

UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA



TESE DE DOUTORADO

**Aplicação clínica do sequenciamento e análise
bioinformática de exomas**

ORIENTADA: Natália Duarte Linhares

ORIENTADOR: Prof. Dr. Sergio Danilo Junho Pena

BELO HORIZONTE

Dezembro – 2014

Natália Duarte Linhares

TESE DE DOUTORADO

**Aplicação clínica do sequenciamento e análise
bioinformática de exomas**

Tese apresentada ao Programa de Pós-Graduação em Genética do Departamento de Biologia Geral da Universidade Federal de Minas Gerais, como requisito parcial à obtenção do título de Doutora em Genética, área de concentração “Genômica e Bioinformática”.

Orientador: Prof. Dr. Sergio Danilo Junho Pena

BELO HORIZONTE

Dezembro – 2014

043

Linhares, Natália Duarte.

Aplicação clínica do sequenciamento e análise bioinformática de exomas
[manuscrito] / Natália Duarte Linhares. – 2014.

113 f. : il. ; 29,5 cm.

Orientador: Sergio Danilo Junho Pena.

Tese (doutorado) – Universidade Federal de Minas Gerais, Departamento de
Biologia Geral.

1. Exoma. 2. Sequenciamento. 3. Bioinformática - Teses. 4. Doenças
mendelianas. 5. Genótipo - Teses. 6. Fenótipo – Teses. 7. Cromossomo 1- Teses.
8. Miofibromatose infantil. 9. Cérebro – Calcificação. 10. Genética – Teses. I.
Pena, Sergio Danilo Junho. II. Universidade Federal de Minas Gerais.
Departamento de Biologia Geral. III. Título.

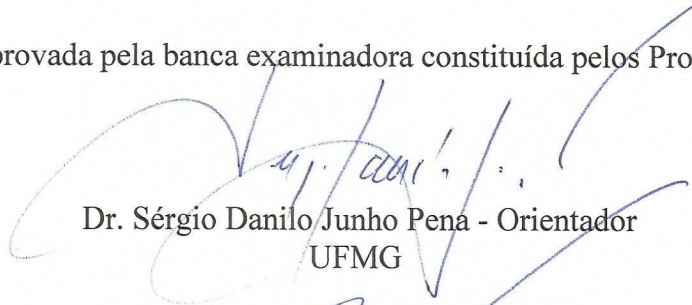
CDU: 575



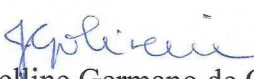
"Aplicação clínica do sequenciamento e análise bioinformática de exomas"

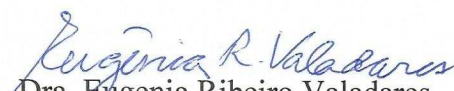
Natália Duarte Linhares

Tese aprovada pela banca examinadora constituída pelos Professores:


Dr. Sérgio Danilo Junho Pena - Orientador
UFMG


Dra. Renata Nascimento de Freitas
UFOP


Dra. Jaqueline Germano de Oliveira
Fiocruz/MG


Dra. Eugenia Ribeiro Valadares
UFMG


Dra. Juliana Gurgel Giannetti
UFMG

Belo Horizonte, 16 de dezembro de 2014.

*Aos pacientes estudados e às suas famílias, pela preciosa colaboração para realização
deste trabalho.*

AGRADECIMENTOS

Ao meu orientador, prof. Sergio Pena, pelo grande apoio e incentivo, pelo exemplo de dedicação à ciência e principalmente por me presentear com esta pesquisa de doutorado. Por toda a sua atenção, disponibilidade e ajuda na minha formação.

Às médicas do Hospital das Clínicas, Dra. Magda Bahia, Dra. Juliana Gurgel e Dra. Eleonora Druve pela colaboração aos nos indicar as famílias aqui estudadas. Agradeço também à Dra. Priscila Ferri Liu pelos esclarecimentos e artigos sobre hepatite autoimune e ao Dr. Shinfay Liu pela ajuda nas coletas de sangue dos pacientes. À Dra. Ana Carla Goldberg, do Hospital Israelita Albert Einstein (São Paulo), pela colaboração para as investigações de HLA de classe II dos pacientes com hepatite autoimune. À Dra. Heloísa Pena e à Carolina Nunes, do Laboratório Gene, pela ajuda com a extração de DNA dos pacientes e envio de amostras para o Canadá. Ao Dr. Sérgio Pereira do “Centre for Applied Genomics” (Hospital for Sick Children, Toronto, Canadá) pela assistência no sequenciamento do exoma dos pacientes aqui estudados e ajuda na interpretação dos dados. Ao Juliano de Paula pela ajuda nos experimentos de sequenciamento de Sanger. Ao Dr. Moisés Pedrosa pela ajuda na realização das fotomicrografias das biopsias de um dos pacientes estudos neste trabalho. Ao prof. Dawidson Gomes, que sempre esteve disposto a esclarecer minhas dúvidas sobre os experimentos de *western blot*.

Às agências de fomento CNPq, pelo fornecimento da bolsa de estudo, e FAPEMIG, pelo financiamento deste projeto. À secretária da Pós-graduação de Genética, Mary Santos, que sempre me recebeu prontamente, com disposição e carinho.

Aos membros da banca de qualificação, Prof. Eduardo Tarazona, Dra. Eugênia Valadares e Dra. Maíra Freire, pelas contribuições que auxiliaram na continuidade do trabalho.

Aos amigos do Laboratório de Genômica Clínica da Faculdade de Medicina da UFMG, Raony Cardenas, Maíra Freire, Raquel Liboredo, Michele Penna e Tiago Magalhães, pelo suporte, ajuda com os experimentos de biologia molecular e análises de bioinformática, e pelas enriquecedoras e motivadoras discussões.

Aos professores, pesquisadores e amigos do Laboratório de Genética e Bioquímica da UFMG. Agradeço especialmente à profa. Andréa Mara Macedo, profa. Glória Regina Franco, prof. Carlos Renato Machado, Neuza Antunes, Danielle Durso, Grazielle Vieira, Pricila Cunha, Fernanda Soardi, Ceres Alves e João Pedro Rocha, pela

agradável convivência, pelo apoio constante e pelos conselhos científicos, que foram muitos.

À Marlene de Miranda, pela amizade e pela ajuda nas coletas do meu sangue, necessários para os testes preliminares de novas técnicas. Ao Ítalo do Valle e Fernanda Soardi, que também forneceram sangue para os meus testes.

Ao meu noivo Gustavo, por todo apoio e motivação. Por me escutar sempre com muita atenção e carinho e pelos inúmeros conselhos. À minha família, em especial aos meus pais, Anastácia Duarte e Valter Linhares e ao meu irmão, Fernando Linhares, por me darem suporte na minha carreira científica. Pelo amor incondicional.

A todos, meus sinceros agradecimentos.

ÍNDICE

	Página
LISTA DE FIGURAS	VIII
LISTA DE TABELAS	X
LISTA DE ABREVIATURAS	XI
RESUMO	1
ABSTRACT	4
ESTRUTURA DA TESE	6
CAPÍTULO 1 - INTRODUÇÃO	7
Aplicações na pesquisa e no diagnóstico do sequenciamento do exoma	8
Princípios da técnica de sequenciamento do exoma	9
Análise bioinformática de exomas	13
Etapa 1: alinhamento	14
Etapa 2: recalibração do escore de qualidade e detecção de	15
variantes e genótipos	
Etapa 3: anotação	15
Etapa 4: uso de filtros para identificar a variante responsável	17
Softwares utilizados para a detecção de variantes candidatas	22
Limitações e vantagens da técnica de sequenciamento do exoma	27
Referências	28
CAPÍTULO 2 - OBJETIVOS	35
CAPÍTULO 3 - ARTIGOS	36
ARTIGO 1: “1p13.2 deletion displays clinical features overlapping	37
Noonan syndrome, likely related to NRAS gene haploinsufficiency”	
Abstract	41
Introduction	42
Clinical report	43
Methods	44
Results	45
Discussion	47
Figure Titles and Legends	52
Table	53
References	57

<i>Figures</i>	56
ARTIGO 2: “ <i>Exome sequencing identifies a novel homozygous variant in NDRG4 in a family with infantile myofibromatosis</i> ”	60
<i>Abstract</i>	61
<i>Introduction</i>	61
<i>Clinical Report</i>	62
<i>Methods</i>	62
<i>Results</i>	63
<i>Discussion</i>	63
<i>References</i>	66
ARTIGO 3: “ <i>Modulation of expressivity in PDGFRB-related infantile myofibromatosis: a role for PTPRG?</i> ”	67
<i>Abstract</i>	68
<i>Introduction</i>	69
<i>Materials and Methods</i>	69
<i>Results</i>	70
<i>Discussion</i>	71
<i>References</i>	73
ARTIGO 4: Análise de exomas de pacientes com microcefalia, convulsões e calcificações no tegmento pontino do cérebro	74
Introdução	75
Relato clínico	76
Métodos	78
Resultados parciais e discussão	82
Referências	97
CAPÍTULO 4 - CONCLUSÕES E PERSPECTIVAS	90
ANEXOS	96
ANEXO 1 - Carta de submissão do artigo 1	97
ANEXO 2 - Artigo 1 - <i>Supplementary Tables</i>	98
ANEXO 3 - Termos de Consentimento Livre e Esclarecido	104
ANEXO 4 - Termos de Assentimento	108

LISTA DE FIGURAS

	Página
CAPÍTULO 1 - INTRODUÇÃO	
FIG. 1. Funcionamento dos <i>kits</i> comerciais para captura do exoma.	10
FIG. 2. Funcionamento da PCR em emulsão.	11
FIG. 3. Metodologia da plataforma SOLiD: sequenciamento por ligação.	12
FIG. 4. Exemplo de uma sequência no formato FASTQ.	13
FIG. 5. Etapas da análise bioinformática de exomas.	14
FIG. 6. Exemplo de um arquivo VCF.	15
FIG. 7. Interface da ferramenta de filtrar as variantes do <i>software Enlis Genome</i> .	23
FIG. 8. Interface da ferramenta <i>One Click</i> do <i>software Mendel,MD</i> .	25
FIG. 9. Interface do <i>software Alamut Visual</i> .	26
FIG. 10. Interface do <i>software Alamut Visual</i> após fazer o upload do arquivo BAM do paciente.	27
 CAPÍTULO 3 - ARTIGOS	
ARTIGO 1:	
FIG. 1. <i>Frontal view of patients with deleted NRAS.</i>	56
FIG. 2. <i>Copy number profile of chromosome 1 of our patient obtained by aCGH.</i>	57
FIG. 3. <i>Schematic representation of the deleted segments in our patient and those previously reported with isolated 1p13.2 microdeletions.</i>	58
FIG. 4. <i>Genomic variants listed in the Database of Genomic Variants (DGVbeta) in the smallest region of overlap between the patients.</i>	59
 ARTIGO 2:	
FIG. 1. <i>Hematoxylin and eosin (H&E) stained sections of the tumor.</i>	62
FIG. 2. <i>Illustration of the analysis method using progressive filters.</i>	64
FIG. 3. <i>Comparative analysis using RLTPR and NDRG4 orthologs.</i>	65
FIG. 4. <i>Results of the Sanger sequencing and autozygous analysis.</i>	66

ARTIGO 3:

FIG. 1. <i>Sanger sequencing.</i>	70
FIG. 2. <i>Comparative analysis using PTPRG orthologues.</i>	71
FIG. 3. <i>Schematic representation of the mechanisms of interaction of PDGFRB and PTPRG.</i>	72

ARTIGO 4:

FIG. 1. Tomografia computadorizada do cérebro da paciente I realizada com 9 meses (A e B) e do paciente II realizada com 9 anos (C e D).	78
FIG. 2. Região duplicada na paciente I pela análise de genotipagem (chr4:55,138,935–55,162,980) visualizada no <i>Database of Genomic Variants</i> .	83
FIG. 3. Resultado da análise de genotipagem de parte do cromossomo 4.	83

LISTA DE TABELAS

	Página
CAPÍTULO 1 - INTRODUÇÃO	
Tabela 1. Classificação do impacto da variante realizada pelo SnpEff.	16
CAPÍTULO 3 - ARTIGOS	
ARTIGO 1:	
Table 1. <i>Clinical features of ten patients with isolated 1p13.2 microdeletions.</i>	53
ARTIGO 2:	
Table 1. <i>Summary of exome sequencing data for each sample.</i>	63
Table 2. <i>Details of the variants narrowed down using consecutive filters based on autosomal recessive model of inheritance.</i>	64
ARTIGO 4:	
Tabela 1. Sequências dos <i>primers</i> usados na qPCR e tamanho esperado <i>amplicons</i> .	80
Tabela 2. Valores obtidos pela análise de qPCR.	84
Tabela 3. Detalhes das variantes selecionadas através da análise de exoma dos pacientes I e II afetados com calcificação no cérebro.	87

LISTA DE ABREVIATURAS

aCGH – Hibridação genômica comparativa em *array*
ACMG – American College of Medical Genetics and Genomics
ANNOVAR – *Annotate Variation*
ARL5B – *ADP-ribosylation factor-like 5B*
BAM – *Binary Alignment Map*
BFAST – *BLAT-like Fast Accurate Search Tool*
BWA – *Burrows-Wheeler Alignment*
BWT – *Burrows-Wheeler Transformation*
CCBE1 – *Collagen and calcium binding EGF domains 1*
CCDS – Sequências das regiões codificadoras consenso
CGD – *Clinical Genomic Database*
CHD2 – *Chromodomain helicase DNA binding protein 2*
CNV – *Copy number variation*
CONEP – Comissão Nacional de Ética em Pesquisa
CRLS1 – *Cardiolipin synthase 1*
CT Scan – *Computerized Tomography Scan*
dbSNP – *Single Nucleotide Polymorphism database*
DC-VC – Doença comum-variante comum
DC-VR – Doença comum-variante rara
DGV – *Database of Genomic Variants*
ESP – *Exome Sequencing Project*
FAN – Fatores antinucleares
FSS – Síndrome de Freeman-Sheldon
GATK – *Genome Analysis Toolkit*
GERP – *Genetic Evolutionary Rate Profiling*
GWAS – *Genome-Wide Association Studies*
H&E – *Hematoxylin and eosin*
HAI – Hepatite autoimune
HAI-1 – Hepatite autoimune tipo 1
HAI-2 – Hepatite autoimune tipo 2
HC – Hospital das Clínicas
HGMD – *Human Gene Mutation Database*

HGNC – HUGO Gene Nomenclature Committee
 HGVS – Human Genome Variation Society
 HI – *Haploinsufficiency Index*
 HLA – *Human Leukocyte Antigen*
 IM – *Infantile myofibromatosis*
 KEGG – *Kyoto Encyclopedia of Genes and Genomes*
 LKM1 – *Liver/kidney microsome type 1*
 LOH – *Loss of heterozygosity*
 LRR – *Leucine-rich repeats*
 MAF – *Minor Allele Frequency*
 MAPK – *Mitogen activated protein kinase*
 MGI – *Mouse Genome Informatics*
 MHC – Complexo principal de histocompatibilidade
 NCBI – National Center for Biotechnology Information
 NDRG4 – *N-myc downregulated gene family member 4*
 NGS – *Next generation sequencing*
 NHGRI – National Human Genome Research Institute
 NHLBI – National Heart, Lung, and Blood Institute
 NOTCH3 – *Notch 3*
 NRAS – *Neuroblastoma RAS viral (v-ras) oncogene homolog*
 NS6 – *Noonan Syndrome type 6*
 OMIM – *Online Mendelian Inheritance in Man*
 PCR – Reação em Cadeia da Polimerase
 PDGFRA – *Platelet-derived growth factor receptor, alpha polypeptide*
 PDGFRB – *Platelet-derived growth factor receptor, beta polypeptide*
 PTH – Paratormônio
 PTPRG – *Protein tyrosine phosphatase, receptor type, gamma*
 RLTPR – *RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing*
 RSU1 – *Ras suppressor protein 1*
 SAM – *Sequence Alignment Map*
 SCN1A – *Sodium channel, neuronal type I, alpha subunit*
 SIFT – *Sorting Intolerant From Tolerant*
 SNP – Polimorfismo de Nucleotídeo Simples

SNV – Variantes de um único nucleotídeo
SOLiD – *Support Oligonucleotide Ligation Detection*
TCLE – Termo de Consentimento Livre e Esclarecido
UH – *Unmasking heterozygosity*
VCF – *Variant Call Format*
VEP – *Variant Effect Predictor*
VUS – *Variants of Uncertain Significance*
WES – *Whole Exome Sequencing*
WGS – *Whole Genome Sequencing*

RESUMO

Aproximadamente 7.000 doenças mendelianas já foram descritas, porém menos da metade delas foi relacionada com mutações em genes específicos. No presente trabalho, realizamos a análise bioinformática de exomas de pacientes com doenças mendelianas raras, atendidos no Hospital das Clínicas da UFMG ou no Laboratório Gene – Núcleo de Genética Médica, com o objetivo de localizar variantes que pudessem estar causalmente relacionadas com as doenças.

Foi realizado um estudo embasado na análise de exomas, tendo como foco duas aplicações clínicas distintas: (1) aprimoramento da correlação genótipo-fenótipo em cromossomopatias através do estudo de um paciente com microdeleção no cromossomo 1p13.2; (2) estudo de doenças mendelianas de causa pouco esclarecida; neste, analisamos duas famílias com miofibromatose infantil e dois pacientes com calcificações no cérebro e convulsões.

No primeiro caso, avaliamos o fenótipo de pacientes com microdeleções em 1p13.2 e observamos que muitos apresentavam características de síndrome de Noonan. A região deletada em comum entre todos os pacientes contém o gene *NRAS*, e mutações de ganho de função nesse gene foram descritas em casos de síndrome de Noonan tipo 6. Realizamos, então, a análise de sequenciamento do exoma do nosso paciente para examinar a possibilidade de variantes patogênicas recessivas desmascaradas pela deleção hemizigota e também para verificar se o paciente tinha alguma variante patogênica nos genes relacionados com a síndrome de Noonan. Concluímos que o fenótipo dos pacientes pode ser relacionado causalmente com a deleção do gene *NRAS*. Então, havia o paradoxo de que haploinsuficiência de *NRAS* poderia causar características similares a mutações de ganho de função. Uma situação similar foi descrita para a síndrome de Noonan com ganho de função e haploinsuficiência do gene *SHOC2* e também para a síndrome cardio-facio-cutânea, na qual o fenótipo da doença pode aparecer tanto em razão de deleções de *MAP2K2* quanto de mutações de ganho de função. Concluímos, assim, que a desregulação da via Ras/MAPK causada tanto por mutação de ganho de função como por haploinsuficiência pode resultar em fenótipos similares.

Realizamos também a análise de exomas, a fim de estudar doenças mendelianas de causa pouco esclarecida; para tal, analisamos duas famílias com miofibromatose infantil (MI), desordem rara, de patogênese molecular pouco conhecida, caracterizada

pelo desenvolvimento de tumores benignos na pele, no músculo, nos ossos e nas vísceras. Neste estudo, realizamos inicialmente a análise dos exomas de dois irmãos diagnosticados com MI multicêntrica visceral e dos seus pais saudáveis e consanguíneos. Identificamos uma variante homozigota no gene *NDRG4* nos dois irmãos. Coerentemente com o fenótipo de MI, *NDRG4* é um gene relacionado com o desenvolvimento de tumores. Desse modo, sugerimos que a variante em *NDRG4* pode ser a variante causal na família estudada com MI autossômica recessiva. Para confirmar nossos achados, pesquisamos na literatura por famílias com MI com herança possivelmente autossômica recessiva. Estudamos, então, uma família francesa com dois irmãos afetados com pais saudáveis e não consanguíneos. Identificamos, porém, nos dois irmãos, uma variante heterozigota no gene *PDGFRB*. Variantes nesse gene foram relacionadas com a forma autossômica dominante de MI. Para nossa surpresa, a mãe dos pacientes também apresentava a mesma variante em *PDGFRB*. Constatamos que ambos os irmãos herdaram paternalmente uma variante heterozigota em *PTPRG*, uma enzima que sabidamente desfosforila PDGFRB, reduzindo sua atividade. Sugerimos, então, que variantes em *PTPRG* podem explicar a penetrância fenotípica completa nos irmãos afetados. Portanto, propusemos que a variante em *NDRG4* deve ser investigada em outras famílias com MI autossômica recessiva e que variantes no gene *PDGFRB* estão relacionadas apenas com a forma autossômica dominante da doença.

Além disso, estudamos dois pacientes com microcefalia, convulsões e calcificações no tegmento pontino do cérebro, desordem rara com etiologia pouco conhecida. Suspeitamos de variantes provavelmente patogênicas em quatro genes candidatos: *CCBE1*, *ARL5B*, *RSU1* e *CRLS1*. Todavia, pela técnica de sequenciamento de Sanger, verificamos que tais variantes eram erros de leitura do sequenciamento do exoma. Realizamos, então, nova análise dos exomas, procurando por genes que sabidamente causassem epilepsia. Encontramos, por meio dessa análise, variantes candidatas nos genes *SCN1A* e *CHD2*. Mutações heterozigotas nesses genes foram causalmente relacionadas com síndrome de Dravet, caracterizada por crises de convulsões com início no primeiro ano de vida. Porém, estas variantes também não foram confirmadas pela técnica de sequenciamento de Sanger. Análises futuras serão então necessárias para identificar novos genes candidatos.

Concluindo, com o presente trabalho verificamos que a análise de exomas pode melhorar a qualidade de correlações genótipo-fenótipo de síndromes causadas por anormalidades cromossômicas, bem como auxiliar na detecção de genes que, quando

mutados, podem ser causadores de doenças genéticas. Os resultados encontrados contribuíram para o melhor entendimento das doenças genéticas estudadas, auxiliando, desse modo, no diagnóstico de pacientes futuros e contribuindo para o acompanhamento e o aconselhamento genético dos pacientes e de seus familiares.

ABSTRACT

Approximately 7,000 Mendelian diseases are described, but less than half of these disorders were associated with mutations in specific genes. In this work we performed bioinformatics analysis of exomes from patients with rare Mendelian diseases, treated at the Hospital das Clínicas da UFMG or Laboratório Gene – Núcleo de Genética Médica, aiming to find variants that may be causally associated with the diseases.

We conducted a study grounded in exome analysis, which has focused in two distinct clinical applications: (1) to improve the genotype-phenotype correlation in chromosomal disorders, by studying a patient with chromosome 1p13.2 microdeletion; (2) to study Mendelian diseases with undefined cause; we analyzed two families with infantile myofibromatosis and two patients with brain calcifications and seizures.

In the first case we evaluated the phenotype of patients with 1p13.2 microdeletions and we found that they had many features of Noonan Syndrome. The overlap deletion region of all the patients contained the *NRAS* gene, gain-of-function mutations of which have been described as causing Noonan Syndrome, Type 6. We performed thus exome sequencing analysis in our patient in order to examine the possibility of a recessive pathogenic variant unmasked by his hemizygous deletion and also to verify if he had pathogenic variants in any of the Noonan Syndrome genes. We concluded that their phenotype may be causally related to deletion of *NRAS*. Thus, we have the paradox that *NRAS* haploinsufficiency appears to be able to cause similar characteristics to gain-of-function mutations. Similar situation have been described for Noonan syndrome with gain-of-function mutations and haploinsufficiency of the *SHOC2* gene and also for the related Cardio-facio-cutaneous syndrome, in which signs of the disease can appear due to both *MAP2K2* deletions and gain-of-function mutations. Consequently we concluded that dysregulation of the Ras/MAPK pathway caused either by gain-of-function mutations or haploinsufficiency may result in similar phenotypes.

We also performed exome analysis to study Mendelian diseases with poorly defined cause, for this we analyzed two families with infantile myofibromatosis (MI), which is a rare disorder, with molecular pathogenesis poorly understood, characterized by the development of benign tumors in the skin, muscle, bones and viscera. In this study, we analyzed firstly the exomes of two brothers diagnosed with visceral

multicentric IM, and of its healthy and consanguineous parents. We identified a homozygous variant in the gene *NDRG4* in the two brothers. Consistent with the IM phenotype, *NDRG4* is a tumor-related gene. Thus, we suggested that the variant in *NDRG4* may be the causal variant in the studied family with autosomal recessive IM. To confirm our findings, we surveyed the literature for IM families with inheritance possibly autosomal recessive. Then we studied a French family with two affected siblings of healthy and not consanguineous parents. However, we identified in the two brothers a heterozygous variant in the gene *PDGFRB*. Variants in this gene have been associated with autosomal dominant form of IM. To our surprise, the mother of the patients also had the same variant in *PDGFRB*. We also found that both brothers inherited paternally a heterozygous variant in *PTPRG*, an enzyme known to dephosphorylate *PDGFRB*, reducing its activity. Then we suggested that variants in *PTPRG* may explain the full phenotypic penetrance in the affected siblings. Therefore, we propose that the variant in *NDRG4* should be investigated in other families with autosomal recessive IM, and variants in the gene *PDGFRB* are associated only with the autosomal dominant form of the disease.

In addition, we studied two patients with microcephaly, seizures and calcifications in pontine tegmentum of the brain, which is a rare disorder with poorly understood etiology. We suspected of probably pathogenic variants in four candidate genes: *CCBE1*, *ARL5B*, *RSU1* and *CRLS1*. However, by Sanger sequencing technique, we found that such variants were exome sequencing errors. We performed then new analysis of the exomes, looking for genes that are known to cause epilepsy. Doing this analysis we found candidate variants in *SCN1A* and *CHD2*. Heterozygous mutations in these genes have been causally related to Dravet syndrome, characterized by seizures with onset in the first year of life. Nevertheless, these variants were also not confirmed by Sanger sequencing. Future analyzes will then be required to identify novel candidate genes.

In conclusion, with this study we verified that exome analysis can improve the quality of genotype-phenotype correlations of syndromes caused by chromosomal abnormalities and can assist in the detection of genes that, when mutated, may cause genetic diseases. Our results contributed to a better understanding of the studied genetic diseases, supporting, thus, the diagnosis of future patients and contributing to the monitoring and genetic counseling of patients and their families.

ESTRUTURA DA TESE

Esta tese apresenta um capítulo inicial com a introdução geral, abordando aplicações e análise do sequenciamento de exoma. Em seguida, no segundo capítulo são apresentados os objetivos gerais e específicos. No capítulo três, a metodologia, resultados e discussão são apresentados na forma de quatro artigos, descritos abaixo. Finalizamos a tese com um capítulo de conclusões e perspectivas. Ao final de cada capítulo/artigo foram detalhadas as suas referências.

- Artigo 1: “*1p13.2 deletion displays clinical features overlapping Noonan syndrome, likely related to NRAS gene haploinsufficiency*”, submetido para a revista *European Journal of Medical Genetics*;

- Artigo 2: “*Exome sequencing identifies a novel homozygous variant in NDRG4 in a family with infantile myofibromatosis*”, publicado na revista *European Journal of Medical Genetics*;

- Artigo 3: “*Modulation of expressivity in PDGFRB-related infantile myofibromatosis: a role for PTPRG?*”, publicado na revista *Genetics and Molecular Research*;

- Artigo 4: “Análise de exomas de pacientes com microcefalia, convulsões e calcificações no tegmento pontino do cérebro”.

CAPÍTULO 1 – INTRODUÇÃO

Estamos na era da genômica: métodos genômicos refinados como a hibridação genômica comparativa em *array* (aCGH) e o polimorfismo de nucleotídeo simples (SNP) *array* provaram ser ferramentas úteis ao diagnóstico de pacientes com suspeita de rearranjos cromossômicos, como microdeleções e microduplicações^[1]. Entretanto, tais técnicas não são indicadas para diagnóstico de pacientes com suspeita de doenças monogênicas com base genética desconhecida^[2]. Doenças monogênicas são causadas por mutações em genes únicos, herdadas com padrões dominante, recessivo ou ligado ao cromossomo X. Nos casos em que a base molecular da doença é conhecida, pode ser realizado, por exemplo, o sequenciamento do gene alterado ou a Reação em Cadeia da Polimerase (PCR) com *primers* específicos para a região da mutação^[2].

Aproximadamente 7.000 doenças mendelianas foram descritas, porém somente cerca de 30%-50% dessas desordens foram causalmente relacionadas a mutações em genes específicos^[3]. Até recentemente, a identificação do gene responsável era realizada por meio de clonagem posicional, isto é, de análise de ligação, objetivando identificar um intervalo genômico de aproximadamente 0,5 cM-10 cM, que contém cerca de 300 genes candidatos^[4]. O sequenciamento de um grande número de genes candidatos era caro e demorado. Adicionalmente, outros fatores limitavam o descobrimento do gene responsável, tais como a disponibilidade de poucos pacientes para estudo, o *fitness* reprodutivo reduzido, a penetrância incompleta e a heterogeneidade genética de algumas doenças^[5, 6].

O sequenciamento genômico completo como prática rotineira se tornou possível somente após a invenção do chamado “sequenciamento de nova geração” (*next-generation sequencing* – NGS). Esse método é 50.000 vezes mais eficiente do que o sequenciamento de Sanger, conhecido como “sequenciamento de primeira geração”, o qual foi utilizado para o sequenciamento original do genoma humano^[7]. A título de comparação, o custo do sequenciamento do genoma humano de referência, publicado em 2004 pelo International Human Genome Sequencing Consortium, foi estimado em 300 milhões de dólares^[8]; já o custo atual do sequenciamento total do genoma por meio de NGS é de aproximadamente 8.000 dólares^[9].

Todavia, a análise do genoma completo é complexa, demorada e trabalhosa^[10]. Foi, então, desenvolvido um método que sequencia seletivamente apenas a parte do genoma que é transcrita, ou seja, os exons^[11]. Nesse método, chamado de

“sequenciamento do exoma”, são sequenciados todos os ~180.000 exons do genoma, que correspondem a aproximadamente 1%-2% das 3,4 bilhões de bases do genoma^[11]. Esta porção do genoma, apesar de pequena, é sede de 85% das variantes relacionadas às doenças mendelianas^[12]. Desse modo, foi possível a redução do tempo de análise e do custo do exame: atualmente, o preço do sequenciamento do exoma é de aproximadamente 1.000 dólares em empresas especializadas^[13].

A demonstração inicial de que o exoma pode ser utilizado como ferramenta para a identificação de variantes genéticas relacionadas com doenças, realizado por Ng *et al.*,^[11] em 2009, foi vista como uma revolução na genética humana e o início da era do sequenciamento genômico. Desde então, o método de sequenciamento do exoma se tornou amplamente utilizado para identificar genes relacionados a doenças mendelianas em circunstâncias em que os métodos convencionais falharam^[14, 15]. Isso ocorre porque a maioria das variantes relacionadas com doenças mendelianas está localizada em regiões codificadoras de proteínas^[12]. Além disso, uma grande fração das variantes que alteram proteínas é patogênica, como substituições de uma única base que causam troca de aminoácidos ou que causam códons finalizadores e pequenas inserções/deleções (indels)^[12, 16].

APLICAÇÕES NA PESQUISA E NO DIAGNÓSTICO DO SEQUENCIAMENTO DO EXOMA

O sequenciamento do exoma pode ser realizado em dois contextos diferentes: pesquisa e diagnóstico^[13, 17]. Além de ser utilizado para identificar genes candidatos em doenças monogênicas, conforme exposto previamente, o sequenciamento do exoma vem sendo usado com sucesso na pesquisa para identificar variantes causais em desordens heterogêneas (como perda auditiva e deficiência intelectual), em doenças comuns e desordens complexas (como doenças cardiovasculares e obesidade) e em estudos relacionados com o câncer^[revisado por 18, 17]. Já a análise de exoma para diagnóstico é indicada pelo American College of Medical Genetics and Genomics (ACMG) para detecção de variantes raras em pacientes com fenótipo de doença mendeliana nos seguintes casos^[19]:

a) Os dados fenotípicos ou a história familiar sugerem fortemente etiologia genética, mas o fenótipo não corresponde a uma desordem genética específica, para a qual um teste genético que tem como alvo um gene específico esteja disponível clinicamente;

b) O paciente apresenta uma desordem genética definida que demonstra elevado grau de heterogeneidade genética, tornando a análise de exoma uma abordagem mais prática e barata (múltiplos genes são analisados simultaneamente);

c) O paciente (ou o feto) apresenta uma doença genética provável, mas testes genéticos específicos disponíveis para o fenótipo não conseguiram chegar a um diagnóstico.

O sequenciamento do exoma para diagnóstico pode levar a descoberta de uma variante patogênica conhecida ou a revelação de uma variante nova e aparentemente patogênica em genes que sabidamente podem causar a doença em estudo^[20]. Nos casos de descoberta de variantes novas aparentemente patogênicas em genes ainda não relacionados com doenças humanas, pesquisas devem ser realizadas para estabelecer a patogenicidade da variante^[20]. Dessa forma, há casos de análises de exoma para diagnóstico que acabam se tornando pesquisa.

Ao realizar o sequenciamento de exoma para diagnóstico ou para pesquisa, existe também a possibilidade de identificação de variantes que podem causar susceptibilidade genética para uma doença não relacionada com a indicação para o teste. Tais variantes são descritas como achados secundários ou incidentais. As recomendações atuais da ACMG são de incluir na análise um conjunto de 56 genes que representem 24 desordens clinicamente acionáveis, ou seja, para as quais tratamento ou prevenção estejam disponíveis^[21]. Entre essas desordens, estão incluídos, por exemplo, o câncer de mama e de ovário hereditário (relacionados com variantes nos genes *BRCA1* e *BRCA2*) e o retinoblastoma (variantes no gene *RB1*)^[21].

Contudo, essas recomendações para o retorno dos resultados incidentais têm se mostrado controversas; há opiniões divergentes quanto a quais achados incidentais devem ser relatados, quanto a questões do termo de consentimento e quanto a relatar condições que se manifestam apenas na idade adulta para os pais de crianças que realizaram o exame^[22]. Consequentemente, entende-se ser ainda necessário alcançar consenso sobre esse assunto^[23].

PRINCÍPIOS DA TÉCNICA DE SEQUENCIAMENTO DO EXOMA

O sequenciamento do exoma é realizado basicamente em duas etapas: (1) captura do exoma e (2) sequenciamento. Para a realização da captura do exoma, são utilizados *kits* oferecidos por diversas empresas, como Agilent, Illumina e Roche Nimblegen^[10]. Os *kits* tipicamente têm como objetivo capturar os exons, ou seja, as

sequências das regiões codificadoras consenso de humano (CCDS)^[24], que cobrem aproximadamente 29 Mb do genoma, e todos os exons de genes RefSeq (genes RefSeq são genes bem caracterizados, usados como padrões de referência). Adicionalmente, os *kits* podem também incluir sondas para as regiões flanqueadoras e para microRNAs^[25, 26].

Existem distintos métodos de captura, porém os mais utilizados para exoma humano são os protocolos de captura por hibridação “Captura em solução” e “Captura em Array”^[25], descritos na Figura 1.

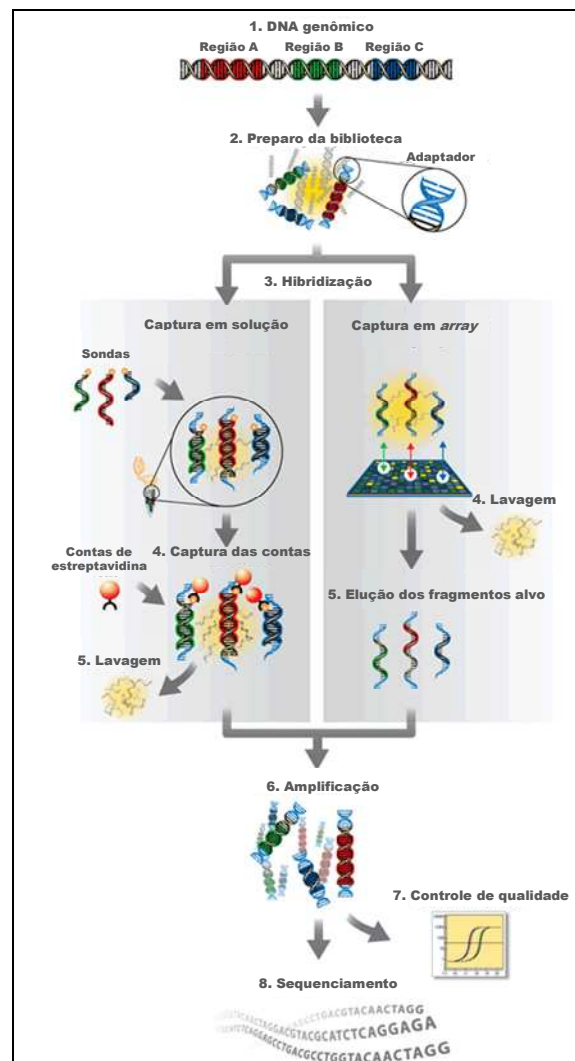


Fig. 1. Funcionamento dos kits comerciais para captura do exoma. Primeiro o DNA genômico total da amostra é fragmentado para construir uma biblioteca. Os fragmentos da biblioteca são, então, flanqueados por adaptadores. No protocolo de “Captura em solução”, sondas de oligonucleotídeos marcadas com biotina são misturadas à biblioteca para capturar as sequências-alvo (exons). As sequências capturadas são assim selecionadas utilizando contas magnéticas de estreptavidina. Já no protocolo de “Captura em array”, a biblioteca é hibridizada

ao *array*, que contém sondas complementares às sequências-alvo (exons). Hibridações não específicas são removidas por lavagem, e as sequências capturadas são recuperadas após eluição do *array*. Os adaptadores são utilizados como *primers* para amplificar as sequências-alvo em uma única PCR. Após análise da qualidade da captura por qPCR, a amostra está pronta para o sequenciamento. Figura retirada do site www.nimblegen.com.

