



**Fundação Oswaldo Cruz
Instituto Nacional de Saúde da Mulher,
da Criança e do Adolescente Fernandes Figueira**

**ESTUDO EXPLORATÓRIO DA SINDROME HIPOPLASIA
CARTILAGEM-CABELO NO BRASIL: UMA ABORDAGEM
GENÔMICA, PROTEÔMICA E DE ANÁLISE ESTRUTURAL DE RNA**

Maria Eduarda Gomes Ferreira Neves

Rio de Janeiro
Outubro de 2022



**Fundação Oswaldo Cruz
Instituto Nacional de Saúde da Mulher,
da Criança e do Adolescente Fernandes Figueira**

**ESTUDO EXPLORATÓRIO DA SINDROME HIPOPLASIA
CARTILAGEM-CABELO NO BRASIL: UMA ABORDAGEM
GENÔMICA, PROTEÔMICA E DE ANÁLISE ESTRUTURAL DE
RNA**

Maria Eduarda Gomes Ferreira Neves

Tese apresentada à Pós-
graduação em Pesquisa
Aplicada à Saúde da Criança e
da Mulher, como parte dos
requisitos para obtenção do
título de Doutora em Ciências

Orientadora: Dra. Sayonara Maria de Carvalho Gonzalez
Co-orientadores: Dr. Juan Clinton Llerena Junior
Dra. Luciana Pizzatti Barboza

Rio de Janeiro
Outubro de 2022

CIP - Catalogação na Publicação

Gomes Ferreira Neves, Maria Eduarda.

Estudo exploratório da Síndrome Hipoplasia Cartilagem-Cabelo no Brasil: uma abordagem genômica, proteômica e de análise estrutural de RNA / Maria Eduarda Gomes Ferreira Neves. - Rio de Janeiro, 2022.

200 f.

Tese (Doutorado Acadêmico em Pesquisa Aplicada à Saúde da Criança e da Mulher) - Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira, Rio de Janeiro - RJ, 2022.

Orientadora: Sayonara Maria de Carvalho Gonzalez.

Co-orientador: Juan Clinton Llerena Jr..

Co-orientadora: Luciana Pizzatti Barboza.

Bibliografia: f. 147-160

1. Osteocondrodisplasias. 2. Técnicas de Diagnóstico Molecular. 3. Efeito Fundador. 4. Reação em Cadeia da Polimerase em Tempo Real. 5. Proteômica. I. Título.

“Um paciente não é só um paciente.
Ele é o amor de alguém.
O pai de alguém
A mãe de alguém
O filho de alguém
Avô, avó de alguém
O melhor amigo de alguém
Nossa missão é, a todo instante,
cuidar e fazer tudo pelo amor da vida de alguém.”

AGRADECIMENTOS

São muitas as pessoas que merecem agradecimento pelo desenvolvimento dessa tese. Mas dessa vez vou me permitir agradecer, primeiramente, a mim. Não é o caso de ingratidão ou de falta de modéstia. Mas é que não posso negar que sinto orgulho da minha trajetória e do quanto aprendi a valorizá-la. Foram pouco mais de quatro anos de profunda transformação e quando olho para trás, quase não reconheço a pessoa que iniciou essa jornada. Tiveram momentos de altos e baixos, com toda certeza, mas foram muitas lições aprendidas, muita resiliência e muito amadurecimento. Somado a tantos acontecimentos paralelos, ou talvez em decorrência deles, finalmente me enxerguei cientista. Então, inicialmente, parabênzo a mim mesma pela força e pelo comprometimento com esse trabalho que se mostrou tão bonito.

Todo esse esforço obviamente implicou no desenvolvimento de uma rabugentice genuína, afinal ninguém é de ferro. E como tenho senso crítico para reconhecer meu esforço, também o tenho para reconhecer a natureza enfadonha das minhas lamentações diárias no laboratório e na vida (quem não me ouviu reclamar de uma clonagem é porque não esteve perto o suficiente). Então, nada mais justo do que dedicar umas belas palavras aos amigos tão queridos que estiveram ao meu lado a todo o tempo (e aproveitando para esclarecer que a ordem não interfere na preferência, afinal cada um contribuiu à sua maneira com a realização desse doutorado).

Para minha querida orientadora, que há 9 anos aguenta minhas inseguranças e minha forma excêntrica de pensar, vai o mais profundo e sincero agradecimento. A vida com certeza não me deu “só” uma chefe, mas uma parceirona para todas as horas. Obrigada pelas oportunidades, pelo presente que é participar desse projeto,

mas principalmente por acolher aquela menina em 2013 e a transformar na cientista que hoje ela é.

Ao meu co-orientador Juan, que sempre incentivou e impulsionou o meu crescimento profissional. Obrigada por sempre me enxergar, apesar de eu sempre querer me “esconder”.

Minha gêmea Natita. Ah, minha gêmea, Natita! Será que preciso dizer que é a irmã que meu coração escolheu? Fica até difícil (e clichê) tentar encontrar as palavras. Você sempre será a sensatez por fora do caos que eu sou.

Minha Bellezita, delicadeza em forma de pessoa. Foi um presente ter você junto a mim nessa caminhada com tanta generosidade e incentivo no nosso dia-a-dia. Obrigada, obrigada! Por tudinho!

Bruna, a única coisa que consigo pensar é que você é uma pessoa muito evoluída para sanar as minhas infinitas dúvidas de proteômica com a maior paciência desse mundo. O quanto eu te perturbei, não dá nem para contar. Mas também não podemos negar que nos divertimos muito nas nossas conversas. Você é muito maravilhosa! Conta comigo sempre!

Ferdi, obrigada por você compartilhar todo o conhecimento técnico que possibilitou iniciar essa jornada da confirmação do efeito fundador, mas também por compartilhar as experiências de vida nos momentos de angústia e aperto no coração. Você foi demais!

Julio, você foi uma grata surpresa. Começou tirando uma dúvida aqui e outra ali e quando menos percebi tinha um novo amigo para tentar desvendar os mistérios dos problemas técnicos ao longo dos experimentos. As divagações renderam muitas “viagens na maionese” e gargalhadas. Além do apoio intelectual e os incentivos, não

posso deixar de agradecer a ajuda nutricional para manter o tecido adiposo em dia com os inúmeros bombons Ouro Branco.

Dani Fins (meu chaveirinho), que me deu a maior força com as PCRs malucas neste final de tese, dedico um grande agradecimento e um abraço bem apertado!

Carlos, amigo querido, obrigada por ser sempre tão gentil, disponível e acima de tudo um bom amigo!

Soczita, sempre me incentivando no dia-a-dia, sou grata pelas palavras de apoio, especialmente nos dias em que tudo dava errado!

À toda equipe do lab e à Neia, agradeço pelo apoio diário.

Um agradecimento especial para às Dras. Denise Cavalcanti e Têmis Félix, pela parceria de anos nesse projeto lindo.

À toda equipe da do LABMOPS/UFRJ e à Dra. Luciana, pelo suporte e pelo acolhimento no desenvolvimento dos ensaios de proteômica.

Ao Dr. Ernesto Cafarena e à Dra. Déborah Antunes do Laboratório de Genômica Funcional e Bioinformática - IOC/Fiocruz, pela colaboração neste trabalho.

À banca examinadora, agradeço pela disponibilidade e por aceitarem contribuir com a melhoria deste trabalho.

Aos pacientes e seus familiares, por todo o empenho em participar das pesquisas desenvolvidas.

Ao Inagemp, CAPES e FIOCRUZ pelo apoio financeiro.

Agradeço às plataformas da Fiocruz (em especial ao LACIFF e à plataforma de Sequenciamento de DNA por eletroforese capilar/IOC), pela estrutura e colaboração de todos esses anos.

A todos os amigos da vida e à família que ganhei com o casório (ixi, é gente à beça!), eu agradeço por me incentivarem durante toda essa jornada!

Ao meu marido, o melhor de todo esse mundo! Que me deu aconchego e trouxe o equilíbrio em todos os momentos que eu vacilava, além de vibrar com todas as conquistas, por menor que elas fossem. Obrigada por tanto, te amo!

Minha família, puxa! Tenho tanto a agradecer que usar muitas palavras acabaria não dizendo nada. O que sinto que resume muito bem o sentimento é exatamente a estrofe da música de Toy Story “eu nada seria se não fosse você(s)!”. Mas aqui ainda vai um especial agradecimento aos meus pais. Espero sempre honrar a vocês!

À Deus e toda a espiritualidade amiga. Não tenho dúvidas que se cheguei até aqui, foi por estar sendo conduzida para este propósito.

RESUMO

A Hipoplasia Cartilagem-Cabelo (do inglês, Cartilage Hair Hypoplasia; CHH) é uma doença autossômica recessiva descrita por McKusick *et al.*, em 1964 em crianças da comunidade Amish. Clinicamente, os pacientes apresentam displasia óssea metafisária em diferentes graus de gravidade. Além disso, também pode ser observado hipotricose e imunodeficiência. Já no âmbito molecular, a condição se caracteriza por variantes patogênicas do gene *RMRP*. Apesar de ter sido delineada há quase 60 anos, ainda não existe uma correlação genótipo-fenótipo bem compreendida. Este trabalho é uma continuação do trabalho de mestrado realizado no IFF/Fiocruz, durante 2017 e 2018, no qual foi descrita uma coorte de 23 pacientes brasileiros com CHH em que foram encontradas diversas variantes patogênicas. Dentre estas, a variante g.196C>T destacou-se por sua elevada frequência no grupo estudado, diferentemente do observado em pacientes de outras nacionalidades, sugerindo um possível efeito fundador (EF) para a população brasileira. Esse trabalho teve por objetivo realizar diferentes ensaios moleculares para auxiliar na compreensão da CHH, além de investigar a hipótese de uma origem ancestral comum da variante g.196C>T. O estudo foi dividido em 4 eixos, sendo três relacionados às pesquisas de caráter exploratório dos mecanismos da doença e um dedicado à análise do EF. Dentro deste último, utilizando um painel de marcadores do tipo TAG SNPs cuidadosamente selecionados, foi observado que cromossomos de diferentes regiões brasileiras carregando o nucleotídeo T na posição 196 do gene *RMRP* compartilharam o haplótipo T/C/G/A (16/17 haplótipos), apontando para uma origem comum desta substituição de base no gene *RMRP*. Adicionalmente, foram realizadas análises de proteômica comparativa, evidenciando que o perfil proteômico dos leucócitos de pacientes e controles expressam proteínas que traduzem vias moleculares distintas, além de apresentar diferenças de expressão de proteínas importantes relacionadas aos fenótipos clínicos da doença. Também foram realizados ensaios de RT-qPCR que mostraram que tanto os níveis do RNA *RMRP*, quanto dos dois pequenos RNAs derivados de *RMRP*, estavam significativamente reduzidos nos pacientes em relação ao grupo controle. Por fim, com o intuito de tentar prever o impacto das variantes na estrutura tridimensional do RNA, foi realizada uma análise *in silico* que mostrou que as alterações patogênicas identificadas nos pacientes ocorreram tanto em regiões conservadas entre espécies de mamíferos quanto em domínios essenciais para o complexo ribonucleoproteico. Duas alterações estão localizadas em regiões associadas à biogênese dos pequenos RNAs derivados de *RMRP*. Além disso, foi possível observar que certas variantes podem alterar o pareamento de bases e a topologia da estrutura das alças da molécula, o que poderia influenciar na montagem do complexo RNase MRP. Em conjunto, os dados desta tese lançam luz sobre diversos pontos ainda não explorados para CHH, além de dar suporte para novos estudos que tenham por objetivo viabilizar uma medicina de precisão, contribuindo para minimizar os impactos da doença e para a promoção da qualidade de vida dos pacientes.

Palavras chave: Osteocondrodiplasias; Técnicas de Diagnóstico Molecular; Efeito Fundador; Reação em Cadeia da Polimerase em Tempo Real; Proteômica

ABSTRACT

Cartilage Hair Hypoplasia (CHH) is an autosomal recessive disease described by McKusick et al. in 1964 in the Amish community. Clinically, patients present metaphyseal bone dysplasia in different degrees of severity. In addition, hypotrichosis and immunodeficiency may also be present. At the molecular level, the condition is characterized by pathogenic variants of the *RMRP* gene. Although CHH has been described more than 60 years ago, the genotype-phenotype correlation is still not well-understood. This work is a continuation of the dissertation carried out at IFF/Fiocruz, during 2017 and 2018, in which a cohort of 23 Brazilian patients with CHH was described and several pathogenic variants were found. Among these, the high frequency of g.196C>T variant in the studied group called our attention, unlike that observed in patients of other nationalities, suggesting a possible founder effect (EF) in the Brazilian population. This work aimed to perform different molecular assays to aid in the understanding of CHH, in addition to investigating the hypothesis of a common ancestral origin of the g.196C>T variant. The study was divided into 4 axes, three related to exploratory research on disease mechanisms and one dedicated to the analysis of EF. Within the latter, using a carefully selected panel of TAG SNPs markers, it was observed that chromosomes from different Brazilian regions carrying T at nucleotide 196 of the *RMRP* gene shared the T/C/G/A haplotype (16/17 haplotypes), indicating a common origin of this base substitution at position 196 of the *RMRP* gene. Additionally, comparative proteomics analysis was performed, showing that the proteomic profile of leukocytes from patients and controls express proteins that translate distinct molecular pathways, in addition to presenting differences in the expression of proteins related to important clinical phenotypes of the disease. RT-qPCR assays were also performed, which showed that both *RMRP* RNA levels and the two *RMRP*-derived small RNAs were significantly reduced in patients compared to the control group. Finally, in order to try to predict the impact of the variants on the three-dimensional structure of the RNA, an *in silico* analysis was performed which showed that the pathogenic alterations identified in the patients occurred both in regions conserved between mammalian species and in domains essential for the ribonucleoprotein complex. Two alterations are located in regions associated with the biogenesis of *RMRP*-derived small RNAs. In addition, it was possible to observe that certain variants can change the base pairing and the topology of the structure of the molecule's loops, which could influence the assembly of the RNase MRP complex. Together, the data from this thesis shed light on several points not yet explored for CHH, in addition to providing support for new studies that aim to enable a precision medicine toward CHH patients, helping to minimize the impacts of the disease and to promote quality of life of patients.

Keywords: Osteochondrodysplasias; Molecular Diagnostic Techniques; Founder Effect; Real-Time Polymerase Chain Reaction; Proteomics

LISTA DE TABELAS

Tabela 1: Tag SNPs selecionados para a investigação do efeito fundador.	78
Tabela 2: Sumarização dos resultados das análises do gene <i>RMRP</i> dos pacientes.	103
Tabela 3: Médias individuais de ancestralidade analisando com 3 populações parentais. geradas <i>pelo Software Structure</i>	109
Tabela 4: Frequências haplotípicas dos TAG SNPS selecionados para as populações Europeia e Africana.....	111
Tabela 5: Diversidade haplotípica calculada para diferentes grupos dentro da coorte.	116
Tabela 6: Expressão dos dois pequenos RNAs derivados de RMRP em pacientes com CHH e em indivíduos controles.....	119
Tabela 7: Principais vias canônicas; reguladores a montante; funções moleculares e celulares; doenças e desordens identificados pelo software IPA a partir dos dados de proteômica em leucócitos de pacientes com CHH.. ..	126

LISTA DE FIGURAS

Figura 1: Linhagens do embrião durante a neurulação.	27
Figura 2: Ossificação intramembranosa..	28
Figura 3: Esquema da ossificação endocondral..	31
Figura 4: Esquema da placa de crescimento.....	32
Figura 5: Vias de sinalização de HH e PTHrP na diferenciação de condrócitos e osteoblastos..	35
Figura 6: Vias de sinalização Wnt e Notch na diferenciação de condrócitos e osteoblastos..	38
Figura 7: Vias de sinalização de TGF β , BMP e FGF na diferenciação de condrócitos e osteoblastos.....	40
Figura 8: Reguladores da diferenciação condrogênica de MSCs identificados até o momento.....	42
Figura 9: Radiografia de membro inferior..	45
Figura 10: Espectro da Hipoplasia Cartilagem-Cabelo.	46
Figura 11: Esquema da estrutura da ribonuclease (RNase RMRP).....	48
Figura 12: Processamento do pré-rRNA pela RNase MRP em <i>S. cerevisiae</i>	49
Figura 13: Funções da RNase MRP.....	50
Figura 14: Estrutura secundária de RMRP indicando a localização de RMRP-S1 e S2..	52
Figura 15: Tipos de alterações que ocorrem no gene <i>RMRP</i>	55
Figura 16: Esquema mostrando a sequência do gene RMRP humano e a sua conservação entre várias espécies categorizada em 4 níveis pelo alinhamento de sequência de 9 espécies	56

Figura 17: Distribuição das concentrações de uniões consanguíneas pelo mundo ⁸⁷	60
Figura 18: Estimativas genéticas ancestrais de estudos de diferentes regiões geográficas.	64
Figura 19: Desenho do estudo delineando as etapas realizadas no trabalho.	69
Figura 20: Esquema dos marcadores genéticos selecionados para montagem dos haplótipos.	78
Figura 21: Mapa de desequilíbrio de ligação dos tag-SNPs selecionados pelo Haploview.	79
Figura 22: Fluxograma das etapas para seleção de marcadores do tipo TAG SNPs para a construção do painel molecular para análise de haplótipos.	80
Figura 24: Esquema do vetor pKS.	83
Figura 23: Esquema da inferência das fases dos alelos para os marcadores rs10972552, rs7021463, rs1339374, rs1361338.	88
Figura 25: Variantes genéticas apontadas na estrutura secundária do RMRP. . .	105
Figura 26: Caracterização das variantes localizadas nos domínios da RNase MRP.	107
Figura 27: Proporções da ancestralidade individual obtidas pelo programa <i>STRUCTURE</i>	108
Figura 28: Distribuição da frequência absoluta dos haplótipos identificados na coorte estudada em relação à frequência dos haplótipos baseado no 1000G.	112
Figura 29: Distribuição dos haplótipos por estado.	113
Figura 30: Rede de haplótipos elaborado no programa Network.	115
Figura 31: Expressão dos transcritos RMRP por qPCR em tempo real.	118

Figura 32: Avaliação da abundância relativa dos pequenos RNAs derivados de RMRP.....	119
Figura 33: Diagrama de Venn das proteínas detectadas por proteômica nos grupos de pacientes e controles.....	121
Figura 34: Gráfico de dispersão <i>Volcano</i> das proteínas diferencialmente expressas entre pacientes e controles.. ..	122
Figura 35: Agrupamento hierárquico com base na distância Euclidiana das proteínas diferencialmente expressas em CHH.. ..	123
Figura 36: Esquema gerado pelo banco de dados Reactome de processos e vias biológicas envolvidas em CHH.. ..	125
Figura 37: Análise da localização de expressão das proteínas identificadas nos pacientes.	127

LISTA DE ABREVIATURAS

ACN	Acetonitrila
°C	Graus Celsius
cDNA	DNA complementar
CHH	Hipoplasia Cartilagem-Cabelo (do inglês, Cartilage Hair Hypoplasia)
Cm	Centímetro
CT	Limiar do ciclo, do inglês, cycle treshold
dATP	Desoxiadenosina trifosfato
dCTP	Desoxicitosina trifosfato
DC	Defeito congênito
DE	Displasia Esquelética
dGTP	Desoxiguanosina trifosfato
DNA	Ácido desoxirribonucleico
dNTP	Desoxinucleotideo trifosfato
DTT	Ditiotreitol
dTTP	Desoxitimina trifosfato
LD	Desequíbrio de ligação (do inglês, Linkage disequilibrium)
ECLAMC	Estudo Colaborativo Latino Americano de Malformações Congênicas
EDTA	Ácido etilenodiaminotetracético
FA	Ácido fórmico do inglês: formic acid
FDR	Taxa de descoberta falsa; do inglês: "False Discovery rate"
G	Gramas
IFF	Instituto Nacional da Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira

IPA	<i>Ingenuity Pathway Analysis software</i>
LD	Desequilíbrio de ligação (do inglês, linkage disequilibrium)
MC	Malformação congênita
Mg	Miligrama
Nin	Min
mL	Mililitro
mM	Milimolar
Ng	Nanograma
Nm	Nanômetro
NCBI	National Center of Biotechnology Information
Pb	Pares de base
PD	<i>Proteome Discoverer software</i>
PCR	Reação em cadeia pela polimerase
pH	Potencial hidrogeniônico
Pmol	Picomol
<i>RMRP</i>	Componente de RNA da endoribonuclease de processamento de RNA Mitochondrial (do inglês, RNA Component Of Mitochondrial RNA Processing Endoribonuclease)
SDS	Dodecilsulfato de sódio
SNP	Polimorfismo de um único nucleotídeo (do inglês, single nucleotide polymorphism)
TAE	Tampão Tris-Acetato EDTA
TE	Tampão Tris-EDTA
Tris	Tris (hidroximetil) aminometano

U	Unidades
V	Volts
X	Vezes
x g	Vezes a força da gravidade
µg	Micrograma
µL	Microlitro
λ Hind III	DNA Lambda digerido com enzima HIND III
EDTA	Ácido etilenodiamino tetra-acético; do inglês: “Ethylenediamine tetraacetic acid”
miRNA	microRNA
NaCl	Cloreto de sódio
RNA	Ácido ribonucleico; do inglês: “Ribonucleic acid”
TFA	Ácido trifluoroacético; do inglês: “trifluoroacetic acid”

SUMÁRIO

1. INTRODUÇÃO.....	20
2. JUSTIFICATIVA.....	23
3. OBJETIVOS.....	25
3.1. Geral:.....	25
3.2. Específicos:	25
4. REFERENCIAL TEÓRICO	26
4.1 Tecido ósseo e osteogênese.....	26
4.2 Sinalização em condrócitos.....	33
4.3 Displasias esqueléticas	42
4.4 Características clínicas da Hipoplasia Cartilagem-Cabelo	44
4.5 Aspectos moleculares da Hipoplasia Cartilagem-Cabelo.....	47
4.6 Equilíbrio genético e história migratória no Brasil.....	57
5. HIPÓTESE	67
6. METODOLOGIA.....	68
6.1 Desenho do estudo	68
6.2 Local do estudo	69
6.3 Critérios de inclusão	69
6.4 Critérios de exclusão.....	70
6.5 Obtenção de amostras	70
6.6 Identificação de variantes no gene <i>RMRP</i> em amostras de pacientes com suspeita de CHH	70
6.6.1 Isolamento de DNA	70
6.6.2 Amplificação de DNA	71

6.6.3 Sequenciamento	72
6.6.4 Análises de sequência	73
6.6.5 Discriminação alélica	73
6.7 Predição da estrutura secundária dos transcritos RMRP contendo variantes por análises <i>in silico</i>	75
6.8 Investigação do possível efeito fundador da variante g.196.C>T no gene <i>RMRP</i> na população Brasileira de pacientes com CHH	76
6.8.1 Caracterização genética da população estudada	76
6.8.2 Construção de um painel de marcadores adequado para investigação do possível efeito fundador	76
6.8.3 Análise de haplótipos	80
6.9 Análise de expressão gênica por qPCR em tempo real	89
6.9.1 Obtenção dos leucócitos	89
6.9.2 Extração de RNA	90
6.9.3 Tratamento do RNA com DNaseI	91
6.9.4 Análise de mRNA RMRP	91
6.9.5 Análise dos pequenos RNAs derivados do RNA RMRP	93
6.9.6 Análises estatísticas	95
6.10 Análise proteômica de leucócitos provenientes de pacientes	96
6.10.1 Extração e quantificação de proteínas dos leucócitos	96
6.10.2 Preparo das amostras e digestão de proteínas	97
6.10.3 Purificação e secagem dos peptídeos trípticos	98

6.10.4 Espectrometria de massas.....	99
6.10.5 Análise dos dados brutos e identificação das proteínas	100
6.10.6 Análises de resultados de proteoma <i>in silico</i>	100
7. Aspectos Éticos	102
8. Resultados.....	103
8.1 Análise de variantes gene RMRP das amostras de pacientes com suspeita clínica de CHH	103
8.2 Análise estrutural das variantes no gene <i>RMRP</i>	105
8.3 Investigação do efeito fundador da variante g.196C>T no gene <i>RMRP</i> na população brasileira de pacientes com CHH.....	108
8.3.1 Análise do perfil de ancestralidade genética dos pacientes com CHH .	108
8.3.2 Análise de haplótipos	109
8.4 Análise de expressão	117
8.4.1 Expressão do gene <i>RMRP</i>	117
8.4.2 Expressão dos pequenos RNAs derivados dos transcritos RMRP em leucócitos de pacientes com CHH	118
8.5 Análise de proteômica em leucócitos de pacientes com CHH	120
8.5.1 Análises qualitativas e quantitativas.....	120
8.5.2 Análise funcional <i>in silico</i>	124
9. DISCUSSÃO.....	128
10. CONCLUSÃO	145
REFERÊNCIAS	147
11. ANEXOS.....	161

11.1 Aprovação do comitê de ética em pesquisa com seres humanos	161
11.2 Lista de oligonucleotídeos sintéticos	166
11.3 Médias de ancestralidade a nível individual analisando com 3 populações parentais geradas pelo Software Structure	167
11.4 Lista de haplótipos encontrados nos pacientes	168
11.5 Cromatogramas da análise por LC-MS/MS das amostras de pacientes e controles em duplicatas técnicas (A-V)	170
11.6 Proteínas diferencialmente expressas em leucócitos de pacientes com CHH em relação aos controles saudáveis.	178
11.7 Proteínas diferencialmente expressas em leucócitos de pacientes com CHH do presente relacionadas às vias de sinalização	195
11.8 Proteínas diferencialmente expressas em leucócitos de pacientes com CHH do presente estudo também detectadas em ensaio de transcriptoma da literatura (Rogler <i>et al.</i> , 2013).....	198
11.9 Artigo publicado.....	200

1.INTRODUÇÃO

Defeitos Congênitos (DC) referem-se a alterações de origem genética e/ou ambiental na morfogênese humana que ocorrem no período embrionário-fetal, envolvendo comprometimento estrutural, funcional ou ambos¹. O registro de DCs no Brasil vem sendo aprimorado, e tem mostrado as malformações congênitas como a segunda causa de mortalidade infantil. Dados do DATASUS² indicam que os DCs são responsáveis por 21% do total de mortes infantis e esses números demonstram a necessidade de se priorizar essas condições nos programas de Saúde Pública.

O Centro de Genética Médica do Instituto Nacional da Saúde da Mulher da Criança e do Adolescente Fernandes Figueira (IFF) faz o acompanhamento clínico de patologias genéticas raras, de caráter evolutivo, crônico e de alta complexidade médica. Há anos esse serviço vem aprimorando e desenvolvendo metodologias laboratoriais diagnósticas, como técnicas de genética molecular no estudo etiológico dos DCs, especialmente de malformações congênitas (MC), de displasias esqueléticas (DE) e de doenças do desenvolvimento na criança. As DEs, condições geneticamente determinadas e que envolvem o desenvolvimento e crescimento da cartilagem e do tecido ósseo, destacam-se entre os atendimentos. Devido à grande variedade fenotípica, a determinação clínico-radiológica destas doenças pode ser desafiadora e os estudos moleculares, além de auxiliarem no diagnóstico, permitem uma melhor compreensão dos mecanismos envolvidos no desenvolvimento do fenótipo³. Dentro do grupo das DEs, em 2015 o IFF se tornou uma referência nacional para a síndrome Hipoplasia Cartilagem-Cabelo (do inglês, CHH)^{4,5}.

A síndrome Hipoplasia Cartilagem-Cabelo (CHH; OMIM #250250) é uma doença caracterizada por baixa estatura com membros curtos e anormalidades

esqueléticas predominantemente metafisárias⁶. Também pode fazer parte do quadro clínico cabelos finos e esparsos, frouxidão ligamentar, alterações gastrointestinais, deficiência imunológica e até evolução para alguns tipos de câncer^{6,7,8}.

A CHH é causada por mutações recessivas no gene *RMRP*, que não é traduzido em proteína⁶. Este é um gene de expressão ubiquitária que codifica um RNA de 268 nucleotídeos, sendo um componente estrutural do complexo RNase MRP, uma enzima que atua em diversos processos celulares importantes como o processamento do rRNA de 5.8S, a replicação do DNA mitocondrial, a progressão do ciclo celular, entre outros^{9,10}. Tem sido descrito na literatura que o elevado grau da heterogeneidade das mutações em *RMRP* encontradas em pacientes com CHH pode afetar a expressão, a estabilidade e a função do gene *RMRP* em diferentes níveis, podendo relacionar-se com o agravamento do fenótipo da doença¹¹. Além disso, também foi descrita a biogênese de dois pequenos RNAs semelhantes a miRNAs, a partir de transcritos de *RMRP*, que atuam como reguladores de diversos genes relacionados a processos envolvidos em fenótipos dessa síndrome¹².

Em 2015, um estudo colaborativo sobre CHH foi iniciado no Instituto Nacional de Saúde da Mulher, da Criança e do Adolescente Fernandes Figueira^{4,5} com o objetivo de ampliar os conhecimentos acerca desta patologia através da pesquisa e da facilitação de acesso a testes genéticos aos pacientes para confirmação do diagnóstico clínico. Desde então, foram confirmadas alterações no gene *RMRP* em 32 pacientes com CHH, recrutados por três importantes centros de genética médica no Brasil: IFF/Fiocruz (Rio de Janeiro), UNICAMP (Campinas) e HCPA (Porto Alegre). A colaboração, além de fornecer dados precisos para o aconselhamento genético, estabeleceu um fluxo para acompanhamento clínico periódico dos pacientes. Do ponto de vista epidemiológico-molecular, os resultados

obtidos diferenciam os pacientes com CHH brasileiros de outras populações, uma vez que a variante patogênica g.71A>G, frequentemente observada em indivíduos com esta síndrome na literatura¹³, foi encontrada em apenas um indivíduo da nossa casuística. Por outro lado, mais de 60% dos pacientes analisados no presente estudo apresentavam a alteração g.196C>T, sendo uma frequência diferente da observada em outros trabalhos, levando à hipótese da ocorrência de um possível feito fundador para a população brasileira^{4,5}.

Diante destas novas perspectivas de pesquisa, o presente trabalho teve o objetivo de explorar os possíveis mecanismos moleculares da doença; assim como, melhor caracterizá-los nos pacientes brasileiros. Desse modo, a análise da ancestralidade genômica da coorte e a verificação do compartilhamento de haplótipos entre os cromossomos que apresentam variante g.196C>T foram empregadas para investigar o possível efeito fundador desta alteração no Brasil. Complementando o estudo, abordagens *in silico* também foram utilizadas com o intuito de avaliar os possíveis efeitos na estrutura secundária do RNA em decorrência das variantes detectadas. Por fim, foram realizados testes de PCR quantitativo em tempo real para avaliar a expressão do RNA RMRP e de seus pequenos RNAs e análises de proteômica para dissecção do perfil molecular dos pacientes.

2. JUSTIFICATIVA

O Centro de Genética Médica do IFF/Fiocruz é referência no Estado do Rio de Janeiro para o estudo das displasias esqueléticas, incluindo síndrome CHH, o que permite um maior acesso aos pacientes com esta patologia^{4,5,14}. Deve-se ressaltar que os pacientes com CHH podem apresentar comprometimento na linhagem hematopoiética, no sistema imune e no intestino, além de predisposição para alguns tipos de câncer, necessitando um programa de monitoramento clínico contínuo. Além disso, dada a variada manifestação clínica e radiológica, juntamente à sobreposição fenotípica com outras DEs, a análise molecular além de ser o *padrão ouro* para o seu diagnóstico de certeza, também é importante para o aconselhamento genético apropriado dos pacientes e de seus familiares.

A ampla variabilidade fenotípica e genotípica dificulta a compreensão do efeito das variantes no curso da doença. Nesse sentido, a compreensão dos mecanismos fisiopatológicos que levam ao desenvolvimento de CHH torna-se um importante campo de pesquisa a partir do gene *RMRP*. Somado a isto, o estudo publicado em 2014 por Rogler e colaboradores¹², que descreveu dois pequenos RNAs derivados do transcrito RMRP responsáveis pela regulação da expressão de diversos genes relacionados a fenótipos associados à esta síndrome, introduziu uma abordagem molecular adicional ao tema.

Em estudo prévio, nosso grupo estabeleceu uma coorte com 23 pacientes com CHH, em que foram identificadas 12 variantes novas⁵. Desta forma, a análise *in silico* para investigar o efeito das variantes na estrutura secundária do RNA, faz-se uma ferramenta complementar à avaliação da patogenicidade de variantes. Paralelamente, a investigação do padrão de

expressão do RNA RMRP e dos pequenos RNAs derivados desse transcrito, na ampla gama de perfis genéticos observados nos pacientes, em adição à abordagem proteômica, até então não explorada nessa patologia, são técnicas com potencial de ampliar a compreensão sobre os mecanismos de desenvolvimento da doença.

Os resultados anteriores da nossa coorte mostraram uma elevada frequência alélica da alteração g.196C>T (60,9%), diferentemente do observado em outras populações de CHH, apontando para um possível efeito fundador na população brasileira. Algumas variantes fundadoras já foram descritas para CHH, como a substituição, g.71A>G, que ocorre em 92% dos pacientes finlandeses e 48% dos pacientes de outras regiões da Europa e as variantes g.219A>G e uma duplicação de 17 pb na região promotora do gene *RMRP* em pacientes japoneses^{13,15}. Nesse sentido, a investigação da ancestralidade genética deste grupo tornou-se relevante para a compreensão da elevada prevalência da variante g.196C>T em pacientes com CHH no Brasil.

Vale ressaltar que este projeto contempla parcerias nacionais com o Hospital de Clínicas de Porto Alegre (HCPA), a Universidade Estadual de Campinas (UNICAMP), o Laboratório de Biologia Molecular e Proteômica do Sangue (UFRJ), os Laboratórios de Hanseníase e juntamente ao Programa de Computação Científica (PROCC), ambos da FIOCRUZ. Assim, a expertise dos pesquisadores envolvidos neste projeto colaborativo também contribuiu para propiciar intercâmbio científico, importante para a formação de recursos humanos.

3. OBJETIVOS

3.1. Geral:

Investigar o possível efeito fundador e a ancestralidade genética da alteração g.196C>T em pacientes com síndrome Hipoplasia Cartilagem-Cabelo, bem como avaliar o efeito das variantes previamente identificadas no nível de expressão do transcrito de RMRP, dos pequenos RNAs e das proteínas alvos; e por fim, na estrutura secundária do RNA RMRP.

3.2: Específicos:

- Descrever a presença de variantes patogênicas em pacientes com CHH;
- Realizar análises de bioinformática para avaliar o efeito das variantes identificadas nos pacientes na estrutura secundária do RNA;
- Inferir a ancestralidade genômica individual de pacientes com CHH na população brasileira através dos marcadores informativos de ancestralidade do tipo indels;
- Construir os haplótipos dos pacientes incluídos neste estudo, baseado nos dados do sequenciamento dos marcadores selecionados;
- Comparar a distribuição dos diferentes tipos de haplótipos entre grupos de variantes identificadas;
- Avaliar os níveis de expressão do gene *RMRP* e dos pequenos RNAs derivados de transcritos de RMRP por RT-qPCR em leucócitos de pacientes com CHH.
- Descrever o perfil de expressão de proteínas provenientes de leucócitos de pacientes com CHH através da abordagem proteômica.

4. REFERENCIAL TEÓRICO

4.1 Tecido ósseo e osteogênese

O corpo humano apresenta mais de 200 ossos e juntamente com cartilagens, ligamentos e tendões formam o sistema esquelético, responsável por diversas funções no organismo¹⁶. A principal função do sistema ósseo é auxiliar a sustentação do corpo e a proteção dos órgãos internos, além de participar da realização dos movimentos corporais em coordenação com os músculos e articulações. Outra função importante é a de armazenamento de sais minerais, como cálcio e fósforo¹⁷.

A formação óssea, ou osteogênese, é um processo finamente regulado de forma espaço-temporal, envolvendo a atividade de diversos genes^{18,19}. Esse processo se inicia no feto no final da terceira semana da gestação, quando o mesoderma paraxial se torna segmentado, formando os somitos, e é finalizado por volta dos 20 anos, quando cessa o crescimento ósseo²⁰. No embrião, o esqueleto é gerado por três linhagens distintas: 1) esclerótomo (figura 1), que é derivado dos somitos e são responsáveis pelo desenvolvimento do esqueleto axial; 2) crista neural cranial (figura 1), que dá origem ao arco faríngeo e aos ossos e cartilagem craniofaciais e, 3) mesoderma lateral, que origina os ossos e cartilagem dos membros^{21,22}.

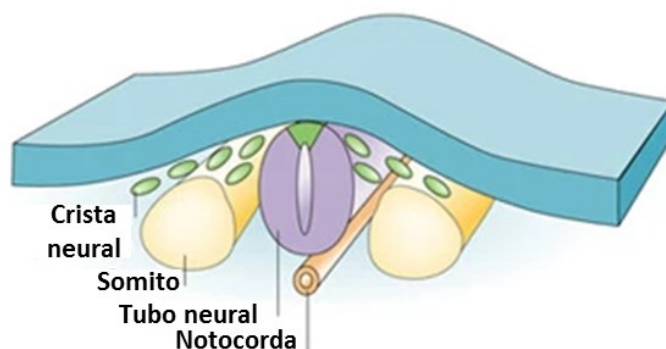


Figura 1: Linhagens do embrião durante a neurulação. A imagem mostra o tubo neural, os somitos e as células da crista neural que darão origem à maioria dos ossos do esqueleto. Figura adaptada de Gammill, L. and Bronner-Fraser, 2003.

Morfologicamente, os ossos podem ser classificados em quatro tipos principais: longos, curtos, planos e irregulares e suas formações podem ocorrer através de dois mecanismos: a ossificação intramembranosa e a ossificação endocondral¹⁷. No primeiro, ocorre a conversão direta do tecido conjuntivo primitivo em osso. No segundo, as células mesenquimais se diferenciam em cartilagem, formando um molde cartilaginoso, que é progressivamente substituído por células ósseas¹⁸.

Ossificação Intramembranosa

Através da ossificação intramembranosa são formados alguns dos ossos do crânio e de algumas partes das clavículas. Esta também contribui para o crescimento de ossos curtos e para o espessamento dos ossos longos. Nesse processo de ossificação o tecido ósseo não é depositado sobre um intermediário cartilaginoso^{18,23}.

A ossificação intramembranosa começa no centro de ossificação primária, no interior de membranas de tecido conjuntivo. Primeiramente, ocorre a diferenciação de células mesenquimais em osteoblastos agrupados que sintetizam uma matriz ainda não mineralizada, rica em colágeno tipo I, chamada

osteóide (figura 1A). Com a progressiva deposição de fosfato e cálcio nessa matriz, alguns osteoblastos ficam embebidos nesta, tornando-se osteócitos (figura 2)^{18,20,23}.

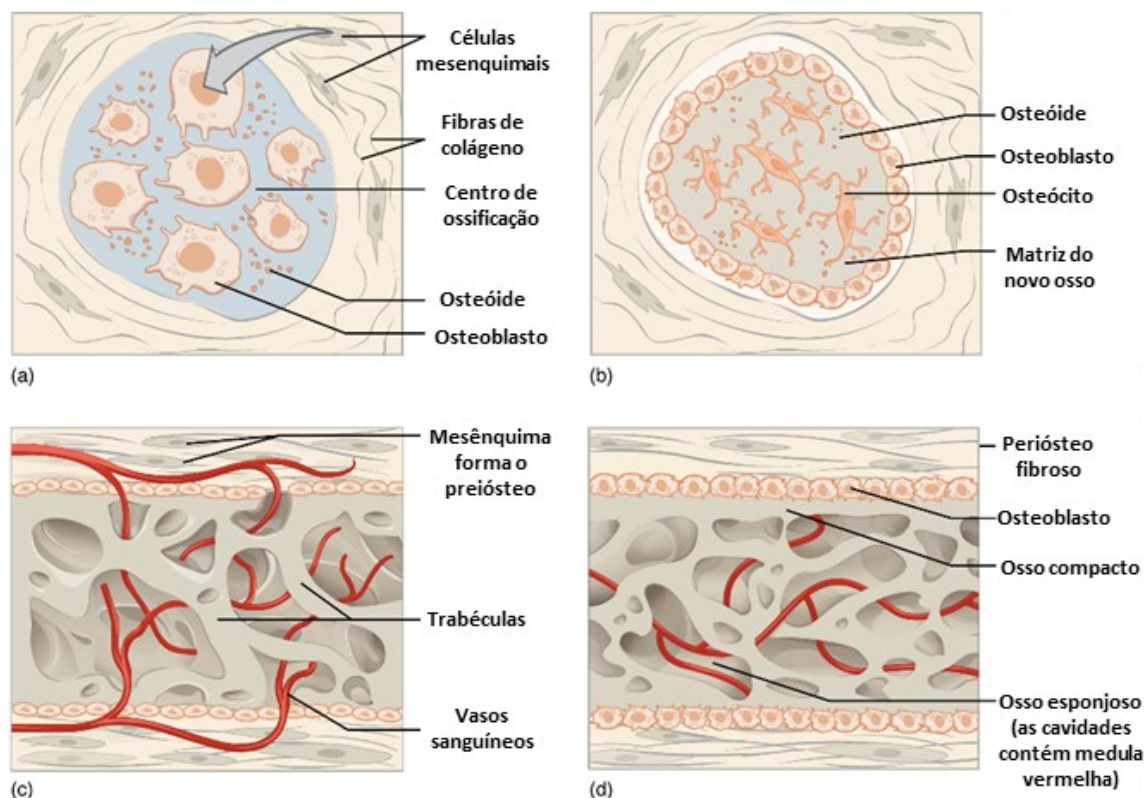


Figura 2: Ossificação intramembranosa. A ossificação intramembranosa segue quatro etapas: (a) as células mesenquimais se reúnem em aglomerados, diferenciam-se em osteoblastos e formam centros de ossificação. (b) O osteóide secretado retém os osteoblastos, que posteriormente se tornam osteócitos. (c) Matriz trabecular e formação de perióstio. (d) O osso compacto se desenvolve superficialmente ao osso trabecular e os vasos sanguíneos agrupados se condensam na medula óssea vermelha²⁴.

Durante a diferenciação das células-tronco mesenquimais em osteoblastos, a sinalização de Wnt é responsável pelo aumento dos níveis de β -catenina nas células tronco-mesenquimais, levando a expressão de *Runx2*. Este é um dos mais importantes fatores de transcrição atuantes neste processo e induz a posterior expressão do fator de transcrição *Osterix*, também envolvido na diferenciação osteoblástica²³.

Alguns osteoblastos permanecem nas periferias e continuam formando placas de osso compacto nas superfícies. No centro de ossificação é possível observar que vários desses grupos surgem simultaneamente. Entre as placas ósseas da superfície que se confluem, o osso permanece esponjoso e esse aspecto é acentuado pela ação dos osteoclastos que reabsorvem o osso. As cavidades formadas entre as trabéculas são invadidas por vasos sanguíneos e células mesenquimais indiferenciadas, formando assim a medula óssea. O crescimento radial de vários centros de ossificação leva à substituição gradual da membrana conjuntiva preexistente por tecido ósseo e parte dessa membrana permanece inalterada, constituindo o endóstio e o perióstio (Figura 2C e 2D)^{18,22}.

Ossificação Endocondral

A ossificação endocondral é o mecanismo pelo qual cerca de 80% dos ossos de mamíferos são formados, incluindo ossos longos e curtos, além de alguns ossos da base do crânio²⁴. Esse mecanismo é caracterizado pela deposição de tecido ósseo sobre um molde de cartilagem hialina, que tem formato similar ao produto final, porém com tamanho reduzido (figura 3)¹⁹.

Em um osso longo, por exemplo, a endocondrogênese se inicia no centro de ossificação primário na diáfise (parte do osso longo entre suas extremidades, chamadas epífises). Inicialmente ocorre a condensação de células mesenquimais, que em seguida diferenciam-se em condrócitos²⁵. Uma matriz rica em colágeno tipo II é secretada por essas células para formar o modelo de cartilagem. A cartilagem, por ser um tecido avascular, precisa receber nutrientes e remover os resíduos metabólicos por difusão, através dos vasos do pericôndrio, uma membrana circundante que cobre a cartilagem²³.

À medida que os condrócitos sofrem hipertrofia, é liberado colágeno tipo X, controlando a mineralização da matriz circundante e favorecendo a formação de trabéculas. A condição de hipóxia gerada estimula a secreção de fator de crescimento endotelial vascular que atrai vasos sanguíneos²⁵. Os condrócitos sofrem apoptose e os vasos sanguíneos invadem essas cavidades levando osteoclastos e osteoblastos, que estimulam remoção da cartilagem calcificada e sua substituição por osso trabecular e medula óssea. Os osteoblastos depositam a matriz óssea sobre os tabiques de cartilagem calcificada, formando espículas ósseas com uma parte central de cartilagem calcificada e uma parte superficial de tecido ósseo primário. Os osteoclastos digerem a matriz da cartilagem e pelo processo de reabsorção óssea criam uma cavidade longitudinal, formando o canal medular. As células mesenquimais que chegam a este canal formam a medula óssea hematogênica e dão início a produção de células sanguíneas. Após a ossificação da diáfise, o processo se expande em direção às epífises^{18,20,23}.

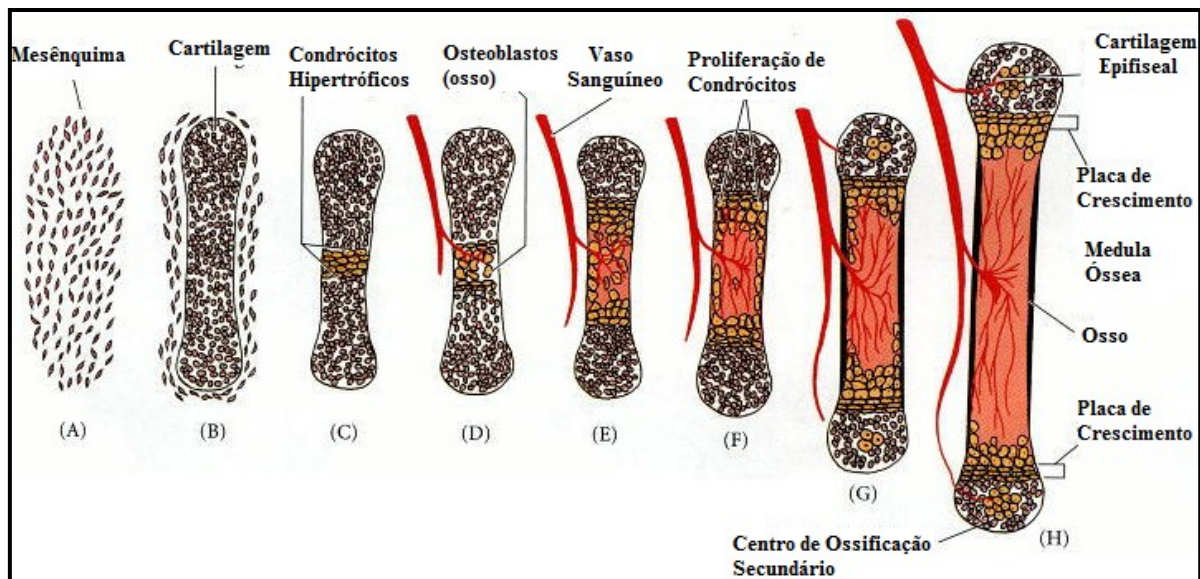


Figura 3: Esquema da ossificação endocondral. (A, B) Células mesenquimais condensam e se diferenciam em condrócitos para formar o molde cartilaginoso dos ossos. (C) Condrócitos no centro do eixo sofrem hipertrofia. Ocorre a redução e mineralização da matriz extracelular e morte por apoptose dos condrócitos. A morte dos condrócitos permite a entrada de vasos sanguíneos. (D, E) Os vasos sanguíneos trazem osteoblastos, que se ligam à matriz cartilaginosa em degeneração e ao depósito de matriz óssea. (F, G) Formação da medula óssea. (G, H) Crescimento ósseo a partir da placa de crescimento. Centros de ossificação secundários também são formados quando os vasos sanguíneos entram nas pontas dos ossos¹⁸.

