

M

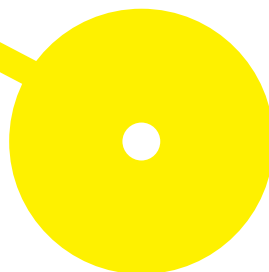
MESTRADO

EM ANÁLISES CLÍNICAS E SAÚDE PÚBLICA - MICROBIOLOGIA E SAÚDE PÚBLICA

Deteção de variantes genómicas usando duas metodologias de sequenciação massiva paralela – análise de exoma e genoma – Um estudo de caso

Ana Isabel Gonçalves de Oliveira

07/2021





**ESCOLA
SUPERIOR
DE SAÚDE**

**Deteção de variantes genómicas usando duas metodologias de sequenciação massiva
paralela – análise de exoma e genoma – um estudo de caso**

Autor

Ana Isabel Gonçalves de Oliveira

Orientadores

Prof. Nuno Maia, Escola Superior de Saúde do Instituto Politécnico do Porto (ESSIP.Porto),
Unidade Multidisciplinar de Investigação Biomédica (UMIB) / ITR – Laboratório para a Investigação
Integrativa e Translacional em Saúde Populacional

Prof. Doutora Sandra Marlene Mota, Centro de Investigação em Saúde e Ambiente (CISA), Escola
Superior de Saúde do Instituto Politécnico do Porto (ESSIP.Porto)

**Dissertação apresentada para cumprimento dos requisitos
necessários à obtenção do grau de Mestre em Análises Clínicas e
Saúde Pública – Ramo de Especialização em Microbiologia e
Saúde Pública pela Escola Superior de Saúde do Instituto
Politécnico do Porto.**

Agradecimentos

Finalizando o meu mestrado em Análises Clínicas e Saúde Pública, aproveito para agradecer a todas as pessoas que me ajudaram na realização desta tese.

Ao meu orientador, Dr. Nuno Maia, um especial e sincero agradecimento por ter aceite ser meu orientador, por se mostrar sempre disponível, por toda a paciência e apoio prestado. Obrigada pelos ensinamentos, foi um privilégio ser sua orientanda. À Dr. Sandra Marlene Mota por todo o apoio prestado.

Aos meu pais e ao meu irmão, um especial agradecimento por todo o apoio incondicional e por me ajudarem sempre a ultrapassar todos os obstáculos, mostrando que tudo é possível basta acreditar. Sem eles nada disto seria possível.

Às minhas amigas mais próximas, Mónica Garcia, Ana Ribeiro e Sandra Abreu, pela força enorme que me transmitiram e pela confiança que depositaram em mim.

E a todos aqueles que de uma forma ou de outra me auxiliaram no meu percurso, o meu mais sincero obrigada.

Resumo

A identificação das causas genéticas de uma determinada doença é crucial para uma mais precisa caracterização, diagnóstico e prognóstico, contribuindo em última estância para a possibilidade de um acompanhamento personalizado dos doentes e aconselhamento genético, incluindo diagnóstico pré-natal a familiares em risco. Ao interrogar várias regiões do genoma em simultâneo, a sequenciação massiva paralela assume um papel preponderante na identificação de variantes patogénicas, encurtando a odisseia do diagnóstico clínico e genético. A sequenciação massiva paralela pode apresentar diversas abordagens, destacando-se a sequenciação do exoma e do genoma pelo seu contributo na identificação de novas variantes e genes associados a doença.

Este estudo, centrado numa família com indivíduos afetados com défice intelectual, tem como objetivo a identificação de variantes provavelmente patogénicas, através da análise de dados de sequenciação massiva paralela. Estes dados foram sujeitos a um processo de filtragem recorrendo a diversas bases de dados e ferramentas bioinformáticas, tendo sido possível identificar um conjunto de variantes possivelmente envolvidas no fenótipo dos doentes. A sequenciação do genoma revelou ser mais eficaz na deteção de variantes genómicas quando comparado com a sequenciação do exoma, nomeadamente naquelas localizadas em regiões não codificantes, alterações do número de cópias e alterações cromossómicas estruturais. Este trabalho permite direcionar a investigação genética desta família, focada na caracterização funcional das variantes identificadas.

Palavras-chave: Variantes genéticas; Sequenciação massiva paralela; Sequenciação do exoma; Sequenciação do genoma.

Abstract

Identifying the genetic causes of a given disease is crucial for a more accurate characterization, diagnosis and prognosis, ultimately contributing for the possibility of personalized monitoring of the patients and genetic counselling, including prenatal diagnosis, to family members at risk. By simultaneously interrogating several regions of the genome, the massive parallel sequencing plays a key role in the identification of pathogenic variants, shortening the odyssey of clinical and genetic diagnosis. Massive parallel sequencing can present several approaches, standing out the exome and genome sequencing for their contribution to the identification of new variants and genes associated with disease.

This study, focused on a family with individuals affected with intellectual disability, aims to identify probably pathogenic variants through the analysis of massive parallel sequencing data. These data were subjected to a filtering process using several databases and bioinformatics tools, making it possible to identify a set of variants possibly involved in the phenotype. Genome sequencing showed to be more capable to detect genomic variants, compared to exome sequencing, especially those located in non-coding regions, copy number variations and structural chromosomal abnormalities. This work allows narrowing the clinical investigation of this family, focusing on the functional characterization of the identified variants.

Keywords: Genetic variants; Massive Parallel Sequencing; Exome Sequencing; Genome Sequencing.

Índice

| | |
|-------------------------------------------------------------------------------------------------------|----|
| 1. Introdução | 1 |
| 1.1. Variantes Genéticas | 1 |
| 1.2. Metodologias Chave no Diagnóstico Genético na Era Pré Sequenciação Massiva Paralela | 2 |
| 1.2.1 Citogenética Convencional | 2 |
| 1.2.2 Sequenciação de Sanger | 3 |
| 1.2.3 Microarrays | 3 |
| 1.3. Diagnóstico Genético por Sequenciação Massiva Paralela | 4 |
| 1.3.1 Sequenciação do exoma | 7 |
| 1.3.2 Sequenciação do genoma | 7 |
| 1.4 Objetivos | 8 |
| 2. Métodos | 9 |
| 2.1 História Clínica e Investigação Genética | 9 |
| 2.2 Processo de Filtragem | 9 |
| 2.2.1 Variantes de sequência – sequenciação do exoma | 11 |
| 2.2.2 Variantes de sequência – sequenciação do genoma | 11 |
| 2.2.3 Variantes do número de cópias (CNVs) | 12 |
| 2.2.4 Anomalias cromossômicas estruturais (SCVs) | 12 |
| 2.3 Softwares, bases de dados e outras ferramentas bioinformáticas | 12 |
| 2.3.1 Mutalyzer | 12 |
| 2.3.2 Single Nucleotide Polymorphism Database (dbSNP) | 13 |
| 2.3.3 Genome Aggregation Database (gnomAD) | 13 |
| 2.3.4 DECIPHER | 13 |
| 2.3.5 Database of Genomic Variants (DGV) | 13 |
| 2.3.6 ClinVar | 14 |
| 2.3.7 Online Mendelian Inheritance in Man (OMIM) | 14 |
| 2.3.8 SysID database | 14 |
| 2.3.9 Genotype–Tissue Expression (GTEx) | 14 |
| 2.3.10 Encyclopedia of DNA Elements (ENCODE) | 15 |
| 2.3.11 UCSC genome browser | 15 |

| | |
|----------------------------------------------------------------------------------------------------------|----|
| 2.3.12 <i>NNsplice</i> | 15 |
| 2.3.13 <i>Integrative Genomics Viewer (IGV)</i> | 15 |
| 3. Resultados | 17 |
| 3.1. Cobertura e variantes detetadas | 17 |
| 3.1.1 Variantes de sequência em regiões codificantes e regiões intrónicas flanqueadoras | 17 |
| 3.1.2 Variantes de sequência em regiões não codificantes | 18 |
| 3.1.3 Variantes do número de cópias e anomalias cromossómicas estruturais | 19 |
| 3.2. Variantes seleccionadas pelo processo de filtragem | 19 |
| 3.2.1 Análise das variantes de sequência | 19 |
| 3.2.2 Análise das variantes do número de cópias | 20 |
| 3.2.3 Análise das anomalias cromossómicas estruturais | 20 |
| 4. Discussão | 47 |
| 5. Conclusão | 51 |
| Referências Bibliográficas | 52 |
| Anexos | 58 |

Índice de Figuras

| | |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------|
| Figura 1 – Representação esquemática das etapas básicas da sequenciação de DNA, utilizando diferentes plataformas MPS (adaptado de Gupta N., Verma V.K. (2019) Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality)..... | 4 |
| Figura 2 – Fluxograma alusivo ao processo de filtragem utilizado neste estudo. | 10 |
| Figura 3 – Número de variantes de sequência detetadas em regiões codificantes e regiões intrónicas flanqueadoras, na análise do exoma e do genoma por tipo. | 18 |
| Figura 4 – Número de variantes de sequência detetadas em regiões não codificantes. | 18 |

Índice de Tabelas

| | |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Tabela 1 – Variantes de sequência detetadas pela análise de sequenciação do exoma (filtragem de acordo com a sua frequência)..... | 21 |
| Tabela 2 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a sua frequência e informação da base de dados ENCODE)..... | 22 |
| Tabela 3 – Variantes de sequência detetadas pela análise de sequenciação do exoma (filtragem de acordo com a informação da base de dados ClinVar)..... | 24 |
| Tabela 4 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação da base de dados ClinVar)..... | 25 |
| Tabela 5 – Variantes de sequência detetadas pela análise de sequenciação do exoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEX)..... | 28 |
| Tabela 6 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEX)..... | 29 |
| Tabela 7 – Variante de sequência final selecionada pela análise de sequenciação do exoma.... | 34 |
| Tabela 8 – Variantes de sequência finais selecionadas pela análise de sequenciação do genoma..... | 34 |
| Tabela 9 – CNVs detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência)..... | 35 |
| Tabela 10 – CNVs finais selecionadas pela análise dos dados de sequenciação do genoma..... | 39 |
| Tabela 11 – Inversões detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência)..... | 40 |
| Tabela 12 – Translocações detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência)..... | 45 |
| Tabela 13 – Translocação final detetada pela análise dos dados de sequenciação do genoma. | 46 |

Lista de abreviaturas

aCGH – *Microarray-based Comparative Genomic Hybridization*

AF – *Allele Frequency*

BAM – *Binary Alignment Mapping*

BCL11A – *Baf Chromatin Remodeling Complex Subunit*

CNVs – *Copy Number Variations*; Variantes do Número de Cópias

dbSNP – *Single Nucleotide Polymorphism Database*

DGV – *Database of Genomic Variants*

DI – Intellectual disability; Déficit Intelectual

DNA – Ácido desoxirribonucleico

DPYD – *Dihydropyrimidine Dehydrogenase*

ELOVL4 – *Elongation of Very Long Chain Fatty Acids-Like 4*

ENCODE – *Encyclopedia of DNA Elements*

ES – *Exome Sequencing*; Sequenciação do Exoma

FISH – *Fluorescent in situ hybridization*

FXS – *Fragile X syndrome*

FUT8 – *Fucosyltransferase 8*

gnomAD – *Genome Aggregation Database*

GS – *Genome Sequencing*; Sequenciação do Genoma

GTEx – *Genotype-Tissue Expression*

HGVS – *Human Genome Variation Society*

IGV – *Integrative Genomics Viewer*

KMT2C – *Lysine-Specific Methyltransferase 2C*

MYT1L – *Myelin Transcription Factor 1-Like*

MPS – *Massive Parallel Sequencing*; Sequenciação Massiva Paralela

NGS – *Next Generation Sequencing*

OMIM – *Online Mendelian Inheritance in Man*

PCR – *Polymerase Chain Reaction*

RORA – *Rar-Related Orphan Receptor A*

SETBP1 – *Set-Binding Protein 1*

SCVs – *Structural Chromosomal Variants*; Anomalias Cromossômicas Estruturais

SNP – *Single Nucleotide Polymorphism*

SNV – *Single Nucleotide Variant*

TACR1 – *Tachykinin Receptor 1*

TPM – valor médio de transcritos por milhão

TUSC 3 – *Tumor Suppressor Candidate 3*

VCF – *Variant Call Format*

ZC3H14 – *Zinc Finger CCCH Domain-Containing Protein 14*

1. Introdução

1.1. Variantes Genéticas

As variantes genéticas provocam desvios à sequência normal do genoma, contribuindo para a diversidade humana. O desenvolvimento dos métodos de sequenciação resultou na realização de mais testes genéticos, sendo possível a identificação de uma maior quantidade de variantes patogênicas e de genes associados a doença (1,2).

São inúmeras as variantes genéticas que podem estar presentes no genoma humano, podendo em alguns dos casos conduzir a doença. Um exemplo são as anomalias cromossômicas. Estas, têm origem em erros durante a gametogênese (meiose) ou em fases iniciais do desenvolvimento embrionário (mitose), e podem ser numéricas ou estruturais (*structural chromosomal variants, SCVs*). As variantes numéricas consistem no ganho ou perda de um único cromossoma ou de vários cromossomas, por outro lado, as estruturais incluem rearranjos equilibrados (translocações e inversões) ou desequilibrados (deleções, duplicações, inserções, cromossoma em anel e isocromossomas) (3–5). Outro exemplo são as variantes de sequência, que consistem em alterações de uma ou mais bases nucleotídicas na sequência do DNA, onde se incluem substituições, deleções, inserções, duplicações, inversões e translocações. Estas alterações podem afetar a sequência de aminoácidos e alteração no processo de *splicing* do mRNA (2,6). A substituição de uma base na sequência do genoma humano, designada como *Single Nucleotide Variant (SNV)* pode ser classificada dependendo do seu efeito na proteína como variante sinónima, *missense* ou *nonsense*, quando ocorrem em regiões codificantes. É considerada uma variante sinónima quando a substituição da base nucleotídica não conduz à troca do aminoácido. Uma variante *missense* ocorre quando há alteração de uma base na sequência do genoma, codificando um aminoácido diferente. Uma variante *nonsense* ocorre quando há alteração de uma base na sequência do genoma codificando um codão de terminação, originando uma proteína truncada. As SNVs estão localizadas em todo o genoma tanto em regiões codificantes como em regiões não codificantes, sendo por este motivo as variantes genéticas mais comuns detetadas na análise do genoma (7,8). As deleções e inserções de um ou mais pares de bases levam a que a sequência do genoma humano seja alterada comparativamente à sequência de referência, originando proteínas com dimensões diferentes e, podendo resultar numa função distinta da original (9).

Outra classe de variantes genéticas são as repetições nucleotídicas, onde mais de 40 doenças, afetando principalmente o sistema nervoso central, são causadas por expansões de regiões repetitivas tri-, tetra- e pentanucleotídicas. Estas mutações são dinâmicas e podem alterar de tamanho quando transmitidas para as gerações futuras, estando associadas a fenótipos variados dentro da própria família. Quanto maior for a repetição, maior será a gravidade da doença e mais precoce será o início dos sintomas (10).

A descoberta da causa genética de diversas doenças humanas tem especial foco em regiões codificantes de proteínas, os exões, uma vez que estas variantes podem ter um impacto direto na estrutura e consequentemente na função da proteína. Contudo, as variantes patogénicas não são apenas detetadas em regiões codificantes, sendo importante o seu estudo noutras regiões, nomeadamente em regiões não codificantes, como por exemplo os intrões, que podem ter um efeito deletério se essas variantes interferirem no processo de *splicing*. O *splicing* é o processo de remoção do intrão do mRNA, e alterações nesse processo podem resultar na perda ou inclusão de sequências codificantes. O efeito das variantes potencialmente envolvidas no processo de *splicing* não é de fácil interpretação, daí serem necessários estudos *in silico* para avaliar as possíveis consequências do *splicing* na variante em questão. Estes estudos são apenas previsões, sendo necessário análises de transcritos para a sua confirmação (11–13). Outras variantes localizadas em regiões não codificantes, por exemplo em regiões promotoras e reguladoras, podem desempenhar um papel relevante em doenças humanas. Estas são as mais prováveis de terem um impacto funcional e podem ainda alterar a expressão de um determinado gene (14,15). As variantes genéticas podem ser classificadas de acordo com o padrão de hereditariedade Mendeliano associado: (i) autossómicas recessivas, se são necessárias duas variantes alélicas num gene localizado num dos 22 autossomas para causar doença; (ii) autossómicas dominantes, se apenas uma variante num gene localizado num dos autossomas é suficiente para provocar doença ou (iii) variantes ligadas ao cromossoma X (16,17).

1.2. Metodologias Chave no Diagnóstico Genético na Era Pré Sequenciação Massiva Paralela

1.2.1 Citogenética Convencional

A citogenética convencional inclui ensaios de cariótipo e hibridização fluorescente *in situ* (*fluorescent in situ hybridization*, FISH). Estas técnicas marcaram o início dos testes genéticos com a descoberta da trissomia 21 como causa da síndrome de *Down* em 1959, sendo ainda hoje utilizadas no diagnóstico de várias doenças. A título de exemplo, no caso de doenças associadas ao défice intelectual (DI), uma revisão da literatura realizada por *Michelson* e seus colaboradores em 2011 (18) através de dados publicados entre 1980 e 2009 sobre a aplicação destas técnicas no diagnóstico, sugere que (i) a análise do cariótipo é capaz de identificar anomalias cromossómicas entre 4 % a 18,6 % dos doentes e (ii) a análise FISH apresenta um rendimento de diagnóstico de 3,5 % a 4,2 %. Conforme descrito por outros autores, o rendimento de diagnóstico das duas técnicas de citogenética convencional pode variar de 8 a 10 %, podendo atingir os 18 % quando aplicadas a doentes com sinais clínicos compatíveis com a síndrome de *Down*, a causa genética mais comum de DI (18,19).

A identificação do gene *FMR1* em 1991, como associado ao síndrome X-frágil (*Fragile X syndrome*, FXS), constituiu um marco na mudança de análises de citogenética convencional para os testes moleculares. O

FXS foi também o primeiro distúrbio genético identificado, causado por uma expansão de repetição de trinucleotídeos, uma classe de mutações recém-identificada à data. (20,21). A mesma revisão da literatura descrita anteriormente realizada por *Michelson* e seus colaboradores em 2011 (18) revelou um rendimento de diagnóstico molecular de FXS de cerca de 2 % para pacientes do sexo masculino com DI, e com um provável padrão de hereditariedade ligado ao cromossoma X (18).

1.2.2 Sequenciação de Sanger

A seleção do gene alvo a ser estudado é uma tarefa árdua, nomeadamente quando se trata de doenças com elevada heterogeneidade fenotípica e genotípica, e/ou quando existe dificuldade em identificar o padrão de hereditariedade subjacente. Este facto é ainda mais notório quando existe apenas um doente afetado na família, muitas vezes devido a uma variante *de novo*, ou seja, em que nenhum dos progenitores é portador. Não obstante, alguns projetos desenvolvidos na última década demonstraram que a técnica de sequenciação de *Sanger* é uma importante ferramenta na identificação de diversos genes associados a doença, como por exemplo genes associados a DI (22). Vários outros estudos foram publicados (22–24) baseados em técnicas de sequenciação de *Sanger*, tendo sido descobertos novos genes envolvidos em doença, contudo os estudos são muito restritos por se basearem apenas no estudo de regiões alvo e não de todo o genoma (22).

1.2.3 Microarrays

A identificação de grandes regiões em homozigotia através de *microarrays* de polimorfismos de um único nucleótido (*single nucleotide polymorphisms*, SNPs, ou *single nucleotide variants*, SNVs), permitiu a identificação de novos genes associados a doenças autossómicas recessivas. Um estudo publicado por *Schuurs-Hoeijmakers* e seus colaboradores em 2011, demonstra que esta metodologia é uma mais valia mesmo em famílias não consanguíneas, especialmente quando combinado com outras metodologias, como por exemplo a sequenciação massiva paralela (25).

A implementação de outras técnicas de *microarrays*, nomeadamente de hibridização genómica comparativa baseada em *microarray* (*microarray-based comparative genomic hybridization*, aCGH) para a identificação de variantes de número de cópias (*copy number variations*, CNVs) permitiu a identificação de anomalias cromossómicas, nomeadamente deleções e duplicações, com uma taxa de resolução superior. O aCGH apresenta algumas limitações, nomeadamente na deteção de anomalias cromossómicas balanceadas que ocorrem em cerca de 1,50 % dos nascimentos. Para além disso, o aCGH não é capaz de detetar duplicações e deleções menores que aproximadamente 200 kb. Estas limitações seriam facilmente ultrapassadas por outras metodologias como o FISH, ou técnicas baseadas na reação da polimerase em cadeia (*polymerase chain reaction*, PCR), contudo são técnicas demoradas e trabalhosas e/ou específicas apenas para uma região alvo do genoma. Desta forma, tornou-se necessário

o desenvolvimento de metodologias moleculares que abrangam todo o genoma e que apresentem uma resolução mais alta (22).

1.3. Diagnóstico Genético por Sequenciação Massiva Paralela

As metodologias que interrogam todo o genoma, como a sequenciação massiva paralela (*massive parallel sequencing*, MPS), vulgarmente conhecida como sequenciação de nova geração (*next generation sequencing*, NGS), tornaram-se uma alternativa aos métodos de sequenciação já existentes. A sua capacidade de realização de diagnóstico clínico num curto espaço de tempo e análise de pequenos fragmentos de DNA em simultâneo, contribuiu para avanços no diagnóstico genético de doenças hereditárias (26,27).

Atualmente, existem quatro plataformas principais de MPS. Apesar das especificidades técnicas de cada plataforma, todas possuem características comuns inerentes à metodologia de MPS (figura 1) (28).

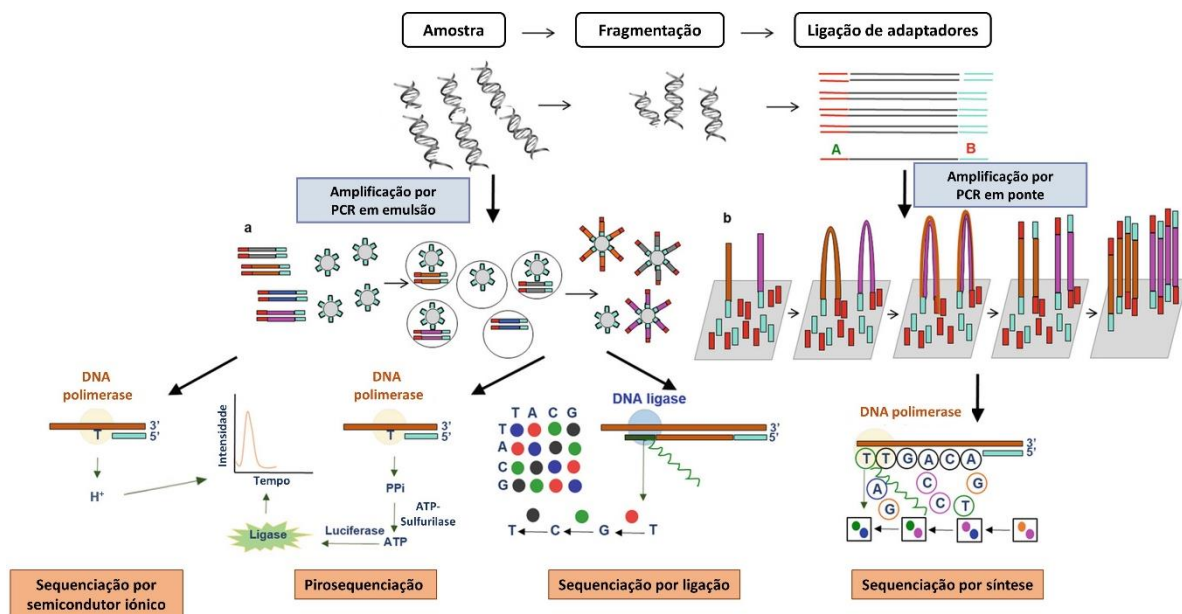


Figura 1 – Representação esquemática das etapas básicas da sequenciação de DNA, utilizando diferentes plataformas MPS (adaptado de Gupta N., Verma V.K. (2019) *Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality*).

Relativamente ao processamento das amostras, estas plataformas necessitam da preparação de uma biblioteca, que inclui fragmentação de DNA e ligação de adaptadores. Estes são sequências universais e específicas de cada plataforma que permitem a ligação de cada fragmento a um suporte sólido e amplificação por PCR (PCR em emulsão e PCR em ponte). Do processo de amplificação obtêm-se como resultado *clusters* de DNA, onde cada um é originário de um único fragmento da biblioteca. Da amplificação de cada *cluster* é obtido um sinal por luz, por fluorescência ou por modificação do potencial iônico,

dependendo do princípio técnico subjacente a cada plataforma: pirosequenciação, sequenciação por síntese, sequenciação por ligação ou sequenciação por semiconductor iônico. Deste modo, é possível sequenciar em simultâneo, centenas de milhares de *clusters* numa única corrida, maximizando assim o rendimento do processo, e por sua vez diminuindo consideravelmente os custos de sequenciação (29).

Na pirosequenciação, um único nucleótido é adicionado à reação. Se esse nucleótido for incorporado por complementaridade pela polimerase, é libertada uma molécula de pirofosfato. O sinal final obtido por luz é gerado após uma série de reações químicas e detetado através de uma câmara que regista a sequência do *cluster* (30). A sequenciação por síntese é baseada na incorporação de nucleótidos separadamente. Os nucleótidos utilizados nas reações de sequenciação estão ligados a uma única molécula fluorescente, específicos para cada nucleótido, e a uma molécula terminadora que impede a ligação de mais do que um nucleótido à sequência molde. Após o sinal fluorescente ser lido e registado, a molécula de fluorescência e o grupo terminador são clivados e lavados, permitindo a ligação de um novo nucleótido. Este processo é repetido até que a reação de sequenciação seja concluída (31). Na sequenciação por ligação não ocorre a incorporação de nucleótidos pela polimerase, mas sim a ligação de uma de 16 sondas oligonucleotídicas. Cada sonda possui duas bases específicas, seis bases degeneradas e uma das quatro moléculas fluorescentes. A primeira sonda liga-se à cadeia molde pela DNA *ligase* através da complementaridade das duas bases específicas da mesma, após a hibridização do iniciador com o adaptador da biblioteca. O sinal fluorescente é lido e registado. No final de aproximadamente 7 ciclos, a cadeia de DNA é desnaturada e é inserido um novo *primer* para reiniciar um novo ciclo de reações, sendo utilizados no total 5 *primers* de sequenciação (32). Na sequenciação por semicondutores iônicos, a sequência do *cluster* é determinada através das mudanças de pH da solução de sequenciação, resultando na libertação de iões de hidrogénio (H^+) durante a incorporação de nucleótidos. O restante processo de sequenciação é semelhante ao descrito para a pirosequenciação (33).