Após a captura, é realizado o sequenciamento do exoma em uma plataforma de sequenciamento de nova geração. Fundamentalmente, o que diferencia o sequenciamento de nova geração do sequenciamento de Sanger é que, no sequenciamento de nova geração, milhares de reações de sequenciamento são realizadas ao mesmo tempo, em paralelo, e por isso essas plataformas têm a capacidade de produzir grandes conjuntos de dados em um reduzido intervalo de tempo^[27].

Peculiarmente, cada plataforma disponível no mercado se baseia em distintos preceitos. A plataforma 454 da Roche utiliza PCR em emulsão e se fundamenta no método de pirosequenciamento. A plataforma Genome Analyzer da Illumina realiza amplificação em fase sólida e sequenciamento por síntese utilizando terminadores reversíveis. Por sua vez, a plataforma Ion Torrent Proton da Life Technologies utiliza PCR em emulsão e sequenciamento por monitoramento de alterações no pH. A plataforma SOLiD (*support oligonucleotide ligation detection*) da Applied Biosystems também usa PCR em emulsão e se baseia no método de sequenciamento por ligação^[revisado por 28, 8, 27]. A metodologia da plataforma SOLiD está sendo descrita com mais detalhes nas Figuras 2 e 3, uma vez que tal plataforma foi utilizada no presente trabalho.

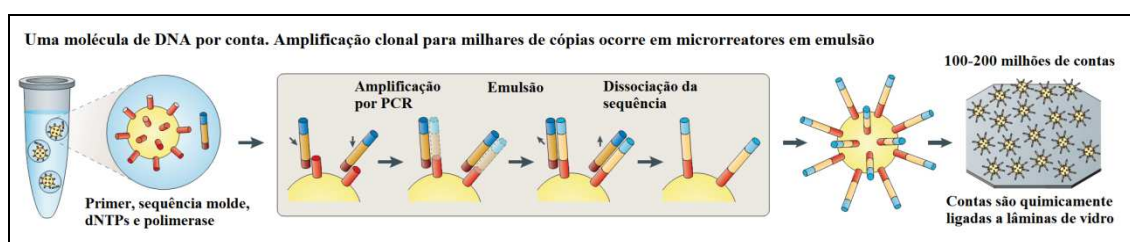


Fig. 2. Funcionamento da PCR em emulsão. Na PCR em emulsão (emPCR), a reação ocorre em uma emulsão óleo-aquosa criada para encapsular os complexos DNA-conta em gotas aquosas individuais. A PCR acontece dentro das gotas; dessa forma, são formadas contas que contêm milhares de cópias da mesma sequência molde. As contas podem ser fixadas quimicamente em lâminas de vidro. Figura modificada de Metzker, 2010.

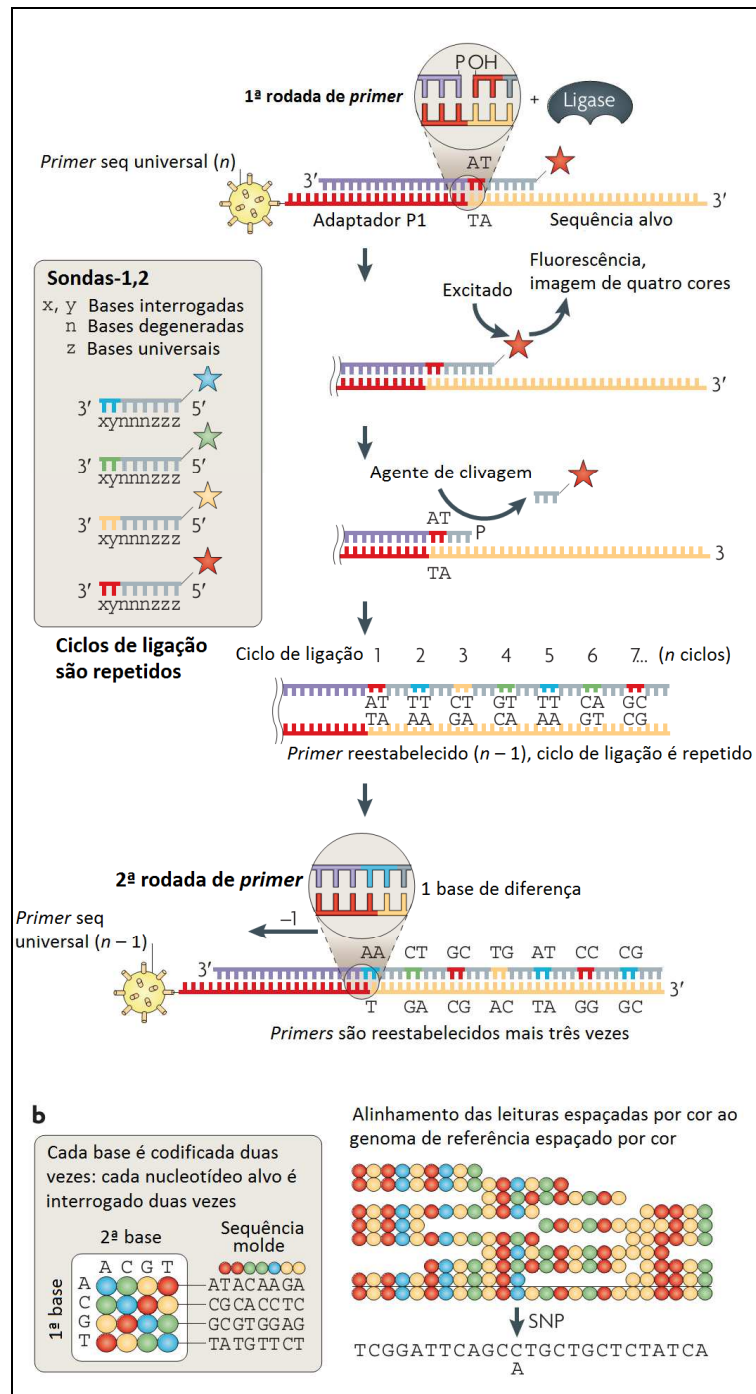


Fig. 3. Metodologia da plataforma SOLiD: sequenciamento por ligação. Utiliza uma biblioteca de sondas nomeada de “Sondas-1,2”, em que uma sequência de oligonucleotídeos com duas bases na região 3’ (bases interrogadas) são associadas com um fluorocromo específico (por exemplo AA, CC, GG e TT são marcados com fluorocromo azul). A biblioteca de sondas-1,2 é adicionada à reação após anelamento do *primer* universal com a sequência do adaptador. Condições apropriadas permitem a hibridação seletiva e a ligação das sondas nas regiões complementares. Após a captura da imagem, a ligação das sondas-1,2 é quimicamente clivada com íons de prata, que geram um grupo 5’-PO₄. O ciclo do SOLiD é repetido mais nove vezes. Após a extensão, o *primer* é retirado, e são realizadas mais quatro rodadas, cada uma com dez

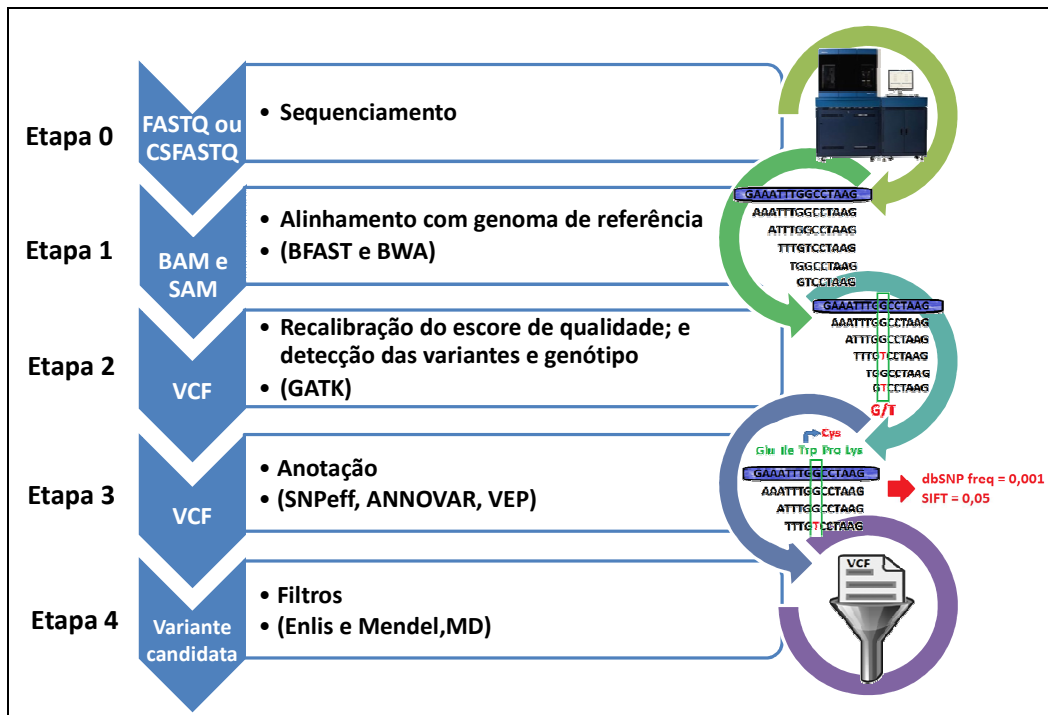


Fig. 5. Etapas da análise bioinformática de exomas. Dentro das setas azuis, são mostrados os formatos dos arquivos gerados em cada etapa. Exemplos de programas utilizados em cada etapa são detalhados entre parênteses, abaixo da descrição do processo realizado em cada etapa.

➤ Etapa 1: alinhamento

A primeira etapa de análise bioinformática utilizando o arquivo FASTQ ou CSFASTQ é fazer o alinhamento com o genoma de referência (hg19/GRCh37). Existem vários *softwares* que fazem esse processo, entre esses o BFAST e o BWA^[29, 30]. O *software Burrows-Wheeler Alignment* (BWA) é um alinhador de sequências que implementa o algoritmo *Burrows-Wheeler Transformation* (BWT), com o objetivo de encontrar o melhor alinhamento com a referência em um tempo computacional aceitável^[34]. Já o *software BLAT-like Fast Accurate Search Tool* (BFAST) realiza o alinhamento com o genoma de referência e utiliza o método de *Smith-Waterman*, o qual faz a inclusão de *gaps* para realizar a detecção de indels^[35]. Após efetuar o alinhamento, esses *softwares* geram arquivos no formato *Sequence Alignment Map* (SAM) e/ou no formato *Binary Alignment Map* (BAM). O SAM é um arquivo *tab-delimited* de texto que contém dados de alinhamento das sequências. Já o BAM é uma versão binária do arquivo SAM e que, por isso, possui tamanho reduzido em relação ao SAM original. O arquivo BAM geralmente é acompanhado de um arquivo index do BAM com extensão BAI; a indexação visa à recuperação rápida de alinhamentos sobrepostos em uma região determinada.

➤ Etapa 2: recalibração do escore de qualidade e detecção de variantes e genótipos

A segunda etapa é a recalibração do escore de qualidade e a detecção das variantes e do genótipo. Esta etapa pode ser realizada utilizando o *software Genome Analysis ToolKit* (GATK), *framework* desenvolvido na plataforma Java[®] pelo Broad Institute, que executa diversos processos, incluindo: (i) o mapeamento inicial de leitura; (ii) o realinhamento local em torno indels; (iii) a recalibração do escore de qualidade das bases; (iv) a detecção das variantes e do genótipo (esta etapa é denominada *call* ou “chamada das variantes”); e (v) a separação de variantes verdadeiras dos artefatos de sequenciamento comuns em tecnologias de sequenciamento de nova geração^[36]. Após realizar todos esses processos utilizando os arquivos SAM/BAM, o *software* GATK gera um arquivo no formato *Variant Call Format* (VCF), ou seja, um arquivo de texto que contém todas as variantes do paciente, quando comparado ao genoma de referência. Um exemplo de arquivo VCF pode ser visualizado na Figura 6.

```
##fileformat=VCFv4.1
##FILTER=<ID=GATKStandard,Description="QD < 10.0 || ReadPosRankSum < -20.0 || SB > -0.01">
##FILTER=<ID=LowQual,Description="Low quality">
##FILTER=<ID=TruthSensitivityTranche99.00to99.90,Description="Truth sensitivity tranche level at VQS Lod: -3.5581 <= x < 1.9202">
##FILTER=<ID=TruthSensitivityTranche99.90to100.00+,Description="Truth sensitivity tranche level at VQS Lod: -181.5813">
##FILTER=<ID=TruthSensitivityTranche99.90to100.00,Description="Truth sensitivity tranche level at VQS Lod: -181.5813 <= x < -3.5581">
##FORMAT=<ID=AB,Number=1,Type=Float,Description="Allele balance for each het genotype">
##FORMAT=<ID=AD,Number=1,Type=Integer,Description="Allelic depths for the ref and alt alleles in the order listed">
##FORMAT=<ID=DP,Number=1,Type=Integer,Description="Read Depth (only filtered reads used for calling)">
##FORMAT=<ID=GQ,Number=1,Type=Float,Description="Genotype Quality">
##FORMAT=<ID=GT,Number=1,Type=String,Description="Genotype">
##FORMAT=<ID=PL,Number=G,Type=Integer,Description="Normalized, Phred-scaled likelihoods for genotypes as defined in the VCF specification">
##INFO=<ID=AB,Number=1,Type=Float,Description="Allele Balance for hets (ref/(ref+alt))">
##INFO=<ID=AC,Number=A,Type=Integer,Description="Allele count in genotypes, for each ALT allele, in the same order as listed">
##INFO=<ID=AF,Number=A,Type=Float,Description="Allele Frequency, for each ALT allele, in the same order as listed">
#CHROM POS ID REF ALT QUAL FILTER INFO FORMAT Exome_Patient_1
chr1 865757 . A T 36.06 TruthSensitivityTranche99.00to99.90 AB=0.772; GT:AB:AD:DP:GQ:PL 0/1:0.77:44,13:59:66:05:66,0,1192
chr1 865776 . C T 42.40 TruthSensitivityTranche99.90to100.00 AB=0.862; GT:AB:AD:DP:GQ:PL 0/1:0.86:50,8:60:72:40:72,0,583
chr1 865777 . C T 67.98 TruthSensitivityTranche99.90to100.00 AB=0.790; GT:AB:AD:DP:GQ:PL 0/1:0.79:49,13:63:97:97:98,0,814
chr1 876499 rs4372192 A G 694.53 PASS AC=2;AF=; GT:AD:DP:GQ:PL 1/1:0,25:25:63.15:728,63,0
chr1 878784 . C G 54.74 PASS AB=0.483; GT:AB:AD:DP:GQ:PL 0/1:0.48:14,15:29:84.73:85,0,299
chr1 880238 rs3748592 A G 1618.26 PASS AC=2;AF=; GT:AD:DP:GQ:PL 1/1:0,43:43:99:1651,129,0
chr1 881627 rs2272757 G A 710.38 PASS AC=2;AF=; GT:AD:DP:GQ:PL 1/1:0,24:24:63.15:743,63,0
chr1 883625 rs4970378 A G 242.23 PASS AC=2;AF=; GT:AD:DP:GQ:PL 1/1:0,7:7:21.07:275,21,0
chr1 887560 rs3748595 A C 686.24 PASS AC=2;AF=; GT:AD:DP:GQ:PL 1/1:0,29:29:66.13:719,66,0
```

Fig. 6. Exemplo de arquivo VCF. O arquivo VCF contém linhas de metainformação (linhas começando com ##), uma linha de cabeçalho (começando com #) e, em seguida, linhas com as informações sobre cada variante, como, por exemplo, sua posição no genoma, os valores de qualidade do sequenciamento naquela posição e os genótipos da amostra para cada posição.

➤ Etapa 3: anotação

A terceira etapa é a anotação das variantes. Nesta etapa, o *software* SnpEff pode ser utilizado para fazer a predição dos efeitos funcionais causados pelas variantes (por

exemplo, se a variante causa troca de aminoácido ou se causa códons finalizadores). O *software* SnpEff também classifica as variantes com base no seu impacto (Tab. 1)^[37].

Tabela 1. Classificação do impacto da variante realizada pelo SnpEff. Retirado de http://snpeff.sourceforge.net/SnpEff_manual.html

Impacto	Significado	Exemplo
Alto	Variantes com impacto alto (perturbador) na proteína, causando provavelmente a truncagem da proteína, perda de função ou provocando o <i>nonsense mediated decay</i> .	Variantes que causam mudança na matriz de leitura (<i>frameshift_variant</i>) ou códons finalizadores (<i>stop_gained</i>)
Moderado	Variantes que possam alterar a eficácia da proteína.	Variantes que causam troca de aminoácido (<i>missense_variant</i>) ou deleções de um ou mais códons (<i>inframe_deletion</i>)
Baixo	Variantes que são assumidas como na sua maioria inofensivas ou improváveis de mudar o comportamento de proteínas.	Variantes sinônimas (sem troca de aminoácido) (<i>synonymous_variant</i>)
Modificador	Normalmente variantes em posições não codificantes ou variantes que afetam genes não codificantes, em que as previsões são difíceis ou não há evidência de impacto.	Variantes em introns (<i>intron_variant</i>) ou nas regiões 3'UTR (<i>3_prime_UTR_variant</i>)

Já o *software Annotate Variation* (ANNOVAR)^[38] pode ser utilizado para detectar se as variantes estão presentes em bancos de dados públicos, como o *Single Nucleotide Polymorphism database* (dbSNP)^[39], *1000 Genomes Project*^[40] e National Heart, Lung, and Blood Institute – *Exome Sequencing Project* (NHLBI ESP)^[41]:

– dbSNP é um banco de dados público desenvolvido pelo National Center for Biotechnology Information (NCBI) em colaboração com o National Human Genome Research Institute (NHGRI), criado em 1998, com o objetivo de armazenar dados sobre variantes genéticas encontradas em diferentes espécies^[39]. Pesquisadores de instituições públicas e privadas podem incluir variantes no dbSNP. Após a inclusão, o NCBI vai adicionar um identificador de referência rsID (por exemplo, rs25) em cada variante. A última *build* do dbSNP para humanos é a versão 142, publicada em outubro de 2014, a qual contém aproximadamente 112,7 milhões de variantes.

– O *1000 Genomes Project* é um consórcio internacional, criado em 2008, que têm como objetivo estabelecer um catálogo extensivo de SNPs, variantes e variações estruturais do genoma humano, utilizando técnicas de sequenciamento de nova geração^[40]. Esse projeto foi baseado no modelo de colaboração internacional usado no *HapMap Project*, lançado em 2002, que detectou variantes comuns no genoma humano

através de técnicas de genotipagem^[42]. No projeto piloto do *1000 Genomes Project*, foi realizado o sequenciamento de aproximadamente 1.000 indivíduos, o que originou o nome do projeto (entre esses indivíduos, estão também amostras utilizadas no *HapMap Project*). Atualmente, porém, na fase 3 do projeto, já estão disponíveis online dados de 2.504 indivíduos de 26 populações diferentes.

– ESP é um projeto que realizou o sequenciamento de 6.503 amostras; por isso, a sigla da fase final do projeto é ESP6500^[41]. As amostras são originadas de populações dos Estados Unidos bem estudadas fenotipicamente; em geral, são amostras de estudos de desordens cardíacas, vasculares e de doenças pulmonares. O conjunto completo de SNPs do projeto *NHLBI ESP* está incluído no dbSNP *build* 138.

O *software* ANNOVAR também disponibiliza escores pré-computados de preditores que realizam estimativas quantitativas sobre o impacto que as variantes podem causar na função da proteína, tais como o *Sorting Intolerant From Tolerant* (SIFT)^[43] e o *Polymorphism Phenotyping v2* (PolyPhen2)^[44] (descritos na próxima etapa). O *software* *Variant Effect Predictor* (VEP)^[45], ferramenta disponível pelo Ensembl, também pode ser utilizado para, entre outras funções, disponibilizar o escore das variantes pelo SIFT e pelo PolyPhen2.

➤ **Etapa 4: uso de filtros para identificar a variante responsável**

A quarta etapa é a de filtrar as variantes do exoma, ou seja, analisar os exomas para encontrar as variantes candidatas. Esta é a etapa mais trabalhosa e minuciosa. Isso porque a análise dos dados do exoma é um grande desafio, principalmente em razão da falta de conhecimento sobre as funções e interações dos genes, resultando em um grande número de variantes de significado desconhecido (*variants of uncertain significance – VUS*)^[46]. Outra limitação é a grande quantidade de polimorfismos não patogênicos encontrados no exoma de um indivíduo e os erros de sequenciamento. Yang *et al.*^[47] utilizaram duas plataforma da Illumina para sequenciar exomas de 250 pacientes e encontraram aproximadamente 200.000 a 400.000 variantes de um único nucleotídeo (SNVs) e indels em cada exoma analisado. Das variantes eleitas como potencialmente candidatas a serem causadoras da doença, 14% se mostraram falso positivas pelo sequenciamento de Sanger^[47]. Já Wall *et al.*^[48] estimaram uma taxa de erros de sequenciamento de aproximadamente 0,1% para SNVs utilizando a plataforma Illumina HiSeq para o sequenciamento de exoma. Os autores observaram que a taxa de erros de sequenciamento geralmente é maior (até 6%) para variantes raras.

A maioria dos erros de leitura do sequenciamento de nova geração são as indels^[49]. Isso porque a sensibilidade e a especificidade dos *softwares* disponíveis para a detecção de variantes não são satisfatórias, especialmente nos casos das indels^[50]. Ao analisar um exoma, há três indicadores que podem ser considerados ao avaliar se a variante candidata em questão é provavelmente real ou se é um falso-positivo: profundidade de leitura, qualidade do sequenciamento na posição da variante e o balanço alélico (variantes heterozigotas têm um balanço alélico com aproximadamente 50% das bases iguais à referência e 50% mutadas; já variantes homozigotas têm um balanço alélico com aproximadamente 100% das bases mutadas). Todavia, é importante salientar que outros métodos de biologia molecular, como o sequenciamento de Sanger, devem ser realizados para confirmar as variantes candidatas.

Outro fato que dificulta a análise de exomas é a “promiscuidade” nas relações genótipo-fenótipo, a qual foi revelada recentemente com a utilização das técnicas de sequenciamento de nova geração: altos níveis de heterogeneidade alélica (diferentes mutações em um gene), heterogeneidade de locus (mutações em diferentes genes causando fenótipos similares) e heranças digênicas ou oligogênicas com alelos modificadores ou causais secundários (várias variantes que, em conjunto, causam ou agravam o fenótipo)^[revisado por 51].

Para facilitar a análise de exomas, existem vários filtros que podem ser utilizados para selecionar a variante candidata possivelmente relacionada com a doença em estudo^[52, 53], os quais são descritos abaixo:

- **Comparação de exomas de indivíduos com o mesmo fenótipo e uso de controles**

A maioria dos estudos faz comparações das variantes encontradas nos exomas de alguns indivíduos sem parentesco afetados pela doença. Essa comparação tem o objetivo de identificar variantes novas e raras no mesmo gene. Para identificar se a variante é nova, as variantes do indivíduo são comparadas com polimorfismos não patogênicos disponíveis em bancos de dados públicos, tais como o dbSNP^[39], o *1000 Genomes Project*^[40] e o *NHLBI ESP*^[41], já mencionados anteriormente. Esses tipos de filtros são utilizados para eliminar genes candidatos; os polimorfismos encontrados na amostra controle provavelmente não são relacionados com a doença em estudo. Tal estratégia é muito usada, uma vez que somente cerca de 2% das SNVs identificadas no exoma de um indivíduo são novas^[53].

Ao considerar as variantes disponíveis em bancos de dados, uma maneira de distinguir polimorfismos comuns na população de variantes raras é observando o valor da frequência do alelo menos comum (*MAF* – *minor allele frequency*). O *MAF* é a frequência em que ocorre o alelo menos comum em uma população. Alelos muito comuns apresentam $MAF \geq 5\%$, e alelos raros apresentam $MAF \leq 1\%$ ^[53, 54].

Dois exemplos de estudos que utilizaram esses filtros para identificar as variantes responsáveis por doenças monogênicas são os estudos de *Ng et al.*^[11] em 2009 e de *Ng et al.*^[6] em 2010. *Ng et al.*^[11] sequenciaram o exoma de 12 indivíduos: oito indivíduos previamente estudados no *HapMap Project* e quatro indivíduos sem parentesco afetados pela síndrome de Freeman-Sheldon (FSS), desordem rara com padrão de herança dominante. Utilizando os exomas dos oito indivíduos do *HapMap* e as variantes do dbSNP *build* 129 como controles, os autores demonstraram que foi possível a identificação correta do gene *MYH3*, identificado previamente como responsável por FSS^[55].

Já *Ng et al.*^[6] demonstraram a primeira aplicação do sequenciamento do exoma para descobrir o gene relacionado com uma doença mendeliana de causa desconhecida; a síndrome de Miller, que tem padrão de herança recessivo. Foram sequenciados os exomas de quatro indivíduos afetados de três famílias diferentes. A identificação do gene candidato *DHODH* foi possível ao filtrar os resultados utilizando os exomas de oito indivíduos do *HapMap* e as variantes do dbSNP129. O sequenciamento de Sanger de três famílias adicionais com a síndrome de Miller confirmou a presença de mutações de troca de aminoácidos heterozigotas compostas em *DHODH*.

O aumento do número de dados de exomas ou genomas bem documentados disponíveis publicamente vai permitir o uso de milhares de controles, o que pode facilitar a detecção das variantes responsáveis, mesmo se o número de pacientes analisados com a doença for limitado. De fato, vários estudos foram bem-sucedidos ao identificar as causas de doenças genéticas vistas em somente um paciente^[56]. Esse tipo de análise, porém, só pode ser realizada se as amostras controle ou do banco de dados não apresentarem a doença estudada. Os pesquisadores precisam estar cientes, por exemplo, de que os participantes do ESP, assim como explicado anteriormente, não são todos completamente saudáveis. Outro fato que deve ser considerado é que o dbSNP não contém apenas polimorfismos neutros, e sim uma pequena porção de variantes patogênicas raras^[39, 57]. Além disso, o banco de dados *1000 Genomes Project* pode considerar como indivíduos controle pessoas portadoras de doenças recessivas ou com

penetrância incompleta, sendo que essas vão carregar um dos alelos patogênicos e não apresentarão fenótipo.

Em alguns estudos, são sequenciadas amostras de vários indivíduos controle, que certamente não são afetados pela doença, em vez de usar os bancos de dados públicos. Ao utilizar indivíduos controle oriundos da mesma população que a amostra, podemos filtrar também por variantes comuns na população em estudo^[53]. Além disso, deve-se considerar que exomas de indivíduos sequenciados utilizando diferentes *kits* de captura/plataformas de sequenciamento e analisados com diferentes *softwares* de anotação e de detecção de variantes vão apresentar diferenças sistemáticas em seus resultados^[58]. Desta forma, o ideal é empregar como controles exomas de indivíduos sequenciados com o mesmo *kit* de captura/plataforma e analisados com os mesmos *softwares*.

- **Seleção de variantes candidatas com base no seu impacto**

As variantes candidatas também podem ser selecionadas com base na sua patogenicidade, ou seja, na alteração causada na proteína em questão. Por exemplo, variantes que causam mudança na matriz de leitura, códons finalizadores e alterações em sítios de *splice* canônicos são mais patogênicas que variantes silenciosas^[6]. Ademais, os genes candidatos podem ser selecionados de acordo com a função do gene em questão, e genes com funções relacionadas com o fenótipo da doença (ou que codificam proteínas que interagem com outras em um *pathway* relacionado com o fenótipo) apresentam maior probabilidade de serem os genes responsáveis. Para fazer a busca pela função dos genes candidatos, podem-se utilizar sites como GeneCards^[59] e UniProt^[60]. Em alguns casos, a função do gene candidato ainda não é conhecida (ou é pouco conhecida), e mutações nesse gene não são descritas em humanos; nestes casos, pode-se procurar por mutações descritas em outros organismos modelos e se tais mutações causam um fenótipo similar nesses organismos. O banco de dados *Mouse Genome Informatics* (MGI)^[61] é útil para procurar por mutações em camundongos e seu fenótipo.

No caso da realização do sequenciamento do exoma para diagnóstico, pode-se realizar a filtragem procurando por variantes em genes que são causalmente relacionados com doenças genéticas humanas. Podem-se utilizar filtros que selecionam genes descritos nos bancos de dados, tais como: *Online Mendelian Inheritance in Man* (OMIM)^[62], *Human Gene Mutation Database* (HGMD)^[63], ClinVar^[64] ou *Clinical*

Genomic Database (CGD)^[65]. Alguns artigos também oferecem listas de genes relacionados com doenças mendelianas, e essas listas também podem ser empregadas^[66].

Outra maneira de selecionar as variantes candidatas é realizando estimativas quantitativas do impacto da variante na função da proteína. Alguns *softwares* disponíveis na internet, como o phyloP^[67] e o *Genetic Evolutionary Rate Profiling* (GERP)^[68], analisam o impacto da variante de acordo com a observação de que mutações patogênicas se localizam em regiões muito conservadas evolutivamente. Existem também as ferramentas que analisam as mutações não sinônimas, considerando também as alterações físico-químicas da proteína causadas pela substituição do aminoácido, tais como o SIFT^[43], o PolyPhen2^[44], o *Mutation Taster*^[69] e o Align-GVGD^[70].

Os escores do SIFT e do PolyPhen2 variam de 0,00 a 1,00, sendo que variantes com escores próximos de 0,00 pelo SIFT e com escores próximos de 1,00 no PolyPhen2 têm mais probabilidade de serem patogênicas. Já o *Mutation Taster* prediz que as variantes são “polimorfismos” ou “causadoras de doenças”, e o Align-GVGD classifica as variantes em categorias de C0 a C65; variantes classificadas na categoria C65 têm mais probabilidade de serem patogênicas. Basicamente, variantes têm grande probabilidade de serem patogênicas quando apresentam escores menores que 0,05 pelo SIFT, maiores que 0,85 pelo PolyPhen2, são classificadas como “causadora de doenças” pelo *Mutation Taster* e são classificadas na categoria C65 pelo Align-GVGD.

É importante salientar que o desempenho geral desses preditores está aquém do esperado; por exemplo, os preditores SIFT e Polyphen2 apresentam sensibilidade relativamente alta (cerca de 70%-80%), mas apresentam baixa especificidade (cerca de 15%-40%)^[71, 72]. Desta forma, tais preditores geralmente erram mais ao predizer que variantes neutras são deletérias. Para melhorar o nível de acertos das predições, uma técnica comumente utilizada é a de adotar múltiplos preditores na avaliação do impacto da variante. Consequentemente, foram desenvolvidos *softwares* que combinam os resultados de vários preditores gerando um escore de probabilidade consenso; assim sendo, esses *softwares* têm demonstrado resultados mais confiáveis^[73, 74, 75].

Adicionalmente, as análises computacionais de modelagem de proteínas *in silico* também podem ser realizadas de forma a predizer o efeito da mutação na estabilidade da proteína, nas ligações de hidrogênio, na dinâmica conformacional e na atividade, entre outras propriedades fisiologicamente importantes para a proteína^[revisado por 76].

- **Efeito do tipo de herança no estudo e considerações ao decidir quais membros da família serão sequenciados**

O tipo de herança da desordem monogênica em estudo tem grande influência no *design* do experimento e na parte analítica dele, influenciando, por exemplo, o número de casos que devem ser sequenciados e na seleção dos casos mais informativos. Estudos de desordens recessivas, em geral, necessitam menos indivíduos sequenciados do que os estudos de desordens dominantes para localizar o gene responsável. Isso ocorre porque o genoma de qualquer indivíduo tem aproximadamente 50 vezes menos genes com duas variantes novas que alteram a estrutura e/ou função da proteína do que com apenas uma^[53].

O sequenciamento de alguns indivíduos do mesmo *pedigree* afetados pela doença estudada pode estreitar a busca pela variante responsável. A escolha dos indivíduos afetados mais informativos para realizar o sequenciamento é muito importante. Por exemplo, em famílias em que múltiplos indivíduos são afetados por uma doença rara com herança possivelmente autossômica dominante, uma estratégia é a de sequenciar amostras de indivíduos genealogicamente mais distantes: quanto mais distantes os indivíduos, menos variantes e polimorfismos eles terão em comum^[20].

Quando dados de mapeamento são disponíveis, a estratégia mais eficiente é a de sequenciar um par de indivíduos afetados nos quais seus haplótipos sobrepostos produzem a menor região genômica em comum. Para *pedigrees* consanguíneos com suspeita de padrão de herança recessiva, provavelmente é suficiente o sequenciamento de uma única pessoa com a menor região de homozigosidade determinada pela genotipagem^[77, 78]. Nessas duas estratégias, o sequenciamento do exoma é utilizado para substituir o sequenciamento de Sanger de todos os genes no intervalo crítico.

O sequenciamento do exoma de trios compostos por pais e filho é uma metodologia efetiva para identificar mutações codificantes *de novo*. Isso ocorre porque eventos *de novo* múltiplos que acontecem em um gene específico (ou dentro de uma família de genes ou *pathways*) são extremamente raros^[79].

SOFTWARES UTILIZADOS PARA A DETECÇÃO DE VARIANTES CANDIDATAS

A fim de filtrar os milhares de variantes presentes nos exomas e detectar variantes candidatas, encontram-se disponíveis comercialmente *softwares* como o *Enlis Genome* (fornecido pela empresa Enlis Genomics)^[80]. Esse *software* permite a análise de múltiplos exomas, genomas e de resultados de painéis de sequenciamento de genes

específicos e de SNP array. Para fazer a análise de exomas, o usuário (pesquisador ou médico) deve fazer o *upload* do arquivo VCF do seu paciente. O *software Enlis Genome* vai fazer uma nova anotação das variantes presentes no VCF e, então, converterá o arquivo VCF em um arquivo GENOME, que é o arquivo de leitura do *Enlis Genome*.

Esse *software* apresenta várias ferramentas, como: filtrar as variantes, procurar por variantes heterozigotas compostas e procurar por regiões de homozigose. A interface da ferramenta para filtrar variantes está exemplificada na Figura 7. Com essa ferramenta, o usuário pode selecionar as variantes em um determinado paciente utilizando vários filtros, como: selecionar as variantes em genes que causem doenças descritas no OMIM; variantes com MAF < que 1% nos bancos de dados *1000 Genomes Project* e *Exome Sequencing Project*; selecionar variantes com base no impacto que vão causar na proteína (por exemplo, excluindo variantes sinônimas); selecionar variantes que causem mudanças de aminoácidos, preditas como deletérias (baseado em escore de SIFT <= 0,05), entre outros.

