MARIANA FERREIRA DE ASSIS FUNARI

**Investigação genético-molecular da baixa estatura isolada
através de sequenciamento paralelo em larga escala**

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

Programa de Endocrinologia

Orientador: Prof. Dr. Alexander Augusto de
Lima Jorge

São Paulo

2019

MARIANA FERREIRA DE ASSIS FUNARI

**Investigação genético-molecular da baixa estatura isolada
através de sequenciamento paralelo em larga escala**

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

Programa de Endocrinologia

Orientador: Prof. Dr. Alexander Augusto de
Lima Jorge

São Paulo

2019

Dados Internacionais de Catalogação na Publicação (CIP)

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

©reprodução autorizada pelo autor

Funari, Mariana Ferreira de Assis
Investigação genético-molecular da baixa estatura
isolada através de sequenciamento paralelo em larga
escala / Mariana Ferreira de Assis Funari. -- São
Paulo, 2019.

Tese(doutorado)--Faculdade de Medicina da
Universidade de São Paulo.
Programa de Endocrinologia.
Orientador: Alexander Augusto de Lima Jorge.

Descritores: 1.Insuficiência de
crescimento/genética 2.Lâmina de
crescimento/crescimento e desenvolvimento 3.SHOX
4.Biologia molecular/métodos 5.Sequenciamento de
nucleotídeos em larga escala 6.Mutação/genética

USP/FM/DBD-082/19

Responsável: Erinalva da Conceição Batista, CRB-8 6755

AGRADECIMENTOS

Quero agradecer imensamente a todos que contribuíram de alguma forma para a execução e finalização deste trabalho.

Agradeço ao meu orientador pela oportunidade de executar esse projeto e de participar de outros tantos trabalhos durante esse período. Agradeço ainda por todos os ensinamentos ao longo desses anos.

Agradeço à Dra. Berenice pelo apoio de sempre e pela oportunidade de ir aos congressos apresentar os nossos resultados.

Agradeço aos pacientes que participaram da casuística desse trabalho, seus respectivos familiares e a toda a equipe do ambulatório que atende e cuida muito bem deles.

Agradeço à Bruna por toda ajuda nos longos dias de preparo de biblioteca e a Thais por estar sempre disposta a me ajudar com os dados clínicos.

Agradeço ao Antônio e ao Lucas pelas análises complicadas da bioinformática e pela paciência com as minhas perguntas infundáveis.

Agradeço ao nosso grupo do NGS, a Anna Flavia, Amanda, Luciana, Mirian e a Monica, pelas discussões semanais que nos ajudaram a entender melhor os desafios dessa metodologia. A Amanda agradeço ainda pela ajuda no SELA. A Mirian agradeço por toda a ajuda e conselhos de sempre. E a Monica agradeço pela amizade e pelos puxões de orelha necessários.

Agradeço aos colegas do LIM42, em especial a todas as pessoas que tornaram o meu trabalho mais fácil, a Cidinha, a Cristina, a Fran, a Mara, a Nilda, a Rosangele e a Suelen. E agradeço também aos colegas do LIM25 pela ajuda essencial nos momentos em que precisei deles, em especial a Elis, Marco, Eliete, Geni e a Graça.

Agradeço especialmente a minha família, meus pais queridos, meu irmão e minha cunhada pelos incentivos constantes e injeções de ânimo nos momentos mais difíceis. Apesar de estarem distantes, estiveram sempre presentes. Agradeço muito por todo o apoio que sempre me deram. E ao meu irmão agradeço ainda pela paciência de me ajudar com o *photoshop*.

E, por fim, agradeço ao meu querido Raul por ser o meu maior incentivador. Agradeço a ele pela companhia, por me fazer feliz e por não me deixar desistir em nenhum momento.

“Em algum lugar, alguma coisa incrível
está esperando para ser descoberta”

Sharon Begley

SUMÁRIO

Lista de abreviaturas e símbolos

Lista de figuras

Lista de tabelas

Lista de anexos

Resumo

Abstract

1. Introdução	1
1.1. Vias de regulação do crescimento	2
1.2. O gene <i>SHOX</i> e a baixa estatura	4
1.3. Correlação entre o genótipo e a baixa estatura	6
1.4. Investigação molecular da baixa estatura	7
2. Objetivos	10
3. Casuística.....	12
3.1. Considerações éticas.....	13
3.2. Critérios de seleção de pacientes	13
3.2.1. Critérios de inclusão.....	14
3.2.2. Critérios de exclusão	14
3.3. Grupo de indivíduos controles	14
4. Métodos.....	15
4.1. Estratégia do estudo	16
4.2. Extração do DNA genômico.....	16
4.3. Sequenciamento paralelo em larga escala	17
4.3.1. Preparo das bibliotecas.....	17
4.3.1.1. <i>Kit</i> customizado para o preparo das bibliotecas do painel	18
4.3.1.2. <i>Kit</i> comercial para o preparo das bibliotecas do exoma	19
4.3.1.3. Protocolo de preparo das bibliotecas	19
4.3.2. Amplificação clonal.....	21

4.3.3. Sequenciamento	21
4.3.4. Bioinformática	22
4.3.4.1. Identificação de variantes de ponto ou pequenas indels	24
4.3.4.2. Identificação de CNVs	27
4.3.5. Determinação da patogenicidade das variantes identificadas	28
5. Resultados	29
5.1. Sequenciamento paralelo em larga escala do painel	30
5.1.1. Caracterização dos pacientes selecionados para o estudo do painel	30
5.1.2. Métricas do sequenciamento do painel	31
5.1.3. Avaliação do gene <i>SHOX</i> pelo painel	32
5.1.4. Variantes identificadas em outros genes pelo painel	44
5.1.5. Caracterização fenotípica dos pacientes com variantes patogênicas ou provavelmente patogênicas	45
5.1.6. Detalhes das variantes de ponto identificadas	48
5.1.6.1. Variantes que afetam fatores parácrinos da cartilagem de crescimento	48
5.1.6.2. Variantes que afetam vias intracelulares	50
5.1.6.3. Variantes que afetam o eixo GH/IGF-1	51
5.2. Sequenciamento paralelo em larga escala do exoma	51
5.2.1. Caracterização dos pacientes selecionados para o estudo do exoma	51
5.2.2. Métricas do sequenciamento do exoma	51
5.2.3. Variantes identificadas pelo exoma	53
6. Discussão	55
6.1. Análise dos resultados do painel	56
6.1.1. Análise de eficiência do painel na identificação dos defeitos do <i>SHOX</i>	56
6.1.2. Variantes identificadas em outros genes pelo painel	59
6.1.2.1. O gene <i>FGFR3</i> e as variantes identificadas	59
6.1.2.2. O gene <i>GNAS</i> e a variante identificada	60

6.1.2.3. O gene <i>NPR2</i> e a variante identificada.....	61
6.1.2.4. O gene <i>IHH</i> e as variantes identificadas.....	62
6.1.2.5. O gene <i>BRAF</i> e a variante identificada.....	64
6.1.2.6. O gene <i>NF1</i> e a variante identificada.....	65
6.1.2.7. O gene <i>IGF1R</i> e a variante identificada.....	66
6.2. Análise dos resultados do exoma	67
6.3. Considerações gerais	68
7. Conclusões.....	71
8. Referências	73
9. Anexos	83

LISTA DE ABREVIATURAS E SÍMBOLOS

<i>ABraOM</i>	banco de dados populacional brasileiro, Arquivo Brasileiro Online de Mutações
ACMG/AMP	Colégio Americano de Genética Médica e Associação de Patologia Molecular
<i>ACAN</i>	gene <i>aggrecan</i>
aCGH	<i>array</i> de hibridação genômica comparativa, de <i>comparative genomic hybridization array</i>
AHO	osteodistrofia hereditária de <i>Albright</i>
AS:A	relação altura sentada dividida pela altura total
BE	baixa estatura
BEI	baixa estatura isolada
BMPs	proteínas morfogenéticas ósseas, de <i>bone morphogenetic proteins</i>
BNP	peptídeo natriurético tipo B
<i>BRAF</i>	gene <i>B-Raf proto-oncogene, serine/threonine kinase</i>
BWA	programa de bioinformática <i>Burrows-Wheeler Aligner</i>
CADD	ferramenta de predição <i>in silico</i> <i>Combined Annotation Dependent Depletion</i>
CAPPesq	Comissão de Ética para Análise em Pesquisa
<i>ClinVar</i>	banco de dados que possui informações de significado clínico de variantes descritas
cm	centímetros
CNE	elementos não codificadores altamente conservados
CNP	peptídeo natriurético tipo C
CNVs	variantes de número de cópias
<i>COL10A1</i>	gene <i>collagen type X alpha 1 chain</i>
CONTRA	programa de bioinformática para análise de CNVs <i>COpy Number Targeted Resequencing Analysis</i>
<i>dbscSNV</i>	ferramenta de predição <i>in silico</i> para variantes em regiões consenso para <i>splicing</i>
<i>dbSNP</i>	banco de dados de variantes
DLW	discondrosteose de <i>Léri-Weill</i>
DNA	ácido desoxirribonucléico
DO	densidade óptica
ECR	regiões evolutivamente conservadas

ECS	sequências evolutivamente conservadas
ExAC	banco de dados populacional <i>Exome Aggregation Consortium</i>
FAPESP	Fundação de Amparo à Pesquisa do Estado de São Paulo
FATHMM	ferramenta de predição <i>in silico</i> <i>Functional Analysis Through Hidden Markov Models</i>
FGFR3	gene do receptor 3 do fator de crescimento de fibroblasto, de <i>fibroblast growth factor receptor 3</i>
FGFs	fatores de crescimento dos fibroblastos, de <i>fibroblast growth factors</i>
g	gramas
GERP	ferramenta de predição <i>in silico</i> <i>Genomic Evolutionary Rate Profiling</i>
GH	hormônio do crescimento
GH1	gene do hormônio do crescimento, de <i>growth hormone 1</i>
gnomAD	banco de dados populacional <i>Genome Aggregation Database</i>
GWAS	estudo de associação de genômica ampla, de <i>genome-wide association study</i>
HOXA9	gene <i>homeobox A9</i>
HOXD13	gene <i>homeobox D13</i>
IGF1	gene do fator 1 de crescimento insulino-símile, de <i>insulin like growth factor 1</i>
IGF-1	fator 1 de crescimento insulino-símile, de <i>insulin like growth factor 1</i>
IGFBP-3	proteína ligadora 3 do fator de crescimento insulino-símile, de <i>insulin-like growth factor binding protein 3</i>
IGV	programa de bioinformática <i>Integrative Genomics Viewer</i>
IHH	gene <i>indian hedgehog</i>
indels	pequenas deleções e inserções
JAK/STAT1	via de sinalização JAK/STAT1, de <i>janus protein tyrosine kinase/signal transducer and activator of transcription 1</i>
Kb	kilobases
kg	quilogramas
L	litro
LIM	laboratório de investigação médica
LoF	perda de função, de <i>loss of function</i>
lowDNA	marcador de peso molecular <i>low DNA mass ladder</i>
M	molar
MAF	frequência do alelo menos frequente, de <i>minor allele frequency</i>
MAPK	proteínas quinase ativadas por mitógenos, de <i>mitogen-activated protein kinases</i>

<i>md5sum</i>	programa de bioinformática <i>MD5 Message-Digest Algorithm</i>
mL	mililitros
MLPA	amplificação dependente da ligação de múltiplas sondas, de <i>multiplex ligation-dependent probe amplificação</i>
mRNA	ácido ribonucleico mensageiro
<i>NF1</i>	gene <i>neurofibromin 1</i>
ng	nanogramas
NGS	sequenciamento de nova geração, de <i>next-generation sequencing</i>
nm	nanômetro
nM	nanomolar
nº	número
<i>NPPB</i>	gene <i>natriuretic peptide B</i>
<i>NPPC</i>	gene <i>natriuretic peptide C</i>
<i>NPR2</i>	gene <i>natriuretic peptide receptor 2</i>
NPR-B	receptor dos peptídeos natriuréticos, principalmente do tipo C
ØX	marcador de peso molecular <i>ØX174 DNA-Hae III digest</i>
<i>p</i>	valor-p ou probabilidade de significância
PAR1	região pseudoautosômica 1
pb	pares de base
PCR	reação em cadeia da polimerase
pH	potencial hidrogeniônico
PHP	pseudohipoparatiroidismo
PHP1A	pseudohipoparatiroidismo tipo 1A
PHP1C	pseudohipoparatiroidismo tipo 1C
POH	heteroplasia óssea progressiva
<i>Polyphen2</i>	ferramenta de predição <i>in silico Polymorphism Phenotyping v2</i>
PPHP	pseudopseudohipoparatiroidismo
<i>PROVEAN</i>	ferramenta de predição <i>in silico PROtein Variation Effect ANalyzer</i>
PTH	paratormônio
PTHrP	peptídeo relacionado ao paratormônio
qPCR	reação em cadeia da polimerase quantitativa
rhGH	hormônio do crescimento recombinante
<i>RUNX2</i>	gene <i>runt related transcription factor 2</i>
SELA	Laboratório de Sequenciamento em Larga Escala – multiusuário da Faculdade de Medicina da USP

<i>SHOX</i>	gene <i>short stature homeobox containing gene</i>
SIC	segundo informações colhidas
<i>SIFT</i>	ferramenta de predição <i>in silico SIFT Human Coding SNPs</i>
SNVs	variantes de um único nucleotídeo
<i>SOX5</i>	gene <i>SRY-box 5</i>
<i>SOX6</i>	gene <i>SRY-box 6</i>
<i>SOX9</i>	gene <i>SRY-box 9</i>
SPLE	sequenciamento paralelo em larga escala
SS	baixa estatura, de <i>short stature</i>
TAE	tampão tris - acetato de sódio - HCl
TSH	hormônio tireoestimulante
U/kg/dia	unidades por quilograma por dia
UTR	região não traduzida, de <i>untranslated region</i>
USP	Universidade de São Paulo
VCF	arquivo <i>variant call format</i> , que contém a lista das variantes identificadas
WT	indivíduo com genótipo normal, de <i>wild type</i>
x	vezes
Xp	braço curto do cromossomo X
Yp	braço curto do cromossomo Y
Z	escore-Z
α S	subunidade alfa da proteína Gs
λ DNA	marcador de peso molecular lambda DNA
λ <i>HindIII</i>	marcador de peso molecular lambda DNA digerido com a enzima <i>HindIII</i> , de λ <i>DNA-HindIII digest</i>
μ g	micrograma
μ L	microlitro
$^{\circ}$ C	graus <i>Celsius</i>

LISTA DE FIGURAS

Figura 1 - Representação esquemática da cartilagem de crescimento de um osso longo com suas diferentes zonas e condrócitos em diferentes fases de maturação...	4
Figura 2 - Representação esquemática do gene <i>SHOX</i> e suas regiões regulatórias na PAR1.....	5
Figura 3 - Resumo da estratégia do estudo e seleção da casuística de pacientes com baixa estatura isolada.....	16
Figura 4 - Passo a passo do sequenciamento paralelo em larga escala.....	18
Figura 5 - Fluxograma com a estratégia de filtragem das variantes identificadas no SPLE do painel.....	26
Figura 6 - Imagem do programa <i>IGV</i> mostrando a cobertura real do sequenciamento do gene <i>SHOX</i> no painel.....	32
Figura 7 - Mapa de CNVs dos casos com alterações previamente conhecidas e dos pacientes sem diagnóstico que tiveram CNVs identificadas pelo painel..	33
Figura 8 - Visualização da variante c.440G>A: p.Arg147His do <i>SHOX</i> em heterozigose no sequenciamento do painel, pelo programa <i>IGV</i> , e no sequenciamento por <i>Sanger</i> ..	34
Figura 9 - Análises do <i>CONTRA</i> e do <i>IGV</i> da Paciente 12, com duplicação dos exons 1, 2 e 3.....	36
Figura 10 - Análises do <i>CONTRA</i> , <i>IGV</i> e <i>Sanger</i> da deleção intragênica envolvendo os exons 4, 5 e 6a da Paciente 11.....	37
Figura 11 - Análises do <i>CONTRA</i> , <i>IGV</i> e <i>Sanger</i> da deleção a 3' do <i>SHOX</i> da Paciente 13.	38
Figura 12 - Análises do <i>CONTRA</i> das Pacientes 5 e 15.....	40
Figura 13 - Análises do <i>CONTRA</i> e do <i>IGV</i> da Paciente 16, com deleção isolada do exon 1	41
Figura 14 - Análises do <i>CONTRA</i> , <i>IGV</i> e <i>Sanger</i> da deleção intragênica envolvendo parte do exon 3 do <i>SHOX</i> da Paciente 17	43
Figura 15 - Raio-X de mão e punho do Paciente 22 com variante no <i>IHH</i>	49
Figura 16 - Heredograma das nove famílias submetidas ao exoma.....	52

LISTA DE TABELAS

Tabela 1 - Características clínicas dos 77 pacientes com baixa estatura isolada submetidos ao painel	30
Tabela 2 - Métricas das corridas que continham os pacientes do painel realizadas nos equipamentos <i>NextSeq500 Illumina</i> , com kits <i>NextSeq MidiOutput V2 2x150</i> e <i>MiSeq Illumina</i> , com kits <i>MiSeq V3 2x300</i>	31
Tabela 3 - Detalhamento da filtragem das variantes identificadas pelo painel	44
Tabela 4 - Detalhes das principais variantes alélicas identificadas no sequenciamento do painel com a classificação de acordo com os critérios do ACMG/AMP	46
Tabela 5 - Características clínicas dos pacientes portadores das variantes alélicas classificadas como patogênicas ou provavelmente patogênicas identificadas no sequenciamento do painel	47
Tabela 6 - Características clínicas dos nove casos índice com baixa estatura isolada submetidos ao exoma	52
Tabela 7 - Métricas das corridas de exoma que continham os pacientes selecionados realizadas no equipamento <i>HiSeq2500 Illumina</i> , com kits <i>HiSeq SBS V4 2x125</i>	53
Tabela 8 - Detalhamento da filtragem das variantes identificadas pelo exoma	54

LISTA DE ANEXOS

ANEXO A - Lista dos genes presentes nas duas versões do painel para investigação da baixa estatura.....	84
ANEXO B - Coordenadas dos cromossomos X e Y correspondentes às regiões regulatórias do <i>SHOX</i> incluídas na versão 2 do painel para investigação da baixa estatura e suas respectivas coberturas (prevista e obtida).	85
ANEXO C - Endereço dos programas e ferramentas utilizados na análise de bioinformática.....	86
ANEXO D - Critérios de classificação de patogenicidade dos programas <i>in silico</i> utilizados nas filtragens da planilha VCF.....	87
ANEXO E - Critérios do ACMG/AMP para classificação das variantes alélicas quanto à patogenicidade	88
ANEXO F - Normas para combinação dos critérios do ACMG/AMP para classificação das variantes alélicas identificadas	90
ANEXO G - Critérios ACMG/AMP utilizados para a classificação das principais variantes alélicas de ponto com potencial patogênico identificadas no sequenciamento do painel	91
ANEXO H - Detalhes das principais variantes classificadas como variantes de significado incerto identificadas no sequenciamento do painel que segregaram com o fenótipo da baixa estatura na família ou estavam ausentes nos familiares com altura dentro da normalidade	92
ANEXO I - Detalhes das principais variantes classificadas como variantes de significado incerto identificadas no sequenciamento do painel que não foram segregadas nas famílias dos pacientes por ausência de amostra de DNA dos familiares.....	93
ANEXO J - Artigos em que participei como colaboradora durante o período do doutorado que utilizaram o sequenciamento paralelo em larga escala como metodologia.....	94

RESUMO

Funari MFA. *Investigação genético-molecular da baixa estatura isolada através de sequenciamento paralelo em larga escala* [tese]. São Paulo: Faculdade de Medicina, Universidade de São Paulo; 2019.

O crescimento é um processo complexo influenciado por múltiplos fatores, entre eles fatores genéticos e ambientais. Os defeitos do gene *SHOX* têm sido descritos como uma das principais causas monogênicas da baixa estatura (BE), sendo responsáveis por 1,1 a 22,2% dos casos de BE idiopática. Estudos recentes têm demonstrado que defeitos em outros genes (que codificam proteínas importantes para a fisiologia da cartilagem de crescimento) têm sido apontados como causas de BE. A estratégia atual para o estudo molecular da BE isolada envolve metodologias que investigam a presença de mutações ou deleções no próprio *SHOX* e suas regiões regulatórias, assim como mutações em outros genes, como o *NPR2* e o *FGFR3*. Em muitos casos, entretanto, nenhuma variante alélica que justifique o fenótipo dos pacientes é detectada e eles permanecem sem diagnóstico molecular. Baseado nesse contexto, o objetivo principal deste projeto foi tentar identificar uma possível etiologia molecular para BE desses pacientes, estudando vários genes simultaneamente. Para isso, foi desenvolvido um painel customizado de sequenciamento paralelo em larga escala, incluindo toda a região genômica do *SHOX*, suas regiões regulatórias e outros genes relacionados ao crescimento. Este painel foi testado para o estudo dos defeitos do *SHOX* a partir da análise de 15 pacientes com alterações previamente conhecidas nesse gene. Em seguida, 77 pacientes com BE isolada foram avaliados. O painel apresentou boa eficiência para o estudo do *SHOX*, sendo capaz de identificar todas as alterações conhecidas através da análise de mutações de ponto e de variantes de número de cópias, esta última realizada pela inspeção visual do sequenciamento e a utilização do programa *CONTRA*. Em relação aos casos sem diagnóstico, o painel identificou 15 variantes com potencial patogênico, sendo capaz de definir a etiologia da BE em 14 (14/77; 18%) pacientes. Entre os pacientes que permaneceram sem diagnóstico, foram selecionados nove casos para o estudo do exoma. Essa análise, porém, não identificou variantes candidatas. Sendo assim, concluímos que o painel desenvolvido apresentou bom desempenho para o diagnóstico molecular da baixa estatura isolada, inclusive para a identificação das alterações do *SHOX*, tão comuns nesse grupo de pacientes, podendo ser utilizado na análise molecular inicial desses indivíduos. Além do diagnóstico, abordagem de sequenciamento simultâneo de vários genes possibilitou a ampliação do espectro fenotípico associado a alguns genes já conhecidos, aumentando também o nosso conhecimento em relação às vias de regulação do crescimento.

Descritores: insuficiência de crescimento/genética; lâmina de crescimento/crescimento e desenvolvimento; *SHOX*; biologia molecular/métodos; sequenciamento de nucleotídeos em larga escala; mutação/genética.

ABSTRACT

Funari MFA. *Molecular investigation of isolated short stature using targeted massive parallel sequencing* [thesis]. São Paulo: "Faculdade de Medicina, Universidade de São Paulo"; 2019.

Growth is a complex process influenced by multiple factors, including genetic and environmental factors. *SHOX* defects have been described as one of the main monogenic causes of short stature (SS), accounting for 1.1 to 22.2% of cases of idiopathic short stature. Recent studies have shown that defects in other genes also associated with the growth plate development have been pointed as causes of SS. The current approach for the molecular study of isolated short stature involves methodologies that investigate mutations or deletions in *SHOX* itself and its regulatory regions, as well as mutations in some other genes like *NPR2* and *FGFR3*. In many cases, however, no allelic variant is identified and these patients remain undiagnosed. Based on this context, the aim of this project is to identify a molecular etiology for the short stature of these patients, investigating several genes simultaneously. A customized panel of targeted massive parallel sequencing was developed, including the entire genomic region of *SHOX*, as well as its regulatory regions and other genes related to the growth process. We evaluated 15 patients with previously known *SHOX* defects, including point mutations and copy number variants, to assess the panel efficiency in detecting mutations in this gene. Other 77 patients with isolated SS were submitted to the panel. Our panel demonstrated a good efficiency for the study of *SHOX* gene. It was able to identify all known defects with *CONTRA* analysis and visual inspection of sequencing coverage. Regarding the cases without diagnosis, the panel identified 15 variants with pathogenic potential, being able to define the etiology of the isolated SS in 14 (14/77; 18%) patients. As expected, most patients presented variants in genes associated with the growth plate physiology. Among the patients who remained undiagnosed, nine were selected to the exoma analysis. This study generated a broad list of genes for each family, but it did not identify any interesting candidate gene for any of them. Therefore, we conclude that the developed panel presented a good performance for the molecular diagnosis of the short stature of unknown cause, including the identification of *SHOX* defects, which are so common in this group of patients. It can be used in the initial molecular analysis of these individuals. In addition to the diagnosis, the simultaneous sequencing approach of several genes allowed the expansion of the phenotypic spectrum associated with some known genes, also increasing our knowledge about the pathways which regulate growth.