Ao final da deposição de matriz óssea nas epífises, o tecido cartilaginoso existirá apenas nos pontos de articulação e na placa de crescimento, que é responsável pelo crescimento longitudinal dos ossos longos. Ela está localizada entre as epífises e a diáfise e é um disco cartilaginoso que não é penetrado pelo osso em expansão durante sua formação. Nela, os condrócitos se distribuem em cinco zonas distintas a partir da epífise: zona de repouso, zona de proliferação, zona de cartilagem hipertrófica, zona de cartilagem calcificada e zona de ossificação (figura 4)¹⁸.

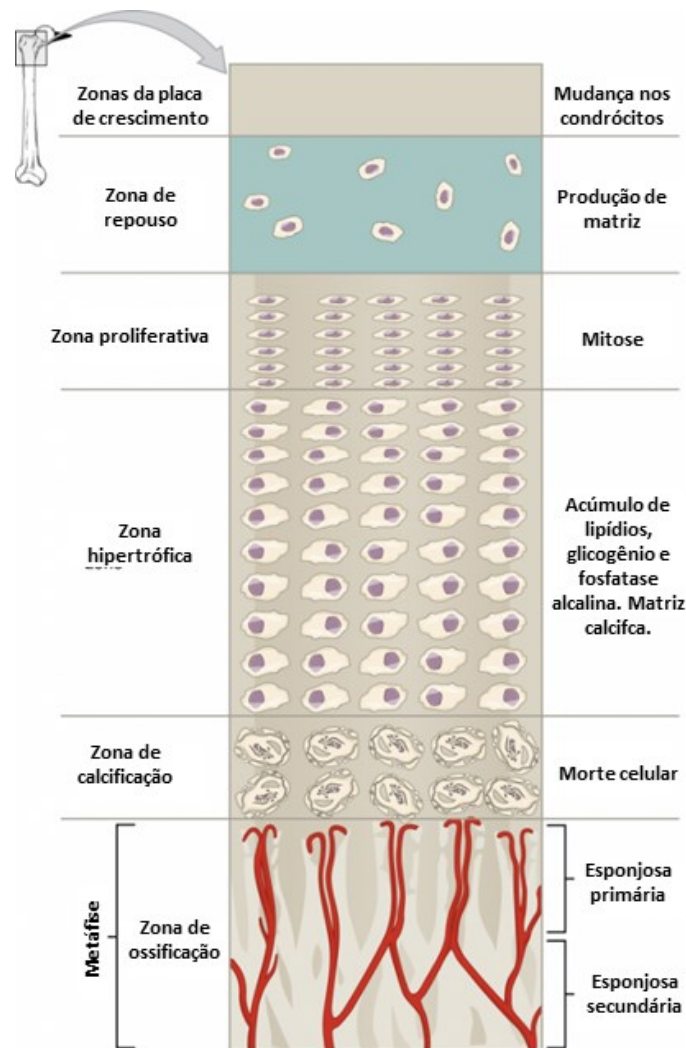


Figura 4: Esquema da placa de crescimento. A figura evidencia as zonas da distribuição dos condrócitos a partir da epífise: zona germinativa, zona proliferativa, zona hipertrófica e zona de ossificação endocondral²⁴.

Uma região da placa de crescimento próximo à epífise mantém a cartilagem hialina sem qualquer alteração em sua morfologia, sendo essa região correspondente à zona de repouso ou zona germinativa. Adjacente a esta zona, os condrócitos dividem-se rapidamente e formam fileiras paralelas no sentido longitudinal do osso, formando a zona proliferativa. Ocorre então a modificação da cartilagem hialina, havendo hipertrofia dos condrócitos e redução da matriz cartilaginosa a finos tabiques, sendo esta a zona de cartilagem hipertrófica. Em seguida, a matriz cartilaginosa é mineralizada e os condrócitos sofrem morte por apoptose¹⁸.

A largura da placa de crescimento é equilibrada pela proliferação e degeneração dos condrócitos, e substituição da cartilagem calcificada por osso na placa de crescimento. Esse processo também resulta no crescimento do comprimento ósseo longitudinal que continua durante a infância e a adolescência. No final da puberdade, a proliferação de condrócitos cessa e a cartilagem hialina remanescente será removida e substituída por osso, tornando-se uma linha epifisária^{23,24}.

A gênese e remodelamento do tecido ósseo por ossificação endocondral são processos sutilmente controlados por várias vias de sinalização²⁶.

4.2 Sinalização em condrócitos

O complexo sistema de regulação das etapas sucessivas de maturação dos condrócitos na placa de crescimento (repouso, proliferativa, pré-hipertrófica e hipertrófica) depende de diversas respostas celulares coordenadas espaço-temporalmente. Através da integração de múltiplas cascatas de sinalização como WNT, NOTCH, FGF, IHH, PTHrP, TGF- β , BMP e Rho GTPases^{27,28}. Embora avanços significativos sobre a compressão dos processos reguladores moleculares na placa de crescimento tenham sido alcançados, esse tópico ainda permanece não completamente esclarecido. A seguir serão abordadas as principais sinalizações que ocorrem em condrócitos.

PTHrP: a proteína relacionada ao hormônio da paratireóide (PTHrP) é um fator de crescimento parácrino importante na regulação da função dos condrócitos (figura 5)²⁷. Sua sinalização é induzida por IHH e previne a diferenciação hipertrófica prematura. O mecanismo de *feedback* negativo entre essas duas sinalizações preserva os condrócitos no estado proliferativo, controla

a proliferação de condrócitos e mantém o comprimento das colunas de células na placa de crescimento²⁹. O PTHrP liga-se ao receptor específico PTHR1 que é acoplado a uma proteína G e a sinalização pode seguir por dois caminhos antagônicos na diferenciação de condrócitos: ativando a proteína cinase A e C (PKA e PKC) ou interagindo com a sinalização PLC/IP3, promovendo o aumento de Ca²⁺ intracelular. No primeiro, a ligação de PTH ao seu receptor mantém os condrócitos em estado proliferativo. Essa sinalização também impede a diferenciação dos condrócitos através da atuação sobre fatores de transcrição chave desse processo. Por um lado, SOX9 é ativado e por outro, o potenciador de miócitos 2 (MEF2) e RUNX2 são inibidos por PKA em decorrência do recrutamento da histona desacetilase HDAC4 classe II no núcleo. Já na segunda, a proliferação dos condrócitos é inibida e a progressão para a diferenciação hipertrófica é estimulada pela sinalização de PLC/IP3 através de PTHrP^{27,30}.

Ihh: IHH é um ligante membro da família de proteínas hedgehog de mamíferos composta por: Indian (IHH), Sonic (SHH) e Desert Hedgehog (DHH)³¹. A via de Ihh está associada à fase de proliferação e maturação de condrócitos (figura 5), e é regulada positivamente pelo fator de transcrição RUNX2³¹. Sua sinalização é iniciada pela interação de IHH com os receptores transmembrana PTCH 1 e 2, o que promove a ativação da proteína de ação intracelular SMO, levando à transcrição por Gli (GLI 1, 2 e 3)²⁷. A sua regulação pode ocorrer de forma direta, promovendo a proliferação de condrócitos pela inibição da atividade de GLI3 ou indiretamente, através da alça de retroalimentação negativa com PTHrP, controlando a hipertrofia dessas células²⁷.

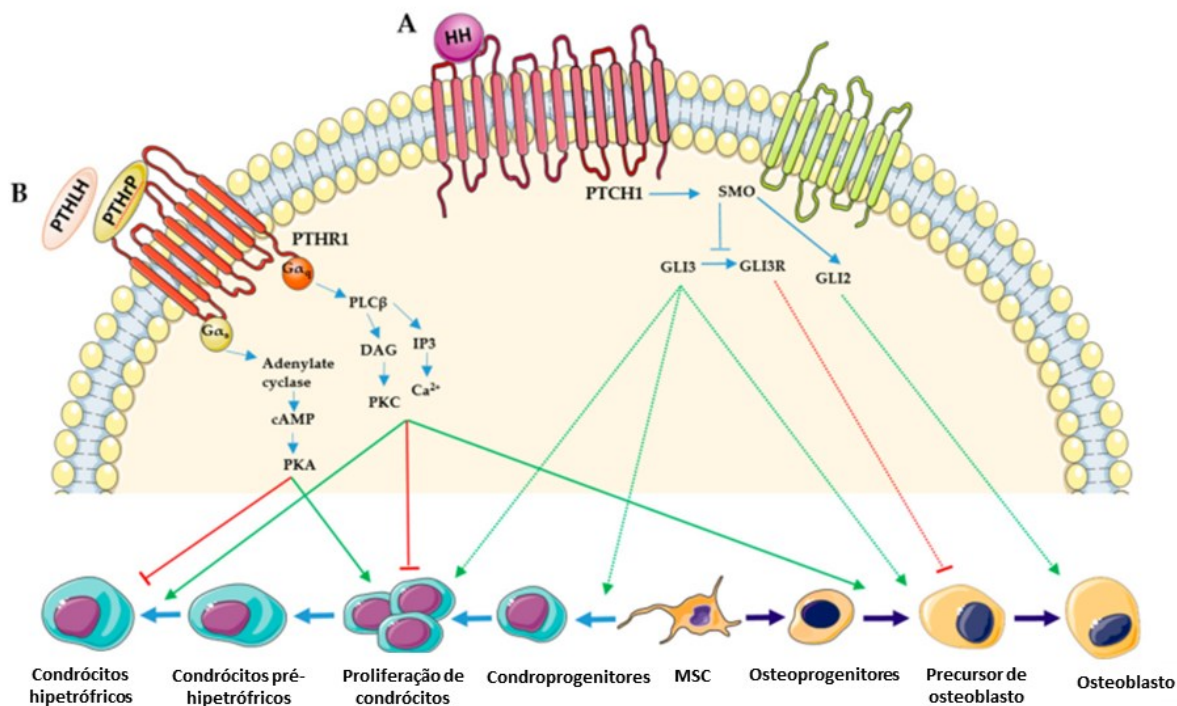


Figura 5: Vias de sinalização de HH e PTHrP na diferenciação de condrócitos e osteoblastos **A)** Os ligantes HH interagem com o receptor PTCH 1 causando a ativação da proteína SMO e, conseqüentemente, a ativação dos fatores de transcrição GLI. A sinalização HH regula positivamente os primeiros passos da condrogênese a partir de células-tronco mesenquimais (do inglês, *mesenchymal stem cells* - MSC) e proliferação de condrócitos. Ele induz a especificação de precursores de osteoblastos e a maturação em osteoblastos maduros. A ativação do repressor GLI3 (GLI3R) inibe a proliferação de precursores de osteoblastos. **B)** PTHrP ou PTHLH se ligam a PTHR1 ativando proteínas $G_{\alpha s}$ e $G_{\alpha q}$. Por meio de $G_{\alpha s}$, o PTHrP ativa a adenilato ciclase e a produção de AMPc, que por sua vez ativa proteínas PKA induzindo a proliferação de condrócitos e inibindo sua diferenciação hipertrofica. No entanto, através do $G_{\alpha q}$, o PTHrP ativa o PLC β (fosfolipase β) e a produção de DAG (diacil glicerol) e IP3 (inositol trifosfato), levando à ativação de PKC (proteína cinase C) e a um aumento de Ca^{2+} , respectivamente. A sinalização PLC β inibe a proliferação de condrócitos e induz sua hipertrofia e diferenciação osteoblástica²⁷.

WNT: A rede de sinalização intracelular de Wnt está presente em vários contextos do desenvolvimento dos condrócitos (figura 6)^{29,32}. A família de ligantes WNT é composta por 19 membros que se ligam ao receptor transmembrana Frizzled (FZD), processo facilitado pelos co-receptores, LRP ou ROR, ativando então uma cascata de sinalização^{27,33}. A via de Wnt pode ocorrer por três vias, sendo uma via canônica (WNT/ β -catenina) e duas não canônicas (Polaridade Celular Planar - PCP e a Wnt/ Ca^{2+})³³. A sinalização canônica de Wnt é bem caracterizada, na qual, em resposta aos agonistas extracelulares, ocorre

a translocação de β -catenina para o núcleo, que se liga aos fatores de transcrição TCF/LEF, ativando genes-alvo que controlam a proliferação e diferenciação celular. Na ausência de WNT, o complexo composto por GSK3 β , axina e APC é responsável por fosforilar a β -catenina, sinalizando para o proteossoma promover a sua degradação. A via não canônica PCP, estimulada pela ligação de WNT, sinaliza para organização do citoesqueleto ou ativação da transcrição por c-jun e AP1. A cascata envolve a ativação de pequenas GTPases, como RAC, Cdc42 e Rho, por meio de da interação com Dishevel (DVL)^{28,34}. Finalmente, na via não canônica Wnt/Ca²⁺, ocorre liberação intracelular de Ca²⁺ e a ativação de algumas enzimas sensíveis ao cálcio, após a ligação de WNT aos receptores FZD e ROR. Esse evento induz a expressão gênica através da ativação de fatores de transcrição como o NF-AT, CREB e NF-KB^{27,34}. A sinalização canônica de Wnt e conseqüentemente a ativação de β -catenina, promove hipertrofia de condrócitos. Por outro lado, a inativação da β -catenina causa um aumento da expressão de SOX9, levando a uma diminuição da proliferação de condrócitos e uma maturação hipertrófica atrasada^{27,35}. WNT também pode promover hipertrofia de condrócitos pela inibição da atividade de PTHrP. Além disso, estudos relatam que a via não canônica Wnt/PCP desempenha um papel importante na formação da coluna de condrócitos da placa de crescimento³¹.

NOTCH: o papel da sinalização Notch nas placas de crescimento demonstra-se bastante complexo, apresentando expressão diferencial de seus integrantes tanto no estágio inicial, quanto tardio da condrogênese (figura 6)^{31,32}. Essa via é altamente conservada e envolve um dos quatro receptores transmembrana NOTCH1-4 e ligantes como JAG1 e JAG2; DLL1, DLL3 e

DLL4^{27,31}. A via canônica de Notch modula a sinalização pela comunicação célula-célula²⁸. Nela, o ligante interage com o receptor NOTCH, que sofre clivagens proteolíticas sequenciais, iniciando pelo domínio extracelular, para liberar o domínio intracelular de NOTCH (NICD). Este é translocado para o núcleo, onde se liga com a proteína RBPjk, ativando a expressão de genes a jusante, entre eles fatores de transcrição das famílias HES e da família HEY, envolvidos em muitos processos celulares, incluindo proliferação e diferenciação celular^{33,36,37}. A sinalização de Notch é modulada pelo contexto, sendo mais intensa durante o início da condrogênese e diminuída com a progressão da diferenciação dos condrócitos. Por fim, quando estas células iniciam a fase de hipertrofia, Notch novamente potencializa esse processo através da regulação da expressão de SOX9³⁶.

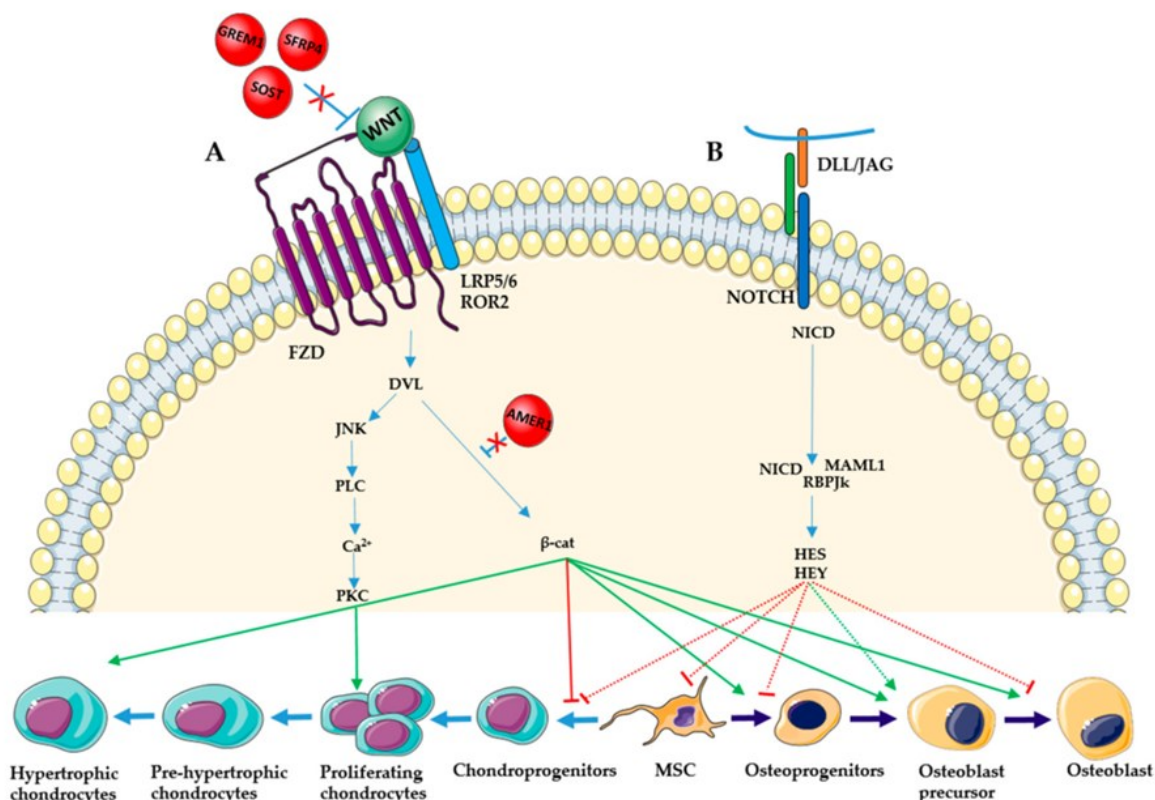


Figura 6: Vias de sinalização Wnt e Notch na diferenciação de condrocitos e osteoblastos.

A) os ligantes WNT se ligam aos receptores FZD com ou sem co-receptores específicos como LRP5, LRP6 e ROR2. Na via canônica, a ligação de WNT ativa a translocação de β -catenina no núcleo. Isso inibe a primeira etapa da condrogênese, mas induz a diferenciação hipertrofica dos condrocitos e a diferenciação e maturação dos osteoblastos. As vias não canônicas, através a ativação das proteínas DVL, JNK e PLC, ativam PKC e induzem a proliferação de condrocitos. A sinalização WNT pode ser regulada negativamente por inibidores extracelulares, como SOST, GREM1 e SFRP4, e por inibidores intracelulares, como AMER1. **B)** nas vias NOTCH, DLL ou JAG, ligantes ligam-se ao receptor NOTCH e causam sucessivas clivagens proteolíticas, resultando na liberação de NICD que se translocam no núcleo e interagem com as proteínas MAML1 e RBPJk. A sinalização inibe a condrogênese e o primeiro e o último estágio da diferenciação dos osteoblastos, mas promove a diferenciação de osteoprogenitores em precursores de osteoblastos²⁷.

FGF: Os fatores de crescimento de fibroblastos (FGFs) são uma grande família de mais de 18 proteínas que se ligam a quatro receptores de FGF do tipo tirosina cinase (FGFR 1-4)²⁷. A ligação dos FGFs aos seus receptores induz a homodimerização desta proteína e os domínios tirosina cinase dos FGFRs são fosforilados de forma cruzada. Em consequência, a proteína adaptadora FRS2 α também é fosforilada e ocorre o recrutamento subsequente de outras proteínas do mesmo tipo como STAT1, PLC γ e GRB2, que podem interagir com múltiplas proteínas envolvidas em vias de sinalizações, como JAK-STAT, PKC,

MAPK e PI3K, que regulam diversos processos do desenvolvimento dos condrócitos²⁷ (figura 7).

Todas as fases do desenvolvimento ósseo contam com a expressão dos FGFs e de seus receptores e sua modulação ocorre de forma espaço-temporal²⁷. Os ligantes FGFs 1, 2, 5, 8–14, 16–19, 21 e 23 já foram detectados na placa de crescimento. Tratando de receptores, a expressão de FGFR1 ocorre em condrócitos pré-hipertróficos e hipertróficos. Já FGFR2 foi observado como o primeiro marcador de condensação pré-condrogênica, sendo também detectado na zona de repouso da placa de crescimento. FGFR3 é expresso em todos os tipos de condrócitos, exceto nos hipertróficos. Por fim, FGFR4 está presente em baixas quantidades nas zonas de repouso e proliferativas das placas de crescimento³⁹.

TGF- β e BMP: A sinalização do fator de crescimento transformador β (TGF- β) e da proteína morfogenética óssea (BMP) também desempenham um papel relevante em todas as etapas da condrogênese embrionária (figura 7). As proteínas envolvidas nessas vias constituem uma superfamília altamente conservada de mais de 40 membros^{31,32}. Em ambas as sinalizações, após a combinação dos ligantes da superfamília TGF- β a um complexo formado por dois receptores tipos I (TGFBR1, BMPRI ou ALK) e dois receptores do tipo II (TGFBR2 ou BMPRII), ocorre a autofosforilação dos receptores tipo II. Em seguida, o receptor I e a proteína R-SMAD (SMAD4) também são fosforilados, e o sinal subsequente induz formação de um outro complexo das proteínas R-SMAD com Co-SMAD (SMAD2 e 3 majoritariamente relacionadas à sinalização por TGF- β e SMAD1, 5 e 8 à sinalização por BMP). Este é então acumulado no núcleo, desencadeando uma série de reações relacionadas a transcrição gênica, como

a regulação de *SOX9* e *RUNX2*, que participam do controle da proliferação e diferenciação de condrócitos^{31,33}. A fosforilação de TGF- β e BMP também pode promover a ativação não canônica da cascata MAPK. Essa via não depende de Smad e pode ativar as proteínas JNK e p38, que podem acelerar a diferenciação terminal dos condrócitos pela alteração do equilíbrio entre a sinalização de TGF- β BMP^{27,32,33}.

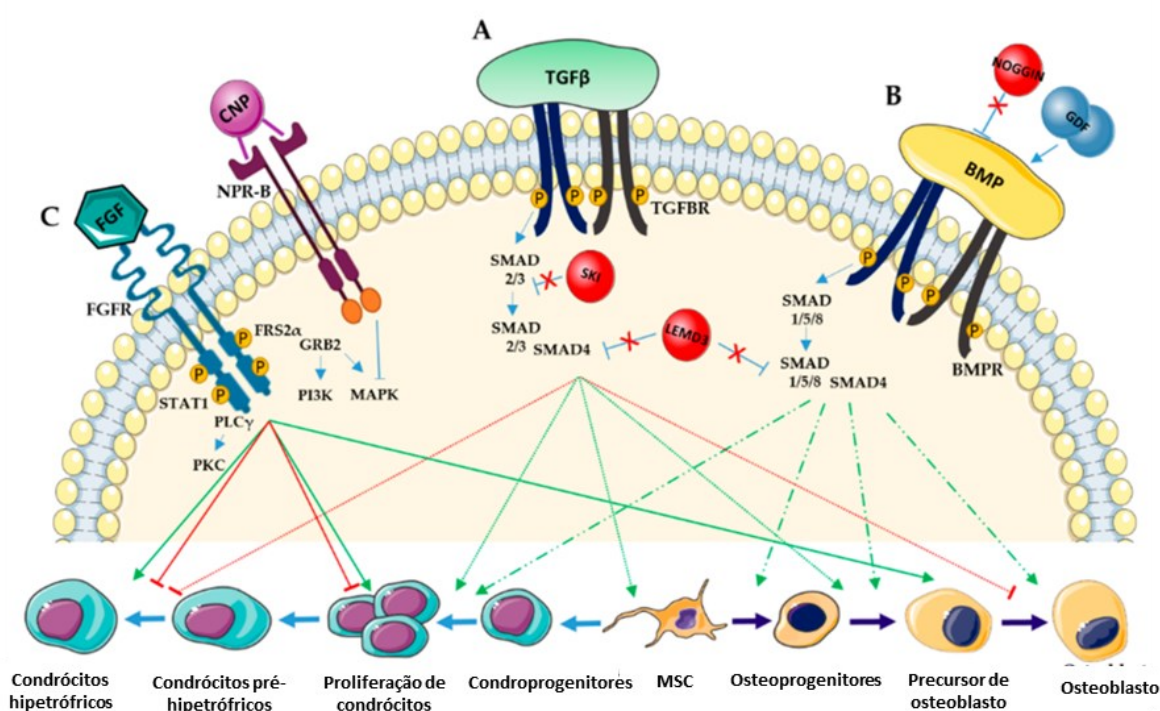


Figura 7: Vias de sinalização de TGF β , BMP e FGF na diferenciação de condrócitos e osteoblastos. **A)** Ligantes de TGF- β ligam-se a TGFBR tipo 1 e 2 fosforilando os receptores e proteínas SMAD 2 e 3 que interagem com SMAD4. Este complexo se acumula no núcleo e induz condrogênese e osteoblastogênese, mas inibe a última etapa de condrócitos e maturação dos osteoblastos. A atividade SMAD pode ser bloqueada por inibidores intracelulares como SKI e LEMD3. **B)** ligantes BMP ligam-se a BMPR tipo 1 e 2 causando a fosforilação dos receptores e de proteínas SMAD 1, 5 e 8 que interagem com SMAD4. Esse complexo se acumula no núcleo e induz a proliferação de condrócitos e todas as etapas da diferenciação dos osteoblastos. Ligantes GDF e o antagonista NOGGIN também se ligam ao BMPR. **C)** os ligantes do FGF ligam-se ao receptor FGFR promovendo a transfosforilação e a fosforilação da proteína adaptadora FRS2 α . Ocorre então a ativação de STAT1, PLC γ e GRB2, e este último, por sua vez, ativa as proteínas PI3K e MAPK. Por essas vias, o FGF promove a proliferação e diferenciação de condrócitos na primeira etapa do desenvolvimento e promove a proliferação e diferenciação de osteoblastos. Em contraste, nas fases posteriores de desenvolvimento, inibe a proliferação e diferenciação de condrócitos. A sinalização MAPK, a jusante do FGFR, pode ser regulada negativamente pela via CNP-NPR²⁷.

Rho GTPases: Pequenas GTPases da família Rho também são conhecidas como proteínas essenciais para o desenvolvimento da cartilagem e do osso³⁹. A modulação da sinalização de RhoA e Rac1/Cdc42 demonstram papéis antagônicos no controle da proliferação, hipertrofia e apoptose de condrócitos. A sinalização Rac1 e Cdc42 apresentam papéis importantes na diferenciação hipertrófica de condrócitos e RhoA, por sua vez, atua na proliferação e supressão da hipertrofia de condrócitos^{40,41}. Em estudos prévios, a supressão de Rac1 em células ectodérmicas do broto do membro, mostram uma disrupção da via de Wnt e um grave truncamento dos membros. Da mesma forma, o silenciamento de Cdc42 no mesênquima resultou em membros mais curtos e várias anormalidades esqueléticas e a placa de crescimento apresentava uma cartilagem hipertrófica mais larga e a redução na expressão de *COL0* e *MMP13*, sugerindo inibição da diferenciação de condrócitos. Em queratinócitos, observou-se retardo de crescimento, perda de pelos na 4ª semana pós-natal, proliferação reduzida e mudança de destino celular⁴².

Em resumo o processo de esqueletogênese decorre do equilíbrio entre vários processos envolvendo diferentes células, genes e vias de sinalização (figura 8).

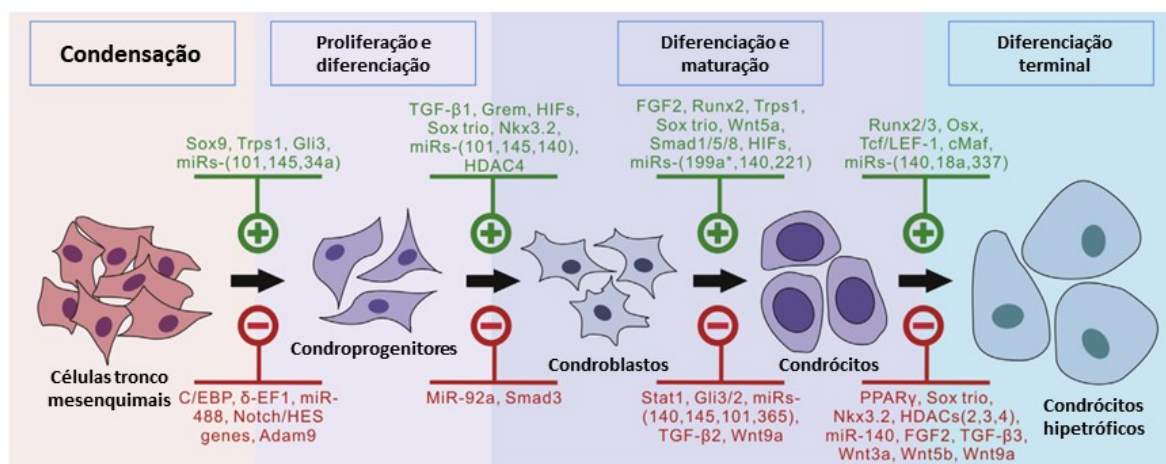


Figura 8: Reguladores da diferenciação condrogênica de MSCs identificados até o momento. Moléculas sinalizadoras, fatores de transcrição e outros reguladores são listados. Os reguladores positivos são mostrados no painel superior em verde, enquanto os reguladores inibitórios são mostrados em vermelho no painel inferior³². (Multifaceted signaling regulators of chondrogenesis: Implications in cartilage regeneration and tissue engineering).

A desregulação em qualquer etapa desse processo pode levar a fenótipos compatíveis com as displasias²⁶.

4.3 Displasias esqueléticas

Displasias esqueléticas (DEs), ou osteocondrodismplasias, integram um grupo de doenças genéticas e clinicamente heterogêneas, decorrentes de anomalias nos processos de desenvolvimento, crescimento e homeostase da cartilagem e do esqueleto^{27,44}. A prevalência global das DEs é estimada em 1/5.000 nascimentos, embora individualmente cada tipo de doença óssea seja considerado raro²⁷. Muitas dessas condições têm início no período pré-natal e podem ser identificadas antes do nascimento através da avaliação de ultrassonografia de rotina⁴⁴.

Algumas das características das displasias esqueléticas incluem encurtamento nos ossos das pernas e/ou braços, ossos curvados ou fraturados, tórax pequeno, encurtamento das costelas, crescimento ósseo assimétrico, duplicação dos dedos das mãos ou dos pés, ossos irregulares (espessos ou finos) e desmineralização⁴⁵. A gravidade da manifestação clínica pode variar de leve,

em que pode haver apenas um comprometimento ósseo isolado e sem associação a demais sintomas clínicos, ou até mesmo uma apresentação letal da doença. A diversidade clínica das DEs evidencia a complexidade dos mecanismos etiopatogênicos dessa classe de doenças²⁶.

De acordo com a versão do ano de 2019 da classificação de displasias esqueléticas, existem 461 patologias classificadas em 42 grupos seguindo critérios clínicos, radiológicos, bioquímicos moleculares. A maioria dessas condições (92%) tem base genética determinada. Até o momento, 437 genes já foram associados com essas condições e apenas 36 das 461 permanecem sem definição etiológica conhecida. Com a crescente aplicação do sequenciamento de nova geração (NGS) para fins diagnósticos, muitas das causas genéticas das displasias esqueléticas puderam ser elucidadas ou até reclassificadas³.

É interessante observar a variedade de genes causadores das DEs, que podem desencadear o fenótipo por diferentes mecanismos moleculares. São genes que atuam em diversos processos biológicos, em diferentes tecidos, e que produzem diferentes classes de produtos gênicos com diferentes funções no organismo. Até mesmo a localização das alterações em um mesmo gene pode levar a diferentes fenótipos, como observado nas doenças relacionadas ao gene *FGFR3*, onde são relatados 6 tipos de condrodisplasias (Displasia Tanatofórica tipos I e II, SADDAN, Acondroplasia, Hipocondroplasia, CATSHL), 2 tipos de craniossinostoses sindômicas (Crouzon, Muenke), além da Síndrome lacrimo-auriculo-dento-digital³.

Muitos genes têm papéis claros no desenvolvimento ou manutenção de ossos e cartilagens (por exemplo, os envolvidos na sinalização de *NOTCH*, *WNT*, *TGF β* , *BMP*). Existem também os genes que têm um papel mais geral, como a

regulação da transcrição gênica, divisão celular ou transporte intracelular³. Por outro lado, para alguns genes desta lista, o papel na biologia do esqueleto ainda não está claro, como em casos de alguns genes que codificam para proteínas mitocondriais⁴⁶. Além disso, também foi sugerido que em pacientes com DE, as células podem apresentar um possível comportamento alterado desencadeado por alterações de estresse oxidativo, como uma capacidade limitada de produção de ATP, uma proliferação diminuída ou um aumento de indução de morte celular⁴⁷. Até mesmo sequências reguladoras, que residem fora de regiões gênicas, também são causas de distúrbios esqueléticos. Ademais, genes que codificam para miRNA e para RNAs que não são traduzidos também podem ser responsáveis por displasias esqueléticas³. Um exemplo bem conhecido é o gene *RMRP*, que codifica o RNA componente da endoribonuclease mitocondrial de processamento de RNA. Esse gene, quando acometido por variantes patogênicas, leva à Hipoplasia Cartilagem-Cabelo⁷.

4.4 Características clínicas da Hipoplasia Cartilagem-Cabelo

A Hipoplasia Cartilagem-Cabelo (do termo em inglês, *Cartilage-Hair Hypoplasia* – CHH; OMIM 250250) é uma displasia esquelética, inicialmente descrita em 1965 por Victor McKusick. A incidência desta patologia na Ordem Amish é de 1:1000, ao passo que na Finlândia é de 1:23000⁴⁸.

Clinicamente, CHH se caracteriza por uma baixa estatura, encurtamento de membros, displasia metafisária e crescimento deficiente do cabelo, resultando numa aparência de fios finos e ralos. A placa de crescimento desses pacientes é caracterizada por ossificação tardia e presença de poucos condrócitos hipertróficos⁵⁸. Em alguns casos a expressão fenotípica pode ser

ainda mais pleiotrópica⁴⁹, podendo ocorrer também frouxidão dos ligamentos, alterações gastrointestinais, deficiência imunológica e evolução para alguns tipos de câncer, como leucemia, neoplasias de pele, fígado e linfomas; sendo o mais frequente o linfoma do tipo não-Hodgkin, que pode acometer cerca de 6-11% dos pacientes^{8,48,50}.

Os achados radiográficos característicos são a ossificação tardia e a trabeculação dos ossos longos. Também é possível observar metáfises ósseas recortadas, ampliadas e irregularmente escleróticas (figura 9)⁵¹.

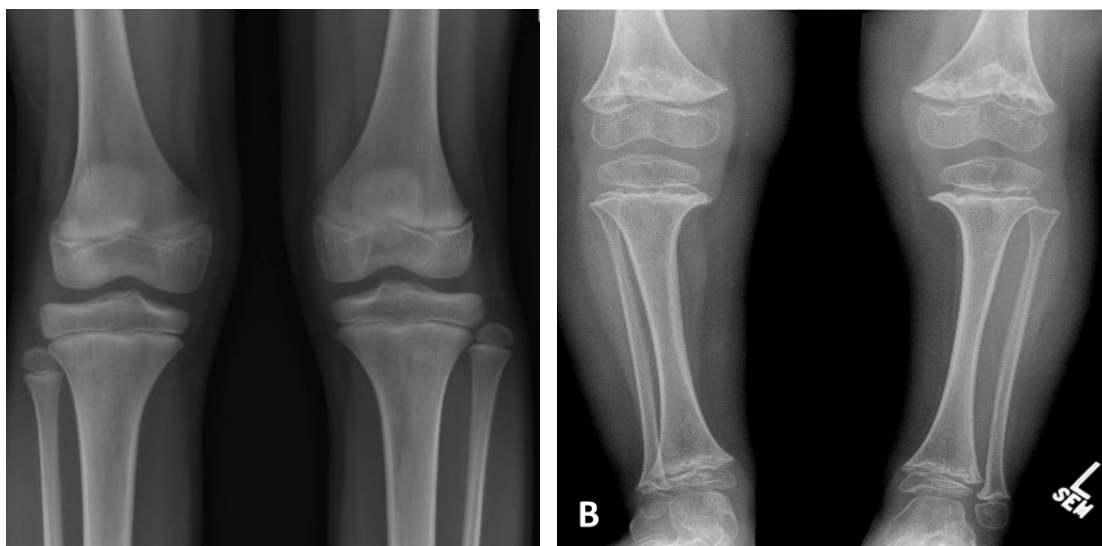


Figura 9: Radiografia de membro inferior. A) Radiografia de indivíduo sem alterações metafisárias aos 8 anos **B)** Paciente com CHH aos 4 anos de idade. As imagens demonstram alterações radiológicas típicas mostrando irregularidades metafisárias com margens de aspecto fragmentado e escleróticas⁵².

O comprometimento do sistema imune é principalmente caracterizado por anemia, linfopenia e neutropenia, diminuição da produção de anticorpos e taxa reduzida de CD4/CD8. Os indivíduos imunodeficientes podem apresentar maior propensão a infecções por vírus e bactérias, podendo se destacar a infecção por varicela-zóster⁵⁰.

Classificações mais antigas distribuíam as características clínicas em 3 subtipos baseados no nível de baixa estatura e características radiográficas adicionais: Displasia Metafisária sem Hipotricose (MDWH), Hipoplasia Cartilagem-Cabelo (CHH) e Displasia Anauxética (AD)^{48,53}. Os pacientes com MDWH eram caracterizados apenas pela displasia metafisária leve. Já os pacientes com CHH além do comprometimento esquelético moderado, apresentavam hipotricose, problemas intestinais e defeitos imunológicos. Os pacientes classificados com AD apresentavam displasia metafisária grave com comprometimento da coluna (displasia espondiloepimetafisária), hipodontia, características faciais e atraso mental leve (figura 10)^{48,53}.

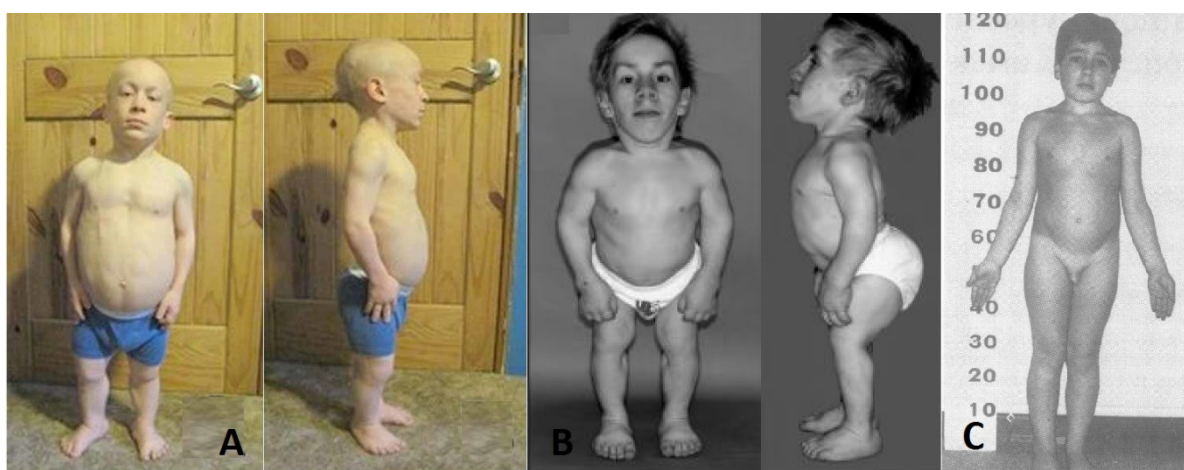


Figura 10: Espectro da Hipoplasia Cartilagem-Cabelo A) Paciente com CHH aos 14 anos de idade (ancestralidade europeia)⁵⁴; B) Paciente com AD aos 16,5 anos de idade (ancestralidade europeia)¹¹; C) Pacientes com MDWH aos 8,5 anos (ancestralidade desconhecida)⁵⁵.

Atualmente, ainda não se sabe se existe relação entre a ocorrência de anormalidades extra-esqueléticas com a gravidade das manifestações ósseas e a variabilidade dos fenótipos têm sido grande obstáculo para o diagnóstico clínico dos pacientes⁴⁸.

4.5 Aspectos moleculares da Hipoplasia Cartilagem-Cabelo

Variantes patogênicas com padrão de herança autossômico recessivo no gene *RMRP*, localizado na região 9p13, foram associadas à Hipoplasia Cartilagem-Cabelo¹³. O RNA produzido pelo gene *RMRP* é classificado como RNA longo não codificante (do inglês, *long non-coding RNA* - lncRNA) e tem expressão ubiquitária⁵⁶. Nesse grupo estão incluídos transcritos que apresentam mais de 200 pares de base de comprimento e não são traduzidos em proteína. Os lncRNA correspondem a grande parte do transcriptoma de mamíferos e a eles têm sido atribuídas diversas funções celulares, incluindo a regulação da transcrição gênica⁵⁷.

O RNA RMRP possui 268 pares de bases e participa de uma pequena ribonucleoproteína nucleolar (snoRNP) chamada RNase MRP (do inglês, *mitochondrial RNA processing*) (figura 11)⁵⁸. Esse complexo ribonucleoproteico é formado pela associação do transcrito com diferentes subunidades proteicas (Rpp14, Rpp20, Rpp21, Rpp25, Rpp30, Rpp38, Rpp40, hPop1, hPop4 e hPop5) e é altamente conservado entre diversas de espécies, incluindo humanos, camundongo, rato, vaca, *Xenopus*, levedura, *Arabidopsis* e tabaco⁵⁹. Oito proteínas da endorribonuclease RNase MRP são compartilhadas com o complexo RNase P, envolvido principalmente na maturação de precursores de tRNA. A RNase MRP atua como uma endorribonuclease com função de clivagem sobre diversos alvos e é encontrada predominantemente no nucléolo, embora tenha sido inicialmente isolada na mitocôndria^{9,58}.

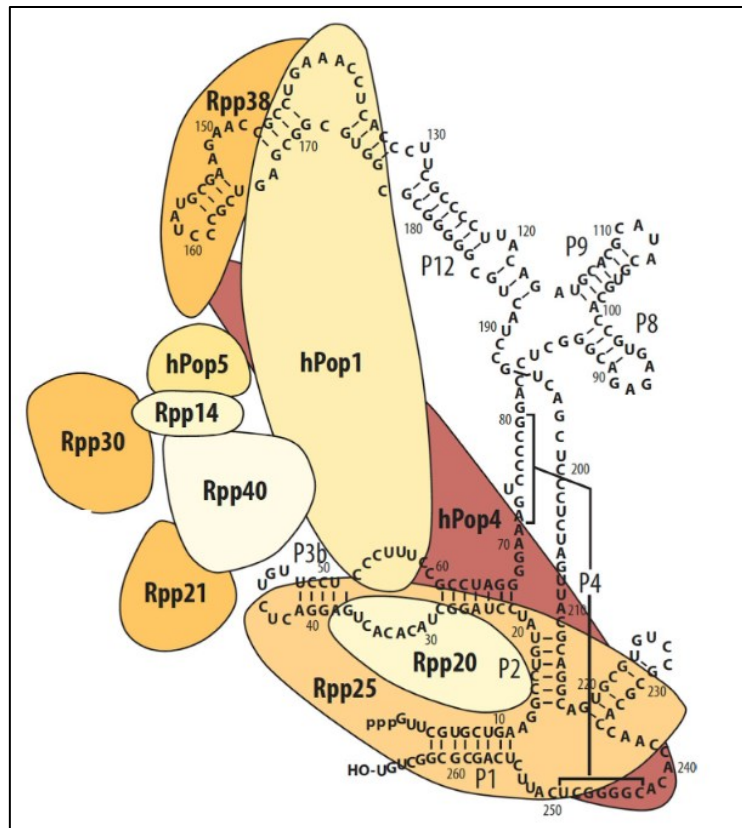


Figura 11: Esquema da estrutura da ribonuclease (RNase MRP). Para formar a ribonuclease, diferentes proteínas se ligam a domínios específicos da estrutura secundária do RNA. As proteínas Rpp20, Rpp21, Rpp25, Rpp30, Rpp38, hPop1 e hPop4 interagem diretamente com a estrutura de RNA formada pela transcrição de MRP, enquanto hPop5 e Rpp14 não. Os domínios P1, P2, P8, P9, P12 e P19 participação predominante, na função de clivagem de rRNA, enquanto partes dos domínios P3 e P4 parecem estar envolvidos principalmente na clivagem do mRNA^{9,53}.

Uma das funções da RNase MRP é o processamento do pré-rRNA no nucléolo para gerar o rRNA de 5.8S. Estudos em leveduras demonstraram que o complexo RNase MRP cliva o pre-rRNA no sítio A3 e ajuda a converter a forma longa do rRNA de 5.8S em sua forma curta e ativa (figura 12)⁹.

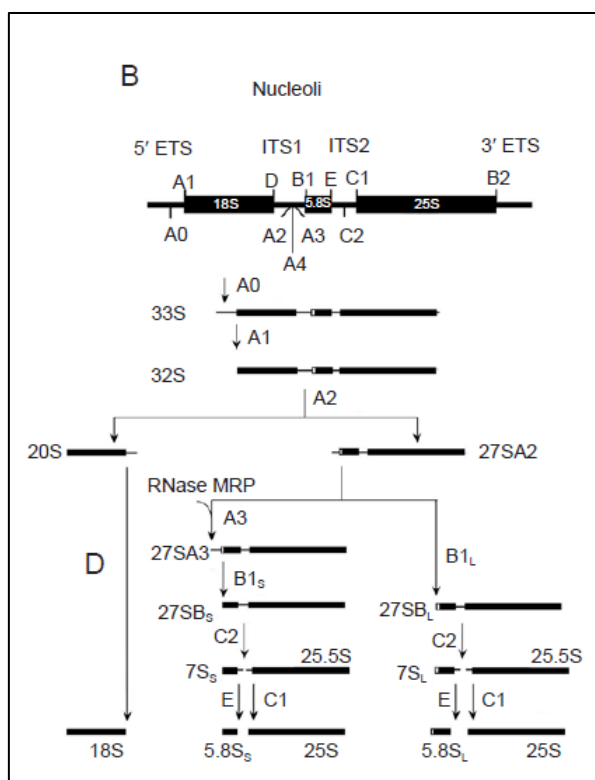


Figura 12: Processamento do pré-rRNA pela RNase MRP em *S. cerevisiae*. Uma das funções do complexo RNase MRP é o processamento do precursor do rRNA 5.8S, subunidade da partícula ribossômica 60S. O transcrito primário 35S é processado em rRNAs maduros 25S, 18S e 5.8S. Na imagem, são indicados os sítios de clivagem (A0 - E), os espaçadores transcritos externos (5'-ETS e 3'-ETS) e os espaçadores transcritos internos (ITS1 e ITS2). O pequeno retângulo branco marca a sequência na forma longa de 5.8S rRNA (5.8SL), que está ausente na forma curta, 5.8SS. A RNase MRP processa o sítio A3 na região espaçadora interna (ITS)⁹.