The screenshot displays the 'Variation Filter' interface of the Enlis Genome software. At the top, there are buttons for 'Load Filter Set' and 'Save Current Filter Set'. Below this, several filter panels are visible: 'IN Single Genome' (set to '40: mm13304'), 'Omin Genes' (search within OMIM database), 'Allele Frequency, Global' (set to '<' and '1 %'), 'Protein Impact' (with a 'Select all' button), and 'Protein Predicted Deleterious' (set to 'Yes'). Below these panels, there's a section for 'Add Variation Filter' with various checkboxes and buttons. At the bottom, a table shows the results of the filter: 'Result: 766 variations at 756 positions'. The table has columns for 'Chromosome Pos.', 'Var Type (zyg)', 'Ref Seq', 'Var Seq', 'Gene(s)', 'Gene Component', 'Protein Impact', 'AlleleFreq', 'M.Cons', and 'Clinical Sig.'. The first few rows of the table are as follows:

Chromosome Pos.	Var Type (zyg)	Ref Seq	Var Seq	Gene(s)	Gene Component	Protein Impact	AlleleFreq	M.Cons	Clinical Sig.
1: 3,342,705	SNP (x1)	A	G	PRDM16	CDS	MISSENSE	0.00%	Yes (4.4)	unk.
1: 9,782,661	SNP (x1)	T	G	PIK3CD	CDS	MISSENSE	0.00%	Yes (5.3)	unk.
1: 10,318,636	SNP (x1)	A	G	KIF1B	CDS	MISSENSE	0.00%	Yes (5.6)	unk.
1: 16,375,613	SNP (x1)	A	G	CLCNKB	ACCEPTOR CDS	DISRUPT SYNONYMOUS	0.00%	Yes (4.0)	unk.
1: 16,378,297	SNP (x1)	G	A	CLCNKB	CDS	MISSENSE	0.00%	Yes (3.2)	unk.
1: 20,964,396	SNP (x1)	G	A	PINK1	CDS	MISSENSE	0.00%	Yes (5.4)	unk.
1: 22,154,620	SNP (x1)	A	G	HSPG2	CDS	MISSENSE	0.00%	Yes (4.4)	unk.
1: 22,180,667	SNP (x1)	A	G	HSPG2	CDS	MISSENSE	0.00%	Yes (5.1)	unk.
1: 22,192,227	SNP (x1)	G	A	HSPG2	CDS	MISSENSE	0.00%	Yes (4.9)	unk.
1: 22,202,444	SNP (x1)	A	G	HSPG2	CDS	MISSENSE	0.00%	Yes (5.5)	unk.
1: 23,236,942	SNP (x1)	A	G	EPHB2	CDS	MISSENSE	0.00%	Yes (4.6)	unk.
1: 25,883,684	SNP (x1)	T	C	LDLRAP1	CDS	MISSENSE	0.00%	Yes (5.6)	unk.

At the bottom of the table, it says 'Showing result 1 - 756 out of 756' and 'Page number: 1 of 1 Previous Next'.

Fig. 7. Interface da ferramenta de filtrar as variantes do *software Enlis Genome*. Nesse exemplo, foram filtradas 766 variantes no paciente “mm13304” que são raras, causam impacto na proteína e são preditas como deletérias e em genes descritos no OMIM.

No Laboratório de Genômica Clínica, desenvolvemos o *software Mendel,MD* (RGCCCL Cardenas e SDJ Pena, manuscrito sendo preparado para publicação), *software* similar ao *Enlis Genome*, já que tem como objetivo principal filtrar as variantes do exoma. Ao contrário, porém, do *Enlis Genome*, o *Mendel,MD* será uma ferramenta

open-source e gratuita. Atualmente utilizamos o Mendel,MD online; contudo, futuramente, ele poderá também ser baixado e instalado no computador do usuário e executado localmente. Para utilizá-lo, o usuário deverá fazer o *upload* do arquivo VCF do exoma a ser analisado. O *software* fará, então, a anotação das variantes do arquivo VCF utilizando diferentes ferramentas (como o SnpEff^[37] e o VEP^[45]), vai encontrar as MAFs das variantes nos bancos de dados dbSNP141, *1000 Genomes Project* e ESP6500 e irá verificar se as variantes estão presentes nos bancos OMIM, HGMD, ClinVar ou CGD. Ao final da anotação, o Mendel,MD vai gerar um novo arquivo VCF, que conterá todas as informações de cada variante.

O Mendel,MD oferece várias ferramentas como a “Análise de filtros” (*Filter Analysis*), “Um click” (*One Click*), “Análise de famílias” (*Family Analysis*), “Análise de *pathway*” (*Pathway Analysis*) e “Comparação” (*Comparison*). Com a ferramenta *Filter Analysis*, o usuário pode selecionar vários filtros e valores para fazer a seleção das variantes. Ao utilizar a ferramenta *One Click*, os filtros e os valores já estão preestabelecidos, e o usuário precisa apenas selecionar os indivíduos que quer analisar e clicar em “*submit*” para filtrar as variantes. Os filtros e os valores preestabelecidos para selecionar variantes candidatas no *One Click* são:

- Incluir apenas as variantes presentes em genes comuns entre todos os indivíduos selecionados;
- Excluir variantes que foram adicionadas no dbSNP até a *build* 129 (já que a *build* 129 é considerada a última versão do dbSNP contendo apenas variantes comuns; as variantes raras do *1000 Genomes Project* foram incluídas a partir da *build* 130)^[81];
- Mostrar apenas variantes com profundidade de leitura igual ou superior a 10 (para excluir erros de leitura);
- Incluir apenas variantes com impacto moderado ou alto pelo SnpEff;
- E selecionar apenas variantes com MAF menor de 0,5% nos bancos de dados *1000 Genomes Project*, dbSNP141 e ESP6500.

Os filtros e os valores preestabelecidos são indicados para começar a análise dos exomas, mas o usuário pode modificar os valores de tais filtros ou incluir novos para filtrar ainda mais as variantes (por exemplo, utilizando os preditores de patogenicidade SIFT e PolyPhen2). Adicionalmente, o usuário pode optar por utilizar os filtros e campos preestabelecidos, de acordo com o tipo de herança em estudo (recessiva homozigota, recessiva heterozigota composta, dominante heterozigota, ligada ao X

recessiva homozigota e ligada ao X dominante heterozigota). A Figura 8 mostra a interface da ferramenta *One Click*.

Fig. 8. Interface da ferramenta *One Click* do software Mendel,MD. Neste exemplo, serão selecionadas as variantes com padrão de herança recessivo homozigoto do paciente “mm13304”.

A ferramenta *Family Analysis* do Mendel,MD é utilizada para comparar as variantes de uma família, sendo que podem ser analisados vários pacientes afetados. É uma ferramenta muito útil para procurar variantes heterozigotas compostas no filho (uma variante é herdada do pai, e a outra, da mãe), variantes em homozigose no filho e em heterozigose nos pais, variantes *de novo* no filho (não estão presentes nos pais), e ligadas ao cromossomo X no filho (em hemizigose no filho, não presente no pai e em heterozigose na mãe).

Utilizando a ferramenta *Pathway Analysis*, o usuário pode selecionar por variantes em genes em um *pathway* específico do banco de dados *Kyoto Encyclopedia of Genes and Genomes* (KEGG)^[82]. Já a ferramenta *Comparison* pode ser utilizada para comparar dois exomas e analisar quantas variantes eles têm em comum. O software Mendel,MD oferece também estatísticas de quantidade de variantes em cada indivíduo, sobre a qualidade das variantes e a cobertura do sequenciamento.

Outro software comercial muito útil para a análise das exomas é o Alamut Visual (fornecido pela empresa Interactive Biosoftware)^[83]. O Alamut Visual é um software que integra a informação genética de diferentes fontes, como o ClinVar, o

dbSNP, o ESP, entre outros, em um ambiente consistente, que pode ser utilizado para descrever as variantes usando nomenclatura recomendada pela Human Genome Variation Society (HGVS)^[84] e ajudar a interpretar o seu *status* patogênico.

Desse modo, esse *software* pode ser empregado para avaliar as variantes candidatas selecionadas utilizando os *softwares* *Enlis Genome* ou *Mendel,MD*. É uma ferramenta muito fácil de utilizar: o usuário precisa apenas escrever a posição da variante a ser analisada na janela de busca (exemplo chr3:62180793) e selecionar o transcrito do gene que ele quer analisar. O Alamut Visual vai mostrar, então, uma interface gráfica única, revelando todos os dados disponíveis para aquela variante (Fig. 9). Dessa maneira, o usuário pode criar uma mutação na posição e visualizar os dados da variante, como frequência nos bancos de dados e escores de predição de patogenicidade (como os escores do Align-GVGD, do SIFT, do *Mutation Taster*, entre outros). Para fazer a predição de patogenicidade, o Alamut Visual utiliza sequências de organismos ortólogos, o que faz com que as predições sejam mais confiáveis^[72]. No Alamut Visual, também é possível visualizar as *reads* na posição da variante e analisar o balanço alélico da mutação; para tal, é necessário fazer o *upload* do arquivo BAM no *software* (Fig. 10).

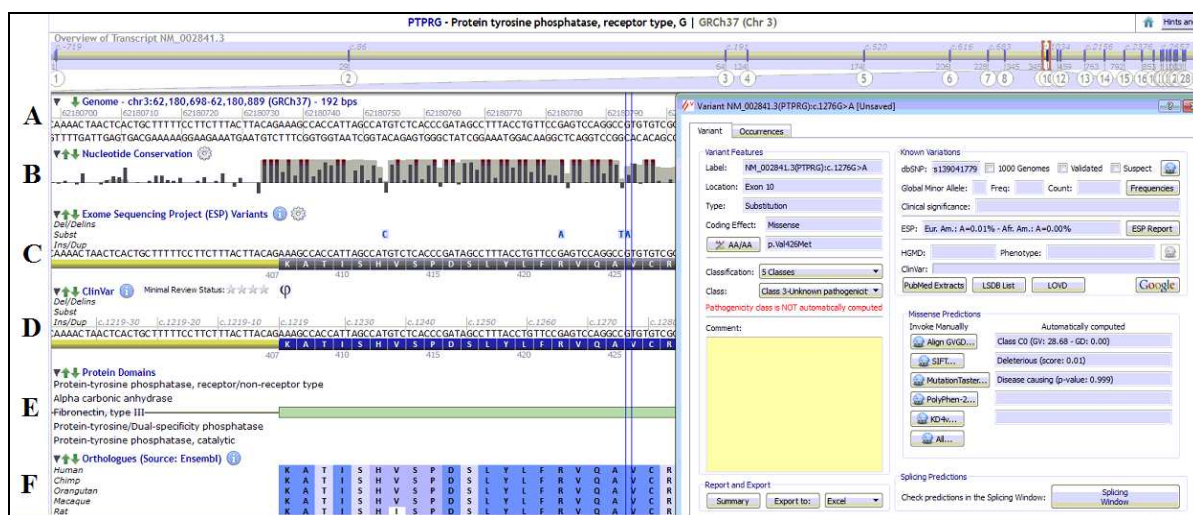


Fig. 9. Interface do software Alamut Visual. Nesse exemplo, é mostrada a variante c.1276G>A na posição chr3:62180793 do gene *PTPRG*. É possível visualizar, à esquerda da figura, de cima para baixo: A) Posição genômica da variante; B) Conservação dos nucleotídeos por preditores como o PhyloP; C) variantes reportadas no ESP; D) variantes reportadas pelo ClinVar; E) domínios proteicos da proteína *PTPRG*; F) conservação dos aminoácidos em organismos ortólogos. Já, à direita da figura, está aberta uma janela que informa vários dados da

variante em análise, como frequência nos bancos de dados e escores de predição de patogenicidade.

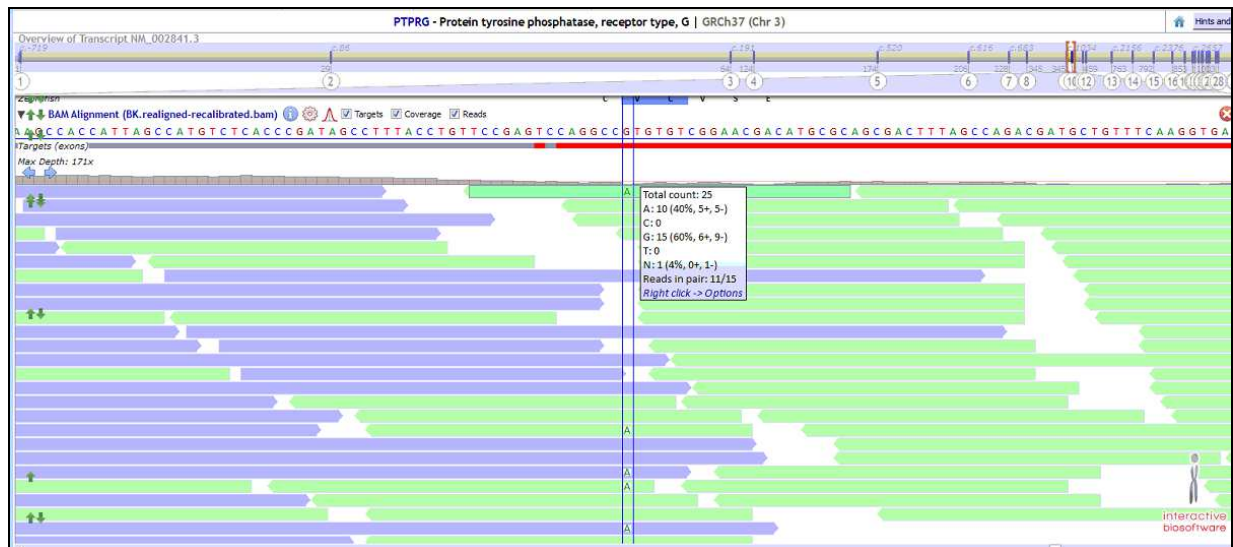


Fig. 10. Interface do software Alamut Visual após fazer o upload do arquivo BAM do paciente. Nesse exemplo, é mostrada que a variante c.1276G>A na posição chr3:62180793 do gene *PTPRG* tem o seguinte balanço alélico: 15 bases G iguais à referência e 10 bases A mutadas.

LIMITAÇÕES E VANTAGENS DA TÉCNICA DE SEQUENCIAMENTO DO EXOMA

Desde a sua introdução, em 2009, o exoma foi utilizado em vários projetos visando à identificação de genes que, quando sofrem algum tipo de mutação, causam doenças mendelianas^[53, 56]. É importante salientar, porém, que toda técnica tem sua limitação, e, em algumas situações, a técnica de sequenciamento do exoma pode falhar em encontrar os genes causais. Isso pode ocorrer em razão de vários motivos: o gene causal pode não apresentar uma sonda complementar no *kit* utilizado (por exemplo, se o gene não for conhecido como gene); pode haver cobertura inadequada da região que contém a variante causal (por exemplo, devido à captura ou sequenciamento inadequado); e a variante causal pode estar localizada fora da região dos exons^[53]. Além disso, o sequenciamento do exoma é menos preciso ou não detecta alguns tipos de variações genômicas, como variantes de número de cópia (CNVs), deleções de um exon inteiro, mutações por expansão de microssatélites, variações estruturais (como translocações balanceadas), aneuploidias e variações epigenéticas^[20].

Entretanto, entre as diversas vantagens previamente descritas, se destacam: é realizada uma análise refinada do genoma, permitindo a detecção de SNVs e indels de

oito a dez nucleotídeos ou menores; e não é necessário que se defina um gene-alvo para que esse seja testado isoladamente (já que são analisados, em teoria, todos os 20.000 genes humanos). Além disso, sua eficácia foi comprovada, uma vez que a técnica vem sendo aplicada com sucesso na pesquisa e no diagnóstico de doenças genéticas^[13].

O diagnóstico correto dos pacientes por meio da análise de exomas poderá fornecer às famílias elementos científicos sólidos para lidar com a doença, orientando quanto à sua evolução e eventuais complicações e ajudando a definir um prognóstico reprodutivo para a família (com base no risco recorrência em gestações futuras do casal). Em alguns casos, o diagnóstico correto pode levar a um tratamento específico ou a uma estratégia de conduta que muda drasticamente a evolução clínica do paciente^[6, 85]. Ademais, nos casos em que a descoberta do gene causador não altera a conduta clínica, o tratamento ou o prognóstico, o diagnóstico ainda pode ser útil porque gera conforto para a família ao descobrir o “motivo da doença” e encerra uma odisseia de exames caros, potencialmente invasivos e estressantes^[20].

Recentemente, Yang *et al.*^[86] mostraram que, dos 2000 exomas solicitados na clínica para diagnóstico de pacientes com fenótipo sugestivo de doença genética, foram encontradas as variantes possivelmente responsáveis em 25% dos casos (504 pacientes). Essa taxa de sucesso no diagnóstico de 25% é a maior no campo dos testes genômicos utilizados na clínica: exames com base em *array* (incluindo o *array*-CGH e o SNP *array*) possibilitam o diagnóstico de 15% a 20% dos casos, e o cariótipo detecta alterações em cerca de 3% dos casos não associados a síndromes cromossômicas reconhecíveis clinicamente (como a síndrome de Down)^[87].

REFERÊNCIAS

1. LINHARES, N. D.; SVARTMAN, M.; VALADARES, E. R. Diagnóstico citogenético de pacientes com retardo mental idiopático. *J Bras Patol Med Lab*, v. 48, n. 1, p. 33-39, 2012.
2. KORF, B. R.; REHM, H. L. New approaches to molecular diagnosis. *JAMA*, v. 309, n. 14, p. 1511-21, 2013.
3. KAISER, J. Human genetics. Affordable 'exomes' fill gaps in a catalog of rare diseases. *Science*, v. 330, n. 6006, p. 903, 2010.
4. ROBINSON, P. N.; KRAWITZ, P.; MUNDLOS, S. Strategies for exome and genome sequence data analysis in disease-gene discovery projects. *Clin Genet*, v. 80, n. 2, p. 127-32, 2011.

5. CHOI, M. *et al.* Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. *Proc Natl Acad Sci U S A*, v. 106, n. 45, p. 19096-101, 2009.
6. NG, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet*, v. 42, n. 1, p. 30-5, 2010.
7. International Human Genome Consortium. Finishing the euchromatic sequence of the human genome. *Nature*, v. 431, n. 7011, p. 931-45, 2004.
8. METZKER, M. L. Sequencing technologies - the next generation. *Nat Rev Genet*, v. 11, n. 1, p. 31-46, 2010.
9. WETTERSTRAND, K. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP). Disponível em: www.genome.gov/sequencingcosts.
10. MAMANOVA, L. *et al.* Target-enrichment strategies for next-generation sequencing. *Nat Methods*, v. 7, n. 2, p. 111-8, 2010.
11. NG, S. B. *et al.* Targeted capture and massively parallel sequencing of 12 human exomes. *Nature*, v. 461, n. 7261, p. 272-6, 2009.
12. BOTSTEIN, D.; RISCH, N. Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet*, v. 33 Suppl, p. 228-37, 2003.
13. KU, C. S. *et al.* Exome sequencing: dual role as a discovery and diagnostic tool. *Ann Neurol*, v. 71, n. 1, p. 5-14, 2012.
14. KU, C. S.; NAIDOO, N.; PAWITAN, Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet*, v. 129, n. 4, p. 351-70, 2011.
15. DAUBER, A. *et al.* Whole exome sequencing reveals a novel mutation in CUL7 in a patient with an undiagnosed growth disorder. *J Pediatr*, v. 162, n. 1, p. 202-4 e1, 2013.
16. KRYUKOV, G. V.; PENNACCHIO, L. A.; SUNYAEV, S. R. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. *Am J Hum Genet*, v. 80, n. 4, p. 727-39, 2007.
17. RABBANI, B.; TEKIN, M.; MAHDIEH, N. The promise of whole-exome sequencing in medical genetics. *J Hum Genet*, v. 59, n. 1, p. 5-15, 2014.
18. MAJEWSKI, J. *et al.* What can exome sequencing do for you? *J Med Genet*, v. 48, n. 9, p. 580-9, 2011.
19. American College of Medical Genetics and Genomics (ACMG) - Policy Statement - Points to consider in the clinical application of genomic sequencing. *Genet Med*, v. 14, n. 8, p. 759-61, 2012.

20. BIESECKER, L. G.; GREEN, R. C. Diagnostic clinical genome and exome sequencing. *N Engl J Med*, v. 370, n. 25, p. 2418-25, 2014.
21. GREEN, R. C. *et al.* ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med*, v. 15, n. 7, p. 565-74, 2013.
22. YU, J. H. *et al.* Attitudes of genetics professionals toward the return of incidental results from exome and whole-genome sequencing. *Am J Hum Genet*, v. 95, n. 1, p. 77-84, 2014.
23. BURKE, W. *et al.* Recommendations for returning genomic incidental findings? We need to talk! *Genet Med*, v. 15, n. 11, p. 854-9, 2013.
24. PRUITT, K. D. *et al.* The consensus coding sequence (CCDS) project: Identifying a common protein-coding gene set for the human and mouse genomes. *Genome Res*, v. 19, n. 7, p. 1316-23, 2009.
25. TEER, J. K.; MULLIKIN, J. C. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet*, v. 19, n. R2, p. R145-51, 2010.
26. CHILAMAKURI, C. S. *et al.* Performance comparison of four exome capture systems for deep sequencing. *BMC Genomics*, v. 15, p. 449, 2014.
27. MARDIS, E. R. Next-generation sequencing platforms. *Annu Rev Anal Chem (Palo Alto Calif)*, v. 6, p. 287-303, 2013.
28. MARDIS, E. R. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet*, v. 9, p. 387-402, 2008.
29. ALTMANN, A. *et al.* A beginners guide to SNP calling from high-throughput DNA-sequencing data. *Hum Genet*, v. 131, n. 10, p. 1541-54, 2012.
30. BAO, R. *et al.* Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing. *Cancer Inform*, v. 13, n. Suppl 2, p. 67-82, 2014.
31. COCK, P. J. *et al.* The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*, v. 38, n. 6, p. 1767-71, 2010.
32. MOORTHIE, S.; HALL, A.; WRIGHT, C. F. Informatics and clinical genome sequencing: opening the black box. *Genet Med*, v. 15, n. 3, p. 165-71, 2013.
33. KASSAHN, K. S.; SCOTT, H. S.; CARAMINS, M. C. Integrating massively parallel sequencing into diagnostic workflows and managing the annotation and clinical interpretation challenge. *Hum Mutat*, v. 35, n. 4, p. 413-23, 2014.

34. LI, H.; DURBIN, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, v. 25, n. 14, p. 1754-60, 2009.
35. HOMER, N.; MERRIMAN, B.; NELSON, S. F. BFAST: an alignment tool for large scale genome resequencing. *PLoS One*, v. 4, n. 11, p. e7767, 2009.
36. DEPRISTO, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, v. 43, n. 5, p. 491-8, 2011.
37. CINGOLANI, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)*, v. 6, n. 2, p. 80-92, 2012.
38. WANG, K.; LI, M.; HAKONARSON, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*, v. 38, n. 16, p. e164, 2010.
39. SHERRY, S. T. *et al.* dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res*, v. 29, n. 1, p. 308-11, 2001.
40. The 1000 Genomes Project Consortium - A map of human genome variation from population-scale sequencing. *Nature*, v. 467, n. 7319, p. 1061-73, 2010.
41. Exome Variant Server, NHLBI Exome Sequencing Project (ESP). Disponível em: <http://evs.gs.washington.edu/EVS/>.
42. The International HapMap Consortium - A haplotype map of the human genome. *Nature*, v. 437, n. 7063, p. 1299-320, 2005.
43. KUMAR, P.; HENIKOFF, S.; NG, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, v. 4, n. 7, p. 1073-81, 2009.
44. ADZHUBEI, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods*, v. 7, n. 4, p. 248-9, 2010.
45. MCLAREN, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics*, v. 26, n. 16, p. 2069-70, 2010.
46. JACOB, H. J. Next-generation sequencing for clinical diagnostics. *N Engl J Med*, v. 369, n. 16, p. 1557-8, 2013.
47. YANG, Y. *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med*, v. 369, n. 16, p. 1502-11, 2013.
48. WALL, J. D. *et al.* Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res*, 2014.

49. YANG, X.; CHOCKALINGAM, S. P.; ALURU, S. A survey of error-correction methods for next-generation sequencing. *Brief Bioinform*, v. 14, n. 1, p. 56-66, 2013.
50. O'RAWE, J. *et al.* Low concordance of multiple variant-calling pipelines: practical implications for exome and genome sequencing. *Genome Med*, v. 5, n. 3, p. 28, 2013.
51. LU, J. T.; CAMPEAU, P. M.; LEE, B. H. Genotype-phenotype correlation--promiscuity in the era of next-generation sequencing. *N Engl J Med*, v. 371, n. 7, p. 593-6, 2014.
52. BIESECKER, L. G. Exome sequencing makes medical genomics a reality. *Nat Genet*, v. 42, n. 1, p. 13-4, 2010.
53. BAMSHAD, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*, v. 12, n. 11, p. 745-55, 2011.
54. LI, M. X. *et al.* A comprehensive framework for prioritizing variants in exome sequencing studies of Mendelian diseases. *Nucleic Acids Res*, v. 40, n. 7, p. e53, 2012.
55. TOYDEMIR, R. M. *et al.* Mutations in embryonic myosin heavy chain (MYH3) cause Freeman-Sheldon syndrome and Sheldon-Hall syndrome. *Nat Genet*, v. 38, n. 5, p. 561-5, 2006.
56. NEED, A. C. *et al.* Clinical application of exome sequencing in undiagnosed genetic conditions. *J Med Genet*, v. 49, n. 6, p. 353-61, 2012.
57. JIANG, R. *et al.* Genome-wide evaluation of the public SNP databases. *Pharmacogenomics*, v. 4, n. 6, p. 779-89, 2003.
58. WU, L. *et al.* Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. *J Med Genet*, 2014.
59. GeneCards. Disponível em: www.genecards.org.
60. UniProt. Disponível em: <http://www.uniprot.org/>.
61. Mouse Genome Informatics. Disponível em: <http://www.informatics.jax.org/>.
62. Online Mendelian Inheritance in Man, OMIM®. McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University (Baltimore, MD). Disponível em: <http://omim.org/>.
63. STENSON, P. D. *et al.* The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics,

- diagnostic testing and personalized genomic medicine. *Hum Genet*, v. 133, n. 1, p. 1-9, 2014.
64. LANDRUM, M. J. *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*, v. 42, n. Database issue, p. D980-5, 2014.
 65. SOLOMON, B. D. *et al.* Clinical genomic database. *Proc Natl Acad Sci U S A*, v. 110, n. 24, p. 9851-5, 2013.
 66. BOONE, P. M. *et al.* Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. *Genome Res*, v. 23, n. 9, p. 1383-94, 2013.
 67. POLLARD, K. S. *et al.* Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res*, v. 20, n. 1, p. 110-21, 2010.
 68. COOPER, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res*, v. 15, n. 7, p. 901-13, 2005.
 69. SCHWARZ, J. M. *et al.* MutationTaster evaluates disease-causing potential of sequence alterations. *Nat Methods*, v. 7, n. 8, p. 575-6, 2010.
 70. TAVTIGIAN, S. V. *et al.* Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *J Med Genet*, v. 43, n. 4, p. 295-305, 2006.
 71. FLANAGAN, S. E.; PATCH, A. M.; ELLARD, S. Using SIFT and PolyPhen to predict loss-of-function and gain-of-function mutations. *Genet Test Mol Biomarkers*, v. 14, n. 4, p. 533-7, 2010.
 72. HICKS, S. *et al.* Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat*, v. 32, n. 6, p. 661-8, 2011.
 73. GONZALEZ-PEREZ, A.; LOPEZ-BIGAS, N. Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel. *Am J Hum Genet*, v. 88, n. 4, p. 440-9, 2011.
 74. FROUSIOS, K. *et al.* Predicting the functional consequences of non-synonymous DNA sequence variants--evaluation of bioinformatics tools and development of a consensus strategy. *Genomics*, v. 102, n. 4, p. 223-8, 2013.
 75. BENDL, J. *et al.* PredictSNP: robust and accurate consensus classifier for prediction of disease-related mutations. *PLoS Comput Biol*, v. 10, n. 1, p. e1003440, 2014.

76. STEFL, S. *et al.* Molecular mechanisms of disease-causing missense mutations. *J Mol Biol*, v. 425, n. 21, p. 3919-36, 2013.
77. BILGUVAR, K. *et al.* Whole-exome sequencing identifies recessive WDR62 mutations in severe brain malformations. *Nature*, v. 467, n. 7312, p. 207-10, 2010.
78. KRAWITZ, P. M. *et al.* Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome. *Nat Genet*, v. 42, n. 10, p. 827-9, 2010.
79. NACHMAN, M. W.; CROWELL, S. L. Estimate of the mutation rate per nucleotide in humans. *Genetics*, v. 156, n. 1, p. 297-304, 2000.
80. Enlis Genome - Research Edition. Disponível em: <http://www.enlis.com/>.
81. [dbSNP-announce] dbSNP Human Build 130. Disponível em: <http://www.ncbi.nlm.nih.gov/mailman/htdig/dbsnp-announce/2009q2/000088.html>.
82. KANEHISA, M.; GOTO, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res*, v. 28, n. 1, p. 27-30, 2000.
83. Alamut Visual. Disponível em: <http://www.interactive-biosoftware.com/alamut-visual/>.
84. Human Genome Variation Society - Nomenclature for the description of sequence variants. Disponível em: <http://www.hgvs.org/mutnomen/>.
85. WORTHEY, E. A. *et al.* Making a definitive diagnosis: successful clinical application of whole exome sequencing in a child with intractable inflammatory bowel disease. *Genet Med*, v. 13, n. 3, p. 255-62, 2011.
86. YANG, Y. *et al.* Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*, v. 312, n. 18, p. 1870-9, 2014.
87. MILLER, D. T. *et al.* Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*, v. 86, n. 5, p. 749-64, 2010.

CAPÍTULO 2 – OBJETIVOS

Objetivo geral

O objetivo principal deste trabalho foi realizar a análise bioinformática de exomas de pacientes com doenças mendelianas raras com a finalidade de localizar possíveis variantes causalmente relacionadas com as doenças. Estudamos um paciente que apresenta uma microdeleção no cromossomo 1p13.2, duas famílias com miofibromatose infantil e dois pacientes com calcificações no cérebro e convulsões.

Objetivos específicos

- Comparar o fenótipo do paciente com deleção em 1p13.2 com o fenótipo de pacientes relatados previamente com deleções que sobrepõem a do nosso paciente;
- Correlacionar o fenótipo do paciente com deleção em 1p13.2 com os genes localizados na região deletada;
- Verificar se há possíveis variantes patogênicas na região deletada 1p13.2 através da análise do exoma, o que implicaria no desmascaramento de uma doença recessiva e causaria variabilidade fenotípica;
- Identificar, através da análise do exoma, as variantes responsáveis pela miofibromatose infantil nos dois pacientes de cada uma das duas famílias estudadas, e pela calcificação no cérebro e convulsões nos dois pacientes estudados;
- Confirmar a existência das variantes encontradas através do sequenciamento de Sanger.

CAPÍTULO 3 - ARTIGOS

ARTIGO 1

1p13.2 deletion displays clinical features overlapping Noonan syndrome, likely related to NRAS gene haploinsufficiency

Manuscript Number:

Title: 1p13.2 deletion displays clinical features overlapping Noonan syndrome, likely related to NRAS gene haploinsufficiency

Article Type: Research Paper

Section/Category: Clinical Reports

Keywords: 1p13.2 deletion, exome sequencing, Noonan syndrome type 6, NRAS gene, RASopathy, unmasking heterozygosity.

Corresponding Author: Mrs. Natália Linhares,

Corresponding Author's Institution:

First Author: Natália Linhares

Order of Authors: Natália Linhares; Máira Freire; Raony Cardenas; Heloísa Pena; Katherine Lachlan; Bruno Dallapiccola; Seema Lalani; Carlos Bacino; Bruno Delobel; Paul James; Ann-Charlotte Thuresson; Göran Annerén; Sergio Pena

Abstract: In this article we report on genotype-phenotype correlations in patients with chromosome 1p13.2 microdeletions. We describe a new patient presenting with clinical features suggestive of Noonan Syndrome (NS). Exome sequencing analysis examined the possibility of a recessive pathogenic variant unmasked by his hemizygous deletion and also verified if he had pathogenic variants in any of the Noonan Syndrome genes. We concluded that the clinical characteristics of the patient were attributed to the 1p13.2 haploinsufficiency. Retrospective analysis of published patients with 1p13.2 microdeletions showed that many of them displayed characteristics of NS. In all cases the deleted segment included NRAS gene. Gain-of-function mutations of NRAS gene were causally related to NS type 6. Thus, NRAS haploinsufficiency and gain-of-function mutations have similar clinical consequences. Overlapping NS phenotypes have been reported also in subjects displaying either gain-of-function mutations or haploinsufficiency of SHOC2 gene, while Cardio-Facio-Cutaneous syndrome resulted either from deletion and gain-of-function mutations of MAP2K2 gene. In conclusion, both gain and loss of function of Ras/MAPK pathway result in overlapping clinical consequences.

Clinical report

1p13.2 deletion displays clinical features overlapping Noonan syndrome, likely related to *NRAS* gene haploinsufficiency

Natália D. Linhares ^a, Maíra C. M. Freire ^b, Raony G. C. C. L. Cardenas ^a, Heloísa B. Pena ^c, Katherine Lachlan ^d, Bruno Dallapiccola ^e, Seema R. Lalani ^f, Carlos A. Bacino ^f, Bruno Delobel ^g, Paul A. James ^h, Ann-Charlotte Thuresson,ⁱ Göran Annerén,ⁱ Sergio D. J. Pena ^{a,b,c*}

^a Laboratório de Genômica Clínica, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

^b Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil.

^c Laboratório Gene – Núcleo de Genética Médica, Belo Horizonte, Brazil.

^d Wessex Clinical Genetics Service, Princess Anne Hospital, Southampton, United Kingdom.

^e Bambino Gesù Children Hospital, Rome, Italy.

^f Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA.

^g Centre de génétique chromosomique. GH de l'Institut Catholique de Lille - Hopital Saint Vincent de Paul, Lille Cedex, France.

^h Victorian Clinical Genetics Service, Melbourne, Victoria, Australia.

ⁱ Department of Immunology, Genetics and Pathology, The Rudbeck laboratory, Uppsala University, Uppsala, Sweden.

Running title: Linhares et al., 1p13.2 microdeletion

***Correspondence to:**

Sergio Danilo Junho Pena, MD, PhD, FRCP(C)

Universidade Federal de Minas Gerais, Faculdade de Medicina. Av. Alfredo Balena, 190,
sala 321, Bairro Santa Efigênia. 30130-100 - Belo Horizonte, MG – Brazil; Telephone: +55
31 3409-8053

E-mail: spena@dcc.ufmg.br

Abstract

In this article we report on genotype-phenotype correlations in patients with chromosome 1p13.2 microdeletions. We describe a new patient presenting with clinical features suggestive of Noonan Syndrome (NS). Exome sequencing analysis examined the possibility of a recessive pathogenic variant unmasked by his hemizygous deletion and also verified if he had pathogenic variants in any of the Noonan Syndrome genes. We concluded that the clinical characteristics of the patient were attributed to the 1p13.2 haploinsufficiency. Retrospective analysis of published patients with 1p13.2 microdeletions showed that many of them displayed characteristics of NS. In all cases the deleted segment included *NRAS* gene. Gain-of-function mutations of *NRAS* gene were causally related to NS type 6. Thus, *NRAS* haploinsufficiency and gain-of-function mutations have similar clinical consequences. Overlapping NS phenotypes have been reported also in subjects displaying either gain-of-function mutations or haploinsufficiency of *SHOC2* gene, while Cardio-Facio-Cutaneous syndrome resulted either from deletion and gain-of-function mutations of *MAP2K2* gene. In conclusion, both gain and loss of function of Ras/MAPK pathway result in overlapping clinical consequences.

Keywords: 1p13.2 deletion, exome sequencing, Noonan syndrome type 6, *NRAS* gene, RASopathy, unmasking heterozygosity.

1. Introduction

Patients with chromosomal deletions may present with variable clinical phenotypes for different reasons. First, the deletions frequently vary in size and the phenotypic alterations may depend on which loci are deleted in the specific case, as exemplified by the contiguous gene syndromes such as Langer-Giedion syndrome (OMIM #150230) and Williams-Beuren syndrome (OMIM #194050). Second, the phenotypic effect of the deleted segment may be modulated by the rest of the genome, in analogy with autosomal dominant diseases with variable expressivity and as, for instance, is the case in digenic inheritance [Schaffer, 2013]. Third, early post zygotic events, and/or environmental factors may influence phenotypic discordances, such as in cases of monozygotic twins with discordant phenotype and chromosome 22q11.2 microdeletion [Yamagishi et al., 1998]. Fourth, the clinical phenotype may be influenced by the presence of a pathogenic variant in a gene located in the region homologous to the deleted segment, in the intact member of the chromosome pair. This phenomenon has been termed “unmasking heterozygosity” (UH) and when it occurs the phenotype of a recessive disease might emerge from the microdeletion due to hemizygozity [Coman and Gardner, 2007; Poot, 2012]. To our knowledge, the earliest example of this phenomenon was described in retinoblastoma (OMIM #180200), which may occur because a chromosomal deletion may cause hemizygozity for an *RBI* (OMIM *614041) pathogenic variant on chromosome 13q14.

Next generation sequencing (NGS) techniques, such as whole genome sequencing (WGS) and whole exome sequencing (WES) allow the screening of the genome including the non-deleted allele for variants that may contribute to variable phenotypic expression in deletion syndromes. Accordingly, NGS has been reported to reveal deleterious variants unmasked by hemizygous deletions in individuals with 22q11.2 deletion syndrome (OMIM #192430; #188400) [McDonald-McGinn et al., 2013], with mental retardation or multiple

congenital abnormalities and hemizygous deletions [Hochstenbach et al., 2012], and with thrombocytopenia absent radius syndrome (OMIM #274000) [Albers et al., 2012].