Descriptors: failure to thrive/genetics; growth plate/growth and development; *SHOX*; molecular biology/methods; high-throughput nucleotide sequencing; mutation/genetics.

1. INTRODUÇÃO

O crescimento é um processo complexo e dinâmico, que compreende a replicação e diferenciação de células em diferentes tecidos e envolve a interação de múltiplos fatores, inclusive de fatores ambientais. Ele apresenta um padrão previsível, acontece de forma semelhante entre os indivíduos e é um bom indicador do estado de saúde de uma criança. Além da desnutrição, doenças endócrinas, gastrointestinais, renais ou genéticas podem resultar em diminuição do crescimento e baixa estatura (BE).

Há muito tempo vem se estudando os fatores genéticos relacionados a altura e provavelmente milhares de variantes genéticas distribuídas ao longo do genoma contribuem para a determinação da estatura de um indivíduo (1). Estudos genéticos de algumas síndromes e displasias esqueléticas realizados em paralelo com estudos de associação de genômica ampla (GWAS, de *genome-wide association study*) têm levado ao reconhecimento de regiões ou genes envolvidos no crescimento, no desenvolvimento ósseo e na determinação da altura. Até momento, mais de 697 variantes independentes, distribuídas por 423 loci do genoma, já foram associadas à altura (1-5) e existem descritos no OMIM mais de 2.400 fenótipos, com um ou mais genes associados, que incluem a baixa estatura ou a diminuição do crescimento como um dos seus achados clínicos (<https://www.omim.org/>; em fevereiro de 2019). Apesar disso, defeitos moleculares que justificam o fenótipo dos pacientes com baixa estatura são identificados em uma pequena parcela dos casos, enquanto a maioria deles permanece idiopática.

1.1. Vias de regulação do crescimento

O eixo GH/IGF-1 (hormônio do crescimento/ fator 1 de crescimento insulino-símile) sempre foi considerado o principal sistema regulador do crescimento linear pós-natal (4). Entretanto defeitos nesse eixo são encontrados em uma minoria das crianças com baixa estatura (4). Mutações que resultam na deficiência de GH ou de IGF-1, ou ainda na insensibilidade aos mesmos, são defeitos muito raros. Esses dados, juntamente com achados recentes que reforçam a importância de processos inerentes à cartilagem de crescimento, deixam claro que esse sistema é apenas um dos muitos que controlam o complexo processo do crescimento (4, 5). Por esse motivo, recentemente foi sugerida uma nova classificação dos distúrbios de

crescimento, que apresenta a cartilagem de crescimento como o principal referencial (5).

A determinação da altura depende de um processo denominado ossificação endocondral, que acontece na cartilagem de crescimento dos ossos longos. Nesse processo extremamente coordenado, as células mesenquimais se condensam originando os condrócitos, que por sua vez, se multiplicam e se diferenciam, tornando-se hipertróficos (6). Simultaneamente é gerada uma matriz extracelular e um ambiente propício para vascularização e chegada dos osteoblastos que, por fim, vão resultar no alongamento do osso (6).

A cartilagem de crescimento é um elemento complexo. Nela, os condrócitos encontram-se divididos entre as zonas de reserva, proliferativa e hipertrófica (Figura 1). A regulação da diferenciação dos mesmos, assim como toda a fisiologia da cartilagem, envolve uma série de fatores hormonais, parácrinos, fatores de transcrição, proteínas intracelulares e moléculas de sinalização, e da matriz extracelular (6, 7). Defeitos nos genes que codificam qualquer um desses fatores ou seus respectivos receptores podem resultar em doenças com alterações esqueléticas, crescimento alterado e estatura abaixo ou acima da média (7).

Os recentes estudos de sequenciamento paralelo em larga escala (SPLE), principalmente aqueles que envolvem sequenciamento do exoma, têm identificado variantes em novos genes, o que tem ampliado essa rede de interações da cartilagem de crescimento (4, 5). Alguns paradigmas relacionados às displasias esqueléticas estão sendo quebrados. Variantes anteriormente associadas a fenótipos graves têm sido identificadas em pacientes com fenótipos mais leves e alguns padrões de herança estão se modificando. Aliado a isso, diversos estudos funcionais têm conseguido demonstrar o papel desses novos genes dentro da fisiologia do crescimento, fundamentando os novos achados (4, 5). Sistemas e vias de sinalização envolvendo todos esses fatores têm sido esclarecidas.

Baseada nesses constantes achados, a nova classificação sugerida por Wit e colaboradores (5) aponta como causas da baixa estatura os defeitos genéticos que afetam:

- os fatores parácrinos, que envolvem o sistema *IHH/PTHrP* (*indian hedgehog*/peptídeo relacionado ao paratormônio), o sistema *CNP/NPR2*, a

sinalização dos fatores de crescimento dos fibroblastos (*FGFs*), em especial via *FGFR3*, e a sinalização das proteínas morfogenéticas ósseas (*BMPs*);

- a matriz extracelular, que envolve as proteínas da matriz e genes como o *ACAN*, que codifica o principal proteoglicano da cartilagem, o agregan;
- as vias intercelulares, que envolvem fatores de transcrição (*SOX9*, *RUNX2* e *SHOX*) e vias de sinalização, como *RAS/MAPK*;
- e alguns processos celulares fundamentais.

Figura 1- Representação esquemática da cartilagem de crescimento de um osso longo com suas diferentes zonas e condrócitos em diferentes fases de maturação. Do lado direito, uma representação das principais proteínas envolvidas no processo de proliferação e diferenciação desses condrócitos. Fonte: Adaptado de Marchini e colaboradores, 2016 (7).

1.2. O gene *SHOX* e a baixa estatura

Um dos genes que atuam na cartilagem de crescimento é o *SHOX* (*short stature homeobox-containing gene*; MIM 312865) (8, 9). Considerado um dos principais reguladores do crescimento, ele tem seu papel dentro da ossificação endocondral ainda não totalmente esclarecido. Seus defeitos, porém, são

considerados a principal causa monogênica da baixa estatura, sendo responsáveis por cerca de 1,1 a 22,2% dos casos de baixa estatura idiopática (7).

O *SHOX* ocupa uma região de aproximadamente 40 Kb dentro da região pseudoautossômica 1 (PAR1) dos cromossomos sexuais e é altamente expresso nas células osteogênicas (8-10). Ele codifica um fator de transcrição que atua diretamente na cartilagem (11, 12), participando ativamente no processo de diferenciação dos condrócitos (10). Foi sugerido que, além de atuar como um repressor da diferenciação dos condrócitos retardando a fusão das cartilagens de crescimento, sua proteína atua na organização colunar das células dentro cartilagem (10). Defeitos no *SHOX* ou em suas regiões regulatórias (área do *SHOX*; regiões *enhancer*), localizadas a 3' e a 5' do gene (Figura 2) (13-16), resultariam na diferenciação prematura dos condrócitos, o que aceleraria a fusão da cartilagem e provocaria a parada prematura do crescimento (10).

Figura 2 - Representação esquemática do gene *SHOX* e suas regiões regulatórias na PAR1. As regiões regulatórias, localizadas a 3' e a 5' do gene, são regiões evolutivamente conservadas identificadas por diferentes trabalhos (13-16). A linha superior indica a distância do telômero do Xp (*GRCh37/hg19*). Os asteriscos representam as sondas de MLPA (de *multiplex ligation-dependent probe amplification*) presentes nos kits mais recentes. Tel: telômero; PAR1: região pseudoautossômica 1; CNE: elementos não codificadores altamente conservados; ECR: regiões evolutivamente conservadas; ECS: sequências evolutivamente conservadas. Fonte: Adaptado de Marchini e colaboradores, 2016 (7).

A proteína *SHOX* ainda modula a expressão de genes alvos também envolvidos no processo do crescimento (17). Muitos destes genes, assim como as vias de sinalização reguladoras do crescimento ósseo, não estão bem definidos. Em 2007, o gene *NPPB*, que codifica o peptídeo natriurético tipo B (BNP), foi descrito com o primeiro alvo da regulação transcricional do *SHOX* (18). A expressão do *SHOX* regula positivamente a expressão do *NPPB* em linhagens de osteosarcoma e condrosarcoma. Atualmente alguns outros genes, como o *SOX5*, *SOX6*, *FGFR3* e o *HOXA9*, já são conhecidos por interagirem com *SHOX* (7, 19-21).

São necessárias duas cópias ativas do *SHOX* para o crescimento normal (10). Tanto a diminuição, quanto o aumento do número de cópias ativas resulta em alterações significativas no crescimento (7, 8, 22). As deleções relacionadas ao *SHOX*, que correspondem a 80% dos defeitos do gene, podem ser detectadas por *multiplex ligation-dependent probe amplification* (MLPA), enquanto as mutações de ponto, que correspondem aos 20% restante, são identificadas por sequenciamento por *Sanger* (7). Os defeitos em homozigose no *SHOX* resultam em uma displasia esquelética grave, a displasia mesomélica de *Langer* (23), enquanto os defeitos em heterozigose no gene ou em regiões que regulam a sua expressão resultam em fenótipos que variam de displasias esqueléticas menos graves, como a discondrosteose de *Léri-Weill* (DLW) (24-31) até baixa estatura isolada (31-36).

A frequência dos defeitos em pacientes com DLW e BE idiopática varia bastante na literatura. Mutações no *SHOX* são encontradas em 70 a 90% das casuísticas de DLW (7, 24-31). Essa frequência no grupo de pacientes com BE idiopática varia ainda mais, de 1,1 a 22,2%, refletindo diferenças nos critérios de seleção dos pacientes e na metodologia utilizada na investigação genético-molecular (7, 31-36). Quando a desproporção corpórea é um dos critérios de seleção, essa frequência costuma ser geralmente mais alta. Um estudo realizado há alguns anos pelo nosso grupo identificou defeitos no *SHOX* em 19% dos pacientes com baixa estatura desproporcional (37).

1.3. Correlação entre o genótipo e a baixa estatura

Os estudos envolvendo o *SHOX* deixam claro que mutações afetando um único gene envolvido no desenvolvimento podem levar a múltiplos fenótipos. Uma determinada mutação, inclusive dentro da mesma família, pode resultar em fenótipos distintos (31, 38). Enquanto alguns indivíduos apresentam quadros clínicos discretos, com desproporção corpórea, outros podem apresentar alterações esqueléticas importantes, como a deformidade de *Madelung*, característica da DLW. Esta variabilidade, que dificulta a correlação genótipo-fenótipo (30-32), é um dos aspectos que mais chama atenção em relação aos defeitos desse gene. É possível que variantes em regiões regulatórias ou em outros genes que interagem com a

proteína do *SHOX* estejam atuando como moduladores do fenótipo nesses casos, sendo responsáveis por essa variabilidade.

Assim como acontece com os defeitos do *SHOX*, a variabilidade fenotípica também tem sido descrita nos defeitos que afetam outros genes como o *NPR2*, *ACAN* e *FGFR3* (4, 39-41). Nessa nova realidade desvendada pelos trabalhos mais recentes, defeitos em genes importantes para a cartilagem, podem resultar tanto em displasias esqueléticas, como em baixa estatura isolada, duas categorias antes consideradas bem distintas.

Muitas vezes, o fenótipo mais grave das displasias tende a ocorrer quando a mutação está presente em homozigose ou quando a mesma afeta de forma mais agressiva a função da proteína, prejudicando um domínio importante, por exemplo. Já o fenótipo mais leve tende a acontecer quando a mutação está presente em heterozigose, quando ela afeta a proteína de forma parcial ou quando o próprio gene possui um papel menos crítico para o crescimento (4).

Além disso, estudos de GWAS revelaram que, em muitos casos, a altura final dos indivíduos pode ser influenciada ainda por polimorfismos presentes nos mesmos genes que levam a displasias esqueléticas (*COL10A1*, *ACAN*, *HOXD13*, *CHSY1*, *IHH*) (1, 42). Esses polimorfismos teriam um pequeno efeito sobre a estatura final, mas quando somados resultariam em uma estatura na parte inferior da normalidade da curva ou até mesmo em baixa estatura.

1.4. Investigação molecular da baixa estatura

O diagnóstico molecular da BE permite estabelecer a causa do déficit de crescimento, diminuindo a ansiedade da família em relação à doença. O diagnóstico permite também estabelecer o prognóstico adequado para altura adulta, indicar o tratamento com hormônio do crescimento recombinante (rhGH) em casos específicos e oferecer à família um aconselhamento genético (43).

A estratégia atual para o estudo molecular da BE isolada envolve a investigação dos principais genes associados, como o próprio *SHOX*, o *NPR2* e o *FGFR3*. São utilizadas metodologias como o sequenciamento por *Sanger*, para identificação de mutações de ponto nas regiões codificadoras e flanqueadoras dos

exons, e no caso específico do *SHOX*, metodologias que também investigam a presença de deleções no gene e em suas regiões regulatórias, como o MLPA (44).

Apesar dessa estratégia, em muitos casos nenhuma variante alélica que justifique o fenótipo é detectada e os pacientes permanecem sem diagnóstico molecular, possivelmente pela presença de mutações em regiões não investigadas do próprio *SHOX* ou em outros genes envolvidos no desenvolvimento ósseo. A necessidade de utilização de múltiplas técnicas torna a investigação molecular demorada, trabalhosa e onerosa, dificultando, de certa forma, a sua aplicação na rotina laboratorial. Além disso, existem perguntas relacionadas à variabilidade fenotípica, ainda não explicada molecularmente. Por estas razões, é essencial que se estude mais profundamente toda a região do *SHOX*, os genes relacionados a ele e à própria cartilagem.

O SPLE, também conhecido como sequenciamento de nova geração (NGS, de *Next-Generation Sequencing*), é uma metodologia que começou a ser desenvolvida por volta de 2005 (45, 46) e permite o sequenciamento de várias regiões de interesse de muitos indivíduos simultaneamente. No caso do exoma, sequenciamos todas as regiões exônicas de todos os genes do genoma humano (aproximadamente 19.000 genes) (47), enquanto que no estudo dos painéis, sequenciamos apenas os genes de interesse. O sequenciamento tradicional por *Sanger*, que usamos na abordagem de gene candidato, acontece em uma escala muito menor (48). Em cada reação sequenciamos um fragmento específico de um único paciente, por exemplo, um exon de um determinado gene. Para sequenciar um gene inteiro de um paciente precisamos fazer, em média, uma ou duas reações de sequenciamento por exon desse gene. Sendo assim, para sequenciar todos os exons dos genes contidos nos painéis ou no próprio genoma por *Sanger* levaríamos muito tempo e gastaríamos muito dinheiro. O Projeto Genoma Humano, que teve início em 1990, antes do surgimento do NGS, exemplifica bem isso. O consórcio internacional, que contou com o apoio de inúmeros laboratórios espalhados pelo mundo, levou cerca de 13 anos para concluir o sequenciamento de todo o genoma humano com um investimento, de aproximadamente, 3 bilhões de dólares (49). Atualmente, utilizando a técnica de SPLE é possível realizar o sequenciamento do genoma por cerca de 1.100 dólares em alguns dias, com a análise em poucas semanas (50).

Dentro desse contexto, este projeto pretende dar continuidade à linha de investigação molecular da BE desenvolvendo uma metodologia, baseada no SPLE, para investigação de defeitos nos principais genes envolvidos na ossificação endocondral, com custo relativo baixo e rápida avaliação simultânea várias regiões.

Sendo assim, desenvolvemos um painel customizado, onde foi incluído toda a região genômica do *SHOX*, genes que o regulam ou são regulados por ele, como o *SOX5*, *SOX6*, *FGFR3*, *HOXA9* e *NPR2* (7, 18-21, 51, 52), além dos principais genes associados aos distúrbios do crescimento.

A avaliação inicial de eficiência da metodologia foi realizada a partir da utilização de amostras de pacientes com defeitos previamente identificados no *SHOX*. Análises de sensibilidade da nova tecnologia foram realizadas e comparadas com as técnicas tradicionalmente empregadas (sequenciamento por *Sanger* e MLPA) (44) visando uma possível aplicação da mesma na rotina de investigação laboratorial da baixa estatura.

Adicionalmente, alguns pacientes com baixa estatura familiar que continuaram sem uma possível causa genética, mesmo após o sequenciamento dessas regiões alvo, foram selecionados para o sequenciamento de todo o exoma, em uma tentativa de identificar novos genes e defeitos moleculares associados com o fenótipo de crescimento atenuado. Pretendíamos com esta estratégia, ajudar no diagnóstico da baixa estatura da família em questão, e contribuir para a ampliação do conhecimento das bases moleculares do crescimento, podendo também ajudar no desenvolvimento de novas aplicações diagnósticas para essa metodologia.

2. OBJETIVOS

2.1. Objetivo principal

- Identificar uma possível causa genética para o distúrbio de crescimento de pacientes com baixa estatura isolada sem achados esqueléticos específicos utilizando o sequenciamento paralelo em larga escala.

2.2. Objetivos específicos

- Desenvolver um painel customizado de sequenciamento paralelo em larga escala, envolvendo toda a região genômica do *SHOX*, suas regiões regulatórias e outros genes relacionados ao crescimento, para investigação molecular da baixa estatura isolada;
- Verificar a eficiência do painel desenvolvido para o estudo dos defeitos que envolvem o gene *SHOX*, inclusive as variantes de número de cópias;
- Realizar o sequenciamento exômico em famílias selecionadas, ainda sem diagnóstico, buscando a identificação de novos genes associados aos distúrbios de crescimento.

3. CASUÍSTICA

Este estudo foi realizado no Laboratório de Hormônios e Genética Molecular - LIM42, da Unidade de Endocrinologia do Desenvolvimento, no Laboratório de Endocrinologia Celular e Molecular - LIM25, da Unidade de Endocrinologia Genética, ambos da Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo e no Laboratório de Sequenciamento em Larga Escala (SELA), Multiusuário FAPESP, Rede PREMiUM, da Faculdade de Medicina da Universidade de São Paulo

3.1. Considerações éticas

O estudo foi conduzido de acordo com princípios éticos e orientações contidas na declaração de *Helsinki* e com termos descritos pela Portaria 196/96 do Conselho Nacional de Saúde.

O projeto foi submetido à avaliação pela Comissão de Ética em Pesquisa (CAPPesq). O consentimento por escrito foi obtido de todos os pacientes ou pais/tutores antes que os procedimentos de pesquisas fossem iniciados.

A confidencialidade dos resultados obtidos foi garantida a todos os pacientes. Os achados de defeitos genéticos relacionados à doença em investigação foram informados aos pacientes e/ou seus responsáveis legais que tiveram aconselhamento genético e estão sendo acompanhados conforme o recomendado pela Sociedade Brasileira de Genética Médica no projeto Diretrizes do Conselho Federal de Medicina.

3.2. Critérios de seleção de pacientes

Foram estudados 77 pacientes com diagnóstico de baixa estatura isolada, acompanhados no Ambulatório da Unidade de Endocrinologia do Desenvolvimento da Disciplina de Endocrinologia da Faculdade de Medicina da Universidade de São Paulo. Esses pacientes foram submetidos ao protocolo de investigação padrão de crianças com baixa estatura (53). Adicionalmente, os pacientes com desproporção corpórea, definida como escore-Z da relação altura sentada dividida pela altura total superior a 2, foram avaliados radiograficamente e/ou por geneticista para afastar o diagnóstico de displasias esqueléticas conhecidas (37).

3.2.1. Critérios de inclusão

- Baixa estatura caracterizada por escore-Z da altura inferior ou igual a menos 2 sem causa definida após investigação clínica, laboratorial e radiográfica (53);
- Ou baixa estatura em relação ao seu padrão familiar caracterizada por escore-Z da altura do paciente menos escore-Z da altura alvo inferior ou igual a menos 2;
- Ou previsão de baixa altura ao término do crescimento, caracterizada por escore-Z da altura adulta prevista pela idade óssea inferior ou igual a menos 2.

3.2.2. Critérios de exclusão

- Causas conhecidas de baixa estatura como presença de distúrbios genéticos, cromossômicos, desnutrição, doenças hepáticas, patologias sistêmicas ou endocrinológicas.

3.3. Grupo de indivíduos controles

- Para avaliação da eficiência da metodologia para o diagnóstico dos defeitos do *SHOX*, foram incluídos pacientes com variantes no número de cópias (CNVs) e mutações de ponto identificadas no gene em estudos prévios do grupo.
- Para a análise de frequência das variantes, foi utilizado como controle o banco de variantes brasileiro, o *Brazilian Genomic Variants (ABraOM)* (<http://abraom.ib.usp.br/>) (54), constituído de variantes exônicas de um grupo de 609 idosos aparentemente normais da cidade de São Paulo. Foi utilizado também um grupo de pacientes do SELA, constituído por variantes também exônicas de 387 indivíduos, entre pacientes com fenótipos diversos e seus familiares afetados ou não.

4. MÉTODOS

4.1. Estratégia do estudo

A estratégia do estudo e da seleção dos pacientes foi resumida na Figura 3.

Figura 3 - Resumo da estratégia do estudo e seleção da casuística de pacientes com baixa estatura isolada. BEI: baixa estatura isolada; SPLE: sequenciamento paralelo em larga escala.

4.2. Extração do DNA genômico

As amostras de DNA dos pacientes selecionados e seus familiares foram obtidas a partir de leucócitos de sangue periférico ou de células da mucosa oral extraídas por *swab*. O DNA foi extraído de acordo com procedimento padronizado no Laboratório de Hormônios e Genética Molecular - LIM42 da Unidade de Endocrinologia do Desenvolvimento, da Faculdade de Medicina da Universidade de São Paulo, baseado no método de *Miller* e colaboradores (55).

A concentração do DNA extraído foi obtida por leitura em espectrofotômetro (*Biophotometer*, *Eppendorf*, Hamburgo, Alemanha) no comprimento de onda de 260

nm (1 unidade DO 260 = 50 µg/mL). A relação ideal entre as leituras em 260 e 280 nm para a caracterização da pureza do material deveria ser superior a 1,75 (56).

As amostras de DNA foram submetidas à eletroforese em gel de agarose (*Thermo Fisher Scientific*, Waltham, MA, EUA) a 1% em TAE (Tris 0,004 M; Ácido Acético Glacial; EDTA 0,001 M pH 8,0) contendo *SYBR Safe DNA gel stain* (*Invitrogen*, *Thermo Fisher Scientific*, Waltham, MA, EUA) na concentração de 0,1 µL/mL de gel e observadas em um transiluminador com luz ultravioleta a fim de verificar sua integridade. Como padrões de massa foram utilizados 500 ng do marcador de peso molecular λ *DNA-HindIII digest* (250 ng/µL) e 20 ng do λ DNA (10 ng/µL) (ambos *Invitrogen*, *Thermo Fisher Scientific*, Waltham, MA, EUA). Em seguida, as amostras foram mantidas congeladas a menos 20 °C até sua utilização para o preparo das bibliotecas para o SPLE.

4.3. Sequenciamento paralelo em larga escala

A metodologia do sequenciamento paralelo em larga escala pode ser dividida em algumas etapas: (a) o preparo das bibliotecas, (b) a amplificação clonal, (c) o sequenciamento e (d) a bioinformática para análise dos dados. Um passo a passo simplificado de todo o processo está esquematizado na Figura 4.

4.3.1. Preparo das bibliotecas

As bibliotecas do painel e do exoma foram preparadas de acordo com o protocolo *SureSelect^{XT} Target Enrichment System for Illumina Paired-End Sequencing Library* (*Agilent Technologies*, Santa Clara, CA, EUA), a diferença foi o *kit* de sondas utilizado na captura dos dois casos. Para o preparo das bibliotecas do painel, utilizamos um *kit* customizado e para os exomas, utilizamos um *kit* comercial.

Figura 4 - Passo a passo do sequenciamento paralelo em larga escala. Os passos 1 e 2 esquematizam as etapas do preparo das bibliotecas. Em 1, está esquematizada a fragmentação mecânica, realizada no *Covaris*, e adição dos adaptadores. No passo 2, a hibridação com sondas específicas, seguida da captura das regiões de interesse (genes do painel ou todas regiões exônicas, no caso do exoma) com *beads* magnéticas. No passo 3, a amplificação clonal dos fragmentos das bibliotecas que acontece na *flowcell* do equipamento. E o passo 4 esquematiza a reação de sequenciamento propriamente dita, com incorporação dos nucleotídeos e escaneamento da *flowcell*. Fonte: Adaptado de Metzker e colaboradores, 2010 (57) e Adey e colaboradores, 2010 (58).

