Guilherme Lopes Yamamoto

**Aplicabilidade clínica da técnica de
sequenciamento de nova geração com enfoque
em displasias esqueléticas**

Tese apresentada à Faculdade de Medicina da
Universidade de São Paulo para obtenção do
título de Doutor em Ciências

Programa de Pediatria

Orientadora: Dra. Débora Romeo Bertola

São Paulo

2017

Dados Internacionais de Catalogação na Publicação (CIP)

Preparada pela Biblioteca da
Faculdade de Medicina da Universidade de São Paulo

©reprodução autorizada pelo autor

Yamamoto, Guilherme Lopes

Aplicabilidade clínica da técnica de sequenciamento de nova geração com
ênfase em displasias esqueléticas / Guilherme Lopes Yamamoto -- São Paulo,
2017.

Tese(doutorado)--Faculdade de Medicina da Universidade de São Paulo.
Programa de Pediatria.

Orientadora: Débora Romeo Bertola.

Descritores: 1.Análise de sequência de DNA 2.Sequenciamento de nucleotídeos
em larga escala 3.Genética 4.Genética médica 5.Doenças genéticas inatas 6.Doenças
ósseas/genética 7.Triagem neonatal

USP/FM/DBD-326/17

Dedicatória

Dedico esta tese aos pacientes e suas famílias, que no Brasil se doam genuinamente aos trabalhos de pesquisa por simples altruísmo.

Agradecimentos

1. À minha esposa, sem a qual sentido não existiria, por amar me aguentar e tantas mais coisas.
 2. Aos meus pais, sem os quais eu não existiria, por me amar e me aguentar há mais de 35 anos.
 3. AChong, Débora e Lilian, sem as quais a genética seria um caminho não trilhado, por me acolherem e aguentaremnesses últimos 8 anos.
 4. Aos médicos e residentes da Unidade de Genética do ICr-HC-FMUSP, sem os quais esta tese não teria casuística, por me desafiarem e me aguentarem às segundas, quartas e quintas feiras.
 5. Às mulheres do serviço de sequenciamento do CEGH-CEL, Kátia, Kelly, Maria Rita, Meire, Mônica, Monize, Naila e Vanessa, sem as quais esta tese não teria casuística, métodos, resultados, discussão ou anexos, por fazerem do laboratório uma segunda casa e me aguentarem às terças e sextas feiras.
 6. Ao Michel, sem o qual as pedras seriam maiores e mais frequentes, por ser co-primeiro autor em muito mais que um artigo e me aguentar nas horas vagas.
- Último. À Debora Romeo Bertola, sem a qual esta tese não existiria, por ser lastro sem ser âncora, ser norte sem ser cabresto e por me aguentar mais que os agradecidos 3 a 6 somados durante quase todos os dias dos últimos 8 anos, inclusive nas férias, me estimulando, desafiando e estabilizando, por mais que eu teime em discordar.

Agradeço também à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por financiar a pesquisa relacionada a esta tese.

Sumário

Lista de Siglas.....	vii
Lista de Abreviações	vii
Lista de Figuras	viii
Lista de Quadros	viii
Resumo.....	ix
Summary	xi
1. Introdução.....	1
1.1. A técnica de NGS.....	3
1.1.1. Preparo das bibliotecas.....	4
1.1.2. Método de sequenciamento NGS.....	6
1.1.3. Análise de bioinformática	8
1.2. Desafio diagnóstico em doenças genéticas.....	10
1.3. Aplicação da técnica de NGS no contexto de pesquisa	11
1.4. Aplicação da técnica de NGS no contexto clínico.....	13
1.5. Grupos de doenças monogênicas e aplicação do NGS	15
1.5.1. Doenças ósseas genéticas.....	16
1.5.2. RASopatias	21
1.6. Contexto atual.....	23
1.7. Triagem neonatal	24
2. Objetivos.....	28
2.1. Objetivo principal.....	28
2.2. Objetivos gerais.....	28
2.3. Objetivos específicos.....	28
3. Casuística e Métodos	30
3.1. Casuísticas.....	30
3.1.1. Casuística das etapas A e B	30
3.1.2. Casuística da etapa C	32
3.2. Técnicas Moleculares.....	34

3.2.1.	Extração de DNA.....	34
3.2.2.	Sequenciamento Sanger.....	34
3.2.3.	Sequenciamento de nova geração (NGS)	35
3.3.	Análise de bioinformática	38
3.4.	Interpretação das variantes	40
3.5.	Escolha do teste genético.....	44
3.6.	Relato de achados incidentais/secundários.....	44
3.7.	Análise estatística	45
4.	Resultados.....	46
4.1.	Validação da técnica de NGS (etapa A).....	46
4.2.	Aplicação da técnica de NGS em doenças ósseas genéticas e RASopatias (etapa B).....	50
4.2.1.	Doenças ósseas genéticas.....	50
4.2.2.	RASopatias	63
4.2.3.	Relato de achados incidentais	67
4.3.	Aplicação da técnica de NGS em amostras sem diagnóstico (etapa C).....	67
4.3.1.	Análise de amostras sem conhecimento da hipótese diagnóstica clínica (simulação de Triagem Neonatal).....	67
5.	Discussão	77
5.1.	Validação da técnica de NGS (etapa A).....	77
5.2.	Aplicação da técnica de NGS em doenças ósseas genéticas e RASopatias (etapa B).....	84
5.3.	Aplicação da técnica de NGS em amostras sem diagnóstico (etapa C).....	95
5.4.	Novas técnicas de sequenciamento e opções de testes genéticos.....	100
6.	Conclusões.....	103
7.	Anexos	105
8.	Referências	116

Lista de Siglas

ACMG – Colégio Americano de Genética Médica e Genômica (American College of Medical Genetics and Genomics)

CEGH-CEL – Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo

CNV - Variação no número de cópias (*Copy Number Variation*)

EUA – Estados Unidos da América

MS - Ministério da saúde

NGS – Sequenciamento de nova geração (Next Generation Sequencing)

OCM - Osteólise Carpotarsal multicêntrica

OMIM - Banco de dados *Online Mendelian Inheritance in Men*

SED - Displasia espondiloepifisária (*Spondyloepiphyseal Dysplasia*)

SMD - Displasia espondiloepifisária (*Spondylometaphyseal Dysplasia*)

SMD-CRD - Displasia espondiloepifisária com distrofia de cones e bastonetes (*Spondylometaphyseal Dysplasia with Cone-Rod Dystrophy*)

SMRT-Seq - Sequenciamento de moléculas únicas em tempo real (*Single Molecule Real Time Sequencing*)

SNV - Variante de nucleotídeo único ou de ponto (*Single Nucleotide Variant*)

WES – Sequenciamento do exoma completo (Whole-exome sequencing)

WGS – Sequenciamento do genoma completo (Whole-genome sequencing)

Lista de Abreviações

Vs. – *Versus*

Comp. - Composto

Hemi. – Hemizigoto

Het. – Heterozigoto

Homo. – Homozigoto

Lista de Figuras

Figura 1: Comparação entre as técnicas de sequenciamento Sanger e sequenciamento de nova geração (NGS).....	10
Figura 2: Via RAS/MAPK e respectivas doenças decorrentes de mutações germinativas associadas	22
Figura 3: Casuística das etapas A, B e C	33
Figura 4: Identificação de variantes nos genes <i>SOS2</i> e <i>LZTR1</i> como etiologia de síndrome de Noonan em múltiplas famílias da casuística de RASopatias	66
Figura 5: Região visualizada no IGV da única variante falso positiva identificada por NGS não confirmada por Sanger.....	80

Lista de Quadros

Quadro 1: Número de fenótipos com base molecular conhecida no OMIM, de acordo com os seus padrões de herança (adaptado de Koboldt et al., 2013 e atualizado em julho de 2017)....	13
Quadro 2: Genes relacionados a doenças ósseas genéticas que não tinham etiologia conhecida na nosologia de 2015, incluindo aqueles descritos previamente e não incorporados na nosologia (Bonafe et al., 2015).....	19
Quadro 3: Critérios maiores para Triagem Neonatal.....	26
Quadro 4: Critérios de patogenicidade ou benignidade de variantes, classificação da ACMG (Richards et al., 2015)	42
Quadro 5: Classificação de variantes segundo os critérios* da ACMG 2015**	43
Quadro 6: Lista de variantes conhecidas por Sanger e avaliadas por NGS	47
Quadro 7: Lista de variantes identificadas por NGS e avaliadas por Sanger.	48
Quadro 8: Variantes e hipóteses diagnósticas dos probandos da casuística de doenças ósseas genéticas investigados por Sanger.	51
Quadro 9: Variantes, hipóteses diagnósticas e grupo da nosologia dos probandos da casuística de doenças ósseas genéticas investigados por Painel NGS.....	55
Quadro 10: Variantes e hipóteses diagnósticas dos probandos da casuística de doenças ósseas genéticas investigados por exoma completo (WES).	61
Quadro 11: Variantes raras em homozigose identificadas por WES nas famílias com displasia espondilometafisária com distrofia de cones e bastonetes.....	62
Quadro 12: Variantes e hipótese dos probandos de RASopatias investigados por WES.	64
Quadro 13: Hipótese clínica, hipótese do teste cego e variantes encontradas por painel NGS em amostras analisadas sem conhecimento da hipótese diagnóstica clínica.	69
Quadro 14: Resumo dos achados positivos e negativos nas 45 amostras "casos" e 45 amostras "controles" de simulação de triagem neonatal.	76
Quadro 15: Critérios propostos no presente estudo para qualidade mínima de variantes detectadas por NGS para que possam ser consideradas de qualidade equivalente a padrão-ouro e prescindir de confirmação por outro método de sequenciamento.	84
Quadro 16: Sensibilidade diagnóstica da técnica de NGS em diferentes estudos.....	88

Resumo

Yamamoto GL. Aplicabilidade clínica da técnica de sequenciamento de nova geração com enfoque em displasias esqueléticas [Tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2017.

INTRODUÇÃO: Na última década surgiu uma nova técnica, o sequenciamento de nova geração, que, contrário ao método tradicional de Sanger, permite o sequenciamento em paralelo e em larga escala de múltiplos genes, ou até mesmo todos os genes humanos, a menor custo e com uma análise mais acelerada. Essa técnica possibilitou a descoberta de novos genes responsáveis por diversas doenças mendelianas, sendo rapidamente incorporada no contexto clínico. **OBJETIVOS:** comparar os resultados das técnicas de Sanger e sequenciamento de nova geração em amostras controle; introduzir a técnica de sequenciamento de nova geração no contexto clínico nas casuísticas de doenças ósseas genéticas e RASopatias; avaliar a sensibilidade diagnóstica desta técnica em amostras sem dados clínicos fornecidos. **MÉTODOS:** o sequenciamento de nova geração (sob a forma de um painel de genes customizado ou do exoma) foi realizado em amostras com mutações identificadas previamente por Sanger e em dois grupos de doenças mendelianas, 144 pacientes com doenças ósseas e 79 com RASopatias, além de 90 amostras sem dados clínicos conhecidos (45 casos e 45 controles). A técnica de Sanger foi aplicada em 29 amostras de doenças ósseas e em 81 amostras para confirmação de variantes identificadas pelo sequenciamento de nova geração. **RESULTADOS:** A sensibilidade da técnica de sequenciamento de nova geração foi estimada em 95,92% e a especificidade em 98,77%. Na casuística de doenças ósseas, a sensibilidade diagnóstica das amostras sequenciadas por Sanger foi de 69% (20/29), por painel customizado, 60% (75/125) e por exoma, 63% (12/19). Na casuística de RASopatias, a sensibilidade diagnóstica através do exoma foi de 46% (36/79). Como resultado deste trabalho, dois genes novos associados a RASopatias (*LZTR1* e *SOS2*) e um associado a uma displasia esquelética (*PCYT1A*) foram identificados. Na análise das amostras sem conhecimento prévio da hipótese clínica foi obtida uma sensibilidade diagnóstica de 46,67% (21/45), mas que chegou a 73,08% (14/26) para as hipóteses de erros inatos do metabolismo.

CONCLUSÕES: Foi demonstrado que a sensibilidade e a especificidade da técnica do sequenciamento de nova geração são altas e correspondentes a valores encontrados por outros grupos na literatura. Essa técnica foi considerada apropriada não apenas no contexto de pesquisa, demonstrado aqui pela descoberta de três novos genes associados a doenças mendelianas, como também para análises clínicas. Neste estudo a técnica foi aplicada com sucesso no contexto clínico, seja pelo painel customizado, seja pelo exoma, com uma positividade semelhante à encontrada pela técnica de Sanger. Mesmo na análise de amostras sem história clínica prévia foi possível identificar variantes patogênicas em quase metade dos casos, e numa porcentagem ainda maior quando a doença era um erro inato do metabolismo. Essa sensibilidade é comparável à obtida pela espectroscopia de massas em Tandem aplicada à triagem de múltiplas condições simultaneamente, o que sugere que a técnica do sequenciamento de nova geração poderá ser incorporada ao programa de triagem neonatal no futuro, ampliando o emprego de testes genéticos em complementaridade aos testes bioquímicos tradicionais.

Descritores: análise de sequência de DNA; sequenciamento de nucleotídeos em larga escala; genética; genética médica; doenças genéticas inatas; doenças ósseas/genética; triagem neonatal

Summary

Yamamoto GL. Clinical applicability of the next generation sequencing technique with a focus on skeletal dysplasias [Thesis]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2017.

INTRODUCTION: In the last decade a new technique, the next generation sequencing, has emerged, which, contrary to the traditional Sanger method, performs parallel and high-throughput sequencing of multiple genes, or even all human genes, at a lower cost and with a faster analysis. This technique allowed the discovery of new genes responsible for several Mendelian diseases and has been quickly incorporated into the clinical context. **OBJECTIVES:** to compare the results of Sanger technique and next generation sequencing in control samples; to apply the next generation sequencing technique in the clinical practice to the cases of genetic skeletal disorders and RASopathies; to evaluate the diagnostic yield of this technique in samples without clinical data provided. **METHODS:** Next generation sequencing (in the form of a customized gene panel or exome) was performed in samples with mutations previously identified by Sanger sequencing and in two groups of Mendelian diseases, 144 patients with skeletal disorders and 79 patients with RASopathies, besides 90 samples with unknown clinical data (45 cases and 45 controls). The Sanger technique was applied in 29 samples of skeletal disorders and in 81 samples for confirmation of variants identified by next generation sequencing. **RESULTS:** The sensitivity of the next generation sequencing technique was estimated at 95.92% and the specificity at 98.77%. In the case of skeletal disorders, the diagnostic yield of the samples sequenced by Sanger was 69% (20/29), by customized panel, 60% (75/125), and by exome, 63% (12/19). In the individuals with RASopathies, the diagnostic yield through exome sequencing was 46% (36/79). As a result of this study, two new genes associated with RASopathies (*LZTR1* and *SOS2*) and one associated with a skeletal dysplasia (*PCYT1A*) were identified. In the analysis of the samples without previous knowledge of the clinical hypothesis, a total diagnostic yield of 46.67% (21/45) was obtained, but it was up to 73.08% (14/26) in the group with hypothesis of inborn errors of metabolism. **CONCLUSIONS:** It was demonstrated that the sensitivity and specificity

of the next generation sequencing technique are high and are similar to values found by other groups in the literature. The technique was considered appropriate not only for research, as demonstrated here by the identification of three new genes associated to Mendelian diseases, but also for clinical analysis. In this study the technique was successfully applied in the clinical context, both by customized panel and by exome, with positivity similar to that obtained by the Sanger technique. Even in the analysis of samples with no previous clinical history, it was possible to identify pathogenic variants in almost half of the cases, and an even greater percentage was obtained when the disease was an inborn error of metabolism. This sensitivity is comparable to that obtained by Tandem mass spectroscopy applied to multi-condition screening, suggesting that the next generation sequencing technique may be incorporated in the future into the neonatal screening program, increasing the use of genetic testing in complementarity to the biochemical tests.

Descriptors: sequence analysis, DNA; high-throughput nucleotide sequencing; genetics; genetics, medical; genetic diseases, inborn; bone diseases/genetic; neonatal screening.

1. Introdução

Os conhecimentos sobre o material genético humano vêm aumentando de forma exponencial nas últimas décadas. Embora as bases teóricas da hereditariedade tenham sido delineadas por Mendel há 150 anos, somente a partir da década de 1960 técnicas laboratoriais permitiram o estudo do material genético e o mapeamento dos cromossomos e, nas década seguintes, dos genes associados à herdabilidade dos fenótipos (Crotwell e Hoym, 2012; Heather e Chain, 2016).

As primeiras visualizações dos cromossomos no núcleo de uma célula sob microscópio óptico (citogenética) ocorreram no final dos anos 50 e permitiram à medicina associar quadros como a síndrome de Down e a de Klinefelter a anomalias cromossômicas numéricas - trissomia do cromossomo 21 e presença de três cromossomos sexuais (XXY), respectivamente. Com o desenvolvimento de colorações específicas na década de 60, chamadas técnicas de bandamento cromossômico, alterações estruturais nos cromossomos, acometendo parte dos cromossomos, como as deleções de pelo menos 5 a 10 megabases (Mb) e as translocações, puderam ser identificadas. No entanto, foi apenas a partir do final da década de 70, com o advento das técnicas de biologia molecular e reação da polimerase em cadeia (PCR), que testes para identificar alterações submicroscópicas se tornaram possíveis (Crotwell e Hoym, 2012). As técnicas de biologia molecular se dividiram em dois principais grupos: as técnicas de sequenciamento e as de citogenômica. As técnicas de citogenômica se propõem a detectar deleções e duplicações cromossômicas submicroscópicas chamadas de “CNV” (do inglês *Copy Number Variation*), que não podem ser observadas devido à limitação de resolução do cariótipo com bandamento. Iniciam-se com a técnica de FISH (*Fluorescent In Situ Hybridization*) e evoluem para as técnicas atuais de *array* cromossômico e MLPA (*Multiplex Ligation-dependent Probe*

Amplification). Dessa forma, essas diferentes técnicas são direcionadas para o estudo das anomalias cromossômicas (Crotwell e Hoym, 2012).

Já as técnicas de sequenciamento se propõem a determinar cada uma das bases dos genes humanos para a identificação de mutações de ponto (trocas de bases ou deleções e duplicações de poucas bases), sendo particularmente empregada(s) para as doenças monogênicas ou mendelianas. A principal técnica desenvolvida na década de 1970, ainda com uso de material radioativo e que se fundamentava na utilização de nucleotídeos terminadores (dideoxinucleotídeos) para obtenção de fragmentos curtos de DNA, foi a técnica de Sanger. O sequenciamento Sanger foi aprimorado entre as décadas de 1980 a 2000, trocando-se o método de detecção de radioatividade para o de fluorescência (Heather e Chain, 2016). O sequenciamento Sanger foi também a técnica predominantemente utilizada para a identificação completa do genoma humano (Projeto Genoma Humano), publicado em 2003, ao custo de bilhões de dólares. Esse projeto permitiu concluir que há nos seres humanos aproximadamente 20.000 genes, número inferior às estimativas iniciais (Chong et al., 2015; de Chadarevian, 2017). Devido ao grande avanço tecnológico associado a esse projeto, o valor para sequenciar um novo genoma humano completo (WGS – do inglês *Whole Genome Sequencing*) em 2003, ao término do projeto, seria algo na casa de 50-100 milhões de dólares em vez de bilhões de dólares (NIH - <https://www.genome.gov/sequencingcostsdata/>). Ainda assim, naquela época, os custos de sequenciamento para um genoma completo, ou mesmo para o sequenciamento do exoma completo (WES – do inglês *Whole Exome Sequencing*), eram completamente inviáveis para a aplicação rotineira na prática clínica. O exoma corresponde ao conjunto das sequências responsáveis por codificar moléculas de RNA, que em última instância codificam as proteínas. Embora estas regiões genômicas correspondam a uma fração entre 1 a 2% do genoma humano, elas abrigam em torno de 80 a 90% das alterações patogênicas encontradas nas doenças mendelianas (Chong et al., 2015). Dessa forma, o sequenciamento do exoma constitui uma abordagem de menor custo, mas com alta probabilidade de identificação das variantes patogênicas, quando comparado ao sequenciamento do genoma. Assim, no ano de 2000, quando havia necessidade de diagnóstico por sequenciamento na genética

médica, este ainda era feito pela técnica de Sanger. Em torno de 2007 surgiu uma nova técnica de sequenciamento denominada “sequenciamento de nova geração” (NGS – do inglês *Next Generation Sequencing*) ou “sequenciamento de segunda geração” (Heather e Chain, 2016). Em relação à técnica de sequenciamento Sanger, a técnica de NGS prometia apresentar escala de custo de 10 a milhares de vezes menor. O mecanismo que permitia tal redução de custo foi a introdução da estratégia de sequenciamento em larga escala e em paralelo, ou seja, métodos de sequenciamento automático de milhões de sequências de DNA simultaneamente. Assim, um gene inteiro, um grupo de genes, ou mesmo todos os genes de um indivíduo podiam ser sequenciados simultaneamente. Essa estratégia só se tornou possível devido ao grande desenvolvimento, no século XXI, de sensores / câmeras fotográficas para detecção dos múltiplos sinais em alta resolução e dos computadores para o posterior processamento dos dados obtidos pelos sensores. Diversas empresas desenvolveram sequenciadores baseados nessas estratégias com pequenas diferenças técnicas, como os sequenciadores da Roche (454), Life/Ion Torrent (Ion Torrent, Ion Proton), e Illumina (MiSeq, NextSeq, HiSeq), entre outros (Heather e Chain, 2016). Com essa nova técnica, o custo do WGS caiu em poucos anos, entre 2007 e 2011, de 10 milhões de dólares para menos de 10 mil dólares (<https://www.genome.gov/sequencingcostsdata/>). Com a estratégia de captura de regiões de interesse, tais como o WES, ou mais focadas ainda, como painéis de genes de interesse para determinados fenótipos (por exemplo, doenças ósseas genéticas ou síndrome de Noonan e síndromes relacionadas à síndrome de Noonan, decorrentes de mutações em genes pertencentes à via de sinalização RAS/MAPK - RASopatias), o custo de sequenciamento pode ser reduzido ainda mais, para valores compatíveis com custos de testes diagnósticos clínicos.

1.1. A técnica de NGS

A grande diferença entre a técnica de NGS e a técnica anterior de sequenciamento Sanger é que a técnica de NGS permite a análise de múltiplos genes simultaneamente, em um tempo e com um custo viáveis. Se tentássemos sequenciar o

exoma por Sanger, os custos seriam provavelmente na casa de dezenas de milhões de dólares e o tempo necessário, em torno de 200 semanas ou quase quatro anos (supondo 100 dólares por éxon e análise de 1000 éxons por semana), enquanto que o WES por NGS custa hoje aproximadamente 1000 dólares e fica pronto em menos de uma semana. A drástica redução do custo por base sequenciada e a diminuição e otimização do tempo para se realizarem os protocolos de laboratório e análise dos arquivos gerados pelo sequenciador são decorrentes da estratégia de paralelização do sequenciamento, bem como do uso de computação, tanto na etapa de sequenciamento como na etapa de análise de dados. Podemos dividir a técnica de NGS em três etapas fundamentais, que, embora possam apresentar variações de sequenciador para sequenciador, estão presentes em todos os tipos de NGS, quer seja com captura de painel de genes, quer com captura do exoma completo (regiões codificadoras de todos os genes), quer sem captura, gerando o WGS. A primeira etapa, chamada de “preparo das bibliotecas”, consiste na preparação do DNA para que ele seja “reconhecido” pelos sequenciadores NGS. A segunda etapa é o sequenciamento propriamente dito, que utiliza químicas inovadoras dentro de um robô de pipetagem com sensores e computadores acoplados. E, finalmente, a terceira etapa é a de análise dos dados gerados no sequenciador. Na técnica de NGS, devido ao imenso volume, tais dados são impossíveis de serem analisados pelos olhos humanos apenas (Heather e Chain, 2016). Uma comparação entre as diferentes etapas das técnicas de NGS e de Sanger pode ser observada na Figura 1, localizada no término do item 1.1 da introdução.

1.1.1.Preparo das bibliotecas

O preparo das amostras de DNA (ou RNA) com as etapas necessárias para que possam ser sequenciadas pela técnica de NGS é chamado de “preparo de bibliotecas”. Múltiplas empresas oferecem kits de preparo de bibliotecas e de captura de regiões de interesse. Embora, em tese, todo laboratório de biologia molecular possa comprar separadamente os reagentes e realizar o preparo das bibliotecas de forma “caseira” (*in-house*), a maioria dos laboratórios de pesquisa e de testes clínicos do mundo utiliza kits

pré-fabricados, opção de custo quase equivalente e que assegura maior rastreabilidade dos lotes de reagentes (Aigrain et al., 2016).

Independentemente do kit de preparo e captura a ser utilizado (painel de genes customizado ou captura do exoma completo), as etapas de preparo das amostras de DNA são as mesmas, com pequenas variações entre os kits de cada fabricante, a depender da origem e quantidade de amostra biológica disponível (por exemplo, sangue, saliva, bloco de tecido em parafina).

Após a extração do DNA, a primeira etapa de preparo da biblioteca consiste em se obter uma boa quantificação do DNA dupla fita das amostras, pois, se pouco concentradas, estas podem apresentar rendimento muito baixo ou nulo nas etapas seguintes.

Uma vez quantificado, o DNA precisa ser fragmentado em tamanhos de 100 a 600 pares de bases. A fragmentação pode ser mecânica, por meio de ultrassom, utilizando sonicadores (como, por exemplo, os da empresa Covaris – Covaris, Inc., Woburn, Massachusetts, USA), ou química, com múltiplos pares de enzimas de restrição (como nos kits Nextera da empresa Illumina – Illumina, San Diego, California, USA), para garantir uma fragmentação aleatória.

Após a fragmentação, é necessário ligar os adaptadores e *indexes* de identificação da amostra. Os adaptadores são sequências que não têm homologia ao DNA humano. Eles são utilizados para: a) permitir amplificação por PCR de todos os fragmentos com um único par de primer homólogo aos adaptadores, b) permitir a ligação de um primer de sequenciamento (também homólogo ao adaptador) e c) proporcionar um ponto de ligação dos fragmentos a uma placa de vidro (Illumina) ou esfera em emulsão (sequenciadores Ion Proton / Ion Torrent Life Technologies / Thermo Fisher Scientific – Thermo Fisher Scientific, Carlsbad, Califórnia, EUA). Os *indexes* são necessários para identificar os futuros fragmentos de DNA gerados, quando mais de uma amostra é sequenciada ao mesmo tempo. Para a ligação dos adaptadores e *indexes*, múltiplas técnicas podem ser utilizadas, como *end-repair* + adenilação + ligase (Illumina Truseq) ou retrotransposases (Illumina Nextera), por exemplo.

Uma vez realizada a ligação entre adaptadores e fragmentos de DNA da amostra, a maioria dos kits recomenda uma etapa de amplificação de todos os fragmentos por PCR, para aumentar a quantidade de fragmentos de DNA antes da captura. Essa etapa de amplificação não é necessária em kits de sequenciamento do genoma completo, pois não será realizada a captura. À amplificação, segue-se uma etapa de lavagem das enzimas contidas na amostra (normalmente realizada com etanol 80% e *beads* de purificação de DNA Agencourt AMPure – *beads* da empresa Beckman Coulter, Brea, Califórnia, EUA).

Múltiplas estratégias de captura podem ser empregadas (entre elas, o preparo de bibliotecas de amplicons de PCR específicos, que dispensa as etapas de fragmentação, ligação de adaptadores e amplificação inespecífica, previamente descritas). No entanto, uma das técnicas mais utilizadas - devido aos fatos de gerar menor quantidade de duplicatas de PCR e maior fidelidade à representatividade alélica original da amostra - é a técnica de captura por sonda de homologia ligada a *bead* magnética com estreptavidina. Essa captura consiste em ligar uma *bead* magnética apenas aos fragmentos de interesse, segurá-los na parede do tubo de plástico por meio de um ímã forte e lavar o resto de DNA da amostra com tampão.

Capturada a biblioteca, segue-se a uma nova etapa de amplificação inespecífica, lavagem do DNA e quantificação final da biblioteca para introdução no sequenciador. Esta etapa de quantificação final é crítica, principalmente nos sequenciadores Illumina, devido ao risco de, no momento da clusterização, poder gerar hipo- (e baixa quantidade de sequenciamento) ou hiper-clusterização (e perda da lâmina de sequenciamento por falta de discriminação entre os clusters) (Caruccio , 2011; Aigrain et al., 2016).

1.1.2.Método de sequenciamento NGS

Diferentemente do sequenciamento Sanger, que é baseado quase que exclusivamente em nucleotídeos terminadores com marcação fluorescente, o sequenciamento de nova geração não é realizado com o mesmo princípio químico nos

diferentes sequenciadores disponíveis. Múltiplas empresas investiram na tecnologia de sequenciamento de nova geração ou de segunda geração, entre elas ABI – SOLiD, Complete Genomics – (que vende sequenciamento, mas não sequenciadores), Illumina – MiSeq, HiSeq (entre outros), Life Technologies – Ion Torrent e Ion Proton, Qiagen – GeneReader e Roche – 454. Todavia, nem todas elas se mantiveram no mercado. O objetivo principal do sequenciamento de segunda geração é sequenciar em paralelo milhões de fragmentos curtos de DNA. Tais fragmentos podem ser sequenciados por dois métodos principais: sequenciamento por ligação (SBL) – SOLiD e Complete Genomics; e sequenciamento por síntese (SBS) – Illumina, GeneReader, Ion torrent/proton e 454. Os dois métodos de sequenciamento de segunda geração mais utilizados atualmente são os da Life Technologies e da Illumina, sendo que a tecnologia da Illumina foi a mais amplamente adotada no mercado, por apresentar menor taxa de erro e bom custo por base sequenciada (Goodwin et al., 2016).

Embora os diferentes sequenciadores apresentem diferenças na técnica, grande parte das etapas são comuns. A corrida de sequenciamento em si é iniciada pela etapa de clusterização (em lâmina de vidro nos sequenciadores Illumina ou *bead* em emulsão nos sequenciadores da Life Technologies – família Ion). A clusterização é uma nova amplificação inespecífica dos fragmentos de DNA para garantir sinal suficiente para a detecção da base sequenciada (luminoso – Illumina ou pH – Life), mas nessa amplificação é necessário que cada fragmento forme clones individualizados para que se possam sequenciar milhões de fragmentos simultaneamente. Nos sequenciadores Illumina, a estratégia utilizada é realizar o PCR em ponte na superfície de uma lâmina de vidro, garantindo que cada halo de DNA amplificado seja proveniente de um único fragmento original de DNA.

Os ciclos de sequenciamento são determinados pela inclusão sequencial, seguida de lavagem, das bases nitrogenadas A, T, C e G (com marcação de fluoróforo no caso de sequenciadores Illumina) que se ligarão por complementaridade aos clones do fragmento de DNA original e emitirão sinal a ser detectado pelo sequenciador (câmeras fotográficas – Illumina ou pHmetro – Life Technologies).

O sinal é gravado pelo computador do sequenciador e armazenado em imagens que serão transformadas em arquivos de texto (formato “*basecall*” – ‘.bcl’), incluindo a

sequência de DNA de cada fragmento e seus respectivos *indexes* de individualização das amostras.

1.1.3. Análise de bioinformática

Diferentemente do que ocorre no sequenciamento por Sanger, em que a análise da sequência pode ser realizada visualmente pelo homem observando diretamente os picos de intensidade de cores relacionados a cada base sequenciada, a análise das milhões de sequências de fragmentos curtos de DNA demanda o uso de computadores e processos de bioinformática.

Os arquivos ‘.bcl’ contendo as bases de DNA de cada fragmento são processados no computador por softwares das próprias empresas de sequenciamento (bcl2fastq ou HiSeq Analysis Software no caso da Illumina) para gerar os arquivos fastq. Os arquivos fastq são arquivos em formato tradicional ‘fasta’ que contêm todas as leituras de todos os fragmentos de uma determinada amostra/paciente. Nesses arquivos, no entanto, os fragmentos estão ainda em disposição completamente aleatória (da maneira que aleatoriamente foram fragmentados, capturados e clusterizados). Isso impossibilita a pesquisa de mutações, pois não se sabe a posição de cada fragmento no DNA humano (ou em outra espécie sequenciada).

A etapa necessária para associar os fragmentos de DNA contidos no arquivo fastq às suas respectivas posições originais nos cromossomos humanos é a de alinhamento. Múltiplos *softwares* gratuitos estão disponíveis para realizar o alinhamento de sequências de DNA. Um dos mais utilizados e recomendados pelo Instituto Broad (grupo de genômica da Harvard Medical School) é o bwa (Burrows-Wheeler-Aligner) e sua atualização bwa-mem (Li & Durbin, 2009). Essa etapa de alinhamento compara os fragmentos com a referência humana e aloca fragmentos com alta semelhança (que podem ter algumas diferenças em relação à referência, justamente pelo fato de o objetivo do sequenciamento ser a pesquisa de mutações) a posições cromossômicas, gerando um arquivo SAM (*sequence alignment map*) ou BAM (*binary alignment map*). Algumas etapas de “limpeza” dos arquivos BAM podem

ajudar na remoção de falso-positivos e na detecção de variantes: marcação das possíveis duplicatas de PCR, realinhamento com base em banco de variações de pequenas deleções e duplicações (indel) conhecidas e recálculo da qualidade do sequenciamento de cada base após o realinhamento. Esses arquivos BAM já podem ser visualizados em browsers como o IGV (*Integrative Genomic Viewer* – Broad Institute), para pesquisa e avaliação de mutações.

No entanto, a pesquisa manual de milhares ou milhões de variantes em centenas ou milhares de genes é inviável do ponto de vista prático e, portanto, é necessário que *softwares* de identificação de variantes sejam utilizados para a detecção automática e gravação das variantes em um arquivo chamado “VCF” (*Variant Call Format*). Múltiplos *softwares* de chamada de variantes foram desenvolvidos, como o GATK (*Genome Analysis Tool Kit* – Broad Institute) (GATK Best Practices - <https://software.broadinstitute.org/gatk/best-practices/>).

Mesmo com o arquivo VCF, o trabalho de identificação de variantes patogênicas ainda seria praticamente impossível, pois no arquivo VCF não há nenhuma informação sobre a variante em si, além de sua posição cromossômica e por vezes o número de registro no dbSNP (<https://www.ncbi.nlm.nih.gov/snp>). Assim, a última etapa da análise de bioinformática é a de anotação de inúmeras características biológicas, populacionais e preditivas de cada variante alterar ou não a proteína de forma deletéria. *Softwares* como o ANNOVAR (Wang et al., 2010) podem ser utilizados juntamente com múltiplos bancos de dados públicos: banco de transcritos de cada gene (REFSEQ - NCBI); bancos de frequências alélicas em populações, como ABraOM, ExAC, gnomAD, 1000 genomes, ESP6500; bancos de associação com doenças, como ClinVar, OMIM e HGMD; bancos de predição de dano à proteína, como SIFT, PolyPhen2, Mutation Taster e LRT (Seaby et al., 2016).

Figura 1: Comparação entre as técnicas de sequenciamento Sanger e sequenciamento de nova geração (NGS).

1.2. Desafio diagnóstico em doenças genéticas

No catálogo *Online Mendelian Inheritance in Man* (OMIM), que lista todos os fenótipos mendelianos, atualizado em julho de 2017, estão incluídas mais de 8.000 doenças, das quais aproximadamente 5.000 mil têm base molecular conhecida (www.omim.org/statistics/entry). Em conjunto, as doenças com padrão de herança mendeliano contribuem com uma porcentagem considerável das doenças humanas conhecidas (0,4% dos nascidos vivos). Quando consideradas também as anomalias

congênitas de causa multifatorial, aproximadamente 8% dos nativos apresentam uma doença com componente genético diagnosticada até a idade de adulto jovem. As malformações congênitas não só são frequentes, como também constituem a principal causa de óbito no primeiro ano de vida em países desenvolvidos (Chong et al., 2015). O conhecimento das bases moleculares das doenças pode contribuir para um manejo terapêutico mais adequado e direcionado, além de permitir estimar, de forma mais precisa, o risco de recorrência para as famílias.

O diagnóstico de doenças monogênicas permanece um desafio grande. Em ambulatórios especializados de Genética Médica, a taxa de diagnóstico fica ao redor de 50%. Somam-se a isso a demora na conclusão do diagnóstico, devido à necessidade de avaliações em diferentes serviços e à possibilidade de diagnósticos iniciais equivocados (Chong et al., 2015).

Nesse contexto, a técnica do sequenciamento de nova geração surge como ferramenta que permite identificar a base molecular em doenças sem etiologia definida. Foi inicialmente utilizada no âmbito de pesquisa e, posteriormente, aplicada na prática clínica, como ferramenta para baratear e acelerar os diagnósticos genéticos.

1.3. Aplicação da técnica de NGS no contexto de pesquisa

Historicamente, a identificação de genes responsáveis por um determinado fenótipo mendeliano foi possível a partir da identificação de loco (mapeamento gênico) por meio de estudos de ligação. O primeiro gene identificado por clonagem posicional foi o *CFTR*, responsável pela fibrose cística, doença de herança autossômica recessiva (Tsai e Dorfman, 2013). Menos comumente, uma anomalia cromossômica pode ser a causa de um fenótipo mendeliano e auxiliar na identificação do loco, como no caso da distrofia muscular de Duchenne – DMD (Koenig et al., 1987).

A partir de 2003, com a publicação do genoma humano completo, eliminou-se a necessidade de clonar cada gene ainda não associado a doenças. Mas, embora o genoma humano tivesse sido sequenciado, o fato é que muitos fenótipos ainda não

tinham etiologia genética conhecida. Isso se devia ao fato de que muitas das doenças raras não apresentavam famílias de afetados grandes o suficiente para viabilizar estudos de ligação, além do que, o sequenciamento Sanger permanecia muito custoso, em termos de trabalho e de tempo de execução. A descoberta da técnica de NGS permitiu reduzir tanto os custos de sequenciamento como o tempo e trabalho humano necessários para o sequenciamento de múltiplos genes. Essa análise simultânea de múltiplos genes em múltiplos pacientes facilita os processos de estudo de ligação entre indivíduos da mesma família, comparação de genes mutados entre indivíduos de famílias diferentes e identificação de mutações *de novo* em qualquer gene do exoma. Dessa maneira, nos últimos 10 anos (2007-2017), graças à técnica de NGS, observa-se a maior taxa da história de descoberta de genes associados a fenótipos. Em 2007, o banco de dados OMIM (*Online Mendelian Inheritance in Man*) registrava 2048 fenótipos com base molecular conhecida. Em 2013, este número subiu para 3834, e em julho de 2017, estão catalogados 5024 fenótipos com etiologia definida. Quadro 1 (Koboldt et al., 2013, Boycott et al., 2013, Chong et al., 2015, Goodwin et al., 2016, www.omim.org).

Quadro 1: Número de fenótipos com base molecular conhecida no OMIM, de acordo com os seus padrões de herança (adaptado de Koboldt et al., 2013 e atualizado em julho de 2017)

Descrição de fenótipos com base molecular conhecida - OMIM Padrão de Herança	Janeiro 2007	Julho 2013	Julho 2017
Autossômico	1851	3525	4669
Ligado ao X	169	277	320
Ligado ao Y	2	4	4
Mitocondrial	26	28	31
Total	2048	3834	5024

1.4. Aplicação da técnica de NGS no contexto clínico

Devido à grande redução de custo trazida pela técnica de NGS, rapidamente sua aplicação saiu do contexto de pesquisa e começou a ser incorporada nos laboratórios de análises clínicas. O CEGH-CEL (Centro de Pesquisas sobre o Genoma Humano e Células-Tronco, Instituto de Biociências, Universidade de São Paulo), utilizando-se de experimentos a serem apresentados na presente tese, foi um dos laboratórios pioneiros no Brasil a implementar a técnica na rotina de testes clínicos a partir de 2012, em paralelo aos laboratórios internacionais (Lee et al., 2014, Yang et al., 2014). Os principais aspectos da aplicação da técnica de NGS no contexto clínico não

necessariamente são uma preocupação quando a técnica é realizada no contexto de pesquisa. Tais aspectos foram e continuam a ser os seguintes: a) indicação para pacientes com suspeita de doença mendeliana; b) decisão entre teste focado ou ampliado (painel vs. WES ou WGS); c) aconselhamento pré-teste com termo de consentimento e aconselhamento pós-teste; d) critérios de qualidade do teste; e) interpretação das variantes encontradas; f) inclusão no relatório de achados secundários (Jamuar e Tan, 2015).

Vale ressaltar que, uma vez introduzida a técnica em laboratórios clínicos a custo mais acessível, médicos não geneticistas começaram a solicitar testes genéticos. Isso explica, em parte, a estimativa atual de que até 1/3 dos testes genéticos solicitados não têm indicação adequada (Ruzzo et al., 2017). Uma das maiores dificuldades na solicitação de testes clínicos continua a ser a escolha da abrangência do teste (principalmente painéis de genes vs. exoma completo), que envolve questões de custo e questões éticas em relação a achados secundários (Klein e Foroud, 2017). Em relação à solicitação do teste, há consenso mundial de que, quando se trata de uma questão de saúde, há necessidade de a solicitação ser realizada por médico e acompanhada de aconselhamento e consentimento adequados (Jamuar e Tan, 2015, CREMESP parecer 15036, 2013). Os critérios de qualidade necessários para a realização da técnica de NGS foram abordados a fundo tanto pela comunidade europeia como pela norte-americana e o fato de a transição entre pesquisa e contexto clínico ter sido muito rápida é um fator de preocupação (Rehm et al., 2013, Matthijs et al., 2016). Mas talvez a questão mais fundamental do teste clínico esteja no fato de o resultado não ser submetido a um conjunto de revisores especializados (como no caso de testes de pesquisa), embora tenha imediata relevância para a saúde do paciente. Nesse sentido a etapa possivelmente mais crítica do processo é a elaboração do relatório com a devida interpretação e inclusão ou não de variantes relacionadas ao quadro clínico do paciente e a achados secundários. O questionamento sobre quando relatar achados incidentais ou secundários sempre existiu em testes genéticos, nos quais, por exemplo, uma não-paternidade pode ser inadvertidamente identificada. Embora até hoje não haja um consenso mundial sobre quais achados relatar e a quem cabe a responsabilidade por estas informações, em 2015 o Colégio Americano de

Genética Médica e Genômica (*American College of Medical Genetics and Genomics – ACMG*) publicou recomendações para o relato de achados secundário em testes genéticos nos Estados Unidos da América (EUA). Essa recomendação serve como base para discussão, mesmo fora dos EUA (Richards et al., 2015, Matthijs et al., 2016, Kalia et al., 2017).