Da sequenciação massiva paralela resultam grandes quantidades de dados, exigindo uma infraestrutura de bioinformática sofisticada para a sua análise e interpretação. A bioinformática é uma disciplina que desenvolve e aplica técnicas computacionais para entender, organizar e analisar as informações geradas pela MPS (26,34). O processo de análise bioinformática inicia-se pela verificação da qualidade das leituras obtidas na sequenciação, seguido do alinhamento dos fragmento sequenciados no genoma de referência. Posteriormente, realiza-se a identificação e anotação das variantes para inferir a sua relevância. Por fim, pode ainda ser efetuada a priorização e filtragem das variantes identificadas (17,26,34).

O formato padrão para geração de dados na MPS é o FASTQ. Estes arquivos contêm informações sobre cada leitura, ou seja, identificação do arquivo, identificação da sequência e valores de qualidade. Para cada nucleótido sequenciado é atribuída uma pontuação de qualidade, sendo que as pontuações mais altas indicam maior qualidade de leitura. As leituras de sequenciação com qualidade baixa são sujeitas a

exclusão e as leituras de qualidade suficiente estão prontas para serem alinhadas com o genoma de referência (17,34).

Após o processo de validação da qualidade das leituras, estas podem ser alinhadas de duas formas distintas (i) alinhamento dos fragmentos sequenciados com o genoma de referência ou (ii) montagem *de novo*, que significa a montagem de um genoma do início sem ajuda de dados externos. A escolha principal para MPS é a primeira, onde os fragmentos sequenciados são mapeados na sequência de referência conhecida, normalmente na versão do genoma humano hg19/GRCh37. A montagem *de novo* ainda é restrita a determinados projetos mais específicos. O processo de alinhamento de sequências pode apresentar alguns desafios pois a maioria das leituras provenientes da sequenciação massiva paralela originam milhões de fragmentos curtos, dificultando o seu alinhamento, sobretudo em regiões repetitivas (17).

Após o alinhamento, a informação é organizada em ficheiros com o formato *Binary Alignment Mapping* (BAM), onde a informação genética é compactada e protegida (17). Os alinhamentos podem ser observados através de *softwares* específicos, sendo possível uma visualização das variantes genéticas como substituições, deleções e inserções, e rearranjos como inversões e translocações (34,35).

A anotação das variantes é a etapa que sucede ao alinhamento das leituras, onde são identificados os locais, tipos e conteúdo das mesmas. O objetivo principal desta anotação é fornecer o contexto da variante, apresentando a localização e posição do gene dentro do genoma e o impacto da variante (*missense, nonsense*, sinónimas, entre outras) (17). Para esta análise são incorporadas informações de outras base de dados como *Online Mendelian Inheritance in Man* (OMIM), *UCSC genome browser* e *SysID database*, esta última especificamente para doenças associadas ao DI. Todas estas informações são usadas para priorizar as variantes a estudar, para a sua classificação e interpretação. Os ficheiros resultantes desta análise são os *Variant Call Format* (VCF) (17,34). OS VCFs permitem a visualização das variantes presentes e prosseguir para a sua filtragem e priorização (17,36). Os arquivos VCF registam um número elevado de variantes, por exemplo cerca de 100 000 variantes por exoma. Deste modo, um processo de filtragem das variantes é indispensável para selecionar aquelas que poderão estar associadas à doença (17). No final, todos os algoritmos de filtragem devem ser devidamente validados para garantir que apenas as variantes que correspondem aos critérios usados sejam mantidas. Caso contrário, podemos excluir variantes que são clinicamente significativas. Para uma melhor identificação e compreensão das variantes existem ferramentas bioinformáticas que nos permitem localizá-las (posição cromossómica em locais funcionais de interesse), compará-las, efetuar análises *in silico* que permitem prever os seus efeitos a nível dos transcritos e da proteína, obter frequências alélicas, entre outras (34).

A MPS é amplamente utilizada no diagnóstico de doenças infecciosas, imunológicas e hereditárias, diagnóstico pré-natal não invasivo e ainda em cancros somáticos, tendo contribuído em grande escala para a identificação de novas variantes e novos genes causadores de doenças (17,26,37). No caso

particular do DI, já foram identificados mais de 1400 genes comprovadamente envolvidos e mais de 1200 genes candidatos (base de dados SysID atualizada a 10 de abril de 2021, consultada a 01 de julho de 2021) (38,39).

A MPS, tal como qualquer tecnologia molecular também apresenta algumas limitações, nomeadamente relacionadas com: (i) necessidade de profissionais bem treinados para análise e interpretação dos dados; (ii) necessidade de uma boa infraestrutura computacional e (iii) comprimento curto das leituras, que leva a uma maior dificuldade de alinhamento, nomeadamente em regiões de elevado teor repetitivo de bases (37,40,41).

As plataformas MPS permitem diferentes abordagens, de onde se destacam a sequenciação do exoma (*exome sequencing*, ES) e do genoma (*genome sequencing*, GS). Ambas permitem o estudo apenas do doente, ou com os progenitores paralelamente (sequenciação em trio). A primeira apresenta um custo mais baixo contudo, a segunda permite avaliar o alelismo das variantes através da sua origem (materna, paterna e *de novo*). Quando adicionados os dados dos pais à análise reduz-se em 10 vezes o número de variantes provavelmente patogénicas, favorecendo o processo de análise das variantes (22).

1.3.1 Sequenciação do exoma

A sequenciação do exoma permite o estudo de regiões codificantes de proteínas, os exões, que representam cerca de 2 % do genoma humano, e regiões intrónicas flanqueadoras. É nestas regiões que se localizam a grande maioria das variantes patogénicas descritas até agora. A descoberta de novos genes candidatos envolvidos em doenças e a possibilidade de reanalisar os dados à medida que novas informações no campo da genética humana vão surgindo, são algumas das principais vantagens desta técnica (22,42). A maioria dos estudos utiliza a sequenciação do exoma como técnica de eleição, pelo facto de (i) possuir um custo mais baixo, (ii) ser analiticamente menos desafiadora do que a sequenciação do genoma e (iii) pela maior probabilidade de identificação de variantes patogénicas em regiões codificantes (43).

Uma revisão publicada por *Srivastava* e seus colaboradores em 2019 (44), sobre a comparação entre ES e a técnica de *microarrays*, revelou que a ES apresenta capacidade de detetar os tipos de variantes patogénicas mais prevalentes, as SNVs, em indivíduos com doenças do neurodesenvolvimento. A ES demonstrou um rendimento de diagnóstico (36 % no geral) superior à técnica de *microarray* (15 – 20 %). Os autores prevêm que a ES se torne um teste de primeira linha para este grupo de doenças (44).

1.3.2 Sequenciação do genoma

A sequenciação do genoma abrange não só as regiões codificantes, como também as não codificantes sendo por isso o teste de diagnóstico mais abrangente, sendo esta a sua principal vantagem sobre o ES (42). Esta técnica permite a deteção, num único teste, da maioria dos tipos de variantes genéticas como

variantes de sequência, tanto em regiões codificantes como em regiões não codificantes, bem como variantes do número de cópias e anomalias cromossômicas (numéricas ou estruturais) (22).

Gillissen e seus colaboradores (45) realizaram pela primeira vez em 2014 um estudo utilizando como método a sequenciação do genoma completo para identificação de CNVs e SNVs com origem *de novo*. Nesse estudo foram considerados indivíduos com DI com resultado negativo de aCGH e ES. No final foi obtido um rendimento de 42 % utilizando GS. Com base em dados publicados anteriormente, o autor sugere um rendimento de diagnóstico cumulativo de 62 % (45).

A sequenciação do genoma completo é uma técnica recente que ainda requer o aperfeiçoamento dos algoritmos de identificação de determinadas variantes, como CNVs e SCVs, soluções de armazenamento de dados, e maior desenvolvimento no conhecimento sobre a função das regiões não codificantes do genoma. Não obstante, espera-se que a GS se torne a tecnologia de primeira linha para o diagnóstico de doenças hereditárias (46).

1.4 Objetivos

O presente estudo tem como objetivo identificar os defeitos genéticos potencialmente implicados no fenótipo apresentado pelos doentes da família em estudo, através da análise de dados resultantes de duas abordagens de sequenciação massiva paralela.

2. Métodos

2.1 História Clínica e Investigação Genética

Este estudo assenta na investigação genética de uma família composta por um casal não consanguíneo com cinco filhos, sendo que três deles, dois do sexo feminino e um do sexo masculino, foram referenciados à consulta de Genética Médica por défice intelectual.

Os indivíduos afetados apresentam um atraso no desenvolvimento e défice intelectual, sem dismorfias faciais, mas com alterações estruturais cerebrais detetadas na ressonância magnética: (i) caso *índice* (sexo feminino) – atrofia cerebral, mais acentuada no tronco e cerebelo, particularmente no vérmis; irmã – atrofia do vermis cerebelar e aumento da glândula pineal; irmão – atrofia cortical-subcortical discreta. A investigação genética de primeira linha anteriormente realizada incluiu análises de citogenética (cariótipo, análise de deleções e duplicações em regiões subteloméricas, e aCGH), genética molecular (síndrome de X- frágil) e bioquímica genética (ácido guanidinoacético, aminoácidos no sangue e na urina, ácidos orgânicos na urina, ácidos gordos de cadeia muito longa), todas sem alterações. Assim, o DNA extraído a partir de sangue periférico do caso *índice* foi sujeito a uma análise de sequenciação de exoma e genoma, cuja filtragem e análise dos dados será desenvolvida ao longo deste trabalho. De ressaltar que a sequenciação do genoma foi também efetuada no DNA de ambos os progenitores (análise de trio). Foram disponibilizados os dados de sequenciação do exoma e do genoma nos formatos BAM e VCF¹.

2.2 Processo de Filtragem

O processo de filtragem das variantes é efetuado nos ficheiros VCF, de modo a serem selecionadas as variantes mais relevantes para o estudo em questão (figura 2). Contudo, é necessária uma análise manual dos ficheiros BAM para a validação das variantes detetadas (anexos 1 a 7). Assim, de seguida, serão apresentados os processos de filtragem das variantes de sequência obtidas a partir da sequenciação do exoma e do genoma separadamente. Posteriormente, será também indicado o modo de filtragem das CNVs e SCVs identificadas pela sequenciação do genoma.

As siglas dos genes estão denominadas de acordo com as recomendações do *HUGO Gene Nomenclature Committee* (47) e as variantes descritas de acordo com as recomendações da *Human Genome Variation Society* (HGVS) (48). A ferramenta *Mutalyzer* foi utilizada para a conversão da nomenclatura das variantes de acordo com a sequência de DNA codificante (49).

¹ A sequenciação foi realizada no *Department of Human Genetics, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center*, em Nijmegen, Holanda, tendo sido assegurados os parâmetros de qualidade.

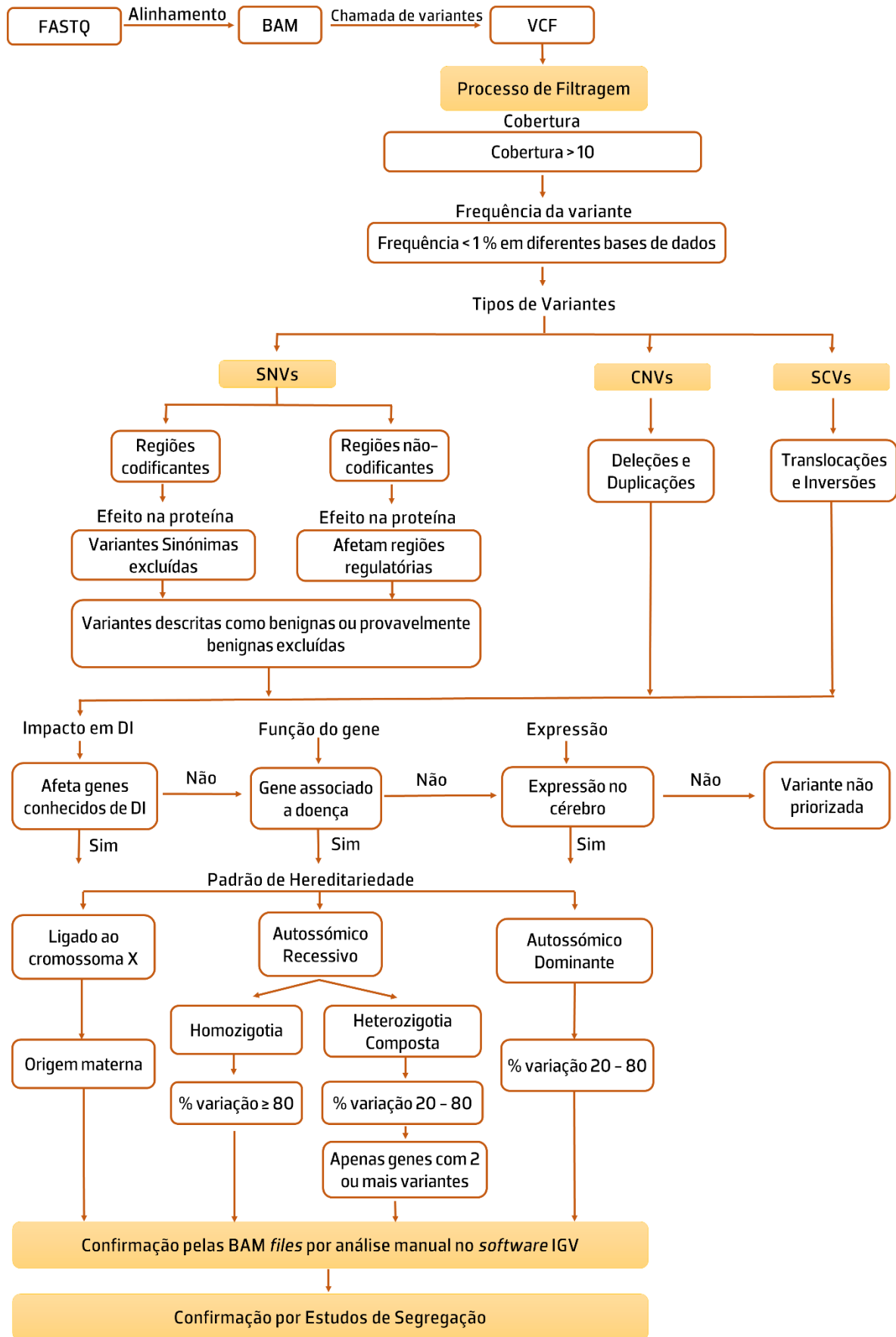


Figura 2 – Fluxograma alusivo ao processo de filtragem utilizado neste estudo.

2.2.1 Variantes de sequência – sequenciação do exoma

As variantes de sequência referentes à análise da sequenciação do exoma foram inicialmente filtradas de acordo com: (i) cobertura mínima de 10 vezes; (ii) percentagem de variação mínima de 20 %; (iii) frequência < 1 % em diferentes bases de dados, nomeadamente dbSNP (8), gnomAD (50) e base de dados local²; (iii) localização da variante, ex.: exão, locais aceitadores e dadores de *splicing*; (v) efeito na proteína (ex.: variantes sinónimas excluídas).

Numa segunda fase de filtragem, foi considerada a informação relativamente ao envolvimento das variantes em doença na base de dados ClinVar (51), tendo sido eliminadas todas as variantes descritas como benignas ou provavelmente benignas.

De seguida, foram avaliadas as características dos genes onde se localizam as variantes, nomeadamente a sua função e a sua associação a doença. Os genes foram gradualmente filtrados de acordo com: (i) genes implicados em défice intelectual (de acordo com a base de dados SysID (38) e OMIM (52)), (ii) implicação em doenças que afetam o sistema nervoso central (base de dados OMIM (52)); (ii) função e expressão génica significativa no cérebro (>0,1 TPM, GTEx *Project*(53)). Foi ainda considerada a zigtotia das variantes e a sua origem, ou seja, se a variante é de origem materna, paterna ou *de novo* (informação retirada da análise da GS). Neste passo, a filtragem foi realizada de acordo com o envolvimento dos genes relacionados com (i) doenças autossómicas recessivas: homozigotia (% de variação > 80) ou em heterozigotia composta (% de variação entre 20–80); (ii) doenças autossómicas dominantes: variantes *de novo* em heterozigotia (% de variação 20–80) e (iii) doenças ligadas ao cromossoma X: variantes de origem materna. Posteriormente, foram priorizadas as variantes através de uma análise *in silico* considerando um valor > 0,4 segundo o *NNSPLICE* para variantes em locais dadores e aceitadores de *splicing* (54). Por fim, todas as variantes foram validadas pela análise manual dos ficheiros BAM através da utilização do *software* IGV (55) (anexos 1, 2 e 3).

2.2.2 Variantes de sequência – sequenciação do genoma

A sequenciação do genoma permite a deteção de variantes de sequência em regiões codificantes e não codificantes. Estas foram filtradas de acordo com os critérios já descritos anteriormente na análise do exoma. A única diferença prende-se com o facto de que no processo de filtragem das variantes em regiões não codificantes, foram consideradas aquelas que se localizavam em regiões descritas como reguladoras na base de dados ENCODE (56).

² Base de dados do *Department of Human Genetics, Donders Institute for Brain, Cognition and Behavior, Radboud University Medical Center, Nijmegen, Holanda.*

2.2.3 Variantes do número de cópias (CNVs)

As CNVs, ou seja, deleções e duplicações, foram identificadas pela análise de dados de sequenciação do genoma e foram filtradas de acordo com: (i) frequência < 1 % DECIPHER (57), (ii) ausência na *Database of Genomic Variants* (DGV) (58) e (iii) gene implicado em défice intelectual. A validação ou exclusão das CNVs foi efetuada através de uma análise manual dos ficheiros BAM recorrendo ao *software* IGV (55) (anexos 4 e 5). Para aferir o possível efeito patogénico das CNVs identificadas, procedeu-se à comparação das suas características com outras já descritas, nomeadamente tamanho e genes envolvidos. A base de dados *USCS Genome Browser* (59,60), foi utilizada para obter os dados das bases de dados DECIPHER e DGV numa única interface, facilitando assim a análise das CNVs.

2.2.4 Anomalias cromossómicas estruturais (SCVs)

As anomalias cromossómicas estruturais, ou seja, inversões e translocações, identificadas pela sequenciação do genoma foram inicialmente filtradas de acordo com a presença de genes envolvidos em défice intelectual, com especial foco nos *breakpoints*. Posteriormente, foi efetuada a análise manual dos ficheiros BAM através do *software* IGV (55) (anexos 6 e 7).

2.3 Softwares, bases de dados e outras ferramentas bioinformáticas

De modo a proceder à análise e interpretação das variantes foram utilizados *softwares*, bases de dados e outras ferramentas bioinformáticas, que serão descritas neste capítulo.

2.3.1 Mutalyzer

O *Mutalyzer* foi utilizado para a conversão da nomenclatura das variantes de acordo com a sequência de DNA codificante. Esta ferramenta é uma base de dados abrangente que oferece suporte para a verificação da nomenclatura das variantes de sequência. Apresenta algumas funcionalidades tais como, (i) *Name Checker* que recebe a descrição da variante de sequência e verifica se esta está correta; (ii) *Syntaz Checker* para verificar se a sintaxe da variantes de sequência está correta; (iii) *Position Converter* para converter posições cromossómicas em posições para transcrição e vice-versa; (iv) *SNP Converter* para converter uma variante dbSNP numa anotação HGVS; (v) *Name Generator* para descrever uma variante de acordo com as recomendações do HGVS; (vi) *Description Extractor* para gerar a descrição de uma variante HGVS a partir de uma sequência de referência; (vii) *Reference File Loader* que permite a utilização na nossa própria sequência de referência; (viii) *Batch Checkers* para aceitar lista de entradas de variantes e (ix) *Web Services* para fornecer instruções para serviços *online* (49).

2.3.2 Single Nucleotide Polymorphism Database (dbSNP)

A base de dados dbSNP foi lançada pelo *National Center for Biotechnology Information* (NCBI) em resposta à necessidade de uma plataforma central que incorporasse várias informações relativas às variantes existentes no genoma. O dbSNP é acessível gratuitamente e apresenta dados relativos a: (i) substituições de nucleótido de base única; (ii) pequenas inserções e deleções e (iii) microssatélites, decorrentes de uma variedade de fontes, incluindo laboratórios individuais, indústria privada, centros de sequenciação de genoma e ainda de outras bases de dados. A plataforma dbSNP facilita a pesquisa de informação das variantes em termos de: (i) localização da sequência; (ii) frequência populacional; (iii) consequência molecular; (iv) função do gene; (v) homologia entre espécies; (vi) qualidade SNP e (vii) grau de variação da população (8,61). A base de dados dbSNP foi utilizada no processo de filtragem para informação relativa à frequência de cada variante.

2.3.3 Genome Aggregation Database (gnomAD)

A base de dados gnomAD foi desenvolvida em 2016 com o objetivo de agregar e harmonizar todos os dados provenientes de processos de sequenciação, quer do exoma quer do genoma, originários de projetos em grande escala. Deste modo, os dados são apresentados de forma mais resumida e tornam-se acessíveis para toda a comunidade científica. Esta ferramenta veio substituir o *Exome Aggregation Consortium* (ExAC), incluindo informações relativas a regiões não codificantes, bem como, variantes estruturais (50). Esta ferramenta foi utilizada no processo de filtragem de modo a fornecer informação sobre a frequência alélica.

2.3.4 DECIPHER

A base de dados DECIPHER é utilizada para partilhar e comparar dados fenotípicos e genotípicos contendo informações de mais de 39 mil doentes, utilizando o navegador *Ensembl* para definir os genes envolvidos em diferentes mutações genéticas. Esta plataforma é dinâmica e auxilia na colaboração entre investigadores para acelerar a descoberta de novas síndromes e de novas funções dos genes, uma vez que não só exibe as alterações detetadas na sequência do genoma, como também disponibiliza rapidamente informações clinicamente relevantes (57). Esta plataforma foi utilizada no processo de filtragem das variantes do número de cópias, uma vez que permite comparar as CNVs e SCVs identificadas com outras já descritas associadas a doença, através da comparação de tamanho entre elas.

2.3.5 Database of Genomic Variants (DGV)

A base de dados DGV descreve a prevalência de CNVs e SCVs e outras variantes no genoma humano em indivíduos saudáveis e clinicamente não afetados. A maioria dos dados presentes nesta plataforma dizem

respeito a dados de sequências genômicas individuais provenientes da MPS, melhorando consideravelmente a precisão da detecção de variantes estruturais (58). O DGV foi utilizado no processo de filtragem com o propósito de comparar as CNVs e SCVs identificadas com outras já descritas associadas a indivíduos normais.

2.3.6 ClinVar

O *ClinVar* é uma base de dados acessível publicamente para identificação de variantes genéticas humanas e para interpretação da sua associação a doença, englobando SNVs, pequenas inserções ou deleções, CNVs e SCVs. A confirmação do significado clínico de cada variante ou conjunto de variantes é realizada por laboratórios clínicos e de investigação, entre outros. O *ClinVar* é orientado apenas para variantes que apresentam um significado clínico ou funcional (62). Esta plataforma foi utilizada no processo de filtragem para verificar se as variantes se encontram ou não descritas como associadas a doença.

2.3.7 Online Mendelian Inheritance in Man (OMIM)

O OMIM, uma ferramenta acessível publicamente, foi lançado em 1987 tornando-se no principal arquivo de informações abrangentes sobre genes, fenótipos e sobre a sua relação. Contrariamente a outras bases de dados primárias, o OMIM resume informações novas e importantes, desempenhando também um papel fundamental na nomenclatura e classificação de fenótipos genéticos (52,63). Esta base de dados foi utilizada no processo de filtragem com o principal objetivo de pesquisar individualmente cada gene, determinando a sua função e a sua associação a doença, sendo um passo importante para compreender o fenótipo associado aos genes em questão.

2.3.8 SysID database

A base de dados SysID é uma ferramenta vinculada para o défice intelectual, sendo atualizada regularmente com o número de genes primários e candidatos. Através do SysID é possível a designação do gene consoante a sua severidade, ou seja, DI de moderado a grave, e também consoante o seu potencial síndromico (38). Esta base de dados foi utilizada no processo de filtragem para determinar através da pesquisa individual de cada gene a sua associação a DI.

2.3.9 Genotype-Tissue Expression (GTEx)

Em 2018 foi lançado o *Genotype-Tissue Expression*, um recurso público em grande escala de pesquisa rápida, que permite visualizar o resultado de estudos de expressão e regulação de genes humanos e a sua relação com a variação genética, englobando vários tecidos de origem humana (64). A base de dados

GTEX foi utilizada no processo de filtragem para visualização da expressão de cada gene a nível cerebral, onde foram retirados os valores correspondentes ao valor médio de transcritos por milhão (TPM).

2.3.10 *Encyclopedia of DNA Elements (ENCODE)*

A base de dados ENCODE foi lançada em 2003 devido ao esforço coletivo da comunidade científica pelo processo de sequenciação e anotação do genoma humano. O objetivo principal desta plataforma é agrupar um conjunto de experiências laboratoriais para identificar regiões bioquimicamente ativas nos genomas humanos e também nos genomas de ratos, numa grande variedade de amostras (65). O ENCODE é uma base de dados onde estão descritas alterações em algumas regiões não codificantes sendo por este motivo utilizada no processo de filtragem dessas regiões, onde foram apenas consideradas as variantes que se localizavam em regiões descritas como reguladoras ou envolvidas no processo de *splicing*.

2.3.11 *UCSC genome browser*

O *genome browser* da Universidade da Califórnia em Santa Cruz foi lançado em 2000, sendo uma base de dados de anotações do genoma. Embora seja possível a apresentação da sequência de DNA, o *genome browser* é utilizado principalmente para expor anotações do genoma, onde a maioria deriva do processo de alinhamento de sequências com o genoma de referência. Os dados são posteriormente armazenados e apresentados como coordenadas inicial e final no respetivo cromossoma (ex.: chr10:123400–123499). O UCSC permite obter os dados das bases de dados *Decipher* e DGV, entre outras, numa única interface, sendo por isso utilizado no processo de filtragem das CNVs (59,60).