4.3.1.1. *Kit* customizado para o preparo das bibliotecas do painel

O painel customizado foi elaborado a partir da ferramenta *Agilent SureDesign 2.0* (*Agilent Technologies*, Santa Clara, CA, EUA), que desenha sondas específicas para seleção e captura de regiões genômicas de interesse, com base na metodologia *SureSelect^{XT}* (*Agilent Technologies*, Santa Clara, CA, EUA).

Foram desenhadas duas versões desse painel. Na primeira versão (Painel versão 1) foram incluídas as regiões exônicas de 74 genes relacionados ao crescimento, muitos deles relacionados à cartilagem de crescimento e alguns ligados diretamente à função do *SHOX*. A região alvo total desse Painel versão 1 foi de aproximadamente 481 Kb. Na segunda versão, o Painel versão 2, foram mantidos os genes da versão 1 e foram acrescentados alguns genes descritos mais recentemente em associação com os distúrbios do crescimento, totalizando 98 genes. A região alvo dessa versão 2 foi de aproximadamente 489 Kb. Nas duas

versões, o gene *SHOX* foi incluído em sua totalidade, ou seja, regiões intrônicas e exônicas. Além disso, na versão 2, foram incluídas algumas regiões intrônicas, localizadas a 5' e a 3' do gene, descritas como regiões com potencial regulatório do mesmo.

A lista completa dos genes presentes nas duas versões do painel está descrita no Anexo A, enquanto as coordenadas das regiões regulatórias estão descritas no Anexo B. Alguns genes incluídos nas duas versões, apesar de estarem, de alguma forma, relacionados ao crescimento, não fizeram parte especificamente desse projeto. Eles foram incluídos por uma questão de custo-benefício na síntese dos painéis.

4.3.1.2. Kit comercial para o preparo das bibliotecas do exoma

As bibliotecas para o sequenciamento exômico dos indivíduos selecionados (sem diagnóstico molecular após a utilização do painel customizado) foram preparadas com um *kit* comercial disponível para essa finalidade, o *SureSelect^{XT} Human All Exon V6* (Agilent Technologies, Santa Clara, CA, EUA).

4.3.1.3. Protocolo de preparo das bibliotecas

O protocolo de preparo de bibliotecas *SureSelect^{XT}* (Agilent Technologies, Santa Clara, CA, EUA), também envolve uma série de etapas:

Fragmentação do DNA. Duzentos ng de DNA genômico de cada paciente, quantificados pelo método fluorimétrico do *Qubit v2.0* (Invitrogen, Thermo Fisher Scientific, Waltham, MA, EUA), foram fragmentados de forma mecânica, por sonicação, no equipamento *Covaris* (Covaris Inc, Woburn, MA, EUA) de acordo com o protocolo de preparo de biblioteca, gerando fragmentos de tamanho entre 150 e 550 pb (Figura 4, passo 1).

As próximas quatro etapas foram realizadas em um termociclador *Veriti 96-well thermal cycler* (Thermo Fisher Scientific, Waltham, MA, EUA).

Reparo das extremidades. As extremidades dos fragmentos de DNA foram reparadas e posteriormente foi realizada uma adenilação das extremidades 3', o que possibilita a ligação de adaptadores (Figura 4, passo 1).

Ligação dos adaptadores e amplificação. Adaptadores foram adicionados a cada extremidade dos fragmentos, para a posterior adição dos indexes e amplificação clonal. Os fragmentos contendo os adaptadores foram amplificados em uma reação em cadeia da polimerase (PCR) (Figura 4, passo 1).

Hibridação e captura. Nessas etapas, as bibliotecas do painel foram enriquecidas em sequências provenientes dos genes previamente selecionados, enquanto que nos exomas, estas bibliotecas foram altamente enriquecidas em sequências provenientes das regiões codificadoras de todos os genes do genoma. O enriquecimento foi obtido através da captura das regiões de interesse com a utilização de *beads* magnéticas acopladas a estreptavidina, após uma hibridação *overnight* do DNA fragmentado com sondas biotiniladas complementares a essas regiões. Estas sondas são específicas para os genes alvos desenhados no painel ou para os genes do exoma (Figura 4, passo 2).

Indexação. A cada biblioteca foi adicionado um índice diferente (sequência específica de 8 pb que foi adicionada a todos os fragmentos do paciente, que permite a identificação dos dados desse paciente no *pool* de amostras sequenciadas). Os fragmentos indexados foram então amplificados em uma nova PCR.

Quantificação. O tamanho final das bibliotecas foi determinado no equipamento *TapeStation 2200*, com o *kit D1000 Screen Tape System* (ambos *Agilent Technologies*, Santa Clara, CA, EUA). Em seguida, as bibliotecas foram quantificadas por PCR quantitativo (qPCR) em tempo real, com o *kit Kapa for Illumina Library Quantification*, KK4824 (*Kapa Biosystems, Inc*, Wilmington, MA, EUA) no equipamento *AriaMx* (*Agilent Technologies*, Santa Clara, CA, EUA), de acordo com o protocolo do *kit* de qPCR.

Preparo dos *pools*. Antes de serem sequenciadas, as bibliotecas foram organizadas em *pools* com concentração equimolar de 10 nM. Os cálculos levaram em consideração as concentrações obtidas no qPCR e o tamanho final das bibliotecas. No caso dos painéis, foram feitas duas estratégias. Para a versão 1 foram feitos *pools* com 96 bibliotecas, enquanto que para a versão 2 foram feitos *pools* de 32 bibliotecas. Para os exomas foram feitos *pools* com 12 bibliotecas. Esses *pools* foram novamente quantificados por qPCR em tempo real, antes de serem preparados para o sequenciamento.

4.3.2. Amplificação clonal

Nesta etapa, também chamada de clusterização, as bibliotecas foram diluídas e depositadas em uma lâmina especial específica para o sequenciador, denominada *flowcell*. Na superfície da mesma encontram-se fixados de maneira paralela duas espécies de oligonucleotídeos, que são complementares aos adaptadores acoplados às extremidades dos fragmentos de DNA das bibliotecas. Após a ligação dos fragmentos da biblioteca a esses oligonucleotídeos foi realizada uma amplificação, onde cada molécula foi amplificada várias vezes, através de uma reação conhecida como PCR em ponte (*bridge PCR*). Ao final desta etapa, foram gerados clones (*clusters*) de moléculas de DNA idênticas à molécula original, ligados covalentemente à superfície da *flowcell* (Figura 4, passo 3).

4.3.3. Sequenciamento

Após a amplificação clonal, as moléculas antisense foram removidas enzimaticamente e o processo de sequenciamento foi iniciado com o acoplamento de um oligonucleotídeo iniciador especial. Em seguida, nucleotídeos modificados com terminadores reversíveis marcados com fluoróforos específicos foram adicionados ao meio. A cada ciclo de incorporação, foram geradas imagens de toda a superfície da *flowcell* através de um *scanner* de fluorescência em cada um dos comprimentos de onda específicos para os fluoróforos. Os *clusters* presentes na superfície da *flowcell* foram então identificados e mapeados (Figura 4, passo 4). A sobreposição das imagens produzidas a cada ciclo de incorporação propicia a

identificação da sequência dos nucleotídeos de cada *cluster* de moléculas, ou seja, a sequência do fragmento que deu origem ao *cluster*.

Na metade dos ciclos (ao final do sequenciamento do primeiro *read*) acontece uma inversão dos fragmentos ligados na *flowcell* e uma reclusterização para que aconteça o sequenciamento do mesmo fragmento no sentido reverso (do segundo *read*). Isso caracteriza a configuração *paired end* de SPLE, que possibilita leituras de fragmentos de até 300 pares de base, para os painéis, e de até 250 pb para os exomas, dependendo do *kit* de sequenciamento utilizado. Ao final dos ciclos, o equipamento gerou um arquivo com todas as sequências com a extensão *.fastq.

As duas últimas etapas de “Amplificação clonal” e “Sequenciamento” foram realizadas em equipamentos da plataforma *Illumina* (*Illumina, Inc, San Diego, CA, EUA*) do Laboratório de Sequenciamento em Larga Escala (SELA) da Rede PREMiUM, Multiusuário FAPESP, localizado na Faculdade de Medicina da Universidade de São Paulo. As bibliotecas de painel foram sequenciadas em seis corridas distintas, sendo três realizadas na plataforma *NextSeq 500 Illumina* com *kits NextSeq Kit MidiOutput V2 2x150* (em *pools* de 96 bibliotecas) (Painel versão 1) e três, na plataforma *MiSeq Illumina* com *kits MiSeq Kit V3 2x300* (em *pools* de 32 bibliotecas) (Painel versão 2) (todos *Illumina, Inc, San Diego, CA, EUA*). A amplificação clonal nesses casos foi realizada no próprio sequenciador. Os exomas foram sequenciados em quatro corridas distintas, realizadas na plataforma *HiSeq 2500 Illumina* com *kits HiSeq SBS V4 2x125* (ambos *Illumina, Inc, San Diego, CA, EUA*) (em *pools* de 12 bibliotecas distribuídas em duas *lanes*, totalizando 48 bibliotecas por corrida). A amplificação clonal dos mesmos foi realizada em um equipamento auxiliar denominado *CBot System* (*Illumina, Inc, San Diego, CA, EUA*), em função da maior quantidade de fragmentos. Os protocolos utilizados para essas duas etapas são específicos para os *kits* de sequenciamento e equipamentos utilizados.

4.3.4. Bioinformática

O grande volume de dados gerados pelos sequenciadores de nova geração requer o emprego de ferramentas de bioinformática sofisticadas, a fim de que poucas alterações genéticas de interesse sejam identificadas em meio a centenas

ou milhares de polimorfismos não relacionados ao fenótipo. Este processo foi realizado por diversos programas que funcionam em cadeia (*pipelines*). Os endereços dos *sites* relativos a essa análise estão disponíveis nos Anexos C e D.

As principais etapas da análise dos dados, tanto dos painéis como dos exomas, encontram-se descritas de forma resumida a seguir:

Demultiplexagem. A primeira etapa da análise consistiu em separar as sequências de cada paciente a partir da sequência dos índices, uma vez que as amostras foram sequenciadas de forma agrupada, em *pools*.

Checagem de consistência. Checagem da consistência dos arquivos *.fastq, cujo tamanho varia de dezenas a centenas de *gigabytes*, e cálculo de estatísticas básicas, como número de sequências produzidas. Nesta etapa, foram utilizados programas padrão de cálculo de integridade como *md5sum*.

Alinhamento com a sequência referência. Alinhamento das sequências das amostras a uma sequência referência do genoma humano (*GRCh37/hg19*). Nesta etapa, a origem de cada fragmento de DNA da amostra foi determinada. Esse alinhamento foi realizado com o programa *BWA*.

Recalibragem das sequências alinhadas. Como o alinhamento foi feito individualmente, sequência a sequência, tornou-se necessário recalibrar o alinhamento realizado na fase anterior, visto que, em algumas situações, diferentes sequências oriundas da mesma região cromossômica apresentam resultados ligeiramente diferentes de mapeamento (por exemplo, quando existem pequenas inserções, deleções ou repetições no genoma alvo). A recalibragem pode ser considerada um realinhamento, mas agora de forma contextualizada. Nesta etapa ainda foi utilizado o programa *BWA*, e como resultado gerou-se um arquivo com extensão *.bam.

4.3.4.1. Identificação de variantes de ponto ou pequenas indels

Chamada das variantes. A partir do arquivo *.bam, foi realizada a genotipagem das amostras, que consiste em determinar, com bases estatísticas, todos os alelos das mesmas (as bases de cada par do cromossomo para uma determinada posição). Nessa “chamada das variantes” (*variant call*) foi utilizado o programa *Freebayes*, que identifica variantes de um único nucleotídeo (SNVs) e pequenas deleções e inserções (indels) e libera o arquivo em formato *.vcf (*variant call format*).

Recalibragem da genotipagem. Na fase anterior, os genótipos foram calculados levando-se em conta apenas os dados disponíveis do alinhamento. Uma correção foi realizada a fim de que outros parâmetros fossem levados em conta, principalmente dados populacionais. Resumidamente, falsos positivos foram detectados e sinalizados para que fossem descartados. Novamente, módulos do programa *Freebayes* realizaram esta tarefa para SNVs e indels. As CNVs identificadas não são recalibradas e passam diretamente para a próxima etapa da análise.

Anotação dos genótipos variantes detectados. Esta etapa consiste em: (a) anotar as variantes encontradas, com programa *ANNOVAR* (59); (b) levantar as frequências das mesmas nos bancos públicos de variantes [como *Genome Aggregation Database (gnomAD)*, *1000 Genomes Project* e o *Exome Aggregation Consortium (ExAC)*] e (c) estimar o efeito de cada uma delas na proteína sintetizada, com a utilização de ferramentas de predição *in silico* como: *SIFT*, *Polyphen2*, *MutationTaster*, *Mutation Assessor*, *PROVEAN*, *FATHMM*, *CADD* e *GERP*, para mutações do tipo *missense*, e *dbSCSNV* para variantes que possivelmente alteram o *splicing*. Após essa etapa foi gerada a VCF anotada que contém todas as variantes identificadas em todos os pacientes com todas as informações inerentes a elas (localização, tipo, frequência, consequência, etc). Os Anexos C e D tem a lista dos programas de predição *in silico*, seus respectivos *sites* e critérios de patogenicidade.

Priorização das variantes candidatas. Nessa etapa, que envolve uma ampla interação com o pesquisador e depende do conhecimento do mesmo em relação ao fenótipo analisado, foi aplicada uma sequência de filtros na planilha VCF anotada

com a intenção de gerar uma lista limitada de variantes potencialmente relacionadas ao fenótipo de cada paciente e, se possível, identificar a variante associada à doença. A Figura 5 traz um resumo dos critérios de filtragem utilizados para análise dos resultados do painel. Eles são baseados em evidências populacionais, nos tipos de variantes, nas consequências das mesmas para os genes e seus produtos (evidência computacional) e no padrão de herança esperado para o fenótipo do paciente. Os critérios para a filtragem dos exomas seguem o mesmo fluxograma, com a diferença de que ela é realizada por núcleos familiares. Isso inclui uma nova etapa onde as variantes são filtradas de acordo com o fenótipo de cada membro da família.

Confirmação e segregação das variantes identificadas. As variantes alélicas consideradas relevantes foram confirmadas visualmente através do programa *Integrative Genomics Viewer (IGV)*. Essa visualização do sequenciamento a partir do arquivo *.bam permite a identificação de falsas variantes chamadas incorretamente pelo programa *Freebayes*, principalmente por problemas na cobertura do sequenciamento. Uma vez confirmadas no *IGV*, as variantes de interesse foram confirmadas e segregadas nas famílias pelo sequenciamento por *Sanger*. Resumidamente, foram sintetizados oligonucleotídeos iniciadores específicos para amplificar por PCR as regiões que flanqueiam as variantes. As amplificações foram acompanhadas de um controle negativo e os produtos amplificados foram submetidos a uma purificação enzimática com a enzima *ExoSAP-IT (Affymetrix, Thermo Fisher Scientific, Waltham, MA, EUA)* conforme a recomendação do fabricante. O sequenciamento foi realizado com o *kit ABI Prism™ Big Dye Terminator Cycle Sequencing Ready Reaction Kit 3.1 (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, EUA)* e os produtos desta reação foram submetidos a uma eletroforese capilar no sequenciador automático *ABI Prism 3130xl Genetic Analyzer (Applied Biosystems, Thermo Fisher Scientific, Waltham, MA, EUA)*. A leitura dos eletroferogramas foi realizada visualmente com o auxílio do programa *Sequencher 4.10.1 (GeneCodes Corporation, Ann Arbor, MI, EUA)*, sempre em comparação à sequência referência do gene e a um eletroferograma de um indivíduo normal.

Figura 5 - Fluxograma com a estratégia de filtragem das variantes identificadas no SPLE do painel. MAF: frequência do alelo menos frequente (de *minor allele frequency*); SELA: Laboratório de Sequenciamento em Larga Escala da Faculdade de Medicina da USP; D: deletério; P: possivelmente patogênico; A: deletério; H: efeito alto; M: efeito moderado; *IGV*: programa *Integrative Genomics Viewer*; ACMG/AMP: Colégio Americano de Genética Médica e Associação de Patologia Molecular; MAF<0,01 nesse caso significa uma frequência inferior a 1% nos bancos de dados populacionais (*gnomAD*, *1000Genomes*, *ExAC* e *ESP*) e uma frequência inferior a 1% nos indivíduos sequenciados no SELA.

4.3.4.2. Identificação de CNVs

Chamada das CNVs. A partir do mesmo arquivo *.bam utilizado para a chamada das variantes de ponto e pequenas indels, foi realizado uma análise de CNVs com o programa *COPY Number Targeted Resequencing Analysis (CONTRA)* (60). Esse programa é capaz de identificar a presença de deleções ou duplicações de grandes regiões a partir da comparação da cobertura do sequenciamento do paciente como uma linha de base criada a partir da análise de cobertura de vários indivíduos. Para isso, ele fragmenta todas as regiões sequenciadas em regiões menores e libera o resultado de cada região em *log-ratio* juntamente com uma análise de significância (valores de *p*). *Log-ratios* próximos de -1,0, com valores de *p* ajustados inferiores a 0,05, indicam a presença de deleções em heterozigose. As regiões com deleções em homozigose apresentam *log-ratios* muito baixos, cerca de -7,0. E *log-ratios* próximos de +1,0, indicam a presença de duplicações. O critério utilizado para a filtragem das planilhas foi de $\pm 0,7$.

Confirmação das deleções relacionadas ao gene *SHOX* por MLPA. As CNVs relacionadas ao *SHOX*, identificadas pela análise do *CONTRA*, quando possível foram confirmadas e segregadas pela metodologia de MLPA. Todas as reações foram realizadas de acordo com as instruções do protocolo do *kit SALSA MLPA P018-G1 SHOX* (versão G01-02) (*MRC Holland*, Amsterdam, Holanda). Esse *kit* possibilita a análise simultânea de CNVs de 48 regiões, sendo 26 referentes à PAR1 (incluindo o gene *SHOX* e suas regiões regulatórias), 13 referentes a diferentes regiões do cromossomo X e nove, a regiões de referência (controles). As amostras processadas com o *kit* acima foram submetidas a uma eletroforese capilar no sequenciador automático *ABI Prism 3130xl Genetic Analyzer* e processadas pelo programa *GeneMapper 5.0* (ambos *Applied Biosystems*, *Thermo Fisher Scientific*, Waltham, MA, EUA). Os dados obtidos foram analisados pelo programa *MRC Coffalyzer* (*MRC Holland*, Amsterdam, Holanda), que determina a relação das áreas dos picos de interesse com as áreas dos picos das regiões controle, considerando um pico como sendo normal quando a razão entre essas áreas estiver entre 0,7 e 1,3, deletado quando for inferior a 0,7 e duplicado quando for superior a 1,3. Quatro

indivíduos com estatura normal foram utilizados como grupo controle para a validação da reação e análise dos pacientes.

As análises de bioinformática desse projeto foram realizadas em colaboração com o Dr. Antonio Marcondes Lerário (do Departamento de Medicina Interna da Divisão de Metabolismo, Endocrinologia e Diabetes da Universidade de Michigan, nos Estados Unidos), e com o Lucas Santos de Santana (da Unidade de Endocrinologia Genética, Laboratório de Endocrinologia Celular e Molecular - LIM25). Eles foram responsáveis pelo processamento das métricas das corridas e dos dados brutos até a geração dos arquivos analisáveis (VCF anotada e arquivo final do *CONTRA*).

4.3.5. Determinação da patogenicidade das variantes identificadas

As principais variantes identificadas foram classificadas de acordo com os critérios definidos, em 2015, pelo Colégio Americano de Genética Médica e pela Associação de Patologia Molecular (ACMG/AMP) (61).

Esses critérios, baseados em evidências populacionais, genéticas, computacionais e funcionais, têm como objetivo parametrizar a avaliação das variantes em relação seu efeito sobre o fenótipo dos pacientes. Desta maneira as variantes podem ser classificadas em cinco grupos: (a) patogênicas, (b) provavelmente patogênicas, (c) provavelmente benignas, (d) benignas e (e) de significado incerto. Os Anexos E e F trazem todos os critérios definidos pela ACMG/AMP e as normas para combinação dos mesmos e classificação das variantes.

Na maioria dos casos, a classificação foi realizada com o auxílio da ferramenta *InterVar* (<http://wintervar.wglab.org/>) (62). Essa ferramenta faz uma classificação automática das variantes do tipo *missense* baseada em um levantamento da literatura e nas informações presentes nos bancos de dados disponíveis *on line*. Em seguida, essa classificação pode ser então modificada com a inclusão de informações específicas do paciente, como dados de segregação da variante na família, resultados de estudos funcionais realizados, etc.

5. RESULTADOS

5.1. Sequenciamento paralelo em larga escala do painel

5.1.1. Caracterização dos pacientes selecionados para o estudo do painel

Foram selecionados 15 pacientes com defeitos já conhecidos no gene *SHOX* e/ou suas regiões regulatórias e 77 pacientes com BE isolada sem causa definida. Todos eles foram submetidos ao SPLE e tiveram suas bibliotecas preparadas com as sondas da versão 1 ou 2 do painel. Algumas características clínicas dos 77 pacientes com BE isolada estão resumidas na Tabela 1. Os dados foram obtidos na avaliação inicial dos pacientes.

Tabela 1 - Características clínicas dos 77 pacientes com baixa estatura isolada submetidos ao painel

Número de pacientes	77
Sexo (feminino/masculino)	26/51
Idade cronológica (anos)	9,0 ± 3,6
História familiar de BEI	44 (57%)
Mãe com baixa estatura	24 (31%)
Pai com baixa estatura	15 (19%)
Mãe e pai com baixa estatura	5 (6%)
Pacientes com altura dentro da altura alvo	57 (74%)
História familiar de consanguinidade	4 (5%)
História de atraso puberal	12 (16%)
Pacientes com avanço da idade óssea em pré-puberdade	6 (8%)
Z da altura	-2,8 ± 0,6
Pacientes com desproporção corpórea	30 (39%)
Z da altura sentada/altura total	1,3 ± 1,6

BEI: baixa estatura isolada; Z: escore-Z.

Muitos desses 77 pacientes já tinham sido previamente submetidos a alguma avaliação molecular de pelo menos um gene considerado candidato: 31 (40%) tinham sido avaliados para defeitos no gene *SHOX*, 19 (25%) para mutações no gene *NPR2*, e 15 (19%) para mutações na região *hot-spot* do gene *FGFR3*. Oito (8/77; 10%) deles tinham o estudo desses três genes. Trinta e oito (38/77; 49%) pacientes não tinham sido submetidos a nenhum tipo de análise molecular prévia.

5.1.2. Métricas do sequenciamento do painel

As métricas das corridas realizadas estão resumidas na Tabela 2. A cobertura média dos principais genes do projeto foi de 278 vezes. Em média, 99,94% das regiões codificadoras dos genes alvo do painel apresentaram uma cobertura de mais de 10 vezes, e 99,85%, mais de 20 vezes.

Em uma análise voltada para o *SHOX*, a cobertura horizontal específica para a região codificadora do transcrito principal (NM_000451) foi de 100%. A cobertura mínima nessa região foi de 44x. Em relação à região genômica, também do transcrito principal, 81% dela foi coberta acima de 10x. A Figura 6 mostra a cobertura real do *SHOX* nas duas versões do painel. Nela é possível identificar algumas das regiões do gene que não foram cobertas.

Em relação as regiões regulatórias, também não conseguimos uma cobertura de pelo menos 10x em 100% delas, conforme já previsto inicialmente pela ferramenta que desenhou as sondas (*Agilent SureDesign 2.0*). Apesar disso, conseguimos realizar a análise dessas regiões. Das 12 regiões regulatórias incluídas para a captura e sequenciamento, cinco tiveram 100% da sua região alvo coberta com mais de 10x de cobertura, e oito apresentaram boa cobertura em pelo menos 95% da sua região alvo. As coberturas previstas e obtidas para essas regiões estão descritas no anexo B.

Tabela 2 - Métricas das corridas que continham os pacientes do painel realizadas nos equipamentos *NextSeq500 Illumina*, com kits *NextSeq MidiOutput V2 2x150* e *MiSeq Illumina*, com kits *MiSeq V3 2x300*

	Média	Mínimo	Máximo
nº de reads gerados	1.610.683	557.504	4.055.173
nº de bases sequenciadas (pb)	124.844.186	50.849.203	255.854.551
Cobertura média (x)*	278	120	546
% com cobertura maior que 10x*	99,94	98,83	99,99
% com cobertura maior que 20x*	99,85	99,50	99,99

As bibliotecas do painel versão 1 foram incluídas em pools de 96 bibliotecas e cada pool foi sequenciado com kits *NextSeq MidiOutput V2 2x150*. As bibliotecas da versão 2 foram incluídas em pools de 32 bibliotecas e sequenciadas com kits *MiSeq V3 2x300*. Foram feitas ao todo seis corridas. *Para esses cálculos foram consideradas apenas as regiões codificadoras dos genes relacionados diretamente ao projeto. pb: pares de base; x: vezes; %: porcentagem.