Observadas as questões anteriores, é fato inegável que os testes genéticos baseados na tecnologia de NGS contribuíram muito para a melhoria da prática clínica e a “saga diagnóstica” de pacientes com doenças raras. A sensibilidade diagnóstica (*diagnostic yield*) tem aumentado ao longo dos anos e depende de alguns fatores fundamentais: a qualidade da hipótese diagnóstica de quem indica o teste; a doença apresentar forte evidência de um padrão mendeliano clássico e se os genes associados são ou não conhecidos; e, finalmente, a disponibilidade de bancos de dados públicos de variantes patogênicas ou presentes em populações-controle. Os valores de sensibilidade diagnóstica em múltiplos estudos variam em torno de 25 a 40% para WES e de 20 a mais de 70% para painéis (Lee et al., 2014; Shashi et al., 2014; Wang et al., 2014; Yang et al., 2014; Jamuar e Tan, 2015; Jamie et al., 2016; Retterer et al., 2016; Yubero et al., 2016; Celestino-Soper et al., 2017).

1.5. Grupos de doenças monogênicas e aplicação do NGS

A escolha de modelos clínicos para validação da técnica nas múltiplas instituições teve como enfoque dois princípios norteadores: grupos de doenças que apresentassem um padrão de herança claramente mendeliano (ou seja, nas quais uma ou no máximo duas mutações em geral são as responsáveis pelo quadro clínico); e a disponibilidade de coortes de pacientes previamente estudadas por sequenciamento Sanger, sem identificação de sua base molecular. O grupo de doenças ósseas genéticas e o grupo constituído por mutações germinativas em genes da via RAS/MAPK (RASopatias) constituem dois exemplos que se enquadram nesses requisitos.

1.5.1. Doenças ósseas genéticas

As doenças ósseas genéticas constituem um conjunto heterogêneo de doenças caracterizadas por: 1) afetarem de maneira significativa o sistema esquelético, correspondendo à definição de displasias esqueléticas, doenças metabólicas ósseas, disostoses e malformações e/ou reduções do esqueleto; 2) possuírem uma base genética comprovada ou muito provável; e 3) apresentarem distinção entre si ao nível molecular (genes) ou ao menos um conjunto de características diagnósticas distintas observado em múltiplos indivíduos ou famílias (Superti-Furga et al., 2007).

Embora individualmente raras, a incidência conjunta das doenças ósseas genéticas ao nascimento pode chegar a 2-5/10.000 nascimentos (Rasmussen et al., 1996, Barbosa-Buck et al., 2012). A apresentação pode variar de um quadro letal a um leve acometimento ósseo, sem outros sinais clínicos, a depender da doença ou até mesmo do tipo de mutação dentro de uma mesma doença (Carter et al., 2009).

Uma nomenclatura internacional de doenças constitucionais do osso (*International Nomenclature of Constitutional Diseases of Bone*) foi inicialmente proposta em 1970 e constantemente atualizada por especialistas em 1977, 1983, 1991, 1997, 2001 e 2006, 2010 e 2015 (Mortier, 2001; Superti-Furga et al., 2007; Warman et al, 2011; Bonafe et al, 2015). Na última revisão, em 2015, foram incluídas 436 doenças, agrupadas em 42 grupos de acordo com critérios clínicos, moleculares, bioquímicos e/ou radiográficos. A grande maioria das doenças ósseas genéticas decorre de variantes em genes únicos, compatível com um padrão de herança mendeliano. A última classificação compreende desde as doenças mais clássicas, com padrão de herança monogênico e genes conhecidos, como a acondroplasia (autossômica dominante, associada ao *FGFR3*), o nanismo diastrófico (autossômica recessiva, associada ao gene *SLC26A2*) e a displasia frontometafisária (ligada ao X, associada ao gene *FLNA*), até afecções sem gene ou padrão de herança claramente definidos, como a síndrome fêmur-fíbula-ulna, incluindo ainda doenças com aumento ou diminuição da densidade óssea e um grande grupo de erros inatos do metabolismo (Bonafe et al,

2015). Um fato para os quais os revisores chamaram a atenção é o crescente conhecimento das bases moleculares das doenças ósseas genéticas nos últimos anos. Em 2006, quase metade (157 de 372) das osteocondrodisplasias listadas não tinha gene identificado; em 2010, 140 de 456 das afecções não tinham gene identificado e, em 2015, após o emprego da técnica do NGS, apenas 39 delas não tinham base genética definida. Entretanto, a identificação de novos genes (para um total de 364 em 2015) não necessariamente simplificou a classificação dessas doenças, dado que, por vezes, múltiplos genes são associados a uma única doença e, por outras, duas ou mais doenças são associadas a um único gene. Por exemplo: mutações no gene *FGFR3* estão associadas a acondroplasia, hipocondroplasia, displasia tanatofórica, síndrome de Crouzon com *acantosis nigricans*, síndrome de Muenke e neoplasias, como câncer colo-retal e mieloma múltiplo (nestes casos por ocasião de um evento somático). Já o gene do transportador de sulfato da displasia diastrófica da família de transportadores de solutos (*solute linked carrier* - *SLC26A2*) está associado a displasia diastrófica, atelosteogênese tipo 2, acondrogênese tipo 1B e displasia epifisária múltipla. No entanto, a síndrome de Crouzon (sem *acantosis nigricans*) é principalmente causada por mutações em *FGFR2*, enquanto a ateleosteogênese tipo 1 é causada por mutações em *FLNB* (Bonafe et al., 2015).

Ainda que a associação entre um determinado gene e um fenótipo específico não seja fácil em diversas circunstâncias, a identificação da base genética das doenças ósseas é de extrema importância, pois leva à possibilidade de um diagnóstico preciso, de aconselhamento genético mais acurado às famílias e de terapêuticas para os afetados. É o que ocorre nos casos de ampliação do uso de hormônio de crescimento, mesmo para síndromes sem sua deficiência *a priori* (Mazzanti et al., 2009), de aplicação de bisfosfonatos na tentativa de melhorar a qualidade óssea em situações osteopênicas (Batch et al. 2003) e de terapias específicas, como, por exemplo, a que está em desenvolvimento para acondroplasia (Lorget et al. 2012).

De acordo com as nosologias de doenças ósseas genéticas de 2006, 2010 e 2015 (Superti-Furga et al., 2007, Warman et al., 2011, Bonafe et al., 2015) em 2006 215 condições possuíam etiologia conhecida, em 2010 somavam 316 e, em 2015, 397 das 436. Esse aumento expressivo na identificação das bases moleculares coincide

com a descoberta da técnica de NGS. A primeira publicação que identificou um novo gene por NGS foi o relato, em 2010, de que a síndrome de Miller - uma disostose mandibulofacial com defeitos de redução de membros pós-axial - está associada a mutações no gene *DHODH* (Ng et al., 2010). Uma revisão da literatura com busca no Pubmed com os termos “skeletal dysplasia”, “genetic skeletal disorder” e “whole-exome sequencing” associada à busca específica por cada fenótipo da última nosologia sem etiologia conhecida resultou em mais 43 fenótipos cuja etiologia molecular já foi elucidada ou ao menos proposta, sendo que a maioria absoluta dessas etiologias foi descoberta por meio da técnica de NGS (Quadro 2). Em 2016, Chen e colegas publicaram um banco de dados chamado 'SkeletonGenetics' para armazenar as variantes já descritas associadas a doenças ósseas genéticas. O banco está disponível no site <http://101.200.211.232/skeletongenetics/> e até o momento conta com 8225 variantes em 357 genes (Chen et al., 2016).

Quadro 2: Genes relacionados a doenças ósseas genéticas que não tinham etiologia conhecida na nosologia de 2015, incluindo aqueles descritos previamente e não incorporados na nosologia (Bonafe et al., 2015).

Ano	Doença	OMIM Fenótipo	Gene	Referência
2012	Osteoartropatia hipertrófica primária (Paquidermoperiostose)	167100	<i>SLCO2A1</i>	Zhang et al.
	Estenose medular diafisária com histiocitoma fibroso maligno	112250	<i>MTAP</i>	Camacho-Vanegas et al.
2013	Displasia torácica de costelas curtas com ou sem polidactilia 8	615503	<i>WDR60</i>	McInerney-Leo et al.
	Displasia gnato-diafisária	166260	<i>ANO5</i>	Marconi et al.
2014	Defeito congênito da glicosilação (CDG) tipo Ig	607143	<i>ALG12</i>	Muraliet al.
	Síndrome Orofaciodigital 14	615948	<i>C2CD3</i>	Thauvin-Robinet et al.
	Imunodeficiência 23 (com displasia esquelética)	615816	<i>PGM3</i>	Stray-Pedersen et al.
	Displasia cerebrofaciotorácica	213980	<i>TMCO1</i>	Pehlivanet al.
2015	Síndrome de Steel	615155	<i>COL27A1</i>	Gonzaga-Jauregui et al.
	Displasia torácica de costelas curtas com ou sem polidactilia 15	617088	<i>DYNC2LI1</i>	Taylor et al.
	Meier-Gorlin 6	616835	<i>GMNN</i>	Burrage et al.
	Kabuki-like (com defeito de segmentação de vértebras) / síndrome Au-Kline	616580	<i>HNRNPK</i>	Au et al.
	Displasia Singleton-Merten	182250	<i>IFIH1</i>	Rutsch et al.
	Displasia torácica de costelas curtas com ou sem polidactilia 14	616546	<i>KIAA0586</i>	Malicdanet al.
	Platispondilia com amelogênese imperfeita	601216	<i>LTBP3</i>	Huckert et al.
	Osteogênese imperfeita tipo XVII	616507	<i>SPARC</i>	Mendoza-Londono et al.

Quadro 2, continuação

Ano	Doença	OMIM Fenótipo	Gene	Referência
2016	Osteopetrose-like	-	<i>RELA</i>	Frederiksen et al.
	Displasia de Gillesen-Kaesch- Nishimura	263210	<i>ALG9</i>	Tham et al.
	Meier-Gorlin 7	617063	<i>CDC45</i>	Fenwick et al.
	Displasia frontometáfisária	617137	<i>MAP3K7 TAK1</i>	Wade et al.
	Síndrome cardioespondilocarpofacial	157800	<i>MAP3K7</i>	Le Goff et al.
	Síndrome com múltiplas luxações	-	<i>EXOC6B</i>	Girisha et al.
	Displasia torácica de costelas curtas com ou sem polidactilia 16	617102	<i>IFT52</i>	Zhang et al.
	Síndrome Orofaciodigital	-	<i>IFT57</i>	Thevenon et al.
	Síndrome Orofaciodigital 15	617127	<i>KIAA0753</i>	Chevrier et al.
	Síndrome Orofaciodigital (com hamartomas da língua e polisindactilia)	217085	<i>INTU WDPCP</i>	Toriyama et al.
	Síndrome letal de contraturas congênitas 10	617022	<i>NEK9</i>	Casey et al.
	Displasia espondiloepifisária axial	601171	<i>C21orf2</i>	Wang et al.
	Displasia metáfisária osteoesclerótica	615198	<i>LRRK1</i>	Iida et al.
	Displasia geofísica/acromiocrítica	231050/614185 102370	<i>LTBP3</i>	McInerney-Leo et al.
	Progeróide neonatal / Wiedemann- Rautenstrauch	264090	<i>POLR3A</i>	Jay et al.
	Doença de Pyle	265900	<i>SFRP4</i>	Kiper et al.
	Disostose acrofacial tipo Rodriguez (alélica a Nager)	201170	<i>SF3B4</i>	Marques et al.
	Síndrome Orofaciodigital 16	617563	<i>TMEM107</i>	Shylo et al.
2017	Defeitos cardíacos e malformações esqueléticas	-	<i>ABL1</i>	Wang et al.
	Displasia espondiloepimetáfisária com neurodegeneração	-	<i>AIFM1</i>	Mierzevska et al.
	Displasia Desbuquois-Like	-	<i>CSGALNACT1</i>	Vodopivec et al.
	Displasia imuno-esquelética com anormalidades de neurodesenvolvimento	617425	<i>EXTL3</i>	Oud et al.
	Displasia espondiloepimetáfisária tipo Shohat	602557	<i>DDRKG1</i>	Egunsola et al.
	Larsen-like	-	<i>GZF1</i>	Patel et al.
	Meier-Gorlin 8	617564	<i>MCM5</i>	Vetro et al.
	Larsen-like com fraqueza muscular	-	<i>P4HA1</i>	Zou et al.
	Esclerose endosteal com hipoplasia cerebelar	213002	<i>POLR3B</i>	Ghoumid et al.

1.5.2.RASopatias

As RASopatias constituem um grupo de doenças decorrentes de mutações germinativas em genes da via de sinalização RAS/MAPK, incluindo a síndrome de Noonan (SN), a síndrome cardiofaciocutânea (CFC), a síndrome de Costello, a síndrome de Noonan com manchas lentiginosas múltiplas, a síndrome de Noonan-like com cabelos anágenos frouxos (SN-like), a neurofibromatose tipo I, a síndrome de Legius, a fibromatose gengival hereditária e a malformação capilar e arteriovenosa. Embora cada uma dessas doenças apresente um fenótipo específico, há uma sobreposição de suas características clínicas, particularmente entre as cinco primeiras. Elas se caracterizam por um comprometimento do crescimento, dismorfismos faciais, pescoço curto e/ou alado, anomalias cardiovasculares, deformidades torácicas, criptorquidia nos pacientes do sexo masculino, diátese hemorrágica, anomalias ectodérmicas, deficiência intelectual de grau variável e predisposição aumentada para o desenvolvimento de câncer. Entre elas, a SN é a mais prevalente e apresenta a maior heterogeneidade genética (Aoki et al., 2016; Tydman e Rauen, 2016). Uma representação dos múltiplos genes na via RAS/MAPK, bem como as doenças associadas a cada um deles, pode ser observada da Figura 2.

Figura 2: Via RAS/MAPK e respectivas doenças decorrentes de mutações germinativas associadas

NF1, Neurofibromatose tipo 1; SN, síndrome de Noonan; CFC, CardioFacioCutânea; * Indica proteínas/genes possivelmente causais que foram reportados a partir de 2013. Adaptado de Aoki et al., 2016.

O primeiro gene associado à SN, o gene *PTPN11*, foi descoberto em 2001, alterado em 50% da casuística estudada, o que demonstrava haver heterogeneidade genética na síndrome. Pela evidente sobreposição clínica entre a SN e a síndrome CFC, inicialmente os pesquisadores testaram o gene *PTPN11* em casuísticas de pacientes com CFC, mas não obtiveram êxito. Na sequência, a estratégia foi avaliar outros genes pertencentes à via de sinalização RAS/MAPK, permitindo a associação dos genes *KRAS* e *BRAF* à síndrome CFC. Com isso, o caminho natural foi testar outros genes da mesma via em indivíduos com SN e negativos para mutação no gene *PTPN11*, tendo sido descobertos sete outros genes distintos (*SOS1*, *RAF1*, *KRAS*, *NRAS*, *SHOC2* e *CBL*) pelo sequenciamento tradicional de Sanger até o ano de 2010. A positividade do estudo conjunto de todos esses genes em grupos de pacientes com SN, incluindo o mais frequente, o *PTPN11*, era de 70% (Tartaglia et al., 2010). A descoberta de novos genes responsáveis pela SN na porcentagem restante tem sido uma tarefa árdua, pois procuram-se genes responsáveis por frações muito pequenas. No entanto, com o emprego da técnica de NGS, partindo de casuísticas de indivíduos já previamente triados e sem mutações em genes conhecidos e/ou casos familiares, em apenas dois anos, entre 2013 e 2015, foi possível associar um número expressivo de genes à SN: *RIT1*, *RRAS*, *RASA2*, *A2ML1*, *SPRY1*, *MAP3K8*, *SOS2* e *LZTR1* (Aoki et al., 2016). Finalmente, em 2016, o último gene a ser identificado por NGS associado a SN com cabelos anágenos frouxos foi o *PPP1CB* (Gripp et al., 2016).

1.6. Contexto atual

Com a revisão apresentada da literatura e os exemplos dados dos dois grupos de doenças - doenças ósseas genéticas e RASopatias - pode-se observar que de 2011 até os dias atuais a técnica de NGS está em primeiro plano nas pesquisas genéticas e laboratórios clínicos em todo o mundo, sendo de fundamental importância para gerar novos conhecimentos. Mais especificamente, de 2011 a 2013 houve um foco maior na questão da qualidade das sequências obtidas por NGS e da comparabilidade entre a

técnica de NGS e a técnica de Sanger (Chin et al., 2013, Sikkema-Raddatz et al., 2013). Embora a qualidade da técnica de NGS ainda seja alvo de estudos (Goldfeder et al., 2016; Schenkel et al., 2016), de 2013 em diante os estudos passaram a dar mais ênfase à questão da sensibilidade diagnóstica (*diagnostic yield*) da técnica e de seus subconjuntos de exames (painel de genes vs. WES vs. WGS) em múltiplos contextos clínicos (Klein e Foroud., 2017). Desde 2016, um dos principais questionamentos em relação à técnica de NGS diz respeito à acurácia clínica na ausência de hipótese diagnóstica. Essa característica é de extrema relevância para a aplicação da técnica no contexto de triagem neonatal, que se torna cada vez mais uma necessidade premente (Francescatto e Katsanis, 2015; Valdez et al. 2016).

1.7. Triagem neonatal

A triagem neonatal é um conjunto de testes realizados em todos os recém-nascidos com o objetivo de identificar o mais precocemente possível doenças congênitas e aplicar tratamentos ou intervenções para tentar melhorar ou evitar o curso da doença (Pitt, 2010). O primeiro teste de triagem neonatal no mundo surgiu na década de 1960 para a triagem de fenilcetonúria, desenvolvido em grande parte a partir do trabalho de Robert Guthrie nos Estados Unidos. Nas décadas subsequentes, o programa se espalhou para quase todos os países do mundo e incorporou testes para outras doenças, como o hipotireoidismo congênito e a fibrose cística (Pitt, 2010). No entanto, a decisão de incluir um teste (ou doença a ser pesquisada) em um painel de triagem neonatal não é simples. Desde o início da triagem neonatal, quando apenas testes bioquímicos compunham o painel, a testagem de todas as crianças indiscriminadamente gerou várias preocupações. Os principais aspectos considerados são os custos para o sistema; os benefícios para a saúde do indivíduo; os riscos de erros nos testes, gerando angústia desnecessária nos pais; bem como os riscos éticos (detecção de falsa paternidade, por exemplo). Esses aspectos foram abordados na publicação, levada a cabo em 1968 pela Organização Mundial de Saúde, dos critérios de Wilson e Jungner para viabilidade da inclusão de doenças em programas de triagem

neonatal. Segundo esses autores, a doença a ser triada deve constituir importante problema de saúde para a população; a história natural da doença deve ser bem conhecida; deve existir um estágio precoce de doença identificável; o tratamento em estágio precoce deve trazer benefícios maiores do que em estágios posteriores; um teste adequado deve ser desenvolvido ou estar disponível para o estágio precoce; o teste e o tratamento devem ser aceitáveis para a população; intervalos para repetição do teste devem ser determinados; a provisão dos serviços de saúde deve ser adequada e suficiente, tendo em vista os custos extras de ações de saúde resultantes da triagem; os riscos, tanto físicos quanto psicológicos, devem ser menores do que os benefícios (Quadro 3) (Andermann et al., 2008).

No Brasil, os primeiros testes de triagem neonatal tiveram início no final da década de 1970 na Associação dos Pais e Amigos dos Excepcionais (APAE) de São Paulo. Mas o programa de triagem só se disseminou para outros estados na década de 90, depois de o Estatuto da Criança e do Adolescente e a portaria 22 de 1992 do Ministério da Saúde (MS) tornarem obrigatória a triagem da fenilcetonúria e do hipotireoidismo congênito. Em 2001, quando da criação do Programa Nacional de Triagem Neonatal (PNTN), a anemia falciforme e a fibrose cística foram adicionadas à lista de doenças de triagem neonatal. O MS estabeleceu a implantação do PNTN em fases: fase I - fenilcetonúria e hipotireoidismo congênito; fase II - anemia falciforme; fase III - fibrose cística. Em 2012, o PNTN foi revisado e a PT/GM nº 2.829 de 14 de dezembro de 2012 propôs a criação da fase IV, para incorporação da hiperplasia adrenal congênita (HAC) e a deficiência de biotinidase (DBT). (Souza et al., 2002; de Carvalho et al., 2007; <http://portalsaude.saude.gov.br>).

Os programas de triagem neonatal são também o primeiro e maior exemplo de triagem genética populacional. A triagem para fibrose cística nos Estados Unidos da América (EUA), com testes genéticos para mutações em *CFTR*, foi a pioneira em implementar testes de DNA em um programa de triagem. Mas é importante lembrar que mesmo por meio de testes bioquímicos apenas, o conhecimento do padrão de herança autossômico recessivo da maioria das doenças investigadas em triagem neonatal permite inferir o genótipo tanto da criança afetada (homozigoto ou heterozigoto composto) como de seus pais (heterozigotos obrigatórios) e, portanto, a

maioria dos testes de triagem neonatal disponíveis são, de uma maneira ou de outra, triagem de “genótipos/genética da população” (Green et al., 2006).

Quadro 3: Critérios maiores para Triagem Neonatal

1. A condição é um problema importante de saúde e o diagnóstico precoce traz benefício ao neonato.
2. Um tratamento aceitável está disponível, bem como a infraestrutura para confirmação, aconselhamento e tratamento
3. Um teste apropriado está disponível.
4. A história natural da condição é conhecida e existe um estágio latente ou de sintomas precoces reconhecível.
5. Os custos de triagem, confirmação e tratamento devem ser comparados aos custos totais da não-realização da triagem.

Adaptado de Pitt, 2010 e Andermann et al., 2008; Resumo dos critérios de Wilson e Jungner

Com o advento de técnicas que permitiam a triagem de múltiplas condições simultaneamente, sobretudo espectrometria de massas em Tandem, algumas condições passaram a ter custo zero para serem adicionadas ao painel de triagem neonatal. No entanto, o custo do teste não é o único critério de inclusão. Não há um consenso sobre a melhor maneira de se autorizar ou vetar a inclusão de novos testes em cada país ou sistema de saúde e estudos de efetividade de custo/benefício muitas vezes não são de simples realização ou interpretação, sobretudo quando as doenças são raras ou os benefícios, subjetivos (Pitt, 2010). Após discussão entre pesquisadores e gestores de saúde, o Departamento de Saúde e Serviços Humanos dos EUA (HHS), baseado nas recomendações da ACMG, concluiu que se a condição tem custo zero de teste e constitui diagnóstico diferencial de pelo menos uma outra condição que

preencha os critérios para figurar a lista de triagem neonatal, ambas podem ser incluídas na lista (*American College of Medical Genetics Newborn Screening Expert Group*, 2006). Nesse contexto, embora a alta sensibilidade de determinado teste constitua fator relevante para a possibilidade de sua aplicação em triagem neonatal, os testes para múltiplas condições simultâneas podem apresentar baixa sensibilidade diagnóstica para algumas ou várias das condições triadas (por exemplo tirosinemia tipo 1) (Pitt, 2010).

Percebe-se que, nos EUA, parte da triagem neonatal com uso de testes genéticos para fibrose cística já estava padronizada quando sobreveio a técnica de NGS que possibilita a investigação simultânea de múltiplos genes. Nesse contexto, surgiu rapidamente a questão de quando e como aplicar esta técnica para a triagem neonatal. (Valdez et al. 2016). Já no Brasil, embora nem mesmo a espectroscopia de massas em Tandem esteja no horizonte para implantação no SUS, tanto a triagem neonatal ampliada, como a tecnologia de NGS para testes diagnósticos, já são amplamente utilizadas na saúde suplementar (inclusive com previsão pela ANS - Resolução Normativa Nº 387, de 28 de outubro de 2015, atualizada em junho de 2016). Assim, os mesmos questionamentos sobre o uso de NGS em triagem neonatal também são válidos no Brasil.

2. Objetivos

2.1. Objetivo principal

Avaliar a aplicabilidade clínica da técnica de NGS.

2.2. Objetivos gerais

- A. Comparar os resultados das técnicas de Sanger e NGS em amostras controle.
- B. Introduzir e caracterizar a técnica de NGS no contexto clínico nas casuísticas de doenças ósseas genéticas e RASopatias.
- C. Avaliar a sensibilidade diagnóstica da técnica de NGS em amostras “cegas” (pacientes com doenças congênitas cujo quadro clínico o analisador não foi informado, como modelo para triagem neonatal).

2.3. Objetivos específicos

A.1. Determinar a sensibilidade do método de NGS com o re-sequenciamento de amostras positivas por Sanger.

A.2. Determinar a especificidade do método de NGS com o re-sequenciamento por Sanger de mutações encontradas primariamente por NGS.

A.3. Comparar as vantagens e desvantagens de cada técnica em diferentes situações.

B.1. Estabelecer a sensibilidade diagnóstica (*diagnostic yield*) da técnica de NGS em casuística de doenças ósseas genéticas, tanto nas amostras sequenciadas com captura de painel de genes, como nas amostras sequenciadas por WES.

B.2. Procurar novos genes associados a doenças ósseas genéticas e RASopatias em pacientes e familiares das casuísticas estudadas.

C.1. Determinar a sensibilidade diagnóstica (*diagnostic yield*) e a taxa de falso-positivos em estudo de casos e controles com análise cega que simule a triagem neonatal.

C.2. Estabelecer critérios para reportar mutações patogênicas em amostras com análise cega, a fim estabelecer o ponto ideal na relação entre sensibilidade e taxa de falso-positivos.

3. Casuística e Métodos

Para facilitar a referência aos diferentes procedimentos e resultados, as etapas de métodos (e posteriormente resultados e discussão) foram divididas e nomeadas etapas “A”, “B” e “C”, em correspondência aos objetivos de mesmo nome. Ou seja, a etapa A compreende os experimentos para comparação direta da sensibilidade e especificidade entre as técnicas de Sanger e NGS. A etapa B compreende a aplicação de NGS às casuísticas de doenças ósseas genéticas e RASopatias. E a etapa C, a aplicação da técnica de NGS sem conhecimento prévio de diagnóstico ou hipótese diagnóstica, em contexto análogo ao da triagem neonatal.

3.1. Casuísticas

3.1.1. Casuística das etapas A e B

A Unidade de Genética do Instituto da Criança (ICr), hospital terciário ligado à Faculdade de Medicina da Universidade de São Paulo (FMUSP), atende em média 1.500 pacientes por ano acometidos por diversas doenças genéticas. As displasias esqueléticas, incluídas na nova classificação das doenças ósseas genéticas de 2015 (Bonafe et al., 2015), compõem aproximadamente 15% destes atendimentos, enquanto as RASopatias compõem aproximadamente 10%, ambas parcelas consideráveis dos atendimentos. Ao todo, foi coletado DNA de aproximadamente 250 pacientes com doenças ósseas genéticas ou respectivos parentes e 200 de pacientes com RASopatias ou respectivos parentes (Figura 3).

De 2010 a 2012, antes do início desta tese, foi realizado o sequenciamento Sanger, em parceria com o Laboratório de Cardiologia e Biologia Molecular do InCor,

dos genes *FGFR3*, *TRPV4*, *GDF5*, *GJA1*, *SOST*, *LIFR*, *CANT1*, *LMNA*, *WNT7A* e *PAPSS2* em 75 pacientes com doenças ósseas genéticas, com elucidação da etiologia molecular em 46 deles, o que corresponde a uma sensibilidade diagnóstica (*diagnostic yield*) em torno de 61%. Aos 40% dos pacientes sem elucidação molecular e aos mais de 100 pacientes integrantes da coorte em que a suspeita diagnóstica recai sobre genes não incluídos entre os sequenciados por Sanger, propôs-se a inclusão nesta tese, para que fossem submetidos, quando apropriado, à investigação por sequenciamento de nova geração.

Até 2012, como resultado de teses anteriores do departamento de Genética Médica ICr-FMUSP e de parceria com o Laboratório de Cardiologia e Biologia Molecular do InCor, as principais mutações nos principais genes associados a RASopatias [*PTPN11* (ex. 2–15), *SOS1* (ex. 1–23), *RAF1* (ex 7,14,17), *KRAS* (ex. 2–6), *SHOC2* (ex. 2) e *CBL* (ex. 8–9)] haviam sido testadas em mais de 200 pacientes. Foram selecionados 79 probandos com diagnóstico clínico de RASopatias (em sua maior parte síndrome de Noonan) e respectivos familiares para realização de WES.

A inclusão de pacientes neste estudo foi condicionada à assinatura, pelo próprio indivíduo ou por responsável, de termo de consentimento livre e esclarecido. Os projetos foram aprovados pelo Comitê de Ética do Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo sob os números 0558/10 e 0843/08.

Os principais procedimentos realizados em ambas as coortes foram:

A) Seguimento clínico dos pacientes em ambulatório de genética clínica com coleta e arquivamento de:

A1) História clínica, exame físico completo, heredograma e resultado de exames complementares necessários ao paciente (bioquímica, imagem, avaliação auditiva ou oftalmológica, etc). Quando possível, os pais foram avaliados clinicamente, para se constatar se a doença era isolada ou herdada;

A2) Radiografias de esqueleto completo (crânio e coluna – frente e perfil –; tórax, ossos longos, quadril, mãos e pés – frente), realizadas no Setor de Radiologia do Instituto da Criança;

A3) Fotografias do paciente, com permissão dos responsáveis;

B) Coleta de material para testes moleculares.

B1) Foram coletados 4 mililitros (mL) de sangue em tubo contendo ácido etileno diaminotetracético (EDTA), para extração do DNA dos leucócitos do sangue periférico e realização do teste molecular. Coletou-se sangue dos pais e/ou outros parentes quando havia mais de um afetado na família ou quando a realização do sequenciamento por NGS visava a avaliar a segregação das variantes presentes apenas no indivíduo afetado. Quando o familiar não podia comparecer ao atendimento, enviou-se um kit de coleta de saliva, para posterior extração de DNA, conforme protocolo do fabricante (Oragene – DNA Genotek, subsidiária de OraSure Technologies, Bethlehem, Pensilvânia, EUA).

Além dos pacientes das casuísticas de doenças ósseas genéticas e RASopatias, foram incluídos na etapa A (de comparação direta entre Sanger e NGS) alguns pacientes do CEGH-CEL que disponibilizaram seus dados para realização de pesquisa por meio de opção em termo de consentimento informado.

Ao todo, na etapa A, foi realizado sequenciamento Sanger seguido de NGS em 49 pacientes. Além disso, realizou-se sequenciamento NGS seguido de confirmação por Sanger em 81 pacientes (Figura 3).

Na etapa B foi realizado NGS em 144 probandos da casuística de doenças ósseas genéticas, dos quais 125 por painel de genes e 19 por WES. Adicionalmente, 29 indivíduos foram avaliados pelo sequenciamento tradicional de Sanger. Da casuística de RASopatias, foram sequenciados 79 probandos por WES (Figura 3).

3.1.2. Casuística da etapa C

Para a análise de uma coorte sem dados clínicos informados, estabeleceu-se colaboração com o Laboratório de testes genéticos do Centro de Pesquisas sobre o Genoma Humano e Células Tronco (CEGH-CEL) do Instituto de Biociências da Universidade de São Paulo e com o Departamento de Genética da Universidade Federal do Rio Grande do Sul (UFRGS). Foram selecionadas e sequenciadas por painel NGS ('P4') 90 amostras de pacientes (45 de cada centro) para a análise "cega" de dados de NGS. Os 45 pacientes do CEGH-CEL haviam sido testados, por meio do

painel de genes customizado nesse serviço, para doenças de padrão de herança mendeliana, com resultado negativo, e foram portanto selecionados como controles negativo para doenças genéticas. Todos eles disponibilizaram dados para a realização de pesquisa por meio de opção em termo de consentimento informado. Os 45 pacientes da UFRGS eram acompanhados com hipótese de doenças congênitas e forneceram amostra para que fossem incluídos em casuística de pesquisas que investigassem a etiologia molecular de suas doenças (Figura 3). Após a análise molecular realizada neste estudo, as seguintes informações foram obtidas da casuística da UFRGS: dos 45 indivíduos, 14 obtiveram confirmação molecular pela técnica do Sanger (sendo todos erros inatos do metabolismo). Dos 31 restantes, 10 apresentavam hipótese de erros inatos do metabolismo; quatro, de uma síndrome epiléptica; 16, de doenças neuromusculares e um, de imunodeficiência.

Figura 3: Casuística das etapas A, B e C

3.2. Técnicas Moleculares

3.2.1.Extração de DNA

A maioria das amostras utilizadas na tese tiveram DNA extraído a partir de sangue periférico no robô de extração automático QIA Symphony com o kit de extração de DNA da própria Qiagen (Qiagen, Hilden, Alemanha). Algumas amostras foram extraídas a partir de saliva com o kit de extração de DNA de saliva da Oragene (Oragene – DNA Genotek, subsidiária de OraSure Technologies, Bethlehem, Pensilvânia, EUA). Amostras provenientes de outros laboratórios foram extraídas pela técnica de *'salting out'* ou protocolos validados nos respectivos laboratórios.

Todas as amostras foram analisadas em NanoDrop (Thermo Fischer Scientific, Carlsbad, Califórnia, EUA) para quantificação do DNA e avaliação da pureza (razão das absorbâncias em comprimentos de onda 260/280nm maior que 1,80).

3.2.2.Sequenciamento Sanger

Ensaio para a identificação de mutações com sequenciamento bidirecional do gene em questão foram realizados nos sequenciadores automáticos ABI Prism Genetic Analyzer ABI3500XL (Applied Biosystems, Thermo Fischer Scientific, Carlsbad, Califórnia, EUA) do Laboratório de Cardiologia e Biologia Molecular do InCor ou no ABI3730 Genetic Analyzer (Applied Biosystems, Thermo Fischer Scientific, Carlsbad, Califórnia, EUA) do CEGH-CEL. Foram desenhados e solicitados primers para a região de interesse. Seguiram-se amplificação por PCR convencional, purificação com ExoSAP-IT (Applied Biosystems, Thermo Fischer Scientific, Carlsbad, Califórnia, EUA), reação de sequenciamento com BigDye Terminator (Applied Biosystems, Thermo Fischer Scientific, Carlsbad, Califórnia, EUA), seguida de purificação final em coluna de sílica e aplicação do produto final no sequenciador.

3.2.3. Sequenciamento de nova geração (NGS)

O sequenciamento de nova geração foi realizado em sua maior parte no CEGH-CEL, em colaboração para a implementação da técnica na rotina de testes genéticos clínicos oferecidos pelo laboratório do centro. Uma pequena parte das amostras para teste de sensibilidade da técnica de NGS foi realizada no contexto de outra tese de doutorado, sob a orientação da Dra. Débora Bertola, com amostras de RASopatias, em colaboração para implementação da técnica no laboratório Fleury.

3.2.3.1. Paineis de genes customizados

Em 2012, com o *software* online da empresa Illumina (Illumina Design Studio), foi desenvolvido para este trabalho um painel que incluía 82 genes responsáveis por diferentes doenças ósseas genéticas. Esses genes foram selecionados tendo em vista as doenças ósseas genéticas mais prevalentes no ambulatório de genética médica do ICr-HC-FMUSP, à luz da classificação das doenças ósseas genéticas adotada na época (Warman et al., 2011). Os genes de doenças ósseas genéticas desse painel inicial (2012-2014) foram adicionados a um total de 230 genes, contemplando outras doenças de interesse do laboratório, como transtorno do espectro autista, doenças neuromusculares e síndromes com anomalias craniofaciais. Após um piloto inicial para 96 amostras ('P1' - 230 genes, Anexo I), o painel foi revisado e vários genes e regiões foram acrescentados, relacionados a câncer hereditário, doenças de padrão de herança recessivo, com possível interesse para triagem neonatal ou pré-concepcional de casais, outras síndromes com distúrbio do desenvolvimento e RASopatias ('P2' - 465 genes, Anexo II). Tanto o P1 como o P2 foram realizados utilizando o kit Illumina *TruSeq custom enrichment*, que necessita de 1µg de DNA inicial e fragmentação via ultrassom com uso de sonicador Covaris (Covaris, Inc., Woburn, Massachusetts, USA). Posteriormente, em 2014 e 2015, o painel de genes foi revisado e a tecnologia de preparo de biblioteca e captura foi atualizada para o kit Illumina *Nextera Rapid Capture Custom Enrichment*, que demanda menor quantidade de DNA inicial (50ng) e possibilita fragmentação direta por método enzimático, prescindindo do sonicador. Em

2014 e 2015, à medida que foram sendo reportados na literatura, incluíram-se alguns poucos genes dos mesmos grupos de doenças de interesse já existentes ('P3' - 485 genes, Anexo III; 'P4' - 496 genes, Anexo IV) . O painel mais atualizado ('P4') inclui 142 genes para doenças craniofaciais, doenças ósseas genéticas e doenças do tecido conectivo. O painel inicial ('P1') foi utilizado majoritariamente nas amostras da etapa A, enquanto os painéis revisados e ampliados foram utilizados para amostras da etapa A, majoritariamente na etapa B e para todas as amostras da etapa C. Para esta última etapa, analisaram-se os genes associados às 34 doenças primárias recomendadas na triagem neonatal dos EUA, bem como às doenças que compõem diagnósticos diferenciais [*Uniform Screening Panel* do *Department of Health and Human Services* (HHS) dos EUA].

3.2.3.2. Captura do exoma completo (WES)

O kit de captura utilizado para WES na maioria das amostras foi o Illumina *Nextera Rapid Capture Expanded Exome Kit*, que apresenta 62Mb de região alvo, incluindo todos os exons codificantes e UTRs de genes descritos no RefSeq, além de miRNAs. No início (2012-2013), algumas amostras foram realizadas com o kit Illumina *TruSeq exome enrichment* (com região alvo de 45Mb, que não inclui UTR e miRNA). A partir de 2015, algumas amostras foram processadas com o kit Agilent *SureSelect Human All Exon V6* (com região alvo de 60Mb que não inclui miRNA).

3.2.3.3. Preparo das bibliotecas

As bibliotecas foram preparadas seguindo as recomendações do fabricante dos respectivos kits. Todos os kits de painel foram preparados em lotes de 12 amostras por captura, o que também se aplica aos kits de WES da empresa Illumina. Com o kit de WES da empresa Agilent, as amostras foram preparadas individualmente e adicionadas a um lote (*pool*) comum logo antes da etapa de sequenciamento. O lote final, pronto para sequenciamento, foi quantificado tanto pelo kit *High Sensitivity DNA Analysis* do Bioanalyser (Agilent Technologies, Santa Clara, Califórnia, EUA), para

determinação do tamanho das bibliotecas, como por qPCR no instrumento LightCycler 480 (Roche - F. Hoffmann–La Roche AG, Basileia, Suíça), utilizando kits da empresa Kapa, KAPA SYBR FAST (Kapa Biosystems, subsidiária da Roche - F. Hoffmann–La Roche AG Basileia, Suíça), específicos para bibliotecas a serem utilizadas em sequenciadores Illumina. Após quantificação, os lotes de bibliotecas foram diluídos para 4nM e injetadas nos sequenciadores Illumina para a etapa de clusterização em quantidades de 17 a 20pMol.

3.2.3.4. Sequenciamento NGS

O sequenciamento de nova geração no CEGH-CEL foi realizado nas plataformas MiSeq (painel de genes) e HiSeq (WES) da Illumina (Illumina, San Diego, Califórnia, EUA). As amostras foram preparadas para sequenciamento de acordo com os protocolos dos fabricantes e os kits de sequenciamento foram comprados da empresa Illumina de acordo com a quantidade necessária de sequenciamento por amostra (Illumina – TruSeq library preparation, TruSeq exome enrichment, TruSeq custom enrichment, Nextera Rapid Capture Expanded Exome Kit, Agilent SureSelect Human All Exon V6, MiSeq sequencing by synthesis, HiSeq sequencing by synthesis, Paired-End Cluster Generation). A clusterização das amostras corridas no MiSeq ou HiSeq modo '*Rapid*' foi realizada no próprio sequenciador, enquanto a clusterização das amostras de WES realizadas no HiSeq modo '*High Output*' foi realizada no cBot (robô de clusterização da própria Illumina).

O sequenciamento de nova geração no Fleury foi realizado na plataforma IonTorrent (Life Technologies / Thermo Fischer Scientific, Carlsbad, Califórnia, EUA), com preparo das bibliotecas por meio de kit de captura customizado para genes de RASopatias da própria Life Technologies e chips de sequenciamento compatíveis com o sequenciador IonTorrent (Life Technologies / Thermo Fischer Scientific, Carlsbad, Califórnia, EUA).

Todas as amostras foram sequenciadas com o intuito de uma cobertura final nas regiões alvo de pelo menos 30x. Para garantir esse valor mínimo, os WES foram sequenciados em quantidade estimada a gerar 60-100x de cobertura média por

amostra e os painéis foram sequenciados em quantidade estimada a gerar 100-200x de cobertura média por amostra. A conta realizada para estimar a cobertura final nas regiões alvo é: total de sequência gerada pelo kit de sequenciamento (número de ciclos e bases sequenciadas por fragmento X número total de fragmentos sequenciados) dividido pelo tamanho da região capturada (em torno de 60Mb para WES e 1 a 2Mb para painéis) multiplicada pela eficiência de captura (quantidade de DNA das regiões alvo nas amostras / quantidade total de DNA). Diferentes kits de sequenciamento foram usados para cada amostra e sequenciador, mas todos geram sequências pareadas (*paired-end*) de tamanho 100x100 a 300x300 pares de bases.