2.3.12 *NNsplice*

NNsplice é uma ferramenta *online* de previsão de alteração de *splicing* baseado numa rede neural. Esta é utilizada para identificação de variantes potencialmente patogénicas com impacto nos locais de *splicing*, uma vez que, alterações nestes locais podem resultar na perda ou inclusão de sequências codificantes (54,66). Esta base de dados foi utilizada no processo de filtragem para uma análise *in silico* de cada variante relativa ao processo de *splicing*.

2.3.13 *Integrative Genomics Viewer (IGV)*

O IGV é uma ferramenta *desktop* escrita na linguagem de programação Java, com um elevado desempenho na visualização e pesquisa de dados genómicos, mesmo quando estes são demasiado amplos, requerendo sempre uma seleção de um genoma de referência antes da introdução dos dados. Este *software* permite a interação em tempo real de todas as escalas de resolução, quer do genoma

completo quer de pares de bases. Comparativamente a outras ferramentas com o mesmo propósito, o IGV destaca-se pela sua amplitude, suporte numa ampla gama de tipos de dados e ainda pela capacidade de visualização de várias regiões genómicas em simultâneo, lado a lado. O IGV suporta uma variedade de formatos de arquivo genómico, nomeadamente o formato BAM para alinhamento de sequências. Através desta plataforma é possível a visualização de rearranjos intra-cromossómicos, como deleções, inserções, duplicações e inversões, e ainda de rearranjos inter-cromossómicos como translocações, utilizando um esquema de cores para sinalizar os pares de *reads* não alinhadas (35).

3. Resultados

3.1. Cobertura e variantes detetadas

Apesar de partilharem o mesmo princípio metodológico, a sequenciação do exoma e do genoma apresentam diferenças sobretudo no que diz respeito à cobertura das diferentes regiões genómicas e ao tipo de variantes que detetam. Enquanto que a sequenciação do exoma mostrou ser eficaz na cobertura das regiões codificantes e regiões intrónicas flanqueadoras (cobertura média de 40 vezes), a sequenciação do genoma revelou capacidade para cobrir todas as regiões genómicas (cobertura média de 50 vezes). Assim, enquanto que a sequenciação do exoma permitiu a identificação de variantes pontuais e pequenas deleções e inserções com sucesso, a sequenciação do genoma revelou capacidade adicional para identificar variantes do número de cópias e anomalias cromossómicas estruturais, envolvendo regiões codificantes e não codificantes.

3.1.1 Variantes de sequência em regiões codificantes e regiões intrónicas flanqueadoras

A análise da sequenciação do exoma resultou num total de 100 177 variantes de sequência, enquanto que na análise da sequenciação do genoma esse número foi de 87 861, sendo uma diferença de 12 316 variantes. Quer a análise do exoma quer a do genoma detetam variantes de sequência que depois serão traduzidas em variantes *nonsense*, *missense*, sinónimas e variantes de *splicing*, (com especial impacto nos locais canónicos), pequenas deleções, inserções e deleções com inserções (*delins*), *in frame* e *out-of-frame*, e outras. Como pode ser observado na figura 3, na sequenciação do exoma foram detetadas 89 809 SNVs (103 *nonsense*, 10 750 *missense*, 11 518 sinónimas, 3 385 em locais de *splicing* não canónicos, 41 nos locais canónicos de *splicing* e 64 012 em regiões não codificantes, sem previsão de alteração do processo de *splicing*), 5 273 deleções, 4 972 inserções e 123 *delins*. Na sequenciação do genoma foram detetadas 81 056 SNVs (138 *nonsense*, 13 127 *missense*, 12 957 sinónimas, 2 102 em locais de *splicing* não canónicos, 197 nos locais canónicos de *splicing* e 52 535 em regiões não codificantes, sem previsão de alteração do processo de *splicing*), 3 592 deleções e 3 213 inserções.

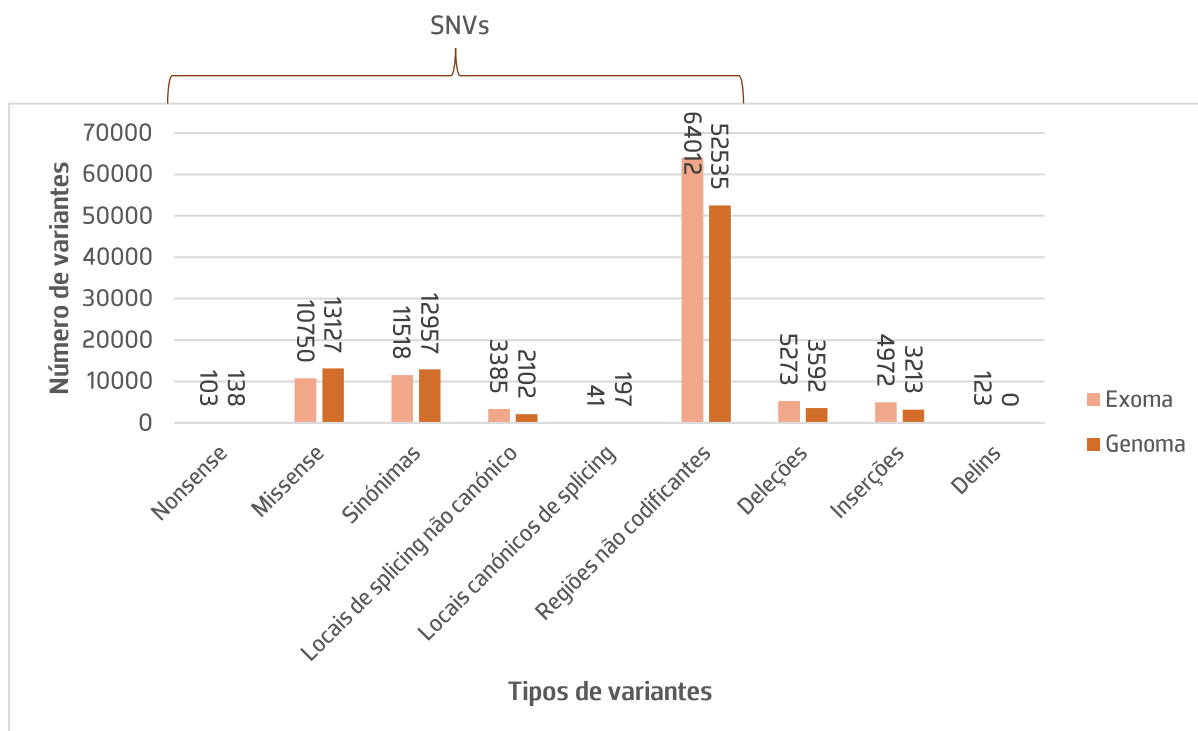


Figura 3 – Número de variantes de sequência detetadas em regiões codificantes e regiões intrônicas flanqueadoras, na análise do exoma e do genoma por tipo.

3.1.2 Variantes de sequência em regiões não codificantes

Dada a dificuldade de interpretar o possível efeito das variantes em regiões não codificantes, a sua análise ficou limitada àquelas localizadas em regiões descritas como reguladoras, ou envolvidas no processo de *splicing*, segundo a base de dados ENCODE (65). Estas variantes foram detetadas através da sequenciação do genoma, tendo sido identificadas substituições, deleções e inserções. No total, foram contabilizadas 30 594 variantes, sendo que 27 173 dessas são substituições, 1 849 são deleções e 1 572 são inserções. Estas variantes não estão incluídas na análise anterior.

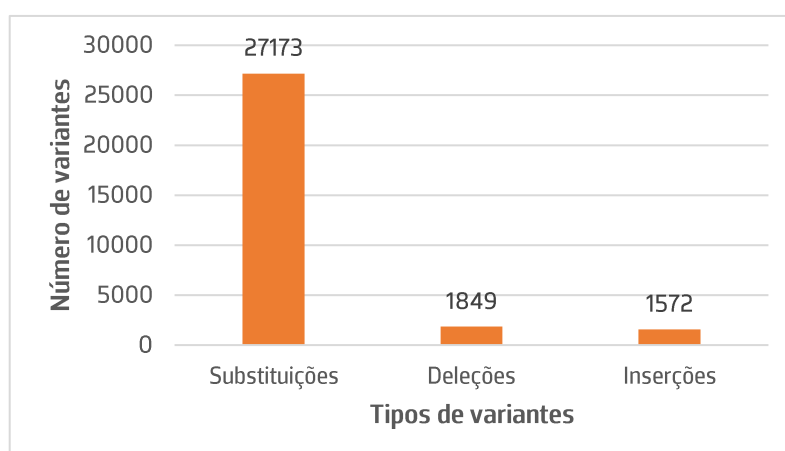


Figura 4 – Número de variantes de sequência detetadas em regiões não codificantes.

3.1.3 Variantes do número de cópias e anomalias cromossômicas estruturais

As variantes do número de cópias, isto é, deleções e duplicações, podem ser observadas tanto na análise do exoma como do genoma, contudo na análise do exoma apenas foi detetada 1 variante a qual não foi posteriormente validada pelas *BAM files*. Assim, na análise de genoma foram identificadas 757 deleções e 455 duplicações.

Relativamente às anomalias cromossômicas estruturais, isto é, inversões e translocações, detetadas a partir da análise dos dados de sequenciação de genoma, foram contabilizadas 17 inversões e 8 translocações.

3.2. Variantes selecionadas pelo processo de filtragem

A filtragem das variantes foi efetuada de acordo com o procedimento apresentado no ponto 2.2 do capítulo anterior. As variantes selecionadas serão descritas de acordo com as diferentes características consideradas para os diferentes níveis de filtragem (figura 2). O número de variantes vai reduzindo à medida que vão sendo eliminadas por não cumprimento dos critérios de seleção considerados. Serão apresentados os resultados do processo de filtragem das variantes de sequência da análise da sequenciação do exoma e do genoma separadamente, e das CNVs e SCVs identificadas pela análise do genoma. A análise das CNVs a partir da análise do exoma não será apresentada, uma vez que, como foi descrito anteriormente, apenas uma variante desse tipo foi identificada, não tendo sido validada pela análise dos ficheiros BAM.

3.2.1 Análise das variantes de sequência

O processo de filtragem teve como ponto de partida a cobertura e a percentagem de variação mínima de cada variante, assim como a sua frequência, localização no gene e efeito na proteína, como já descrito anteriormente (tabela 1 e tabela 2). Relativamente às variantes em regiões não codificantes, foram apenas consideradas aquelas que se localizam em regiões reguladoras descritas na base de dados ENCODE (tabela 2).

De seguida, a consulta da base de dados ClinVar levou à exclusão de 3 variantes classificadas como benignas, ou provavelmente benignas no exoma e 4 variantes no genoma (tabela 3 e tabela 4, respetivamente).

Numa fase posterior de filtragem, foram avaliadas as características dos genes, a zigotia das variantes e a sua origem (materna, paterna ou *de novo*) (tabela 5 e tabela 6). Para este último critério, foi utilizada a informação da sequenciação do genoma, uma vez que foi efetuada uma análise de trio. Finalmente, foram consideradas as variantes que cuja zigotia e origem estão de acordo com o padrão de hereditariedade descrito para as doenças reconhecidamente associadas a cada gene. No caso das variantes localizadas em genes ainda não associados a doença, foram consideradas as variantes em heterozigotia composta, e

em heterozigotia com origem *de novo*, resultando em apenas uma variante no exoma e 19 variantes no genoma (tabela 7 e tabela 8). Nas últimas tabelas foi ainda adicionada informação resultante da análise *in silico*, previsão de alteração de *splicing*, para auxiliar na discussão do possível envolvimento das variantes no fenótipo da doente. Todas as variantes foram manualmente validadas pela análise dos ficheiros BAM, no *software* IGV.

3.2.2 Análise das variantes do número de cópias

As variantes do número de cópias identificadas pela sequenciação do genoma foram inicialmente filtradas de acordo com a sua frequência e envolvimento de genes implicados em défice intelectual, como descrito anteriormente (tabela 9). Numa segunda fase de análise, foram avaliadas todas as variantes numa análise manual dos ficheiros BAM, resultando em apenas duas CNVs, tabela 10.

3.2.3 Análise das anomalias cromossómicas estruturais

As anomalias cromossómicas estruturais identificadas pela sequenciação do genoma foram igualmente seleccionadas de acordo com a sua frequência e envolvimento de genes implicados em défice intelectual, com especial foco nos *breakpoints* (tabela 11 e tabela 12). Numa segunda fase de análise, os ficheiros BAM foram manualmente analisados usando o *software* IGV para validação ou exclusão das variantes identificadas, resultando em apenas uma translocação, tabela 13.

Tabela 1 – Variantes de sequência detetadas pela análise de sequenciação do exoma (filtragem de acordo com a sua frequência).

| gDNA | cDNA | Gene | Efeito na proteína | Frequência | | |
|--------------------------------------------|-------------------------------|-----------------|----------------------|------------|----------|--------------|
| | | | | dbSNP | Local | gnomAD AF |
| Chr1(GRCh37):g.22206977C>T | NM_005529.5:c.2074G>A | <i>HSPG2</i> | p.(Val692Met) | 0,000199 | 0,002287 | 0,002882 |
| Chr1(GRCh37):g.156621322A>G | NM_021948.4:c.1138A>G | <i>BCAN</i> | p.(Ile380Val) | 0 | 0 | 0 |
| Chr1(GRCh37):g.202092567A>G | NM_004767.4:c.476A>G | <i>GPR37L1</i> | p.(Tyr159Cys) | 0 | 4,08E-05 | 0,00000398 |
| Chr1(GRCh37):g.220335594C>T | NM_012414.3:c.3171G>A | <i>RAB3GAP2</i> | p.(Met1057Ile) | 0 | 0,000531 | 0,00009557 |
| Chr1(GRCh37):g.235993606G>A | NM_000081.3:c.112C>T | <i>LYST</i> | p.(His38Tyr) | 0 | 0 | 0,00007431 |
| Chr1(GRCh37):g.235993696G>C | NM_000081.3:c.22C>G | <i>LYST</i> | p.(Leu8Val) | 0 | 0 | 0 |
| Chr2(GRCh37):g.51254843T>C | NM_001135659.1:c.569A>G | <i>NRXN1</i> | p.(Asn190Ser) | 0,000199 | 0,001429 | 0,0006691 |
| Chr3(GRCh37):g.54952510_54952511delCT | NM_020678.3:c.1013_1014del | <i>LRTM1</i> | p.(Glu338Alafs*3) | 0,00379 | 0,002532 | 0 |
| Chr3(GRCh37):g.54958975C>G | NM_020678.3:c.275G>C | <i>LRTM1</i> | p.(Gly92Ala) | 0,000399 | 0,00147 | 0,001274 |
| Chr5(GRCh37):g.137759929A>T | NM_016604.3:c.4138A>T | <i>KDM3B</i> | p.(Thr1380Ser) | 0 | 0 | 0,000007953 |
| Chr6(GRCh37):g.80629160G>A | NM_022726.3:c.646C>T | <i>ELOVL4</i> | p.(Arg216*) | 0 | 0 | 0,000003979 |
| Chr8(GRCh37):g.117661094T>C | NM_003756.2:c.779A>G | <i>EIF3H</i> | p.(Tyr260Cys) | 0 | 0 | 0 |
| Chr10(GRCh37):g.13693960T>A | NM_001318337.1:c.3164A>T | <i>FRMD4A</i> | p.(Asn1055Ile) | 0,005391 | 0,000531 | 0,001665 |
| Chr10(GRCh37):g.13698924_13698929delCGCCGC | NM_001318337.1:c.2759_2764del | <i>FRMD4A</i> | p.(Gly920_Gly921del) | 0 | 0 | 0 |
| Chr17(GRCh37):g.27910045G>A | NM_001085454.1:c.197C>T | <i>GIT1</i> | p.(Thr66Met) | 0,000998 | 0,00147 | 0,0006979 |
| Chr19(GRCh37):g.12774196G>A | NM_000528.3:c.844C>T | <i>MAN2B1</i> | p.(Pro282Ser) | 0,000399 | 0,003226 | 0,002641 |
| Chr20(GRCh37):g.2736395G>T | NM_001110514.1:c.1651G>T | <i>EBF4</i> | p.(Ala551Ser) | 0 | 0 | 0,000006492 |

AF – allele frequency, frequência alélica

Tabela 2 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a sua frequência e informação da base de dados ENCODE).

| gDNA | cDNA | Gene | Efeito na proteína | ENCODE | Frequência | | |
|--------------------------------------|-------------------------------------------|----------|--------------------|-------------------------|------------|-------------|--------------|
| | | | | | dbSNP | Local | gnomAD AF |
| Chr1(GRCh37):g.22206977C>T | NM_005529.5:c.2074G>A | HSPG2 | p.(Val692Met) | Heterochrom | 0,000199 | 0,002286834 | 0,002882 |
| Chr1(GRCh37):g.44341625A>C | NM_001350619.1:c.348-18430A>C | ST3GAL3 | p.? | Heterochrom | 0 | 0 | 0 |
| Chr1(GRCh37):g.97599568A>T | NM_000110.3:c.2623-35380T>A | DPYD | p.? | Heterochrom | 0 | 0 | 0 |
| Chr1(GRCh37):g.97599570A>T | NM_000110.3:c.2623-35382T>A | DPYD | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr1(GRCh37):g.97690599A>G | NM_000110.3:c.2442+9809T>C | DPYD | p.? | Heterochrom | 0 | 0 | 0 |
| Chr1(GRCh37):g.97786785C>A | NM_000110.3:c.2059-14932G>T | DPYD | p.? | Heterochrom | 0 | 0 | 0 |
| Chr1(GRCh37):g.97786951A>C | NM_000110.3:c.2059-15098T>G | DPYD | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr1(GRCh37):g.144854574T>C | NM_001350520.1:c.7382A>G | PDE4DIP | p.(His2461Arg) | Elongation | 0 | 0,0004492 | 0,000482 |
| Chr1(GRCh37):g.144855876C>G | NM_001350520.1:c.7163G>C | PDE4DIP | p.(Gly2388Ala) | Elongation | 0 | 0 | 0,0004701 |
| Chr1(GRCh37):g.144857705G>A | NM_001350520.1:c.6835C>T | PDE4DIP | p.(Arg2279Trp) | Elongation | 0 | 0 | 0,001917 |
| Chr1(GRCh37):g.144881547G>A | NM_001350520.1:c.4138C>T | PDE4DIP | p.(Arg1380Cys) | Elongation | 0 | 0,007473048 | 0,001909 |
| Chr1(GRCh37):g.144994670C>A | NM_001350520.1:c.-62962G>T | PDE4DIP | p.? | Enhancer | 0 | 0,007513884 | 0,001551 |
| Chr1(GRCh37):g.156621322A>G | NM_021948.4:c.1138A>G | BCAN | p.(Ile380Val) | Heterochrom | 0 | 0 | 0 |
| Chr1(GRCh37):g.202092567A>G | NM_004767.4:c.476A>G | GPR37L1 | p.(Tyr159Cys) | Heterochrom | 0 | 4,08363E-05 | 0,00000398 |
| Chr1(GRCh37):g.202748940T>A | NM_001314042.1:c.205-2718A>T | KDM5B | p.? | Weak_Txn | 0 | 0 | 0 |
| Chr1(GRCh37):g.220335594C>T | NM_012414.3:c.3171G>A | RAB3GAP2 | p.(Met1057Ile) | Elongation | 0 | 0,000530872 | 0,00009557 |
| Chr1(GRCh37):g.235993606G>A | NM_000081.3:c.112C>T | LYST | p.(His38Tyr) | Heterochrom | 0 | 0 | 0,00007431 |
| Chr1(GRCh37):g.235993696G>C | NM_000081.3:c.22C>G | LYST | p.(Leu8Val) | Heterochrom | 0 | 0 | 0 |
| Chr2(GRCh37):g.1801701_1801703delCCT | NM_001329844.1:c.3276+3765_3276+3767del | MYT1L | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr2(GRCh37):g.1801705insCA | NM_001329844.1:c.3276+3763_3276+3764insCA | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.1801719C>T | NM_001329844.1:c.3276+3749G>A | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.1801968_1801969delTT | NM_001329844.1:c.3276+3499_3276+3500del | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.1801992T>C | NM_001329844.1:c.3276+3476A>G | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.1815183C>T | NM_001329844.1:c.3081-2244G>A | MYT1L | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr2(GRCh37):g.1815184A>G | NM_001329844.1:c.3081-2245T>C | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.1815187A>G | NM_001329844.1:c.3081-2248T>C | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.1956513T>C | NM_001329844.1:c.153-9407A>G | MYT1L | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr2(GRCh37):g.2009269G>T | NM_001329844.1:c.-157-8150C>A | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.2009272A>G | NM_001329844.1:c.-157-8153T>C | MYT1L | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.51254843T>C | NM_001135659.1:c.569A>G | NRXN1 | p.(Asn190Ser) | Simple Repeat | 0,000199 | 0,001429271 | 0,0006691 |
| Chr2(GRCh37):g.60683103C>T | NM_022893.3:c.*4436G>A | BCL11A | p.? | Heterochrom | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694543A>G | NM_022893.3:c.487+1324T>C | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694672A>G | NM_022893.3:c.487+1195T>C | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694677A>G | NM_022893.3:c.487+1190T>C | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694678C>G | NM_022893.3:c.487+1189G>C | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694765G>A | NM_022893.3:c.487+1102C>T | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694775T>A | NM_022893.3:c.487+1092A>T | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694839A>G | NM_022893.3:c.487+1028T>C | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694840C>G | NM_022893.3:c.487+1027G>C | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.60694907A>T | NM_022893.3:c.487+960T>A | BCL11A | p.? | Simple Repeat | 0 | 0 | 0 |

AF – allele frequency, frequência alélica

Tabela 2 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a sua frequência e informação da base de dados ENCODE) (continuação).

| gDNA | cDNA | Gene | Efeito na proteína | ENCODE | Frequência | | |
|-------------------------------------------|---------------------------------|--------|----------------------|-------------------------|------------|-------------|--------------|
| | | | | | dbSNP | Local | gnomAD AF |
| Chr2(GRCh37):g.98764312C>G | NM_144992.4:c.988+13910C>G | VWA3B | p.? | Heterochrom | 0 | 0 | 0 |
| Chr2(GRCh37):g.138728515G>A | NM_006895.2:c.190+728G>A | HNMT | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.179309165G>A | NM_003690.4:c.380C>T | PRKRA | p.(Pro127Leu) | Elongation | 0 | 0,001429271 | 0,005223 |
| Chr2(GRCh37):g.210803410C>G | NM_001371986.1:c.6466-786C>G | UNC80 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr2(GRCh37):g.232102494A>T | NM_001291656.1:c.781-2162A>T | ARMC9 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr3(GRCh37):g.54952510_54952511delCT | NM_020678.3:c.1013_1014del | LRTM1 | p.(Glu338Alafs*3) | Heterochrom | 0,00379 | 0,002531852 | 0 |
| Chr3(GRCh37):g.54958975C>G | NM_020678.3:c.275G>C | LRTM1 | p.(Gly92Ala) | Heterochrom | 0,000399 | 0,001470108 | 0,001274 |
| Chr3(GRCh37):g.171121411A>G | NM_015028.3:c.58-33937T>C | TNIK | p.? | Heterochrom | 0 | 0 | 0 |
| Chr4(GRCh37):g.94049178C>A | NM_001510.3:c.735+17074C>A | GRID2 | p.? | Enhancer | 0 | 0 | 0 |
| Chr4(GRCh37):g.94531340C>T | NM_001510.3:c.2194-16080C>T | GRID2 | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr4(GRCh37):g.94531390A>G | NM_001510.3:c.2194-16030A>G | GRID2 | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr4(GRCh37):g.94599870G>C | NM_001510.3:c.2360+52284G>C | GRID2 | p.? | Heterochrom | 0 | 0 | 0 |
| Chr4(GRCh37):g.183542815T>A | NM_001080477.3:c.750-6989T>A | TENM3 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr4(GRCh37):g.183542883A>T | NM_001080477.3:c.750-6921A>T | TENM3 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr4(GRCh37):g.183542886A>G | NM_001080477.3:c.750-6918A>G | TENM3 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr4(GRCh37):g.183542890A>T | NM_001080477.3:c.750-6914A>T | TENM3 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr5(GRCh37):g.137759929A>T | NM_016604.3:c.4138A>T | KDM3B | p.(Thr1380Ser) | Elongation | 0 | 0 | 0,000007953 |
| Chr6(GRCh37):g.80629160G>A | NM_022726.3:c.646C>T | ELOVL4 | p.(Arg216*) | Heterochrom | 0 | 0 | 0,000003979 |
| Chr7(GRCh37):g.82628855T>C | NM_033026.5:c.3301-33052A>G | PCL0 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr8(GRCh37):g.15414124C>T | NM_001356429.1:c.138+16047C>T | TUSC3 | p.? | Heterochrom | 0 | 0 | 0 |
| Chr8(GRCh37):g.15430711C>A | NM_001356429.1:c.138+32634C>A | TUSC3 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr8(GRCh37):g.15543421T>C | NM_001356429.1:c.798+12076T>C | TUSC3 | p.? | Heterochrom | 0 | 0 | 0 |
| Chr8(GRCh37):g.117661094T>C | NM_003756.2:c.779A>G | EIF3H | p.(Tyr260Cys) | Elongation | 0 | 0 | 0 |
| Chr9(GRCh37):g.6644050A>C | NM_000170.2:c.334+564T>G | GLDC | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr10(GRCh37):g.13693960T>A | NM_001318337.1:c.3164A>T | FRMD4A | p.(Asn1055Ile) | Elongation | 0,005391 | 0,000530872 | 0,001665 |
| Chr10(GRCh37):g.13698924_13698929delCGCCG | NM_001318337.1:c.2759_2764del | FRMD4A | p.(Gly920_Gly921del) | Elongation | 0 | 0 | 0 |
| Chr14(GRCh37):g.20918208T>G | NM_017807.3:c.412-783A>C | OSGEP | p.? | Elongation | 0 | 0 | 0 |
| Chr14(GRCh37):g.32266388G>A | NM_025152.2:c.607+9309G>A | NUBPL | p.? | Heterochrom | 0 | 0 | 0 |
| Chr14(GRCh37):g.89074213C>G | NM_024824.4:c.1868+506C>G | ZC3H14 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr14(GRCh37):g.89074233T>C | NM_024824.4:c.1868+526T>C | ZC3H14 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr15(GRCh37):g.40952126T>G | NM_170589.4:c.6914+467T>G | KNL1 | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr15(GRCh37):g.60808170A>T | NM_134260.2:c.382-1214T>A | RORA | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr15(GRCh37):g.60808172T>A | NM_134260.2(RORA):c.382-1214T>A | RORA | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr15(GRCh37):g.60808218A>T | NM_134260.2:c.382-1262T>A | RORA | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr15(GRCh37):g.60889818C>T | NM_134260.2:c.137+29619G>A | RORA | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr15(GRCh37):g.61325608A>C | NM_134260.2:c.-406035T>G | RORA | p.? | Heterochrom | 0 | 0 | 0 |
| Chr15(GRCh37):g.61325613A>C | NM_134260.2(RORA):c.-406040T>G | RORA | p.? | Potential introduced ss | 0 | 0 | 0 |
| Chr15(GRCh37):g.61363779A>G | NM_134260.2:c.-444206T>C | RORA | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr17(GRCh37):g.19380486G>C | NM_001383.4:c.293+912G>C | DPH1 | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr17(GRCh37):g.27910045G>A | NM_001085454.1:c.197C>T | GIT1 | p.(Thr66Met) | Elongation | 0,000998 | 0,001470108 | 0,0006979 |

