos. Para cada gene é seleccionado o SNP com o sinal de associação mais forte tornando-se no sinal representativo do gene. Os genes são ordenados por ordem decrescente do valor de significância. Para cada grupo de genes ou vias biológicas é calculado o valor de enriquecimento (*enrichment score*, ES) através do teste estatístico Kolmogorov-Smirnov, para os valores reais e para os valores de permutação. O resultado reflecte a sobre-representação dos genes dentro de uma via no topo da lista prioritizada de genes. Contudo, se o cálculo de ES se baseia no maior sinal estatístico dentro de cada gene, então os genes com um maior número de SNPs terão mais probabilidade de terem um valor estatístico superior aos genes com poucos SNPs. O método faz o ajustamento para o tamanho do gene através de dois procedimentos descritos em Wang *et al.* (2007): permutações e cálculo de NES (*Normalized Enrichment Score*). O NES é calculado para que vias biológicas/grupos de genes com diferentes tamanhos, possam ser comparadas entre si. O algoritmo GSEA adaptado (programa `calculate_gsea.pl`) no

final dá um *p-value* de enriquecimento de associação para cada uma das vias biológicas em estudo (Wang *et al.*, 2007).

O programa corrige os resultados para *multiple testing* através de *False Discovery Rate* (FDR) e FWER (*Familywise Error Rate*) que permitem identificar falsos positivos (erros do tipo I) assumindo que os *p-values* correspondentes a hipóteses nulas verdadeiras são independentes e distribuídos uniformemente. O valor de FDR é calculado baseado nos valores de NES permutados (Wang *et al.*, 2007).

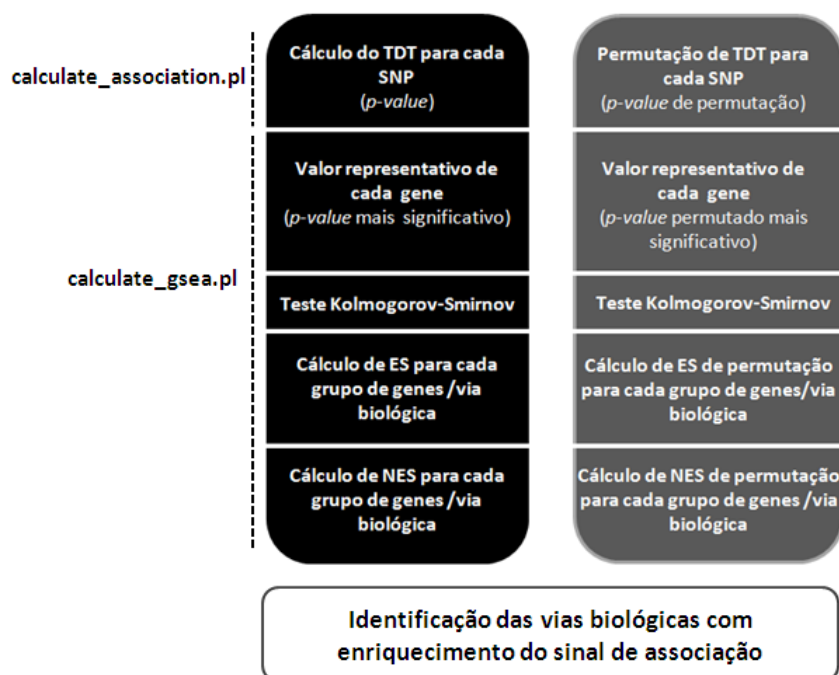


Figura 9. Esquema representativo do *software* GenGen. O *software* realiza em simultâneo os processos de cálculo dos valores reais (rectângulo preto) e o cálculo dos valores de permutação (rectângulo a cinzento). A comparação dos valores permite dar o sinal de significância das vias biológicas. A descrição detalhada encontra-se descrita no texto.

IV. RESULTADOS

1. RASTREIO DE VARIAÇÕES GENÉTICAS NO GENE *NTRK2*

O rastreio de variações presentes nos exões 6/7, 8, 22 e 23 (segundo a descrição feita por Stoilov *et al.*, 2002) do gene *NTRK2* foi realizado em 100 pacientes portugueses com autismo (dos quais 86 são do sexo masculino e 14 do sexo feminino). Foram encontradas duas alterações em duas regiões intrónicas distintas: uma nova variante no intrão 6-7 e um SNP (rs3739807) no intrão 21-22, já descrito na base de dados dbSNP (NCBI) (tabela 12). Em nenhum dos 100 indivíduos sequenciados foram encontrados as duas variantes em simultâneo. A confirmação das variações encontradas fez-se através da repetição do processo de sequenciação.

Tabela 12. Alterações de DNA encontradas nas regiões sequenciadas no gene *NTRK2*.

Alteração do nucleótido	Domínio	Tipo de alteração da sequência	Número de indivíduos afectados	Mutação / Polimorfismo	Referências
c.288-15T>C ^(a)	Intrónico	Varição intrónica (Intrão 6-7)	3	Alteração nucleotídica	Identificado neste estudo
c.1938-55C>T ^(b) rs3739807	Intrónico	Varição intrónica (Intrão 21-22)	4	<i>Single Nucleotide Polymorphism (SNP)</i>	Descrito na base de dados dbSNP (NCBI)

^(a) Nomenclatura proposta para identificar a alteração nucleotídica encontrado neste estudo.

^(b) NM_006180.3

1.1. Variação no intrão 6-7

A região sequenciada compreende os exões 6 e 7 e as respectivas fronteiras intrónicas (figura 10). Foi encontrada uma alteração na base 15 do intrão 6-7, antes do início do exão 7, em três pacientes com autismo. A nomenclatura proposta para a descrição deste polimorfismo foi criada de acordo com as regras descritas por den Dunnen and Antonarakis (2001). Na nomenclatura c.288-15T>C \underline{c} refere-se à sequência de cDNA, 288 é a posição do primeiro nucleótido do exão seguinte (ou seja, do exão 7) e 15 é a posição do nucleótido alterado no intrão, ou seja, o nucleótido original T altera para um C.

```

AAAAATGGCAAGGCTCAGAGTGGTGTGGAGGGAGCACCTTGGACACCTGGGTGAGGTGGATGGG
AGTTCTTCATGGTGTCTGCCTACATTCTGAGTCACTGCGATTCACTCTCTGCTTTGTACAG
TTTCATCGCAAACCAGAAAAGGTTAGAAATCATCAACGAAGATGATGTTGAAGCTTATG
TGGGACTGAGAAATCTGTGAGTACTCAGGACCAGGGCACATTATCTCAGAGAATTTTCCTGT
TGTCTGCTCTGGTCAGGCAGGCATTCACTGGTTCGTTCTAATGTGCATGAAATTATGTGTTTTC
ACAGGACAATTGTGGATTCTGGATTAAAATTTGTGGCTCATAAAGCATTTCTGAAAAAC
AGCAACCTGCAGCACATGTAAGTAGAGATTGATTTCTTTTGCTTCCCAGGACCCATTTTATTC
AATATTTCCCCCCTCTGTTTATTTCTTTCTTTTCCAACCTGAAACCTCCCTTTTTAAAGTTAG
TATAGCTGTGATAATAGCTTAGAAACATTAGCTCTGATTTG

```

Figura 10. Sequência nucleotídica dos exões 6 e 7 (a negrito) e respectivas regiões intrônicas do gene *NTRK2*. O nucleótido assinalado pelo círculo verde corresponde à nova alteração nucleotídica c.288-15T>C. As sequências dos *primers* utilizados estão sublinhadas a cinzento.

A alteração foi identificada através do programa de análises de sequências *STADEN* (figura 11), quando se emparelham as cadeias *forward* e *reverse* com a sequência original, isto é, a sequência descrita na base de dados genómica *ENSEMBL*. A figura mostra a alteração encontrada, sendo a primeira sequência correspondente à sequência original e as seguintes linhas correspondem à cadeia *forward* (F) e *reverse* (R), dos três indivíduos sequenciados. A letra Y encontrada na posição onde originalmente deveria estar um T é interpretada, à luz do código de Ambiguidade, como dois possíveis nucleótidos: T ou C (figura 11). A confirmação surge com a identificação de dois picos, em ambas as cadeias, um azul indicativo da presença de um C e o pico vermelho indicativo da presença de um T (figura 12), o que sugere que os três indivíduos são heterozigóticos para esta alteração (genótipo CT).

Sequencia original	CTAATGTGCATGAAATTATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA
313R	CTAATGTGCATGAAAT Y ATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA
85F	CTAATGTGCATGAAAT Y ATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA
237R	CTAATGTGCATGAAAT Y ATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA
85R	CTAATGTGCATGAAAT Y ATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA
313F	CTAATGTGCATGAAAT Y ATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA
237F	CTAATGTGCATGAAAT Y ATGTGTTTTTCACAGGACAATTGTGGATTCTGGATTA

Figura 11. Resultado da análise de sequenciação dos exões 6/7, no programa *STADEN*. A primeira sequência representa a sequência original, sendo as restantes sequências (F e R) dos pacientes com autismo onde se encontra a alteração c.288-15T>C, representada pela letra Y (pertencente ao código de ambiguidade). F representa a sequência *forward* e R representa a sequência *reverse*.

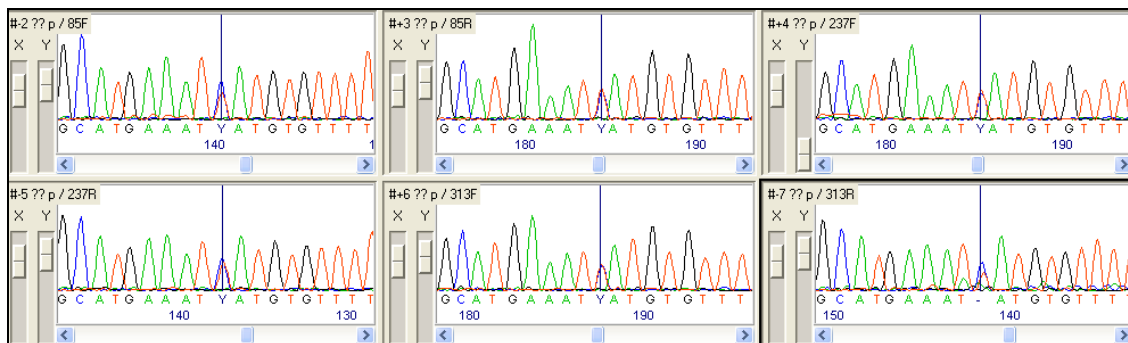


Figura 12. Cromatograma do resultado de sequenciação dos exões 6/7 que mostra em detalhe a região onde ocorre alteração do nucleótido representado por Y (adaptado de *STADEN Package*), nos três pacientes com autismo.

1.1.1. Caracterização clínica

Sendo esta uma variante rara analisaram-se os dados clínicos dos pacientes com o objectivo de identificar características clínicas comuns e relevantes entre os indivíduos afectados. A caracterização clínica dos três indivíduos onde foi encontrado a alteração c.288-15T>C é mostrada na tabela 13.

Tabela 13. Características clínicas referentes aos três indivíduos com a alteração c.288-15T>C.

Critério	Dados clínicos		
Sexo	Masculino; Caucasiano	Masculino; Caucasiano	Masculino; Caucasiano
Idade	5 anos	5 anos	14 anos
QD Global	Normal	Deficiência mental ligeira	Deficiência mental ligeira a moderada
Epilepsia/EEG anormal	Não	Sim	Não
Dismorfologias	Não	Não	Não
Problemas no parto/neonatais	Não	Icterícia Neonatal	Não
Atraso na linguagem	Sim (primeiras palavras aos 40meses)	Sim (primeiras frases aos 48meses)	Sim (primeiras frases aos 48 meses)
História familiar	Pais e irmã saudáveis	Pai tem Doença de Parkinson	Mãe com dificuldades de aprendizagem
Diagnóstico (ADOS)	Autismo	Autismo	Autismo
Alterações neurológicas	Não	Atraso desenvolvimento psico-motor a partir dos 2 anos	Atraso desenvolvimento psicomotor a partir dos 2 anos
Outros problemas	Problemas de sono; problemas alimentares	Não	Não
Etiologia	Idiopático	Idiopático	Idiopático

Os três indivíduos onde foi encontrada a alteração são do sexo masculino, caucasianos e com idades compreendidas entre os 5 e os 14 anos. A idade considerada na tabela é a idade à data da primeira colheita. Apresentam em comum autismo de etiologia idiopática (segundo os critérios de diagnóstico ADOS) e atrasos na linguagem (a formação das primeiras frases a partir dos 35 meses é considerado atraso na linguagem). Os restantes critérios de avaliação e diagnóstico são díspares. Um dos pacientes não tem deficiência mental (QD global de 79), alterações neurológicas, não apresenta sinais de epilepsia, não teve quaisquer problemas de índole neonatal, os seus familiares mais próximos são saudáveis. Apenas apresenta problemas de sono e alimentares. Os outros dois indivíduos têm deficiência mental ligeira a moderada e atrasos no desenvolvimento psicomotor a partir dos 2 anos. Não existem características clínicas comuns que sejam explicadas pela presença desta variante genética, excepto o atraso na linguagem que pode ser relevante.

1.1.2. Análise preditiva

Como a variante tem localização intrónica pode influenciar o processo de *splicing*. Para a análise bioinformática de c.288-15T>C foram utilizados três ferramentas de análise preditiva (<http://www.ebi.ac.uk/asd-srv/wb.cgi>): o programa *Intron Analysis* que identifica locais *Donor e*

Acceptor de *splicing*, *Branch Point* (BP) e locais de *Polypyrimidine Tract* (PPT) numa sequência intrónica, o programa *The Splicing Rainbow* que identifica os locais de ligação de proteínas necessárias no processo de *splicing* e o programa *Regulatory Sequence Search* que permite identificar sequências regulatórias de *splicing* (tais como *Intronic Splicing Enhancers*, ISE e *Silencers*, ISS, *Exonic Splicing Silencers*, ESS e *enhancers*, ESE) através da comparação com informações obtidas nas bases de dados.

O programa *Intron Analysis* identificou três regiões PPT (figura 13), uma das quais inclui o nucleótido onde ocorre a alteração (T>C), no entanto, não existem alterações nos locais de ligação PPT na presença do polimorfismo (figura 13). A PPT é uma região de ligação de proteínas no mRNA, rica em nucleótidos de pirimidina, com 15 a 20 bp de comprimento e localizada a aproximadamente 5-40 bp do final do intrão. Esta região permite a ligação do spliceossoma e de outras proteínas envolvidas no processo de *splicing* alternativo como PTB (*polypyrimidine tract-binding protein*). O programa *The Splicing Rainbow* identificou locais de ligação de proteínas HnRNP (*Heterogeneous nuclear ribonucleoproteins*), HuR (*ELAV-like protein 1*) e SRp20 (*serine/arginine-rich protein*). Na presença do alelo C do polimorfismo ocorre a alteração de um dos locais de ligação HnRNP, o desaparecimento de um local de ligação HuR e o aparecimento de um novo local de ligação SRp20 (figura 13). As proteínas ribonucleares heterogéneas (HnRNPs) participam na síntese do pré-mRNA e são responsáveis pela supressão do mecanismo de *splicing*, através do bloqueio da ligação do spliceossoma nas PPT. As proteínas SR estão envolvidas na regulação e selecção dos locais de *splicing* no mRNA. No mecanismo de *splicing* alternativo as proteínas HnRNP competem com as SR pelos sítios de ligação no pré-mRNA impedindo o *splicing* nestes locais. As proteínas HuR sintetizadas pelo gene *ELAV1* têm um papel importante na estabilização do mRNA. Os resultados do programa *Regulatory Sequence Search* (não apresentados) indicam que a região onde ocorre a alteração não é uma possível região regulatória de *splicing*.

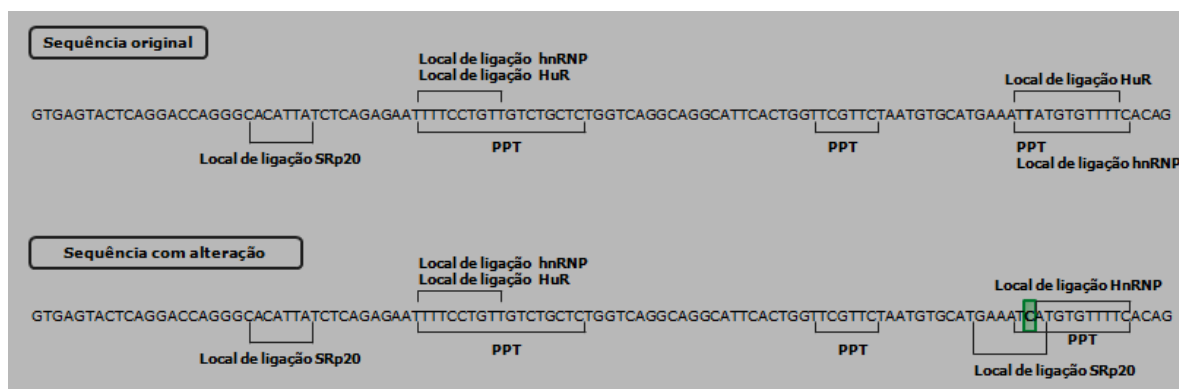


Figura 13. Esquema representativo dos resultados dos programas de análise preditiva *Intron Analysis* e *The Splicing rainbow* no intrão 6/7 do gene *NTRK2*, nas sequências original e com a nova alteração nucleotídica c.288-15T>C (nucleótido C assinalado a verde).

Estes resultados preditivos indicam que na presença da alteração c.288-15T>C não há alterações dos sítios de ligação PPT, onde se liga o spliceossoma, nem nas sequências regulatórias (ISS, ISE, ESS, ESE) de *splicing*, no entanto, há alteração dos locais de ligação das proteínas HnRNP, HuR e SRp20, o que poderá influenciar o mecanismo de *splicing* do gene *NTRK2*.

1.1.3. Frequências alélicas e genotípicas

Por último, calcularam-se as frequências genotípica e alélica na amostra portuguesa controlo. Foram sequenciados 100 indivíduos adultos saudáveis (população controlo) sem nenhuma patologia associada, provenientes de Portugal Continental, 60% pertencentes ao sexo masculino e 40% ao sexo feminino, com idade média de 44 anos. A extracção de DNA, as condições de reacção de amplificação e sequenciação foram idênticas às utilizadas para os indivíduos com autismo (ver secções III.2 e III.3). Apenas foi encontrado a alteração nucleotídica c.288-15T>C num indivíduo, sendo a prevalência na população controlo de 1%. Para verificar se existem diferenças significativas entre as frequências alélicas e genotípicas das populações caso e controlo realizou-se um teste de χ^2 . Os resultados são apresentados na seguinte tabela.