3.3. Análise de bioinformática

Os cromatogramas gerados pelos sequenciadores do tipo Sanger (formato .abi) foram analisados visualmente com auxílio do *software* Sequencher (Gene Codes Corporation, Ann Arbor, Michigan, EUA). As variantes foram anotadas manualmente e informações sobre elas foram conferidas em *websites* de bancos de dados (MutationTaster, HGMD, ClinVar, OMIM, Mutalyzer, SIFT, PolyPhen2).

As sequências geradas por NGS foram transformadas em formato fastq automaticamente pelos próprios sequenciadores ou em servidores com *software* disponibilizado pela empresa fabricante da máquina de sequenciamento (Illumina Bcl2Fastq – HiSeq). As sequências fastq foram consideradas dados crus de sequenciamento e armazenadas como tal. Um conjunto ordenado de comandos e programas de computador (*pipeline*) foi desenvolvido neste estudo por meio de *softwares* disponíveis gratuitamente para a comunidade científica e com base nas “melhores práticas” (*Best Practices*) recomendadas pelo centro de genômica da Harvard – Broad Institute (DePristo et al., 2011). O *pipeline* de análise usou as sequências fastq de entrada para primeiramente alinhar com bwa (Burrows-Wheeler-Aligner) ou bwa-mem (Li & Durbin, 2009) em relação à referência humana hg19, gerando arquivos SAM (Sequence Alignment Map). Em seguida, com o programa *picard* (<http://broadinstitute.github.io/picard>), foi realizada a conversão de SAM para

BAM (*Binary Alignment Map*) e foram marcadas as sequências provavelmente geradas por duplicatas de PCR (*mark duplicates*). A identificação de variantes de ponto e inserções/deleções (in/del) foi realizada com o programa GATK (Genome Analysis Tool Kit) do instituto Broad (McKenna et al., 2010), mais especificamente com a ferramenta ‘*Unified Genotyper*’. Processos de realinhamento baseado em variações de in/del locais e recalibração da qualidade das bases sequenciadas foram realizados com o próprio GATK, seguindo as recomendações de *Best Practices*. As variantes encontradas, armazenadas em arquivos do tipo VCF (variant call format), foram anotadas com o programa Annovar (Wang et al., 2010), utilizando múltiplos bancos de dados públicos, tanto aqueles criados no exterior (ExAC, gnomAD, 6500 exomes ESP, 1000 genomes, HGMD, NCBI ClinVar, NCBI dbSNP, Sift, PolyPhen2, Mutation Taster, NCBI RefSeq), como o desenvolvido em nosso país (ABraOM), segundo a lista abaixo.

Lista de sites com bancos de dados públicos utilizados para anotação das variantes (com seus respectivos usos):

<http://abraom.ib.usp.br/> (frequência de variantes em população controle brasileira)

<http://exac.broadinstitute.org/> (frequência de variantes em população controle)

<http://gnomad.broadinstitute.org/> (frequência de variantes em população controle)

<http://evs.gs.washington.edu/EVS/> (frequência de variantes em população controle)

<http://www.internationalgenome.org/> (frequência de variantes em população controle)

<http://www.hgmd.cf.ac.uk/ac/> (banco de variantes patogênicas)

<https://www.ncbi.nlm.nih.gov/clinvar/> (banco de variantes patogênicas)

<https://www.ncbi.nlm.nih.gov/projects/SNP/> (banco de variantes conhecidas)

<http://sift.jcvi.org/> (programa de predição de dano à proteína)

<http://genetics.bwh.harvard.edu/pph2/> (programa de predição de dano à proteína)

<http://www.mutationtaster.org/> (programa de predição de dano à proteína)

<https://www.ncbi.nlm.nih.gov/refseq/> (banco de genes e transcritos conhecidos)

A anotação das variantes gerou arquivos em texto separados por tabulação ou vírgulas que foram analisados como planilhas de Excel.

Para a pesquisa de deleções ou duplicações grandes (variações do número de cópias, CNV), foram utilizados a análise visual de regiões candidatas no Integrative Genomic Viewer (IGV) do Broad Institute e o programa de computador NextGene da empresa SoftGenetics (SoftGenetics, State College, Pensilvânia, EUA), que aplica um algoritmo de cadeia oculta de Markov (HMM) para estimar o número de cópias de cada região cromossômica a partir da cobertura local normalizada para cada amostra.

3.4. Interpretação das variantes

As variantes foram interpretadas como patogênicas ou não de acordo com critérios clássicos de associação entre fenótipo e genótipo. Estes critérios levam em consideração a raridade das variantes em populações de controles, a presença das variantes em pacientes afetados previamente descritos ou avaliadas por estudos funcionais, a segregação da variante com o fenótipo na família, a predição de alteração na proteína ou *splicing* e a compatibilidade do fenótipo do paciente com o gene identificado. Em 2015, a ACMG publicou recomendações para interpretação de variantes genéticas associadas a doenças mendelianas, recomendações essas que incorporaram e categorizaram os critérios mencionados anteriormente (Richards et al., 2015) (Quadros 4 e 5). A partir de 2015, os critérios utilizados nesta tese para classificação de variantes patogênicas foram aqueles da ACMG. Antes de 2015, foram utilizados os seguintes limiares para filtragem e classificação de patogenicidade de variantes:

-Frequência alélica em populações de controles:

- Menor que 1% em controles externos (1000 genomas e 6500 exomas)
- Menor que 5% em controles internos (amostras do CEGH-CEL)

-Presença da variante em outros pacientes com a mesma doença:

- Presença em bancos de dados públicos (por exemplo HGMD e ClinVar)

- Artigo na literatura com relato de outros pacientes acometidos e/ou estudos funcionais da variante
- Predição de alteração da proteína ou *splicing*:
 - Não codificante < sinônima <*missense* ou *in frame*<*splicing* <*nonsense* ou *frameshift* ou grandes deleções
 - Múltiplos programas *in silico* de predição de dano apontando como deletéria (pelo menos três dos seguintes: Sift, PolyPhen2, Mutation Taster e LRT)
- Compatibilidade com o fenótipo do paciente:
 - Quadro clínico do paciente semelhante ao descrito na literatura
 - Variante presente em heterozigose para doenças com padrão de herança dominante; variante em homozigose ou em heterozigose composta para doenças com padrão de herança autossômico recessivo e variante em hemizigose para doenças com padrão de herança ligado ao X recessivo
- Segregação da variante na família:
 - Variante *de novo* para quadros isolados
 - Variante presente em outros afetados e ausente em familiares não afetados para casos dominantes familiares (com penetrância completa)
 - Variante presente em heterozigose nos pais do afetado ou presente em homozigose/heterozigose composta em outros familiares afetados para casos autossômicos recessivos

Quando necessário, a variante foi avaliada em outros membros da família do probando, tanto em outros afetados, como em indivíduos normais, desde que houvesse disponibilidade e consentimento em participar da pesquisa. A avaliação de qualidade do sequenciamento e da presença da variante na amostra por meio da interpretação visual dos dados de sequenciamento com o visualizador IGV nos respectivos arquivos BAM foi realizada para todas as variantes classificadas como patogênicas ou provavelmente patogênicas pelos critérios anteriores.

Quadro 4: Critérios de patogenicidade ou benignidade de variantes, classificação da ACMG (Richards et al., 2015)

Benigna (<i>Benign</i>)				Patogênica (<i>Pathogenic</i>)			
	Stand alone	Forte (<i>Strong</i>)	Suporte (<i>Supporting</i>)	Suporte (<i>Supporting</i>)	Moderada (<i>Moderate</i>)	Forte (<i>Strong</i>)	Muito Forte (<i>Very Strong</i>)
Dados Populacionais	MAF (<i>minor allele frequency</i>) > 5% em controles BA1	MAF muito alto para doença BS1 OU observação nos controles inconsistente com a penetrância da doença BS2			Ausente nos bancos de dados populacionais PM2	Prevalência nos afetados estatisticamente maior que nos controles PS4	
Dados Computacionais e Preditivos			Múltiplos programas de predição <i>in silico</i> sugerem que não há impacto no gene/proteína BP4 Missense em um gene onde apenas truncante causa doença BP1 Sinônima sem predição de alterar <i>splicing</i> BP7	Múltiplos programas de predição <i>in silico</i> sugerem que há efeito deletério no gene/proteína PP3	Missense nunca descrita no mesmo resíduo de aminoácido de uma variante missense já estabelecida como patogênica PM5 Variante altera o comprimento da proteína PM4	Troca de mesmo aminoácido que em uma variante já estabelecida como patogênica PS1	Variante predita como nula em um gene onde LoF (<i>loss of function</i>) é sabidamente mecanismo de doença PVS1
Dados Funcionais		Estudos funcionais bem estabelecidos não mostram perda de função BS3				Estudos funcionais bem-estabelecidos mostram efeito deletério PS3	
Dados de Segregação		Não-segregação com a doença BS4		Co-segregação da doença em múltiplos afetados PP1	Aumento dos dados de segregação →		
Dados De novo					De novo (SEM paternidade e maternidade confirmados) PM6	De novo (paternidade e maternidade confirmados) PS2	
Dados Alélicos			Observada em <i>trans</i> com uma variante dominante BP2 Observada em <i>cis</i> com uma variante patogênica BP2		Para doenças recessivas, detectada em <i>trans</i> com uma variante patogênica PM3		
Outros Databases			Fonte confiável (dados privados) = benigno BP6	Fonte confiável = patogênico PP5			
Outros dados			Identificada em um caso com uma causa alternativa BP5	Fenótipo do paciente ou família específico para o gene PP4			

Quadro 5: Classificação de variantes segundo os critérios* da ACMG 2015**

<p>Patogênica (Pathogenic)</p> <p>(i) 1 Muito Forte (<i>Very Strong</i> PVS1) <i>E</i></p> <p>(a) ≥ 1 Forte (<i>Strong</i> PS1–PS4) <i>OU</i></p> <p>(b) ≥ 2 Moderada (<i>Moderate</i>) (PM1–PM6) <i>OU</i></p> <p>(c) 1 Moderada (<i>Moderate</i>) (PM1–PM6) <i>E</i> 1 Suporte (<i>Supporting</i>) (PP1–PP5) <i>OU</i></p> <p>(d) ≥ 2 Suporte (<i>Supporting</i>) (PP1–PP5)</p> <p>(ii) ≥ 2 Forte (<i>Strong</i>) (PS1–PS4) <i>OU</i></p> <p>(iii) 1 Forte (<i>Strong</i>) (PS1–PS4) <i>E</i></p> <p>(a) ≥ 3 Moderada (<i>Moderate</i>) (PM1–PM6) <i>OU</i></p> <p>(b) 2 Moderada (<i>Moderate</i>) (PM1–PM6) <i>E</i> ≥ 2 Suporte (<i>Supporting</i>) (PP1–PP5) <i>OU</i></p> <p>(c) 1 Moderada (<i>Moderate</i>) (PM1–PM6) <i>E</i> ≥ 4 Suporte (<i>Supporting</i>) (PP1–PP5)</p> <p>Provavelmente Patogênica (Likely Pathogenic)</p> <p>(i) 1 Muito Forte (<i>Very Strong</i>) (PVS1) <i>E</i> 1 Moderada (<i>Moderate</i>) (PM1–PM6) <i>OU</i></p> <p>(ii) 1 Forte (<i>Strong</i>) (PS1–PS4) <i>E</i> 1–2 Moderada (<i>Moderate</i>) (PM1–PM6) <i>OU</i></p> <p>(iii) 1 Forte (<i>Strong</i>) (PS1–PS4) <i>E</i> ≥ 2 Suporte (<i>Supporting</i>) (PP1–PP5) <i>OU</i></p> <p>(iv) ≥ 3 Moderada (<i>Moderate</i>) (PM1–PM6) <i>OU</i></p> <p>(v) 2 Moderada (<i>Moderate</i>) (PM1–PM6) <i>E</i> ≥ 2 Suporte (<i>Supporting</i>) (PP1–PP5) <i>OU</i></p> <p>(vi) 1 Moderada (<i>Moderate</i>) (PM1–PM6) <i>E</i> ≥ 4 Suporte (<i>Supporting</i>) (PP1–PP5)</p> <p>Benigna (Benign)</p> <p>(i) 1 <i>Stand-Alone</i> (BA1) <i>OU</i></p> <p>(ii) ≥ 2 Forte (<i>Strong</i>) (BS1–BS4)</p> <p>Provavelmente Benigna (Likely Benign)</p> <p>(i) 1 Forte (<i>Strong</i>) (BS1–BS4) <i>E</i> 1 Suporte (<i>Supporting</i>) (BP1–BP7) <i>OU</i></p> <p>(ii) ≥ 2 Suporte (<i>Supporting</i>) (BP1–BP7)</p> <p>Variante de Significado Incerto/Indeterminado (Variant of Unknown/Uncertain Significance) (VUS)</p> <p>Nenhum dos acima ou informações contraditórias</p>

*válidos para doenças monogênicas de penetrância completa

**adaptado de Richards et al., 2015

Além da avaliação de patogenicidade das variantes individualmente, para a confirmação do diagnóstico molecular do paciente é necessário que as variantes encontradas estejam em um gene associado a fenótipo compatível com o apresentado pelo paciente e que o número de variantes encontradas esteja de acordo com o padrão de herança e segregação da doença na família.

3.5. Escolha do teste genético

Os critérios para escolha de qual teste genético realizar inicialmente em cada paciente da casuística de doenças ósseas genéticas visaram à maior probabilidade de diagnóstico com o menor custo. Nos casos em que a hipótese diagnóstica baseada nos dados clínicos apontava para uma única condição relacionada a um único gene pequeno (menos que 10 exons) ou que apresentava *hotspot* mutacional (por exemplo hipocondroplasia – *FGFR3*), a investigação genética foi iniciada por sequenciamento Sanger. Nos casos em que a hipótese diagnóstica apontava para uma única condição ou para um conjunto limitado de diagnósticos diferenciais com genes contidos no painel NGS, quando havia alta confiança na correção da hipótese, a investigação genética foi iniciada por painel de genes NGS. Nos casos em que a hipótese diagnóstica não apontava para uma condição específica, em que havia baixa probabilidade estimada de confirmação das hipóteses principais ou que envolviam condição de etiologia desconhecida, a investigação genética foi iniciada por WES.

3.6. Relato de achados incidentais/secundários

Seguiram-se as recomendações norte-americanas especificadas pela ACMG para o relato de achados incidentais em amostras de análises clínicas. Desta forma, deveriam ser reportadas aos pacientes variantes patogênicas ou provavelmente patogênicas nos genes presentes na lista desta recomendação no momento da análise (56 genes em 2013 e 59 genes em 2016) (Green et al., 2013; Kalia et al., 2017) (Anexo V e Anexo VI).

3.7. Análise estatística

Para cálculo de sensibilidade e especificidade dos testes foi realizada comparação dos resultados de sequenciamento de nova geração com o exame padrão-ouro atual (sequenciamento Sanger), com intervalo de confiança de 95%. O tamanho amostral (N) para avaliar sensibilidade e especificidade foi estimado em 45 pacientes em cada grupo com base no seguinte cálculo: $N = Z * Z (P (1-P)) / (D * D)$; Onde: P= proporção esperada; D= semi-amplitude do intervalo de confiança (utilizado valor de 4%); Z = 1,96 (para $\alpha=0,05$ e IC 95%); assumindo proporção esperada (P) para sensibilidade e especificidade de 98% com base nos trabalhos de Sikkema-Raddatz et al., (2013) e Rattenberry et al., (2013).

Os resultados de sensibilidade diagnóstica (*diagnostic yield*) de cada teste genético foram reportados de maneira descritiva.

4. Resultados

4.1. Validação da técnica de NGS (etapa A)

A técnica do NGS foi aplicada em 49 pacientes com variantes já identificadas por Sanger (Quadro 6). Em dois pacientes (P2455 e CC1074) a técnica de NGS não detectou a variante previamente reportada [*COL18A1*:NM_030582:exon2:c.12-2A>T em homozigose e *BRAF*:NM_004333:exon16:(p.Arg626Lys) em heterozigose]. A amostra com mutação em *BRAF* foi retestada por Sanger e nesse terceiro teste a variante também não foi identificada. Além dos dois resultados falso-negativos, houve um resultado com divergência de anotação na amostra CC2614 [*SOS1*:NM_005633:exon10:c.1655G>C:p.Arg552Thr vs *SOS1*:p.Arg552Lys]. Como a anotação anterior de sequenciamento Sanger continha apenas informação acerca da proteína, não foi possível aferir se o primeiro analisador errou ao anotar a troca de base observada.

Foram realizadas confirmações por Sanger de 81 variantes encontradas por NGS (Quadro 7). Dessas, apenas uma foi considerada resultado falso-positivo [*SOS2*:NM_006939:exon7:c.930T>A:(p.Asn310Lys)] no paciente CC5072. A amostra c10113-III, que apresentava a variante em hemizigose (*DMD*:NM_004006.2:IVS40:c.5740-16T>A), necessitou que a confirmação por Sanger fosse repetida com primers diferentes, pois no primeiro experimento não houve amplificação do produto.

Com os resultados de falso-negativos (2/49) e falso-positivo obtidos (1/81), a sensibilidade da técnica de NGS pode ser estimada em 95,92% (Intervalo de Confiança [IC] 95% = 86,02% a 99,50%) e a especificidade, em 98,77% (IC 95% = 93,31% a 99,97%).

Quadro 6: Lista de variantes conhecidas por Sanger e avaliadas por NGS

Número	Registro	Gene	Transcrito	Variante Sanger				Variante NGS	Obs:
				Exon	cDNA	Proteína	Estado		
1	P1663	CFTR	NM_000492	exon11	c.1521_1523del	p.(Phe508del)	het	Confirmada	
2	P2455	COL18A1	NM_030582	IVS1	c.12-2A>T	? (splicing)	homo	Não identificada	baixa cobertura – região rica em CG
3	C24034	CAPN3	NM_000070	exon4	c.550del	p.(Thr184Argfs*36)	het	Confirmada	
4	C24034	CAPN3	NM_000070	exon10	c.1322del	p.(Gly441Valfs*22)	het	Confirmada	
5	C24416	FKRP	NM_024301	exon4	c.341C>G	p.(Ala114Gly)	het	Confirmada	
6	C24416	FKRP	NM_024301	UTR5	c.1-34C>T	?	het	Confirmada	
7	F5393-1	FGFR2	NM_000141	exon9	c.940-2A>G	? (splicing)	het	Confirmada	
8	F7794-1	TCOF1	NM_000356	exon8	c.885_886del	p.(Ala296Cysfs*47)	het	Confirmada	
9	CC5222	CHST3	NM_004273	exon3	c.673A>G	p.(Ser225Gly)	het	Confirmada	
10	CC5222	CHST3	NM_004273	exon3	c.1148G>C	p.(Arg383Pro)	het	Confirmada	
11	CC5222	CHST3	NM_004273	exon3	c.1070G>A	p.(Arg357Gln)	het	Confirmada	
12	CC5140	COL2A1	NM_033150	exon47	c.3157G>A	p.(G1053Arg)	het	Confirmada	
13	CC5383	COL2A1	NM_033150	exon42	c.2758C>T	p.(Arg920Cys)	het	Confirmada	
14	F8864-1	HES7	NM_032580	exon2	c.123G>A	p.(Glu41Glu)	het	Confirmada	
15	F8474-1	CHST3	NM_004273	exon3	c.1070G>A	p.(Arg357Gln)	het	Confirmada	
16	F9163-1	GDF5	NM_000557	exon2	c.826T>G	p.(Ser276Ala)	het	Confirmada	
17	F8980-1	SMARCA1	NM_014140	exon15	c.2293del	p.(Glu765Argfs*47)	het	Confirmada	
18	CC1437	PTPN11	NM_002834	exon3	c.317A>C	p.(Asp106Ala)	het	Confirmada	
19	CC665	SOS1	NM_005633	exon14	c.2183A>T	p.(Lys728Ile)	het	Confirmada	
20	CC665	SOS1	NM_005633	exon2	c.195A>C	p.(Arg65Arg)	het	Confirmada	
21	CC671	SHOC2	NM_007373	exon2	c.4A>G	p.(Ser2Gly)	het	Confirmada	
22	CC687	KRAS	NM_004985	exon2	c.13A>G	p.(Lys5Glu)	het	Confirmada	
23	CC688	BRAF	NM_004333	exon12	c.1501G>A	p.(Glu501Lys)	het	Confirmada	
24	CC702	PTPN11	NM_002834	exon3	c.184T>G	p.(Tyr62Asp)	het	Confirmada	
25	CC703	BRAF	NM_004333	exon15	c.1741A>G	p.(Asn581Asp)	het	Confirmada	
26	CC712	SHOC2	NM_007373	exon2	c.4A>G	p.(Ser2Gly)	het	Confirmada	
27	CC713	BRAF	NM_004333	exon12	c.1502A>G	p.(Glu501Gly)	het	Confirmada	
28	CC723	PTPN11	NM_002834	exon8	c.854T>C	p.(Phe285Ser)	het	Confirmada	
29	CC723	PTPN11	NM_002834	exon11	c.1232C>T	p.(Thr411Met)	het	Confirmada	
30	CC775	PTPN11	NM_002834	exon13	c.1471C>T	p.(Pro491Ser)	het	Confirmada	
31	CC1437	PTPN11	NM_002834	exon3	c.317A>C	p.(Asp106Ala)	het	Confirmada	
32	CC2614	SOS1	NM_005633	exon10	?	p.(Arg552Lys)	het	NM_005633:exon10: c.1655G>C:p.R552T	anotação divergente
33	CC4385	RAF1	NM_002880	exon7	c.770C>T	p.(Ser257Leu)	het	Confirmada	
34	CC4486	BRAF	NM_004333	exon11	c.1408A>C	p.(Thr470Pro)	het	Confirmada	
35	CC5042	SOS1	NM_005633	exon6	c.806T>C	p.(Met269Thr)	het	Confirmada	
36	CC5060	KRAS	NM_004985	exon2	c.40G>A	p.(Val14Ile)	het	Confirmada	
37	CC5136	BRAF	NM_004333	exon6	c.736G>C	p.(Ala246Pro)	het	Confirmada	
38	CC5188	SHOC2	NM_007373	exon2	c.89C>G	p.(Ala30Gly)	het	Confirmada	
39	CC5324	PTPN11	NM_002834	exon3	c.205G>C	p.(Glu69Gln)	het	Confirmada	
40	CC5313	PTPN11	NM_002834	exon13	c.1504T>G	p.(Ser502Ala)	het	Confirmada	
41	CC5374	PTPN11	NM_002834	exon3	c.182A>G	p.(Asp61Gly)	het	Confirmada	
42	CC5320	HRAS	NM_005343	exon2	c.34G>A	p.(Gly12Ser)	het	Confirmada	
43	CC1512	PTPN11	NM_002834	exon4	c.417G>C	p.(Glu139Asp)	het	Confirmada	
44	CC933	SOS1	NM_005633	exon13	c.2122G>A	p.(Ala708Thr)	het	Confirmada	
45	CC1099	SOS1	NM_005633	exon16	c.2536G>A	p.(Glu846Lys)	het	Confirmada	
46	CC1889	BRAF	NM_004333	exon11	c.1391G>A	p.(Gly464Glu)	het	Confirmada	
47	CC1074	BRAF	NM_004333	exon16	?	p.(Arg626Lys)	het	Não identificada	boa cobertura
48	CC4860	KRAS	NM_004985	exon5	c.531_533del	p.(Lys180del)	het	Confirmada	
49	CC1009	SOS1	NM_005633	exon10	c.1655G>C	p.(Arg552Thr)	het	Confirmada	

Quadro 7: Lista de variantes identificadas por NGS e avaliadas por Sanger.

(parte 1, amostras 1 a 41 de 81)

Número	Registro	Gene	Transcrito	Variante NGS				Variante Sanger	Obs:
				Exon	cDNA	Proteína	Estado		
1	F9568-1	DYNC2H1	NM_001377	IVS35	c.5559-1G>C	? (splicing)	het	Confirmada	
2	F8214-1	COL1A2	NM_000089	exon41	c.2656G>A	p.(Gly886Ser)	het	Confirmada	
3	F9171-1	TRPV4	NM_021625	exon15	c.2389G>A	p.(Glu797Lys)	het	Confirmada	
4	F6025-1	DSC1	NM_004948	exon5	c.559A>G	p.(Ile187Val)	het	Confirmada	
5	F8807-1	HSPG2	NM_005529	exon74	c.9970G>A	p.(G3324Arg)	het	Confirmada	
6	F8123-1	PCYT1A	NM_005017	exon10	c.968dup	p.(S323Argfs*38)	homo	Confirmada	
7	F8069-1	PCYT1A	NM_005017	exon6	c.385G>A	p.(Glu129Lys)	homo	Confirmada	
8	CC2783	LZTR1	NM_006767	exon4	c.356A>G	p.(Tyr119Cys)	het	Confirmada	
9	CC3766	LZTR1	NM_006767	exon8	c.742G>A	p.(Gly248Arg)	het	Confirmada	
10	CC736	LZTR1	NM_006767	exon9	c.850C>T	p.(Arg284Cys)	het	Confirmada	
11	CC5265	LZTR1	NM_006767	exon9	c.859C>T	p.(His287Tyr)	het	Confirmada	
12	CC4671	LZTR1	NM_006767	exon12	c.1341C>A	p.(Phe447Leu)	het	Confirmada	
13	CC1586	LZTR1	NM_006767	exon16	c.1939A>G	p.(Ile647Val)	het	Confirmada	
14	CC1586	SOS2	NM_006939	exon20	c.3275C>T	p.(Pro1092Leu)	het	Confirmada	
15	CC4981	SOS2	NM_006939	exon9	c.1127C>G	p.(Thr376Ser)	het	Confirmada	
16	CC5072	SOS2	NM_006939	exon7	c.930T>A	p.(Asn310Lys)	het	Não identificada	Falso-positivo - Balanço alélico ruim no NGS
17	CC4653	SOS2	NM_006939	exon6	c.800T>A	p.(Met267Lys)	het	Confirmada	
18	CC5208	LRP5	NM_002335	exon14	c.3122C>T	p.(Thr1041Met)	homo	Confirmada	
19	F8413-1	SMARCA1	NM_014140	exon3	c.428T>G	p.(Leu143*)	het	Confirmada	
20	F8413-1	SMARCA1	NM_014140	exon15	c.2293del	p.(Glu765Argfs*47)	het	Confirmada	
21	F9171-1	TRPV4	NM_021625	exon15	c.2389G>A	p.(Glu797Lys)	het	Confirmada	
22	CC5209	FGFR3	NM_000142	exon3	c.188C>T	p.(Pro63Leu)	het	Confirmada	
23	CC245	RIT1	NM_001256820	exon3	c.62C>G	p.(Ala21Gly)	het	Confirmada	
24	CC5054	RIT1	NM_001256820	exon4	c.138T>G	p.(Phe46Leu)	het	Confirmada	
25	P3058	MECP2	NM_004992	exon4	c.502C>T	p.(Arg168*)	het	Confirmada	
26	P3064	MECP2	NM_004992	exon4	c.473C>T	p.(Thr158Met)	het	Confirmada	
27	P3075	MECP2	NM_004992	exon3	c.301C>T	p.(Pro101Ser)	het	Confirmada	
28	P3094	CAPN3	NM_000070	IVS5	c.802-9G>A	? (splicing)	homo	Confirmada	
29	P3098	MECP2	NM_004992	exon4	c.808C>T	p.(Arg270*)	het	Confirmada	
30	P3103	CFTR	NM_000492	exon4	c.350G>A	p.(Arg117His)	het	Confirmada	
31	P3103	CFTR	NM_000492	IVS9	c.1210-12T[7]	?	homo	Confirmada	
32	F9194-1	TWIST1	NM_000474	exon1	c.592A>C	p.(Met198Leu)	het	Confirmada	
33	P3210	MECP2	NM_004992	exon4	c.880C>T	p.(Arg294*)	het	Confirmada	
34	P3216	MECP2	NM_004992	exon4	c.772G>T	p.(Glu258*)	het	Confirmada	
35	P3217	CFTR	NM_000492	exon3	c.220C>T	p.(Arg74Trp)	het	Confirmada	
36	P3217	CFTR	NM_000492	exon6	c.601G>A	p.(Val201Met)	het	Confirmada	
37	P3217	CFTR	NM_000492	exon22	c.3484C>T	p.(Arg1162*)	het	Confirmada	
38	P3217	CFTR	NM_000492	exon23	c.3808G>A	p.(Asp1270Asn)	het	Confirmada	
39	c.26587 I	DMD	NM_004006	exon31	c.4240C>T	p.(Gln1414*)	hemi	Confirmada	
40	c.26609 I	DMD	NM_004006	exon39	c.5551C>T	p.(Gln1851*)	hemi	Confirmada	
41	c.12084	DMD	NM_004006	exon24	c.3268C>T	p.(Gln1090*)	hemi	Confirmada	

Quadro7: Lista de variantes identificadas por NGS e avaliadas por Sanger.

(parte 2, amostras 42 a 81 de 81)

Número	Registro	Gene	Transcrito	Variante NGS				Variante Sanger	Obs:
				Exon	cDNA	Proteína	Estado		
42	c.15518	DMD	NM_004006	exon8	c.822T>A	p.(Tyr274*)	hemi	Confirmada	
43	c.26604 I	DMD	NM_004006	exon9	c.918T>G	p.(Tyr306*)	het	Confirmada	
44	c.26625 I	DMD	NM_004006	exon68	c.9948C>A	p.(Cys3316*)	hemi	Confirmada	
45	c.19865 I	DMD	NM_004006	exon58	c.8652_8653del	p.(Tyr2885Profs*9)	hemi	Confirmada	
46	P2931	DMD	NM_004006	exon30	c.4091T>A	p.(Leu1364*)	hemi	Confirmada	
47	P3250-I	CAPN3	NM_000070	exon5	c.633del	p.(Lys211Asnfs*9)	het	Confirmada	
48	P3250-I	CAPN3	NM_000070	exon5	c.701G>A	p.(Gly234Glu)	het	Confirmada	
49	P3253	MECP2	NM_004992	exon4	c.473C>T	p.(Thr158Met)	het	Confirmada	
50	P2807 I	DMD	NM_004006	exon33	c.4591C>T	p.(Gln1531*)	hemi	Confirmada	
51	P3251	FGFR2	NM_000141	exon8	c.983A>G	p.(Tyr328Cys)	het	Confirmada	
52	P3272	MECP2	NM_004992	exon4b	c.763C>T	p.(Arg255*)	het	Confirmada	
53	c.25103	DMD	NM_004006	exon52	c.7567del	p.(Arg2523Glyfs*15)	hemi	Confirmada	
54	c.26349	DMD	NM_004006	IVS9	c.961-2A>C	? (splicing)	hemi	Confirmada	
55	c.24101	DMD	NM_004006	exon70	c.10171C>T	p.(Arg3391*)	hemi	Confirmada	
56	c.26741 I	DMD	NM_004006	exon67	c.9677_9678del	p.(Thr3226Argfs*4)	hemi	Confirmada	
57	P3316	MECP2	NM_004992	exon4	c.502C>T	p.(Arg168*)	het	Confirmada	
58	P3322	CFTR	NM_000492	exon8	c.1000C>T	p.(Arg334Trp)	het	Confirmada	
59	c.17512 I	DMD	NM_004006	exon17	c.2017C>T	p.(Gln673*)	hemi	Confirmada	
60	c21948-III	DMD	NM_004006	exon33	c.4527T>G	p.(Tyr1509*)	hemi	Confirmada	
61	c25492-III	DMD	NM_004006	exon19	c.2293G>A	p.(Ala765Thr)	hemi	Confirmada	
62	c10113-III	DMD	NM_004006	IVS40	c.5740-16T>A	?	hemi	Confirmada	Allele Dropout – usado primers 40F com 41R
63	c.26806-I	DMD	NM_004006	exon3	c.131dup	p.(Gln45Thrfs*44)	hemi	Confirmada	
64	P3377-I	CFTR	NM_000492	IVS3	c.274-6T>C	?	het	Confirmada	
65	c20099-I	DMD	NM_004006	IVS52	c.7661-2A>G	? (splicing)	hemi	Confirmada	
66	c26572-I	DMD	NM_004006	exon37	c.5159del	p.(Leu1720*)	hemi	Confirmada	
67	c25717-I	DMD	NM_004006	exon31	c.4271T>A	p.(Leu1424*)	hemi	Confirmada	
68	P3378	CAPN3	NM_000070	exon4	c.550del	p.(Thr184Argfs*36)	het	Confirmada	
69	P3378	CAPN3	NM_000070	exon21	c.2242C>T	p.(Arg748*)	het	Confirmada	
70	C.24942-III	DMD	NM_004006	exon44	c.6391C>T	p.(Gln2131*)	hemi	Confirmada	
71	c26786-I	DMD	NM_004006	exon21	c.2665C>T	p.(Arg889*)	hemi	Confirmada	
72	c.26905-I	CAPN3	NM_000070	exon3	c.412dup	p.(Leu138Profs*15)	het	Confirmada	
73	c.26905-I	CAPN3	NM_000070	exon22	c.2362_2363del insTCATCT	p.(Arg788Serfs*14)	het	Confirmada	
74	P3423-I	CFTR	NM_000492	exon15	c.2552G>T	p.(Arg851Leu)	het	Confirmada	
75	P3423-I	CFTR	NM_000492	IVS18	c.2988+1G>A	? (splicing)	het	Confirmada	
76	c-25123-I	DMD	NM_004006	exon65	c.9403C>T	p.(Gln3135*)	hemi	Confirmada	
77	c-26919-I	DMD	NM_004006	exon6	c.488G>A	p.(Trp163*)	hemi	Confirmada	
78	P2522-I	DMD	NM_004006	exon10	c.1132C>G	p.(Gln378Glu)	hemi	Confirmada	
79	P2522-I	DMD	NM_004006	exon55	c.8198_8199del	p.(Glu2733Alafs*11)	hemi	Confirmada	
80	c-26924-I	DMD	NM_004006	exon33	c.4647del	p.(Lys1549Asnfs*11)	hemi	Confirmada	
81	P3453-I	CFTR	NM_000492	exon11	c.1521_1523del	p.(Phe508del)	homo	Confirmada	

4.2. Aplicação da técnica de NGS em doenças ósseas genéticas e RASopatias(etapa B)

4.2.1. Doenças ósseas genéticas

Foram investigados ao todo 173 novos probandos da casuística de doenças ósseas genéticas acompanhados no ICr-HC-FMUSP (144 por NGS e 29 pela técnica do Sanger). O diagnóstico molecular definitivo foi obtido em 107 destes indivíduos e um diagnóstico provável em mais um probando.

4.2.1.1. Sensibilidade diagnóstica dos testes genéticos na casuística de doenças ósseas genéticas

A sensibilidade diagnóstica (*diagnostic yield*) geral na amostra de doenças ósseas genéticas foi de $107/173 = 61,85\%$ (IC 95% = 0,5442 to 0,6876).

No grupo de 29 pacientes nos quais a análise genética foi realizada pela técnica de Sanger, variantes patogênicas ou provavelmente patogênicas foram encontradas em 20 indivíduos (Quadro 8) com uma sensibilidade diagnóstica (*diagnostic yield*) de 68,97% (IC 95% = 0,5063 a 0,8286), valor semelhante ao encontrado anteriormente (de 61%) nos 75 pacientes de doenças ósseas genéticas investigados por Sanger antes do início da tese.

Quadro 8: Variantes e hipóteses diagnósticas dos probandos da casuística de doenças ósseas genéticas investigados por Sanger.

Variantes em negrito foram consideradas patogênicas ou provavelmente patogênicas

No	Registro	Hipótese Diagnóstica	Sanger	Estado
1	CC5427	Desbuquois	CANT1:NM_138793:c.734C>T:p.(Pro245Leu)	homo
2	CC5507	Desbuquois	<i>CANT1</i> normal	
3	CC5490	SED com múltiplas luxações congênicas	CHST3:NM_004273:c.718A>T:p.(Lys240*)	homo
4	F8948-1	Caffey	<i>COL1A1</i> (exon 41) normal	
5	CC5666	Hipocondroplasia	FGFR3:NM_000142:ex13:c.1620C>A:p.(Asn540Lys)	het
6	CC5667	Hipocondroplasia	FGFR3:NM_000142:ex13:c.1620C>A:p.(Asn540Lys)	het
7	F8858-1	Hipocondroplasia	FGFR3:NM_000142:ex13:c.1620C>A:p.(Asn540Lys)	het
8	F9639-1	Hipocondroplasia	FGFR3:NM_000142:ex13:c.1620C>A:p.(Asn540Lys)	het
9	F8897-1	Hipocondroplasia	FGFR3:NM_000142:ex13:c.1620C>A:p.(Asn540Lys)	het
10	F9645-1	Displasia tanatofórica	FGFR3:NM_000142:ex13:c.742:C>T:p.(Arg248Cys)	het
11	CC5533	Acondroplasia	FGFR3:NM_000142:ex9:c.1138G>A:p.(Gly380Arg)	het
12	F9649-1	Acondroplasia	FGFR3:NM_000142:ex9:c.1138G>A:p.(Gly380Arg)	het
13	F9756-1	Acondroplasia	FGFR3:NM_000142:ex9:c.1138G>A:p.(Gly380Arg)	het
14	CC5421	Braquidactilia C	GDF5:NM_000557:c.1198_1200dup:p.(Cys400dup)	het
15	LA	Braquidactilia C	<i>GDF5</i> normal	
16	EF	Displasia oculo-dento-digital	<i>GJA1</i> normal	
17	CC5531	Hajdu-Cheney	NOTCH2:NM_024408:c.6403_6404del:p.(Leu2135Glu fs*2)	homo
18	F8817-1	Hajdu-Cheney	NOTCH2:NM_024408:c.6853C>T:p.(Gln2285*)	het
19	F9895-1	Camurati-Engelmann	TGFB1:NM_000660:c.652C>T:p.(Arg218Cys)	het
20	F9896-1	Camurati-Engelmann	TGFB1:NM_000660:c.652C>T:p.(Arg218Cys)	het
21	F2472-1	Displasia metatrópica	TRPV4:NM_021625:c.2396C>T:p.(Pro799Leu)	het
22	CC5521	SMD tipo Kozlowski	TRPV4:NM_021625:c.2389C>G:p.(Glu797Gln)	het
23	CC5562	SMD tipo Kozlowski	TRPV4:NM_021625:c.2389C>G:p.(Glu797Gln)	het
24	F8979-1	Fuhrman	WNT7A:NM_004625:c.1A>G:p.(Met1?) / c.931C>A:p.(Arg311Ser)	het comp
25	CC5612	Fuhrman	<i>WNT7A</i> normal	
26	CC5633	Fuhrman	<i>WNT7A</i> normal	
27	CC5642	Desbuquois	<i>WNT7A</i> normal	
28	CC5686	Fuhrman	<i>WNT7A</i> normal	
29	F9469-1	Fuhrman	<i>WNT7A</i> normal	

SED = Displasia espondiloepifisária; SMD = Displasia espondilometafisária; het = Heterozigose; het comp = Heterozigose composta; homo = Homozigose; hemi = Hemizigose;

Foi realizada pesquisa por NGS em 144 probandos com hipótese de doença óssea genética. Desses, 125 foram testados por meio de painel com os principais genes de doenças ósseas genéticas e doenças craniofaciais (P1, P2, P3 e P4). Já o WES foi realizado em 19 probandos, dos quais dois foram submetidos ao sequenciamento por painel e em seguida ao WES (CC5209 e F8128-1), ambos sem resultados conclusivos nas duas análises.

Dos 125 probandos testados por painel de genes NGS, o diagnóstico foi conclusivo em 75 e duvidoso em um paciente (F9728-1), o que resulta em uma sensibilidade diagnóstica (*diagnostic yield*) de $75/125 = 60\%$ (IC 95% = 0,5123 a 0,6817) (Quadro 9). No paciente com resultado duvidoso, o qual apresentava uma hipótese clínica de uma doença de padrão de herança autossômico recessivo, foi identificada apenas uma mutação patogênica em heterozigose em um gene (*LFNG*) compatível com a hipótese clínica. É possível que uma segunda mutação não tenha sido detectada pela limitação do método.

Numa análise de CNVs, em quatro pacientes foi possível detectar deleções (uma em homozigose - *LIFR*; e três em heterozigose - *ALPL*, *SHOX* e *TWIST1*) por meio da análise pelo *software* NextGene. A deleção em *LIFR* foi confirmada por *bridge*-PCR, a deleção em *ALPL* foi encontrada em heterozigose composta com uma mutação patogênica de ponto (que foi observada na tabela de variantes como se estivesse em "homozigose") e as deleções em *SHOX* e *TWIST1* foram confirmadas por MLPA.

Em apenas três probandos (F9767-1 Quadro 9 No 65, F9781-1 Quadro 9 No 82 e F10807-1 Quadro 9 No 116) o diagnóstico definitivo obtido após o teste genético foi diferente das hipóteses iniciais para o caso.

Dos 19 probandos testados por WES, o diagnóstico foi conclusivo em 12, o que resultou em sensibilidade diagnóstica (*diagnostic yield*) de 63,16% (IC 95% = 0,4094 a 0,8095) (Quadro 10). Em uma paciente (F6594-1) com hipótese diagnóstica de doença autossômica recessiva, uma mutação provavelmente patogênica foi identificada e uma deleção de alguns exons foi detectada em heterozigose por observação direta do gene candidato no IGV. Essa alteração foi confirmada em laboratório no exterior. Em quatro dos 12 probandos confirmados só foi possível chegar ao diagnóstico definitivo após o teste, o que significa que as hipóteses diagnósticas iniciais não foram confirmadas.