AF – allele frequency, frequência alélica

Tabela 2 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a sua frequência e informação da base de dados ENCODE) (continuação).

| gDNA | cDNA | Gene | Efeito na proteína | ENCODE | Frequência | | |
|------------------------------------------|----------------------------------|--------|--------------------|-------------------------|------------|-------------|-------------|
| | | | | | dbSNP | Local | gnomAD |
| | | | | | | | AF |
| Chr18(GRCh37):g.42408269C>T | NM_015559.2:c.487-40926C>T | SETBP1 | p.? | Heterochrom | 0 | 0 | 0 |
| Chr18(GRCh37):g.59730637G>T | NM_176787.4:c.2672+9269C>A | PIGN | p.? | Simple Repeat | 0 | 0 | 0 |
| Chr19(GRCh37):g.12774196G>A | NM_000528.3:c.844C>T | MAN2B1 | p.(Pro282Ser) | Elongation | 0,0003994 | 0,00322607 | 0,002641 |
| Chr20(GRCh37):g.2736395G>T | NM_001110514.1:c.1651G>T | EBF4 | p.(Ala551Ser) | Repressed | 0 | 0 | 0,000006492 |
| Chr22(GRCh37):g.18909958T>C | NM_016335.4:c.850-41A>G | PRODH | p.? | Potential introduced ss | 0 | 0 | 0,0003901 |
| ChrX(GRCh37):g.54783631G>C | NM_198510.2:c.2876C>G | ITIH6 | p.(Pro959Arg) | Heterochrom | 0 | 0 | 0,00001139 |
| ChrX(GRCh37):g.110250081T>C | NM_001128168.2:c.-89918T>C | PAK3 | p.? | Heterochrom | 0 | 0 | 0 |
| ChrX(GRCh37):g.119075419A>C | NM_024528.3:c.386+1764T>G | NKAP | p.? | Potential introduced ss | 0 | 0 | 0 |
| ChrX(GRCh37):g.146993596_146993598delGCG | NM_002024.5(FMR1):c.-102_-100del | FMR1 | p.? | Simple Repeat | 0 | 0,002940216 | 0 |

AF – allele frequency, frequência alélica

Tabela 3 – Variantes de sequência detetadas pela análise de sequenciação do exoma (filtragem de acordo com a informação da base de dados ClinVar).

| gDNA | cDNA | Gene | Efeito na proteína | ClinVar |
|--------------------------------------------|-------------------------------|----------|----------------------|---------------------------------------|
| Chr1(GRCh37):g.22206977C>T | NM_005529.5:c.2074G>A | HSPG2 | p.(Val692Met) | RCV000757384.6: benigna |
| Chr1(GRCh37):g.156621322A>G | NM_021948.4:c.1138A>G | BCAN | p.(Ile380Val) | NR |
| Chr1(GRCh37):g.202092567A>G | NM_004767.4:c.476A>G | GPR37L1 | p.(Tyr159Cys) | NR |
| Chr1(GRCh37):g.220335594C>T | NM_012414.3:c.3171G>A | RAB3GAP2 | p.(Met1057Ile) | RCV000497690.1: significado incerto |
| Chr1(GRCh37):g.235993606G>A | NM_000081.3:c.112C>T | LYST | p.(His38Tyr) | RCV000817010.2: significado incerto |
| Chr1(GRCh37):g.235993696G>C | NM_000081.3:c.22C>G | LYST | p.(Leu8Val) | NR |
| Chr2(GRCh37):g.51254843T>C | NM_001135659.1:c.569A>G | NRXN1 | p.(Asn190Ser) | RCV000473936.8: significado incerto |
| Chr3(GRCh37):g.54952510_54952511delCT | NM_020678.3:c.1013_1014del | LRTM1 | p.(Glu338Alafs*3) | NR |
| Chr3(GRCh37):g.54958975C>G | NM_020678.3:c.275G>C | LRTM1 | p.(Gly92Ala) | NR |
| Chr5(GRCh37):g.137759929A>T | NM_016604.3:c.4138A>T | KDM3B | p.(Thr1380Ser) | NR |
| Chr6(GRCh37):g.80629160G>A | NM_022726.3:c.646C>T | ELOVL4 | p.(Arg216*) | RCV000023451.4: patogénica |
| Chr8(GRCh37):g.117661094T>C | NM_003756.2:c.779A>G | EIF3H | p.(Tyr260Cys) | NR |
| Chr10(GRCh37):g.13693960T>A | NM_001318337.1:c.3164A>T | FRMD4A | p.(Asn1055Ile) | RCV000967151.2: benigna |
| Chr10(GRCh37):g.13698924_13698929delCGCCGC | NM_001318337.1:c.2759_2764del | FRMD4A | p.(Gly920_Gly921del) | NR |
| Chr17(GRCh37):g.27910045G>A | NM_001085454.1:c.197C>T | GIT1 | p.(Thr66Met) | NR |
| Chr19(GRCh37):g.12774196G>A | NM_000528.3:c.844C>T | MAN2B1 | p.(Pro282Ser) | RCV001123738.2: provavelmente benigna |
| Chr20(GRCh37):g.2736395G>T | NM_001110514.1:c.1651G>T | EBF4 | p.(Ala551Ser) | NR |

NR – não reportado

Tabela 4 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação da base de dados ClinVar).

| gDNA | cDNA | Gene | Efeito na proteína | ClinVar |
|--------------------------------------|-------------------------------------------|-----------------|--------------------|-------------------------------------|
| Chr1(GRCh37):g.22206977C>T | NM_005529.5:c.2074G>A | <i>HSPG2</i> | p.(Val692Met) | RCV000757384.6: benigna |
| Chr1(GRCh37):g.44341625A>C | NM_001350619.1:c.348-18430A>C | <i>ST3GAL3</i> | p.? | NR |
| Chr1(GRCh37):g.97599568A>T | NM_000110.3:c.2623-35380T>A | <i>DPYD</i> | p.? | NR |
| Chr1(GRCh37):g.97599570A>T | NM_000110.3:c.2623-35382T>A | <i>DPYD</i> | p.? | NR |
| Chr1(GRCh37):g.97690599A>G | NM_000110.3:c.2442+9809T>C | <i>DPYD</i> | p.? | NR |
| Chr1(GRCh37):g.97786785C>A | NM_000110.3:c.2059-14932G>T | <i>DPYD</i> | p.? | NR |
| Chr1(GRCh37):g.97786951A>C | NM_000110.3:c.2059-15098T>G | <i>DPYD</i> | p.? | NR |
| Chr1(GRCh37):g.144854574T>C | NM_001350520.1:c.7382A>G | <i>PDE4DIP</i> | p.(His2461Arg) | NR |
| Chr1(GRCh37):g.144855876C>G | NM_001350520.1:c.7163G>C | <i>PDE4DIP</i> | p.(Gly2388Ala) | NR |
| Chr1(GRCh37):g.144857705G>A | NM_001350520.1:c.6835C>T | <i>PDE4DIP</i> | p.(Arg2279Trp) | NR |
| Chr1(GRCh37):g.144881547G>A | NM_001350520.1:c.4138C>T | <i>PDE4DIP</i> | p.(Arg1380Cys) | NR |
| Chr1(GRCh37):g.144994670C>A | NM_001350520.1:c.-62962G>T | <i>PDE4DIP</i> | p.? | NR |
| Chr1(GRCh37):g.156621322A>G | NM_021948.4:c.1138A>G | <i>BCAN</i> | p.(Ile380Val) | NR |
| Chr1(GRCh37):g.202092567A>G | NM_004767.4:c.476A>G | <i>GPR37L1</i> | p.(Tyr159Cys) | NR |
| Chr1(GRCh37):g.202748940T>A | NM_001314042.1:c.205-2718A>T | <i>KDM5B</i> | p.? | NR |
| Chr1(GRCh37):g.220335594C>T | NM_012414.3:c.3171G>A | <i>RAB3GAP2</i> | p.(Met1057Ile) | RCV000497690.1: significado incerto |
| Chr1(GRCh37):g.235993606G>A | NM_000081.3:c.112C>T | <i>LYST</i> | p.(His38Tyr) | RCV000817010.2: significado incerto |
| Chr1(GRCh37):g.235993696G>C | NM_000081.3:c.22C>G | <i>LYST</i> | p.(Leu8Val) | NR |
| Chr2(GRCh37):g.1801701_1801703delCCT | NM_001329844.1:c.3276+3765_3276+3767del | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1801705insCA | NM_001329844.1:c.3276+3763_3276+3764insCA | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1801719C>T | NM_001329844.1:c.3276+3749G>A | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1801968_1801969delTT | NM_001329844.1:c.3276+3499_3276+3500del | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1801992T>C | NM_001329844.1:c.3276+3476A>G | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1815183C>T | NM_001329844.1:c.3081-2244G>A | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1815184A>G | NM_001329844.1:c.3081-2245T>C | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1815187A>G | NM_001329844.1:c.3081-2248T>C | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.1956513T>C | NM_001329844.1:c.153-9407A>G | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.2009269G>T | NM_001329844.1:c.-157-8150C>A | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.2009272A>G | NM_001329844.1:c.-157-8153T>C | <i>MYT1L</i> | p.? | NR |
| Chr2(GRCh37):g.51254843T>C | NM_001135659.1:c.569A>G | <i>NRXN1</i> | p.(Asn190Ser) | RCV000473936.8: significado incerto |

NR – não reportado

Tabela 4 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação da base de dados ClinVar) (continuação).

| gDNA | cDNA | Gene | Efeito na proteína | ClinVar |
|---------------------------------------|-------------------------------|---------------|--------------------|----------------------------|
| Chr2(GRCh37):g.60683103C>T | NM_022893.3:c.*4436G>A | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694543A>G | NM_022893.3:c.487+1324T>C | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694672A>G | NM_022893.3:c.487+1195T>C | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694677A>G | NM_022893.3:c.487+1190T>C | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694678C>G | NM_022893.3:c.487+1189G>C | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694765G>A | NM_022893.3:c.487+1102C>T | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694775T>A | NM_022893.3:c.487+1092A>T | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694839A>G | NM_022893.3:c.487+1028T>C | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694840C>G | NM_022893.3:c.487+1027G>C | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.60694907A>T | NM_022893.3:c.487+960T>A | <i>BCL11A</i> | p.? | NR |
| Chr2(GRCh37):g.98764312C>G | NM_144992.4:c.988+13910C>G | <i>VWA3B</i> | p.? | NR |
| Chr2(GRCh37):g.138728515G>A | NM_006895.2:c.190+728G>A | <i>HNMT</i> | p.? | NR |
| Chr2(GRCh37):g.179309165G>A | NM_003690.4:c.380C>T | <i>PRKRA</i> | p.(Pro127Leu) | RCV000390494.2:benigna |
| Chr2(GRCh37):g.210803410C>G | NM_001371986.1:c.6466-786C>G | <i>UNC80</i> | p.? | NR |
| Chr2(GRCh37):g.232102494A>T | NM_001291656.1:c.781-2162A>T | <i>ARMC9</i> | p.? | NR |
| Chr3(GRCh37):g.54952510_54952511delCT | NM_020678.3:c.1013_1014del | <i>LRTM1</i> | p.(Glu338Alafs*3) | NR |
| Chr3(GRCh37):g.54958975C>G | NM_020678.3:c.275G>C | <i>LRTM1</i> | p.(Gly92Ala) | NR |
| Chr3(GRCh37):g.171121411A>G | NM_015028.3:c.58-33937T>C | <i>TNFK</i> | p.? | NR |
| Chr4(GRCh37):g.94049178C>A | NM_001510.3:c.735+17074C>A | <i>GRID2</i> | p.? | NR |
| Chr4(GRCh37):g.94531340C>T | NM_001510.3:c.2194-16080C>T | <i>GRID2</i> | p.? | NR |
| Chr4(GRCh37):g.94531390A>G | NM_001510.3:c.2194-16030A>G | <i>GRID2</i> | p.? | NR |
| Chr4(GRCh37):g.94599870G>C | NM_001510.3:c.2360+52284G>C | <i>GRID2</i> | p.? | NR |
| Chr4(GRCh37):g.183542815T>A | NM_001080477.3:c.750-6989T>A | <i>TENM3</i> | p.? | NR |
| Chr4(GRCh37):g.183542883A>T | NM_001080477.3:c.750-6921A>T | <i>TENM3</i> | p.? | NR |
| Chr4(GRCh37):g.183542886A>G | NM_001080477.3:c.750-6918A>G | <i>TENM3</i> | p.? | NR |
| Chr4(GRCh37):g.183542890A>T | NM_001080477.3:c.750-6914A>T | <i>TENM3</i> | p.? | NR |
| Chr5(GRCh37):g.137759929A>T | NM_016604.3:c.4138A>T | <i>KDM3B</i> | p.(Thr1380Ser) | NR |
| Chr6(GRCh37):g.80629160G>A | NM_022726.3:c.646C>T | <i>ELOVL4</i> | p.(Arg216*) | RCV000023451.4: patogénica |
| Chr7(GRCh37):g.82628855T>C | NM_033026.5:c.3301-33052A>G | <i>PCL0</i> | p.? | NR |
| Chr8(GRCh37):g.15414124C>T | NM_001356429.1:c.138+16047C>T | <i>TUSC3</i> | p.? | NR |
| Chr8(GRCh37):g.15430711C>A | NM_001356429.1:c.138+32634C>A | <i>TUSC3</i> | p.? | NR |
| Chr8(GRCh37):g.15543421T>C | NM_001356429.1:c.798+12076T>C | <i>TUSC3</i> | p.? | NR |

NR – não reportado

Tabela 4 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação da base de dados ClinVar) (continuação).

| gDNA | cDNA | Gene | Efeito na proteína | ClinVar |
|-----------------------------------------|----------------------------------|---------------|----------------------|---------------------------------------|
| Chr8(GRCh37):g.117661094T>C | NM_003756.2:c.779A>G | <i>EIF3H</i> | p.(Tyr260Cys) | NR |
| Chr9(GRCh37):g.6644050A>C | NM_000170.2:c.334+564T>G | <i>GLDC</i> | p.? | NR |
| Chr10(GRCh37):g.13693960T>A | NM_001318337.1:c.3164A>T | <i>FRMD4A</i> | p.(Asn1055Ile) | RCV000967151.2: benigna |
| Chr10(GRCh37):g.13698924_13698929delCGC | NM_001318337.1:c.2759_2764del | <i>FRMD4A</i> | p.(Gly920_Gly921del) | NR |
| Chr14(GRCh37):g.20918208T>G | NM_017807.3:c.412-783A>C | <i>OSGEP</i> | p.? | NR |
| Chr14(GRCh37):g.32266388G>A | NM_025152.2:c.607+9309G>A | <i>NUBPL</i> | p.? | NR |
| Chr14(GRCh37):g.89074213C>G | NM_024824.4:c.1868+506C>G | <i>ZC3H14</i> | p.? | NR |
| Chr14(GRCh37):g.89074233T>C | NM_024824.4:c.1868+526T>C | <i>ZC3H14</i> | p.? | NR |
| Chr15(GRCh37):g.40952126T>G | NM_170589.4:c.6914+467T>G | <i>KNL1</i> | p.? | NR |
| Chr15(GRCh37):g.60808170A>T | NM_134260.2:c.382-1214T>A | <i>RORA</i> | p.? | NR |
| Chr15(GRCh37):g.60808172T>A | NM_134260.2(RORA):c.382-1214T>A | <i>RORA</i> | p.? | NR |
| Chr15(GRCh37):g.60808218A>T | NM_134260.2:c.382-1262T>A | <i>RORA</i> | p.? | NR |
| Chr15(GRCh37):g.60889818C>T | NM_134260.2:c.137+29619G>A | <i>RORA</i> | p.? | NR |
| Chr15(GRCh37):g.61325608A>C | NM_134260.2:c.-406035T>G | <i>RORA</i> | p.? | NR |
| Chr15(GRCh37):g.61325613A>C | NM_134260.2(RORA):c.-406040T>G | <i>RORA</i> | p.? | NR |
| Chr15(GRCh37):g.61363779A>G | NM_134260.2:c.-444206T>C | <i>RORA</i> | p.? | NR |
| Chr17(GRCh37):g.1938048G>C | NM_001383.4:c.293+912G>C | <i>DPH1</i> | p.? | NR |
| Chr17(GRCh37):g.27910045G>A | NM_001085454.1:c.197C>T | <i>GIT1</i> | p.(Thr66Met) | NR |
| Chr18(GRCh37):g.42408269C>T | NM_015559.2:c.487-40926C>T | <i>SETBP1</i> | p.? | NR |
| Chr18(GRCh37):g.59730637G>T | NM_176787.4:c.2672+9269C>A | <i>PIGN</i> | p.? | NR |
| Chr19(GRCh37):g.12774196G>A | NM_000528.3:c.844C>T | <i>MAN2B1</i> | p.(Pro282Ser) | RCV001123738.2: provavelmente benigna |
| Chr20(GRCh37):g.2736395G>T | NM_001110514.1:c.1651G>T | <i>EBF4</i> | p.(Ala551Ser) | NR |
| Chr22(GRCh37):g.18909958T>C | NM_016335.4:c.850-41A>G | <i>PRODH</i> | p.? | NR |
| ChrX(GRCh37):g.54783631G>C | NM_198510.2:c.2876C>G | <i>ITIH6</i> | p.(Pro959Arg) | NR |
| ChrX(GRCh37):g.110250081T>C | NM_001128168.2:c.-89918T>C | <i>PAK3</i> | p.? | NR |
| ChrX(GRCh37):g.119075419A>C | NM_024528.3:c.386+1764T>G | <i>NKAP</i> | p.? | NR |
| ChrX(GRCh37):g.146993596_146993598delG | NM_002024.5(FMR1):c.-102_-100del | <i>FMR1</i> | p.? | NR |

NR – não reportado

Tabela 5 – Variantes de sequência detetadas pela análise de sequenciação do exoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEX).

| ... | Gene | Efeito na proteína | Zigotia | Hereditariedade | Gene associado a doença | | Função do gene | Expressão no cérebro (TPM) | Observações |
|-----|-----------------|----------------------|---------------|-----------------|----------------------------------------------------------------------------------------------------------------------------|-------------|------------------------------------------------------------------------|----------------------------|-------------|
| | | | | | OMIM | SysID | | | |
| ... | <i>BCAN</i> | p.(Ile380Val) | Heterozigotia | Paterna | Sem fenótipo associado | Não | Adesão celular, migração, orientação do axônio e plasticidade neuronal | 110,3 | a) |
| ... | <i>GPR37L1</i> | p.(Tyr159Cys) | Heterozigotia | Paterna | Sem fenótipo associado | Não | Influência no desenvolvimento e na função neuronal | 56,28 | a) |
| ... | <i>RAB3GAP2</i> | p.(Met1057Ile) | Heterozigotia | Materna | Síndrome de <i>Martsof</i> e Síndrome 2 de <i>Warburg micro</i> – AR | Sim – AR | Papel fundamental no neurodesenvolvimento | 5,909 | b) |
| ... | <i>LYST</i> | p.(His38Tyr) | Heterozigotia | Paterna | Síndrome de Chediak-Higashi – AR | Sim – AR | Conservação da co-regulação lipídica entre espécies | 3,619 | c) |
| ... | <i>LYST</i> | p.(Leu8Val) | Heterozigotia | Paterna | Síndrome de Chediak-Higashi – AR | Sim – AR | Conservação da co-regulação lipídica entre espécies | 3,619 | c) |
| ... | <i>NRXN1</i> | p.(Asn190Ser) | Heterozigotia | Materna | Esquizofrenia, suscetibilidade a17; Síndrome 2 semelhante a <i>Pitt-Hopkins</i> – AR | Sim – AR | Bloqueio da formação de sinapses | 11,75 | b) |
| ... | <i>LRTM1</i> | p.(Glu338Alafs*3) | Heterozigotia | Materna | - | Não | - | 0,0368 | a) |
| ... | <i>LRTM1</i> | p.(Gly92Ala) | Heterozigotia | Materna | - | Não | - | 0,0368 | a) |
| ... | <i>KDM3B</i> | p.(Thr1380Ser) | Heterozigotia | Paterna | Síndrome de <i>Diets-Jongmans</i> – AD | Sim – AD | Interrupção da diferenciação das células HL-60 | 18,22 | d) |
| ... | <i>ELOVL4</i> | p.(Arg216*) | Heterozigotia | Paterna | Ictiose, tetraplegia espástica, atraso no desenvolvimento, ataxia espinocerebelar 34, Doença de <i>Stargardt 3</i> – AD/AR | Sim – AD/AR | Síntese de ácidos gordos de cadeia longa no cérebro | 11,49 | * |
| ... | <i>EIF3H</i> | p.(Tyr260Cys) | Heterozigotia | Materna | Sem fenótipo associado | Não | Controlo do processo de tradução | 31,8 | a) |
| ... | <i>FRMD4A</i> | p.(Gly920_Gly921del) | Heterozigotia | Paterna | Agnesia do corpo caloso, com anomalias faciais e ataxia cerebelar – AR | Sim – AR | Regulação da polaridade da célula epitelial | 8,498 | b) |
| ... | <i>GIT1</i> | p.(Thr66Met) | Heterozigotia | Paterna | Sem fenótipo associado | Não | Pode afetar a morfologia da sinapse | 132,4 | a) |
| ... | <i>EBF4</i> | p.(Ala551Ser) | Heterozigotia | Materna | Sem fenótipo associado | Não | Desenvolvimento neural e maturação de células B | 12,66 | a) |

TPM – valor médio de transcritos por milhão

- a) Variante em gene não associado a doença. Se autossómico recessivo: apenas uma variante ou variantes alélicas. Se autossómico dominante: variante herdada;
- b) Apenas uma variante num gene associado a doença autossómica recessiva;
- c) Variantes alélicas associadas a doença autossómica recessiva;
- d) Variante herdada em gene associado a doença autossómica dominante.

*Penetrância incompleta

Tabela 6 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEx).

| ... | Gene | Efeito na proteína | Zigotia | Hereditariedade | Gene associado a doença | | Função do gene | Expressão no cérebro (TPM) | Observações |
|-----|-----------------|--------------------|---------------|-----------------|--------------------------------------------------------------------------|----------|-----------------------------------------------------------------------------|----------------------------|-------------|
| | | | | | OMIM | SysID | | | |
| ... | <i>ST3GAL3</i> | p.? | Heterozigotia | Paterna/Materna | Encefalopatia epiléptica; Déficit intelectual autossômico recessivo - AR | Sim - AR | Catalisação da transferência de ácido siálico para substratos com galactose | 22,59 | b) |
| ... | <i>DPYD</i> | p.? | Heterozigotia | Materna | Défice de dihidropirimidina desidrogenase - AR | Sim - AR | Atividade em células mononucleares do sangue periférico | 1,446 | - |
| ... | <i>DPYD</i> | p.? | Heterozigotia | Materna | Défice de dihidropirimidina desidrogenase - AR | Sim - AR | Atividade em células mononucleares do sangue periférico | 1,446 | - |
| ... | <i>DPYD</i> | p.? | Heterozigotia | De novo | Défice de dihidropirimidina desidrogenase - AR | Sim - AR | Atividade em células mononucleares do sangue periférico | 1,446 | - |
| ... | <i>DPYD</i> | p.? | Heterozigotia | Paterna/Materna | Défice de dihidropirimidina desidrogenase - AR | Sim - AR | Atividade em células mononucleares do sangue periférico | 1,446 | - |
| ... | <i>DPYD</i> | p.? | Heterozigotia | Paterna/Materna | Défice de dihidropirimidina desidrogenase - AR | Sim - AR | Atividade em células mononucleares do sangue periférico | 1,446 | - |
| ... | <i>PDE4DIP</i> | p.(His2461Arg) | Heterozigotia | Paterna/Materna | Sem fenótipo associado | Sim - AD | Ligação enzimática | 12,37 | d) |
| ... | <i>PDE4DIP</i> | p.(Gly2388Ala) | Heterozigotia | Paterna/Materna | Sem fenótipo associado | Sim - AD | Ligação enzimática | 12,37 | d) |
| ... | <i>PDE4DIP</i> | p.(Arg2279Trp) | Heterozigotia | Paterna | Sem fenótipo associado | Sim - AD | Ligação enzimática | 12,37 | d) |
| ... | <i>PDE4DIP</i> | p.(Arg1380Cys) | Heterozigotia | Paterna | Sem fenótipo associado | Sim - AD | Ligação enzimática | 12,37 | d) |
| ... | <i>PDE4DIP</i> | p.? | Heterozigotia | Paterna | Sem fenótipo associado | Sim - AD | Ligação enzimática | 12,37 | d) |
| ... | <i>BCAN</i> | p.(Ile380Val) | Heterozigotia | Paterna | Sem fenótipo associado | Não | Adesão celular, migração, orientação do axônio e plasticidade neuronal | 110,3 | a) |
| ... | <i>GPR37L1</i> | p.(Tyr159Cys) | Heterozigotia | Paterna | Sem fenótipo associado | Não | Influência no desenvolvimento e na função neuronal | 56,28 | a) |
| ... | <i>KDM5B</i> | p.? | Heterozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico recessivo - AR | Sim - AR | Ativação do genoma embrionário | 5,925 | b) |
| ... | <i>RAB3GAP2</i> | p.(Met1057Ile) | Heterozigotia | Materna | Síndrome de <i>Martsof</i> e Síndrome 2 de <i>Warburg</i> micro - AR | Sim - AR | Papel fundamental no neurodesenvolvimento | 5,909 | b) |
| ... | <i>LYST</i> | p.(His38Tyr) | Heterozigotia | Paterna | Síndrome de <i>Chediak-Higashi</i> - AR | Sim - AR | Conservação da co-regulação lipídica entre espécies | 3,619 | c) |
| ... | <i>LYST</i> | p.(Leu8Val) | Heterozigotia | Paterna | Síndrome de <i>Chediak-Higashi</i> - AR | Sim - AR | Conservação da co-regulação lipídica entre espécies | 3,619 | c) |

TPM – valor médio de transcritos por milhão

- Variante em gene não associado a doença. Se autossômico recessivo: apenas uma variante ou variantes alélicas. Se autossômico dominante: variante herdada;
- Apenas uma variante num gene associado a doença autossômica recessiva;
- Variantes alélicas associadas a doença autossômica recessiva;
- Variante herdada em gene associado a doença autossômica dominante.