Tabela 14. Frequências genotípicas absolutas e frequências alélicas para a nova alteração c.288-15T>C. Resultados do teste de χ^2 em 100 indivíduos das populações de casos e controlo.

	Genótipos (%)			Alelos (%)	
	TT	CT	CC	T	C
Casos	97 (97%)	3 (3%)	0	197 (98,5%)	3 (1,5%)
Controlos	99 (99%)	1 (1%)	0	199 (99,5%)	1 (0,5%)
χ^2	0,255			0,253	
<i>P-value</i>	0,614			0,615	

Não existem diferenças significativas ($0,614 < P < 0,615$) entre as frequências alélicas e genotípicas das populações de casos e controlos. Provavelmente, não será esta a variante responsável pelo sinal de associação anteriormente encontrado por Correia *et al.*, (2010) na região dos exões 6/7, mas pode estar envolvido no mecanismo de *splicing* do gene, independentemente de os indivíduos serem ou não afectados pela doença.

A nova alteração nucleotídica c.288-15T>C foi submetida à base de dados genómica dbSNP (NCBI) com as respectivas frequências genotípicas absolutas, na amostra populacional portuguesa de pacientes com autismo. Os dados estarão disponíveis para consulta no dbSNP Build B132.

1.2. Variação no intrão 21-22

Foi identificada uma variação no intrão 21-22, a cinquenta e cinco pares de bases do início do exão 22, em quatro indivíduos com autismo (figura 14). A variação é o SNP rs3739807 (C>T) (Chr9:87,570,143) descrito na base de dados dbSNP (NCBI) (NM_006180.3) e no ENSEMBL (287278:C/T). O alelo mais comum, C, ocorre com uma frequência de 0,861 e a frequência do alelo T é de 0,139, no entanto esta informação é referente à população japonesa (NCBI).

```
TCCAGACTCTTTGGGAAACAGAGCAGCCTGGGTGCTCTCAGGACTGCAGAAGTACAACACTGTGTG
AGCACCTTTGTCTCCTCTTCATGCTAAGTCAGGCAGCATCTTTTAGCACCAGCAGCTACAGGGT
GGGGGTGAGGAGCTTAGCAAGAGGGACGGGGAGGGGCAGGGGCAAAGGGCCCTGGAGTGAAAA
TGCTGAGGCCCCAGCTTCATTCTCCATGTCCTTCCCCAGGGCACACGGCCCTGATGCCGTGCT
GATGGCTGAGGGCAACCCGCCACGGAAC TGACGCAGTCCGAGATGCTGCATATAGCCAGCAG
ATCGCCGGGGCATGGTCTACCTGGCGTCCCAGCACTTCGTGCACC GCGATTGGCCACCAGGA
ACTGCCTGGTCGGGGAGAACTTGTGGTGAAAATCGGGGACTTTGGGATGTCCCGGGACGTGTA
CAGCACTGACTACTACAGGTGAGTAGCTGTGCAGATCAGAGACCCAGGGACCTCTTCCCTG
CGGGACCCCTGCTGTATTGTTGCTGAGGCTGTCATAACAAAATATCATGACTGGTGGCTTA
AACGACAGACATGTGTTTCTCATGGCTCTGGAGGCTGGAAGTCCAAGCTCAAGGTGTGAGTGGG
ATTGGTTCCTCCTGAGGCCT
```

Figura 14. Sequência nucleotídica do exão 22 (a negrito) e respectivas regiões intrônicas do gene *NTRK2*. As sequências dos *primers* utilizados estão sublinhadas a cinzento. O SNP rs3739807 está assinalado pelo círculo verde.

As sequências obtidas para os 100 indivíduos com autismo foram analisadas pelo programa *STADEN*. As figuras mostram os resultados da sequenciação para os quatros indivíduos onde foi encontrado o SNP rs3739807. A figura 15 é o resultado do emparelhamento das cadeias *forward* e *reverse* com a sequência de DNA original. A letra Y identifica a alteração no nucleótido C para T como é mostrado em detalhe na figura 16.

Sequencia original	CAGGGGCAAAGGGCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
22R	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
183R	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
387R	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
194R	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
194F	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
387F	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
22F	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI
183F	CAGGGGCAAAGGGYCCCTGGAGTGAAAAATGCTGAGGCCCCAGCTI

Figura 15. Resultado da sequenciação do exão 22 que mostra a alteração de um nucleótido, na região intrônica, descrito como rs3739807 (adaptado de STADEN). A primeira sequência representa a sequência original, sendo as restantes sequências (F e R) dos pacientes com autismo onde se encontra o SNP. Alteração do nucleótido é representada pela letra Y (pertencente ao código de ambiguidade).

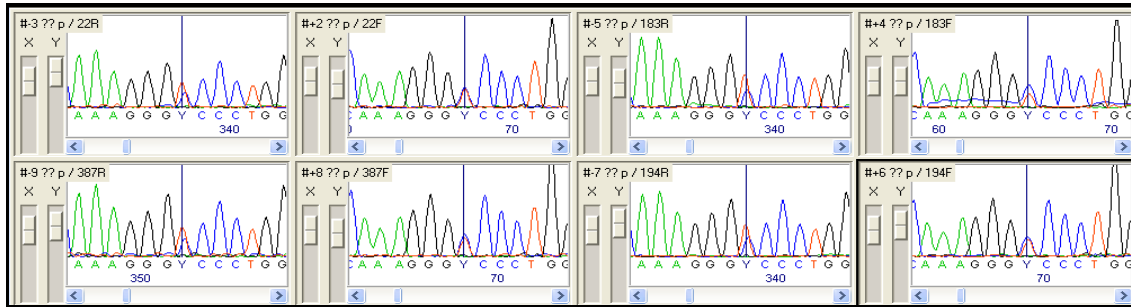


Figura 16. Resultado do cromatograma de sequenciação que mostra em detalhe a região onde ocorre a alteração de um nucleótido, rs3739807, nos quatro indivíduos com autismo (F e R correspondem, respectivamente à cadeia *forward* e *reverse*) (adaptado de *Staden Package Gap4*).

Foram analisadas as características clínicas dos quatro indivíduos com o objectivo de identificar fenótipos comuns. A caracterização clínica dos é mostrada na tabela 15.

Tabela 15. Caracterização clínica dos quatro indivíduos com o polimorfismo no intrão 21-21, rs3739807.

Critério	Dados clínicos			
Sexo	Masculino; Caucasiano	Masculino; Caucasiano	Masculino; Caucasiano	Masculino; Caucasiano
Idade	9 anos	7 anos	4 anos	7 anos
QD Global	Deficiência mental moderada)	Deficiência mental ligeira)	Deficiência mental ligeira)	Deficiência mental moderado)
Epilepsia/EEG anormal	Sim	Não	Não	Sim / EEC alterado no foco central direito
Dismorfologias	Hipotonia	Não	Não	Não
Problemas no parto/neonatais	Não	Não	Não	Não
Atraso na linguagem	Sim (primeiras frases aos 48meses)	Sim (primeiras frases aos 60meses)	Sim	Sim (primeiras frases aos 50 meses)
História familiar	Quatro irmãos, dois dos quais têm dificuldades de aprendizagem	Pais e irmãos normais	Pais e irmã normais	Pais normais
Diagnóstico (ADOS)	Autismo	Autismo	Autismo	ASD (Distúrbio do espectro do Autismo)
Alterações neurológicas	Atrofia Sub-cortical (hemisfério direito); Atraso desenvolvimento psico-motor desde a nascença	Não tem atraso desenvolvimento motor	Atraso desenvolvimento psico-motor desde a nascença	Atraso desenvolvimento psico-motor desde a nascença Ressonância magnética alterada; discreto alargamento frontal; assimetria do hipocampo, atrofia cortical
Outros problemas	Não	Não	Capacidade especial de memória visual	Não
Etiologia	Idiopático	Idiopático	Idiopático	Idiopático

Os quatro indivíduos são do sexo masculino e caucasianos e com idades, à data da primeira colheita, compreendidas entre 4 e 9 anos. A três destes indivíduos foi-lhes diagnosticado autismo (típico) tendo o quarto paciente um dos distúrbios abrangidos pelo espectro do Autismo (segundo, os critérios de Diagnóstico ADOS), mas todos com etiologia idiopática. Todos têm deficiências mentais ligeiras a moderadas (QD Global entre 35 e 70) e atrasos na linguagem, com o desenvolvimento das primeiras frases após os 48 meses. Não houve qualquer tipo de complicação neonatal a registar e, à excepção de um dos pacientes que tem dois irmãos com dificuldades de aprendizagem, os restantes não têm na família parentes afectados ou com outro tipo de patologias. Apenas um paciente apresenta hipotonia, assim como atrasos no

desenvolvimento psico-motor à semelhança de mais dois pacientes. Os dois indivíduos que têm epilepsia têm algumas alterações neurológicas, indicadas na tabela. No entanto, não é detectado nenhum padrão fenotípico comum nos indivíduos com este SNP.

As frequências genotípicas absolutas de rs3739807 na amostra populacional portuguesa com autismo foram submetidas à base de dados genómica dbSNP (NCBI) (tabela 16). Os dados estarão disponíveis para consulta com a dbSNP Build B132.

Tabela 16. Frequências genotípicas absolutas de rs3739807 em 100 indivíduos com autismo, na população portuguesa.

Genótipo	Frequência genotípica (%)
TT	0
CT	4 (4%)
CC	96 (96%)

O SNP rs3739807 tem localização intrónica, não tem influência no processo de *splicing* e não é uma variante funcional. Provavelmente esta não será a variante responsável pelo sinal de associação anteriormente encontrado por Correia *et al.*, 2010 na região do exão 22, na população portuguesa.

2. TESTES DE ASSOCIAÇÃO GENÉTICA

2.1. Genes *homeobox DLX1, DLX2, HOXA1, PITX1 e EN1*

Os resultados significativos para o teste de TDT em 520 marcadores dos genes *EN1, PITX1, HOXA1, DLX1* e *DLX2* em 1369 famílias da população AGP encontram-se na tabela 17.

Foi encontrada evidência nominal para associação do gene *EN1* com o autismo na população AGP, em oito dos 148 marcadores genotipados ($0,009 < P < 0,036$). Todos os SNP estão localizados a jusante (*downstream*) da região 3'UTR, excepto rs893572 localizado na região 5'UTR. Não foi encontrada associação genética dos genes *HOXA1, PITX1, DLX1* e *DLX2* com o autismo, na população AGP.

Tabela 17. Resultados significativos do teste de associação genética TDT, na população AGP, para os genes *ENI*, *PITX1*, *HOXA1*, *DLX1* e *DLX2*.

Gene	Marcador	Coordenadas	Alelo Menor/Maior	T	NT	χ^2	P-value	OR	MAF	Localização
ENI	rs17838804	2:119235336	G/A	220	278	6,755	0,009	0,791	0,104	Downstream 3'UTR
	rs4849706	2:119292114	T/C	528	454	5,576	0,018	1,163	0,228	Downstream 3'UTR
	rs572585	2:119354478	T/C	536	469	4,467	0,035	1,143	0,238	Downstream 3'UTR
	rs162679	2:119407150	T/C	539	470	4,719	0,030	1,147	0,239	Downstream 3'UTR
	rs332068	2:119417538	T/C	542	469	5,271	0,022	1,156	0,239	Downstream 3'UTR
	rs332111	2:119435703	G/T	591	521	4,406	0,036	1,134	0,272	Downstream 3'UTR
	rs332103	2:119449172	A/G	563	491	4,918	0,027	1,147	0,255	Downstream 3'UTR
	rs893572	2:119607470	T/G	285	339	4,673	0,031	0,841	0,133	5'UTR

T – genótipo transmitido
 NT – genótipo não transmitido
 χ^2 – resultado do teste de chi-quadrado
 HWE – teste de equilíbrio de Hardy-Weinberg
 MAF – *Minor allele frequency*
 OR – *Odds ratio*

Na sub-população PT (tabela 18) observou-se associação no marcador rs893572 do gene *ENI* com o autismo, também associado na população AGP. O marcador rs2456162 (genotipado para *PITX1*) está associado na população portuguesa, no entanto, a sua localização física actual (segundo a base de dados ENSEMBL) situa-se no intrão 2-3 do novo gene AC026691.1. O gene AC026691.1 tem 2474bp e seis exões que codificam uma proteína de 145 aminoácidos sobre a qual ainda se desconhece o tipo e função (UniProt:Q9H5L9).

Tabela 18. Resultados significativos ($p \leq 0,05$) do teste de associação genética TDT na sub-população PT para os genes *ENI*, *PITX1*, *HOXA1*, *DLX1* e *DLX2*. Assinalado a cinzento o SNP associado na população AGP e sub-população PT.

Gene	Marcador	Coordenadas	Alelo Menor/Maior	T	NT	χ^2	P-value	OR	MAF	Localização
ENI	rs893572	2:119607470	T/G	46	69	4,600	0,032	0,667	0,160	5'UTR
PITX1	rs2456162	5:134414122	T/C	103	72	5,491	0,019	1,431	0,258	Intrónico (2-3) gene AC026691.1

T – nº de indivíduos em que o alelo menor é transmitido
 NT – nº de indivíduos em que o alelo menor não é transmitido
 χ^2 – resultado do teste de chi-quadrado
 HWE – teste de equilíbrio de Hardy-Weinberg
 MAF – *Minor allele frequency*
 OR – *Odds Ratio*

2.2. Genes envolvidos nas vias biológicas activadas por BDNF/TrkB

2.2.1. População AGP

Os resultados significativos do teste de TDT de 1592 marcadores em 1369 famílias da população AGP encontram-se na tabela 19.

Foi encontrada associação dos genes *NTRK2*, *SOS2*, *SHC3*, *PLCG2*, *PIK3R1*, *PIK3R5*, *PIK3CG*, *PIK3CA*, *MAP2K2* e *FRS2* com o autismo. Seis marcadores do gene *NTRK2* encontram-se associados ($0,008 < P < 0,048$) ao autismo: três dos quais têm localização a montante da região

5'UTR e a jusante da região 3'UTR e os restantes estão situados nos intrões 14-15 e 19-20. Nos genes *SOS2* e *SHC3* foi encontrada associação nominal ($P=0,022$ e $0,015 < P < 0,045$) em três dos seus marcadores localizados nas regiões intrónicas e a jusante da região 3'UTR. O gene *PLCG2* apresenta evidência de associação nominal com o autismo nos marcadores localizados nos intrões 2-3 e 8-9.

Os sinais de associação nominal dos genes *PIK3R5* e *PIK3CA* estão localizados nas regiões intrónicas 1-2 e 16-17, respectivamente. Foi identificada associação ($0,005 < P < 0,049$) com o autismo no gene *PIK3R1* nas regiões a montante e a jusante de 5'UTR e 3'UTR e dois destes marcadores comuns têm valores de associação muito significativos, no entanto, um baixo valor de OR. O marcador intrónico rs12536419 genotipado para o gene *PIK3CG* está associado ao autismo, contudo, a sua localização física actual (segundo a base de dados ENSEMBL) corresponde ao novo gene RP5-884M6.1. Este gene de 2981bp e dois exões codifica uma proteína de 116 aminoácidos sobre a qual se desconhece o tipo e função (UniProt: *Putative uncharacterized protein* LOC340340). Os restantes marcadores associados ($0,004 < P < 0,044$) estão localizados na região a montante de 5'UTR, nos intrões 5-6 e 10-11, na 3'UTR e na região a jusante desta. Os valores de associação são bastante significativos em dois marcadores comuns, contudo o valor de OR é de 1,2 para ambos. No gene *PIK3R3* foi identificada associação nominal em sete dos seus marcadores ($0,027 < P < 0,045$) localizados nas regiões intrónicas 1-2 e 6-7, exão7, 5'UTR e 3'UTR.

O gene *MAP2K2*, pertencente à via de sinalização MEK/ERK, apresenta uma região de susceptibilidade no intrão 2-3, associada ao autismo ($P=0,015$). Doze marcadores do gene *FRS2*, participante nas vias PI3K-AKT e MEK/ERK encontram-se associados ao autismo ($0,001 < P < 0,049$) e os marcadores estão localizados a montante de 5'UTR, na região 5'UTR e nas regiões intrónicas 1-2, 4-5, 5-6.

Tabela 19. Resultados significativos do teste de associação genética TDT, na população AGP, para os genes envolvidos nas vias biológicas activadas por BDNF/TrkB.