Since UH can be revealed using exome sequencing, we used it in a genotype-phenotype correlation study of chromosome 1p13.2 microdeletions. We here report the case of a 21-year-old man with a multisystem phenotype containing some features of Noonan syndrome (OMIM #613224) and 1p13.2 deletion extending from 1:112,096,417 to 1:115,805,157 (hg19) diagnosed by aCGH. Exome sequencing analysis examined the possibility of a recessive pathogenic variant unmasked by his hemizygous deletion and also verified if he had pathogenic variants in any of the Noonan Syndrome genes. We concluded that the clinical characteristics of the patient were attributed to the 1p13.2 haploinsufficiency. Additionally, we compared his clinical characteristics with the phenotype of six patients with deletions overlapping the one of our patient reported previously in DECIPHER database and three individuals reported in the literature [Bisgaard et al., 2007; Fitzgibbon et al., 2009; Mattia et al., 1992]. On these bases we propose that haploinsufficiency of *NRAS* gene (OMIM *164790) may be important in determining the clinical phenotype of this microdeletion.

2. Clinical report

Our patient was a 21-year-old man born of healthy unrelated parents (DECIPHER patient no. 274660). He was born at 37 weeks of pregnancy by cesarean delivery for breech presentation. Birth weight was 2,720 g (3rd-10th centile), with length of 47 cm (10th centile) and occipital frontal circumference (OFC) 34.5 cm (25-50th centile). He presented with hypotonia and feeding difficulties in infancy. His motor development was delayed, having walked independently at 22 months of age. At 3 years-old he spoke single words and at 4 years-old he spoke sentences. He never learned to write or to read. Since infancy he has had recurring bouts of vomiting. Colonoscopy revealed a polypoid mucosa in the terminal ileum

and chronic enteritis. Ophthalmologic exams showed myopia, with normal fundus.

Echocardiography and nuclear magnetic resonance of the brain were normal.

He was evaluated by a geneticist at 20 years of age, when his weight was 68.5 kg (25-50th centile), height 161 cm (<3rd centile) and OFC 53 cm (< 3rd centile). Besides the short stature, his phenotype had several features reminiscent of Noonan Syndrome, including intellectual disability, ptosis, low hairline at the nape, broad neck, excess of pigmented nevi resembling lentigines, *pectus excavatum* and scoliosis (see Fig. 1). Complementary examinations showed patellar chondromalacia, osteoporosis, asymmetry of the lower limbs, nephrolithiasis with increased loss of K^+ and Ca^{++} in urine, hypercholesterolemia and primary hypothyroidism. The karyotype of the patient and his parents were reported as being normal. Subsequent aCGH analysis (Agilent SurePrint G3 4x180K) revealed a 3.71 Mb deletion on chromosome 1p13.2 (chr1:112,096,417–115,805,157; UCSC Genome Browser hg19) (Fig. 2).

3. Methods

3.1. Samples and DNA isolation

This study was approved by the Research Ethics Committee of Universidade Federal de Minas Gerais and by the Brazilian National Committee of Ethics in Research (CONEP – Brazil) with number 854.709. The written informed consent including permission for publication was provided by the parents. Genomic DNA was isolated from peripheral blood of the patient using a modified salting out procedure [Miller et al., 1988].

3.2. Whole exome sequencing and analysis

Whole exome sequencing was performed in the patient's sample by The Centre for Applied Genomics, The Hospital for Sick Children (Toronto, Canada) using the Agilent SureSelect Human All Exon V4 Kit (Agilent Technologies) and SOLiD 5500xl platform

(Applied Biosystems). All data were aligned to the hg19/GRCh37 reference genome build via BFAST and BWA aligner. Variants were quality trimmed using the Genome Analysis Toolkit (GATK 1.1.28) and they were annotated for functional effect by SnpEff 2.0.5. SnpEff also provided a simple assessment of the putative impact of the variant (e.g. High, Moderate, or Low impact). Alignment, calling and annotation of the variants against databases such as 1000 Genomes (April 2012 release), NHLBI Exome Sequencing Project (ESP6500) and Single Nucleotide Polymorphism database (dbSNP137) were done using a software developed in-house called *Mendel,MD* (RGCCL Cardenas and SDJ Pena, manuscript in preparation). The prediction software SIFT (Sorting Intolerant From Tolerant) was used in order to estimate whether a given amino acid substitution affected protein function. SIFT prediction is based on the degree of conservation of amino acid residues in sequence alignments derived from closely related sequences and on the physical properties of amino acids.

4. Results

Whole exome sequencing of the patient sample identified 76,822 variants. Approximately 26,863 of these variants passed in the GATK quality filters and had a Minor Allele Frequency (MAF) ≤ 0.01 in the 1000 Genomes, ESP6500 and dbSNP137. In the deleted region, which extended from 1:112,096,417 to 1:115,805,157 (hg19), exome analysis revealed 35 hemizygous variants that passed in GATK quality filters. Only ten of these variants were considered to have moderate or high impact by SnpEff, which means that they are assumed to have high disruptive impact in the protein or might change protein effectiveness (examples: stop gained, frameshift, missense and splice donor variants). Based on the assumption that the potential damaging variants are rare, the variants were then filtered considering an allele frequency <0.01 in the 1000 Genomes database, ESP6500 and

dbSNP137, however no plausibly pathogenic variant was retained. Besides, all of these ten variants were predicted as tolerated by SIFT (score ≥ 0.05) (Supplementary Table 1). In addition, we verified by exome analysis that the patient did not have any pathogenic variants in the genes associated with Noonan Syndrome and other rasopathies (*PTPN11*, *SHOC2*, *KRAS*, *CBL*, *SOS1*, *RAF1*, *HRAS*, *NRAS*, *RASA1*, *SPRED1*, *BRAF*, *MAP2K1* and *MAP2K2*) [Rauen, 2013].

The deleted 1p13.2 region of our patient contained 69 genes according to the NCBI Map Viewer Annotation Release 105, of which 28 were listed in Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders (Supplementary Table 2). However, only eight of these genes were known to be associated with Mendelian disease phenotypes (Supplementary Table 3). Among the recessive disease genes there was the *TSHB* gene (OMIM *188540), in which homozygous mutations had been reported in patients with congenital nongoitrous hypothyroidism-4 (OMIM #275100). Our patient had hypothyroidism, but it was not congenital.

In order to perform a genotype-phenotype correlation study of chromosome 1p13.2 microdeletions, we compared the clinical characteristics of our patient to those of previously described cases (Table 1 and Fig. 1). Isolated chromosome 1p13.2 microdeletions are not common; we could find only three individuals reported in the literature [Bisgaard et al., 2007; Fitzgibbon et al., 2009; Mattia et al., 1992]. Additionally, six patients with isolated deletions partially overlapping the one of our patient had been reported previously in DECIPHER database. All these nine patients were included in our analysis.

The deletions in the several patients ranged from 2.19 Mb to 17.93 Mb in size and eight of them included a region extending from 115,018,964 to 115,713,872 (hg19). The region of overlap of all patients contained ten genes: *TRIM33*, *BCAS2*, *DENND2C*, *AMPD1*, *NRAS*, *CSDE1*, *SIKE1*, *SYCP1*, *TSHB*, *TSPAN2* (Fig. 3). The Database of Genomic Variants

(DGVbeta) lists numerous copy-number variations (CNVs) in individuals with normal phenotypes within this region. The segment with the least number of CNVs was the one containing the genes *AMPD1*, *NRAS* and *CSDE1* (Fig. 4).

The patient reported by Fitzgibbon et al. [2009] presented congenital insensitivity to pain and temperature, which led the authors to propose that this might be a consequence of haploinsufficiency of the *NGF* gene (OMIM *162030), which is related to hereditary sensory and autonomic neuropathy type V, an autosomal recessive disease (sequencing of the undelated homologous chromosome had not shown any pathogenic variants). However, in disagreement with this suggestion, *NGF* was also deleted in three patients described in DECIPHER, none of whom presented congenital insensitivity to pain and temperature (patients no. 250335, 253793 and 260230). In relation to coloboma, Bisgaard et al. [2007] hypothesized that haploinsufficiency of the *WNT2B* gene (OMIM *601968) might have caused it in their patient. Indeed, this gene was deleted in two other patients with ocular coloboma (Patients no. 250335 and 257066), but it was also deleted in our patient, who did not have this malformation.

5. Discussion

We here report a patient with isolated 1p13.2 microdeletion diagnosed by aCGH analysis. We performed exome sequencing in the patient in order to check for possible unmasking heterozygosity and to exclude variants in any of the genes related to Noonan Syndrome or other RASopathies. The analysis of complete exome sequencing (WES) identified 35 hemizygous variants in the deleted region. However, evaluation of frequency in the population and bioinformatics protein predictions did not indicate pathogenicity of these variants. Moreover, we did not find pathogenic variants in any of the genes known to be

associated with Noonan Syndrome. Consequently, we conclude that the phenotype of our patient is exclusively related to his hemizygous 1p13.2 state.

We then tried to identify other genes possibly related with the phenotype of our patient and others. Our attention was drawn to the presence of deletion of *NRAS* for four reasons:

1. Variants of *NRAS* have been described as causing Noonan Syndrome, Type 6 (OMIM #613224), an autosomal dominant disorder [Cirstea et al., 2010]. However, these have been shown to be gain-of-function mutations;
2. The phenotype of our patient and also of other patients with the 1p13.2 deletion includes several features of Noonan Syndrome (Table 1 and Fig. 1);
3. Haploinsufficiency Index predictions (HI) shown in the DECIPHER database for all the genes deleted in our patient, revealed that the *NRAS* gene was the one with higher rank of HI (which suggested that this gene is more likely to exhibit haploinsufficiency) [Huang et al., 2010];
4. The region surrounding *NRAS* is rarely seen in polymorphic CNVs (Fig. 4).

NRAS is a member of the Ras subfamily, which includes 36 members; *KRAS*, *HRAS* and *NRAS* have been the most studied because of their critical roles in human oncogenesis [Wennerberg et al., 2005]. Ras proteins belong to a class of signal-transducing GTPases which cycle between an active GTP-bound and an inactive GDP-bound conformation. Alterations of conditions, including gene variants, may result in stabilization of Ras proteins in their active state leading to malignant transformation [Scheffzek et al., 1997]. Ras proteins activate several pathways, including the Ras/mitogen activated protein kinase (MAPK) pathway [Wennerberg et al., 2005]. This pathway plays an essential role in regulating the cell survival, differentiation, proliferation and apoptosis, among many others [Pearson et al., 2001].

Germline variants in genes that encode components or regulators of the Ras/MAPK pathway have been shown to cause developmental syndromes collectively referred to as RASopathies. These disorders include several forms of Noonan syndrome, including Costello syndrome (OMIM #218040), cardio-facio-cutaneous syndrome (OMIM #115150), and capillary malformation–arteriovenous malformation syndrome (OMIM #608354) [Rauen, 2013]. In all of these cases the mutations have been shown to involve gain-of-function, increasing activity in the Ras/MAPK pathway. It is believed that this is the reason why they are clinically similar, with overlapping phenotypic features including characteristic facial features, cardiac defects, cutaneous abnormalities, neurocognitive delay and an increased cancer risk [Rauen, 2013; Tidyman and Rauen, 2009].

Few heterozygous mutations have been described as cause of the Noonan type 6 (NS6) and they have all been associated with gain-of-function in *NRAS* [Cirstea et al., 2010; De Filippi et al., 2009; Denayer et al., 2012; Kraoua et al., 2012; Runtuwene et al., 2011]. Cirstea et al. [2010] expressed each mutant as yellow fluorescent protein–NRAS fusion proteins in cells from the COS-7 line and showed for example that NRAS substitutions p.Thr50Ile or p.Gly60Glu resulted in enhanced phosphorylation of MEK and ERK in the presence of serum or after epidermal growth factor (EGF) stimulation. In addition, the authors observed that the p.Gly60Glu NRAS mutant accumulated constitutively in the active, GTP-bound form [Cirstea et al., 2010]. Runtuwene et al. [2011] showed that similar effects occurs in GTP-bound in p.Ile24Asn mutants and they revealed that p.Gly60Glu, p.Ile24Asn and the positive control (p.Gly12Val) mutants also enhanced MAPK activation. These patients had typical and variable clinical features of Noonan syndrome, including hypertelorism, low-set ears, short stature, webbed neck, low hairline, thorax deformities, motor delay/muscular hypotonia and lentigenes/cafe-au-lait spots [Reviewed by Kraoua et al., 2012].

On the other hand, our patient and other cases with pure chromosome 1p13.2 microdeletions and hemizygozity for *NRAS* reviewed by us from the literature or from the DECIPHER database all seemed to present several clinical features of Noonan Syndrome (Table 1 and Fig. 1). The presence of such features related to Noonan Syndrome in patients with deletion of *NRAS* and presumably haploinsufficiency of the gene product appears *prima facie* paradoxical.

We searched the literature and we have found three similar examples. In Noonan syndrome, Edwards et al. [2014] described a 5-year-old male with two *de novo* pathogenic *PTPN11* variants in *cis* (OMIM *176876). The double-mutant gene product (SHP-2) was found to be catalytically impaired. More recently, Chen et al. [2014] described a patient with a Noonan-like phenotype associated with a deletion of the 10q25.2 chromosomal region that included the *SHOC2* gene (OMIM *602775). Gain-of-function mutations in this gene have been shown to cause a Noonan Syndrome-like syndrome. In the same vein, Nowaczyk et al. [2014] reported seven patients with deletions of chromosome 19p13.3 including the *MAP2K2* gene (OMIM *601263) and phenotype features of Cardio-facio-cutaneous syndrome, which is a RASopathy known to be caused by activating mutations of the *BRAF*, *MAP2K1*, *MAP2K2*, or *KRAS* genes. On these bases, these authors proposed that haploinsufficiency of *MAP2K2/MEK2* “appears to be a new model of a RASopathy, where a deletion of one of the components of the pathway, MEK2, results in a RASopathy-like phenotype” [Nowaczyk et al., 2014]. We believe that the association here described of 1p13.2 microdeletion involving hemizygozity for *NRAS* and clinical features of Noonan syndrome, represent a further example of the model proposed by Nowaczyk et al. [2014], according to which, haploinsufficiency of a gene in the Ras/MAPK pathway may cause dysregulation of the pathway and produce a RASopathy phenotype.

Conflict of interest statement

The authors declare no conflict of interest.

Acknowledgements

The authors are grateful to the patients and their family for their precious cooperation in this study. We thank Dr. Sérgio Pereira of The Centre for Applied Genomics, The Hospital for Sick Children (Toronto, Canada) for assistance with exome data interpretation. This work was supported by FAPEMIG grant to SDJP (process CDS-30/11). MCMF and RGCCLC were supported by a fellowship from CAPES and NDL was supported by a fellowship from CNPq. This study makes use of data generated by the DECIPHER Consortium. A full list of centres who contributed to the generation of the data is available from the website and via email from decipher@sanger.ac.uk. Funding for DECIPHER project was provided by the Wellcome Trust.

Web Resources

1000 Genomes, <http://browser.1000genomes.org>

Database of Genomic Variants (DGVbeta) <http://dgvbeta.tcag.ca/dgv/app/home>

DECIPHER database (<https://decipher.sanger.ac.uk>)

Ensembl Genome Browser, <http://www.ensembl.org/index.html>

Genome Analysis Toolkit (GATK), <http://www.broadinstitute.org/gatk/>

NCBI dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>

NHLBI Exome Sequencing Project (ESP), <http://evs.gs.washington.edu/EVS/>

Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org/>

SIFT, <http://sift.jcvi.org/>

SnEff, <http://snpeff.sourceforge.net/>

Figure Titles and Legends

Fig. 1. Frontal view of patients with deleted *NRAS*. Patients present features of Noonan Syndrome 6, including macrocephaly, short/webbed neck, low hairline, skin abnormalities, triangular face with age, low-set ears, arched eyebrows, hypertelorism, ptosis, downslating papebral fissures and epicanthal folds. Considering the patients with *NRAS* deletion, we did not have a picture of patient 253793 and 258063.

Fig. 2. Copy number profile of chromosome 1 of our patient obtained by aCGH. The chromosome 1 copy number imbalances are indicated on the left panel and shown in detail on the right panel: the alterations marked by the square show a ~3.71 Mb 1p13.2 deletion.

Fig. 3. Schematic representation of the deleted segments in our patient and those previously reported with isolated 1p13.2 microdeletions. Except for the patient described by Mattia et al.1992, who has a microdeletion from 1p13 to 1p22.3, all other patients have their breakpoints defined by molecular methods. Ideogram of chromosome 1, physical map and deleted segments are indicated according to their placement on the Ensembl Genome Browser.

Fig. 4. Genomic variants listed in the Database of Genomic Variants (DGVbeta) in the smallest region of overlap between the patients (chr1:115,018,964–115,713,872) (hg19). The black arrow indicate the region less populated with CNVs.

Table

Table 1. Clinical features of ten patients with isolated 1p13.2 microdeletions

	Mattia <i>et al.</i> 1992	Bisgaard <i>et al.</i> 2007	Fitzgibbon <i>et al.</i> 2009	Patient 250335	Patient 253793	Patient 256753	Patient 257066	Patient 258063	Patient 260230	This report 274660
Extension ^a	ND	~103804508-115747737	~114609456-116035987	105976094-120529725	113268038-117119460	102041407-112285318	95286548-113219619	113525828-115713872	115018964-118929584	112096417-115805157
Size (Mb)	ND	11.9-14	1.4-3.1	14.55	3.85	10.24	17.93	2.19	3.91	3.71
<i>NRAS</i> deleted	+	+	+	+	+	-	-	+	+	+
Age at last examination	2 years	13 years	5 years	3 years	38 years	4 years	2 years	8 years	8 years	21 years
Gender	Male	Female	Female	Female	Male	Male	Female	Female	Female	Male
Infantile feeding difficulties	ND	ND	+	No	No	+	+	No	No	+
Facial features of Noonan Syndrome ^b	+	+	+	No ^c	^d +	No	No	No ^c	No ^c	+
Short stature	No	+	ND	+	+	No	+	No	+	+
Macrocephaly	No	No	ND	Relative	+	No	Relative	+	+	No
Webbed neck	Short neck	+	ND	Short neck	Short neck	No	Short and broad neck	No	No	Broad neck
Ophthalmological abnormalities	No	Iris coloboma	ND	Iris coloboma	No	No	Papillary coloboma	Astigmatism	Strabismus	Myopia
Motor delay/muscular hypotonia	+	+	+	+	+	+	+	+	+	+
Intellectual disability	NA	+	+	+	+	+	NA	+	+	+
Speech delay	+	+	+	+	+	+	+	+	+	+
Low hairline	ND	+	ND	+	No	No	ND	No	No	+
Skin abnormalities	Forehead nevus flammeus	ND	ND	No	Several lentigenes	No	No	No	Several vitiligo, dry skin	Several lentigenes
Congenital heart defects	No	ND	ND	+ ^e	No	No	ND	No	No	No

NA, not applicable; ND, not determined. a- UCSC Genome Browser hg19 coordinates (except for Bisgaard et al. 2007). Mattia et al. (1992) did

not perform molecular analysis. b- Facial features of Noonan Syndrome 6 include: triangular face with age, low-set ears, hypertelorism,

palpebral ptosis, downslanting palpebral fissures, epicanthal folds. c- Overall gestalt is not strongly reminiscent of Noonan syndrome. d- Patient

253793 is an adult, and facial features of Noonan syndrome are less recognizable in adult individuals. e - Ventricular and atrial septum defect

References

- 1
2 Albers CA, Paul DS, Schulze H, Freson K, Stephens JC, Smethurst PA et al. Compound
3 inheritance of a low-frequency regulatory SNP and a rare null mutation in exon-
4 junction complex subunit RBM8A causes TAR syndrome. *Nat Genet* 2012; 44:435-9,
5 S1-2.
6
7
8
9 Bisgaard AM, Rasmussen LN, Moller HU, Kirchhoff M, Bryndorf T. Interstitial deletion of
10 the short arm of chromosome 1 (1p13.1p21.1) in a girl with mental retardation, short
11 stature and colobomata. *Clin Dysmorphol* 2007; 16:109-12.
12
13
14 Chen JL, Zhu X, Zhao TL, Wang J, Yang YF, Tan ZP. Rare copy number variations
15 containing genes involved in RASopathies: deletion of SHOC2 and duplication of
16 PTPN11. *Mol Cytogenet* 2014; 7:28.
17
18
19 Cirstea IC, Kutsche K, Dvorsky R, Gremer L, Carta C, Horn D et al. A restricted spectrum of
20 NRAS mutations causes Noonan syndrome. *Nat Genet* 2010; 42:27-9.
21
22
23 Coman DJ, Gardner RJ. Deletions that reveal recessive genes. *Eur J Hum Genet* 2007;
24 15:1103-4.
25
26
27 De Filippi P, Zecca M, Lisini D, Rosti V, Cagioni C, Carlo-Stella C et al. Germ-line mutation
28 of the NRAS gene may be responsible for the development of juvenile
29 myelomonocytic leukaemia. *Br J Haematol* 2009; 147:706-9.
30
31
32 Denayer E, Peeters H, Sevenants L, Derbent M, Fryns JP, Legius E. NRAS Mutations in
33 Noonan Syndrome. *Mol Syndromol* 2012; 3:34-8.
34
35
36 Edwards JJ, Martinelli S, Pannone L, Lo IF, Shi L, Edelmann L et al. A PTPN11 allele
37 encoding a catalytically impaired SHP2 protein in a patient with a Noonan syndrome
38 phenotype. *Am J Med Genet A* 2014; 164A:2351-5.
39
40
41 Fitzgibbon GJ, Kingston H, Needham M, Gaunt L. Haploinsufficiency of the nerve growth
42 factor beta gene in a 1p13 deleted female child with an insensitivity to pain. *Dev Med*
43 *Child Neurol* 2009; 51:833-7.
44
45
46 Hochstenbach R, Poot M, Nijman IJ, Renkens I, Duran KJ, Van't Slot R et al. Discovery of
47 variants unmasked by hemizygous deletions. *Eur J Hum Genet* 2012; 20:748-53.
48
49
50 Huang N, Lee I, Marcotte EM, Hurles ME. Characterising and predicting haploinsufficiency
51 in the human genome. *PLoS Genet* 2010; 6:e1001154.
52
53
54 Kraoua L, Journal H, Bonnet P, Amiel J, Pouvreau N, Baumann C et al. Constitutional NRAS
55 mutations are rare among patients with Noonan syndrome or juvenile
56 myelomonocytic leukemia. *Am J Med Genet A* 2012; 158A:2407-11.
57
58
59
60
61
62
63
64
65

- 1 Mattia FR, Wardinsky TD, Tuttle DJ, Grix A, Jr., Smith KA, Walling P. Interstitial deletion
2 of the short arm of chromosome 1 (46XY, del(1)(p13p22.3)). Am J Med Genet 1992;
3 44:551-4.
4
- 5 McDonald-McGinn DM, Fahiminiya S, Revil T, Nowakowska BA, Suhl J, Bailey A et al.
6 Hemizygous mutations in SNAP29 unmask autosomal recessive conditions and
7 contribute to atypical findings in patients with 22q11.2DS. J Med Genet 2013; 50:80-
8 90.
9
- 10 Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from
11 human nucleated cells. Nucleic Acids Res 1988; 16:1215.
12
- 13 Nowaczyk MJ, Thompson BA, Zeeman S, Moog U, Sanchez-Lara PA, Magoulas PL et al.
14 Deletion of MAP2K2/MEK2: a novel mechanism for a RASopathy? Clin Genet 2014;
15 85:138-46.
16
- 17 Pearson G, Robinson F, Beers Gibson T, Xu BE, Karandikar M, Berman K et al. Mitogen-
18 activated protein (MAP) kinase pathways: regulation and physiological functions.
19 Endocr Rev 2001; 22:153-83.
20
- 21 Poot M. Beware of Hemizygous Deletions That May Unmask Deleterious Variants. Mol
22 Syndromol 2012; 3:45-6.
23
- 24 Rauen KA. The RASopathies. Annu Rev Genomics Hum Genet 2013; 14:355-69.
25
- 26 Runtuwene V, van Eekelen M, Overvoorde J, Rehmann H, Yntema HG, Nillesen WM et al.
27 Noonan syndrome gain-of-function mutations in NRAS cause zebrafish gastrulation
28 defects. Dis Model Mech 2011; 4:393-9.
29
- 30 Schaffer AA. Digenic inheritance in medical genetics. J Med Genet 2013; 50:641-52.
31
- 32 Scheffzek K, Ahmadian MR, Kabsch W, Wiesmuller L, Lautwein A, Schmitz F et al. The
33 Ras-RasGAP complex: structural basis for GTPase activation and its loss in
34 oncogenic Ras mutants. Science 1997; 277:333-8.
35
- 36 Tidyman WE, Rauen KA. The RASopathies: developmental syndromes of Ras/MAPK
37 pathway dysregulation. Curr Opin Genet Dev 2009; 19:230-6.
38
- 39 Wennerberg K, Rossman KL, Der CJ. The Ras superfamily at a glance. J Cell Sci 2005;
40 118:843-6.
41
- 42 Yamagishi H, Ishii C, Maeda J, Kojima Y, Matsuoka R, Kimura M et al. Phenotypic
43 discordance in monozygotic twins with 22q11.2 deletion. Am J Med Genet 1998;
44 78:319-21.
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Mattia
et al. 1992



Bisgaard
et al. 2007



Fitzgibbon
et al. 2009



Patient
250335



Patient
260230



This report
274660



Fig. 2_color

[Click here to download high resolution image](#)

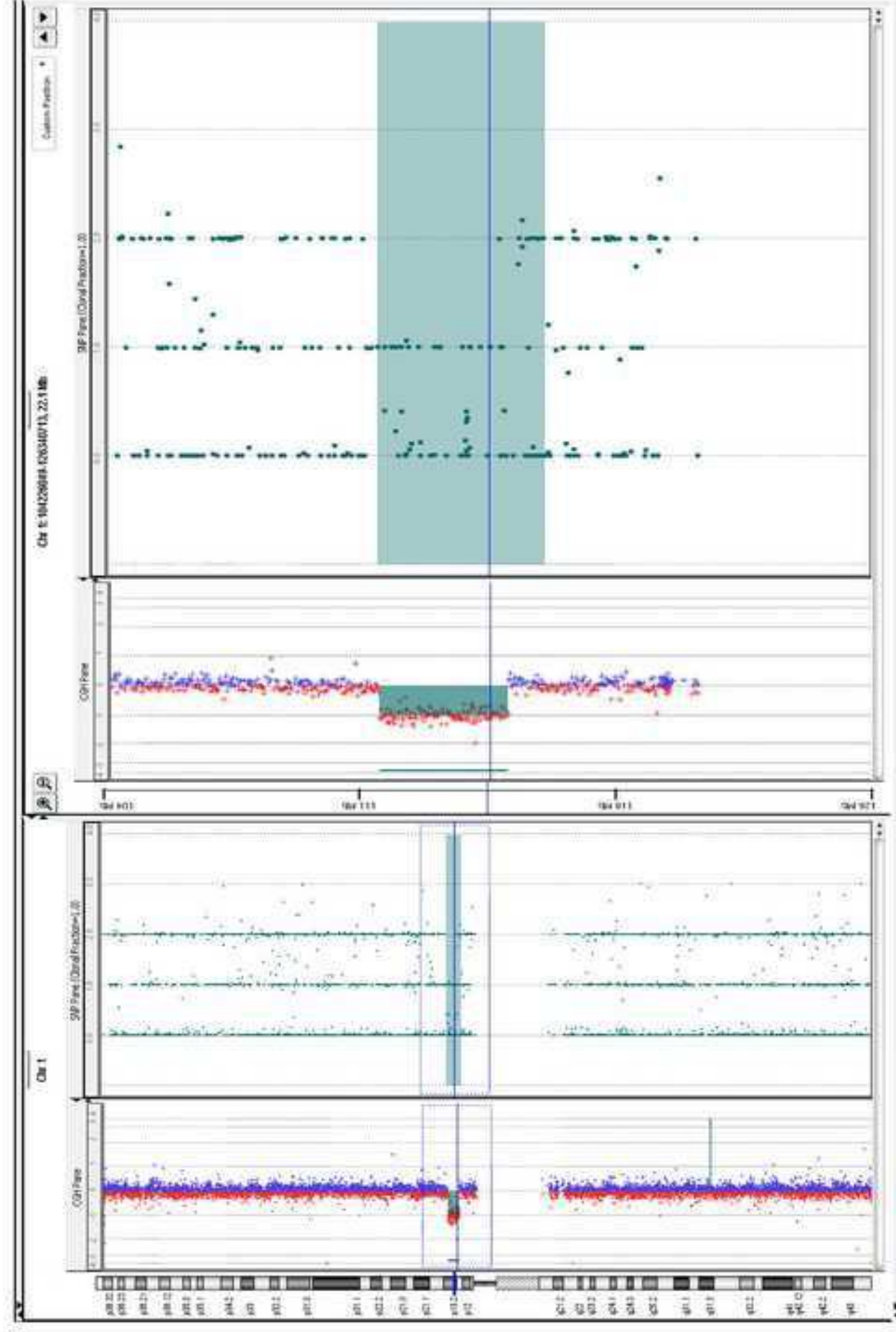


Fig. 3_color

[Click here to download high resolution image](#)

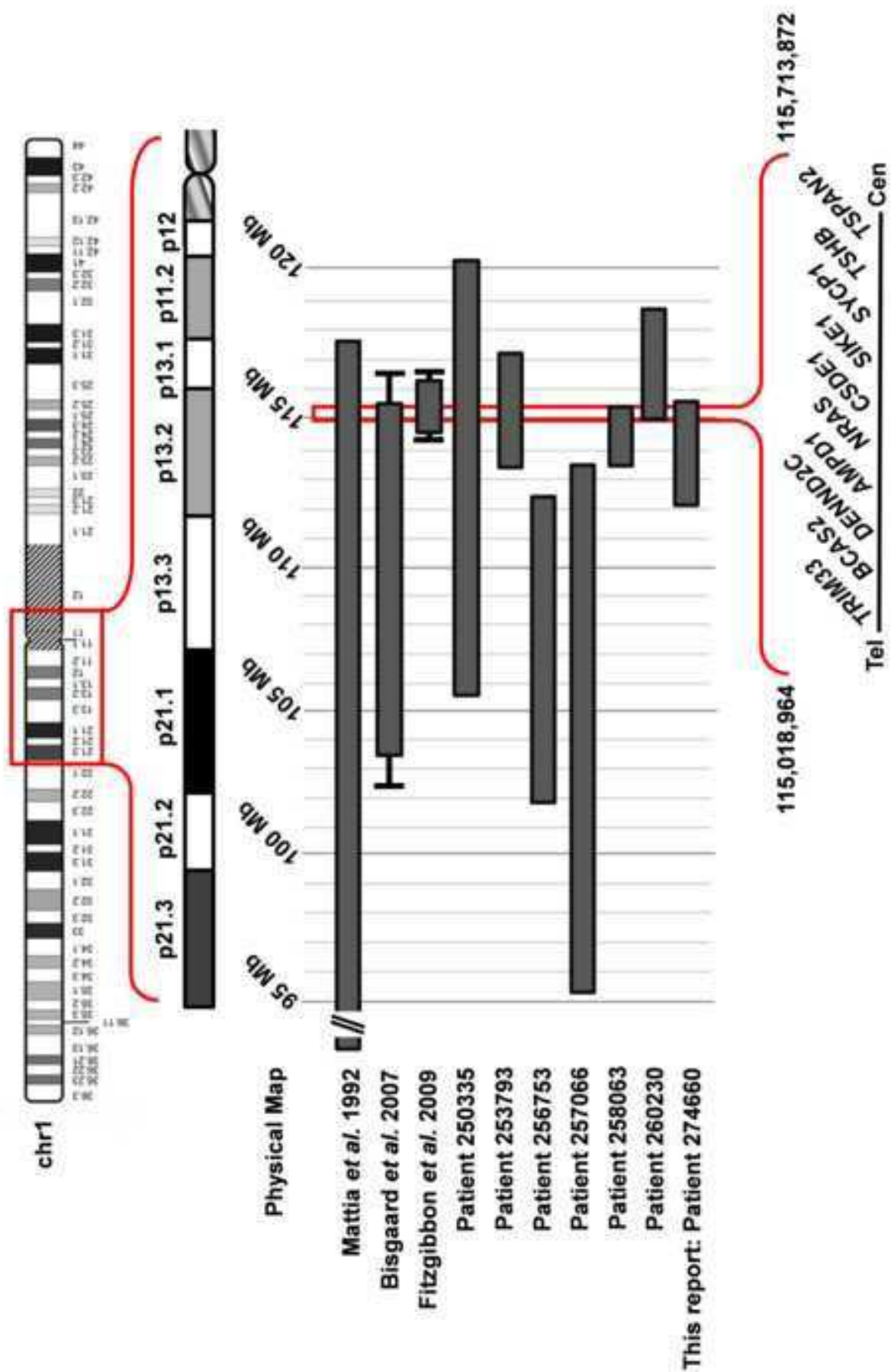
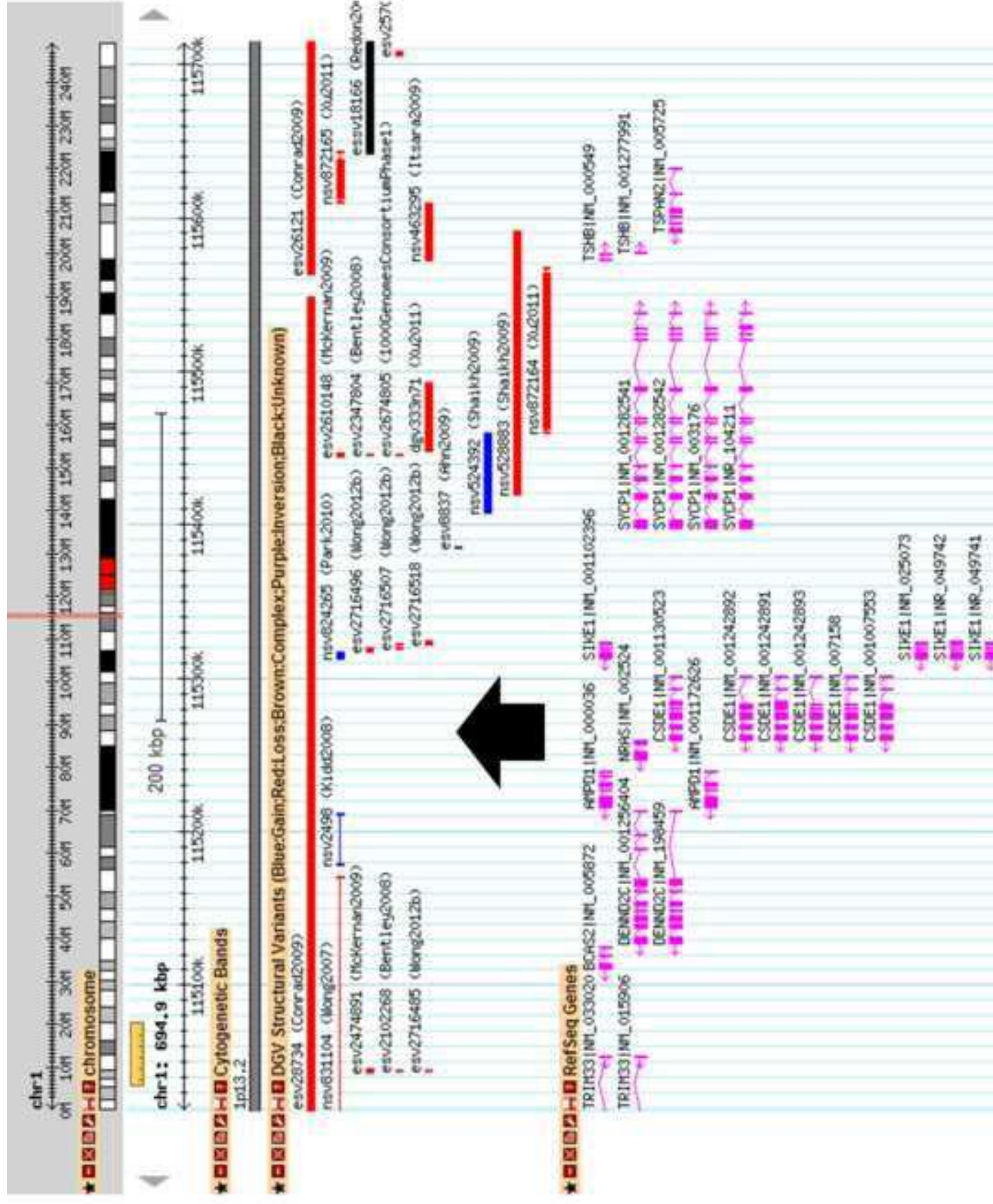


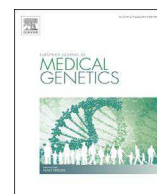
Fig. 4_color

[Click here to download high resolution image](#)



ARTIGO 2

***Exome sequencing identifies a novel homozygous variant in NDRG4
in a family with infantile myofibromatosis***



Exome report

Exome sequencing identifies a novel homozygous variant in *NDRG4* in a family with infantile myofibromatosis

Natália D. Linhares^a, Maíra C.M. Freire^b, Raony G.C.C.L. Cardenas^a, Heloísa B. Pena^c,
Magda Bahia^d, Sergio D.J. Pena^{a,b,c,*}

^a Laboratório de Genômica Clínica, Faculdade de Medicina, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

^b Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil

^c Laboratório Gene – Núcleo de Genética Médica, Belo Horizonte, Brazil

^d Divisão de Gastroenterologia Pediátrica, Hospital das Clínicas da UFMG, Belo Horizonte, Brazil

ARTICLE INFO

Article history:

Received 22 February 2014

Accepted 31 August 2014

Available online 18 September 2014

Keywords:

NDRG4 gene

Myofibromatosis

Exome sequencing

Congenital generalized fibromatosis

Neoplasms

ABSTRACT

Infantile myofibromatosis (IM) is a rare disorder characterized by the development of benign tumors in the skin, muscle, bone, and viscera. The incidence is 1/150,000 live births and the disease is the most common cause of fibrous tumors in infancy. Cases which lack visceral involvement generally have a more benign course, usually with spontaneous regression of the tumors. On the other hand, the prognosis tends to be unfavorable when there is involvement of vital organs, which can lead to significant mortality.