Figura 6 - Imagem do programa *IGV* mostrando a cobertura real do sequenciamento do gene *SHOX* no painel. Na parte inferior da figura, com a indicação “*SHOX*”, as caixas pretas enumeradas de 1 a 6b, correspondem aos exons do gene. Na região com indicação da “cobertura”, quanto maior a altura da barra cinza, maior é a cobertura. Na região com indicação de “reads”, as barras na horizontal correspondem aos vários *reads* dessa região. Esses *reads* continuam para baixo, se formos rolando a barra lateral para baixo. Cada *read* corresponde a uma vez que a região foi sequenciada. As regiões em branco, sem barras horizontais, indicadas pelas setas pretas, correspondem às regiões que não foram cobertas, e as setas brancas indicam regiões onde há uma diminuição significativa da cobertura.

5.1.3. Avaliação do gene *SHOX* pelo painel

Quinze pacientes com alterações moleculares já conhecidas foram selecionados para avaliar o desempenho do painel em detectar alterações no gene *SHOX*. Selecionamos dois pacientes com mutações de ponto em heterozigose no *SHOX* (c.105C>A: p.Tyr35* e c.440G>A: p.Arg147His), sendo uma paciente do sexo feminino e outro do sexo masculino (Pacientes 1 e 2). Foram selecionados ainda 13 pacientes com diferentes CNVs envolvendo o gene *SHOX* e/ou suas regiões regulatórias. Cinco pacientes apresentavam grandes deleções na PAR1, envolvendo o *SHOX*, toda a área do *SHOX* e outros genes (Pacientes 3 a 7); três pacientes apresentavam deleções um pouco menores, envolvendo o gene e parte da área do *SHOX* (Pacientes 8 a 10); uma paciente com uma deleção intragênica do *SHOX*, envolvendo os exons 4, 5 e 6a, porém com pontos de quebra ainda indefinidos (Paciente 11) (38); uma paciente com uma duplicação envolvendo os exons 1, 2 e 3 (Paciente 12); e três pacientes com deleções isoladas de regiões localizadas a 3’ do gene (Pacientes 13 a 15). A Figura 7 contém todos esses pacientes e suas respectivas regiões de deleção ou duplicação, identificadas previamente por MLPA.

Figura 7 - Mapa de CNVs dos casos com alterações previamente conhecidas e dos pacientes sem diagnóstico que tiveram CNVs identificadas pelo painel. As sondas do MLPA de diversos kits utilizados e suas localizações estão descritas na parte superior da figura. Os quadros brancos indicam as regiões preservadas, os cinzas escuros, as regiões com deleção em heterozigose, os pretos, com deleção em homozigose e os quadros cinza claro indicam regiões duplicadas. A esquerda está identificação dos indivíduos com CNVs. Os pacientes de 3 a 15 correspondem aos pacientes com alteração já conhecidas utilizados para a análise de eficiência do painel, enquanto dos pacientes 16 e 17, correspondem aos casos pacientes sem diagnóstico. PAR1: região pseudoautossômica 1.

Com essa análise pretendíamos avaliar: (a) a capacidade do painel de capturar e sequenciar o gene *SHOX*, que tem sido descrito como um gene difícil de ser estudado por SPLE (63), (b) a eficiência, tanto do kit de preparo de bibliotecas, quanto dos programas de análise, em identificar diferentes tipos de variantes alélicas (desde variantes de ponto até CNVs de diferentes tamanhos), e (c) a capacidade do painel de identificar variantes alélicas localizadas em regiões pseudoautosômicas dos cromossomos sexuais, localizadas tanto no cromossomo X, como no Y, e por esse motivo, selecionamos pacientes dos dois sexos (cinco do sexo masculino e oito do feminino).

A análise da VCF identificou as duas mutações de ponto (2/2; 100%) previamente conhecidas. A Figura 8 mostra uma imagem do programa IGV com a mutação de ponto c.440G>A: p.Arg147His em heterozigose, do Paciente 2, para exemplificar a visualização desse tipo de variante no SPLE.

Figura 8 - Visualização da variante c.440G>A: p.Arg147His do *SHOX* em heterozigose no sequenciamento do painel, pelo programa IGV, e no sequenciamento por Sanger. A: A imagem, gerada pelo programa IGV, mostra o sequenciamento do indivíduo com a mutação c.440G>A: p.Arg147His em heterozigose na parte inferior (Paciente 2) e de um indivíduo sem a mutação na parte superior (Controle). No indivíduo com a mutação, vemos, na região indicada pelo círculo preto, a presença de duas cores, que correspondem à presença de duas bases diferentes (G e A). O mesmo não é visualizado no indivíduo sem a mutação (região indicada pela seta branca). B: Eletroferograma do sequenciamento por Sanger do indivíduo com a mutação, mostrando a mesma em heterozigose (região indicada pela seta).

A partir da análise visual do sequenciamento do *SHOX* no *IGV* foi possível identificar as 13 CNVs previamente conhecidas (Pacientes 3 a 15), independente do sexo dos pacientes. Nos pacientes com deleções foi possível visualizar uma diminuição da cobertura, com a identificação dos pontos de quebra em alguns casos, e na paciente com a duplicação (Paciente 12), foi possível observar um aumento considerável da cobertura (Figura 9; 9B e 9C). Foi observado também nessa paciente um desequilíbrio entre a razão das SNVs localizadas na região duplicada de 70:30, e não de 50:50, conforme esperado para uma região bialélica (Figura 9C).

A análise do *CONTRA* foi capaz de identificar 12 (12/13; 92%) CNVs. O programa não foi capaz de identificar a duplicação da Paciente 12. Nesse caso, o gráfico do *CONTRA* mostrou uma tendência de duplicação (Figura 9A), mas os *log-ratios* ajustados observados na região duplicada ficaram com uma média de +0,48, muito distantes do +1,0 esperado, e por isso acabaram sendo excluídos durante a filtragem da planilha.

O painel foi capaz ainda de identificar os pontos de quebra da deleção intragênica do Paciente 11, uma vez que o painel sequenciou “toda” a região genômica do gene. Diante dos resultados obtidos pelos programas *CONTRA* e *IGV* para essa paciente, conseguimos sintetizar oligonucleotídeos específicos para amplificar e confirmar os pontos de quebra da deleção por *Sanger*. Pelo MLPA, tínhamos uma deleção envolvendo três exons (4, 5 e 6a) com um intervalo mínimo de 4.000 pb e um intervalo máximo de 14.000 pb. Após o sequenciamento do painel e confirmação por *Sanger*, o intervalo deletado foi definido em 11.722 pb. A Figura 10 traz os detalhes desses resultados.

O sequenciamento das regiões regulatórias do *SHOX* também permitiu a identificação dos pontos de quebra em dois (2/3) pacientes com deleções isoladas na região localizada 3' do gene (Pacientes 13 e 14) (Figura 11). A visualização desses pontos de quebra no *IGV* permitiu o desenho de oligonucleotídeos iniciadores específicos e o sequenciamento da região. Foi possível então concluir que esses dois pacientes eram portadores de uma mesma deleção de 47.543 pb, localizada em um sítio regulatório importante, anteriormente descrita por Benito-Sanz e colaboradores (16).

Figura 9 - Análises do *CONTRA* e do *IGV* da Paciente 12, com duplicação dos exons 1, 2 e 3. A: Gráfico do *CONTRA* mostrando alguns pontos acima da linha do *log-ratio* igual a zero (indicados pela seta), mas muito distantes do +1,0, que indicaria uma duplicação. Apesar da tendência observada, o programa não foi capaz de identificar a duplicação. B e C: Imagens do *IGV* mostrando a cobertura do sequenciamento do *SHOX* e a duplicação da paciente. As duas imagens mostram o sequenciamento da paciente na parte inferior (Paciente 12) e de um indivíduo normal na parte superior (Controle). B: Nessa imagem mais ampla, vemos na paciente um aumento cobertura nas regiões duplicadas (região marcada com a caixa). C: Essa imagem traz alguns detalhes de parte da região duplicada. As linhas pretas na base do gráfico de cobertura (indicadas pelas setas em B e C) indicam um excesso de *reads* nessas regiões. Além disso, podemos observar em C um desequilíbrio entre a razão das SNVs localizadas nessa região duplicada, de aproximadamente 70:30 e não de 50:50, conforme esperado para uma região bialélica (as cores que indicam as variantes não correspondem a 50% dos alelos).

Figura 10 - Análises do *CONTRA*, *IGV* e *Sanger* da deleção intragênica envolvendo os exons 4, 5 e 6a da Paciente 11. A: Gráfico do *CONTRA* da paciente mostrando alguns pontos com *log-ratios* próximos de -1,0 na região do gene *SHOX*, que indicam a presença de deleção em heterozigose (pontos indicados pela seta). B e C: Imagens do *IGV* mostrando com detalhe os pontos de quebra da deleção. As duas imagens, mostram o sequenciamento da paciente com a deleção intragênica na parte inferior (Paciente 11) e de um indivíduo normal na parte superior (Controle). Na paciente vemos uma queda brusca de cobertura nas regiões indicadas pelas setas pretas, que correspondem aos pontos de quebra da deleção (B: ponto de quebra 5'; C: ponto de quebra 3'). O mesmo não é visualizado no indivíduo sem a deleção (regiões indicadas pelas setas brancas). D: Sequenciamento por *Sanger* da região que contém o ponto de quebra da deleção, que possibilitou a definição exata do fragmento deletado em 11.722 pb.

Figura 11 - Análises do *CONTRA*, *IGV* e *Sanger* da deleção a 3' do *SHOX* da Paciente 13. Resultados semelhantes foram identificados na Paciente 14. A: Gráfico do *CONTRA* da paciente mostrando alguns pontos com *log-ratios* próximos de -1,0 na região relacionada ao gene *SHOX*, que indicam a presença de deleção em heterozigose (pontos indicados pela seta). B, C e D: Imagens do *IGV* mostrando com detalhe os pontos de quebra da deleção. As imagens mostram o sequenciamento da paciente com a deleção na parte inferior (Paciente 13) e de um indivíduo normal na parte superior (Controle). C: Na paciente vemos uma queda brusca de cobertura na região indicada pela seta preta, que corresponde ao ponto de quebra 5' da deleção. O mesmo não é visualizado no indivíduo sem a deleção (região indicada pela seta branca). D: É possível ver um pedaço pequeno de alguns *reads* em uma região teoricamente sem cobertura no sequenciamento da paciente (dentro da caixa). A presença desses *reads*, que deveriam ter se alinhado a outra região, indica o ponto de quebra 3' da deleção. E: Sequenciamento por *Sanger* da região que contém o ponto de quebra da deleção, que possibilitou a definição exata do fragmento deletado em 47.543 pb.

O painel também não enfrentou dificuldades para identificar a deleção a 3' linha do gene das Pacientes 5 e 15, em homozigose e heterozigose, respectivamente (Figura 12). Essas duas pacientes são filha e mãe. A filha apresenta uma grande deleção na PAR1, envolvendo o *SHOX*, sua região regulatória e alguns outros genes, no alelo paterno, e essa deleção menor, localizada a 3' do gene, no alelo herdado da mãe. Como consequência disso, ela apresenta uma deleção em homozigose dessa região.

A avaliação prospectiva do gene *SHOX* nos 77 pacientes com BE isolada sem diagnóstico ainda identificou duas deleções intragênicas através da associação da análise do *CONTRA* com a inspeção visual do sequenciamento no IGV.

Paciente 16: deleção do exon 1 do *SHOX*

A análise de CNVs pelo programa *CONTRA* foi capaz de identificar uma deleção intragênica, em heterozigose, envolvendo o exon 1 do *SHOX* em uma paciente do sexo feminino com baixa estatura desproporcional, que ainda não havia sido submetida ao estudo do *SHOX*.

O programa detectou a deleção de três fragmentos na região do *SHOX*, com *log-ratios* médios de -0,97 a -1,0, e valores de *p* ajustados inferiores a 0,01. O intervalo mínimo de deleção predito, com base nos fragmentos deletados, foi de 4.586 pb. Essa deleção foi confirmada por MLPA, onde a sonda correspondente ao exon 1 (01145-L00702; kit P018-G1) também estava deletada em heterozigose. Pelo MLPA, o intervalo máximo de deleção previsto, baseado nas sondas conservadas foi de 11.400 pb. Nesse caso, não conseguimos delimitar os pontos de quebra da deleção. Os resultados, tanto do *CONTRA* quanto do MLPA, estão disponíveis na Figura 13. A deleção também foi identificada na mãe da paciente que é desproporcional apesar de apresentar altura dentro da normalidade.

Figura 13 - Análises do CONTRA e do IGV da Paciente 16, com deleção isolada do exon 1. A: A imagem mostra o resultado da análise de CNVs pelo programa CONTRA da paciente com a deleção isolada do exon 1. Na região correspondente ao *SHOX*, visualizamos alguns pontos com log-ratios próximos de -1,0, o que indica a presença de deleção em heterozigose (região indicada pela seta). B e C: Imagens do IGV mostrando a cobertura do sequenciamento do *SHOX* e a deleção do exon 1 da paciente. As duas imagens mostram o sequenciamento da paciente com a deleção na parte superior (Paciente 16) e de um indivíduo normal na parte inferior (Controle). C: Essa imagem traz os detalhes das regiões demarcadas pelas caixas em B. Podemos ver uma queda de cobertura nas regiões deletadas da paciente. Vemos também que não há SNVs em heterozigose nessa região, o que sugere a presença de apenas um alelo.

Paciente 17: deleção de parte do exon e do intron 3

A análise visual do *SHOX* através do *IGV* mostrou uma diminuição da cobertura do sequenciamento no final do exon 3 e levou à identificação de uma deleção parcial envolvendo os 20 nucleotídeos finais desse exon e parte do intron 3 em uma paciente do sexo feminino com baixa estatura desproporcional familiar. Essa paciente apresentava um estudo prévio do *SHOX* negativo, tanto para deleções como para mutações de ponto.

A partir das coordenadas observadas no *IGV*, foram desenhados oligonucleotídeos iniciadores específicos para amplificar a região que continha os pontos de quebra, e foi possível definir o tamanho exato da deleção em 679 pb. A deleção foi segregada na família da paciente, estando presente tanto no pai, quanto na irmã, que apresentam estatura dentro da normalidade, mas são desproporcionais. Essa deleção não foi identificada pela análise do *CONTRA* provavelmente por se tratar de uma deleção pequena, que não envolve um exon inteiro. Além disso, nessa análise específica, a região deletada ficou localizada na extremidade de dois fragmentos gerados pelo programa, o que pode ter dificultado a identificação da mesma. A Figura 14 traz os resultados das análises dessa paciente.

Figura 14 - Análises do *CONTRA*, *IGV* e *Sanger* da deleção intragênica envolvendo parte do exon 3 do *SHOX* da Paciente 17. A: Gráfico do *CONTRA* da paciente indicando ausência de deleções na região do *SHOX* (região apontada pela seta em branco sem pontos indicando deleções). B: Imagem do *IGV* mostrando com detalhe os pontos de quebra da deleção. A imagem mostra o sequenciamento da paciente com a deleção na parte inferior (Paciente 17) e de um indivíduo normal na parte superior (Controle). No sequenciamento da paciente vemos uma queda brusca de cobertura na região entre as setas pretas, que correspondem aos pontos de quebra da deleção. O mesmo não é visualizado no indivíduo sem a deleção (regiões indicadas pelas setas brancas). C: Sequenciamento por *Sanger* da região que contém o ponto de quebra da deleção, que possibilitou a definição exata do fragmento deletado em 679 pb.

5.1.4. Variantes identificadas em outros genes pelo painel

O sequenciamento dos 77 pacientes submetidos ao painel (46 pacientes na versão 1, e 31 na versão 2) identificou, em média, 478 variantes por paciente, sendo 25 variantes consideradas raras. Essas variantes foram selecionadas de acordo com a localização, sendo priorizadas aquelas localizadas em regiões exônicas ou em sítios de *splice*. Após a exclusão das variantes sinônimas, restaram três variantes, em média, por paciente, totalizando 184 variantes selecionadas, todas elas em heterozigose simples. A Tabela 3 apresenta o detalhamento da filtragem acima.

Tabela 3 - Detalhamento da filtragem das variantes identificadas pelo painel

Crítérios de filtragem	Média	Mínimo	Máximo
Variantes na planilha	1757	1230	2520
Variantes por paciente	478	384	619
Variantes raras na população (MAF<0,01)*	48	16	118
Variantes raras nos indivíduos do SELA (MAF<0,01)**	25	8	44
Variantes em regiões codificadoras e sítios de <i>splice</i>	6	0	12
Variantes de interesse após a exclusão das variantes sinônimas	3	0	7

MAF: frequência do alelo menos frequente (de *minor allele frequency*); SELA: Laboratório de Sequenciamento em Larga Escala da Faculdade de Medicina da USP; *MAF<0,01 nesse caso significa uma frequência inferior a 1% nos bancos de dados populacionais (*gnomAD*, *1000Genomes*, *ExAC* e *ESP*); **MAF<0,01 nesse caso significa uma frequência inferior a 1% nos indivíduos sequenciados no SELA.

A classificação final das 184 variantes foi realizada de acordo com os critérios do ACMG/AMP. Um total de 50 (50/184; 27%) variantes foram selecionadas para a confirmação e segregação por *Sanger*. Após a análise das segregações e reclassificação das variantes, seis foram consideradas patogênicas e sete provavelmente patogênicas, em 12 pacientes (Tabela 4). Os critérios do ACMG/AMP utilizados para a classificação dessas variantes estão descritos no Anexo G.

Se somarmos a essa análise de variantes de ponto, os resultados obtidos pela análise de CNVs, que identificou duas deleções intragênicas no *SHOX*, descritas acima, teremos 15 variantes em 14 pacientes (Tabela 4). Nove (9/15; 60%) variantes afetam fatores parácrinos que regulam a cartilagem de crescimento (3x*FGFR3*, 1x*GNAS*, 1x*NPR2* e 4x*IHH*), quatro (4/15; 26%) afetam vias intracelulares (2x*SHOX*, 1x*BRAF* e 1x*NF1*) e outras duas (2/15; 13%) afetam genes do eixo GH/IGF-1 (1x*IGF1R* e 1x*GH1*). Onze (11/15; 73%) das variantes

identificadas afetam genes de alguma forma associados à fisiologia da cartilagem de crescimento (3x*FGFR3*, 1x*GNAS*, 4x*IHH* e 1x*NPR2* e 2x*SHOX*). Com esses resultados foi possível estabelecer o diagnóstico molecular em 18% (14/77) dos pacientes com baixa estatura isolada. Lembrando que 39 (39/77; 51%) pacientes já tinham sido submetidos a alguma avaliação molecular prévia.

Entre as 184 variantes selecionadas, ainda foram identificadas 18 variantes benignas, 21 provavelmente benignas e 123 de significado incerto (VUS). As nove variantes restantes ou não foram confirmadas no IGV ou não estavam presentes no transcrito principal dos seus genes. Após o estudo de segregação de 32 VUS, 11 variantes foram consideradas variantes candidatas, uma vez que segregaram com o fenótipo da baixa estatura na família e/ou estavam ausentes nos familiares com estatura normal (Anexo H). Outras oito VUS, localizadas em genes interessantes que poderiam explicar o fenótipo dos pacientes, também foram consideradas candidatas apesar de não terem sido segregadas por falta de amostras de DNA dos familiares (Anexo I). Sendo assim, 19 VUS de interesse foram identificadas em 15 pacientes (15/77; 19%). Essas variantes acabaram sendo incluídas nesse grupo porque, apesar de raras e preditas como patogênicas nos modelos *in silico*, algumas estavam localizadas em genes candidatos novos, que não possuem estudos suficientes para serem associados a determinados quadros clínicos, algumas variantes foram identificadas em um estado diferente do esperado para o padrão de herança da doença, em alguns casos o fenótipo do paciente era diferente do fenótipo já descrito para o gene e em outros não foi possível realizar o estudo de segregação.

5.1.5. Caracterização fenotípica dos pacientes com variantes patogênicas ou provavelmente patogênicas

Na Tabela 5 estão descritas as principais características clínicas dos pacientes portadores das variantes classificadas como patogênicas ou provavelmente patogênicas identificadas pelo painel.

Tabela 4 - Detalhes das principais variantes alélicas identificadas no sequenciamento do painel com a classificação de acordo com os critérios do ACMG/AMP (61)

Paciente	Gene	Sequência referência	Variante	Tipo	dbSNP ID	Frequência gnomAD	Frequência ABraOM	Classificação ACMG/AMP
16	SHOX	NM_000451.3	Deleção do exon 1	Deleção	ND	NA	Ausente	Patogênica
17	SHOX	NM_000451.3	Deleção de 679 pb	Deleção	ND	NA	Ausente	Patogênica
18	FGFR3	NM_000142.4	c.667C>T: p.Arg223Cys	Missense	ND	Ausente	Ausente	Provavelmente patogênica
19	FGFR3	NM_000142.4	c.1620C>A: p.Asn540Lys	Missense	rs28933068	Ausente	Ausente	Patogênica
20	FGFR3	NM_000142.4	c.1657G>T: p.Val553Leu	Missense	rs199544087	8,224x10 ⁻⁶	Ausente	Provavelmente patogênica
21	NPR2	NM_003995.3	c.1711-1G>C	Splice site	ND	Ausente	Ausente	Patogênica
22	IHH	NM_002181.3	c.172G>A: p.Glu58Lys	Missense	ND	Ausente	Ausente	Patogênica
23	IHH	NM_002181.3	c.532G>C: p.Val178Leu	Missense	ND	Ausente	Ausente	Patogênica
24	IHH	NM_002181.3	c.685G>A: p.Val229Met	Missense	rs975084944	Ausente	Ausente	Provavelmente patogênica
	GH1	NM_000515.4	c.124C>T: p.Arg42Cys	Missense	ND	1,167x10 ⁻⁴	Ausente	Provavelmente patogênica
25	IHH	NM_002181.3	c.949G>A: p.Val317Met	Missense	ND	4,066x10 ⁻⁶	Ausente	Provavelmente patogênica
26	GNAS	NM_080425.3	c.2682C>G: p.Ser894Arg	Missense	ND	Ausente	Ausente	Patogênica
27	IGF1R	NM_000875.4	c.143G>A: p.Arg48His	Missense	rs774734018	3,979x10 ⁻⁶	Ausente	Provavelmente patogênica
28	BRAF	NM_004333.4	c.712G>A: p.Val238Ile	Missense	rs373935661	4,062x10 ⁻⁶	Ausente	Provavelmente patogênica
29	NF1	NM_000267.3	c.3447G>A: p.Met1149Ile	Missense	rs1064794277	Ausente	Ausente	Patogênica

Os detalhes da classificação da ACMG/AMP estão descritos no Anexo G. As frequências das variantes foram obtidas em fevereiro de 2019. NA: não aplicável; ND: não descrito.