Gene	Marcador	Coordenadas	Minor/ Major	T	NT	OR	IC 95%	P-value	MAF	Localização
NTRK2	rs17087472	9:87246051	G/A	200	253	0,790	(0,65, 0,95)	0,013	0,188	Upstream 5'UTR
	rs1659412	9:87278152	C/T	197	253	0,779	(0,65, 0,94)	0,008	0,090	Upstream 5'UTR
	rs716893	9:87419117	T/C	661	585	1,130	(1,01, 1,26)	0,031	0,345	Intrónico (14-15)
	rs1036915	9:87437849	C/T	659	578	1,140	(1,01, 1,27)	0,021	0,326	Intrónico (14-15)
	rs3860945	9:87585624	G/A	387	334	1,159	(1,01, 1,34)	0,048	0,152	Intrónico (19-20)
	rs1931109	9:87734520	G/C	147	114	1,289	(1,00, 1,65)	0,041	0,049	Downstream 3'UTR
SOS2	rs10149742	14:50624784	G/A	221	272	0,813	(0,68, 0,97)	0,022	0,110	Intrónico (11-12)
SHC3	rs7864748	9:91713387	C/T	391	449	0,871	(0,76, 1,00)	0,045	0,198	Intrónico (2-3)
	rs2316182	9:91813985	C/T	335	275	1,218	(1,04, 1,93)	0,015	0,122	Downstream 3'UTR
PLCG2	rs4456499	16:81850916	A/G	388	331	1,172	(1,01, 1,36)	0,034	0,153	Intrónico (2-3)
	rs6420427	16:81852307	T/C	707	629	1,124	(1,01, 1,25)	0,033	0,439	Intrónico (2-3)

Gene	Marcador	Coordenadas	Minor/ Major	T	NT	OR	IC 95%	P-value	MAF	Localização
PLCG2	rs4325546	16:81873821	T/C	558	631	0,884	(0,79, 0,99)	0,034	0,306	Intrónico (2-3)
	rs11865395	16:81914583	T/C	509	592	0,859	(0,76, 0,97)	0,012	0,316	Intrónico (8-9)
PIK3R5	rs2106342	17:8840933	A/C	659	582	1,132	(1,01, 1,27)	0,029	0,3391	Intrónico (1-2)
PIK3R1	rs13182765	5:67151065	A/G	322	274	1,175	(1,1, 1,38)	0,049	0,126	Upstream 5'UTR
	rs1490801	5:67166008	C/T	684	611	1,119	(1,00, 1,25)	0,043	0,382	Upstream 5'UTR
	rs13358352	5:67306754	G/A	667	581	1,148	(1,03, 1,28)	0,015	0,366	Upstream 5'UTR
	rs13158718	5:67342328	C/T	322	274	1,175	(1,00, 1,38)	0,049	0,122	Upstream 5'UTR
	rs11953134	5:67348542	C/T	313	365	0,858	(1,06, 1,35)	0,046	0,145	Upstream 5'UTR
	rs176003	5:67386576	G/A	562	471	1,193	(0,78, 0,96)	0,005	0,255	Upstream 5'UTR
	rs282280	5:67404615	G/A	628	728	0,863	(0,78, 0,96)	0,007	0,448	Upstream 5'UTR
	rs853811	5:67742704	T/G	511	581	0,879	(0,78, 0,99)	0,034	0,362	Downstream 3'UTR
	rs706729	5:67857888	G/A	191	233	0,819	(0,68, 0,99)	0,041	0,082	Downstream 3'UTR
	rs706733	5:67883394	C/T	293	348	0,842	(0,72, 0,98)	0,029	0,133	Downstream 3'UTR
PIK3CG	rs12536419	7:106419296	C/A	318	373	0,853	(0,73, 0,99)	0,036	0,149	Intrónico (1-2) ENSG00000228742
	rs4727663	7:106482966	G/A	653	579	1,128	(1,01, 1,26)	0,035	0,331	Upstream 5'UTR
	rs757902	7:106499858	G/A	545	617	0,883	(0,79, 0,99)	0,034	0,313	Upstream 5'UTR
	rs849367	7:106515654	C/A	216	262	0,824	(0,69, 0,99)	0,035	0,099	Intrónico (5-6)
	rs4460309	7:106518289	T/C	551	482	1,143	(1,01, 1,29)	0,032	0,239	Intrónico (5-6)
	rs4730205	7:106529749	T/C	591	517	1,143	(1,02, 1,29)	0,026	0,279	Intrónico (10-11)
	rs11766675	7:106547853	G/A	480	397	1,209	(1,06, 1,38)	0,005	0,189	3'UTR
	rs13247349	7:106569720	G/T	534	445	1,200	(1,06, 1,36)	0,004	0,227	Downstream 3'UTR
	rs17153654	7:106586436	T/C	238	284	0,838	(0,71, 1,00)	0,044	0,105	Downstream 3'UTR
	rs711440	7:106587464	T/C	297	360	0,825	(0,71, 0,96)	0,014	0,148	Downstream 3'UTR
	rs1724278	7:106588730	T/C	322	375	0,859	(0,74, 1,00)	0,044	0,151	Downstream 3'UTR
	rs13245742	7:106590613	A/G	451	386	1,168	(1,02, 1,34)	0,025	0,176	Downstream 3'UTR
	rs13438464	7:106590592	G/T	448	375	1,195	(1,04, 1,37)	0,011	0,178	Downstream 3'UTR
	rs1636804	7:106591791	T/C	308	364	0,846	(0,73, 0,98)	0,030	0,148	Downstream 3'UTR
rs17315922	7:106605978	T/C	516	600	0,860	(0,76, 0,97)	0,012	0,289	Downstream 3'UTR	
PIK3CA	rs9823044	3:178943472	T/G	144	180	0,800	(0,64, 1,00)	0,045	0,069	Intrónico (16-17)
PIK3R3	rs7538978	1:46505554	G/A	627	551	1,138	(1,01, 1,28)	0,027	0,306	3'UTR
	rs1707322	1:46505647	T/C	624	549	1,137	(1,01, 1,28)	0,029	0,302	3'UTR
	rs785468	1:46521517	C/T	624	552	1,13	(1,01, 1,27)	0,036	0,303	Sinónimo (Exão 7)
	rs785465	1:46522577	T/G	625	553	1,13	(1,01, 1,27)	0,036	0,303	Intrónico (6-7)
	rs1768802	1:46547868	T/C	623	552	1,129	(1,01, 1,27)	0,038	0,303	Intrónico (1-2)
	rs785484	1:46573515	T/C	608	537	1,132	(1,01, 1,27)	0,036	0,299	Intrónico (1-2)
	rs1707303	1:46598273	T/G	609	541	1,126	(1,04, 1,26)	0,045	0,301	5'UTR
MAP2K2	rs4525614	19:4115184	G/A	193	148	1,304	(1,05, 1,62)	0,015	0,063	Intrónico (2-3)
FRS2	rs10748128	12:69827158	T/G	556	626	0,888	(0,79, 1,00)	0,042	0,329	Upstream 5'UTR
	rs17225279	12:69831537	A/G	264	311	0,848	(0,72, 1,00)	0,049	0,114	Upstream 5'UTR
	rs10878988	12:69832969	G/A	162	202	0,802	(0,65, 0,99)	0,036	0,072	Upstream 5'UTR
	rs17813377	12:69836133	G/A	347	414	0,838	(0,73, 0,99)	0,015	0,165	Upstream 5'UTR
	rs221068	12:69860206	C/T	646	728	0,887	(0,80, 0,99)	0,027	0,450	5'UTR
	rs221072	12:69865064	A/C	307	372	0,825	(0,71, 0,96)	0,013	0,139	Intrónico (1-2)
	rs221098	12:69886787	G/T	309	375	0,824	(0,71, 0,96)	0,012	0,140	Intrónico (1-2)
	rs522368	12:69899403	T/C	312	377	0,827	(0,71, 0,96)	0,013	0,140	Intrónico (1-2)

Gene	Marcador	Coordenadas	Minor/ Major	T	NT	OR	IC 95%	P-value	MAF	Localização
FRS2	rs520866	12:69916493	G/A	307	371	0,827	(0,71, 0,96)	0,014	0,139	Intrónico (1-2)
	rs487536	12:69940705	C/T	270	352	0,767	(0,65, 0,90)	0,001	0,134	Intrónico (4-5)
	rs579046	12:69943775	T/C	304	366	0,830	(0,71, 0,97)	0,017	0,138	Intrónico (4-5)
	rs516391	12:69958202	A/C	371	428	0,866	(0,75, 1,00)	0,043	0,173	Intrónico (5-6)

T – nº de indivíduos em que o alelo menor é transmitido

NT – nº de indivíduos em que o alelo menor não é transmitido

HWE – teste de equilíbrio de Hardy-Weinberg

MAF – *Minor allele frequency*

OR – *Odds ratio*

IC 95% - Intervalo de confiança a 95%

Minor/Major – Alelo com a menor frequência alélica / Alelo com a maior frequência alélica

Devido ao elevado número de marcadores em estudo, este teste de associação deveria ser corrigido para *multiple testing*. No entanto, decidimos não corrigir nenhum destes resultados para não sermos demasiado conservativos (por exemplo, utilizando a correcção de Bonferroni), e consequentemente desprezarmos resultados potencialmente relevantes para a patologia da doença, tendo em consideração que esperamos encontrar genes de baixo ou moderado efeito para esta doença complexa. Deste modo, considerou-se um valor de significância estatística de *P-value* igual ou inferior a 0,05 para associação do marcador genético com a doença.

2.2.2. Sub-população PT

Os resultados significativos para o teste de TDT de 1592 marcadores, em 229 famílias da sub-população PT encontram-se na tabela 20.

Foi encontrada associação dos genes *NTRK2*, *SOS1*, *SOS2*, *SHC3*, *PLCG2*, *PIK3R5*, *PIK3R1*, *PIK3CG*, *PIK3CD*, *AKT1* e *GAB1* com o autismo, na população portuguesa. Dois marcadores do gene *PIK3CG* e um marcador do gene *NTRK2* encontram-se associados na população portuguesa e na população AGP: rs1724278 ($P=0,045$) (*PIK3CG*), rs1636804 ($P=0,031$) (*PIK3CG*) e rs3860945 ($P=0,048$) (*NTRK2*).

No gene *NTRK2*, em treze marcadores foi identificada evidência de associação com o autismo ($0,008 < P < 0,046$), três dos quais encontram-se a montante de 5'UTR e apresentam os valores mais significativos de associação. Os dois marcadores de baixa frequência têm um *effect size* moderado, que se traduz na probabilidade de ter estes genótipos aumenta o risco de autismo em 2,1 vezes comparativamente à população em geral. Oito dos marcadores associados (associação nominal) estão localizados no intrão 14-15 e no intrão 19-20, regiões de susceptibilidade também identificadas na população total. Por último, o marcador rs526454, genotipado para este gene pertencente actualmente (segundo a base de dados genética ENSEMBL) ao recém descrito pseudogene ENSG00000226590.

Os genes *SOS1* e *SOS2*, que codificam moléculas transdutoras de sinal, apresentam associação nominal com o autismo em três dos seus marcadores ($0,008 < P < 0,029$), localizados nas regiões 3'UTR e intrónicas. No entanto, a região associada no gene *SOS2* na população AGP não é a mesma que na população portuguesa.

O gene *SHC3*, participante na via PI3K-AKT, está associado (associação nominal) ao autismo: rs3750339 localizado no exão 10 é um SNP sinónimo, rs7048604, rs2039307, ambos intrónicos (6-7 e 1-2, respectivamente) e rs11137545 localizado a montante de 5'UTR ($0,022 < P < 0,046$).

O gene *PLCG2* que codifica uma sub-unidade da molécula PLC γ tem uma participação fulcral na via biológica com a mesma denominação. Neste gene, oito marcadores estão associados com o autismo ($0,011 < P < 0,047$) nos intrões 5-6, 7-8, 21-22, 22-23 e 26-27, no entanto, estas regiões de susceptibilidade são diferentes das associadas na população AGP.

Em quatro dos genes que codificam as sub-unidades da molécula PI3K foram encontradas evidências de associação dos genes *PIK3R1* ($0,004 < P < 0,049$), *PIK3R5*, *PIK3CD* e *PIK3CG* ($0,002 < P < 0,049$) com o autismo nas regiões a montante e a jusante de 3'UTR e 5'UTR, também associadas na população AGP. Alguns destes marcadores comuns têm um sinal significativo de associação, no entanto, os seus valores de OR não são superiores a 1.

No gene *AKT1*, participante na via de sinalização de PI3K-AKT foi encontrada associação nominal ($0,002 < P < 0,020$). O snp rs4983387 (A>G), genotipado para este gene, está actualmente localizado no exão 2 do gene *ZBTB42* (*Zinc finger and BTB domain 42*) codificante de uma proteína de 422 aminoácidos. Por ser um polimorfismo não sinónimo, a presença do alelo G altera o codão original AAA, que codifica o aminoácido lisina (Lys), para GAA que codifica o ácido glutâmico ou glutamato. Os restantes marcadores estão localizados na região intrónica 2-3 e na 5'UTR.

Por fim, no gene *GAB1* que codifica a molécula gab1 presente na via biológica anterior, foi identificada associação nominal com o autismo em quatro dos seus marcadores ($0,019 < P < 0,039$) localizados na região 5'UTR e no intrão 1-2.

Nos genes *BDNF*, *SHC1*, *GRB2*, *HRAS*, *RAF1*, *MAP2K1*, *MAP2K2*, *PIK3R2*, *PIK2R3*, *PIK3CA*, *PIK3CB*, *PIK3CD*, *AKT2* e *PLCG1* não foi encontrada de associação dos seus marcadores com o autismo, na população portuguesa.

Tabela 20. Resultado significativos do teste de associação genética TDT, na sub-população PT, para os genes envolvidos nas vias biológicas activadas por BDNF/TrkB (a cinzento os SNPs associados na população AGP e sub-pop. PT).

Gene	Marcador	Coordenadas	Minor/ Major	T	NT	OR	IC 95%	p-value	MAF	Localização
NTRK2	rs12115546	9:87177985	G/A	38	18	2,111	(1,20, 3,69)	0,0075	0,063	Upstream 5'UTR
	rs10512149	9:87179562	G/A	38	18	2,111	(1,20, 3,69)	0,0075	0,063	Upstream 5'UTR
	rs1147199	9:87275895	T/C	58	83	0,698	(0,50, 0,98)	0,035	0,216	Upstream 5'UTR
	rs1187321	9:87283031	T/A	38	60	0,633	(0,42, 0,95)	0,026	0,229	Região regulatória
	rs1047896	9:87425973	T/C	82	55	1,491	(1,06, 2,09)	0,021	0,195	intrónico (14-15)
	rs10512154	9:87441475	A/G	88	62	1,419	(1,02, 1,96)	0,034	0,222	intrónico (14-15)
	rs1624327	9:87429290	T/C	94	66	1,424	(1,03, 1,95)	0,027	0,246	Intrónico (14-15)
	rs2277193	9:87574009	C/T	98	69	1,420	(1,04, 1,93)	0,025	0,245	Intrónico (19-20)
	rs6559840	9:87575499	T/C	104	72	1,444	(1,07, 1,95)	0,016	0,242	Intrónico (19-20)
	rs17087959	9:87575920	C/T	84	60	1,400	(1,00, 1,95)	0,046	0,194	Intrónico (19-20)
	rs1899637	9:87576008	T/C	49	28	1,750	(1,10, 2,70)	0,017	0,088	Intrónico (19-20)
	rs3860945	9:87585624	G/A	68	39	1,744	(1,76, 2,58)	0,005	0,127	Intrónico (19-20)
rs526454	9:87772807	G/A	24	40	0,600	(0,36, 0,99)	0,046	0,072	ENSG00000226590 (pseudogene)	
SOS2	rs8020226	14:50603129	C/T	25	46	0,543	(0,33, 0,88)	0,013	0,084	Intrónico (18-19)
SOS1	rs1059310	2:39212637	A/G	131	98	1,337	(1,03, 1,74)	0,029	0,400	3'UTR
	rs2888586	2:39247289	T/C	127	88	1,443	(1,1, 1,84)	0,008	0,441	Intrónico (11-12)
SHC3	rs3750399	9:91656963	T/C	120	91	1,319	(1,00, 1,05)	0,046	0,341	Sinónimo (exão10)
	rs7048604	9:91670399	T/C	117	88	1,330	(1,01, 1,9)	0,043	0,322	Intrónico (6-7)
	rs2039307	9:91769759	T/C	71	101	0,703	(0,52, 1,07)	0,022	0,266	Intrónico (1-2)
	rs11137545	9:91841972	G/A	68	95	0,715	(-)	0,034	0,262	Upstream 5'UTR
PLCG2	rs3935877	16:81900853	T/C	74	46	1,609	(1,11, 2,32)	0,011	0,155	Intrónico (5-6)
	rs4888184	16:81907481	G/A	75	47	1,596	(1,11, 2,30)	0,011	0,153	Intrónico (7-8)
	rs11644436	16:81912278	T/C	70	98	0,714	(0,53, 0,97)	0,031	0,235	Intrónico (7-8)
	rs8062633	16:81912983	T/C	43	70	0,614	(0,42, 0,90)	0,011	0,156	Intrónico (7-8)
	rs3922849	16:81955177	C/T	94	128	0,734	(0,56, 0,96)	0,022	0,465	Intrónico (21-22)
	rs9938623	16:81958173	T/C	88	118	0,745	(0,57, 0,98)	0,037	0,366	Intrónico (22-23)
	rs7201045	16:81959952	A/G	73	99	0,737	(0,54, 1,00)	0,047	0,292	Intrónico (22-23)
	rs8043619	16:81969436	T/C	110	81	1,358	(1, 02, 1,81)	0,036	0,291	Intrónico (26-27)
PIK3R5	rs442502	17:8773840	T/C	72	98	0,734	(-)	0,046	0,240	5'UTR
PIK3R1	rs10054692	5:67095390	G/A	45	28	1,607	(1,00, 1,38)	0,047	0,088	Upstream 5'UTR
	rs12234107	5:67221604	G/T	51	33	1,545	(1,00, 1,20)	0,049	0,102	Downstream 3'UTR
	rs1344367	5:67845698	T/G	92	135	0,681	(0,52, 1,09)	0,004	0,395	Downstream 3'UTR
	rs2607374	5:67959925	C/T	64	43	1,488	(1,01, 1,06)	0,042	0,147	Downstream 3'UTR
PIK3CG	rs757903	7:106500436	T/C	75	103	0,728	(0,54, 1,28)	0,036	0,282	Upstream 5'UTR
	rs11976223	7:106550161	C/T	133	102	1,304	(1,01, 1,15)	0,043	0,457	3'UTR
	rs17153616	7:106568489	G/T	37	60	0,616	(0,41, 1,02)	0,020	0,114	Downstream 3'UTR
	rs711440	7:106587464	T/C	42	75	0,560	(0,88, 1,02)	0,002	0,161	Downstream 3'UTR
	rs10953522	7:106588926	G/A	132	92	1,435	(1,00, 1,10)	0,008	0,461	Downstream 3'UTR
	rs1724278	7:106588730	T/C	45	80	0,562	(0,39, 1,03)	0,002	0,164	Downstream 3'UTR
	rs1636804	7:106591791	T/C	43	74	0,581	(0,40, 1,36)	0,004	0,159	Downstream 3'UTR
rs6942461	7:106612197	G/T	132	102	1,294	(1,00, 1,14)	0,049	0,482	Downstream 3'UTR	
PIK3CD	rs4076986	1:9688512	T/C	87	54	1,611	(-)	0,005	0,184	Upstream 5'UTR
AKT1	rs2498789	14:105249545	C/T	35	62	0,564	(0,37, 0,85)	0,006	0,119	Intrónico (2-3)

Gene	Marcador	Coordenadas	Minor/ Major	T	NT	OR	IC 95%	P-value	MAF	Localização
AKT1	rs2494743	14:105251720	C/T	30	52	0,576	(0,37, 0,90)	0,015	0,104	Intrónico (2-3)
	rs2498787	14:105256081	A/G	26	54	0,481	(0,30, 0,77)	0,002	0,097	Intrónico (2-3)
	rs2494752	14:105263608	A/G	27	47	0,574	(0,36, 0,92)	0,020	0,087	5'UTR
	rs2498784	14:105264963	T/C	27	47	0,574	(0,36, 0,92)	0,020	0,087	5'UTR
	rs4983387	14:105267728	G/A	15	34	0,441	(0,24, 0,81)	0,007	0,092	Não-sinónimo ZBTB42(Exão 2)
GAB1	rs13105156	4:144256713	C/T	26	44	0,590	(0,36, 0,96)	0,031	0,092	5'UTR
	rs3805253	4:144289502	C/T	98	129	0,759	(0,58, 0,99)	0,039	0,457	Intrónico (1-2)
	rs6537155	4:144293053	A/G	96	129	0,744	(0,57, 0,97)	0,028	0,456	Intrónico (1-2)
	rs1027851	4:144298152	A/G	95	130	0,730	(0,56, 0,95)	0,019	0,405	Intrónico (1-2)

T – nº de indivíduos em que o alelo menor é transmitido

NT – nº de indivíduos em que o alelo menor não é transmitido

HWE – teste de equilíbrio de Hardy-Weinberg

MAF – *Minor allele frequency*

OR – *Odds ratio*

IC 95% - Intervalo de confiança de 95%

Minor/Major – Alelo com a menor frequência alélica / Alelo com a maior frequência alélica

Devido ao elevado número de marcadores em estudo, este teste de associação deveria ser corrigido para *multiple testing*. No entanto, decidimos não corrigir nenhum destes resultados para não sermos demasiado conservativos (por exemplo, utilizando a correcção de Bonferroni), e consequentemente desprezarmos resultados potencialmente relevantes para a patologia da doença, tendo em consideração que esperamos encontrar genes de baixo ou moderado efeito para esta doença complexa. Deste modo, considerou-se um valor de significância estatística de *P-value* igual ou inferior a 0,05 para associação do marcador genético com a doença.