Tabela 6 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEx) (continuação).

| ... | Gene | Efeito na proteína | Zigotia | Hereditariedade | Gene associado a doença | | Função do gene | Expressão no cérebro (TPM) | Observações |
|-----|---------------|--------------------|---------------|-----------------|-----------------------------------------------------------------------------------------|----------|-----------------------------------|----------------------------|-------------|
| | | | | | OMIM | SysID | | | |
| ... | <i>MYT1L</i> | p.? | Homozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Homozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Homozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | De novo | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | - |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | De novo | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | - |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | De novo | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | - |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>MYT1L</i> | p.? | Heterozigotia | Paterna/Materna | Atraso no desenvolvimento autossômico dominante – AD | Sim – AD | Manutenção da identidade neuronal | 13,73 | d) |
| ... | <i>NRXN1</i> | p.(Asn190Ser) | Heterozigotia | Materna | Esquizofrenia, suscetibilidade a, 17 e Síndrome 2 semelhante a <i>Pitt-Hopkins</i> – AR | Sim – AR | Bloqueio da formação de sinapses | 11,75 | b) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | De novo | Síndrome de Dias-Logan – AD | Sim – AD | Mediador da troca de hemoglobina | 10,06 | - |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna | Síndrome de Dias-Logan – AD | Sim – AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | De novo | Síndrome de Dias-Logan – AD | Sim – AD | Mediador da troca de hemoglobina | 10,06 | - |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna/Materna | Síndrome de Dias-Logan – AD | Sim – AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | De novo | Síndrome de Dias-Logan – AD | Sim – AD | Mediador da troca de hemoglobina | 10,06 | - |

TPM – valor médio de transcritos por milhão

- b) Apenas uma variante num gene associado a doença autossômica recessiva;
- d) Variante herdada em gene associado a doença autossômica dominante.

Tabela 6 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTeX) (continuação).

| ... | Gene | Efeito na proteína | Zigotia | Hereditariedade | Gene associado a doença | | Função do gene | Expressão no cérebro (TPM) | Observações |
|-----|---------------|--------------------|---------------|-----------------|-----------------------------------------------------------------------------------------------------------------------------|-------------|------------------------------------------------------------|----------------------------|-------------|
| | | | | | OMIM | SysID | | | |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna | Síndrome de Dias-Logan - AD | Sim - AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna | Síndrome de Dias-Logan - AD | Sim - AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna/Materna | Síndrome de Dias-Logan - AD | Sim - AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna/Materna | Síndrome de Dias-Logan - AD | Sim - AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>BCL11A</i> | p.? | Heterozigotia | Paterna/Materna | Síndrome de Dias-Logan - AD | Sim - AD | Mediador da troca de hemoglobina | 10,06 | d) |
| ... | <i>VWA3B</i> | p.? | Heterozigotia | Paterna/Materna | Ataxia espinocerebelar autossômica recessiva 22 - AR | Sim - AR | Atua na transcrição e correção do DNA | 0 | b) |
| ... | <i>HNMT</i> | p.? | Heterozigotia | Paterna/Materna | Asma e atraso no desenvolvimento autossômico recessivo 51 - AR | Sim - AR | Inibe a atividade endógena de metilação de histamina COS-7 | 4,494 | b) |
| ... | <i>UNC80</i> | p.? | Heterozigotia | Paterna/Materna | Hipotonia infantil com retardo psicomotor - | Sim - AR | Sensibilidade extracelular ao Ca (2+) das correntes de | 13,44 | b) |
| ... | <i>ARMC9</i> | p.? | Heterozigotia | Paterna | Síndrome de <i>Joubert</i> 30 - AR | Sim - AR | Formação de um módulo de proteína | 4,391 | b) |
| ... | <i>LRTM1</i> | p.(Glu338Alafs*3) | Heterozigotia | Materna | - | Não | - | 0,0368 | e) |
| ... | <i>LRTM1</i> | p.(Gly92Ala) | Heterozigotia | Materna | - | Não | - | 0,0368 | e) |
| ... | <i>TNIK</i> | p.? | Heterozigotia | Materna | Atraso no desenvolvimento autossômico recessivo 54 - AR | Sim - AR | Efector específico para regular o citoesqueleto de actina | 16,8 | b) |
| ... | <i>GRID2</i> | p.? | Heterozigotia | Paterna | Ataxia espinocerebelar autossômica recessiva 18 - AR | Sim - AR | Organização sináptica | 0,4796 | f) |
| ... | <i>GRID2</i> | p.? | Heterozigotia | De novo | Ataxia espinocerebelar autossômica recessiva 18 - AR | Sim - AR | Organização sináptica | 0,4796 | f) |
| ... | <i>GRID2</i> | p.? | Heterozigotia | De novo | Ataxia espinocerebelar autossômica recessiva 18 - AR | Sim - AR | Organização sináptica | 0,4796 | f) |
| ... | <i>GRID2</i> | p.? | Heterozigotia | Materna | Ataxia espinocerebelar autossômica recessiva 18 - AR | Sim - AR | Organização sináptica | 0,4796 | f) |
| ... | <i>TENM3</i> | p.? | Heterozigotia | Paterna/Materna | Microftalmia síndrômica 15 - AR | Não | Adesão de células heterofílicas | 4,242 | f) |
| ... | <i>TENM3</i> | p.? | Homozigotia | Paterna/Materna | Microftalmia síndrômica 15 - AR | Não | Adesão de células heterofílicas | 4,242 | f) |
| ... | <i>TENM3</i> | p.? | Heterozigotia | Paterna | Microftalmia síndrômica 15 - AR | Não | Adesão de células heterofílicas | 4,242 | f) |
| ... | <i>TENM3</i> | p.? | Heterozigotia | Paterna | Microftalmia síndrômica 15 - AR | Não | Adesão de células heterofílicas | 4,242 | f) |
| ... | <i>KDM3B</i> | p.(Thr1380Ser) | Heterozigotia | Paterna | Síndrome de <i>Diets-Jongmans</i> - AD | Sim - AD | Interrupção da diferenciação das células HL-60 | 18,22 | d) |
| ... | <i>ELOVL4</i> | p.(Arg216*) | Heterozigotia | Paterna | Ictiose, tetraplegia espástica e atraso no desenvolvimento, Ataxia espinocerebelar 34, Doença de <i>Stargardt</i> 3 - AD/AR | Sim - AD/AR | Síntese de ácidos gordos de cadeia longa no cérebro | 11,49 | - |

TPM – valor médio de transcritos por milhão

- b) Apenas uma variante num gene associado a doença autossômica recessiva;
- d) Variante herdada em gene associado a doença autossômica dominante;
- e) Baixa expressão no cérebro;
- f) Fenótipo não encaixa.

Tabela 6 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEx) (continuação).

| ... | Gene | Efeito na proteína | Zigotia | Hereditariedade | Gene associado a doença | | Função do gene | Expressão no cérebro (TPM) | Observações |
|-----|---------------|----------------------|---------------|-----------------|-----------------------------------------------------------------------------------------|----------|------------------------------------------------------------|----------------------------|-------------|
| | | | | | OMIM | SysID | | | |
| ... | <i>PCLO</i> | p.? | Heterozigotia | Paterna/Materna | Hipoplasia pontocerebelar tipo 3 - AR | Sim - AR | Sensor de Ca (2+) na excitose em células beta pancreáticas | 5,092 | b) |
| ... | <i>TUSC3</i> | p.? | Heterozigotia | Paterna | Atraso no desenvolvimento autossómico recessivo 7 - AR | Sim - AR | Mediação da captação celular de Mg (2+) | 13,68 | - |
| ... | <i>TUSC3</i> | p.? | Heterozigotia | Paterna | Atraso no desenvolvimento autossómico recessivo 7 - AR | Sim - AR | Mediação da captação celular de Mg (2+) | 13,68 | - |
| ... | <i>TUSC3</i> | p.? | Heterozigotia | De novo | Atraso no desenvolvimento autossómico recessivo 7 - AR | Sim - AR | Mediação da captação celular de Mg (2+) | 13,68 | - |
| ... | <i>EIF3H</i> | p.(Tyr260Cys) | Heterozigotia | Materna | Sem fenótipo associado | Não | Controlo do processo de tradução | 31,8 | a) |
| ... | <i>GLDC</i> | p.? | Heterozigotia | Paterna/Materna | Encefalopatia glicina - AR | Sim - AR | Clivagem da glicina no sistema nervoso central | 6,555 | b) |
| ... | <i>FRMD4A</i> | p.(Gly920_Gly921del) | Heterozigotia | Paterna | Agnesia do corpo caloso, com anomalias faciais e ataxia cerebelar - AR | Sim - AR | Regulação da polaridade da célula epitelial | 8,498 | b) |
| ... | <i>OSGEP</i> | p.? | Heterozigotia | Paterna/Materna | Síndrome de <i>Galloway-Mowat</i> 3 - AR | Sim - AR | Linhas celulares de podócitos humanos | 22,15 | b) |
| ... | <i>NUBPL</i> | p.? | Heterozigotia | Paterna | Défiência do complexo mitocondrial I, tipo nuclear 21 - AR | Sim - AR | Montagem do complexo I da cadeia respiratória | 3,101 | b) |
| ... | <i>ZC3H14</i> | p.? | Heterozigotia | Paterna/Materna | Atraso no desenvolvimento autossómico recessivo 56 - AR | Sim - AR | Ligação especificamente ao RNA poli (A) | 4,054 | - |
| ... | <i>ZC3H14</i> | p.? | Heterozigotia | De novo | Atraso no desenvolvimento autossómico recessivo 56 - AR | Sim - AR | Ligação especificamente ao RNA poli (A) | 4,054 | - |
| ... | <i>KNL1</i> | p.? | Heterozigotia | De novo | Microcefalia primária 4, autossómica recessiva - AR | Sim - AR | Pertence ao complexo de centrómero | 0,03177 | b) |
| ... | <i>RORA</i> | p.? | Heterozigotia | Paterna | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | d) |
| ... | <i>RORA</i> | p.? | Heterozigotia | Paterna | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | d) |
| ... | <i>RORA</i> | p.? | Heterozigotia | Paterna | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | d) |

TPM – valor médio de transcritos por milhão

- a) Variante em gene não associado a doença. Se autossómico recessivo: apenas uma variante ou variantes alélicas. Se autossómico dominante: variante herdada;
- b) Apenas uma variante num gene associado a doença autossómica recessiva;
- d) Variante herdada em gene associado a doença autossómica dominante.

Tabela 6 – Variantes de sequência detetadas pela análise de sequenciação do genoma (filtragem de acordo com a informação das bases de dados OMIM, SysID e GTEx) (continuação).

| ... | Gene | Efeito na proteína | Zigotia | Hereditariedade | Gene associado a doença | | Função do gene | Expressão no cérebro (TPM) | Observações |
|-----|---------------|--------------------|---------------|-----------------|--------------------------------------------------------------------------------------------------------|----------|-----------------------------------------------------------------------|----------------------------|-------------|
| | | | | | OMIM | SysID | | | |
| ... | <i>RORA</i> | p.? | Heterozigotia | Materna | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | d) |
| ... | <i>RORA</i> | p.? | Heterozigotia | De novo | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | - |
| ... | <i>RORA</i> | p.? | Heterozigotia | Paterna/Materna | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | d) |
| ... | <i>RORA</i> | p.? | Heterozigotia | Paterna/Materna | Transtorno do desenvolvimento intelectual com ou sem epilepsia ou ataxia cerebelar - AD | Sim - AD | Papel no desenvolvimento cerebelar | 3,728 | d) |
| ... | <i>DPH1</i> | p.? | Heterozigotia | Paterna | Atraso no desenvolvimento com baixa estatura, características faciais dismórficas e cabelos ralos - AR | Sim - AR | Função na biossíntese da diftamida como dímero ou multímero | 28,13 | b) |
| ... | <i>GIT1</i> | p.(Thr66Met) | Heterozigotia | Paterna | Sem fenótipo associado | Não | Pode afetar a morfologia da sinapse | 132,4 | a) |
| ... | <i>SETBP1</i> | p.? | Heterozigotia | De novo | Síndrome de retração da face média de <i>Schinzel-Giedion</i> e atraso no desenvolvimento AD 29 - AD | Sim - AD | Interação com SET endógeno numa linha celular de osteossarcoma humano | 3,135 | - |
| ... | <i>PIGN</i> | p.? | Heterozigotia | Materna | Múltiplas anomalias congénitas-hipotonia-síndrome de convulsões 1 - AR | Sim - AR | Papel central na replicação da geração de NIC estruturais e numéricos | 1,063 | b) |
| ... | <i>EBF4</i> | p.(Ala551Ser) | Heterozigotia | Materna | Sem fenótipo associado | Não | Desenvolvimento neural e maturação de células B | 12,66 | a) |
| ... | <i>PRODH</i> | p.? | Heterozigotia | Materna | Esquizofrenia e Hiperprolinemia tipo I - AD/AR | Sim - AR | Envolvido em fenótipos psiquiátricos e comportamentais | 50,8 | b) |
| ... | <i>ITIH6</i> | p.(Pro959Arg) | Heterozigotia | Paterna | - | Sim - XL | - | 0 | g) |
| ... | <i>PAK3</i> | p.? | Heterozigotia | Paterna | Atraso no desenvolvimento ligado ao X 30/47 - XL | Sim - XL | Papel no crescimento pós-natal do cérebro | 12,23 | g) |
| ... | <i>NKAP</i> | p.? | Heterozigotia | Materna | Transtorno do desenvolvimento intelectual ligado ao X, síndromico tipo <i>Hackman-Di Donato</i> - XL | Sim - XL | Proteína nuclear com papel no splicing e processamento de RNA | 3,546 | f) |
| ... | <i>FMR1</i> | p.? | Heterozigotia | Paterna | Síndrome de X-frágil - XL | Sim - XL | Papel no splicing | 15,52 | g) |

TPM – valor médio de transcritos por milhão

- a) Variante em gene não associado a doença. Se autossómico recessivo: apenas uma variante ou variantes alélicas. Se autossómico dominante: variante herdada;
- b) Apenas uma variante num gene associado a doença autossómica recessiva;
- d) Variante herdada em gene associado a doença autossómica dominante;
- f) Fenótipo não encaixa;
- g) Variante no cromossoma X de origem paterna.

Tabela 7 – Variante de sequência final selecionada pela análise de sequenciação do exoma.

| gDNA | cDNA | Gene | Efeito na proteína | Análise <i>in silico</i> |
|----------------------------|----------------------|---------------|--------------------|--------------------------|
| | | | | <i>Splicing</i> |
| Chr6(GRCh37):g.80629160G>A | NM_022726.3:c.646C>T | <i>ELOVL4</i> | p.(Arg216*) | SA |

SA – sem alteração

Tabela 8 – Variantes de sequência finais selecionadas pela análise de sequenciação do genoma.

| gDNA | cDNA | Gene | Efeito na proteína | Análise <i>in silico</i> |
|-----------------------------|-------------------------------|---------------|--------------------|--------------------------|
| | | | | <i>Splicing</i> |
| Chr1(GRCh37):g.97599568A>T | NM_000110.3:c.2623-35380T>A | <i>DPYD</i> | p.? | SA |
| Chr1(GRCh37):g.97599570A>T | NM_000110.3:c.2623-35382T>A | <i>DPYD</i> | p.? | SA |
| Chr1(GRCh37):g.97690599A>G | NM_000110.3:c.2442+9809T>C | <i>DPYD</i> | p.? | SA |
| Chr1(GRCh37):g.97786785C>A | NM_000110.3:c.2059-14932G>T | <i>DPYD</i> | p.? | SA |
| Chr1(GRCh37):g.97786951A>C | NM_000110.3:c.2059-15098T>G | <i>DPYD</i> | p.? | SA |
| Chr2(GRCh37):g.1815183C>T | NM_001329844.1:c.3081-2244G>A | <i>MYT1L</i> | p.? | SA |
| Chr2(GRCh37):g.1815184A>G | NM_001329844.1:c.3081-2245T>C | <i>MYT1L</i> | p.? | SA |
| Chr2(GRCh37):g.1815187A>G | NM_001329844.1:c.3081-2248T>C | <i>MYT1L</i> | p.? | SA |
| Chr2(GRCh37):g.60683103C>T | NM_022893.3:c.*4436G>A | <i>BCL11A</i> | p.? | SA |
| Chr2(GRCh37):g.60694672A>G | NM_022893.3:c.487+1195T>C | <i>BCL11A</i> | p.? | SA |
| Chr2(GRCh37):g.60694678C>G | NM_022893.3:c.487+1189G>C | <i>BCL11A</i> | p.? | SA |
| Chr6(GRCh37):g.80629160G>A | NM_022726.3:c.646C>T | <i>ELOVL4</i> | p.(Arg216*) | SA |
| Chr8(GRCh37):g.15414124C>T | NM_001356429.1:c.138+16047C>T | <i>TUSC3</i> | p.? | SA |
| Chr8(GRCh37):g.15430711C>A | NM_001356429.1:c.138+32634C>A | <i>TUSC3</i> | p.? | SA |
| Chr8(GRCh37):g.15543421T>C | NM_001356429.1:c.798+12076T>C | <i>TUSC3</i> | p.? | SA |
| Chr14(GRCh37):g.89074213C>G | NM_024824.4:c.1868+506C>G | <i>ZC3H14</i> | p.? | SA |
| Chr14(GRCh37):g.89074233T>C | NM_024824.4:c.1868+526T>C | <i>ZC3H14</i> | p.? | SA |
| Chr15(GRCh37):g.61325608A>C | NM_134260.2:c.-406035T>G | <i>RORA</i> | p.? | SA |
| Chr18(GRCh37):g.42408269C>T | NM_015559.2:c.487-40926C>T | <i>SETBP1</i> | p.? | SA |

SA – sem alteração

Tabela 9 – CNVs detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência).

| Coordenada genómica inicial | Coordenada genómica final | Tipo de CNV | Genes implicados | Frequência | | Observações |
|-----------------------------|---------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| Chr7(GRCh37):g.151976663 | Chr7(GRCh37):g.151988690 | Duplicação | <i>KMT2C</i> | 0 | 0 | a) |
| Chr7(GRCh37):g.152074968 | Chr7(GRCh37):g.152081131 | Duplicação | <i>KMT2C</i> | 0 | 0 | - |
| Chr7(GRCh37):g.152101032 | Chr7(GRCh37):g.152113665 | Duplicação | <i>KMT2C</i> | 0 | 0 | - |
| Chr8(GRCh37):g.100307460 | Chr8(GRCh37):g.100346565 | Deleção | <i>VPS13B</i> | 0 | 0 | a) |
| Chr9(GRCh37):g.2147036 | Chr9(GRCh37):g.2150768 | Deleção | <i>SMARCA2</i> | 0 | 0 | a) |
| Chr12(GRCh37):g.99993641 | Chr12(GRCh37):g.100009508 | Deleção | <i>ANKS1B</i> | 0 | 0 | a) |
| Chr15(GRCh37):g.76884496 | Chr15(GRCh37):g.76895003 | Deleção | <i>SCAPER</i> | 0 | 0 | a) |
| Chr16(GRCh37):g.1785947 | Chr16(GRCh37):g.1789009 | Deleção | <i>MAPK8IP3</i> | 0 | 0 | a) |
| Chr17(GRCh37):g.44213566 | Chr17(GRCh37):g.44292956 | Duplicação | <i>KANSL1</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.2699929 | ChrX(GRCh37):g.3748751 | Deleção | <i>ARSD,ARSE,ARSF,ARSH,CXorf28,GYG2,MXRA5,PRKX,RP11-706015.1,XG</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.5058184 | ChrX(GRCh37):g.23094122 | Deleção | <i>ACO02365.1,ACE2,AMELX,AP1S2,ARHGAP6,ASB11,ASB9,ATXN3L,BEND2,BMX,CA5B,CDKL5,CLCN4,CNKSR2,CTPS2,CXorf23,DDX53,EGFL6,EIF1AX,FAM9A,FAM9B,FAM9C,FANCB,FIGF,FRMPD4,GEMIN8,GLRA2,GPM6B,GPR143,GPR64,GRPR,HCCS,HDHD1,KAL1,KLHL34,MAGEB17,MAP3K15,MAP7D2,MBTPS2,MID1,MOSPD2,MSL3,NHS,NLGN4X,OFD1,PDHA1,PHEX,PHKA2,PIGA,PIR,PNPLA4,PPEF1,PRPS2,RAB9A,RAI2,RBBP7,REPS2,RPS6KA3,RS1,S100G,SCML1,SCML2,SH3KBP1,SHROOM2,SMPX,SMS,STS,SYAP1,TBL1X,TCEANC,TLR7,TLR8,TMEM27,TMSB4X,TRAPPC2,TXLNG,VCX,VCX2,VCX3A,VCX3B,WWC3,YY2,ZNF645,ZRSR</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.23101209 | ChrX(GRCh37):g.32986725 | Deleção | <i>ACOT9,APOO,ARX,CXorf21,CXorf58,DCAF8L1,DCAF8L2,DMD,EIF2S3,FTHL17,GK,IL1RAPL1,KLHL15,MAGEB1,MAGEB10,MAGEB18,MAGEB2,MAGEB3,MAGEB4,MAGEB5,MAGEB6,NROB1,PCYT1B,PK3,POLA1,PRDX4,PTCHD1,SAT1,TAB3,ZFX</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.32989219 | ChrX(GRCh37):g.33837773 | Deleção | <i>DMD</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.34445797 | ChrX(GRCh37):g.44300537 | Deleção | <i>ATP6AP2,BCOR,CASK,CHDC2,CXorf22,CXorf27,CXorf30,CXorf38,CYBB,DDX3X,DYNLT3,EFHC2,FAM47B,FAM47C,GPR34,GPR82,LANCL3,MAGEB16,MAOA,MAOB,MED14,MID1IP1,NDP,NYX,OTC,PRRG1,RPGR,SRPX,SYTL5,TM4SF2,TMEM47,TSPAN7,USP9X,XK</i> | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 9 – CNVs detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial | Coordenada genómica final | Tipo de CNV | Genes implicados | Frequência | | Observações |
|-----------------------------|---------------------------|-------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| ChrX(GRCh37):g.44306762 | ChrX(GRCh37):g.48230327 | Deleção | <i>ARAF, CDK16, CFP, CHST7, CXorf24, CXorf31, CXorf36, DUSP21, ELK1, FUNDCl, KDM6A, KRBOX4, NDUFB11, PHF16, RBM10, RGN, RP2, SLC9A7, SPACA5, SPACA5B, SSX1, SSX3, SSX5, SYN1, TIMP1, UBA1, USP11, UXT, ZNF157, ZNF182, ZNF41, ZNF630, ZNF674, ZNF81</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.48245846 | ChrX(GRCh37):g.49173829 | Deleção | <i>AC115618.1, AF196779.12, CACNA1F, CCDC120, CCDC22, EBP, ERAS, FOXP3, FTSJ1, GAGE10, GATA1, GLOD5, GPKOW, GRIPAP1, HDAC6, KCND1, MAGIX, OTUD5, PCSK1N, PIM2, PLP2, PORCN, PPP1R3F, PQBP1, PRAF2, PRICKLE3, RBM3, SLC35A2, SLC38A5, SSX4, SSX4B, SUV39H1, SYP, TBC1D25, TFE3, TIMM17B, WAS, WDR13, WDR45</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.49371400 | ChrX(GRCh37):g.52103225 | Deleção | <i>AKAP4, BMP15, CCNB3, CLCN5, CXorf67, GSPT2, MAGED1, MAGED4, MAGED4B, NUDT10, NUDT11, PAGE1, PAGE4, RP11-114H20.1, RP11-363G10.2, SHROOM4, USP27X</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.52892023 | ChrX(GRCh37):g.56793641 | Deleção | <i>AL353698.1, ALAS2, APEX2, FAM104B, FAM120C, FAM156A, FAM156B, FGD1, FOXR2, GNL3L, GPR173, HSD17B10, HUWE1, IQSEC2, ITIH6, KDM5C, KLF8, MAGED2, MAGEH1, MTRNR2L10, PAGE2, PAGE2B, PAGE3, PAGE5, PFKFB1, PHF8, RIBC1, RRAGB, SMC1A, TRO, TSPYL2, TSR2, UBQLN2, USP51, WNK3, XAGE3</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.61806734 | ChrX(GRCh37):g.67122710 | Deleção | <i>AMER1, AR, ARHGFE9, ASB12, EDA2R, HEPH, LAS1L, MSN, MTMR8, SPIN4, VSIG4, ZC3H12B, ZC4H2</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.67130396 | ChrX(GRCh37):g.70137081 | Deleção | <i>ARR3, AWAT1, AWAT2, DGAT2L6, DLG3, EDA, EFN1, FAM155B, GDDP2, IGBP1, KIF4A, OPHN1, OTUD6A, P2RY4, PDZD11, PJA1, RAB41, STARD8, TEX11, YIPF6</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.70140115 | ChrX(GRCh37):g.72063253 | Deleção | <i>ACRC, CITED1, CXCR3, CXorf65, ERCC6L, FOXO4, GJB1, HDAC8, IL2RG, ITGB1BP2, MED12, NHSL2, NLGN3, NONO, OGT, PHKA1, PIN4, RGAG4, RPS4X, SLC7A3, SNX12, TAF1, ZMYM3</i> | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 9 – CNVs detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial | Coordenada genómica final | Tipo de CNV | Genes implicados | Frequência | | Observações |
|-----------------------------|---------------------------|-------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| ChrX(GRCh37):g.72069656 | ChrX(GRCh37):g.78921417 | Deleção | <i>ABCB7, ATP7A, ATRX, CDX4, CHIC1, COX7B, CYSLTR1, DMRTC1, FGF16, GPR174, ITM2A, KIAA2022, LPAR4, MAGEE1, MAGEE2, MAGT1, NAP1L2, NAP1L6, P2RY10, PABPC1L2A, PABPC1L2B, PBDC1, PGAM4, PGK1, RLIM, SLC16A2, TAF9B, UPRT, ZCCHC13, ZCCHC5, ZDHHC15</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.78926829 | ChrX(GRCh37):g.81981818 | Deleção | <i>BRWD3, FAM46D, HMGN5, SH3BGR1, TBX22</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.81988678 | ChrX(GRCh37):g.93860715 | Deleção | <i>APOOL, CHM, CPXCR1, CYLC1, DACH2, FAM133A, HDX, KHLHL4, NAP1L3, PABPC5, PCDH11X, POF1B, POU3F4, RP56KA6, SATL1, TGIF2LX, ZNF711</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.93864808 | ChrX(GRCh37):g.114542151 | Deleção | <i>ACSL4, ALG13, AMMECR1, AMOT, ARL13A, ARMCX1, ARMCX2, ARMCX3, ARMCX4, ARMCX5, ARMCX6, ATG4A, BEX1, BEX2, BEX4, BEX5, BHLHB9, BTK, CAPN6, CENPI, CHRDL1, CLDN2, COL4A5, COL4A6, CSTF2, CXorf57, DCX, DIAPH2, DRP2, ESX1, FAM199X, FRMPD3, GLA, GLRA4, GNG5P2, GPRASP1, GPRASP2, GUCY2F, H2BFM, H2BFWT, HNRNPH2, HTR2C, IL13RA2, IL1RAPL2, IRS4, KCN E1L, LHFPL1, LRCH2, LUZP4, MID2, MORC4, MORF4L2, MUM1L1, NCBP2L, NGFRAP1, NOX1, NRK, NUP62CL, NXF2, NXF2B, NXF3, NXF5, NXT2, PAK3, PCDH19, PIH1D3, PLP1, PRPS1, PSMD10, RAB40A, RAB40AL, RAB9B, RBM41, RBMXL3, RGAG1, RIPPLY1, RNF128, RP6-24A23.6, RPA4, RPL36A, RPL36A-HNRNPH2, SERPINA7, SLC25A53, SRPX2, SYTL4, TAF7L, TBC1D8B, TCEAL1, TCEAL2, TCEAL3, TCEAL4, TCEAL5, TCEAL6, TCEAL7, TCEAL8, TCP11X1, TCP11X2, TEX13A, TEX13B, TIMM8A, TMEM164, TMEM31, TMEM35, TMSB15A, TMSB15B, TNMD, TRMT2B, TRPC5, TSC22D3, TSPAN6, VSIG1, WBP5, XKRX, ZCCHC16, ZCCHC18, ZMAT1</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.115005873 | ChrX(GRCh37):g.118873175 | Deleção | <i>AGTR2, CXorf56, CXorf61, DOCK11, IL13RA1, KIAA1210, KLHL13, LONRF3, NKRF, PGRMC1, SEPT6, SLC25A43, SLC25A5, SLC6A14, UBE2A, WDR44, ZCCHC12</i> | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 9 – CNVs detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial | Coordenada genómica final | Tipo de CNV | Genes implicados | Frequência | | Observações |
|-----------------------------|---------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| ChrX(GRCh37):g.118876861 | ChrX(GRCh37):g.121303936 | Deleção | <i>AKAP14,ATP1B4,C1GALT1C1,CT47A1,CT47A10,CT47A11,CT47A12,CT47A2,CT47A3,CT47A4,CT47A5,CT47A6,CT47A7,CT47A8,CT47A9,CT47B1,CUL4B,GLUD2,LAMP2,MCTS1,NDUFA1,NKAP,RHOXF1,RHOXF2,RHOXF2B,RNF113A,RPL39,SOWAHD,TMEM255A,UPF3B</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.121316138 | ChrX(GRCh37):g.128500326 | Deleção | <i>ACTRT1,CXorf64,DCAF12L1,DCAF12L2,GRIA3,SH2D1A,STAG2,TENM1,THOC2,XIAP</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.128502341 | ChrX(GRCh37):g.131939262 | Deleção | <i>AIFM1,APLN,ARHGAP36,BCORL1,ELF4,ENOX2,FRMD7,GPR119,HS6ST2,IGSF1,MBNL3,MST4,OCRL,OR13H1,RAB33A,RAP2C,RBMX2,SASH3,SLC25A14,SMARCA1,UTP14A,XPNPEP2,ZDHC9,ZNF280C</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.131942286 | ChrX(GRCh37):g.134852128 | Deleção | <i>CCDC160,CXorf48,DDX26B,FAM122B,FAM122C,FAM127A,FAM127B,FAM127C,GPC3,GPC4,HPRT1,HS6ST2,MOSPD1,PHF6,PLAC1,SMIM10,TFDP3,USP26,ZNF449,ZNF75D</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.134867099 | ChrX(GRCh37):g.135299251 | Deleção | <i>CT45A2,CT45A3,CT45A4,CT45A5,CT45A6,FHL1,MMGT1,SAGE1,SLC9A6</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.135302717 | ChrX(GRCh37):g.139094417 | Deleção | <i>ARHGEF6,ATP11C,BRS3,CD40LG,CXorf66,F9,FGF13,GPR101,GPR112,HTATSF1,MAP7D3,MCF2,RBMX,VGLL1,ZIC3</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.139096550 | ChrX(GRCh37):g.140082137 | Deleção | <i>CDR1,LINC00632,SOX3</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.146369340 | ChrX(GRCh37):g.148633825 | Deleção | <i>AFF2,CXorf40A,FMR1,FMR1NB,IDS</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.150295881 | ChrX(GRCh37):g.151653980 | Deleção | <i>CNGA2,FATE1,GABRA3,GABRE,GPR50,MAGEA10,MAGEA4,PASD1,PRRG3,VMA21</i> | 0 | 0 | a) |
| ChrX(GRCh37):g.151865696 | ChrX(GRCh37):g.152229366 | Deleção | <i>CETN2,CSAG1,MAGEA12,MAGEA2,MAGEA2B,MAGEA3,MAGEA6,NSDHL,PNMA3,PNMA3,ZNF185</i> | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 9 – CNVs detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial | Coordenada genómica final | Tipo de CNV | Genes implicados | Frequência | | Observações |
|-----------------------------|---------------------------|-------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| ChrX(GRCh37):g.152349585 | ChrX(GRCh37):g.154572647 | Deleção | ABCD1, ARHGAP4, ATP2B3, ATP6AP1, AVPR2, BCAP31, BGN, BRCC3, CLIC2, CMC4, CTAG1A, CTAG1B, CTAG2, DKC1, DNASE1L1, DUSP9, EMD, F8, F8A1, FAM3A, FAM50A, FAM58A, FLNA, FUNDC2, G6PD, GAB3, GDI1, H2AFB1, HAUS7, HCF1, IDH3G, IKBKG, IRAK1, L1CAM, LAGE3, LCA10, MAGEA1, MECP2, MPP1, MTCP1, NAA10, OPN1LW, OPN1MW, OPN1MW2, PDZD4, PLXNA3, PLXNB3, PNCK, RAB39B, RENBP, RPL10, SLC10A3, SLC6A8, SMIM9, SRPK3, SSR4, TAZ, TEX28, TEX28P1, TEX28P2, TKTL1, TME118, TREX2, U82695.9, UBL4A, VBP1, ZFP92, ZNF275 | 0 | 0 | a) |
| ChrX(GRCh37):g.154588727 | ChrX(GRCh37):g.154789553 | Deleção | F8A2, F8A3, H2AFB2, H2AFB3, TMLHE | 0 | 0 | a) |
| ChrX(GRCh37):g.154790554 | ChrX(GRCh37):g.154797172 | Deleção | TMLHE-AS1 | 0 | 0 | a) |
| ChrX(GRCh37):g.154803752 | ChrX(GRCh37):g.154917276 | Deleção | TMLHE-AS1 | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 10 – CNVs finais selecionadas pela análise dos dados de sequenciação do genoma.