Tabela 5 - Características clínicas dos pacientes portadores das variantes alélicas classificadas como patogênicas ou provavelmente patogênicas identificadas no sequenciamento do painel

Paciente	Gene	Sexo	IC (anos)	IO-IC (anos)	Z da altura	Z da AS:A	Z do IMC	Z da altura pai	Z da altura mãe	Achados clínicos adicionais
16	<i>SHOX</i>	Fem.	9,3	-0,4	-2,0	3,0	1,2	-0,9	-0,3	Ausentes
17	<i>SHOX</i>	Fem.	12,2	-1,7	-2,9	5,2	-1,2	-1,0*	0,1	Ausentes
18	<i>FGFR3</i>	Fem.	17,2	ND	-4,1	3,6	1,4	-0,7	-3,9	Perímetro cefálico no limite superior da normalidade e acentuação da lordose lombar
19	<i>FGFR3</i>	Fem.	6,4	0,5	-2,6	3,7	1,5	ND	0,0**	Acentuação da lordose lombar e geno varo
20	<i>FGFR3</i>	Fem.	11,6	-1,2	-2,7	1,1	-0,5	-0,6	-2,6	Ausentes
21	<i>NPR2</i>	Masc.	9,7	-3,7	-2,5	-1,3	1,2	-2,2	-1,5	Pavilhão auricular proptótico de implantação normal, epicanto discreto, base do nariz rasa, bochechas proeminentes e filtro nasal longo e raso
22	<i>IHH</i>	Masc.	3,4	-0,7	-2,8	3,1	1,6	-1,2	-2,2	Palato ogival, hipertelorismo mamário, hipertrofia muscular, pés chatos, defeito de fusão posterior de T12 e L1 e 5º metacarpo curto
23	<i>IHH</i>	Masc.	7,1	-1,3	-2,5	2,2	0,8	-1,2	-0,2	Sobrelhas marcadas
24	<i>IHH/GH1</i>	Fem.	12,9	-1,9	-3,8	1,9	-0,8	-2,2	0,0	Estrabismo, implantação do cabelo em tridente e cúbito valgo. Escore-Z do IGF-1 = 0,1
25	<i>IHH</i>	Masc.	7,5	-2,5	-2,3	0,0	0,3	-1,8***	0,5	Clinodactilia
26	<i>GNAS</i>	Fem.	11,8	1,2	-3,4	2,8	0,2	-0,8	0,3	<i>Pectus excavatum</i> , polegar hipoplásico e falange distal alargada, palato ogival, 4º metacarpo curto e calcificações subcutâneas. Dosagens de cálcio, fósforo e PTH normais
27	<i>IGF1R</i>	Masc.	8,8	-1,8	-3,1	2,1	0,5	-1,8	-1,8	Pequeno para idade gestacional para peso e microcefalia
28	<i>BRAF</i>	Masc.	8,7	-2,2	-3,0	-0,7	-0,9	-2,0	-3,1	Acentuação da cifose torácica
29	<i>NF1</i>	Masc.	9,7	-5,9	-4,1	1,5	0,0	-3,0	-2,2	Alteração no pavilhão auricular, <i>pectus excavatum</i> e manchas café com leite

*O pai da Paciente 17, é desproporcional (Z da AS:A: 5,5), apesar da estatura normal. **A altura da mãe da Paciente 19 foi obtida a partir de informações colhidas. ***O pai do Paciente 25 possui algumas alterações faciais e alguns familiares com baixa estatura, apesar de apresentar estatura dentro da normalidade. Fem.: feminino; Masc.: masculino; IC: idade cronológica; IO-IC: diferença entre a idade óssea a idade cronológica; Z: escore-Z; IMC: índice de massa corpórea; AS:A: relação da altura sentada dividida pela altura total; ND: não disponível; IGF-1: fator 1 de crescimento insulino-símile; PTH: paratormônio

5.1.6. Detalhes das variantes de ponto identificadas

5.1.6.1. Variantes que afetam fatores parácrinos da cartilagem de crescimento

No painel identificamos três pacientes com variantes em heterozigose no gene *FGFR3*, duas delas em pacientes com desproporção corpórea importante e alterações esqueléticas leves que poderiam estar associadas às alterações nesse gene.

A primeira variante c.667C>T: p.Arg223Cys (Paciente 18) está localizada na porção extracelular do receptor, no segundo domínio imunoglobulina *like*, que juntamente com o terceiro domínio interage diretamente com os fatores de crescimento dos fibroblastos (FGFs) e proteoglicanos de sulfato de heparano. A variante segregou com o fenótipo de baixa estatura desproporcional da família, uma vez que foi herdada da mãe que apresentava fenótipo semelhante ao da paciente.

A segunda e a terceira variante, c.1620C>A: p.Asn540Lys (Paciente 19) e c.1657G>T: p.Val553Leu (Paciente 20), estão localizadas no domínio tirosina-quinase da proteína. A p.Asn540Lys, já descrita no *dbSNP* com o rs28933068, é considerada patogênica pelo *ClinVar*, um banco de dados que possui informações clínicas relacionadas a variantes alélicas. Ela é considerada uma das mutações mais comuns associadas à hipocondroplasia (63, 64). Já a variante p.Val553Leu foi a única variante no *FGFR3* identificada em um paciente sem desproporção corpórea e com ausência de alterações esqueléticas. A variante segregou com o fenótipo, estando presente na mãe, também baixa, e ausente no pai e no irmão, ambos com estatura normal.

Foi identificada uma paciente com uma variante em heterozigose no gene *GNAS*. A variante c.2682C>G: p.Ser894Arg (Paciente 26) estava ausente nos pais, com estatura normal, e presente na irmã gêmea univitelina da paciente, que apresenta fenótipo semelhante ao dela. Ambas apresentam dificuldade escolar, alguns estigmas leves e calcificações subcutâneas. Laboratorialmente elas não apresentam alterações no metabolismo do cálcio. A paternidade e maternidade foram confirmadas por se tratar de uma variante *de novo*.

Identificamos um paciente com uma variante em heterozigose no gene *NPR2*. A variante tipo *splice site* c.1711-1G>C (Paciente 21) foi predita como deletéria em

pelo programa *dbscSNV* da planilha VCF, que avalia mutações desse tipo. A análise de predição de *splicing* foi realizada ainda em outros dois programas, o *Human splicing finder* e o *BDGP Splice site prediction*, e ambos apontaram que essa variante, localizada um nucleotídeo antes do início do exon 11 do *NPR2*, modifica o sítio de *splice*. Um deles indicou que há perda do sítio aceptor preferencial, e que outro sítio, 95 nucleotídeos a frente, passaria a ser o sítio preferencial, resultando na perda do exon 11 e inserção de regiões intrônicas no RNA mensageiro (mRNA). O endereço desses dois programas está disponível no Anexo C. A variante também não foi identificada na mãe e no irmão do paciente que apresentam estatura normal. Por falta de amostra de DNA do pai, que também apresenta baixa estatura, não foi possível concluir a análise de segregação na família.

Foram identificados ainda quatro pacientes com variantes em heterozigose no gene *IHH* (Pacientes 22 a 25), sendo que dois deles apresentavam baixa estatura com desproporção corpórea. A reavaliação das radiografias de mãos e pés desses e de outros pacientes com variantes no *IHH* (65) após o achado molecular identificou leves alterações metafisárias nas falanges médias principalmente do 2º e 5º dedo e na falange distal do polegar (Figura 15).

Figura 15 - Raio-X de mão e punho do Paciente 22 com variante no *IHH*. Podemos visualizar as alterações metafisárias nas falanges médias do 2º e 5º dedo e na falange distal do polegar.

A primeira variante, c.172G>A: p.Glu58Lys (Paciente 22), foi identificada inicialmente no caso índice de uma grande família com vários membros afetados. Ela segregou perfeitamente com o fenótipo de baixa estatura e desproporção corpórea na família. Esse paciente foi um dos que apresentou as alterações nas radiografias (Figura 15). A segunda variante, c.532G>C: p.Val178Leu (Paciente 23), foi identificada em um paciente sem história familiar de baixa estatura. A variante estava ausente nos pais. A paternidade e a maternidade também foram confirmadas nesse caso.

As outras duas variantes, c.685G>A: p.Val229Met (Paciente 24) e c.949G>A: p.Val317Met (Paciente 25), foram identificadas em pacientes sem desproporção corpórea. A Paciente 24 apresentou ainda uma segunda variante provavelmente patogênica, também em heterozigose, no gene *GH1* (c.124C>T:p.Arg42Cys). Ambas as variantes foram herdadas do pai, que também é baixo. Tanto ele, como a paciente não apresentam quadro clínico, nem exames laboratoriais compatíveis com deficiência de GH. Sendo assim, essa variante no *GH1*, apesar de classificada como provavelmente patogênica, não é suficiente para causar esse quadro clínico em heterozigose. A variante do Paciente 25 também foi herdada do pai, que apresenta estatura no limite inferior da normalidade e possui alguns familiares com baixa estatura.

5.1.6.2. Variantes que afetam vias intracelulares

Além dos dois pacientes descritos anteriormente com deleções no *SHOX*, foi identificado um paciente com uma variante em heterozigose no gene *BRAF*. A variante c.712G>A:p.Val238Ile (Paciente 28) foi identificada em um paciente com baixa estatura proporcional e acentuação da cifose torácica. Ele apresentou peso e comprimentos normais ao nascimento e desenvolvimento neuropsicomotor também normal. Entre os achados laboratoriais foi identificado níveis de IGF-1 baixos aos 9 anos, apesar de dosagens normais de IGFBP3 e GH após teste estímulo. A variante foi herdada da mãe que também apresenta uma baixa estatura importante.

Foi identificado também um paciente com uma variante em heterozigose no gene *NF1*, a variante c.3447G>A:p.Met1149Ile; rs373935661 (Paciente 29). O paciente apresenta baixa estatura familiar sem desproporção corpórea. Tanto o pai

quanto a mãe do paciente apresentam baixa estatura. Essa variante foi herdada da mãe e estava ausente no pai e no irmão do paciente, que apresenta estatura normal.

5.1.6.3. Variantes que afetam o eixo GH/IGF-1

Além da variante no gene *GH1*, que foi descrita anteriormente, identificamos uma variante em heterozigose no gene *IGF1R*, c.143G>A: p.Arg48His (Paciente 27). O paciente com baixa estatura desproporcional (família de ancestralidade oriental), nasceu pequeno para idade gestacional para o peso com microcefalia. Ele evoluiu com desenvolvimento neuropsicomotor normal e perímetro cefálico no limite inferior da normalidade. A deficiência de GH foi afastada por teste de liberação e foi observado valor elevado de IGF-1 em uma dosagem (413 µg/L, referência 67-373 µg/L). A variante foi herdada do pai com altura no limite inferior da normalidade, estando ausente na mãe e no irmão com estatura normal.

5.2. Sequenciamento paralelo em larga escala do exoma

5.2.1. Caracterização dos pacientes selecionados para o estudo do exoma

Entre os 63 pacientes que continuaram sem diagnóstico, mesmo após serem submetidos ao sequenciamento do painel, foram selecionados nove para o estudo do exoma. Entre pacientes e familiares foram realizados 23 exomas. A Figura 16 traz o heredograma dessas famílias. Algumas características clínicas dos pacientes submetidos ao exoma estão resumidas na Tabela 6. Grande parte dos dados foi obtida na avaliação inicial dos pacientes.

5.2.2. Métricas do sequenciamento do exoma

As métricas das corridas realizadas estão resumidas na Tabela 7. A cobertura média foi de 96,1 vezes. Em média, 97,4% das regiões alvo do exoma apresentaram uma cobertura maior que 10 vezes, e 94,4% maior que 20 vezes.

Figura 16 - Heredograma das nove famílias submetidas ao exoma. Os casos índices dessas famílias apresentaram resultado negativo no painel. As Famílias de 1 a 5 apresentam baixa estatura desproporcional, enquanto as Famílias de 6 a 9 apresentam apenas baixa estatura, sem desproporção corpórea. As figuras pretas indicam os indivíduos com baixa estatura. WES: representam os indivíduos submetidos ao estudo do exoma.

Tabela 6 - Características clínicas dos nove casos índice com baixa estatura isolada submetidos ao exoma

Sexo (feminino/masculino)	3/6
Idade cronológica (anos)	6,2 ± 3,1
História familiar de BEI	5 (56%)
Mãe com baixa estatura	4 (44%)
Pai com baixa estatura	1 (11%)
Pacientes com altura dentro da altura alvo	7 (78%)
História familiar de consanguinidade	1 (11%)
História de atraso puberal	0 (0%)
Pacientes com avanço da idade óssea em pré-puberdade	1 (11%)
Z da altura	-2,7 ± 0,4
Pacientes com desproporção corpórea	5 (56%)

Tabela 7 - Métricas das corridas de exoma que continham os pacientes selecionados realizadas no equipamento *HiSeq2500 Illumina*, com kits *HiSeq SBS V4 2x125*

	Média	Mediana	Mínimo	Máximo
Cobertura (x)	96,1	95,7	54,2	125,2
% com cobertura maior que 10x	97,4	98,4	88,1	98,9
% com cobertura maior que 20x	94,4	96,6	73,7	97,9

As bibliotecas do exoma foram incluídas em *pools* de 12 amostras e cada *pool* foi sequenciado em duas *lanes* com kits *HiSeq SBS V4 2x125*, totalizando 48 bibliotecas por corrida. Foram feitas ao todo quatro corridas. x: vezes; %: porcentagem.

5.2.3. Variantes identificadas pelo exoma

O sequenciamento dos 23 pacientes submetidos ao estudo do exoma identificou, em média, 104.492 variantes por paciente, sendo em 3.419 variantes consideradas raras. Essas variantes foram filtradas nas respectivas famílias de acordo com o(s) padrão(ões) de herança esperado(s) e selecionadas de acordo com a localização, sendo priorizadas aquelas localizadas em regiões exônicas ou em sítios de *splice*. Em seguida, foram priorizadas de acordo com o potencial patogênico. Após a exclusão das variantes sinônimas, restaram 135 variantes, em média, por paciente. A Tabela 8 apresenta o detalhamento dessa filtragem. Lembrando que a grande discrepância entre os valores mínimos e máximos da tabela está, de certa forma, associada aos diferentes tipos de herança esperados em cada família.

Após a análise das filtrações, não foram encontradas variantes raras consideradas responsáveis pela baixa estatura em nenhuma das nove famílias analisadas.

Tabela 8 - Detalhamento da filtragem das variantes identificadas pelo exoma

CrITÉRIOS de filtragem	MÉDIA	MÍNIMO	MÁXIMO
Variantes na família	171.384	143.600	195.753
Variantes por paciente	104.492	43.668	148.775
Variantes raras na população (MAF<0,01)*	8.114	1.083	12.666
Variantes raras nos indivíduos do SELA (MAF<0,01)**	3.419	162	5.388
Variantes após a segregação na família	910	11	2.297
Variantes em regiões codificadoras e sítios de <i>splice</i>	189	1	420
Variantes de interesse após a exclusão das variantes sinônimas	135	1	306

MAF: frequência do alelo menos frequente (de *minor allele frequency*); SELA: Laboratório de Sequenciamento em Larga Escala da Faculdade de Medicina da USP. *MAF<0,01 nesse caso significa uma frequência inferior a 1% nos bancos de dados populacionais (*gnomAD*, *1000Genomes*, *ExAC* e *ESP*); **MAF<0,01 nesse caso significa uma frequência inferior a 1% nos indivíduos sequenciados no SELA.

6. DISCUSSÃO

Esse trabalho teve como objetivo principal aumentar a taxa de diagnóstico molecular de pacientes com baixa estatura isolada sem causa definida através da identificação de mutações por sequenciamento paralelo em larga escala. Inicialmente foi utilizado um painel customizado com genes envolvidos nas principais vias do crescimento, e posteriormente, algumas famílias específicas, com resultado negativo no painel, foram submetidas ao estudo do exoma.

6.1. Análise dos resultados do painel

6.1.1. Análise de eficiência do painel na identificação dos defeitos do *SHOX*

O painel apresentou boa eficiência para a análise dos defeitos do gene *SHOX*. A estratégia de tentar capturar toda a região genômica do gene permitiu que toda a região codificadora do transcrito principal fosse sequenciada com cobertura satisfatória, com menos 40x em todos os pacientes. A captura das regiões regulatórias aconteceu conforme a previsão do *kit*, e apesar de algumas regiões não terem sido cobertas de forma ideal em sua totalidade, a análise dessas regiões foi informativa, principalmente naqueles pacientes com grandes deleções já conhecidas.

A análise das variantes, através da filtragem da planilha VCF, associada à análise de CNVs pelo *CONTRA* e pela inspeção visual da cobertura do gene no *IGV*, foi capaz de identificar todos os 15 defeitos do *SHOX* previamente identificados, sendo duas mutações de ponto e 13 CNVs.

Uma dessas 13 CNVs não foi identificada pela análise do *CONTRA*. Na duplicação da Paciente 12, o gráfico do programa mostrou uma tendência de duplicação, porém os *log-ratios* da planilha ficaram muito distantes do +1,0, conforme esperado para esse tipo de CNV, o que fez com que a informação fosse perdida no processo de filtragem. A sensibilidade do programa para identificação de CNVs, portanto, foi de 92% (12/13) nessa análise de eficiência. A análise visual do sequenciamento do *SHOX* através do *IGV*, por sua vez, foi capaz de identificar todas as alterações previamente conhecidas.

As grandes deleções envolvendo o *SHOX*, sua região regulatória e outros genes da *PAR1* foram identificadas com facilidade pelo *CONTRA*, que detectou

vários fragmentos com *log-ratios* próximos de -1,0 e valores de p significativos ao longo das regiões deletadas. Apesar disso, o programa não conseguiu diminuir o intervalo de deleção previsto em relação aos resultados prévios do MLPA, em função da baixa densidade de sondas do painel desenhadas nas regiões 5' e 3' do *SHOX*. Possivelmente a inclusão de regiões maiores e de outros genes vizinhos a ele na captura melhora a capacidade do painel em estimar o tamanho das deleções encontradas.

A análise visual do sequenciamento do *SHOX* possibilitou a identificação dos pontos de quebra de uma deleção intragênica previamente identificada em uma família com DLW (Paciente 11) (38). Conseguimos definir o tamanho exato do fragmento deletado em 11.722 pb e verificamos que a deleção da nossa família é diferente de uma deleção descrita, em 2008, por *Fukami* e colaboradores, apesar de envolver os mesmos exons (66). Eles descreveram uma deleção de 5.594 pb, envolvendo os exons 4, 5 e 6a, em uma outra família com DLW (66).

Em relação às deleções localizadas na região 3' do *SHOX*, o painel possibilitou a definição dos pontos de quebra de duas pacientes (Pacientes 13 e 14), que apresentavam exatamente a mesma deleção que foi descrita por *Benito-Sanz* e colaboradores em 2012 (16). Na ocasião, estudos funcionais demonstraram uma função regulatória dessa região em relação ao *SHOX*, apesar de uma penetrância reduzida das deleções em relação ao fenótipo dos pacientes. Em uma das nossas famílias a segregação não foi realizada por falta do DNA dos pais e na outra a deleção foi herdada da mãe da paciente, com estatura normal, e estava ausente no pai que é baixo. Sendo assim, a deleção não segregou com a baixa estatura dessa família. Achado que confirma a penetrância incompleta da variante.

O painel identificou duas novas deleções intragênicas no *SHOX* entre os 77 pacientes com BE isolada sem diagnóstico molecular: uma deleção envolvendo o exon 1, na Paciente 16, e uma deleção envolvendo parte do exon 3, na Paciente 17. Microdeleções envolvendo um ou mais exons do *SHOX* já foram descritas, porém são raras, correspondendo a cerca de 5% dos casos com defeitos no *SHOX* (7). A deleção específica do exon 1 já foi descrita outras duas vezes em relatos de caso sem avaliações funcionais (67, 68). Ambos os trabalhos identificaram deleções isoladas do exon 1 em pacientes com baixa estatura idiopática sem desproporção corpórea ou outros achados dismórficos significativos. No primeiro trabalho, a

deleção foi identificada por qPCR (67), enquanto no segundo, foi detectada por MLPA e confirmada por *array* de hibridação genômica comparativa (aCGH, de *comparative genomic hybridization array*) (68). Este último trabalho conseguiu delimitar o tamanho da deleção em 8,2 Kb. No nosso caso, não conseguimos identificar os pontos de quebra e delimitar o fragmento deletado. Os dois trabalhos especulam a provável consequência dessa deleção com base na presença de elementos importantes para a transcrição do *SHOX* na região do exon 1. Neste exon está localizado um dos dois promotores descritos do *SHOX* (30). Além disso, geralmente a região 5' não traduzida (UTR) do mRNA possui um papel essencial na regulação pós-transcricional, controlando a localização do mRNA, sua estabilidade e eficiência de tradução. Problemas nessa região poderiam modificar a eficiência do *splicing*, aumentar a degradação do mRNA e dificultar o transporte do mesmo para o citoplasma (69). Baseado nesses pontos, ambos concluem que a haploinsuficiência do exon 1 resultaria na haploinsuficiência do gene e, consequentemente na baixa estatura dos pacientes. Seguindo esse mesmo raciocínio, acreditamos que a deleção do exon 1 identificada na nossa paciente (Paciente 16), que segrega com a desproporção na família, possa ser a causa da sua baixa estatura.

Uma deleção parcial do *SHOX* envolvendo parte do exon 3, que codifica parte do domínio *homeobox* da proteína, foi identificada em uma paciente com baixa estatura desproporcional (Paciente 17). O pai e a irmã da paciente, apesar de não serem baixos, são desproporcionais. A irmã, inclusive, apresenta deformidade de *Madelung*, característica comum em pacientes com defeitos no *SHOX*. Essa deleção não foi identificada na análise inicial por MLPA, pois está localizada fora das regiões analisadas pelas sondas do *kit*. Ela também não foi identificada na análise do *CONTRA*. Apesar da falha do *CONTRA*, a deleção foi vista na análise visual do sequenciamento do *SHOX*, sendo assim, o estudo pelo o painel foi a única metodologia capaz de identificar a deleção dessa família.

Considerando as CNVs previamente conhecidas e as duas identificadas prospectivamente, temos 15 casos, sendo que dois (2/15; 13%) deles não foram identificados pelo *CONTRA* (Paciente 12 e Paciente 17). Essa sensibilidade de 87% reforça as limitações dos programas para identificação desse tipo de alteração a partir dos dados de sequenciamento de regiões alvo por SPLE.

6.1.2. Variantes identificadas em outros genes pelo painel

Além das duas deleções do *SHOX*, foram identificadas outras 13 variantes patogênicas ou provavelmente patogênicas em 12 pacientes do grupo de 77 pacientes com BE isolada sem diagnóstico molecular, sendo nove variantes que afetam fatores parácrinos da cartilagem de crescimento (3x*FGFR3*, 1x*GNAS*, 1x*NPR2* e 4x*IHH*), duas que afetam vias intracelulares, mais especificamente a via RAS/MAPK (1x*BRAF* e 1x*NF1*) e outras duas que afetam genes do eixo GH/IGF-1 (1x*IGF1R* e 1x*GH1*).

A seguir serão discutidos alguns aspectos em relação às variantes e aos genes nos quais elas foram identificadas.

6.1.2.1. O gene *FGFR3* e as variantes identificadas

A sinalização dos FGFs é fundamental para a proliferação e diferenciação dos condrócitos (6). A sinalização especificamente via *FGFR3* (receptor 3 do fator de crescimento do fibroblasto, *de fibroblast growth factor receptor 3*; MIM 134934; NM_000142.4) inibe a proliferação dos condrócitos através da ativação da JAK/STAT1 (70), diminui a produção de matriz extracelular, acelera a diferenciação hipertrófica e reduz o tamanho dos condrócitos hipertróficos (71, 72). Essas ações fazem do *FGFR3* um regulador negativo do crescimento.

Mutações inativadoras nesse gene em humanos estão associadas ao ganho de estatura independente do seu estado (73, 74). Mutações ativadoras prejudicam o crescimento, podendo resultar, entre outros fenótipos, em acondroplasia (MIM 100800), hipocondroplasia (MIM 146000) e displasia tanatofórica tipo I (MIM 187600) ou tipo II (MIM 187601), todas com padrão de herança autossômico dominante (75). A acondroplasia e a displasia tanatofórica apresentam quadros clínicos mais graves com importante baixa estatura de início pré-natal entre outros achados, enquanto na hipocondroplasia, causada por mutações menos deletérias, a baixa estatura tem início pós-natal (4). Essas mutações levam à ativação constitucional do receptor e o grau de ativação do mesmo correlaciona-se com a gravidade da apresentação clínica.

Nesse trabalho, nós identificamos três casos de baixa estatura com variantes raras, patogênicas ou provavelmente patogênicas, em heterozigose no *FGFR3*,

sendo dois deles com presença de desproporção corpórea. Uma das variantes foi identificada na Paciente 18 com baixa estatura importante e presença lordose lombar, que podem ser compatíveis com um quadro discreto de hipocondroplasia. Além disso, a variante, localizada no segundo domínio imunoglobulina *like*, onde já existem outras variantes patogênicas descritas (76, 77), segrega com a baixa estatura da família. A segunda variante identificada na Paciente 19, que também apresenta lordose lombar, é uma das variantes mais comuns associadas a essa displasia esquelética, a p.Asn540Lys. Sendo assim, apesar das duas pacientes não apresentarem um quadro clínico típico de hipocondroplasia, nós acreditamos que essas variantes podem ser a causa molecular da baixa estatura delas.