The identification of rare variants in genes that may cause IM is the first step towards the possibility of targeted treatments; however, the molecular pathogenesis of IM is poorly understood. In the present study, we report the results of exome sequence analysis of two brothers diagnosed with visceral multicentric infantile myofibromatosis, and their healthy consanguineous parents. In the two brothers we identified novel homozygous variants in *NDRG4* gene (N-myc downregulated gene family member 4) and in *RLTPR* gene (RGD motif, leucine rich repeats, tropomodulin domain and proline-rich containing). The healthy parents were heterozygous for both variants. Consistent with the phenotype of IM, *NDRG4* is a tumor-related gene; its expression has been shown to be decreased in numerous tumor types, suggesting that it might be a tumor suppressor gene. Additionally, studies have demonstrated that *NDRG4* may have a role in cell survival and tumor invasion. We thus propose that this homozygous variant in *NDRG4* may be the causative variant of the autosomal recessive form of IM in the studied family and that it should be investigated in other cases of autosomal recessive infantile myofibromatosis.

© 2014 Elsevier Masson SAS. All rights reserved.

1. Introduction

Infantile myofibromatosis (IM; OMIM 228550 and 615293) is a rare disorder characterized by the development of benign tumors in soft tissues and striated muscles, involving occasionally visceral organs and bones. The tumors are present at birth or develop shortly thereafter, with 90% of cases occurring before two years of age [Orbach, 2013]. The incidence is 1/150,000 live births, although this condition is the most common cause of fibrous tumors in infancy [Orbach, 2013; Wiswell et al., 1988]. Stout [1954] first

described this entity based on histology and suggested the term “congenital generalized fibromatosis”. Years later, Chung and Enzinger [1981], noted that these benign tumors consist of cells that have morphological characteristics of both smooth muscle and fibroblasts, and they renamed the disease “infantile myofibromatosis”.

Infantile myofibromatosis may occur as one of two forms: solitary lesions or multicentric tumors. The former tend to occur in skin, especially in the soft tissues of the head and neck, followed by the upper extremities and trunk. The latter is characterized by multiple nodules in soft tissues, skeleton and internal organs [Tamburrini et al., 2003]. Infants with a solitary lesion, or with multiple lesions but without visceral involvement, generally have a benign course, frequently with spontaneous regression of the tumors [Huang et al., 2005]. On the other hand, the prognosis tends to be unfavorable when there is involvement of the

* Corresponding author. Universidade Federal de Minas Gerais, Faculdade de Medicina. Av. Alfredo Balena, 190, sala 321, Bairro Santa Efigênia, 30130-100 Belo Horizonte, MG, Brazil. Tel.: +55 31 3409 8053.

E-mail address: spena@dcc.ufmg.br (S.D.J. Pena).

gastrointestinal, cardiac, or pulmonary system; the mortality then can be as high as 73% [Wiswell et al., 1988]. The treatment of choice for cases that affect the vital functions has been surgical excision [Gatibelza et al., 2012]. In aggressive cases, chemotherapy and corticosteroids have been used as therapeutic strategies – nevertheless, the literature is unclear about the overall success of such treatments [Gandhi et al., 2003].

Little is known about the molecular pathogenesis of IM. Although most reported cases are simplex, both autosomal recessive and dominant inheritance have been reported in multiplex cases [Zand et al., 2004]. Recently, variants in the genes *PDGFRB* (platelet-derived growth factor receptor, beta polypeptide; OMIM 173410) and *NOTCH3* (notch 3; OMIM 600276) have been reported as being the cause of the autosomal dominant forms of IM [Cheung et al., 2013; Martignetti et al., 2013].

In this report we describe the results of exome sequence analysis of two brothers diagnosed with multicentric IM and their consanguineous healthy parents. We suggest that the variant c.511G>C (p.Val171Leu) in the gene *NDRG4* (N-myc downregulated gene family member 4; OMIM 614463) present in homozygosity in the two brothers and in heterozygosity in their parents is a strong candidate to be the causative variant in this family.

2. Clinical report

2.1. The family

The studied family is composed of four individuals; both sons are affected with multicentric infantile myofibromatosis. The parents are healthy and report distant consanguinity.

2.2. Proband

The patient was first seen at the Pediatric Clinic of Hospital das Clínicas of UFMG at age five. The history had been healthy until 12 months of age, when he presented recurrent abdominal pain and chronic diarrhea. At four years, cow's milk was substituted by milk based on soy; however the symptoms were not eliminated.

From then on his growth was slow; at eight years of age his weight was 14 kg and height 105 cm (both below the 3rd centile). Colonoscopy detected ulcerations and polyps suggestive of subepithelial lesions and raising the clinical suspicion of inflammatory bowel disease. He was thus treated with sulfasalazine and prednisone, with improvement of the abdominal pain. The dose of corticoid was then tapered down and finally the drug was stopped, which was followed by recurrence of the abdominal pain. Another colonoscopy was performed at nine years and showed multiple subepithelial lesions with a central ulceration, some of them with fibrin. Video capsule endoscopy showed two lesions with 4 mm and 7 mm in the jejunum, which were described in detail in a previously publication about the case [Alberti et al., 2012].

Biopsies were taken from the colonic lesions and a pathological diagnosis of infantile myofibromatosis was made. H&E stained sections showed that the tumors were composed of myofibroblasts (Fig. 1). Immunohistochemistry demonstrated that the tissue was immunoreactive for vimentin, actin and desmin but were uniformly negative for the vascular markers, CD34 and S-100.

Also at nine years of age, the patient was hospitalized twice with pneumonia. Chest computed tomography showed eleven bilateral pulmonary nodules (measuring 2.0 mm–1.1 cm), mediastinal lymphadenomegaly and a tissue mass infiltrating the bronchus to the lower lobe of the left lung. This mass also infiltrated the left inferior pulmonary vein, leading the inference that it was probably a neoplastic lesion. The abdominopelvic CT Scan showed a rounded lesion with well-defined edges, located at the liver, measuring approximately 2.1 × 1.9 cm, solid nodular lesions attached to the walls of the descending and sigmoid colon, cholecystolithiasis and nephrolithiasis. The patient underwent left lower lobectomy. Pathological analysis of the product of the resection showed myofibroblastic proliferation (Fig. 1B), bronchiectasis and bronchopneumonia. Immunohistochemistry confirmed the diagnosis of infantile myofibromatosis.

2.3. Proband's brother

This second patient, proband's brother, was first seen at the Pediatric Clinic of Hospital das Clínicas of UFMG at six years of age because of crusty skin lesions in the trunk, head segment, scalp, upper and lower limbs, which had begun to appear since one year of age, diagnosed as seborrheic dermatitis.

Total abdominal ultrasonography was performed because of his brother diagnosis and showed at least four hepatic nodules and in the hepatic hilum. Esophagus, stomach and duodenum were normal at endoscopy. Colonoscopy showed indurated lesions with subepithelial aspect with erosion and apical ulcerations of various sizes located in all colonic segments, with endoscopic appearance suggestive of infantile myofibromatosis. Presently, at age nine, his weight and height are below the 3rd percentile. He has had a good performance in school.

3. Methods

3.1. Samples and DNA isolation

This study was approved by the Research Ethics Committee of the Universidade Federal de Minas Gerais. The written informed consent was undersigned by participants and/or their caretakers. Genomic DNA was isolated from peripheral blood of the patients and their parents using a modified salting out procedure [Miller et al., 1988].

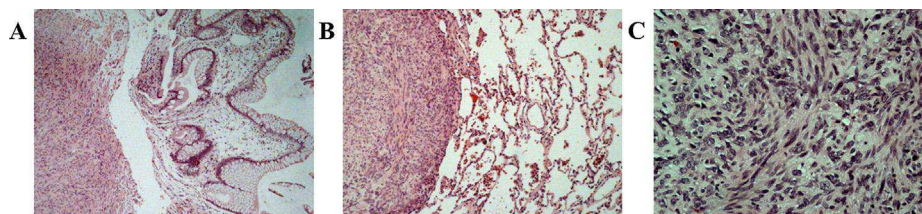


Fig. 1. Hematoxylin and eosin (H&E) stained sections of the tumor. (A) Bowel biopsy at magnification of 100× showing intestine villi at right of the figure and myofibroblasts at the left side. (B) Lung biopsy at 100× magnification showing alveolar septum at right of the figure and myofibroblasts at the left side. (C) Tumor at magnification of 400× showing that it was composed of spindle cells, which resembled normal smooth muscle and fibroblast cells arranged in a whorled and fascicular pattern.

3.2. Genotyping analysis

Genotyping on the Affymetrix Genome-Wide Human SNP Array 6.0 was performed at Aros Applied Biotechnology, Aarhus, Denmark.

3.3. Exome sequencing and analysis

Exome sequencing was performed in samples from the two brothers and their healthy parents by The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada using the Agilent SureSelect Human All Exon V4 Kit (Agilent Technologies), which targeted 20,965 genes and 334,378 exons. Enriched genomic DNA was massively parallel sequenced on the SOLiD 5500xl (Applied Biosystems). Approximately 130 million of 75 bp × 35 bp paired-end reads per individual were returned. The average coverage was 103-fold, with *circa* 96% of the target bases being covered at least at 10× (Table 1). All data were aligned to the GRCh37/hg19 reference genome build via BFAST and BWA aligner. Variants were quality trimmed using Genome Analysis Toolkit (GATK 1.1.28) and they were annotated for functional effect by SnpEff 2.0.5. Alignment, calling and annotation of the variants against databases such as 1000 Genomes (April 2012 release), NHLBI Exome Sequencing Project (ESP6500) and Single Nucleotide Polymorphism database (dbSNP137) were done using a software developed in-house called *Mendel*, MD (RGCLL Cardenas and SDJ Pena, manuscript in preparation).

3.4. Sanger sequencing

Sanger sequencing was performed for validation of the variants of interest identified in exome analysis using the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems®) and the Applied Biosystems (ABI) 3130 Genetic Analyzer. Sequencing data was analyzed using the software *Sequencher* version 4.1.4 (Genes Code Corporation).

4. Results

Analysis of the Affymetrix 6.0 SNP-array genotype data of the patients and their parents did not show presence of any micro-deletions or microduplications. A coefficient of genetic relationship of 0.0134 was estimated between the two parents using the software *IBDelphi* [Carr et al., 2011], confirming the history of consanguinity.

Over 80,000 variants were identified in each individual by exome sequencing. Approximately 14,000 of these variants passed in the GATK quality filters and had a Minor Allele Frequency

(MAF) ≤ 0.03 in the 1000 Genomes and ESP6500 (Supplementary Table 1). For the filtering process we had to test different inheritance models. The fact that both parents were phenotypically normal and that two children are affected rendered dominant inheritance (autosomal or X-linked) unlikely.

We first tested an X-linked recessive inheritance. For that, variants were narrowed down using five consecutive filters: (1) variants on the X chromosome hemizygous in the patients, heterozygous in the mother and not present in the father; (2) variants which passed in GATK quality filters; (3) variants with read depth > 10; (4) rare variants in the population (MAF ≤ 0.005 in the 1000 Genomes, ESP6500 and dbSNP137); (5) variants with impact moderate or high by SnpEff. None of the variants were predicted as deleterious, i.e., all had SIFT score above 0.05.

We then tested the autosomal recessive model of inheritance, which received support from the presence of distant consanguinity between the parents. Variants were narrowed down using the previously described five consecutive filters except for the first of them, which was replaced by “variants homozygous in the patients and heterozygous in the parents” (Fig. 2). It resulted in the identification of candidate variants in two genes: *NDRG4* and *RLTPR* (OMIM 610859). The brothers were homozygous for both variants and the healthy parents were heterozygous (Table 2).

The variant in *RLTPR* (NM_001013838.1; Chr16:67681506; Intron 11; c.871+1G>T) was a splice site donor mutation. This gene has 12 transcripts described in Ensembl Genome Browser (splice variants). However, none of them retained the intron 11. Alignment of *RLTPR* orthologues were made using the software *Alamut* version 2.4.5 (Interactive Biosoftware) and showed that the variant is located in a region not well conserved throughout evolution (Fig. 3A). This variant had not been described previously by the 1000 Genomes Project, ESP6500 and dbSNP137.

The variant in *NDRG4* (NM_001130487.1; Chr16:58538536; Exon 9; c.511G>C) was a missense substitution of p.Val171Leu. The variant is located in a well conserved region throughout evolution in *NDRG4* orthologues; except for the *Drosophila melanogaster* and *Caenorhabditis elegans* sequences (Fig. 3B). The variant in *NDRG4* was validated by Sanger sequencing in samples from the parents and both brothers (Fig. 4A). It was predicted to be deleterious by SIFT (score 0.00), possibly damaging by PolyPhen-2 (score 0.717) and disease causing by Mutation Taster (*p*-value 1.00). This variant also had not been described in the previously described databases. Additionally, this G/C variant had not been seen in any of TCAG's exome or whole genome databases, consisting of 436 autism samples sequenced on SOLiD, 1083 whole genome samples sequenced by Complete Genomics, 37 Clinical exome samples on an Illumina HiSeq platform and 80 samples sequenced on an Illumina HiSeq platform. However, a heterozygous G/A variant had been detected in one Complete Genomics samples.

Autozygosity mapping was performed on the exome sequence data of the children and their parents using *AgileGenotyper* and *AgileVariantMapper* [Carr et al., 2013]. It revealed a large run of homozygosity at chromosome 16 extended from 58,319,834 to 70,680,838 in the first son and from 58,319,950 to 71,101,288 in the second son (hg19). Interestingly, this autozygous region is exclusively shared by the affected brothers and includes the *NDRG4* gene (which is located at chr16:58,497,548–58,547,522) and the *RLTPR* gene (chr16:67,679,029–67,691,471) (Fig. 4B).

5. Discussion

In the present study, exome sequencing identified novel homozygous variants in *NDRG4* and *RLTPR* in two brothers described herein with multicentric IM. The *NDRG4* gene seems to be the better candidate as discussed below.

Table 1
Summary of exome sequencing data for each sample.

	1st son	2nd son	Father	Mother
Total captured regions size	51 Mb	51 Mb	51 Mb	51 Mb
% of captured regions with coverage >10	96.2%	95.8%	96.2%	96.4%
Average coverage of captured region (×)	107×	96×	102×	107×
Total number of SNPs	77,154	77,482	76,566	79,324
Total number of INDELs	6743	6562	6705	6965
N rare homozygous (n confirmed by Sanger) ^a	694 (1)	785 (1)	767 (0)	616 (0)
N compound heterozygous (n confirmed by Sanger) ^a	754 (0)	1029 (0)	—	—
N X linked (n confirmed by Sanger) ^a	61 (0)	53 (0)	—	—
N de novo events (n confirmed by Sanger) ^a	4253 (0)	8615 (0)	—	—

Frequency [MAF] ≤ 0.03 in the 1000 Genomes and ESP6500.

^a Only considered variants which passed in GATK quality filters and with Minor Allele.

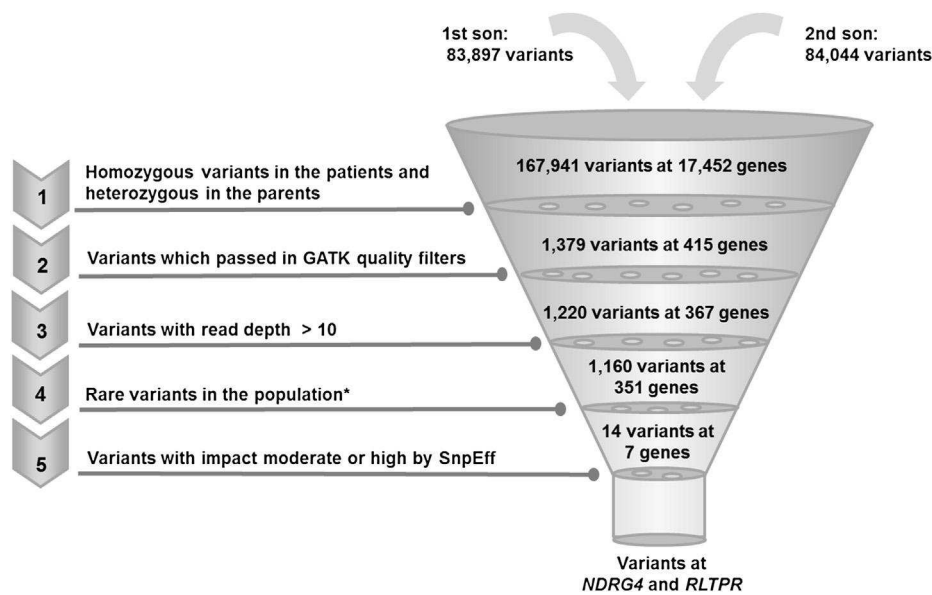


Fig. 2. Illustration of the analysis method using progressive filters. Candidate variants were narrow down using five consecutive filters based on autosomal recessive model of inheritance. *Minor allele frequency [MAF] ≤ 0.005 in the 1000 Genomes, ESP6500 and dbSNP137.

The RLTPR protein is a member of the leucine-rich repeats (LRRs) family with a RGD motif, a tropomodulin region and a proline-rich domain [Matsuzaka et al., 2004]. Its function is unclear, but it might be involved with T cell-dependent autoimmunity: its

Table 2
Details of the variants narrowed down using consecutive filters based on autosomal recessive model of inheritance.

Chromosome	chr16	chr16
Position	58538536	67681506
Gene name	<i>NDRG4</i>	<i>RLTPR</i>
Reference	G	G
Number of reads with reference	3	1
1st SON		
Alternative 1st SON	C	T
Number of reads with alternative	126	111
1st SON		
Number of reads with reference	2	6
2nd SON		
Alternative 2nd SON	C	T
Number of reads with alternative	98	101
2nd SON		
Number of reads with reference	66	65
MOTHER		
Alternative MOTHER	C	T
Number of reads with alternative	38	54
MOTHER		
Number of reads with reference	60	63
FATHER		
Alternative FATHER	C	T
Number of reads with alternative	55	49
FATHER		
Mutation type	Missense	Intron
Refseq	NM_001130487.1	NM_001013838.1
Mutation DNA	c.511G>C	c.871+1G>T
(HGVS nomenclature_c.)		
Mutation protein	p.Val171Leu	unknown
(HGVS nomenclature_p.)		
Prediction from (SIFT, Polyphen2, Mutation Taster, etc.)	SIFT: Deleterious PP2: Possibly damaging Mutation Taster: Disease causing	Not predicted
Sanger verification	YES	NO

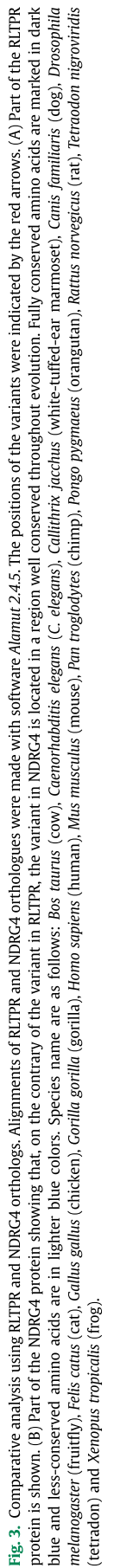
expression had been shown to be downregulated in psoriatic skin and it had been demonstrated that in the mouse, the protein Rltpr is essential for costimulation via CD28 and the development of regulatory T cells [Liang et al., 2013; Matsuzaka et al., 2004]. Additionally, expression of Rltpr had been detected exclusively in lymphoid tissues [Liang et al., 2013].

On the other hand, *NDRG4* is a member of the N-myc down-regulated gene family (NDRG), which is composed of four members named NDRG1–4. The isoforms of *NDRG4* mRNA in human organs/tissues had been identified exclusively in the brain and heart [Reviewed by Yang et al., 2013]. Consistent with the phenotype of IM, *NDRG4* is a tumor-related gene; expression of *NDRG4* has been shown to be decreased in numerous cancer cell lines and tumor types, such as in human gliomas and colorectal cancer tissue, suggesting that it may exert function as a tumor suppressor gene in these tissues [Reviewed by Yang et al., 2013]. Furthermore, it had been demonstrated that *NDRG4* has roles in aggressive aspects of the tumor biology such as colony formation, cell proliferation and tumor invasion [Reviewed by Yang et al., 2013].

Nishimoto et al., [2003] demonstrated that *NDRG4* (also known as *SMAP8*) is involved in the regulation of mitogenesis via MAPK signaling in aortic smooth muscles cells (A10 cells) of rats. The authors demonstrated that *smap8* overexpression decreased during proliferation and migration of A10 cells, whereas the sensitivity to platelet-derived growth factor (PDGF) was enhanced. Prior to ERK1/2 activation, *smap8* protein was phosphorylated by the staurosporine-sensitive kinase, which is specifically activated by PDGF.

Remarkably in line with our results, variants in *PDGFRB* and *NOTCH3* were recently related with the autosomal dominant form of IM [Cheung et al., 2013; Martignetti et al., 2013]. Similarly to *NDRG4*, *PDGFRB* promotes growth of mesenchymal cells, including blood vessels and smooth muscles, which are affected in IM. In addition, the expression of *PDGFRB* is regulated by NOTCH signaling in vascular smooth muscle cells [Jin et al., 2008]. Moreover, as previously described, there is a functional association between the *NDRG4* gene and PDGF, suggesting that these variants are possibly involved in the same pathway leading to IM.

In conclusion, our findings suggest that variants in *NDRG4* may be involved with the autosomal recessive form of IM. The variant in



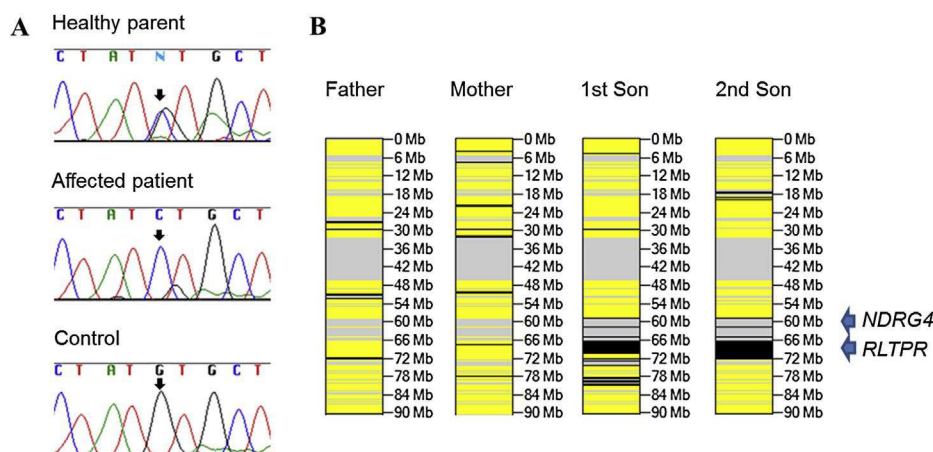


Fig. 4. Results of the Sanger sequencing and autozygous analysis. (A) Fragments of sequence chromatograms are shown of *NDRG4* from a healthy parent (heterozygous mutation), an affected patient (homozygous mutation) and a control individual (normal homozygous). Black arrows indicate the position of the variant >C. (B) Genotype data for chromosome 16 created by *AgileGenotyper* and *AgileVariantMapper* using heterozygous cut-off ratio of 40%. The columns represent the autozygous profile of the father, mother and sons, respectively. Each row represents the call of a single SNP on chromosome 16 (yellow, heterozygous; black, homozygous; gray, no-call). Note the large run of homozygosity from ~58 Mb to 71 Mb that is exclusively shared by the affected brothers; it is an autozygous region including *NDRG4* and *RLTPR*.

NDRG4 gene, in contrast of the one in *RLTPR* gene, is located in a region well conserved throughout evolution, is predicted as deleterious and, especially, is associated with tumor development. In addition, we highlight that we have demonstrated a putative evidence of autosomal recessive causality of IM and, unless other families are identified with the mutation, future functional studies will be needed for confirmation.

Conflict of interest statement

Authors do not have any conflict of interest.

Acknowledgments

The authors are grateful to the patients and their family for their precious cooperation in this study. We thank Dr. Moisés Pedrosa for the photomicrographs of patient's biopsies and Dr. Sérgio Pereira of The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada for assistance with exome data interpretation. This work was supported by FAPEMIG grant to SDJP (process CDS-30/11). NDL was supported by a fellowship from CNPq and MCMF and RGCLC were supported by a fellowship from CAPES.

Appendix A. Supplementary data

Supplementary data related to this article can be found at <http://dx.doi.org/10.1016/j.ejmg.2014.08.010>.

References

- Alberti LR, Souto Bittencourt PF, Rodrigues Ferreira A, Rodrigues Da Silva RR, Diniz Carvalho S, Nunes Cachicolo FM, et al. Multicentric infantile myofibromatosis of the small bowel detected by video capsule endoscopy in a child. *Endoscopy* 2012;44(Suppl. 2 UCTN):E258–9.
- Carr IM, Bhaskar S, O'Sullivan J, Aldahmesh MA, Shamseldin HE, Markham AF, et al. Autozygosity mapping with exome sequence data. *Hum Mutat* 2013;34:50–6.
- Carr IM, Markham SA, Pena SD. Estimating the degree of identity by descent in consanguineous couples. *Hum Mutat* 2011;32:1350–8.
- Cheung YH, Gayden T, Campeau PM, Leduc CA, Russo D, Nguyen VH, et al. A recurrent PDGFRB mutation causes familial infantile myofibromatosis. *Am J Hum Genet* 2013;92:996–1000.
- Chung EB, Enzinger FM. Infantile myofibromatosis. *Cancer* 1981;48:1807–18.
- Gandhi MM, Nathan PC, Weitzman S, Levitt GA. Successful treatment of life-threatening generalized infantile myofibromatosis using low-dose chemotherapy. *J Pediatr Hematol Oncol* 2003;25:750–4.

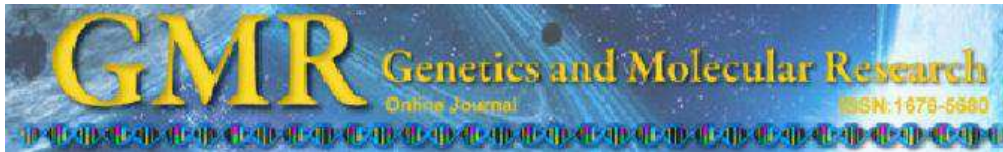
- Gatibelza ME, Vazquez BR, Bereni N, Denis D, Bardot J, Degardin N. Isolated infantile myofibromatosis of the upper eyelid: uncommon localization and long-term results after surgical management. *J Pediatr Surg* 2012;47:1457–9.
- Huang CJ, Lin KL, Jung SM, Wu CT, Wang HS. Infantile myofibromatosis presenting with scalp dermoid cyst. *Pediatr Neurol* 2005;33:296–9.
- Jin S, Hansson EM, Tikka S, Lanner F, Sahlgren C, Farnebo F, et al. Notch signaling regulates platelet-derived growth factor receptor-beta expression in vascular smooth muscle cells. *Circ Res* 2008;102:1483–91.
- Liang Y, Cucchetti M, Roncagalli R, Yokosuka T, Malzac A, Bertoso E, et al. The lymphoid lineage-specific actin-uncapping protein Rltpr is essential for costimulation via CD28 and the development of regulatory T cells. *Nat Immunol* 2013;14:858–66.
- Martignetti JA, Tian L, Li D, Ramirez MC, Camacho-Vanegas O, Camacho SC, et al. Mutations in PDGFRB cause autosomal-dominant infantile myofibromatosis. *Am J Hum Genet* 2013;92:1001–7.
- Matsuzaka Y, Okamoto K, Mabuchi T, Iizuka M, Ozawa A, Oka A, et al. Identification, expression analysis and polymorphism of a novel RLTPR gene encoding a RGD motif, tropomodulin domain and proline/leucine-rich regions. *Gene* 2004;343:291–304.
- Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res* 1988;16:1215.
- Nishimoto S, Tawara J, Toyoda H, Kitamura K, Komurasaki T. A novel homocysteine-responsive gene, smap8, modulates mitogenesis in rat vascular smooth muscle cells. *Eur J Biochem* 2003;270:2521–31.
- Orbach D. Infantile myofibromatosis. *Orphanet Encyclopedia*; November 2013. Available from: http://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=2591.
- Stout AP. Juvenile fibromatoses. *Cancer* 1954;7:953–78.
- Tamburrini G, Gessi M, Colosimo Jr C, Lauriola L, Giangaspero F, Di Rocco C. Infantile myofibromatosis of the central nervous system. *Child's Nerv Syst* 2003;19:650–4.
- Wiswell TE, Davis J, Cunningham BE, Solenberger R, Thomas PJ. Infantile myofibromatosis: the most common fibrous tumor of infancy. *J Pediatr Surg* 1988;23:315–8.
- Yang X, An L, Li X. NDRG3 and NDRG4, two novel tumor-related genes. *Biomed Pharmacother* 2013;67:681–4.
- Zand DJ, Huff D, Everman D, Russell K, Saitta S, McDonald-McGinn D, et al. Autosomal dominant inheritance of infantile myofibromatosis. *Am J Med Genet A* 2004;126A:261–6.

Web resources

- 1000 Genomes, <http://browser.1000genomes.org>.
- Ensembl Genome Browser, <http://www.ensembl.org/index.html>.
- Genome Analysis Toolkit (GATK), <http://www.broadinstitute.org/gatk/>.
- Mutation Taster, <http://www.mutationtaster.org/>.
- NCBI dbSNP, <http://www.ncbi.nlm.nih.gov/projects/SNP/>.
- NHLBI Exome Sequencing Project (ESP), <http://evs.gs.washington.edu/EVS/>.
- Online Mendelian Inheritance in Man (OMIM), <http://www.omim.org/>.
- PolyPhen-2, <http://genetics.bwh.harvard.edu/pph2/>.
- SIFT, <http://sift.jcvi.org/>.
- SnPEff, <http://snpeff.sourceforge.net/>.
- UCSC Genome Browser, <http://genome.ucsc.edu>.

ARTIGO 3

Modulation of expressivity in PDGFRB-related infantile myofibromatosis: a role for PTPRG?



Short Communication

Modulation of expressivity in *PDGFRB*-related infantile myofibromatosis: a role for *PTPRG*?

N.D. Linhares¹, M.C.M. Freire², R.G.C.C.L. Cardenas¹, M. Bahia³,
E. Puzenat⁴, F. Aubin⁴ and S.D.J. Pena^{1,2,5}

¹Laboratório de Genômica Clínica, Faculdade de Medicina,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

²Departamento de Bioquímica e Imunologia, Instituto de Ciências Biológicas,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil

³Divisão de Gastroenterologia Pediátrica,
Hospital das Clínicas da Universidade Federal de Minas Gerais,
Belo Horizonte, MG, Brasil

⁴Service de Dermatologie, Centre Hospitalier Universitaire,
Université de Franche Comté, Besançon, France

⁵Laboratório GENE, Núcleo de Genética Médica, Belo Horizonte, MG, Brasil

Corresponding author: S.D.J. Pena

E-mail: spena@dcc.ufmg.br

Genet. Mol. Res. 13 (3): 6287-6292 (2014)

Received June 26, 2014

Accepted June 30, 2014

Published August 15, 2014

DOI <http://dx.doi.org/10.4238/2014.August.15.11>

ABSTRACT. Infantile myofibromatosis is a rare genetic disorder characterized by the development of benign tumors in the skin, muscle, bone, and viscera. The molecular pathogenesis is still incompletely known. An autosomal dominant form had been reported as causally related with mutations in the gene for platelet-derived growth factor receptor beta (*PDGFRB*). We report here two siblings with infantile myofibromatosis and with a *PDGFRB* mutation identified by exome sequence analysis. However, the unaffected mother also had the same

PDGFRB mutation. We showed that both children had also inherited from their healthy father a heterozygous mutation in the gene for receptor protein tyrosine phosphatase gamma (*PTPRG*), an enzyme known to dephosphorylate *PDGFRB*. We suggest that in this family, the additional mutation in *PTPRG* may explain the full phenotypic penetrance in the siblings affected, in comparison with the unaffected mother.

Key words: Infantile myofibromatosis; *PDGFRB* gene; *PTPRG* gene; Incomplete penetrance; Whole-exome sequence analysis

INTRODUCTION

Infantile myofibromatosis (IM; MIM 228550) is a rare disorder characterized by the development of benign tumors in the skin, striated muscles, bones, and in exceptional cases, visceral organs (Orbach, 2013). Recently, two studies suggested that mutations in *PDGFRB* (gene for platelet-derived growth factor receptor, beta polypeptide; MIM 173410) were causally related to autosomal dominant IM (Cheung et al., 2013; Martignetti et al., 2013). In one of these studies, Cheung et al. (2013) identified the same c.1681C>T (p.Arg561Cys) mutation in *PDGFRB* in all the 11 affected individuals from four distinct families with IM. None of the unaffected parents had the mutation. In the second and larger study, Martignetti et al. (2013) performed whole-exome sequencing in 11 affected members of nine unrelated IM families. In eight of these they identified one of two pathogenic mutations in *PDGFRB*: c.1681C>T (p.Arg561Cys) or c.1978C>A (p.Pro660Thr).

While trying to establish the culpable gene in cases of autosomal recessive infantile myofibromatosis, we studied a French family with two affected children and their healthy non-consanguineous parents, previously described in detail (Puzenat et al., 2009). Whole-exome sequencing identified a *PDGFRB* mutation in both siblings with IM but, to our surprise, also in the unaffected mother. This is the first documentation of incomplete penetrance in *PDGFRB*-related IM.

MATERIAL AND METHODS

Whole-exome sequencing and analysis

Whole-exome sequencing was performed by the Centre for Applied Genomics, Hospital for Sick Children, Toronto, Canada, using the Agilent SureSelect Human All Exon V4 kit (Agilent Technologies, Santa Clara, CA, USA) and the SOLiD 5500xl platform (Applied Biosystems, Foster City, CA, USA).

All data were aligned via BFAST and BWA aligner to the hg19/GRCh37 reference genome. Variants were quality trimmed using Genome Analysis Toolkit (GATK 1.1.28) and they were annotated for functional effect by SnpEff 2.0.5. Alignment, calling and annotation of the variants against databases such as 1000 Genomes (April 2012 release), NHLBI Exome Sequencing Project (ESP6500) and Single Nucleotide Polymorphism database (dbSNP137) were done using a software developed in-house called Mendel, MD (RGCCCL Cardenas and SDJ Pena, manuscript in preparation).

Sanger sequencing

Sanger sequencing was performed for validation of the variants of interest identified in exome analysis using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, CA, USA) and the Applied Biosystems (ABI) 3730 Genetic Analyzer. Sequencing data was analyzed using the Sequencer version 4.1.4 software (Genes Code Corporation, Ann Arbor, MI, USA).

RESULTS

Over 100,000 variants were identified in each individual by exome sequencing. After filtering for common polymorphisms and non-pathogenic variants, we identified, in both siblings with IM, the same mutation c.1681C>T (p.Arg561Cys; NM_002609.3) in *PDGFRB*, which had been previously described by Cheung et al. (2013) and Martignetti et al. (2013). Sanger sequencing confirmed the presence of the heterozygous mutation (Figure 1A). However, to our surprise, we found that the unaffected mother also showed the same *PDGFRB* c.1681C>T heterozygous mutation.