O complexo RNase MRP também é responsável pela clivagem de iniciadores de RNA necessários para a replicação do DNA na mitocôndria (figura 13A), além de poder influenciar na progressão do ciclo celular no final da mitose, através da degradação de mRNA ciclina B2 (figura 13B)^{9,11}. A síntese da ciclina B2 é essencial para a entrada em mitose e a sua degradação marca a saída desta fase. Em leveduras, foi observado que as mutações afetam a clivagem pela RNase MRP da região 5'UTR do mRNA de CLB2, o ortólogo da ciclina B2 em humanos, promovendo o aumento dos seus níveis e levando assim, a parada na fase final da mitose⁶¹. O transcrito *RMRP* promove uma interrupção da

progressão do ciclo celular em células que se dividem rapidamente, como as da placa de crescimento, medula óssea e folículos capilares⁶¹.

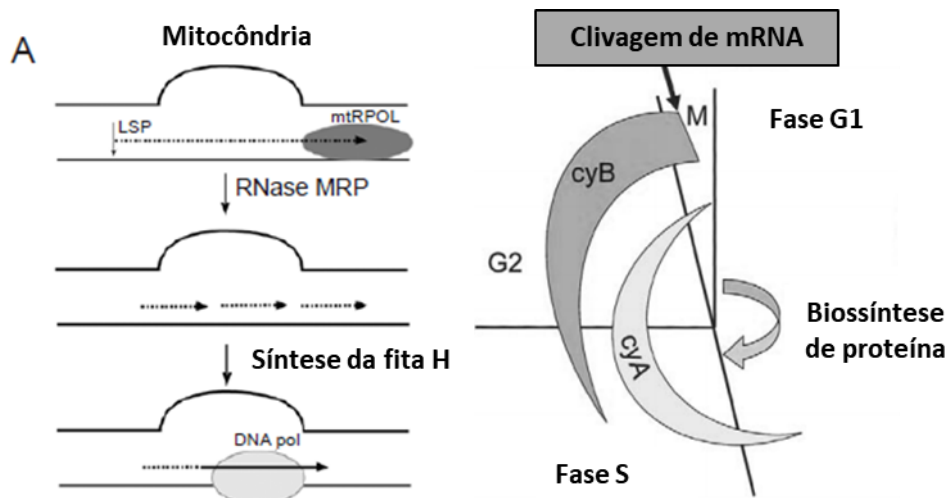


Figura 13: Funções da RNase MRP. Atividade catalítica **A)** no processo de replicação do DNA mitocondrial, através da clivagem de iniciadores. O processo se inicia com a transcrição da fita leve feita pela RNA polimerase mitocondrial, a partir do promotor. Em seguida a fita pesada também é transcrita. O transcrito permanece ligado ao duplex de DNA e é clivado pela RNase MRP para formar os iniciadores que são usados para a iniciar a replicação do DNA pela DNA polimerase gama. LSP: promotor da fita leve mitocondrial; mtRPTOL: RNA polimerase mitocondrial; DNA: DNA polimerase. **B)** na progressão do ciclo celular. O complexo RNase MRP atua na clivagem do mRNA da Ciclina B2, regulando a transição da fase G2 para a fase de mitose. As células de levedura mutantes mostram níveis aumentados de mRNA de ciclina B2 e se acumulam em mitose tardia^{9,11}. cyB: ciclina B; cyA: ciclina A.

O mRNA da viperina, que codifica para uma proteína antiviral celular, também é um substrato para a RNase MRP, podendo estar relacionada com a fisiopatologia da CHH, especialmente quanto à imunidade dos pacientes⁶².

O RNA RMRP parece ter um papel importante durante a diferenciação dos condrócitos e, em especial, na diferenciação hipertrófica dos condrócitos⁵⁸. As placas de crescimento de pacientes com CHH são caracterizadas por atraso na ossificação e a presença de poucas células hipertróficas. A transcrição de *RMRP* em condrócitos parece ser controlada por vias de sinalização condrogênicas bem conhecidas. Em ensaios *in vitro* com células ATCD5, os níveis de expressão de *RMRP* são especialmente induzidos em resposta a WNT-

5A, WNT-3A e BMP-2, morfógenos já bem estabelecidos como associados à hipertrofia de condrócitos. Por outro lado, a atividade do promotor *RMRP* foi reduzida após o tratamento com PTHrP e bFGF, responsáveis retardar a hipertrofia e a diferenciação terminal de condrócitos. Em modelos de *zebrafish* para CHH, os condrócitos se encontravam de maneira desorganizada nas múltiplas camadas de células. Apesar desses achados, ainda não são completamente conhecidos os mecanismos de controle transcricional de *RMRP* em condrócitos e em outros tipos celulares⁵⁸.

Rogler e colaboradores (2014)¹² descreveram que o RNA *RMRP* é fonte para a biogênese de dois pequenos RNAs, através de um mecanismo dependente de Dicer, que atuam como miRNA: *RMRP-S1* e *RMRP-S2* (figura 14). Os miRNAs são pequenas moléculas endógenas de ácido ribonucleico, não codificantes, com cerca de 22 nucleotídeos, que atuam como reguladores da expressão gênica de uma variedade de processos fisiológicos e celulares tanto em nível transcricional como pós-transcricional, através da clivagem de um RNA mensageiro (mRNA) alvo ou através da repressão da tradução.

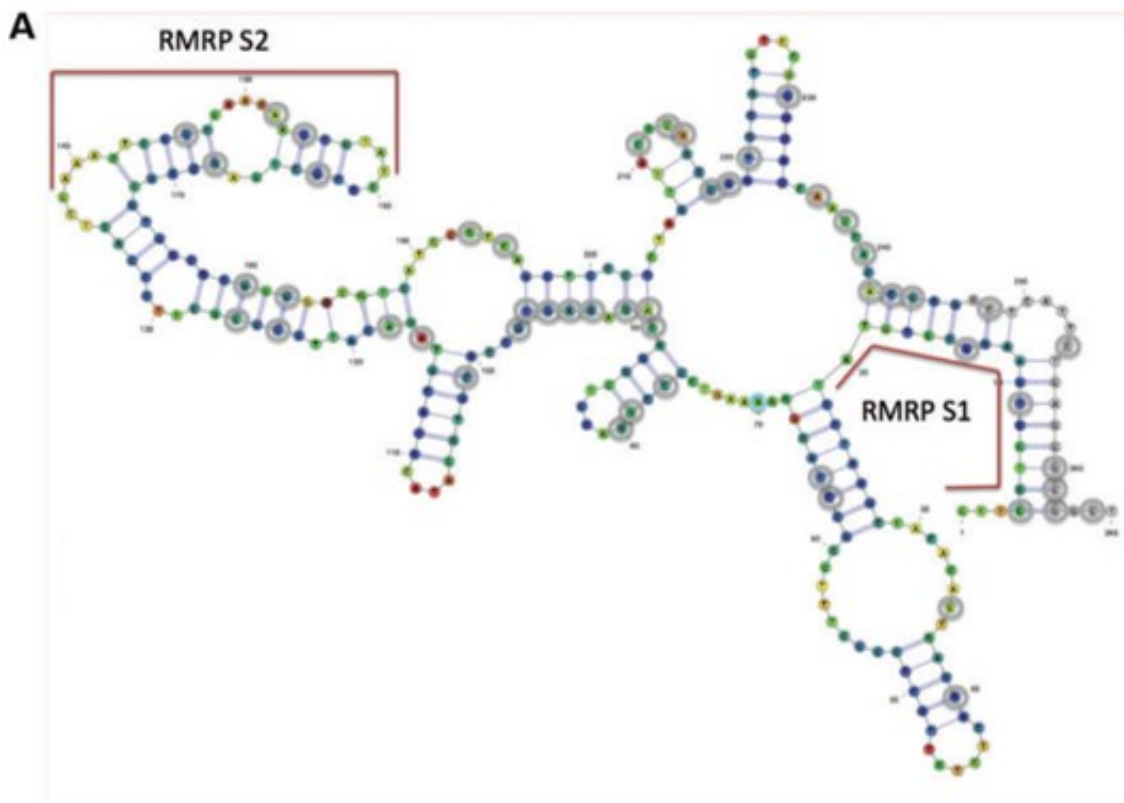


Figura 14: Estrutura secundária de RMRP indicando a localização de RMRP-S1 e S2. As barras vermelhas ao lado da estrutura indicam as posições que dão origem aos pequenos RNAs derivados de RMRP. Os círculos cinza indicam os nucleotídeos com variantes já descritas como associadas à CHH¹².

Foi observado que essas duas moléculas derivadas de RMRP regulam positivamente e negativamente diversos alvos, dentre eles, genes de proliferação, crescimento/desenvolvimento ósseo e relacionados à hematopoese, como *SOX4*, *PTCH2*; *MTDH*, *LEF1*, *HOXD13*; *RBBP9* e *HMOX1*. Em seus ensaios, os autores descrevem que as quantificações de RMRP-S1 e S2 variam de acordo com a substituição estudada, sendo identificadas oscilações na expressão até mesmo na presença de variantes que não ocorrem nas regiões de RMRP-S1 e S2, propriamente dita. Foi então sugerido que as alterações na estrutura secundária causada pelas variantes de CHH, independente de sua localização, podem afetar a biogênese dos dois pequenos RNAs. Desse modo, na hipótese de alterações do nível celular de transcrito RMRP, devido às

diferentes mutações, a regulação gênica exercida por RMRP-S1 e S2 poderia ser afetada, podendo também alterar o fenótipo dos pacientes¹².

Além disso, é provável que o RNA codificado pelo gene *RMRP* desempenhe outras funções que ainda não foram bem definidas⁶³. Estudos moleculares do gene *RMRP* demonstraram a presença de uma alta densidade de polimorfismos de nucleotídeo único (SNPs) e uma extensa série de variantes patogênicas. Essas alterações genéticas podem ser classificadas em duas categorias, sendo a primeira a que inclui inserções, duplicações, ou triplicações no promotor do gene *RMRP* entre o elemento TATA e o sítio de iniciação da transcrição (figura 15)⁶². Variantes na região regulatória do gene resultam na redução significativa da expressão do gene⁷. Acredita-se que isto possa estar relacionado ao fato do gene *RMRP* ser transcrito pela RNA polimerase III. Nesse caso, a distância entre o elemento TATA e sítio de iniciação da transcrição é essencial para a transcrição eficiente⁶³. Diversos estudos relatam que a presença de alterações dessa classe diminui a expressão do gene *RMRP* e é sugerido que variantes bialélicas nessa região talvez sejam incompatíveis com a vida^{61,64}. Reforçando essa hipótese, quase que a totalidade dos pacientes relatados com alterações desse tipo são heterozigotos compostos (apresentam uma alteração na região promotora em um alelo e uma alteração na região transcrita em outro alelo), embora já tenha sido descrito duas alterações no promotor em dois pacientes^{65,66}. Além disso, nenhum estudo conseguiu obter modelos animais apresentando alelos nulos para *RMRP*, indicando que estas construções são letais no período embrionário⁶⁴. Até o momento, já foram identificadas cerca de 30 diferentes mutações deste tipo associadas com a CHH³⁵.

A segunda categoria contempla variantes na região a ser transcrita e, portanto, na molécula de RNA, que está associada com diversas proteínas que irão formar o complexo proteico⁴⁸. Estas alterações são de um único nucleotídeo ou pequenas duplicações e estão possivelmente associadas a alterações na estrutura do RNA e a uma função anormal do RNA. Pacientes com variantes em homozigose ou heterozigose composta nessa região são comuns na literatura e mais de 70 mutações desse tipo foram identificadas⁴⁹.

A redução drástica da quantidade de RNA RMRP em decorrência de mutações na região promotora do gene parece ser mais deletéria do que quantidades normais deste RNA contendo mudanças estruturais sutis, decorrentes de mudanças na região transcrita⁶³. Tem sido apontado que o elevado grau da heterogeneidade das mutações em *RMRP* encontradas em pacientes com CHH pode afetar em diferentes níveis tanto a expressão do gene *RMRP*, quanto a sua função, e que, por conseguinte, pode estar relacionada com o agravamento do fenótipo da doença⁶⁷. Poucos trabalhos na literatura exploram a temática dos mecanismos de transcrição de *RMRP*. Estudos anteriores demonstraram que uma região de 84 bp na região promotora do gene *RMRP* é suficiente para permitir a sua transcrição pela RNA polimerase III. Entretanto, estudos de predição *in silico* de sítios putativos de fatores de transcrição na região de 1.500 pb a montante do local de início da transcrição do gene *RMRP* sugerem a presença de CREB, MEF2, TCF/LEF e SMAD, que também podem participar da regulação espaço-temporal da expressão de *RMRP*. Em células de câncer de cólon, foi observado que a sinalização induzida por WNT-3A é capaz de regular a transcrição de *RMRP*, através da ligação dos cofatores de transcrição de β -catenina e YAP em sítio TCF4 e TBX5⁶⁸. Além dos controles de

transcrição, os níveis de RNA RMRP também podem ser controlados por feedback negativo através da formação de um complexo entre a proteína de transcriptase reversa associada à telomerase (TERT) e o lncRMRP, tendo como resultado siRNAs gerados por um mecanismo dependente de Dicer⁶⁹.

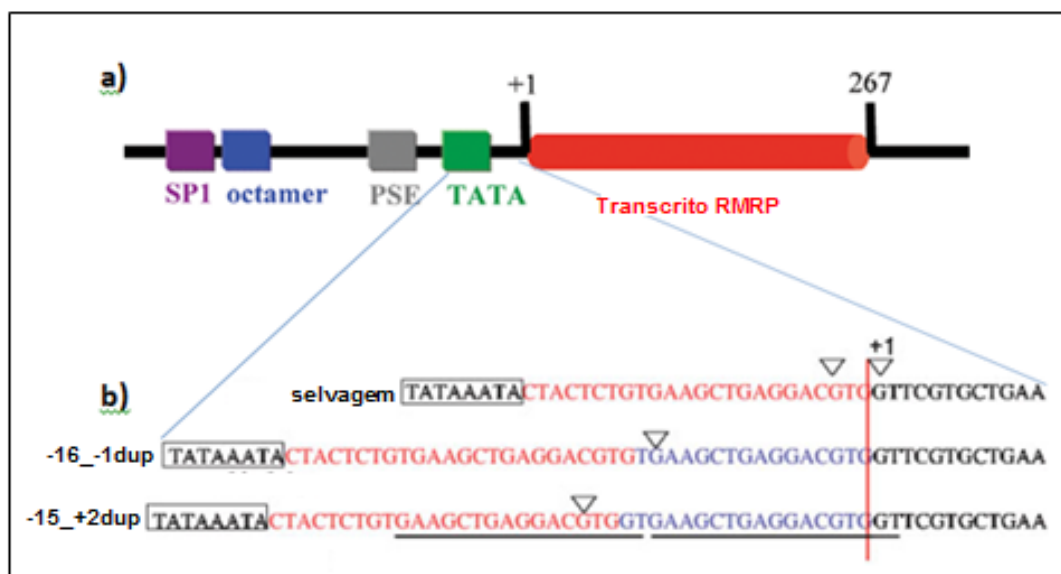


Figura 15: Tipos de alterações que ocorrem no gene RMRP. A) Esquema da estrutura do gene RMRP. O gene não contém íntrons, não é traduzido em proteína e seu transcrito apresenta 267b de comprimento (vermelho). A região promotora contém um sítio de ligação SP1 (roxo), uma octâmero (azul), um elemento de sequência proximal (do inglês PSE) (cinza) e uma caixa TATA (verde); **B)** sequência do promotor do gene selvagem (do inglês WT) e dos promotores mutantes. Os triângulos mostram os locais de iniciação da transcrição. O retângulo mostra a caixa TATA. As sequências vermelhas e azuis mostram o promotor central e as mutações de duplicação. As sequências sublinhadas são as que foram duplicadas (indicadas pela numeração seguidas da sigla dup). Figura adaptada de Hermanns *et al.*, (2005)⁵⁶ e Nakashima *et al.*, (2007)⁶⁷.

Em geral, para o diagnóstico molecular, a correta segregação das mutações nas famílias, a ausência de mutações em homozigose ou heterozigose composta nos controles, a frequência das variantes na população e no caso de mutações não-sinônimas, o grau do efeito danoso na proteína são critérios comumente usados na definição de patogenicidade de uma variante⁶³. Como o gene *RMRP* codifica para um RNA que não é traduzido, a interpretação do efeito de alguma alteração torna-se mais difícil. Neste caso, a classificação da patogenicidade de alterações no gene leva em consideração outros critérios tais

como: a localização do nucleotídeo alterados, ou seja, se estão localizados na região de pareamento das bases para formar a estrutura secundária do RNA, na região de interação com as proteínas que irão formar o complexo RNase MRP ou então, em sequências conservadas evolutivamente (figura 16)^{63,70}. Esta última tem sido utilizada como uma análise importante para identificar essas regiões funcionais que foram conservadas ao longo da evolução, pois parte da premissa de que a conservação de sequências decorre da importância funcional⁷¹.

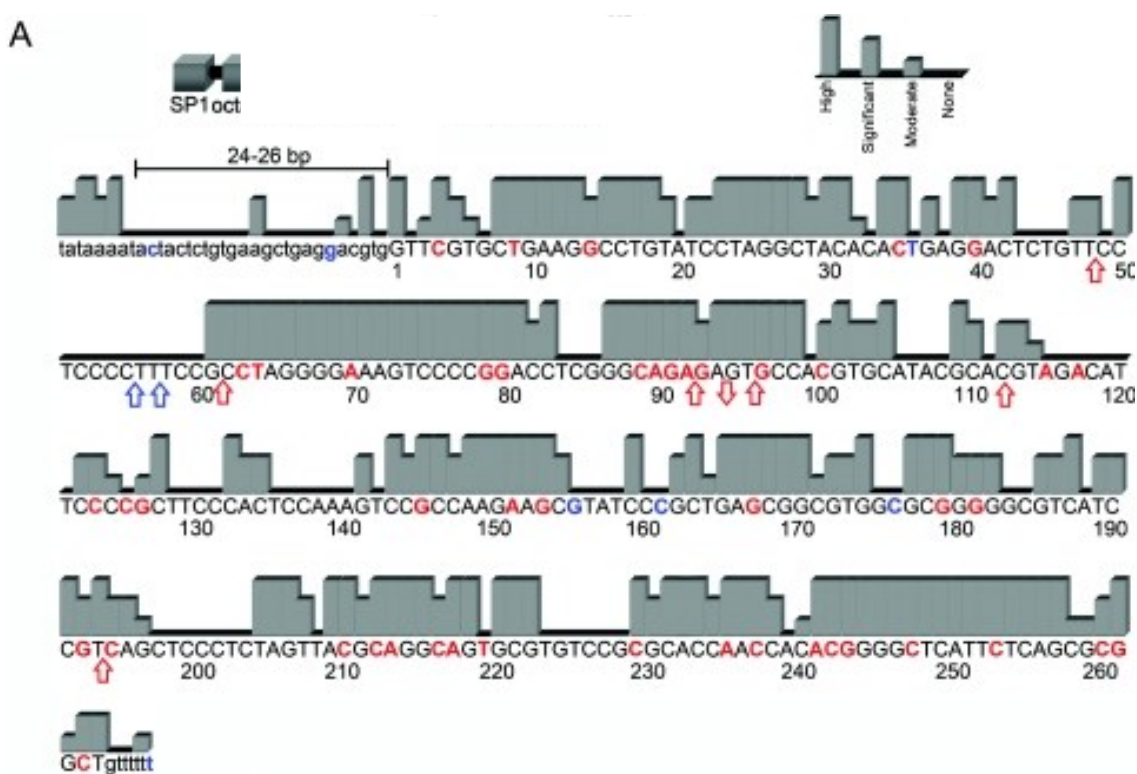


Figura 16: Esquema mostrando a sequência do gene RMRP humano e a sua conservação entre várias espécies categorizada em 4 níveis pelo alinhamento de sequência de 9 espécies. A distância entre o TATA box e o sítio de início da transcrição é conservado em 24-26pb. A distância é conhecida por ser importante para a ligação correta do complexo basal de transcrição da RNA polimerase III (TFIIIB) e é comumente alterada por inserções, duplicações e triplicações, levando a uma expressão reduzida da transcrição. Diversas alterações descritas no gene estão destacadas em vermelho e polimorfismos estão assinalados em azul. Figura adaptada de Thiel *et al.*, 2007⁵³.

Alterações fundadoras relacionadas à CHH já foram identificadas para algumas populações. A substituição g.71A>G em *RMRP* é a mais

frequentemente associada à esta patologia no mundo, e está presente na maioria dos pacientes de origem Amish e em 92% de pacientes de origem finlandesa. Em outras populações europeias sua ocorrência é estimada em 48%⁴⁰. Já em pacientes japoneses, ao invés destas alterações, foram identificadas duas mutações fundadoras: uma duplicação de 17 nucleotídeos na região transcrita e a substituição g.219A>G. Esta última também foi encontrada em outros grupos étnicos^{15,72}.

4.6 Equilíbrio genético e história migratória no Brasil

A frequência de SNPs em populações naturais é condicionada pelo equilíbrio entre a seleção natural, a migração e a deriva genética⁷⁴.

A seleção natural é a mudança nas frequências alélicas dependente do efeito desse alelo sobre o sucesso reprodutivo e não ao acaso. Os alelos adaptativos são positivamente selecionados, pois, suas características tornam os indivíduos mais capazes de sobreviver e de se reproduzir, aumentando o número de descendentes em relação a outros membros da população⁷⁵. Por outro lado, alelos desfavoráveis podem ser eliminados.

O efeito das migrações decorre da capacidade de alteração das frequências gênicas quando determinadas populações recebem imigrantes ou quando uma população diminui pelo efeito do fluxo de migração. Isto é, se uma amostra não representativa da população emigrar, a frequência de certos alelos pode diminuir, dependendo do tamanho e da intensidade da corrente migratória. O estabelecimento de imigrantes em um território já ocupado por outra população, pode também estabelecer novas frequências gênicas neste novo local⁷⁵.

Por fim, a deriva genética é caracterizada pela forma randômica pela qual ocorre uma alteração da distribuição da frequência alélica na população, sendo descritos dois mecanismos pelo qual esse fenômeno pode acontecer: o efeito gargalo e o efeito fundador. O efeito gargalo acontece quando há uma redução extrema da população por efeito de desastres naturais, capazes de dizimar quase todos os indivíduos de um grupo, sobrevivendo apenas uma pequena amostra aleatória de pessoas. Ao final, a composição genética da população passa a ser apenas dos sobreviventes, que pode ser bastante diferente em comparação com a população parental. Alguns alelos podem ser completamente perdidos. Já o efeito fundador se distingue deste mecanismo, pois a alteração aleatória das frequências alélicas não decorre de uma catástrofe, mas da fundação de uma nova colônia por um pequeno número de fundadores, formando pequenos isolados. Nesses grupos, a predominância de um ou mais alelos na população pode ocorrer devido a um indivíduo altamente fecundo que transmite seu material genético a um maior número de descendentes, ainda que um desses alelos confira baixo valor adaptativo a seus portadores⁷⁵.

Em determinadas populações, algumas doenças hereditárias podem estar anormalmente prevalentes⁷⁶. Uma doença autossômica recessiva com elevada frequência em uma comunidade isolada sugere a ocorrência de um efeito fundador e as variantes fundadoras podem surgir por eventos *de novo* ou serem trazidas para população por imigrantes^{77,78}. Em sociedades tradicionais isso é muito comum devido ao fluxo gênico bastante restrito, refletindo em mutações fundadoras locais. Os agrupamentos sociais isolados favorecem o aumento de casamentos endogâmicos, como os observados entre a antiga ordem Amish residente na América do Norte⁷⁹. Este grupo religioso cristão conservador é

descendente de suíços que passaram brevemente pela França em 1853 e migraram para a América do Norte. Diversas doenças genéticas estão presentes nessa comunidade como a síndrome de Ellis-Van Creveld^{80,81,82}.

Um casamento é considerado como consanguíneo quando ocorre no máximo entre primos em segundo grau, pois além desse grau de parentesco as consequências da consanguinidade se comportam quase que como a população geral. Em casamentos consanguíneos, os cônjuges compartilham genes herdados de um ou mais ancestrais comuns, diminuindo, portanto, a variabilidade genética em pequenas populações. Por exemplo, em um casamento consanguíneo entre primos em primeiro grau, o casal compartilha em média 12,5% dos loci gênicos, já primos em segundo grau compartilham em média 3,3%⁸³.

A frequência de casamentos entre familiares varia entre diferentes populações humanas com culturas e religiões distintas e vem diminuindo em decorrência de diversos fatores como urbanização, conquistas femininas e melhorias nos meios de locomoção. Além disso, o advento da internet facilitou a comunicação entre pessoas de diferentes origens (figura 17)⁷⁵. Estima-se que cerca de 10,4% dos indivíduos de todo o mundo estão de alguma forma relacionados como primos de segundo grau ou de maior grau de parentesco. As taxas de consanguinidade para a América do Norte ocorrem entre 0,5% e 4% do total de casamentos e podem atingir 22 e 55% no norte da África e no leste e sudoeste da Ásia, respectivamente⁸³. Na América Latina os registros de consanguinidade correspondem a 1% e 2% do total de uniões, sendo que no Brasil a estimativa média é em torno de 1,87% e varia de acordo com a região analisada⁸⁵. Apesar da consanguinidade não ocorrer em uma parcela tão

expressiva nas uniões conjugais, vale lembrar que os casamentos ocorrem de forma preferencial e não aleatória mesmo em sociedades menos conservadoras, sendo o grupo racial considerado um importante quesito nas escolhas⁸⁶.

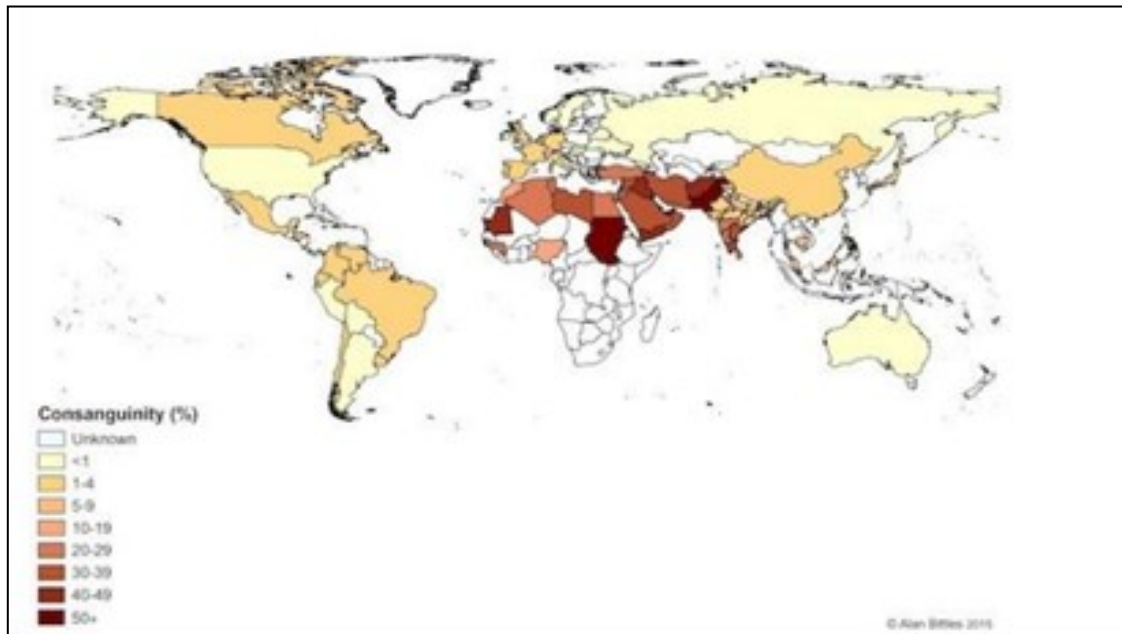


Figura 17: Distribuição das concentrações de uniões consanguíneas pelo mundo⁸⁷.

A população brasileira é resultado de mistura entre diferentes grupos étnicos. A informação genética desta população reúne características de nativos americanos e também carrega as marcas de diversas ondas imigratórias recebidas no país, especialmente de origem europeia. Além disso, o genoma dos brasileiros apresenta uma forte contribuição de indivíduos africanos que aportaram no país pela intensa diáspora Africana. Mesmo com cinco séculos de história, a população ainda pode ser considerada recente e é esperado que as mutações fundadoras encontradas aqui tenham se originado em outros continentes e tenham sido trazidas por imigrantes⁸⁸.

Apesar da expedição de Pedro Álvares chegar ao Brasil em 1500, a efetiva e sistemática ocupação portuguesa da costa litorânea só teve início em

1534 com a implantação do regime de capitânicas hereditárias. Com as incursões dos bandeirantes a partir de 1585, a região interior do Brasil foi gradativamente ocupada. É importante lembrar que antes da ocupação lusitana, os indígenas já habitavam o atual território brasileiro, sendo estimado que cerca de 2,4 milhões de nativos americanos viviam no Brasil quando chegaram os primeiros colonizadores. Entretanto, os conflitos desencadeados pela colonização, juntamente com a ação escravizadora de índios e a incidência de doenças para as quais eles não possuíam anticorpos (como sarampo, gripes e “febres malignas”), levaram a um trágico decréscimo da população nativa⁸⁹.

Apesar de a presença europeia ser extremamente significativa para a ocupação das terras brasileiras, um aporte de quatro milhões de africanos, motivado pela demanda colonial de mão-de-obra, corroborou com o povoamento do Brasil, durante mais de 300 anos, entre os séculos XVI e meados do XIX⁹⁰. O Brasil foi a região que mais recebeu escravos de origem africana, equivalente a mais de um terço de todo o tráfico transatlântico no continente americano. A distribuição de africanos se deu de forma heterogênea no vasto território. As principais regiões de origem desses indivíduos africanos eram Angola, Congo e Moçambique; e os estados de maior importação dessa força de trabalho eram o Rio de Janeiro e a Bahia, abrigando 38% e 25% do total de afrodescendentes, respectivamente⁹¹. Outras regiões também receberam um contingente em menores proporções: Pernambuco (13%), São Paulo (12%), Maranhão (7%) e Pará (5%)⁹². A presença de escravos nos estados estava muito atrelada a regiões com importância econômica, como o litoral, locais de cultivo de produtos agrários e regiões de exploração de minério⁸⁶.

O fluxo de europeus que chegou ao Brasil desde o início da colonização foi bastante expressivo e acredita-se que até 1808 cerca de 500.000 lusos chegaram ao país⁹³, sendo eles majoritariamente indivíduos do sexo masculino⁸⁹. A estimativa é que, do total de imigrantes que desembarcaram no Brasil entre 1500 e 1972, 58% fossem europeus, 40% africanos e 2% asiáticos⁹⁴.

Em 1808, um novo cenário se estabeleceu quando a família real portuguesa chegou ao país, juntamente com a corte composta por 15.000 pessoas⁸⁹. No mesmo ano, Dom João determinou a abertura dos portos, fato que atraiu imigrantes, em especial portugueses, italianos e alemães. Em 1870, o principal grupo étnico do Brasil eram os negros e após a abolição da escravidão em 1888, uma nova onda de imigração europeia foi iniciada, por estímulo do Estado com o objetivo de “branquear” a população⁹⁵. Neste momento foi negada a entrada de imigrantes negros e imposto uma série de restrições para a entrada de asiáticos, não podendo a entrada de indivíduos destas etnias ultrapassarem 3% do total de imigrantes. A estimativa é que quatro milhões de imigrantes se fixaram no Brasil até meados de 1920, incluindo espanhóis, alemães, sírios, libaneses e japoneses^{89,96}.

Como consequência de todos esses processos, a população brasileira se tornou extremamente miscigenada. Atualmente, a população brasileira possui proporções diferentes do componente genético ancestral de acordo com cada região e isso decorre da distribuição geográfica dos grupos étnicos no Brasil, que foi diferenciada pela dispersão seletiva durante a colonização. Segundo a revisão de Souza e cols. 2019⁸⁸, os brasileiros possuem uma alta prevalência de ascendência europeia em todas as regiões (68,1% geral), percentual que flutua de acordo com a região. A segunda mais prevalente é a ascendência Africana

atingindo 19,6% e por último se apresenta o componente nativo americano (11,6%). Vale ressaltar que a mistura entre homens europeus e mulheres nativas americanas teve um início concomitantemente com o estabelecimento da colonização, no entanto, houve uma redução considerável dessa população indígena pelos conflitos e doenças trazidas da Europa^{92,97}. A distribuição por região mostra que cada local tem sua particularidade. Na região Norte, é onde a população que apresenta a maior proporção indígena em seu material genético (27%), comparada às demais. A região Nordeste é que concentra a maior contribuição de origem negra no genoma (22%). Por fim, os estados que possuem maior contingente de ascendentes europeus, atingindo 80%, são do Sul do país (figura 18)⁸⁸. A ocorrência de histórias demográficas variadas durante a colonização levou a uma miscigenação intensa do povo brasileiro e a população resultante acabou por ser considerada como uma das mais heterogêneas do mundo⁹⁸.

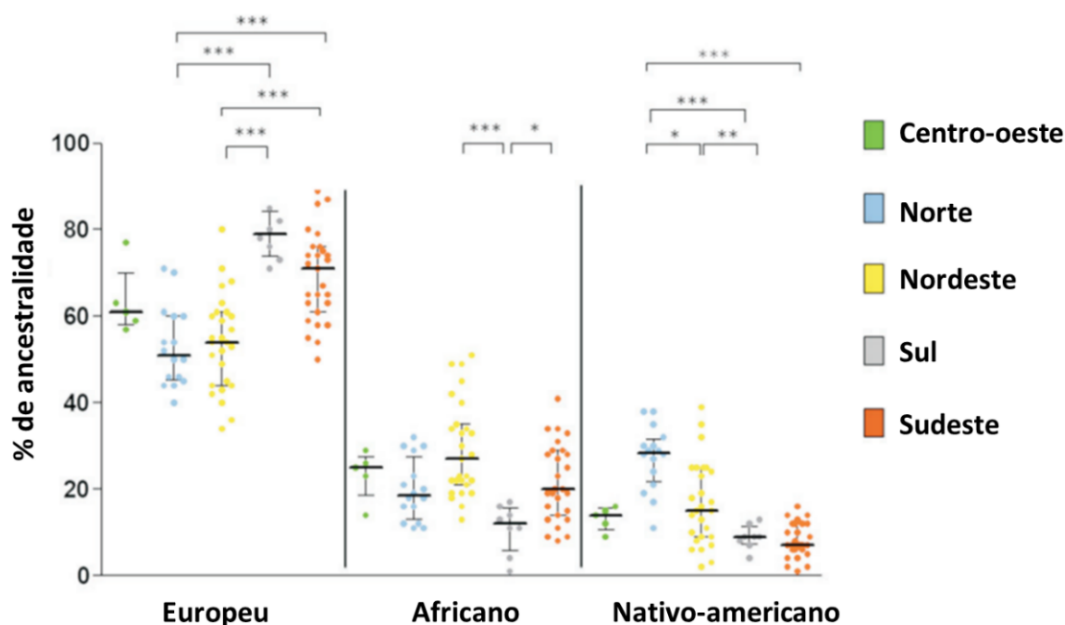


Figura 18: Estimativas genéticas ancestrais de estudos de diferentes regiões geográficas. Asteriscos indicam as diferenças estatísticas (Kruskal- Teste de Wallis, * $p < 0,01$; ** $p < 0,001$, *** $p < 0,0001$).⁸⁸

O estudo da genealogia tem por objetivo inferir relações biológicas entre indivíduos. Métodos tradicionais incluem a análise de dados familiares, como sobrenomes, para determinar o parentesco. George Darwin utilizava esse método em 1875 para estimar as frequências de primos a partir dos dados de incidência de nomes familiares⁹⁹. O uso de dados familiares ainda é uma ferramenta utilizada na genética médica para estudos de ancestralidade, apesar de suas limitações em termos de registro, armazenamento e análise de dados da família¹⁰⁰. Com o desenvolvimento e aprimoramento de técnicas de sequenciamento de DNA, houve um grande avanço no âmbito da genealogia genética, permitindo uma maior compreensão sobre ascendência pré-histórica, origem étnica e geográfica dos ascendentes e de fatores hereditários que afetam a saúde⁹⁹.

É possível determinar o grau de parentesco de dois indivíduos através da probabilidade de eles apresentarem genes idênticos (dominantes ou recessivos), herdados de um ancestral comum a ambos, independente do fenótipo normal ou afetado. Essa medida de correlação genética entre dois indivíduos é denominada coeficiente de parentesco (r) e em populações onde esse parâmetro é muito elevado a probabilidade de um indivíduo ser homocigoto é maior⁷⁵.

Quando variantes em loci diferentes tem a tendência de serem herdadas juntas, elas se encontram em desequilíbrio de ligação (do inglês, *linkage disequilibrium*; LD) e pode ser medida através do coeficiente de correlação entre dois loci ao quadrado (r^2)¹⁰¹. A proximidade física de duas variantes em um cromossomo influencia diretamente nesse parâmetro, variando de 0 para loci extremamente proximais, quase sempre herdados juntos, a 0,5 para loci em extremidades opostas de um cromossomo ou em cromossomos separados. Os padrões de desequilíbrio de ligação são variáveis entre populações e podem ser úteis nos estudos sobre a história das populações humanas¹⁰².

As estimativas de ancestralidade genética podem empregar diversos tipos de marcadores genéticos, tais como as repetições em tandem, inserção/deleções (INDELs) e polimorfismos de nucleotídeo único (SNPs)⁸⁸. Painéis de variantes no cromossomo Y e no DNA mitocondrial frequentemente são utilizados para fornecer informações sobre as linhagens paternas e maternas do probando. Já os painéis de marcadores autossômicos não relacionados, que contém informações derivadas de ambos os genitores, são comumente usados para avaliar padrões de variação genética nas populações humanas em todo o mundo, como por exemplo, os marcadores informativos de ascendência (AIMs)¹⁰³. Estes marcadores podem ser do tipo SNPs ou INDELs e

têm sido empregados para inferir a composição genética de populações miscigenadas, uma vez que são selecionados para maximizar a diferença absoluta na frequência alélica entre duas populações ancestrais⁸⁸.

A prevalência elevada de variantes associadas à CHH em algumas populações do mundo, como nas populações Amish, finlandesa e Japonesa tem sido explicada pelo fenômeno do efeito fundador^{13,10486}. Semelhantemente, os dados genéticos obtidos em nosso estudo anterior mostraram a alta prevalência da alteração g.196C>T, consistente com um efeito fundador para a população brasileira.

5. HIPÓTESE

Duas hipóteses são apresentadas nesta tese:

1) A elevada frequência da alteração g.196C>T na coorte estudada decorre de um possível efeito fundador para a população brasileira de pacientes com CHH.

2) As alterações genéticas observadas no gene *RMRP* dos pacientes com CHH podem levar a diferentes alterações em nível celular, se expressando na variabilidade fenotípica observada nestes pacientes.

6. METODOLOGIA

6.1 Desenho do estudo

Foi realizado um estudo descritivo, retrospectivo e prospectivo a partir de uma análise multidimensional da síndrome CHH. Este trabalho foi desenvolvido a partir da colaboração com três centros hospitalares e estruturado em quatro eixos estratégicos (figura 19): diagnóstico molecular de CHH; análises *in silico* para avaliar o efeito das variantes encontradas na estrutura secundária do transcrito RMRP; estudo molecular do possível efeito fundador e da ancestralidade genética das famílias com CHH; análise de expressão gênica (por qPCR e proteômica). A logística de envio de amostras com qualidade foi um limitante para a realização dos ensaios. Dessa forma, cada análise foi realizada com o número de amostras disponíveis no momento da sua execução, conforme descrito abaixo:

- Etapa retrospectiva de análise de variantes no gene RMRP na coorte histórica: 23 pacientes (Gomes *et al*, 2019)
- Etapa de análise de variantes no gene *RMRP*: amostras de 9 novos pacientes.
- Etapa de análise molecular do efeito fundador: amostras de 27 pacientes.
- Etapa de análise de expressão gênica por qPCR: amostras de 18 pacientes (lncRNA) e 8 amostras (pequenos RNA).
- Etapa de análise de proteômica utilizando amostras de 18 pacientes.

As amostras dos pacientes foram encaminhadas ao Laboratório de Medicina Genômica do IFF pelo ambulatório de Genética Médica e pelas

instituições coparticipantes HCPA e UNICAMP, onde foram realizadas as análises conforme os métodos descritos nessa seção.

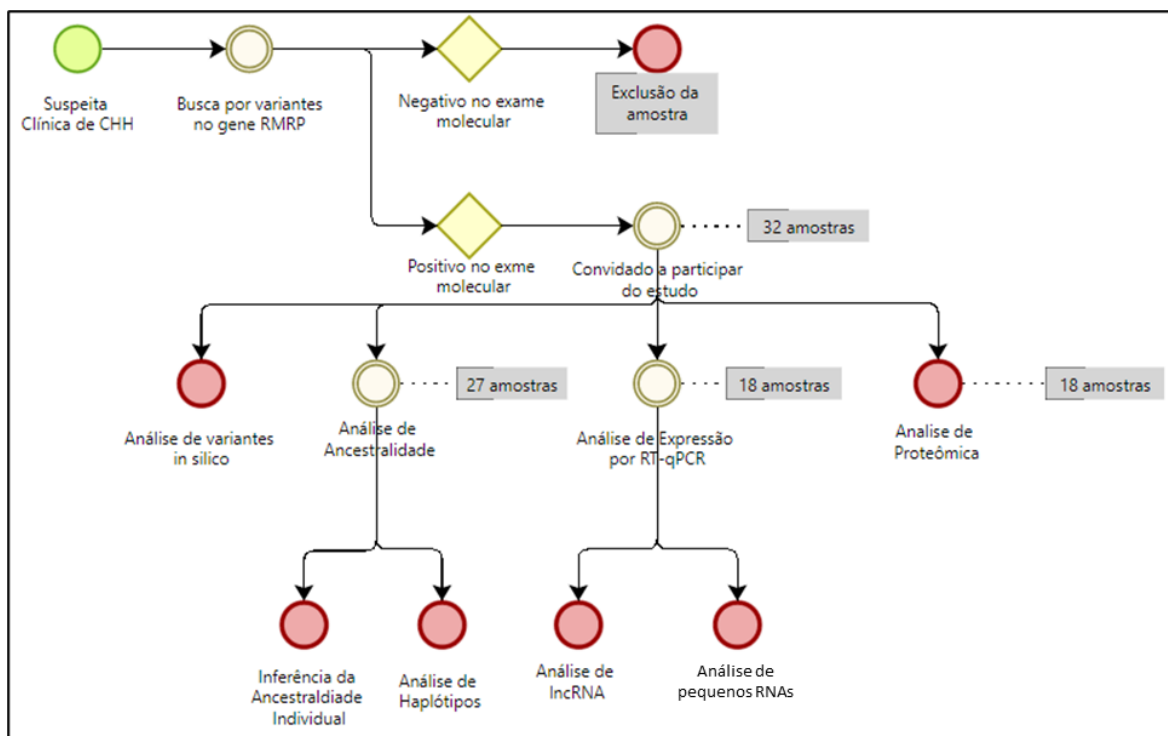


Figura 19: Desenho do estudo delineando as etapas realizadas no trabalho.

6.2 Local do estudo

Os ensaios e análises moleculares deste estudo foram realizados no Laboratório de Medicina Genômica (IFF/Fiocruz) e no Laboratório de Biologia Molecular e Proteômica do Sangue (LABMOPS/LADETEC/UFRJ).

6.3 Critérios de inclusão

Foram considerados apenas indivíduos com diagnóstico de CHH baseado em avaliação clínica e radiológica, sem restrição de idade. Também foram coletadas amostras biológicas dos genitores, quando disponível.

6.4 Critérios de exclusão

Foram excluídos do estudo indivíduos que apresentavam osteocondrodisplasias que não configurassem CHH, indivíduos portadores de aberrações cromossômicas e síndromes genéticas.

6.5 Obtenção de amostras

O estudo foi realizado a partir de amostras de conveniência provenientes de pacientes apresentando fenótipo compatível com CHH atendidos no IFF, no Hospital de Clínicas da Universidade Estadual de Campinas (UNICAMP) e no Hospital de Clínicas de Porto Alegre (HCPA). Após obtenção do termo de consentimento, foram coletadas amostras de sangue periférico (3-8 mL) em tubos contendo anticoagulante EDTA (ácido etilenodiaminotetracético) para a extração de DNA, RNA e proteínas. As amostras foram enviadas ao Laboratório de Medicina Genômica seguindo as normas vigentes de biossegurança, onde foram identificadas por códigos numéricos com o objetivo de preservar a identidade dos indivíduos.

6.6 Identificação de variantes no gene *RMRP* em amostras de pacientes com suspeita de CHH

6.6.1 Isolamento de DNA

O DNA genômico foi extraído a partir de sangue total em um prazo máximo de 48 horas após sua coleta, utilizando o kit comercial *ReliaPrep™ Blood gDNA Miniprep System* (Promega).

Brevemente, utilizando um volume de 200µL de amostra por purificação, foi adicionada Proteinase K e um tampão de lise, para a digestão das proteínas

e rompimento das células, à 56°C por 10 min. Em seguida, foi utilizado um tampão de ligação genômica que permite alta ligação de DNA à membrana da coluna, que foi centrifugada à 14.000 x g para remoção do líquido resultante. A coluna foi então lavada três vezes para remoção das impurezas utilizando tampão de lavagem e centrifugações de 14.000 x g. Por fim, a amostra foi transferida para um novo tubo e o DNA genômico foi eluído em 50 µL de tampão de eluição.

Após a extração, a concentração e a pureza do DNA foram determinadas por absorvância a 230, 260 e 280nm pelo método espectrofométrico utilizando o equipamento Nanodrop 2000 (Thermo Scientific). A integridade do DNA foi analisada através de eletroforese em gel de agarose 0,8 dissolvida em TAE (tampão Tris-Acetato EDTA) 1x, contendo 0,5µg/ml de brometo de etídeo, utilizando o padrão de peso molecular *Lambda HindIII*. A corrida eletroforética foi realizada em aproximadamente 30 min a 10V/cm e o gel foi revelado em sistema de imagem iBright (Invitrogen). Em seguida a amostra foi armazenada a 4°C até a sua utilização.

6.6.2 Amplificação de DNA

O gene *RMRP* e a região regulatória foram amplificados por PCR utilizando oligonucleotídeos sintéticos, cuja sequência foi gentilmente cedida pelos pesquisadores Dr. Andrea Superti-Furga e Dra. Luiza Bonafé do Hospital Universitário de Lausanne. As reações foram realizadas em volume final de 50 µL, utilizando 100ng de DNA, 0,2 µM de cada oligonucleotídeo, 1 unidade de GoTaq DNA polimerase (Promega), 1,5mM de MgCl₂, 0,2mM de cada desoxiribonucleotídeo (dATP, dCTP, dGTP, dTTP), tampão 1X apropriado para

Taq. As condições de ciclagem foram constituídas de uma etapa de 5 minutos a 94°C seguida de 35 ciclos compostos de 3 etapas, sendo a primeira de 30 segundos a 94°C, a segunda de 45 segundos a 54°, e a terceira etapa de 45 segundos a 72°C. Por fim, a reação foi completada com uma etapa de 5 minutos a 72°C em termociclador Veriti (ThermoFisher). Os produtos das reações de PCR foram analisados em gel de agarose 1,2% utilizando o padrão de peso molecular *1 Kb Plus DNA ladder* (ThermoFisher).