A última variante, p.Val553Leu, identificada na Paciente 20, também segregou com a baixa estatura na família. Temos então uma variante rara provavelmente patogênica no gene *FGFR3* em uma paciente com baixa estatura isolada sem desproporção corpórea. Esse achado se soma a um caso descrito em 2015 de uma variante patogênica (c.1150T>C; p.Met528Ile) identificada em uma paciente com baixa estatura proporcional, contribuindo para a ampliação do espectro fenotípico das variantes nesse gene (78). O estudo funcional *in vitro* dessa variante descrita demonstrou que ela foi capaz de promover a ativação do FGFR3, STAT e ERK/MAPK conforme esperado para as variantes causadoras de displasia esquelética nesse gene. Nossa variante, assim como a descrita por Kant e colaboradores, está localizada no domínio proteína quinase do receptor. Acreditamos, portanto, que ela possa justificar a baixa estatura da nossa família.

6.1.2.2. O gene *GNAS* e a variante identificada

O *GNAS* (*GNAS complex locus*; MIM 139320; NM_080425.3) é o gene que codifica a subunidade alfa da proteína Gs. Esse gene apresenta um padrão de expressão diferencial dos alelos nos diferentes tecidos. Mecanismos epigenéticos de *imprinting* controlam esse padrão, definindo o alelo que será expresso. Por esse motivo, as mutações nesse gene resultam em uma grande variabilidade fenotípica. Os efeitos das mutações dependem da origem do alelo mutado (se a mesma é materna ou paterna) (79).

Em alguns tecidos, como os rins, a expressão é exclusivamente materna. Mutações inativadoras em heterozigose no alelo materno, sejam elas herdadas ou *de novo*, resultam em resistência ao paratormônio (PTH) e estão associadas ao fenótipo do pseudohipoparatiroidismo (PHP; MIM 103580) (80). Nas subcategorias PHP1A e PHP1C temos a presença de osteodistrofia hereditária de *Albright* (AHO), que além da resistência ao PTH, é caracterizada por uma série de achados clínicos relacionados ao desenvolvimento ósseo, inclusive braquidactilia e baixa estatura (80, 81).

Mutações no alelo paterno, também em heterozigose, podem resultar em heteroplasia óssea progressiva (POH; MIM 166350), caracterizada por ossificações heterotrópicas que se expandem para músculos e tecidos conectivos, ou em pseudopseudohipoparatiroidismo (PPHP; MIM 612463), que apresenta um fenótipo semelhante à AHO, porém sem resistência ao PTH (81, 82).

Todas essas combinações de fenótipos tornam difícil a caracterização das doenças desse grande grupo que envolve a via de sinalização PTH/PTHrP. Um trabalho recente propôs uma nova classificação que tenta abranger o máximo de fenótipos possíveis, levando em consideração os defeitos genéticos que resultam nos diferentes quadros clínicos (81).

Nós identificamos uma variante patogênica *de novo* no *GNAS* em duas irmãs gêmeas univitelinas com baixa estatura desproporcional. Elas possuem algumas características compatíveis com o quadro clínico de pseudopseudohipoparatiroidismo, onde são descritas alterações cognitivas, metatarsos curtos, presença de ossificações subcutâneas e baixa estatura sem resistência ao PTH. Sendo assim, essa variante deve ser o defeito molecular que explica o fenótipo das duas.

6.1.2.3. O gene *NPR2* e a variante identificada

O *NPR2* (*natriuretic peptide receptor 2*; MIM 108961; NM_003995.3) é o gene que codifica o NPR-B, que é um dos receptores que mediam a ação dos peptídeos natriuréticos, principalmente do peptídeo natriurético tipo C (CNP), que é codificado pelo gene *NPPC* (*natriuretic peptide precursor C*; MIM 600296). O sistema CNP/NPR2 é um dos principais reguladores parácrinos da ossificação endocondral e

do crescimento. A ligação do NPR-B com o CNP desencadeia uma cascata de fosforilação intercelular responsável por mediar as ações biológicas do CNP, que entre outras coisas, leva à inibição da via MAPK (83) e antagoniza com a sinalização via FGFR3 (84).

Mutações inativadoras em homozigose no *NPR2* são associadas à displasia acromesomélica tipo *Maroteaux* (MIM 602875), uma displasia esquelética grave, caracterizada por uma baixa estatura pós-natal importante, de até menos 5 desvios padrão, que afeta os segmentos médios e distais dos membros (85). Mutações em heterozigose já foram descritas em associação com vários fenótipos, que vão desde condrodisplasias, como a condrodisplasia epifiseal tipo *Miura* (MIM 615923), até quadros de baixa estatura sem anormalidades esqueléticas (MIM 616255). Os estudos publicados nos últimos anos sugerem que 2% das crianças com BE idiopática podem apresentar mutações no *NPR2* (40, 86, 87). Em paralelo a isso, as mutações ativadoras (88, 89) ou a superexpressão do CNP (90), como era de se esperar, resultam em estatura elevada.

Diante disso, acreditamos que a variante identificada no nosso paciente é o defeito molecular responsável pela sua baixa estatura. Trata-se de uma variante rara, classificada como patogênica pelos critérios do ACMG/AMP, que pela predição *in silico* altera o *splicing* no exon 11 desse gene que possui 22 exons.

6.1.2.4. O gene *IHH* e as variantes identificadas

O *IHH* (*indian hedgehog*; MIM 600726; NM_002181.3) é um gene que codifica uma proteína, produzida pelos próprios condrócitos, envolvida na via IHH de sinalização de fatores parácrinos da cartilagem de crescimento (6). Essa proteína, juntamente com o sistema PTHrP, desempenha um importante papel na cartilagem, alimentando um *feedback* negativo, que coordena a proliferação e hipertrofia dos condrócitos. Agindo diretamente sobre os mesmos, ela estimula sua proliferação, e estimulando a via do PTHrP, ela mantém os condrócitos na zona proliferativa inibindo sua diferenciação (6).

Mutações no gene *IHH* durante muito tempo foram associadas com a displasia acrocapitofemoral (MIM 607778), com padrão de herança autossômico

recessivo, e a braquidactilia tipo A1 (MIM 112500), com padrão de autossômico dominante, ambas com baixa estatura entre os achados clínicos.

Nesse trabalho, tivemos quatro casos de baixa estatura (dois deles com desproporção corpórea) com variantes raras com potencial patogênico em heterozigose no *IHH*. Nos dois achados iniciais, dos Pacientes 22 e 23 (p.Glu58Lys e p.Val178Leu), que apresentam baixa estatura desproporcional, as variantes estão localizadas no domínio de sinalização N-terminal da proteína, podendo impactar sua estrutura e função biológica, alterando a interação com alguns ligantes.

Esse achado de variantes no *IHH* em pacientes sem sinais de displasia esquelética nos chamou a atenção e por esse motivo resolvemos participar de um estudo em colaboração com um grupo espanhol que avaliou o gene *IHH* em um total de 290 pacientes, sendo 111 deles com BE idiopática, 19 nascidos pequenos para idade gestacional que continuaram baixos, 60 com braquidactilia inespecífica e 100 pacientes com displasia esquelética leve (65). Esse estudo identificou 10 variantes em heterozigose com potencial patogênico nesse gene. Os pacientes portadores das mutações apresentavam, de forma geral, baixa estatura levemente desproporcional com um achado frequente de encurtamento da falange média do 5º dedo. Nenhum deles apresentava as características clássicas da braquidactilia tipo A1, previamente associada aos defeitos em heterozigose do *IHH*.

Os outros dois achados, dos Pacientes 24 e 25 (p.Val229Met e p.Val317Met), que não apresentam desproporção corpórea, foram posteriores a essa colaboração. As variantes estão localizadas no domínio C-terminal, região onde seria mais difícil avaliar seu efeito funcional (65). A variante p.Val317Met, da Paciente 25, foi uma das 10 variantes identificadas no trabalho em colaboração com o grupo espanhol. Ela foi identificada em um paciente do grupo deles, com altura no limite inferior da normalidade (Z da altura = -1,8), sem informações a respeito das proporções corpóreas, porém com alterações no raio X de mão e punho. Esse paciente apresentava encurtamento grave da falange média do 5º dedo e um encurtamento leve falange média do 2º dedo.

Os nossos quatro pacientes não apresentam achados compatíveis com as displasias esqueléticas previamente associadas com as mutações no *IHH*, porém acreditamos que as variantes identificadas são a causa molecular da baixa estatura deles. Com esses achados nós colaboramos para a ampliação do espectro

fenotípico associado às mutações em heterozigose no gene *IHH* (65). Agora, além da braquidactilia tipo A1, elas podem resultar em quadros de baixa estatura sem achados esqueléticos específicos.

6.1.2.5. O gene *BRAF* e a variante identificada

O gene *BRAF* (*B-Raf proto-oncogene, serine/threonine kinase*; MIM 164757; NM_004333.4) codifica uma proteína da família RAF de proteínas quinases serina/treonina, que participa da regulação da via de sinalização MAPK/ERK, controlando, entre outras coisas, o crescimento e diferenciação celular (91).

Mutações somáticas localizadas no domínio proteína quinase estão associadas a melanomas malignos e outros tipos de neoplasias (92), enquanto que as mutações germinativas do tipo *missense* localizadas ao longo da proteína estão associadas à síndrome cardiofaciocutâneo (MIM 115150), à síndrome de *Noonan* (MIM 613706) e síndrome de *Leopard* (MIM 613707) (93). Essas síndromes apresentam fenótipos que se sobrepõem e são caracterizadas, de forma geral, por baixa estatura, aspectos faciais dismórficos, defeitos cognitivos variáveis e presença de alterações cardíacas.

Nós identificamos uma variante *missense* com potencial patogênico em um paciente que apresenta baixa estatura sem outros achados clínicos significativos, além de acentuação da cifose torácica. Na literatura tem sido descritos pacientes com rasopatias sem sinais clínicos típicos das síndromes (5). O nosso paciente apresentou níveis IGF-1 baixos e ausência de resposta ao tratamento com rhGH, características muitas vezes observadas em pacientes com síndrome de *Noonan*. A variante p.Val238Ile, herdada da mãe também baixa, está localizada em uma importante região de dedo de zinco da proteína. No *ClinVar* existem descrições de outras variantes nessa mesma região associadas às síndromes de *Noonan* e cardiofaciocutâneo. Por todos esses achados, acreditamos que essa variante por ser a responsável pelo quadro de baixa estatura dele.

6.1.2.6. O gene *NF1* e a variante identificada

O *NF1* (*Neurofibromin 1*; MIM 613113; NM_000267.3) é um gene que codifica a neurofibromina, uma proteína citoplasmática ubíqua predominantemente expressa nos neurônios, células *Schwann*, oligodendrócitos e leucócitos. Trata-se de uma molécula com muitos domínios e a capacidade de regular vários processos celulares, incluindo a via RAS-AMP cíclico, MAPK/ERK, adenilil ciclase, etc (94). Mutações em heterozigose nesse gene, que é considerado um supressor tumoral, resultam na neurofibromatose tipo 1 (MIM 162200). Os pacientes com essa doença apresentam uma grande variabilidade fenotípica e o diagnóstico clínico é feito com a identificação de alguns sinais: manchas café com leite, sardas axilares ou inguinais, neurofibromas cutâneos, nódulos de Lisch, lesão esquelética específica (displasia da asa do esfenóide, redução da espessura do córtex dos ossos longos, pseudo-artrite) e glioma óptico. Grande parte dos pacientes apresenta um quadro clínico leve, porém entre as principais complicações da doença está o desenvolvimento de tumores malignos.

Pacientes com neurofibromatose podem apresentar alguns sinais clínicos da síndrome de *Noonan*, como baixa estatura, ptose, hipoplasia da face média, pescoço alado, dificuldade de aprendizado e fraqueza muscular, caracterizando a síndrome neurofibromatose-*Noonan* (95).

Camundongos com nocaute para o *NF1* em células mesenquimais indiferenciadas de membros em desenvolvimento apresentam arqueamento da tíbia, crescimento diminuído, vascularização anormal dos tecidos esqueléticos, fusão das articulações do quadril, entre outras anormalidades. A redução do crescimento desses animais é consequência da diminuição da proliferação e diferenciação dos condrócitos.

Identificamos uma variante patogênica no gene *NF1* em um paciente com características inespecíficas. Além das manchas café com leite, ele apresenta *pectus excavatum* e um pavilhão auricular dismórfico, que de certa forma poderia compor um quadro clínico da síndrome neurofibromatose-*Noonan*. A variante p.Met1149Ile está localizada em uma região *hotspot* para mutações patogênicas. Na literatura existem dois trabalhos que descrevem mutações nesse mesmo códon, o 1149 (96, 97). O primeiro deles identificou exatamente a mesma variante encontrada

no nosso paciente (c. 3447G>A; p.Met1149Ile) em um paciente dentro de um grupo de 169 pacientes com suspeita clínica de neurofibromatose tipo 1 (96). A variante, localizada em um aminoácido bem conservado, era *de novo* e, na época, não foi identificada em um grupo de 100 indivíduos normais, o que fala a favor do seu potencial patogênico. O trabalho, entretanto, não possui informações detalhadas a respeito do quadro clínico do paciente que apresentava a variante. No *ClinVar* existem três descrições dessa variante como patogênica ou provavelmente patogênica. O segundo trabalho é um relato de caso que descreve um paciente com neurofibromatose associada com uma deficiência do complexo mitocondrial I (97). Esse paciente, com sinais clássicos de neurofibromatose, apresentava baixa estatura, além de atraso no desenvolvimento, e possuía a variante c.3445A>G (p.Met1149Val), que envolve o mesmo aminoácido da nossa. Com base nesses achados, acreditamos que essa variante deva ser responsável pelo quadro clínico do nosso paciente. Nós perdemos o acompanhamento dele, e não sabemos se ele evoluiu com alguma das complicações associadas às mutações no gene *NF1*.

6.1.2.7. O gene *IGF1R* e a variante identificada

O gene *IGF1R* (*insulin-like growth factor I receptor*, MIM 147370; NM_000875.4) codifica o receptor 1 do fator de crescimento insulino-símile (IGF-1R), o receptor 1 do IGF-1. A ligação entre eles promove a ativação da atividade tirosina-quinase intrínseca do receptor, que acaba ativando algumas cascatas de sinalização que resultam em efeitos proliferativos e metabólicos dos IGFs. Esse sistema é o principal regulador endócrino do crescimento pré-natal, que é pouco influenciado pelo efeito do GH (98). Sendo assim, defeitos tanto no *IGF1* como no *IGF1R* podem resultar em retardo do crescimento intrauterino com baixa estatura de início pré-natal.

Mutações nesse gene resultam em resistência ao IGF-1 (MIM 270450) que apresenta padrão de herança normalmente autossômico dominante com fenótipo variável. Essa variabilidade depende da atividade remanescente da sinalização do IGF-1 (99). Já foram descritas mutações dos tipos *nonsense* e *missense* em heterozigose simples, em heterozigose composta ou deleções (99, 100). Essas mutações diminuem a autofosforilação do receptor, reduzindo a ativação das vias de

sinalização. Os portadores das mutações costumam apresentar principalmente baixa estatura de início pré-natal, atraso na idade óssea e microcefalia. Alguns achados dismórficos também podem estar presentes (99). Os exames laboratoriais em muitos casos não ajudam porque os resultados variam tanto quanto as alterações fenotípicas (99). A dosagem de IGF-1, nem sempre está aumentada, como se esperaria. O IGFBP-3 costuma ficar dentro dos limites da normalidade. O teste de estímulo de GH quando aumentado, ajuda a descartar a deficiência do mesmo.

Nosso paciente que possui a variante no *IGF1R* apresentava desproporção corpórea, um achado não muito comum em pacientes com defeitos nesse gene. Ele não apresentava nenhum sinal importante, além de atraso na idade óssea e baixo peso ao nascimento. O GH pós-estímulo foi normal e ele apresentou uma única dosagem de IGF-1 elevada, o que poderia ter levantado suspeitas de um quadro de insensibilidade ao IGF-1. Porém, seu quadro clínico inespecífico, tornava difícil essa suspeita antes do estudo molecular.

6.2. Análise dos resultados do exoma

O estudo do exoma foi realizado em nove famílias selecionadas que apresentaram resultado negativo no painel. Essa análise não identificou nenhuma variante específica que pudesse esclarecer a causa da baixa estatura das mesmas. Sendo assim, a abordagem desses pacientes pelo exoma não aumentou a taxa de diagnóstico nesse grupo. Isso pode ter acontecido por algumas razões. O padrão de herança esperado em grande parte das famílias era autossômico dominante e o fato delas serem pequenas (trios com membros de apenas duas gerações) contribuíram para um grande número de variantes raras ao final das filtragens, dificultando a interpretação dos dados.

Além disso, podemos pensar em outras possibilidades. A altura é um traço poligênico e em alguns casos várias variantes, raras ou não, podem se associar resultando em um quadro de baixa estatura, dessa forma nunca chegaríamos a uma variante específica. Variantes patogênicas também podem estar presentes em genes ainda não associados ao desenvolvimento ósseo e ao crescimento. Por esse motivo, os dados do exoma dessas famílias devem ser reavaliados futuramente.

6.3. Considerações gerais

O sequenciamento do painel customizado desenvolvido gerou métricas satisfatórias. Ele apresentou boa cobertura no sequenciamento dos principais genes já associados à baixa estatura isolada (*ACAN*, *FGFR3*, *GH1*, *GHR*, *GHSR*, *IGFALS*, *IGF1*, *IGF1R*, *IHH*, *PAPPA2*, *NPPC*, *NPR2*, *SHOX*, *STAT5B*) e à fisiologia da cartilagem de crescimento. Com exceção do gene *ACAN*, que apresenta uma região repetitiva no exon 12, que dificulta a captura e o alinhamento das sequências em aproximadamente 2,3% da sua região alvo, os demais genes de interesse apresentaram cobertura maior que 10x em toda a região codificadora dos seus transcritos principais em cerca de 100% dos nossos pacientes.

Utilizando uma associação de análises, que envolveu a filtragem das planilhas VCF, a análise do *CONTRA* e a análise visual da cobertura do sequenciamento do gene *SHOX*, o painel foi capaz de identificar todos os defeitos previamente conhecidos, se mostrando uma boa ferramenta para a investigação desse gene.

A sensibilidade final do *CONTRA* para a detecção de CNVs foi de 87%. A sensibilidade descrita no artigo que descreveu o programa é de 68,0% para deleções de 50 a 200 pb e de 96,4% para CNVs envolvendo exons inteiros (60). A deleção que não foi detectada possui 679 pb, mas não envolve um exon inteiro. Já a duplicação envolve três exons. Esse resultado deixa claro a necessidade de se testar novos programas para a análise de CNVs com abordagens diferentes e, se necessário, desenvolver ferramentas de bioinformática mais acuradas para essa aplicação, em uma tentativa de melhorar essa taxa de sensibilidade. Enquanto isso, ressaltamos a importância da análise visual do sequenciamento do *SHOX*, principalmente nos casos com suspeita de alterações nesse gene. Ainda em relação a essa análise visual, graças a ela foi possível a definição dos pontos de quebra em quatro de 15 (4/15; 27%) CNVs.

A análise de variantes de pontos e pequenas *indels* no grupo de 77 pacientes sem diagnóstico identificou variantes candidatas em 12 pacientes, que somados aos outros dois com deleções identificadas no *SHOX*, resultam em 14 pacientes com resultado positivo. Onze (11/14; 78%) pacientes apresentaram variantes em genes, de alguma forma, associados à fisiologia da cartilagem de crescimento. Achado já esperado devido a uma grande quantidade de pacientes com desproporção

corpórea na nossa casuística (30/77; 39%). Nesse grupo de 11 de pacientes com variantes em genes relacionados à cartilagem, 64% (7/11) deles eram desproporcionais.

A partir desses resultados foi possível estabelecer o diagnóstico molecular em 18% (14/77) dos pacientes com baixa estatura isolada da nossa casuística. Essa taxa de positividade não aumentou mesmo após a realização do estudo do exoma em nove (9/63; 14%) famílias com resultado negativo no painel. É importante ressaltar que nesse trabalho, 51% dos pacientes já havia sido submetido a uma análise molecular prévia de pelo menos um dos três principais genes associados à baixa estatura isolada (*SHOX*, *NPR2* e *FGFR3*). Esse fato pode ter diminuído a nossa taxa de positividade.

Os estudos prévios que utilizaram metodologias de sequenciamento paralelo em larga escala para investigar a causa de pacientes com baixa estatura apresentam taxas de positividade muito diferentes, variando de 2 a 46% (101-105). Essa grande variação pode ser consequência das diferenças nas metodologias aplicadas e nos grupos de pacientes avaliados nesses trabalhos, tanto em relação ao número de indivíduos estudados como ao fenótipo que eles apresentavam. Um trabalho realizado recentemente pelo nosso grupo avaliou 55 pacientes com baixa estatura isolada nascidos pequenos para a idade gestacional, por SPLE, sendo 39 deles por um painel customizado com genes relacionados ao crescimento, e os demais por exoma (106). A taxa de positividade nesse grupo de pacientes, cuja maioria foi submetida a um estudo por painel, foi de 15%, que é bem próxima da nossa. O trabalho também identificou uma predominância de variantes patogênicas nos genes relacionados à cartilagem de crescimento.

Os trabalhos de SPLE tem modificado paradigmas em relação a etiologia genética de algumas doenças, identificando mutações em alguns genes em pacientes com fenótipos diferentes do esperado. Por esse motivo, acreditamos que a abordagem proposta nesse projeto, que envolve o sequenciamento simultâneo de vários genes e que ainda permite a identificação de CNVs, principalmente no gene *SHOX*, é uma abordagem interessante para a investigação inicial de pacientes com baixa estatura isolada sem achados esqueléticos específicos. Nesse grupo apenas o estudo molecular é capaz de estabelecer o diagnóstico, e a ausência de

características clínicas específicas dificulta a abordagem molecular por gene candidato.

A identificação do gene responsável pelo fenótipo da baixa estatura do paciente fecha o diagnóstico daquela família, abrindo algumas possibilidades em relação ao tratamento e ao aconselhamento genético da mesma. Além disso, em alguns casos, esse diagnóstico contribui para a ampliação do nosso conhecimento em relação às complexas vias que regulam o crescimento.

7. CONCLUSÕES

- O painel customizado desenvolvido se mostrou eficiente para identificação dos defeitos do gene *SHOX*.
- O painel conseguiu estabelecer o diagnóstico etiológico em 18% dos nossos casos de baixa estatura isolada, identificando principalmente variantes em genes relacionados à fisiologia da cartilagem de crescimento.
- A abordagem por estudo do exoma não aumentou a taxa de diagnóstico nesse grupo de pacientes.

8. REFERÊNCIAS

1. Guo MH, Hirschhorn JN, Dauber A. Insights and Implications of Genome-Wide Association Studies of Height. *J Clin Endocrinol Metab*. 2018;103(9):3155-68.
2. Lango Allen H, Estrada K, Lettre G, Berndt SI, Weedon MN, Rivadeneira F, et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*. 2010;467(7317):832-8.
3. Durand C, Rappold GA. Height matters-from monogenic disorders to normal variation. *Nat Rev Endocrinol*. 2013;9(3):171-7.
4. Baron J, Säwendahl L, De Luca F, Dauber A, Phillip M, Wit JM, et al. Short and tall stature: a new paradigm emerges. *Nat Rev Endocrinol*. 2015;11(12):735-46.
5. Wit JM, Oostdijk W, Losekoot M, van Duyvenvoorde HA, Ruivenkamp CA, Kant SG. MECHANISMS IN ENDOCRINOLOGY: Novel genetic causes of short stature. *Eur J Endocrinol*. 2016;174(4):R145-73.
6. Kronenberg HM. Developmental regulation of the growth plate. *Nature*. 2003;423(6937):332-6.
7. Marchini A, Ogata T, Rappold GA. A Track Record on SHOX: From Basic Research to Complex Models and Therapy. *Endocr Rev*. 2016;37(4):417-48.
8. Rao E, Weiss B, Fukami M, Rump A, Niesler B, Mertz A, et al. Pseudoautosomal deletions encompassing a novel homeobox gene cause growth failure in idiopathic short stature and Turner syndrome. *Nat Genet*. 1997;16(1):54-63.
9. Ellison JW, Wardak Z, Young MF, Gehron Robey P, Laig-Webster M, Chiong W. PHOG, a candidate gene for involvement in the short stature of Turner syndrome. *Hum Mol Genet*. 1997;6(8):1341-7.
10. Munns CJ, Haase HR, Crowther LM, Hayes MT, Blaschke R, Rappold G, et al. Expression of SHOX in human fetal and childhood growth plate. *J Clin Endocrinol Metab*. 2004;89(8):4130-5.
11. Clement-Jones M, Schiller S, Rao E, Blaschke RJ, Zuniga A, Zeller R, et al. The short stature homeobox gene SHOX is involved in skeletal abnormalities in Turner syndrome. *Hum Mol Genet*. 2000;9(5):695-702.
12. Thomas NS, Maloney V, Bass P, Mulik V, Wellesley D, Castle B. SHOX mutations in a family and a fetus with Langer mesomelic dwarfism. *Am J Med Genet A*. 2004;128A(2):179-84.
13. Fukami M, Kato F, Tajima T, Yokoya S, Ogata T. Transactivation function of an approximately 800-bp evolutionarily conserved sequence at the SHOX 3' region: implication for the downstream enhancer. *Am J Hum Genet*. 2006;78(1):167-70.