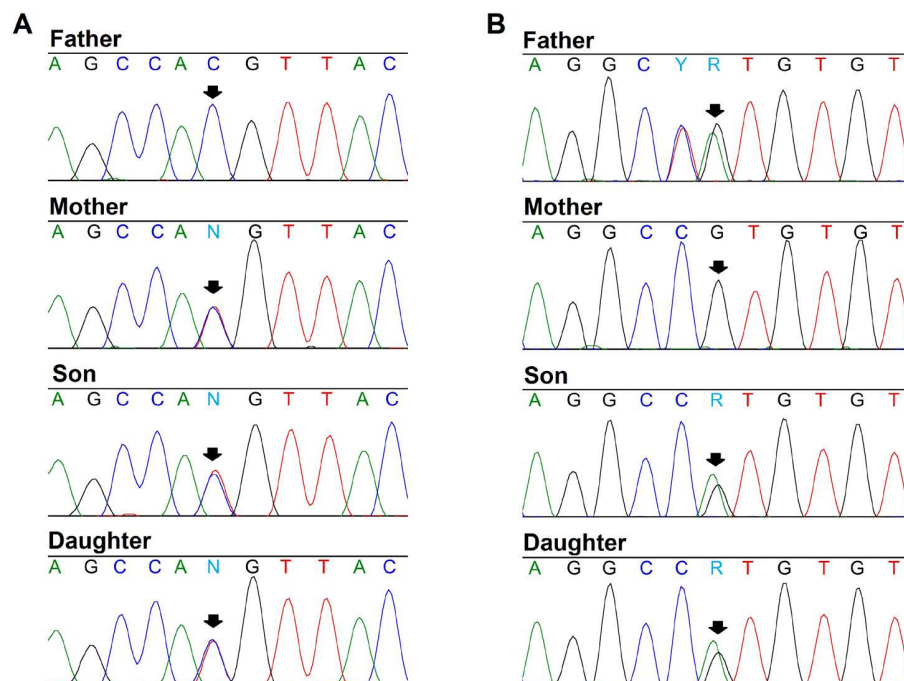


Figure 1. Sanger sequencing. **A.** Fragments of sequence chromatograms of *PDGFRB* from the healthy father (normal homozygous), the healthy mother (heterozygous mutation), and the affected son and daughter (heterozygous mutation). Black arrows indicate the position of the variant c.1681C>T (p.Arg561Cys). **B.** Fragments of sequence chromatograms are shown of *PTPRG* from the healthy father (heterozygous mutation), the healthy mother (normal homozygous), and the affected son and daughter (heterozygous mutation). Black arrows indicate the position of the variant c.1276G>A (p.Val426Met).

We know that some variants may require the presence of additional pathogenic variants at other loci for full penetrance (Cooper et al., 2013; Schaffer, 2013). We thus searched the affected siblings of our family for paternal mutations in genes known to interact with *PDGFRB*. Indeed, we found that both siblings had inherited from their healthy father a heterozygous mutation in *PTPRG* (protein tyrosine phosphatase, receptor type, gamma; MIM 176886), which codes for an enzyme known to dephosphorylate PDGFRB, reducing its activity (Barr et al., 2009). We believe that the *PTPRG* mutation might have contributed to the phenotype penetrance in the children and non-penetrance in the mother. The existence of the *PTPRG* mutation (NM_002841.3; p.Val426Met; c.1276G>A; rs139041779) was confirmed by Sanger sequencing (Figure 1B). According to the NHLBI Exome Sequencing Project (ESP6500), it has an allele frequency of 0.0001. The mutation is located in a region well conserved throughout evolution in *PTPRG* orthologues (Figure 2). It was predicted to be deleterious by SIFT (score 0.00), probably damaging by PolyPhen-2 (score 0.979) and disease causing by Mutation Taster (P value of 0.999).

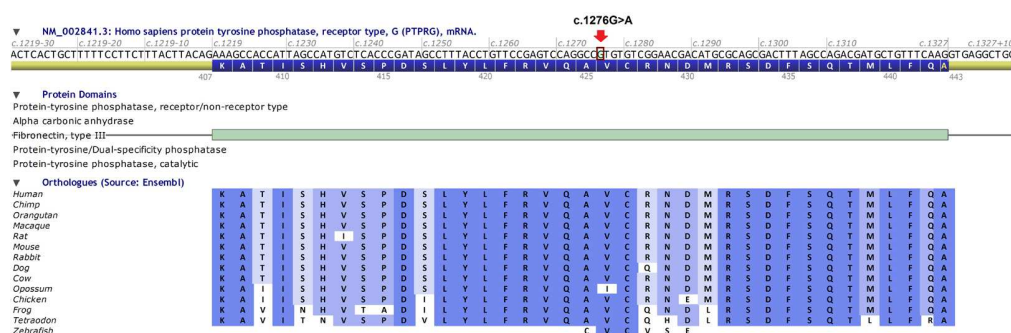


Figure 2. Comparative analysis using PTPRG orthologues. Part of the PTPRG protein showing that the variant is located in a region well conserved throughout evolution. The position of the variant was indicated by the red arrow. Alignments of PTPRG orthologues were made with the Alamut 2.4 software. Fully conserved amino acids are marked in dark blue and less-conserved amino acids are in lighter blue colors. Species name are as follows: *Homo sapiens* (human), *Pan troglodytes* (chimpanzee), *Pongo pygmaeus* (orangutan), *Macaca mulatta* (macaque), *Rattus norvegicus* (rat), *Mus musculus* (mouse), *Oryctolagus cuniculus* (rabbit), *Canis familiaris* (dog), *Bos Taurus* (cow), *Monodelphis domestica* (opossum), *Gallus gallus* (chicken), *Xenopus tropicalis* (frog), *Tetraodon nigroviridis* (tetraodon), *Danio rerio* (zebrafish).

DISCUSSION

Incomplete penetrance has been previously proposed to occur in families with the autosomal dominant form of IM (Bracko et al., 1992; Kulkarni et al., 2012). In fact, in two of the families described by Martignetti et al. (2013) [family IM-4 and the previously published family IM-6 (Zand et al., 2004)], there might have been non-penetrance in individuals I-1 and I-2 from both pedigrees.

It is imaginable that the *PDGFRB* and *PTPRG* mutations in the children described here are additive events that could cooperatively lead to tumorigenesis through the Ras/MAPK pathway and cause IM. The p.Arg561Cys mutation appears to lead to increased PDGFRB kinase activity after stimulation by the ligand (Cheung et al., 2013). The *PTPRG* mutation seen

in our patients would probably cause a decrease in its efficiency to dephosphorylate PDGFRB, thus enhancing its activation and stimulating downstream signaling pathways that may induce cellular hyperproliferation (Figure 3). It remains to be established whether the *PTPRG* mutation is a private feature of our family, or whether it is a more general phenomenon. Thus, it will certainly be of interest to investigate the integrity of *PTPRG* in the cases described by Cheung et al. (2013) and Martignetti et al. (2013).

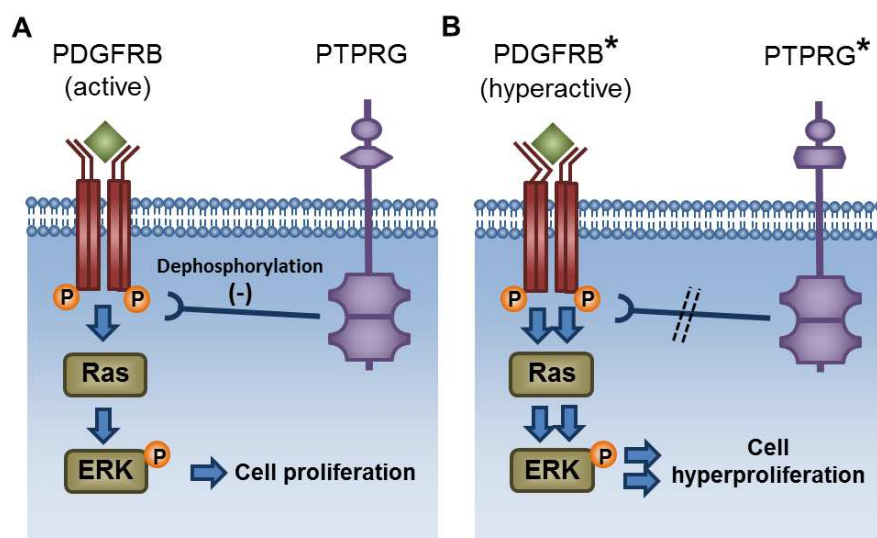


Figure 3. Schematic representation of the mechanisms of interaction of PDGFRB and PTPRG. Mutant proteins are indicated by an asterisk. **A.** Normal individual. Upon binding by PDGF, PDGFRB dimerizes and is autophosphorylated in multiple specific tyrosine residues, leading to the initiation of signal transduction pathways such as Ras/MAPK pathway. PTPRG modulates PDGFRB activation by dephosphorylation. **B.** Our patients. Mutant PDGFRB is excessively activated. Mutant PTPRG has decreased dephosphorylation activity, enhancing PDGFRB activation. This leads to cell hyperproliferation through Ras/MAPK pathway.

In conclusion, our data confirm that mutations in *PDGFRB* cause IM. Additionally, we suggest that a mutation in *PTPRG* may act as an aggravating factor, and we believe this explains the phenotypic penetrance in the two children studied here. Mutations in other interacting genes, especially *PTPRG*, as shown in the present study, may modulate the degree of expressivity of autosomal dominant infantile myofibromatosis.

ACKNOWLEDGMENTS

Research supported by Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) to SDJ Pena (process CDS-30/11). We thank Dr. Sérgio Pereira of the Centre for Applied Genomics, Hospital for Sick Children, Toronto, Canada, for assistance with exome sequencing and data interpretation and Juliano Leal de Paula for assistance with Sanger sequencing.

REFERENCES

- Barr AJ, Ugochukwu E, Lee WH, King ON, et al. (2009). Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* 136: 352-363.
- Bracko M, Cindro L and Golouh R (1992). Familial occurrence of infantile myofibromatosis. *Cancer* 69: 1294-1299.
- Cheung YH, Gayden T, Campeau PM, LeDuc CA, et al. (2013). A recurrent PDGFRB mutation causes familial infantile myofibromatosis. *Am. J. Hum. Genet.* 92: 996-1000.
- Cooper DN, Krawczak M, Polychronakos C, Tyler-Smith C, et al. (2013). Where genotype is not predictive of phenotype: towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* 132: 1077-1130.
- Kulkarni K, Desai S, Grundy P and Sergi C (2012). Infantile myofibromatosis: report on a family with autosomal dominant inheritance and variable penetrance. *J. Pediatr. Surg.* 47: 2312-2315.
- Martignetti JA, Tian L, Li D, Ramirez MC, et al. (2013). Mutations in PDGFRB cause autosomal-dominant infantile myofibromatosis. *Am. J. Hum. Genet.* 92: 1001-1007.
- Orbach D (2013). Infantile Myofibromatosis. Available at [http://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=2591]. Accessed April 9, 2014.
- Puzenat E, Marioli S, Algros MP, Faivre B, et al. (2009). Familial infantile myofibromatosis. *Ann. Dermatol. Venereol.* 136: 346-349.
- Schaffer AA (2013). Digenic inheritance in medical genetics. *J. Med. Genet.* 50: 641-652.
- Zand DJ, Huff D, Everman D, Russell K, et al. (2004). Autosomal dominant inheritance of infantile myofibromatosis. *Am. J. Med. Genet. A* 126A: 261-266.

ARTIGO 4

Análise de exomas de pacientes com microcefalia, convulsões e calcificações no tegmento pontino do cérebro

INTRODUÇÃO

Calcificações intracraniais são frequentemente observadas nos gânglios de base e foram relacionadas com mais de 50 doenças genéticas, neurológicas e de metabolismo de cálcio, incluindo doença celíaca, desordens mitocondriais, leucoencefalopatia com cistos, hipoparatiroidismo e pseudohipoparatiroidismo^[1, 2, 3, 4]. Achados adicionais frequentes em pacientes com calcificações intracraniais de etiologia desconhecida incluem microcefalia, convulsões, surdez neurosensorial e anormalidades de retina^[4]. Em crianças, calcificações intracerebrais são mais frequentemente adquiridas em associação com a infecção intra-uterina, hemorragia intra-uterina ou pós-natal precoce, ou neoplasia^[5].

Estudos patológicos mostram que o cálcio é o elemento principal presente nas calcificações intracranianas e caracteriza a aparência radiológica da doença, entretanto mucopolissacarídeos e traços de zinco, fósforo, ferro, alumínio e magnésio também estão presentes^[6, 7]. A calcificação ocorre geralmente nas paredes dos vasos, no espaço perivascular das arteríolas, capilares e também nas veias^[6].

Calcificações no tegmento pontino sem o envolvimento dos gânglios de base ou do cerebelo são muito raras, foram descritos apenas dois casos, e sua etiologia é pouco conhecida^[8]. Stamelou *et al.*^[8] descreveram uma mulher de 53 anos com ataxia progressiva com tremor palatal associada à uma calcificação acentuada no tegmento pontino. Livingston *et al.*^[9] descreveram uma criança do sexo masculino de 7 anos com atraso de desenvolvimento motor, convulsões, diplegia espástica leve e calcificação no tegmento pontino e substância branca profunda e periventricular. O sequenciamento do gene *COL4A1* (que codifica a subunidade alfa-1 do colágeno tipo IV) detectou uma mutação heterozigota em c.1973G>A (p.Gly658Asp), que os autores concluíram estar relacionada com o fenótipo do paciente. Mutações no gene *COL4A1* foram recentemente relacionadas com doenças de pequenos vasos que afetam o cérebro, olhos e rins, com hemorragias intracerebrais e danos isquêmicos (OMIM *120130).

No presente trabalho estudamos duas crianças sem parentesco (um menino de 10 anos e uma menina de 2 anos) que apresentavam microcefalia e convulsões recorrentes unilaterais seguidas de hemiparasia pós ictal prolongada transitória. Ambos apresentavam achados de neuroimagem peculiares: calcificações grosseiras no tegmento pontino médio e calcificações muito sutis na transição córtico-subcortical. Uma investigação extensiva foi realizada, excluindo infecções congênicas e desordens metabólicas.

Realizamos então análise de genotipagem com o objetivo de identificar variantes estruturais responsáveis pela desordem neurológica encontrada. Detectamos uma duplicação parcial do gene *PDGFRA* (*platelet-derived growth factor receptor, alpha polypeptide*; OMIM *173490) em ambos os pacientes. Porém, após análise por PCR quantitativa em tempo real, constatamos que esta duplicação era falso-positiva. Depois de excluídas as microdeleções e microduplicações pela análise de genotipagem, prosseguimos para a análise dos exomas. Descobrimos que os pacientes apresentam variantes heterozigotas nos genes *SCN1A* (*sodium channel, neuronal type I, alpha subunit*; OMIM *182389) e *CHD2* (*chromodomain helicase DNA binding protein 2*; OMIM *602119), os quais são causalmente relacionados com encefalopatia epiléptica infantil (síndrome de Dravet; OMIM #607208). Pretendemos realizar o sequenciamento de Sanger desses genes para confirmar as variantes antes de proceder análises futuras.

RELATO CLÍNICO

Paciente I

Primeira filha de casal saudável não consanguíneo nasceu de parto cesárea com 2.925 g de peso e 48 cm de comprimento. Com quatro meses foi observado que sua fontanela anterior estava quase fechada e que seus marcos de desenvolvimento foram ficando atrasados. Com nove meses seu desenvolvimento era de uma criança de seis meses. A tomografia computadorizada do cérebro mostrou calcificação grosseira no tegmento pontino médio e calcificações muito sutis na transição córtico-subcortical (Fig. 1A e B). Com 12 meses e com 13 meses apresentou convulsões parciais, ficando hemiparética à esquerda na primeira crise e à direita no pós-comicial da segunda crise.

Com um ano e dois meses, foi avaliada pelo serviço de Neurologia do Hospital das Clínicas da UFMG (HC-UFMG); seu perímetro cefálico era 42 cm (abaixo do percentil 2.5) e o exame neurológico mostrou uma criança alerta, com bom contato visual, capaz de localizar sons e com moderado atraso no desenvolvimento. Sentava sem suporte e engatinhava de uma maneira diferente (com impulsão dos membros inferiores e sustentação com os membros superiores). Não apresentava deficiências motoras e seu tônus muscular, reflexos e coordenação motora eram normais. Estava tomando fenobarbital 50 mg ao dia e não teve mais convulsões.

A triagem metabólica foi normal, incluindo ácidos orgânicos da urina, aminoácidos do plasma, perfil de acilcarnitinas, lactato sanguíneo, amônia, função hepática e glicemia. Anti-gliadina e anti-endomísio foram negativos. Sorologia para

citomegalovírus e toxoplasmose (IgM e IgG) foram ambos negativos. Foram realizadas análises dos níveis de paratormônio (PTH), análise do líquido, cariótipo, eletroencefalografia (realizada três vezes), avaliação oftalmológica (incluindo eletrorretinografia), todos normais. A tomografia computadorizada do cérebro, realizada com oito anos, não mostrou informações adicionais. Ressonância magnética do cérebro e angiorressonância magnética não mostraram evidências de anormalidades de substância branca e nem formação de cistos.

Paciente II

Primeiro filho de casal saudável não consanguíneo nasceu de parto normal com peso de 3.220 g, 47 cm de comprimento e perímetro cefálico de 32 cm. Seu desenvolvimento motor foi normal no primeiro ano de vida, sentou com seis meses de idade e andou com 13 meses. Com seis meses de idade apresentou hipotonia episódica com cianose perioral. Com 11 meses ficou muito irritado e hipotônico, com episódios de vômitos e olhar fixo desviado. Foi internado no hospital, onde foi realizada a tomografia computadorizada que mostrou calcificação grosseira no tegmento pontino médio e calcificações sutis na transição cortico-subcortical. Foi receitado carbamazepina.

Com 2 anos de idade apresentou convulsões parciais com generalização secundária, com hemiparesia direita. Depois de um ano apresentou os mesmos sintomas, porém com hemiparesia do lado esquerdo. Durante os três anos seguintes os episódios ocorreram uma vez ao ano, sendo que o último episódio ocorreu com seis anos.

Foi avaliado pelo serviço de Neurologia do HC-UFMG com nove anos; seu exame físico foi normal, com perímetro cefálico de 48 cm (abaixo do percentil 2.5). Desenvolveu comportamento hiperativo e atraso cognitivo, sem qualquer disfunção motora. Exames oftalmológicos foram normais. Foram realizados os exames a seguir, todos normais: ácidos orgânicos da urina, aminoácidos do plasma, perfil de acilcarnitinas, lactato sanguíneo, amônia, função hepática, glicemia, análise do líquido (incluindo folato, interferon-alfa e neurotransmissores), eletroencefalografia, sorologia para citomegalovírus e toxoplasmose (IgG e IgM) e cariótipo. Anticorpos anti-gliadina e os níveis de paratormônio foram normais.

Ressonância magnética do encéfalo não mostrou informações adicionais (sem evidência de cistos ou leucoencefalopatia). A espectroscopia de prótons mostrou leve redução no pico de N-acetil-aspartato e leve aumento no pico da colina. Foi realizada a

tomografia computadorizada do cérebro, mantendo as alterações do exame realizado com 11 meses (Fig. 1C e D).

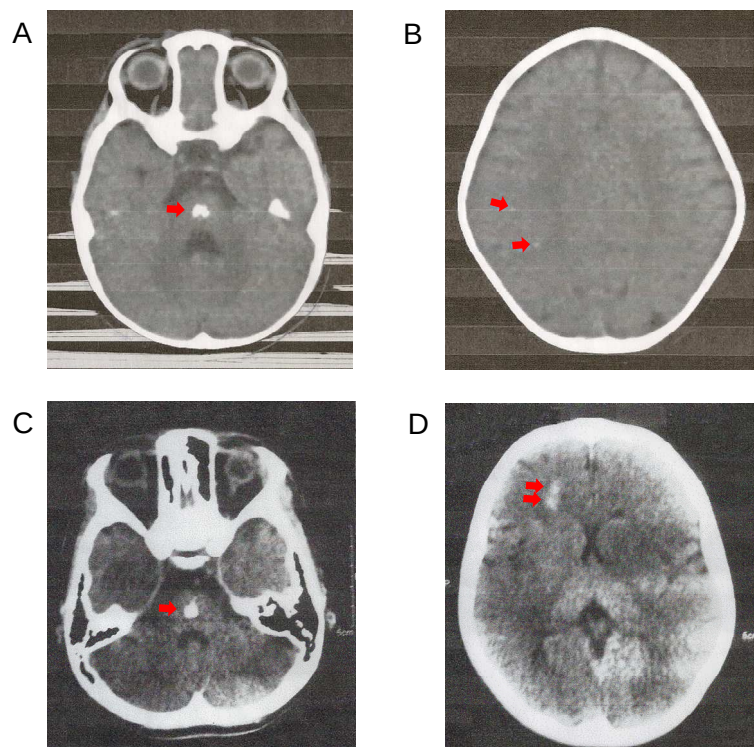


Fig. 1. Tomografia computadorizada do cérebro da paciente I realizada com 9 meses (A e B) e do paciente II realizada com 9 anos (C e D). A e C) Note a calcificação grosseira no tegmento pontino médio, indicadas pelas setas. B e D) Note as calcificações sutis na transição córtico-subcortical, na esquerda, indicadas pelas setas.

MÉTODOS

Amostras e extração de DNA

Este estudo foi aprovado pelo Comitê de Ética e Pesquisa da Universidade Federal de Minas Gerais e pela Comissão Nacional de Ética em Pesquisa (CONEP), com número de parecer 854.709. O Termo de Consentimento Livre e Esclarecido (TCLE) incluindo permissão para publicação foi assinado pelos pais. O DNA genômico foi isolado de amostra de sangue periférico dos pacientes e de seus pais utilizando procedimento *salting out*^[10].

Análise de genotipagem

Análises de genotipagem utilizando o *Affymetrix Genome-Wide Human SNP Array 6.0* foram realizadas pelo serviço terceirizado da Aros Applied Biotechnology

(Aarhus, Dinamarca) em amostras de ambos os pacientes e dos pais da paciente I. O *Affymetrix Genome-Wide Human SNP Array 6.0* fornece 950 mil marcadores para CNV, incluindo 202.000 sondas destinadas para 5.677 regiões de CNVs do Banco de Dados *Toronto Database of Genomic Variants* e 744.000 sondas uniformemente espaçadas ao longo do genoma. Além disso, contêm 900 mil marcadores para SNPs que permitem detectar perdas de heterozigose, dissomias uniparentais e regiões do genoma que são idênticas por descendência. A análise dos resultados foi realizada utilizando o *software CNViewer*^[11]. Este *software* foi projetado para comparar o número de cópias (Estado CN ou *CN state*) e perda de heterozigosidade (*Loss of heterozigosity* - LOH) dos resultados de Affymetrix SNP array 6.0 entre os membros afetados e não afetados de um *pedigree*.

PCR quantitativa em tempo real (qPCR)

Para validar a duplicação do gene *PDGFRA* encontrada pela análise de genotipagem, foi realizada a PCR quantitativa em tempo real (qPCR) em amostras dos pacientes conforme metodologia de Weksberg *et al.*^[12] e Hughes *et al.*^[13] e descrita detalhadamente a seguir.

Cada qPCR continha 5 µL de *SYBR® Green PCR Master Mix 2X* (Applied Biosystems); 10 ng de DNA genômico; cada par de *primers* em sua concentração ótima e água *nuclease-free*, totalizando um volume final de 10 µL. As condições de amplificação consistiram de uma etapa inicial a 50°C por 2 minutos, uma etapa de desnaturação a 95°C por 10 minutos, seguido de 50 ciclos de: 95°C por 15 segundos (desnaturação) e uma etapa única de anelamento e extensão a 60°C por 1 minuto. Após os 50 ciclos de amplificação, todas as amostras foram submetidas à desnaturação gradual para determinação da curva de dissociação (*melting curve*). Em todas as reações foi feito um controle branco de amplificação para cada par de *primers*, onde água estéril foi acrescentada no lugar de DNA. As corridas de qPCR foram feitas no equipamento *Rotor-Gene Q* (QIAGEN) e a análise dos resultados foi realizada utilizando o *Rotor-Gene Q Software* (QIAGEN).

Os *primers* para o gene alvo *PDGFRA* foram desenhados utilizando o programa *Primer3*, versão 4.0.0 (<http://primer3.ut.ee/>), e os *primers* para amplificação do gene de referência *HMBS* foram obtidos a partir de Weksberg *et al.*^[12] (nesse trabalho, o gene *HMBS* foi denominado *HEM3*, no entanto, optamos por utilizar no presente estudo os símbolos de genes aprovados pelo HUGO Gene Nomenclature Committee (HGNC))

(<http://www.genenames.org>)). As sequências dos *primers* e o tamanho esperado de cada *amplicon* são apresentados na Tabela 1. Reações de amplificação por PCR foram feitas com o objetivo de confirmar a especificidade dos *primers* para amplificação do gene *PDGFRA* e, posteriormente, o produto da amplificação foi visualizado em gel de agarose (1%) corado com brometo de etídeo.

Tabela 1. Sequências dos *primers* usados na qPCR e tamanho esperado dos *amplicons*.

<i>Primer</i>	Sequência (5'→3')	Tamanho do <i>amplicon</i> (pb)
PDGFRA-F	AACCCTGCTGATGAAAGCAC	136
PDGFRA-R	GCTACCTTGGGCACTGAATC	
HMBS-F	TGCACGGCAGCTTAACGAT	201
HMBS-R	AGGCAAGGCAGTCATCAAGG	

Inicialmente, foram realizados testes de concentração dos *primers* para determinar a concentração ótima de cada *primer* nas qPCRs; as concentrações testadas foram 0,2; 0,4; 0,6; 0,8 e 1 μ M. Nesta etapa deve-se escolher a menor concentração de *primer* que resulta no menor valor de *Ct* (*threshold cycle*) e em um sinal de fluorescência (ΔR_n) significativo, com pouca ou nenhuma formação de dímeros de *primers*.

Posteriormente, foram realizados testes de eficiência de amplificação da qPCR utilizando diluições logarítmicas seriadas de DNA genômico: 10^1 , 10^0 , 10^{-1} e 10^{-2} nM. A partir desse teste foram geradas curvas padrão e o valor de *slope* (coeficiente angular da reta) foi utilizado para o cálculo da eficiência de amplificação da qPCR para cada gene [Eficiência = $(10^{-1/\text{slope}} - 1) \times 100$] e também na normalização dos dados da qPCR.

Um controle negativo (DNA de um indivíduo sem a duplicação) foi utilizado em todas as corridas de qPCR. Todas as amostras foram analisadas em triplicata. As equações utilizadas para normalização dos dados de qPCR (KC_t) e para o cálculo da alteração no número de cópias alélicas (ΔKC_t) de cada gene foram obtidas a partir de Weksberg *et al.*^[12] e Hughes *et al.*^[13], e são mostradas logo abaixo (equação 1 e equação 2, respectivamente). De acordo com Weksberg *et al.*^[12] e Hughes *et al.*^[13], um valor de $\Delta KC_t = 0 \pm SD$ (SD: desvio padrão) indica nenhuma alteração no número de cópias alélicas e, portanto, nenhuma anormalidade genética; valores de $\Delta KC_t = -1 \pm SD$ indicam perda de uma cópia alélica (microdeleção) e finalmente, valores de $\Delta KC_t = 1 \pm SD$ indicam ganho de uma cópia alélica (microduplicação).

$$KC_t = \left(\frac{AC_{tR} - C_{tR}}{S_R} \right) \times S_T + C_{tT} \quad \text{(Equação 1)}$$

Onde:

- o KC_t : C_t normalizado do gene alvo.
- o AC_{tR} : Média dos valores de C_t para o gene de referência considerando todas as amostras incluídas em uma mesma corrida de qPCR.
- o C_{tR} : C_t do gene de referência para a amostra a ser normalizada (no caso de triplicata usa-se como C_{tR} a média dos três valores de C_t do gene de referência para a amostra a ser normalizada).
- o S_R : *Slope* da curva padrão para o gene de referência.
- o S_T : *Slope* da curva padrão para o gene alvo.
- o C_{tT} : C_t do gene alvo para a amostra a ser normalizada (no caso de triplicata usa-se como C_{tT} a média dos três valores de C_t do gene alvo para a amostra a ser normalizada).

$$\Delta KC_t = KC_{t \text{ (controle)}} - KC_{t \text{ (afetado)}} \quad \text{(Equação 2)}$$

Onde:

- o ΔKC_t : Representa o ganho ou a perda no número de cópias alélicas (*fold copy number change*), ou seja, indica se há ou não alteração no número de cópias de um gene alvo (microdeleção/microduplicação/normal).
- o $KC_{t \text{ (controle)}}$: C_t normalizado do gene alvo (KC_t do gene alvo) para a amostra correspondente ao controle negativo.
- o $KC_{t \text{ (afetado)}}$: C_t normalizado do gene alvo (KC_t do gene alvo) para a amostra correspondente ao paciente com suspeita clínica de uma determinada síndrome genética.

Sequenciamento do exoma e análise

O sequenciamento do exoma foi realizado em amostras de ambos os pacientes pelo “The Centre for Applied Genomics”, The Hospital for Sick Children (Toronto, Canadá) utilizando o kit de captura *Agilent SureSelect Human All Exon V4 Kit* (Agilent

Technologies), que tem sondas para capturar 20,965 genes e 334,378 exons, totalizando 51 Mb do genoma. O DNA capturado foi sequenciado na plataforma SOLiD 5500xl (Applied Biosystems). Foram sequenciadas aproximadamente 53 milhões de 35 bp x 75 bp *paired-end reads* em cada exoma. A cobertura média foi de 30x para o exoma da Paciente I e de 33x para o exoma do Paciente II, com cerca de 60% das bases sequenciadas cobertas por pelo menos 10x. Os dados foram alinhados ao genoma de referencia hg19/GRCh37 via BFAST and BWA. As variantes foram selecionadas quanto à sua qualidade utilizando o *Genome Analysis Toolkit* (GATK 1.1.28; <http://www.broadinstitute.org/gatk/>) e foram anotadas quanto ao seu efeito funcional pelo SnpEff 2.0.5 (<http://snpeff.sourceforge.net/>). O SnpEff também classifica as variantes com relação ao seu impacto (exemplo: impacto alto, moderado ou leve). A anotação das variantes contra banco de dados como *1000 Genomes* (April 2012 release; <http://browser.1000genomes.org>), *NHLBI Exome Sequencing Project* (ESP6500; <http://evs.gs.washington.edu/EVS/>) e *Single Nucleotide Polymorphism database* (dbSNP137. <http://www.ncbi.nlm.nih.gov/projects/SNP/>) foram feitas utilizando o *software* desenvolvido *in-house* chamado Mendel,MD (RGCCCL Cardenas and SDJ Pena, artigo sendo preparado). A predição dos *softwares* SIFT (*Sorting Intolerant From Tolerant*; <http://sift.jcvi.org/>) e Polyphen2 (*Polymorphism Phenotyping v2*; <http://genetics.bwh.harvard.edu/pph2/>) foram utilizadas para estimar se uma determinada troca de aminoácidos afetaria a função da proteína. Estes *softwares* analisam o impacto da variante considerando as alterações físico-químicas causadas pelas substituições de aminoácidos na proteína e avaliam também a conservação evolutiva da posição da variante^[14, 15].

Sequenciamento de Sanger

Para validar as variantes candidatas identificadas pela análise de exoma foi realizado o sequenciamento de Sanger utilizando o *BigDye Terminator v3.1 Cycle Sequencing Kit* (Applied Biosystems®) e o *Applied Biosystems (ABI) 3130 Genetic Analyzer*. A análise dos resultados foi realizada utilizando o programa *Sequencher* version 4.1.4 (Genes Code Corporation).

RESULTADOS PARCIAIS E DISCUSSÃO

A análise dos resultados de genotipagem dos pacientes mostrou uma duplicação parcial do gene *PDGFRA* em ambos os pacientes: a paciente I apresentou uma

duplicação em chr4:55138935-55162980 (24,05 Kb de extensão) e o paciente II apresentou uma duplicação em chr4:55140652-55153205 (12,55 Kb de extensão). Como mutações/amplificações no gene *PDGFRA* foram previamente relacionadas com gliomas no tegmento pontino^[16] e mutações no gene *PDGFRB* foram associadas com calcificação nos gânglios de base^[17], decidimos avaliar detalhadamente esta duplicação parcial. Segundo o banco de dados DGVbeta (*Database of Genomic Variants*) não foram descritas CNVs na região (Fig. 2). Foi realizada então a genotipagem dos pais da paciente I e eles não apresentaram a duplicação parcial (Fig. 3).

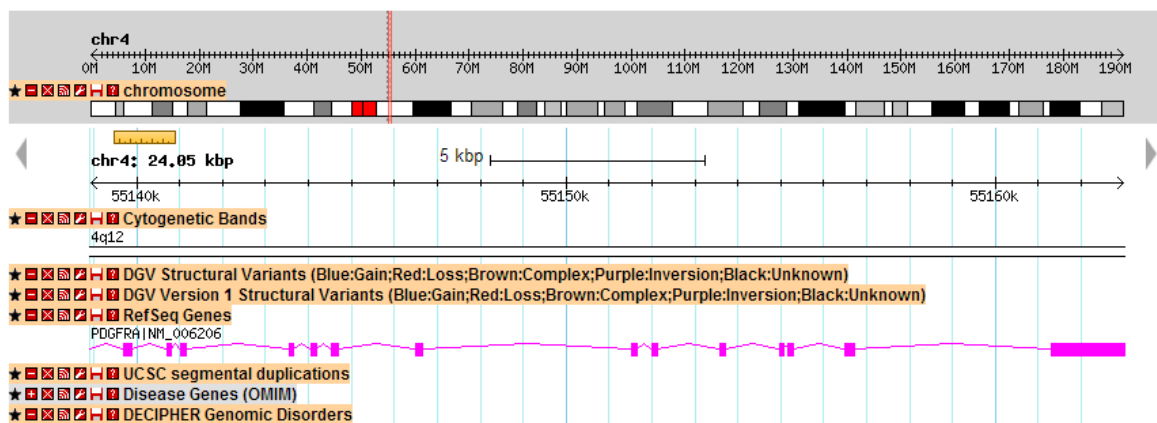


Fig. 2. Região duplicada na paciente I pela análise de genotipagem (chr4:55,138,935–55,162,980) visualizada no *Database of Genomic Variants*. Não foram descritas variantes estruturais na região. Note que a região inclui parte do gene *PDGFRA* (de rosa).

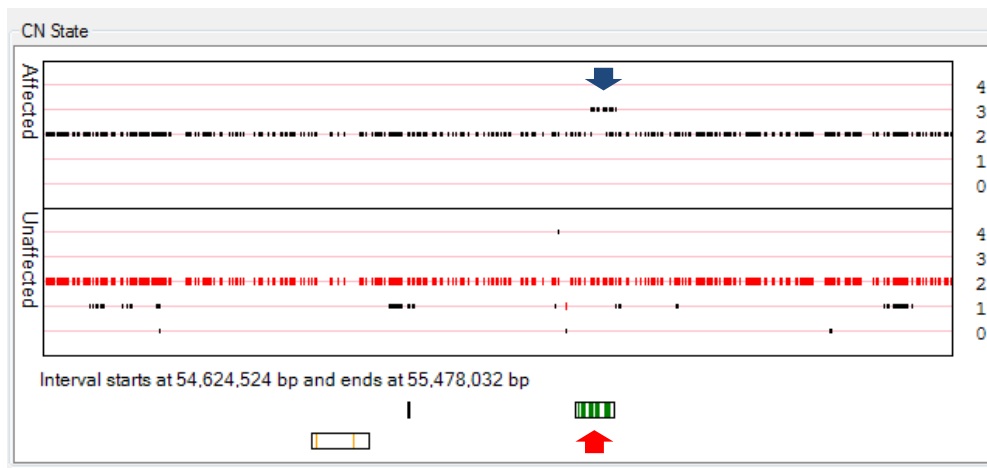


Fig. 3. Resultado da análise de genotipagem mostrando parte do cromossomo 4. Na parte superior da figura é mostrado o resultado dos pacientes analisados e na parte inferior (de vermelho) é mostrado o resultado da mãe da paciente I. A localização do gene *PDGFRA* está indicada pela seta vermelha. A seta azul indica a duplicação parcial do gene *PDGFRA* nos

pacientes. Note que a mãe da paciente não apresenta a duplicação. A análise foi realizada utilizando o programa CNViewer.