O DNA proveniente das reações de PCR foi purificado utilizando *High Pure PCR Product Purification Kit* (Roche), de acordo com as especificações do fabricante. Para isso, o volume do produto de PCR foi completado com água destilada para um volume de 100µL. Um tampão de ligação foi adicionado e a mistura foi transferida para uma coluna posicionada sobre um tubo de coleta. A coluna foi centrifugada a 14.000 x g por 1 min e o filtrado foi descartado. Foram feitas duas lavagens subsequentes da coluna utilizando-se um tampão de lavagem e centrifugações a 14.000 x g por 1 min. A coluna foi então centrifugada vazia a 14.000 x g por 1 min para permitir que todo etanol residual evaporasse. Por fim, a eluição do DNA foi feita em 50µL de água livre de nucleasse seguida de incubação por 1 min a temperatura ambiente e centrifugação a 16.000 x g por 1 min. O produto de PCR purificado foi quantificado utilizando o espectrofotômetro Nanodrop 2000 e armazenado a -20°C até a sua utilização.

6.6.3 Sequenciamento

Os produtos de PCR foram sequenciados no equipamento *ABI PRISM TM 3100 Capillary DNA Sequencer Applied Biosystems* da Plataforma de Sequenciamento da Fundação Oswaldo Cruz. As reações de sequenciamento

foram realizadas de acordo com Sanger *et al.*, (1997)¹⁰⁵ utilizando-se o kit *BigDye™ Terminator Cycle Sequencing Kit 3.1 Ready Reaction* (Applied Biosystems), conforme o protocolo descrito em Otto *et al.*, 2008¹⁰⁶. O sequenciamento de cada amostra foi realizado separadamente com oligonucleotídeos senso e antisenso.

6.6.4 Análises de sequência

A qualidade dos eletroferogramas foi avaliada utilizando-se o *software BioEdit* (versão 7.1.9.0)¹⁰⁷. Em seguida, as sequências resultantes do sequenciamento foram alinhadas com a sequência de referência do gene *RMRP* e de sua região regulatória, obtida no banco de dados *GeneBank* (<https://www.ncbi.nlm.nih.gov/genbank/>), para identificação de variantes. Para esta análise foi utilizado o programa *Clone Manager* (versão 9.0). As variantes identificadas foram descritas com base na montagem GRCh38 do genoma humano.

Quando os eletroferogramas apresentavam padrão compatível com alterações de inserção/deleção (indel), os produtos de PCR eram clonados para confirmar variantes individualmente em cada alelo, de acordo com o descrito abaixo.

6.6.5 Discriminação alélica

Para aquelas amostras de pacientes que apresentavam picos sobrepostos no eletroferograma, indicativo da presença de indels no gene *RMRP* (mecanismo frequente relacionado com a doença), foi feita a discriminação

alélica através da inserção dos fragmentos de PCR de interesse em plasmídeo pGEM-T (Promega). Os *amplicons* purificados foram inseridos em vetores comerciais destinados à clonagem direta de produto de PCR através de uma reação de ligação seguindo a relação molar de inserto/vetor de 3:1. Alíquotas de 5 µL desta reação foram utilizadas para transformar 50 µL bactérias competentes da cepa DH5α. O processo é iniciado incubando a mistura de células com ligação em gelo durante 30 min, procedendo em seguida para o choque térmico das células a 42°C por 90 seg. Para a recuperação das bactérias, 900 µL de LB foram adicionados e a mistura incubada a 37°C sob agitação por 1h (modelo C24KC Incubator Shaker, marca New Brunswick). O plaqueamento foi feito com 50 e 100 µL de cada transformação separadamente em placas de cultura descartáveis contendo 1,5% de ágar, 100µg/mL de ampicilina, 20µg/mL de X-gal (5-bromo-4-cloro-3-indolil-B-D-galactosídeo) e 200µg/mL de IPTG (isopropil-1-thio-B-D-galactosídeo) em um volume final de 25 mL de LB. As placas foram mantidas por 16-20h a 37°C em estufa de incubação 347 CD (FANEM) e posteriormente transferidas para 4°C. A confirmação da presença do plasmídeo recombinante foi feita através da PCR de colônia. Cinco colônias foram transferidas separadamente para microtubos contendo água de injeção e aquecidas a 99°C por 6 min para o rompimento das células e exposição do material genético. O restante das condições da reação e de ciclagem foram preparados conforme item 6.72. Os produtos das reações de PCR foram analisados por eletroforese em gel de agarose e purificados, conforme item 6.6.2; sequenciados e analisados conforme itens 6.6.3 e 6.6.4.

6.7 Predição da estrutura secundária dos transcritos RMRP contendo variantes por análises *in silico*

Foram realizadas análises *in silico* no Laboratório de Genômica Funcional e Bioinformática (IOC/Fiocruz) para determinar os efeitos das mutações encontradas na região transcrita do gene *RMRP* na estrutura secundária do RNA de pacientes recrutados a partir de 2015, utilizando os *software* Varna, LocARNA.e RNAalifold^{108,109,110}.

A estrutura secundária foi obtida do banco de dados Rfam database (nº de acesso RF00030). O banco de dados Rfam¹¹⁰ oferece alinhamento, estrutura secundária consenso e modelos de covariância para famílias de RNA. A anotação dos domínios foi obtida de *Walker and Avis, 2005*¹¹². A partir dos alinhamentos múltiplos, que englobam 66 espécies, disponíveis no banco de dados Rfam, os nucleotídeos conservados foram mapeados. O desenho e a anotação foram realizados com o *software* Varna. Um alinhamento de múltiplas seqüências foi construído usando o LocARNA, disponível no *website* The Freiburg RNA tools (<http://rna.informatik.uni-freiburg.de/LocARNA/Input.jsp>). Os múltiplos alinhamentos do LocARNA são mostrados em conjunto com a estrutura prevista, por RNAalifold. O LocARNA calcula alinhamentos aos pares por programação dinâmica. Alinhamentos múltiplos são construídos a partir de alinhamentos aos pares usando uma estratégia de alinhamento progressivo.

6.8 Investigação do possível efeito fundador da variante g.196.C>T no gene *RMRP* na população Brasileira de pacientes com CHH

6.8.1 Caracterização genética da população estudada

As inferências das porcentagens genômicas de diferentes ancestralidades (Africana, Europeia e Nativo Americana) de 23 pacientes estudados foi realizada através da genotipagem de marcadores informativos de ancestralidade (AIMs- *ancestry informative markers*) utilizando o multiplex AIM-Indels. A detecção dos fragmentos foi feita pelo sequenciador ABI3500® e os produtos gerados foram analisados pelo *software* GeneMapper ID v4.1. O cálculo das estimativas de ancestralidade individual e global da população estudada foi realizado pelo programa STRUCTURE v2.3.4¹¹³, usando como amostras de referência as populações ancestrais africanas, europeias e ameríndias do painel de diversidade HGDP-CEPH¹¹⁴.

6.8.2 Construção de um painel de marcadores adequado para investigação do possível efeito fundador

Para avaliar se a mutação g.196C>T em indivíduos de diferentes famílias no Brasil era devido ao efeito fundador foi construído um painel de marcadores genéticos do tipo TAG SNPs para montagem de haplótipos e comparação dos índices de diversidade entre os portadores e não-portadores desta variante.

A primeira etapa consistiu em realizar uma busca por variantes genéticas em regiões flangeadoras do gene *RMRP* (Chr9: 35657748 – 35658015), abrangendo uma extensão de 7.000pb nas direções 5` e 3` do gene, disponíveis do banco de dados 1000Genomes (banco público com dados depositados de 1.354 indivíduos da fase 3 do projeto, sendo 504 africanos, 503 europeus e 347

nativo-americanos). Todos os SNPs foram extraídos utilizando o *software* VCF tools, obtendo-se um total de 452 SNPs. Foram então calculadas as frequências alélicas de todos esses SNPs nas populações parentais do banco de dados 1000G utilizando o *software* PLINK¹¹⁵. Em seguida, estes foram filtrados com base na anotação genômica funcional realizada pelo *software* Annovar¹¹⁶, seguindo os seguintes critérios: frequência acima de 0,1 na população africana e europeia; apenas substituições de uma única base e ocorrência em região intergênica, intrônica, 5' não traduzida ou 3' não traduzida. No total, foram retidos 23 SNPs, compreendendo uma região total de 12.337 pb. Em seguida, foram selecionados os tag SNPs, utilizando o programa Haploview¹¹⁷, tendo como referência a população europeia já que esta representa a raiz ancestral com maior contribuição para a população brasileira.

Quatro marcadores polimórficos foram selecionados, abrangendo uma região de 7.749 pb (Chr9: 35656875-35664624), incluindo um marcador no gene *RMRP* escolhido para representar a variante g.196C>T, que não estava presente no banco de dados utilizado (rs7021463, localizado uma distância de 39 pb da g.196C>T) (figura 20 e 21). Os marcadores rs10972552, rs7021463, rs1339374, rs1361338 preenchem os critérios de seleção utilizados, tanto para a população europeia, quanto africana. Os SNPs selecionados a partir dessa metodologia estão descritos na tabela 1.

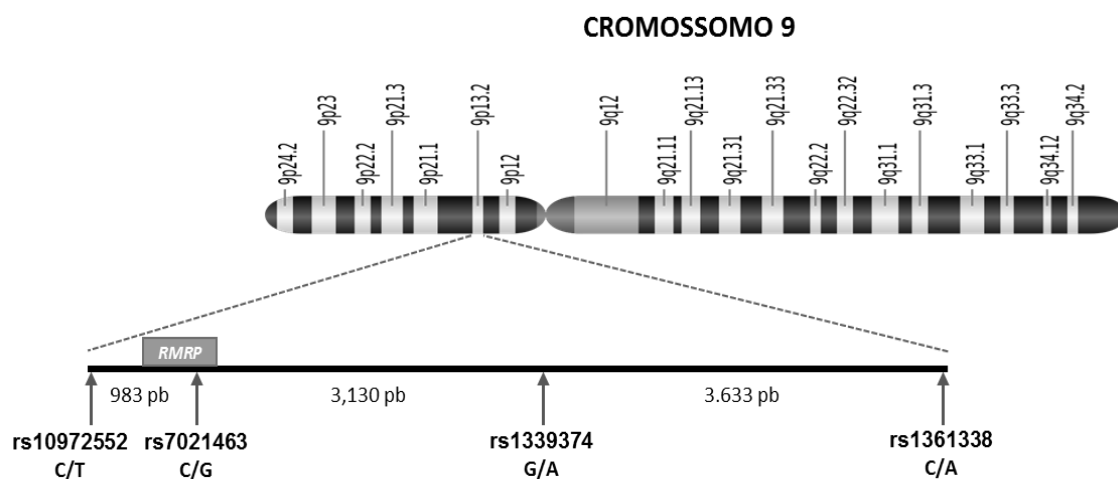


Figura 20: Esquema dos marcadores genéticos selecionados para montagem dos haplótipos. A figura mostra as regiões flangeadoras do gene *RMRP* exibindo a localização dos quatro SNPs (rs10972552, rs7021463, rs1339374, rs1361338) usados para a construção dos haplótipos.

Tabela 1: Tag SNPs selecionados para a investigação do efeito fundador.

SNP	A1	A2	Freq. A1 EUR	Freq. A1 AFR	Região
rs10972552	C	T	0,543	0,485	A montante
rs7021463	G	C	0,133	0,351	A jusante
rs1339374	A	G	0,182	0,407	A jusante
rs1361338	A	C	0,507	0,164	A jusante

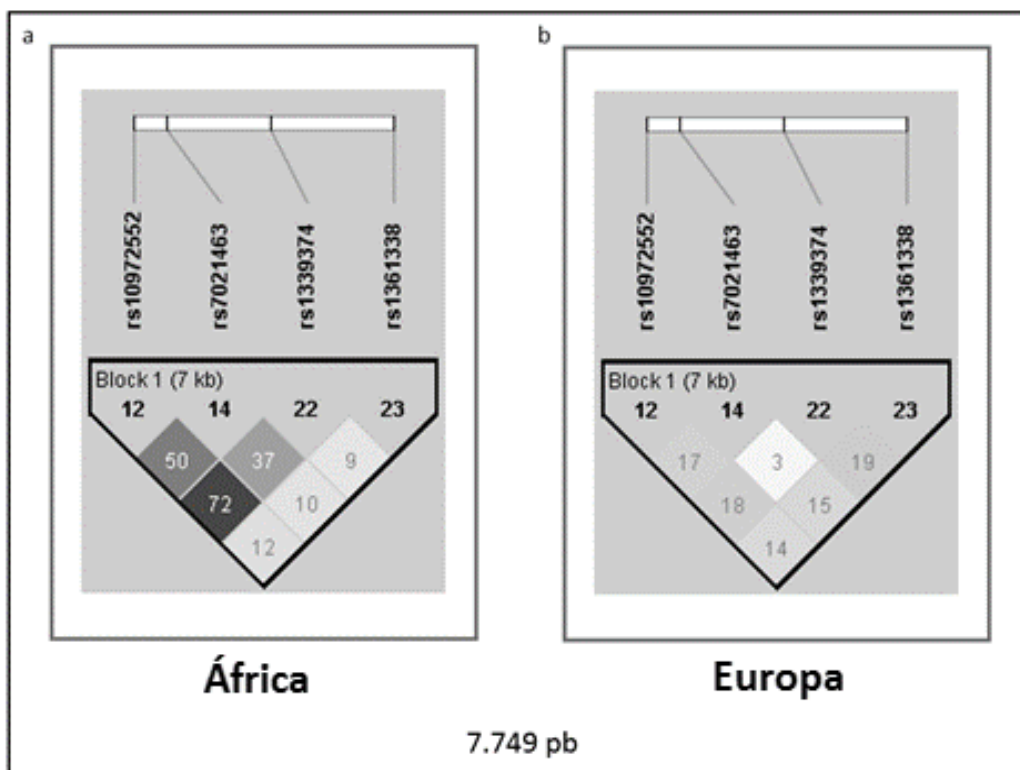


Figura 21: Mapa de desequilíbrio de ligação dos tag-SNPs selecionados pelo Haploview na a) África b) Europa

Ainda, como última análise, foi investigada a frequência dos haplótipos em Europeus e Africanos a partir do *software* Haploview com base nos SNPs selecionados.

O fluxograma da estratégia empregada para a seleção dos SNPs está representando na figura 22. Os TAG SNPs selecionados que melhor representavam a região genômica estudada, foram sequenciados e analisados para verificar se os pacientes compartilhavam o mesmo haplótipo.

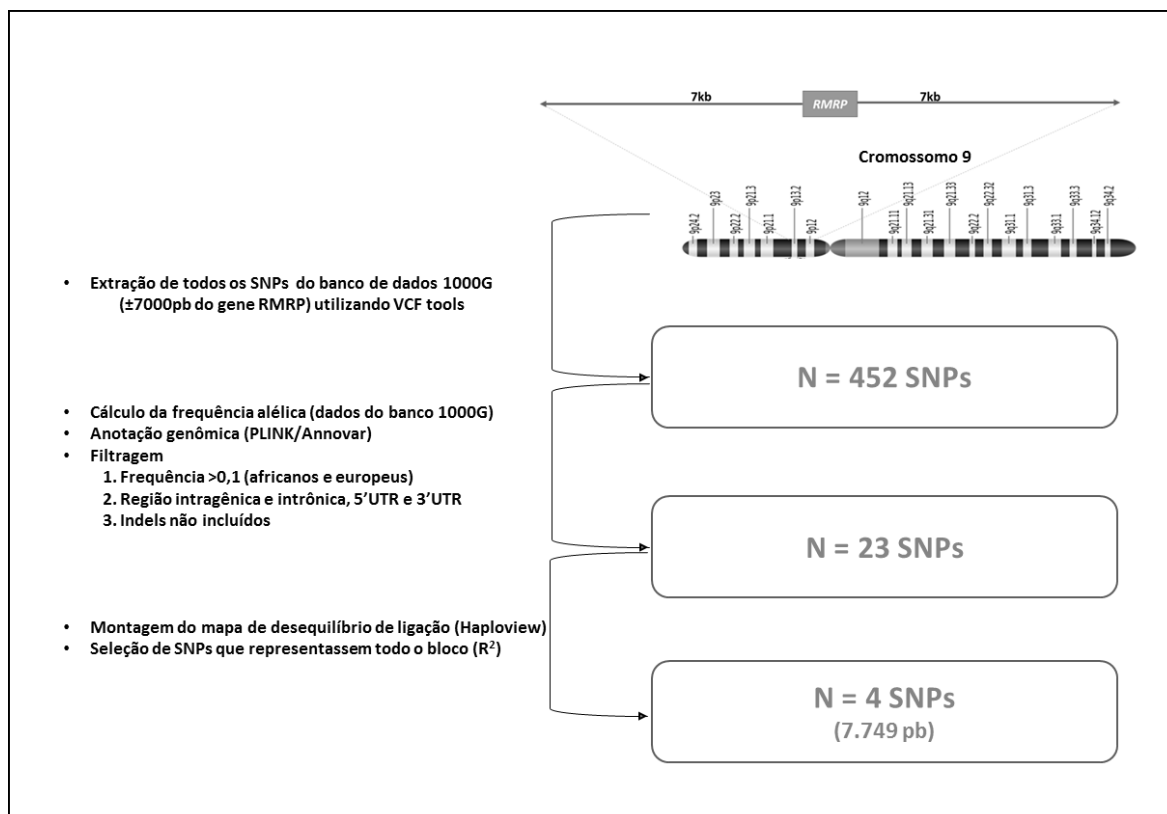


Figura 22: Fluxograma das etapas para seleção de marcadores do tipo TAG SNPs para a construção do painel molecular para análise de haplótipos.

6.8.3 Análise de haplótipos

A montagem dos haplótipos foi realizada através da análise de trio para determinação da fase gamética dos marcadores selecionados. No entanto, foi necessário também fazer o uso da técnica de clonagem para a discriminação alélica, nos casos em que o material dos genitores não estava disponível. Ambas as estratégias estão descritas a seguir.

6.8.3.1 Análise de trio

Para essa abordagem, as regiões contendo os tag SNPs selecionados foram diretamente amplificadas e sequenciadas utilizando amostras dos pacientes e de seus progenitores, quando disponíveis.

As reações foram realizadas em volume final de 50 μ L, utilizando 100ng de DNA, 0,2 μ M de cada oligonucleotídeo, 1 unidade de GoTaqDNA polimerase (Promega), 1,5mM de $MgCl_2$, 0,2mM de cada desoxiribonucleotídeo (dATP, dCTP, dGTP, dTTP), tampão 1X apropriado para Taq. Para tal, as reações foram preparadas utilizando 100ng de DNA, 0,2 μ M de cada oligonucleotídeo (anexo 11.2), 1,5mM de $MgCl_2$, tampão 1X apropriado para Taq e o volume total foi de 50 μ L. A reação foi incubada então em termociclador Veriti (ThermoFisher) e foi submetida às seguintes alternâncias de temperatura: 5 minutos a 94°C para desnaturação; 35 ciclos compostos de 3 etapas: sendo a primeira de 30 segundos a 94°C, a segunda de 30 segundos a 58° e a terceira etapa de 30 segundos a 72°C; e para finalizar, a reação foi completada com uma etapa de 5 minutos a 72°C.

Os produtos das reações de PCR foram analisados em gel de agarose 1,2% utilizando o padrão de peso molecular *1 Kb Plus DNA ladder* (ThermoFisher). Em seguida, as reações foram purificadas conforme item 6.7.2, e quantificadas utilizando o espectrofotômetro Nanodrop 2000. Os sequenciamentos foram preparados conforme 6.7.3, utilizando os oligonucleotídeos desenhados para sequenciar os tags SNPs (anexo 11.2).

6.8.3.2 Separação alélica por clonagem

6.8.3.2.1 Amplificação da região alvo:

O DNA genômico previamente extraído para genotipagem do gene *RMRP* conforme o item 6.7.1 foi utilizado como molde para amplificar toda a região alvo contendo os SNPs selecionados pelo *software* Haploview. Foram utilizados oligonucleotídeos sintéticos específicos acrescidos de sítios de restrição apropriados para a clonagem dos fragmentos no vetor pKS (pBluescript SK - Addgene) (figura 24).

As reações foram padronizadas utilizando tampão 1X, 0,2 mM de mix de dNTP, 0,2 μ M de oligonucleotídeo senso e antisenso, em volume final de 50 μ L, 100ng de DNA, 0,2 μ M de cada oligonucleotídeo, Enhancer GC Superfi 1X e 0,5 U da enzima Platinum™ SuperFi DNA polimerase, específica para amplificação de produtos longos, em um total de 50 μ L. O perfil térmico da reação era composto de uma incubação inicial a 98°C por 30 segundos seguido de 35 ciclos de alternância de temperatura (98°C por 10 seg, 60°C por 10 seg, e 72°C por 6 min) e uma extensão final de 72°C por 5 min. A qualidade e especificidade dos fragmentos obtidos foram visualizadas ppor eletroforese em gel de agarose 0,8%, utilizando padrão de peso molecular *1 Kb plus DNA Ladder* (ThermoFisher). Em seguida, o DNA proveniente das reações de PCR foi purificado conforme item 6.7.2, e quantificado utilizando o espectrofotômetro Nanodrop 2000.

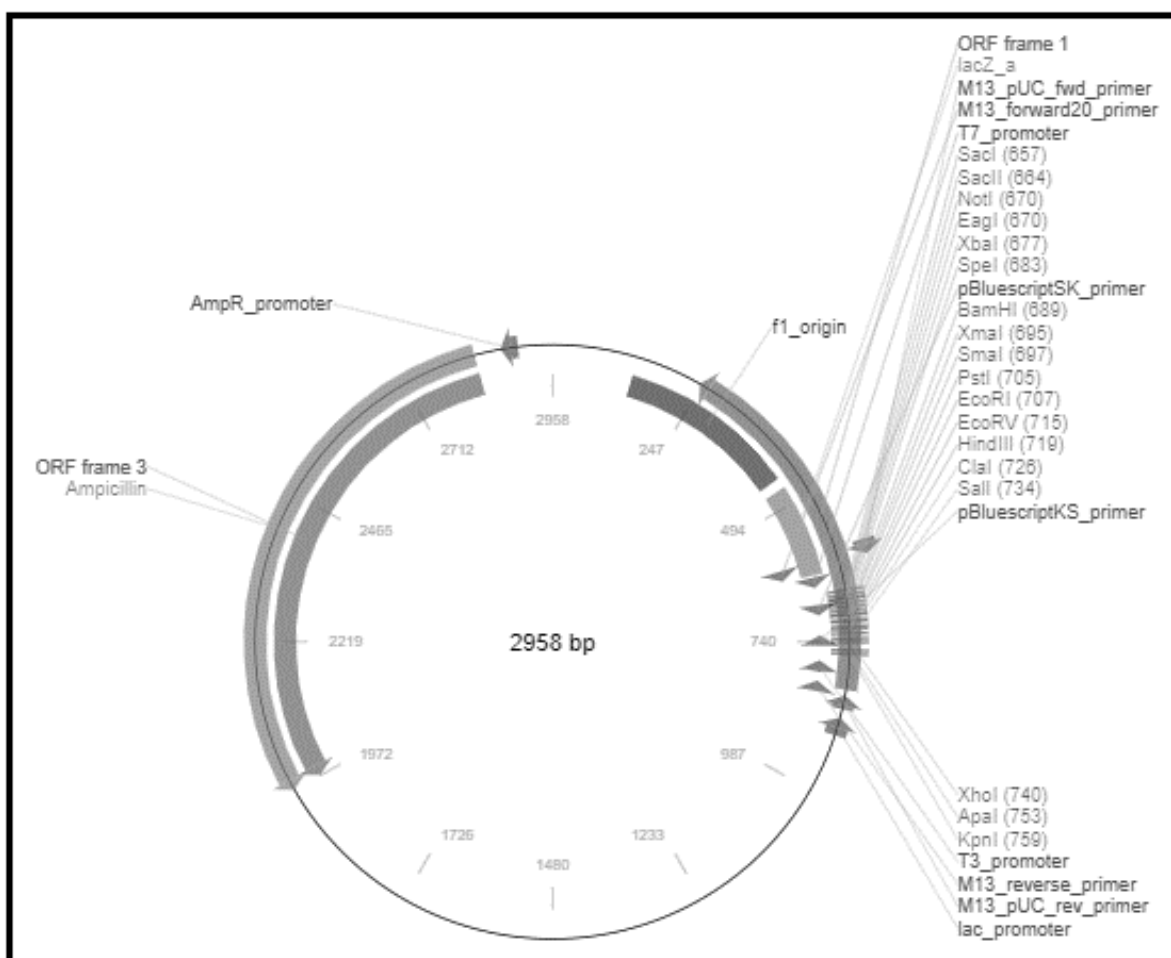


Figura 23: Esquema do vetor pKS. Vetor utilizado para gerar as construções empregadas na investigação do efeito fundador (Addgene)

6.8.3.2.2 Digestões de DNA com enzimas de restrição

O plasmídeo pKS e o produto de PCR foram submetidos a reações de digestão com a enzima de restrição *HindIII*, utilizando-se uma proporção de 1 U de enzima para cada micrograma de DNA com tampão E, apropriado para a enzima, BSA 1X e volume de água suficiente para ajustar o volume para 20 μ L. Após a digestão, a enzima será inativada por temperatura (65 $^{\circ}$ /15 min).

6.8.3.2.4 Defosforilação do plasmídeo pKS digerido com *HindIII*

Para evitar a autoligação dos plasmídeos pKS digeridos com a endonuclease *HindIII*, foi realizada uma reação de defosforilação dos vetores com a enzima fosfatase alcalina de camarão termossensível (TSAP, do inglês Thermosensitive Shrimp Alkaline Phosphatase) (Promega). As reações foram realizadas utilizando 1 µg do DNA plasmidial digerido com *HindIII*, 1 U TSAP e água estéril até completar o volume de 20 µL. As reações foram incubadas à 37°C durante 15 minutos e após esse tempo, a enzima TSAP foi inativada por incubação da reação à 74 °C, durante 15 minutos.

6.8.3.2.5 Ligações de fragmentos de DNA a vetores de clonagem para discriminação alélica

Após a digestão dos fragmentos de interesse, o material foi submetido a reações de ligação, seguindo-se a relação molar de inserto/vetor de 3:1, utilizando-se uma concentração típica de 50 ng de DNA do vetor digerido e defosforilado. Para a reação de ligação foram adicionados à mistura inserto/vetor, 1 U de T4 – DNA ligase (ThermoFisher), tampão apropriado 1X e água para um volume final de 20 µL. As reações foram incubadas a 37°C por 1h e/ou 4°C durante a noite. Alíquotas de 5 µL desta reação foram utilizadas para transformar bactérias competentes da cepa Trans1-T1 Phage Resistant Chemically Competent Cells (Transgen Biotech).

6.8.3.2.6 Transformação das Células Competentes

Para a transformação, 10% da ligação foi adicionado em 50 μ L de célula competente e incubado em gelo durante 30 min. Após o período de incubação as células foram submetidas a um choque térmico, em banho-maria, a 42°C por 30 segundos e imediatamente incubadas em gelo por 2 minutos. Após esse período, foram adicionados 500 μ L de LB (Luria Bertani) e a mistura foi incubada a 37°C com agitação de 200 rpm (rotações por minuto) por 1h (Incubadora *Shaker* Refrigerada com movimento orbital modelo NT735 - marca Nova Técnica). Cerca de 100 e 150 μ L de cada transformação foram transferidos separadamente para placas de cultura descartáveis contendo 1,5% de bacto-ágar, 100 μ g/mL de ampicilina, 20 μ g/mL de X-gal (5-bromo-4-cloro-3-indolil-B-D-galactosídeo) e 200 μ g/mL de IPTG (isopropil-1-thio-B-D-galactosídeo) em um volume final de 25 mL de LB. As placas foram mantidas por 16-20h a 37°C em estufa de incubação (modelo 347 CD, marca FANEM) e em seguida transferida para 4°C para que eventuais colônias brancas se tornem azuis. Em seguida, foram realizadas reações de PCR para confirmação da presença do plasmídeo recombinante.

6.8.3.2.7 PCR de Colônias

Para a seleção por PCR das colônias contendo os plasmídeos de interesse foram utilizados os mesmos oligonucleotídeos do sequenciamento do gene *RMRP*. Dez colônias brancas foram identificadas e parcialmente transferidas separadamente para microtubos contendo 34,4 μ L de água Milli-Q autoclavada e aquecidas a 99°C por 6 min, para que o DNA dessas colônias fosse utilizado como molde para as reações de PCR. Para isso, a mistura

água/bactéria foi mantida em gelo e em seguida foram adicionados os reagentes necessários para a reação de amplificação conforme item 6.7.2. Os produtos das reações de PCR foram analisados por eletroforese em gel de agarose 1,2 %, purificados e sequenciados conforme item 6.7.2.

Logo após a corrida eletroforética, as colônias positivas para PCR foram cultivadas em 600 µL de meio LB contendo antibiótico ampicilina por 16h a 37° C, sob agitação de 250 rpm. Após esse período, foram adicionados 20% de glicerol estéril, constituindo assim estoques de bactérias recombinantes que foram mantidos a -70° C.

6.8.3.2.8 Análise dos sequenciamentos

As sequências obtidas foram alinhadas com a sequência de referência do genoma humano conforme item 6.7.4, de modo a determinar o alelo presente em cada uma das colônias.

6.8.3.2.9 Purificação de DNA plasmidial em pequena escala (MINIPREP)

Após o resultado do sequenciamento, os estoques das colônias contendo um plasmídeo correspondente a cada alelo dos pacientes, foram descongelados. Após chegarem à temperatura ambiente, 100 µL foram inoculados em 5 mL de meio LB contendo ampicilina e incubados a 37° C por 16 h a 200 rpm. Após a incubação, foram então transferidos 200 µL da cultura para um microtubo contendo glicerol 20% para preparar um novo estoque e o restante da cultura foi utilizado para a extração e purificação do plasmídeo utilizando-se *Kit Wizard® Plus SV Minipreps DNA Purification Systems* (Promega), de acordo com

especificações do fabricante. O DNA plasmideal purificado foi quantificado no Nanodrop 2000 e utilizado como molde para as reações de sequenciamento conforme item 6.7.3, utilizando os oligonucleotídeos desenhados para sequenciar os tags SNPs (anexo 11.2).

6.8.3.3 Análises de Sequência

Os sequenciamentos foram alinhados com a sequência de referência do fragmento delimitado pelo painel de SNPs selecionado conforme item 6.10.2, obtida no banco de dados de nucleotídeos do *GeneBank*. Foram utilizados os programas *Clone Manager* (versão 9.0) e *BioEdit* (versão 7.1.9.0). Os alelos de cada marcador foram anotados para cada um dos pacientes para a posterior montagem dos haplótipos. Para a abordagem de trio, a fase de cada marcador foi inferida usando dados da análise dos dados genéticos dos progenitores, quando necessário (figura 23).

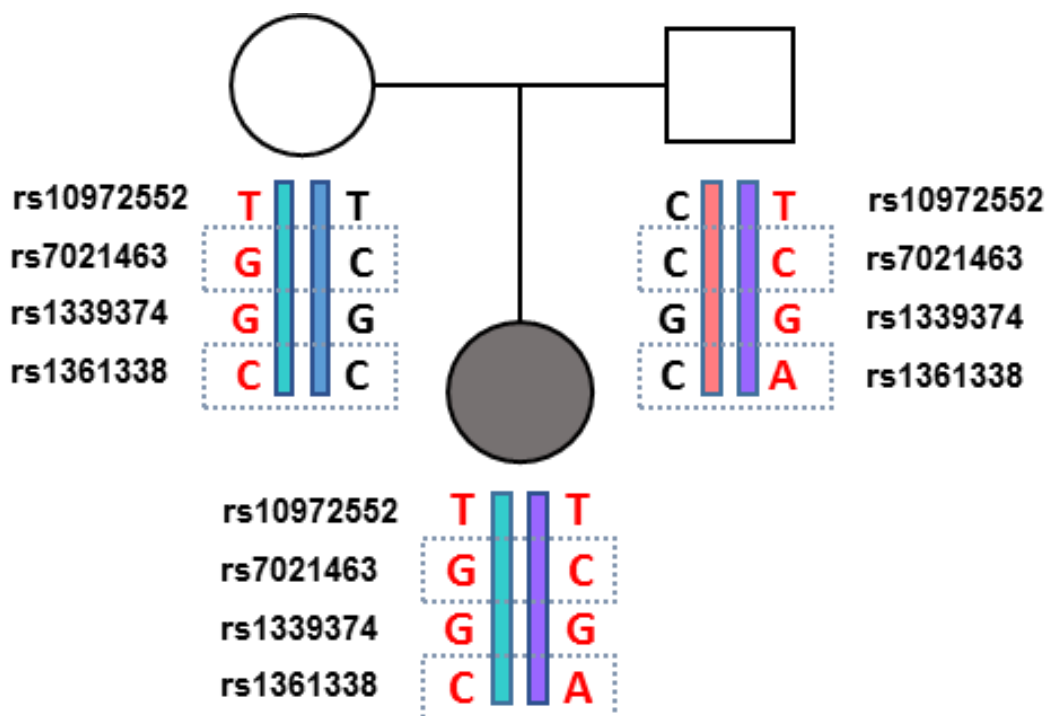


Figura 24: Esquema da inferência das fases dos alelos para os marcadores rs10972552, rs7021463, rs1339374, rs1361338. No exemplo acima, existem dois marcadores heterocigotos, que, na ausência de informações dos genitores, não seria possível determinar suas fases gaméticas (retângulos pontilhados). Para o marcador rs7021463, é possível observar que a paciente em questão apresenta apenas os alelos “G” e “C”, assim como sua genitora. No entanto, seu pai apresenta apenas o alelo “C” neste locus, de modo que podemos inferir que o “G” presente na probanda foi herdado de sua mãe. O mesmo ocorre para o último marcador (rs1361338), onde se pode inferir que o “A” foi herdado do seu genitor.

6.8.3.4 Análise estatística

As frequências alélicas foram determinadas por contagem direta. A diversidade haplotípica foi calculada conforme descrito em Nei (1987)¹¹⁸, no qual os resultados variam de zero a um, significando, o limite mínimo, que todos os haplótipos são iguais e o máximo, que cada indivíduo tem um haplótipo diferente. A equação utilizada está descrita a seguir:

$$h = 1 - \sum p^2$$

Em que:

- p é a frequência de cada haplótipo observado na amostra
- n é o número total de indivíduos analisados.

O teste do exato de Fisher foi usado para comparar as frequências entre haplótipos portadores ou não da alteração g.196C>T. A hipótese nula foi rejeitada utilizando como parâmetro $p < 0,05$.

Foi construída a rede de haplótipos empregando o programa Network Software 4.6.1.3 através do método *median-joining network* para observar a relação filogenética da distribuição dos haplótipos. Esta análise elabora redes haplotípicas com mínima abrangência ("*minimum spanning*"), utilizando todas as sequências inseridas, através da união de árvores que apresentem o menor número de alterações entre os nós¹¹⁹.

6.9 Análise de expressão gênica por qPCR em tempo real

Para analisar as alterações da expressão do gene *RMRP* e dos pequenos RNAs derivado deste transcrito em pacientes, foram realizados ensaios de RT-qPCR utilizando leucócitos provenientes de sangue periférico conforme descrito a seguir.

6.9.1 Obtenção dos leucócitos

A lise de hemácias das amostras de sangue periférico foi realizada com adição de tampão de lise de Hemácias (KHCO₃ 0.2M; NH₄Cl 3M; EDTA 1mM) completando o volume para 20mL. Após homogeneização por vórtex, o tubo foi incubado por 30 minutos no gelo. Em seguida, o tubo foi centrifugado a 1.500 rpm por 15 minutos e o sobrenadante foi descartado. Esse procedimento foi repetido até a ser obtido um pellet de leucócitos brancos. Por fim, o pellet foi lavado com 1mL de PBS (tampão Fosfato Salino) 1X e dividido em 2 tubos para extração de RNA e proteína.

6.9.2 Extração de RNA

O tubo contendo os leucócitos em PBS obtidos conforme o item anterior foi centrifugado a 1.000 rpm por 5 minutos e o sobrenadante foi descartado. O precipitado de células foi ressuspenso em 1 mL de Trizol (Invitrogen) e armazenado em freezer -80°C até o momento da extração do RNA total. Uma vez descongelada, a solução obtida foi incubada por 5 min temperatura ambiente de forma a permitir a completa dissociação das proteínas. Logo após, adicionou-se 200 µL de Clorofórmio (Merck) e o conteúdo foi homogeneizado vigorosamente por inversão. Após uma breve incubação de 2 min à temperatura ambiente, o tubo foi centrifugado a 12.000 x g por 15 min em centrífuga refrigerada à 4°C. Finalizada a centrifugação, foi obtida uma solução contendo 3 fases: uma superior incolor, a qual continha o RNA, uma inferior orgânica de coloração rosa (contendo DNA e proteínas) e uma interfase. A fase superior foi então transferida para um novo tubo e foram adicionados 500 µL de isopropanol (Merck). A solução foi homogeneizada e incubada por 10 min à temperatura ambiente. Para precipitar o RNA, o tubo foi centrifugado à 12.000 x g por 10 min a 4°C e ao final o sobrenadante foi descartado. Ao pellet, foi adicionado 1 mL de etanol 75% para lavá-lo por meio de centrifugação a 7500 x g por 5 min. Por fim, o sobrenadante foi removido com auxílio de uma pipeta e o precipitado foi eluído em água livre de nucleases.

A quantificação do RNA foi realizada pelo espectrofotômetro Nanodrop 2.000 e a integridade do RNA foi analisada em gel de agarose 0,8% dissolvida em TAE 1X (0,04M Tris-acetato, 0,001M EDTA, água contendo 0,01% de Dietilpirocarbonato (DEPC), contendo 0,5µg/ml de brometo de etídeo (Invitrogen). A corrida eletroforética foi realizada a 5V/cm.

6.9.3 Tratamento do RNA com DNaseI

Com o objetivo de degradar possíveis contaminações por DNA genômico, o RNA extraído por foi submetido ao tratamento com DNaseI livre de RNase (New EnglandBiosciences). As reações foram preparadas utilizando 2µg de RNA; 2,6 µL de tampão de DNase 10X (New EnglandBiosciences), 1,25 U de DNase I (2 U/µL) em um volume final de 26 µL. O tubo foi então incubado à 37°C por 3 horas e para inativar a enzima, a reação foi incubada a 75°C por 10 min com a adição de 2,6 µL de EDTA 0,05 M.

6.9.4 Análise de mRNA RMRP

6.9.4.1 Síntese de cDNA (DNA complementar)

Para quantificar os transcritos de RMRP por RT-qPCR, foi sintetizada a fita de DNA complementar a partir de 2 µg do RNA tratado com DNase, utilizando-se a enzima *SuperScriptIII* (Invitrogen) e oligonucleotídeos randômicos (Invitrogen).

Ao RNA, foi adicionado 1 µL de oligonucleotídeo randômico (50ng/µL, Invitrogen) e 1 µL de mix de dNTP (10 mM); em seguida, o volume foi completado com água livre de nucleases até 10 µL. A mistura foi então incubada a 65°C por 5 min e logo após, transferida para o gelo. A reação foi preparada utilizando 2 µL de tampão apropriado 5X para *SuperScriptIII*, 4 µL de MgCl₂ (25 mM), 2 µL de DTT (0,1 M), 1 µl RNaseOUT (40U/µl) e 1 µL (200 unidades/µl) de SuperScript™ III RT. O conteúdo foi gentilmente homogeneizado e adicionado à mistura de RNA/oligonucleotídeos. Para realizar a reação, o tubo foi incubado em termociclador em uma ciclagem de 10 min a 25°C, 50 min a 50°C e a 85°C

por 5 min. Ao final, para eliminar o RNA residual, adicionou-se 1 μ L de RNase H e o tubo foi incubado a 37°C por 20 min. Os cDNAs foram armazenados em freezer -20°C até sua utilização.

6.9.4.2 Quantificação da expressão por PCR em tempo real

A técnica de PCR em tempo real foi padronizada, utilizando o equipamento *ABI 7500 Fast* (Applied Biosystems) com o objetivo de analisar e quantificar a expressão do gene *RMRP*. Para isto, conjuntos de oligonucleotídeos foram desenhados utilizando os *softwares* Clone Manager v.9.2e OligoAnalyzer com base nas sequências do RNA mensageiro disponíveis no banco de dados GenBank (www.ncbi.gov). Após a seleção dos iniciadores, foi feita uma análise utilizando a ferramenta Primer BLAST disponível em NCBI (<http://www.ncbi.nlm.nih.gov/>) para analisar a especificidade dos iniciadores.

Previamente ao estudo, as eficiências de amplificação do gene alvo e dos genes de referência foram determinadas pelo *7.500 Software* através da curva de eficiência de amplificação, visando eficiências similares para todos os genes. Foram testados cinco genes para controles internos: *GPI*, *GAPDH*, *TOP1*, *HPRT1* e *UBC*. As curvas padrão foram preparadas com cinco pontos, em triplicata e com fator 5 de diluição, partindo de 20 ng de cDNA. Os dois genes com eficiência mais próxima do gene alvo foram selecionados para serem utilizados como normalizadores. Todos os iniciadores usados e as respectivas concentrações na reação estão descritos na tabela do anexo 11.2.

Após a escolha dos genes de normalização, as reações foram feitas em triplicata com 20 ng de cDNA dos pacientes e de indivíduos controles, 7,5 μ L de *Power SYBR Green PCR Master Mix* (Applied Biosystems), 0,5 μ M de cada

oligonucleotídeo, em volume final de 15 μL . O perfil térmico da reação consistiu em 2 min a 50°C, 10 min a 95°C, e 40 ciclos de dois passos: 15 seg a 95°C e 1 min a 60°C. Ao final da ciclagem, foi incluída uma etapa de dissociação das curvas dos produtos finais a para avaliar a especificidade das amplificações. A normalização dos dados de expressão foi dada pela média aritmética dos valores de expressão dos mRNAs de GAPDH e TOP1 nas amostras do grupo controle.

6.9.5 Análise dos pequenos RNAs derivados do RNA RMRP

6.9.5.1 Síntese de cDNA (DNA complementar)

O RNA tratado com DNaseI no item 6.11.3 também foi utilizado para sintetizar a o cDNA específico dos alvos RMRP-S1 e RMRP-S2; e do normalizador miR-16, utilizando o *kit TaqMan™ MicroRNA Reverse Transcription* (ThermoFisher).

Primeiramente, foi preparada uma mistura contendo 0,15 μL de mix de dNTP (100 mM), 1 μL de enzima MultiScribe™ Reverse Transcriptase (50 U/ μL), 1,5 μL de tampão de transcrição reversa (10 \times), 0,19 μL de inibidor de RNase (20 U/ μL), completando com água livre de nucleasse para o volume de 7 μL . Em seguida foi adicionado 100 ng do RNA em um volume de 5 μL . Para completar o preparo, 3 μL de oligonucleotídeo alvo-específico integrante do ensaio customizado (ThermoFisher) foi pipetado junto ao conteúdo do tubo e homogeneizado gentilmente. A ciclagem foi programada conforme descrito a seguir:

Tempo	Temperatura
30 min	16°C
30 min	42°C
5 min	85°C
∞	4°C

Por fim, o cDNA foi armazenado em freezer – 20°C.

6.9.5.2 PCR quantitativo em tempo real

Os níveis de expressão dos dois pequenos RNAs derivados de RMRP previamente identificados em publicação, conhecidos por regular a expressão de diversos genes, foram avaliados pela técnica de PCR em tempo real utilizando o equipamento ABI 7500 Fast (Applied Biosystems). Para esta análise foi utilizado o ensaio customizado TaqMan desenvolvido pela empresa Applied Biosystems, que é constituído por um par de iniciadores em *stem loop* e uma sonda marcada com um fluoróforo. As RT-qPCRs foram normalizadas em relação aos níveis de transcritos normalizadores de miR-16, baseado no ensaio descrito por Rogler et al., 2015¹².

Previamente ao estudo, as eficiências de amplificação do gene alvo e do gene de referência foram determinadas por curva-padrão de amplificação, visando eficiências similares para todos os iniciadores. Para esses experimentos foram realizadas diluições seriadas de quantidades conhecidas de cDNA (6 ng, 3 ng, 1,5 ng, 0,75 ng, 0,375 ng). As reações para construção da curva padrão foram realizadas, utilizando 1 µL do cDNA, 5 µL de TaqMan™ Universal Master Mix II com UNG (Applied Biosystems) e 0,5 µL do ensaio customizados TaqMan (20X) contendo sonda e iniciadores específicos em volume final de 10 µL. Inicialmente, as condições de ciclagem para todos

os alvos foram: 2 min a 50°C, 10 min a 95°C, e 40 ciclos de duas etapas, sendo uma de 15 seg a 95°C e uma de 1 min a 60°C. Entretanto, as eficiências de amplificação dos pequenos RNAs RMRP-S1 e RMRP-S2 precisaram ser otimizadas através de um ajuste no perfil térmico da reação. Foi incluída uma etapa de anelamento separadamente da extensão, sendo definido: uma etapa inicial de 2 min a 50°C, desnaturação a 95°C/10 min, seguida de 40 ciclos de desnaturação a 95°C/15s, anelamento a 60°C/30 seg e extensão a 72°C/30s.

Após a determinação da eficiência dos iniciadores, os ensaios de quantificação dos pequenos RNAs derivados de RMRP nas amostras dos pacientes e de indivíduos controles foram realizados seguindo as mesmas condições estabelecidas. Todas as reações, inclusive da curva de eficiência, foram feitas em triplicatas.

6.9.6 Análises estatísticas

A normalização dos dados de expressão tanto para os pequenos RNAs, quanto para lncRNA, foi dada pela média aritmética dos valores de expressão dos genes de referência nas amostras do grupo controle. Os cálculos de quantificação relativa foram baseados no método do DDC_T de Livak e Schmittgenb (2001)¹²⁰, conforme a fórmula a seguir:

$$FC = E^{-\Delta\Delta C_T}$$

Em que:

- $\Delta C_T = C_T$ alvo - C_T do gene normalizador
- $\Delta\Delta C_T = \Delta C_T$ amostra - ΔC_T do calibrador
- E = eficiência do oligonucleotídeo
- FC = razão de expressão relativa (do inglês, fold change)

As análises estatísticas foram realizadas utilizando-se o programa GraphPad Prism 8 (GraphPad Software, CA). Primeiramente, o teste de Grubbs foi aplicado para detectar possíveis outliers em cada grupo. Para verificar a normalidade da distribuição dos dados foi utilizado o teste de Shapiro-Wilk. Por fim, os dados foram analisados através do teste não paramétrico de Mann Whitney para comparação de dois grupos, com um nível de significância $\leq 0,05$ (* $p \leq 0,05$; ** $p \leq 0,01$; *** $p \leq 0,001$; **** $p \leq 0,0001$).

6.10 Análise proteômica de leucócitos provenientes de pacientes

Com o objetivo de investigar possíveis alterações da expressão de genes alvos dos pequenos RNAs derivados de RMRP, foi feita a análise de proteômica a partir de leucócitos.