14. Sabherwal N, Bangs F, Röth R, Weiss B, Jantz K, Tiecke E, et al. Long-range conserved non-coding SHOX sequences regulate expression in developing chicken limb and are associated with short stature phenotypes in human patients. *Hum Mol Genet.* 2007;16(2):210-22.
15. Durand C, Bangs F, Signolet J, Decker E, Tickle C, Rappold G. Enhancer elements upstream of the SHOX gene are active in the developing limb. *Eur J Hum Genet.* 2010;18(5):527-32.
16. Benito-Sanz S, Royo JL, Barroso E, Paumard-Hernández B, Barreda-Bonis AC, Liu P, et al. Identification of the first recurrent PAR1 deletion in Léri-Weill dyschondrosteosis and idiopathic short stature reveals the presence of a novel SHOX enhancer. *J Med Genet.* 2012;49(7):442-50.
17. Rao E, Blaschke RJ, Marchini A, Niesler B, Burnett M, Rappold GA. The Leri-Weill and Turner syndrome homeobox gene SHOX encodes a cell-type specific transcriptional activator. *Hum Mol Genet.* 2001;10(26):3083-91.
18. Marchini A, Häcker B, Marttila T, Hesse V, Emons J, Weiss B, et al. BNP is a transcriptional target of the short stature homeobox gene SHOX. *Hum Mol Genet.* 2007;16(24):3081-7.
19. Aza-Carmona M, Shears DJ, Yuste-Checa P, Barca-Tierno V, Hisado-Oliva A, Belinchón A, et al. SHOX interacts with the chondrogenic transcription factors SOX5 and SOX6 to activate the aggrecan enhancer. *Hum Mol Genet.* 2011;20(8):1547-59.
20. Decker E, Durand C, Bender S, Rödelisperger C, Glaser A, Hecht J, et al. FGFR3 is a target of the homeobox transcription factor SHOX in limb development. *Hum Mol Genet.* 2011;20(8):1524-35.
21. Durand C, Decker E, Roeth R, Schneider KU, Rappold G. The homeobox transcription factor HOXA9 is a regulator of SHOX in U2OS cells and chicken micromass cultures. *PLoS One.* 2012;7(9):e45369.
22. Iughetti L, Capone L, Elsedfy H, Bertorelli R, Predieri B, Bruzzi P, et al. Unexpected phenotype in a boy with trisomy of the SHOX gene. *J Pediatr Endocrinol Metab.* 2010;23(1-2):159-69.
23. Langer LO. Dyschondrosteosis, a hereditary bone dysplasia with characteristic roentgenographic features. *Am J Roentgenol Radium Ther Nucl Med.* 1965;95:178-88.
24. Grigeliuniene G, Eklöf O, Ivarsson SA, Westphal O, Neumeyer L, Kedra D, et al. Mutations in short stature homeobox containing gene (SHOX) in dyschondrosteosis but not in hypochondroplasia. *Hum Genet.* 2000;107(2):145-9.

25. Schiller S, Spranger S, Schechinger B, Fukami M, Merker S, Drop SL, et al. Phenotypic variation and genetic heterogeneity in Léri-Weill syndrome. *Eur J Hum Genet.* 2000;8(1):54-62.
26. Grigelioniene G, Schoumans J, Neumeyer L, Ivarsson A, Eklöf O, Enkvist O, et al. Analysis of short stature homeobox-containing gene (SHOX) and auxological phenotype in dyschondrosteosis and isolated Madelung deformity. *Hum Genet.* 2001;109(5):551-8.
27. Ross JL, Scott C, Marttila P, Kowal K, Nass A, Papenhausen P, et al. Phenotypes Associated with SHOX Deficiency. *J Clin Endocrinol Metab.* 2001;86(12):5674-80.
28. Cormier-Daire V, Huber C, Munnich A. Allelic and nonallelic heterogeneity in dyschondrosteosis (Leri-Weill syndrome). *Am J Med Genet.* 2001;106(4):272-4.
29. Flanagan SF, Munns CF, Hayes M, Williams B, Berry M, Vickers D, et al. Prevalence of mutations in the short stature homeobox containing gene (SHOX) in Madelung deformity of childhood. *J Med Genet.* 2002;39(10):758-63.
30. Blaschke RJ, Töpfer C, Marchini A, Steinbeisser H, Janssen JW, Rappold GA. Transcriptional and translational regulation of the Leri-Weill and Turner syndrome homeobox gene SHOX. *J Biol Chem.* 2003;278(48):47820-6.
31. Jorge AA, Souza SC, Nishi MY, Billerbeck AE, Libório DC, Kim CA, et al. SHOX mutations in idiopathic short stature and Leri-Weill dyschondrosteosis: frequency and phenotypic variability. *Clin Endocrinol (Oxf).* 2007;66(1):130-5.
32. Rappold GA, Fukami M, Niesler B, Schiller S, Zumkeller W, Bettendorf M, et al. Deletions of the homeobox gene SHOX (short stature homeobox) are an important cause of growth failure in children with short stature. *J Clin Endocrinol Metab.* 2002;87(3):1402-6.
33. Binder G, Ranke MB, Martin DD. Auxology is a valuable instrument for the clinical diagnosis of SHOX haploinsufficiency in school-age children with unexplained short stature. *J Clin Endocrinol Metab.* 2003;88(10):4891-6.
34. Morizio E, Stuppia L, Gatta V, Fantasia D, Guanciali Franchi P, Rinaldi MM, et al. Deletion of the SHOX gene in patients with short stature of unknown cause. *Am J Med Genet A.* 2003;119A(3):293-6.
35. Huber C, Rosilio M, Munnich A, Cormier-Daire V, Module FSG. High incidence of SHOX anomalies in individuals with short stature. *J Med Genet.* 2006;43(9):735-9.
36. Rappold G, Blum WF, Shavrikova EP, Crowe BJ, Roeth R, Quigley CA, et al. Genotypes and phenotypes in children with short stature: clinical indicators of SHOX haploinsufficiency. *J Med Genet.* 2007;44(5):306-13.

37. Malaquias AC, Scalco RC, Fontenele EG, Costalonga EF, Baldin AD, Braz AF, et al. The sitting height/height ratio for age in healthy and short individuals and its potential role in selecting short children for SHOX analysis. *Horm Res Paediatr*. 2013;80(6):449-56.
38. Funari MF, Jorge AA, Pinto EM, Arnhold IJ, Mendonca BB, Nishi MY. Cryptic intragenic deletion of the SHOX gene in a family with Léri-Weill dyschondrosteosis detected by Multiplex Ligation-Dependent Probe Amplification (MLPA). *Arq Bras Endocrinol Metabol*. 2008;52(8):1382-7.
39. Liu Z, Lavine KJ, Hung IH, Ornitz DM. FGF18 is required for early chondrocyte proliferation, hypertrophy and vascular invasion of the growth plate. *Dev Biol*. 2007;302(1):80-91.
40. Vasques GA, Amano N, Docko AJ, Funari MF, Quedas EP, Nishi MY, et al. Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature in patients initially classified as idiopathic short stature. *J Clin Endocrinol Metab*. 2013;98(10):E1636-44.
41. Nilsson O, Guo MH, Dunbar N, Popovic J, Flynn D, Jacobsen C, et al. Short stature, accelerated bone maturation, and early growth cessation due to heterozygous aggrecan mutations. *J Clin Endocrinol Metab*. 2014;99(8):E1510-8.
42. Lui JC, Nilsson O, Chan Y, Palmer CD, Andrade AC, Hirschhorn JN, et al. Synthesizing genome-wide association studies and expression microarray reveals novel genes that act in the human growth plate to modulate height. *Hum Mol Genet*. 2012;21(23):5193-201.
43. Scalco RC, Melo SS, Pugliese-Pires PN, Funari MF, Nishi MY, Arnhold IJ, et al. Effectiveness of the combined recombinant human growth hormone and gonadotropin-releasing hormone analog therapy in pubertal patients with short stature due to SHOX deficiency. *J Clin Endocrinol Metab*. 2010;95(1):328-32.
44. Funari MF, Jorge AA, Souza SC, Billerbeck AE, Arnhold IJ, Mendonca BB, et al. Usefulness of MLPA in the detection of SHOX deletions. *Eur J Med Genet*. 2010;53(5):234-8.
45. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bembien LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*. 2005;437(7057):376-80.
46. Shendure J, Porreca GJ, Reppas NB, Lin X, McCutcheon JP, Rosenbaum AM, et al. Accurate multiplex polony sequencing of an evolved bacterial genome. *Science*. 2005;309(5741):1728-32.
47. Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. *Hum Mol Genet*. 2014;23(22):5866-78.

48. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-7.
49. Consortium IHGS. Finishing the euchromatic sequence of the human genome. *Nature*. 2004;431(7011):931-45.
50. Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) 2017 [Available from: www.genome.gov/sequencingcostsdata].
51. Tamura N, Doolittle LK, Hammer RE, Shelton JM, Richardson JA, Garbers DL. Critical roles of the guanylyl cyclase B receptor in endochondral ossification and development of female reproductive organs. *Proc Natl Acad Sci U S A*. 2004;101(49):17300-5.
52. Olney RC. C-type natriuretic peptide in growth: a new paradigm. *Growth Horm IGF Res*. 2006;16 Suppl A:S6-14.
53. Oostdijk W, Grote FK, de Muinck Keizer-Schrama SM, Wit JM. Diagnostic approach in children with short stature. *Horm Res*. 2009;72(4):206-17.
54. Naslavsky MS, Yamamoto GL, de Almeida TF, Ezquina SAM, Sunaga DY, Pho N, et al. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Hum Mutat*. 2017;38(7):751-63.
55. Miller SA, Dykes DD, Polesky HF. A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Res*. 1988;16(3):1215.
56. Sambrook J, Fritsch EF, Maniatis T. *Molecular cloning: A laboratory manual*. 2nd ed. New York: Cold Spring Harbor; 1989.
57. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet*. 2010;11(1):31-46.
58. Adey A, Morrison HG, Asan, Xun X, Kitzman JO, Turner EH, et al. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density in vitro transposition. *Genome Biol*. 2010;11(12):R119.
59. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
60. Li J, Lupat R, Amarasinghe KC, Thompson ER, Doyle MA, Ryland GL, et al. CONTRA: copy number analysis for targeted resequencing. *Bioinformatics*. 2012;28(10):1307-13.
61. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus

- recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med*. 2015;17(5):405-24.
62. Li Q, Wang K. InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines. *Am J Hum Genet*. 2017;100(2):267-80.
63. Bellus GA, McIntosh I, Smith EA, Aylsworth AS, Kaitila I, Horton WA, et al. A recurrent mutation in the tyrosine kinase domain of fibroblast growth factor receptor 3 causes hypochondroplasia. *Nat Genet*. 1995;10(3):357-9.
64. Xue Y, Sun A, Mekikian PB, Martin J, Rimo DL, Lachman RS, et al. FGFR3 mutation frequency in 324 cases from the International Skeletal Dysplasia Registry. *Mol Genet Genomic Med*. 2014;2(6):497-503.
65. Vasques GA, Funari MFA, Ferreira FM, Aza-Carmona M, Sentchordi-Montané L, Barraza-García J, et al. IHH Gene Mutations Causing Short Stature With Nonspecific Skeletal Abnormalities and Response to Growth Hormone Therapy. *J Clin Endocrinol Metab*. 2018;103(2):604-14.
66. Fukami M, Dateki S, Kato F, Hasegawa Y, Mochizuki H, Horikawa R, et al. Identification and characterization of cryptic SHOX intragenic deletions in three Japanese patients with Léri-Weill dyschondrosteosis. *J Hum Genet*. 2008;53(5):454-9.
67. Tan YM, Loke KY. Isolated haploinsufficiency of exon 1 of the SHOX gene in a patient with idiopathic short stature. *J Clin Pathol*. 2006;59(7):773-4.
68. Sandoval GT, Jaimes GC, Barrios MC, Cespedes C, Velasco HM. SHOX gene and conserved noncoding element deletions/duplications in Colombian patients with idiopathic short stature. *Mol Genet Genomic Med*. 2014;2(2):95-102.
69. Pesole G, Mignone F, Gissi C, Grillo G, Licciulli F, Liuni S. Structural and functional features of eukaryotic mRNA untranslated regions. *Gene*. 2001;276(1-2):73-81.
70. Sahni M, Ambrosetti DC, Mansukhani A, Gertner R, Levy D, Basilico C. FGF signaling inhibits chondrocyte proliferation and regulates bone development through the STAT-1 pathway. *Genes Dev*. 1999;13(11):1361-6.
71. Mancilla EE, De Luca F, Uyeda JA, Czerwiec FS, Baron J. Effects of fibroblast growth factor-2 on longitudinal bone growth. *Endocrinology*. 1998;139(6):2900-4.
72. Minina E, Kreschel C, Naski MC, Ornitz DM, Vortkamp A. Interaction of FGF, Ihh/Pthlh, and BMP signaling integrates chondrocyte proliferation and hypertrophic differentiation. *Dev Cell*. 2002;3(3):439-49.

73. Toydemir RM, Brassington AE, Bayrak-Toydemir P, Krakowiak PA, Jorde LB, Whitby FG, et al. A novel mutation in FGFR3 causes camptodactyly, tall stature, and hearing loss (CATSHL) syndrome. *Am J Hum Genet.* 2006;79(5):935-41.
74. Makrythanasis P, Temtamy S, Aglan MS, Otaify GA, Hamamy H, Antonarakis SE. A novel homozygous mutation in FGFR3 causes tall stature, severe lateral tibial deviation, scoliosis, hearing impairment, camptodactyly, and arachnodactyly. *Hum Mutat.* 2014;35(8):959-63.
75. Vajo Z, Francomano CA, Wilkin DJ. The molecular and genetic basis of fibroblast growth factor receptor 3 disorders: the achondroplasia family of skeletal dysplasias, Muenke craniosynostosis, and Crouzon syndrome with acanthosis nigricans. *Endocr Rev.* 2000;21(1):23-39.
76. Tavormina PL, Shiang R, Thompson LM, Zhu YZ, Wilkin DJ, Lachman RS, et al. Thanatophoric dysplasia (types I and II) caused by distinct mutations in fibroblast growth factor receptor 3. *Nat Genet.* 1995;9(3):321-8.
77. Heuertz S, Le Merrer M, Zabel B, Wright M, Legeai-Mallet L, Cormier-Daire V, et al. Novel FGFR3 mutations creating cysteine residues in the extracellular domain of the receptor cause achondroplasia or severe forms of hypochondroplasia. *Eur J Hum Genet.* 2006;14(12):1240-7.
78. Kant SG, Cervenкова I, Balek L, Trantirek L, Santen GW, de Vries MC, et al. A novel variant of FGFR3 causes proportionate short stature. *Eur J Endocrinol.* 2015;172(6):763-70.
79. Hayward BE, Bonthron DT. An imprinted antisense transcript at the human GNAS1 locus. *Hum Mol Genet.* 2000;9(5):835-41.
80. Albright F, Burnett CH, Smith PH, Parson W. Pseudohypoparathyroidism: An example of "Seabright-Bantam" Syndrome. *Endocrinology.* 1942;30:922-32.
81. Thiele S, Mantovani G, Barlier A, Boldrin V, Bordogna P, De Sanctis L, et al. From pseudohypoparathyroidism to inactivating PTH/PTHrP signalling disorder (iPPSD), a novel classification proposed by the EuroPHP network. *Eur J Endocrinol.* 2016;175(6):P1-P17.
82. Albright F, Forbes AP, Henneman PH. Pseudo-pseudohypoparathyroidism. *Trans Assoc Am Physicians.* 1952;65:337-50.
83. Teixeira CC, Agoston H, Beier F. Nitric oxide, C-type natriuretic peptide and cGMP as regulators of endochondral ossification. *Dev Biol.* 2008;319(2):171-8.
84. Krejci P, Masri B, Fontaine V, Mekikian PB, Weis M, Prats H, et al. Interaction of fibroblast growth factor and C-natriuretic peptide signaling in regulation of chondrocyte proliferation and extracellular matrix homeostasis. *J Cell Sci.* 2005;118(Pt 21):5089-100.

85. Bartels CF, Bükülmez H, Padayatti P, Rhee DK, van Ravenswaaij-Arts C, Pauli RM, et al. Mutations in the transmembrane natriuretic peptide receptor NPR-B impair skeletal growth and cause acromesomelic dysplasia, type Maroteaux. *Am J Hum Genet.* 2004;75(1):27-34.
86. Amano N, Mukai T, Ito Y, Narumi S, Tanaka T, Yokoya S, et al. Identification and functional characterization of two novel NPR2 mutations in Japanese patients with short stature. *J Clin Endocrinol Metab.* 2014;99(4):E713-8.
87. Wang SR, Jacobsen CM, Carmichael H, Edmund AB, Robinson JW, Olney RC, et al. Heterozygous mutations in natriuretic peptide receptor-B (NPR2) gene as a cause of short stature. *Hum Mutat.* 2015;36(4):474-81.
88. Hannema SE, van Duyvenvoorde HA, Premisler T, Yang RB, Mueller TD, Gassner B, et al. An activating mutation in the kinase homology domain of the natriuretic peptide receptor-2 causes extremely tall stature without skeletal deformities. *J Clin Endocrinol Metab.* 2013;98(12):E1988-98.
89. Miura K, Kim OH, Lee HR, Namba N, Michigami T, Yoo WJ, et al. Overgrowth syndrome associated with a gain-of-function mutation of the natriuretic peptide receptor 2 (NPR2) gene. *Am J Med Genet A.* 2014;164A(1):156-63.
90. Bocciardi R, Giorda R, Buttgerit J, Gimelli S, Divizia MT, Beri S, et al. Overexpression of the C-type natriuretic peptide (CNP) is associated with overgrowth and bone anomalies in an individual with balanced t(2;7) translocation. *Hum Mutat.* 2007;28(7):724-31.
91. Sithanandam G, Kolch W, Duh FM, Rapp UR. Complete coding sequence of a human B-raf cDNA and detection of B-raf protein kinase with isozyme specific antibodies. *Oncogene.* 1990;5(12):1775-80.
92. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, et al. Mutations of the BRAF gene in human cancer. *Nature.* 2002;417(6892):949-54.
93. Sarkozy A, Carta C, Moretti S, Zampino G, Digilio MC, Pantaleoni F, et al. Germline BRAF mutations in Noonan, LEOPARD, and cardiofaciocutaneous syndromes: molecular diversity and associated phenotypic spectrum. *Hum Mutat.* 2009;30(4):695-702.
94. Trovó-Marqui AB, Tajara EH. Neurofibromin: a general outlook. *Clin Genet.* 2006;70(1):1-13.
95. Allanson JE, Hall JG, Van Allen MI. Noonan phenotype associated with neurofibromatosis. *Am J Med Genet.* 1985;21(3):457-62.
96. Griffiths S, Thompson P, Frayling I, Upadhyaya M. Molecular diagnosis of neurofibromatosis type 1: 2 years experience. *Fam Cancer.* 2007;6(1):21-34.

97. Domingues S, Isidoro L, Rocha D, Sales Marques J. Neurofibromatosis Type 1: A Novel NF1 Mutation Associated with Mitochondrial Complex I Deficiency. *Case Rep Genet.* 2014;2014:423071.
98. Gluckman PD. Clinical review 68: The endocrine regulation of fetal growth in late gestation: the role of insulin-like growth factors. *J Clin Endocrinol Metab.* 1995;80(4):1047-50.
99. Leal AeC, Canton AP, Montenegro LR, Coutinho DC, Arnhold IJ, Jorge AA. [Mutations in insulin-like growth factor receptor 1 gene (IGF1R) resulting in intrauterine and postnatal growth retardation]. *Arq Bras Endocrinol Metabol.* 2011;55(8):541-9.
100. Abuzzahab MJ, Schneider A, Goddard A, Grigorescu F, Lautier C, Keller E, et al. IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. *N Engl J Med.* 2003;349(23):2211-22.
101. Wang SR, Carmichael H, Andrew SF, Miller TC, Moon JE, Derr MA, et al. Large-scale pooled next-generation sequencing of 1077 genes to identify genetic causes of short stature. *J Clin Endocrinol Metab.* 2013;98(8):E1428-37.
102. Guo MH, Shen Y, Walvoord EC, Miller TC, Moon JE, Hirschhorn JN, et al. Whole exome sequencing to identify genetic causes of short stature. *Horm Res Paediatr.* 2014;82(1):44-52.
103. Hauer NN, Popp B, Schoeller E, Schuhmann S, Heath KE, Hisado-Oliva A, et al. Clinical relevance of systematic phenotyping and exome sequencing in patients with short stature. *Genet Med.* 2018;20(6):630-8.
104. Hattori A, Katoh-Fukui Y, Nakamura A, Matsubara K, Kamimaki T, Tanaka H, et al. Next generation sequencing-based mutation screening of 86 patients with idiopathic short stature. *Endocr J.* 2017;64(10):947-54.
105. Kim YM, Lee YJ, Park JH, Lee HD, Cheon CK, Kim SY, et al. High diagnostic yield of clinically unidentifiable syndromic growth disorders by targeted exome sequencing. *Clin Genet.* 2017;92(6):594-605.
106. Freire BL, Homma TK, Funari MFA, Lerario AM, Vasques GA, Malaquias AC, et al. Multigene sequencing analysis of children born small for gestational age with isolated short stature. *J Clin Endocrinol Metab.* 2019.

9. ANEXOS

ANEXO A - Lista dos genes presentes nas duas versões do painel para investigação da baixa estatura

<u>ACAN</u>	<u>FGF8</u>	<u>IGF2BP2</u>	<u>OBSL1</u>	<u>SHOC2</u>
<u>ADAMTS10</u>	<u>FGFR1</u>	IGFALS	OTX2	<u>SHOX</u>
ADAMTS17	<u>FGFR3</u>	IGSF1	PAPPA2*	<u>SHOX2</u>
ARNT2	<u>GDF5*</u>	IGSF10*	PAX6*	SIX3*
<u>BMP2*</u>	GH1	<u>IHH</u>	PITX2	SMO*
<u>BMPR1B*</u>	GHR	KAL1	PNPLA6*	SOS1
BRAF	GHRH	KMT2D	POU1F1	SOS2
CBL	GHRHR	KRAS	PRKG2	SOX2
CCDC53*	GHSR	LHX3	PROKR2	SOX3
CCDC8	GLI2	LHX4	PROP1	<u>SOX5</u>
CDH2	<u>GNAS</u>	LZTR1	PTCH1*	<u>SOX6</u>
CDON*	GPR161*	MAP2K1	PTPN11	<u>SOX9</u>
<u>COL2A1*</u>	HDAC6	MAP2K2	RAB3IP	SRCAP
<u>COMP*</u>	HESX1	MEF2C*	RAF1	STAT5B
CREBBP*	HHIP	NF1	RASA2*	TCF7L1*
CUL7	HMGA2*	<u>NPPB</u>	RIT1	TGIF1
DGCR8*	<u>HOXA9</u>	<u>NPPC</u>	RNPC3	WNT5A
DMXL2	HRAS	<u>NPR2</u>	<u>ROR2</u>	ZIC2*
EP300*	IGF1	NRAS	<u>RUNX2</u>	
<u>FBN1**</u>	IGF1R	NSUN2*	SHH	

Os genes sublinhados e em negrito são aqueles relacionados diretamente à cartilagem de crescimento e/ou ao gene *SHOX*. *Os genes marcados com um asterisco estão presentes apenas na versão 2 do painel.

**Nessa versão 2, foram incluídos apenas a região *hotspot* do gene *FBN1*, os exons 41, 42 e 43, por se tratar de um gene muito grande.

ANEXO B - Coordenadas dos cromossomos X e Y correspondentes às regiões regulatórias do *SHOX* incluídas na versão 2 do painel para investigação da baixa estatura e suas respectivas coberturas (prevista e obtida).