Realizamos então a PCR quantitativa em tempo real com amostras dos pacientes para confirmar a duplicação parcial. Primeiramente, verificamos por meio de experimentos de eletroforese em gel de agarose que o produto da PCR resultante da amplificação de *PDGFRA* com os *primers* customizados exclusivamente para esse trabalho era único e do tamanho esperado. Realizamos a análise da curva de dissociação nos ensaios de qPCR para verificar a especificidade das reações, a formação de dímeros de *primers* e determinar a temperatura de *melting* (T_m) de cada *amplicon*. Foi detectada a presença de um pico único e bem definido na curva de dissociação para cada gene, e não foi gerado sinal de fluorescência significativo no controle branco de amplificação, indicando que não houve formação de dímeros de *primers* e que as amplificações foram específicas. O teste de concentração de *primers* mostrou que a concentração ótima para todos os *primers* foi de 0,2 μ M. A partir do teste de eficiência de amplificação da qPCR, a quantidade de DNA a ser utilizada foi padronizada em 10,0 ng por amostra. A partir dos valores de C_t normalizados (KC_t), foi feito o cálculo da alteração no número de cópias alélicas (ΔKC_t) do gene alvo *PDGFRA*. A Tabela 2 mostra os valores obtidos. A análise dos resultados mostrou que os pacientes não apresentaram duplicação do gene *PDGFRA*, com valores de ΔKC_t próximos de zero, mostrando que esta variante estrutural é uma variante falso-positiva da análise de genotipagem.

Tabela 2. Valores obtidos pela análise de qPCR.

	HMBS (referência)	PDGFRA (alvo)		
Amostras	C_{tR}	C_{tT}	KC_t	ΔKC_t
Controle	21,9367	20,9167	21,2208	0,0000
Paciente I	21,9500	21,0467	21,3379	-0,1171
Paciente II	21,9600	21,0767	21,3581	-0,1374

Depois de excluídas as duplicações e deleções pela análise de genotipagem, realizamos então a análise dos exomas dos pacientes. Foram identificadas aproximadamente 80.000 variantes em cada paciente. As variantes foram filtradas de modo a incluir: variantes heterozigotas em ambos os pacientes; variantes que passaram nos filtros de qualidade do GATK; variantes com impacto moderado ou alto pelo SnpEff; variantes raras na população (frequência do alelo menos comum [MAF] $\leq$ 0.005 em todos os bancos de dados analisados); variantes com número de *reads* acima

de 10; e variantes preditas como patogênicas (pelo SIFT e Polyphen2). Como resultado, obtivemos variantes em 172 genes.

Destes genes foram selecionados quatro genes candidatos com base na sua patogenicidade/ impacto na proteína:

1) *CCBE1* (*collagen and calcium binding EGF domains 1*; OMIM *612753); impacto moderado pelo software SnpEff, são variantes preditas como deletérias pelos softwares SIFT e Polyphen2;

2) *ARL5B* (*ADP-ribosylation factor-like 5B*; OMIM *608909); impacto alto pelo software SnpEff: mutação do tipo *Splice site donor*;

3) *RSU1* (*Ras suppressor protein 1*; OMIM *179555); impacto alto pelo software SnpEff: mutação do tipo *Splice site acceptor*;

4) *CRLS1* (*cardiolipin synthase 1*; OMIM *608188); impacto alto pelo software SnpEff: mutação do tipo *Splice site donor*.

A tabela 3 mostra detalhes sobre cada uma das variantes nos genes candidatos. Segundo os resultados do exoma os pacientes eram heterozigotos compostos para mutações no gene *CCBE1*, sendo que ambos os pacientes apresentavam mutações heterozigotas no exon 8 posição chr18:57107022, a paciente I tinha uma mutação no exon 3 na posição chr18:57147424 e o paciente II tinha uma mutação no exon 8 posição chr18:57107007. Porém, quase todas estas variantes não foram confirmadas pelo sequenciamento de Sanger; eram erros de leitura do sequenciamento do exoma. A única variante confirmada foi a variante no exon 3 do gene *CCBE1* na paciente I. Consequentemente, estes quatro genes foram excluídos como candidatos.

Recentemente, realizamos nova análise dos exomas utilizando uma nova abordagem: procuramos por variantes em genes que causam epilepsia segundo o banco de dados de doenças genéticas humanas OMIM (*Online Mendelian Inheritance in Man*). As epilepsias são um conjunto diversificado de desordens neurológicas que envolvem crises recorrentes de convulsões, que são a manifestação de atividade elétrica anormal no sistema nervoso central, e representam uma das desordens neurológicas mais comuns, com incidência de até 3% na população geral^[18, 19]. Observamos que ambos os pacientes apresentavam variantes heterozigotas nos genes *SCN1A* e *CHD2*.

O gene *SCN1A* codifica a subunidade alfa do canal de sódio, essencial para a excitabilidade dos neurônios^[19]. De todos os genes envolvidos com epilepsia, *SCN1A* é atualmente o de maior relevância clínica, com o maior número de mutações relacionadas com a epilepsia: cerca de 900 variantes distintas já foram relatadas^[20, 21].

Variantes no gene *SCN1A* estão relacionadas às seguintes doenças no OMIM: *Dravet syndrome* ou *early infantile epileptic encephalopathy-6* (OMIM #607208); *Epilepsy, generalized, with febrile seizures plus, type 2* (OMIM #604403); *Febrile seizures, familial, 3A* (OMIM #604403); e *Migraine, familial hemiplegic, 3* (OMIM #609634).

A paciente I tem uma variante em chr2:166929897 e o paciente II tem uma variante em chr2:166895959 (a tabela 3 mostra os detalhes das variantes). A variante heterozigota na posição chr2:166929897 foi previamente relatada em uma paciente com Síndrome de Dravet clássica^[22]. A Síndrome de Dravet tem incidência de 1/4.000 nascimentos, e é caracterizada por crises de convulsões geralmente prolongadas com início no primeiro ano de vida, provocadas geralmente por febre, e os pacientes apresentam desenvolvimento cognitivo e motor normal antes do início das crises com desaceleração subsequente^[22, 23]. Mutações heterozigotas em *SCN1A* são encontradas em cerca de 75% dos casos de pacientes com síndrome de Dravet, sendo que 90% destas mutações são *de novo*^[24, 23].

Variantes heterozigotas no gene *CHD2* são relacionadas com *Epileptic encephalopathy, childhood-onset* (OMIM #615369). Ambos os pacientes têm uma variante que causa um códon finalizador prematuro na posição p.Leu470*. O RNAm produzido provavelmente será alvo do *nonsense mediated decay*, que é um processo celular que realiza a degradação de transcritos que contém códons finalizadores prematuros^[25]. A encefalopatia epiléptica com início na infância é uma forma grave de epilepsia caracterizada pelo aparecimento de vários tipos de crises nos primeiros anos de vida e os indivíduos afetados têm regressão cognitiva e deficiência intelectual^[26]. Mutações que geram códons finalizadores prematuros no gene *CHD2* (nas posições p.Trp1657* e p.Arg466*) foram descritas em pacientes com encefalopatia epiléptica e com características de síndrome de Dravet^[24].

Acreditávamos que as variantes nos genes *SCN1A* e *CHD2* seriam fortes candidatas, porém, estas variantes poderiam ser erros de leitura do sequenciamento do exoma, devido ao balanço alélico desequilibrado (Tabela 3). Realizamos então o sequenciamento de Sanger e verificamos que as variantes eram de fato erros de leitura do sequenciamento do exoma. Análises futuras serão então necessárias para identificar novos genes candidatos possivelmente relacionados com o fenótipo dos pacientes.

Tabela 3. Detalhes das variantes selecionadas através da análise de exoma dos pacientes I e II afetados com calcificação no cérebro

Cromossomo	18	18	18	10	10	20	2	2	15
Posição	57147424	57107022	57107007	18961636	16737155	5990559	166929897	166895959	93492213
Nome do gene	<i>CCBE1</i>	<i>CCBE1</i>	<i>CCBE1</i>	<i>ARL5B</i>	<i>RSU1</i>	<i>CRLS1</i>	<i>SCN1A</i>	<i>SCN1A</i>	<i>CHD2</i>
Localização	Exon 3	Exon 8	Exon 8	Intron 4	Intron 7	Intron 3	Exon 3a	Exon 16	Exon 13a
Referencia	C	G	C	T	C	G	G	T	T
Número de reads igual a referencia PACIENTE I	11	4	9	19	10	29	51	29	22
Alternativa PACIENTE I	T	C	A	A	T	T	A	A	A
Número de reads igual a alternativa PACIENTE I	9	7	5	7	12	11	13	0	6
Número de reads igual a referencia PACIENTE II	35	8	9	31	18	38	67	27	32
Alternativa PACIENTE II	T	C	A	A	T	T	A	A	A
Número de reads igual a alternativa PACIENTE II	0	12	8	10	9	10	2	12	9
Tipo de mutação	<i>Missense</i>	<i>Missense</i>	<i>Missense</i>	<i>Splice site donor</i>	<i>Splice site acceptor</i>	<i>Splice site donor</i>	<i>Missense</i>	<i>Missense</i>	<i>Stop gained</i>
Refseq	NM_133459.3	NM_133459.3	NM_133459.3	NM_178815.3	NM_012425.3	NM_019095.4	NM_001202435.1	NM_001202435.1	NM_001271.3
Mutação no DNA (HGVS nomenclatura_c.)	c.259C>T	c.802G>C	c.817C>A	c.339+2T>A	c.599-1C>T	c.444+1G>T	c.235G>A	c.2563T>A	c.1409T>A
Mutação na proteína (HGVS nomenclatura_p.)	p.Pro87Ser	p.Gly268Arg	p.Pro273Thr	-	-	-	p.Asp79Asn	p.Leu855Ile	p.Leu470*
Predição pelo SIFT	Deleterious	Deleterious	Deleterious	-	-	-	Deleterious	Deleterious	-
Predição pelo Polyphen2	Probably damaging	Probably damaging	Possibly damaging	-	-	-	Probably damaging	Possibly damaging	-
Predição pelo <i>Mutation Taster</i>	Disease causing	Disease causing	Disease causing	-	-	-	Disease causing	Disease causing	-
Confirmado por Sanger	Sim	Não	Não	Não	Não	Não	Não	Não	Não

REFERÊNCIAS

1. BABA, Y. *et al.* Heredofamilial brain calcinosis syndrome. *Mayo Clin Proc*, v. 80, n. 5, p. 641-51, 2005.
2. GOBBI, G. Coeliac disease, epilepsy and cerebral calcifications. *Brain Dev*, v. 27, n. 3, p. 189-200, 2005.
3. SAITO, Y. *et al.* Cerebellopontine calcification: a new entity of idiopathic intracranial calcification? *Acta Neuropathol*, v. 110, n. 1, p. 77-83, 2005.
4. LIVINGSTON, J. H. *et al.* Intracranial calcification in childhood: a review of aetiologies and recognizable phenotypes. *Dev Med Child Neurol*, v. 56, n. 7, p. 612-26, 2014.
5. RAJAB, A. *et al.* Recessive developmental delay, small stature, microcephaly and brain calcifications with locus on chromosome 2. *Am J Med Genet A*, v. 149A, n. 2, p. 129-37, 2009.
6. DUCKETT, S. *et al.* Presence of zinc, aluminum, magnesium in striopallidodentate (SPD) calcifications (Fahr's disease): electron probe study. *Acta Neuropathol*, v. 38, n. 1, p. 7-10, 1977.
7. KOZIK, M.; KULCZYCKI, J. Laser-spectrographic analysis of the cation content in Fahr's syndrome. *Arch Psychiatr Nervenkr*, v. 225, n. 2, p. 135-42, 1978.
8. STAMELOU, M. *et al.* Progressive ataxia and palatal tremor associated with dense pontine calcification: A unique case. *Mov Disord*, v. 28, n. 8, p. 1155-7, 2013.
9. LIVINGSTON, J. *et al.* COL4A1 mutations associated with a characteristic pattern of intracranial calcification. *Neuropediatrics*, v. 42, n. 6, p. 227-33, 2011.
10. MILLER, S. A.; DYKES, D. D.; POLESKY, H. F. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*, v. 16, n. 3, p. 1215, 1988.
11. CARR, I. M. *et al.* Rapid visualisation of microarray copy number data for the detection of structural variations linked to a disease phenotype. *PLoS One*, v. 7, n. 8, p. e43466, 2012.
12. WEKSBERG, R. *et al.* A method for accurate detection of genomic microdeletions using real-time quantitative PCR. *BMC Genomics*, v. 6, p. 180, 2005.
13. HUGHES, S. *et al.* *PCR: Methods Express*. Bloxham, UK: Scion Publishing Ltd, 2007. (Use of quantitative PCR for the detection of genomic microdeletions or microduplications).

14. KUMAR, P.; HENIKOFF, S.; NG, P. C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*, v. 4, n. 7, p. 1073-81, 2009.
15. ADZHUBEI, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nat Methods*, v. 7, n. 4, p. 248-9, 2010.
16. PUGET, S. *et al.* Mesenchymal transition and PDGFRA amplification/mutation are key distinct oncogenic events in pediatric diffuse intrinsic pontine gliomas. *PLoS One*, v. 7, n. 2, p. e30313, 2012.
17. SANCHEZ-CONTRERAS, M. *et al.* Genetic Screening and Functional Characterization of PDGFRB Mutations Associated With Basal Ganglia Calcification of Unknown Etiology. *Hum Mutat*, 2014.
18. HAUSER, W. A.; KURLAND, L. T. The epidemiology of epilepsy in Rochester, Minnesota, 1935 through 1967. *Epilepsia*, v. 16, n. 1, p. 1-66, 1975.
19. ESCAYG, A.; GOLDIN, A. L. Sodium channel SCN1A and epilepsy: mutations and mechanisms. *Epilepsia*, v. 51, n. 9, p. 1650-8, 2010.
20. MULLEY, J. C. *et al.* SCN1A mutations and epilepsy. *Hum Mutat*, v. 25, n. 6, p. 535-42, 2005.
21. PARIHAR, R.; GANESH, S. The SCN1A gene variants and epileptic encephalopathies. *J Hum Genet*, v. 58, n. 9, p. 573-80, 2013.
22. ZUBERI, S. M. *et al.* Genotype-phenotype associations in SCN1A-related epilepsies. *Neurology*, v. 76, n. 7, p. 594-600, 2011.
23. HELBIG, I. Dravet syndrome. Disponível em: http://www.orpha.net/consor/cgi-bin//OC_Exp.php?lng=EN&Expert=33069.
24. SULS, A. *et al.* De novo loss-of-function mutations in CHD2 cause a fever-sensitive myoclonic epileptic encephalopathy sharing features with Dravet syndrome. *Am J Hum Genet*, v. 93, n. 5, p. 967-75, 2013.
25. NAGY, E.; MAQUAT, L. E. A rule for termination-codon position within intron-containing genes: when nonsense affects RNA abundance. *Trends Biochem Sci*, v. 23, n. 6, p. 198-9, 1998.
26. CARVILL, G. L. *et al.* Targeted resequencing in epileptic encephalopathies identifies de novo mutations in CHD2 and SYNGAP1. *Nat Genet*, v. 45, n. 7, p. 825-30, 2013.

CAPÍTULO 4 – CONCLUSÕES E PERSPECTIVAS

O advento das tecnologias de sequenciamento de nova geração tem promovido avanços nos estudos de doenças humanas nas áreas de genoma, transcriptoma e epigenoma e interatoma^[1, 2]. O sequenciamento do exoma já provou ser um método com custo baixo, quando estratificados os benefícios, além de eficaz para detectar variantes causadoras de doenças humanas^[3, 4]. Entretanto, a análise dos exomas ainda é um grande desafio, principalmente em razão da falta de conhecimento sobre as funções e interações dos genes, de erros de sequenciamento, de “promiscuidade” nas relações genótipo-fenótipo e de dificuldades ao utilizar ferramentas bioinformáticas para filtrar os milhares de variantes encontradas^[5, 6, 7, 8].

Realizamos com sucesso duas aplicações clínicas diferentes para a análise de exomas e verificamos que a análise de exomas pode: (1) melhorar a qualidade de correlações genótipo-fenótipo de síndromes causadas por anormalidades cromossômicas, como no caso do paciente com microdeleção no cromossomo 1p13.2; (2) auxiliar na detecção do gene causador de doenças genéticas, como nos estudos das duas famílias com miofibromatose infantil e dos dois pacientes com calcificação no cérebro.

Concluimos que foram obtidos resultados promissores e inovadores nesse trabalho sendo que descrevemos novas variantes patogênicas humanas e interações na gênese de doenças. Na primeira família estudada, com dois filhos afetados com miofibromatose infantil, propusemos que mutações homozigotas no gene *NDRG4* podem ser as variantes causais para a forma autossômica recessiva da doença. Já na segunda família, identificamos que variantes nos genes *PDGFRB* e *PTPRG* podem conjuntamente causar a penetrância fenotípica completa nos dois irmãos com miofibromatose infantil.

Nesse estudo, também analisamos genes envolvidos em vias que contribuem para o fenótipo das doenças, como no caso dos pacientes com microdeleção em 1p13.2, em que identificamos que seus fenótipos apresentam muitas das características da síndrome de Noonan e podem estar causalmente relacionados com deleções de *NRAS*. Pretendemos ainda prosseguir com a análise dos exomas dos pacientes com calcificação no cérebro, com o objetivo de identificar as variantes causais do fenótipo dos pacientes.

Como perspectivas futuras, estamos aplicando a análise de exomas para uma terceira finalidade: estudar pacientes com herança genética complexa, multifatorial, com

componente genético provavelmente poligênico ou oligogênico. Um exemplo de herança oligogênica ocorre nas ciliopatias, ou seja, grupo de 15 desordens de disfunção nos cílios, que causam anormalidades de desenvolvimento e fenótipos degenerativos, hipoteticamente causados por combinações de mais de 50 *locis* (variantes deletérias com frequências <0,1%), interagindo com alelos modificadores e causais secundários^[9]. Outro exemplo é a doença de Crohn, sobre a qual foi demonstrado que mais de 30 variantes promovem suscetibilidade a essa doença inflamatória do trato gastrointestinal^[10].

São duas as hipóteses consideradas nos estudos genéticos de doenças complexas: doença comum-variante comum (DC-VC) ou doença comum-variante rara (DC-VR). A hipótese DC-VC postula que doenças complexas resultam de efeitos cumulativos de um grande número de variantes comuns com efeitos modestos. Já a hipótese DC-VR presume que múltiplas variantes raras com grandes impactos são as principais determinantes da herdabilidade das doenças complexas^[11]. A hipótese DC-VR está de acordo com o modelo “*clan genomics*” ou “*mutational burden*”, que prediz que diversas variantes raras e deletérias em conjunto causam ou agravam o fenótipo^[12].

Os estudos de associação genômica ampla (GWAS – *Genome-Wide Association Studies*), utilizando, por exemplo, o SNP *array*, testam principalmente a hipótese DC-VC, já que analisam variantes comuns^[11]. Em aproximadamente uma década, os GWAS revelaram associação de muitos SNPs com suscetibilidade a doenças^[13, 14, 15]. Ao contrário do que se esperava da metodologia, porém, os achados de análises de GWAS explicaram apenas uma proporção modesta de herdabilidade genética em casos de doenças complexas^[16]. Uma variedade de hipóteses foram propostas para explicar esse fato, incluindo variantes raras não testadas, e interações gene-gene e gene-ambiente^[16]. Suportando a hipótese de que variantes raras têm um papel nas doenças complexas, foi revelado que os alelos identificados através de GWAs não são tipicamente os alelos causadores, mas estão em desequilíbrio de ligação com as variantes causais^[11].

Atualmente, com o advento do sequenciamento de nova geração, é possível confirmar tal hipótese ao avaliar desde variantes comuns até variantes raras^[17]. Análises de exoma já se mostraram eficientes para a descoberta de genes de susceptibilidade em doenças complexas, mas desvendá-las é considerado ainda um desafio^[17, 18].

Para testar essa terceira aplicação, estamos estudando casos familiares com hepatite autoimune tipo I. Acreditamos que a hepatite autoimune tem componente genético provavelmente poligênico. A hepatite autoimune (HAI) é uma doença hepática

inflamatória crônica de causa desconhecida, que afeta um grupo de pacientes que perderam a tolerância imunológica a antígenos do fígado^[19]. A sua prevalência está entre 0,5 e 1/100.000^[20]. É caracterizada por elevação flutuante e crônica da atividade das transaminases séricas, hipergamaglobulinemia, presença de autoanticorpos e alterações histológicas^[20, 19]. Foram distinguidos dois tipos principais de HAI, de acordo com a presença de autoanticorpos: o tipo 1 é o mais comum e é caracterizado pela presença de anticorpos antimúsculo liso e/ou antinucleares (também conhecidos como fatores antinucleares – FAN) e pode suceder a qualquer idade, mas é mais frequente nos adultos; o tipo 2 é caracterizado pela presença de anticorpos anti-LKM1 (*anti-liver/kidney microsome type 1*) e ocorre quase exclusivamente nas crianças e apenas raramente após a puberdade^[21, 20, 19]. A reatividade de anti-LKM1 em HAI-2 é dirigida principalmente contra uma região do citocromo P450IID6 (CYP2D6)^[22]. Entre 10% e 15% dos casos de cirrose biliar e colangite esclerosante primária, estão associados à HAI^[20].

O mecanismo fisiopatológico conhecido na HAI consiste em uma resposta inflamatória com células de linfócitos T e linfócitos B, macrófagos e células *natural killer*. O fator desencadeante ou os fatores que estimulam essa resposta inflamatória ainda são desconhecidos^[19]. Apesar de a patogênese de HAI ser pouco conhecida, a predisposição genética tem sido associada principalmente aos genes do complexo principal de histocompatibilidade (MHC) classe II, mais especificamente para o locus de *Human Leukocyte Antigen* (HLA)-DR^[23]. Em pacientes caucasianos europeus e norte-americanos, HAI-1 está associada com a presença de alelos DRB1*0301, DRB3*0101 e DRB1*0401, ao passo que HAI-2 está associada com DRB1*0301 ou DRB1*07^[23].

No Brasil, o alelo de susceptibilidade primário para HAI-1 é DRB1*1301 (presente em mais de 2/3 dos pacientes); entretanto, uma associação secundária com DRB1*0301 também foi identificada; a predisposição a HAI-2 está associada com os alelos DRB1*07 ou DRB1*03^[24]. Já os haplótipos de HLA-DQ são variados em brasileiros^[25]. Foi observado que há aumento da frequência do alelo DQB1*06 em pacientes HAI-1 (68% vs 41% dos controles), em grande desequilíbrio de ligação com HLA-DRB1*13^[24, 25].

Além de demonstrar que uma doença de herança genética complexa, a hepatite autoimune, pode ser dissecada através de estudos exômicos de casos familiares, pretendemos comprovar, por meio de estudos de quantificação, expressão gênica e de

localização intracelular e pelo uso de camundongos transgênicos, que as variantes do gene *NDRG4* têm relação causal com a miofibromatose infantil autossômica recessiva.

Desse modo, as próximas etapas, essenciais para provar essa hipótese, envolvem estudos de quantificação dos produtos gênicos por *western blots*, estudos de localização intracelular por imunofluorescência que usam anticorpos anti-*NDRG4*, comercialmente disponíveis, e a extração de RNA e estudos de expressão gênica. Entretanto, o estudo que acreditamos ser o mais incisivo e mais promissor é a criação de camundongos transgênicos com a mesma variante p.Val171Leu encontrada nos nossos pacientes (o resíduo de valina é conservado na mesma posição do gene ortólogo de camundongos).

Finalmente, é importante reconhecer que os fenótipos clínicos dos pacientes são o resultado de variações e interações genéticas, epigenéticas, ambientais, metabólicas, entre outras. Assim, a informação incorporada no exoma deve ser analisada com sabedoria para cumprir a promessa de "*primum non nocere*"^[17].

REFERÊNCIAS

1. HAWKINS, R. D.; HON, G. C.; REN, B. Next-generation genomics: an integrative approach. *Nat Rev Genet*, v. 11, n. 7, p. 476-86, 2010.
2. XUAN, J. *et al.* Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett*, v. 340, n. 2, p. 284-95, 2013.
3. KU, C. S.; NAIDOO, N.; PAWITAN, Y. Revisiting Mendelian disorders through exome sequencing. *Hum Genet*, v. 129, n. 4, p. 351-70, 2011.
4. DAUBER, A. *et al.* Whole exome sequencing reveals a novel mutation in *CUL7* in a patient with an undiagnosed growth disorder. *J Pediatr*, v. 162, n. 1, p. 202-4 e1, 2013.
5. BAMSHAD, M. J. *et al.* Exome sequencing as a tool for Mendelian disease gene discovery. *Nat Rev Genet*, v. 12, n. 11, p. 745-55, 2011.
6. JACOB, H. J. Next-generation sequencing for clinical diagnostics. *N Engl J Med*, v. 369, n. 16, p. 1557-8, 2013.
7. LU, J. T.; CAMPEAU, P. M.; LEE, B. H. Genotype-phenotype correlation--promiscuity in the era of next-generation sequencing. *N Engl J Med*, v. 371, n. 7, p. 593-6, 2014.
8. WALL, J. D. *et al.* Estimating genotype error rates from high-coverage next-generation sequence data. *Genome Res*, 2014.

9. DAVIS, E. E.; KATSANIS, N. The ciliopathies: a transitional model into systems biology of human genetic disease. *Curr Opin Genet Dev*, v. 22, n. 3, p. 290-303, 2012.
10. BARRETT, J. C. *et al.* Genome-wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat Genet*, v. 40, n. 8, p. 955-62, 2008.
11. MARIAN, A. J. Molecular genetic studies of complex phenotypes. *Transl Res*, v. 159, n. 2, p. 64-79, 2012.
12. LUPSKI, J. R. *et al.* Clan genomics and the complex architecture of human disease. *Cell*, v. 147, n. 1, p. 32-43, 2011.
13. HINDORFF, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A*, v. 106, n. 23, p. 9362-7, 2009.
14. MANOLIO, T. A. Genomewide association studies and assessment of the risk of disease. *N Engl J Med*, v. 363, n. 2, p. 166-76, 2010.
15. VISSCHER, P. M. *et al.* Five years of GWAS discovery. *Am J Hum Genet*, v. 90, n. 1, p. 7-24, 2012.
16. STRANGER, B. E.; STAHL, E. A.; RAJ, T. Progress and promise of genome-wide association studies for human complex trait genetics. *Genetics*, v. 187, n. 2, p. 367-83, 2011.
17. MARIAN, A. J. Challenges in medical applications of whole exome/genome sequencing discoveries. *Trends Cardiovasc Med*, v. 22, n. 8, p. 219-23, 2012.
18. WU, L. *et al.* Case-only exome sequencing and complex disease susceptibility gene discovery: study design considerations. *J Med Genet*, 2014.
19. FERRI LIU, P. M. *et al.* Autoimmune hepatitis in childhood: the role of genetic and immune factors. *World J Gastroenterol*, v. 19, n. 28, p. 4455-63, 2013.
20. POUPON, R. Chronic autoimmune hepatitis. Disponível em: http://www.orpha.net/consor/cgi-bin/OC_Exp.php?Expert=2137.
21. CZAJA, A. J.; FREESE, D. K. Diagnosis and treatment of autoimmune hepatitis. *Hepatology*, v. 36, n. 2, p. 479-97, 2002.
22. MANNS, M. P. *et al.* LKM-1 autoantibodies recognize a short linear sequence in P450IID6, a cytochrome P-450 monooxygenase. *J Clin Invest*, v. 88, n. 4, p. 1370-8, 1991.
23. OLIVEIRA, L. C. *et al.* Autoimmune hepatitis, HLA and extended haplotypes. *Autoimmun Rev*, v. 10, n. 4, p. 189-93, 2011.

24. BITTENCOURT, P. L. *et al.* Genetic heterogeneity in susceptibility to autoimmune hepatitis types 1 and 2. *Am J Gastroenterol*, v. 94, n. 7, p. 1906-13, 1999.
25. GOLDBERG, A. C. *et al.* Analysis of HLA haplotypes in autoimmune hepatitis type 1: identifying the major susceptibility locus. *Hum Immunol*, v. 62, n. 2, p. 165-9, 2001.

ANEXOS

ANEXO 1

CARTA DE SUBMISSÃO DO ARTIGO 1

European Journal of Medical Genetics

Jul 4 em 10:54 AM

Para nataliaduli@yahoo.com.br

Ms. Ref. No.: EJMG-D-15-00276

Title: 1p13.2 deletion displays clinical features overlapping Noonan syndrome, likely related to NRAS gene haploinsufficiency
European Journal of Medical Genetics

Dear Mrs. Natália Linhares,

Your manuscript

"1p13.2 deletion displays clinical features overlapping Noonan syndrome, likely related to NRAS gene haploinsufficiency"

has been received by European Journal of Medical Genetics and will be evaluated for possible publication.

Your manuscript was assigned nr EJMG-D-15-00276. Please refer to this number for any inquiries or questions you may have.

You may check on the progress of your paper by logging on to the Elsevier Editorial System as an author. The URL

is <http://ees.elsevier.com/ejmg/>.

Your username is: nataliaduli@yahoo.com.br

If you need to retrieve password details, please go to: http://ees.elsevier.com/ejmg/automail_query.asp

European Journal of Medical Genetics is trying very hard to provide a rapid review process. If you should not have heard of your manuscript within 8 weeks after receipt of this email, please contact the Editorial office.

For further assistance, please visit our customer support site at <http://help.elsevier.com/app/answers/list/p/7923>. Here you can search for solutions on a range of topics, find answers to frequently asked questions and learn more about EES via interactive tutorials. You will also find our 24/7 support contact details should you need any further assistance from one of our customer support representatives.

Yours sincerely,

Pr Alain VERLOES, M.D., Ph.D.
Editor-in-Chief

Center for Human Genetics
Herestraat 49
B-3000 Leuven
Belgium

Tel : 016/34 59 03
Fax : 016/34 60 51
email : ejmg@uz.kuleuven.ac.be

ANEXO 2

ARTIGO 1 - Supplementary Tables

Supplementary Table 1. Hemizygous variants within the deleted region that passed in GATK quality filters.

Pos	Variant_ID	Ref	Alt	Qual	Read depth	Gene	Genomes1k MAF	dbSNP MAF	ESP MAF	SIFT	Impact	Func_class	Consequence
115231459	rs386500536	C	T	516.14	16	AMPD1	0.000998403	0.002296					intron_variant
114438951	rs1217401	A	G	1548.19	44	AP4B1	0.376997	0.3466	0.437952	0.07	MODERATE	MISSENSE	missense_variant
114424601	rs1217381	C	T	2843.61	78	BCL2L15	0.999401	0.9986	0.000154	0.63	MODERATE	MISSENSE	missense_variant
114423653	rs1217382	C	T	793.61	29	BCL2L15	0.401158	0.3691	0.456198				3_prime_UTR_variant
115128087	rs386561521	A	G	377.98	11	DENND2C	0.828275	0.7989					3_prime_UTR_variant
115144703	rs386480540	A	G	1210.74	33	DENND2C	0.998403	0.9972	0.007535				intron_variant
115168097	rs12136548	T	C	1069.38	29	DENND2C	0.125599	0.163	0.221821	0.73	MODERATE	MISSENSE	missense_variant
112269872	rs197434	G	A	3156.61	90	FAM212B	0.860623	0.8476	0.14255		LOW	SILENT	synonymous_variant
113657092	rs1216793	A	C	1495.13	40	LRIG2	0.901757	0.9118	0.096878		LOW	SILENT	synonymous_variant
113658887	rs1216791	G	A	962.27	28	LRIG2	0.789936	0.7934	0.23543				intron_variant
114191994	rs1743594	G	A	817.72	23	MAGI3	0.338658	0.3242	0.410503				intron_variant
114092306	rs10858004	A	G	1179.21	31	MAGI3	0.16254	0.1543					intron_variant
114161116	rs1217200	C	A	122.01	5	MAGI3	0.671925	0.6396					intron_variant
114267347	rs1230640	C	T	1441.05	40	PHTF1	0.703474	0.6662	0.194679				intron_variant
114240319	rs3835263	T	TA	1107.46	29	PHTF1	0.678115		0.214936				3_prime_UTR_variant
114249382	rs6698586	C	T	2900.02	78	PHTF1	0.679313	0.6488	0.221086				intron_variant
114281438	.	C	A	104.77	8	PHTF1							intron_variant
113254648	rs78366259	A	G	987.75	28	PPM1J	0.0109824	0.01286	0.021298	0.19	MODERATE	MISSENSE	missense_variant
113255389	rs386481312	C	T	804.48	22	PPM1J	0.967851	0.9706	0.039136	1.0	MODERATE	MISSENSE	missense_variant
114377568	rs2476601	A	G	1492.69	45	PTPN22	0.972644	0.9578	0.068353	1.0	MODERATE	MISSENSE	missense_variant
114377093	rs386573918	A	G	797.01	22	PTPN22	0.701278	0.663	0.197324				intron_variant

Pos	Variant_ID	Ref	Alt	Qual	Read depth	Gene	Genomes1k MAF	dbSNP MAF	ESP MAF	SIFT	Impact	Func_class	Consequence
112106444	rs1591712	G	A	3174.29	99	RAP1A	0.894768	0.9132	0.105172				intron_variant
112251983	rs3738300	C	G	3001.04	78	RAP1A	0.322484	0.3251					intron_variant
115316998	.	A	G	186.53	9	SIKE1							splice_region_variant
113456546	rs1049434	A	T	4655.27	122	SLC16A1	0.676717	0.6575	0.333769	1.0	MODERATE	MISSENSE	missense_variant
113098534	rs6658555	C	T	1723.01	48	ST7L	0.1248	0.1593	0.204444	0.07	MODERATE	MISSENSE	missense_variant
113083439	rs10745330	C	T	635.67	19	ST7L	0.489816	0.5303					intron_variant
115398044	rs60659641	GC	G	1068.49	0	SYCP1	0.9998	0.9844	0.007352				intron_variant
114680540	rs17032441	C	T	3158.91	85	SYT6	0.207668	0.2112	0.203983		LOW	SILENT	synonymous_variant
114948281	rs6537825	A	G	4024.46	106	TRIM33	0.85643	0.8384	0.057666	1.0	MODERATE	MISSENSE	missense_variant
114940632	rs5777192	CA	C	241.46	3	TRIM33							intron_variant
115576023	rs386515121	A	G	947.38	27	TSHB	0.983626	0.9853	0.024758	0.58	MODERATE	MISSENSE	missense_variant
115631924	rs4240540	C	T	136.87	5	TSPAN2	0.973043	0.95					intron_variant
115601615	rs7523818	A	G	1152.42	30	TSPAN2	0.965655	0.955	0.065893				intron_variant
113063125	rs910697	A	G	4607.74	119	WNT2B	0.483427	0.5239	0.47578		LOW	SILENT	synonymous_variant

Supplementary Table 2. Genes deleted in the patient according to the NCBI Map Viewer Annotation Release 105.