6.10.1 Extração e quantificação de proteínas dos leucócitos

O tubo contendo os leucócitos em PBS obtidos conforme o item 6.11.1 foram centrifugados a 1.000 rpm por 5 minutos e o sobrenadante foi descartado. Foram então adicionados 200 μ L de Tampão de Extração de Proteínas Solúveis ao pellet (KCl 500mM; MgCl₂ 5mM; DTT 1mM; EDTA 0.5mM; Glycerol 35%; HEPES 100mM) e 12 μ L de inibidor de protease. (cOmplete™ protease inhibitor cocktail - Roche). As amostras foram armazenadas em freezer -80°C até o momento de sua utilização. Para finalizar a extração de proteínas, os extratos proteicos foram descongelados, promovendo o rompimento das células¹²¹ e os tubos foram vortexados (LP Vortex Mixer - ThermoFisher) por 1 minuto e centrifugados por 30 minutos a 14.000 x g a 4°C (Heraeus™ Fresco™ 17 Microcentrifuge - ThermoFisher). Em seguida, o sobrenadante foi separado em

um novo microtubo e as proteínas foram quantificadas pelo método fluométrico utilizando o kit de quantificação de proteínas *Qubit™ Protein Assay Kit* (Life Technologies), seguindo o protocolo do fabricante.

6.10.2 Preparo das amostras e digestão de proteínas

Após a quantificação, foi realizada a troca do tampão de extração de proteínas pelo tampão de digestão. Para isso, cerca de quinhentos microgramas de extrato proteico dos leucócitos foram adicionados em colunas Amicon® Ultra 2 mL Centrifugal filters ultracel 3k (Millipore), cujo pH foi previamente equilibrado. Em seguida, o volume foi ajustado para 2mL com tampão bicarbonato de amônio [50 mM] e os tubos foram submetidos à centrifugação a 4.000 x g por 10 minutos a 16°C (Heraeus Cryofuge 5500i - ThermoFisher), de acordo com as instruções do fabricante. O eluato foi transferido para um novo microtubo e novamente quantificados, utilizando o kit de quantificação de proteínas *Qubit™ Protein Assay Kit* (Life Technologies).

Foram digeridos 42,6 µg de proteínas, em solução, de acordo com o descrito por PIZZATTI et al., 2012¹²¹ conforme descrito a seguir. Para a desnaturação de proteínas, adicionou-se 25 µL do reagente RapiGest SF Surfactant (Waters) às amostras que, em seguida, foram incubadas a 80°C por 15 minutos (ThermoMixer® C - Eppendorf). Logo após, foram adicionados 10 mM do reagente Dithiothreitol - DTT (GE Healthcare) em cada tubo e estes foram mantidos a 60°C por 30 minutos para reduzir as pontes dissulfeto das proteínas. A alquilação foi realizada adicionando 40 mM de Iodoacetamida - IAA (BioRad) e as amostras foram incubadas por 30 minutos à temperatura ambiente, sob abrigo da luz. Posteriormente, foi adicionada tripsina (Promega) em uma

proporção 1:100 do volume total da amostra em cada tubo, seguida de incubação a 37°C por 21 horas sob agitação de 1000 rpm. A reação de digestão foi interrompida adicionando ácido trifluoroacético (TFA, do inglês: Trifluoroacetic acid) 1%. Por fim, as amostras foram vortexadas e centrifugadas a 10.000 x g por 10 minutos e o sobrenadante foi transferido para um novo tubo.

6.10.3 Purificação e secagem dos peptídeos trípticos

Para remoção de impurezas e concentração dos peptídeos, foram montadas mini-colunas de fase reversa inserindo um disco de membrana C18 no interior de uma ponteira de volume de 200 µL sem barreira (uma coluna por amostra). Logo após, foram adicionados 10 µL de resina POROS™ R2 (ThermoFisher) diluída em acetonitrila (ACN) 100%, e o líquido foi retirado por pressão com o auxílio de uma seringa acoplada a ponteira. Posteriormente, adicionou-se 100 µL de ACN 100% para auxiliar na retirada de eventuais bolhas formadas na resina retida na ponteira. Novamente o líquido foi removido com a seringa. Para equilibrar o pH da coluna, foi adicionado 100 µL de TFA 0,1% e o líquido foi eliminado com utilizando a seringa. Esse processo foi repetido três vezes. Ao final dessa etapa, a coluna foi acondicionada em um microtubo para dar início a purificação das proteínas. A amostra digerida foi então aplicada na coluna e filtrada com o auxílio da seringa. O volume recuperado foi aplicado novamente na coluna, a fim de garantir a aderência dos peptídeos da amostra. Em seguida, a coluna foi lavada três vezes com 100 µL de TFA 0,1%. A eluição dos peptídeos de interesse aderidos à coluna foi feita em duas etapas. Na primeira, 100µL de uma solução composta por 50% ACN/0,1% TFA foram adicionados na coluna posicionada sobre um novo microtubo e pressionados a

passar sobre a coluna utilizando a seringa. A segunda eluição foi realizada com 100µL de uma solução constituída de 70% ACN/0,1% TFA para eluir os peptídeos mais hidrofóbicos. Por fim, as amostras foram submetidas à secagem na centrifuga a vácuo (Savant® DNA Speed Vac® Model 120 - ThermoFisher) e quando completamente secas, foram armazenadas no freezer -30°C até o momento de injeção no espectrômetro de massas.

6.10.4 Espectrometria de massas

As amostras foram ressuspendidas na concentração final de 0,2 µg/µL com ácido fórmico 0,1% e analisadas em um cromatógrafo easynanoLC1000 acoplado ao espectrômetro de massas de alta resolução Q-Exactive Plus com analisador Orbitrap (ThermoFisher). Foram injetados 2 µg de amostra por injeção em duplicatas técnicas. Foi utilizada a coluna trap (Easy column C18, 2cm x 100µm i. d. x 5µm, 120Å) para carregar os peptídeos. Já a separação foi feita em fluxo constante de 300 nL/min de solventes A (5% ACN / 0,1% FA) e B (95% ACN / 0,1% FA) na coluna analítica (ReprosilPur C18, 25cm x 75µm i. d. x 3µm). O gradiente teve uma etapa inicial de 133 minutos em que o percentual do solvente B foi gradualmente aumentado de 2% até 20%. Já a segunda etapa consistiu no aumento de concentração do mesmo solvente de 40% para 40% em um período de 34 minutos. Esse gradiente passou por uma terceira etapa, aumentando o percentual até 95% em 5 minutos. Para finalizar, o gradiente foi estabilizado por 8 minutos, finalizando o processo em 180 minutos. Foi utilizado o eletrospray de 2,7kV de voltagem a 250°C como fonte de ionização. O modo de aquisição dependente de dados (do inglês, data dependent acquisition, DDA) foi operado pelo equipamento com exclusão de 45ms. Os espectros MS full scan

foram gerados pelo analisador Orbitrap em resolução de 70.000 (FWHM a 400 m/z) e os 15 íons mais intensos com energia de colisão de 30% e resolução de 17.500 em scans MS/MS foram submetidos a uma fragmentação. Foram excluídas da análise MS/MS as espécies com carga +1.

6.10.5 Análise dos dados brutos e identificação das proteínas

Os dados brutos resultantes da corrida no espectrômetro de massas foram processados e analisados no *software* Proteome Discoverer 2.1 (PD; ThermoFisher) utilizando como referência as sequências humanas do banco de dados Uniprot (<https://www.uniprot.org/>)¹²².

Os peptídeos foram identificados pelo algoritmo Sequest HT com os seguintes parâmetros: clivagem específica de tripsina com no máximo 2 sítios perdidos (do inglês, missed cleavages), tamanho mínimo de seis aminoácidos por peptídeo e tolerância de massa de peptídeos de 10 ppm e de MS/MS de 0,05 Da. As modificações químicas foram consideradas fixas para carbamidometilação da cisteína e dinâmicas para oxidação de metionina e acetilação N-terminal da proteína. As taxas de falso descobrimento (do inglês, false discovery rate, FDR) foram obtidas utilizando o Percolator node e os critérios foram ajustados para selecionar apenas peptídeos com q-valor igual ou inferior a 0,01. Baseado nesse parâmetro foi considerado como limite um único peptídeo com FDR de 0,01 por proteína.

6.10.6 Análises de resultados de proteoma *in silico*

Após o processamento, as proteínas identificadas foram analisadas qualitativamente, utilizando a ferramenta Venny 2.1

(<https://bioinfogp.cnb.csic.es/tools/venny/>)¹²³ e quantitativamente, utilizando o *software* Perseus Computational Plataform.

Inicialmente, os dados resultantes da análise do PD (item 6.12.5) foram incluídos no programa Perseus para geração de planilhas com o número de identificação e nome das proteínas para cada replicata. Essas informações foram utilizadas para a construção do diagrama de Venn, possibilitando a identificação de proteínas exclusivamente presentes nos pacientes e nos controles, além daquelas comuns a ambos os grupos.

Para enriquecimento da análise funcional e das redes de interação de genes e proteínas foi empregado o *software* FunRich (*Functional Enrichment Analysis Tool*)^{124,125,126} e sua base de dados própria, que é uma compilação de diversos bancos de dados funcionais e de anotação de proteínas de forma curada. Utilizando as proteínas exclusivas e que apresentaram diferença de expressão de cada grupo foram feitas análises funcionais quanto a processos biológicos, vias de sinalização, locais de expressão e fatores de transcrição reguladores do perfil proteômico identificado. O banco de dados Reactome (<https://reactome.org/>)¹²⁷ e o *software* IPA (*Ingenuity Pathway Analysis*) também foram utilizados para verificar as diversas vias e processos biológicos aos quais as proteínas diferencialmente expressas estavam associadas. Ao final, as informações funcionais mais relevantes de cada grupo foram destacadas e discutidas.

7. Aspectos Éticos

Este projeto foi aprovado no Comitê de Ética em Pesquisa do IFF (nº do parecer de aprovação: 1.339.799) e das instituições coparticipantes, UNICAMP (nº do parecer de aprovação: 1.383.346) e HCPA (nº do parecer de aprovação: 1.569.797), objetivando seguir a Resolução 466/12, que trata das diretrizes e normas de pesquisas envolvendo seres humanos (anexo 1).

8. Resultados

8.1 Análise de variantes gene RMRP das amostras de pacientes com suspeita clínica de CHH

A coorte estudada atualmente é composta por 32 pacientes brasileiros, provenientes de diferentes regiões do Brasil (nordeste, sudeste, distrito federal, sul; com diagnóstico clínico-radiológico de CHH, sendo 23 avaliados previamente por nosso grupo^{4,5} e 9 novos pacientes recrutados no presente trabalho. Do total, 13 amostras eram provenientes do IFF, 7 da UNICAMP e 12 do HCPA. Foi observada recorrência familiar em 2 famílias, cada uma com 2 irmãos afetados; e, um caso de consanguinidade declarada (primos em primeiro grau). O resumo dos genótipos de cada paciente está descrito na tabela abaixo.

Tabela 2: Sumarização dos resultados das análises do gene RMRP dos pacientes.

ID da amostra	Instituição de origem	Região de nascimento	Alelo 1	Alelo 2
1	IFF	SE	g.-21_-2dup20(ACTCTGTGAAGCTGAGGACG)	g.196C>T
2	IFF	SE	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
3	IFF	SE	g.-19_-3trip17(TCTGTGAAGCTGAGGAC)	g.194 G>A
4	IFF	SE	g.-24_-15dup10(ACTACTCTGT)	g.97_98dup2(TG)
5	IFF	SE	g.-22_-14dup9(TACTCTGTG)	g.196C>T
6	IFF	SE	g.97_98dup2(TG)	g.196C>T
7	IFF	SE	g.10T>A	g.196C>T
8	IFF	SE	g.-22_-3dup20(TACTCTGTGAAGCTGAGGAC)	g.196C>T
9	IFF	SE	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
10	IFF	SE	g.-6_+4dup10(GGACGTGGTT)	g.71A>G
11	IFF	SE	g.196C>T	g.245G>C
12	IFF	SE	g.-19_-3dup17(TCTGTGAAGCTGAGGAC)	g.196C>T
13	IFF	SE	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
14	UNICAMP	SE	g.-19_-3dup17(TCTGTGAAGCTGAGGAC)	g.181G>A

ID da amostra	Instituição de origem	Região de nascimento	Alelo 1	Alelo 2
15	UNICAMP	SE	g.6G>T	g.196C>T
16*	UNICAMP	NE	g.36C>G	g.252A>C
17*	UNICAMP	NE	g.36C>G	g.252A>C
18	UNICAMP	SE	g.-13_+1dup14(AAGCTGAGGACGTG)	g.196C>T
19	UNICAMP	NE	g.-8_-7ins20(GGGACTACTCTGTGAAGCTG)	g.196C>T
20	UNICAMP	S	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.97_98dup2(TG)
21	HCPA	S	g.97_98dup2(TG)	g.147G>C
22	HCPA	S	g.6G>T	g.196C>T
23	HCPA	S	g.-22_-1dup20(TACTCTGTGAAGCTGAGAAC) + g.261_331dup70	g.196C>T
24*	HCPA	S	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
25*	HCPA	S	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
26	HCPA	S	g.36C>G	g.98G>A
27	HCPA	S	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
28	HCPA	S	g.69_70ins1(T)	g.196C>T
29	HCPA	S	g.-20_+3dup23(CTCTGTGAAGCTGAGAACGTGG T)	g.5C>T
30	HCPA	S	g.36C>T	g.196C>T
31	HCPA	S	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)	g.196C>T
32	HCPA	S	g.-24_-10dup15(ACTACTCTGTGAAGC)	g.196C>T

Legenda: *paciente que possui irmão afetado com mesmo genótipo; em negrito estão sinalizados os pacientes novos da coorte. SE: sudeste; NE: nordeste; S: Sul.

O genótipo predominante dos 9 novos pacientes estudados foi uma heterozigose composta de uma alteração na região promotora ou na região transcrita associada à variante g.196C>T; exceto, pelo paciente 10, que apresentou a alteração g.71A>g associada à uma duplicação na porção regulatória do gene *RMRP*. Desse modo, a frequência da variante g.196C>T na nossa coorte passou de 62% (13/21) para de 70% (21/30), considerando apenas

um propósito nos casos familiares. Além disso, não foi observado nenhum caso de homozigose.

8.2 Análise estrutural das variantes no gene *RMRP*

Treze mutações foram mapeadas na estrutura secundária do RNA RMRP selvagem. Domínios e nucleotídeos conservados da RMRP estão descritos na figura 25. Entre as mutações identificadas, sete estão em regiões de emparelhamentos de bases do tipo Watson-Crick: 5C-U, 6G-U, 10U-A, 98G-A, 97_98dupUG, 147G-C, 181G-A. Com exceção da alteração 5G-U, todas encontram-se em nucleotídeos conservados. A variante 35C-G está no domínio P3.

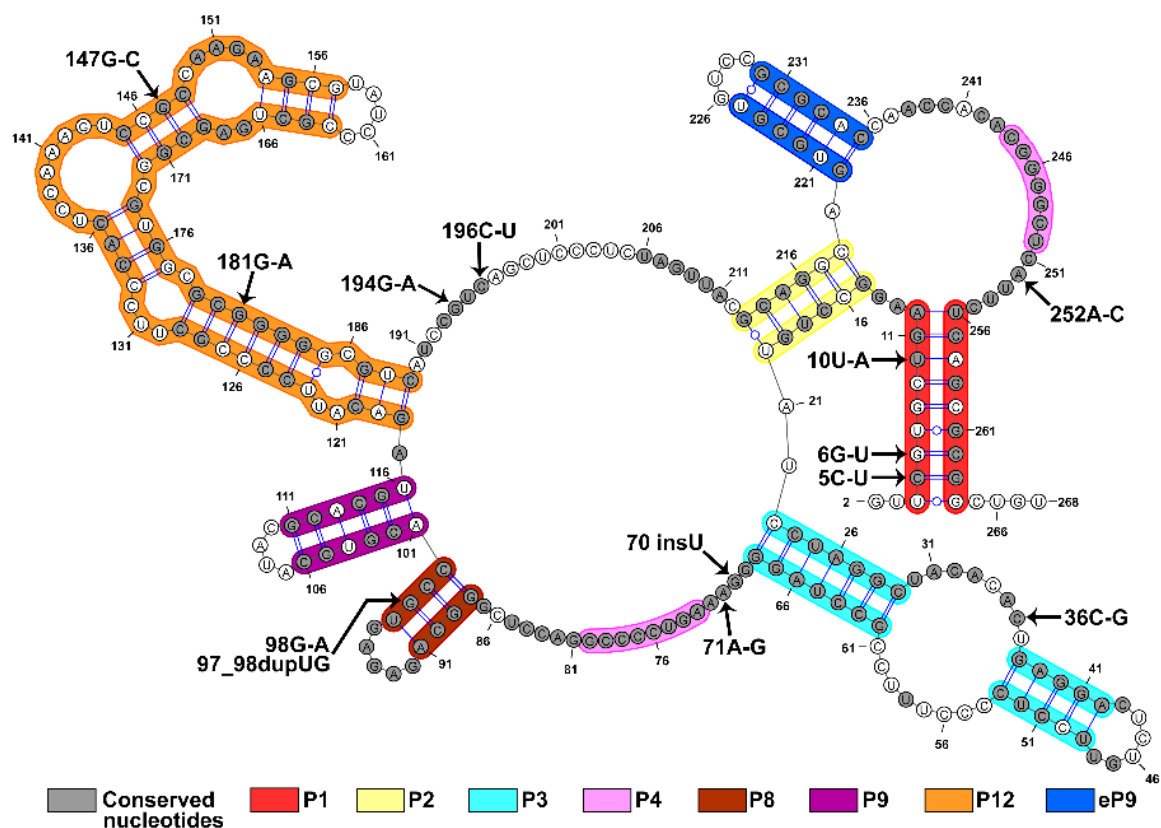


Figura 25: Variantes genéticas apontadas na estrutura secundária do RMRP. A anotação dos domínios da estrutura estão sinalizadas em cores diferentes.

Neste trabalho, foi utilizada uma abordagem de simulação do possível dobramento de RNA para analisar as consequências das variantes genéticas encontradas nos pacientes na estrutura do transcrito RMRP.

Em razão das trocas de nucleotídeos, as variantes 6G-U e 10U-A (figura 26A), 98G-A (figura 26C), 147G-C e 181G-A (figura 26D) mostraram desfazer ligações de Watson-Crick nas regiões de pareamento, comprometendo a topologia da estrutura de *hairpin* (grampo). Por outro lado, a alteração 5C-U mantém a interação entre as bases, porém torna-a mais fraca devido à perda de uma ligação de hidrogênio (o pareamento C-G apresenta três ligações; enquanto, o pareamento U-G apresenta apenas duas ligações; figura. 26A). Já a variante 97_98dupUG (figura 26C), provoca um aumento da alça (loop) da estrutura em grampo do domínio P8.

As alterações 36C-G (figura 26B), 70insU, 71A-G, 194G-A, 196C-U e 252A-C não demonstraram alterar a topologia da estrutura secundária, apesar de ocorrerem em regiões conservadas.

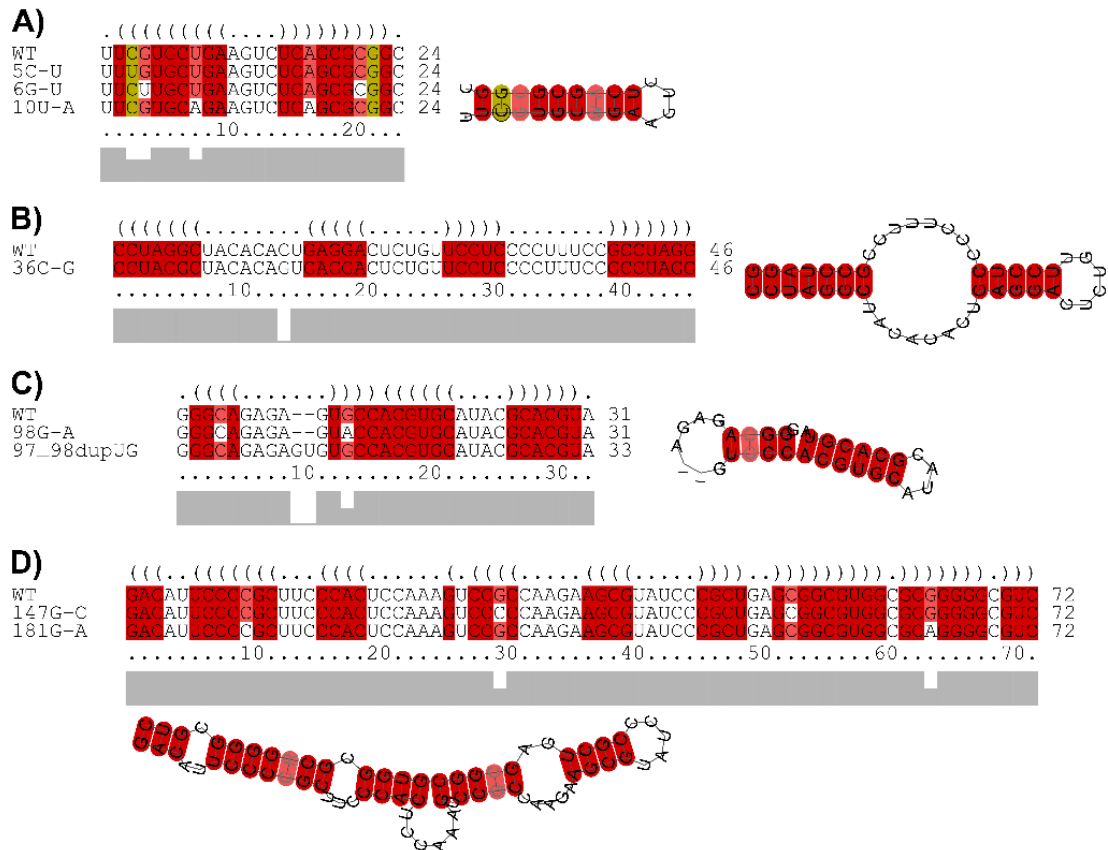


Figura 26: Caracterização das variantes localizadas nos domínios da RNase MRP. Esquerda está o alinhamento entre a sequência selvagem e mutante. Direita está a estrutura secundária consenso do alinhamento. **A)** Domínio P1; **B)** Domínio P3; **C)** Domínio P8 e P9; e **D)** Domínio P12.

Por fim, as alterações 5C-U, 6G-U e 10U-A (presente nos pacientes 7, 15, 22, 29) ocorrem na região que gera o pequeno RNA RMRP-S1 e a alteração 147G-C (paciente 16) está contida na região relacionada a biogênese do pequeno RNA RMRP-S2.

8.3 Investigação do efeito fundador da variante g.196C>T no gene *RMRP* na população brasileira de pacientes com CHH

8.3.1 Análise do perfil de ancestralidade genética dos pacientes com CHH

Com o objetivo de inferir das porcentagens genômicas de diferentes ancestralidades (Africana, Europeia e Nativo Americana) de forma individual e global na coorte estuda, foram genotipados os marcadores do tipo AIMS (*ancestry informative markers*) e comparados com as amostras de referência das populações ancestrais do painel de diversidade HGDP-CEPH utilizando o *software Structure*. Essa análise foi realizada utilizando amostras de 21 pacientes e a contribuição genômica média de cada população parental (Africana, Europeia e Nativo Americana) no grupo total de pacientes foi de 24%, 58% e 18%, respectivamente (figura 27; tabela 3). Os dados de ancestralidade genômica a nível individual encontram-se na tabela do anexo 11.3.

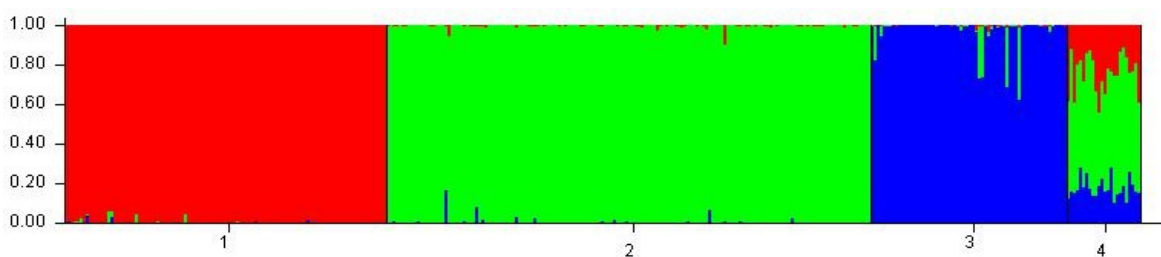


Figura 27: Proporções da ancestralidade individual obtidas pelo programa *STRUCTURE*. As cores no gráfico representam o percentual de marcadores AIM-Indels associados à cada população ancestral (africanos: vermelho; europeia: verde; nativo americana: azul) presente no material genético dos indivíduos (representados por linhas verticais). As cores das linhas dependem da fração de ancestralidade estimada para cada indivíduo. 1-3) amostras de referência do painel diversidade HGDP-CEPH (Africanos; Europeus; Nativo Americanos, respectivamente); 4) pacientes do estudo.

Neste tipo de análise, o *software* utiliza populações de referência para inferir o componente ancestral de cada população parental nos indivíduos teste. Os dados também foram utilizados para verificar se havia diferença do

percentual de marcadores genéticos das populações ancestrais entre pacientes com e sem a alteração molecular de interesse g.196C>T, observando-se que as médias de ancestralidade entre esses dois grupos não apresentaram diferenças com significância estatística ($p=0,4929$; $p=0,2187$; $p=0,2673$; para africanos, europeus e ameríndios, respectivamente).

Tabela 3: Médias individuais de ancestralidade analisando com 3 populações parentais, geradas pelo Software Structure.

Amostras da população		AFR	EUR	AME	Nº de indivíduos
População pura	AFRICANOS (AFR)	0,996	0,002	0,002	105
	EUROPEUS (EUR)	0,002	0,994	0,004	158
	AMERÍNDIOS (AME)	0,002	0,024	0,975	64
População de estudo	PACIENTES CHH (GERAL)	0,240	0,585	0,175	21
	Pacientes com g.196C>T	0,251	0,564	0,185	13
	Pacientes sem g.196C>T	0,221	0,619	0,160	8

Legenda: A tabela mostra as médias de proporções ancestrais das diferentes populações ancestrais (africana, europeia e ameríndia) e da população teste (pacientes com e sem a alteração g.196C>T).

8.3.2 Análise de haplótipos

Como a presença da variante g.196C>T foi recorrente em pacientes de diferentes regiões do Brasil, além de ser pouco frequente em outras regiões do mundo, a hipótese era de que essa mutação teria uma única origem ancestral. Para determinar se essa variante era derivada de um ancestral comum, a análise dos haplótipos foi realizada em amostras de 27 pacientes, sendo 18 deles carreadores da variante g.196C>T (tabela 2). Era esperado que irmãos da mesma família tivessem o mesmo haplótipo; e, por isso, foram contabilizados uma única vez. Além disso, a amostra de uma paciente que apresentou a variante g.196C>T não estava disponível para o estudo e o sequenciamento dos

marcadores para a montagem do haplótipo de duas pacientes (também carreadoras da variante g.196C>T) não foi concluído até a finalização da redação desta tese. Inicialmente, foi feita uma análise descritiva das frequências alélicas e haplotípicas dos marcadores bialélicos selecionados nas populações africanas e europeias do 1000Genomas (amostras controle de bancos de dados). Esta primeira análise sugeriu a existência de 10 haplótipos diferentes e revelou que a região delimitada de interesse seria discriminatória entre europeus e africanos: o haplótipo majoritário em indivíduos europeus (C/C/G/C) apresentou frequência de 0,356 na Europa e 0,075 na África; enquanto, o haplótipo majoritário na África (C/C/A/C) apresentou frequência de 0,393 em africanos e 0,005 em europeus (tabela 4).

O resultado do sequenciamento dos tag SNPs na coorte estudada foi utilizado para montagem dos haplótipos e para o cálculo das frequências alélicas e haplotípicas. Esses dados estão sumarizados na tabela 4, tabela do anexo 11.4 e na figura 28. A montagem do haplótipo foi determinada pela análise da origem parental dos alelos para 14 pacientes e por clonagem para 13 pacientes.

Tabela 4: Frequências haplotípicas dos TAG SNPS selecionados para as populações Europeia e Africana.

rs10972552	rs7021463	rs1339374	rs1361338	Nº de observações na coorte			Frequência na coorte	Frequência em europeus*	Frequência em africanos*
				Total	Com g.196C>T	Sem g.196C>T			
T	C	G	A	30	17	13	0,537	0,325	0,148
T	C	G	C	1	0	1	0,037	0,001	0,016
T	G	G	C	9	0	9	0,148	0,129	0,350
T	G	G	A	0	0	0	0,037	0,002	0,000
C	C	A	A	4	0	4	0,074	0,177	0,013
C	C	G	C	8	1	7	0,130	0,356	0,075
C	C	G	A	1	0	1	0,0185	0,003	0,004
C	C	A	C	1	0	1	0,0185	0,005	0,393
C	G	G	C	0	0	0	0,000	0,002	0,001
T	C	A	C	0	0	0	0,000	0,000	0,001
Total					18	36	1,000	1,000	1,000

Legenda: Em cinza estão marcados os haplótipos mais frequentes na população europeia e africana. *dados do banco de dados 1000G (N=504 EUR; N=503 AFR).

Os resultados da análise indicaram que um total de 7 haplótipos distintos ocorrem no cromossomo 9 nas famílias brasileiras de pacientes com CHH estudadas. Dos 10 haplótipos estimados pelo Haploview, três não foram encontrados: T/G/G/A (frequência de 0,02% apenas em europeus), C/G/G/C (frequência de 0,1% apenas em africanos) e T/C/A/C (frequência de 0,2% europeus e 0,1% em africanos).

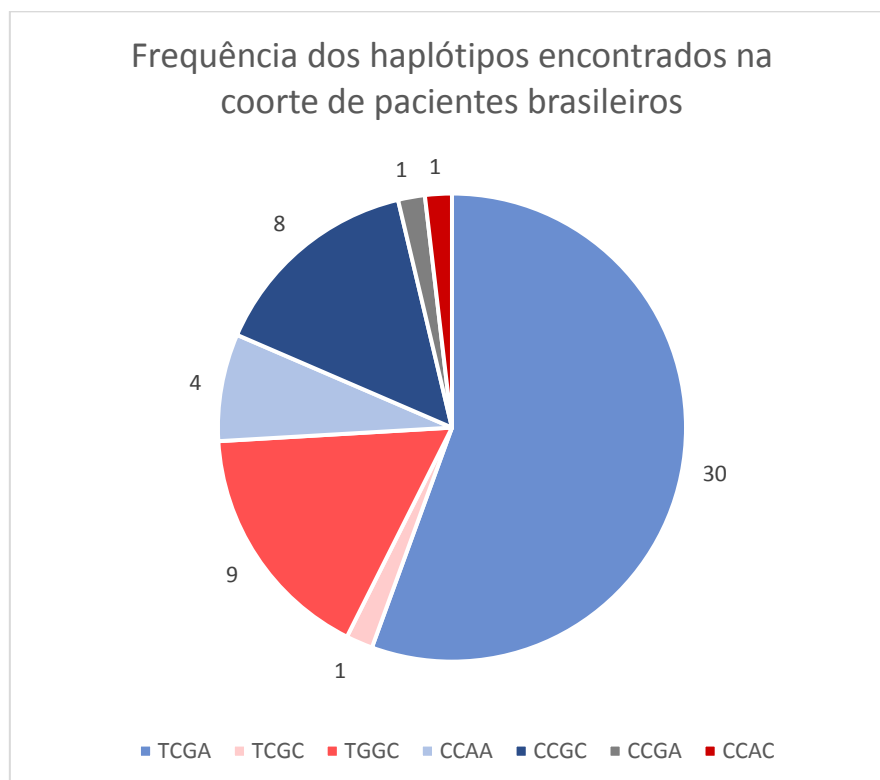


Figura 28: Distribuição da frequência absoluta dos haplótipos identificados na coorte estudada em relação à frequência dos haplótipos baseado no 1000G. As frequências absolutas na coorte estão representadas pelas proporções entre as seções do círculo. O gradiente de cores representa as frequências dos haplótipos no banco de dados, sendo as mais claras referentes às frequências maiores e as mais escuras, referentes às menores. Em azul: frequências mais comuns em europeus. Em vermelho: frequências dos haplótipos mais comuns em africanos. Em cinza, frequência semelhante para ambas as populações.

Com o objetivo de entender a distribuição dos diferentes haplótipos no Brasil, foram representados no mapa as proporções de cada tipo de haplótipo observado de acordo com o estado de origem dos pacientes (figura 29). Vale

ressaltar, que não foi identificado nenhum “cluster” (aglomerado) ou comunidade com alta prevalência de CHH.

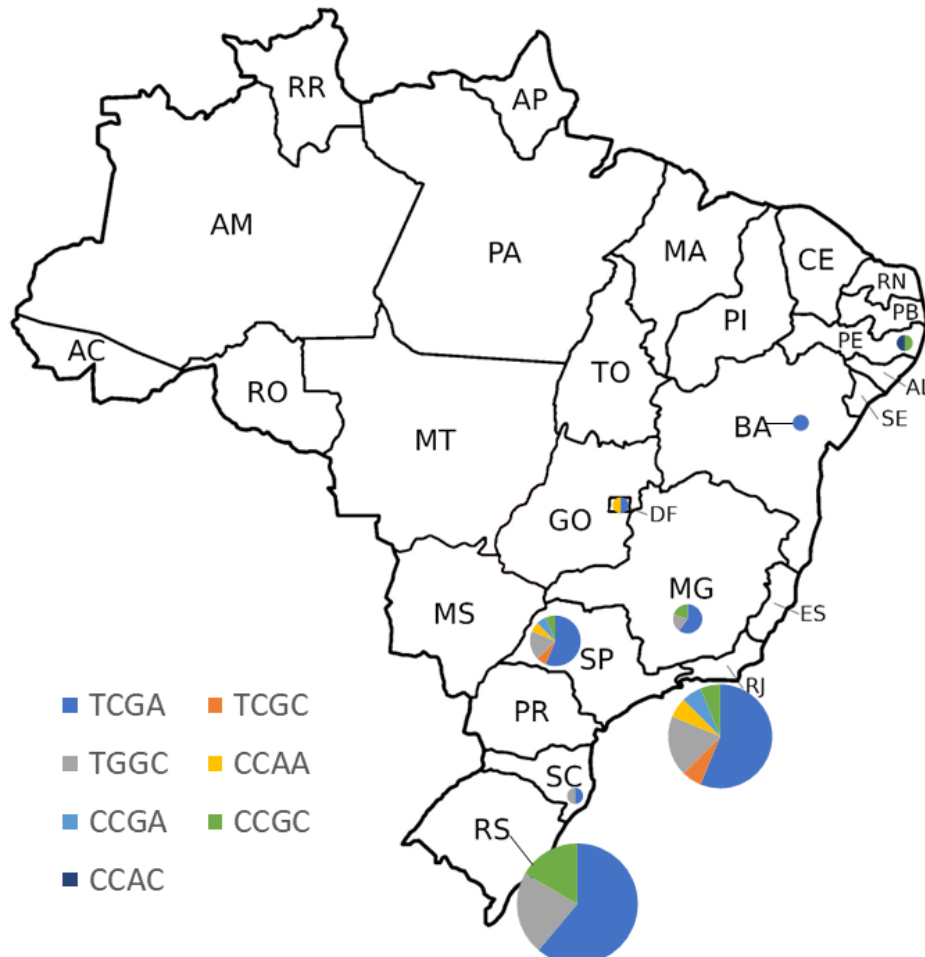


Figura 29: Distribuição dos haplótipos por estado. O tamanho dos círculos é proporcional ao número total de haplótipos de acordo com o estado de origem dos pacientes. As cores dos gráficos indicam a proporção de cada tipo de haplótipo observado em cada estado.

Dois haplótipos estavam em associação à variante g.196C>T; enquanto, as demais variantes ocorriam distribuídas entre os demais haplótipos identificados (tabela 4. Todos os haplótipos associados à variante g.196C>T estudados em pacientes brasileiros foram do tipo T/C/G/A (n = 17; 94,4%), exceto para o indivíduo 30, que apresentou o tipo C/C/G/C (5,6%). O haplótipo T/C/G/A foi o mais frequente nas famílias estudadas (30/54; 55,5%), estando

também presente em 13 dos 36 cromossomos carreadores de outras variantes analisados (36,1%). A frequência do haplótipo T/C/G/A nos bancos de dados foi estimada em 32,5% para europeus, sendo esse valor muito semelhante ao observado para os haplótipos que não contém a variante g.196C>T (13/36; 36,1%).

Agrupando os haplótipos por tipo de alteração no gene *RMRP*, foi possível fazer algumas observações interessantes. Dentre as seis contagens da variante g.-26_-5dup22(TACTACTCTGTGAAGCTGAGAA), todos estavam presentes em cromossomos que apresentam o haplótipo T/G/G/C (haplótipo mais frequente em africanos). Além disso, tanto a variante g.97_98dup2(TG), quanto a variante g.6G>T estavam em completa associação com o haplótipo T/C/G/A (N=4 e N=2, respectivamente). Por fim, a alteração g.36C>T foi associada ao haplótipo C/C/G/C (N=2), majoritário na Europa e também ao haplótipo T/C/G/A (N=1).

Considerando que foram observadas 18 alterações heterozigotas na região regulatória do gene *RMRP*, buscamos inferir se as variantes deste tipo poderiam estar preferencialmente associadas à alguma etnia. Entretanto, pela contagem direta dos cromossomos, agrupando para cada população ancestral, não parece haver uma correlação específica (N=9 para EUR e N=9 para AFR).

Com o objetivo de entender a correlação filogenética entre os haplótipos, utilizamos o *software Network* para construir a rede haplotípica (figura 30). Esse diagrama facilita a visualização da distância genética entre haplótipos, especialmente nos casos em que os cromossomos contêm uma mesma variante, mas apresentam combinações distintas para o painel de marcadores selecionados.

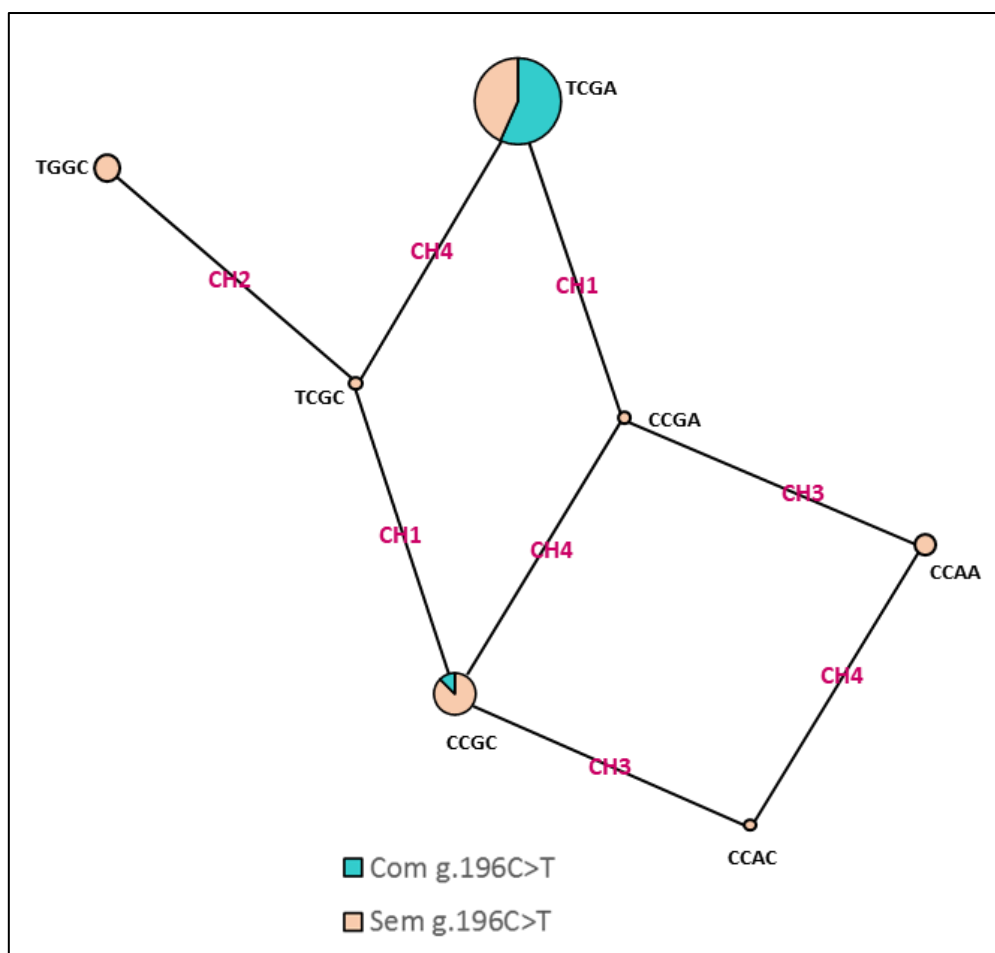


Figura 30: Rede de haplótipos elaborado no programa Network. O gráfico é composto por círculos que representam os haplótipos observados (devidamente identificados), os quais são ligados por ramos. O tamanho dos círculos é proporcional ao número de observações de cada haplótipo e foram divididos por cores em função da presença ou ausência da variante g.196C>T. Os caracteres que aparecem nos ramos correspondem aos passos mutacionais entre os haplótipos e suas numerações mostram a posição da variação de nucleotídeo nos haplótipos. CH1: representa alteração no primeiro nucleotídeo do haplótipo; CH2: representa alteração no segundo nucleotídeo do haplótipo; CH3: representa alteração no terceiro nucleotídeo do haplótipo; CH4: representa alteração no último nucleotídeo do haplótipo.

Os três haplótipos mais abundantes (T/C/G/A; T/G/G/C e C/C/G/C) estão separados por dois passos mutacionais entre si. Os marcadores que variam nessa tríade são os marcadores CH1, CH2 E CH4, permanecendo o marcador CH3 constante nos três haplótipos (guanina). Os dois haplótipos associados com a variante g.196C>T, localizados nos ramos centrais, diferenciados pelo primeiro (CH1) e último nucleotídeos (CH4).

Objetivando avaliar o quão diversa era a região selecionada no material genético da nossa coorte, foi calculada a diversidade dos haplótipos totais da coorte. Este cálculo também foi realizado na estratificação entre os grupos que abrigam a alteração g.196C>T e os que não a contêm, além das populações ancestrais africana e europeia (baseado nos dados do banco de dados 1000G), estando estes dados discriminados na tabela abaixo (tabela 5).

Tabela 5: Diversidade haplotípica calculada para diferentes grupos dentro da coorte.

Grupo	Diversidade	Total de haplótipos
Total dos haplótipos	0,64	54
Haplótipos com alteração g.196C>T	0,11	18
Haplótipos sem a alteração g.196C>T	0,76	36
Diversidade de haplótipos Europeus*	0,72	-
Diversidade de haplótipos Africanos*	0,70	-

Legenda: *calculada no item 6.10.2/6.10.6

Foi observado um nível de 0,64 de diversidade de haplótipos entre as sequências derivadas de cada amostra analisada no presente estudo. Por outro lado, quando observado especificamente o grupo de cromossomos que continham a variante g.196C>T, esse valor foi consideravelmente inferior (0,11), especialmente quando comparado ao grupo de haplótipos que não abrigavam a variante em questão (0,76). A significância estatística da associação do haplótipo T/C/G/A com a variante g.196C>T para população brasileira de pacientes com CHH foi demonstrada pelo teste exato de Fisher ($p < 0,001$). Nas populações parentais a diversidade haplotípica do painel construído foi de 0,72 na Europa e 0,70 na África.

8.4 Análise de expressão

8.4.1 Expressão do gene *RMRP*

O método quantitativo de PCR em tempo real foi realizado, utilizando amostras de leucócitos de 17 pacientes e 19 indivíduos controles saudáveis, com o objetivo de investigar se as variantes em *RMRP* interferem nos níveis de expressão deste gene. Os níveis de expressão de RNA foram normalizados usando o gene *GAPDH* e *TOP1*. Os resultados mostraram que houve uma redução significativa no perfil de expressão do gene *RMRP* no grupo dos pacientes em comparação ao grupo controle (**p=0,0093; figura 31).

A abundância relativa do lncRNA *RMRP* está reduzida em todos os pacientes analisados independentemente do genótipo detectado nesses pacientes.

A mediana de expressão nos pacientes foi de 0,125 (média e DP: 0,1374±0,105); enquanto, nos controles a mediana foi de 1,914 (média e DP: 10,55±14,600).

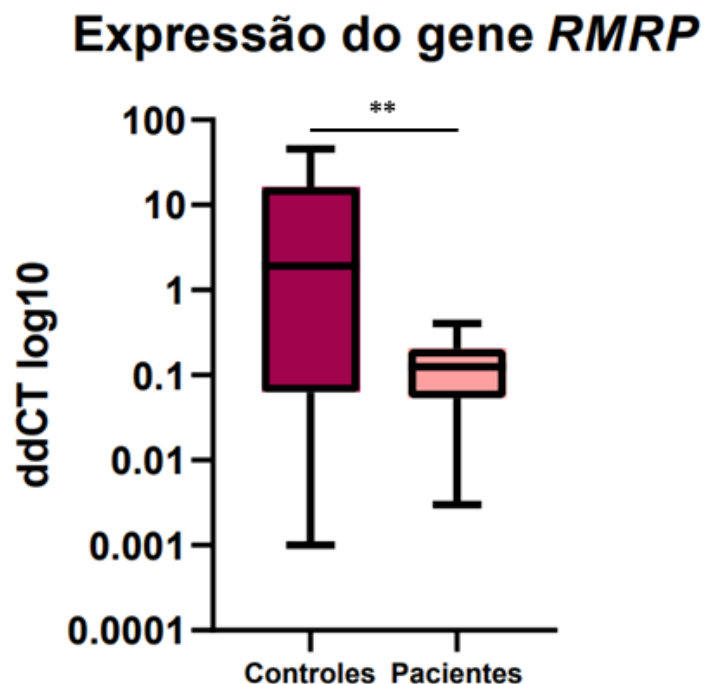


Figura 31: Expressão dos transcritos RMRP por qPCR em tempo real. Comparação da quantificação relativa dos níveis de RNA investigados no sangue periférico de pacientes com CHH (N=17) e indivíduos controles (N=19). Os valores do cálculo obtido pelo método do “CT comparativo” foram convertidos para logaritmos de base 10 (log10). No boxplot, a linha interna representa a mediana (ou quartil 50) dos valores de expressão gênica e os limites da caixa representam os quartis 25 e 75; as hastes externas indicam os valores máximos e mínimo da distribuição. Para avaliação estatística, foi utilizado o Teste Mann Whitney e as diferenças dos níveis de expressão entre o grupo de pacientes e o grupo controle foram estatisticamente significantes. (**p=0,0093).

8.4.2 Expressão dos pequenos RNAs derivados dos transcritos RMRP em leucócitos de pacientes com CHH

Com o intuito de avaliar os efeitos das variantes encontradas nos pacientes nos níveis de expressão dos pequenos RNAs, conduzimos ensaios TaqMan de PCR quantitativo em tempo real utilizando miR-16 como normalizador.

A avaliação dos pequenos RNAs foi feita nas amostras de sangue total de 9 pacientes e de 16 controles. Os dados obtidos indicam que tanto o RMRP-S1, quanto o RMRP-S2 foram diferencialmente expressos, tendo suas expressões significativamente reduzidas no grupo dos pacientes com CHH em comparação

ao grupo controle (altamente significativo, $p < 0,0001$ para ambos; tabela 6, figura 32).

Tabela 6: Expressão dos dois pequenos RNAs derivados de RMRP em pacientes com CHH e em indivíduos controles.