4.2.1.2. Identificação de genes novos na casuística de doenças ósseas genéticas

A probanda F8124-1 apresentava a hipótese de displasia espondilometafisária (SMD) de etiologia desconhecida. Seus pais eram primos de primeiro grau e havia uma prima também acometida, filha de casamento consanguíneo, o que é compatível com um padrão de herança autossômico recessivo. Uma segunda família consanguínea incluída neste estudo, com dois irmãos acometidos, foi avaliada inicialmente no serviço de genética de Ribeirão Preto. Essa família apresentava alterações esqueléticas iguais às da primeira família, associadas a alterações oftalmológicas, com hipótese clínica da displasia espondilometafisária com distrofia de cones e bastonetes (SMD-CRD). Desta forma, na hipótese de que os afetados nas duas famílias apresentavam a mesma doença, à semelhança dos estudos clássicos de ligação, foi realizado WES nas duas famílias (dois pacientes de cada família). O teste resultou em apenas um gene (*PCYT1A*) com mutações raras em homozigose, com provável perda de função, comum a ambas as famílias (Quadro 11). Embora a probanda F8124-1 ainda fosse assintomática em relação a alterações oculares, a investigação com eletroretinograma mostrou uma distrofia leve de cones e bastonetes (SMD-CRD). Dessa maneira foi identificado o gene responsável por uma doença que até então era desconhecido. Esses resultados foram publicados no periódico *American Journal of Human Genetics* em 2014 (Yamamoto et al., 2014).

O probando F8978-1 também apresentava um quadro de SMD, mas com aspecto de "fraturas" nas regiões metafisárias e sem acometimento ocular. A hipótese de SMD tipo "*corner fracture*", também chamada tipo Sutcliffe, foi levantada para o caso. O gene associado a essa displasia esquelética era desconhecido e os casos descritos eram únicos nas famílias, sugerindo um padrão de herança autossômico dominante. Essa circunstância fundamentou a realização de WES no paciente e em seus genitores não afetados. Uma única variante *de novo* foi encontrada em um gene associado a uma proteína do tecido conectivo e um estudo colaborativo foi firmado com um pesquisador no exterior, para a realização de estudos funcionais. Com o

conhecimento da etiologia molecular de SMD tipo Sutcliffe, um segundo probando da casuística desta tese (CC5209) também foi identificado como portador de uma variante provavelmente patogênica em heterozigose no mesmo gene (Quadro 10).

Quadro 9: Variantes, hipóteses diagnósticas e grupo da nosologia dos probandos da casuística de doenças ósseas genéticas investigados por Painei NGS.

Número	Registro	Hipótese Diagnóstica	Grupo da nosologia (confirmados)	Painei	Variantes	Estado
1	CC5222	SED com múltiplas luxações congênitas	4	P1	CHST3:NM_004273:exon3:c.673A>G:p.(Ser225Gly) / CHST3:NM_004273:exon3:c.1148G>C:p.(Arg383Pro)	het comp
2	CC5208	Síndrome da osteoporose-pseudoglioma	25	P1	LRP5: NM_002335:exon14:c.3122C>T:p.(Thr1041Met)	hom o
3	CC5261	Síndrome de Stüve-Wiedemann	18	P1	DELEÇÃO LIFR homozigose exons 15-20	hom o
4	CC5140	SEMD tipo Strudwick	2	P1	COL2A1:NM_033150:exon47:c.3157GA:p.(Gly1053Arg)	het
5	CC5209	SMD		P1	? FGFR3:NM_000142:c.188C>T:p.(Pro63Leu) presente no pai normal	het
6	CC5232	Pseudoacondroplasia	10	P1	COMP:NM_000095:exon13:c.1417_1419del:p.(Asp473del)	het
7	CC5383	SED congênita	2	P1	COL2A1:NM_033150:exon42:c.2758C>T:p.(Arg920Cys)	het
8	CC5536	Displasia diastrófica	4	P1	SLC26A2:NM_000112:exon3:c.835C>T:p.(Arg279Trp) / SLC26A2:NM_000112:exon3:c.2124_2125dup:p.(Phe709Serfs*27)	het comp
9	F8214-1	Síndrome de Ehlers-Danlos / OI tipo I	25	P1	COL1A2:NM_000089:exon41:c.2656G>A:p.(Gly886Ser)	het
10	F8127-1	Hipofosfatasia	26	P1	ALPL:NM_000478:exon6:c.542C>T:p.(Ser181Leu) / (heterozigoto composto desta mutação com deleção do outro alelo)	het comp
11	F8245-1	SED congênita	2	P1	COL2A1:NM_033150:exon42c.2758C>T:p.(Arg920Cys)	het
12	F8413-1	Displasia imuno-óssea de Schimke	13	P1	SMARCAL1:NM_001127207:exon3:c.428T>G:p.(Leu143*) / SMARCAL1:NM_001127207:exon15:c.2293del:p.(Glu765Argfs*47)	het comp
13	F8633-1	Acondroplasia	1	P1	FGFR3:NM_000142:exon9:c.1138G>A:p.(Gly380Arg)	het
14	F8291-1	Displasia geleofísica/acromicrica		P1	Não identificadas variantes patogênicas	
15	F3515-1	Displasia frontonasal		P1	? ALX1: NM_006982:c.541C>A:p.(Gln181Lys)	het
16	F8599-1	Síndrome de Jeune	9	P1	IFT80:NM_020800:exon8:c.721G>C:p.(Gly241Arg) / IFT80:NM_020800:exon18:c.2048G>T:p.(Gly683Val)	het comp

17	F9182-1	Braquiolmia	4	P1	PAPSS2:NМ_004670:exon2:c.145+1G>C / PAPSS2:NМ_004670:exon7:c.809G>A:p.(Gly270Asp)	het comp
18	F9270-1	Progeria	28	P1	LMNA:NМ_170707:exon11:c.1824C>T:p.(Gly608Gly)	het
19	F6033-1	Síndrome de Ehlers-Danlos com escoliose congênita		P1	Não identificadas variantes patogênicas	
20	F9172-1	Síndrome de Rothmund-Thomson		P1	Não identificadas variantes patogênicas	
21	F9543-1	Displasia frontometafisária		P2	Não identificadas variantes patogênicas	
22	F4479-1	Síndrome de Marshall		P2	Não identificadas variantes patogênicas	
23	F8792-1	SED congênita		P2	Não identificadas variantes patogênicas	
24	5583	Pseudoacondroplasia	10	P2	COMP:NМ_000095:exon13:c.1450T>G:p.(Cys484Gly)	het
25	5589	Doença de Caffey		P2	Não identificadas variantes patogênicas	
26	F8212-1	Displasia associada a COL2A1	2	P2	COL2A1:NМ_001844:exon53:c.4139del:p.(Leu1380Argfs*55)	het
27	F8128-1	Condrodisplasia punctata		P2	Não identificadas variantes patogênicas	
28	F8171-1	Síndrome de Ehlers-Danlos musculocontratural		P2	Não identificadas variantes patogênicas	
29	F9230-1	OI tipo III		P2	Não identificadas variantes patogênicas	
30	F5968-1	Síndrome de Marshall	3	P2	COL11A1:NМ_001854:exon50:c.3709-2A>G	het
31	CC5627	SED Tarda	13	P2	TRAPPC2:NМ_014563:IVS2:c.93+5C>A	hemi
32	F8980-1	Displasia imuno-óssea de Schimke	13	P2	SMARCAL1:NМ_014140:exon7:c.1191del:p.(Thr399Argfs*38) / SMARCAL1:NМ_014140:exon15:c.2293del:p.(Glu765Argfs*47)	het comp
33	F8793-1	Síndrome de Jeune	9	P2	DYNC2H1:NМ_001377:exon57:c.9044A>G:p.(Asp3015Gly) / DYNC2H1:NМ_001377:exon74:c.10942T>A:p.(Trp3648Arg)	het comp
34	F8826-1	Síndrome de Jeune		P2	Não identificadas variantes patogênicas	
35	F8821-1	Síndrome de Jeune		P2	Não identificadas variantes patogênicas	
36	F8791-1	SEMD tipo Strudwick	2	P2	COL2A1:NМ_001844:exon26:c.1718G>A:p.(Gly573Asp)	het
37	F8771-1	Displasia Diastrófica	4	P2	SLC26A2:NМ_000112:exon3:c.835C>T:p.(Arg279Trp) / SLC26A2:NМ_000112:exon3:c.1724del:p.(Lys575Serfs*10)	het comp
38	F9171-1	SMD tipo Kozlowsky	8	P2	TRPV4:NМ_021625:exon15:c.2389G>A:p.(Glu797Lys)	het
39	F9162-1	Displasia geleofísica	15	P2	FBN1:NМ_000138:exon42:c.5183C>A:p.(Ala1728Glu)	het
40	F8396-1	Síndrome de Ehlers-Danlos musculocontratural	4	P2	CHST14:NМ_130468:exon1:c.784G>A:p.(Glu262Lys)	hom o
41	F8625-1	Síndrome de Ehlers-Danlos musculocontratural		P2	Não identificadas variantes patogênicas	

42	F8429-1	Mesomelia		P2	deleção em SHOX???	
43	F8611-1	OI	25	P2	COL1A2:NM_000089:exon35:c.2116_2117insCTCCTGCTG:p.(Ala705_Gly706insAlaProAla)	het
44	F8760-1	OI	25	P2	COL1A2:NM_000089:exon49:c.3278G>A:p.(Gly1093Asp)	het
45	F9163-1	Displasia "Grebe-Like"		P2	Não identificadas variantes patogênicas	
46	F8864-1	Disostose espondilocostal		P2	Não identificadas variantes patogênicas	
47	F8810-1	Sinostose espôndilo carpotarsal		P2	Não identificadas variantes patogênicas	
48	F9227-1	mãe de RN falecido com hipótese de OPD	7	P2	FLNA:NM_001456:exon3:c.583G>A:p.(Gly195Ser)	het
49	F9246-1	Disostose mandíbulo-facial		P2	Não identificadas variantes patogênicas	
50	F9247-1	Esclerose óssea generalizada + mieloma múltiplo		P2	Não identificadas variantes patogênicas	
51	F9256-1	Esclerosteose	24	P2	SOST:NM_025237:exon2:c.372G>A:p.Trp124*	hom o
52	F9483-1	Síndrome de Saethre-Chotzen	33	P2	deleção TWIST1	het
53	F9517-1	Displasia craniometafisária		P2	Não identificadas variantes patogênicas	
54	F9519-1	Ateleosteogênese	7	P2	MOSAICO (20%) - FLNB:NM_001457:exon3:c.596G>C:p.(Arg199Pro)	mosa ico
55	F9532-1	Displasia Diastrófica	4	P2	SLC26A2:NM_000112:exon2:c.532C>T:p.(Arg178*) / SLC26A2:NM_000112:exon3:c.835C>T:p.(Arg279Trp)	het comp
56	F9533-1	Craniossinostose síndrômica		P2	Não identificadas variantes patogênicas	
57	F9534-1	Síndrome craniofrontonasal		P2	Não identificadas variantes patogênicas	
58	F9666-1	SMD Kozlowski - Metatrópica		P2	Não identificadas variantes patogênicas	
59	F9714-1	Hipocondroplasia	1	P2	FGFR3:NM_000142:exon7:c.746C>T:p.(Ser249Phe)	het
60	F9715-1	Displasia epifisária múltipla	10	P2	COMP:NM_000095:exon9:c.887C>T:p.(Pro296Leu)	het
61	F9727-1	Síndrome de Loeys-Dietz	30	P2	TGFBR2:NM_003242:exon7:c.1582C>A:p.(Arg528Ser)	het
62	F9737-1	Síndrome de Marshall		P2	Não identificadas variantes patogênicas	
63	F9763-1	Fibrocondrogênese	3	P2	COL11A1:NM_001854:exon1:c.75_76del:p.(Phe26Profs*5)	hom o
64	F9175-1	Displasia frontonasal		P2	? EFNB1:NM_004429:exon4:c.509C>T:p.(Ala170Val)	hemi
65	F9767-1	Displasia tanatofórica? → OI tipo II	25	P2	COL1A2:NM_000089:exon42:c.2720G>A:p.(Gly907Asp)	het
66	F9779-1	Acondrogênese tipo 2	2	P2	COL2A1:NM_001844:exon46:c.3221G>A:p.(Gly1074Asp)	het
67	OCD37	Esclerosteose		P2	Não identificadas variantes patogênicas	

68	F8759-1	Progeria		P2	Não identificadas variantes patogênicas	
69	F9667-1	Hipocondroplasia		P3	? COL9A3:NM_001853:exon30:c.1726G>A:p.(Ala576Thr)	het
70	F9159-1	Disostose espondilocostal + acidemia metilmalônica		P3	MMAA:NM_172250:exon2:c.397C>T:p.(Gln133*) (Não identificada a responsável pela disostose)	hom o
71	F8277-1	Disostose espondilocostal		P3	Não identificadas variantes patogênicas	
72	F8824-1	Disostose espondilocostal		P3	Não identificadas variantes patogênicas	
73	F8439-1	Disostose espondilocostal		P3	Não identificadas variantes patogênicas	
74	F8415-1	Disostose espondilocostal		P3	Não identificadas variantes patogênicas	
75	F2693-1	Disostose espondilocostal	35	P3	DLL3:NM_016941:exon4:c.599_603dup:p.(Pro202Alafs*41) / DLL3:NM_016941:exon7:c.1096G>A:p.(Gly366Arg)	het comp
76	F8860-1	Disostose espondilocostal		P3	Não identificadas variantes patogênicas	
77	F9226-1	Sequência de Klippel-Feil+deformidade de Sprengel		P3	? SALL1:NM_002968:exon2:c.787G>T:p.(Asp263Tyr)	
78	F9544-1	Hipocondroplasia	1	P3	FGFR3:NM_000142:exon8:c.940G>A:p.(Ala314Thr)	het
79	F9701-1	Hipocondroplasia	1	P3	FGFR3:NM_000142:exon7:c.833A>G:p.(Tyr278Cys)	het
80	F9728-1	Disostose espondilocostal	35	P3	LNFG:NM_001166355:exon2:c.219+1G>A	het
81	F9742-1	Disostose espondilocostal	35	P3	DLL3:NM_016941:exon4:c.517T>C:p.(Ser173Pro) / DLL3:NM_016941:exon7:c.1096G>A:p.(Gly366Arg)	het
82	F9781-1	COL2A1? → Acromesomelia tipo Maroteaux	16	P3	NPR2:NM_003995:exon1:c.131G>A:p.(Arg44Gln)	hom o
83	F9888-1	OI tipo III	25	P3	COL1A2:NM_000089:exon47:c.3142G>T:p.(Gly1048Cys)	het
84	F9905-1	Disostose espondilocostal	35	P3	DLL3:NM_016941.3:exon4:c.599_603dup:p.(Pro202Alafs*41)	hom o
85	OCD34	SMD Kozlowski - Metatrópica	8	P3	TRPV4:NM_021625:exon15:c.2389G>C:p.(Glu797Gln)	het
86	OCD2363	SMD Kozlowski - Metatrópica	8	P3	TRPV4:NM_021625:exon12:c.1847G>A:p.(Arg616Gln)	het
87	F9941-1	Displasia Diastrófica	4	P3	SLC26A2:NM_000112:exon3:c.835C>T:p.(Arg279Trp) / SLC26A2:NM_000112:exon3:c.2124_2125dup:p.(Phe709Serfs*27)	het comp
88	F9931-1	OI grave		P3	Não identificadas variantes patogênicas	
89	F9956-1	OI grave		P3	Não identificadas variantes patogênicas	
90	F7307-1	OI grave	25	P3	COL1A2:NM_000089:exon24:c.1361G>T:p.(Gly454Val)	het
91	F9982-1	OI grave	25	P3	COL1A2:NM_000089:exon38:c.2314G>A:p.(Gly772Ser)	hom o

92	F9994-1	Disostose espondilocostal		P3	Não identificadas variantes patogênicas	
93	F10054-1	Síndrome de CURRARINO		P3	Não identificadas variantes patogênicas	
94	F10177-1	Ateleosteogênese tipo 1	7	P3	FLNB:NM_001457:exon3:c.605T>C:p.(Met202Thr)	het
95	F10207-1	Disostose espondilocostal	35	P3	DLL3:NM_016941:exon7:c.1094-2A>T	hom o
96	F10254-1	Displasia associada a <i>COL2A1</i> / MPS tipo IV	27	P3	GALNS:NM_000512:exon2:c.181C>T:p.(Arg61Trp)	hom o
97	F10860-1	Síndrome de Rothmund-Thomson		P3	Não identificadas variantes patogênicas	
98	F10862-1	SMD Kozlowski - Metatrópica	8	P3	TRPV4:NM_021625:exon15:c.2396C>T:p.(Pro799Leu)	het
99	F10998-1	Pseudoacondroplasia	10	P4	COMP:NM_000095:exon13:c.1318G>C:p.(Gly440Arg)	het
100	F10861-1	Displasia acromesômica de Du Pan	16	P4	GDF5:NM_000557:exon2:c.1198_1200dup:p.(Cys400dup)	het
101	F9321-1	Síndrome de Rothmund-Thomson	39	P4	RECQL4:NM_004260:exon16:c.2547_2548del:p.(Phe850Profs*33) / RECQL4:NM_004260:exon14:c.2269C>T:p.(Gln757*)	het comp
102	F9782-1	Displasia associada a COL2A1	2	P4	COL2A1:NM_001844:exon17:c.1043G>T:p.(Gly348Val)	het
103	F10262-1	OI	25	P4	COL1A2:NM_000089:exon11:c.497G>A:p.(Gly166Asp)	het
104	F10273-1	Espondilocarpotarsal? / Disrupção?		P4	Não identificadas variantes patogênicas	
105	F10304-1	OI tipo III	25	P4	P3H1:NM_022356:exon4:c.838C>T:p.(Gln280*)	hom o
106	F10308-1	Displasia Sattan		P4	Não identificadas variantes patogênicas	
107	F10397-1	SED / Desbuquois	20	P4	CANT1:NM_138793:exon2:c.344_345del:p.(Thr115Argfs*67)	hom o
108	F10424-1	Dyggve Melchior Clausen	13	P4	DYM:NM_017653:exon15:c.1624T>C:p.(Cys542Arg) / DYM:NM_017653:exon10:c.1000del:p.(Ser334Valfs*66)	het comp
109	F10329-1	SED		P4	Não identificadas variantes patogênicas	
110	F10538-1	Espondilocarpotarsal / Sinostoses múltiplas	7	P4	FLNB:NM_001457:exon13:c.1945C>T:p.(Arg649*)	het
111	NACP- ma	Displasia diastrófica	4	P4	SLC26A2:NM_000112:exon2:c.532C>T:p.(Arg178*) / SLC26A2:NM_000112:exon3:c.835C>T:p.(Arg279Trp)	het comp
112	F10575-1	Síndrome de Rothmund-Thomson	39	P4	deleção 11bp IVS8 RECQL4 chr8 145740497-145740507 GCCTCAGCCCAdeI	hom o
113	F10576-1	SED / MPS tipo IV		P4	Não identificadas variantes patogênicas	
114	F10629-1	Síndrome de Rothmund-Thomson	39	P4	deleção 11bp IVS8 RECQL4 chr8 145740497 145740507 GCCTCAGCCCAdeI	hom o

115	F10751-1	SED com múltiplas luxações	4	P4	CHST3:NM_004273:exon3:c.141-1G>C / CHST3:NM_004273:exon3:c.660C>A:p.(Phe220Leu)	het comp
116	F10807-2 F10807-3	mãe e pai de F10807-1 displasia letal → Bruck tipo 2	25	P4	PLOD2:NM_000935:exon15:c.1619G>A:p.(Trp540*)	het
117	F4474-1	síndrome progeróide marfanóide		P4	Não identificadas variantes patogênicas	
118	F10308-1	FGFR3 mosaico		P4	Não identificadas variantes patogênicas	
119	F10610-1	SED congênita	2	P4	COL2A1:NM_001844:exon27:c.1754G>T:p.(Gly585Val)	het
120	F10539-1	Pseudoacondroplasia - displasia epifisária múltipla	10	P4	COMP:NM_000095:exon10:c.1052G>T:p.(Cys351Phe)	het
121	F10533-1	Displasia campomélica	18	P4	SOX9:NM_000346:exon2:c.507C>G:p.(His169Gln)	het
122	P4201	Acrodisostose	15	P4	PRKAR1A:NM_002734:exon11:c.1004G>C:p.(Arg335Phe)	het
123	F10845-1	Displasia associada a <i>COL2A1</i>		P4	Não identificadas variantes patogênicas	
124	F8600-1	Blefaroqueilodôntica? / AEC?	40	P4	TP63:NM_003722:exon14:c.1765A>T:p.(Ile589Phe)	het
125	F11146-1	Síndrome de Rothmund-Thomson		P4	Não identificadas variantes patogênicas	

AEC = Ankyloblepharon-Ectodermal defects-Cleft lip/palate; MPS = Mucopolissacaridose; OI = Osteogenese imperfeita; OPD = OtoPalatoDigital; SED = displasia espondiloepifisária; SEMD = displasia espondiloepimetáfisária; SMD = displasia espondilometáfisária; ? = hipótese duvidosa;

/ = separa duas hipóteses (hipótese em **negrito** confirmada pelo teste);

→ = hipótese após teste;

Variantes em **negrito** foram classificadas como patogênicas ou provavelmente patogênicas; Variantes com ? foram classificadas como de significado incerto/indeterminado (VUS)

het = Heterozigose; het comp = Heterozigose composta; homo = Homozigose; hemi = Hemizigose;

Quadro 10: Variantes e hipóteses diagnósticas dos probandos da casuística de doenças ósseas genéticas investigados por exoma completo (WES).

Número	DNA	Hipótese Diagnóstica	Exoma completo (WES) NGS	Estado
1	F8474-1	Síndrome orelha-patela?	Não identificadas variantes patogênicas	
2	F5389-1	Displasia frontonasal	Não identificadas variantes patogênicas	
3	CC5209	SMD tipo Sutcliffe	Mutação missense em gene novo	het
4	F8124-1	SMD → SMD CRD	PCYT1A:NM_005017:exon9:c.968dup:(p.Ser323Argfs*38)	homo
5	F9568-1	OFD tipo IV	DYNC2H1:NM_001377:exon36:c.5559-1G>C / DYNC2H1:NM_001377:exon56:c.8888_8890del:p.(Glu2963del)	het comp
6	CC5568	Displasia associada a COL2A1 → Unha patela	LMX1B:NM_002316:exon1:c.139+1G>C	het
7	F8128-1	Condrodisplasia punctata	Não identificadas variantes patogênicas	
8	F8978-1	SMD? / SMD tipo Sutcliffe?	Mutação missense em gene novo	het
9	F8539-1	Displasia acromesomélica	Não identificadas variantes patogênicas	
10	F6025-1		Não identificadas variantes patogênicas	
11	F8613-1	Feingold?	Não identificadas variantes patogênicas	
12	F8807-1	Displasia de Rolland-Desbuquois	HSPG2:NM_005529:exon74:c.9970G>A:p.(Gly3324Arg) / HSPG2:NM_005529:exon12:c.1429_1430insGGCCGTGGCACACTGATC ATCCGTGATGTGAAGGAGTCAGACCAGGGTGCCTT:p.(Tyr477Trpfs*9)	
13	F6594-1	Síndrome progeróide neonatal	POLR3A:NM_007055:exon26:c.3392A>G:p.(Lys1131Arg) / deleção de 3 exons no outro alelo	het comp
14	F9893-1	Condrodisplasia punctata rizomélica	PEX7:NM_000288:exon7:c.649G>A:p.(Gly217Arg)	homo
15	F9992-1	Defeito de segmentação vertebral	Não identificadas variantes patogênicas	
16	F10865-1	Desbuquois II? → SED tipo Kimberley	ACAN:NM_001135:exon13:c.6923G>A:p.(Arg2308His) / Intrônica: chr15:89414577C>A	het comp
17	F9278-3	Síndrome genitopatelar	KAT6A:NM_006766:exon17:c.3385C>T:p.(Arg1129*)	het
18	F9187-1	anomalia do carpo → OCM	MAFB:NM_005461:exon1:c.211C>G:p.(Pro71Ala)	het
19	580_YS	anomalia do carpo → OCM	MAFB:NM_005461:exon1:c.176C>T:p.(Pro59Leu)	het

OCM = Osteólise carpotarsal multicêntrica; OFD = OroFacioDigital; SED = displasia espondiloepifisária; SMD = displasia espondilometafisária; CRD = com distrofia de cones e bastonetes; ? = hipótese duvidosa; / = separa duas hipóteses (hipótese em **negrito** confirmada pelo teste); → = hipótese após teste; Variantes em **negrito** foram classificadas como patogênica ou provavelmente patogênica; het = Heterozigose; het comp = Heterozigose composta; homo = Homozigose; hemi = Hemizigose;

Quadro 11: Variantes raras em homozigose identificadas por WES nas famílias com displasia espondilometafisária com distrofia de cones e bastonetes.

Número de variantes em ambos afetados da Família 1 (F8124)		17341	
Variantes em homozigose		5243	
Raras (<0.01 freq. alélica em controles externos)		326	
<0.05 freq. alélica em controles internos		17	
Gene	Variante	Tipo	dbSNP137
<i>PNRC2</i>	NM_017761:c.237A>C:p.(Ser79Ser)	<i>sense</i>	rs10493018
<i>ATP13A5</i>	NM_198505:c.1831G>A:p.(Ala611Thr)	<i>missense</i>	rs143477252
<i>ATP13A4</i>	NM_032279:c.2895G>T:p.(Arg965Arg)	<i>sense</i>	rs150429290
<i>MUC4</i>	NM_018406:c.10220T>C:p.(Met3407Tjr)	<i>missense</i>	.
<i>MUC4</i>	NM_018406:c.8733G>C:p.(Thr2911Thr)	<i>sense</i>	rs28385894
<i>MUC4</i>	NM_018406:c.8405C>T:p.(Pro2802Leu)	<i>missense</i>	.
<i>MUC4</i>	NM_018406:c.7320G>A:p.(Ser2440Ser)	<i>sense</i>	rs71321826
<i>MUC4</i>	NM_018406:c.7313T>C:p.(Val2438Ala)	<i>missense</i>	.
<i>MUC4</i>	NM_018406:c.7312G>A:p.(Val2438Ile)	<i>missense</i>	.
<i>MUC4</i>	NM_018406:c.7295G>C:p.(Arg2432Pro)	<i>missense</i>	.
<i>MUC4</i>	NM_018406:c.7275T>A:p.(Ala2425Ala)	<i>sense</i>	.
<i>MUC4</i>	NM_018406:c.6111T>C:p.(Asp2037Asp)	<i>sense</i>	rs199853476
<i>MUC4</i>	NM_018406:c.6096G>T:p.(Pro2032Pro)	<i>sense</i>	.
<i>MUC4</i>	NM_018406:c.6081T>A:p.(Thr2027Thr)	<i>sense</i>	.
PCYT1A	NM_005017:exon9:c.968dup:(p.Ser323Argfs*38)	frameshift	.
<i>RNF168</i>	NM_152617:c.1481C>T:p.(Pro494Leu)	<i>missense</i>	rs35132476
<i>POTEG</i>	NM_001005356:c.869T>A:p.(Phe290Tyr)	<i>missense</i>	rs200845849
Número de variantes em ambos afetados da Família 2		20160	
Variantes em homozigose		6873	
Raras (<0.01 freq. alélica em controles externos)		307	
<0.05 freq. alélica em controles internos		18	
Gene	Variante	Tipo	dbSNP137
<i>APOD</i>	NM_001647:c.383A>G:p.(Tyr128Cys)	<i>missense</i>	rs150664052
<i>PCYT1A</i>	NM_005017:c.996C>G:p.(Pro332Pro)	<i>sense</i>	rs61737911
PCYT1A	NM_005017:c.385G>A:p.(Glu129Lys)	missense	.
<i>FAM170A</i>	NM_001163991:c.665G>T:p.(Gly222Val)	<i>missense</i>	rs200846075
<i>USP17L8</i>	NM_001256872:c.534C>T:p.(Pro178Pro)	<i>sense</i>	.
<i>USP17L8</i>	NM_001256872:c.146T>A:p.(Val49Asp)	<i>missense</i>	rs201927052
<i>IL33</i>	NM_001199641:c.273C>T:p.(Phe91Phe)	<i>sense</i>	.
<i>CCDC73</i>	NM_001008391:c.3050T>A:p.(Ile1017Lys)	<i>missense</i>	rs199847743
<i>DACT1</i>	NM_001079520:c.1193T>C:p.(Leu398Pro)	<i>missense</i>	rs114859205
<i>SYNE2</i>	NM_015180:c.15021C>T:p.(Ala5007Ala)	<i>sense</i>	rs148113884
<i>PPP1R36</i>	NM_172365:c.1016T>C:p.(Val339Ala)	<i>missense</i>	rs149477152
<i>PPP1R36</i>	NM_172365:c.1166T>C:p.(Phe389Ser)	<i>missense</i>	rs149234804
<i>PLEKHG3</i>	NM_015549:c.461G>A:p.(Arg154Gln)	<i>missense</i>	rs149822835
<i>PLEK2</i>	NM_016445:c.588G>T:p.(Val196Val)	<i>sense</i>	.
<i>ARG2</i>	NM_001172:c.1051C>T:p.(Arg351Cys)	<i>missense</i>	rs112748595
<i>ZFYVE26</i>	NM_015346:c.5799C>T:p.(Ser1933Ser)	<i>sense</i>	rs145858625
<i>CBLN2</i>	NM_182511:c.20G>A:p.(Gly7Glu)	<i>missense</i>	.
<i>KIR3DL3</i>	NM_153443:c.229C>T:p.(Arg77Trp)	<i>missense</i>	rs202170060

4.2.2.RASopatias

4.2.2.1. Sensibilidade diagnóstica de WES na casuística de RASopatias

Foram investigados por WES 79 probandos com suspeita de RASopatias nos quais os principais genes e mutações associados às formas mais comuns de RASopatias já haviam sido excluídos por sequenciamento Sanger. Em 36 probandos foi identificada uma variante patogênica ou provavelmente patogênica que confirma o diagnóstico do paciente, com uma sensibilidade diagnóstica (*diagnostic yield*) de 36/79 = 45,57 % (IC 95% = 0,3505 to 0,5650). Destes, 33 foram confirmados com um dos tipos de RASopatias, enquanto que três foram identificados com diagnósticos diferenciais de RASopatias (Quadro 12).

4.2.2.2. Identificação de genes novos

Analisando 50 probandos com hipótese de síndrome de Noonan nos quais nenhuma variante patogênica havia sido encontrada nos genes conhecidos até 2015, foi possível identificar dois novos genes nos quais variantes, ou estavam presentes em um determinado gene em mais de um paciente na casuística de SN (*LZTR1*), ou estavam presentes em um gene sabidamente pertencente à via RAS/MAPK (*SOS2*). Para ambos os genes, as variantes supostamente patogênicas segregavam com pacientes acometidos em pelo menos duas famílias. Além disto, em pelo menos um caso isolado (tanto de *LZTR1* como de *SOS2*) foi provado que o evento era *de novo*. Colaboradores dos EUA e Polônia contribuíram com mais uma família cada um, e o trabalho foi publicado no periódico *Journal of Medical Genetics* em 2015 (Figura 3) (Yamamoto et al., 2015).

Quadro 12: Variantes e hipótese dos probandos de RASopatias investigados por WES.

Número	Registro	Hipótese	Variantes em Heterozigose identificadas por WES
1	CC3766	Noonan	LZTR1 - NM_006767:exon8:c.742G>A:p.(Gly248Arg)
2	CC245	Noonan	RIT1 - NM_001256820:exon3:c.62C>G:p.(Ala21Gly)
3	CC736	Noonan	LZTR1 - NM_006767:exon9:c.850C>T:p.(Arg284Cys)
4	CC1586	Noonan	
5	CC2783	Noonan	LZTR1 - NM_006767:exon4:c.356A>G:p.(Tyr119Cys)
6	CC4981	Noonan	SOS2 - NM_006939:exon9:c.1127C>G:p.(Thr376Ser)
7	CC5100	Noonan	NRAS - NM_002524:exon5:c.179G>A:p.(Gly60Glu)
8	CC5265	Noonan	LZTR1 - NM_006767:exon9:c.859C>T:p.(His287Tyr)
9	CC883	CFC	KRAS - NM_004985:exon5:c.458A>T:p.(Asp153Val)
10	CC722	Noonan	PTPN11- NM_002834:exon4:c.A172C:p.N58H
11	CC726	Noonan	PTPN11- NM_002834:exon6:c.236A>G:p.(Gln79Arg)
12	CC941	Noonan	PTPN11 - NM_002834:exon4:c.124A>G:p.(Thr42Ala)
13	CC5399	Noonan	RIT1 - NM_001256820:exon4:c.176G>C:p.(Gly59Ala)
14	CC2492	CFC	BRAF - NM_004333:exon11:c.1390G>A:p.(Gly464Arg)
15	CC1216	Noonan	
16	CC4653	Noonan	SOS2 - NM_006939:exon6:c.800T>A:p.(Mey267Lys)
17	CC3682	Noonan	? HRAS - NM_176795:exon5:c.500C>T:p.(Pro167Leu)
18	CC5219	Noonan	RAF1 - NM_002880:exon7:c.782C>T:p.(Pro261Leu)
19	CC5099	Noonan	
20	CC4671	Noonan	? LZTR1 - NM_006767:exon12:c.1341C>A:p.(Phe447Leu)
21	CC493	Noonan	RIT1 - NM_001256820:exon3:c.62C>G:p.(Ala21Gly)
22	CC3324	Noonan	
23	CC1547	Noonan	
24	CC1246	Noonan	
25	CC956	Noonan	RIT1 - NM_001256820:exon4:c.176G>C:p.(Gly59Ala)
26	CC5105	NFNS	NF1 - NM_000267:exon28:c.C3721T:p.R1241X
27	CC5200	Noonan	
28	CC3770	Noonan	PTPN11 - NM_002834:exon13:c.A1510G:p.M504V
29	CC4640	Noonan	RIT1 - NM_001256821:exon2:c.G155C:p.S52T
30	CC1264	Noonan	
31	CC962	Noonan	
32	CC5198	Noonan	
33	CC5284	Noonan	
34	CC5114	Noonan	
35	CC5059	CFC	MAP2K1 - NM_002755:exon3:c.389A>G:p.(Tyr130Cys)
36	CC5098	CFC?	? MAP2K2 - NM_030662:exon1:c.86C>A:p.(Ala29Asp)
37	CC1599	Noonan	
38	CC893	CFC	BRAF - NM_004333:exon12:c.1502A>G:p.(Glu501Gly)
39	CC963	CFC	
40	CC5097	Noonan	? MYBPC3 - NM_000256:exon14:c.1321G>A:p.(Glu441Lys)

41	CC5299	Noonan	
42	CC4902	Noonan	PTPN11 - NM_002834:exon5:c.G181A:p.D61N
43	CC5119	Noonan	
44	CC5064	Noonan	
45	CC4965	Noonan	
46	CC1471	Noonan → Schuurs-Hooijmakers (RMD 17)	PACS1 - NM_018026:exon4:c.607C>T:p.(Arg203Trp)
47	CC3533	Noonan	
48	CC1169	Noonan	PTPN1 - :NM_002834:exon3:c.178G>T:p.(Gly60Cys)
49	CC884	Noonan	
50	CC1162	Noonan	
51	CC804	Noonan	
52	CC5305	CFC → RMD 32	KAT6A - NM_006766:exon17:c.3385C>T:p.(Arg1129*)
53	CC5120	Noonan	
54	CC5071	Noonan → Helsmoortel-van der Aa (RMD 28)	ADNP - NM_015339:exon5:c.2491_2494del:p.(Leu831Ilefs*82)
55	CC5308	CFC	
56	CC4388	Noonan	PTPN11 - NM_002834:exon13:c.1510A>G:p.(Met504Val)
57	CC3546	Noonan	
58	CC1980	Noonan	? MYBPC3 - NM_000256:exon22:c.2179G>A:p.(Val727Met)
59	CC851	Noonan / King	
60	CC1139	CFC	? A2ML1 - NM_144670:exon15:c.1829G>A:p.(Arg610His)
61	CC5054	Noonan	RIT1 - NM_001256820:exon4:c.138T>G:p.(Phe46Leu)
62	CC5407	Noonan	PTPN11 - NM_002834:exon4:c.417G>T:p.(Glu139Asp)
63	CC5133	Noonan	
64	CC5072	Noonan	RASA1 - NM_002890:exon21:c.2698_2701del:p.(Val900Phefs*10)
65	CC815	Noonan	
66	CC4565	Noonan / Marfan	
67	CC5035	Noonan	
68	CC2016	Noonan	PTPN11 - NM_002834:exon8:c.923A>G:p.(Asn308Ser)
69	CC1371	Noonan	
70	CC1155	Noonan	
71	CC5188	Noonan	
72	CC5088	NF1	NF1 - NM_000267:exon30:c.3989_3992del:p.(Glu1330Alafs*12)
73	CC4996	Noonan	
74	CC4602	CFC	NRAS - NM_002524:exon3:c.149C>T:p.(Thr50Ile)
75	CC3680	Noonan	
76	CC2207	Noonan	PTPN11 - NM_002834:exon4:c.473G>A:p.(Gly158Glu)
77	CC1413	Noonan	
78	CC1158	Noonan	PTPN11 - NM_002834:exon13:c.1507G>A:p.(Gly503Arg)
79	CC5217	Noonan	PPP1CB - NM_002709:exon2:c.146C>G:p.(Pro49Arg)

RMD = Retardo mental autossômico dominante; → = hipótese após teste; **Negrito**= patogênicas ou provavelmente patogênicas; ? = significado incerto/indeterminado (VUS)

Famílias *SOS2* Famílias *LZTR1*

Figura 4: Identificação de variantes nos genes *SOS2* e *LZTR1* como etiologia de síndrome de Noonan em múltiplas famílias da casuística de RASopatias (Yamamoto et al., 2015)

4.2.3. Relato de achados incidentais

Não foi encontrada nenhuma variante patogênica ou provavelmente patogênica nos genes contidos na lista de achados secundários da ACMG nos 19 probandos de doenças ósseas genéticas investigados por WES.

4.3. Aplicação da técnica de NGS em amostras sem diagnóstico (etapa C)

4.3.1. Análise de amostras sem conhecimento da hipótese diagnóstica clínica (simulação de Triagem Neonatal)

Na análise de 90 indivíduos sem hipótese clínica *a priori*, as variantes de cada indivíduo foram classificadas como patogênica, provavelmente patogênica ou VUS de acordo com os critérios da ACMG (Quadro 13). Para classificar o indivíduo como positivo na triagem cega, foi considerada a presença de variante patogênica ou provavelmente patogênica em um alelo, nos casos de genes associados a doenças de herança dominante, ou variante em hemizigose nos meninos para genes associados a doenças recessivas ligadas ao X. Para as variantes identificadas em genes associados a doenças autossômicas recessivas, as amostras foram analisadas de dois modos: a) amostras classificadas como positivas na presença de duas variantes patogênicas ou provavelmente patogênicas em um mesmo gene; b) amostras classificadas como positivas na presença de uma variante patogênica ou provavelmente patogênica e uma segunda variante VUS em um mesmo gene. Após a desidentificação das amostras e a avaliação de melhor relação entre sensibilidade e especificidade das duas possibilidades de análise para doenças autossômicas recessivas, foi optado pela classificação descrita no modo a.

4.3.1.1. Sensibilidade diagnóstica de NGSna casuística de amostras “cegas”

Foram encontrados resultados verdadeiro-positivos (coincidentes com as hipóteses diagnósticas) em 22 dos 45 casos, o que corresponde a uma sensibilidade diagnóstica (*diagnostic yield*) total de 48,89% (IC 95% = 0,3496 a 0,6300). Variantes do tipo CNV foram encontradas em duas amostras, ambas com deleções de exons consideradas patogênicas (Quadro 13).

Em um controle com hipótese clínica da síndrome de Marfan foi encontrado um resultado verdadeiro positivo com duas variantes patogênicas ou provavelmente patogênicas em um gene responsável por uma doença considerada um diagnóstico diferencial da hipótese inicial. Portanto, foi encontrado resultado falso-positivo em um de 44 controles, o que corresponde a uma especificidade de 97,73% (IC 95% = 0,8430 a >0,9999) (Quadro 14).

Observando a sensibilidade diagnóstica apenas nos casos com hipótese de erros inatos do metabolismo, 19 de 26 casos foram corretamente diagnosticados, ou seja, 73,08% (IC 95% = 0,4615 to 0,9148). Os resultados resumidos estão descritos no Quadro 14 e os resultados completos com as variantes encontradas são apresentados no Quadro 13.

Quadro 13: Hipótese clínica, hipótese do teste cego e variantes encontradas por painel NGS em amostras analisadas sem conhecimento da hipótese diagnóstica clínica.