3. ESTUDOS DE INTERACÇÃO GENÉTICA

3.1. Genes *NTRK2*, *DLX1*, *DLX2*, *HOXA1*, *PITX1* e *EN1*

O grande desafio que se coloca actualmente na compreensão das doenças complexas consiste na detecção de interacções entre geneótipos que poderão estar a contribuir para um determinado fenótipo mais do que o esperado. Foi realizado o teste de interacção genética, através do programa PGMDR (v.07). O teste foi efectuado em 229 famílias da sub-população PT. Os melhores modelos para cada par de genes (*NTRK2* x *PITX1*, *NTRK2* x *HOXA1*, *NTRK2* x *DLX1*, *NTRK2* x *DLX2*, *NTRK2* x *EN1*) e para cada combinação de marcadores (2 a 2 e 3 a 3) são apresentados na tabela 21. A representação gráfica para cada “melhor modelo” identificado está no anexo F.

Tabela 21. Resultado de PGMDR para a interacção genética entre o gene *NTRK2* e os genes *DLX1*, *DLX2*, *HOXA1*, *PITX1* e *EN1*.

Ensaio	Melhor Modelo	TBA	Sign test (p-value)	CVC
DLX1xNTRK2	rs7042038 [<i>NTRK2</i>], rs4877909[<i>NTRK2</i>]	0,5019	3 (0,9453)	4
	rs11140659[<i>NTRK2</i>], rs1837304[<i>NTRK2</i>], rs10121443[<i>NTRK2</i>]	0,4834	6 (0,3770)	2
PITX1xNTRK2	rs7042038[<i>NTRK2</i>], rs4877909[<i>NTRK2</i>]	0,5769	7 (0,1719)	8
	rs985543[<i>NTRK2</i>], rs7042038[<i>NTRK2</i>], rs4877909[<i>NTRK2</i>]	0,5008	6 (0,3770)	3
EN1xNTRK2	rs332073[<i>EN1</i>], rs7856130[<i>NTRK2</i>]	0,4633	4 (0,8281)	2
	rs1344896[<i>EN1</i>], rs332665[<i>EN1</i>], rs7563505[<i>EN1</i>]	0,4619	4 (0,8281)	1
HOXA1xNTRK2	rs1020120[<i>NTRK2</i>], rs6559838[<i>NTRK2</i>]	0,5027	5 (0,6230)	5
	rs2918626[<i>NTRK2</i>], rs7855888[<i>NTRK2</i>], rs10435949[<i>NTRK2</i>]	0,5014	5 (0,6230)	3
DLX2xNTRK2	rs11140676[<i>NTRK2</i>], rs1020120[<i>NTRK2</i>]	0,5386	7 (0,1719)	4
	rs1020120[<i>NTRK2</i>], rs2579371[<i>NTRK2</i>], rs2200968[<i>NTRK2</i>]	0,4950	5 (0,6230)	2

TBA – Testing Balanced Accuracy

CVC – Crossvalidation consistency

Correia e colegas (Correia *et al.*, 2010) observaram que na presença do alelo T de rs1187321 (localizado na região do promotor do gene *NTRK2*) podem ocorrer alterações na ligação dos factores de transcrição e por conseguinte, influenciar a regulação do gene no autismo. Os resultados deste estudo de interacção estatística mostram que não há modelos de interacção significativos ($P > 0,05$), o que significa que a presença de dois ou três combinações de génotipos dos genes *NTRK2* e genes *homeobox* não contribuem para a multiplicidade do fenótipo da doença.

3.2. Genes *BDNF* e *NTRK2*

No seguimento da hipótese de que a via de sinalização BDNF/TrkB poderá estar alterada nos pacientes com autismo (Correia *et al.*, 2010) testou-se a acção de diferentes génotipos dos genes *BDNF* e *NTRK2*, na sub-população PT. Foram utilizados 323 marcadores para análise de interacção no programa PGMDR. Os resultados do melhor modelo de interacção estatística seleccionado para cada uma das combinações de marcadores estão representados na tabela 22. Não foi encontrado nenhum modelo de interacção significativo ($P < 0,05$). Apesar de as moléculas BDNF e TrkB interagirem fisicamente (como ligando e o respectivo receptor específico) estes resultados sugerem que os diferentes génotipos de cada gene não influenciam a função do outro gene, em pacientes com autismo. A combinação de dois ou três génotipos não contribui para a heterogeneidade do fenótipo.

Tabela 22. Resultados de PGMDR para a interacção entre os marcadores dos genes *NTRK2* e *BDNF*.

Ensaio	Modelo	TBA	Sign Test (<i>p</i> -value)	CVC
BDNF x NTRK2	rs4111912[<i>NTRK2</i>], rs7048015[<i>NTRK2</i>]	0,4643	3 (0,9453)	3
	rs1147199[<i>NTRK2</i>], rs1187286[<i>NTRK2</i>], rs10435949[<i>NTRK2</i>]	0,5226	6 (0,3770)	2

TBA – *Testing Balanced Accuracy*

CVC – *Crossvalidation consistency*

4. ENRIQUECIMENTO DO SINAL DE ASSOCIAÇÃO DE VIAS BIOLÓGICAS NO AUTISMO

A análise de enriquecimento do sinal de associação de vias biológicas tem dois objectivos: identificar se as vias biológicas activadas por BDNF/TrkB têm enriquecimento do sinal de associação no autismo e, de forma mais abrangente, identificar novas vias biológicas associadas na susceptibilidade à doença.

O resultado da análise de enriquecimento por sinal de associação dos três grupos de genes participantes nas três vias activadas a jusante de BDNF/TrkB (denominadas no estudo por “PI3K-AKTsignaling”, “MAPKsignaling” e “PLCGsignaling”) não foi significativo ($P=0,270$; $P=0,328$; $P=0,917$, respectivamente) (tabela 23). Os resultados da análise de enriquecimento dos sinais de associação dos genes das vias biológicas pré-definidas identificaram três vias biológicas, ordenadas pelo valor significativo de enriquecimento do sinal de associação (*P*-value) e por correcção para *multiple testing* de FDR (tabela 23): a via biológica do metabolismo de compostos isoprenóides (GO0006720) (NES= 2,218; $P<0,0001$; FDR=0,110), a via biológica da Aminoacilação de tRNA na síntese proteica (GO0006418) (NES=3,139; $P=0,002$; FDR=0,021) e a via biológica da m-Calpaína na mobilidade celular (mCalpain *Pathway*) (NES= 3,480; $P=0,002$ FDR=0,026).

Tabela 23. Resultados significativos do enriquecimento por sinal de associação dos genes de 1655 vias biológicas/grupos de genes em 1369 famílias da população AGP, através do software GenGen. Os resultados estão ordenados por *p*-value ($P<0,05$) e por FDR.

Via biológica	Descrição	Nº Genes	NES	<i>P</i> -value	FDR
GO0006720	Metabolismo dos compostos isoprenóides	17	2,218	0,000	0,110
GO0006418	Aminoacilação de tRNA na síntese proteica	32	3,139	0,002	0,021
mCalpainPathway	mCalpaína e mobilidade celular	23	3,480	0,002	0,026
(...)					
PLCGsignaling	Via PLCγ	11	0,607	0,270	0,651
MAPKsignaling	Via biológica MAPK-MEK-ERK	7	0,418	0,328	0,734
PI3K-AKTsignaling	Via biológica PI3K-AKT	15	-1,328	0,917	0,987

NES - *Normalized Enrichment Score*

FDR - *False Discovery Rate*

A via biológica do metabolismo dos compostos isoprenóides inclui os seus metabolitos: digoxina, dolichol e ubiquinona. Os isoprenóides são uma classe de lípidos sintetizados através da transformação de unidades de isopreno. A função dos metabolitos desta via biológica já foi avaliada no autismo (Kurup and Kurup, 2003). Os resultados do estudo revelam níveis elevados de digoxina, molécula secretada no hipotálamo e inibidor das bombas sódio-potássio (Na^+/K^+ -ATPase), estas últimas presentes em quase todas as células e responsáveis por manter o potencial de acção das mesmas. A inibição das bombas sódio-potássio resulta no aumento de Ca^{2+} intracelular e na diminuição dos níveis de magnésio.

O processo de aminoacilação do tRNA (RNA de transferência) na síntese proteica estabelece as regras do código genético (Park *et al.*, 2008). A aminoacilação do tRNA é catalizada pela enzima aminoacil de tRNA sintetase (*Aminoacyl tRNA synthetase*, AARS) ocorre entre o grupo carboxilo COOH do aminoácido e o grupo 3'OH do nucleotido terminal. Alguns estudos na literatura revelam que mutações nos genes que codificam AARSs podem estar ligadas à susceptibilidade de doenças como o cancro, patologias neuronais, doenças autoimunes, entre outras (Park *et al.*, 2008). No entanto, não há registo na literatura sobre a influência destas enzimas na etiologia das ASD.

A via de sinalização m-Calpaína na mobilidade celular (BioCarta) é a quarta via biológica com maior evidência de enriquecimento por sinal de associação no autismo. A família das Calpaínas (proteases intracelulares) está envolvida em processos de apoptose, proliferação e migração celular. No SNC (função menos estudada) estão implicadas em processos de neurodegeneração, plasticidade sináptica, memória e aprendizagem (Franco and Huttenlocher, 2005). Duas das isoformas são expressas no cérebro - μ -Calpain e m-Calpain - que se encontram nos axónios e dendrites dos neurónios. A m-Calpaína é activada nos neurónios e duas formas: dependente da concentração de cálcio (Ca^{2+}) e poder ser activada por EGF e BDNF, através de fosforilação catalizada por MAPK (Zadran *et al.*, 2010). A alteração da actividade destas calpaínas no cérebro já foi associada à doença de Alzheimer (Tsuji *et al.*, 1998).

Para as vias biológicas de topo criámos novos conjuntos de genes, no qual foi excluído um gene diferente, de cada vez que analisámos. Este processo foi realizado para verificar se o *software* GenGen está a utilizar todos os genes de uma mesma via biológica e se todos estão a contribuir de igual forma para o sinal de associação da via. A análise realizada denominou-se por "*Leave one out*". Tomando como exemplo a via biológica da m-Calpaína na mobilidade celular foram criadas novas vias biológicas excluindo, para cada uma, um gene diferente. Estas novas vias biológicas foram re-analisadas no *software* GenGen e os seus resultados estão na seguinte tabela 24:

Tabela 24. Resultado da análise de "Leave One Out" na via biológica mCalpaína.

Via biológica mCalpaína	Nº genes	NES	P-value	FDR
mCalpainPathway_15	22	3,68	0,0000	0,0350
mCalpainPathway_21	22	3,66	0,0010	0,0290
mCalpainPathway_22	22	3,66	0,0010	0,0260
mCalpainPathway	23	3,48	0,0020	0,0260
mCalpainPathway_10	22	3,20	0,0020	0,0270
mCalpainPathway_13	22	3,40	0,0020	0,0250
mCalpainPathway_17	22	3,34	0,0020	0,0250
mCalpainPathway_3	22	2,85	0,0020	0,0280
mCalpainPathway_4	22	3,38	0,0020	0,0250
mCalpainPathway_5	22	3,42	0,0020	0,0270
mCalpainPathway_6	22	3,54	0,0020	0,0280
mCalpainPathway_9	22	3,42	0,0020	0,0250
(...)				
mCalpainPathway_11	22	3,03	0,0030	0,0250
mCalpainPathway_18	22	2,74	0,0030	0,0260

O resultado desta análise mostra que o sinal de significância da via biológica (a negrito) não ficou alterado quando excluímos, um a um, cada gene do grupo. Isto significa que todos os genes da via biológica contribuíram de igual forma para este sinal de significância.

Por último, observemos os parâmetros obtidos pela análise bioinformática. O *P-value* é definido como a probabilidade da diferença que observamos ser pelo menos tão grande como a que acontece se a hipótese nula for verdade. Este parâmetro dá-nos o nível de significância do teste estatístico. O *P-value* de permutação representa o número de vezes que determinada via biológica teve um melhor valor de ES ao acaso do que o valor real do teste. Por exemplo, o grupo de genes "PLCysignaling" tem um valor de permutação de $P=0,270$ o que significa que este grupo teve 270 vezes um valor de teste ao acaso superior ao valor de teste real, no conjunto das 1000 permutações ($P=270/1000$). Este resultado não é significativo porque se deve ao acaso e não ao valor real de teste. A correcção dos valores de NES para *multiple testing* foi realizada por FDR (*False Discovery Rate*) (Benjamini and Hochberg, 1995). O teste FDR estima a probabilidade que um conjunto com um determinado NES tem de representar um evento falso positivo.

V. DISCUSSÃO DE RESULTADOS

1. RASTREIO DE VARIANTES NO GENE *NTRK2*

O resultado do rastreio de variantes genéticas nas regiões do gene *NTRK2* anteriormente associadas ao autismo identificou uma nova variante genética no intrão 6-7 e um SNP rs3739807 no intrão 21-22 (anteriormente descrito). Nenhuma das variantes identificadas tem um fenótipo associado, excepto o atraso na linguagem que poderá ser relevante (tabelas 13 e 15), característica comum a todos os indivíduos onde foi encontrada a alteração. No caso de rs3739807 (intrão 21-22), provavelmente este SNP não seria a variante genética que estaria em LD com os marcadores de risco (*tag* SNP) anteriormente associados ao autismo, na sub-população PT (Correia *et al.*, 2010) porque não tem significado funcional e por se encontrar na população controlo, isto é, não é exclusiva dos pacientes com autismo. O próximo passo poderá ser o rastreio de um maior número de indivíduos e alargar a região a sequenciar.

A presença da nova variante rara (c.288-15T>C) poderá ter influência no mecanismo de *splicing* do gene *NTRK2*. A alteração de um local de ligação de proteínas HnRNP e o aparecimento de um novo local de ligação de proteínas SRp20 na mesma região, aliados ao facto destas proteínas terem funções antagonistas, podem influenciar o processo de *splicing* do gene. O desaparecimento de um local de ligação das proteínas HuR pode influenciar a estabilidade do mRNA e comprometer o mecanismo de *splicing* (figura 13). Desta forma, a presença desta variante intrónica, poderá alterar a ligação de proteínas envolvidas no mecanismo de *splicing* do gene *NTRK2*. Para confirmação destes resultados deverão ser realizados ensaios funcionais, assim como o rastreio de um maior número de indivíduos com autismo e controlos para compreender de que forma esta variante poderá influenciar a expressão do gene *NTRK2*.

A identificação e submissão da nova variante na base de dados dbSNP deu mais um contributo para o conhecimento do genoma humano.

2. TESTES DE ASSOCIAÇÃO GENÉTICA

Os testes de associação foram realizados na população AGP e na sub-população PT por ter sido observado que esta sub-população tem um *background* genético distinto das restantes populações que integraram o estudo GWAS realizado (anexo D).

2.1. População AGP

Os factores de transcrição com homeodomínio estão envolvidos no desenvolvimento cerebral e na diferenciação celular constituindo por isso, bons genes candidatos nas doenças de desenvolvimento neuronal. Neste estudo, alguns marcadores do gene *homeobox EN1* encontram-se associados ao autismo na população AGP. Assim sendo, o próximo passo será o rastreio das regiões 5'UTR e 3'UTR do gene com o intuito de identificar variantes genéticas de susceptibilidade ao autismo. Dado que a proteína *Engrailed-1* desempenha um papel importante na sobrevivência dos neurónios dopaminérgicos (Haubenberger, 2009) será interessante compreender se a expressão do gene que o codifica estará afectada nos pacientes com autismo e tentar identificar/compreender se a ligação deste factor de transcrição ao gene *EN1* e, posteriormente, realizar ensaios de expressão da proteína.

Ao nível da via de sinalização BDNF/TrkB, os resultados revelam associação dos marcadores do gene *NTRK2* com o autismo e a ausência de associação dos marcadores do gene *BDNF* com o autismo. Neste sentido, será necessário rastrear as regiões de susceptibilidade identificadas no gene *NTRK2* - região 5'UTR, 3'UTR, intrões (e respectivos exões) 14, 15, 19 e 20 - com o objectivo de encontrar as variantes genéticas de risco. No caso do gene *BDNF* no presente estudo não está associado ao autismo, contudo, já foi encontrada associação de um haplótipo numa amostra populacional chinesa de pacientes com autismo (Cheng *et al.*, 2009) e também noutras doenças como a doença bipolar e a ADHD (Dmitrzak-Weglarz *et al.*, 2008; Conner *et al.*, 2008), assim como elevados níveis da neurotrofina em pacientes com autismo.

Os restantes resultados revelam associação de diferentes marcadores de genes participantes nas três vias biológicas a jusante da ligação BDNF/TrkB, com o autismo: MEK/ERK (genes *SOS2*, *SHC3*, *FRS2* e *MAP2K2*), PI3K-AKT (genes *SCH3*, *FRS2*, *PIK3R5*, *PIK3R1*, *PIK3CG*, *PIK3CA* e *PIK3R3*) e PLC γ (gene *PLCG2*). Os genes *SOS2*, *SHC3* e *FRS2* codificam as proteínas fosforiladas por TrkB e são importantes intermediárias entre a via de sinalização BDNF/TrkB e as vias biológicas por si induzidas. Consequentemente, mutações nestes genes poderão conduzir a uma função proteica deficiente e, deste modo, influenciar o funcionamento normal das vias biológicas. O mesmo se aplica para os genes que codificam algumas das sub-unidades da molécula de PI3K nas quais a presença de variantes funcionais nas regiões catalíticas e regulatórias poderão influenciar a função da molécula e, consequentemente, influenciar o funcionamento da via biológica PI3K. A enzima *mitogen-activated protein kinase kinase 2*, codificada pelo gene *MAP2K2*, tem um papel importante na transdução de sinal de *Mitogen growth factor*, assim como, fosforila e activa MAPK1/ERK2 e MAPK2/ERK3. Mutações no gene *MAP2K2* estão associadas ao síndrome Cardio-Facio-Cutânea, à deficiência mental, entre outros (Rodriguez-Viciano and Rauen, 2008). Será necessário rastrear as diferentes regiões de susceptibilidade de todos os genes associados para identificar as variantes genéticas de risco e,

no caso de existirem variantes funcionais, realizarem-se posteriormente estudos de expressão e funcionais. A alteração da expressão de cada um dos genes poderá comprometer o funcionamento destas importantes vias biológicas no desenvolvimento e funcionamento cerebral e investigar se estes genes influenciam, de facto, o risco à doença e de que forma é observado as suas alterações a nível fenotípico.