Grupo		Expressão de RMRP-S1	Expressão de RMRP-S2	Teste Mann Whitney
Pacientes	média (\pm DP)	0,052 (\pm 0,03)	0,047 (\pm 0,02)	**** $p < 0,0001$
	mediana	0,040	0,040	
Controles	média (\pm DP)	6,688 (\pm 13,270)	16,810 (\pm 34,21)	
	mediana	0,635	0,400	

Expressão de RMRP-S1 e RMRP-S2

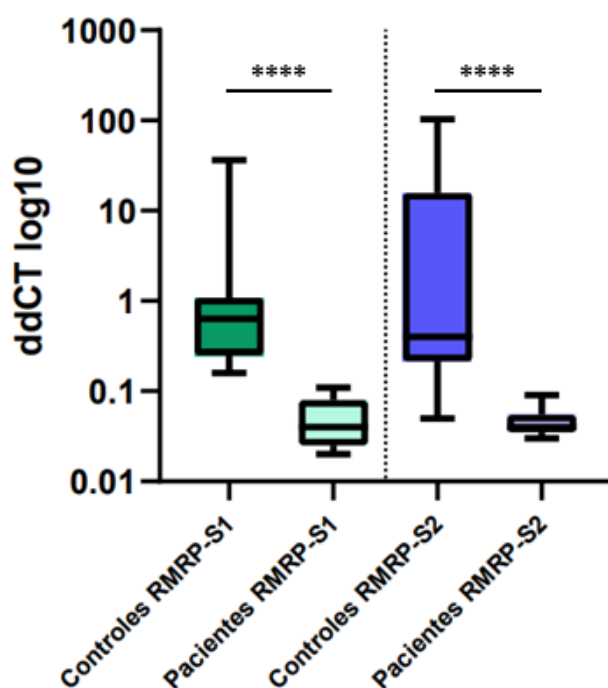


Figura 32: Avaliação da abundância relativa dos pequenos RNAs derivados de RMRP. Os gráficos mostram os valores de *fold change* ($2^{-\Delta\Delta CT}$) transformados em \log_{10} dos alvos analisados em indivíduos afetados ($N=9$) e em indivíduos controle ($N=16$). A barra dentro da caixa representa a mediana de expressão de cada gene, os limites da caixa à dispersão dos valores entre os quartis 25 e 75 e os extremos do gráfico os valores mínimo e máximo do conjunto de dados. Do lado esquerdo da linha pontilhada estão representadas as observações referentes ao RMRP-S1 e do lado direito estão representadas as observações referentes ao RMRP-S2. Os níveis de expressão nos pacientes apresentaram uma redução altamente significativa detectada pelo teste de Mann Whitney para ambos os alvos (**** $p < 0,0001$).

8.5 Análise de proteômica em leucócitos de pacientes com CHH

A fim de se obter maiores informações sobre a patogênese da CHH causada por mutações no gene *RMRP*, foram realizados experimentos de espectrometria de massa para se obter o perfil proteômico em leucócitos dos pacientes CHH. Para este estudo, 16 amostras de pacientes com CHH estavam disponíveis, além de 5 amostras controle; o perfil qualitativo de proteínas foi comparado entre esses dois grupos. Os cromatogramas resultantes da análise por LC-MS/MS das amostras de controles e pacientes em duplicatas está no anexo 11.5.

8.5.1 Análises qualitativas e quantitativas

Após o processamento dos dados brutos, foram identificadas um total de 2.784 proteínas. Utilizando como critério somente proteínas presentes em ambas as replicatas, o somatório das amostras de pacientes com CHH apresentou 1.931 proteínas; enquanto, o somatório do grupo de controle saudáveis apresentou 2.008 proteínas. Esses dados foram utilizados para a construção do diagrama de Venn, para visualizar as proteínas exclusivamente presentes em cada grupo; sendo contabilizadas 144 proteínas presentes somente nos pacientes, 186 apenas nos controles e 1.822 presentes na interseção dos dois grupos, totalizando 2.152 proteínas (figura 33). A tabela do anexo 11.6 apresenta as proteínas identificadas em cada grupo analisado nesta análise

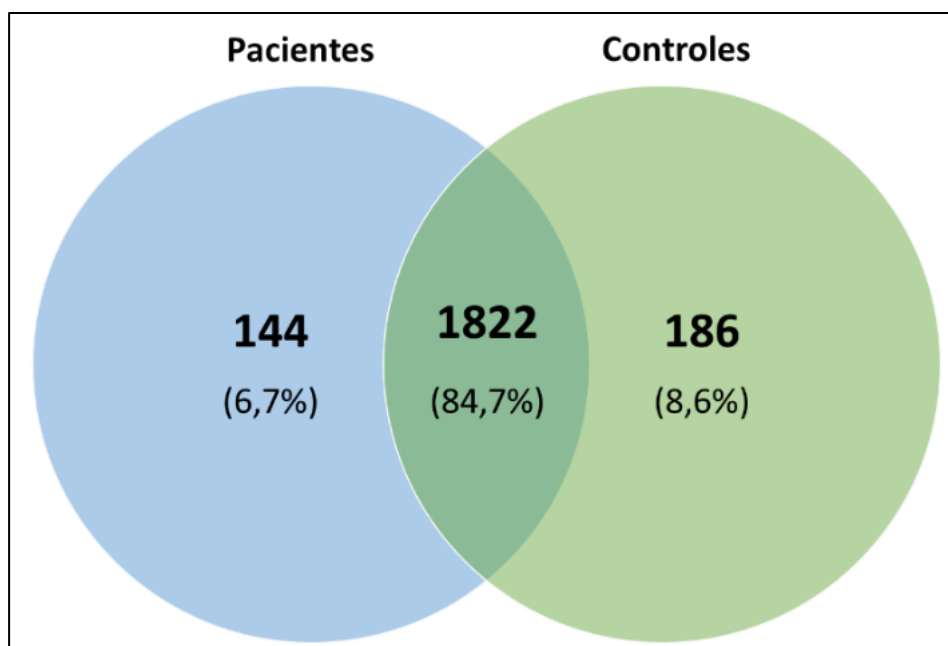


Figura 33: Diagrama de Venn das proteínas detectadas por proteômica nos grupos de pacientes e controles. Gráfico da distribuição das proteínas identificadas no grupo de pacientes e no grupo de controles saudáveis.

Para caracterizar a assinatura proteômica de pacientes com CHH, primeiramente foram determinadas as proteínas que apresentaram diferença de expressão entre o grupo de pacientes e de controles. Utilizando o *software* PERSEUS, a área dos picos cromatográficos correspondentes a cada proteína identificada foi calculada e submetida a transformação com $\log_2(x)$; e, em seguida, esses dados foram normalizados. Do total de 2.152 proteínas, 631 foram desconsideradas por não atenderem aos pré-requisitos de qualidade para o cálculo. Os valores resultantes foram considerados como abundância relativa de cada proteína. O resultado da análise baseada no teste-t mostrou um total de 274 proteínas com expressão significativamente alterada; e, destas, 28 proteínas apresentaram razões de abundância aumentada e 246 proteínas apresentaram razões de abundância diminuída em relação ao grupo controle (tabela do anexo 11.6). Esses dados foram utilizados para a construção do gráfico de dispersão

do tipo *Volcano* (figura 34), sendo possível observar uma clara diferença no volume de proteínas negativamente reguladas, em relação ao grupo controle.

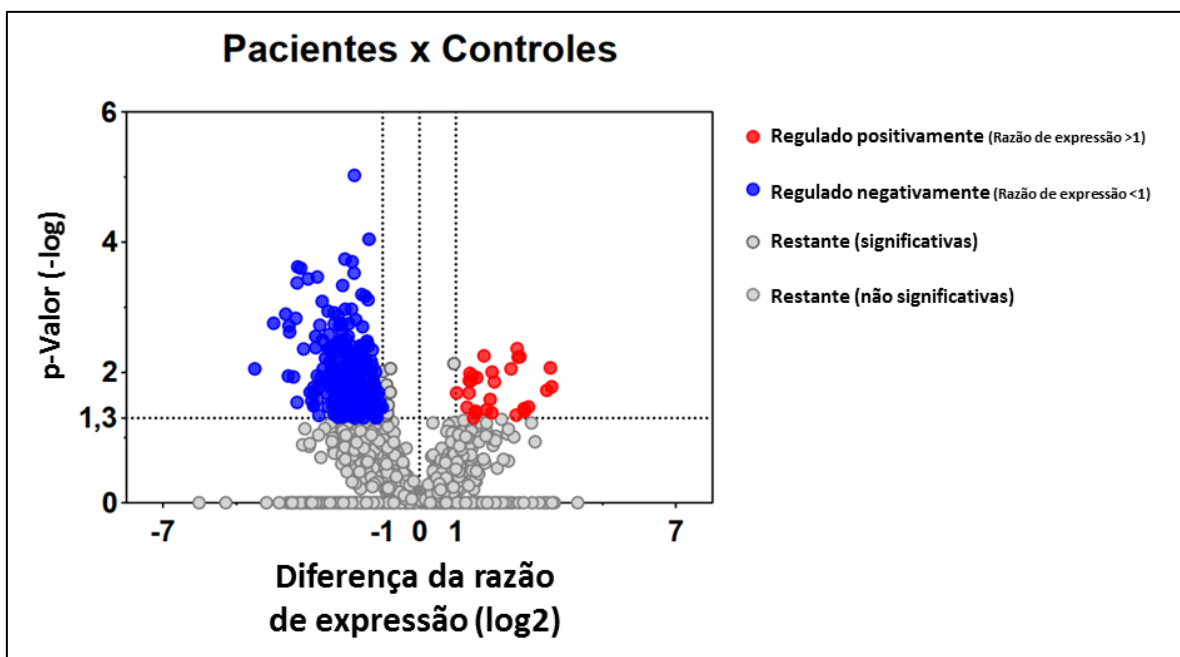


Figura 34: Gráfico de dispersão *Volcano* das proteínas diferencialmente expressas entre pacientes e controles. Esse gráfico mostra as diferenças de expressão com significância estatística das proteínas quantificadas nos pacientes em comparação aos controles saudáveis. As linhas pontilhadas destacam os pontos de significância estatística: $p < 0.05$ / $-\text{Log } p\text{-valor} > 1.3$; $\text{FDR} = 0.05$.

Uma análise de agrupamento hierárquico também foi realizada para investigar os padrões de expressão entre as proteínas do grupo de pacientes e do grupo controle (figura 35). Essa análise compara as informações de expressão das proteínas com base no teste-t ($p < 0,05$), gerando um gráfico em estrutura em forma de árvore, denominado dendograma, em que os dados se agrupam por similaridade. A figura também é combinada com um gráfico de calor (do inglês, *heat map*), que mostra o padrão de expressão de proteínas para cada uma das amostras com base em uma escala de cor.

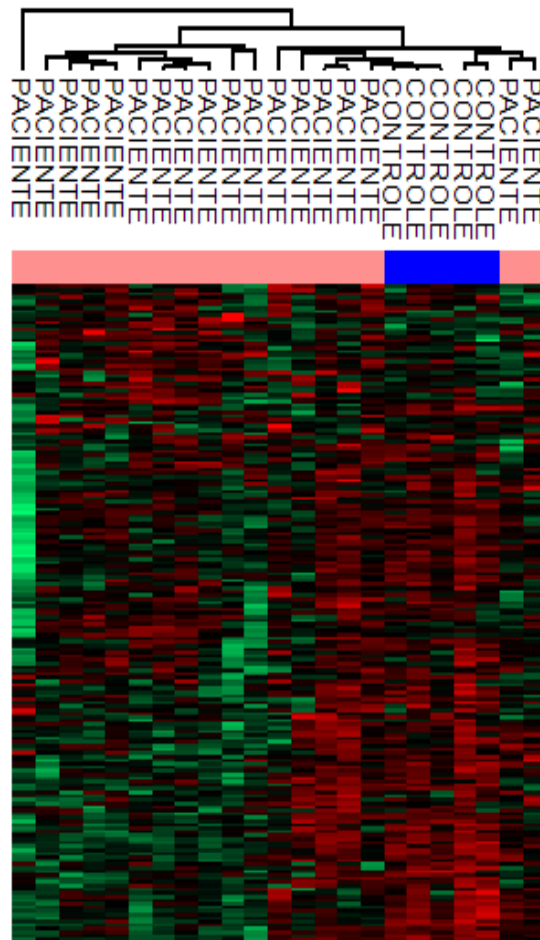


Figura 35: Agrupamento hierárquico com base na distância Euclidiana das proteínas diferencialmente expressas em CHH. Essa análise gera um gráfico dividido em duas partes: dendograma e mapa de calor (do inglês, *heat map*), de forma que é possível observar o quão similar são os indivíduos de cada grupo e também observar os padrões gerais de proteínas no conjunto de dados. A identificação das amostras está localizada na parte inferior do dendograma, sendo representado na cor salmão o grupo de paciente e em azul o grupo controle. A escala de cores corresponde à razão de expressão (fold change), de modo que proteínas reguladas positivamente com maior diferença de expressão em relação ao controle são representadas em vermelho mais intenso e proteínas negativamente reguladas com maior diferença de expressão são representadas em verde mais intenso. A ordem das amostras no mapa de calor: pacientes 3, 28, 9, 23, 6, 24, 25, 8, 1, 2, 5, 32, 29, controles 1-5, pacientes 30, 31, 4, 27 e 10.

O grupo de proteínas diferencialmente expressas foi determinado pela soma das proteínas com expressão significativamente diferente entre pacientes e controles; e as proteínas presentes exclusivamente em cada grupo, totalizando 604 proteínas. Dessas, 71,5% (432/604) estavam reguladas positivamente e 28,5% (172/604) estavam reguladas negativamente.

8.5.2 Análise funcional *in silico*

Com o objetivo de interpretar o perfil proteômico e sugerir prováveis vias de sinalizações/processos biológicos presentes de forma majoritária nas amostras dos pacientes com CHH, os dados foram analisados no *software FunRich* e no banco de dados interativo online *Reactome*. Para isso, foram utilizados os dados das proteínas diferencialmente expressas.

A avaliação funcional *in silico* auxilia na compreensão global dos dados gerados pela análise proteômica de larga escala, considerando que esta é uma técnica de alta demanda que produz resultados robustos e extremamente numerosos. A contextualização das informações do proteoma com o significado biológico tem potencial de gerar hipóteses importantes para a melhor compreensão da doença e identificação de possíveis alvos terapêuticos.

As proteínas com expressão significativamente alterada foram classificadas de acordo com a anotação do banco *Reactome*, permitindo obter um amplo panorama dos processos biológicos relacionados às proteínas diferencialmente expressas nos pacientes em relação ao perfil de indivíduos saudáveis, tais como: replicação do DNA, ciclo celular, expressão gênica (transcrição), transporte mediado por vesícula, resposta celular, metabolismo, homeostase, morte celular programada, metabolismo de proteínas, transdução de sinal, contração muscular, metabolismo de RNA, biologia do desenvolvimento, doença e sistema imune

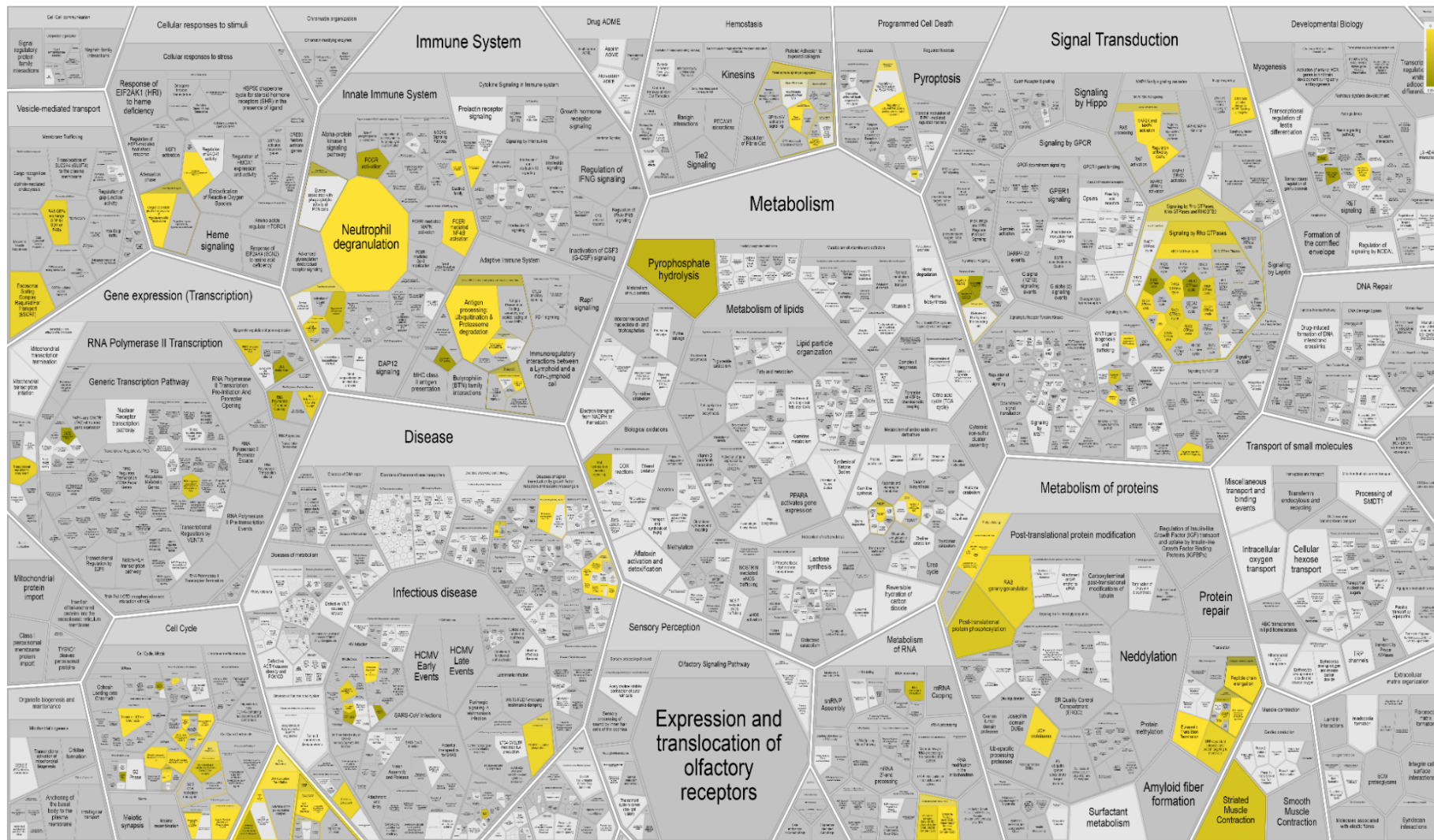


Figura 36: Esquema gerado pelo banco de dados Reactome de processos e vias biológicas envolvidas em CHH. Em amarelo são sinalizadas as vias relacionadas às proteínas diferencialmente expressas identificadas nos pacientes por proteômica. Em cinza estão representadas as demais vias presentes em células humanas que não foram associadas às proteínas detectadas no presente trabalho.

Análises mais detalhadas no banco de dados *Reactome* mostraram categorias altamente enriquecidas que se relacionam com processos já relacionados às funções do gene *RMRP*, tais como: (1) metabolismo de RNA; (2) progressão no ciclo celular (fase G2) e (3) metabolismo de proteínas. Estendendo a busca com o objetivo de obter informações sobre as redes regulatórias em leucócitos de pacientes com CHH, foram observadas alterações nas proteínas relacionadas às seguintes vias de sinalização celular: *WNT*, *TGF- β* e *Rho GTPases*. As proteínas relacionadas à cada uma das vias estão descritas na tabela do anexo 11.7.

Os dados também foram analisados no *software Ingenuity Pathway Analysis* (IPA) e as principais categorias de genes observadas foram as relacionadas a processos e respostas do sistema imunológico, genes importantes no controle da manutenção celular e genes que desempenham papel na formação de tumor (tabela 7).

Tabela 7: Principais vias canônicas; reguladores a montante; funções moleculares e celulares; doenças e desordens identificados pelo software IPA a partir dos dados de proteômica em leucócitos de pacientes com CHH. Para a análise das vias, foram utilizados os dados dos genes diferencialmente expressos.

Principais vias canônicas (mais significativas)	Nº de moléculas
Principais reguladores a montante(?)	
MMP12	
mir-122	
NFE2L2	
Lh	
HSPA5	
Câncer	577
Morte celular e sobrevivência	205
Manutenção e função celular	170
Resposta inflamatória	96
Síntese de proteínas	89
Desenvolvimento do organismo	58
Tráfego de células imunes	39
Modificações pós-traducionais	23

Por fim, a análise de distribuição espacial das proteínas mostrou que os polipeptídeos identificados nos leucócitos dos pacientes já foram identificados anteriormente em mais de 200 tipos celulares/tecidos, sendo alguns já relacionados com fenótipos de CHH, incluindo linfócitos T/B e diversos tipos de células neoplásicas (figura 37).

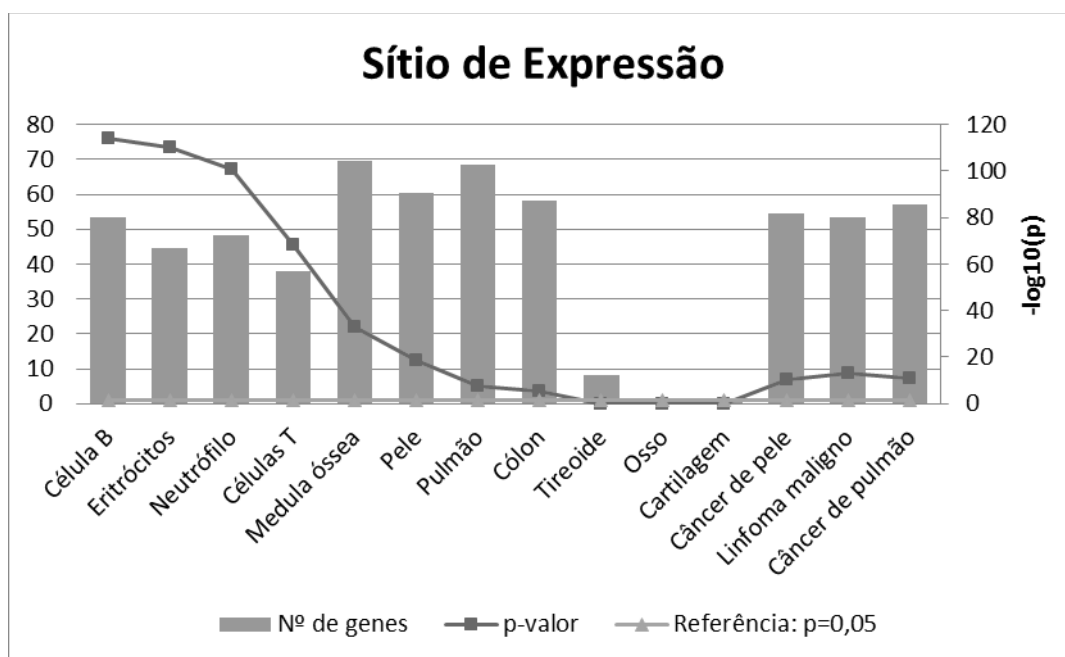


Figura 37: Análise da localização de expressão das proteínas identificadas nos pacientes. As bases de dados utilizadas para a obtenção da localização de expressão fora *Proteoma Map*; *ProteomicsDB* e *Human Proteinpedia*.

9. DISCUSSÃO

Cartilage-Hair Hypoplasia é uma doença multissistêmica causada por alterações recessivas no gene *RMRP*. Este gene é amplamente expresso em diversos tecidos humanos e codifica o RNA estrutural da ribonucleoproteína RNase MRP; uma molécula que atua como endoribonuclease sobre diversos substratos importantes para o funcionamento biológico. Apesar de ser um gene de apenas 268 pb, apresenta uma série extensa de variantes patogênicas, sendo relatadas mais de 100 mutações, tanto na região transcrita quanto em sua região promotora⁴⁹. Uma alta densidade de polimorfismos de único nucleotídeo (SNPs) benignos também já foram descritos⁴⁹. Nossos achados se correlacionam com essas observações, considerando que desde o início das investigações até o presente foram identificadas 12 novas variantes em nossa coorte de pacientes que ainda não haviam sido descritas previamente na literatura⁵; sendo observada uma maior ocorrência de duplicações localizadas entre o TATA box e o sítio +1 de transcrição, dentre estas novas variantes. Bonafé *et al.*, 2005⁶³, levanta a hipótese de que uma possível eficiência reduzida dos mecanismos de reparo do DNA específicos para o gene *RMRP* possa justificar a elevada ocorrência de SNPs e mutações patogênicas diversas em *RMRP*, que não condizem com a alta conservação dessa sequência entre espécies. No entanto, essa hipótese ainda não foi testada.

A análise da estrutura secundária pode ser uma maneira de inferir a patogenicidade de variantes no gene *RMRP* que ainda não foram descritas na literatura. Usando simulações computacionais, foram caracterizadas no RNA os possíveis impactos estruturais das variantes genéticas encontradas em nossos pacientes. Cinco alterações (6G>U, 10U>A, 98G>A, 147G>C e 181G>A) ocorreram

em nucleotídeos envolvidos no pareamento de bases na estrutura secundária da molécula de RNA, possivelmente comprometendo a topologia da estrutura em grampo (*hairpin*). A variante 6C>U manteve o pareamento entre os nucleotídeos, mas produziu uma interação mais fraca após a perda de uma ligação H (C>G, três ligações vs. U>G, duas ligações), provavelmente resultando em uma molécula mais instável e prejudicando a montagem do complexo. A alteração da estrutura secundária do RNA pode afetar a interação com outras moléculas, como as proteínas do complexo RNase MRP além do reconhecimento dos substratos deste complexo⁴⁹. O modelo de interações proteína-RNA da RNase MRP mostra que alguns domínios do transcrito são críticos para a ligação de diferentes subunidades de proteínas^{128,129,130,131,60}.

Além disso, a variante 36C>G ocorreu no domínio P3, importante para interações entre a molécula de RNA e as proteínas POP1, RPP20, RPP25 e POP4¹³². Assim, a modificação do domínio P3 pode comprometer a montagem do complexo RNase MRP. Os resultados também demonstraram que a variante 97_98dup2(UG) aumentou a alça em grampo do domínio P8, importante para a ligação de RPP20. Além disso, as variantes 147G>C/181G>A ocorreram no domínio P12, que é uma região importante para a ligação de POP1 e RPP38 à molécula de RNA⁶⁰.

Seis variantes, 36C>G, 70insU, 71A>G, 194G>A, 196C>U e 252A>C, não afetaram a estrutura secundária. Todas as variantes mencionadas acima, com exceção de 5G>U, estavam localizadas em uma região evolutivamente conservada, sugerindo sua relevância funcional⁶³. Regiões de conservação entre espécies sugerem importância no desempenho das funções biológicas da molécula.

Interessantemente, as alterações 5C>U, 6G>U e 10U>A (presente nos pacientes 7, 15, 22 e 29) ocorreram na região que gera o pequeno RNA RMRP-S1;

e, a alteração 147G-C (paciente 21) está localizada na região relacionada ao RMRP-S2, possivelmente interferindo na biogênese dessas moléculas; e, conseqüentemente, na regulação dos seus alvos. Um trecho da sequência relacionada ao RMRP-S2 (nucleotídeos 118-175 em camundongos, correspondente a 118-167 em humanos) já foi associado com a sinalização responsável pela localização mitocondrial do RNA RMRP¹³. Vale ressaltar que o paciente 16 desenvolveu um quadro grave de linfoma não Hodgkin e veio a óbito. Nesse sentido, a realização de estudos funcionais seria interessante para esclarecer a influência da variante 147G-C na localização mitocondrial e se alterações nessa posição podem estar associadas a um pior prognóstico. Os nucleotídeos localizados na posição 23-62 já foram relacionados à sinalização para a localização nucleolar do RMRP¹³. Três pacientes apresentaram a variante g.36C>G localizada nessa região, porém sem relatos de sinais clínicos mais expressivos.

Em nosso estudo prévio, reunimos um grupo de 23 pacientes com CHH⁵. A inesperada alta prevalência da variante g.196C>T nos indivíduos analisados (60,9%), sugeria um possível efeito fundador para essa variante nestes pacientes. De fato, a inclusão de 9 novos pacientes na coorte, reforçou essa hipótese, visto que 8 dos 9 novos indivíduos com CHH apresentavam essa alteração em um dos alelos, aumentando a frequência da alteração g.196C>T para 70% nesta coorte. Vale ressaltar que não foi detectado nenhum caso de consanguinidade nas famílias carreadoras desta variante e todos genótipos heterozigotos compostos.

O primeiro artigo relatando a variante g.196C>T (rs948931144) foi publicado em 2002 por Luiza Bonafé e colaboradores⁷⁰, sendo um relato de dois casos, em que um dos pacientes, de origem suíça/dinamarquesa, apresentava essa alteração em heterozigose composta com uma duplicação no promotor. Neste mesmo ano,

outros três indivíduos carregando a variante (norte-americano, brasileiro e israelense) de um total de 44 famílias não finlandesas também foram descritos¹³. Posteriormente, em 2005, essa alteração foi novamente identificada por Bonafé e colaboradores em 4 casos (provenientes da Itália, Suíça e EUA) de um estudo com 36 indivíduos, por Hermanns *et al.*, 2006¹⁰, em um único caso proveniente dos EUA em um grupo de 27 pacientes e por Thiel *et al.*, 2007⁵³ em um paciente de origem espanhola/francesa. Em todos esses trabalhos a frequência não ultrapassava 11,1% das coortes estudadas.

Em bancos de dados públicos de indivíduos saudáveis *Gnomad* (GRCh37) foram localizados 9/129.268 alelos com a alteração g.196C>T (frequência = 0,00006939), sendo 5 em latino-americanos, 1 em africanos/afro-americanos, 2 em europeus não finlandeses e 1 em outra etnia, reforçando a baixa ocorrência dessa alteração fora do Brasil e possivelmente mais representada em indivíduos de origem Latino Americana. No banco de dados Abraom (pessoas idosas do estado de São Paulo), só foi detectado um em 1.780 alelos analisados (frequência = 0,000427). Por outro lado, uma consulta aos dados da empresa brasileira de análise genômica Mendelics Diagnósticos, revelou 68/ 65.000 alelos contendo essa variante (frequência = 0,001), sendo o indivíduo em homozigose. Apesar de ser um banco de dados que contém um viés clínico, por se tratar de um laboratório diagnóstico, essas informações reiteram que a frequência desta variante ainda é maior do que a esperada entre brasileiros, reforçando a hipótese da possível existência de um efeito fundador.

Em humanos, este fenômeno genético consiste na redução da variabilidade genômica que ocorre quando um pequeno grupo de indivíduos se estabelece em outro ambiente, se separando de uma população maior¹³³. As características

genóticas e fenóticas serão derivadas do grupo inicial dessa subpopulação que foi separada e, dependendo da representação dos alelos fundadores, a frequência destes traços pode ser muito diferente da população original maior. Como resultado, certas doenças hereditárias podem ter uma maior frequência em alguns grupos populacionais limitados¹³⁴. Esse fenômeno geralmente causa desequilíbrio de ligação entre marcadores de determinados loci que foram herdados dos primeiros imigrantes, modificando as frequências alélicas que seriam esperadas por uma associação aleatória de alelos para um determinado haplótipo¹³³.

Os haplótipos dos indivíduos da nossa coorte CHH foram construídos utilizando como base os dados de sequenciamento da região genômica de interesse e identificação das variantes dos tag SNPs da amostra de DNA dos pacientes e de seus progenitores. A diversidade haplotípica calculada de todos os cromossomos analisados foi relativamente alta ($H_d=0,64$). Esse parâmetro varia de zero a um e, quanto mais próximo a 1, maior é a variação entre os haplótipos estudados; por outro lado, quanto mais próximo a 0, a variação entre a combinação de marcadores é quase nula¹³⁵. Esse índice estima a probabilidade de selecionar ao acaso dois haplótipos de uma amostra e estes serem diferentes (Nei, 1987)¹¹⁸, sendo frequentemente utilizado em estudos de filogeografia.

Nossa análise mostrou que nas famílias de pacientes com CHH brasileiras, a variante g.196C>T está segregando junto a dois haplótipos distintos, sendo quase em sua totalidade associada com o haplótipo T/C/G/A. Analisando especificamente os cromossomos que apresentam a variante g.196C>T, observou-se uma redução significativa na diversidade haplotípica ($H_d=0,11$) quando comparada aos cromossomos que não apresentam essa alteração ($H_d=0,76$). Esse valor de H_d não é compatível com a variabilidade esperada para a região considerando que esta foi

selecionada para ser altamente diversificada, o que sugere que provavelmente trata-se de uma alteração que teve origem em um ancestral comum e não de um *hotspot* mutacional.

Curiosamente, outras duas variantes estão associadas à cromossomos que mostram o mesmo padrão tag SNPs. A variante g.97_98dup2(TG) segrega com o mesmo haplótipo da variante g.196C>T em todos os quatro pacientes que apresentam esta duplicação na sequência transcrita do gene. Já a alteração g.-26_-05dup22(TACTACTCTGTGAAGCTGAGAA) é outra variante recorrente em nossa série em 6 cromossomos carreadores desta variante apresentam o haplótipo T/G/G/C. A região regulatória do gene *RMRP* é alvo de diversas variantes, especialmente duplicações, já tendo sido reportadas mais de 30 tipos diferentes. A literatura traz relatos em que essas alterações aumentam a distância entre o TATA box e o sítio +1, interferindo no correto assentamento do complexo basal de transcrição e levando ao silenciamento do gene⁶³. Alterações deste tipo podem ser tão deletérias que, quando presentes em ambos os alelos, provavelmente são letais no período embrionário⁵⁶. Diante das nossas observações, pode-se questionar se um efeito fundador pode estar contribuindo para a distribuição de outras variantes em nosso país. Entretanto, é difícil fazer suposições devido ao baixo número amostral que pode enviesar a análise.

Dois passos mutacionais separam o haplótipo mais frequentemente associado a g.196C>T (T/C/G/A) do observado em um único cromossomo com esta variante (C/C/G/C). É possível que a esses marcadores tenham sofrido mutação, modificando o haplótipo associado à essa variante. Uma outra hipótese a se considerar é a de recombinação, dependendo do tempo transcorrido desde o surgimento dessa substituição. A proximidade entre os marcadores determina uma

menor taxa de recombinação. Por outro lado, quando a distância da mutação, aumenta, existe uma maior probabilidade de decaimento da associação deste locus com o marcador de interesse¹³⁶. Por isso, no caso dos dois haplótipos associados à variante g.196C>T, dependendo do tempo transcorrido desde o surgimento dessa substituição, é razoável considerar que os dois loci separados por maiores distâncias possam ter sofrido recombinação.

Com base nos resultados, podemos sugerir que a variante g.196C>T ocorreu inicialmente no haplótipo mais frequente (T/C/G/A), o qual foi introduzido por um ascendente em comum das famílias no Brasil e posteriormente os dois marcadores díspares sofreram recombinações (C/G/G/C), sendo considerado um haplótipo ainda pouco associado à esta variante. A análise dos haplótipos de pacientes não brasileiros com a alteração g.196C>T nos ajudaria a entender se esta variante teve uma única ocorrência que foi distribuída mundialmente. Esses ensaios também poderiam ajudar a rastrear o local de sua origem e como ela foi mais expressivamente expandida no Brasil.

O cálculo das frequências haplotípicas nas populações ancestrais proporcionou novos apontamentos sobre a possível origem da variante estudada. As informações mostram que o bloco de marcadores que flanqueiam a mutação g.196C>T estudado está provavelmente em uma porção de origem europeia do cromossomo de pacientes brasileiros, uma vez que o haplótipo T/C/G/A é o segundo mais frequente em indivíduos deste continente (tabela 4). Além disso esses cálculos mostraram que a frequência do haplótipo em pacientes que não apresentam a variante g.196C>T é semelhante a distribuição na população europeia (tabela 4), demonstrando que possivelmente não é efeito de amostragem. Também vale ressaltar que os resultados do *software Structure* reforçam essa hipótese, visto que

as inferências das porcentagens genômicas ancestrais foram predominantemente europeias em nossa coorte. Em conjunto, esses dados sugerem fortemente a ocorrência do efeito fundador, em que os cromossomos analisados provavelmente derivam de um evento isolado ocorrido em um indivíduo de origem europeia de nossos antepassados.

Interessantemente, a variante g.71A>g, considerada uma variante fundadora para a população finlandesa, presente em mais de 90% dos pacientes deste país e em 48% dos pacientes de origem europeia de modo geral¹³, foi observada em apenas um dos pacientes do presente estudo. O haplótipo relacionado à esta variante (C/C/A/G) no paciente 10 é o mais frequente em indivíduos da população europeia, de acordo com os cálculos de frequência alélica baseados nos dados do 1000Genome.

O presente estudo também mostrou que alguns pacientes apresentam combinações haplotípicas notáveis que reforça a extensa miscigenação da população brasileira. A revisão de Souza et al., 2019⁸⁸ nos mostra que a composição tri-híbrida da população resulta de mais de 500 anos de mistura entre nativos americanos, europeus e africanos, sendo uma das populações heterogêneas do mundo, o que pode explicar a presença de haplótipos com diferentes frequências nessas populações ancestrais em nossa coorte. O uso de marcadores no DNA mitocondrial e no cromossomo Y na população brasileira mostrou que, em geral, a proporção de ascendência materna é igualmente europeia, africana e nativo americana enquanto a ascendência paterna é predominantemente de origem europeia, reflexo da assimetria sexual típica de países colonizados onde há exploração sexual do homem colonizador sobre mulheres colonizadas^{137,138}. Em referência a esse dado, pode-se destacar o paciente 23 de ascendência africana

autodeclarada, apresenta o segundo haplótipo mais frequente em europeus (carreando a variante g.196C>T) associado ao segundo mais frequente em africanos (carreando a variante g.-23_-4dup20(TACTCTGTGAAGCTGAGAAC)). Similarmente, os pacientes nascidos em Paudalho/PE, que apresentaram tanto o haplótipo mais frequente em africanos quanto o mais frequente em europeu. De maneira geral, o Nordeste é a região geográfica do Brasil que apresenta o maior percentual de origem africana e se tratando de média de ascendência europeia, os maiores valores são observados no Sul e no Sudeste. Já os traços genéticos de origem ameríndia foram identificados em maiores proporções no Norte.

A alteração g.196C>T foi identificada em pacientes de diferentes regiões do Brasil. Considerando que a dinâmica da nossa população passou por muitas influências considerando a escala temporal, desde a chegada de diversos colonizadores e imigrantes, até a da intensificação do fluxo migratório interno, especialmente no sentido das zonas rurais para as áreas urbanas, é possível que essa não tenha ocorrido há tão poucas gerações atrás, explicando assim a sua dispersão por diversos estados da federação.

A detecção de variantes com efeito fundador, que apresentam maior frequência em indivíduos de determinadas etnias ou de regiões geográficas específicas, torna-se de grande importância para o rastreamento das localidades onde este fenômeno foi observado. Nesse sentido, a realização de análises *in silico* para estimar a datação da variante, agregadas com informações históricas, podem auxiliar no mapeamento dos eventos que levaram à ocorrência do efeito fundador no Brasil, permitindo a implantação de políticas de diagnóstico pré-nupcial e de aconselhamento genético nas localidades em que a frequência da alteração g.196C>T possa estar mais elevada.

A displasia metafisária associada à baixa estatura é a característica clínica mais notável da CHH, entretanto, outras características clínicas também podem estar associadas como hipotricose, frouxidão ligamentar, anemia, doença de Hirschsprung, defeitos na imunológicos e uma maior suscetibilidade ao câncer^{8,48,49,50}. As manifestações clínicas desta condição genética são altamente variáveis, até mesmo entre irmãos que apresentam o mesmo genótipo^{5,13}. O envolvimento em processos biológicos tão importantes e variados podem estar relacionados à dificuldade de se correlacionar a gravidade das manifestações esqueléticas e a ocorrência de anomalias extra esqueléticas com os genótipos identificados, reforçando a necessidade de se compreender melhor a fisiopatologia da doença.

Ensaio para a análise de expressão gênica foram amplamente empregados neste trabalho para melhor compreender os mecanismos subjacentes à CHH. É consenso na literatura que as alterações no promotor afetam mais drasticamente a expressão do gene *RMRP*^{56,63,67}. Neste trabalho, 15/17 amostras de pacientes analisadas por RT-qPCR tinham alterações nesta sequência em um dos alelos. Apenas dois relatos de caso na literatura descrevem pacientes com CHH apresentando duas alterações nesta região^{65,66}. Além disso, a ausência de produto gênico de *RMRP* é letal no período embrionário em camundongos⁵⁶. O RNA *RMRP* é transcrito pela RNA polimerase III no nucleoplasma através do reconhecimento do promotor do tipo III (sem elementos promotores intragênicos)^{139,140} caracterizado pela caixa TATA e outros elementos adicionais localizados a montante ou a 5' desta (elemento PSE, elemento octâmero e SP1 local de ligação)¹⁴¹. Os espaçamentos entre os elementos do promotor tipo III são fixos e importantes neste processo celular. O TATA box está localizado entre a região -33 e -25 do sítio de início da transcrição

e é afetado por variantes do tipo duplicação/inserção em alguns pacientes com CHH que alongam essa distância, interferindo assim na síntese do RNA RMRP. Kavadas e colegas (2008)⁶⁵ descrevem um paciente com CHH portador de duas alterações na região promotora em heterozigose composta e sugere que um sítio alternativo de iniciação de transcrição possa ser utilizado, de modo que ambos os alelos sejam expressos, ainda que em níveis mínimos. Bonafé *et al.*, 2005⁶³ sugere que o efeito possa ser menos deletério quando há quantidades normais de RNA RMRP contendo mudanças estruturais sutis do que em situações de redução acentuada de transcritos RMRP com estrutura normal.

Inicialmente, alguns trabalhos descreveram que a expressão era reduzida apenas para os alelos contendo alterações na região promotora do gene *RMRP*, como em Ridanpää (2001)⁶, em que a expressão total em pacientes com heterozigose composta (alteração na região 5' combinada com alteração pontual no transcrito) mostrou redução de 35%–40% em relação aos controles e aos pacientes com duas alterações intragênicas. Reforçando essa suspeita, Nakashima (2003/2007)^{67,72}, através do sequenciamento direto do DNA complementar de RMRP, detectou apenas a presença do alelo com variante no transcrito, indicando que os alelos com as alterações na sequência do promotor poderiam não ser expressos. Posteriormente, outros trabalhos também detectaram redução dos transcritos com alterações pontuais, relatando que esse tipo de variante pode prejudicar a estrutura secundária do RNA diminuindo sua estabilidade, o que resultaria em um aumento da taxa de degradação da molécula de RNA^{53,56,63,67}. Nossos resultados são consistentes com esses dados mais recentes, considerando que houve redução altamente significativa (** $p < 0,0093$) na expressão em todos os pacientes, independente do genótipo, reforçando a hipótese de que todos os tipos de mutações

envolvidas em CHH podem afetar a eficiência da transcrição e/ou estabilidade do RNA *in vivo*^{56,67}.

Em 2013, um trabalho trouxe à luz do conhecimento sobre CHH a existência de dois pequenos RNAs derivados de RMRP, responsáveis por regular positiva e negativamente mais de 900 genes relacionados a importantes fenótipos desta condição¹². O perfil de expressão de RMRP-S1 e RMRP-S2 encontrados nesta tese se assemelha ao observado por Rogler e cols. A quantificação de ambas as moléculas avaliada em leucócitos de pacientes com diferentes tipos genótipos, estava significativamente reduzida em relação ao grupo controle (**** $p < 0,0001$). Em seu estudo, o autor analisou esses transcritos em amostras provenientes de fibroblastos de dois pacientes homocigotos para a variante g.71A>G, mostrando uma redução de 60% e 70% para RMRP-S1 e RMRP-S2, respectivamente; e em uma linhagem de células B de um paciente com alterações em heterocigose composta (g.-26-6 dup e g.147G>A), reduzida em 80% e 60%, respectivamente. Foi sugerido que o fato da variante g.147G>A estar localizada na sequência relacionada à RMRP-S2, é possível que esta troca tenha influenciado na biogênese desse pequeno RNA. Apesar de termos avaliado a expressão dos pequenos RNAs em amostras de pacientes com genótipos variados, o que poderia ajudar a entender o efeito de cada combinação genotípica individualmente em relação à variabilidade fenotípica, sendo, portanto, um diferencial deste trabalho, infelizmente não tivemos acesso às amostras com alterações que ocorrem na região de onde derivam os pequenos RNAs, que poderiam evidenciar reduções ainda mais expressivas na quantificação dessas moléculas.

O estudo de Rogler¹² também avaliou a expressão de genes na presença de moléculas mimetizadoras e inibidoras de RMRP-S1 e RMRP-S2 em cultura de

células HEK293 por transcriptoma. Baseando-se neste trabalho, foi realizado um ensaio de proteômica utilizando amostras de 18 pacientes CHH de nossa coorte e de 5 controles. Das 604 proteínas diferencialmente expressas nos pacientes do presente estudo, 45 eram comuns ao trabalho de Rogler (tabela do anexo 11.8). Dessas, 3 estavam associadas à regulação de RMRP-S1 e 42 estavam associadas à regulação de RMRP-S2. Apesar do trabalho de Rogler mostrar regulação negativa para todas as 45 proteínas observadas em comum às observadas no presente trabalho, 5 delas estavam reguladas positivamente em relação ao grupo controle nos nossos ensaios, sendo 1 associada à regulação por RMRP-S1 e 4 associadas à regulação por RMRP-S2: RBBP9 (RMRP-S1); RAB44, ACYP1, PRDX3, NUCKS1 (RMRP-S2). Isso pode ser explicado pelas diferenças inerentes a técnica utilizada, pelo uso de amostras de tecidos diferentes e também pela influência de regulações pós transcricionais. Todas essas proteínas, exceto RAB44 já foram relatadas como prováveis biomarcadores em diferentes tipos de tecidos neoplásicos, apresentando expressão elevada, inclusive em linfomas^{142,143,144}. Levando em consideração a predisposição para diferentes tipos de câncer em pacientes com CHH, especialmente linforeticular, pode ser interessante o estabelecimento de um fluxo de análises para vigilância sob o ponto de vista do desenvolvimento de neoplasias, como por exemplo por meio de biópsias líquidas, visando garantir intervenção precoce. Além disso, uma análise mais aprofundada do conjunto total de proteínas identificadas por proteômica também pode detectar outros polipeptídeos que estejam associados às comorbidades da CHH e que, em situações de desregulação dos níveis de expressão, podem auxiliar no manejo clínico desses pacientes, como por exemplo a proteína RAC2, específica de células hematopoéticas e envolvida na regulação da sinalização das células do sistema imune¹⁴⁵.

Estudos de proteômica têm se tornado populares no estudo de doenças humanas, porém até o momento não existem trabalhos utilizando essa metodologia aplicada à CHH na literatura. O processamento dos dados de proteômica realizada no presente estudo, permitiu a identificação de um conjunto de proteínas diferencialmente expressas nos pacientes, evidenciando um perfil proteômico distinto do perfil dos controles. A semelhança do padrão de expressão de proteínas dentro de cada grupo (pacientes e controles) ficou evidenciada nos dendrogramas no topo dos heatmaps. A proteômica é uma análise global de proteínas que tem por objetivo obter uma visão integrada da biologia, estudando todas as proteínas de célula de forma simultânea¹⁴⁶. Nos últimos anos, a interdisciplinariedade entre proteômica e genômica, conhecida como proteogenômica, permitiu identificar pacientes afetados pela mesma doença e até mesmo estratificá-los em subgrupos dentro de uma mesma patologia, além de contribuir para a elucidação dos mecanismos da doença e estabelecer alvos de drogas que podem melhorar a qualidade de vida dos pacientes¹⁴⁷.