Molecular prévio	Registro	Hipótese Desconhecida	Hipótese do teste cego; número OMIM; padrão de herança	Variante 1 (Estado: homozigose; heterozigose; hemizigose)	Variante 2 (Estado: homozigose; heterozigose; hemizigose)	Observa ção	Sexo (hemizi gose chrX e SRY)
Sim	SGM1	Doença de McArdle	Doença de McArdle; 232600; Autossômico recessivo	homo <i>PYGM</i> :NM_005609:exon1:c.148C>T :p.(Arg50*)			M
Sim	SGM2	Xantomatose Cerebrotendínea	Xantomatose Cerebrotendínea; 213700; Autossômico recessivo	homo <i>CYP27A1</i> :NM_000784:exon8:c.142 1G>A:p.(Arg474Gln)			F
Não	SGM3	Distrofia de cinturas	Negativo	het <i>FTCD</i> :NM_006657:exon2:c.136C>T: p.(Arg46Cys)			F
Sim	SGM5	Fenilcetonúria	Fenilcetonúria; 261600; Autossômico recessivo	homo <i>PAH</i> :NM_000277:exon12:c.1222C> T:p.(Arg408Trp)			F
Não	SGM6	Miopatia congénita	Miopatia congênita com desproporção de tipo de fibras; 255310; Autossômico recessivo; Autossômico dominante	het <i>TPM3</i> :NM_152263:exon1:c.103G>T :p.(Glu35*)			F
Não	SGM7	Doença de McArdle	Negativo				F
Não	SGM9	Imunodeficiência	Negativo	het <i>ATP7B</i> :NM_000053:exon6:c.1922T >C:p.(Leu641Ser)	het <i>MUTYH</i> :NM_001128425:c.536A> G:p.(Tyr179Cys)		M
Não	SGM10	Neurônio motor/Miopatia	Negativo	het <i>CACNA1S</i> :NM_000069:exon25:c.31 33_3135del:p.(Phe1045del)			M
Não	SGM11	Doença de Menkes	Negativo	het <i>MMACHC</i> :NM_015506:exon2:c.271 dup:p.(Arg91Lysfs*14)	het <i>GLB1</i> :NM_000404:exon15:c.1511 A>G:p.(Asn504Ser)		F
Não	SGM13	Miopatia congénita ligada ao X	Miopatia miotubular ligada ao X; 310400; Ligado ao X recessivo	hemi <i>MTM1</i> :NM_000252:exon12:c.1262G >A:p.(Arg421Gln)			M
Não	SGM14	Neurônio motor/Miopatia	Negativo	het <i>GCDH</i> :NM_000159:c.1204C>T:(p.Ar g402Trp)			M

Não	SGM17	Distrofia muscular congênita	Negativo	hemi MID1:NM_001193278:exon3:c.880G>A:p.(Glu294Lys)			M
Sim	SGM18	Mucopolissacaridose IV	Mucopolissacaridose IVA; 253000; Autossômico recessivo	homo GALNS:NM_000512:exon5:c.491A>C:p.(Asn164Thr)	het ACADVL:NM_000018:exon14:c.1406G>A:p.(Arg469Gln)		F
Sim	SGM19	Fenilcetonúria	Fenilcetonúria; 261600; Autossômico recessivo	het PAH:NM_000277:exon12:c.1241A>G:p.(Tyr414Cys)	het PAH:NM_000277:exon3:c.194T>C:p.(Ile65Thr)		M
Não	SGM21	Rett	Holoprosencefalia 4; 142946; Dominante penetrância incompleta	het TGIF1:NM_170695:c.872C>T:p.(Ser291Phe)			F
Não	SGM22	Doença de Menkes	Doença de Menkes, 309400; Ligado ao X recessivo	ATP7A del exons 17-22			M
Sim	SGM23	Tirosinemia	Tirosinemia tipo I; 276700; Autossômico recessivo	homo FAH:NM_000137:exon7:c.554-1G>T			M
Não	SGM25	Rett	Negativo	het GALNS:NM_000512:c.1156C>T:p.(Arg386Cys)	het MUTYH:NM_001128425:c.1187G>A:p.(Gly396Asp)		F
Não	SGM26	Danon	Negativo	het BRCA2:NM_000059:exon20:c.8488-1G>A			M
Não	SGM27	Paralisia Hipocalêmica	Paralisia Hipercalêmica periódica tipo 2; 170500; Autossômico dominante	het SCN4A:NM_000334:c.2006G>A:p.(Arg669His)	het FANCA:NM_000135:exon37:c.3706C>T:p.(Gln1236*)		M
Sim	SGM29	Doença de Canavan	Doença de Canavan; 271900; Autossômico recessivo	homo ASPA:NM_001128085:c.914C>A:p.(Ala305Glu)	het CFTR:NM_000492:c.1210-11T>G		M
Não	SGM31	Doença de Wilson	Doença de Wilson 277900; Autossômico recessivo	het ATP7B:NM_000053:c.3809A>G:p.(Asn1270Ser)	het ATP7B:NM_000053:c.3207C>A:p.(His1069Gln)		F
Sim	SGM33	Deficiência de Ornitina Transcarbamilase	Deficiência de Ornitina Transcarbamilase; 311250; Ligado ao X recessivo	het OTC:NM_000531:exon4:c.343_344 del:p.(Asp115Tyrfs*7)			F
Não	SGM38	Distrofia muscular	Negativo	het PDSS2:NM_020381:exon3:c.484C>T:p.(His162Tyr)			F
Não	SGM39	Distrofia merosina negativa	Negativo	het PCDH15:NM_033056:exon33:c.5381C>T:p.(Tyr1794Ile)	het LAMA2:NM_000426:exon9:c.1255del:p.(Ile419Leufs*4)		F
Não	SGM40	Rett	Negativo	het DLAT:NM_001931:exon4:c.523	het		F

				G>C:p.(Ala175Pro)	SLC22A5:NM_003060:exon1:c.211C>A:p.(Leu71Met)		
Não	SGM41	Deficiência de PDH	Negativo	hemi ABCD1:NM_000033:IVS8:chrX:153008637G>C	het ASPA:NM_000049:exon2:c.326A>G:p.(Tyr109Cys)	PDHA1, DLAT nl (falta: PDHB; PDP1; PDX1)	M
Não	SGM42	Galactosemia	Galactosemia; 230400; Autossômico recessivo	homo GALT:NM_000155:exon2:c.90dup:p. (His31Alafs*10)			F
Não	SGM43	Aciduria Glutárica tipo I	Rett; 312750; Ligado ao X dominante	het MECP2:NM_004992:exon3:c.274_275dup:p.(Pro93Aspfs*33)			F
Não	SGM44	Segawa	Negativo	het MCCC2:NM_022132:exon1:c.79T>C:p.(Ser27Pro)			M
Sim	SGM45	Aciduria 3-OH-3-metil glutárica	Deficiência de HMG-CoA liase; 246450; Autossômico recessivo	homo HMGCL deleção exon 7 e 8			M
Não	SGM49	Miopatia congênita	Negativo				M
Sim	SGM50	Xantomatose cerebrotendínea	Xantomatose cerebrotendínea; 213700; Autossômico recessivo	homo CYP27A1:NM_000784:c.1421G>A:p. (Arg474Gln)			F
Sim	SGM52	Maple Syrup Urine Disease	Maple Syrup Urine Disease tipo Ib; 248600; Autossômico recessivo	het BCKDHB:NM_000056:exon8:c.853C>T:p.(Arg285*)	hetBCKDHB:NM_000056:exon3:c.343+2T>G		F
Sim	SGM53	Def primária de carnitina	Negativo			painelnã o tem CPT2	F
Não	SGM54	Miopatia congênita	Negativo	het CPS1:NM_001875:exon14:c.1424del:p.(Gly475Alafs*2)			F
Não	SGM56	Acidos orgânicos alterados	Holoprosencefalia 7; 610828; Autossômico dominante	het PTCH1:NM_000264:exon9:c.1328C>G:p.(Ala443Gly)	het PCCA:NM_000282:exon22:c.2017G>C:p.(Val673Leu)		M
Não	SGM57	Doença de Canavan	Negativo	het SLC22A5:NM_003060:exon1:c.254_264dup:p.(Ile89Glyfs*45)	het POR:NM_000941:c.859G>C:p.(Ala287Pro)		F
Sim	SGM58	Gangliosidose GM1	GM1-gangliosidose; tipo I 230500; Autossômico recessivo	homo GLB1:NM_000404:exon15:c.1577dup:p.(Trp527Leufs*5)			F

Não	SGM62	Porfíria eritropoética congênita	Negativo	het LMNA:NM_005572:exon3:c.565C>T :p.(Arg189Trp)	het MCCC1:NM_020166:exon10:c.98 4A>T:p.(Lys328Asn)	painel o tem UROS	M
Não	SGM63	Miopatia congênita	Negativo	het FTCD:NM_006657:exon9:c.990dup: p.(Pro331Alafs*2)	het EIF2AK3:NM_004836:c.154G>A: p.(Ala52Thr)		M
Não	SGM64	Rett	Rett; 312750; Ligado ao X dominante	het MECP2:NM_004992:exon4:c.397C> T:p.(Arg133Cys)			F
Sim	SGM65	Xantomatose Cerebrotendínea	Xantomatose Cerebrotendínea; 213700; Autossômico recessivo	het CYP27A1:NM_000784:exon4:c.825 G>A:p.(Trp275*)	het CYP27A1:NM_000784:c.1016C> T:p.(Thr339Met)		M
Não	SGM66	Paralisia Hipocalêmica	Negativo	het ATP7B:NM_000053:c.3809A>G:p.(A sn1270Ser)			M
Não	SGM67	Miopatia com contraturas	Negativo	het ALDH7A1:NM_001182:c.1279G>C: p.(Glu427Gln)			F
	P3611-I	Controle	Negativo	het MCCC2:NM_022132:exon16:c.1545 G>C:p.(Glu515Asp)			M
	P3721-I	Controle	Negativo	het HADHA:NM_000182:exon17:c.1808 T>C:p.(Leu603Pro)			M
	P3805-I	Controle	Negativo	het DLAT:NM_001931:exon4:c.605C>T: p.(Ser202Leu)			F
	P3809-I	Controle	Negativo	het GLI2:NM_005270:exon11:c.1811A> C:p.(His604Pro)	het SETBP1:NM_001130110:exon4:c. 703C>T:p.(Pro235Ser)		F
	P3811-I	Controle	Negativo	het SLC2A1:NM_006516:exon5:c.607G >A:p.(Val203Met)			F
	P3823-I	Controle (Marfan)	Homocistinúria; 236200; Autossômico recessivo	het CBS:NM_000071:c.833T>C:p.(Ile27 8Thr)	het CBC:NM_000071:exon9:c.828+1 G>A		F
	P3833	Controle	Negativo	het DHCR7:NM_001360:c.964- 1G>C	het MUT:NM_000255:exon4:c.836A> G:p.(Tyr279Cis)		M
	P3836	Controle	Negativo	het CFTR:NM_000492:exon12:c.1624G >T:p.(Gly542*)	hemi ATP7A:NM_000052:exon5:c.1516 A>G:p.(Ile506Val)		M
	P3842	Controle	Negativo	homo DLL3:NM_016941:exon5:c.674G>A:	het ATP7B:NM_000053:c.3101A>G:p		F

				p.(Ser225Asn)	.(His1034Arg)		
	P3855	Controle	Negativo	het ACADM:NM_000016:c.985A>C:p.(Lys329Gln)	het CFTR:NM_000492:c.1210-11T>G		F
	P3884-I	Controle	Negativo	het GJB2:NM_004004:c.35G>T:p.(Gly12Val)			F
	P3896	Controle	Negativo	het SETX:NM_015046:exon4:c.203G>A:p.(Arg68His)			F
	P3897-I	Controle	Negativo	het ETHE1:NM_014297:exon6:c.695C>T:p.(Pro232Leu)			M
	P3906-I	Controle	Negativo				F
	P3912-I	Controle	X Frágil; 300624; Ligado ao X dominante	hemi FMR1:NM_002024:exon5:c.413G>A:p.(Arg138Gln)	het MTR:NM_000254:exon19:c.1982T>C:p.(Ile661Tyr)		M
	P3924-I	Controle	Negativo				M
	P3925-I	Controle	Negativo	het FOXE1:NM_004473:exon1:c.613C>A:p.(Leu205Met)	het ATP7B:NM_000053:exon21:c.4151A>G:p.(Tyr1384Cys)		M
	P3933-I	Controle	Negativo	het PAH:NM_000277:c.688G>A:p.(Val230Ile)			F
	P3935-I	Controle	Negativo	het TBX22:NM_016954:exon8:c.1544C>A:p.(Ala515Glu)			M
	P3940-I	Controle	Negativo	het ATP7B:NM_000053:c.1607T>C:p.(Val536Ala)	het GLB1:NM_000404:exon15:c.1588C>T:p.(Arg530Cys)		F
	P3941-I	Controle	Negativo	het TG:NM_003235:c.6725G>A:p.(Arg242His)			F
	P3945-I	Controle	Negativo	het HERC2:NM_004667:exon93:c.1434A>G:p.(Lys4782Glu)	het DMD:NM_004006:exon39:c.5550A>C:p.(Lys1850Asn)		F
	P3952-I	Controle	Negativo	het DUOX2:NM_014080:exon6:c.602dup:p.(Gln202Thrfs*99)			M
	P3955-I	Controle	Negativo	het CYP21A2:NM_000500:c.1360C>T:p.(Pro454Ser)	het IDUA:NM_000203:exon10:c.1474C>T:p.(Arg492Trp)		F
	P3962-I	Controle	Negativo	het FOXG1:NM_005249:exon1:c.206C>			F

				A:p.(Pro69Gln)			
	P3969-I	Controle	Negativo	het CUBN:NM_001081:exon17:c.2275A >G:p.(Ser759Gly)	het CUBN:NM_001081:exon9:c.910G >A:p.(Glu304Lys)		F
	P3982-I	Controle	Negativo	het CFTR:NM_000492:exon11:c.1521_1523del:p.(Phe508del)	het ATP7B:NM_000053:exon7:c.1995G>A:p.(Met665Ile)		M
	P3983-I	Controle	Negativo	het ETHE1:NM_014297:c.278C>T:p.(Ser93Phe)	het PCDH15:NM_033056:exon33:c.4681_4684dup:p.(Ser1562*)		F
	P3986-I	Controle	Negativo				M
	P3987-I	Controle	Negativo				F
	P3993-I	Controle	Negativo	het CBS:NM_000071:c.1105C>T:p.(Arg369Cys)			F
	P3998-I	Controle	Negativo	het IVD:NM_002225:c.941C>T:p.(Ala314Val)			M
	P4003-I	Controle	Negativo	het RTTN:NM_173630:exon29:c.3926A>C:p.(Glu1309Ala)			M
	P4091-I	Controle	Negativo	het GLI2:NM_005270:exon12:c.2119T>C:p.(Ser707Pro)	het RAI1:NM_030665:exon3:c.2170G>A:p.(Ala724Thr)		M
	P4099-I	Controle	Negativo	het CBS:NM_000071:c.1105C>T:p.(Arg369Cys)			M
	P4100-I	Controle	Negativo (provavelmente patogênica + VUS)	het ATP7B:NM_000053:exon4:c.1607T>C:p.(Val536Ala)	het ATP7B:NM_000053:exon14:c.3086C>G:p.(Thr1029Ser)		F
	P4102-I	Controle	Negativo	het MMAA:NM_172250:exon6:c.947G>A:p.(Arg316His)			F
	P4107-I	Controle	Negativo (provavelmente patogênica + VUS)	het ASS1:NM_000050:exon5:c.323G>T:p.(Arg108Leu)	het ASS1:NM_000050:exon5:c.322C>T:p.(Arg108Trp)		M
	P4125-I	Controle	Negativo	het MTRR:NM_002454:exon7:c.1021G>A:p.(Val341Ile)	het HLC5:NM_000411:exon5:c.133C>G:p.(Pro45Ala)		M
	P4128-I	Controle	Negativo	het GALNS:NM_000512:c.1127G>A:p.(Arg376Gln)			F

	P4129-I	Controle	Negativo	het NSD1:NM_022455:c.4892A>G:p.(Lys1631Arg)	het FECH:NM_000140:exon2:c.163G>A:p.(Gly55Ser)		F
	P4131-I	Controle	Negativo (patogênica + VUS)	het GALNS:NM_000512:exon11:c.1162G>A:p.(Asp388Asn)	het GALNS:NM_000512:exon4:c.372G>A:p.(Leu124Leu)		F
	P4145-I	Controle	Negativo				M
	P4150-I	Controle	Negativo	het SLC22A5:NM_003060:c.136C>T:p.(Pro46Ser)			M
	P4151-I	Controle	Negativo	het CBS:NM_000071:c.1330G>A:p.(Asp444Asn)			F
	VP	Erro inato do metabolismo		Patogênica (herança recessiva)			
	FP	Miopatia		Provavelmente patogênica (herança recessiva)			
	VN	Distrofia muscular		VUS (herança recessiva)			
	FN	Síndrome de Rett		Patogênica (herança dominante)			
		Imunodeficiência congênita		Provavelmente patogênica (herança dominante)			
				VUS (herança dominante)			
				Patogênica (doenças musculares)			
				Provavelmente patogênica (doenças musculares)			
				Patogênica ou Provavelmente patogênica (surdez)			

VP = Verdadeiro-positivo; FP = Falso-positivo; VN = Verdadeiro-negativo; FN = Falso-Negativo; VUS = Variante de significado incerto/indeterminado

Quadro 14: Resumo dos achados positivos e negativos nas 45 amostras "casos" e 45 amostras "controles" de simulação de triagem neonatal.

45 “casos” UFRGS:
<ul style="list-style-type: none"> •14 com confirmação molecular prévia 13 verdadeiro-positivos 1 caso o gene não consta no painel Sensibilidade $13/14 = 93\%$ ou $13/13 = 100\%$
<ul style="list-style-type: none"> •31 sem confirmação 9 verdadeiro-positivos 2 casos as doenças não constam no painel Sensibilidade $9/31 = 29\%$ ou $9/29 = 31\%$
•Sensibilidade total $22/45 = 49\%$ ou $22/42 = 52\%$
<ul style="list-style-type: none"> •26 EIM – 19 verdadeiro-positivos, 2 doenças não constam no painel – Sensibilidade EIM $19/26 = 73\%$ ou $19/24 = 79\%$ •10 miopatias – 2 verdadeiro-positivos •4 distrofias musculares – 0 verdadeiro-positivo •4 hipóteses de Rett – 1 verdadeiro-positivo •1 hipótese de imunodeficiência – 0 verdadeiro-positivo
45 “controles” laboratório de análises clínicas CEGH-CEL:
<p>43 verdadeiro-negativos</p> <p>1 falso-positivo para variante descrita na literatura como patogênica (com estudo funcional) em <i>FMR1</i></p> <p>1 verdadeiro-positivo (teste realizado para Marfan, paciente possui 1 variante patogênica e 1 variante provavelmente patogênica em <i>CBS</i> para homocistinúria – confirmada clinicamente)</p> <p>•Especificidade $43/44 = 98\%$</p>

5. Discussão

5.1. Validação da técnica de NGS (etapa A)

No presente estudo, a técnica de NGS apresentou alta sensibilidade [95,92% (Intervalo de Confiança [IC] 95% = 86,02% a 99,50%)] e alta especificidade [98,77% (IC 95% = 93,31% a 99,97%)]. Esses resultados são comparáveis com os encontrados por outros grupos na literatura, que identificaram sensibilidade e especificidade acima de 99% (Chin et al., 2013, Sikkema-Raddatz et al., 2013) e que chegam a desaconselhar validação, pela técnica de Sanger, de variantes identificadas por NGS (Beck et al., 2016).

Apenas duas variantes conhecidas por Sanger apresentaram resultado falso-negativo na pesquisa por NGS (de um total de 49). Além de podermos afirmar que a técnica de NGS apresenta uma alta sensibilidade em relação à técnica de Sanger, considerada o padrão-ouro, podemos também observar que os resultados falso-positivos encontrados não são completamente inesperados e há medidas para evitar que ocorram. A variante em *COL18A1*, não identificada pela técnica de NGS, tem os exons 1 e 2 ricos em bases CG. Regiões ricas em bases CG apresentam maior dificuldade de amplificação durante as etapas de PCR do preparo de bibliotecas; por isso, são comumente regiões de baixa cobertura. Esse exemplo evidencia uma limitação da técnica de NGS, que nem sempre cobre homogeneamente todas as regiões exônicas. Entretanto, essas regiões de baixa cobertura podem ser identificadas tanto por análise computacional como por visualização direta em “browsers” como o IGV. Portanto, quando existem regiões de baixa cobertura em genes possivelmente associados ao quadro clínico do paciente, o resultado negativo do teste pela técnica de NGS pode não ser definitivo e a região deve ser investigada por outro método ou sequenciada novamente, para aumentar a cobertura por NGS. Em relação ao outro resultado falso-negativo, a região da variante em *BRAF* está bem coberta pela técnica

de NGS no paciente CC1074. A principal possibilidade para explicar a discordância entre os resultados é que esta se deva à troca de amostra em alguma das etapas. Erros de identificação de amostra em laboratórios clínicos ocorrem na faixa de 1 a 2% (Lippi et al., 2017). Devido à alta qualidade do sequenciamento NGS na região, era provável que a amostra testada não apresentasse a variante procurada. A amostra enviada para NGS foi retestada por Sanger e a variante em *BRAF* realmente não foi identificada. Não se pode assegurar que o primeiro teste de Sanger tenha sido realizado na mesma amostra testada por NGS. Tampouco se pode excluir a possibilidade de que se tenha registrado o resultado de outro paciente no momento do registro do primeiro resultado do teste de Sanger.

Em relação à variante com anotação divergente no gene *SOS1* (p.Arg552Thr no NGS e p.Arg552Lys no Sanger), embora não seja possível afirmar com certeza qual foi a mutação previamente encontrada por Sanger, uma vez que a anotação da mutação recuperada estava escrita apenas no nível da proteína, é altamente provável que se esteja diante da mesma mutação encontrada por NGS. Isso porque se trata de uma mutação rara presente no mesmo códon. Nesse exemplo podemos ressaltar que, uma vez padronizada e validada, a anotação por computador é menos sujeita a erros do que a anotação pelo analisador humano.

Em relação à variante falso-positiva encontrada no gene *SOS2* (p.Asn310Lys), sabemos que a técnica de NGS, quando comparada à de Sanger, apresenta um alto número de variantes falso-positivas na lista de variantes não filtradas, o que se deve principalmente ao fato de as regiões sequenciadas serem muito maiores. Entretanto, além da aplicação de filtros de frequência alélica populacional, que são extremamente úteis no caso de doenças raras, devemos também avaliar indicadores de qualidade das variantes encontradas por NGS. Essa análise de qualidade já era realizada no sequenciamento Sanger quando o interpretador, por exemplo, não confiava em um cromatograma cujo pico de intensidade de sinal se apresentasse desviado da proporção 1:1 para variantes em heterozigose. No caso da variante em *SOS2* identificada neste estudo, levando em conta os dados de qualidade atribuídos à variante, balanço alélico, viés de fita de leitura (*strand bias*) e cobertura local, bem como a visualização da variante no IGV, pode-se classificá-la como uma variante

provavelmente falsa, mesmo antes da confirmação por uma técnica distinta. Na amostra CC5072, essa variante apresentou uma cobertura adequada de 24x; no entanto, há um desvio no balanço alélico de 18 bases referência para 6 bases alternativas (0,25 ou 3:1), sendo que valores esperados para variantes em heterozigose em geral se situam entre 0,3 e 0,7. Finalmente, há um desvio total no viés da fita de leitura (*strand bias*) apenas para as bases lidas como alternativas, circunstância que, nesse caso, é mais importante que o balanço alélico. O resultado de sequenciamento da base específica foi [7+,11-/6+,0-], o que significa que a base referência 'T' foi lida 7 vezes em fitas positivas e 11 vezes em fitas negativas, enquanto a base alternativa 'A' foi lida apenas 6 vezes em fitas positivas. O fato de não haver leitura da base alternativa em fitas negativas indica com alta probabilidade que não existe uma base alternativa no DNA da amostra. Esse erro provavelmente se deveu ao fato de a região apresentar um pequeno homopolímero com repetição de três bases logo antes do nucleotídeo lido erroneamente, na fita de sequenciamento positiva e, somado a isso, a má qualidade de sequenciamento dos seis fragmentos que apontaram a base alternativa 'A', tanto que várias outras bases desses seis fragmentos foram lidas como mutadas. Essas duas últimas constatações podem ser visualizadas no IGV e corroboram a suspeita de que se trata de falsa variante (Figura 4). Dessa forma, é sempre recomendado que, diante de uma possível variante patogênica, haja uma avaliação de todos estes parâmetros, incluindo a visualização pelo IGV.

Figura 5: Região visualizada no IGV da única variante falso positiva identificada por NGS não confirmada por Sanger. A amostra na qual foi atribuída (CC5072) a variante está na metade superior da figura. A amostra da metade inferior é um controle realizado no mesmo experimento. Notar que as seis leituras de base alternativa ‘T’ estão em fragmentos que “apontam para a direita” (fita positiva) e cuja qualidade de sequenciamento está ruim como um todo. O mesmo padrão ocorreu na amostra controle, o controle não foi chamado como mutado pois há menor número de bases alternativas em boa qualidade

Além disso, é importante ressaltar que a técnica de Sanger, embora considerada padrão-ouro, não apresenta sensibilidade de 100%, como foi possível demonstrar com a amostra c10113-III e com a variante em hemizigose DMD:NM_004006.2:IVS40:c.5740-16T>A. Nesse exemplo, a primeira tentativa de confirmação por Sanger não conseguiu amplificar a região-alvo e identificar a variante encontrada por NGS. No entanto, como a qualidade da variante pelo NGS estava alta [0+,0-/19+,21-], ou seja, mais de 30 fragmentos apontavam apenas o alelo alternativo em hemizigose no cromossomo X de um menino e sem *strand bias*, foi repetido o sequenciamento Sanger com um novo par de primers. O segundo experimento por Sanger confirmou a variante detectada por NGS e constatou que na primeira reação havia ocorrido perda da representatividade do alelo alternativo (*allele dropout*). Esse é um risco a que está sujeito o sequenciamento Sanger, por depender de “uma única sonda” (o par específico de primers) para abordar a região. Estão menos sujeitas ao risco de falsa representatividade alélica as bibliotecas para NGS capturadas com múltiplas sondas para a mesma região (*tiled*) ou com uma sonda grande (>90 nucleotídeos), o que evita que variantes individuais interfiram na afinidade de hibridação. Bibliotecas para NGS capturadas por “*Amplicon*” (que gera múltiplos produtos de PCR para regiões de interesse) podem correr o mesmo risco que o sequenciamento Sanger (Silva et al., 2017).

Atualmente (2016-2017) vários pesquisadores têm questionando se a técnica de NGS não deveria ser considerada padrão-ouro (mesmo quando comparada a outras técnicas que não apenas Sanger) (Lytsy et al., 2017; Kotrova et al., 2017) e quais seriam os critérios de qualidade que justificariam a necessidade de confirmação de variantes por outros métodos (Beck et al. 2016). Em artigo de 2016 baseado na avaliação de milhares de variantes (>5800), Beck et al. afirmaram que a técnica de NGS não só apresenta as características necessárias para ser considerada padrão-ouro no contexto clínico, como também tem performance superior à da técnica de Sanger. O principal critério de qualidade que o grupo utilizou e recomendou foi criado pelo próprio grupo e é denominado ‘*Most Probable Genotype*’ (MPG). Esse critério aplica um algoritmo matemático que utiliza a cobertura local, o balanço entre os alelos e a qualidade das bases sequenciadas para atribuir um escore que exprime a

probabilidade do genótipo chamado em relação a outros possíveis genótipos (Beck et al., 2016). Tal critério é semelhante ao algoritmo utilizado pelo GATK, que atribui ao genótipo chamado um escore de possibilidade genotípica denominado *Normalized Phred-scaled likelihoods of the possible genotypes* (PL) (McKenna et al., 2010). O ponto de corte para esses critérios, conquanto baseado em múltiplas observações, ainda é arbitrário. O próprio artigo de Beck et al. (2016) afirma que, em geral, um score de MPG=10 (considerado como ponto de corte para alta qualidade) é conseguido com sequências de NGS com coberturas de pelo menos 10x a 20x. Com base nos dados da literatura e na experiência obtida com o processamento de dados genéticos da população brasileira (Naslavsky, Yamamoto et al., 2017), o autor desta tese sugere um critério análogo aos anteriores, só que de mais fácil obtenção, dependente apenas de limiares dos dados brutos de sequenciamento, dispensadas equações matemáticas mais complexas que tanto divergem entre laboratórios e pesquisadores. O critério proposto é que variantes em homozigose ou hemizigose que apresentem cobertura mínima de 10x ou variantes em heterozigose com cobertura mínima de 20x, com boa qualidade de bases (escore de sequenciamento maior que 30), bom balanço alélico (0,3 a 0,7 no caso de heterozigose ou >0,90 para homo/hemizigose) e ausência de *strand bias*, poderiam ser consideradas padrão-ouro para análises clínicas (Quadro 15). O limiar proposto pelo autor para considerar ausência de *strand bias* é que a diferença entre a proporção de bases referência na fita positiva e negativa e a proporção de bases alternativas nas fitas positivas e negativas seja menor que 100% para variantes em heterozigose. Um método simples de calcular essa diferença é dividir a proporção que for maior pela proporção que for menor e checar se o resultado é menor que 2,0 ou, no caso de um rigor matemático mais formal, calcular o módulo do logaritmo na base 10 da razão entre as proporções (independentemente de denominador ou numerador) e obter valor menor que 0,30103 (Quadro 15). Em uma variante hipotética com as seguintes proporções: [12+,10-/10+,8-], a diferença seria de $|\log_{10}(12/10)/(10/8)|=0,0177$ ou $(10/8)/12/10=1,0417$ (ou seja, sem *strand bias*). Já em uma variante com as seguintes proporções [12+,12-/10+,4-], a diferença seria de $|\log_{10}(12/12)/(10/4)|=0,39794$ ou $(10/4)/(12/12)=2,5$ (ou seja, com *strand bias*). No caso de variantes em homozigose ou hemizigose, nas quais a quantidade de leituras da

base referência são próximas de 0 (ou pelo menos menor que 10% da quantidade total de bases), pode-se aplicar diretamente um balanço entre leituras positivas e negativas do alelo alternativo e exigir o mesmo limiar aplicável ao balanço alélico (Quadro 15). É importante ressaltar que nos sequenciadores NGS de segunda geração (incluindo todos os utilizados nesta tese), as variantes in/del apresentam consistentemente uma qualidade pior que as variantes de ponto (*single nucleotide variants* – SNV). Dessa maneira, é válido questionar se os critérios e limiares sugeridos seriam igualmente aplicáveis a qualquer tipo de variante. Embora, como via de regra, a principal estratégia para melhorar a qualidade de variantes seja impor limiares de cobertura mais elevados, no caso de variantes in/del essa medida talvez não seja suficiente e tampouco a mais apropriada, dado o caráter sistemático dos erros de alinhamento relacionados a tais variantes. Tendo em mente a busca de solução para variantes falsas provindas de erros sistemáticos, pode-se considerar que há um segundo critério extremamente útil para a avaliação de qualidade de qualquer tipo de variante, mas de especial importância no caso de in/del: o critério de comparação da variante encontrada em uma amostra com dados de amostras-controle realizadas em um mesmo lote. A comparação pode ser feita tanto por frequência alélica esperada na tabela VCF, como por observação direta no IGV de múltiplas amostras simultaneamente (como na Figura 3). Com o crescimento dos bancos de dados genômicos, é possível saber com precisão a frequência alélica de variantes presentes em mais de 5% da população. Com esse dado podemos inferir que variantes presentes nos controles de um mesmo lote de sequenciamento com frequência superior a 5%, mas ausentes (ou pouco frequentes) em populações-controle ou em lotes realizados por outra técnica ou tipo de sequenciador, apresentam alto risco de serem falso-positivas devido a erros sistemáticos. A desvantagem desse critério comparativo está em depender da disponibilidade de amostras-controle e de outros dados populacionais, não lhe bastando o dado bruto resultante do sequenciamento de apenas uma amostra.

Quadro 15: Critérios propostos no presente estudo para qualidade mínima de variantes detectadas por NGS para que possam ser consideradas de qualidade equivalente a padrão-ouro e prescindir de confirmação por outro método de sequenciamento.

	Heterozigose	Homo/Hemizigose
Cobertura de bases com qualidade de sequenciamento ≥ 30	20x	10x
Balanço alélico (alelo alt. / alelo ref.)	0,3 a 0,7	>0,90 (alternativo) <0,10 (referência)
Balanço das fitas de leitura (<i>strand bias</i>)	(maior proporção pos/neg) / (menor proporção pos/neg) < 2,0 ou $ \log_{10}(\text{pos/neg alelo ref.})/(\text{pos/neg alelo alt.}) < 0,30103$	[(alelo alt. pos) / (alelo alt. neg)] entre 0,3 a 0,7

Variantes com balanço alélico entre 0,1 e 0,3 e entre 0,7 e 0,9 não necessariamente são decorrentes de erros e podem indicar presença de mosaicismo ou CNV na amostra. No entanto, para efeito de relatório clínico, é recomendado que sejam confirmadas por outro método ou que o CNV seja identificado por outro método. Esta proposta requer validação em coortes maiores.

5.2. Aplicação da técnica de NGS em doenças ósseas genéticas e RASopatias (etapa B)

Nesta etapa do projeto, o sequenciamento de nova geração foi empregado, tanto no contexto de pesquisa como no contexto clínico, em dois grupos de doenças monogênicas, as doenças ósseas, particularmente as displasias esqueléticas, e as RASopatias.

Na aplicação como pesquisa, o objetivo foi desvendar a base molecular de doenças sem etiologia definida (displasia espondilometafisária com distrofia de cones e

bastonetes e displasia espondilometafisária tipo *corner fracture* / tipo Sutcliffe) ou com etiologia parcialmente conhecida (síndrome de Noonan, dentre as RASopatias). Foi possível identificar três novos genes associados a essas doenças – *LZTR1* e *SOS2*, para a síndrome de Noonan, e *PCYT1A* para a displasia espondilometafisária com distrofia de cones e bastonetes. Um quarto gene, associado a displasia espondilometafisária tipo *corner fracture* ou Sutcliffe, aguarda validação em estudo colaborativo com um pesquisador do exterior.

O êxito no emprego da técnica de NGS (WES) com o intuito de identificar genes associados a fenótipos mendelianos é um fato incontestável. Na última década, o conhecimento das bases moleculares das doenças mendelianas aumentou de forma vertiginosa (Chong et al., 2015). Por exemplo, uma pesquisa com os termos “skeletal dysplasia” e “whole-exome sequencing” no banco de publicações médicas (PubMed) revela que, em 2011, dois novos genes foram identificados; em 2013, esse número subiu para nove e, em 2015 e 2016, chegou a dez por ano.

No entanto, nem sempre é fácil identificar genes novos associados a fenótipos mendelianos. Muitas vezes é preciso estudar casos familiares ou grupo de indivíduos que compartilham o mesmo fenótipo, para aumentar a probabilidade de identificação do gene responsável pela doença (Chong et al., 2015). No presente estudo, a identificação do gene *PCYT1A* partiu da análise de duas famílias consanguíneas que apresentavam mais de um indivíduo afetado, com padrão de herança reconhecido como autossômico recessivo para a displasia esquelética. Nesse caso, o fato da análise ter por fim identificar mutações em homozigose facilitou o estudo, tendo sido identificadas mutações raras em homozigose em apenas um gene compartilhado pelas duas famílias. No entanto, as dificuldades podem persistir quando se encontra um gene cuja função é pouco conhecida ou que não apresenta uma função específica que justifique o fenótipo estudado. Foi o que ocorreu com o gene *PCYT1A*, que faz parte da via da fosfatidilcolina, até então não associada diretamente ao metabolismo ósseo (Yamamoto et al., 2014). Nesses casos, muitas vezes é necessário realizar estudos funcionais adicionais para provar a função específica do gene no que tange ao fenótipo avaliado. Alternativamente, encontrar mutações neste gene em múltiplos pacientes de diferentes famílias com o mesmo fenótipo também é fortemente sugestivo da

associação causal destas variantes. No caso do gene novo associado, à displasia esquelética espondilometafisária tipo *corner fracture*, a estratégia para encontrar a causa genética foi realizar o WES do trio (probando e genitores não afetados). Com isso, para uma doença cujo padrão de herança sugerido era autossômico dominante, buscavam-se especialmente mutações presentes apenas no probando (mutações *de novo*). Essa estratégia de pesquisa visa a diminuir de forma drástica o número de variantes analisadas, de 1000 a 2000 mutações raras por indivíduo para 1 a 2 mutações *de novo* WES (Acuna-Hidalgo et al., 2016). A mutação identificada nesse caso, presente apenas no probando, altera um gene que codifica uma proteína da matriz extracelular, componente importante no sistema esquelético.

O desafio foi maior no que concerne às RASopatias, em especial à síndrome de Noonan, que é de todas a mais prevalente e a de maior heterogeneidade genética, com padrão de herança autossômico dominante. Na época, o estudo conjunto dos genes da via RAS/MAPK associados à síndrome permitia a confirmação diagnóstica de aproximadamente 70% dos casos (Tartaglia et al., 2010). O fato da síndrome apresentar grande heterogeneidade genética lócica dificulta ainda mais a procura de novos genes, pois é grande a probabilidade de os genes ainda não identificados serem responsáveis por frações pequenas dos restantes 30% de pacientes negativos para os genes conhecidos. Nesse caso, o estudo demandaria uma casuística grande de afetados para viabilizar a busca por mutações em um gene presente num pequeno grupo de pacientes. A estratégia utilizada neste estudo para identificar dois genes novos na casuística de 79 probandos com RASopatias consistiu em partir dos casos familiares, avaliar clinicamente indivíduos afetados e normais das famílias e obter amostra deles, para demonstrar a segregação da variante identificada com o fenótipo nessas famílias. Além disso, nas famílias em que o caso era isolado, provar que o evento era *de novo* foi essencial para o êxito do projeto.

Atualmente, a maioria dos fenótipos mendelianos tem base molecular total ou parcialmente conhecida, circunstância que possibilita compreender melhor as vias biológicas envolvidas no desenvolvimento da doença. Trata-se de passo importante para viabilizar o desenvolvimento de terapias específicas, visando ao manejo mais adequado das doenças genéticas.

No decorrer deste estudo, a técnica de NGS rapidamente extrapolou do contexto de pesquisa para ser incorporada à prática clínica.

Da mesma forma, o presente estudo avaliou a aplicação de NGS no contexto clínico. Foi elaborado um painel customizado para doenças ósseas com mais de 80 genes associados a diferentes fenótipos de doenças do sistema esquelético, tendo sido empregado em 125 probandos com diversos diagnósticos de doenças ósseas. O WES foi realizado em uma parcela menor de probandos com doenças ósseas (19). Já para RASopatias, foi realizado apenas o WES em 79 probandos que já haviam sido testados por Sanger para os principais genes e mutações de RASopatias, com resultado negativo.

Os resultados na casuística total de doenças ósseas genéticas, com 144 indivíduos, mostraram uma sensibilidade diagnóstica de $107/173 = 61,85\%$ (IC 95% = 0,5182 a 0,7121), sendo de $75/125 = 60\%$ (IC 95% = 0,4810 a 0,7111) para os casos realizados por painel NGS e de $12/19 = 63,16\%$ (IC 95% = 0,3191 a 0,8793) para os casos avaliados por WES. Na casuística de RASopatias, a sensibilidade diagnóstica obtida foi de $36/79 = 45,57\%$ (IC 95% = 0,3505 to 0,5650) para os 79 casos por WES (Quadro 16).

No período, também foram realizados 29 testes pela técnica tradicional de Sanger em doenças ósseas genéticas, com sensibilidade diagnóstica de $20/29 = 69\%$ (IC 95% = 0,4349 a 0,8815), o que permite comparar o NGS à técnica considerada padrão-ouro para a pesquisa de mutações pontuais.

A sensibilidade diagnóstica de pouco mais de 60% obtida na casuística de doenças ósseas genéticas foi homogênea para os três métodos (Sanger, painel NGS e WES) no presente estudo e acima da média reportada na literatura, tanto para amostras de doenças ósseas (Polla et al., 2015; Bae et al., 2016; Retterer et al., 2016), como para amostras com outros tipos de doenças genéticas realizadas por NGS. Já a sensibilidade diagnóstica na casuística de RASopatias foi de 46% para WES, que é próxima à média da literatura (Lee et al., 2014; Lepri et al., 2014; Wang et al., 2014; Yang et al., 2014; Taillandier et al., 2015; Kuhn et al., 2016; Jamie et al., 2016; Retterer et al., 2016; Yubero et al., 2016; Celestino-Soper et al., 2017; Bhoj et al., 2017) (Quadro 16).