Dos resultados de associação há quatro genes que apresentam os sinais de associação mais significativos na população AGP: *NTRK2*, *PIK3R1*, *PIK3CG* e *FRS2* (SNPs com $P < 0,01$). No caso dos genes *NTRK2* e *PIK3R1*, que codificam respectivamente o receptor TrkB e a sub-unidade regulatória p85 α de PI3K, os marcadores mais associados encontram-se na região a montante da região não codificada 5'UTR, provavelmente, em LD com variantes genéticas na região do promotor dos genes. Assim, se ocorrerem alterações na regulação da expressão dos genes, as respectivas proteínas poderão estar alteradas. O gene *PIK3CG* que codifica a sub-unidade catalítica p110 γ de PI3K, foi anteriormente encontrado associado ao autismo (Serajee *et al.*, 2003) e juntamente com a associação encontrada no presente estudo sugere que alterações na proteína PI3K poderão implicar a via biológica PI3K-AKT no autismo. Por fim, o gene *FRS2* codifica uma molécula com a mesma denominação que é fosforilada pelo receptor TrkB e que posteriormente vai activar, juntamente com outras moléculas, as vias biológicas MAPK e PI3K-AKT. O rastreio de variantes genéticas nas regiões de susceptibilidade poderá ser um primeiro passo para se compreender se a sua expressão estará alterada nos pacientes com autismo. Se a função desta molécula não for normal, então a activação de algumas vias biológicas poderá estar alterada.

Este será um ponto de partida para o rastreio de novas regiões de susceptibilidade de genes que poderão dar novas pistas sobre as vias biológicas que provavelmente estarão envolvidas na etiologia do autismo e contribuem para as características fenotípicas da doença.

2.2. Sub-População PT

Partindo da associação anteriormente observada entre o autismo e os marcadores dos intrões 5-6-7-8 e 22-23 e região 5' UTR (rs1187321) do gene *NTRK2* na população portuguesa (Correia *et al.*, 2010), os resultados de associação do presente estudo para a sub-população PT, revelam novas regiões de susceptibilidade nos intrões 14-15 e 19-20. Os sinais de associação mais fortes encontram-se nos SNPs localizados a montante da região 5' UTR que aliado à anterior associação de rs1187321 (5' UTR) observada leva a considerá-la como uma região forte de susceptibilidade à doença. Será necessário rastrear estas novas regiões para identificar as variantes genéticas de risco. À semelhança do que foi encontrado na população AGP, alguns marcadores dos genes que codificam duas sub-unidades da região catalítica e uma sub-unidade da região regulatória de PI3K também se encontram associados ao autismo. Será importante

rastrear, por sequenciação, estas regiões de susceptibilidade e compreender se a proteína PI3K terá a sua função alterada, estando implicada na etiologia da doença.

É ainda de notar, o caso especial dos genes *AKT1*, *ZBTB42* e *PIK3CD* nos quais se observou associação do autismo com alguns dos seus marcadores apenas na sub-população portuguesa: dois marcadores no intrão 2-3, um SNP não sinónimo no exão 2 e na região a montante de 5' UTR, respectivamente. O gene *ZBTB42* é um gene que codifica uma proteína *zinc finger*, isto é, codifica domínios proteicos que se ligam ao DNA e que coordenam os íons zinco, essenciais para a estabilidade das cadeias de DNA. Recentemente, o gene *ZNF804A* que codifica um outro *zinc finger* está descrito na literatura como associado a várias doenças, como a esquizofrenia, doença bipolar e ASD (O'Donovan *et al.*, 2009; Willemsen *et al.*, 2009). De futuro, seria interessante rastrear estas regiões para encontrar variantes funcionais de risco. A presença destas associações genéticas apenas nesta sub-população advém da heterogeneidade genética entre as populações (anexo D) e de diferentes padrões de LD existentes entre elas. Cada população poderá ter diferentes variantes genéticas que poderão ser responsáveis por diferentes genótipos de risco no autismo. Desta forma, foi interessante observar que diferentes variantes genéticas podem contribuir para um mesmo fenótipo, o que nos indica que é necessário olhar também para as sub-populações de forma individual e identificar as suas variantes raras. Ao realizarem-se estudos com uma amostra populacional grande como os GWAS, por exemplo, as variantes genéticas "características" das diferentes sub-populações poderão ficar "diluídas" quando se analisam os resultados como um todo.

3. TESTES DE INTERACÇÃO GENÉTICA

Os estudos de interacção genética foram realizados apenas na sub-população PT porque o programa utilizado (GMDR) não comporta o enorme volume de dados relativos à população AGP. Num futuro próximo, novos programas estatísticos de interacção genética surgirão com uma maior capacidade de processamento de dados. Embora com esta limitação, este foi um primeiro passo nos estudos de interacção genética com o gene *NTRK2*. A interacção genética, ou epistasia, pressupõe que a ocorrência do efeito multiplicativo de diferentes genótipos em diferentes *loci* contribui para um determinado fenótipo. Neste estudo, analisámos se a combinação de dois ou de três genótipos altera consideravelmente o fenótipo. No entanto, não foi detectada nenhuma combinação de genótipos que revelasse elevado risco na doença. Em particular, a combinação dos marcadores dos genes *homeobox* com o gene *NTRK2* não interagem de forma a contribuir para o risco da doença. O mesmo sucede entre os marcadores dos genes *NTRK2* e *BDNF*, a sua combinação de genótipos não aumenta consideravelmente o risco para a doença. Esta estratégia de interacção estatística para este conjunto de genes não

se revelou informativa, mas no contexto das doenças complexas faz sentido analisar-se interacções entre genótipos de diferentes genes. Num trabalho futuro poderão ser realizadas análises de interacção estatística entre marcadores de genes das vias biológicas activadas pela ligação BDNF/TrkB.

4. ANÁLISE DE ENRIQUECIMENTO DO SINAL DE ASSOCIAÇÃO

O *software* Gengen permitiu identificar três vias biológicas cujo sinal de associação está mais enriquecido no autismo e que estão envolvidas em processos biológicos diferentes, tais como, proliferação e migração neuronal, metabolismo de isoprenóides e aminoacilação do tRNA. Apesar das suas diferenças funcionais, os genes que participam nestas vias são fortes candidatos na etiologia do autismo. Esta complexa metodologia de análise bioinformática revelou ser uma ferramenta útil de análise em doenças complexas, como o autismo, na medida em que permitiu criar novas hipóteses biológicas.

No caso dos três grupos de genes de interesse no estudo, a análise de enriquecimentos não foi informativa. À luz destes resultados e apesar dos marcadores dos genes se encontrarem positivamente associados ao autismo, o fenótipo não está mais enriquecido nestes conjuntos biológicos de genes, não constituindo factores de susceptibilidade à doença.

Este estudo permitiu compreender o envolvimento da via de sinalização BDNF/TrkB na susceptibilidade ao autismo, mas a mais-valia de possuir resultados de um GWAS permitiu identificar novos grupos de genes associados ao risco da doença. Este abordagem gerou novas hipóteses sobre novos genes candidatos e novas vias biológicas na susceptibilidade ao autismo.

Um dos componentes resultantes do metabolismo dos compostos isoprenóides, a digoxina hipotalâmica parece ter os seus níveis mais elevados nos indivíduos com autismo (Kurup and Kurup, 2003). Este facto aliado à sua função inibidora das bombas de sódio-potássio, que por sua vez são responsáveis pelos potenciais de acção das células coloca a seguinte questão: será que a alteração dos níveis desta molécula poderá ser uma das causas que contribuem para a etiologia do autismo? Assim, será interessante em futuros estudos avaliar os níveis de digoxina produzidas no hipotálamo num maior número de pacientes com autismo e compreender quais os mecanismos celulares que regulam a expressão desta molécula.

Não há registo científico de que a via biológica do processo de aminoacilação do tRNA na síntese proteica esteja envolvida na etiologia do autismo. No entanto, pelo facto de existirem mutações nos genes AARSs que estão associados a outras patologias neuronais, será

interessante compreender se a expressão dos genes que codificam as AARSs estarão alterados nos pacientes com autismo (Park *et al.*, 2008).

Por fim, a via biológica da m-calpaína e a sua função na mobilidade celular. A m-calpaína regula a forma e mobilidade celular (em vários tipos de células), mas no SNC está implicada em processos de neurodegeneração, plasticidade sináptica, memória e aprendizagem. As funções que desempenha tornam-na como uma candidata pertinente na etiologia do autismo. Como referido no capítulo anterior a activação de m-Calpaína nos neurónios é dependente dos níveis de cálcio ou das moléculas EGF (Glading, 2004) e BDNF (Zadran, 2010) que activam m-Calpaína através de fosforilação catalizada por MAPK. Logo, se BDNF e MAPK participam em funções relevantes como plasticidade sináptica, memória e aprendizagem (entre outras) então alterações nas suas vias biológicas poderão alterar o funcionamento normal das calpaínas no sistema nervoso. Será que a expressão da m-calpaína está alterada nos pacientes com autismo e alterações na sua expressão influenciam os processos sinápticos?

As três vias biológicas em discussão participam em processos que estão relacionados com o desenvolvimento e funcionamento no SNC o que vai de encontro com a literatura científica, que defende a disrupção de mecanismos de neurotransmissão como estando envolvidos na etiologia do autismo (Levitt and Campbell, 2009). Estas vias poderão ser exploradas através de estudos de associação e interacção genética, funcionais e de expressão para compreender o seu contributo na etiologia da doença. Este estudo deu um enorme contributo na análise dos dados provenientes do GWAS gerando novas hipóteses que envolvem múltiplos genes, ao invés da estratégia de análise de um único marcador de cada vez. No entanto, este tipo de análise bioinformática ainda levanta algumas questões pertinentes de delineamento experimental, complexidade na preparação de ficheiros, execução informática e análise dos resultados, e que exige bastantes conhecimentos de informática e programação. Através deste estudo e da literatura científica depreendo que a análise de grupos de genes/vias biológicas será a abordagem a seguir no futuro para todas as doenças complexas, incluindo o autismo.

5. ANÁLISE ESTATÍSTICA: ABORDAGENS, VANTAGENS E LIMITAÇÕES

A análise estatística do estudo foi baseada nos genótipos do GWAS, obtidos no âmbito do consórcio AGP. O GWAS representa um grande avanço no estudo de doenças complexas porque utiliza um grande número de indivíduos e de marcadores genéticos uniformemente distribuídos por todo o genoma o que aumenta a probabilidade de identificar variantes comuns de risco. A utilização de dados de famílias foi também uma vantagem porque permitiu controlar o problema da estratificação populacional (que ocorre quando se realizam estudos de

associação caso-controlo). Esta metodologia permitiu, para além dos resultados de associação entre cada marcador e a doença, analisar interacções entre genes e grupos de genes mais enriquecidos no autismo. A sua maior valia foi gerar novas hipóteses.

Apesar de promissores os GWAS apresentam limitações conhecidas. As variantes genéticas detectadas no GWAS são, na sua grande maioria, variantes com MAF superior a 5% não detectando variantes com MAF baixa ($MAF < 1\%$), isto é, variantes raras e difíceis de serem detectadas nas plataformas de genotipagem de alta densidade. Estudos anteriores revelam que apesar de 90% do autismo idiopático ser explicado pela componente genética, as variáveis comuns identificadas só explicam 5% a 10% da etiologia da doença (Maher, 2008; Manolio *et al*, 2009). A restante heritabilidade pode ser explicada pela presença de variantes raras com elevado risco relativo (*effect size*) ($3 < OR < 5$), ou seja, variantes que aumentam o risco de ter a doença três a cinco vezes em relação à população em geral mas que dificilmente são detectáveis nos GWAS. Os resultados da análise de associação revelam variantes genéticas de risco mas com OR baixo ($OR < 1,2$), ou seja, estas variantes contribuem para o risco da doença mas explicam apenas uma pequena parte da variabilidade por si só. Para contornar este problema é necessário provavelmente partir para outras metodologias como a sequenciação do genoma (WGS, *Whole Genome Sequencing*).

Num ensaio individual há um risco associado ao mesmo de se obter um falso positivo (erro tipo I). Os resultados falsos positivos podem ocorrer por múltiplas razões, tais como o enviesamento do delineamento experimental, erros de genotipagem e *multiple testing*. Para minimizar os erros recorreu-se a diferentes métodos durante o estudo. A escolha dos marcadores de genotipagem foi realizada pelo fabricante da plataforma de genotipagem o que evita o enviesamento do delineamento experimental por parte do experimentador. Apesar da plataforma de genotipagem utilizada ter uma grande precisão na detecção dos genótipos foram efectuados testes de controlo de qualidade, com recurso aos indivíduos HapMap e controlos internos (duplicados), análise de *call rate*, erros de genotipagem, erros mendelianos e desvios ao HWE para evitar os falsos positivos. Contudo, não foi realizada correcção para *multiple testing* para os resultados de associação genética porque este ainda é um assunto controverso. Existem testes demasiado conservativos (como por exemplo a correcção de Bonferroni) que assumem cada teste como independente. Desta forma, são necessários métodos mais robustos e fiáveis. Mesmo que se tivesse utilizado a correcção para *multiple testing*, não esperaríamos que os nossos resultados continuassem a ser significativos após a correcção, uma vez que o GWAS realizado inclui uma quantidade enorme de marcadores. No entanto, no âmbito de uma doença complexa, resultados menos significativos não podem ser excluídos à partida e devem ser incluídos em estudos futuros uma vez que esperamos que variantes comuns de pequenos efeitos por si só estejam a contribuir para o fenótipo precipitando a doença, conjuntamente com factores ambientais não identificados.

Os resultados do estudo de enriquecimento do sinal de associação de vias biológicas foram os únicos resultados corrigidos para *multiple testing* através do método FDR porque esta correção está incluída no *software* GenGen. O método FDR tem poder para identificar falsos positivos e é muito semelhante à correção de Bonferroni (Benjamini and Hochberg, 1995). Assume que os *p-values* correspondentes a hipóteses nulas verdadeiras são independentes e distribuídos uniformemente, uma suposição que é provavelmente violada quando há enorme LD entre SNP.

VI. DISCUSSÃO GERAL E PERSPECTIVAS FUTURAS

O presente estudo analisou o envolvimento da via de sinalização BDNF/TrkB e das vias biológicas por si activadas no autismo através de diferentes abordagens: sequenciação de exões num gene de susceptibilidade, estudos de associação genética, estudos de interacção genética e análise de enriquecimento do sinal de associação de vias biológicas.

Os resultados de associação e de interacção genética sugerem que, provavelmente, a desregulação da via de sinalização BDNF/TrkB ocorre ao nível do receptor TrkB e não ao nível do ligando, BDNF. As alterações estruturais, funcionais e de afinidade do receptor poderão levar à deficiente ligação de BDNF e consequentes alterações na activação das vias biológicas a jusante desta ligação. Provavelmente, são diferentes os factores que estão a influenciar a região regulatória do gene *NTRK2* razão pela qual é importante o rastreio desta região e a realização de estudos de expressão. Para além da região 5'UTR do gene *NTRK2*, as novas regiões de susceptibilidade nos intrões 14 e 15 (ver anexo A) codificam a região extracelular e transmembranar da proteína e o exão 19, um exão alternativo, é usado numa das isoformas truncadas do receptor (TrkB-T-Shc). Assim, é possível que alterações nestas regiões possam influenciar a função e estrutura do receptor e, consequentemente, alterar a sinalização mediada por TrkB de outras vias biológicas. Esta observação aliada ao elevado número de isoformas descritas na literatura (Stoilov *et al.*, 2002; Luberg *et al.*, 2010) sugere uma elevada complexidade na síntese, regulação e função do gene. No entanto, nenhum estudo publicado identificou qual ou quais as isoformas expressas no tecido cerebral, em indivíduos com autismo. Desta forma, torna-se pertinente avaliar este gene a três níveis: o rastreio, por sequenciação, das regiões de susceptibilidade para encontrar variantes genéticas de risco, identificar as moléculas que regulam a expressão do gene *NTRK2* e ao nível da proteína identificar e quantificar as isoformas expressas nos indivíduos com autismo no cérebro *post-mortem* e nas células do plasma.

Como observado anteriormente (Correia *et al.*, 2010) não foi encontrada evidência de associação do gene *BDNF* ao autismo, embora os níveis desta neurotrofina sejam mais elevados nos pacientes com autismo relativamente à população controlo (na população PT) (Correia *et al.*, 2010). Neste sentido, seria interessante identificar os genes que poderão regular a expressão do gene *BDNF* e secreção da proteína, assim como alargar a amostra populacional em estudo, de forma a compreender o seu envolvimento na etiologia do autismo.

A jusante da ligação BDNF/TrkB, das três vias analisadas as alterações nos genes que participam na via biológica PI3K-AKT parecem contribuir significativamente para a

susceptibilidade ao autismo. Os estudos de associação identificaram novas regiões de susceptibilidade dos genes *PIK3CG*, *PIK3R1*, *FRS2*, *AKT1* e *PIK3CD*. Foi corroborada a hipótese inicial de que alterações em genes nesta via estariam associadas ao autismo e o seu contributo torna-se mais relevante porque a via em questão desempenha um importante papel nos processos de desenvolvimento cerebral, aprendizagem, memória, sobrevivência e diferenciação neuronal. Neste sentido, as alterações ou desregulação da via biológica de PI3K-AKT poderá estar envolvida na etiologia da doença. Anteriores associações dos seus genes sublinham o papel desta via: o gene *AKT1*, que participa na regulação de proteínas apoptóticas e factores de transcrição, está associado à esquizofrenia e doença bipolar e o gene *PIK3CG* já foi associado anteriormente ao autismo (Serajee *et al.*, 2003). Seria importante o rastreio das regiões de susceptibilidade para identificar novas variantes funcionais, assim como estudos de interacção genética, de expressão e consequentes ensaios funcionais para compreender o papel desta via biológica no autismo idiopático.

Este estudo salientou a importância de se analisarem, em doenças complexas, os genes não de forma isolada mas sim compreender como podem interagir, porque biologicamente todos os genes interagem entre si e é esta contribuição múltipla que poderá explicar as características clínicas da doença. A contribuição para o fenótipo pode ser perdida se se analisarem os efeitos genéticos de forma independente. Desta forma, o estudo de vias biológicas continua a ser um importante veículo de estudo no autismo podendo abrir novos caminhos na descoberta de novas terapias de intervenção precoce e/ou diagnóstico na doença.