As análises quanto à significância funcional das proteínas diferencialmente expressas mostraram processos enriquecidos compatíveis com as funções de *RMRP* e com as manifestações clínicas da CHH. Os genes diferencialmente expressos apresentavam funções celulares relacionadas a sistema imunológico, regulação e controle do ciclo celular, apoptose, transdução de sinal, organização cromossômica e replicação do DNA, como já observado em outros relatos da literatura para ensaios de transcriptoma^{56,148,149}.

Vale destacar que as categorias relacionadas à via de metabolismo de RNA e à via de metabolismo de proteínas estavam reguladas negativamente nos pacientes. A síndrome CHH está incluída no grupo das ribossomopatias, uma classe

heterogênea de doenças desencadeada por defeitos no processo de produção de ribossomos¹⁵⁰. Uma das funções do complexo RNase MRP é a clivagem do RNA ribossomal policistrônico, levando à formação final do rRNA 5.8S, integrante da subunidade 60S dessa maquinaria de síntese proteica¹⁵¹. Considerando que 71,5% das proteínas diferencialmente expressas estavam reguladas negativamente, esse resultado pode ser um reflexo de processos envolvidos nesta doença, uma vez que a eficiência da produção do rRNA pode ser reduzida em função das variantes em *RMRP* observada nos pacientes; e, por consequência, a função do ribossomo também estaria prejudicada, assim como a produção de proteínas. O trabalho de Thiel *et. al*, 2007⁵³, utilizando fibroblastos contendo diferentes variantes em *RMRP*, mostra um acúmulo de RNA ribossomal não clivado, reforçando essa hipótese.

Os resultados do Reactome também indicaram que um grande número de proteínas envolvidas em transdução de sinal apresentava uma baixa razão de expressão nos pacientes. Os mecanismos de sinalização que levam à transcrição de *RMRP* e as cascatas subjacentes que são desencadeadas por esse RNA ainda não são bem conhecidos. Mandy e colaboradores demonstraram que variantes patogênicas em *RMRP* aumentam a sinalização canônica de Wnt/ β -catenina. Ademais, nesse mesmo trabalho, a inibição dessa via por uso de fármacos pareceu diminuir parcialmente a condrodisplasia em modelos de *Danio rerio* para CHH¹⁵². Park e Jeong, 2015 também relataram a relação de *RMRP* com a via Wnt/ β -catenina e com a Hippo/YAP por experimentos em células de câncer de cólon, devido aos elementos β -catenina/TCF e YAP/TBX5 detectados por preditores *in silico* em regiões à montante do gene *RMRP*⁶⁸. No presente estudo, as vias de Wnt, Hedgehog, Notch e MAPK estavam negativamente reguladas nos pacientes em relação ao grupo controle. Todas essas vias já foram relacionadas com a complexa regulação

da expressão gênica durante o desenvolvimento e a manutenção da placa de crescimento. Inclusive, alterações em diversos genes que participam dessas sinalizações levam à fenótipos de displasias esqueléticas²⁷. Embora deva ser destacado que o ensaio de proteômica foi realizado em leucócitos e que possivelmente existem diferenças no perfil de expressão de proteínas desse tipo celular quando comparado aos condrócitos, é interessante considerar que a expressão constitutiva diminuída de *RMRP* nos pacientes também possa influenciar essas redes de sinalização em outros tecidos. Além disso, a alteração dessas vias nos leucócitos pode indicar os meios pelos quais os pacientes desenvolvem fenótipos hematológicos.

Interessantemente, um grande número de proteínas relacionadas à via de sinalização mediada por Rho GTPases estava negativamente regulada nos pacientes. Essa via converge para funções já relacionadas a fenótipos específicos presentes em CHH, como o desenvolvimento da cartilagem na ossificação endocondral e alterações no cabelo⁴²; porém, até o presente momento essa via não foi relacionada à expressão do gene *RMRP*. O estudo de Wu et al., 2006, detectou perda de pelos em modelos *knockouts* condicionais para o gene *Ccdc42*¹⁵³. Outro relato, utilizando camundongos *knockouts* condicionais para *Rac1*, mostrou nanismo, cifose, viabilidade reduzida de condrócitos, desorganização da placa de crescimento e morte dos neonatos¹⁵⁴. Além disso, *knockouts* para diferentes genes da família Rho-GTPases também mostraram a importância do papel dessa via em processos como desenvolvimento, ativação, diferenciação e migração dos linfócitos T, intimamente relacionado com o desempenho adequado das funções imunológicas¹⁵⁵.

Embora este seja um estudo preliminar e o primeiro a associar a via de Rho com CHH, esses resultados nos permitem sugerir que essa sinalização pode estar

relacionada com o curso da doença, podendo ser considerado um alvo promissor para futuras investigações e confirmações ortogonais, trazendo uma nova perspectiva sobre sinalização celular que conduz à expressão de *RMRP*.

A estratégia de proteômica pode ser uma poderosa ferramenta para auxiliar na compreensão molecular de doenças, podendo ser aplicada tanto com o intuito de se obter respostas a perguntas direcionadas, quanto em estudos exploratórios para a detecção de novas moléculas e/ou conexões inesperadas. Essa foi uma abordagem inédita ainda não explorada para esta condição e os resultados trouxeram um panorama geral do comportamento molecular desta condição, que pode ser mais bem estudado com a integração de resultados de ensaios utilizando amostras de diferentes tipos celulares, especialmente considerando que a doença é multissistêmica. O fato de o gene ser ubiquitariamente expresso e desta condição envolver alterações hematológicas incentivou o uso de leucócitos, que são acessados periféricamente e de forma minimamente invasiva. A utilização do método de indução de diferenciação celular a partir de células mesenquimais dos pacientes para diferentes tipos celulares envolvidos nesta condição, além levar em consideração os aspectos genéticos apresentados pelos pacientes individualmente, pode auxiliar na compreensão do efeito das diversas combinações de variante na relação genótipo-fenótipo. Por fim, a associação de diversas técnicas de *high throughput* para análise de expressão em diferentes tipos tecidos também pode auxiliar na compreensão da interação entre RMRP, RMRP-S1 e RMRP-S2 e futuramente esse conhecimento pode se estender para elucidação da influência dessas redes em diversos processos biológicos e suas consequências sob os aspectos do desenvolvimento e homeostase do organismo em pacientes com CHH.

10. CONCLUSÃO

O diagnóstico molecular, através da análise de variantes no gene *RMRP*, permitiu estabelecer um diagnóstico preciso entre as famílias; possibilitando, desta forma, uma maior precisão do aconselhamento genético das famílias. Como consequência, foi possível o estabelecimento de uma conduta clínica adequada dos pacientes, especialmente na questão da vigilância imunológica e oncológica.

A avaliação *in silico* das variantes permitiu inferir os efeitos das trocas de nucleotídeos na estrutura secundária do RNA RMRP. Essa abordagem pode ser uma ferramenta importante para avaliar os possíveis danos na molécula que podem acarretar em um desempenho anômalo de sua função biológica, particularmente em casos de variantes ainda não descritas na literatura.

A análise dos haplótipos sugere fortemente a ocorrência de um efeito fundador da variante patogênica g.196C>T para a população brasileira de pacientes com CHH. Futuramente, em associação com estudos *in silico* de datação da variante e com análise de heredogramas tabulados, esses resultados poderão auxiliar na compreensão histórica da introdução da alteração g.196C>T no Brasil, na documentação de dados epidemiológicos da doença e na implementação de políticas de saúde pública envolvendo aconselhamento genético pré-nupcial.

As diversas metodologias empregadas no presente estudo tiveram como objetivo obter mais informações sobre os mecanismos subjacentes à CHH, levando em consideração as especificidades dos pacientes brasileiros. Os resultados relacionados à expressão gênica, tanto a nível de transcritos, quanto de proteínas, além de reforçarem os dados descritos na literatura, contribuem para uma descrição mais ampla do perfil molecular de pacientes com CHH. Além disso, os estudos

originais funcionais qualitativos de proteômica apontaram que a via de sinalização das Rho GTPases está abrangentemente alterada nos pacientes com CHH, trazendo uma nova perspectiva sobre os mecanismos de sinalização celular que conduzem a expressão do gene *RMRP*. Essa nova possível via de sinalização precisa ser amplamente explorada e poderá, futuramente, direcionar o desenvolvimento de alvos terapêuticos, em consonância com o crescimento da medicina personalizada.

Em conjunto, os resultados deste estudo contribuíram para a ampliação do conhecimento sobre a diversos aspectos da síndrome CHH, especialmente nos pacientes brasileiros, trazendo recursos que podem auxiliar no esclarecimento da fisiopatologia desta condição.

REFERÊNCIAS

1. Horovitz DDG, Llerena Jr. JC, Mattos RA. (2005). Atenção aos defeitos congênitos no Brasil: panorama atual. *Cadernos de Saúde Pública*, 21(4), 1055-1064.
2. DATASUS [<https://datasus.saude.gov.br/>]. Dados sobre mortalidade [acesso em 12 de setembro de 2022]. Disponível em: <http://tabnet.datasus.gov.br/cgi/deftohtm.exe?sim/cnv/obt10uf.def>.
3. Mortier GR, Cohn DH, Cormier-Daire V, Hall C, Krakow D, Mundlos S, Nishimura G, Robertson S, Sangiorgi L, Savarirayan R, Sillence D, Superti-Furga A, Unger S, Warman ML. Nosology and classification of genetic skeletal disorders: 2019 revision. *Am J Med Genet A*. 2019 Dec;179(12):2393-2419.
4. Gomes M. Avaliação fenotípica, genotípica e dos níveis de transcritos de RMRP e pré-rRNA 5.8S em pacientes com Hipoplasia Cartilagem-Cabelo. 2017. Dissertação (Mestrado em Pesquisa Aplicada à Saúde da Criança e da Mulher, IFF/FIOCRUZ. Rio de Janeiro, 2017.
5. Gomes ME, Paternostro LC, Moura VR, Antunes D, Caffarena ER, Horovitz D, Sanseverino MT, Leal GF, Felix TM, Cavalcanti DP, Llerena Jr JC, Gonzalez S (2019). Identification of Novel and Recurrent RMRP Variants in a Series of Brazilian Patients With Cartilage-Hair Hypoplasia: McKusick Syndrome. *Mol Syndromol* 2020;10(5):255-263.
6. Ridanpää M, van Eenennaam H, Pelin K, Chadwick R, Johnson C, Yuan B, vanVenrooij W, Pruijn G, Salmela R, Rockas S, Mäkitie O, Kaitila I, de la Chapelle A. Mutations in the RNA component of RNase MRP cause a pleiotropic human disease, cartilage-hair hypoplasia. *Cell*. 2001. 26;104(2):195-203.
7. Riley P Jr, Weiner DS, Leighley B, Jonah D, Morton DH, Strauss KA, Bober MB, Dicintio MS. Cartilage hair hypoplasia: characteristics and orthopaedic manifestations. *J Child Orthop*. Apr;9(2):145-52.
8. Vakkilainen S, Taskinen M, Mäkitie O. Immunodeficiency in cartilage-hair hypoplasia: Pathogenesis, clinical course and management. *Scand J Immunol*. 2020;92(4):e12913.
9. Martin AN, Li Y. RNase MRP RNA and human genetic diseases. *Cell Res*. 2007;17(3):219-26.
10. Hermanns P, Tran A, Munivez E, Carter S, Zabel B, Lee B, Leroy JG. RMRP mutations in cartilage-hair hypoplasia. *Am J Med Genet A*. 2006. 1;140(19):2121-30.

11. Thiel CT, Horn D, Zabel B, Ekici AB, Salinas K, Gebhart E, Rüschemdorf F, Sticht H, Spranger J, Müller D, Zweier C, Schmitt ME, Reis A, Rauch A. Severely incapacitating mutations in patients with extreme short stature identify RNA-processing endoribonuclease RMRP as an essential cell growth regulator. *Am J Hum Genet.* 2005. ;77(5):795-806.
12. Rogler LE, Kosmyrna B, Moskowitz D, Bebawee R, Rahimzadeh J, Kutchko K, Laederach A, Notarangelo LD, Giliani S, Bouhassira E, Frenette P, Roy-Chowdhury J, Rogler CE. Small RNAs derived from lncRNA RNase MRP have gene-silencing activity relevant to human cartilage-hair hypoplasia. *Hum Mol Genet.* 2014. 15;23(2):368-82.
13. Ridanpää M, Sistonen P, Rockas S, Rimoin DL, Mäkitie O, Kaitila I. Worldwide mutation spectrum in cartilage-hair hypoplasia: ancient founder origin of the major 70A-->G mutation of the untranslated RMRP. *Eur J Hum Genet.* 2002;10(7):439-47.
14. Gomes MES, Kanazawa TY, Riba FR, Pereira NG, Zuma MCC, Rabelo NC, Sanseverino MT, Horovitz DDG, Llerena JC Jr, Cavalcanti DP, Gonzalez S. Novel and Recurrent Mutations in the FGFR3 Gene and Double Heterozygosity Cases in a Cohort of Brazilian Patients with Skeletal Dysplasia. *Mol Syndromol.* 2018;9(2):92-99.
15. Hirose Y, Nakashima E, Ohashi H, Mochizuki H, Bando Y, Ogata T, Adachi M, Toba E, Nishimura G, Ikegawa S. Identification of novel RMRP mutations and specific founder haplotypes in Japanese patients with cartilage-hair hypoplasia. *J Hum Genet.* 2006;51(8):706-710.
16. Toosi S, Behravan J. Osteogenesis and bone remodeling: A focus on growth factors and bioactive peptides. *Biofactors.* 2020 May;46(3):326-340.
17. Clarke B. Normal bone anatomy and physiology. *Clin J Am Soc Nephrol.* 2008;3 Suppl 3(Suppl 3):S131-9.
18. Junqueira L. C.; Carneiro J. In: *Tecido Ósseo. Histologia Básica.* 10^a ed. Rio de Janeiro: Guanabara Koogan, 2004.
19. Karsenty G. The complexities of skeletal biology. *Nature.* 2003.15;423(6937):316-8.
20. Moore KL, Persaud TVN, Torchia MG. *Embriologia clínica.* 10.ed. Rio de Janeiro: Elsevier, 2016
21. White A, Wallis G. Endochondral ossification: a delicate balance between growth and mineralisation. *Curr Biol.* 2001. 7;11(15):R589-91.
22. Gilbert SF. *Developmental biology.* 6th ed. Sunderland: Sinauer Associates; 2000.

23. Lopes D, Martins-Cruz C, Oliveira MB, Mano JF. Bone physiology as inspiration for tissue regenerative therapies. *Biomaterials*. 2018;185:240-275.
24. Biga LM, Dawson S, Harwell A, Hopkins R, Kaufmann J, LeMaster M, Matern P, Morrison-Graham K, Quick D, Runyeon J, Whittier L. *Anatomy & Physiology*. Capítulo 6.4 Bone Formation and Development. 2022. Uma versão adaptada de OpenStax Anatomy & Physiology disponível em (<https://openstax.org/details/books/anatomy-and-physiology>) com revisão do conteúdo e desenho de arte de Open Oregon State, Oregon State University.
25. Jing Y, Wang Z, Li H, Ma C, Feng J. Chondrogenesis Defines Future Skeletal Patterns Via Cell Transdifferentiation from Chondrocytes to Bone Cells. *Curr Osteoporos Rep*. 2020;18(3):199-209.
26. Marzin P, Cormier-Daire V. New perspectives on the treatment of skeletal dysplasia. *Ther Adv Endocrinol Metab*. 2020 3;11:2042018820904016.
27. Guasto A, Cormier-Daire V. Signaling Pathways in Bone Development and Their Related Skeletal Dysplasia. *Int J Mol Sci*. 2021.21;22(9):4321.
28. Zhan Jing, Zhihui Liang, Liang Yang, Wenyan Du, Tao Yu, Huayu Tang, Chang de Li, Wei Wei. Bone formation and bone repair: The roles and crosstalk of osteoinductive signaling pathways. *Process Biochemistry* 118 (2022) 252–262
29. Ağirdil Y. The growth plate: a physiologic overview. *EFORT Open Rev*. 2020 Sep 10;5(8):498-507.
30. Studer D, Millan C, Öztürk E, Maniura-Weber K, Zenobi-Wong M. Molecular and biophysical mechanisms regulating hypertrophic differentiation in chondrocytes and mesenchymal stem cells. *Eur Cell Mater*. 2012 24;24:118-35; discussion 135.
31. Samsa WE, Zhou X, Zhou G. Signaling pathways regulating cartilage growth plate formation and activity. *Semin Cell Dev Biol*. 2017;62:3-15.
32. Green JD, Tollemar V, Dougherty M, Yan Z, Yin L, Ye J, Collier Z, Mohammed MK, Haydon RC, Luu HH, Kang R, Lee MJ, Ho SH, He TC, Shi LL, Athiviraham A. Multifaceted signaling regulators of chondrogenesis: Implications in cartilage regeneration and tissue engineering. *Genes Dis*. 2015;2(4):307-327.
33. Chen H, Tan XN, Hu S, Liu RQ, Peng LH, Li YM, Wu P. Molecular Mechanisms of Chondrocyte Proliferation and Differentiation. *Front Cell Dev Biol*. 2021 May 28;9:664168.
34. Lojk J, Marc J. Roles of Non-Canonical Wnt Signalling Pathways in Bone Biology. *Int J Mol Sci*. 2021.7;22(19):10840.

35. Li J, Dong S. The Signaling Pathways Involved in Chondrocyte Differentiation and Hypertrophic Differentiation. *Stem Cells Int.* 2016;2016:2470351.
36. Zieba JT, Chen YT, Lee BH, Bae Y. Notch Signaling in Skeletal Development, Homeostasis and Pathogenesis. *Biomolecules.* 202019;10(2):332.
37. Mead TJ, Yutzey KE. Notch pathway regulation of chondrocyte differentiation and proliferation during appendicular and axial skeleton development. *Proc Natl Acad Sci U S A.* 200925;106(34):14420-5.
38. Xie Y, Zhou S, Chen H, Du X, Chen L. Recent research on the growth plate: Advances in fibroblast growth factor signaling in growth plate development and disorders. *J Mol Endocrinol.* 2014;53(1):T11-34.
39. Ikehata M, Yamada A, Fujita K, Yoshida Y, Kato T, Sakashita A, Ogata H, Iijima T, Kuroda M, Chikazu D, Kamijo R. Cooperation of Rho family proteins Rac1 and Cdc42 in cartilage development and calcified tissue formation. *Biochem Biophys Res Commun.* 20187;500(3):525-529.
40. Wang G, Beier F. Rac1/Cdc42 and RhoA GTPases antagonistically regulate chondrocyte proliferation, hypertrophy, and apoptosis. *J Bone Miner Res.* 2005;20(6):1022-31.
41. Wang G, Woods A, Sabari S, Pagnotta L, Stanton LA, Beier F. RhoA/ROCK signaling suppresses hypertrophic chondrocyte differentiation. *J Biol Chem.* 200426;279(13):13205-14.
42. Duquette PM, Lamarche-Vane N. Rho GTPases in embryonic development. *Small GTPases.* 2014;5(2):8.
43. Ozono K, Namba N, Kubota T, Kitaoka T, Miura K, Ohata Y, Fujiwara M, Miyoshi Y, Michigami T. Pediatric aspects of skeletal dysplasia. *Pediatr Endocrinol Rev.* 2012;10 Suppl 1:35-43.
44. Krakow D, Lachman RS, Rimoin DL. Guidelines for the prenatal diagnosis of fetal skeletal dysplasias. *Genet Med.* 2009;11(2):127-33.
45. Campeau P. GeneReviews [<https://www.ncbi.nlm.nih.gov/books/NBK1116/>]. Skeletal Dysplasias [acesso em 12 de setembro de 2022]. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK279130/>.
46. Royer-Bertrand B, Castillo-Taucher S, Moreno-Salinas R, Cho TJ, Chae JH, Choi M, Kim OH, Dikoglu E, Campos-Xavier B, Girardi E, Superti-Furga G, Bonafé L, Rivolta C, Unger S, Superti-Furga A. Mutations in the heat-shock protein A9 (HSPA9) gene cause the EVEN-PLUS syndrome of congenital malformations and skeletal dysplasia. *Sci Rep.* 2015 24;5:17154.
47. Dikoglu E, Alfaiz A, Gorna M, Bertola D, Chae JH, Cho TJ, Derbent M, Alanay Y, Guran T, Kim OH, Llerenar JC Jr, Yamamoto G, Superti-Furga G, Raymond

- A, Xenarios I, Stevenson B, Campos-Xavier B, Bonafé L, Superti-Furga A, Unger S. Mutations in LONP1, a mitochondrial matrix protease, cause CODAS syndrome. *Am J Med Genet A*. 2015;167(7):1501-9.
48. Mäkitie O, Vakkilainen S. GeneReviews [https://www.ncbi.nlm.nih.gov/books/NBK1116/]. Cartilage-Hair Hypoplasia – Anauxetic Dysplasia Spectrum Disorders [acesso em 15 de setembro de 2022]. Disponível em: https://www.ncbi.nlm.nih.gov/books/NBK84550/.
49. Mattijssen S, Welting TJ, Puijn GJ. RNase MRP and disease. *Wiley Interdiscip Rev RNA*. 2010;1(1):102-16.
50. Rider NL, Morton DH, Puffenberger E, Hendrickson CL, Robinson DL, Strauss KA. Immunologic and clinical features of 25 Amish patients with RMRP 70 A->G cartilage hair hypoplasia. *Clin Immunol*. 2009 Apr;131(1):119-28.
51. Mäkitie O, Marttinen E, Kaitila I. Skeletal growth in cartilage-hair hypoplasia. A radiological study of 82 patients. *Pediatr Radiol*. 1992;22(6):434-9.
52. Faitelson Y, Manson D. Cartilage–hair hypoplasia: a spectrum of clinical and radiological findings. *LymphoSign Journal*. 2015, 2(3): 157-164.
53. Thiel CT, Mortier G, Kaitila I, Reis A, Rauch A. Type and level of RMRP functional impairment predicts phenotype in the cartilage hair hypoplasia-anauxetic dysplasia spectrum. *Am J Hum Genet*. 2007;81(3):519-29.
54. Kwan A, Manning MA, Zollars LK, Hoyme HE. Marked variability in the radiographic features of cartilage-hair hypoplasia: case report and review of the literature. *Am J Med Genet A*. 2012;158A(11):2911-6.
55. Verloes A, Pierard GE, Le Merrer M, Maroteaux P. Recessive metaphyseal dysplasia without hypotrichosis. A syndrome clinically distinct from McKusick cartilage-hair hypoplasia. *J Med Genet*. 1990;27(11):693-6.
56. Hermanns P, Bertuch AA, Bertin TK, Dawson B, Schmitt ME, Shaw C, Zabel B, Lee B. Consequences of mutations in the non-coding RMRP RNA in cartilage-hair hypoplasia. *Hum Mol Genet*. 2005;14(23):3723-40.
57. Engreitz JM, Ollikainen N, Guttman M. Long non-coding RNAs: spatial amplifiers that control nuclear structure and gene expression. *Nat Rev Mol Cell Biol*. 2016;17(12):756-770.
58. Steinbusch MF, Caron MMJ, Surtel DAM, Friedrich F, Lausch E, Puijn GJM, Verhesen W, Schroen BLM, van Rhijn LW, Zabel B and Welting TJM. Expression of RMRP RNA is regulated in chondrocyte hypertrophy and determines chondrogenic differentiation. *Sci Rep*. 2017 25;7(1):6440.
59. Schmitt ME, Clayton DA. Yeast Site-Specific Ribonucleoprotein Endoribonuclease MRP Contains an RNA Component Homologous to

- Mammalian RNase MRP RNA and Essential for Cell Viability. *Genes Dev.* 1992 ;6(10):1975-85.
60. Welting TJ, van Venrooij WJ, Pruijn GJ. Mutual interactions between subunits of the human RNase MRP ribonucleoprotein complex. *Nucleic Acids Res.* 2004 19;32(7):2138-46.
61. Cai T, Aulds J, Gill T, Cerio M, Schmitt ME. The *Saccharomyces cerevisiae* RNase mitochondrial RNA processing is critical for cell cycle progression at the end of mitosis. *Genetics.* 2002;161(3):1029-42.
62. Mattijssen S, Hinson ER, Onnekink C, Hermanns P, Zabel B, Cresswell P, Pruijn GJ. Viperin mRNA is a novel target for the human RNase MRP/RNase P endoribonuclease. *Cell Mol Life Sci.* 2011;68(14):2469-80.
63. Bonafé L, Dermitzakis ET, Unger S, Greenberg CR, Campos-Xavier BA, Zankl A, Ucla C, Antonarakis SE, Superti-Furga A, Reymond A. Evolutionary comparison provides evidence for pathogenicity of RMRP mutations. *PLoS Genet.* 2005;1(4):e47.
64. Rosenbluh J, Nijhawan D, Chen Z, Wong KK, Masutomi K, Hahn WC. RMRP is a non-coding RNA essential for early murine development. *PLoS One.* 2011;6(10):e26270.
65. Kavadas FD, Giliani S, Gu Y, Mazzolari E, Bates A, Pegoiani E, Roifman CM, Notarangelo LD. Variability of clinical and laboratory features among patients with ribonuclease mitochondrial RNA processing endoribonuclease gene mutations. *J Allergy Clin Immunol.* 2008;122(6):1178-84.
66. Vatanavicharn N, Visitsunthorn N, Pho-iam T, Jirapongsananuruk O, Pacharn P, Chokephaibulkit K, Limwongse C, Wasant P. An infant with cartilage-hair hypoplasia due to a novel homozygous mutation in the promoter region of the RMRP gene associated with chondrodysplasia and severe immunodeficiency. *J Appl Genet.* 2010;51(4):523-8.
67. Nakashima E, Tran JR, Welting TJ, Pruijn GJ, Hirose Y, Nishimura G, Ohashi H, Schurman SH, Cheng J, Candotti F, Nagaraja R, Ikegawa S, Schlessinger D. Cartilage hair hypoplasia mutations that lead to RMRP promoter inefficiency or RNA transcript instability. *Am J Med Genet A.* 2007, 15;143A(22):2675-81.
68. Park J, Jeong S. Wnt activated β -catenin and YAP proteins enhance the expression of non-coding RNA component of RNase MRP in colon cancer cells. *Oncotarget.* 2015, 27;6(33):34658-68.
69. Maida Y, Yasukawa M, Furuuchi M, Lassmann T, Possemato R, Okamoto N, Kasim V, Hayashizaki Y, Hahn WC, Masutomi K. An RNA-dependent RNA polymerase formed by TERT and the RMRP RNA. *Nature.* 2009, 10;461(7261):230-5.

70. Bonafé L, Schmitt K, Eich G, Giedion A, Superti-Furga A. RMRP gene sequence analysis confirms a cartilage-hair hypoplasia variant with only skeletal manifestations and reveals a high density of single-nucleotide polymorphisms. *Clin Genet.* 2002;61(2):146-51.
71. Koonin, E.V., Galperin, M.Y. Evolutionary Concept in Genetics and Genomics. In: *Sequence — Evolution — Function.* Springer, Boston, MA, 2003.
72. Nakashima E, Mabuchi A, Kashimada K, Onishi T, Zhang J, Ohashi H, Nishimura G, Ikegawa S. RMRP mutations in Japanese patients with cartilage-hair hypoplasia. *Am J Med Genet A.* 2003, 15;123A(3):253-6.
73. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, Gauthier LD, Brand H, Solomonson M, Watts NA, Rhodes D, Singer-Berk M, England EM, Seaby EG, Kosmicki JA, Walters RK, Tashman K, Farjoun Y, Banks E, Poterba T, Wang A, Seed C, Whiffin N, Chong JX, Samocha KE, Pierce-Hoffman E, Zappala Z, O'Donnell-Luria AH, Minikel EV, Weisburd B, Lek M, Ware JS, Vittal C, Armean IM, Bergelson L, Cibulskis K, Connolly KM, Covarrubias M, Donnelly S, Ferriera S, Gabriel S, Gentry J, Gupta N, Jeandet T, Kaplan D, Llanwarne C, Munshi R, Novod S, Petrillo N, Roazen D, Ruano-Rubio V, Saltzman A, Schleicher M, Soto J, Tibbetts K, Tolonen C, Wade G, Talkowski ME; Genome Aggregation Database Consortium, Neale BM, Daly MJ, MacArthur DG. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020;581(7809):434-443.
74. Bourgeois Y, Boissinot S. On the Population Dynamics of Junk: A Review on the Population Genomics of Transposable Elements. *Genes (Basel).* 2019;10(6):419.
75. Beiguelman, B. *Genética de Populações Humanas.* - Ribeirão Preto: SBG, 2008.
76. Freire-Maia N. Inbreeding in Brazil. *Am J Hum Genet.* 1957;9(4):284-98.
77. Zeegers MP, van Poppel F, Vlietinck R, Spruijt L, Ostrer H. Founder mutations among the Dutch. *Eur J Hum Genet.* 2004;12(7):591-600.
78. Zlotogora J. Multiple mutations responsible for frequent genetic diseases in isolated populations. *Eur J Hum Genet.* 2007;15(3):272-8.
79. Bittles AH. Endogamy, consanguinity and community disease profiles. *Community Genet.* 2005;8(1):17-20.
80. McKusick VA. Ellis-van Creveld syndrome and the Amish. *Nat Genet.* 2000 Mar;24(3):203-4.
81. Harlalka GV, Baple EL, Cross H, Kühnle S, Cubillos-Rojas M, Matentzoglou K, Patton MA, Wagner K, Coblenz R, Ford DL, Mackay DJ, Chioza BA, Scheffner

- M, Rosa JL, Crosby AH. Mutation of HERC2 causes developmental delay with Angelman-like features. *J Med Genet.* 2013;50(2):65-73
82. S Pannain 1, R E Weiss, C E Jackson, D Dian, J C Beck, V C Sheffield, N Cox, S Refetoff. Two Different Mutations in the Thyroid Peroxidase Gene of a Large Inbred Amish Kindred: Power and Limits of Homozygosity Mapping. *J Clin Endocrinol Metab.* 1999;84(3):1061-71.
83. Bittles AH, Black ML. Evolution in Health and Medicine Sackler Colloquium: Consanguinity, Human Evolution, and Complex Diseases. *Proc Natl Acad Sci U S A.* 2010;107 Suppl 1(Suppl 1):1779-86.
84. Ribeiro D. *O povo brasileiro: a formação e o sentido do Brasil.* São paulo: Companhia das Letras. 1995.
85. Liascovich R, Rittler M, Castilla EE. Consanguinity in South America: demographic aspects. *Hum Hered.* 2001;51(1-2):27-34. 4.
86. Azevedo, ES.; Fortuna, CMM.; Silva, KMC.; Sousa, MGF; Machado, MA; Lima, AMVMD; Aguiar, ME; Abé, K.; EULÁLIO, MCMN; Conceição, M.M.; Silva, MCBO.; Santos, MG. Spread and diversity of human populations in Bahia, Brazil. *Hum Biol* 54:329-341. 1982.
87. Bittles AH, Black ML. Evolution in health and medicine Sackler colloquium: consanguinity, human evolution, and complex diseases, *Proc Natl Acad Sci USA*, 2010, vol. 107 Suppl 1(pg. 1779-86)
88. Souza AM, Resende SS, Sousa TN, Brito CFA. A systematic scoping review of the genetic ancestry of the Brazilian population. *Genet Mol Biol.* 2019 Jul-Sep;42(3):495-508.
89. Instituto Brasileiro de Geografia e Estatística [www.ibge.gov.br]. 500 Anos de Povoamento [acesso em 27 setembro 2022]. Disponível em: <https://brasil500anos.ibge.gov.br/>
90. Monteiro JM. As “Raças” Indígenas no Pensamento Brasileiro do Império. In: Maio MC & Santos RV, *Raça Ciência e Sociedade.* Rio de Janeiro, FIOCRUZ/CCBB, p.15-21. 1996.
91. Curtin PD. *The Atlantic slave trade: a census.* University of Wisconsin Press, Madison. 1969.
92. Viana Filho L. *O negro na Bahia.* Rio de Janeiro: Nova Fronteira. 1988.
93. Ribeiro D. *O povo brasileiro: a formação e o sentido do Brasil.* São paulo: Companhia das Letras. 1995.
94. Callegari-Jacques SM, Salzano FM. Brazilian Indian/non-Indian interactions and their effects. *Ciênc Cult* 51:166-174. 1999.

95. Ramos JS. Dos males que vêm com o sangue: as representações raciais e a categoria do imigrante indesejável nas concepções sobre imigração da década de 20. In: Maio MC e Santos RV, *Raça Ciência e Sociedade* Rio de Janeiro, FIOCRUZ/CCBB. p. 59-82. 1996.
96. Klein HS. As origens africanas dos escravos brasileiros. Em S. D. J. Pena (Org.), *Homo Brasilis: aspectos genéticos, lingüísticos, históricos e sociotropológicos da formação do povo brasileiro*. Ribeirão Preto: FUNPEC, 2002.
97. Salzano FM.; Freire-Maia NF. *Problems in human biology: a study of Brazilian populations*. Detroit, Wayne State University Press, 1970.
98. Carvalho-Silva DR, Santos FR, Rocha J, Pena SD. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet*. 2001 ;68(1):281-6..
99. Tetushkin Ela. [Genetic genealogy: history and methodology]. *Genetika*. 2011 May;47(5):581-96.
100. Poletta FA, Orioli IM, Castilla EE. Genealogical data in population medical genetics: Field guidelines. *Genet Mol Biol*. 2014;37(1 Suppl):171-85.
101. Park L. Linkage disequilibrium decay and past population history in the human genome. *PLoS One*. 2012;7(10):e46603.
102. Jeffrey C. Barrett, Chapter 2 - Population Genetics and Linkage Disequilibrium, Editor(s): Eleftheria Zeggini, Andrew Morris, *Analysis of Complex Disease Association Studies*, Academic Press, 2011.
103. Royal CD, Novembre J, Fullerton SM, Goldstein DB, Long JC, Bamshad MJ, Clark AG. Inferring genetic ancestry: opportunities, challenges, and implications. *Am J Hum Genet*. 2010 14;86(5):661-73.
104. Hirose Y, Nakashima , Ohashi H, Mochizuki H, Bando Y, Ogata T, Adachi M, Toba E, Nishimura G, Ikegawa S. Identification of Novel RMRP Mutations and Specific Founder Haplotypes in Japanese Patients With Cartilage-Hair Hypoplasia. *J Hum Genet*. 2006;51(8):706-710.
105. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A*. 1977;74(12):5463-7.
106. Otto EA, Helou J, Allen SJ, O'Toole JF, Wise EL, et al. Mutation analysis in nephronophthisis using a combined approach of homozygosity mapping, CEL I endonuclease cleavage, and direct sequencing. *Hum Mutat*. 2008;29:418–426.
107. Hall TA. BioEdit: A User-Friendly Biological Sequence Alignment Editor and Analysis Program for Windows 95/98/NT. *Nucleic Acids Symposium Series*, 1999, 41, 95-98.

108. Blin G, Denise A, Dulucq S, Herrbach C, Touzet H. VARNA: Interactive drawing and editing of the RNA secondary structure. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 2010;7:309–22.
109. Smith C, Heyne S, Richter AS, Will S, Backofen R. Freiburg RNA Tools: A web server integrating IntaRNA, ExpaRNA and LocARNA. *Nucleic Acids Res.* 2010;38:373–7.
110. Bernhart SH, Hofacker IL, Will S, Gruber AR, Stadler PF. RNAalifold: improved consensus structure prediction for RNA alignments. *BMC Bioinformatics.* 2008;9:474.
111. Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, et al. Rfam 12.0: Updates to the RNA families database. *Nucleic Acids Res.* 2015;43:D130–7.
112. Walker SC, Avis JM. Secondary structure probing of the human RNase MRP RNA reveals the potential for MRP RNA subsets. *Biochem. Biophys. Res. Commun.* 2005;335:314–21.
113. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. *Genetics.* 2000. 155:945–959.
114. Pereira R, Phillips C, Pinto N, Santos C, dos Santos SE, Amorim A, Carracedo A, Gusmão L. Straightforward inference of ancestry and admixture proportions through ancestry-informative insertion deletion multiplexing. *PLoS One.* 2012;7(1):e29684.
115. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015 25;4:7.
116. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005 15;21(2):263-5.
117. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.
118. Nei M. *Molecular evolutionary genetics.* Columbia University Press, 1987, New York, NY, USA
119. Bandelt, H. J.; Forster, P. and Röhl, A. 1999. Median-joining networks for inferring intraspecific phylogenies. *Molecular Biology and Evolution* 16:37-48.
120. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods.* 2001;25(4):402-8.

121. Pizzatti L, Panis C, Lemos G, Rocha M, Cecchini R, Souza GH, Abdelhay E. Label-free MSE proteomic analysis of chronic myeloid leukemia bone marrow plasma: disclosing new insights from therapy resistance. *Proteomics*. 2012;12(17):2618-31.
122. Wang Y, Wang Q, Huang H, Huang W, Chen Y, McGarvey PB, Wu CH, Arighi CN; UniProt Consortium. A crowdsourcing open platform for literature curation in UniProt. *PLoS Biol*. 2021 6;19(12):e3001464.
123. Oliveros, J.C. (2007-2015) Venny. An interactive tool for comparing lists with Venn's diagrams. Disponível em: <https://bioinfogp.cnb.csic.es/tools/venny/index.html>. Acessado em: 13 de junho de 2022.
124. Fonseka P, Pathan M, Chitti SV, Kang T, Mathivanan S. FunRich enables enrichment analysis of OMICs datasets. *J Mol Biol*. 2021 28;433(11):166747.
125. Pathan, M., Keerthikumar, S., Chisanga, D., Alessandro, R., Ang, C.S., Askenase, P., Batagov, A.O., Benito-Martin, A., Camussi, G., Clayton, A., Collino, F., Di Vizio, D., Falcon-Perez, J.M., Fonseca, P., Fonseka, P., Fontana, S., Gho, Y.S., Hendrix, A., Nolte-'t Hoen, E., Iraci, N., Kastaniegaard, K., Kislinger, T., Kowal, J., Kurochkin, I.V., Leonardi, T., Liang, Y., Llorente, A., Lunavat, T.R., Maji, S., Monteleone, F., Overbye, A., Panaretakis, T., Patel, T., Peinado, H., Pluchino, S., Principe, S., Ronquist, G., Royo, F., Sahoo, S., Spinelli, C., Stensballe, A., Thery, C., van Herwijnen, M., Wauben, M., Welton, J., Zhao, J. and Mathivanan, S. (2017) A novel community driven software for functional enrichment analysis of extracellular vesicles data. *J Extracellular Vesicles*. 1:1321455.
126. Pathan, M., Keerthikumar, S., Ang, C.-S., Gangoda, L., Quek, C.Y., Williamson, N.A., Mouradov, D., Sieber, O.M., Simpson, R.J., Salim, A., Bacic, A., Hill, A.F., Stroud, D.A., Ryan, M.T., Agbinya, J.I., Mariadason, J.M., Burgess, A.W. and Mathivanan, S. (2015), FunRich: An open access standalone functional enrichment and interaction network analysis tool. *Proteomics*, 15: 2597-2601.
127. Marc Gillespie, Bijay Jassal, Ralf Stephan, Marija Milacic, Karen Rothfels, Andrea Senff-Ribeiro, Johannes Griss, Cristoffer Sevilla, Lisa Matthews, Chuqiao Gong, Chuan Deng, Thawfeek Varusai, Eliot Ragueneau, Yusra Haider, Bruce May, Veronica Shamovsky, Joel Weiser, Timothy Brunson, Nasim Sanati, Liam Beckman, Xiang Shao, Antonio Fabregat, Konstantinos Sidiropoulos, Julieth Murillo, Guilherme Viteri, Justin Cook, Solomon Shorser, Gary Bader, Emek Demir, Chris Sander, Robin Haw, Guanming Wu, Lincoln Stein, Henning Hermjakob, Peter D'Eustachio, The reactome pathway knowledgebase 2022, *Nucleic Acids Research*, 2021; gkab1028

128. van Eenennaam H, Jarrous N, van Venrooij WJ, Pruijn GJ. Architecture and function of the human endonucleases RNase P and RNase MRP. *IUBMB Life*. 2000;49(4):265-72.
129. van Eenennaam H, Vogelzangs JHP, Lugtenberg D, Van Den Hoogen FHJ, van Venrooij WJ, et al. (2002) Identity of the RNase MRP- and RNase P-associated Th/To autoantigen. *Arthritis Rheum* 46: 3266–3272.
130. Pluk H, van Eenennaam H, Rutjes SA, Pruijn GJ, van Venrooij WJ. RNA-protein interactions in the human RNase MRP ribonucleoprotein complex. *RNA*. 1999 ;5(4):512-24.
131. van Eenennaam H, van der Heijden A, Janssen RJ, van Venrooij WJ, Pruijn GJ. Basic domains target protein subunits of the RNase MRP complex to the nucleolus independently of complex association. *Mol Biol Cell*. 2001;12(11):3680-9
132. Dávila López M, Rosenblad MA, Samuelsson T. Conserved and variable domains of RNase MRP RNA. *RNA Biol*. 2009 ;6(3):208-20.
133. Barton NH and Charlesworth B. Genetic Revolutions, Founder Effects, and Speciation. *Annual Review of Ecology and Systematics* 1984 15:1, 133-164
134. Barton NH and Charlesworth B. Genetic Revolutions, Founder Effects, and Speciation. *Annual Review of Ecology and Systematics* 1984 15:1, 133-164
135. GIUSTINA, L. D. et al. Population structure and genetic diversity in natural populations of *Theobroma speciosum* Willd. Ex Spreng (Malvaceae). *Genetics and molecular Research*, v.13, n.5, p.47-53, 2014.
136. Michael H. Cho MD, MPH David A. Schwartz MD Murray & Nadel's. *Textbook of Respiratory Medicine*, Volume 2, 7ª edição, 9, 106-115.e5. Elsevier, 2022
137. Alves-Silva J, da Silva Santos M, Guimarães PE, Ferreira AC, Bandelt HJ, Pena SD, Prado VF. The ancestry of Brazilian mtDNA lineages. *Am J Hum Genet*. 2000;67(2):444-61.
138. Carvalho-Silva DR, Santos FR, Rocha J, Pena SD. The phylogeography of Brazilian Y-chromosome lineages. *Am J Hum Genet*. 2001;68(1):281-6.
139. Paule MR, White RJ. Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res*. 2000 15;28(6):1283-98.
140. Murphy S, Tripodi M, Melli M. A sequence upstream from the coding region is required for the transcription of the 7SK RNA genes. *Nucleic Acids Res*. 1986 9;14(23):9243-60.

141. Yuan Y, Reddy R. 5' flanking sequences of human MRP/7-2 RNA gene are required and sufficient for the transcription by RNA polymerase III. *Biochim Biophys Acta*. 1991 2;1089(1):33-9.
142. Ismail T, Kim Y, Lee H, Lee DS, Lee HS. Interplay Between Mitochondrial Peroxiredoxins and ROS in Cancer Development and Progression. *Int J Mol Sci*. 2019 Sep 7;20(18):4407.
143. Zhou L, Fu Z, Wang S, Jia J, Cheng Y, Zheng Y, Zhang N, Lu W, Yao Z. ACYP1 Is a Pancancer Prognostic Indicator and Affects the Immune Microenvironment in LIHC. *Front Oncol*. 2022 2; 12:875097.
144. Huang P, Cai Y, Zhao B, Cui L. Roles of NUCKS1 in Diseases: Susceptibility, Potential Biomarker, and Regulatory Mechanisms. *Biomed Res Int*. 2018 29; 2018:7969068.
145. Arrington ME, Temple B, Schaefer A, Campbell SL. The molecular basis for immune dysregulation by the hyperactivated E62K mutant of the GTPase RAC2. *J Biol Chem*. 2020 21;295(34):12130-12142.
146. Chandramouli K, Qian PY. Proteomics: challenges, techniques and possibilities to overcome biological sample complexity. *Hum Genomics Proteomics*. 2009 8; 2009:239204.
147. Braconi D, Bernardini G, Spiga O, Santucci A. Leveraging proteomics in orphan disease research: pitfalls and potential. *Expert Rev Proteomics*. 2021;18(4):315-327.
148. Chabronova A, van den Akker GGH, Meekels-Steinbusch MMF, Friedrich F, Cremers A, Surtel DAM, Peffers MJ, van Rhijn LW, Lausch E, Zabel B, Caron MMJ, Welting TJM. Uncovering pathways regulating chondrogenic differentiation of CHH fibroblasts. *Noncoding RNA Res*. 2021 12;6(4):211-224.
149. Vakkilainen S, Skoog T, Einarsdottir E, Middleton A, Pekkinen M, Öhman T, Katayama S, Krjutškov K, Kovanen PE, Varjosalo M, Lindqvist A, Kere J, Mäkitie O. The human long non-coding RNA gene RMRP has pleiotropic effects and regulates cell-cycle progression at G2. *Sci Rep*. 2019 24;9(1):13758.
150. Venturi G, Montanaro L. How Altered Ribosome Production Can Cause or Contribute to Human Disease: The Spectrum of Ribosomopathies. *Cells*. 2020 15;9(10):2300.
151. Lygerou Z, Allmang C, Tollervey D, Séraphin B. Accurate processing of a eukaryotic precursor ribosomal RNA by ribonuclease MRP in vitro. *Science*. 1996 12;272(5259):268-70.
152. Sun X, Zhang R, Liu M, Chen H, Chen L, Luo F, Zhang D, Huang J, Li F, Ni Z, Qi H, Su N, Jin M, Yang J, Tan Q, Du X, Chen B, Huang H, Chen S, Yin L, Xu X, Deng C, Luo L, Xie Y, Chen L. Rmrp Mutation Disrupts Chondrogenesis and

- Bone Ossification in Zebrafish Model of Cartilage-Hair Hypoplasia via Enhanced Wnt/ β -Catenin Signaling. *J Bone Miner Res.* 2019 ;34(11):2101-2116.
153. Wu X, Quondamatteo F, Lefever T, Czuchra A, Meyer H, Chrostek A, Paus R, Langbein L, Brakebusch C. Cdc42 controls progenitor cell differentiation and beta-catenin turnover in skin. *Genes Dev.* 2006 1;20(5):571-85
154. Wang G, Woods A, Agoston H, Ulici V, Glogauer M, Beier F. Genetic ablation of Rac1 in cartilage results in chondrodysplasia. *Dev Biol.* 2007 15;306(2):612-23.
155. Saoudi A, Kassem S, Dejean A, Gaud G. Rho-GTPases as key regulators of T lymphocyte biology. *Small GTPases.* 2014;5:e28208.
156. Gammill LS, Bronner-Fraser M. Neural crest specification: migrating into genomics. *Nat Rev Neurosci.* 2003 ;4(10):795-805.

11. ANEXOS

11.1 Aprovação do comitê de ética em pesquisa com seres humanos



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Caracterização de mutações, análise da expressão do gene RMRP e avaliação do perfil imunológico de pacientes com doença de McKusick

Pesquisador: Sayonara Maria de Carvalho Gonzalez

Área Temática: Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP.);

Versão: 2

CAAE: 50019315.8.0000.5269

Instituição Proponente: Instituto Fernandes Figueira - IFF/ FIOCRUZ - RJ/ MS

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 1.339.799

Apresentação do Projeto:

Este projeto tem como objetivo identificar mutações no gene RMRP em pacientes encaminhados ao IFF com características clínicas e radiográficas compatíveis com doença de McKusick. Esse estudo também visa avaliar a expressão do gene RMRP e a imunidade celular e humoral desses pacientes.