| Coordenada genómica inicial | Coordenada genómica final | Tipo de CNV | Genes | Componente do gene (ponto de quebra) | Gene associado a doença | |
|-----------------------------|---------------------------|-------------|-------|--------------------------------------|---------------------------|-------|
| | | | | | OMIM | SysID |
| Chr7(GRCh37):g.152074968 | Chr7(GRCh37):g.152081131 | Duplicação | KMT2C | intrónico | Kleefstra syndrome 2 – AD | Sim |
| Chr7(GRCh37):g.152101032 | Chr7(GRCh37):g.152113665 | Duplicação | KMT2C | intrónico | Kleefstra syndrome 2 – AD | Sim |

Tabela 11 – Inversões detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência).

| Coordenada genómica inicial (ponto de quebra 1) | Coordenada genómica inicial (ponto de quebra 2) | Genes implicados | Componente do gene (ponto de quebra) | Frequência | | Observações |
|-------------------------------------------------|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| Chr1(GRCh37):g.11047461 | Chr1(GRCh37):g.223352892 | A3GALT2,AADAACL3,AADAACL4,ABCA4,ABCD3,ABL2,AC004824.2,ACO93157.1,AC096644.1,AC096677.1,AC114494.1,AC119673.1,ACADM,ACBD6,ACOT11,ACP6,ACTL8,ADAM15,ADAM30,ADAMTS4,ADAMTSL4,ADAMTSL4-AS1,ADAR,ADC,ADCY10,ADIPOR1,ADORA1,ADORA3,ADPRHL2,AGBL4,AGLAGMAT | exónico | 0 | 0 | a) |
| Chr1(GRCh37):g.11047572 | Chr1(GRCh37):g.223352832 | A3GALT2,AADAACL3,AADAACL4,ABCA4,ABCD3,ABL2,AC004824.2,ACO93157.1,AC096644.1,AC096677.1,AC114494.1,AC119673.1,ACADM,ACBD6,ACOT11,ACP6,ACTL8,ADAM15,ADAM30,ADAMTS4,ADAMTSL4,ADAMTSL4- | exónico | 0 | 0 | a) |
| Chr1(GRCh37):g.35267697 | Chr1(GRCh37):g.54976380 | ADPRHL2,AGBL4,AGO1,AGO3,AGO4,AKIRIN1,AKR1A1,AL109659.1,AL161915.1,AL357673.1,AL645730.2,AL929472.1,ARTN,ATP6VOB,ATPAF1,B4GALT2,BEND5,BEST4,BMP8A,BMP8B,BTBD19,BTF3L4,C1orf109,C1orf122,C1orf123,C1orf185,C1orf210,C1orf216,C1orf228,C1orf50,CAP1,CC2D1B,CCDC163P,CCDC17,CCDC23,CCDC24,CCDC30,CDCC20,CDCA8,CDCP2,CDKN2C,CITED4,CLDN19,CLSPN,CMPK1,COL8A2,COL9A2,CPT2,CSF3R,CTPS1,CYB5RL,CYP4A11,CYP4A22,CYP4B1,CYP4X1,CYP4Z1,DIO1,DLGAP3 | exónico | 0 | 0 | a) |
| Chr1(GRCh37):g.35268049 | Chr1(GRCh37):g.54976339 | ADPRHL2,AGBL4,AGO1,AGO3,AGO4,AKIRIN1,AKR1A1,AL109659.1,AL161915.1,AL357673.1,AL645730.2,AL929472.1,ARTN,ATP6VOB,ATPAF1,B4GALT2,BEND5,BEST4,BMP8A,BMP8B,BTBD19,BTF3L4,C1orf109,C1orf122,C1orf123,C1orf185,CMPK1,COL8A2,COL9A2,CPT2,CSF3R,CTPS1,CYB5RL,CYP4A11,CYP4A22,CYP4B1,CYP4X1,CYP4Z1,DIO1,DLGAP3,DMAPI1,DMBX1,DMRTA2,DMRTB1,DNAL11,DPH2,EBNA1BP2,ECHDC2,EDN2,EFCAB14,EIF2B3,ELAVL4 | exónico | 0 | 0 | a) |
| Chr2(GRCh37):g.87449818 | Chr2(GRCh37):g.112500660 | AC012360.1,AC012360.2,AC012493.2,AC092675.3,AC108868.1,ACOXL,ACTR1B,ADRA2B,AFF3,ANKRD23,ANKRD36,ANKRD36C,ANKRD39,ARID5A,ASTL,BCL2L11,BUB1,C2ORF15,C2orf15,C2orf40,C2orf49,GKC,IGKJ1,IGKJ2,IGKJ3,IGKJ4,IGKJ5,IGKV1-12,IGKV1-16,IGKV1-17,IGKV1-27,IGKV1-5,IGKV1-6,IGKV1-8,IGKV1-9,IGKV1D-12,IGKV1D-13,IGKV1D-16,IGKV1D-17,IGKV1D-33 | exónico | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 11 – Inversões detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial (ponto de quebra 1) | Coordenada genómica inicial (ponto de quebra 2) | Genes implicados | Componente do gene (ponto de quebra) | Frequência | | Observações |
|-------------------------------------------------|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| Chr2(GRCh37):g.87449928 | Chr2(GRCh37):g.112500566 | AC012360.1,AC012360.2,AC012493.2,AC092675.3,AC108868.1,AC0XL,ACTR1B,ADRA2B,AFF3,ANKRD23,ANKRD36,ANKRD36C,ANKRD39,ARID5A,ASTL,BCL2L11,BUB1,C2ORF15,C2orf15,C2orf40,C2orf49,CCDC138,CH17-132F21.1,CHST10,CIA01,CNGA3,CNNM3,CNNM4,CNOT11,COA5,COX5B,CREG2,DUSP2,EDAR,EIF2AK3,EIF5B,FABP1,FAHD2A | exónico | 0 | 0 | a) |
| Chr2(GRCh37):g.130942995 | Chr2(GRCh37):g.132246845 | AC013269.5,AC140481.2,AMER3,ARHGEF4,CCDC115,CFC1,CFC1B,CYP4F31P,FAM168B,GPR148,IMP4,MZT2A,MZT2B,PLEKHB2,POTEE,POTEI,POTEJ,PTPN18,TUBA3D,TUBA3E | exónico | 0 | 0 | a) |
| Chr2(GRCh37):g.130943322 | Chr2(GRCh37):g.132246503 | AC013269.5,AC140481.2,AMER3,ARHGEF4,CCDC115,CFC1,CFC1B,CYP4F31P,FAM168B,GPR148,IMP4,MZT2A,MZT2B,PLEKHB2,POTEE,POTEI,POTEJ,PTPN18,TUBA3D,TUBA3E | exónico | 0 | 0 | a) |
| Chr3(GRCh37):g.72551426 | Chr3(GRCh37):g.189255466 | A4GNT,AADAC,AADACL2,ABCC5,ABCF3,ABHD10,ABI3BP,ABTB1,AC007919.2,AC022498.1,AC092964.1,AC092964.2,AC104472.1,AC107021.1,AC117395.1,ACAD11,ACAD9,ACKR4,ACPL2,ACPP,ACTL6A,ACTRT3,ADCY5,ADIPOQ,ADPRH,AGTR1,AHSG,ALCAM,ALDH1L1,ALG1L,ARGFX,ARHGAP31,ARHGEF26,ARL13B,ATP1B3,ATP2C1,ATP6V1A,ATR,B3GALNT1,B3GNT5,B4GALT4,BBX,BCHE,BCL6,BFSP2,BOC,BTLA,C3orf17,C3orf22,C3orf27,C3orf30,C3orf33,C3orf36,C3orf38,C3orf52,C3orf55,C3orf56,C3orf58,C3orf65,C3orf70,C3orf72,C3orf79, | exónico | 0 | 0 | a) |
| Chr3(GRCh37):g.72551561 | Chr3(GRCh37):g.189255439 | A4GNT,AADAC,AADACL2,ABCC5,ABCF3,ABHD10,ABI3BP,ABTB1,AC007919.2,AC022498.1,AC092964.1,AC092964.2,AC104472.1,AC107021.1,AC117395.1,ACAD11,ACAD9,ACKR4,ACPL2,ACPP,ACTL6A,ACTRT3,ADCY5,ADIPOQ,ADPRH,AGTR1,AHSG,ALCAM,ALDH1L1,ALG1L,ALG3,AMOTL2,ANAPC13,ANKUB1,AP2M1,ARGFX,ARHGAP31,ARHGEF26,ARL13B,ARL14,ARL6,ARMC8,ASTE1,ATG3,ATP11B,ATP1B3,ATP2C1,ATP6V1A,ATR,B3GALNT1,B3GNT5,B4GALT4,BBX,BCHE,BCL6,BFSP2,BOC,BTLA,C3orf17,C3orf22,C3orf27,C3orf30,C3orf33,C3orf36 | exónico | 0 | 0 | a) |
| Chr3(GRCh37):g.187860343 | Chr3(GRCh37):g.193080602 | AC022498.1,ATP13A5,CCDC50,CLDN1,CLDN16,FGF12,GMNC,HRASLS,IL1RAP,LEPREL1,LPP,MB21D2,OSTN,PYDC2,TMEM207,TP63,TPRG1,UTS | exónico | 0 | 0 | a) |
| Chr3(GRCh37):g.187860453 | Chr3(GRCh37):g.193080591 | AC022498.1,ATP13A5,CCDC50,CLDN1,CLDN16,FGF12,GMNC,HRASLS,IL1RAP,LEPREL1,LPP,MB21D2,OSTN,PYDC2,TMEM207,TP63,TPRG1,UTS | exónico | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 11 – Inversões detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial (ponto de quebra 1) | Coordenada genómica inicial (ponto de quebra 2) | Genes implicados | Componente do gene (ponto de quebra) | Frequência | | Observações |
|----------------------------------------------------|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| Chr5(GRCh37):g.175336068 | Chr5(GRCh37):g.177357141 | ARL10,B4GALT7,CDHR2,CLTB,DBN1,DDX41,DOK3,EIF4E1B,F12,FAF2,FAM153A,FAM153B,FAM193B,FGFR4,GPRIN1,GRK6,HIGD2A,HK3,KIAA1191,LMAN2,MXD3,NOP16,NSD1,PDLIM7,PFN3,PRELID1,PRR7,RAB24,RGS14,RNF44,RP11-1026M7.2,RP11-826N14.2,SIMC1,SLC34A1,SNCB,THOC3,TMED9,TSPAN17,UIMC1,UNC5A,ZNF346 | exónico | 0 | 0 | a) |
| Chr5(GRCh37):g.175336085 | Chr5(GRCh37):g.177357162 | ARL10,B4GALT7,CDHR2,CLTB,DBN1,DDX41,DOK3,EIF4E1B,F12,FAF2,FAM153A,FAM153B,FAM193B,FGFR4,GPRIN1,GRK6,HIGD2A,HK3,KIAA1191,LMAN2,MXD3,NOP16,NSD1,PDLIM7,PFN3,PRELID1,PRR7,RAB24,RGS14,RNF44,RP11-1026M7.2,RP11-826N14.2,SIMC1,SLC34A1,SNCB,THOC3,TMED9,TSPAN17,UIMC1,UNC5A,ZNF346 | exónico | 0 | 0 | a) |
| Chr6(GRCh37):g.25709140 | Chr6(GRCh37):g.32667577 | ABCF1,ABHD16A,ABT1,AGER,AGPAT1,AIF1,AL645922.1,APOM,ATAT1,ATF6B,ATP6V1G2,ATP6V1G2- DDX39B,BAG6,BTN1A1,BTN2A1,BTN2A2,BTN3A1,BTN3A2,BTN3A3,BTNL2,C2,C4A,C4B,C6orf10,C6orf100,C6orf136,C6orf15,C6orf25,C6orf47,C6orf48,CCHCR1,CDSN,CFB,CLIC1,CSNK2B | exónico | 0 | 0 | a) |
| Chr6(GRCh37):g.25709241 | Chr6(GRCh37):g.32667472 | ABCF1,ABHD16A,ABT1,AGER,AGPAT1,AIF1,AL645922.1,APOM,ATAT1,ATF6B,ATP6V1G2,ATP6V1G2- DDX39B,BAG6,BTN1A1,BTN2A1,BTN2A2,BTN3A1,BTN3A2,BTN3A3,BTNL2,C2,C4A,C4B,C6orf10,C6orf100,C6orf136,C6orf15,C6orf25,C6orf47,C6orf48,CCHCR1,CDSN,CFB,CLIC1,CSNK2B | exónico | 0 | 0 | a) |
| Chr13(GRCh37):g.43443282 | Chr13(GRCh37):g.51747989 | AL136218.1,ARL11,CAB39L,CCDC122,CDADC1,COG3,CPB2,CYSLTR2,DLEU1,DLEU7,DNAJC15,EBPL,ENOX1,EPST11,ESD,FAM194B,FNDCC3A,GPALPP1,GTF2F2,HTR2A,ITM2B,KCNRG,KCTD4,KIAA0226L,KPNA3,LACC1,LCPI,LPAR6,LRCH1,LRRC63,MED4,MLNR,NUDT15,NUFIP1,PHF11,RB1,RCBTB1,RCBTB2,RNASEH2B,SERP2,SETDB2,SIAH3,SLC25A30,SMIM2,SPERT,SPRYD7,SUCLA2,TPT1,TRIM13,TSC22D1,ZC3H13 | exónico | 0 | 0 | a) |
| Chr13(GRCh37):g.43443769 | Chr13(GRCh37):g.51747521 | AL136218.1,ARL11,CAB39L,CCDC122,CDADC1,COG3,CPB2,CYSLTR2,DLEU1,DLEU7,DNAJC15,EBPL,ENOX1,EPST11,ESD,FAM194B,FNDCC3A,GPALPP1,GTF2F2,HTR2A,ITM2B,KCNRG,KCTD4,KIAA0226L,KPNA3,LACC1,LCPI,LPAR6,LRCH1,LRRC63,MED4,MLNR,NUDT15,NUFIP1,PHF11,RB1,RCBTB1,RCBTB2,RNASEH2B,SERP2,SETDB2,SIAH3,SLC25A30,SMIM2,SPERT,SPRYD7,SUCLA2,TPT1,TRIM13,TSC22D1,ZC3H13 | exónico | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 11 – Inversões detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial (ponto de quebra 1) | Coordenada genómica inicial (ponto de quebra 2) | Genes implicados | Componente do gene (ponto de quebra) | Frequência | | Observações |
|----------------------------------------------------|----------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| Chr14(GRCh37):g.47117868 | Chr14(GRCh37):g.63226283 | ABHD12B,ACTR10,AL132989.1,AL139099.1,AL358333.1,AL391152.1,AL627171.1,AL627171.2,AP5M1,ARF6,ARID4A,ATG14,ATL1,ATP5S,BMP4,C14orf105,C14orf166,C14orf182,C14orf183,C14orf37,C14orf39,CCDC175,CDKL1,CDKN3,CGRRF1,CNIH1,DAAM1,DACT1,DDHD1,DHRS7,DLGAP5,DNAAF2,ERO1L,EXOC5,FBXO34,FERMT2,FRMD6,GCH1,GMFB,GNG2,GNPNA1,GPR135,GPR137C,HIF1A,JKAMP,KCNH5,KIAA0586,KLHDC1,KLHDC2,KTN1,L2HGDH,L3HYPDH,LGALS3,LRR1,LRRC9 | exónico | 0 | 0 | a) |
| Chr14(GRCh37):g.47118177 | Chr14(GRCh37):g.63226297 | ABHD12B,ACTR10,AL132989.1,AL139099.1,AL358333.1,AL391152.1,AL627171.1,AL627171.2,AP5M1,ARF6,ARID4A,ATG14,ATL1,ATP5S,BMP4,C14orf105,C14orf166,C14orf182,C14orf183,C14orf37,C14orf39,CCDC175,CDKL1,CDKN3,CGRRF1,CNIH1,DAAM1,DACT1,DDHD1,DHRS7,DLGAP5,DNAAF2,ERO1L,EXOC5,FBXO34,FERMT2,FRMD6,GCH1,GMFB | exónico | 0 | 0 | a) |
| Chr15(GRCh37):g.56962090 | Chr15(GRCh37):g.95007473 | AAGAB,ABHD17C,ABHD2,ACO19294.1,ACO27228.1,ACO69368.3,AC105020.1,AC112693.2,ACAN,ACSBG1,ADAM10,ADAMTS7,ADAMTSL3,ADPGK,AEN,AGBL1,AKAP13,ALDH1A2,ALPK3,ANKDD1A,ANKRD34C,ANP32A,ANPEP,ANXA2,AP3B2,AP3S2,APH1B,AQP9,ARID3B,ARIH1,ARNT2,BBS4,BCL2A1,BLM,BNC1,BNIP2,BTBD1,C15ORF31,C15ORF37,C15orf26,C15orf27,C15orf32,C15orf37,C15orf38 | exónico | 0 | 0 | a) |
| Chr15(GRCh37):g.56962221 | Chr15(GRCh37):g.95007495 | AAGAB,ABHD17C,ABHD2,ACO19294.1,ACO27228.1,ACO69368.3,AC105020.1,AC112693.2,ACAN,ACSBG1,ADAM10,ADAMTS7,ADAMTSL3,ADPGK,AEN,AGBL1,AKAP13,ALDH1A2,ALPK3,ANKDD1A,ANKRD34C,ANP32A,ANPEP,ANXA2,AP3B2,AP3S2,APH1B,AQP9,ARID3B,ARIH1,ARNT2,BBS4,BCL2A1,BLM,BNC1,BNIP2,BTBD1,C15ORF31,C15ORF37,C15orf26,C15orf27,C15orf32,C15orf37,C15orf38 | exónico | 0 | 0 | a) |
| Chr15(GRCh37):g.100338092 | Chr15(GRCh37):g.102306884 | ADAMTS17,ALDH1A3,ASB7,CERS3,CHSY1,CTD-2054N24.2,LINS,LRRK1,PCSK6,RP11-89K11.1,SNRPA1,TARSL2,TM2D3,VIMP | exónico | 0 | 0 | a) |
| Chr15(GRCh37):g.100338376 | Chr15(GRCh37):g.102306616 | ADAMTS17,ALDH1A3,ASB7,CERS3,CHSY1,CTD-2054N24.2,LINS,LRRK1,PCSK6,RP11-89K11.1,SNRPA1,TARSL2,TM2D3,VIMP | exónico | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 11 – Inversões detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência) (continuação).