Quadro 16: Sensibilidade diagnóstica da técnica de NGS em diferentes estudos

Doença	N	Teste	Sensibilidade diagnóstica	Referência
Doenças genéticas (geral)	814	WES	26,17% (213/814; IC 95%, 23,26%-29,30%)	Lee et al., 2014
Doenças genéticas (geral)	2000	WES	25,20% (504/2000; IC 95%, 23,35%-27,15%)	Yang et al., 2014
Doenças genéticas (geral)	3040	WES	28,8%	Retterer et al, 2016
Doenças genéticas (doenças ósseas)	54	WES	39%	Retterer et al, 2016
Doenças genéticas (doenças metabólicas)	84	WES	26%	Retterer et al, 2016
Doenças genéticas (autismo)	130	WES	12%	Retterer et al, 2016
Epilepsia	28	Painel NGS (38 genes)	21,43% (6/28; IC 95%, 9,86%-39,89%)	Wang et al., 2014
Hipofosfatasia ou diferenciais de fragilidade óssea	46	Painel NGS (12 genes)	36,96% (17/46; IC 95%, 24,49%-51,43%)	Taillandier et al., 2015
Doenças ósseas	36	Painel NGS (306 genes)	22,22% (8/36; IC 95%, 11,47%-38,33%)	Polla et al., 2015
Displasias esqueléticas (todas hipóteses)	160	Painel NGS (255 genes)	52,50% (84/160; IC 95%, 44,79%-60,09%)	Bae et al., 2016
Displasias esqueléticas (hipóteses alta qualidade)	96	Painel NGS (255 genes)	73,96% (71/96; IC 95%, 64,33%-81,74%)	Bae et al., 2016
Displasias esqueléticas (hipóteses duvidosas)	64	Painel NGS (255 genes)	20,31% (13/64; IC 95%, 12,14%-31,85%)	Bae et al., 2016
Distrofias musculares de cinturas	58	Painel NGS (38 genes)	32,76% (19/58; IC 95%, 22,04%-45,62%)	Kuhn et al., 2016
Doença retiniana hereditária	562	Painel NGS (105 genes)	50,00% (281/562; IC 95%, 45,88%-54,12%)	Jamie et al., 2016
Doença retiniana hereditária (casos negativos após Painel NGS -105 genes)	33	WGS	33,33%* (11/33; IC 95%, 19,66%-50,48%)	Jamie et al., 2016
Erros inatos do metabolismo	146	Painel NGS (171 genes)	50,00% (73/146; IC 95%, 41,99%-58,01%)	Yubero et al., 2016
Doenças cardiovasculares selecionadas (principalmente cardiomiopatia)	33	Painel NGS (61 genes)	39,39% (13/33; IC 95%, 24,65%-56,35%)	Celestino-Soper et al., 2017
RASopatias	80	Painel NGS (11 genes)	47,50% (38/80; IC 95%, 36,92%-58,30%)	Lepri et al., 2014
RASopatias	126	Painel NGS (12 genes)	19,05% (24/126; IC 95%, 13,09%-26,84%)	Bhoj et al., 2017
Doenças ósseas	125	Painel NGS (142 genes)	60,00% (75/125; IC 95%, 51,23%-68,17%)	Presente estudo
Doenças ósseas	19	WES	63,16% (12/19; IC 95%, 40,94%-80,95%)	Presente estudo
RASopatias (casos negativos após investigação tradicional)	79	WES	45,57% (36/79; IC 95%, 35,05%-56,50%)	Presente estudo

* possivelmente indicando sensibilidade diagnóstica total para WGS de $[50\% + (0.33 \times 50\%)] = 66,66\%$

Diferentes variáveis interferem na positividade de um teste genético clínico, o que pode explicar uma variabilidade tão grande nos resultados obtidos nos diferentes estudos na literatura. Em primeiro lugar, a casuística ou o tipo de doença investigada tem um grande peso nesses resultados. O sequenciamento gênico, seja pelo método tradicional de Sanger, seja pelo sequenciamento de nova geração, é destinado à pesquisa de mutações pontuais e pequenas deleções/inserções. Grupos de pacientes que compartilham diagnóstico eminentemente monogênicos têm maior probabilidade de êxito. Nesse quesito, doenças de fácil diagnóstico clínico têm maior chance de resultado positivo. Mesmo em doenças monogênicas, o quanto se conhece dos mecanismos moleculares também influencia na positividade. Para uma doença em que a alteração molecular conhecida é responsável por apenas 30% dos casos, este será o valor máximo de sensibilidade diagnóstica na casuística testada, seja qual for a abrangência do teste molecular. As displasias esqueléticas (que compõem a maior parte das doenças ósseas genéticas) e as RASopatias são dois bons exemplos de doenças monogênicas de bases moleculares bem conhecidas. Isso talvez justifique em parte a alta sensibilidade diagnóstica encontrada neste estudo para as doenças ósseas. Já no que tange às RASopatias, é preciso ter em vista que a casuística estudada neste trabalho não é uma casuística “virgem”, ou seja, não corresponde ao conjunto total de RASopatias do serviço de genética médica do ICr-HC-FMUSP: antes, abrange apenas os casos que, tendo sido triados por sequenciamento Sanger em abordagem focada, apresentaram resultados negativos. Tal circunstância reduz a probabilidade de identificação de amostras positivas, ainda que se aplique um teste mais ampliado. Situação semelhante pode ser observada no estudo de Jamie et al. (2016) em uma casuística de doenças retinianas hereditárias: a sensibilidade diagnóstica de WGS em 33 amostras negativas para um painel de genes NGS com 105 genes relacionados a doenças retinianas foi de 33,33%, enquanto que a sensibilidade do painel na casuística “virgem” foi de 50%. Esses autores também realizaram WGS em 11 amostras que já tinham resultado positivo no painel NGS e as mesmas variantes patogênicas foram identificadas por WGS, o que nos autoriza a supor que se o WGS tivesse sido aplicado diretamente a todas as amostras da casuística “virgem”, a

sensibilidade teria sido de 66,67% (Quadro 16). Além disso, o quadro clínico das RASopatias apresenta uma variabilidade clínica maior do que o das doenças ósseas, sendo que outras síndromes malformativas ou doenças multifatoriais podem ser erroneamente classificadas como RASopatias. As três amostras identificadas neste estudo com síndromes associadas à deficiência intelectual autossômica dominante são evidência disso. Essa casuística não homogênea seria um fator para diminuir a sensibilidade do teste. Tal situação pode ser observada na casuística descrita por Bhoj et al. (2017), na qual apenas 19% dos casos referidos com hipótese de RASopatias foram positivos em um painel NGS com 12 genes da via RAS/MAPK (Quadro 16). Posteriores investigações genéticas identificaram uma etiologia diferente de RASopatias em 28 dos 96 pacientes (29%) que obtiveram resultados negativos para o painel de genes da via RAS/MAPK (Bhoj et al., 2017).

Além da casuística, um segundo fator que interfere na sensibilidade diagnóstica é uma questão premente nos dias de hoje e uma das tarefas mais difíceis na prática clínica: a escolha de qual teste genético indicar para cada paciente. Estimativas levantadas recentemente apontam que até 1/3 das indicações de testes genéticos não são adequadas (Ruzzo et al., 2017). Na prática, não há como criar um algoritmo exato para saber que teste genético deve ser realizado em cada paciente, ou mesmo se nenhum teste deve ser solicitado. Isso se deve ao fato de que a qualidade da hipótese diagnóstica depende intimamente da quantidade de informações clínicas disponíveis e da qualidade e experiência do médico que avalia o caso. O contexto da casuística de doenças ósseas genéticas desta tese é extremamente privilegiado no que diz respeito à avaliação clínica, pois os pacientes são acompanhados em hospital terciário que lhes dá acesso a múltiplos exames de alta qualidade, tais como exames radiológicos, ultrassonografia, tomografia e ressonância magnética. Além disso, o acompanhamento é feito por uma equipe especializada, com discussão inter-especialidades e avaliação de cada caso por múltiplos profissionais (de residentes a médicos com experiência em doenças ósseas). Assim, quando a hipótese clínica é de alta qualidade e a estimativa da probabilidade pré-teste é acurada, há maior probabilidade do teste genético ter resultado positivo. Isso vale tanto para os testes direcionados (Sanger e painel NGS), nos quais há necessidade de escolha prévia de genes para adequar a solicitação,

como para um teste mais abrangente (como WES), no qual a análise consegue ser mais aprofundada em genes que sejam compatíveis com a hipótese clínica. Essa diferença de sensibilidade diagnóstica entre hipóteses de alta qualidade e hipóteses duvidosas pode ser observada no estudo de Bae et al. (2016) em uma casuística de displasias esqueléticas (Quadro 16). Embora os testes mais abrangentes tenham por princípio uma maior probabilidade de cobrir o gene com a alteração de interesse, existe um fator importante a ser considerado: a cobertura média atingida por painel focado em comparação a WES. Em razão de o custo de sequenciamento ser mais baixo para o painel, os laboratórios costumam oferecer painel NGS com cobertura maior que 100x ou até 200x. Já a média de cobertura de exoma clínico é de 70-100x. Isto significa que, ainda que o gene esteja presente no alvo de captura do WES, a probabilidade de ocorrência de regiões de baixa cobertura é maior que no painel NGS. No final, para um gene de interesse para a hipótese diagnóstica, a variante que seria identificada pelo painel devido à cobertura adequada pode não ser identificada por WES (Klein e Foroud, 2017). Além disso, em termos de custo/benefício, a opção entre Sanger, painel NGS e WES depende de alguns fatores, como o custo de cada método, a disponibilidade dos exames e a probabilidade de positividade em cada um dos testes. Muitas vezes a escolha não é tão diretiva, pois a positividade dos testes nunca é de 100% e, iniciando-se pelo exame mais focado e com menor custo, se o resultado for negativo pode ser necessário realizar na sequência um segundo exame mais abrangente. Nessa hipótese, questiona-se se não faria mais sentido iniciar a investigação pelo teste de maior abrangência. Uma regra objetiva é a seguinte: para se solicitar um exame focado, a probabilidade pré-teste, estimada pelo médico solicitante, dele ser positivo deve ser maior que a razão de custos entre o teste focado (painel NGS) e o teste ampliado (WES ou WGS). Vale ressaltar que o conhecimento das bases moleculares de cada doença genética é fundamental para a escolha do teste genético. Se uma doença apresenta uma única mutação conhecida, ou uma região pequena (um exon) com todas ou com grande parte das mutações descritas, o sequenciamento tipo Sanger é suficiente para um resultado positivo. Já no caso da síndrome de Noonan que apresenta uma grande heterogeneidade genética lócica, com

mais de 12 genes já reconhecidamente associados, o NGS é uma ferramenta excelente pela possibilidade da análise conjunta de todos estes 12 genes.

Um terceiro fator a ser destacado, que pode ter contribuído para uma maior sensibilidade diagnóstica no presente estudo, é que as casuísticas não foram investigadas da mesma maneira que amostras de pacientes em laboratórios de análises clínicas. Em laboratórios de análises clínicas, não se costuma realizar teste genético em vários familiares e tampouco se procede a estudos funcionais para variantes e/ou genes encontrados. Múltiplas amostras desta tese foram submetidas a uma abordagem de teste de pesquisa, não de teste clínico. Com isso, amostras que seriam negativas quando analisadas isoladamente acabaram por ser classificadas como positivas. O teste que mais se beneficiou da abordagem de pesquisa foi o WES, que também foi o que apresentou maior discrepância entre a sensibilidade diagnóstica obtida em doenças ósseas genéticas e a média da literatura (Retterer et al., 2016). A abordagem de realizar o teste do trio (probando e genitores) em vez de restringi-lo ao paciente, sabidamente aumenta a sensibilidade diagnóstica, conquanto seja mais custosa (Lee, et al., 2014; Retterer et al., 2016); por isso, já é amplamente oferecida nos laboratórios de análises clínicas. Os critérios de indicação do teste do trio são análogos à comparação anterior entre solicitação de teste focado ou ampliado, mas de especial utilidade na busca de variantes *de novo*, quando se suspeita de quadro isolado dominante.

Além de apresentar bons resultados de sensibilidade diagnóstica, em quatro pacientes da casuística de doenças ósseas genéticas a técnica de NGS permitiu identificar variantes de número de cópias que não necessariamente teriam sido identificadas pelo sequenciamento Sanger. No caso da deleção em homozigose no *LIFR* (Quadro 9 - paciente CC5261), tanto algoritmos computacionais como a própria visualização direta no IGV facilmente identificaram a mutação. Ressalte-se apenas que a visualização direta depende de hipótese diagnóstica de alta qualidade que possa direcionar a análise, pois é inviável percorrer visualmente no IGV todos os genes do WES ou um painel de 500 genes. Já nas deleções ou duplicações em heterozigose, a identificação da variante não é tão imediata. A hipótese de deleção em *ALPL* (Quadro 9 - paciente F8127-1) só pode ser levantada e confirmada porque a variante *missense*

em *trans* apresentada pela paciente foi (erroneamente) identificada, em um primeiro momento, em homozigose, quando se sabia não haver história de consanguinidade entre os pais. Assim, utilizou-se o *software* NextGene para identificar a deleção. Por fim, para as deleções em *SHOX* e *TWIST1* (Quadro 10 - pacientes F8429-1 e F9483-1, respectivamente), a ausência de mutações de ponto patogênicas levou à investigação com o *software* NextGene, que apontou possíveis deleções pequenas em heterozigose, deleções estas confirmadas pela técnica do MLPA. Embora a identificação de CNV ainda não seja considerada padrão-ouro quando realizada a partir de dados de NGS, a possibilidade de detectar grandes deleções e duplicações sem a necessidade de um segundo teste é muito vantajosa (Hayes et al., 2013; Shen et al., 2016).

Com esses resultados, pode-se afirmar que a técnica de sequenciamento NGS é apropriada como teste diagnóstico em ambas as casuísticas estudadas e que, além de apresentar sensibilidade para variantes de ponto semelhante à de Sanger, permite a identificação de CNVs, embora não como padrão-ouro.

No entanto, permanece ainda um grande desafio, que não diz respeito à técnica em si, mas sim à interpretação das variantes encontradas. Até hoje, é difícil classificar determinadas variantes como patogênicas ou provavelmente patogênicas com acurácia. Muitas delas são classificadas como de significado incerto (VUS), não sendo possível atribuir a esta variante um determinismo quanto ao desenvolvimento de doenças. Esse fator é particularmente sensível no emprego de testes abrangentes, como o WES, em que se obtém para cada indivíduo um número na ordem de milhares de variantes raras. Para contornar o problema, as variantes encontradas em estudos realizados ao redor do mundo têm sido depositadas em bancos públicos, o que se aplica tanto a variantes relacionadas a doenças, como às encontradas na população aparentemente sadia de diferentes etnias. No Brasil, foi recentemente publicado o primeiro banco de dados genômicos com 609 amostras de WES representativas da população idosa da cidade de São Paulo. Esse banco auxilia na identificação de variantes provavelmente benignas, pois, além das amostras serem representativas da população, o fato de estas serem provenientes de idosos, praticamente garante que eles não são portadores de doenças genéticas congênitas graves. Além disso, sempre

é melhor comparar amostras de pacientes a controles da mesma população, uma vez que a representatividade das variantes pode depender da ancestralidade (Naslavsky, Yamamoto et al., 2017). No caso das doenças ósseas genéticas, foi criado recentemente um banco de dados chamado 'SkeletonGenetics', para armazenar as variantes já descritas associadas a doenças ósseas genéticas (Chen et al., 2016).

Fora do grupo das doenças monogênicas, nas quais um único fator genético (uma variante rara) é responsável pelo fenótipo, os exames de NGS (WES ou WGS) também permitem avaliar variantes comuns, que em conjunto, associadas a fatores ambientais, são responsáveis por doenças multifatoriais de alta prevalência, como o câncer, a doença coronariana e o diabetes melito. Nesse padrão de herança multifatorial, uma variante aumenta de forma modesta a probabilidade de se desenvolver a doença. Contudo, ainda não sabe quantas variantes são necessárias para o desenvolvimento da doença; tampouco se conhece o peso de cada variante individualmente. À medida que se realizam mais e mais testes genéticos e se compartilham esses dados, tende a ficar mais fácil classificar uma variante rara como patogênica ou não patogênica. Todavia, a interpretação conjunta das variantes comuns demandará esforço adicional por parte dos pesquisadores envolvidos nas pesquisas de etiologia das doenças complexas/multifatoriais.

Na análise dos testes de NGS amplos, o laboratório pode se deparar com variantes em genes associados a doenças que não são o motivo da solicitação do exame (achados incidentais ou secundários). É grande a controvérsia entre os especialistas sobre o que deve ou não ser reportado. Ainda que não haja um consenso universal, a ACMG listou um grupo de doenças potencialmente tratáveis cujas variantes devem ser reportadas, independentemente da razão que levou à realização do exame. Este estudo não encontrou nenhuma variante patogênica em genes responsáveis por tais doenças. Questionamentos ainda mais delicados permanecem, como a possibilidade de serem detectadas, em crianças, variantes que levem ao desenvolvimento de câncer ou de doenças neurodegenerativas na vida adulta, sem que exista tratamento em algumas situações.

5.3. Aplicação da técnica de NGS em amostras sem diagnóstico (etapa C)

Tornou-se atrativo empregar a técnica de NGS na triagem neonatal, pois o exame não só se incorporou à prática clínica a preço acessível, como possibilita a detecção de um amplo espectro de variantes potencialmente causadoras de inúmeras doenças monogênicas. Embora a técnica tenha o potencial de avaliar e diagnosticar, num único experimento, um grande número de doenças, sua aplicação prática continua a esbarrar em questões éticas, psicológicas e econômicas debatidas por diversos especialistas na área de Genética Médica (Reinstein, 2015). Apesar desses questionamentos, projetos-piloto têm sido realizados.

A fibrose cística, doença monogênica de herança autossômica recessiva, é uma das doenças incorporadas aos programas de triagem neonatal em grande parte dos países, inclusive no Brasil. O teste bioquímico realizado para triar casos de fibrose cística, a dosagem do tripsinogênio imunorreativo, apresenta alta sensibilidade, mas baixa especificidade (a taxa de falso-positivos é de 94%). Para contornar esse problema, nos Estados Unidos, a triagem para fibrose cística combina dados moleculares utilizando um painel restrito das mutações mais prevalentes, pela técnica convencional de Sanger. Era natural, portanto, que o gene *CFTR* fosse um dos primeiros a serem avaliados pelo emprego do NGS no contexto da triagem neonatal. Lefterova et al. (2016) utilizaram a nova técnica para avaliar todo o gene *CFTR* em 142 amostras armazenadas de recém-nascidos submetidos à triagem neonatal convencional e 51 amostras com diversas mutações já identificadas em *CFTR*. Após resolverem o problema da realização da técnica de NGS em amostras de papel de filtro, os resultados mostraram uma alta sensibilidade e especificidade, com valor preditivo positivo de 100%. A validação desta técnica no contexto de uma doença na triagem neonatal abre possibilidade para o estudo de outras moléstias. Lefterova et al. (2016) sugerem que se contemplem outras doenças testadas na triagem neonatal, como a deficiência da biotinidase, fenilcetonúria e hemoglobinopatias. No entanto, esses autores afirmam que é preciso vencer vários desafios antes de implementar a

técnica na rotina. Citam, por exemplo, o alto custo de aquisição e manutenção dos equipamentos e a necessidade de capacitar profissionais para aplicarem a técnica e interpretarem os resultados.

Park et al. (2016), na Coreia, desenvolveram um painel customizado com 97 genes para avaliar doenças metabólicas em recém-nascidos submetidos à triagem convencional. Foram selecionados todos os casos positivos na triagem neonatal convencional de 22% dos nascidos vivos no país entre maio de 2013 e julho de 2014. Três grupos foram incluídos no estudo: a) 27 casos positivos com confirmação por Sanger; b) 10 casos negativos e c) 269 amostras com resultados positivos na triagem convencional. No sequenciamento por NGS dos 27 casos positivos, foi obtida uma sensibilidade e especificidade de 100% para 97 variantes alélicas patogênicas. Dentre os 269 pacientes do grupo 'c': c1) em 59 casos foram encontradas mutações patogênicas em genes associados a fenótipos congruentes com as alterações bioquímicas da triagem convencional; c2) em 125 pacientes foram encontradas mutações patogênicas em genes contidos no painel NGS, mas não associados à positividade do teste convencional; c3) em 85 casos não foram encontradas mutações patogênicas, embora a triagem convencional fosse positiva. No grupo de 59 pacientes classificados como positivos (c1), com mutações em genes compatíveis com o resultado da triagem bioquímica convencional, 54 pacientes foram validados por Sanger (o que indica pelo menos 94% de verdadeiro-positivos neste grupo). Um ponto levantado pelos autores foi a possibilidade de diferenciar variantes que alteram o teste bioquímico, mas não causam doença, como a variante Duarte na galactosemia. No grupo de resultados positivos com discrepância entre o teste genético e bioquímico (c2), foram identificados dez casos de mutações bi-alélicas em doenças metabólicas que têm tratamento e que não poderiam ter sido diagnosticadas apenas com a triagem convencional (seriam falso-negativos na triagem convencional).

Devido à inviabilidade logística e financeira, não foi possível iniciar experimentos para validação da técnica de NGS em casuísticas representativas da população brasileira, pois o número necessário de testes para encontrar um verdadeiro positivo é muito grande. Assim, neste estudo, visando à aplicação da técnica do NGS em triagem neonatal, optou-se por um estudo de casos e controles em lugar de coorte. A

desvantagem do estudo de casos e controles é que a escolha dos casos é sempre enviesada de alguma maneira, sobretudo quando o autor necessita de amostras cegas, e por isso, não pode participar da escolha das amostras. O viés das amostras “casos” utilizadas neste estudo é devido principalmente à disponibilidade de amostras de DNA com consentimento para pesquisa no Hospital de Clínicas de Porto Alegre UFRGS; trata-se, portanto, de uma casuística de conveniência. Levando em conta esse viés, é interessante iniciar a interpretação dos resultados de sensibilidade diagnóstica pela análise dos subconjuntos de amostras “casos” com base nas hipóteses iniciais dos pacientes (Quadro 14). A análise realizada no subgrupo de 14 amostras triadas com painel NGS que contavam com confirmação molecular prévia por Sanger é muito semelhante à análise da etapa A; os resultados obtidos também são equivalentes, com 13 confirmações de verdadeiro-positivos e uma amostra cujas variantes patogênicas não poderiam ter sido encontradas, pois o gene mutado não era coberto pelo painel. Já a análise da casuística com hipótese inicial de erros inatos do metabolismo apresentou sensibilidade diagnóstica de 19/26 (73%) ou 19/24 (79%), se forem excluídas as duas amostras com doenças associadas a genes que não constam do painel. A sensibilidade diagnóstica do teste NGS para erros inatos de metabolismo foi mais elevada que para os subgrupos de amostras com distrofias musculares (0/4) e miopatias (2/10). Esse resultado é encorajador, pois os erros inatos do metabolismo constituem grande parte das doenças de interesse para a triagem neonatal, justamente por contarem com tratamento precoce.

Finalmente, a análise da casuística como um todo apresentou sensibilidade diagnóstica de 47% ou 50%, se excluídas as amostras com doenças associadas a genes que não constavam do painel. Essa distinção é importante, pois permite inferir que uma estratégia para aumentar a sensibilidade do teste seria a de utilizar um painel ampliado com mais genes associados a condições de interesse para triagem neonatal. Mas independentemente do percentual considerado, o valor de sensibilidade é baixo se comparado a testes clássicos de triagem neonatal e alto se comparado a testes genéticos diagnósticos. É importante ressaltar que, mesmo com qualidade da hipótese diagnóstica pré-teste nula, dado que a análise foi cega, um bom valor de sensibilidade diagnóstica foi obtido. Para comparar as sensibilidades diagnósticas da pesquisa por

NGS e de um teste de triagem neonatal clássico, como o teste para fenilcetonúria, seria necessário um estudo caso-controle em que todos os casos fossem apenas positivos para fenilcetonúria. No entanto, baseado nas informações prévias de que, em geral, 90% das variantes patogênicas estão na região exônica e de que Shamseldin et al. (2017) demonstram sensibilidade diagnóstica maior que 95% em mais de 100 famílias consanguíneas com estudo de ligação analisadas por WES, é possível estimar que um teste NGS direcionado apenas à fenilcetonúria, uma doença autossômica recessiva na qual sabemos o loco a ser analisado, obteria sensibilidade maior que 90%. Na casuística desta tese, as duas amostras que apresentavam fenilcetonúria foram corretamente identificadas (embora este número seja muito baixo para qualquer possibilidade de análise estatística). A identificação de variantes patogênicas em doenças de padrão autossômico recessivo (ou recessivo ligado ao X) tem uma vantagem muito grande em relação a doenças com padrão de herança dominante devido ao fato de o número de mutações novas, em doenças recessivas, ser praticamente desprezível em relação ao número de mutações herdadas dos pais. Com isso é muito mais provável que variantes que estão presentes na população já tenham sido identificadas e classificadas em bancos de dados públicos de variantes patogênicas do que variantes *de novo*.

Em relação ao número de falso-positivos, tal como nos resultados de Park et al. (2016) e Lefterova et al. (2016), foi encontrado apenas um falso-positivo em 44 controles, com especificidade de 97,73% (IC 95% = 0,8711 a >0,9999). No entanto, diferentemente das análises dos trabalhos citados, em que as buscas de variantes foram focadas apenas em um gene ou em um conjunto de genes de doenças metabólicas, na análise cega do presente estudo foram incluídas todas as doenças com recomendação para triagem neonatal nos EUA e doenças que correspondem a diagnósticos diferenciais com as anteriores. Dessa maneira, era maior o risco de encontrar doenças falso-positivas neste estudo. Além disso, a variante falso-positiva encontrada foi uma variante *missense* no gene *FMR1* [NM_002024:exon5:c.413G>A:p.(Arg138Gln)]. A interpretação de patogenicidade dessa variante, seguindo os critérios da ACMG, foi duvidosa entre provavelmente patogênica e VUS. Por um lado, a variante foi descrita na literatura com estudos

funcionais que sugeriram que levasse à perda de função da proteína. Além disso, é sabido que outras mutações de perda de função no gene *FMR1* estão associadas à síndrome do X-frágil (Myrik et al., 2015). Por outro lado, a frequência da variante é relativamente alta em populações-controle (21 alelos em 199947, sendo 10 hemizigotos no banco de dados gnomAD - <http://gnomad.broadinstitute.org/gene/ENSG00000102081>). No entanto, os critérios da ACMG não são específicos quanto à frequência alélica exata, no caso de doenças recessivas, para se considerar um critério moderado de patogenicidade (PM2 - Quadro 4). Se a variante tivesse sido classificada como VUS, o painel NGS com análise cega teria apresentado especificidade de 100%. Também, é importante ressaltar que mais uma amostra controle foi classificada como positiva na análise cega. No entanto, essa amostra havia sido testada para genes associados à síndrome de Marfan (*FBN1*, *TGFBR1* e *TGFBR2*), com resultado negativo. Na análise aqui realizada, uma variante patogênica e uma variante provavelmente patogênica foram encontradas no gene *CBS*, que está associado à homocistinúria, uma condição que constitui diagnóstico diferencial da síndrome de Marfan. Diante desse novo resultado, o médico solicitante do exame foi re-contactado e informou que, após avaliação por hematologista, foi confirmado o diagnóstico de homocistinúria por dosagem bioquímica. Em consequência, essa amostra foi classificada como verdadeiro-positiva, o que fortalece a tese de que o teste de painel NGS tem alta especificidade mesmo em contexto de análise cega. Finalmente é necessário lembrar que o critério que diminuiu o número de falso-positivos sem diminuir o número de verdadeiro-positivos foi o de relatar como positivos os casos de hipótese de padrão de herança autossômico recessivo que apresentassem duas ou mais variantes raras patogênicas ou provavelmente patogênicas em um mesmo gene. O critério de reportar como positivos os casos de hipótese de padrão de herança autossômico recessivo que apresentassem uma variante patogênica ou provavelmente patogênica e outra variante VUS em um mesmo gene, levaria a três falso-positivos e não incluiria nenhum verdadeiro-positivo, piorando a especificidade sem melhorar a sensibilidade do teste. Dado que um dos critérios da ACMG de patogenicidade de variantes é justamente a variante ser detectada em *trans* com uma variante patogênica (PM3 - Quadro 4), o resultado anterior não era esperado.

Há duas justificativas possíveis para essa discrepância: a) a variante VUS não está em *trans* em relação à variante patogênica (informação que não foi extraída dos dados de NGS neste trabalho); b) em análises de diagnóstico, os pacientes sempre apresentam algum fenótipo e assim partem de uma probabilidade maior de que as variantes patogênicas presentes em seu genoma estejam associadas a este fenótipo.

Os resultados do presente estudo, com sensibilidade de 50 a 70% e especificidade próxima a 100%, sugerem que o teste de sequenciamento NGS é adequado para ser empregado na triagem neonatal em complementação aos testes bioquímicos, que usualmente apresentam alta sensibilidade e baixa especificidade.

No entanto, vale lembrar que a adequação do teste não é condição suficiente para que ele seja incluído em programa de triagem neonatal. Essa inclusão demanda a solução de outras questões, tais como aconselhamento pré- e pós-teste, tratamento pós-teste, análise do custo para o sistema de saúde, definição das condições a serem diagnosticadas pela nova técnica etc. É preciso encorajar debates entre especialistas e gestores sobre tais questões no Brasil e no mundo.

5.4. Novas técnicas de sequenciamento e opções de testes genéticos

Nos últimos anos, surgiram duas opções de sequenciamento genético que não estavam disponíveis em 2013, quando do início desta tese. São elas: o sequenciamento do genoma completo (WGS) (pela técnica de NGS/sequenciamento de segunda geração) e o que vem sendo chamado de “sequenciamento de terceira geração” ou “SMRT *sequencing*” (do inglês *Single Molecule Real Time sequencing*) (Heather e Chain, 2016).

Além de permitir a identificação de variantes intrônicas, a grande vantagem do WGS está em facilitar a análise de CNVs e variantes estruturais. Em contrapartida, o WGS tem custos mais altos e, isoladamente, possibilita a detecção de uma quantidade muito pequena (menor que 5%) de variantes patogênicas além daquelas obtidas pelo WES. Isso porque, além de mais de 90% das variantes patogênicas para doenças

mendelianas estarem nas regiões exônicas ou bordas de *splicing* cobertas pelo WES (Shamseldin et al., 2017), a classificação de patogenicidade para variantes que não alteram proteínas é muito mais difícil e desconhecida. Ainda que a lista final de variantes por WGS seja muito maior em números absolutos, a identificação de variantes patogênicas nesta lista, não seria muito maior do que as variantes patogênicas identificadas por WES. Levando em consideração as justificativas anteriores, uma das propostas atuais de vários laboratórios de análises clínicas, inclusive do CEGH-CEL consiste em oferecer um teste de custo mais acessível que inclui apenas os genes sabidamente associados a doenças mendelianas (em torno de 6700), em vez de aumentar o custo do teste com a oferta do WGS ou de sequenciar todos os genes do exoma (em torno de 20.000). A lógica é semelhante à aplicada para justificar a escolha de WES em detrimento de WGS: o teste de “exoma focado em genes de interesse clínico” tem metade do custo do WES e, com as informações atuais da literatura, encontra praticamente o mesmo número de variantes patogênicas em análises clínicas. Obviamente, essa estratégia não se aplica a amostras de pesquisa e tampouco quando é possível sequenciar mais de um membro da família para tentar identificar variantes e genes ainda não descritos. Outra desvantagem da abordagem focada é que restringe a possibilidade de reanálise de novos genes com os dados do exame já realizado, quando novos genes vêm a ser identificados.

O chamado sequenciamento de terceira geração ou SMRT *sequencing* foi introduzido no mercado pela empresa Pacific Biosciences com o sequenciador PacBio RS (Pacific Biosciences, Menlo Park, Califórnia, EUA). Além dela, uma empresa britânica chamada Oxford Nanopore Technologies está produzindo sequenciadores baseados na tecnologia SMRT seq. (MinION, GridION, PromethION e SmidgION - Oxford Nanopore Technologies, Oxford, Reino Unido), ou seja, sequenciamento de moléculas únicas em tempo real. Embora ainda muito custosa e, no caso da Oxford Nanopore, ainda um tanto quanto experimental, essa tecnologia promete resolver com um único teste diversas limitações apresentadas pelo sequenciamento Sanger e pelo NGS. Esses novos sequenciadores teriam a capacidade de analisar regiões repetitivas do genoma e alterações de metilação, que não são passíveis de identificação pelo sequenciamento de nova geração. Além disso, os *reads* longos facilitam a identificação

de alterações estruturais no genoma e possibilitam estabelecer haplótipos (atribuir fase aos genótipos) e, conseqüentemente, a classificação de *cis* ou *trans* para as variantes identificadas, essencial no contexto da triagem neonatal. Dessa maneira, haveria uma abrangência maior de doenças diagnosticadas, incluindo aquelas com mutações dinâmicas (expansão de trinucleotídeos), como a distrofia muscular tipo Steinert e a síndrome do X-frágil, ou com defeitos na metilação, como as síndromes de Silver-Russell e Beckwith-Wiedemann. Os *reads* longos também auxiliam no diagnóstico da amiotrofia espinhal progressiva, na qual o gene responsável pela doença, *SMN1*, apresenta grande homologia ao gene *SMN2*, impossibilitando a diferenciação entre esses genes pela técnica de NGS. Finalmente essa nova tecnologia também seria de grande utilidade em contexto de triagem neonatal pois permitiria classificar de imediato se duas variantes patogênicas identificadas em um mesmo gene estão em *trans* (e conseqüentemente provavelmente causais de doença autossômica recessiva) ou estão em *cis* e possivelmente não causam doença.

6. Conclusões

A técnica de NGS apresenta alta sensibilidade e alta especificidade na detecção de variantes de ponto, comparável à técnica de Sanger. Tanto a sensibilidade quanto a especificidade podem ser melhoradas se observados critérios de qualidade do sequenciamento NGS para interpretação de variantes e indicação de re-sequenciamento. Os principais critérios de qualidade do sequenciamento NGS são cobertura, balanço alélico e *strand bias*.

A técnica de NGS apresentou diversas vantagens em relação à técnica de Sanger: a) menor risco de *allele dropout*; b) possibilidade de detecção de variantes do tipo CNV; c) maior facilidade e menor custo para análise de genes múltiplos; d) anotação das variantes por *pipeline* de bioinformática mais rápida e automatizada.

Casuísticas de pacientes com doenças de padrão de herança mendelianas podem ser investigadas clinicamente pela técnica de NGS com alta sensibilidade diagnóstica (*diagnostic yield*). Utilizando critérios baseados na qualidade e objetividade da hipótese clínica, na casuística de doenças ósseas genéticas, foram obtidos valores de sensibilidade diagnóstica todos próximos de 60% para teste por Sanger, painel customizado NGS ou WES. A sensibilidade diagnóstica depende mais da qualidade da hipótese clínica, e adequada indicação/abrangência do teste do que da técnica de sequenciamento (Sanger ou NGS) a ser empregada.

O teste de WES é adequado para identificar novas associações entre genes e fenótipos mendelianos utilizando grupos de pacientes ou famílias. Nesse estudo foram identificados três novos genes para fenótipos mendelianos, um em uma displasia esquelética e dois associados à síndrome de Noonan. A identificação de um quarto gene novo, de uma outra displasia esquelética, em colaboração com pesquisadores do exterior, está em processo final de aceitação para publicação.

O teste de painel NGS com genes de interesse de triagem neonatal apresenta sensibilidade diagnóstica de pelo menos 50%, chegando a mais de 70% para as hipóteses de erros inatos do metabolismo, mesmo quando analisado sem informações

da hipótese diagnóstica. A sensibilidade diagnóstica nessa análise é melhor para doenças de padrão recessivo, cujas variantes patogênicas apresentam maior probabilidade de estarem descritas em bancos de variantes populacionais. O critério de análise que permitiu maior especificidade sem prejuízo de sensibilidade foi o de não classificar como positivas as amostras com variantes VUS, mesmo quando a VUS estivesse presente em um gene associado a doença de herança autossômica recessiva com uma outra variante patogênica já identificada. A sensibilidade diagnóstica obtida é alta o suficiente para autorizar a conclusão de que testes por NGS são adequados para programas de triagem neonatal que investiguem múltiplas condições simultaneamente.

7. Anexos

Anexo I: Lista de 230 genes do painel NGS 'P1'

Anexo II: Lista de 465 genes do painel NGS 'P2'

Anexo III: Lista de 485 genes do painel NGS 'P3'

Anexo IV: Lista de 496 genes do painel NGS 'P4'

Anexo V: Lista de 56 genes de achados incidentais ACMG 2013

Anexo VI: Lista de 59 genes de achados secundários ACMG 2016 (Inclusão dos genes: *BMPR1A* e *SMAD4*, associados polipose juvenil; *ATP7B*, associado a doença de Wilson; *OTC*, associado a deficiência de ornitina trasncarbamilase. Removido: *MYLK*, associado a aneurisma de aorta.

Anexo I: Lista de 230 genes do painel NGS 'P1'

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
AARS	601065	CRYAB	123590	FXN	606829	LEMD3	607844	POMGNT1	606822	SHOX	312865
ACP5	171640	CTSK	601105	GABRA5	137142	LFNG	602576	POMT1	607423	SLC17A6	607563
ACTA1	102610	CUL7	609577	GABRB3	137192	LIFR	151443	POMT2	607439	SLC25A12	603667
ADAMTSL2	612277	CYFIP1	606322	GABRG1	137166	LITAF	603795	POR	124015	SLC26A2	606718
ALPL	171760	CYP7B1	603711	GAD1	605363	LMNA	150330	PIIB	123841	SLC6A4	182138
ALX1	601527	DAG1	128239	GAD2	138275	LRP5	603506	PRKCA	176960	SLC9A6	300231
ALX3	606014	DDR2	191311	GALNS	612222	LRSAM1	610933	PRPS1	311850	SMARCAL1	606622
ALX4	605420	DES	125660	GARS	600287	MAOA	309850	PRX	605725	SOST	605740
APTX	606350	DIAPH3	614567	GCH1	600225	MAPK3	601795	PTEN	601728	SOX9	608160
ASMT	300015	DLL3	602768	GDAP1	606598	MARK1	606511	PTH1R	168468	SP7	606633
ASPA	608034	DMD	300377	GDF5	601146	MATN3	602109	RAB23	606144	SPAST	604277
ASTN2	612856	DNAJB6	611332	GDF6	601147	MBD2	603547	RAB7A	602298	SPG11	610844
ATL1	606439	DNM2	602378	GJA1	121014	MDGA2	611128	RAI1	607642	SPG7	602783
BCL2	151430	DPP10	608209	GJB1	304040	MECP2	300005	RBFOX1	605104	SRY	480000
BDNF	113505	DRD4	126452	GLB1	611458	MED25	610197	RECQL4	603780	STS	300747
BIN1	601248	DYM	607461	GLU3	165240	MESP2	605195	REEP1	609139	STXBP6	607958
CACNA1C	114205	DYNC1H1	600112	GNAS	139320	MET	164860	RELN	600514	TCAP	604488
CACNA1H	607904	DYNC2H1	603297	GPR85	605188	MFN2	608507	RIC3	610509	TCOF1	606847
CAPN3	114240	DYSF	603009	GRIA3	305915	MMP13	600108	RMRP	157660	TGFBF1	190181
CAV3	601253	EGR2	129010	GRIK2	138244	MMP9	120361	RUNX2	600211	TGFBF2	190182
CFTR	602421	EIF2AK3	604032	GRIN2A	138253	MPZ	159440	RYR1	180901	TNFRSF11A	603499
CHRNA7	118511	EMD	300384	GRM5	604102	MSX2	123101	SACS	604490	TP63	603273
CHRND	100720	EN2	131310	GRM8	601116	MTM1	300415	SALL1	602218	TPM2	190990
CHRNA7	100730	ERBB4	600543	GRPR	305670	MTMR2	603557	SBDS	607444	TPM3	191030
CHST3	603799	ERG	165080	HES7	608059	MYOT	604103	SBF2	607697	TRAPPC2	300202
CLCN1	118425	EVC	604831	HOXA1	142955	NDRG1	605262	SCN1A	182389	TRIP11	604505
CNTN4	607280	EVC2	607261	HSPB1	602195	NEFL	162280	SELENON	606210	TRPV4	605427
CNTNAP2	604569	FBN1	134797	HSPB8	608014	NIPA1	608145	SEMA3A	603961	TWIST1	601622
COL10A1	120110	FGD4	611104	IFT80	611177	NKX3-2	602183	SEMA5A	609297	UBE3A	601623
COL11A1	120280	FGFR1	136350	IL11RA	600939	NLGN3	300336	SERPINH1	600943	ZNF778	.
COL11A2	120290	FGFR2	176943	IMMP2L	605977	NLGN4X	300427	SETX	608465		
COL18A1	120328	FGFR3	134934	ITGB3	173470	NOG	602991	SEZ6L2	616667		
COL1A1	120150	FHL1	300163	JAG1	601920	NOTCH3	600276	SGCA	600119		
COL1A2	120160	FIG4	609390	JAK2	147796	NPR2	108961	SGCB	600900		
COL2A1	120140	FKBP10	607063	KIF1B	605995	NRXN1	600565	SGCD	601411		
COL9A1	120210	FKRP	606596	KIF5A	602821	NSD1	606681	SGCG	608896		
COL9A2	120260	FKTN	607440	L1CAM	308840	P3H1	610339	SH3BP2	602104		
COL9A3	120270	FLNA	300017	LAMA2	156225	PAFAH1B1	601545	SH3TC2	608206		
COMP	600310	FLNB	603381	LAMB1	150240	PCDH19	300460	SHANK2	603290		
CRTAP	605497	FOXP2	605317	LARGE1	603590	PMP22	601097	SHANK3	606230		

Anexo II: Lista de 465 genes do painel NGS ‘P2’