No contexto de uma doença complexa, os resultados do presente estudo levam à seguinte questão: será que a teoria de CDCV (*Common disease common variance*) ainda se aplica ou a hipótese alternativa de que múltiplas e diferentes variantes raras são as variantes causais na doença? (Pritchard and Cox, 2002; Robinson, 2010) Actualmente, os resultados convergem para a segunda hipótese e defendem que diferentes variantes raras que ocorrem em diferentes indivíduos são responsáveis pela expressão de determinado fenótipo da doença. Por um lado, diferentes indivíduos de uma mesma população terão diferentes variantes causais para um determinado fenótipo e por outro lado, indivíduos pertencentes as duas populações diferentes poderão ter diferentes variantes causais para o mesmo fenótipo. A importância das variantes raras nas doenças complexas torna-se relevante para compreender os seus efeitos sobre o fenótipo da doença. No entanto, será o conjunto das variantes raras, das variantes comuns e factores ambientais que interagindo entre si conduzem ao autismo. Começam a surgir novas tecnologias de sequenciação capazes de rastrear todo o genoma e novos projectos genómicos como o 1000 *Genome Project* ^[3], fruto da cooperação internacional e que permitirá reunir mais informação sobre a variação genética humana, incluindo SNP, variações estruturais (por exemplo, CNV) e haplótipos, provenientes de 2000 indivíduos, de 20 populações distintas. Este é o caminho que se seguirá no futuro das doenças complexas.

Muito ainda permanece por esclarecer no autismo. A complexa arquitectura genética da doença ainda esconde muitos factores de susceptibilidade que continuam por identificar. Os resultados dos GWAS não se revelaram suficientemente robustos para detectar as variantes de risco nas doenças complexas e novas abordagens nascem para satisfazer os novos desafios. Se por um lado se está abandonar a análise de um único *locus*, por outro lado, crescem as metodologias estatísticas capazes de analisar múltiplos *loci* em simultâneo. Após a era do GWAS surge a era do *deep-sequencing* em todo o genoma. Com o aparecimento destas novas tecnologias de topo que irão permitir identificar variantes genéticas, até hoje não detectadas nas plataformas de genotipagem convencionais. Nesta doença, a curto prazo, seria desejável identificar métodos de diagnóstico (para detecção e intervenção precoce) e o desenvolvimento de terapias farmacológicas que possam melhorar a vida das crianças com autismo. Em última análise, este estudo foi mais um importante passo dado na compreensão das vias metabólicas que estão alteradas e envolvidas na etiologia da doença, assim como, um contributo para a detecção de novas regiões de susceptibilidade no autismo.

VII. REFERÊNCIAS BIBLIOGRÁFICAS

Abrahams B, Geschwind DH: Advances in autism genetics: on the threshold of a new neurobiology. *Nature Genetics Reviews* 2008; **9**:341-355.

Alonso P, Gratacòs M, Menchón JM, Saiz-Ruiz J, Segalàs C, Baca-García E, Labad J, Fernández-Piqueras J, Real E, Vaquero C, Pérez M, Dolengevich H, González JR, Bayés M, de Cid R, Vallejo J, Estivill X: Extensive genotyping of the BDNF and NTRK2 genes define protective haplotypes against obsessive-compulsive disorder. *Biological Psychiatry* 2008; Project, **63**(6): 619-28.

Anney R, Klei L, Pinto D, Regan R, Conroy J, Magalhaes T, Correia C, Abrahams B, Sykes N, Pagnamenta A *et al.*: A genome-wide scan for common alleles affecting risk for autism. *Human Molecular Genetics* 2010; **19**(20): 4072-4082.

Autism Genome Project Consortium, Szatmari P, Paterson AD *et al.*: Mapping autism risk loci using genetic linkage and chromosomal rearrangements. *Nature Genetics* 2007; **39**:319-28

Asperger H: Die autistischen psychopathen im kindesalter. *Archiv für Psychiatric und Nervenkrankheiten* 1944; **17**: 76-136

Barrett JC, Hansoul S, Nicolae DL, Cho JH, Duerr RH, Rioux JD, Brant SR, Silverberg MS, Taylor KD, Barmada MM, Bitton A, Dassopoulos T, Datta LW, Green T, Griffiths AM, Kistner EO, Murtha MT, Regueiro MD, Rotter JI, Schumm LP, Steinhardt AH, Targan SR, Xavier RJ; NIDDK IBD Genetics Consortium, Libioulle C, Sandor C, Lathrop M, Belaiche J, Dewit O, Gut I, Heath S, Laukens D, Mni M, Rutgeerts P, Van Gossum A, Zelenika D, Franchimont D, Hugot JP, de Vos M, Vermeire S, Louis E; Belgian-French IBD Consortium; Wellcome Trust Case Control Consortium, Cardon LR, Anderson CA, Drummond H, Nimmo E, Ahmad T, Prescott NJ, Onnie CM, Fisher SA, Marchini J, Ghori J, Bumpstead S, Gwilliam R, Tremelling M, Deloukas P, Mansfield J, Jewell D, Satsangi J, Mathew CG, Parkes M, Georges M, Daly MJ: Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nature Genetics* 2008; **40**: 955-962

Beglinger LJ, Smith TH: A review of subtyping in autism and proposed dimensional classification model. *Journal of Autism and Developmental Disorder* 2001; **31**(4): 411-22

Benjamini Y, Hochberg Y: Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society, Series B (Methodological)* 1995; **57** (1): 289-300.

Bettelheim B. *The empty fortress; Infantile autism and the birth of the self.* 1967. Free Pass Nova Yorque

Bertram L, Tanzi RE: Genome-wide association studies in Alzheimer's disease. *Human Molecular Genetics* 2009; **18**(R2): R137-145

Bibel M, Hoppe E, Barde YA: Biochemical and functional interactions between the neurotrophin receptors trk and p75NTR. *The EMBO Journal* 1999; **18**(3): 616-22

Bodmer W, Bonilla C: Common and rare variants in multifactorial susceptibility to common diseases. *Nature Genetics* 2008; **40**(6): 695-701.

Borecki IB, Suarez BK. Linkage and association: basic concepts. Chp 5: 45-63 in *Genetic Dissection of Complex Traits*. Rao DC, Province MA. 2001. Academic Press

Bourgeron T: A synaptic trek to autism. *Current opinion in Neurobiology* 2009; **19**:231-234

Cantor RM, Lange K, Sinsheimer JS: Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *American Journal of Human Genetic* 2010; **86**(1): 6-22

Castermans D, Wilquet V, Steyaert J, Van de Ven W, Fryns JP, Devriendt K: Chromosomal anomalies in individuals with autism: a strategy towards the identification of genes involved in autism. *Autism* 2004; **2**:141-61.

Chapman NH and Thompson EA. Linkage Disequilibrium Mapping: the role of population history, size and structure. Chp 25: 413-435 in *Genetic Dissection of Complex Traits*. Rao DC and Province MA. 2001. Academic Press

Chen Z, Simmons MS, Perry RT, Wiener HW, Harrell LE, Go RC: Genetic association of neurotrophic tyrosine kinase receptor type 2 (NTRK2) With Alzheimer's disease. *American Journal of Medical Genetics Part B Neuropsychiatry Genetic* 2008; **147**(3): 363-9.

Cheng L, Ge Q, Xiao P, Sun B, Ke X, Bai Y, Lu Z: Association study between BDNF gene polymorphisms and autism by three-dimensional gel-based microarray. *International Journal of Molecular Science* 2009; **10**(6):2487-500.

Conciatori M, Stodgell CJ, Human SL, O'Bara M, Militerni R, Bravaccio C, Trillo S, Montecchi F, Schneider C, Melmed R, Elia M, Crawford L, Spence SJ, Muscarella L, Guarneri V, D'Agruma L, Quattrone A, Zelante I, rabinowitz D, Pascucci T, Puglisi-Allegra S, Reichelt K, Rodier PM, Persico AM: Association between the *HOXA1* A218G polymorphism and increase head circumference in patients with autism. *Biological Psychiatry* 2004; 55:413-419.

Conner AC, Kissling C, Hodges E, Hünnerkopf R, Clement RM, Dudley E, Freitag CM, Rösler M, Retz W, Thome J: Neurotrophic factor-related gene polymorphisms and adult attention deficit hyperactivity disorder (ADHD) score in a high-risk male population. *American Journal of Medical Genetics part B Neuropsychiatry Genetic* 2008;**147B**(8):1476-80.

Cordell HJ: Detecting gene-gene interactions that underlie human diseases. *Nature Genetic Reviews* 2009; **10**:392-404.

Correia C, Coutinho AM, Sousa I, Venda LL, Almeida J, Lontro R, Lobo C, Miguel T, Conroy J, Cochrane L, Gallagher L, Gill M, Ennis S, Oliveira G, Vicente AM: Increased BDNF levels and NTRK2 gene association suggest a disruption of BDNF/TrkB signalling in autism. *Genes, Brain and Behavior* 2010 (*in press*) doi: **10.1111/j.1601-183X.2010.00627.x**

Coutinho AM, Oliveira G, Morgadinho T, Fesel C, Macedo TR, Bento C, Marques C, Ataíde A, Miguel T, Borges L, Vicente AM: Variants of the serotonin transporter gene (*SLC6A4*) significantly contribute to hyperserotonemia in autism. *Molecular Psychiatry* 2004; **9**(3):264-71

Coutinho AM, Sousa I, Martins M, Correia C, Morgadinho T, Bento C, Marques C, Ataíde A, Miguel TS, Moore JH, Oliveira G, Vicente AM: Evidence for epistasis between *SLC6A4* and *ITGB3* in autism etiology and in the determination of platelet serotonin levels. *Human Genetics* 2007; **121**(2):243-56

Dekker MC, van Duijn CM: Prospects of genetic epidemiology in the 21st century. *European Journal of Epidemiology* 2003; **18**(7): 607-616

- Dermitzakis ET, Clark AG: Genetics. Life after GWA studies. *Science* 2009; **326**(5950): 239-40
- Devlin B, Bennett P, Cook Jr. EH, Dawson G, Gonen D, Grigorenko EL, McMahon W, Pauls D, Smith M, Spence MA, CPEA Genetics Network, Schellenberg GD: No evidence for linkage of liability to autism to *HOXA1* in a sample from CPEA Network. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 2002; **114**: 667-672.
- Dmitrzak-Weglarz M, Rybakowski JK, Suwalska A, Skibinska M, Leszczynska-Rodziewicz A, Szczepankiewicz A, Hauser J: Association studies of the *BDNF* and the *NTRK2* gene polymorphisms with prophylactic lithium response in bipolar patients. *Pharmacogenomics*. 2008; **9**(11):1595-603.
- Eisenberg L: The fathers of autistic children. *American Journal Orthopsychiatry* 1957; **127**:715-724
- Franco SJ, Huttenlocher A: Regulation cell migration: calpains make the cut. *Journal of Cell Science* 2005; **118**:3829-3838
- Frazer KA, Murray SS, Schork NJ, Topo EJ: Human genetic variation and its contribution to complex traits. *Nature Reviews Genetics* 2009; **10**, 241-251
- Folstein SE, Rosen-Sheidley B: Genetics of autism: complex aetiology for a heterogeneous disorder. *Nature Genetics Reviews* 2001; **2**:943-955
- Fombonne E: Epidemiology of pervasive developmental disorders. *Pediatric Research* 2009; **65**(6): 591-598
- Gallagher L, Hawi Z, Kearney G, Fitzgerald M and Gill M: No association between allelic variants of *HOXA1/HOXB1* and autism. *American Journal of Medical Genetics Part B (Neuropsychiatric Genetics)* 2004; **124B**:64-67
- Glading A, Bodnar RJ, Reynolds IJ, Shiraha H, Satish L, Potter DA, Blair HC, Wells A: Epidermal growth factor activates m-calpain (calpain II), at least in part, by extracellular signal-regulated kinase-mediated phosphorylation. *Molecular Cell Biology* 2004; **24**(6):2499-512
- Gillberg C: Chromosomal disorders and autism. *Journal of Autism and Developmental Disorders* 1998; **28**(5):415-25.
- Gokce O, Runne H, Kuhn A, Luthi-Carter R: Short-term striatal gene expression responses to brain-derived neurotrophic factor are dependent on MEK and ERK activation. *PLoS One* 2009; **4**(4):e5292
- Gordon D, Finch SJ: Factors affecting statistical power in the detection of genetic association. *The Journal of Clinical Investigation* 2005; **115**(6): 1408-1418
- Griffiths R. 1984. *The abilities of young children*. London. University of London Press.
- Haubenberger D, Reinthaler E, Mueller JC, Pirker W, Katzenschlager R, Froehlich R, Bruecke T, Daniel G, Auff E, Zimprich A: Association of transcription factor polymorphisms *PITX3* and *EN1* with Parkinson's disease. *Neurobiology of Aging* 2009
- Haapasalo A, Sipola I, Larsson K, Akerman KEO, Stoilov P, Stamm S, Wong G, Castrén E: Regulation of TrkB surface expression by brain-derived neurotrophic factor and truncated TRKb isoforms. *The Journal of Biological Chemistry* 2002; **277**(45): 43160-43167.

Hirschhorn JN, Daly MJ: Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics* 2005; **6**(2): 95-108

Ingram JL, Stodgell CJ, Hyman SL, Figlewicz DA, Weitkamp LR and Rodier PM: Discovery of allelic variants of HOXA1 and HOXB1: Genetic susceptibility to autism spectrum disorders. *Teratology* 2000; **62**:393-405

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C: STRING 8--a global view on proteins and their functional interactions in 630 organisms. *Nucleic Acids Research* 2009; **37**(Database issue):D412-6.

Kanner L: Autistic disturbances of affective contact. *Nervous Child* 1943; **2**: 217-250.

Kelleher RJ, Bear MF: The Autistic Neuron: troubled translation? *Cell* 2008; **135**:401-406

Klauck SM: Genetics of autism spectrum disorder. *European Journal of Human Genetics* 2006; **14**:714-720.

Kohli MA, Salyakina D, Pfennig A, Lucae S, Horstmann S, Menke A, Kloiber S, Hennings J, Bradley BB, Ressler KJ, Uhr M, Müller-Myhsok B, Holsboer F, Binder EB: Association of Genetic Variants in the Neurotrophic Receptor-Encoding Gene NTRK2 and a Lifetime History of Suicide Attempts in Depressed Patients. *Archives of General Psychiatry* 2010

Kruglyak L, Nickerson DA: Variation is the spice of life. *Nature Genetics* 2001; **27**:234-236.

Kurup RK, Kurup PA: A Hypothalamic digoxin-mediated model for autism. *Internacional Journal of Neuroscience* 2003; **113**(11):1537-1539.

Lahiri DK, Nurnberger Jr. JI: Non enzymatic extraction of leucocyte DNA directly from blood. *Nucleic Acids Research* 1991; **19**(19):5444.

Leslie NR, Downes CP: PTEN: The down side of PI 3-kinase signalling. *Cell Signaling* 2002; **14**(4): 285-95

Lauritsen MB, Als TD, Dahl HA, Flint TJ, Wang AG, Vang M, Kruse TA, Ewald H, Mors O: A genome-wide search for alleles and haplotypes associated with autism and related pervasive developmental disorders on the Faroe Islands. *Molecular Psychiatry* 2006; **11**(1):37-46.

Levitt P, Campbell DB: The genetic and neurobiologic compass points toward common signaling dysfunctions in ASD. *Journal of Clinical Investigation* 2009; **119**(4): 747-754.

Li J, Tabor HK, Nguyen L, Gleason C, Lotspeich LJ, Spiker D, Risch N, Myers RM: Lack of association between HOXA1 and HOXB1 gene variants and autism in 110 multiplex families. *American Journal of Medical Genetics (Neuropsychiatric Genetics)* 2002; **114**:24-30.

Li B, Leal SM: Discovery of rare variants via sequencing: implications for the design of complex trait association studies. *PLoS Genetic* 2009; **5**(5)

Liu X, Novosedlik N, Wang A, Hudson ML, Cohen IL, Chudley AE, Forster-Gibson CJ, Lewis SME, Holden JJA: The DLX1 and DLX2 genes and susceptibility to autism spectrum disorders. *European Journal of Human Genetics* 2009; **17**:228-235.

Lombard J: Autism: a mitochondrial disorder? *Medical Hypotheses* 1998; **50**(6):497-500.

Lord C, Rutter M, Le Couteur A: Autism Diagnostic Interview-revised: a revised version of a diagnostic interview for caregivers of individuals with possible pervasive development disorders. *Journal Autism Development Disorders* 1994; **24**(5):659-685.

Lord C, Cook EH, Leventhal BL, Amaral DG: Autism spectrum disorders. *Neuron* 2000; **28**(2): 355-63.

Lou XY, Chen GB, Yan L, Ma JZ, Mangold JE, Zhu J, Elston RC, Li MD: A generalized combinatorial approach for detecting gene-by-gene and gene-by- environment interactions with application to nicotine dependence. *The American Journal of Human Genetics* 2007; **80**:1125-1137.

Lou XY, Chen GB, Yan L, Ma JZ, Mangold JE, Zhu J, Elston RC, Li MD: A combinatorial approach to detecting gene-gene and gene-environment interactions in family studies. *The American Journal of Human Genetics* 2008; **83**:457-467.

Luberg K, Wong J, Weickert CS, Timmusk T: Human TrkB gene: novel alternative transcripts, protein isoforms and expression pattern in the prefrontal cerebral cortex during postnatal development. *Journal of Neurochemistry* 2010

Ma D, Salyakina D, Jaworski JM, Konidari I, Whitehead PL, Andersen AN, Hoffman JD, Slifer SH, Hedges DJ, Cukier HN, Griswold AJ, McCauley JL, Beecham GW, Wright HH, Abramson RK, Martin ER, Hussman JP, Gilbert JR, Cuccaro ML, Haines JL, Pericak-Vance MA: A genome-wide association study of autism reveals a common novel risk locus at 5p14.1. *Annals of Human Genetics* 2009; **73**(Pt 3):263-73

Maher B: The case of the missing heritability. *Nature, News Feature Personal genomes* 2008; **456**: 18-21

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttmacher AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM: Finding the missing heritability of complex diseases. *Nature* 2009; **461**(7265): 747-53.

Marques JHF. *Manual da escala de inteligência e aferição para Portugal*. 1970. Lisboa: Centro de Estudos de Psicologia, Faculdade de Letras de Lisboa

McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, Hirschhorn JN: Genome-wide association studies for complex traits: consensus, uncertainty and challenges *Nature Reviews Genetics* 2008; **9**: 356-369

Moore JH: The Ubiquitous nature of epistasis in determining susceptibility to common human diseases. *Human Heredity* 2003; **56**:73-82

Moskvina V, Craddock N, Holmans P, Nikolov I, Pahwa JS, Green E; Wellcome Trust Case Control Consortium, Owen MJ, O'Donovan MC: Gene-wide analyses of genome-wide association data sets: evidence for multiple common risk alleles for schizophrenia and bipolar disorder and for overlap in genetic risk. *Molecular Psychiatry* 2009; **14**(3):252-60

Muhle R, Trentacoste SV, Rapin I: The genetics of autism. *Pediatrics* 2004; **113**(5):472-486.