start	stop	Symbol	<u>Q</u>	Link	Cyto	Description
112025970	112106602	<u>ADORA3</u>	-	<u>OMIM</u>	1p13.2	adenosine A3 receptor
112141629	112150946	<u>RP5-836N10.1</u>	-		1p13.2	uncharacterized LOC100129269
112162405	112256807	<u>RAP1A</u>	+	<u>OMIM</u>	1p13.3	RAP1A, member of RAS oncogene family
112190850	112192219	<u>KRT18P57</u>	-		1p13.2	keratin 18 pseudogene 57
112264686	112298419	<u>FAM212B</u>	-		1p13.2	family with sequence similarity 212, member B

start	stop	Symbol	<u>Q</u>	Link	Cyto	Description
112282463	112290420	<u>FAM212B-AS1</u>	+			FAM212B antisense RNA 1
112287939	112298158	<u>LOC101928718</u>	-			uncharacterized LOC101928718
112298190	112310199	<u>DDX20</u>	+	<u>OMIM</u>	1p21.1-p13.2	DEAD (Asp-Glu-Ala-Asp) box polypeptide 20
112318444	112532147	<u>KCND3</u>	-	<u>OMIM</u>	1p13.3	potassium voltage-gated channel, Shal-related subfamily, member 3
112532593	112541464	<u>LOC643355</u>	+		1p13.2	uncharacterized LOC643355
112715822	112720236	<u>LOC101928754</u>	-			uncharacterized LOC101928754
112905821	112906283	<u>LOC171419</u>	+		1q12	thioredoxin pseudogene
112938800	113003786	<u>CTTNBP2NL</u>	+	<u>OMIM</u>	1p13.2	CTTNBP2 N-terminal like
113004392	113004455	<u>MIR4256</u>	-			microRNA 4256
113010040	113063910	<u>WNT2B</u>	+	<u>OMIM</u>	1p13	wingless-type MMTV integration site family, member 2B
113066140	113162040	<u>ST7L</u>	-		1p13.2	suppression of tumorigenicity 7 like
113162075	113214241	<u>CAPZA1</u>	+	<u>OMIM</u>	1p13.2	capping protein (actin filament) muscle Z-line, alpha 1
113168494	113168847	<u>MRPL53P1</u>	-		1p13.2	mitochondrial ribosomal protein L53 pseudogene 1
113177351	113177403	<u>RNU7-70P</u>	-		1p13.2	RNA, U7 small nuclear 70 pseudogene
113217048	113243368	<u>MOV10</u>	+	<u>OMIM</u>	1p13.2	Mov10, Moloney leukemia virus 10, homolog (mouse)
113243749	113250025	<u>RHOC</u>	-	<u>OMIM</u>	1p13.1	ras homolog family member C
113252616	113257950	<u>PPM1J</u>	-	<u>OMIM</u>	1p13.2	protein phosphatase, Mg ²⁺ /Mn ²⁺ dependent, 1J
113261619	113269857	<u>FAM19A3</u>	+		1p13.2	family with sequence similarity 19 (chemokine (C-C motif)-like), member A3
113290717	113291097	<u>LOC128322</u>	+		1p13.2	nuclear transport factor 2-like
113362791	113393265	<u>RP11-426L16.8</u>	-		1p13.2	uncharacterized LOC100996702

start	stop	Symbol	<u>Q</u>	Link	Cyto	Description
113433214	113433593	<u>RPL39P8</u>	-		1p13.2	ribosomal protein L39 pseudogene 8
113454469	113498975	<u>SLC16A1</u>	-	<u>OMIM</u>	1p12	solute carrier family 16 (monocarboxylate transporter), member 1
113465972	113467295	<u>AKR7A2P1</u>	+		1p12	aldo-keto reductase family 7, member A2 pseudogene 1
113499037	113506694	<u>SLC16A1-AS1</u>	+		1p13.2	SLC16A1 antisense RNA 1
113554309	113615724	<u>RP11-31F15.2</u>	-			uncharacterized LOC100996251
113615792	113667824	<u>LRIG2</u>	+	<u>OMIM</u>	1p13.1	leucine-rich repeats and immunoglobulin-like domains 2
113668114	113669577	<u>RLIMP2</u>	-			ring finger protein, LIM domain interacting pseudogene 2
113711611	113712051	<u>RPS19P2</u>	+		1p13.2	ribosomal protein S19 pseudogene 2
113739403	113748875	<u>LOC643441</u>	-		1p13.2	uncharacterized LOC643441
113741429	113741913	<u>RPS15P1</u>	+		1p13.2	ribosomal protein S15 pseudogene 1
113933087	114228545	<u>MAGI3</u>	+		1p12-p11.2	membrane associated guanylate kinase, WW and PDZ domain containing 3
113992328	113993307	<u>LOC100421402</u>	+			heterogeneous nuclear ribonucleoprotein A1 pseudogene
114119373	114119516	<u>MTND5P20</u>	-			MT-ND5 pseudogene 20
114239824	114302165	<u>PHTF1</u>	-	<u>OMIM</u>	1p13	putative homeodomain transcription factor 1
114241506	114242228	<u>RPS2P14</u>	-		1p13	ribosomal protein S2 pseudogene 14
114304454	114355070	<u>RSBN1</u>	-		1p13.2	round spermatid basic protein 1
114354977	114361461	<u>RP5-1073O3.2</u>	+			uncharacterized LOC101928822
114356433	114414375	<u>PTPN22</u>	-	<u>OMIM</u>	1p13.2	protein tyrosine phosphatase, non-receptor type 22 (lymphoid)
114399257	114443859	<u>AP4B1-AS1</u>	+		1p13.2	AP4B1 antisense RNA 1
114419436	114430230	<u>BCL2L15</u>	-		1p13.2	BCL2-like 15

start	stop	Symbol	<u>Q</u>	Link	Cyto	Description
114437370	114447741	<u>AP4B1</u>	-	<u>OMIM</u>	1p13.2	adaptor-related protein complex 4, beta 1 subunit
114447915	114456708	<u>DCLRE1B</u>	+	<u>OMIM</u>	1p13.2	DNA cross-link repair 1B
114466623	114471880	<u>RP5-1073O3.7</u>	-			uncharacterized LOC101928846
114471819	114520491	<u>HIPK1</u>	+	<u>OMIM</u>	1p13.2	homeodomain interacting protein kinase 1
114522030	114524875	<u>OLFML3</u>	+	<u>OMIM</u>	1p13.2	olfactomedin-like 3
114544055	114545965	<u>RPL13AP10</u>	-		1p13.2	ribosomal protein L13a pseudogene 10
114631914	114696472	<u>SYT6</u>	-	<u>OMIM</u>	1p13.2	OTTHUMP00000195668
114821624	114821908	<u>MRP63P1</u>	-		1p13.1	mitochondrial ribosomal protein 63 pseudogene 1
114896608	114897781	<u>LOC100421116</u>	+			trafficking protein, kinesin binding 2 pseudogene
114935399	115053781	<u>TRIM33</u>	-	<u>OMIM</u>	1p13.1	tripartite motif containing 33
114944063	114953561	<u>LOC101928890</u>	+			uncharacterized LOC101928890
115002456	115002802	<u>RPL26P10</u>	+		1p13.1	ribosomal protein L26 pseudogene 10
115010740	115012266	<u>EIF2S2P5</u>	+		1p13.1	eukaryotic translation initiation factor 2, subunit 2 beta pseudogene 5
115078592	115080077	<u>PKMP1</u>	+		1p13.2	pyruvate kinase, muscle pseudogene 1
115110178	115124265	<u>BCAS2</u>	-	<u>OMIM</u>	1p13.2	breast carcinoma amplified sequence 2
115127155	115212732	<u>DENND2C</u>	-		1p13.2	DENN/MADD domain containing 2C
115215719	115238239	<u>AMPD1</u>	-	<u>OMIM</u>	1p13	adenosine monophosphate deaminase 1
115247085	115259515	<u>NRAS</u>	-	<u>OMIM</u>	1p13.2	neuroblastoma RAS viral (v-ras) oncogene homolog
115259534	115300671	<u>CSDE1</u>	-	<u>OMIM</u>	1p22	cold shock domain containing E1, RNA-binding
115312100	115323308	<u>SIKE1</u>	-	<u>OMIM</u>	1p13.2	suppressor of IKBKE 1
115377077	115394165	<u>NR1H5P</u>	+		1p13.2	nuclear receptor subfamily 1, group H, member 5, pseudogene

start	stop	Symbol	<u>Q</u>	Link	Cyto	Description
115397455	115537991	<u>SYCP1</u>	+	<u>OMIM</u>	1p13-p12	synaptonemal complex protein 1
115572445	115576930	<u>TSHB</u>	+	<u>OMIM</u>	1p13	thyroid stimulating hormone, beta
115590632	115632121	<u>TSPAN2</u>	-	<u>OMIM</u>	1p13.2	tetraspanin 2

Supplementary Table 3. Deleted genes in our patient associated with Mendelian disease phenotypes in OMIM

Gene name	Full gene name	Gene MIM ID	Inheritance	OMIM phenotype(s)
<i>KCND3</i>	Potassium voltage-gated channel, Shal-related subfamily, member 3	*605411	dom	Spinocerebellar ataxia 19, #607346
<i>SLC16A1</i>	Solute carrier family 16 (monocarboxylic acid transporters), member	*600682	dom	Erythrocyte lactate transporter defect, #245340; Hyperinsulinemic hypoglycemia, familial, 7, #610021
<i>LRIG2</i>	Leucine-Rich Repeats- and Immunoglobulin-like Domains-Containing Protein 2	*608869	rec	Urofacial syndrome 2, #615112
<i>AP4B1</i>	Adaptor-related protein complex 4, beta-1 subunit	*607245	rec	Spastic paraplegia 47, autosomal recessive, #614066
<i>TRIM33</i>	Tripartite motif-containing protein 33	*605769	dom	Thyroid carcinoma, papillary, #188550
<i>AMPD1</i>	Adenosine monophosphate deaminase-1, muscle	*102770	rec	Myopathy due to myoadenylate deaminase deficiency, #615511
<i>NRAS</i>	Neuroblastoma RAS viral (v-ras) oncogene homolog	*164790	dom	Colorectal cancer, #114500; Thyroid carcinoma, follicular, #188470; Autoimmune lymphoproliferative syndrome type IV, #614470; Noonan syndrome 6, #613224; Nevus, epidermal #162900
<i>TSHB</i>	Thyroid-stimulating hormone, beta polypeptide	*188540	rec	Hypothyroidism, congenital, nongoitrous 4, #275100

ANEXO 3

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(para os pais ou responsável legal autorizando a participação dos menores de 18 anos e/ou dos participantes legalmente incapazes)

Pesquisa: Aplicação clínica do sequenciamento e análise bioinformática de exomas

Coordenador: Prof. Sérgio Danilo Junho Pena – Tel: (031) 3284-8000 – Horário disponível: de 9:00h às 18:00h. E-mail: spena@dcc.ufmg.br

Pesquisadora: Natália Duarte Linhares – Tel: (031) 3409-8053 - Horário disponível: de 9:00h às 18:00h. E-mail: nataliaduarte@ufmg.br

Comitê de Ética em Pesquisa (COEP) da UFMG: Av. Antônio Carlos, 6627 - Unidade Administrativa II, 2º andar, sala 2005, CEP: 31270-901-BH/MG - Tel: (031) 3409-4592, Horário de atendimento: de 9:00h às 11:00h e de 14:00h às 16:00h. E-mail: coep@prpq.ufmg.br. COEP é o órgão da UFMG que visa proteger o bem-estar dos indivíduos participantes em pesquisas realizadas na Universidade.

Prezado Senhor (a),

Seu filho (a) está sendo convidado (a) para participar de uma pesquisa que tem como objetivo a realização de estudos genéticos, ou seja, estudos dos genes, de pacientes com doenças raras. Em alguns casos, serão realizados também estudos genéticos dos pais com o objetivo de realizar a comparação dos resultados para descobrir o que há de diferente nos filhos. Os genes são pedaços de DNA, que são herdados dos pais e que são responsáveis por armazenar toda a informação necessária para o desenvolvimento e funcionamento apropriado do nosso corpo. Caso o gene tenha uma alteração danosa, ele já não poderá realizar adequadamente a sua função. Nesta pesquisa nós estamos buscando localizar possíveis alterações danosas em genes dos pacientes que podem estar relacionadas com as doenças estudadas. Esta pesquisa contribuirá para o melhor entendimento destas doenças.

Será realizada a análise dos dados do prontuário do seu filho (a) no Hospital das Clínicas da UFMG ou no Laboratório GENE – Núcleo de Genética Médica. Além disto, serão coletadas amostras de sangue do seu filho (a). Iremos utilizar os dados e as amostras de sangue somente para esta pesquisa. Parte do DNA extraído da amostra de sangue coletada do seu filho (a) será enviada para o exterior (Canadá) para a realização de exame genético. As amostras de sangue e de DNA serão descartadas segundo as normas da Vigilância Sanitária após a realização dos exames genéticos desta pesquisa.

Se concordar em participar, permitindo que os dados do prontuário de seu filho (a) sejam analisados e que seja feita a coleta de 5 mL de sangue do mesmo (a), pedimos que assine esse papel dizendo que entendeu as explicações e está concordando. Se necessário, caso não haja obtenção de DNA suficiente para a análise, poderá ser solicitada nova coleta de 5 mL de sangue.

Os riscos envolvidos na participação desta pesquisa são os riscos envolvendo coletas de sangue: a coleta será feita em uma veia do braço do seu filho (a), podendo causar uma leve dor ou incômodo na hora do procedimento e uma pequena mancha roxa no local que desaparecerá em até 3 a 4 dias. Os estudos poderão produzir um resultado útil, mas também poderão não ser informativos. Além disto, os estudos podem demorar aproximadamente dois anos para sua realização, devido às análises genéticas, que são muito complexas. Você não terá nenhum gasto ou ganho financeiro por participar da pesquisa. As consultas e os exames necessários para esta pesquisa serão realizados juntamente com as consultas de rotina do seu filho (a) no Hospital das Clínicas da UFMG ou no Laboratório GENE. Porém, fica garantido ao participante da pesquisa e seu acompanhante o ressarcimento de despesas decorrentes da participação no estudo, tais como transporte e alimentação, se for necessária sua presença para consultas ou exames adicionais às consultas de rotina. Ademais, fica garantido ao participante da

pesquisa assistência, acompanhamento e indenização em caso de danos diretos e indiretos, imediatos ou tardios, decorrentes de sua participação no estudo.

Se o Sr. (a) deseja mais esclarecimentos ou se tem qualquer dúvida sobre a pesquisa nós estamos dispostos a responder qualquer pergunta. Se o Sr. (a) não quiser que seu filho (a) participe ou se quiser desistir em qualquer momento, isso não vai gerar em nenhum prejuízo de qualquer natureza para a sua pessoa ou familiares. Este Termo de Consentimento foi elaborado em duas vias igualmente válidas, que devem ser rubricadas em todas as suas páginas e assinadas, ao seu término, sendo que uma via ficará conosco e a outra com o Sr. (a).

Fica garantido aos participantes deste trabalho o sigilo e a proteção de sua identidade a qualquer momento ao longo do estudo e também após a sua conclusão. A garantia de confidencialidade de dados e respeito à privacidade do participante será assegurada para qualquer apresentação pública de resultados, como em revistas científicas, congressos, simpósios, eventos de outra natureza e meios de comunicação. Após a sua autorização, poderão ser fotografadas as partes do corpo do seu filho (a) que mostrarem as características da sua doença. Nestas fotografias serão utilizadas tarjas pretas sobre os olhos do seu filho (a), exceto nos casos de pacientes onde as características faciais são componentes essenciais ao diagnóstico clínico. Havendo necessidade de fazer fotografias do rosto sem as tarjas pretas, os pais (ou responsáveis legais) serão avisados e poderão decidir se autorizam ou não a realização das fotografias. As fotos dos pacientes envolvidos na pesquisa serão tratadas de forma a minimizar ao máximo, na medida do possível, a possibilidade de identificação dos mesmos.

Os dados genéticos resultantes da pesquisa associados a um indivíduo identificável, bem como informações do prontuário, serão acessíveis somente aos pesquisadores envolvidos e não será permitido o acesso a terceiros, garantindo proteção contra qualquer tipo de discriminação e/ou estigmatização (rotular negativamente). O Sr. (a) tem opção de tomar ou não conhecimento dos resultados desta pesquisa. Ao final da pesquisa nos comprometemos em lhe informar sobre os resultados dos exames do seu filho (a), caso seja de seu interesse.

Durante esta pesquisa analisaremos somente os genes provavelmente relacionados com a doença do paciente em estudo. Porém, é possível que novas doenças genéticas diferentes daquela objeto do estudo sejam identificadas. No caso de novos diagnósticos de doenças genéticas, fica garantido aos participantes aconselhamento genético e acompanhamento clínico, com a indicação dos responsáveis, sem custos.

Eu, _____ autorizo a utilização e publicação dos dados clínicos e laboratoriais analisados e obtidos nesta pesquisa, além das fotografias do meu filho (a) _____

_____, em revistas científicas, livros, pôsteres, apresentações em slides e na tese de doutorado na qual este projeto está vinculado, sem a identificação do seu nome. Estou ciente de que não seremos remunerados pelo uso destas fotografias. Entendo que o artigo científico publicado com as fotos poderá ficar disponível na internet. Declaro que li e entendi o texto acima e concordo em participar deste estudo.

Belo Horizonte, ____ de _____ de _____.

Assinatura dos pais ou responsável legal: _____

Assinatura do coordenador: _____

Assinatura do Pesquisador Responsável/ pessoa por ele delegada: _____

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(para os pais)

Pesquisa: Aplicação clínica do sequenciamento e análise bioinformática de exomas

Coordenador: Prof. Sérgio Danilo Junho Pena – Tel: (031) 3284-8000 – Horário disponível: de 9:00h às 18:00h. E-mail: spena@dcc.ufmg.br

Pesquisadora: Natália Duarte Linhares – Tel: (031) 3409-8053 - Horário disponível: de 9:00h às 18:00h. E-mail: nataliaduarte@ufmg.br

Comitê de Ética em Pesquisa (COEP) da UFMG: Av. Antônio Carlos, 6627 - Unidade Administrativa II, 2º andar, sala 2005, CEP: 31270-901-BH/MG - Tel: (031) 3409-4592, Horário de atendimento: de 9:00h às 11:00h e de 14:00h às 16:00h. E-mail: coep@prpq.ufmg.br. COEP é o órgão da UFMG que visa proteger o bem-estar dos indivíduos participantes em pesquisas realizadas na Universidade.

Prezado Senhor (a),

Você está sendo convidado (a) para participar de uma pesquisa que tem como objetivo a realização de estudos genéticos, ou seja, estudos dos genes, de pacientes com doenças raras. Em alguns casos, serão realizados também estudos genéticos dos pais com o objetivo de realizar a comparação dos resultados para descobrir o que há de diferente nos filhos. Os genes são pedaços de DNA, que são herdados dos pais e que são responsáveis por armazenar toda a informação necessária para o desenvolvimento e funcionamento apropriado do nosso corpo. Caso o gene tenha uma alteração danosa, ele já não poderá realizar adequadamente a sua função. Nesta pesquisa nós estamos buscando localizar possíveis alterações danosas em genes dos pacientes que podem estar relacionadas com as doenças estudadas. Esta pesquisa contribuirá para o melhor entendimento destas doenças.

Serão coletadas amostras de seu sangue. Iremos utilizar estas amostras somente para esta pesquisa. Parte do DNA extraído da amostra do seu sangue será enviada para o exterior (Canadá) para a realização de exame genético. As amostras de sangue e de DNA serão descartadas segundo as normas da Vigilância Sanitária após a realização dos exames genéticos desta pesquisa.

Se concordar em participar, permitindo que seja feita a coleta de 5 mL de seu sangue, pedimos que assine esse papel dizendo que entendeu as explicações e está concordando. Se necessário, caso não haja obtenção de DNA suficiente para a análise, poderá ser solicitada nova coleta de 5 mL de sangue.

Os riscos envolvidos na participação desta pesquisa são os riscos envolvendo coletas de sangue: a coleta será feita em uma veia do seu braço, podendo causar uma leve dor ou incômodo na hora do procedimento e uma pequena mancha roxa no local que desaparecerá em até 3 a 4 dias. Os estudos poderão produzir um resultado útil, mas também poderão não ser informativos. Além disto, os estudos podem demorar aproximadamente dois anos para sua realização, devido às análises genéticas, que são muito complexas. Você não terá nenhum gasto ou ganho financeiro por participar da pesquisa. As consultas e os exames necessários para esta pesquisa serão realizados juntamente com as consultas de rotina do seu filho (a) no Hospital das Clínicas da UFMG ou no Laboratório GENE. Porém, fica garantido ao participante da pesquisa e seu acompanhante o ressarcimento de despesas decorrentes da participação no estudo, tais como transporte e alimentação, se for necessária sua presença para consultas ou exames adicionais às consultas de rotina. Ademais, fica garantido ao participante da pesquisa assistência, acompanhamento e indenização em caso de danos diretos e indiretos, imediatos ou tardios, decorrentes de sua participação no estudo.

Se o Sr. (a) deseja mais esclarecimentos ou se tem qualquer dúvida sobre a pesquisa nós estamos dispostos a responder qualquer pergunta. Se o Sr. (a) não quiser participar ou se quiser desistir em qualquer momento, isso não vai gerar nenhum prejuízo de qualquer natureza para a sua pessoa ou familiares. Este Termo de Consentimento foi elaborado em duas vias igualmente válidas, que devem ser rubricadas em todas as suas páginas e assinadas, ao seu término, sendo que uma via ficará conosco e a outra com o Sr. (a).

Fica garantido aos participantes deste trabalho o sigilo e a proteção de sua identidade a qualquer momento ao longo do estudo e também após a sua conclusão. A garantia de confidencialidade de dados e respeito à privacidade do participante será assegurada para qualquer apresentação pública de resultados, como em revistas científicas, congressos, simpósios, eventos de outra natureza e meios de comunicação.

Os dados genéticos resultantes da pesquisa associados a um indivíduo identificável serão acessíveis somente aos pesquisadores envolvidos e não será permitido o acesso a terceiros, garantindo proteção contra qualquer tipo de discriminação e ou estigmatização (rotular negativamente). O Sr. (a) tem opção de tomar ou não conhecimento dos resultados desta pesquisa. Ao final da pesquisa nos comprometemos em lhe informar sobre os seus resultados, caso seja de seu interesse.

Durante esta pesquisa analisaremos somente os genes provavelmente relacionados com a doença do paciente em estudo. Porém, é possível que novas doenças genéticas diferentes daquela objeto do estudo sejam identificadas. No caso de novos diagnósticos de doenças genéticas, fica garantido aos participantes aconselhamento genético e acompanhamento clínico, com a indicação dos responsáveis, sem custos.

Eu, _____
autorizo a utilização e publicação dos dados clínicos e laboratoriais analisados e obtidos nesta pesquisa em revistas científicas, livros, pôsteres, apresentações em slides e na tese de doutorado na qual este projeto está vinculado, sem a identificação do meu nome. Entendo que o artigo científico poderá ficar disponível na internet. Declaro que li e entendi o texto acima e concordo em participar deste estudo.

Belo Horizonte, ____ de _____ de _____.

Assinatura: _____

Assinatura do coordenador: _____

Assinatura do Pesquisador Responsável/ pessoa por ele delegada: _____

ANEXO 4

TERMO DE ASSENTIMENTO

(para participante com desenvolvimento intelectual correspondente à idade entre 13 e 18 anos)

Pesquisa: Aplicação clínica do sequenciamento e análise bioinformática de exomas

Coordenador: Prof. Sergio Danilo Junho Pena – Tel: (031) 3284-8000 – Horário disponível: de 9:00h às 18:00h. E-mail: spena@dcc.ufmg.br

Pesquisadora: Natália Duarte Linhares – Tel: (031) 3409-8053 - Horário disponível: de 9:00h às 18:00h. E-mail: nataliaduarte@ufmg.br

Comitê de Ética em Pesquisa (COEP) da UFMG: Av. Antônio Carlos, 6627 - Unidade Administrativa II, 2º andar, sala 2005, CEP: 31270-901-BH/MG - Tel: (031) 3409-4592, Horário de atendimento: de 9:00h às 11:00h e de 14:00h às 16:00h. E-mail: coep@prpq.ufmg.br. O Comitê de Ética em Pesquisa da UFMG é formado por um grupo de pessoas que têm como objetivo proteger os participantes de pesquisas realizadas na Universidade.

Você está sendo convidado (a) para participar de uma pesquisa. Seus pais permitiram que você participasse. Esta pesquisa tem como objetivo a realização de estudos genéticos, ou seja, estudos dos genes, de pacientes com doenças raras. Os genes são pedaços de DNA, que são herdados dos pais e que são responsáveis por armazenar toda a informação necessária para o desenvolvimento e funcionamento apropriado do nosso corpo. Caso o gene tenha uma alteração danosa, ele já não poderá realizar adequadamente a sua função. Nesta pesquisa nós estamos buscando localizar possíveis alterações danosas em genes dos pacientes que podem estar relacionadas com as doenças estudadas. Esta pesquisa contribuirá para o melhor entendimento destas doenças.

Será realizada a análise dos dados do seu prontuário (registro de todos os seus exames e internações) no Hospital das Clínicas da UFMG ou no Laboratório GENE – Núcleo de Genética Médica. Além disto, serão coletadas amostras de seu sangue. Iremos extrair (retirar) DNA destas suas amostras de sangue.

Se concordar em participar, permitindo que os dados do seu prontuário sejam analisados e que seja feita a coleta de um pouco do seu sangue, pedimos que assine esse papel dizendo que entendeu as explicações e está concordando. Se necessário, caso não haja obtenção de DNA suficiente para a análise, poderá ser solicitada nova coleta de sangue.

Os riscos envolvidos na participação desta pesquisa são os riscos envolvendo coletas de sangue: a coleta será feita em uma veia do seu braço, podendo causar uma leve dor ou incômodo na hora do procedimento e uma pequena mancha roxa no local que desaparecerá em até 3 a 4 dias. Os estudos poderão produzir um resultado útil, mas também poderão não ser informativos. Além disto, os estudos podem demorar aproximadamente dois anos para sua realização, devido às análises genéticas, que são muito complexas. Você e sua família não ganharão ou perderão dinheiro por participar da pesquisa.

Se você deseja mais esclarecimentos ou se tem qualquer dúvida sobre a pesquisa nós estamos dispostos a responder qualquer pergunta. Se você não quiser participar ou se quiser desistir em qualquer momento, isso não vai gerar nenhum prejuízo de qualquer natureza para a sua pessoa ou familiares. Este Termo de Assentimento foi elaborado em duas vias igualmente válidas, que devem ser assinadas em todas as suas páginas, sendo que uma ficará conosco e a outra com você.

Garantimos que faremos segredo com relação a sua identidade a qualquer momento ao longo do estudo e também após a sua conclusão. Após a autorização de seus pais/ responsáveis, poderão ser fotografadas as partes do seu corpo que mostrarem as características da sua doença. Nestas fotografias serão utilizadas tarjas pretas sobre os seus olhos (retângulos pretos), exceto nos casos de pacientes onde as características do seu rosto sejam importantes para a identificação da doença. Se houver a necessidade de fazer fotografias do seu rosto sem as tarjas pretas, os seus pais (ou responsáveis legais) serão avisados e poderão decidir se autorizam ou não a realização das fotografias. As suas fotos serão tratadas de forma a minimizar ao máximo a possibilidade de sua identificação.

Eu, _____
autorizo a utilização e publicação dos dados clínicos e laboratoriais analisados e obtidos nesta pesquisa, além das minhas fotografias, em revistas científicas, livros, pôsteres, apresentações em slides e na tese de doutorado na qual este projeto está vinculado, sem a identificação do meu nome. Entendo que o artigo científico publicado com as fotos poderá ficar disponível na internet. Declaro que li e entendi o texto acima e concordo em participar deste estudo.

Belo Horizonte, ____ de _____ de _____.

Assinatura: _____

Assinatura do coordenador: _____

Assinatura do Pesquisador Responsável/ pessoa por ele delegada: _____

TERMO DE ASSENTIMENTO

(para participante com desenvolvimento intelectual correspondente à idade entre 10 e 12 anos)

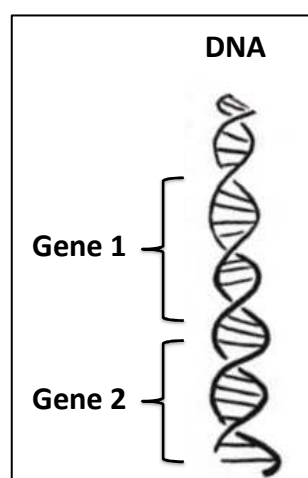
Pesquisa: Aplicação clínica do sequenciamento e análise bioinformática de exomas

Coordenador: Prof. Sergio Danilo Junho Pena – Tel: (031) 3284-8000 – Horário disponível: de 9:00h às 18:00h. E-mail: spena@dcc.ufmg.br

Pesquisadora: Natália Duarte Linhares – Tel: (031) 3409-8053 - Horário disponível: de 9:00h às 18:00h. E-mail: nataliaduarte@ufmg.br

Comitê de Ética em Pesquisa da UFMG: Av. Antônio Carlos, 6627 - Unidade Administrativa II, 2º andar, sala 2005, CEP: 31270-901-BH/MG - Tel: (031) 3409-4592, Horário de atendimento: de 9:00h às 11:00h e de 14:00h às 16:00h. E-mail: coep@prpq.ufmg.br. O Comitê de Ética em Pesquisa da UFMG é formado por um grupo de pessoas que têm como objetivo proteger os participantes de pesquisas realizadas na Universidade.

Você está sendo convidado (a) para participar de uma pesquisa. Seus pais permitiram que você participasse. Nesta pesquisa faremos estudos genéticos de pacientes com doenças raras, ou seja, doenças que acontecem com poucas pessoas. Estudos genéticos são estudos dos genes. Os genes são pedaços de DNA. O DNA fica guardado dentro de todas as partes do nosso corpo e tem o formato de uma escada torcida (veja figura ao lado). Ele é muito pequeno e para você enxergá-lo precisaria de um microscópio.



O DNA é responsável por guardar toda a informação necessária para a formação e funcionamento do nosso corpo. Todos nós recebemos (herdamos) dos nossos pais, no momento da nossa formação na barriga de nossa mãe, uma combinação única de genes. Esta combinação de genes é o que faz as diferenças entre uma pessoa e outra, como características de aparência e algumas doenças... Veja na figura abaixo* que o menino recebeu (herdou) de sua mãe as características de cabelo e recebeu do seu pai o formato do nariz.



Caso o gene tenha uma alteração danosa, ele já não poderá realizar corretamente a sua função. Nesta pesquisa nós estamos buscando localizar possíveis alterações danosas em genes dos pacientes que podem estar relacionadas com as doenças estudadas. Esta pesquisa contribuirá para o melhor entendimento destas doenças.

*A figura utilizada aqui foi retirada do folheto

“Testes genéticos para efeitos de saúde” - Conselho da Europa, 2012.

Nesta pesquisa será realizada a análise dos dados do seu prontuário (registro de todos os seus exames e internações) no Hospital das Clínicas da UFMG ou no Laboratório GENE – Núcleo de Genética Médica. Além disto, serão coletadas amostras (um pouco) do seu sangue. Iremos retirar o DNA destas amostras de sangue.

Se concordar em participar, permitindo que os dados do seu prontuário sejam analisados e que seja feita a coleta de um pouco do seu sangue, pedimos que assine esse papel dizendo que entendeu as explicações e está concordando. Se necessário, caso não haja obtenção de DNA suficiente para a análise, poderá ser solicitada nova coleta de sangue.

Os riscos envolvidos na participação desta pesquisa são os riscos envolvendo coletas de sangue: a coleta será feita em uma veia do seu braço, podendo causar uma leve dor na hora do procedimento e uma pequena mancha roxa no local que desaparecerá em até 3 a 4 dias. Além disto, os estudos podem demorar aproximadamente dois anos para sua realização, devido às análises genéticas, que são muito difíceis.

Se você tem qualquer dúvida sobre a pesquisa, nós podemos responder qualquer pergunta. Se você não quiser participar ou se quiser desistir em qualquer momento, isso não vai gerar nenhum problema para você ou sua família. Este Termo de Assentimento foi elaborado em duas cópias igualmente válidas, que devem ser assinadas em todas as suas páginas, sendo que uma ficará conosco e a outra com você.

Garantimos que faremos segredo com relação a sua identidade a qualquer momento ao longo do estudo e também após a sua conclusão. Após a autorização de seus pais/ responsáveis, poderão ser fotografadas as partes do seu corpo que mostrarem as características da sua doença. Nestas fotografias serão utilizadas tarjas pretas sobre os seus olhos (retângulos pretos), exceto nos casos de pacientes onde as características do seu rosto sejam importantes para a identificação da doença. Se existir a necessidade de fazer fotografias do seu rosto sem as tarjas pretas, os seus pais (ou responsáveis legais) serão avisados e poderão decidir se autorizam ou não a realização das fotografias. As suas fotos serão tratadas de forma a minimizar ao máximo a possibilidade de sua identificação.

Eu, _____
autorizo a utilização e publicação dos dados e exames analisados e obtidos nesta pesquisa, além das minhas fotografias, sem a identificação do meu nome. Entendo que o artigo publicado com as fotos poderá ficar disponível na internet. Declaro que li e entendi o texto acima e concordo em participar deste estudo.

Belo Horizonte, ____ de _____ de _____.

Assinatura: _____

Assinatura do coordenador: _____

Assinatura do Pesquisador Responsável/ pessoa por ele delegada: _____

TERMO DE ASSENTIMENTO

(para participante com desenvolvimento intelectual correspondente à idade entre 6 e 9 anos)

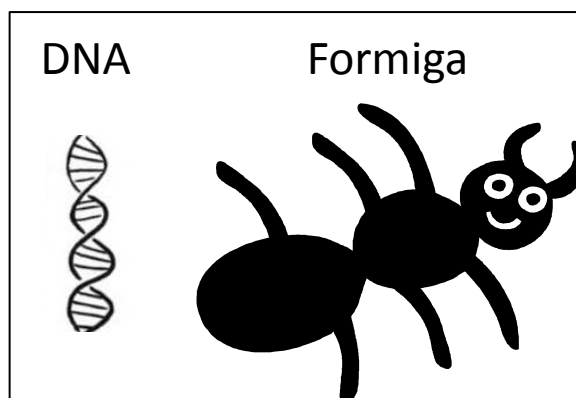
Pesquisa: Aplicação clínica do sequenciamento e análise bioinformática de exomas

Coordenador: Prof. Sergio Danilo Junho Pena – Tel: (031) 3284-8000 – Horário disponível: de 9:00h às 18:00h. E-mail: spena@dcc.ufmg.br

Pesquisadora: Natália Duarte Linhares – Tel: (031) 3409-8053 - Horário disponível: de 9:00h às 18:00h. E-mail: nataliaduarte@ufmg.br

Comitê de Ética em Pesquisa da UFMG: Av. Antônio Carlos, 6627 - Unidade Administrativa II, 2º andar, sala 2005, CEP: 31270-901-BH/MG - Tel: (031) 3409-4592, Horário de atendimento: de 9:00h às 11:00h e de 14:00h às 16:00h. E-mail: coep@prpq.ufmg.br. O Comitê de Ética em Pesquisa da UFMG é formado por um grupo de pessoas que têm como objetivo proteger os participantes de pesquisas realizadas na Universidade.

Você está sendo convidado (a) para participar de uma pesquisa. Nesta pesquisa faremos estudos de pacientes com doenças raras, ou seja, doenças que acontecem com poucas pessoas. Iremos estudar o seu DNA, que é uma coisa que fica guardada em todas as partes do seu corpo e que parece uma escada. O DNA é tão pequeno que você não consegue enxergá-lo. Ele é menor que uma formiga (veja a figura ao lado).



Quando você foi formado (a) na barriga da sua mãe você recebeu uma mistura de DNA do seu pai e da sua mãe. No DNA estão “escritas” as características de uma pessoa e também está “escrito” sobre algumas doenças que ela tem ou pode ter. É por causa desta mistura no DNA que as pessoas são diferentes.

Vea na figura abaixo* que o menino recebeu da sua mãe as características de cabelo e recebeu do seu pai o formato do nariz.



Nesta pesquisa nós estamos buscando localizar possíveis alterações no DNA dos pacientes que podem estar relacionadas com as doenças estudadas. Esta pesquisa irá ajudar a entendermos melhor estas doenças.

*A figura utilizada aqui foi retirada do folheto “Testes genéticos para efeitos de saúde” - Conselho da Europa, 2012.

Para esta pesquisa iremos analisar os dados de seus exames médicos no Hospital das Clínicas da UFMG ou no Laboratório GENE – Núcleo de Genética Médica e iremos coletar um pouco do seu sangue. Se você quiser participar, pedimos que escreva seu nome nesse papel dizendo que entendeu as explicações e está concordando. A coleta de sangue será feita no seu braço, podendo causar um pouco de dor na hora da coleta e uma pequena mancha roxa no local que desaparecerá em até 3 a 4 dias.

Se você tiver qualquer dúvida sobre a pesquisa, nós podemos responder qualquer pergunta. Se você não quiser participar ou se quiser desistir em qualquer hora, isso não vai gerar nenhum problema para você ou sua família.

Ninguém saberá que você está participando da pesquisa, não falaremos a outras pessoas, nem daremos a estranhos as informações que você nos der. Os resultados da pesquisa vão ser publicados, mas sem identificar as pessoas que participaram da pesquisa. Após a autorização de seus pais/ responsáveis, poderemos fotografar as partes do seu corpo que mostrarem as características da sua doença. Nestas fotografias serão utilizados retângulos pretos sobre os seus olhos, a não ser nos casos de pacientes onde as características do seu rosto sejam importantes para a identificação da doença. Se existir a necessidade de fazer fotografias do seu rosto sem os retângulos pretos, os seus pais (ou responsáveis legais) serão avisados e poderão decidir se autorizam ou não a realização das fotografias. As suas fotos serão modificadas, para que, na medida do possível, ninguém saiba que é você na foto.

Eu, _____
aceito participar da pesquisa. Li o texto acima e entendi as coisas ruins e as coisas boas que podem acontecer.

Belo Horizonte, _____ de _____ de _____.

Assinatura: _____

Assinatura do coordenador: _____

Assinatura do Pesquisador Responsável/ pessoa por ele delegada: _____