| Coordenada genómica inicial (ponto de quebra 1) | Coordenada genómica inicial (ponto de quebra 2) | Genes implicados | Componente do gene (ponto de quebra) | Frequência | | Observações |
|-------------------------------------------------|-------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------------------------------|------------|-----|-------------|
| | | | | DECIPHER | DGV | |
| Chr16(GRCh37):g.29238201 | Chr16(GRCh37):g.88226535 | AARS,ABCC11,ABCC12,AC002310.13,AC009060.1,AC010536.1,AC010547.9,AC025287.1,AC026461.1,AC040160.1,AC132186.1,AC135048.1,AC142381.1,ACD,ADAD2,ADAMTS18,ADAT1,ADCY7,AGRP,AHSP,AKTIP,ALDOA,AMFR,APIG1,ARL2BP,ARMCS,ASPHD1,ATMIN,ATP2C2,ATP6VOD1,ATXN1L,B3GNT9,BANP,BBS2,BCAR1,BCKDK,BCL7C,BCMO1,BEAN1,BOLA2,BOLA2B,BRD7,C16orf46,C16orf47,C16orf54,C16orf58,C16orf70,C16orf74,C16orf78,C16orf80,C16orf86,C16orf87,C16orf92,C16orf93,C16orf95,C16orf97,C16orf98,CA5A | exónico | 0 | 0 | a) |
| Chr16(GRCh37):g.29238528 | Chr16(GRCh37):g.88226516 | AARS,ABCC11,ABCC12,AC002310.13,AC009060.1,AC010536.1,AC010547.9,AC025287.1,AC026461.1,AC040160.1,AC132186.1,AC135048.1,AC142381.1,ACD,ADAD2,ADAMTS18,ADAT1,ADCY7,AGRP,AHSP,AKTIP,ALDOA,AMFR,APIG1,ARL2BP,ARMCS,ASPHD1,ATMIN,ATP2C2,ATP6VOD1,ATXN1L,BOLA2B,BRD7,C16orf46,C16orf47,C16orf54,C16orf58,C16orf70,C16orf74,C16orf78,C16orf80,C16orf86,C16orf87,C16orf92,C16orf93,C16orf95,C16orf97,C16orf98,CA5A,CA7,CALB2,CAPNS2,CBFB,CBLN1,CCDC102A,CCDC113 | exónico | 0 | 0 | a) |
| ChrX(GRCh37):g.11608892 | ChrX(GRCh37):g.40416853 | ACE2,ACOT9,AP1S2,APOO,ARHGAP6,ARX,ASB11,ASB9,ATXN3L,BCOR,BEND2,BMX,CA5B,CDKL5,CHDC2,CNKSR2,CTPS2,CXorf21,CXorf22,CXorf23,CXorf27,CXorf30,CXorf58,CYBB,DCAF8L1,DCAF8L2,DDX53,DMD,DYNLT3,EGFL6,EIF1AX,EIF2S3,FAM47A,FAM47B,FAM47C,FAM9C,FANCB,FIGF,FRMPD4,FTHL17,GEMIN8,GK,GLRA2,GPM6B,GPR64,GRPR,IL1RAPL1,KLHL15,KLHL34,LANCL3,MAGEB1,MAGEB10,MAGEB16,MAGEB17,MAGEB18,MAGEB2,MAGEB3,MAGEB4,MAGEB5,MAGEB6 | exónico | 0 | 0 | a) |
| ChrX(GRCh37):g.11609036 | ChrX(GRCh37):g.40416621 | ACE2,ACOT9,AP1S2,APOO,ARHGAP6,ARX,ASB11,ASB9,ATXN3L,BCOR,BEND2,BMX,CA5B,CDKL5,CHDC2,CNKSR2,CTPS2,CXorf21,CXorf22,CXorf23,CXorf27,CXorf30,CXorf58,CYBB,DCAF8L1,DCAF8L2,DDX53,DMD,DYNLT3,EGFL6,EIF1AX,EIF2S3,FAM47A,FAM47B,FAM47C,FAM9C,FANCB,FIGF,FRMPD4,FTHL17,GEMIN8,GK,GLRA2,GPM6B,GPR64,GRPR,IL1RAPL1,KLHL15,KLHL34,LANCL3,MAGEB1,MAGEB10,MAGEB16,MAGEB17 | exónico | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 12 – Translocações detetadas pela análise dos dados de sequenciação do genoma (filtragem de acordo com a sua frequência).

| | coordenada genómica inicial (ponto de quebra 1) hg19 | coordenada genómica inicial (ponto de quebra 2) hg19 | Genes | Componente do gene | Frequência | | Observações |
|----------|------------------------------------------------------------|------------------------------------------------------------|--------------------------------------------------|-----------------------|------------|-----|-------------|
| | | | | | DECIPHER | DGV | |
| t(2/5) | Chr2(GRCh37):g.69323415 | Chr2(GRCh37):g.69323415 | <i>ANTXR1</i> | intrónico | 0 | 0 | a) |
| | Chr2(GRCh37):g.69323539 | Chr2(GRCh37):g.69323539 | <i>ANTXR1</i> | intrónico | 0 | 0 | a) |
| | Chr5(GRCh37):g.109945937 | Chr5(GRCh37):g.109945937 | <i>TMEM232</i> | intrónico | 0 | 0 | a) |
| | Chr5(GRCh37):g.109945957 | Chr5(GRCh37):g.109945957 | <i>TMEM232</i> | intrónico | 0 | 0 | a) |
| t(2/14) | Chr2(GRCh37):g.75381527 | Chr2(GRCh37):g.75381527 | <i>TACR1</i> | intrónico | 0 | 0 | - |
| | Chr2(GRCh37):g.75381615 | Chr2(GRCh37):g.75381615 | <i>TACR1</i> | intrónico | 0 | 0 | - |
| | Chr14(GRCh37):g.66077921 | Chr14(GRCh37):g.66077921 | <i>FUT8</i> | intrónico | 0 | 0 | - |
| | Chr14(GRCh37):g.66078019 | Chr14(GRCh37):g.66078019 | <i>FUT8</i> | intrónico | 0 | 0 | - |
| t(4/8) | Chr4(GRCh37):g.183256370 | Chr4(GRCh37):g.183256370 | <i>TENM3</i> | intrónico | 0 | 0 | a) |
| | Chr4(GRCh37):g.183256669 | Chr4(GRCh37):g.183256669 | <i>TENM3</i> | intrónico | 0 | 0 | a) |
| | Chr8(GRCh37):g.70935933 | Chr8(GRCh37):g.70935933 | <i>AC090574.1(dist=85298),PRDM14(dist=27953)</i> | intergénico | 0 | 0 | a) |
| | Chr8(GRCh37):g.70935937 | Chr8(GRCh37):g.70935937 | <i>AC090574.1(dist=85302),PRDM14(dist=27949)</i> | intergénico | 0 | 0 | a) |
| t(6/20) | Chr6(GRCh37):g.124844006 | Chr6(GRCh37):g.124844006 | <i>NKAIN2</i> | intrónico | 0 | 0 | a) |
| | Chr6(GRCh37):g.124844308 | Chr6(GRCh37):g.124844308 | <i>NKAIN2</i> | intrónico | 0 | 0 | a) |
| | Chr20(GRCh37):g.8443470 | Chr20(GRCh37):g.8443470 | <i>PLCB1</i> | intrónico | 0 | 0 | a) |
| | Chr20(GRCh37):g.8443485 | Chr20(GRCh37):g.8443485 | <i>PLCB1</i> | intrónico | 0 | 0 | a) |
| t(9/16) | Chr9(GRCh37):g.117857652 | Chr9(GRCh37):g.117857652 | <i>TNC</i> | intrónico | 0 | 0 | a) |
| | Chr9(GRCh37):g.117857876 | Chr9(GRCh37):g.117857876 | <i>TNC</i> | intrónico | 0 | 0 | a) |
| | Chr16(GRCh37):g.6306122 | Chr16(GRCh37):g.6306122 | <i>RP11-420N3.2</i> | intrónico | 0 | 0 | a) |
| | Chr16(GRCh37):g.6306215 | Chr16(GRCh37):g.6306215 | <i>RP11-420N3.2</i> | intrónico | 0 | 0 | a) |
| t(13/19) | Chr13(GRCh37):g.110874838 | Chr13(GRCh37):g.110874838 | <i>COL4A1</i> | intrónico | 0 | 0 | a) |
| | Chr13(GRCh37):g.110875145 | Chr13(GRCh37):g.110875145 | <i>COL4A1</i> | intrónico | 0 | 0 | a) |
| | Chr19(GRCh37):g.54097233 | Chr19(GRCh37):g.54097233 | <i>ZNF331(dist=13710),CTB-167G5.5(dist=5649)</i> | intergénico | 0 | 0 | a) |
| | Chr19(GRCh37):g.54097248 | Chr19(GRCh37):g.54097248 | <i>ZNF331(dist=13725),CTB-167G5.5(dist=5634)</i> | intergénico | 0 | 0 | a) |
| t(19/X) | Chr19(GRCh37):g.13655136 | Chr19(GRCh37):g.13655136 | <i>CACNA1A</i> | intrónico | 0 | 0 | a) |
| | Chr19(GRCh37):g.13655224 | Chr19(GRCh37):g.13655224 | <i>CACNA1A</i> | intrónico | 0 | 0 | a) |
| | ChrX(GRCh37):g.86489068 | ChrX(GRCh37):g.86489068 | <i>SNORD45(dist=87261),Z95400.1(dist=251814)</i> | intergénico | 0 | 0 | a) |
| | ChrX(GRCh37):g.86489354 | ChrX(GRCh37):g.86489354 | <i>SNORD45(dist=87547),Z95400.1(dist=251528)</i> | intergénico | 0 | 0 | a) |

a) Não validado pela análise manual dos ficheiros BAM

Tabela 13 – Translocação final detetada pela análise dos dados de sequenciação do genoma.

| | Coordenada genómica inicial (ponto de quebra 1) | Coordenada genómica inicial (ponto de quebra 2) | Genes | Componente do gene | Frequência | | Gene associado a doença | |
|--------|----------------------------------------------------|----------------------------------------------------|--------------|-----------------------|------------|-----|------------------------------------|-------|
| | | | | | DECIPHER | DGV | OMIM | SysID |
| (2/14) | Chr2(GRCh37):g.75381527 | Chr2(GRCh37):g.75381527 | <i>TACR1</i> | intrónico | 0 | 0 | Sem fenótipo associado | Não |
| | Chr2(GRCh37):g.75381615 | Chr2(GRCh37):g.75381615 | <i>TACR1</i> | intrónico | 0 | 0 | Sem fenótipo associado | Não |
| | Chr14(GRCh37):g.66077921 | Chr14(GRCh37):g.66077921 | <i>FUT8</i> | intrónico | 0 | 0 | Desordem congênita de glicosilação | Sim |
| | Chr14(GRCh37):g.66078019 | Chr14(GRCh37):g.66078019 | <i>FUT8</i> | intrónico | 0 | 0 | Desordem congênita de glicosilação | Sim |

4. Discussão

A MPS gera grandes quantidades de informação sobre variantes genéticas, e como tal é necessária uma filtragem criteriosa para identificação das variantes que poderão estar implicadas no fenótipo dos doentes. A MPS pode ser utilizada para sequenciar genomas inteiros, ou então, o exoma completo (46). A sequenciação do exoma e do genoma apresentou resultados diferentes, como seria de esperar, nomeadamente no que diz respeito ao número e tipo de variantes detetadas, complexidade na análise e interpretação dos dados. A sequenciação do exoma mostrou ser capaz de detetar mutações pontuais e pequenas deleções e inserções com sucesso, e a sequenciação do genoma revelou capacidade adicional para identificar variantes do número de cópias e anomalias cromossómicas estruturais, envolvendo regiões codificantes e não codificantes. O facto da sequenciação do genoma detetar um maior número de variantes faz com que aumente a complexidade da análise e interpretação dos dados obtidos, nomeadamente no que diz respeito às variantes em regiões não codificantes, pouco estudadas e das quais ainda não se sabe a sua função (16,67). Para além disso, com a sequenciação do genoma é possível obter uma cobertura mais uniforme de todas as regiões genómicas, o que constitui uma vantagem. A cobertura pode ser definida como a percentagem da região alvo, ou de todo o genoma que é sequenciado a uma profundidade predefinida, ou seja, o número de vezes que um determinado nucleótido é sequenciado. Um estudo publicado em 2017 por *Meytse* e seus colaboradores (16), demonstrou que apesar da GS geralmente apresentar uma cobertura inferior ou idêntica à cobertura do ES, deteta variantes com melhor qualidade. Segundo os autores, esta diferença relativa à deteção das variantes poderá estar relacionada com uma cobertura mais uniforme do genoma. Neste estudo, a GS apresentou uma cobertura superior à do ES. A cobertura é importante principalmente para a deteção de CNVs e SCVs, uma vez que estes na sua maioria não se iniciam nem terminam em exões. A uniformidade da cobertura na ES ainda é claramente uma limitação (16). Apesar da sequenciação do genoma apresentar mais erros nas suas leituras, como possui uma cobertura mais homogénea, na análise final os erros são diminuídos (68). Este facto pode contribuir para a diferença de variantes de sequência detetadas em regiões codificantes/regiões intrónicas flangeadoras pelas duas abordagens (menos 12 316 variantes na análise do genoma). De facto, um estudo publicado em 2015 por *Aziz Belkadi* e seus colaboradores (68) demonstrou que o número de variantes de sequência falso-positivas era de 78 % para ES e de 17 % para GS (68). Após o processo de filtragem descrito anteriormente, apenas uma variante em regiões codificantes foi considerada por ambas as metodologias: NM_022726.3:c.646C>T, p.(Arg216*), no gene *ELOVL4* (OMIM *605512). Este gene está associado a ictiose, tetraplegia espática e défice intelectual, de transmissão autossómica recessiva (OMIM #614457, (69)), à ataxia espinocerebelar 34 (OMIM #133190, (70)) e à doença de Stargardt 3 (OMIM #600110, (71)), ambas de transmissão autossómica dominante. É, portanto, um gene associado a grande variabilidade fenotípica, onde se inclui o défice intelectual e anomalias cerebrais estruturais (associadas às duas primeiras doenças). A variante NM_022726.3:c.646C>T está descrita na base de

dados ClinVar como patogénica (RCV000023451.4), e foi reportada em homozigotia num doente com ictiose, tetraplegia espática e défice intelectual, filho de um casal consanguíneo (69). Relativamente à ataxia espinocerebelar 34, apesar de ser uma doença com padrão de hereditariedade autossómico dominante, está também descrito em casos de penetrância incompleta ((70)). Por esta razão esta variante foi considerada, apesar de ter origem paterna, e do fenótipo não ser completamente sobreponível com os doentes descritos: défice intelectual e anomalias cerebrais estruturais em comum, mas sem ataxia e problemas cutâneos, entre outras. A proteína codificada pelo gene *ELOVL4* participa no sistema de alongamento de ácidos gordos (72). Apesar desta doente ter valores normais destes metabolitos no plasma, é difícil prever a atividade da proteína no cérebro. Este facto é ainda mais pertinente devido à possibilidade de penetrância incompleta da variante em questão. A penetrância incompleta, pode ainda explicar as diferenças encontradas nas anomalias cerebrais estruturais apresentadas pelos três indivíduos afetados.

Com a evolução das técnicas de sequenciação, e com a implementação da sequenciação do genoma, o problema deixa de ser a identificação das variantes e passa a ser a avaliação da sua implicação na saúde humana. Este facto prende-se com a falta de conhecimento da função de determinados genes e de certas regiões do genoma, nomeadamente das regiões não codificantes. Não obstante, existem bases de dados, onde se destaca a ENCODE, que descrevem regiões não codificantes que têm funções de regulação génica ou implicação nos processos de *splicing* (65). Assim, a filtragem destas variantes foi limitada àquelas localizadas nessas regiões descritas na base de dados ENCODE. Desta filtragem, foram selecionadas (i) 5 variantes intrónicas no gene *DPYD* (OMIM *612779), associado a deficiência de dihidropirimidina de transmissão autossómica recessiva (OMIM # 274270, (73)); (ii) 3 variantes intrónicas *de novo* no gene *MYT1L* (OMIM *613084) associado a défice intelectual autossómico dominante 39 (OMIM #616521, (74)); (iii) 3 variantes intrónicas *de novo* no gene *BCL11A* (OMIM *606557), associado ao síndrome Dias-Logan (OMIM #617101, (75)); (iv) 3 variantes intrónicas no gene *TUSC3* (OMIM * 601385), associado a défice intelectual autossómico recessivo 7 (OMIM #611093, (76)); (v) 2 variantes intrónicas no gene *ZC3H14* (OMIM * 613279), associado a défice intelectual autossómico recessivo 56 (OMIM #617125, (77)); (vi) 1 variante *de novo* no gene *RORA* (OMIM *600825), associado a défice intelectual com ou sem epilepsia ou ataxia cerebelar de transmissão autossómica dominante (OMIM #618060, (78)); (vii) 1 variante *de novo* no gene *SETBP1* (OMIM *611060), associado a Síndrome de retração da face média de Schinzel-Giedion (OMIM #269150, (79)) e défice intelectual autossómico dominante 29 (OMIM #616078, (80)). Os fenótipos associados aos síndromes referidos incluem défice intelectual e/ou alterações cerebrais estruturais, considerando-se assim haver alguma sobreposição com o fenótipo apresentado pelos indivíduos afetados desta família.

De notar que não foi priorizada nenhuma variante de sequência em homozigotia, o que é expectável dado tratar-se de uma família não consanguínea.

Relativamente à análise das CNVs, apenas foram apresentadas as detetadas pela análise do genoma, uma vez que, na análise do exoma apenas foi detetada uma variante a qual não foi validada pela análise manual dos ficheiros BAM. Esta disparidade vai de encontro ao que está descrito na literatura, onde se demonstra que a análise do exoma tem menos capacidade para detetar CNVs do que a análise do genoma (16,68). Assim, do processo de filtragem das CNVs foram priorizadas duas possíveis duplicações, ambas intrónicas, no gene *KMT2C* (OMIM *606833), associado à síndrome de *Kleefstra 2* de transmissão autossómica dominante (OMIM #617768, (81)). Estas duplicações abrangem 6163 e 12633 bases, e como tal não foram detetadas pela análise de aCGH efetuada anteriormente (resolução média de 200Kb, (82)). Apesar de não existirem doentes reportados com CNVs semelhantes na base de dados DECIPHER, estas duplicações foram consideradas por envolverem um gene implicado em défice intelectual.

As anomalias cromossómicas estruturais são muito raras e a sua deteção pela MPS ainda permanece pouco desenvolvida, o que faz com que a sua análise e interpretação seja um grande desafio. Estas variantes apenas podem ser detetadas pela sequenciação do genoma, através de leituras emparelhadas que abrangem pontos de interrupção, vulgarmente designados por “*breakpoints*”, que diferem do genoma de referência. As inversões abrangem apenas um cromossoma possuindo dois pontos de quebra, um inicial e um final. Por outro lado, as translocações envolvem a troca de material geónimo entre dois cromossomas diferentes e por esse motivo possuem quatro pontos de quebra, ou seja, quatro *breakpoints* (83). Como já foi referido, do processo de filtragem destas variantes resultou apenas uma translocação, e nenhuma inversão. A translocação detetada envolve dois fragmentos intrónicos (i) de 88 pares de bases do gene *TACR1* (OMIM *162323), localizado em 2p12, e (ii) de 98 pares de bases do gene *FUT8* (OMIM *602589), localizado em 14q23.3. Apesar do gene *TACR1* não estar relacionado com qualquer doença, o gene *FUT8* está associado a uma doença congénita de glicosilação com fucosilação defeituosa 1 (OMIM #602589), sendo o défice intelectual um dos sinais clínicos associados a esta patologia (84). Esta translocação não foi detetada na análise do cariótipo realizada anteriormente, já que este tem uma resolução média de 5Mb (85).

Finalizado o processo de filtragem e de identificação de variantes possivelmente patogénicas, será necessário comprovar que estas se encontram presentes no DNA do caso *index* e verificar se segregam com o fenótipo, isto é, se também estão presentes no DNA dos outros irmãos afetados. Para tal, devem ser utilizadas outras técnicas como: (i) sequenciação de *Sanger* para o estudo de variantes de sequência, e estudo dos *breakpoints* das translocações e (ii) PCR em tempo real genómico quantitativo para confirmação das CNVs. Apenas as variantes que segregarem com o fenótipo serão consideradas para estudos posteriores.

Apesar das análises *in silico* das variantes de sequência e das CNVs que envolvem regiões não codificantes não preverem alteração do processo de *splicing*, devem ser efetuados estudos de transcritos (mRNA) dos genes implicados, através da sequenciação do DNA complementar. Adicionalmente, deve ser

avaliada a implicação destas variantes na expressão génica, por exemplo através de PCR em tempo real quantitativo. O mesmo procedimento deve ser aplicado ao gene *FUT8* para avaliar se a sua interrupção intrónica, causada pela translocação identificada, tem implicações no processo de *splicing* e/ou na expressão do gene. Para os estudos de transcritos, sugere-se a utilização de fibroblastos uma vez que, na impossibilidade de realização de biópsia cerebral, a pele apresenta uma origem embrionária mais próxima do cérebro do que os linfócitos (86).

Após todos os estudos necessários, no caso de nenhuma das variantes ser considerada patogénica, será indispensável: (i) analisar as variantes sinónimas, que foram excluídas na filtragem, uma vez que podem influenciar o processo de *splicing* (85); (ii) analisar genes não associados a doença ou com função desconhecida; (iii) analisar regiões não codificantes de função ainda desconhecida e não apenas as que estão descritas como reguladoras na base de dados ENCODE e (iv) analisar outras CNVs e SCVs e não apenas as que envolvem genes implicados em DI.

5. Conclusão

O MPS revelou ser um método preciso, rápido e eficiente que permite sequenciar vários genes em simultâneo. A sequenciação do genoma revelou ser uma abordagem mais completa do que a sequenciação do exoma, uma vez que permite a deteção de um maior número de variantes, incluindo em regiões não codificantes, CNVs e SCVs, e com menos erros associados. Apesar dos desafios inerentes a esta tecnologia, nomeadamente relacionadas com o armazenamento e interpretação de dados, a comunidade científica defende que esta técnica venha a ocupar um lugar na primeira linha de diagnóstico genético, nomeadamente em doenças com elevada heterogeneidade fenotípica e genotípica como o défice intelectual. Ainda assim, há muitas variantes que este método não deteta, como por exemplo expansões de repetições nucleotídicas, e a tecnologia para a identificação de todos os tipos de variantes genómicas ainda permanece sob investigação. Com o desenvolvimento que assistimos das metodologias de sequenciação, hoje o problema reside na interpretação e na associação das variantes a doença. Para tal, estudos funcionais quer em modelos animais, quer em linhas celulares derivadas dos doentes, são necessários para comprovar a patogenicidade dessas variantes, com especial impacto naquelas localizadas em regiões não codificantes, em genes que ainda não estão associados a doença ou associados a penetrância incompleta.

Não obstante, com este trabalho foi possível identificar um conjunto de variantes que poderão estar associadas ao fenótipo da família em estudo. A caracterização molecular da(s) variante(s) envolvidas no fenótipo dos doentes, permitirá um acompanhamento personalizado destes indivíduos, assim como a disponibilização de testes de diagnóstico e aconselhamento genético, incluindo pré-natal, a familiares em risco.

Referências Bibliográficas

1. Thusberg J, Olatubosun A, Vihinen M. Performance of mutation pathogenicity prediction methods on missense variants. *Hum Mutat.* 2011;32(4):358–68.
2. Richards S, Aziz N, Bale S, Bick D, Das S et al. Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405–24.
3. Balkan M, Akbas H, Isi H, Oral D, Turkyilmaz A, Kalkanli S, et al. Cytogenetic analysis of 4216 patients referred for suspected chromosomal abnormalities in Southeast Turkey. *Genet Mol Res.* 2010;9(2):1094–103.
4. Sharp AJ, Cheng Z, Eichler EE. Structural variation of the human genome. *Annu Rev Genomics Hum Genet.* 2006;7:407–42.
5. Luthardt FW, Keitges E. Chromosomal Syndromes and Genetic Disease. *Encycl Life Sci.* 2001;1–12.
6. Den Dunnen JT, Antonarakis E. Nomenclature for the description of human sequence variations. *Hum Genet.* 2001;109(1):121–4.
7. Ramírez-Bello J, Jiménez-Morales M. [Functional implications of single nucleotide polymorphisms (SNPs) in protein-coding and non-coding RNA genes in multifactorial diseases]. *Gac Med Mex [Internet].* 2017;153(2):238–50.
8. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: The NCBI database of genetic variation. *Nucleic Acids Res.* 2001;29(1):308–11.
9. Veltman JA, Brunner HG. De novo mutations in human genetic disease. *Nat Rev Genet [Internet].* 2012;13(8):565–75.
10. Paulson H, Arbor A. Repeat Expansion Diseases. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine.* 2006. 1627–1627 p.
11. Rogan PK, Caminsky N, Mucaki EJ. Interpretation of mRNA splicing mutations in genetic disease: Review of the literature and guidelines for information-theoretical analysis. *F1000Research.* 2014;3(1).
12. Richard A. Padgett. New connections between splicing and human disease. *Trends Genet [Internet].* 2012;28(4):147–54.
13. Turunen JJ, Niemelä EH, Verma B, Frilander MJ. The significant other: Splicing by the minor spliceosome. *Wiley Interdiscip Rev RNA.* 2013;4(1):61–76.
14. Zhang F, Lupski JR. Non-coding genetic variants in human disease. *Hum Mol Genet.* 2015;24(R1):R102–10.
15. Pérez-Agustín A, Pinsach-Abuin ML, Pagans S. Role of non-coding variants in brugada syndrome.