Nelson KB, Grether JK, Croen LA, Dambrosia JM, Dickens BF, Jelliffe LL, Hansen RL, Phillips TM: Neuropeptides and neurotrophins in neonatal blood of children with autism or mental retardation. *Annals of Neurology* 2001; **49**(5):597-606.

Nishimura K, Nakamura K, Anitha A, Yamada K, Tsujii M, Iwayama Y, Hattori E, Toyota T, Takei N, Miyachi T, Iwata Y, Suzuki K, Matsuzaki H, Kawai M, Sekine Y, Tsuchiya K, Sugihara G, Suda S, Ouchi Y, Sugiyama T, Yoshikawa T, Mori N: Genetic analyses of the brain-derived neurotrophic factor (BDNF) gene in autism. *Biochemical and Biophysical Research Communications* 2007; **356**(1): 200-6

Nordborg M and Tavaré S: Linkage disequilibrium: what history has to tell us. *Trends in Genetics* 2002; **18**(2): 83-90

Oliveira G, Diogo L, Grazina M, Garcia P, Ataíde A, Marques C, Miguel T, Borges L, Vicente AM, Oliveira CR: Mitochondrial dysfunction in autism spectrum disorders: a population-based study. *Developmental Medicine and Child Neurology* 2005; **47**(3): 185-9

Oliveira G, Ataíde A, Marques C, Miguel T, Coutinho AM, Mota-Vieira L, Gonçalves E, Lopes NM, Rodrigues V, Mota HC, Vicente AM: Epidemiology of autism spectrum disorder in Portugal: prevalence, clinical, characterization and medical conditions. *Developmental medicine and child Neurology* 2007; **49**: 726-733

O'Donovan MC, Craddock NJ, Owen MJ: Genetics of psychosis; insights from views across the genome. *Human Genetics* 2009; **126**(1):3-12

Ott J. *Analysis of Human Genetics Linkage*. 3rd Edition (1999). John Hopkins University Press. Baltimore, London.

Pardo CA, Eberhart CG: The neurobiology of autism. *Brain Pathology* 2007; **17**(4):434-447

Park SG, Schimmel P, Kim S: Aminoacyl tRNA synthetases and their connections to disease. *PNAS* 2008; **105**(32): 11043-11049.

Philippi A, Tores F, Carayol J, Rousseau F, Letexier M, Roschmann E, Lindenbaum P, Benajjou A, Fontaine K, Vazart C, Gesnoux P, Brooks P, Hager J: Association of autism with polymorphisms in the paired-like homeodomain transcription factor 1 (PITX1) on chromosome 5q31: a candidate gene analysis. *BMC Medical Genetics* 2007; **8**:74.

Phillips PC: Epistasis – the essential role of gene interactions in the structure and evolution of genetic systems. *Nature Genetics Review* 2008; **9**: 855-867.

Pinto D, Pagnamenta A, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhães T, Correia C, Abrahams B *et al.* : Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* 2010; **466**(7304): 368-372.

Pritchard JK, Cox NJ: The allelic architecture of human disease genes: common disease-common variant...or not? *Human Molecular Biology* 2002; **11**:2417-2423

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ, Sham PC: PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics* 2007; **81**

Rapin I: Autism. *The New England Journal of Medicine* 1997; **337**:97-104

Reichardt LF: Neurotrophin-regulated signalling pathways. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2006; **361**(1473): 1545-64

Ritchie MD, Hahn LW, Roodi N, Bailey LR, Dupont WD, Parl FF, Moore JH: Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *American Journal of Human Genetics* 2001; **69**(1):138-47

Robinson R: Common Disease, Multiple Rare (and Distant) Variants. *PLoS Biology* 2010; **8**(1): e1000293. doi:10.1371/journal.pbio.1000293

Rodger EE, Theibert AB: Function of PI3-Kinase in development of the nervous system. *International Journal Development Neuroscience* 2002; **20**: 187-197.

Rodriguez-Viciano P, Rauen KA: Biochemical characterization of novel germline BRAF and MEK mutations in cardio-facio-cutaneous syndrome. *Methods Enzymol.* 2008; **438**: 277-89.

Romano V, Calí F, Mirisola M, Gambino G, D'Anna R, Di Rosa P, Seidita G, Chiavetta V, Aiello F, Canziani F, De Leo G, Ayala GF, Elia M: Lack of association of HOXA1 and HOXB1 mutations and autism in Sicilian (Italian) patients. *Scientific correspondence Molecular Psychiatry* 2003; **8**: 716-720.

Rutter M: Autism research: lessons from the past and prospects for the future. *Journal of Autism and Developmental Disorders* 2005; **35**:241-57

Rutter M: Diagnosis and definition of childhood autism. *Journal of Autism and Developmental Disorders* 1978; **8**(2):139-161.

Scherer SW, Cheung J, MacDonald JR, Osborne LR, Nakabayashi K, Herbrick JA, Carson AR, Parker-Katirae L, Skaug J, Khaja R, Zhang J, Hudek AK, Li M, Haddad M, Duggan GE, Fernandez BA, Kanematsu E, Gentles S, Christopoulos CC, Choufani S, Kwasnicka D, Zheng XH, Lai Z, Nusskern D, Zhang Q, Gu Z, Lu F, Zeesman S, Nowaczyk MJ, Teshima I, Chitayat D, Shuman C, Weksberg R, Zackai EH, Grebe TA, Cox SR, Kirkpatrick SJ, Rahman N, Friedman JM, Heng HH, Pelicci PG, Lo-Coco F, Belloni E, Shaffer LG, Pober B, Morton CC, Gusella JF, Bruns GA, Korf BR, Quade BJ, Ligon AH, Ferguson H, Higgins AW, Leach NT, Herrick SR, Lemyre E, Farra CG, Kim HG, Summers AM, Gripp KW, Roberts W, Szatmari P, Winsor EJ, Grzeschik KH, Teebi A, Minassian BA, Kere J, Armengol L, Pujana MA, Estivill X, Wilson MD, Koop BF, Tosi S, Moore GE, Boright AP, Zlotorynski E, Kerem B, Kroisel PM, Petek E, Oscier DG, Mould SJ, Döhner H, Döhner K, Rommens JM, Vincent JB, Venter JC, Li PW, Mural RJ, Adams MD, Tsui LC: Human chromosome 7: DNA sequence and biology. *Science* 2003; **300**(5620): 767-72

Schork NJ, Fallin D, Thiel B, Xu X, Broeckel U, Jacob HJ, Cohen D. The Future of Genetic Case-Control Studies. Chp 14: 191-195 in *Genetic Dissection of Complex Traits*. Rao DC and Province MA. 2001. Academic Press

Seng KC, Seng CK: The success of the genome-wide association approach: a brief story of a long struggle. *European Journal of Human Genetics* 2008; **16**: 554-564

Serajee FJ, Nabi R, Zhong H, Mahbubul Huq AH: Association of INPP1, PIK3CG, and TSC2 gene variants with autistic disorder: implications for phosphatidylinositol signalling in autism. *Journal of Medical Genetic* 2003; **40**(11):e119.

Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *American Journal of Human Genetics* 1993; **52**(3): 506-16

Smith EN, Bloss CS, Badner JA, Barrett T, Belmonte PL, Berrettini W, Byerley W, Coryell W, Craig D, Edenberg HJ, Eskin E, Foroud T, Gershon E, Greenwood TA, Hipolito M, Koller DL, Lawson WB, Liu C, Lohoff F, McInnis MG, McMahon FJ, Mirel DB, Murray SS, Nievergelt C, Nurnberger J, Nwulia EA, Paschall J, Potash JB, Rice J, Schulze TG, Scheftner W, Panganiban C, Zaitlen N, Zandi PP, Zöllner S, Schork NJ, Kelsoe JR: Genome-wide association study of bipolar disorder in European American and African American individuals. *Molecular Psychiatry* 2009; **14**(8): 755-763

Stoilov P, Castren E, Stamm S: Analysis of the human TrkB gene genomic organization reveals novel TrkB isoforms, unusual gene length and splicing mechanism. *Biochemical and Biophysical Research Communications* 2002; **290**(3):1054-65.

Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *PNAS* 2005; **102**(43):15545-50

The International HapMap Consortium: The international HapMap project. *Nature* 2003; **426**:789-796.

Thomson G. Significance Levels in Genome Scans. Chp 28 in *Genetic Dissection of Complex Traits*. Rao DC and Province MA. 2001. Academic Press

Tsuji T, Shimohama S, Kimura J, Shimizu K: m-Calpain (calcium-activated neutral proteinase) in Alzheimer's disease brains. *Neuroscience Letters* 1998; **248**(2):109-12

Wang K, Li M, Bucan M: Pathway-based approaches for analysis of Genome-wide association studies. *The American Journal of Human Genetics* 2007; **81**: 1278-1283.

Wang K, Zhang H, Ma D, Bucan M, Glessner JT, Abrahams BS, Salyakina D, Imielinski M, Bradfield JP, Sleiman PM, Kim CE, Hou C, Frackelton E, Chiavacci R, Takahashi N, Sakurai T, Rappaport E, Lajonchere CM, Munson J, Estes A, Korvatska O, Piven J, Sonnenblick LI, Alvarez Retuerto AI, Herman EI, Dong H, Hutman T, Sigman M, Ozonoff S, Klin A, Owley T, Sweeney JA, Brune CW, Cantor RM, Bernier R, Gilbert JR, Cuccaro ML, McMahon WM, Miller J, State MW, Wassink TH, Coon H, Levy SE, Schultz RT, Nurnberger JI, Haines JL, Sutcliffe JS, Cook EH, Minshew NJ, Buxbaum JD, Dawson G, Grant SF, Geschwind DH, Pericak-Vance MA, Schellenberg GD, Hakonarson H: Common genetic variants on 5p14.1 associate with autism spectrum disorders. *Nature* 2009; **459**(7246): 528-33

Weiss LA, Ober C, Cook Jr EH: *ITGB3* shows genetic and expression interaction with *SLC6A4*. *Human Genetics* 2006; **120**(1):93-100

Weiss LA, Arking DE, The Gene Discovery Project of Johns Hopkins and The Autism Consortium: A Genome-wide linkage and association scan reveals novel *loci* for autism. *Letter Nature* 2009; **461**:802-808.

Willemsen MH, Fernandez BA, Bacino CA, Gerkes E, de Brouwer AP, Pfundt R, Sikkema-Raddatz B, Scherer SW, Marshall CR, Potocki L, van Bokhoven H, Kleefstra T: Identification of *ANKRD11* and *ZNF778* as candidate genes for autism and variable cognitive impairment in the novel 16q24.3 microdeletion syndrome. *European Journal of Human Genetics* 2009; **18**(4):429-35

Wing L: The autistic spectrum. *Lancet* 1997; **350**:1761-6

Yamada K and Nabeshima T: Brain-derived Neurotrophic Factor/TrkB signaling in memory process. *Journal of Pharmacological Sciences* 2003; **91**(4):267-70

Yang MS, Gill M: A review of gene linkage, association and expression studies in autism and an assessment of convergent evidence. *International Journal of Developmental Neuroscience* 2007; **25**(2): 69-85

Yoshii A and Constantine-Paton M: Postsynaptic BDNF-TrkB signaling in synapse maturation, plasticity, and disease. *Development Neurobiology* 2010

Zadran S, Jourdi H, Rostamiani K, Qin Q, Bi X and Baudry M: Brain-derived neurotrophic factor and epidermal growth factor activate neuronal m-Calpain via Mitogen-activated Protein Kinase-Dependent Phosphorylation. *The Journal of Neuroscience* 2010; **30**(3): 1086-1095

Zeggini E *et al.*: Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type2 diabetes. *Nature Genetics* 2008; **40**: 638-645.

Cibergrafia

[1] www.genome.gov/gwastudies

[2] www.autismgenome.org

[3] www.1000genomes.org

[4] http://string-db.org/newstring.cgi/show_input_page.pl?UserId=bRmq_LzLhned&sessionId=S7K2gvvD7iYV

ANEXOS

Anexo A. O gene *NTRK2*

A região do promotor deste gene (inclui os exões 1-4) é rica em GC, não tem elementos de iniciação de transcrição (como por exemplo, *TATA box*) e tem nove locais de iniciação de transcrição diferentes (Stoilov *et al.*, 2002).

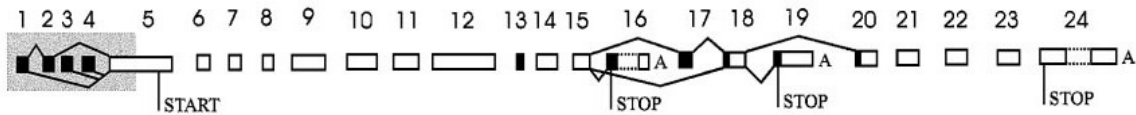


Figura 17. Padrão de *splicing* alternativo do gene *NTRK2*. Os exões constitutivos estão representados por retângulos brancos e os exões alternativos por retângulos pretos. Os padrões de *splicing* alternativo são indicados pelas linhas pretas (retirado de Stoilov *et al.*, 2002).

A complexidade da região promotora deve-se ao mecanismo de *splicing* alternativo do gene que permite produzir 100 isoformas de mRNA que codificam 10 proteínas diferentes e, mais recentemente, estão descritas na literatura 36 novas proteínas (Luberg *et al.*, 2010). O elevado número de isoformas sugere uma elevada complexidade na síntese, regulação e função do gene. Três destas isoformas são mais frequentes: a forma normal (*full length* TrkB) e duas formas truncadas (TrkB-T1 e TrkB-T-Shc). As formas truncadas não têm o domínio intracelular (tirosina cinase) não permitindo a sua ligação ao citoplasma, que por conseguinte, não permite a sinalização de outras vias biológicas (Stoilov *et al.*, 2002).

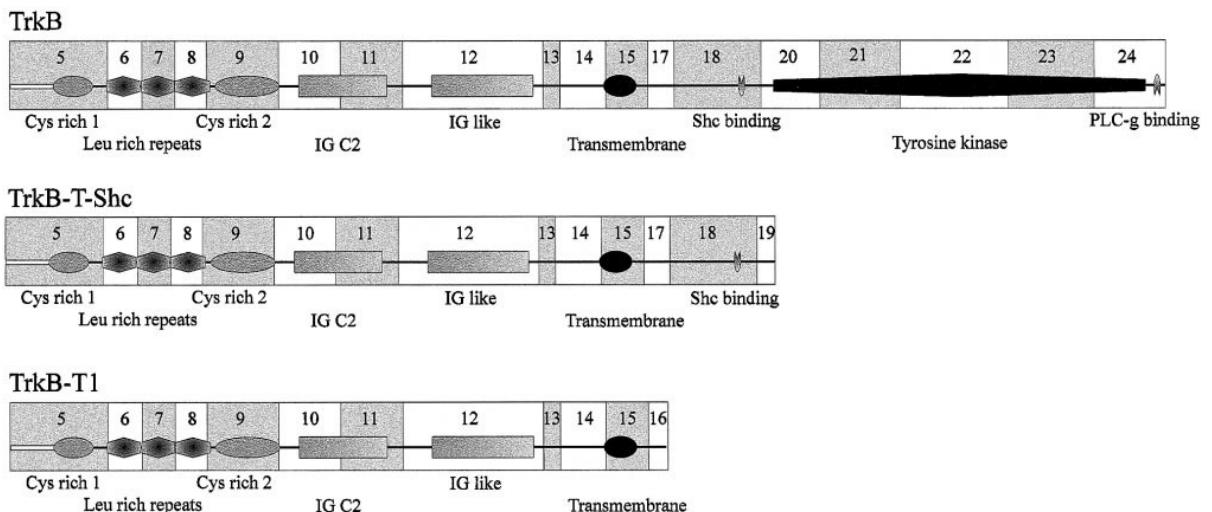


Figura 18. Estrutura de três proteínas: TRkB, TrkB-T1 e TRkB-T-Shc. Os números representam os exões. Os domínios estão indicados por baixo dos quadrados (retirado de Stoilov *et al.*, 2002).