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
AARS	601065	BMPRI1A	601299	COL6A2	120240	DYSF	603009	FMR1	309550	HSPB1	602195
ABCD1	300371	BRCA1	113705	COL6A3	120250	EFNB1	300035	FOLR1	136430	HSPB8	608014
ACADM	607008	BRCA2	600185	COL9A1	120210	EGR2	129010	FOXE1	602617	IDUA	252800
ACADVL	609575	BRIP1	605882	COL9A2	120260	EIF2AK3	604032	FOXG1	164874	IFT80	611177
ACAT1	607809	BTBD	609019	COL9A3	120270	ELN	130160	FOXP2	605317	IL11RA	600939
ACP5	171640	CACNA1S	114208	COMP	600310	EMD	300384	FTCD	606806	IL2RG	308380
ACTA1	102610	CANT1	613165	COQ2	609825	EPCAM	185535	FUS	137070	IL7R	146661
ADA	608958	CAPN3	114240	COQ8A	606980	EPHA3	179611	FXN	606829	IRF6	607199
ADAMTS18	607512	CAV3	601253	COQ9	612837	ERF	611888	GADD45G	604949	ITGA7	600536
ADAMTSL2	612277	CBS	613381	CP	117700	ERG	165080	GALNS	612222	IVD	607036
AFG3L2	604581	CD3D	186790	CPS1	608307	ETHE1	608451	GALNT1	602273	JAG1	601920
ALDH3A2	609523	CD3E	186830	CREBBP	600140	EVC	604831	GALNT12	610290	JAK3	600173
ALDH7A1	107323	CDH1	192090	CRISPLD2	612434	EVC2	607261	GALT	606999	JAM2	606870
ALPL	171760	CDH2	114020	CRTAP	605497	EXO1	606063	GAMT	601240	KANSL1	612452
ALX1	601527	CDH23	605516	CRYAA	123580	EXT1	608177	GARS	600287	KATNAL2	614697
ALX3	606014	CDH7	605806	CRYAB	123590	FAF1	604460	GATM	602360	KCNE1	176261
ALX4	605420	CDH8	603008	CTDP1	604927	FAH	613871	GCDH	608801	KIF22	603213
AMN	605799	CDK4	123829	CTNNB1	116806	FANCA	607139	GCH1	600225	KIF5A	602821
ANO5	608662	CDKL5	300203	CTSK	601105	FANCC	613899	GDAP1	606598	L1CAM	308840
APC	611731	CDKN2A	600160	CUBN	602997	FANCD2	613984	GDF5	601146	LAMA2	156225
APTX	606350	CFTR	602421	CUL7	609577	FANCE	613976	GDF6	601147	LAMP2	309060
ARG1	608313	CHEK2	604373	CYP21A2	613815	FANCF	613897	GIF	609342	LARGE1	603590
ARHGAP29	610496	CHST14	608429	CYP27A1	606530	FANCG	602956	GJA1	121014	LEMD3	607844
ARX	300382	CHST3	603799	CYP7B1	603711	FBLN5	604580	GJB1	304040	LFNG	602576
ASL	608310	CLCN1	118425	DAG1	128239	FBN1	134797	GJB2	121011	LIFR	151443
ASPA	608034	CLDN14	605608	DBT	248610	FECH	612386	GJB3	603324	LIPI	609252
ASS1	603470	CLPTM1L	612585	DCLRE1C	605988	FGD4	611104	GJB6	604418	LITAF	603795
ATL1	606439	CLRN1	606397	DDR2	191311	FGF8	600483	GLB1	611458	LMAN1	601567
ATM	607585	CNTNAP2	604569	DES	125660	FGFR1	136350	GLI2	165230	LMBRD1	612625
ATP7A	300011	COL10A1	120110	DHCR24	606418	FGFR2	176943	GLI3	165240	LMNA	150330
ATP7B	606882	COL11A1	120280	DHCR7	602858	FGFR3	134934	GNAI3	139370	LRP5	603506
ATR	601215	COL11A2	120290	DLAT	608770	FHL1	300163	GNAS	139320	LRSAM1	610933
AXIN2	604025	COL18A1	120328	DLL3	602768	FIG4	609390	GRIK1	138245	MATN3	602109
B4GALT6	604017	COL1A1	120150	DMD	300377	FILIP1L	612993	HADHA	600890	MBD2	603547
BAP1	603089	COL1A2	120160	DNAJB6	611332	FKBP10	607063	HADHB	143450	MBD5	611472
BARD1	601593	COL2A1	120140	DNM2	602378	FKBP14	614505	HBB	141900	MC2R	607397
BCKDHA	608348	COL3A1	120180	DUOX2	606759	FKRP	606596	HERC2	605837	MCCC1	609010
BCKDHB	248611	COL5A1	120215	DYM	607461	FKTN	607440	HES7	608059	MCCC2	609014
BIN1	601248	COL5A2	120190	DYNC2H1	603297	FLNA	300017	HLCS	609018	MECP2	300005
BMP4	112262	COL6A1	120220	DYRK1A	600855	FLNB	603381	HMGCL	613898	MED25	610197

Anexo II: Lista de 465 genes do painel NGS 'P2', continuação

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
<i>MEN1</i>	613733	<i>NAGS</i>	608300	<i>PDSS1</i>	607429	<i>RAG2</i>	179616	<i>SLC25A15</i>	603861	<i>TP53</i>	191170
<i>MESP2</i>	605195	<i>NBN</i>	602667	<i>PDSS2</i>	610564	<i>RAI1</i>	607642	<i>SLC26A2</i>	606718	<i>TP63</i>	603273
<i>MFN2</i>	608507	<i>NDN</i>	602117	<i>PFKL</i>	171860	<i>RARA</i>	180240	<i>SLC26A4</i>	605646	<i>TPM2</i>	190990
<i>MID1</i>	300552	<i>NDRG1</i>	605262	<i>PHF8</i>	300560	<i>RBBP8</i>	604124	<i>SLC2A1</i>	138140	<i>TPM3</i>	191030
<i>MIR1205</i>	.	<i>NDUFV2</i>	600532	<i>PHGDH</i>	606879	<i>RECQL4</i>	603780	<i>SLC5A5</i>	601843	<i>TPO</i>	606765
<i>MIR1206</i>	.	<i>NECTIN1</i>	600644	<i>PLCB4</i>	600810	<i>REEP1</i>	609139	<i>SMAD4</i>	600993	<i>TRAPPC2</i>	300202
<i>MIR1207</i>	.	<i>NEFL</i>	162280	<i>PLOD1</i>	153454	<i>RET</i>	164761	<i>SMARCAL1</i>	606622	<i>TRIM32</i>	602290
<i>MIR1208</i>	.	<i>NF1</i>	613113	<i>PMP22</i>	601097	<i>RGMA</i>	607362	<i>SNAP29</i>	604202	<i>TRIP11</i>	604505
<i>MIR140</i>	611894	<i>NIPA1</i>	608145	<i>PMS1</i>	600258	<i>RHD</i>	111680	<i>SNRPN</i>	182279	<i>TRPC6</i>	603652
<i>MIR200B</i>	612091	<i>NIPBL</i>	608667	<i>PMS2</i>	600259	<i>RMRP</i>	157660	<i>SOD1</i>	147450	<i>TRPS1</i>	604386
<i>MLH1</i>	120436	<i>NKX3-2</i>	602183	<i>POGZ</i>	614787	<i>ROR2</i>	602337	<i>SOST</i>	605740	<i>TRPV4</i>	605427
<i>MMAA</i>	607481	<i>NOG</i>	602991	<i>POLR1C</i>	610060	<i>RPS27L</i>	612055	<i>SOX9</i>	608160	<i>TSC1</i>	605284
<i>MMAB</i>	607568	<i>NOL8</i>	611534	<i>POLR1D</i>	613715	<i>RTTN</i>	610436	<i>SP7</i>	606633	<i>TSC2</i>	191092
<i>MMACHC</i>	609831	<i>NOTCH3</i>	600276	<i>POMGNT1</i>	606822	<i>RUNX1</i>	151385	<i>SPAST</i>	604277	<i>TSHB</i>	188540
<i>MMADHC</i>	611935	<i>NPC1</i>	607623	<i>POMT1</i>	607423	<i>RUNX2</i>	600211	<i>SPG11</i>	610844	<i>TSHR</i>	603372
<i>MMP13</i>	600108	<i>NPC2</i>	601015	<i>POMT2</i>	607439	<i>RYR1</i>	180901	<i>SPG7</i>	602783	<i>TSHZ1</i>	614427
<i>MMP9</i>	120361	<i>NPR2</i>	108961	<i>POR</i>	124015	<i>SALL1</i>	602218	<i>SPR</i>	182125	<i>TSPEAR</i>	612920
<i>MOCs1</i>	603707	<i>NSD1</i>	606681	<i>POTED</i>	607549	<i>SATB2</i>	608148	<i>SRY</i>	480000	<i>TWIST1</i>	601622
<i>MOCs2</i>	603708	<i>NT5C3A</i>	606224	<i>PPIB</i>	123841	<i>SBDS</i>	607444	<i>STK11</i>	602216	<i>UBE3A</i>	601623
<i>MPZ</i>	159440	<i>NTNG1</i>	608818	<i>PQBP1</i>	300463	<i>SBF2</i>	607697	<i>SUMO1</i>	601912	<i>USH1C</i>	605242
<i>MRAP</i>	609196	<i>OFD1</i>	300170	<i>PRPS1</i>	311850	<i>SCN4A</i>	603967	<i>TARDBP</i>	605078	<i>USH1G</i>	607696
<i>MRE11</i>	600814	<i>OGN</i>	602383	<i>PRSS1</i>	276000	<i>SCO2</i>	604272	<i>TBX1</i>	602054	<i>USH2A</i>	608400
<i>MSH2</i>	609309	<i>OTC</i>	300461	<i>PRX</i>	605725	<i>SELENON</i>	606210	<i>TBX22</i>	300307	<i>USHBP1</i>	611810
<i>MSH3</i>	600887	<i>OTOF</i>	603681	<i>PSAT1</i>	610936	<i>SERPINH1</i>	600943	<i>TCAP</i>	604488	<i>USP25</i>	.
<i>MSH6</i>	600678	<i>OXCT1</i>	601424	<i>PSPH</i>	172480	<i>SETBP1</i>	611060	<i>TCF12</i>	600480	<i>VAPB</i>	605704
<i>MSTN</i>	601788	<i>P3H1</i>	610339	<i>PTCH1</i>	601309	<i>SETX</i>	608465	<i>TCF4</i>	602272	<i>VHL</i>	608537
<i>MSX1</i>	142983	<i>PABPN1</i>	602279	<i>PTEN</i>	601728	<i>SGCA</i>	600119	<i>TCOF1</i>	606847	<i>WHRN</i>	607928
<i>MSX2</i>	123101	<i>PAFAH1B1</i>	601545	<i>PTH1R</i>	168468	<i>SGCB</i>	600900	<i>TERT</i>	187270	<i>WNT3</i>	165330
<i>MTHFR</i>	607093	<i>PAH</i>	612349	<i>PTS</i>	612719	<i>SGCD</i>	601411	<i>TFAP2A</i>	107580	<i>XRCC3</i>	600675
<i>MTM1</i>	300415	<i>PALB2</i>	610355	<i>PVT1</i>	165140	<i>SGCG</i>	608896	<i>TG</i>	188450	<i>ZAP70</i>	176947
<i>MTMR2</i>	603557	<i>PAPSS2</i>	603005	<i>PYGM</i>	608455	<i>SH3BP2</i>	602104	<i>TGFBR1</i>	190181		
<i>MTR</i>	156570	<i>PAX8</i>	167415	<i>QDPR</i>	612676	<i>SHANK3</i>	606230	<i>TGFBR2</i>	190182		
<i>MTRR</i>	602568	<i>PCBD1</i>	126090	<i>RAB23</i>	606144	<i>SHH</i>	600725	<i>TGIF1</i>	602630		
<i>MUT</i>	609058	<i>PCCA</i>	232000	<i>RAB7A</i>	602298	<i>SHOX</i>	312865	<i>TH</i>	191290		
<i>MUTYH</i>	604933	<i>PCCB</i>	232050	<i>RAD50</i>	604040	<i>SIX3</i>	603714	<i>TMC1</i>	606706		
<i>MYC</i>	190080	<i>PCDH15</i>	605514	<i>RAD51</i>	179617	<i>SLC17A6</i>	607563	<i>TMIE</i>	607237		
<i>MYO6</i>	600970	<i>PDGFRA</i>	173490	<i>RAD51C</i>	602774	<i>SLC19A3</i>	606152	<i>TMPRSS15</i>	606635		
<i>MYO7A</i>	276903	<i>PDHA1</i>	300502	<i>RAD51D</i>	602954	<i>SLC22A5</i>	603377	<i>TMPRSS3</i>	605511		
<i>MYOT</i>	604103	<i>PDHX</i>	608769	<i>RAG1</i>	179615	<i>SLC25A13</i>	603859	<i>TNFRSF11A</i>	603499		

Anexo III: Lista de 485 genes do painel NGS 'P3'

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
AARS	601065	BCKDHB	248611	COL2A1	120140	DUOX2	606759	FKRP	606596	HADHB	143450
ABCD1	300371	BIN1	601248	COL3A1	120180	DYM	607461	FKTN	607440	HBB	141900
ACADM	607008	BMP4	112262	COL5A1	120215	DYNC2H1	603297	FLNA	300017	HERC2	605837
ACADVL	609575	BMPRI1A	601299	COL5A2	120190	DYRK1A	600855	FLNB	603381	HES7	608059
ACAT1	607809	BRAF	164757	COL6A1	120220	DYSF	603009	FLNC	102565	HLCS	609018
ACPS	171640	BRCA1	113705	COL6A2	120240	EDN1	131240	FMR1	309550	HMGCL	613898
ACTA1	102610	BRCA2	600185	COL6A3	120250	EFNB1	300035	FOLR1	136430	HNRNPDL	607137
ACTB	102630	BRIP1	605882	COL9A1	120210	EGR2	129010	FOXE1	602617	HRA5	190020
ACTG1	102560	BTD	609019	COL9A2	120260	EIF2AK3	604032	FOXG1	164874	HSPB1	602195
ADA	608958	CACNA1S	114208	COL9A3	120270	EIF4A3	608546	FOXP2	605317	HSPB8	608014
ADAMTS18	607512	CANT1	613165	COMP	600310	ELN	130160	FTCD	606806	IDUA	252800
ADAMTS2	612277	CAPN3	114240	COQ2	609825	EMD	300384	FUS	137070	IFT80	611177
ADCK3	.	CAV3	601253	COQ9	612837	EPCAM	185535	FXN	606829	IL11RA	600939
AFG3L2	604581	CBS	613381	CP	117700	EPHA3	179611	GADD45G	604949	IL2RG	308380
ALDH3A2	609523	CD3D	186790	CPS1	608307	ERF	611888	GALNS	612222	IL7R	146661
ALDH7A1	107323	CD3E	186830	CREBBP	600140	ERG	165080	GALNT1	602273	IRF6	607199
ALPL	171760	CDH1	192090	CRISPLD2	612434	ETHE1	608451	GALNT12	610290	ITGA7	600536
ALX1	601527	CDH2	114020	CRTAP	605497	EVC	604831	GALT	606999	IVD	607036
ALX3	606014	CDH23	605516	CRYAA	123580	EVC2	607261	GAMT	601240	JAG1	601920
ALX4	605420	CDH7	605806	CRYAB	123590	EXT1	608177	GARS	600287	JAK3	600173
AMN	605799	CDH8	603008	CTDP1	604927	FAF1	604460	GATM	602360	JAM2	606870
ANO5	608662	CDK4	123829	CTNNB1	116806	FAH	613871	GCDH	608801	KATNAL2	614697
APC	611731	CDKL5	300203	CTSK	601105	FANCA	607139	GCH1	600225	KBTBD13	613727
APTX	606350	CDKN2A	600160	CUBN	602997	FANCC	613899	GDAP1	606598	KCNE1	176261
ARG1	608313	CFL1	601442	CUL7	609577	FANCD2	613984	GDF5	601146	KIF22	603213
ARHGAP29	610496	CFTR	602421	CYP21A2	613815	FANCE	613976	GDF6	601147	KIF5A	602821
ARX	300382	CHEK2	604373	CYP27A1	606530	FANCF	613897	GIF	609342	KRAS	190070
ASL	608310	CHST14	608429	CYP7B1	603711	FANCG	602956	GJA1	121014	L1CAM	308840
ASPA	608034	CHST3	603799	DAG1	128239	FBLN5	604580	GJB1	304040	LAMA2	156225
ASS1	603470	CLCN1	118425	DBT	248610	FBN1	134797	GJB2	121011	LAMP2	309060
ATL1	606439	CLDN14	605608	DCLRE1C	605988	FECH	612386	GJB3	603324	LARGE1	603590
ATM	607585	CLPTM1L	612585	DDR2	191311	FGD4	611104	GJB6	604418	LEMD3	607844
ATP7A	300011	CLRN1	606397	DES	125660	FGF8	600483	GLB1	611458	LFNG	602576
ATP7B	606882	CNTNAP2	604569	DHCR24	606418	FGFR1	136350	GLI2	165230	LIFR	151443
ATR	601215	COL10A1	120110	DHCR7	602858	FGFR2	176943	GLI3	165240	LIPI	609252
AXIN2	604025	COL11A1	120280	DLAT	608770	FGFR3	134934	GNAI3	139370	LITAF	603795
B4GALT6	604017	COL11A2	120290	DLL3	602768	FHL1	300163	GNAS	139320	LMAN1	601567
BAP1	603089	COL18A1	120328	DMD	300377	FIG4	609390	GRHL3	608317	LMBRD1	612625
BARD1	601593	COL1A1	120150	DNAJB6	611332	FKBP10	607063	GRIK1	138245	LMNA	150330
BCKDHA	608348	COL1A2	120160	DNM2	602378	FKBP14	614505	HADHA	600890	LRP5	603506

Anexo III: Lista de 485 genes do painel NGS 'P3', continuação

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
<i>LRSAM1</i>	610933	<i>MTMR2</i>	603557	<i>PAX8</i>	167415	<i>QDPR</i>	612676	<i>SHANK3</i>	606230	<i>TGFBR1</i>	190181
<i>LZTR1</i>	600574	<i>MTR</i>	156570	<i>PCBD1</i>	126090	<i>RAB23</i>	606144	<i>SHH</i>	600725	<i>TGFBR2</i>	190182
<i>MAP2K1</i>	176872	<i>MTRR</i>	602568	<i>PCCA</i>	232000	<i>RAB7A</i>	602298	<i>SHOC2</i>	602775	<i>TGIF1</i>	602630
<i>MATN3</i>	602109	<i>MUT</i>	609058	<i>PCCB</i>	232050	<i>RABEP2</i>	611869	<i>SHOX</i>	312865	<i>TH</i>	191290
<i>MBD2</i>	603547	<i>MUTYH</i>	604933	<i>PCDH15</i>	605514	<i>RAD50</i>	604040	<i>SIX3</i>	603714	<i>TMC1</i>	606706
<i>MBD5</i>	611472	<i>MYC</i>	190080	<i>PDGFRA</i>	173490	<i>RAD51</i>	179617	<i>SLC17A6</i>	607563	<i>TMIE</i>	607237
<i>MC2R</i>	607397	<i>MYO6</i>	600970	<i>PDHA1</i>	300502	<i>RAD51C</i>	602774	<i>SLC19A3</i>	606152	<i>TMPRSS15</i>	606635
<i>MCCC1</i>	609010	<i>MYO7A</i>	276903	<i>PDHX</i>	608769	<i>RAD51D</i>	602954	<i>SLC22A5</i>	603377	<i>TMPRSS3</i>	605511
<i>MCCC2</i>	609014	<i>MYOT</i>	604103	<i>PDSS1</i>	607429	<i>RAF1</i>	164760	<i>SLC25A13</i>	603859	<i>TNFRSF11A</i>	603499
<i>MECP2</i>	300005	<i>NAGS</i>	608300	<i>PDSS2</i>	610564	<i>RAG1</i>	179615	<i>SLC25A15</i>	603861	<i>TP53</i>	191170
<i>MED25</i>	610197	<i>NBN</i>	602667	<i>PFKL</i>	171860	<i>RAG2</i>	179616	<i>SLC26A2</i>	606718	<i>TP63</i>	603273
<i>MEGF10</i>	612453	<i>NDN</i>	602117	<i>PHF8</i>	300560	<i>RAI1</i>	607642	<i>SLC26A4</i>	605646	<i>TPM2</i>	190990
<i>MEN1</i>	613733	<i>NDRG1</i>	605262	<i>PHGDH</i>	606879	<i>RARA</i>	180240	<i>SLC2A1</i>	138140	<i>TPM3</i>	191030
<i>MESPD</i>	605195	<i>NDUFV2</i>	600532	<i>PLCB4</i>	600810	<i>RBBP8</i>	604124	<i>SLC5A5</i>	601843	<i>TPO</i>	606765
<i>MFN2</i>	608507	<i>NECTIN1</i>	600644	<i>PLEC</i>	601282	<i>RECQL4</i>	603780	<i>SMAD4</i>	600993	<i>TRAPPC2</i>	300202
<i>MID1</i>	300552	<i>NEFL</i>	162280	<i>PLOD1</i>	153454	<i>REEP1</i>	609139	<i>SMARCA11</i>	606622	<i>TRIM32</i>	602290
<i>MIR1206</i>	.	<i>NF1</i>	613113	<i>PMP22</i>	601097	<i>RET</i>	164761	<i>SNAP29</i>	604202	<i>TRIP11</i>	604505
<i>MIR1207</i>	.	<i>NIPA1</i>	608145	<i>PMS2</i>	600259	<i>RGMA</i>	607362	<i>SNRPN</i>	182279	<i>TRPC6</i>	603652
<i>MIR1208</i>	.	<i>NIPBL</i>	608667	<i>POGZ</i>	614787	<i>RIT1</i>	609591	<i>SOD1</i>	147450	<i>TRPS1</i>	604386
<i>MIR140</i>	611894	<i>NKX3-2</i>	602183	<i>POLR1C</i>	610060	<i>RMRP</i>	157660	<i>SOS1</i>	182530	<i>TRPV4</i>	605427
<i>MIR200B</i>	612091	<i>NOG</i>	602991	<i>POLR1D</i>	613715	<i>ROR2</i>	602337	<i>SOS2</i>	601247	<i>TSC1</i>	605284
<i>MLH1</i>	120436	<i>NOL8</i>	611534	<i>POMGNT1</i>	606822	<i>RPS27L</i>	612055	<i>SOST</i>	605740	<i>TSC2</i>	191092
<i>MMAA</i>	607481	<i>NOTCH3</i>	600276	<i>POMT1</i>	607423	<i>RITN</i>	610436	<i>SOX9</i>	608160	<i>TSHB</i>	188540
<i>MMAB</i>	607568	<i>NPC1</i>	607623	<i>POMT2</i>	607439	<i>RUNX1</i>	151385	<i>SP7</i>	606633	<i>TSHR</i>	603372
<i>MMACHC</i>	609831	<i>NPC2</i>	601015	<i>POR</i>	124015	<i>RUNX2</i>	600211	<i>SPAST</i>	604277	<i>TSHZ1</i>	614427
<i>MMADHC</i>	611935	<i>NPR2</i>	108961	<i>POTED</i>	607549	<i>RYR1</i>	180901	<i>SPG11</i>	610844	<i>TSPEAR</i>	612920
<i>MMP13</i>	600108	<i>NRAS</i>	164790	<i>PPIB</i>	123841	<i>SALL1</i>	602218	<i>SPG7</i>	602783	<i>TWIST1</i>	601622
<i>MMP9</i>	120361	<i>NSD1</i>	606681	<i>PQBP1</i>	300463	<i>SATB2</i>	608148	<i>SPR</i>	182125	<i>UBE3A</i>	601623
<i>MOC51</i>	603707	<i>NT5C3A</i>	606224	<i>PRKAR1A</i>	188830	<i>SBDS</i>	607444	<i>SRY</i>	480000	<i>USH1C</i>	605242
<i>MOC52</i>	603708	<i>NTNG1</i>	608818	<i>PRPS1</i>	311850	<i>SBF2</i>	607697	<i>STK11</i>	602216	<i>USH1G</i>	607696
<i>MPZ</i>	159440	<i>OFD1</i>	300170	<i>PRSS1</i>	276000	<i>SCN4A</i>	603967	<i>SUMO1</i>	601912	<i>USH2A</i>	608400
<i>MRAP</i>	609196	<i>OGN</i>	602383	<i>PRX</i>	605725	<i>SCO2</i>	604272	<i>TARDBP</i>	605078	<i>USHBP1</i>	611810
<i>MRE11</i>	600814	<i>OTC</i>	300461	<i>PSAT1</i>	610936	<i>SELENON</i>	606210	<i>TBX1</i>	602054	<i>USP25</i>	.
<i>MSH2</i>	609309	<i>OTOF</i>	603681	<i>PSPH</i>	172480	<i>SERPINH1</i>	600943	<i>TBX22</i>	300307	<i>VAPB</i>	605704
<i>MSH3</i>	600887	<i>OXCT1</i>	601424	<i>PTCH1</i>	601309	<i>SETBP1</i>	611060	<i>TCAP</i>	604488	<i>VHL</i>	608537
<i>MSH6</i>	600678	<i>P3H1</i>	610339	<i>PTEN</i>	601728	<i>SETX</i>	608465	<i>TCF12</i>	600480	<i>WHRN</i>	607928
<i>MSTN</i>	601788	<i>PABPN1</i>	602279	<i>PTH1R</i>	168468	<i>SGCA</i>	600119	<i>TCF4</i>	602272	<i>WNT3</i>	165330
<i>MSX1</i>	142983	<i>PAFAH1B1</i>	601545	<i>PTPN11</i>	176876	<i>SGCB</i>	600900	<i>TCOF1</i>	606847	<i>XRCC3</i>	600675
<i>MSX2</i>	123101	<i>PAH</i>	612349	<i>PTS</i>	612719	<i>SGCD</i>	601411	<i>TERT</i>	187270	<i>ZAP70</i>	176947
<i>MTHFR</i>	607093	<i>PALB2</i>	610355	<i>PVT1</i>	165140	<i>SGCG</i>	608896	<i>TFAP2A</i>	107580	<i>ZIC2</i>	603073
<i>MTM1</i>	300415	<i>PAPSS2</i>	603005	<i>PYGM</i>	608455	<i>SH3BP2</i>	602104	<i>TG</i>	188450		

Anexo IV: Lista de 496 genes do painel NGS 'P4'

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
AARS	601065	BIN1	601248	COL1A2	120160	DNM2	602378	FKRP	606596	HBB	141900
ABCD1	300371	BMP1	112264	COL2A1	120140	DUOX2	606759	FKTN	607440	HES7	608059
ACADM	607008	BMP4	112262	COL3A1	120180	DYM	607461	FLNA	300017	HLCS	609018
ACADVL	609575	BMPRI1A	601299	COL5A1	120215	DYNC2H1	603297	FLNB	603381	HMGCL	613898
ACAT1	607809	BRAF	164757	COL5A2	120190	DYRK1A	600855	FLNC	102565	HNRNPDL	607137
ACP5	171640	BRCA1	113705	COL6A1	120220	DYSF	603009	FMR1	309550	HRAS	190020
ACTA1	102610	BRCA2	600185	COL6A2	120240	EDN1	131240	FOLR1	136430	HSPB1	602195
ACTB	102630	BRIP1	605882	COL6A3	120250	EFNB1	300035	FOXE1	602617	HSPB8	608014
ACTG1	102560	BTBD	609019	COL9A1	120210	EGR2	129010	FOXG1	164874	IDUA	252800
ADA	608958	CACNA1S	114208	COL9A2	120260	EIF2AK3	604032	FOXP2	605317	IFITM5	614757
ADAMTS18	607512	CANT1	613165	COL9A3	120270	EIF4A3	608546	FTCD	606806	IFT80	611177
ADAMTSL2	612277	CAPN3	114240	COMP	600310	ELN	130160	FUS	137070	IL11RA	600939
ADCK3	.	CAV3	601253	COQ2	609825	EMD	300384	FXN	606829	IL2RG	308380
AFG3L2	604581	CBL	165360	COQ9	612837	EPCAM	185535	GADD45G	604949	IL7R	146661
ALDH3A2	609523	CBS	613381	CP	117700	EPHA3	179611	GALNS	612222	IRF6	607199
ALDH7A1	107323	CD3D	186790	CPS1	608307	ERF	611888	GALNT1	602273	ITGA7	600536
ALPL	171760	CD3E	186830	CREB3L1	616215	ERG	165080	GALNT12	610290	IVD	607036
ALX1	601527	CDC45	603465	CREBBP	600140	ETHE1	608451	GALT	606999	JAG1	601920
ALX3	606014	CDH1	192090	CRISPLD2	612434	EVC	604831	GAMT	601240	JAK3	600173
ALX4	605420	CDH2	114020	CRTAP	605497	EVC2	607261	GARS	600287	JAM2	606870
AMN	605799	CDH23	605516	CRYAA	123580	EXT1	608177	GATM	602360	JMJD1C	604503
ANO5	608662	CDH7	605806	CRYAB	123590	FAF1	604460	GCDH	608801	KBTBD13	613727
APC	611731	CDH8	603008	CTDP1	604927	FAH	613871	GCH1	600225	KCNE1	176261
APTX	606350	CDK4	123829	CTNNB1	116806	FANCA	607139	GDAP1	606598	KIF22	603213
ARG1	608313	CDKL5	300203	CTSK	601105	FANCC	613899	GDF5	601146	KIF5A	602821
ARHGAP29	610496	CDKN2A	600160	CUBN	602997	FANCD2	613984	GDF6	601147	KRAS	190070
ARX	300382	CFL1	601442	CUL7	609577	FANCE	613976	GIF	609342	L1CAM	308840
ASL	608310	CFTR	602421	CYP21A2	613815	FANCF	613897	GJA1	121014	LAMA2	156225
ASPA	608034	CHD8	610528	CYP27A1	606530	FANCG	602956	GJB1	304040	LAMP2	309060
ASS1	603470	CHEK2	604373	CYP7B1	603711	FBLN5	604580	GJB2	121011	LARGE1	603590
ATL1	606439	CHST14	608429	DAG1	128239	FBN1	134797	GJB3	603324	LEMD3	607844
ATM	607585	CHST3	603799	DBT	248610	FECH	612386	GJB6	604418	LFNG	602576
ATP7A	300011	CLCN1	118425	DCLRE1C	605988	FGD4	611104	GLB1	611458	LIFR	151443
ATP7B	606882	CLDN14	605608	DDR2	191311	FGF8	600483	GLI2	165230	LIP1	609252
ATR	601215	CLPTM1L	612585	DES	125660	FGFR1	136350	GLI3	165240	LITAF	603795
AXIN2	604025	CLRN1	606397	DHCR24	606418	FGFR2	176943	GNAI3	139370	LMAN1	601567
B4GALT6	604017	COL10A1	120110	DHCR7	602858	FGFR3	134934	GNAS	139320	LMBRD1	612625
BAP1	603089	COL11A1	120280	DLAT	608770	FHL1	300163	GRHL3	608317	LMNA	150330
BARD1	601593	COL11A2	120290	DLL3	602768	FIG4	609390	GRIK1	138245	LRP5	603506
BCKDHA	608348	COL18A1	120328	DMD	300377	FKBP10	607063	HADHA	600890	LRSAM1	610933
BCKDHB	248611	COL1A1	120150	DNAJB6	611332	FKBP14	614505	HADHB	143450	LZTR1	600574

Anexo IV: Lista de 496 genes do painel NGS ‘P4’, continuação

Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM	Gene (HGNC)	Número MIM
MAP2K1	176872	MTRR	602568	PCCA	232000	RABEP2	611869	SIX3	603714		
MAP2K2	601263	MUT	609058	PCCB	232050	RAD50	604040	SLC19A3	606152	TH	191290
MATN3	602109	MUTYH	604933	PCDH15	605514	RAD51	179617	SLC22A5	603377	TMC1	606706
MBD2	603547	MYC	190080	PDGFRA	173490	RAD51C	602774	SLC25A13	603859	TMEM38B	611236
MBD5	611472	MYO6	600970	PDHA1	300502	RAD51D	602954	SLC25A15	603861	TMIE	607237
MC2R	607397	MYO7A	276903	PDHX	608769	RAF1	164760	SLC26A2	606718	TMPPSS15	606635
MCCC1	609010	MYOT	604103	PDSS1	607429	RAG1	179615	SLC26A4	605646	TMPPSS3	605511
MCCC2	609014	NAGS	608300	PDSS2	610564	RAG2	179616	SLC2A1	138140	TNFRSF11A	603499
MECP2	300005	NBN	602667	PFKL	171860	RAI1	607642	SLC5A5	601843	TP53	191170
MED25	610197	NDN	602117	PHF8	300560	RARA	180240	SMAD4	600993	TP63	603273
MEGF10	612453	NDRG1	605262	PHGDH	606879	RBBP8	604124	SMARCA1	606622	TPM2	190990
MEN1	613733	NDUFV2	600532	PLCB4	600810	RECQL4	603780	SNAP29	604202	TPM3	191030
MESP2	605195	NECTIN1	600644	PLEC	601282	REEP1	609139	SNRPN	182279	TPO	606765
MFN2	608507	NEFL	162280	PLOD1	153454	RET	164761	SOD1	147450	TRAPPC2	300202
MID1	300552	NF1	613113	PLOD2	601865	RIT1	609591	SOS1	182530	TRIM32	602290
MIR1206	.	NF2	607379	PLS3	300131	RMRP	157660	SOS2	601247	TRIP11	604505
MIR1207	.	NIPA1	608145	PMP22	601097	ROR2	602337	SOST	605740	TRPC6	603652
MIR1208	.	NIPBL	608667	PMS2	600259	RPS27L	612055	SOX9	608160	TRPS1	604386
MIR140	611894	NKX3-2	602183	POLR1C	610060	RTTN	610436	SP7	606633	TRPV4	605427
MIR200B	612091	NOG	602991	POLR1D	613715	RUNX1	151385	SPARC	182120	TSC1	605284
MLH1	120436	NOL8	611534	POMGNT1	606822	RUNX2	600211	SPAST	604277	TSC2	191092
MMAA	607481	NOTCH3	600276	POMT1	607423	RYR1	180901	SPG11	610844	TSHB	188540
MMAB	607568	NPC1	607623	POMT2	607439	SALL1	602218	SPG7	602783	TSHR	603372
MMACHC	609831	NPC2	601015	POR	124015	SATB2	608148	SPR	182125	TSHZ1	614427
MMADHC	611935	NPR2	108961	PPIB	123841	SBDS	607444	SPRED1	609291	TSPEAR	612920
MMP13	600108	NRAS	164790	PQBP1	300463	SBF2	607697	SRY	480000	TWIST1	601622
MMP9	120361	NSD1	606681	PRKAR1A	188830	SCN4A	603967	STK11	602216	UBE3A	601623
MOC51	603707	NT5C3A	606224	PRPS1	311850	SCO2	604272	SUMO1	601912	USH1C	605242
MOC52	603708	NTNG1	608818	PRSS1	276000	SELENON	606210	TARDBP	605078	USH1G	607696
MPZ	159440	OFD1	300170	PRX	605725	SERPINF1	172860	TBX1	602054	USH2A	608400
MRAP	609196	OGN	602383	PSAT1	610936	SERPINH1	600943	TBX22	300307	USHBP1	611810
MRE11A	.	OTC	300461	PSPH	172480	SETBP1	611060	TBX6	602427	USP25	.
MSH2	609309	OTOF	603681	PTCH1	601309	SETX	608465	TCAP	604488	VAPB	605704
MSH3	600887	OXCT1	601424	PTEN	601728	SGCA	600119	TCF12	600480	VHL	608537
MSH6	600678	P3H1	610339	PTH1R	168468	SGCB	600900	TCF4	602272	WHRN	607928
MSTN	601788	PABPN1	602279	PTPN11	176876	SGCD	601411	TCOF1	606847	WNT1	164820
MSX1	142983	PAFAH1B1	601545	PTS	612719	SGCG	608896	TERT	187270	WNT3	165330
MSX2	123101	PAH	612349	PVT1	165140	SH3BP2	602104	TFAP2A	107580	XRCC3	600675
MTHFR	607093	PALB2	610355	PYGM	608455	SHANK3	606230	TG	188450	ZAP70	176947
MTM1	300415	PAPSS2	603005	QDPR	612676	SHH	600725	TGFBR1	190181	ZIC1	600470
MTMR2	603557	PAX8	167415	RAB23	606144	SHOC2	602775	TGFBR2	190182	ZIC2	603073
MTR	156570	PCBD1	126090	RAB7A	602298	SHOX	312865	TGIF1	602630		

Anexo V: Lista de 56 genes de achados incidentais ACMG 2013

Câncer de mama e ovário hereditário	604370, 612555	<i>BRCA1</i>	113705
		<i>BRCA2</i>	600185
Síndrome de Li-Fraumeni	151623	<i>TP53</i>	191170
Síndrome de Peutz-Jeghers	175200	<i>STK11</i>	602216
Síndrome de Lynch	120435	<i>MLH1</i>	120436
		<i>MSH2</i>	609309
		<i>MSH6</i>	600678
		<i>PMS2</i>	600259
Polipose adenomatosa familiar	175100	<i>APC</i>	611731
Polipose associada a MYH; Adenomas múltiplos coloretal, PAF tipo 2; Polipose adenomatosa coloretal, autossômica recessiva, com pilomatricomas	608456, 132600	<i>MUTYH</i>	604933
Síndrome de Von Hippel Lindau	193300	<i>VHL</i>	608537
Neoplasia Endócrina Múltipla Tipo 1	131100	<i>MEN1</i>	613733
Neoplasia Endócrina Múltipla Tipo 2	171400, 162300	<i>RET</i>	164761
Câncer Medula de Tireóide Familiar (FMTC)	1552401	<i>RET</i>	164761
		<i>NTRK1</i>	191315
Síndrome de Tumor Hamartoma PTEN	153480	<i>PTEN</i>	601728
Retinoblastoma	180200	<i>RBI</i>	614041
Síndrome de Paraganglioma- Feocromocitoma Hereditário	168000 (PGL1)	<i>SDHD</i>	602690
	601650 (PGL2)	<i>SDHAF2</i>	613019
	605373 (PGL3)	<i>SDHC</i>	602413
	115310 (PGL4)	<i>SDHB</i>	185470
Complexo Esclerose Tuberosa	191100, 613254	<i>TSC1</i>	605284
		<i>TSC2</i>	191092
Tumor de Wilms associado a WT1	194070	<i>WT1</i>	607102
Neurofibromatose tipo 2	101100	<i>NF2</i>	607379
Síndrome de Ehlers-Danlos - tipo vascular	130050	<i>COL3A1</i>	120180
Síndrome de Marfan, Loeys-Dietz, e Aneurismas e dissecções de aorta familiar	154700, 609192, 608967, 610168, 610380, 613795, 611788	<i>FBN1</i>	134797
		<i>TGFBR1</i>	190181
		<i>TGFBR2</i>	190182
		<i>SMAD3</i>	603109
		<i>ACTA2</i>	102620
		<i>MYLK</i>	600922
Cardiomiopatia Hipertrofica e Dilatada	115197, 192600, 601494, 613690, 115196, 608751, 612098, 600858, 301500, 608758, 115200	<i>MYH11</i>	160745
		<i>MYBPC3</i>	600958
		<i>MYH7</i>	160760
		<i>TNNT2</i>	191045
		<i>TNNI3</i>	191044
		<i>TPM1</i>	191010
		<i>MYL3</i>	160790
		<i>ACTC1</i>	102540
		<i>PRKAG2</i>	602743
		<i>GLA</i>	300644
Taquicardia ventricular polimórfica catecolaminérgica	604772	<i>MYL2</i>	160781
		<i>LMNA</i>	150330
		<i>RYR2</i>	180902
		<i>PKP2</i>	602861
Cardiomiopatia ventricular direita arritmogênica	609040, 604400, 610476, 607450, 610193	<i>DSP</i>	125647
		<i>DSC2</i>	125645
		<i>TMEM43</i>	612048
		<i>DSG2</i>	125671
Síndrome de Romano-Ward de QT longo tipos 1, 2, and 3, síndrome de Brugada	192500, 613688, 603830, 601144	<i>KCNQ1</i>	607542
		<i>KCNH2</i>	152427
		<i>SCN5A</i>	600163
Hipercolesterolemia familiar	143890, 603776	<i>LDLR</i>	606945
		<i>APOB</i>	107730
		<i>PCSK9</i>	607786
Suscetibilidade a hipertermia maligna	145600	<i>RYR1</i>	180901
		<i>CACNA1S</i>	114208

Anexo VI: Lista de 59 genes de achados secundários ACMG 2016

Câncer de mama e ovário hereditário	604370, 612555	<i>BRCA1</i>	113705
		<i>BRCA2</i>	600185
Síndrome de Li-Fraumeni	151623	<i>TP53</i>	191170
Síndrome de Peutz-Jeghers	175200	<i>STK11</i>	602216
Síndrome de Lynch	120435	<i>MLH1</i>	120436
		<i>MSH2</i>	609309
		<i>MSH6</i>	600678
		<i>PMS2</i>	600259
Polipose adenomatosa familiar	175100	<i>APC</i>	611731
Polipose associada a MYH; Adenomas múltiplos coloretal, PAF tipo 2; Polipose adenomatosa coloretal, autossômica recessiva, com pilomatricomas	608456, 132600	<i>MUTYH</i>	604933
Polipose juvenil	174900	<i>BMPRIA</i>	601299
		<i>SMAD4</i>	600993
Síndrome de Von Hippel Lindau	193300	<i>VHL</i>	608537
Neoplasia Endócrina Múltipla Tipo 1	131100	<i>MEN1</i>	613733
Neoplasia Endócrina Múltipla Tipo 2	171400, 162300	<i>RET</i>	164761
Câncer Medula de Tireóide Familiar (FMTC)	1552401	<i>RET</i>	164761
		<i>NTRK1</i>	191315
Síndrome de Tumor Hamartoma PTEN	153480	<i>PTEN</i>	601728
Retinoblastoma	180200	<i>RB1</i>	614041
Síndrome de Paraganglioma- Feocromocitoma Hereditário	168000 (PGL1)	<i>SDHD</i>	602690
	601650 (PGL2)	<i>SDHAF2</i>	613019
	605373 (PGL3)	<i>SDHC</i>	602413
	115310 (PGL4)	<i>SDHB</i>	185470
Complexo Esclerose Tuberosa	191100, 613254	<i>TSC1</i>	605284
		<i>TSC2</i>	191092
Tumor de Wilms associado a WT1	194070	<i>WT1</i>	607102
Neurofibromatose tipo 2	101100	<i>NF2</i>	607379
Síndrome de Ehlers-Danlos - tipo vascular	130050	<i>COL3A1</i>	120180
Síndrome de Marfan, Loeys-Dietz, e Aneurismas e dissecções de aorta familiar	154700, 609192, 608967, 610168, 610380, 613795, 611788	<i>FBNI</i>	134797
		<i>TGFBR1</i>	190181
		<i>TGFBR2</i>	190182
		<i>SMAD3</i>	603109
		<i>ACTA2</i>	102620
		<i>MYH11</i>	160745
Cardiomiopatia Hipertrófica e Dilatada	115197, 192600, 601494, 613690, 115196, 608751, 612098, 600858, 301500, 608758, 115200	<i>MYBPC3</i>	600958
		<i>MYH7</i>	160760
		<i>TNNT2</i>	191045
		<i>TNNI3</i>	191044
		<i>TPM1</i>	191010
		<i>MYL3</i>	160790
		<i>ACTC1</i>	102540
		<i>PRKAG2</i>	602743
		<i>GLA</i>	300644
		<i>MYL2</i>	160781
Taquicardia ventricular polimórfica catecolaminérgica	604772	<i>LMNA</i>	150330
		<i>RYR2</i>	180902
Cardiomiopatia ventricular direita arritmogênica	609040, 604400, 610476, 607450, 610193	<i>PKP2</i>	602861
		<i>DSP</i>	125647
		<i>DSC2</i>	125645
		<i>TMEM43</i>	612048
		<i>DSG2</i>	125671
Síndrome de Romano-Ward de QT longo tipos 1, 2, and 3, síndrome de Brugada	192500, 613688, 603830, 601144	<i>KCNQ1</i>	607542
		<i>KCNH2</i>	152427
		<i>SCN5A</i>	600163
Hipercolesterolemia familiar	143890, 603776	<i>LDLR</i>	606945
		<i>APOB</i>	107730
		<i>PCSK9</i>	607786
Doença de Wilson	277900	<i>ATP7B</i>	606882
Deficiência de Ornitina Transcarbamilase	311250	<i>OTC</i>	300461
Suscetibilidade a hipertermia maligna	145600	<i>RYR1</i>	180901
		<i>CACNA1S</i>	114208

8. Referências

Acuna-Hidalgo R, Veltman JA, Hoischen A. New insights into the generation and role of de novo mutations in health and disease. *Genome Biol.* 2016;17(1):241.