- Int J Mol Sci. 2020;21(22):1–19.
16. Meyts I, Immunology C, Bosch B, Branch R, Bolze A, Branch R, et al. Exome and genome sequencing for inborn errors of immunity. *J Allergy Clin Immunol*. 2017;138(4):957–69.
 17. Pereira R, Oliveira J, Sousa M. Bioinformatics and Computational Tools for Next-Generation Sequencing Analysis in Clinical Genetics. *J Clin Med*. 2020;9(1):132.
 18. Michelson DJ, Shevell MI, Sherr EH, Moeschler JB, Gropman AL, Ashwal S. Evidence Report: Genetic and metabolic testing on children with global developmental delay: Report of the quality standards Subcommittee of the American Academy of Neurology and the Practice Committee of the Child. *Neurology*. 2011;77(17):1629–35.
 19. Vissers LELM, Gilissen C, Veltman JA. Genetic studies in intellectual disability and related disorders. *Nat Rev Genet [Internet]*. 2016;17(1):9–18.
 20. Budworth H, McMurray CT. A brief history of triplet repeat diseases. *Methods Mol Biol*. 2013;1010:3–17.
 21. Ciaccio C, Fontana L, Milani D, Tabano S, Miozzo M, Esposito S. Fragile X syndrome: a review of clinical and molecular diagnoses. *Ital J Pediatr*. 2017;43(1):1–12.
 22. Harripaul R, Noor A, Ayub M, Vincent JB. The Use of Next-Generation Sequencing for Research and Diagnostics for Intellectual Disability. *Cold Spring Harb Perspect Med*. 2017;7(3):1–13.
 23. de Brouwer APM, Yntema HG, Kleefstra T, Lugtenberg D, Oudakker AR, de Vries BBA, et al. Mutation frequencies of X-linked mental retardation genes in families from the EuroMRX consortium. *Hum Mutat*. 2007;28(2):207–8.
 24. Tarpey PS, Smith R, Pleasance E, Whibley A, Edkins S, Hardy C, et al. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. *Nat Genet*. 2009;41(5):535–43.
 25. Schuurs-Hoeijmakers JHM, Hehir-Kwa JY, Pfundt R, Van Bon BWM, De Leeuw N, Kleefstra T, et al. Homozygosity mapping in outbred families with mental retardation. *Eur J Hum Genet*. 2011;19(5):597–601.
 26. Di Resta C, Galbiati S, Carrera P, Ferrari M. Next-generation sequencing approach for the diagnosis of human diseases: Open challenges and new opportunities. *Electron J Int Fed Clin Chem Lab Med*. 2018;29(1):4–14.
 27. Yohe S, Thyagarajan B. Review of clinical next-generation sequencing. *Arch Pathol Lab Med*. 2017;141(11):1544–57.
 28. Gupta N, Verma VK. Next-Generation Sequencing and Its Application: Empowering in Public Health Beyond Reality. 2019. 313–341 p.
 29. Bonetta L. Whole-Genome sequencing breaks the cost barrier. *Cell*. 2010;141(6):917–9.
 30. Harrington CT, Lin EI, Olson MT, Eshleman JR. Fundamentals of pyrosequencing. *Arch Pathol Lab*

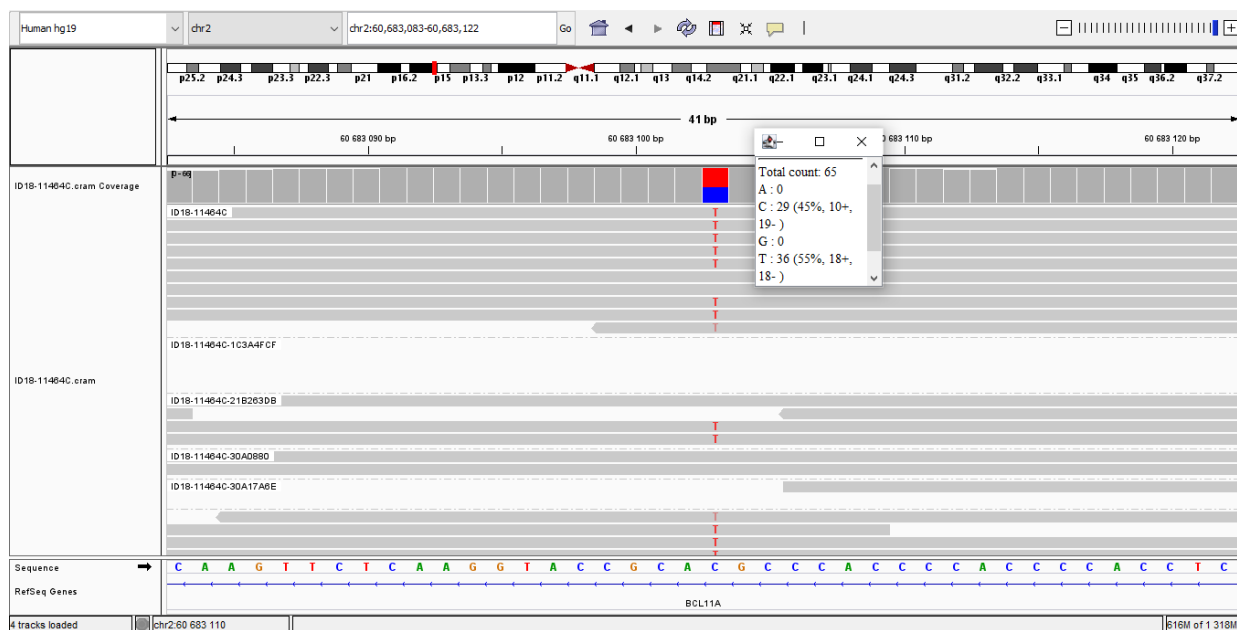
- Med. 2013;137(9):1296–303.
31. Ju J, Kim DH, Bi L, Meng Q, Bai X, Li Z, et al. Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *Proc Natl Acad Sci U S A*. 2006;103(52):19635–40.
 32. Porreca GJ, Shendure J, Church GM. Polony DNA sequencing. *Curr Protoc Mol Biol*. 2006;Chapter 7:1–22.
 33. Budczies J, Bockmayr M, Treue D, Klauschen F, Denkert C. Semiconductor sequencing: How many flows do you need? *Bioinformatics*. 2015;31(8):1199–203.
 34. Roy S, Coldren C, Karunamurthy A, Kip NS, Klee EW, Lincoln SE, et al. Standards and Guidelines for Validating Next-Generation Sequencing Bioinformatics Pipelines: A Joint Recommendation of the Association for Molecular Pathology and the College of American Pathologists. *J Mol Diagnostics*. 2018;20(1):4–27.
 35. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14(2):178–92.
 36. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. *Bioinformatics*. 2011;27(15):2156–8.
 37. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed*. 2013;98(6):236–8.
 38. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. *Am J Hum Genet*. 2016;98(1):149–64.
 39. Vasudevan P, Suri M. A clinical approach to developmental delay and intellectual disability. *Clin Med (Northfield Il)*. 2017;17(6):558–61.
 40. Muzzey D, Evans EA, Lieber C. Understanding the Basics of NGS: From Mechanism to Variant Calling. *Curr Genet Med Rep*. 2015;3(4):158–65.
 41. Elsensohn MH, Leblay N, Dimassi S, Campan-Fournier A, Labalme A, Roucher-Boulez F, et al. Statistical method to compare massive parallel sequencing pipelines. *BMC Bioinformatics*. 2017;18:139.
 42. Petersen BS, Fredrich B, Hoepfner MP, Ellinghaus D, Franke A. Opportunities and challenges of whole-genome and -exome sequencing. *BMC Genet*. 2017;18(1):1–13.
 43. Ku CS, Tan EK, Cooper DN. From the periphery to centre stage: De novo single nucleotide variants play a key role in human genetic disease. *J Med Genet*. 2013;50(4):203–11.
 44. Srivastava S, Love-Nichols JA, Dies KA, Ledbetter DH, Martin CL, Chung WK, et al. Meta-analysis and multidisciplinary consensus statement: exome sequencing is a first-tier clinical diagnostic test for individuals with neurodevelopmental disorders. *Genet Med*. 2019;21(11):2413–21.
 45. Gilissen C, Hehir-Kwa JY, Thung DT, Van De Vorst M, Van Bon BWM, Willemsen MH, et al. Genome

- sequencing identifies major causes of severe intellectual disability. *Nature*. 2014;511(7509):344–7.
46. DG M, TA M, DP D, HL R, Shendure J, GR A, et al. Guidelines for investigating causality of sequence variants in human disease. *Nature*. 2014;508(7497):469–76.
 47. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. Genenames.org: The HGNC and VGNC resources in 2021. *Nucleic Acids Res*. 2021;49(D1):D939–46.
 48. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum Mutat*. 2016;37(6):564–9.
 49. Lefter M, Vis JK, Vermaat M, den Dunnen JT, Taschner PEM, Laros JFJ. Mutalyzer 2: Next Generation HGVS Nomenclature Checker. *Bioinformatics*. 2020;24(68583):1–17.
 50. Koch L. Exploring human genomic diversity with gnomAD. *Nat Rev Genet*. 2020;21(8):448.
 51. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: Improving access to variant interpretations and supporting evidence. *Nucleic Acids Res*. 2018;46(D1):D1062–7.
 52. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM®), an Online catalog of human genes and genetic disorders. *Nucleic Acids Res*. 2015;43(D1):D789–98.
 53. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. *Biopreserv Biobank*. 2015;13(5):311–7.
 54. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved Splice Site Detection in Genie. *J Comput Biol*. 1997;4(3):311–23.
 55. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative Genome Viewer. *Nat Biotechnol*. 2011;29(1):24–6.
 56. Feingold EA, Good PJ, Guyer MS, Kamholz S, Liefer L, Wetterstrand K, et al. The ENCODE (ENCyclopedia of DNA Elements) Project. *Science (80-)*. 2004;306(5696):636–40.
 57. Firth H V., Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources. *Am J Hum Genet*. 2009;84(4):524–33.
 58. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: A curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(D1):958.
 59. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, et al. The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res*. 2019;47(D1):D853–8.
 60. Lee CM, Barber GP, Casper J, Clawson H, Diekhans M, Gonzalez JN, et al. UCSC Genome Browser

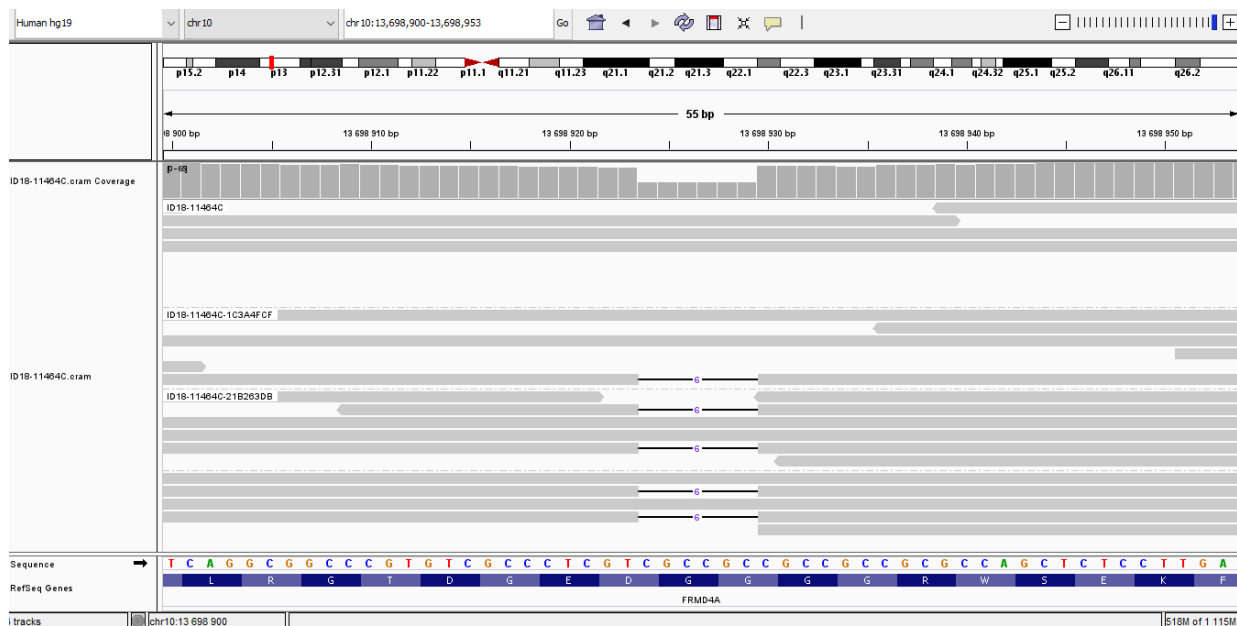
- enters 20th year. *Nucleic Acids Res.* 2020;48(D1):D756–61.
61. Sherry ST, Ward M, Sirotkin K. dbSNP – database for single nucleotide polymorphisms and other classes of minor genetic variation. *Genome Res.* 1999;9(8):677–9.
 62. Landrum MJ, Chitipiralla S, Brown GR, Chen C, Gu B, Hart J, et al. ClinVar: Improvements to accessing data. *Nucleic Acids Res.* 2020;48(D1):D835–44.
 63. Amberger JS, Hamosh A. Searching Online Mendelian Inheritance in Man (OMIM: A Knowledgebase of Human Genes and Genetic Phenotypes. *Curr Protoc Bioinforma.* 2017;58:1.2.1–1.2.12.
 64. The GTEx Consortium. The Genotype–Tissue Expression (GTEx) project. *Nat Genet.* 2013;45(6):580–5.
 65. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA elements (ENCODE): Data portal update. *Nucleic Acids Res.* 2018;46(D1):D794–801.
 66. Acedo A, Sanz DJ, Durán M, Infante M, Pérez-Cabornero L, Miner C, et al. Comprehensive splicing functional analysis of DNA variants of the BRCA2 gene by hybrid minigenes. *Breast Cancer Res.* 2012;14(3):R87.
 67. Meynert AM, Ansari M, FitzPatrick DR, Taylor MS. Variant detection sensitivity and biases in whole genome and exome sequencing. *BMC Bioinformatics.* 2014;15(247):1–11.
 68. Belkadi A, Bolze A, Itan Y, Cobat A, Vincent QB, Antipenko A, et al. Whole–genome sequencing is more powerful than whole–exome sequencing for detecting exome variants. *Proc Natl Acad Sci U S A.* 2015;112(17):5473–8.
 69. Aldahmesh MA, Mohamed JY, Alkuraya HS, Verma IC, Puri RD, Alaiya AA, et al. Recessive mutations in ELOVL4 cause ichthyosis, intellectual disability, and spastic quadriplegia. *Am J Hum Genet.* 2011;89(6):745–50.
 70. Cadieux–Dion M, Turcotte–Gauthier M, Noreau A, Martin C, Meloche C, Gravel M, et al. Expanding the clinical phenotype associated with ELOVL4 mutation: Study of a large French–Canadian family with autosomal dominant spinocerebellar ataxia and erythrokeratoderma. *JAMA Neurol.* 2014;71(4):470–5.
 71. Maugeri A, Meire F, Hoyng CB, Vink C, Van Regemorter N, Karan G, et al. A novel mutation in the ELOVL4 gene causes autosomal dominant Stargardt–like macular dystrophy. *Investig Ophthalmol Vis Sci.* 2004;45(12):4263–7.
 72. Agbaga MP, Brush RS, Mandal MNA, Henry K, Elliott MH, Anderson RE. Role of Stargardt–3 macular dystrophy protein (ELOVL4) in the biosynthesis of very long chain fatty acids. *Proc Natl Acad Sci U S A.* 2008;105(35):12843–8.
 73. Meinsma R, Fernandez Salguero P, Van Kuilenburg ABP, Van Gennip AH, Gonzalez FJ. Human Polymorphism in Drug Metabolism: Mutation in the Dihydropyrimidine Dehydrogenase Gene Results in Exon Skipping and Thymine Uracilurea. *DNA Cell Biol.* 1995;14(1):1–6.

74. de Ligt J, Willemsen MH, van Bon BWM, Kleefstra T, Yntema HG, Kroes T, et al. Diagnostic Exome Sequencing in Persons with Severe Intellectual Disability. *N Engl J Med.* 2012;367(20):1921–9.
75. Dias C, Estruch SB, Graham SA, McRae J, Sawiak SJ, Hurst JA, et al. BCL11A Haploinsufficiency Causes an Intellectual Disability Syndrome and Dysregulates Transcription. *Am J Hum Genet.* 2016;99(2):253–74.
76. Molinari F, Foulquier F, Tarpey PS, Morelle W, Boissel S, Teague J, et al. Oligosaccharyltransferase-Subunit Mutations in Nonsyndromic Mental Retardation. *Am J Hum Genet.* 2008;82(5):1150–7.
77. Pak CH, Garshasbi M, Kahrizi K, Gross C, Apponi LH, Noto JJ, et al. Mutation of the conserved polyadenosine RNA binding protein, ZC3H14/dNab2, impairs neural function in Drosophila and humans. *Proc Natl Acad Sci U S A.* 2011;108(30):12390–5.
78. Guissart C, Latypova X, Rollier P, Khan TN, Stamberger H, McWalter K, et al. Dual Molecular Effects of Dominant RORA Mutations Cause Two Variants of Syndromic Intellectual Disability with Either Autism or Cerebellar Ataxia. *Am J Hum Genet.* 2018;102(5):744–59.
79. Hoischen A, Van Bon BWM, Gilissen C, Arts P, Van Lier B, Steehouwer M, et al. De novo mutations of SETBP1 cause Schinzel-Giedion syndrome. *Nat Genet.* 2010;42(6):483–5.
80. Coe BP, Witherspoon K, Rosenfeld JA, van Bon B, Vulto-van Silfhout A, Bosco P, et al. Refining analyses of copy number variation identifies specific genes associated with developmental delay. *Nat Genet.* 2014;46(10):1063–71.
81. Kleefstra T, Kramer JM, Neveling K, Willemsen MH, Koemans TS, Vissers LELM, et al. Disruption of an EHMT1-associated chromatin-modification module causes intellectual disability. *Am J Hum Genet.* 2012;91(1):73–82.
82. Kucharík M, Budiš J, Hýblová M, Minárik G, Szemes T. Copy Number Variant Detection with Low-Coverage Whole-Genome Sequencing Represents a Viable Alternative to the Conventional Array-CGH. *Diagnostics.* 2021;11(4):708.
83. Weckselblatt B, Rudd MK. Human structural variation: mechanisms of chromosome rearrangements. *Trends Genet.* 2015;31(10):587–99.
84. Ng BG, Xu G, Chandy N, Steyermark J, Shinde DN, Radtke K, et al. Biallelic Mutations in FUT8 Cause a Congenital Disorder of Glycosylation with Defective Fucosylation. *Am J Hum Genet.* 2018;102(1):188–95.
85. Shah MS, Cinnioglu C, Maisenbacher M, Comstock I, Kort J, Lathi RB. Comparison of cytogenetics and molecular karyotyping for chromosome testing of miscarriage specimens. *Fertil Steril.* 2017;107(4):1028–33.
86. Dias M, Partington M. Congenital brain and spinal cord malformations and their associated cutaneous markers. *Pediatrics.* 2015;136(4):e1105–19.

Anexos



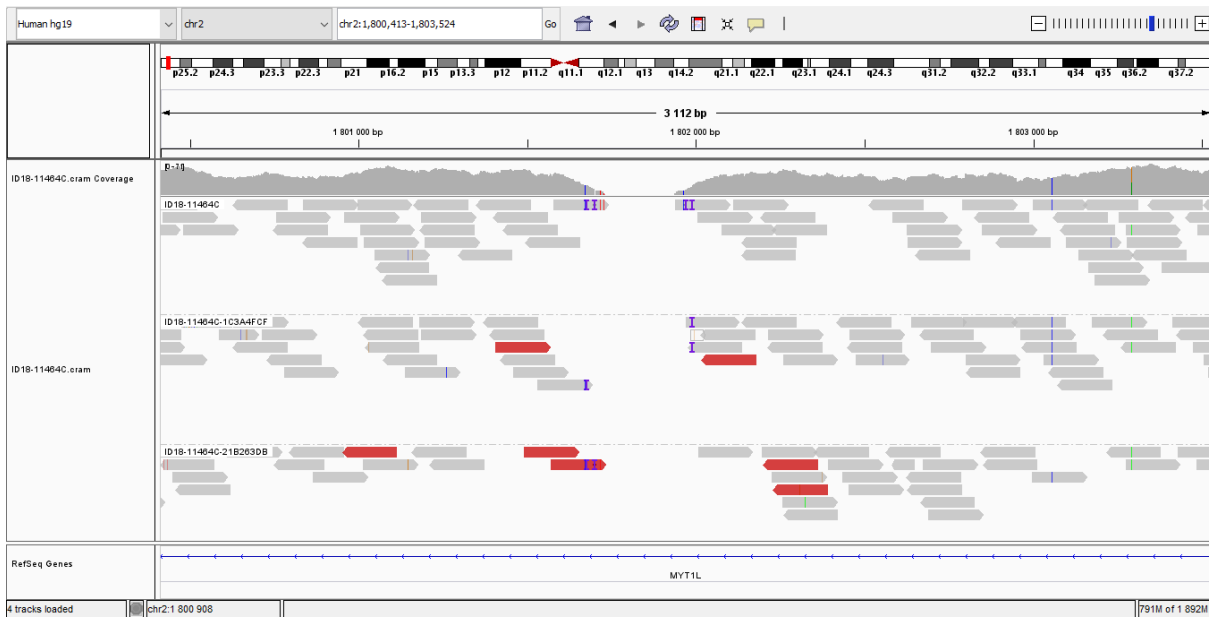
Anexo 1 – Imagem representativa de uma substituição no *software* IGV. É detetada uma substituição de uma citosina (C) por uma timina (T) em 55% das *reads* (retângulo cinzento), ou seja, em heterozigotia.



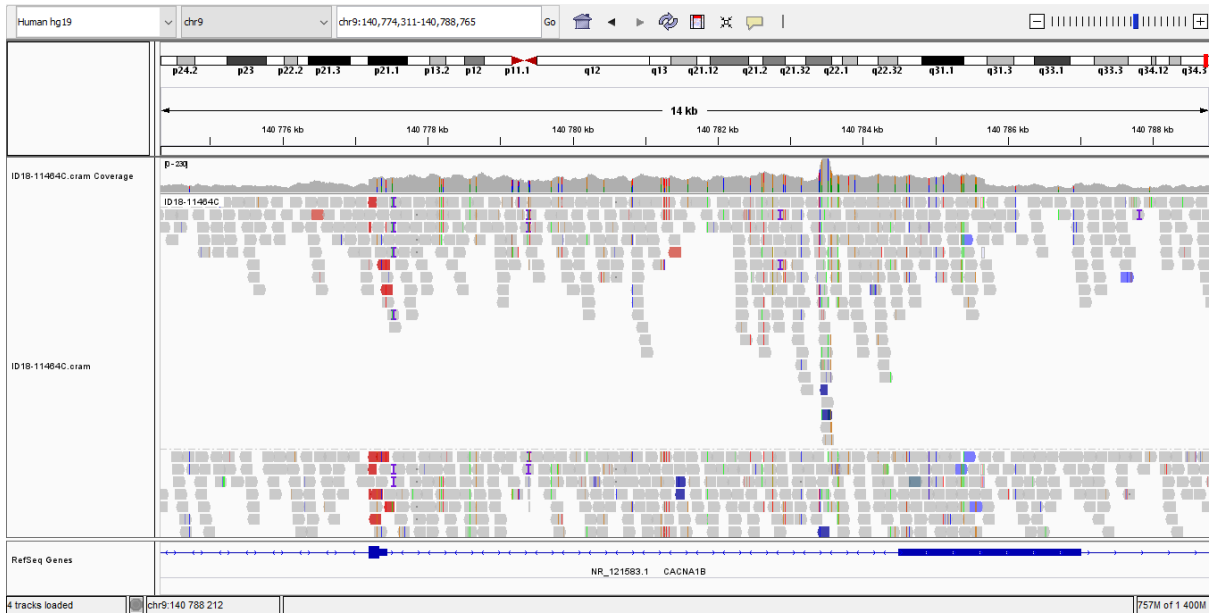
Anexo 2 – Imagem representativa de uma deleção de poucos pares de base (neste caso, seis) no *software* IGV, onde há uma quebra abrupta da cobertura (indicado na linha abaixo do ideograma por “Coverage”).



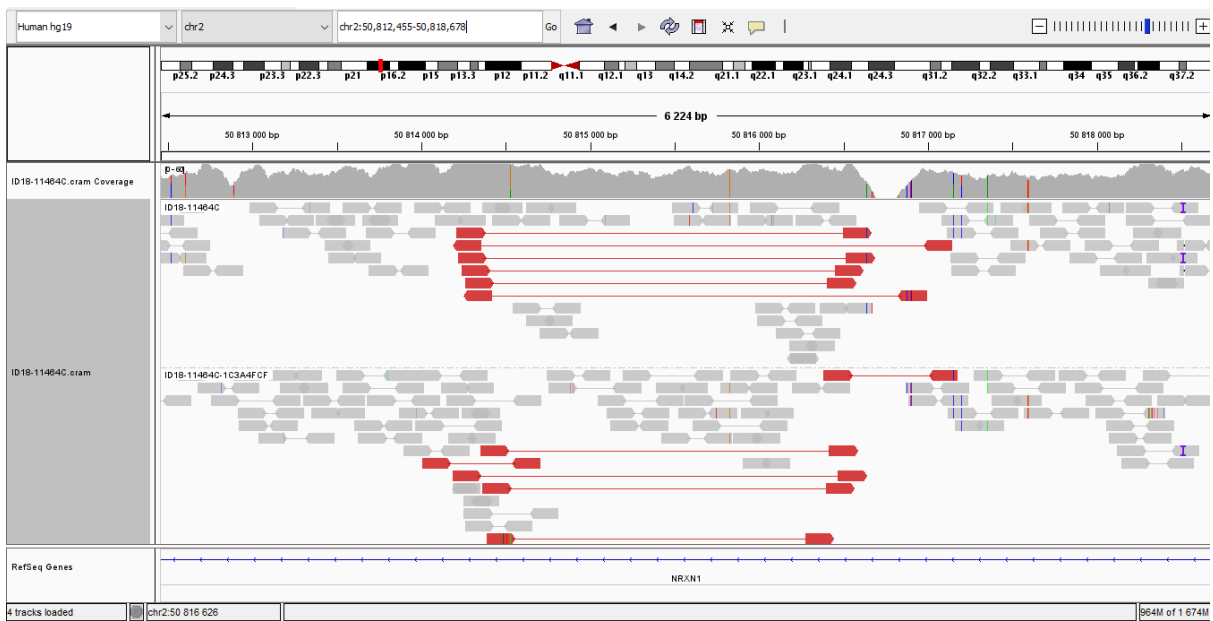
Anexo 3 – Imagem representativa de uma inserção no *software* IGV, apresentada através dos "I" em azul.



Anexo 4 – Imagem representativa de uma deleção no *software* IGV, onde há uma quebra abrupta da cobertura (indicado na linha abaixo do ideograma por "Coverage").



Anexo 5 – Imagem representativa de uma duplicação no *software* IGV, onde há claramente uma alteração da cobertura neste intervalo (indicado na linha abaixo do ideograma por “Coverage”).



Anexo 6 – Imagem representativa de uma inversão no *software* IGV, onde as *reads* que deveriam estar emparelhadas estão no mesmo sentido (indicadas a vermelho), quando deveriam estar em sentidos opostos (cinzento).



Anexo 7 – Imagem representativa de uma translocação no *software* IGV, onde se vê uma redução de cobertura semelhante a uma deleção num dos cromossomas e uma duplicação com aproximadamente o mesmo tamanho no outro cromossoma, com *reads* emparelhadas (indicadas a amarelo).