Anexo B. Via biológica das neurotrofinas (adaptado da base de dados KEGG)

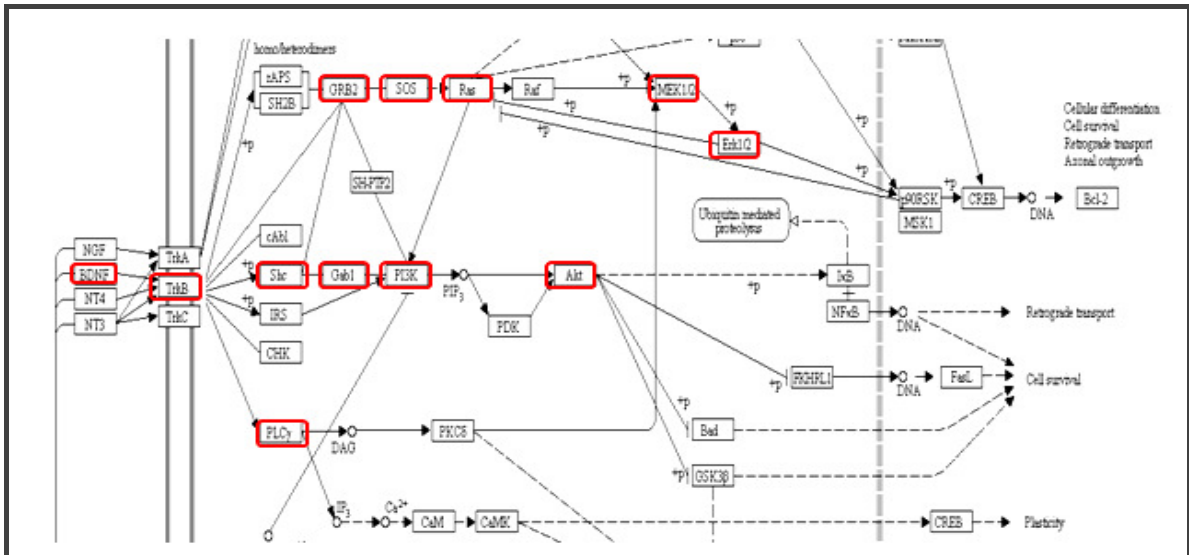


Figura 19. Esquema representativo da via biológica das neurotrofinas onde estão representadas as vias biológicas induzidas pela ligação BDNF/TrkB (assinalados a vermelho). (<http://www.genome.jp/kegg/pathway/hsa/hsa04722.html>)

Anexo C. Protocolo *Salting out* para extracção de DNA (adaptado de Lahiri and Nurnberger, 1991)

O sangue é colhido para tubos Vacutainer de 3ml, contendo EDTA a 7,5% na solução, a fim de evitar a coagulação sanguínea. O volume de sangue é transferido para um tubo *falcon* de 15ml, adicionando-se o mesmo volume de TKM X-100 e 25µl por ml de sangue de IGEPAL (Sigma) para lisar as células. Agitam-se os tubos por inversão quatro a cinco vezes, para que o IGEPAL se dissolva completamente. De seguida, centrifuga-se durante 15 minutos a 2200 rpm, à temperatura ambiente, na centrífuga *Eppendorf* 5810R, com o travão desligado, para que não haja a destruição do sedimento formado. Remove-se o sobrenadante, retendo unicamente o pequeno sedimento formado e adiciona-se 1ml de TKM1 por ml de sangue inicial, que irá inactivar as nucleases e outras enzimas sem destruir o DNA e liberta a hemoglobina que fica em suspensão, dando a cor avermelhada ao sedimento. Agita-se bem para dissolver o sedimento e centrifuga-se a 1600rpm durante 10 minutos, à temperatura ambiente. Elimina-se o sobrenadante e adiciona-se novamente TKM1 ao sedimento, repetindo a centrifugação. Novamente remove-se o sobrenadante e ressuspende-se o sedimento em 160µl de TKM2 (que separa as proteínas e as membranas) por ml de sangue inicial e adiciona-se 10µl (por volume inicial de sangue) de SDS a 10%, a fim de completar a desnaturação das proteínas e membranas, separando o DNA dos restantes resíduos celulares. Deve-se agitar vigorosamente de modo a desfazer o sedimento e incuba-se então a 55°C, durante 10 minutos. De seguida, transfere-se todo o volume para um microtubo de 2ml, contendo 60µl de NaCl 5M por ml de sangue inicial, para precipitar as proteínas. Mistura-se por inversão do tubo e centrifuga-se a 13200rpm durante 20 minutos, na microcentrifuga *Eppendorf* 5415D, à temperatura ambiente. Recolhe-se o sobrenadante que contém o DNA, com cuidado para não o contaminar com o sedimento, para um tubo falcon de 15ml e adiciona-se 460µl de etanol absoluto (100%) por ml de sangue inicial. Inverte-se o tubo várias vezes até o DNA estar completamente precipitado. Retira-se o DNA com o auxílio de uma ansa descartável e lava-se em etanol a 70% a 4°C. Aguarda-se até o DNA estar seco, ou seja, ficar transparente e ressuspende-se em 150µl de tampão TE, permanecendo à temperatura ambiente e no escuro durante 24 horas.

REAGENTES E SOLUÇÕES

TKM X-100

5 ml	Tris-HCl 1M pH7,5
5 ml	KCl 1M
5 ml	MgCl ₂ 1M
2 ml	EDTA 0,5M
12,5 ml	Triton X-100

Perfazer com água até 500ml

TKM1

5 ml	Tris-HCl 1M pH7,5
5 ml	KCl 1M
5 ml	MgCl ₂ 1M
2 ml	EDTA 0,5M

Perfazer com água até 500ml

TKM2

1 ml	Tris-HCl 1M
1 ml	KCl 1M
1 ml	MgCl ₂ 1M
400 µl	EDTA 0,5M
8 ml	NaCl 5M

Perfazer com água até 100ml

TE

1 ml	Tris-HCl 1M pH7,5
200 µl	EDTA 1M pH8,0

Perfazer com água até 100ml

Anexo D. Diagrama do *cluster* das diferentes populações participantes no GWAS

O seguinte diagrama de *cluster* representa o resultado da análise de PCA (*Principal Component Analysis*) das diferentes populações participantes no GWAS. As diferentes populações têm diferente ancestralidade genética, com especial interesse para a mancha correspondente à subpopulação portuguesa (representada pela cor azul-escuro) destaca-se das restantes subpopulações. A diferente ancestralidade das populações pode indicar que os marcadores dos genes associados ao autismo podem diferir entre as populações (Tiago Magalhães, comunicação pessoal).

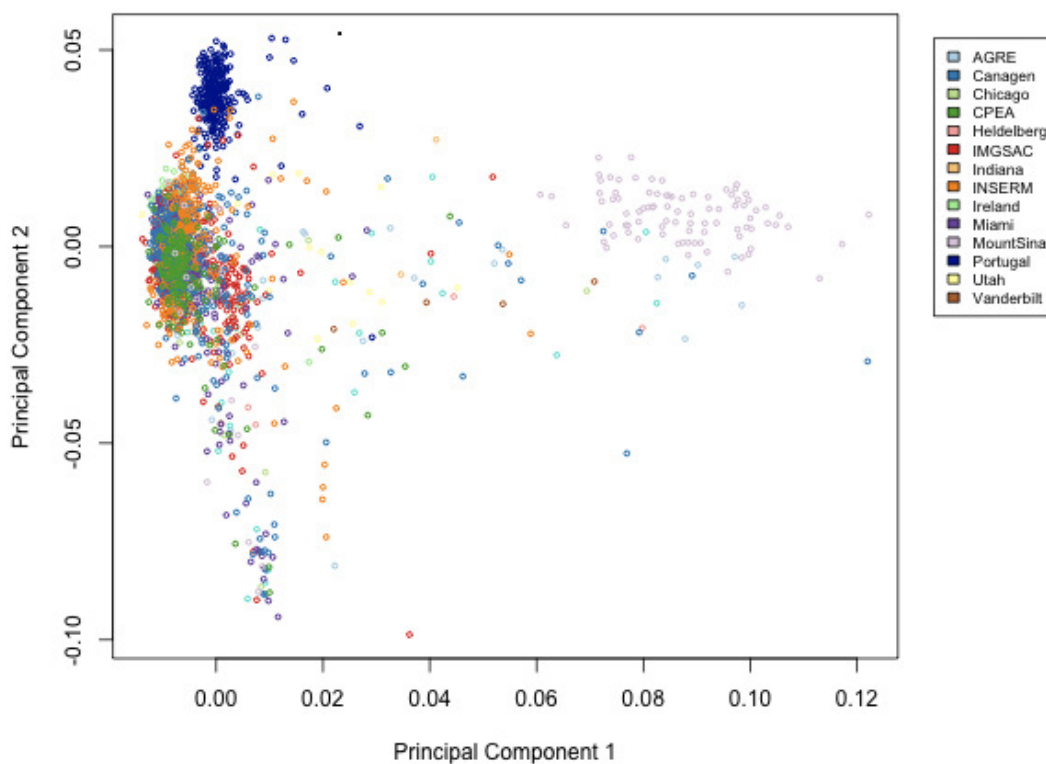


Figura 20. Diagrama de *clustering* dos resultados de *Principal Component Analysis* para as diferentes populações que constituem a amostra AGP. A azul-escuro está representado o sub-grupo de indivíduos portugueses que tende a agrupar-se separadamente na maior parte das restantes amostras e que apresenta uma dispersão em termos de ancestralidade genética muito limitada. As populações que constituem a amostra AGP estão representadas por diferentes cores, como indicado na caixa ao lado do diagrama (Tiago Magalhães, comunicação pessoal).

Anexo E. STRING ^[4]: selecção de genes para análise

O STRING é uma base de dados de interacções proteicas descritas e preditivas que contemplam associações directas (ou físicas) ou indirectas (funcionais) das proteínas. A informação provém de quatro tipos de fontes: contexto genético, experiências laboratoriais, estudos de expressão genica e literatura científica. A base de dados engloba a informação referente a 2590259 proteínas de 630 organismos diferentes.

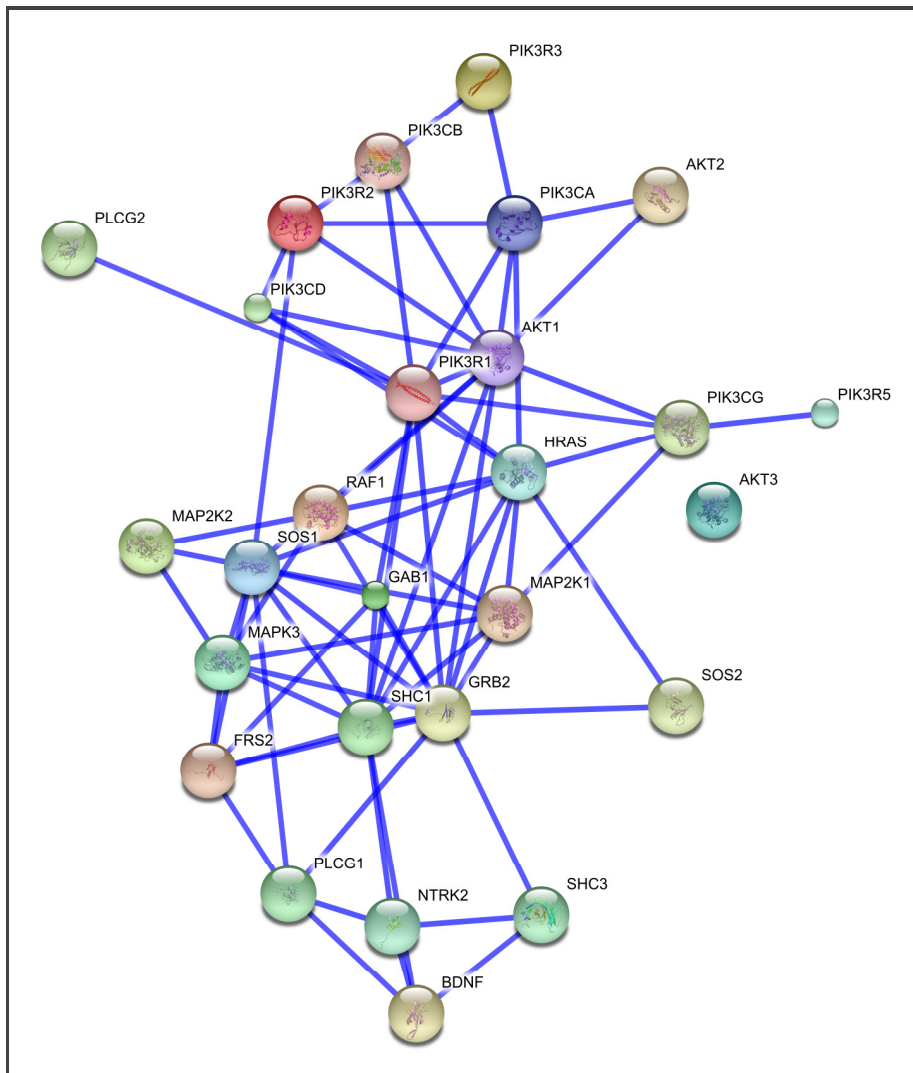


Figura 21. Resultados da selecção de genes através do programa STRING para a lista de 27 genes (tabela 9) com valores de confiança superiores a 95%. Foi excluído o gene *AKT3* por apresentar valores de confiança inferior a 95%.

Anexo F. Estudos de Interação genética

O programa GMDR, reduz a dimensão dos dados e agrupa os genótipos segundo alto (células a cinzento-escuro) ou baixo risco (células a cinzento claro) na susceptibilidade à doença. As seguintes figuras (figuras 22-27) são a representação gráfica dos modelos de interação apresentados na tabela anterior com as combinações de genótipos 2 a 2 e 3 a 3 (esquerda e direita, respectivamente). Em cada célula multifactorial, a barra da esquerda representa um *score* para casos e a barra direita um *score* para controlos, resultado do cálculo do erro de predição do teste estatístico para casos e controlos. Como não existem modelos de interação significativos, não existem combinações de genótipos de risco associados à doença.

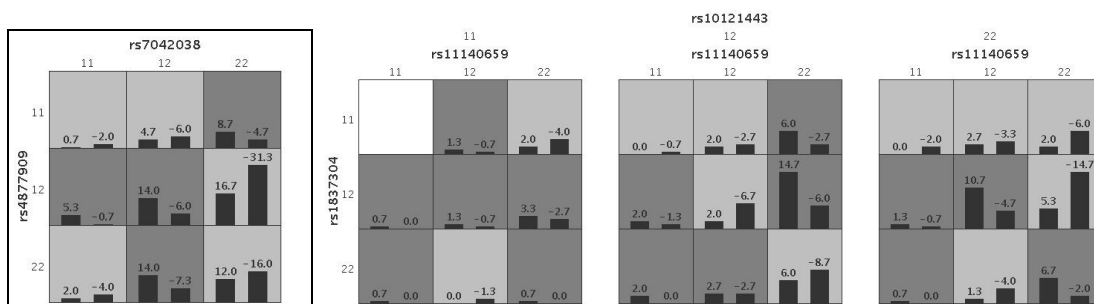


Figura 22. Modelo gráfico multifactorial de *NTRK2* e *DLX1*.

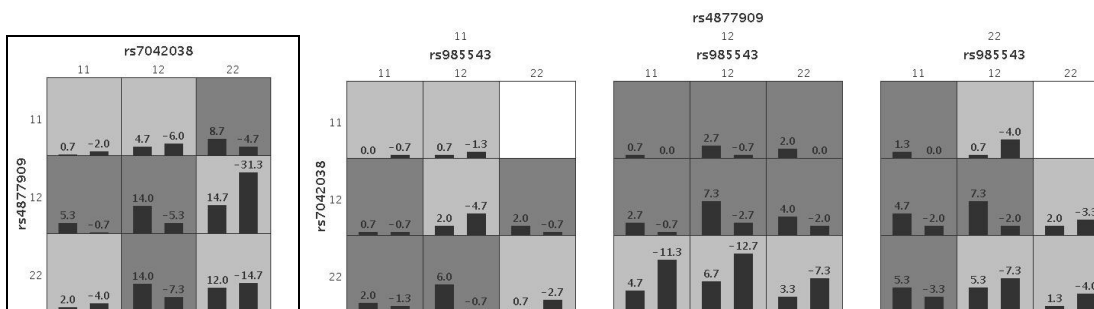


Figura 23. Modelo gráfico multifactorial de *NTRK2* e *PITX1*.

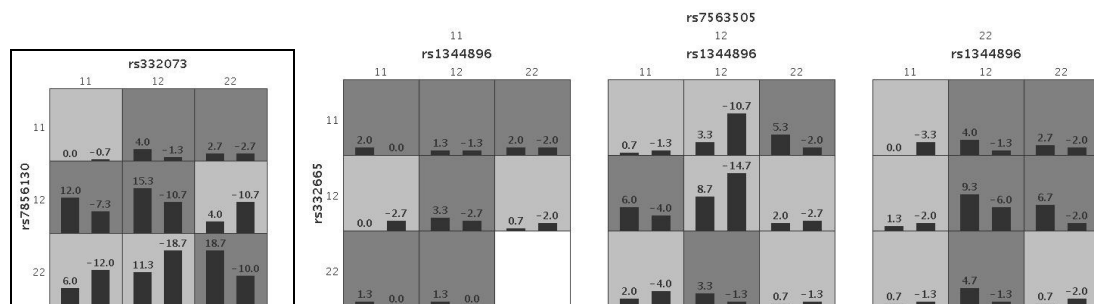


Figura 24. Modelo gráfico multifactorial de *NTRK2* e *EN1*.

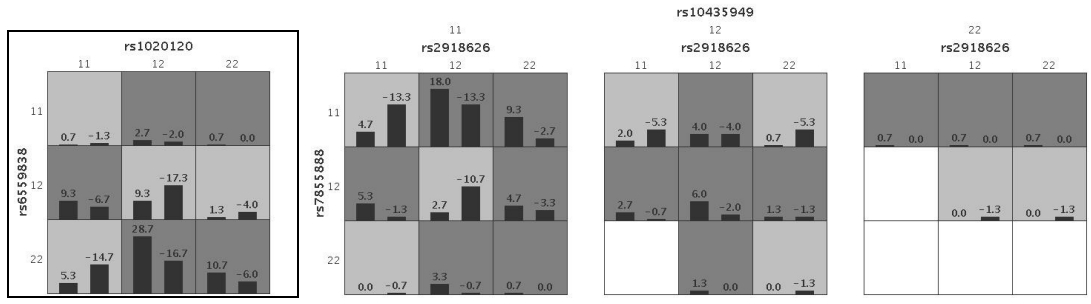


Figura 25. Modelo gráfico multifactorial de *NTRK2* e *HOXA1*.

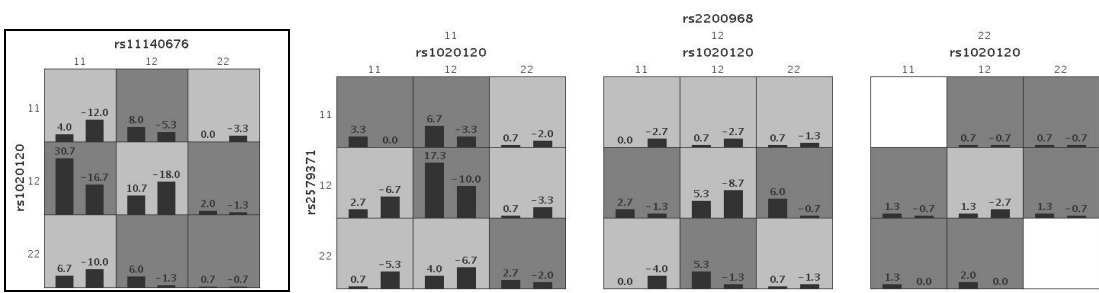


Figura 26. Modelo gráfico multifactorial de *NTRK2* e *DLX2*.

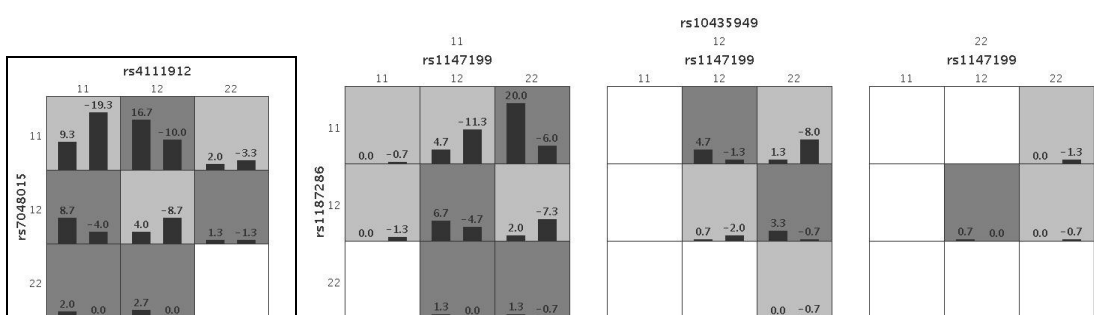


Figura 27. Modelos gráficos multifactoriais de *NTRK2* e *BDNF*.