Posição no ChrX	Posição no ChrY	Tamanho (pb)	Previsão de cobertura (%) [*]	Cobertura obtida (%) ^{**}
chrX:398.100-399.050	chrY:348.100-349.050	950	100,0	97,7
chrX:460.100-460.900	chrY:410.100-410.900	800	100,0	99,4
chrX:516.400-517.400	chrY:466.400-467.400	1000	100,0	100,0
chrX:713.900-714.900	chrY:663.900-664.900	1000	100,0	98,4
chrX:750.700-752.000	chrY:700.700-702.000	1300	100,0	100,0
chrX:763.900-764.900	chrY:713.900-714.900	1000	41,2	44,7
chrX:780.400-781.400	chrY:730.400-731.400	1000	100,0	100,0
chrX:800.700-802.000	chrY:750.700-752.000	1300	79,0	81,5
chrX:809.000-809.500	chrY:759.000-759.500	500	100,0	100,0
chrX:817.500-818.000	chrY:767.500-768.000	500	100,0	100,0
chrX:834.500-835.700	chrY:784.500-785.700	1200	93,7	93,4
chrX:884.500-885.700	chrY:834.500-835.700	1200	40,3	43,3

Posições relativas ao *GRCh37/hg19*. ^{*}A previsão de cobertura corresponde à porcentagem da região alvo prevista para ser sequenciada com boa cobertura pelo programa *Agilent SureDesign 2.0* (*Agilent Technologies, Santa Clara, CA, EUA*). ^{**}A cobertura obtida corresponde a porcentagem da região alvo que foi de fato sequenciada.

ANEXO C - Endereço dos programas e ferramentas utilizados na análise de bioinformática

Programa/ferramenta	Endereço
<i>ABraOM</i>	http://abraom.ib.usp.br/
<i>Agilent SureDesign 2.0</i>	https://earray.chem.agilent.com/suredesign/
<i>ANNOVAR</i>	http://www.openbioinformatics.org/annovar/
<i>BDGP splice site prediction</i>	http://www.fruitfly.org/seq_tools/splice.html
<i>BWA</i>	http://bio-bwa.sourceforge.net/
<i>ClinVar</i>	https://www.ncbi.nlm.nih.gov/clinvar/
<i>CONTRA</i>	http://contra-cnv.sourceforge.net/
<i>dbSNP</i>	https://www.ncbi.nlm.nih.gov/SNP/
<i>ExAC</i>	http://exac.broadinstitute.org/
<i>Freebayes</i>	https://www.msi.umn.edu/sw/freebayes/
<i>gnomAD</i>	https://gnomad.broadinstitute.org/
<i>Human splicing finder</i>	http://www.umd.be/HSF3/
<i>IGV</i>	http://software.broadinstitute.org/software/igv/
<i>OMIM</i>	https://www.omim.org/
<i>1000 Genomes Project</i>	http://www.internationalgenome.org/
<i>Sequencher 4.10.1</i>	http://www.genecodes.com/sequencher

ANEXO D - Critérios de classificação de patogenicidade dos programas *in silico* utilizados nas filtragens da planilha VCF

Ferramenta	Classificação	Critérios patológicos	Endereço
Modelo de predição para alteração do sítio de <i>splice</i>			
<i>dbSNV_ADA_SCORE</i> <i>dbSNV_RF_SCORE</i>	Entre 0 e 1	≥0,6	https://sites.google.com/site/jpopgen/dbNSFP
Modelo de predição de patogenicidade			
<i>SIFT</i>	T: tolerado; D: deletério	D	http://sift.jcvi.org/www/SIFT_chr_coords_submit.html
<i>CADD</i>	Numérico	≥15	http://cadd.gs.washington.edu/
<i>Mutation Acessor</i>	L: efeito leve; M: moderado; H:alto.	M e H	http://mutationassessor.org/
<i>Mutation Taster</i>	N e P: não deletério/polimorfismo; A e D: deletério	A e D	http://www.mutationtaster.org/
<i>PROVEAN</i>	N: neutro; D: deletério	D	http://provean.jcvi.org/genome_submit_2.php
<i>Polyphen2</i>	B: benigno; P: possivelmente patológico; D: deletério	P e D	http://genetics.bwh.harvard.edu/pph2/
<i>FATHMM</i>	T: tolerado; D: deletério	D	http://fathmm.biocompute.org.uk/
Modelo para grau de conservação do aminoácido envolvido			
<i>GERP</i>	Entre -12.3 e 6.17	≥ 2	http://mendel.stanford.edu/SidowLab/downloads/gerp/

Fonte: Adaptado de *Richards* e colaboradores, 2015 (61).

ANEXO E - Critérios do ACMG/AMP para classificação das variantes alélicas quanto à patogenicidade

Critério muito forte para patogenicidade	
PVS1	Variante <i>nonsense</i> , <i>frameshift</i> , em sítios de <i>splice</i> canônicos ± 1 ou 2, códon de iniciação ou deleção de 1 ou vários exons (<i>null variant</i>) em um gene no qual variantes com perda de função (<i>LoF</i>) são mecanismos conhecidos de doença
Critérios fortes para patogenicidade	
PS1	Troca de aminoácido já estabelecida como patogênica, independente da troca do nucleotídeo
PS2	Variante <i>de novo</i> em um paciente com doença sem história familiar (com paternidade confirmada)
PS3	Presença de estudos funcionais (<i>in vitro</i> ou <i>in vivo</i>) bem estabelecidos que suportem o efeito prejudicial no gene ou sua proteína
PS4	A prevalência da variante nos indivíduos afetados é significativamente maior quando comparada com a prevalência em controles
Critérios moderados para patogenicidade	
PM1	Variante localizada em uma região <i>hotspot</i> para mutações e/ou região importante já estabelecida com um domínio funcional da proteína, sem variantes descritas como benignas
PM2	Variante ausente nos bancos de dados populacionais (<i>1000Genomes</i> , <i>ExAC</i>) (ou presente em frequência extremamente baixa, para os casos recessivos)
PM3	Variante detectada em <i>trans</i> com uma variante patogênica para as doenças recessivas
PM4	Modificação no tamanho da proteína causada por deleções ou inserções <i>in-frame</i> em região não repetitiva ou por variantes que perdem o códon de parada (tipo <i>stop-loss</i>)
PM5	Nova variante <i>missense</i> em um aminoácido em que uma variante <i>missense</i> diferente já foi estabelecida como patogênica
PM6	Variante considerada <i>de novo</i> , sem confirmação da paternidade
Critérios fracos para patogenicidade	
PP1	Segregação da variante (em um gene conhecido como causador de doença) com a doença em uma família com vários membros afetados
PP2	Variante <i>missense</i> em um gene com uma baixa taxa de mutações <i>missense</i> benignas e no qual esse tipo de variante é considerado um mecanismo comum de doença
PP3	Várias ferramentas computacionais suportam a evidência de efeito deletério no gene ou proteína (conservação evolutiva, impacto no <i>splicing</i> etc)
PP4	O fenótipo do paciente ou a história da família são bastante específicos de uma doença monogênica
PP5	Variante recém-publicada como patogênica, porém o laboratório não apresenta condições de fazer um estudo funcional independente.
Critérios fracos para benignidade	
BP1	Variante <i>missense</i> em um gene no qual variantes com perda de função (<i>LoF</i>) são mecanismos conhecidos de doença
BP2	Variante detectada em <i>trans</i> com uma variante patogênica dominante com penetrância completa ou observada em <i>cis</i> com uma variante patogênica em qualquer tipo de herança
BP3	Deleções ou inserções <i>in-frame</i> em região repetitiva sem função conhecida

(continua)

ANEXO E (conclusão) - Critérios do ACMG/AMP para classificação das variantes alélicas quanto à patogenicidade

Critérios fracos para benignidade

BP4	Várias ferramentas computacionais sugerindo ausência de efeito deletério no gene ou proteína (conservação evolutiva, impacto no <i>splicing</i> etc)
BP5	Variante encontrada em um caso com uma base molecular alternativa para doença
BP6	Variante recém-publicada como benigna, porém o laboratório não apresenta condições de fazer um estudo funcional independente.
BP7	Variante sinônima cujas ferramentas de predição de <i>splicing</i> não predizem impacto na sequência consenso de <i>splice</i> nem a criação de um novo sítio, e o nucleotídeo não é conservado

Critérios fortes para benignidade

BS1	A frequência do alelo é maior que a esperada para a doença
BS2	Variante para doença recessiva (em homozigose), dominante (em heterozigose) ou ligada ao X (hemizigose), com penetrância completa e de início precoce, observada em indivíduo adulto saudável
BS3	Presença de estudos funcionais (<i>in vitro</i> ou <i>in vivo</i>) bem estabelecidos mostrando ausência de efeito prejudicial para a proteína ou <i>splicing</i>
BS4	Ausência de segregação nos indivíduos afetados da família

Critério benigno por si só

BA1	Frequência alélica maior que 5% nos bancos de dados populacionais (<i>1000Genomes</i> , <i>ExAC</i>)

Fonte: Adaptado de Richards e colaboradores, 2015 (61). PVS: critério muito forte para patogenicidade; PS: critério forte para patogenicidade; PM: critério moderado para patogenicidade; PP: critério fraco para patogenicidade; BP: critério fraco para benignidade; BS: critério forte para benignidade; BA: critério benigno por si só. *Lof*: perda de função (de *loss of function*); *1000Genomes*: *1000Genomes Project*; *ExAC*: *Exome Aggregation Consortium*.

ANEXO F - Normas para combinação dos critérios do ACMG/AMP (61) para classificação das variantes alélicas identificadas

Classificação	Combinação dos critérios
Variante patogênica	(I) 1 critério muito forte para patogenicidade (PVS1) e (a) ≥ 1 critérios fortes para patogenicidade (PS1-PS4) ou (b) ≥ 2 critérios moderados para patogenicidade (PM1-PM6) ou (c) 1 critério moderado para patogenicidade (PM1-PM6) e 1 critério fraco para patogenicidade (PP1-PP5) ou (d) ≥ 2 critérios fracos para patogenicidade (PP1-PP5) ou (II) ≥ 2 critérios fortes para patogenicidade (PS1-PS4) ou (III) 1 critério forte para patogenicidade (PS1-PS4) e (a) ≥ 3 critérios moderados para patogenicidade (PM1-PM6) ou (b) 2 critérios moderados para patogenicidade (PM1-PM6) e ≥ 2 critérios fracos para patogenicidade (PP1-PP5) ou (c) 1 critérios moderados para patogenicidade (PM1-PM6) e ≥ 4 critérios fracos para patogenicidade (PP1-PP5)
Variante provavelmente patogênica	(I) 1 critério muito forte para patogenicidade (PVS1) e 1 critério moderado para patogenicidade (PM1-PM6) ou (II) 1 critério forte para patogenicidade (PS1-PS4) e 1 ou 2 critérios moderado para patogenicidade (PM1-PM6) ou (III) 1 critério forte para patogenicidade (PS1-PS4) e ≥ 2 critérios fracos para patogenicidade (PP1-PP5) ou (IV) ≥ 3 critérios moderados para patogenicidade (PM1-PM6) ou (V) 2 critérios moderados para patogenicidade (PM1-PM6) e ≥ 2 critérios fracos para patogenicidade (PP1-PP5) ou (VI) 1 critério moderado para patogenicidade (PM1-PM6) e ≥ 4 critérios fracos para patogenicidade (PP1-PP5)
Variante benigna	(I) 1 critério benigno por si só (BA1) ou (II) ≥ 2 critérios fortes para benignidade (BS1-BS4)
Variante provavelmente benigna	(I) 1 critério forte para benignidade (BS1-BS4) e 1 critério fraco para benignidade (BP1-BP7) ou (II) ≥ 2 critérios fracos para benignidade (BP1-BP7)
Variante de significado incerto	(I) outra combinação de critérios não mencionada acima ou (II) critérios para benignidade e patogenicidade contraditórios

Fonte: Adaptado de *Richards* e colaboradores, 2015 (61). PVS: critério muito forte para patogenicidade; PS: critério forte para patogenicidade; PM: critério moderado para patogenicidade; PP: critério fraco para patogenicidade; BP: critério fraco para benignidade; BS: critério forte para benignidade; BA: critério benigno por si só.

ANEXO G - Critérios ACMG/AMP utilizados para a classificação das principais variantes alélicas de ponto com potencial patogênico identificadas no sequenciamento do painel

Paciente	Gene	Variante	Critérios da classificação	Classificação ACMG/AMP
18	<i>FGFR3</i>	c.667C>T: p.Arg223Cys	PM1, PM2, PP1, PP3	Provavelmente patogênica
19	<i>FGFR3</i>	c.1620C>A: p.Asn540Lys	PS1, PS4, PM1, PM2, PP3, PP5	Patogênica
20	<i>FGFR3</i>	c.1657G>T: p.Val553Leu	PM1, PM2, PM5, PP1, PP3, PP5	Provavelmente patogênica
21	<i>NPR2</i>	c.1711-1G>C	PVS1, PM2, PP1, PP3	Patogênica
22	<i>IHH</i>	c.172G>A: p.Glu58Lys	PS4, PM1, PM2, PP1, PP3	Patogênica
23	<i>IHH</i>	c.532G>C: p.Val178Leu	PS2, PS4, PM1, PM2, PP3	Patogênica
24	<i>IHH</i>	c.685G>A: p.Val229Met	PM1, PM2, PP1, PP3	Provavelmente patogênica
	<i>GH1</i>	c.124C>T: p.Arg42Cys	PM1, PM2, PP1, PP3, PP5	Provavelmente patogênica
25	<i>IHH</i>	c.949G>A: p.Val317Met	PM1, PM2, PP1, PP3, PP5	Provavelmente patogênica
26	<i>GNAS</i>	c.2682C>G: p.Ser894Arg	PS2, PM1, PM2, PP2, PP3	Patogênica
27	<i>IGF1R</i>	c.143G>A: p.Arg48His	PM1, PM2, PP1, PP3	Provavelmente patogênica
28	<i>BRAF</i>	c.712G>A: p.Val238Ile	PM1, PM2, PP2, PP3	Provavelmente patogênica
29	<i>NF1</i>	c.3447G>A: p.Met1149Ile	PS1, PM1, PM2, PP3, PP5, BP1	Patogênica

PVS: critério muito forte para patogenicidade; PS: critério forte para patogenicidade; PM: critério moderado para patogenicidade; PP: critério fraco para patogenicidade; BP: critério fraco para benignidade. Os detalhes dos critérios da classificação da ACMG/AMP estão descritos no Anexo E.

ANEXO H - Detalhes das principais variantes classificadas como variantes de significado incerto identificadas no sequenciamento do painel que segregaram com o fenótipo da baixa estatura na família ou estavam ausentes nos familiares com altura dentro da normalidade

Paciente	Gene	Sequência referência	Variante	dbSNP ID	Fenótipos já associados ao gene (tipo de herança)
30	ACAN	NM_013227	c.1216G>A: p.Glu406Lys	rs774443075	Displasia pondiloepimetáfiseal tipo <i>aggrecan</i> (AR), baixa estatura com idade óssea avançada (AD)
	GH1	NM_000515	c.470G>A: p.Gly157Asp	rs201165777	Deficiência isolada de hormônio de crescimento (AR, alguns AD) e Síndrome de <i>Kowarski</i> (AR)
31	ADAMTS10	NM_030957	c.1039C>T: p.Arg347Cys	rs782789174	Síndrome de <i>Weill-Marchesani</i> (AR)
32	ADAMTS17	NM_139057	c.C3164T: p.Thr1055Met	rs371570653	Síndrome de <i>Weill-Marchesani</i> (AR)
33	COL2A1	NM_001844	c.550G>A: p.Ala184Thr	rs201817670	Displasias esqueléticas diversas (AD)
34	EP300	NM_001429	c.6928C>G: p.Pro2310Ala	ND	Sem doença no OMIM. Associado a Síndrome de <i>Rubinstein-Taybi</i> (AD) e ao câncer colo retal
35	NPPB	NM_002521	c.145C>T: p.His49Tyr	ND	Sem doença no OMIM. Associado a pericardite constritiva e insuficiência cardíaca sistólica
36	OBSL1	NM_015311	c.1984C>A: p.Gln662Lys	rs761482110	Síndrome de 3M (AR)
	OBSL1	NM_015311	c.4285G>A: p.Val1429Met	rs375101491	Síndrome de 3M (AR)
37	ROR2	NM_004560	c.2381C>T: p.Pro794Leu	rs374141378	Braquidactilia tipo B1 (AD) e Síndrome de <i>Robinow</i> (AR)
38	ROR2	NM_004560	c.2516C>T: p.Pro839Leu	ND	Braquidactilia tipo B1 (AD) e Síndrome de <i>Robinow</i> (AR)

ND: não descrito; AR: herança autossômica recessiva; AD: herança autossômica dominante; OMIM: *Online Mendelian Inheritance in Man* (banco de dados de genes humanos e fenótipos genéticos).

ANEXO I - Detalhes das principais variantes classificadas como variantes de significado incerto identificadas no sequenciamento do painel que não foram segregadas nas famílias dos pacientes por ausência de amostra de DNA dos familiares

Paciente	Gene	Sequência referência	Variante	dbSNP ID	Fenótipos já associados ao gene (tipo de herança)
39	ACAN	NM_013227	c.2902T>C: p.Ser968Pro	rs568435726	Displasia pondiloepimetáfiseal tipo <i>aggrecan</i> (AR), baixa estatura com idade óssea avançada (AD)
40	ACAN	NM_013227	c.6443G>A: p.Gly2148Glu	ND	Displasia pondiloepimetáfiseal tipo <i>aggrecan</i> (AR), baixa estatura com idade óssea avançada (AD)
41	ADAMTS17	NM_139057	c.2965G>A: p.Gly989Arg	rs550242741	Síndrome de <i>Weill-Marchesani</i> (AR)
	NPR2	NM_003995	c.838C>T: p.Arg280Trp	rs771997028	Displasia acromesomélica tipo <i>Maroteaux</i> (AR), condrodysplasia epifiseal tipo <i>Miura</i> (AD) e BE sem achados esqueléticos específicos (AD)
42	ADAMTS17	NM_139057	c.1322A>G: p.Lys441Arg	rs145156568	Síndrome de <i>Weill-Marchesani</i> (AR)
	ROR2	NM_004560	c.2381C>T: p.Pro794Leu	rs374141378	Braquidactilia tipo B1 (AD) e Síndrome de <i>Robinow</i> (AR)
43	EP300	NM_001429	c.2023A>G: p.Met675Val	ND	Sem doença no OMIM. Associado à Síndrome de <i>Rubinstein-Taybi</i> (AD) e ao câncer colo retal
44	EP300	NM_001429	c.2351C>T: p.Pro784Leu	rs201480900	Sem doença no OMIM. Associado à Síndrome de <i>Rubinstein-Taybi</i> (AD) e ao câncer colo retal

ND: não descrito; AR: autossômica recessiva; AD: autossômica dominante; BE: baixa estatura; OMIM: *Online Mendelian Inheritance in Man* (banco de dados de genes humanos e fenótipos genéticos).

ANEXO J - Artigos em que participei como colaboradora durante o período do doutorado que utilizaram o sequenciamento paralelo em larga escala como metodologia.

1. *Multigene sequencing analysis of children born small for gestational age with isolated short stature.*
J Clin Endocrinol Metab. 2019 Jan 2. [Epub ahead of print]
Freire BL, Homma TK, **Funari MFA**, Lerario AM, Vasques GA, Malaquias AC, Arnhold IJP, Jorge AAL.
2. *Novel SUZ12 mutations in Weaver-like syndrome.*
Clin Genet. 2018;94:461-466.
Imagawa E, Albuquerque EVA, Isidor B, Mitsuhashi S, Mizuguchi T, Miyatake S, Takata A, Miyake N, Boguszewski MCS, Boguszewski CL, Lerario AM, **Funari MFA**, Jorge AAL, Matsumoto N.
3. *Two rare loss-of-function variants in the STAG3 gene leading to primary ovarian insufficiency.*
Eur J Med Genet. 2018 Jul 10. [Epub ahead of print]
França MM, Nishi MY, **Funari MFA**, Lerario AM, Baracat EC, Hayashida SAY, Maciel GAR, Jorge AAL, Mendonça BB.
4. *BCL11B mutations in patients affected by a neurodevelopmental disorder with reduced type 2 innate lymphoid cells.*
Brain. 2018 Jul 9. [Epub ahead of print]
Lessel D, Gehbauer C, Bramswig NC, Schluth-Bolard C, Venkataramanappa S, van Gassen KLI, Hempel M, Haack TB, Baresic A, Genetti CA, **Funari MFA**, Lessel I, Kuhlmann L, Simon R, Liu P, Denecke J, Kuechler A, de Kruijff I, Shoukier M, Lek M, Mullen T, Lüdecke HJ, Lerario AM, Kobbe R, Krieger T, Demeer B, Lebrun M, Keren B, Nava C, Buratti J, Afenjar A, Shinawi M, Guillen Sacoto MJ, Gauthier J, Hamdan FF, Laberge AM, Campeau PM, Louie RJ, Cathey SS, Prinz I, Jorge AAL, Terhal PA, Lenhard B, Wieczorek D, Strom TM, Agrawal PB, Britsch S, Tolosa E, Kubisch C.
5. *IHH gene mutations causing short stature with nonspecific skeletal abnormalities and response to growth hormone therapy.*
J Clin Endocrinol Metab. 2018;103(2):604-614.
Vasques GA, **Funari MFA**, Ferreira FM, Aza-Carmona M, Sentchordi-Montané L, Barraza-García J, Lerario AM, Yamamoto GL, Naslavsky MS, Duarte YAO, Bertola DR, Heath KE, Jorge AAL.
6. *Homozygous loss of function BRCA1 variant causing a Fanconi-anemia-like phenotype, a clinical report and review of previous patients.*
Eur J Med Genet. 2018;61(3):130-133.
Freire BL, Homma TK, **Funari MFA**, Lerario AM, Leal AM, Velloso EDRP, Malaquias AC, Jorge AAL.

(continua)

ANEXO J - (conclusão) - Artigos em que participei como colaboradora durante o período do doutorado que utilizaram o sequenciamento paralelo em larga escala como metodologia.

7. *Identification of the first homozygous 1-bp deletion in GDF9 gene leading to primary ovarian insufficiency by using targeted massively parallel sequencing.*
Clin Genet. 2018;93(2):408-411.
França MM, **Funari MFA**, Nishi MY, Narcizo AM, Domenice S, Costa EMF, Lerario AM, Mendonça BB.
8. *A novel homozygous 1-bp deletion in the NOBOX gene in two Brazilian sisters with primary ovarian failure.*
Endocrine. 2017;58(3):442-447.
França MM, **Funari MFA**, Lerario AM, Nishi MY, Pita CC, Fontenele EGP, Mendonça BB.
9. *Mutations in C-natriuretic peptide (NPPC): a novel cause of autosomal dominant short stature.*
Genet Med. 2018;20(1):91-97.
Hisado-Oliva A, Ruzafa-Martin A, Sentchordi L, **Funari MFA**, Bezanilla-López C, Alonso-Bernáldez M, Barraza-García J, Rodriguez-Zabala M, Lerario AM, Benito-Sanz S, Aza-Carmona M, Campos-Barros A, Jorge AAL, Heath KE.
10. *A Novel homozygous missense FSHR variant associated with hypergonadotropic hypogonadism in two siblings from a Brazilian family.*
Sex Dev. 2017;11(3):137-142.
França MM, Lerario AM, **Funari MFA**, Nishi MY, Narcizo AM, de Mello MP, Guerra-Junior G, Maciel-Guerra AT, Mendonça BB.
11. *Clinical characterization of patients with autosomal dominant short stature due to aggrecan mutations.*
J Clin Endocrinol Metab. 2017;102(2):460-469.
Gkourogianni A, Andrew M, Tyzinski L, Crocker M, Douglas J, Dunbar N, Fairchild J, **Funari MF**, Heath KE, Jorge AA, Kurtzman T, LaFranchi S, Lalani S, Lebl J, Lin Y, Los E, Newbern D, Nowak C, Olson M, Popovic J, Pruhová Š, Elblova L, Quintos JB, Segerlund E, Sentchordi L, Shinawi M, Stattin EL, Swartz J, Angel AG, Cuéllar SD, Hosono H, Sanchez-Lara PA, Hwa V, Baron J, Nilsson O, Dauber A.
12. *Two patients with severe short Stature due to a FBN1 mutation (p.Ala1728Val) with a mild form of acromicric dysplasia.*
Horm Res Paediatr. 2016;86(5):342-348.
de Bruin C, Finlayson C, **Funari MF**, Vasques GA, Lucheze Freire B, Lerario AM, Andrew M, Hwa V, Dauber A, Jorge AA.