Aigrain L., Gu Y., Quail MA. Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. *BMC Genomics.* 2016;17:458.

American College of Medical Genetics Newborn Screening Expert Group. Newborn screening: toward a uniform screening panel and system--executive summary. *Pediatrics.* 2006;117(5 Pt 2):S296-307.

Andermann A, Blancquaert I, Beauchamp S, Déry V. Revisiting Wilson and Jungner in the genomic age: a review of screening criteria over the past 40 years. *Bull World Health Organ.* 2008;86:317–9.

Aoki Y, Niihori T, Inoue S, Matsubara Y. Recent advances in RASopathies. *J Hum Genet.* 2016;61(1):33-9.

Au PYB, You J, Caluseriu O, Schwartzentruber J, Majewski J, Bernier FP, Ferguson M; Care for Rare Canada Consortium, Valle D, Parboosingh JS, Sobreira N, Innes AM, Kline AD. GeneMatcher aids in the identification of a new malformation syndrome with intellectual disability, unique facial dysmorphisms, and skeletal and connective tissue abnormalities caused by de novo variants in HNRNPK. *Hum Mutat.* 2015;36(10):1009-14.

Bae JS, Kim NK, Lee C, Kim SC, Lee HR, Song HR, Park KB, Kim HW, Lee SH, Kim HY, Lee SC, Jeong C, Park MS, Yoo WJ, Chung CY, Choi IH, Kim OH, Park WY, Cho TJ. Comprehensive genetic exploration of skeletal dysplasia using targeted exome sequencing. *Genet Med*. 201;18(6):563-9.

Barbosa-Buck CO, Orioli IM, da Graça Dutra M, Lopez-Camelo J, Castilla EE, Cavalcanti DP. Clinical epidemiology of skeletal dysplasias in South America. *Am J Med Genet A*. 2012;158A(5):1038-45.

Batch JA, Couper JJ, Rodda C, Cowell CT, Zacharin M. Use of bisphosphonate therapy for osteoporosis in childhood and adolescence. *J Paediatr Child Health*. 2003;39(2):88-92.

Beck TF, Mullikin JC; NISC Comparative Sequencing Program, Biesecker LG. Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants. *Clin Chem*. 2016;62(4):647-54.

Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat Rev Genet*. 2013;14(10):681-91.

Burrage LC, Charng WL, Eldomery MK, Willer JR, Davis EE, Lugtenberg D, Zhu W, Leduc MS, Akdemir ZC, Azamian M, Zapata G, Hernandez PP, Schoots J, de Munnik SA, Roepman R, Pearing JN, Jhangiani S, Katsanis N, Vissers LE, Brunner HG, Beaudet AL, Rosenfeld JA, Muzny DM, Gibbs RA, Eng CM, Xia F, Lalani SR, Lupski JR, Bongers EM, Yang Y. De Novo GMNN Mutations Cause Autosomal-Dominant Primordial Dwarfism Associated with Meier-Gorlin Syndrome. *Am J Hum Genet*. 2015;97(6):904-13.

Camacho-Vanegas O, Camacho SC, Till J, Miranda-Lorenzo I, Terzo E, Ramirez MC, Schramm V, Cordovano G, Watts G, Mehta S, Kimonis V, Hoch B, Philibert KD, Raabe

CA, Bishop DF, Glucksman MJ, Martignetti JA. Primate genome gain and loss: a bone dysplasia, muscular dystrophy, and bone cancer syndrome resulting from mutated retroviral-derived MTAP transcripts. *Am J Hum Genet.* 201;90(4):614-27.

Carter EM, Raggio CL. Genetic and orthopedic aspects of collagen disorders. *Curr Opin Pediatr.* 2009;21(1):46-54.

Caruccio N. Preparation of next-generation sequencing libraries using Nextera™ technology: simultaneous DNA fragmentation and adaptor tagging by in vitro transposition. *Methods Mol Biol.* 2011;733:241-55.

Casey JP, Brennan K, Scheidel N, McGettigan P, Lavin PT, Carter S, Ennis S, Dorkins H, Ghali N, Blacque OE, Mc Gee MM, Murphy H, Lynch SA. Recessive NEK9 mutation causes a lethal skeletal dysplasia with evidence of cell cycle and ciliary defects. *Hum Mol Genet.* 2016;25(9):1824-35.

Celestino-Soper PB, Gao H, Lynnes TC, Lin H, Liu Y, Spoonamore KG, Chen PS, Vatta M. Validation and Utilization of a Clinical Next-Generation Sequencing Panel for Selected Cardiovascular Disorders. *Front Cardiovasc Med.* 2017;4:11.

Chen C, Jiang Y, Xu C, Liu X, Hu L, Xiang Y et al. Skeleton Genetics: a comprehensive database for genes and mutations related to genetic skeletal disorders. *Database.* 2016;2016:baw127.

Chevrier V, Bruel AL, Van Dam TJ, Franco B, Lo Scalzo M, Lembo F, Audebert S, Baudelet E, Isnardon D, Bole A, Borg JP, Kuentz P, Thevenon J, Burglen L, Faivre L, Rivière JB, Huynen MA, Birnbaum D, Rosnet O, Thauvin-Robinet C. OFIP/KIAA0753 forms a complex with OFD1 and FOR20 at pericentriolar satellites and centrosomes and is mutated in one individual with oral-facial-digital syndrome. *Hum Mol Genet.* 2016;25(3):497-513.

Chin EL, da Silva C, Hegde M. Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations. *BMC Genet.* 2013;14:6.

Chong JX, Buckingham KJ, Jhangiani SN, Boehm C, Sobreira N, Smith JD, Harrell TM, McMillin MJ, Wiszniewski W, Gambin T, Coban Akdemir ZH, Doheny K, Scott AF, Avramopoulos D, Chakravarti A, Hoover-Fong J, Mathews D, Witmer PD, Ling H, Hetrick K, Watkins L, Patterson KE, Reinier F, Blue E, Muzny D, Kircher M, Bilguvar K, López-Giráldez F, Sutton VR, Tabor HK, Leal SM, Gunel M, Mane S, Gibbs RA, Boerwinkle E, Hamosh A, Shendure J, Lupski JR, Lifton RP, Valle D, Nickerson DA; Centers for Mendelian Genomics, Bamshad MJ. The Genetic Basis of Mendelian Phenotypes: Discoveries, Challenges, and Opportunities. *Am J Hum Genet.* 2015;97(2):199-215.

Conselho Regional de Medicina do Estado de São Paulo. Parecer Número: 15036. Data Emissão: 05-11-2013. Ementa: A realização de teste genético em casais assintomáticos com o intuito de identificar possíveis mutações recessivas, com a justificativa de que estes casais poderiam ser beneficiados em conhecer sua condição de portador, só pode ocorrer após avaliação clínica por médico que constate a real necessidade da execução do referido exame, sendo este profissional o responsável pela interpretação dos resultados obtidos e o aconselhamento do casal envolvido.

Crotwell PL, Hoyme HE. Advances in whole-genome genetic testing: from chromosomes to microarrays. *Curr Probl Pediatr Adolesc Health Care.* 2012;42(3):47-73.

de Carvalho TM, dos Santos HP, dos Santos IC, Vargas PR, Pedrosa J. Newborn screening: a national public health programme in Brazil. *J Inherit Metab Dis.* 2007;30:615.

de Chadarevian S. Whose Turn? Chromosome Research and the Study of the Human Genome. *J Hist of Biol.* 2017 Jul 25.

DePristo M, Banks E, Poplin R, Garimella K, Maguire J, Hartl C, Philippakis A, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell T, Kernytsky A, Sivachenko A, Cibulskis K, Gabriel S, Altshuler D, Daly M. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Gen.* 2011;43:491-8

Egunsola AT, Bae Y, Jiang MM, Liu DS, Chen-Evenson Y, Bertin T, Chen S, Lu JT, Nevarez L, Magal N, Raas-Rothschild A, Swindell EC, Cohn DH, Gibbs RA, Campeau PM, Shohat M, Lee BH. Loss of DDRGK1 modulates SOX9 ubiquitination in spondyloepimetaphyseal dysplasia. *J Clin Invest.* 2017;127(4):1475-84.

Fenwick AL, Kliszczak M, Cooper F, Murray J, Sanchez-Pulido L, Twigg SR, Goriely A, McGowan SJ, Miller KA, Taylor IB, Logan C; WGS500 Consortium, Bozdogan S, Danda S, Dixon J, Elsayed SM, Elsobky E, Gardham A, Hoffer MJ, Koopmans M, McDonald-McGinn DM, Santen GW, Savarirayan R, de Silva D, Vanakker O, Wall SA, Wilson LC, Yuregir OO, Zackai EH, Ponting CP, Jackson AP, Wilkie AO, Niedzwiedz W, Bicknell LS. Mutations in CDC45, Encoding an Essential Component of the Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis. *Am J Hum Genet.* 2016;99(1):125-38.

Francescatto L, Katsanis N. Newborn screening and the era of medical genomics. *Semin Perinatol.* 2015;39(8):617-22.

Frederiksen AL, Larsen MJ, Brusgaard K, Novack DV, Knudsen PJ, Schrøder HD, Qiu W, Eckhardt C, McAlister WH, Kassem M, Mumm S, Frost M, Whyte MP. Neonatal High Bone Mass With First Mutation of the NF- κ B Complex: Heterozygous De Novo Missense (p.Asp512Ser) RELA (Rela/p65). *J Bone Miner Res.* 2016;31(1):163-72.

Ghoumid J, Petit F, Boute-Benejean O, Frenois F, Cartigny M, Vanlerberghe C, Smol T, Caumes R, de Roux N, Manouvrier-Hanu S. Cerebellar hypoplasia with endosteal sclerosis is a POLR3-related disorder. *Eur J Hum Genet.* 2017;25(8):1011-4.

Girisha KM, Kortüm F, Shah H, Alawi M, Dalal A, Bhavani GS, Kutsche K. A novel multiple joint dislocation syndrome associated with a homozygous nonsense variant in the EXOC6B gene. *Eur J Hum Genet.* 2016;24(8):1206-10.

Goldfeder RL, Priest JR, Zook JM, Grove ME, Waggott D, Wheeler MT, Salit M, Ashley EA. Medical implications of technical accuracy in genome sequencing. *Genome Med.* 2016 Mar 2;8(1):24.

Gonzaga-Jauregui C, Gamble CN, Yuan B, Penney S, Jhangiani S, Muzny DM, Gibbs RA, Lupski JR, Hecht JT. Mutations in COL27A1 cause Steel syndrome and suggest a founder mutation effect in the Puerto Rican population. *Eur J Hum Genet.* 2015;23(3):342-6.

Goodwin S, McPherson JD, McCombie WR. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet.* 2016;17(6):333-51.

Green NS, Dolan SM, Murray TH. Newborn screening: complexities in universal genetic testing. *Am J Public Health.* 2006;96(11):1955-9.

Green RC, Berg JS, Grody WW, Kalia SS, Korf BR, Martin CL, McGuire AL, Nussbaum RL, O'Daniel JM, Ormond KE, Rehm HL, Watson MS, Williams MS, Biesecker LG; American College of Medical Genetics and Genomics. ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing. *Genet Med.* 2013;15(7):565-74.

Gripp KW, Aldinger KA, Bennett JT, Baker L, Tusi J, Powell-Hamilton N, Stabley D, Sol-Church K, Timms AE, Dobyns WB. A novel rasopathy caused by recurrent de novo missense mutations in PPP1CB closely resembles Noonan syndrome with loose anagen hair. *Am J Med Genet A*. 2016;170(9):2237-47.

Hayes JL, Tzika A, Thygesen H, Berri S, Wood HM, Hewitt S, Pendlebury M, Coates A, Willoughby L, Watson CM, Rabbitts P, Roberts P, Taylor GR. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. *Genomics*. 2013;102(3):174-81.

Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics*. 2016;107(1):1-8.

Huber C, Faqeih EA, Bartholdi D, Bole-Feysot C, Borochowitz Z, Cavalcanti DP, Frigo A, Nitschke P, Roume J, Santos HG, Shalev SA, Superti-Furga A, Delezoide AL, Le Merrer M, Munnich A, Cormier-Daire V Exome sequencing identifies INPPL1 mutations as a cause of opsismodysplasia. *Am J Hum Genet*. 2013;92(1):144-9.

Huckert M, Stoetzel C, Morkmued S, Laugel-Haushalter V, Geoffroy V, Muller J, Clauss F, Prasad MK, Obry F, Raymond JL, Switala M, Alembik Y, Soskin S, Mathieu E, Hemmerlé J, Weickert JL, Dabovic BB, Rifkin DB, Dheedene A, Boudin E, Caluseriu O, Cholette MC, Mcleod R, Antequera R, Gellé MP, Coeuriot JL, Jacquelin LF, Bailleul-Forestier I, Manière MC, Van Hul W, Bertola D, Dollé P, Verloes A, Mortier G, Dollfus H, Bloch-Zupan A. Mutations in the latent TGF-beta binding protein 3 (LTBP3) gene cause brachyolmia with amelogenesis imperfecta. *Hum Mol Genet*. 2015 1;24(11):3038-49.

Iida A, Xing W, Docx MK, Nakashima T, Wang Z, Kimizuka M, Van Hul W, Rating D, Spranger J, Ohashi H, Miyake N, Matsumoto N, Mohan S, Nishimura G, Mortier G, Ikegawa S. Identification of biallelic LRRK1 mutations in osteosclerotic metaphyseal dysplasia and evidence for locus heterogeneity. *J Med Genet*. 2016;53(8):568-74.

Jamie M. Ellingford, Stephanie Barton, Sanjeev Bhaskar, Simon G. Williams, Panagiotis I. Sergouniotis, James O'Sullivan, Janine A. Lamb, Rahat Perveen, Georgina Hall, William G. Newman, Paul N. Bishop, Stephen A. Roberts, Rick Leach, Rick Tearle, Stuart Bayliss, Simon C. Ramsden, Andrea H. Nemeth, Graeme C.M. Black. Whole Genome Sequencing Increases Molecular Diagnostic Yield Compared with Current Diagnostic Testing for Inherited Retinal Disease. *Ophthalmology*. 2016; 123(5): 1143-50.

Jamuar SS, Tan EC. Clinical application of next-generation sequencing for Mendelian diseases. *Hum Genomics*. 2015;9:10.

Jay AM, Conway RL, Thiffault I, Saunders C, Farrow E, Adams J, Toriello HV. Neonatal progeroid syndrome associated with biallelic truncating variants in POLR3A. *Am J Med Genet A*. 2016;170(12):3343-6.

Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, Herman GE, Hufnagel SB, Klein TE, Korf BR, McKelvey KD, Ormond KE, Richards CS, Vlangos CN, Watson M, Martin CL, Miller DT. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(2):249-55.

Klein CJ, Foroud TM. Neurology Individualized Medicine: When to Use Next-Generation Sequencing Panels. *Mayo Clin Proc*. 2017;92(2):292-305.

Kiper POS, Saito H, Gori F, Unger S, Hesse E, Yamana K, Kiviranta R, Solban N, Liu J, Brommage R, Boduroglu K, Bonafé L, Campos-Xavier B, Dikoglu E, Eastell R, Gossiel F, Harshman K, Nishimura G, Girisha KM, Stevenson BJ, Takita H, Rivolta C, Superti-Furga A, Baron R. Cortical-Bone Fragility--Insights from sFRP4 Deficiency in Pyle's Disease. *N Engl J Med*. 2016;374(26):2553-62.

Koboldt, DC, Steinberg, KM, Larson, DE, Wilson, RK, Mardis, E. The Next-Generation Sequencing Revolution and Its Impact on Genomics. *Cell*. 2013;155(1):27-38.

Koenig M, Hoffman EP, Bertelson CJ, Monaco AP, Feener C, Kunkel LM. Complete cloning of the Duchenne muscular dystrophy (DMD) cDNA and preliminary genomic organization of the DMD gene in normal and affected individuals. *Cell*. 1987;50(3):509-17.

Le Goff C, Rogers C, Le Goff W, Pinto G, Bonnet D, Chrabieh M, Alibeu O, Nistchke P, Munnich A, Picard C, Cormier-Daire V. Heterozygous Mutations in MAP3K7, Encoding TGF- β -Activated Kinase 1, Cause Cardiospondylocarpofacial Syndrome. *Am J Hum Genet*. 2016;99(2):407-13.

Lee H, Deignan JL, Dorrani N, Strom SP, Kantarci S, Quintero-Rivera F, Das K, Toy T, Harry B, Yourshaw M, Fox M, Fogel BL, Martinez-Agosto JA, Wong DA, Chang VY, Shieh PB, Palmer CG, Dipple KM, Grody WW, Vilain E, Nelson SF. Clinical exome sequencing for genetic identification of rare Mendelian disorders. *JAMA*. 2014;312(18):1880-7.

Lefterova MI, Shen P, Odegaard JI, Fung E, Chiang T, Peng G, Davis RW, Wang W, Kharrazi M, Schrijver I, Scharfe C. Next-Generation Molecular Testing of Newborn Dried Blood Spots for Cystic Fibrosis. *J Mol Diagn*. 2016;18(2):267-82.

Li H. and Durbin R. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*. 2009;25:1754-60.

Lippi G, Chiozza L, Mattiuzzi C, Plebani M. Patient and Sample Identification. Out of the Maze? *J Med Biochem*. 2017;36(2):107-12.

Lorget F, Kaci N, Peng J, Benoist-Lassel C, Mugniery E, Oppeneer T, Wendt DJ, Bell SM, Bullens S, Bunting S, Tsuruda LS, O'Neill CA, Di Rocco F, Munnich A, Legeai-

Mallet L. Evaluation of the therapeutic potential of a CNP analog in a Fgfr3 mouse model recapitulating achondroplasia. *Am J Hum Genet.* 2012;91(6):1108-14.

Malicdan MC, Vilboux T, Stephen J, Maglic D, Mian L, Konzman D, Guo J, Yildirimli D, Bryant J, Fischer R, Zein WM, Snow J, Vemulapalli M, Mullikin JC, Toro C, Solomon BD, Niederhuber JE; NISC Comparative Sequencing Program, Gahl WA, Gunay-Aygun M. Mutations in human homologue of chicken talpid3 gene (KIAA0586) cause a hybrid ciliopathy with overlapping features of Jeune and Joubert syndromes. *J Med Genet.* 2015;52(12):830-9.

Marconi C, Brunamonti Binello P, Badiali G, Caci E, Cusano R, Garibaldi J, Pippucci T, Merlini A, Marchetti C, Rhoden KJ, Galiotta LJ, Lalatta F, Balbi P, Seri M. A novel missense mutation in ANO5/TMEM16E is causative for gnathodiaphyseal dysplasia in a large Italian pedigree. *Eur J Hum Genet.* 2013;21(6):613-9.

Marques F, Tenney J, Duran I, Martin J, Nevarez L, Pogue R, Krakow D, Cohn DH, Li B. Altered mRNA Splicing, Chondrocyte Gene Expression and Abnormal Skeletal Development due to SF3B4 Mutations in Rodriguez Acrofacial Dysostosis. *PLoS Genet.* 2016;12(9):e1006307.

Matthijs G, Souche E, Alders M, Corveleyn A, Eck S, Feenstra I, Race V, Sistermans E, Sturm M, Weiss M, Yntema H, Bakker E, Scheffer H, Bauer P. Guidelines for diagnostic next-generation sequencing. *Eur J Hum Genet.* 2016;24(10):1515.

Mazzanti L, Tamburrino F, Bergamaschi R, Scarano E, Montanari F, Torella M, Ballarini E, Cicognani A. Developmental syndromes: growth hormone deficiency and treatment. *Endocr Dev.* 2009;14:114-34.

McInerney-Leo AM, Le Goff C, Leo PJ, Kenna TJ, Keith P, Harris JE, Steer R, Bole-Feysot C, Nitschke P, Kielty C, Brown MA, Zankl A, Duncan EL, Cormier-Daire V.

Mutations in LTBP3 cause acromicric dysplasia and geleophysic dysplasia. *J Med Genet.* 2016;53(7):457-64.

McInerney-Leo AM, Schmidts M, Cortés CR, Leo PJ, Gener B, Courtney AD, Gardiner B, Harris JA, Lu Y, Marshall M; UK10K Consortium, Scambler PJ, Beales PL, Brown MA, Zankl A, Mitchison HM, Duncan EL, Wicking C. Short-rib polydactyly and Jeune syndromes are caused by mutations in WDR60. *Am J Hum Genet.* 2013;93(3):515-23.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297-303

Mendoza-Londono R, Fahiminiya S, Majewski J; Care4Rare Canada Consortium, Tétreault M, Nadaf J, Kannu P, Sochett E, Howard A, Stimec J, Dupuis L, Roschger P, Klaushofer K, Palomo T, Ouellet J, Al-Jallad H, Mort JS, Moffatt P, Boudko S, Bächinger HP, Rauch F. Recessive osteogenesis imperfecta caused by missense mutations in SPARC. *Am J Hum Genet.* 2015;96(6):979-85.

Mierzewska H, Rydzanicz M, Biegański T, Kosinska J, Mierzewska-Schmidt M, Ługowska A, Pollak A, Stawiński P, Walczak A, Kędra A, Obersztyn E, Szczepanik E, Płoski R. Spondyloepimetaphyseal dysplasia with neurodegeneration associated with AIFM1 mutation - a novel phenotype of the mitochondrial disease. *Clin Genet.* 2017;91(1):30-7.

Min BJ, Kim N, Chung T, Kim OH, Nishimura G, Chung CY, Song HR, Kim HW, Lee HR, Kim J, Kang TH, Seo ME, Yang SD, Kim DH, Lee SB, Kim JI, Seo JS, Choi JY, Kang D, Kim D, Park WY, Cho TJ. Whole-exome sequencing identifies mutations of KIF22 in spondyloepimetaphyseal dysplasia with joint laxity, leptodactylic type. *Am J Hum Genet.* 2011;89(6):760-6.

Mortier GR. The diagnosis of skeletal dysplasias: A multidisciplinary approach. *Eur J Radiol.* 2001;40(3):161-7.

Murali C, Lu JT, Jain M, Liu DS, Lachman R, Gibbs RA, Lee BH, Cohn D, Campeau PM. Diagnosis of ALG12-CDG by exome sequencing in a case of severe skeletal dysplasia. *Mol Genet Metab Rep.* 2014;1:213-19.

Myrick LK, Deng PY, Hashimoto H, Oh YM, Cho Y, Poidevin MJ, Suhl JA, Visootsak J, Cavalli V, Jin P, Cheng X, Warren ST, Klyachko VA. Independent role for presynaptic FMRP revealed by an FMR1 missense mutation associated with intellectual disability and seizures. *Proc Natl Acad Sci U S A.* 2015;112(4):949-56.

Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, Pho N, Bozoklian D, Sandberg TOM, Brito LA, Lazar M, Bernardo DV, Amaro E Jr, Duarte YAO, Lebrão ML, Passos-Bueno MR, Zatz M. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum Mutat.* 2017;38(7):751-63.

Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet.* 2010;42(1):30-5.

Oud MM, Tuijnenburg P, Hempel M, van Vlies N, Ren Z, Ferdinandusse S, Jansen MH, Santer R, Johannsen J, Bacchelli C, Alders M, Li R, Davies R, Dupuis L, Cale CM, Wanders RJA, Pals ST, Ocaka L, James C, Müller I, Lehmberg K, Strom T, Engels H, Williams HJ, Beales P, Roepman R, Dias P, Brunner HG, Cobben JM, Hall C, Hartley T, Le Quesne Stabej P, Mendoza-Londono R, Davies EG, de Sousa SB, Lessel D, Arts HH, Kuijpers TW. Mutations in EXTL3 Cause Neuro-immuno-skeletal Dysplasia Syndrome. *Am J Hum Genet.* 2017;100(2):281-96.

Park KJ, Park S, Lee E, Park JH, Park JH, Park HD, Lee SY, Kim JW. A Population-Based Genomic Study of Inherited Metabolic Diseases Detected Through Newborn Screening. *Ann Lab Med*. 2016;36(6):561-72.

Patel N, Shamseldin HE, Sakati N, Khan AO, Softa A, Al-Fadhli FM, Hashem M, Abdulwahab FM, Alshidi T, Alomar R, Alobeid E, Wakil SM, Colak D, Alkuraya FS. GZF1 Mutations Expand the Genetic Heterogeneity of Larsen Syndrome. *Am J Hum Genet*. 2017;100(5):831-36.

Payne F, Lim K, Girousse A, Brown RJ, Kory N, Robbins A, Xue Y, Sleight A, Cochran E, Adams C, Dev Borman A, Russel-Jones D, Gorden P, Semple RK, Saudek V, O'Rahilly S, Walther TC, Barroso I, Savage DB. Mutations disrupting the Kennedy phosphatidylcholine pathway in humans with congenital lipodystrophy and fatty liver disease. *Proc Natl Acad Sci U S A*. 2014;111(24):8901-6.

Pehlivan D, Karaca E, Aydin H, Beck CR, Gambin T, Muzny DM, Bilge Geckinli B, Karaman A, Jhangiani SN; Centers for Mendelian Genomics, Gibbs RA, Lupski JR. Whole-exome sequencing links TMCO1 defect syndrome with cerebro-facio-thoracic dysplasia. *Eur J Hum Genet*. 2014 Sep;22(9):1145-8.

Pitt JJ. Newborn screening. *Clin Biochem Rev*. 2010;31(2):57-68.

Rasmussen SA, Bieber FR, Benacerraf BR, Lachman RS, Rimoin DL, Holmes LB. Epidemiology of osteochondrodysplasias: changing trends due to advances in prenatal diagnosis. *Am J Med Genet*. 1996;61(1):49-58.

Rattenberry E, Vialard L, Yeung A, Bair H, McKay K, Jafri M, Canham N, Cole TR, Denes J, Hodgson SV, Irving R, Izatt L, Korbonits M, Kumar AV, Lalloo F, Morrison PJ, Woodward ER, Macdonald F, Wallis Y, Maher ER. A comprehensive next generation sequencing based genetic testing strategy to improve diagnosis of inherited

pheochromocytoma and paraganglioma. *J Clin Endocrinol Metab.* 2013; 98(7):E1248-56.

Rehm HL, Bale SJ, Bayrak-Toydemir P, Berg JS, Brown KK, Deignan JL, Friez MJ, Funke BH, Hegde MR, Lyon E; Working Group of the American College of Medical Genetics and Genomics Laboratory Quality Assurance Committee. ACMG clinical laboratory standards for next-generation sequencing. *Genet Med.* 2013;15(9):733-47.

Reinstein E. Challenges of using next generation sequencing in newborn screening. *Genet Res (Camb).* 2015;97:e21.

Retterer K, Juusola J, Cho MT, Vitazka P, Millan F, Gibellini F, Vertino-Bell A, Smaoui N, Neidich J, Monaghan KG, McKnight D, Bai R, Suchy S, Friedman B, Tahiliani J, Pineda-Alvarez D, Richard G, Brandt T, Haverfield E, Chung WK, Bale S. Clinical application of whole-exome sequencing across clinical indications. *Genet Med.* 2016;18(7):696-704.

Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med.* 2015;17(5):405-24.

Rutsch F, MacDougall M, Lu C, Buers I, Mamaeva O, Nitschke Y, Rice GI, Erlandsen H, Kehl HG, Thiele H, Nürnberg P, Höhne W, Crow YJ, Feigenbaum A, Hennekam RC. A specific IFIH1 gain-of-function mutation causes Singleton-Merten syndrome. *Am J Hum Genet.* 2015;96(2):275-82.

Ruzzo K, Sale T, Willis M, Harding A, Lutgendorf M. Genetic Testing Costs and Compliance With Clinical Best Practices [19OP]. *Obstet Gynecol.* 2017;129:7S.

Seaby EG, Pengelly RJ, Ennis S. Exome sequencing explained: a practical guide to its clinical application. *Brief Funct Genomics*. 2016;15(5):374-84.

Schenkel LC, Kerkhof J, Stuart A, Reilly J, Eng B, Woodside C, Levstik A, Howlett CJ, Rupar AC, Knoll JH, Ainsworth P, Waye JS, Sadikovic B. Clinical Next-Generation Sequencing Pipeline Outperforms a Combined Approach Using Sanger Sequencing and Multiplex Ligation-Dependent Probe Amplification in Targeted Gene Panel Analysis. *J Mol Diagn*. 2016;18(5):657-67.

Shamseldin HE, Maddirevula S, Faqeih E, Ibrahim N, Hashem M, Shaheen R, Alkuraya FS. Increasing the sensitivity of clinical exome sequencing through improved filtration strategy. *Genet Med*. 2017;19(5):593-8.

Shashi V, McConkie-Rosell A, Rosell B, Schoch K, Vellore K, McDonald M, Jiang YH, Xie P, Need A, Goldstein DB. The utility of the traditional medical genetics diagnostic evaluation in the context of next-generation sequencing for undiagnosed genetic disorders. *Genet Med*. 2014;16(2):176-82.

Shen W, Szankasi P, Sederberg M, Schumacher J, Frizzell KA, Gee EP, Patel JL, South ST, Xu X, Kelley TW. Concurrent detection of targeted copy number variants and mutations using a myeloid malignancy next generation sequencing panel allows comprehensive genetic analysis using a single testing strategy. *Br J Haematol*. 2016;173(1):49-58.

Shylo NA, Christopher KJ, Iglesias A, Daluiski A, Weatherbee SD. TMEM107 Is a Critical Regulator of Ciliary Protein Composition and Is Mutated in Orofaciodigital Syndrome. *Hum Mutat*. 2016;37(2):155-9.

Sikkema-Raddatz B, Johansson LF, de Boer EN, Almomani R, Boven LG, van den Berg MP, van Spaendonck-Zwarts KY, van Tintelen P, Sijmons RH, Jongbloed JD, Sinke RJ.

Targeted Next-Generation Sequencing can Replace Sanger Sequencing in Clinical Diagnostics. *Hum Mutat.* 2013;34(7):1035-42.

Souza CF, Schwartz IV, Giugliani R. Triagem neonatal de distúrbios metabólicos. *Cien Saude Coletiva.* 2002;7:129–37.

Stray-Pedersen A, Backe PH, Sorte HS, Mørkrid L, Chokshi NY, Erichsen HC, Gambin T, Elgstøen KB, Bjørås M, Wlodarski MW, Krüger M, Jhangiani SN, Muzny DM, Patel A, Raymond KM, Sasa GS, Krance RA, Martinez CA, Abraham SM, Speckmann C, Ehl S, Hall P, Forbes LR, Merckoll E, Westvik J, Nishimura G, Rustad CF, Abrahamsen TG, Rønnestad A, Osnes LT, Egeland T, Rødningen OK, Beck CR; Baylor-Johns Hopkins Center for Mendelian Genomics, Boerwinkle EA, Gibbs RA, Lupski JR, Orange JS, Lausch E, Hanson IC. PGM3 mutations cause a congenital disorder of glycosylation with severe immunodeficiency and skeletal dysplasia. *Am J Hum Genet.* 2014;95(1):96-107.

Superti-Furga A, Unger S, Beighton P, Bonafé L, Braverman N, Briggs M, Cohn D, Cormier-Dairé V, Francomano C, Hall C, Horton W, Kaitila I, Krakow D, Lachman R, Lee B, LeMerrer M, Mortier G, Mundlos S, Nishimura G, Poznanski A, Rimoin D, Robertson S, Savarirayan R, Spranger J, Sillence D, Warman M, Wilcox W, Wilkie A, Zabel B, Zankl A. Nosology and classification of genetic skeletal disorders: 2006 revision. *Am J Med Genet, Part A.* 2007;143(1):1-18.

Tartaglia M, Zampino G, Gelb BD. Noonan syndrome: clinical aspects and molecular pathogenesis. *Mol Syndromol.* 2010;1(1):2-26.

Tham E, Eklund EA, Hammarsjö A, Bengtson P, Geiberger S, Lagerstedt-Robinson K, Malmgren H, Nilsson D, Grigelionis G, Conner P, Lindgren P, Lindstrand A, Wedell A, Albåge M, Zielinska K, Nordgren A, Papadogiannakis N, Nishimura G, Grigelioniene G. A novel phenotype in N-glycosylation disorders: Gillessen-Kaesbach-Nishimura skeletal dysplasia due to pathogenic variants in ALG9. *Eur J Hum Genet.* 2016;24(2):198-207.

Thauvin-Robinet C, Lee JS, Lopez E, Herranz-Pérez V, Shida T, Franco B, Jegu L, Ye F, Pasquier L, Loget P, Gigot N, Aral B, Lopes CA, St-Onge J, Bruel AL, Thevenon J, González-Granero S, Alby C, Munnich A, Vekemans M, Huet F, Fry AM, Saunier S, Rivière JB, Attié-Bitach T, Garcia-Verdugo JM, Faivre L, Mégarbané A, Nachury MV. The oral-facial-digital syndrome gene C2CD3 encodes a positive regulator of centriole elongation. *Nat Genet.* 2014;46(8):905-11.

Thevenon J, Duplomb L, Phadke S, Eguether T, Saunier A, Avila M, Carmignac V, Bruel AL, St-Onge J, Duffourd Y, Pazour GJ, Franco B, Attie-Bitach T, Masurel-Paulet A, Rivière JB, Cormier-Daire V, Philippe C, Faivre L, Thauvin-Robinet C. Autosomal recessive IFT57 hypomorphic mutation cause ciliary transport defect in unclassified oral-facial-digital syndrome with short stature and brachymesophalangia. *Clin Genet.* 2016;90(6):509-17.

Tidyman WE, Rauen KA. Expansion of the RASopathies. *Curr Genet Med Rep.* 2016 Sep;4(3):57-64.

Toriyama M, Lee C, Taylor SP, Duran I, Cohn DH, Bruel AL, Tabler JM, Drew K, Kelly MR, Kim S, Park TJ, Braun DA, Pierquin G, Biver A, Wagner K, Malfroot A, Panigrahi I, Franco B, Al-Lami HA, Yeung Y, Choi YJ; University of Washington Center for Mendelian Genomics, Duffourd Y, Faivre L, Rivière JB, Chen J, Liu KJ, Marcotte EM, Hildebrandt F, Thauvin-Robinet C, Krakow D, Jackson PK, Wallingford JB. The ciliopathy-associated CPLANE proteins direct basal body recruitment of intraflagellar transport machinery. *Nat Genet.* 2016;48(6):648-56.

Tsui L, Dorfman R. The Cystic Fibrosis Gene: A Molecular Genetic Perspective. *Cold Spring Harb Perspect Med.* 2013;3(2):a009472.

Vodopiutz J, Mizumoto S, Lausch E, Rossi A, Unger S, Janocha N, Costantini R, Seidl R, Greber-Platzer S, Yamada S, Müller T, Jilma B, Ganger R, Superti-Furga A, Ikegawa

S, Sugahara K, Janecke AR. Chondroitin Sulfate N-acetylgalactosaminyltransferase-1 (CSGαINAcT-1) Deficiency Results in a Mild Skeletal Dysplasia and Joint Laxity. *Hum Mutat.* 2017;38(1):34-8.

Vetro A, Savasta S, Russo Raucci A, Cerqua C, Sartori G, Limongelli I, Forlino A, Maruelli S, Perucca P, Vergani D, Mazzini G, Mattevi A, Stivala LA, Salviati L, Zuffardi O. MCM5: a new actor in the link between DNA replication and Meier-Gorlin syndrome. *Eur J Hum Genet.* 2017;25(5):646-650.

Wade EM, Daniel PB, Jenkins ZA, McInerney-Leo A, Leo P, Morgan T, Addor MC, Adès LC, Bertola D, Bohring A, Carter E, Cho TJ, Duba HC, Fletcher E, Kim CA, Krakow D, Morava E, Neuhaus T, Superti-Furga A, Veenstra-Knol I, Wieczorek D, Wilson LC, Hennekam RC, Sutherland-Smith AJ, Strom TM, Wilkie AO, Brown MA, Duncan EL, Markie DM, Robertson SP. Mutations in MAP3K7 that Alter the Activity of the TAK1 Signaling Complex Cause Frontometaphyseal Dysplasia. *Am J Hum Genet.* 2016;99(2):392-406.

Wang J, Gotway G, Pascual JM, Park JY. Diagnostic yield of clinical next-generation sequencing panels for epilepsy. *JAMA Neurol.* 2014;71(5):650-1.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.

Wang X, Charng WL, Chen CA, Rosenfeld JA, Al Shamsi A, Al-Gazali L, McGuire M, Mew NA, Arnold GL, Qu C, Ding Y, Muzny DM, Gibbs RA, Eng CM, Walkiewicz M, Xia F, Plon SE, Lupski JR, Schaaf CP, Yang Y. Germline mutations in ABL1 cause an autosomal dominant syndrome characterized by congenital heart defects and skeletal malformations. *Nat Genet.* 2017;49(4):613-7.

Wang Z, Iida A, Miyake N, Nishiguchi KM, Fujita K, Nakazawa T, Alswaid A, Albalwi MA, Kim OH, Cho TJ, Lim GY, Isidor B, David A, Rustad CF, Merckoll E, Westvik J,

Stattin EL, Grigelioniene G, Kou I, Nakajima M, Ohashi H, Smithson S, Matsumoto N, Nishimura G, Ikegawa S. Axial Spondylometaphyseal Dysplasia Is Caused by C21orf2 Mutations. *PLoS One*. 2016;11(3):e0150555.

Warman, M. L., Cormier-Daire, V., Hall, C., Krakow, D., Lachman, R., LeMerrer, M., Mortier, G., Mundlos, S., Nishimura, G., Rimoin, D. L., Robertson, S., Savarirayan, R., Sillence, D., Spranger, J., Unger, S., Zabel, B. and Superti-Furga, A. Nosology and classification of genetic skeletal disorders: 2010 revision. *Am J Med Genet Part A*, 2011;155:943-68.

Yamamoto GL, Agüena M, Gos M, Hung C, Pilch J, Fahiminiya S, Abramowicz A, Cristian I, Buscarilli M, Naslavsky MS, Malaquias AC, Zatz M, Bodamer O, Majewski J, Jorge AA, Pereira AC, Kim CA, Passos-Bueno MR, Bertola DR. Rare variants in SOS2 and LZTR1 are associated with Noonan syndrome. *J Med Genet*. 2015 Jun;52(6):413-21.

Yamamoto GL, Baratela WA, Almeida TF, Lazar M, Afonso CL, Oyamada MK, Suzuki L, Oliveira LA, Ramos ES, Kim CA, Passos-Bueno MR, Bertola DR. Mutations in PCYT1A cause spondylometaphyseal dysplasia with cone-rod dystrophy. *Am J Hum Genet*. 2014;94(1):113-9.

Yang Y, Muzny DM, Xia F, Niu Z, Person R, Ding Y, Ward P, Braxton A, Wang M, Buhay C, Veeraraghavan N, Hawes A, Chiang T, Leduc M, Beuten J, Zhang J, He W, Scull J, Willis A, Landsverk M, Craigen WJ, Bekheirnia MR, Stray-Pedersen A, Liu P, Wen S, Alcaraz W, Cui H, Walkiewicz M, Reid J, Bainbridge M, Patel A, Boerwinkle E, Beaudet AL, Lupski JR, Plon SE, Gibbs RA, Eng CM. Molecular findings among patients referred for clinical whole-exome sequencing. *JAMA*. 2014;312(18):1870-9.

Yubero D, Brandi N, Ormazabal A, Garcia-Cazorla À, Pérez-Dueñas B, Campistol J, Ribes A, Palau F, Artuch R, Armstrong J; Working Group. Targeted Next Generation

Sequencing in Patients with Inborn Errors of Metabolism. *PLoS One*. 2016;11(5):e0156359.

Zhang W, Taylor SP, Nevarez L, Lachman RS, Nickerson DA, Bamshad M; University of Washington Center for Mendelian Genomics Consortium, Krakow D, Cohn DH. IFT52 mutations destabilize anterograde complex assembly, disrupt ciliogenesis and result in short rib polydactyly syndrome. *Hum Mol Genet*. 2016;25(18):4012-20.

Zhang Z, Xia W, He J, Zhang Z, Ke Y, Yue H, Wang C, Zhang H, Gu J, Hu W, Fu W, Hu Y, Li M, Liu Y. Exome sequencing identifies SLCO2A1 mutations as a cause of primary hypertrophic osteoarthropathy. *Am J Hum Genet*. 2012;90(1):125-32.

Zou Y, Donkervoort S, Salo AM, Foley AR, Barnes AM, Hu Y, Makareeva E, Leach ME, Mohassel P, Dastgir J, Deardorff MA, Cohn RD, DiNonno WO, Malfait F, Lek M, Leikin S, Marini JC, Myllyharju J, Bönnemann CG. P4HA1 mutations cause a unique congenital disorder of connective tissue involving tendon, bone, muscle and the eye. *Hum Mol Genet*. 2017;26(12):2207-17.