Objetivo da Pesquisa:

Objetivo Primário:

Descrever os genótipos, as características fenotípicas e os níveis de expressão do gene RMRP em pacientes com suspeita de Doença de McKusick

Objetivo secundário:

- Investigar a presença de mutações em pacientes com Doença de McKusick.- Avaliar os níveis de expressão do gene RMRP em pacientes com a doença de McKusick.- Descrever o perfil imunológico celular e humoral dos pacientes com Doença de McKusick

Avaliação dos Riscos e Benefícios:

Riscos:

A etapa de exames laboratoriais pode apresentar um possível risco relacionado à coleta de sangue,

Endereço: RUI BARBOSA, 716

Bairro: FLAMENGO

CEP: 22.250-020

UF: RJ

Município: RIO DE JANEIRO

Telefone: (21)2554-1730

Fax: (21)2552-8491

E-mail: cepiff@iff.fiocruz.br

INSTITUTO FERNANDES
FIGUEIRA - IFF/ FIOCRUZ - RJ/
MS



Continuação do Parecer: 1.339.799

com raras complicações e geralmente de pequeno porte. Um pequeno desconforto ou sensação de dor pode ocorrer no paciente em decorrência da inserção da agulha no momento deste procedimento. O extravasamento de sangue da veia no local da punção, caso ocorra, pode levar a um hematoma e um leve desconforto com duração de poucos dias. Com o intuito de diminuir as possibilidades de complicações, as coletas de sangue serão realizadas por pessoas treinadas para este fim. Também há o risco de confidencialidade dos pacientes. Para evitar a quebra de sigilo e o vazamento de informações, todo e qualquer documento será armazenado em arquivo com chave, tendo acesso somente pessoas autorizadas e relacionadas à pesquisa. Também serão utilizados números como forma de identificação dos pacientes, estando o controle dessas informações mantido em caderno de acesso restrito.

Benefícios:

As complicações clínicas associadas à Doença de McKusick podem causar grande impacto nos pacientes. O conhecimento das mutações que originam a doença em pacientes individuais é relevante para a confirmação do diagnóstico clínico e para a realização do aconselhamento genético apropriado. A avaliação imunológica dos pacientes pode permitir o encaminhamento do paciente para acompanhamento a fim de evitar complicações causadas por possíveis infecções e/ou surgimento de neoplasias. A vigilância para infecções por vírus EBV, CMV e/ou HHV-6 (sabidamente associadas com o desenvolvimento de linfoma), bem como o acompanhamento periódico da carga viral nos pacientes com quadro infeccioso, poderá contribuir para detecção precoce desse tipo de neoplasia nos pacientes, melhorando o prognóstico da doença. A caracterização clínica e molecular dos grupos de pacientes afetados no Brasil contribuirá para o melhor entendimento desta doença na nossa população e os dados produzidos poderão servir de subsídio para estudos epidemiológicos da doença no nosso país.

Comentários e Considerações sobre a Pesquisa:

Pesquisa relevante para os objetivos institucionais e contando com colaborações multi-intitucionais que permitirão avanço na área de doenças raras no país.

Endereço: RUI BARBOSA, 716

Bairro: FLAMENGO

CEP: 22.250-020

UF: RJ

Município: RIO DE JANEIRO

Telefone: (21)2554-1730

Fax: (21)2552-8491

E-mail: cepiff@iff.fiocruz.br

**INSTITUTO FERNANDES
FIGUEIRA - IFF/ FIOCRUZ - RJ/
MS**



Continuação do Parecer: 1.339.799

Considerações sobre os Termos de apresentação obrigatória:

Todos os termos foram apresentados.

Recomendações:

Conclusões ou Pendências e Lista de Inadequações:

Projeto atendeu todas as solicitações, sendo recomendado a aprovação.

Considerações Finais a critério do CEP:

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_593187.pdf	25/11/2015 17:11:13		Aceito
Projeto Detalhado / Brochura Investigador	Projeto_versao_2.doc	25/11/2015 16:07:23	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Versao_corrigida_TCLE.doc	25/11/2015 16:04:23	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Instituição e Infraestrutura	Carta_Patologia_IFF.pdf	23/11/2015 18:27:28	Sayonara Maria de Carvalho Gonzalez	Aceito
Orçamento	Orcamento_CHH.docx	12/10/2015 18:29:26	Sayonara Maria de Carvalho Gonzalez	Aceito
Folha de Rosto	Folha_de_rosto_assinada.pdf	12/10/2015 18:07:05	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Pesquisadores	Carta_de_Coparticipacao_HCPA_Dra_Temis_Felix.pdf	07/10/2015 16:03:48	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Pesquisadores	Carta_de_Coparticipacao_HCPA_Dra_Maria_Tereza.pdf	07/10/2015 16:03:19	Sayonara Maria de Carvalho Gonzalez	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_Doenca_de_McKusick_07_10_2015_submissao_CEP.doc	07/10/2015 15:55:40	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Instituição e Infraestrutura	Carta_Infraestrutura_Laboratorio_de_Medicina_Genomica.pdf	07/10/2015 15:52:48	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Instituição e Infraestrutura	Carta_de_Infraestrutura_INCA.doc	07/10/2015 15:48:05	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Instituição e Infraestrutura	Carta_Infraestrutura_IOC.jpg	07/10/2015 15:47:16	Sayonara Maria de Carvalho Gonzalez	Aceito

Endereço: RUI BARBOSA, 716

Bairro: FLAMENGO

CEP: 22.250-020

UF: RJ

Município: RIO DE JANEIRO

Telefone: (21)2554-1730

Fax: (21)2552-8491

E-mail: cepiff@iff.fiocruz.br

**INSTITUTO FERNANDES
FIGUEIRA - IFF/ FIOCRUZ - RJ/
MS**



Continuação do Parecer: 1.339.799

Declaração de Pesquisadores	carta_de_copartipacao_INCA.doc	07/10/2015 15:46:46	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Pesquisadores	Carta_de_Coparticipacao_UNICAMP.pdf	07/10/2015 15:46:28	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_IFF_06_a_12_anos.doc	07/10/2015 15:19:52	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_HCPA_06_a_12_anos.doc	07/10/2015 15:19:13	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_HCPA_12_a_18_anos.doc	07/10/2015 14:02:36	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TALE_IFF_12_a_18_anos.doc	07/10/2015 14:02:21	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_HCPA.doc	06/10/2015 21:13:47	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_IFF.doc	06/10/2015 21:13:36	Sayonara Maria de Carvalho Gonzalez	Aceito
Cronograma	Cronograma.doc	05/10/2015 22:11:26	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Pesquisadores	Carta_Chefe_Departamento_Juan_Llerena.pdf	05/10/2015 21:21:55	Sayonara Maria de Carvalho Gonzalez	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_FCM_UNICAMP.pdf	01/10/2015 14:55:31	Sayonara Maria de Carvalho Gonzalez	Aceito
Declaração de Instituição e Infraestrutura	Carta_Vice_Pesquisa.jpg	01/10/2015 14:51:01	Sayonara Maria de Carvalho Gonzalez	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Endereço: RUI BARBOSA, 716

Bairro: FLAMENGO

CEP: 22.250-020

UF: RJ

Município: RIO DE JANEIRO

Telefone: (21)2554-1730

Fax: (21)2552-8491

E-mail: cepiff@iff.fiocruz.br

INSTITUTO FERNANDES
FIGUEIRA - IFF/ FIOCRUZ - RJ/
MS



Continuação do Parecer: 1.339.799

RIO DE JANEIRO, 26 de Novembro de 2015

Assinado por:
maria elisabeth lopes moreira
(Coordenador)

Endereço: RUI BARBOSA, 716

Bairro: FLAMENGO

CEP: 22.250-020

UF: RJ

Município: RIO DE JANEIRO

Telefone: (21)2554-1730

Fax: (21)2552-8491

E-mail: cepiff@iff.fiocruz.br

11.2 Lista de oligonucleotídeos sintéticos

Oligonucleotídeo	Técnica	Sequência (5'-3')	Tamanho do Fragmento	Concentração na reação
CHH	PCR convencional	F: GCCACAATGCCTCTGAAAGC R: GTTCTTGGCCTTGCTATGAAG	719 pb	200 µM
CHH_HindIII	PCR convencional	F: CTCTCTCTAAGCTTTTTGGCTCCTGGGAGCTCATCTAC R: CTCTCTCTAAGCTTGGGACTGCTTGTCTAGTGGTATTG	8.018 pb	200 µM
rs1361338	PCR convencional	F: TTTGGCTCCTGGGAGCTCATCTAC R: CCCAAGGCATCCACAAAGCTGAG	862pb	200 µM
rs1339374	PCR convencional	F: GCCTGGGCAACACAGTGAGTGAG R: CGGACAAGGATATGGAGGCTTCAG	898pb	200 µM
RS7021463	PCR convencional	F: GCTAGCCACAATGCCTCTGAAAGC R: CAAGGTCACCGCTCTTGAGTACAG	896pb	200 µM
rs10972552	PCR convencional	F: AGTAGCCACCATTCACTGTGTAGC R: GGGACTGCTTGTCTAGTGGTATTG	563pb	200 µM
<i>RMRP</i>	RT-qPCR	F: TGCTGAAGGCCTGTATCC R: CTGCCTGCGTAACTAGAG	213 pb	500 µM
GAPDH	RT-qPCR	F:AATCCCATCACCATCTTCCA R: TGGACTCCACGACTACTCA	82 pb	500 µM
TOP1	RT-qPCR	F: AAGGTCCAGTATTTGCCCCAC R: ATTCATGGTCGAGCATTTTTGC	135pb	500 µM

11.3 Médias de ancestralidade a nível individual analisando com 3 populações parentais geradas pelo Software Structure

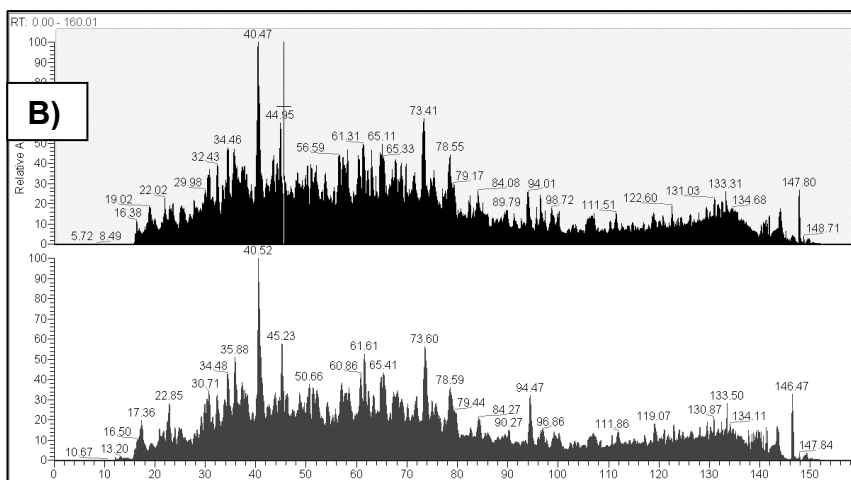
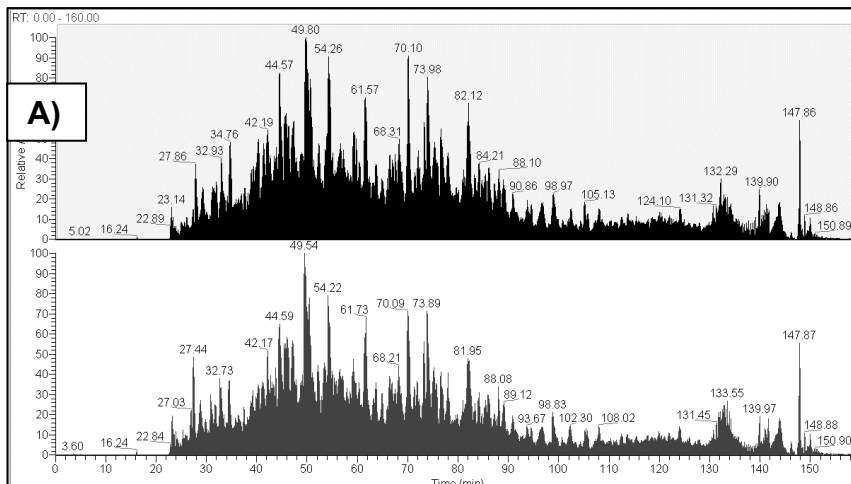
PACIENTE	% AFR	% EUR	% AME
1	0,12	0,719	0,161
2	0,282	0,537	0,181
3	0,178	0,686	0,136
4	0,108	0,706	0,187
5	0,28	0,493	0,226
6	0,238	0,502	0,26
7	0,389	0,458	0,154
14	0,138	0,606	0,257
15	0,199	0,631	0,17
16	0,331	0,532	0,137
17	0,345	0,495	0,159
18	0,257	0,642	0,101
19	0,233	0,57	0,197
20	0,384	0,493	0,123
21	0,127	0,699	0,174
22	0,174	0,546	0,28
23	0,437	0,371	0,192
24	0,217	0,612	0,171
25	0,251	0,605	0,144
26	0,159	0,734	0,107
27	0,191	0,647	0,162

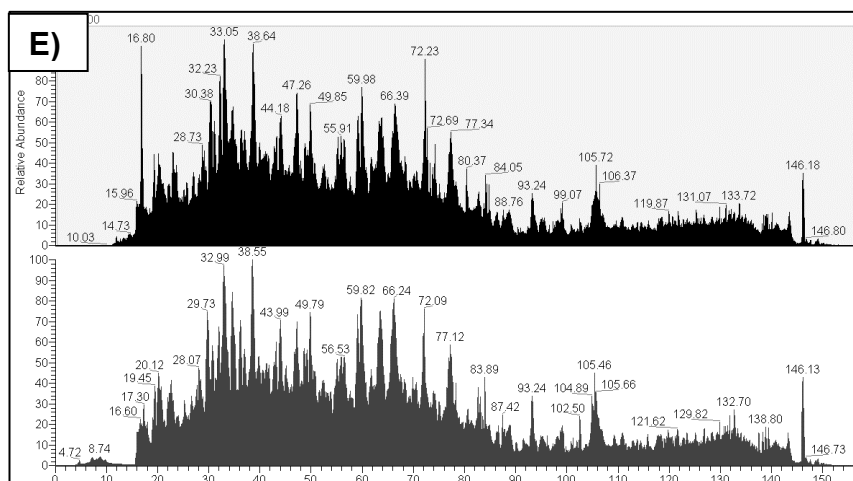
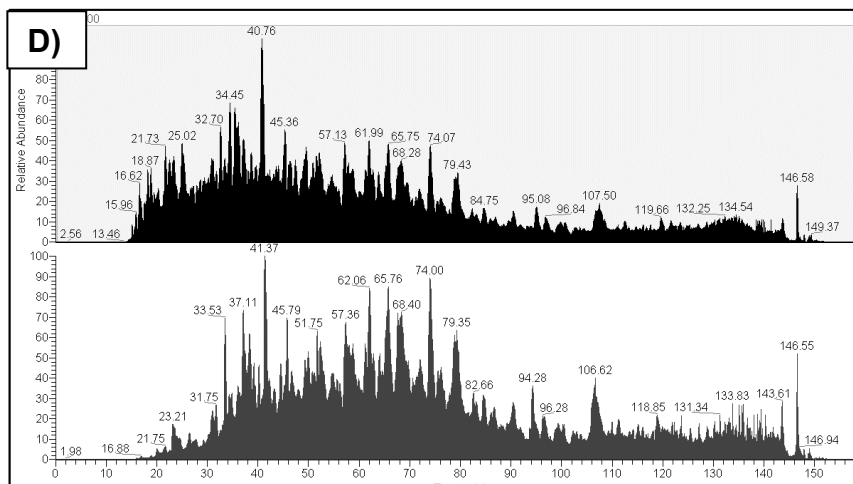
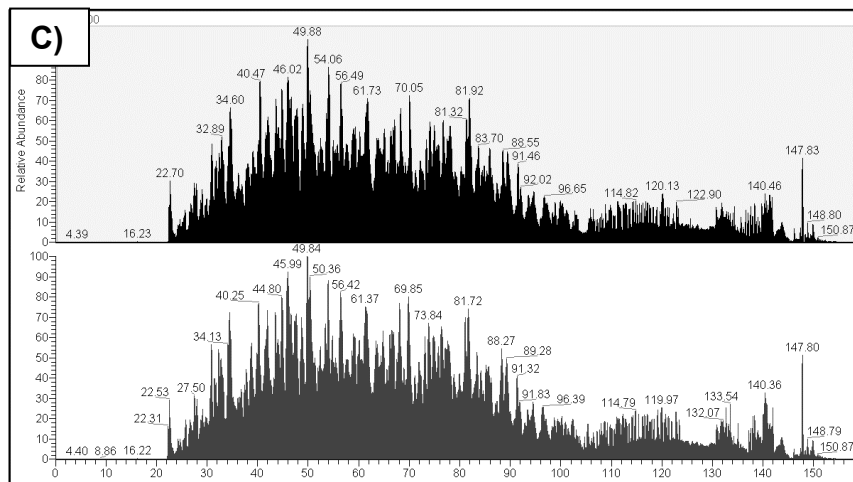
11.4 Lista de haplótipos encontrados nos pacientes

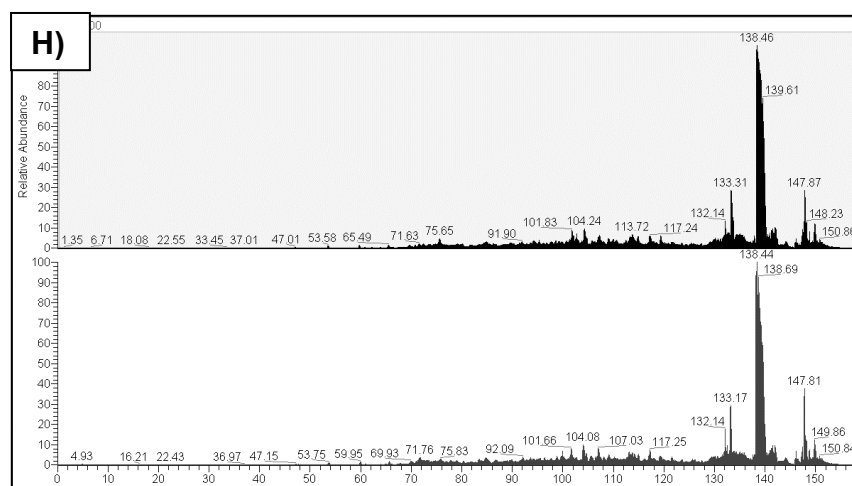
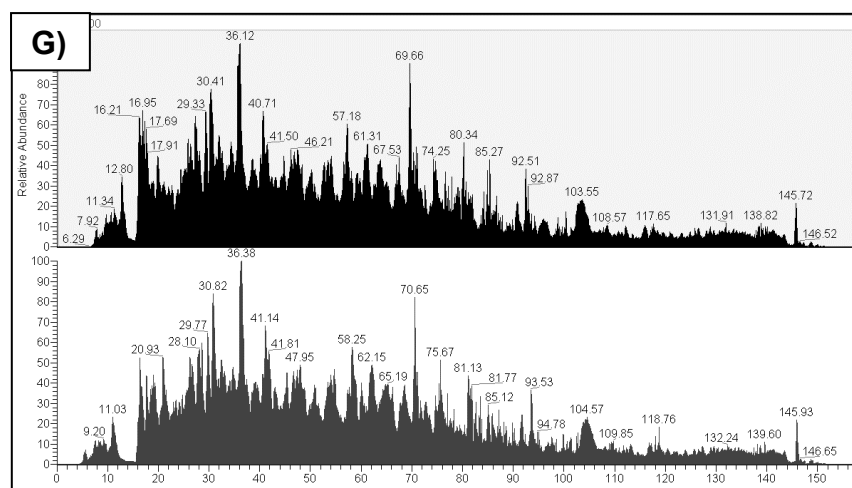
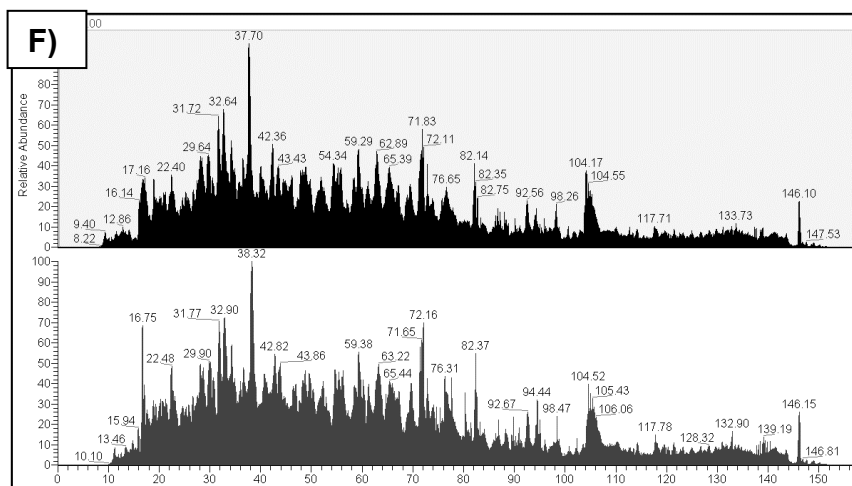
Paciente 1	T	C	G	A	g.196C>T
	T	C	G	C	g.-21_-2dup20(ACTCTGTGAAGCTGAGGACG)
Paciente 2	T	G	G	C	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)
	T	C	G	A	g.196C>T
Paciente 3	T	C	G	A	g.194G>A
	C	C	G	A	g.-19_-3trip17(TCTGTGAAGCTGAGGAC)
Paciente 4	C	C	G	C	g.-24_-15dup10(ACTACTCTGT)
	T	C	G	A	g.97_98dup2(TG)
Paciente 5	T	C	G	A	g.196C>T
	T	G	G	C	g.-22_-14dup9(TACTCTGTG)
Paciente 6	T	C	G	A	g.196C>T
	T	C	G	A	g.97_98dup2(TG)
Paciente 7	C	C	A	A	g.10T>A
	T	C	G	A	g.196C>T
Paciente 8	C	C	A	A	g.-22_-3dup20(TACTCTGTGAAGCTGAGGAC)
	T	C	G	A	g.196C>T
Paciente 9	T	C	G	A/C	g.196C>T
	T	G	G	C/A	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)
Paciente 10	C	C	A	A	g.-6_+4dup10(GGACGTGGTT)
	C	C	G	C	g.71A>G
Paciente 11	T/C	C	G	A/C	g.196C>T
	T/C	C	G	C/A	245G>C
Paciente 14	C	C	A	A	g.181G>A
	T	C	G	A	g.-19_-3dup17(TCTGTGAAGCTGAGGAC)
Paciente 15	T	C	G	A	g.196C>T
	T	C	G	A	g.6G>T
Paciente 16	C	C	G	C	g.36C>G
	C	C	A	C	g.252A>C
Paciente 18	T	C	G	A	g.196C>T
	T	C	G	A	g.-13_+1dup14(AAGCTGAGGACGTG)
Paciente 19	C/T	C	G	A	g.196C>T
	C/T	C	G	C	g.-8_-7ins20(GGGACTACTCTGTGAAGCTG)
Paciente 20	T	C	G	A	g.97_98dup2(TG)
	T	G	G	C	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)
Paciente 21	C	C	G	C	g.147G>C
	T	C	G	A	g.97_98dup2(TG)
Paciente 22	T	C	G	A	g.195C>T
	T	C	G	A	g.6G>T
Paciente 23	T	C	G	A	g.196C>T

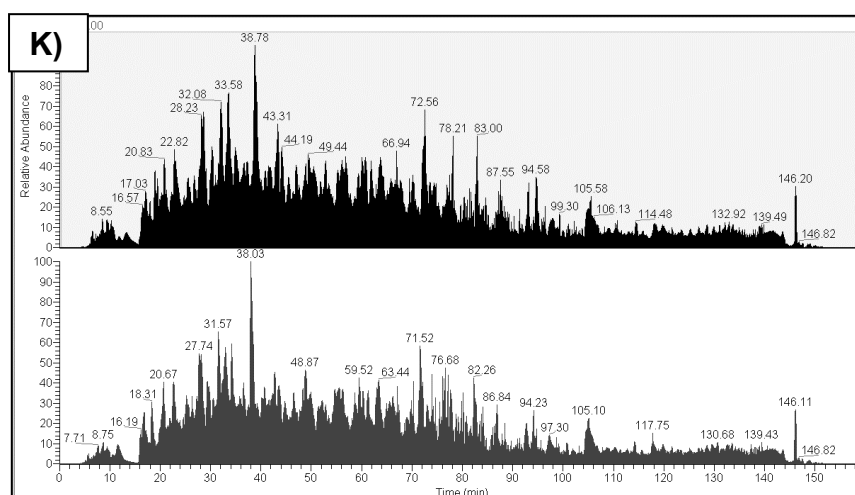
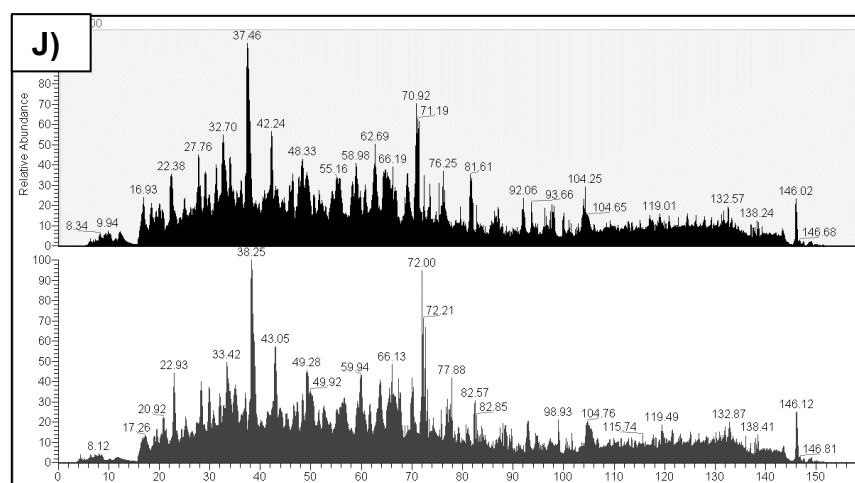
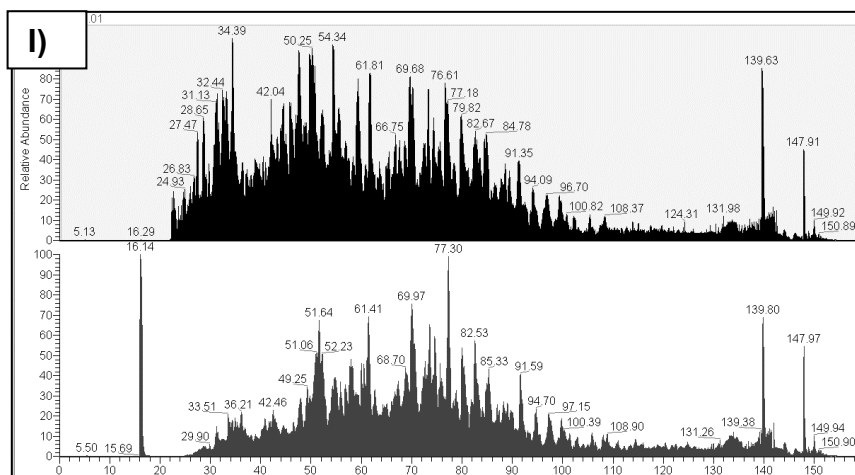
	T	G	G	C	g.-22_-1dup20(TACTCTGTGAAGCTGAGAAC)
Paciente 25	T	C	G	A	g.196C>T
	T	G	G	C	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)
Paciente 26	T	C	G	A	g.98G>A
	T	C	G	A	g.36C>G
Paciente 27	T	G	G	C	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)
	T	C	G	A	g.196C>T
Paciente 29	C	C	G	C	g.36C>T
	C	C	G	C	g.196C>T
Paciente 30	T	C	G	A	g.5C>T
	T	G	G	C	g.-20_+3dup23(CTCTGTGAAGCTGAGAACGTGGT)
Paciente 31	T	C	G	A	g.196C>T
	T	G	G	C	g.-25_-4dup22(TACTACTCTGTGAAGCTGAGAA)
Paciente 32	T	C	G	A	g.196C>T
	T	C	G	A	g.-24_-10dup15(ACTACTCTGTGAAGC)

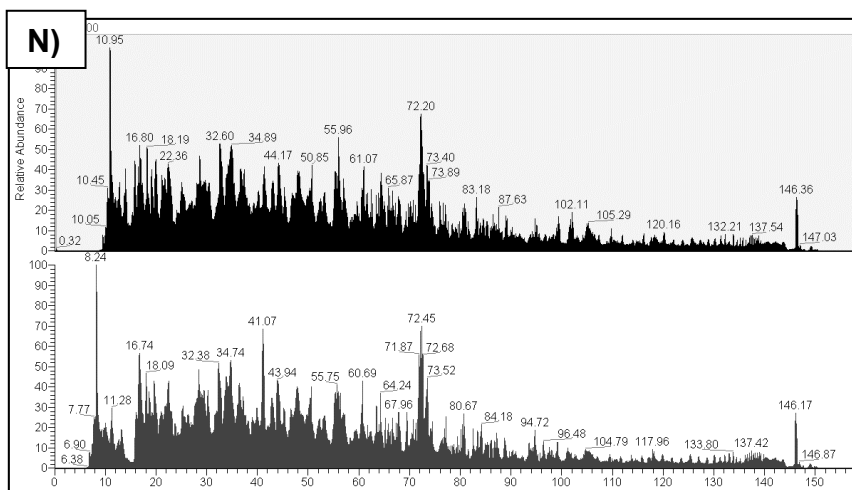
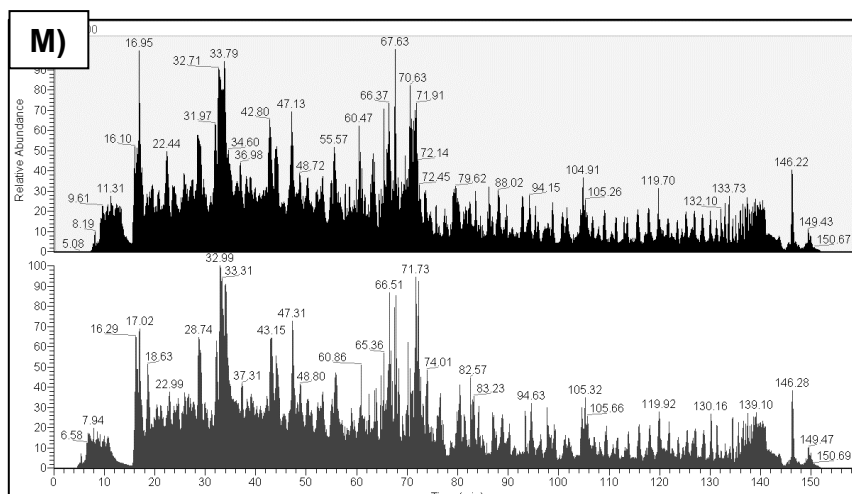
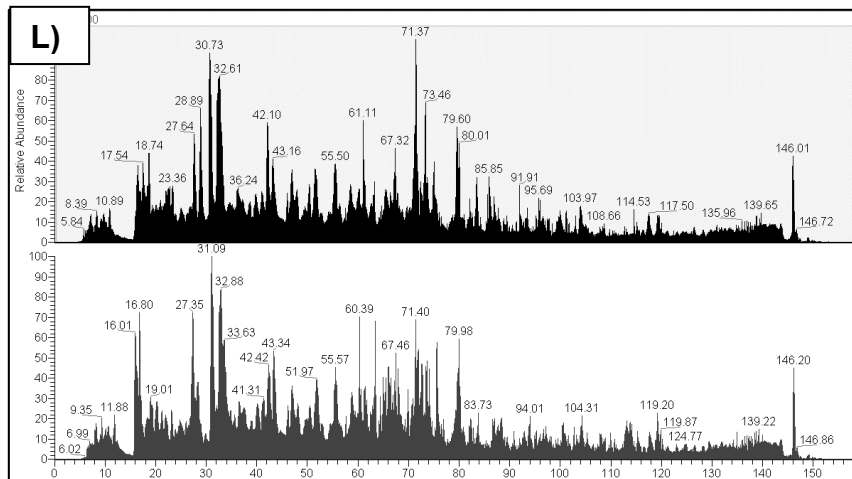
11.5 Cromatogramas da análise por LC-MS/MS das amostras de pacientes e controles em duplicatas técnicas (A-V)

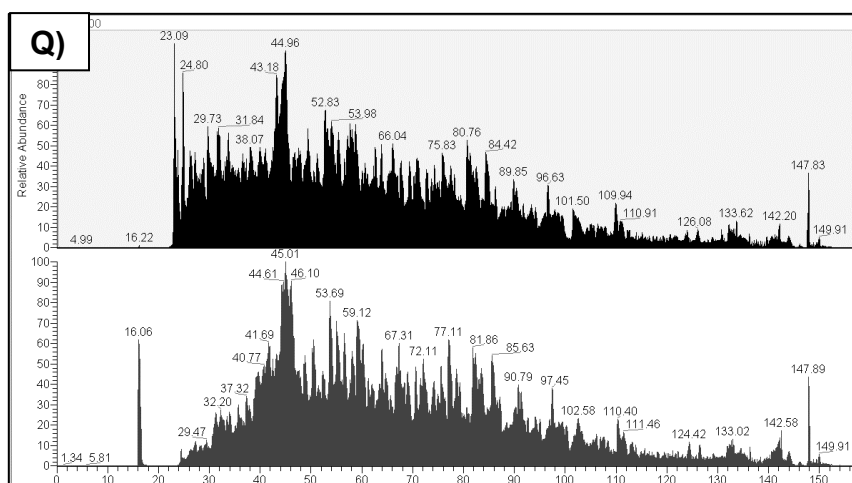
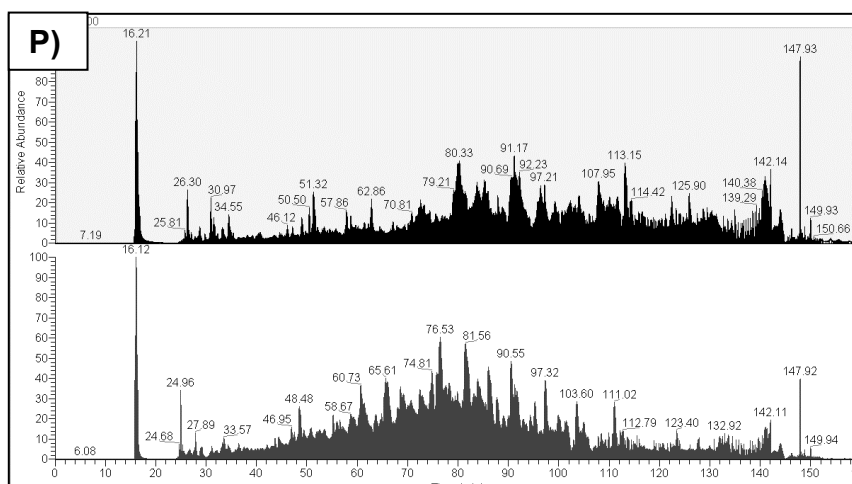
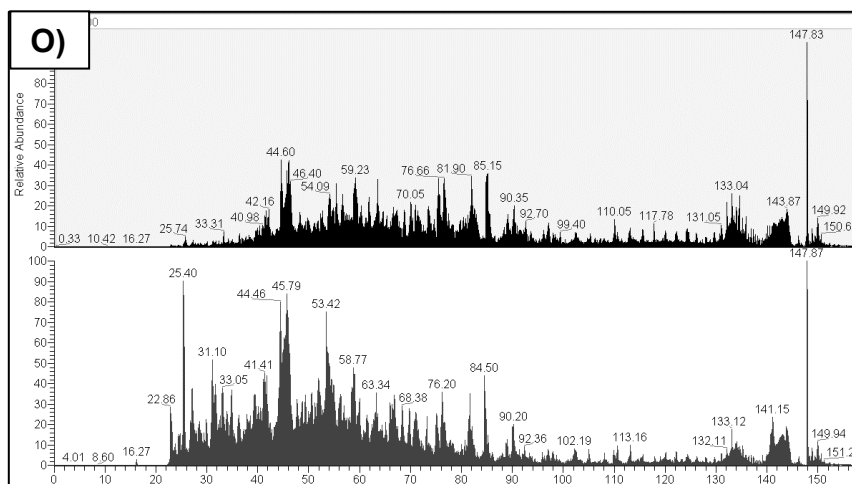


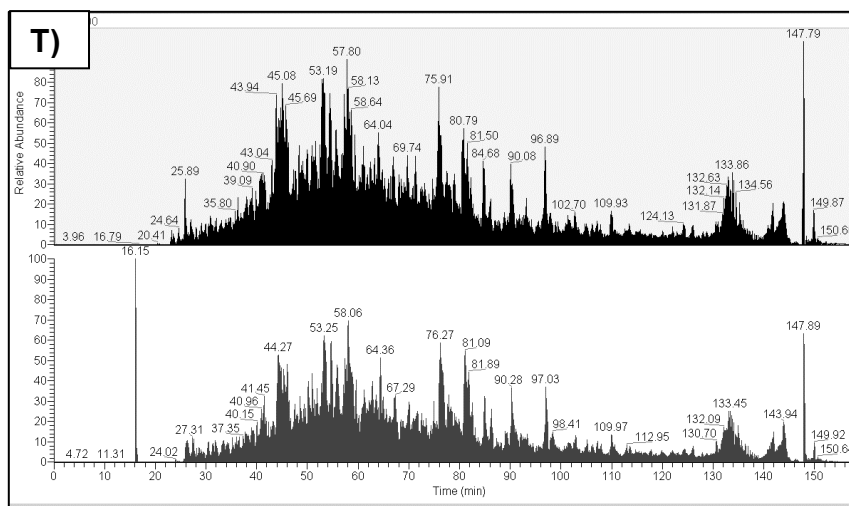
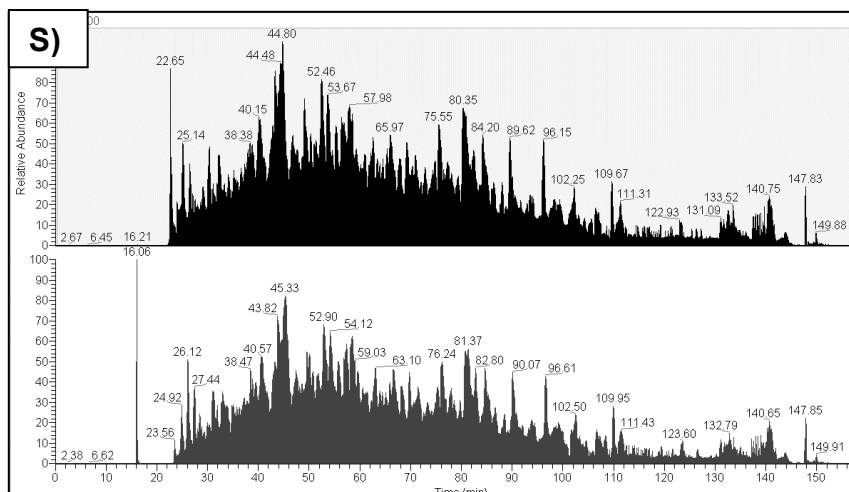
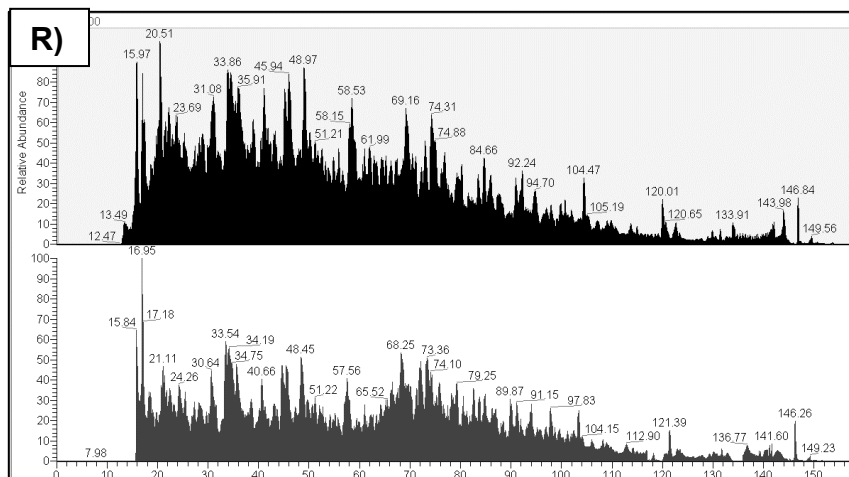












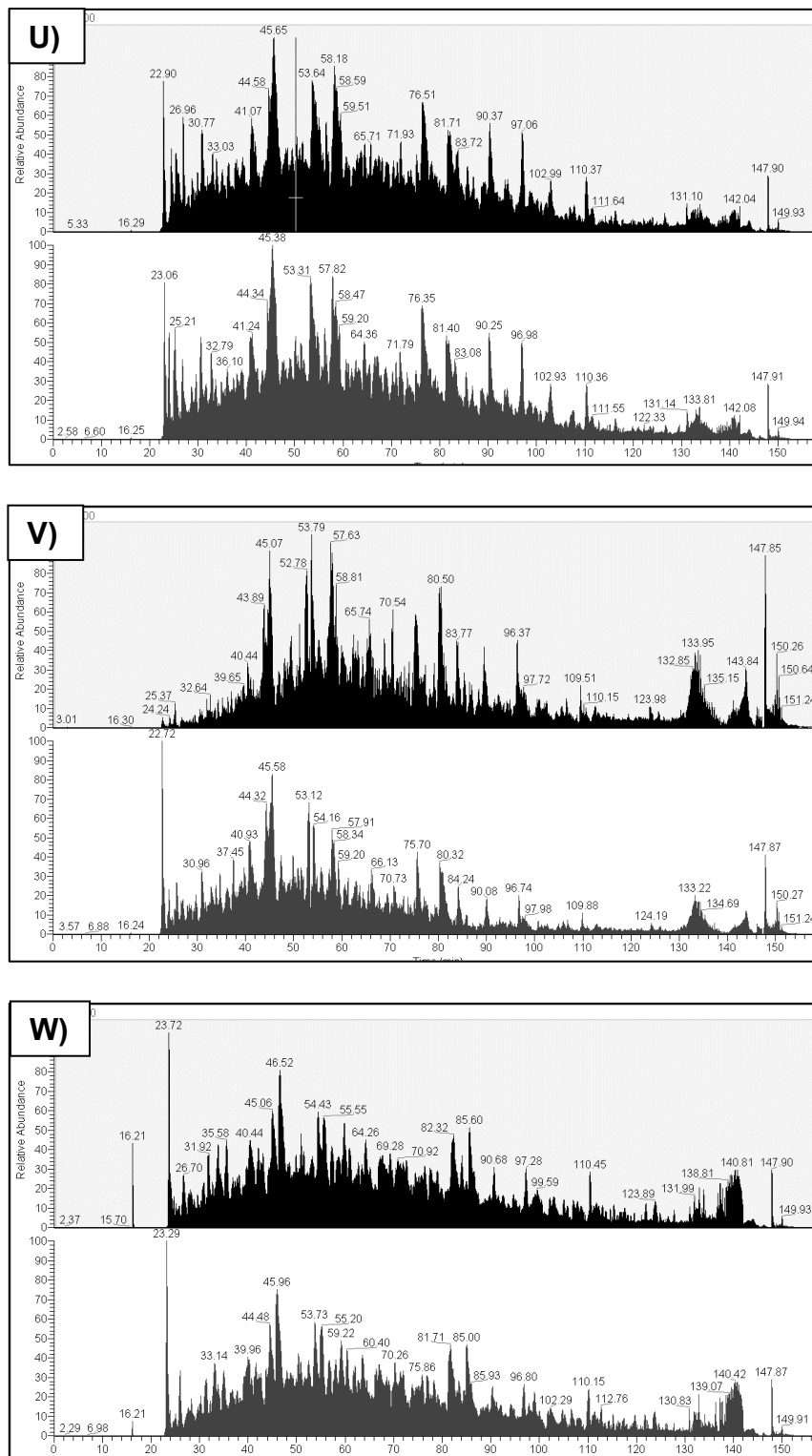


Figura anexo 11.5A-V: Cromatogramas resultantes da análise por LC-MS/MS das amostras de pacientes e controles em duplicatas técnicas (amostra dos pacientes 4, 6, 23, 24, 25, 1, 2, 3, 29, 8, 5, 9, 10, 26, 27, 32, 30 e 31 e dos controles 1, 2, 3, 4 e 5, respectivamente). Os cromatogramas em preto representam a primeira replicata da respectiva amostra e os cromatogramas em vermelho representam a segunda replicata técnica da mesma.

11.6 Proteínas diferencialmente expressas em leucócitos de pacientes com CHH em relação aos controles saudáveis.

Proteínas diferencialmente expressas em leucócitos de pacientes com CHH em relação ao grupo controle				
Proteínas reguladas positivamente				
Número de acesso (UniProt)	Descrição da Proteína	-log P-valor	Diferença de expressão (log2)	
Q05682-4	Isoform 4 of Caldesmon GN=CALD1	-	Exclusiva	
P00450	Ceruloplasmin GN=CP SV=1	-	Exclusiva	
O75636	Ficolin-3 GN=FCN3 SV=2	-	Exclusiva	
P31943	Heterogeneous nuclear ribonucleoprotein H GN=HNRNPH1 SV=4	-	Exclusiva	
P62805	Histone H4 GN=H4C1 SV=2	-	Exclusiva	
P04908	Histone H2A type 1-B/E GN=H2AC4 SV=2	-	Exclusiva	
Q9UHL4	Dipeptidyl peptidase 2 GN=DPP7 SV=3	-	Exclusiva	
P13284	Gamma-interferon-inducible lysosomal thiol reductase GN=IFI30 SV=3	-	Exclusiva	
Q08380	Galectin-3-binding protein GN=LGALS3BP SV=1	-	Exclusiva	
P10619	Lysosomal protective protein GN=CTSA SV=2	-	Exclusiva	
P17096-2	Isoform HMG-Y of High mobility group protein HMG-I/HMG-Y GN=HMGA1	-	Exclusiva	
P07305	Histone H1.0 GN=H1-0 SV=3	-	Exclusiva	
P19827	Inter-alpha-trypsin inhibitor heavy chain H1 GN=ITIH1 SV=3	-	Exclusiva	
P13688	Carcinoembryonic antigen-related cell adhesion molecule 1 GN=CEACAM1 SV=2	-	Exclusiva	
P01624	Immunoglobulin kappa variable 3-15 GN=IGKV3-15 SV=2	-	Exclusiva	
Q9NYF8	Bcl-2-associated transcription factor 1 GN=BCLAF1 SV=2	-	Exclusiva	
Q7Z6P3	Ras-related protein Rab-44 GN=RAB44 SV=4	-	Exclusiva	
O43866	CD5 antigen-like GN=CD5L SV=1	-	Exclusiva	
P01764	Immunoglobulin heavy variable 3-23 GN=IGHV3-23 SV=2	-	Exclusiva	
Q14258	E3 ubiquitin/ISG15 ligase TRIM25 GN=TRIM25 SV=2	-	Exclusiva	
Q9NUL5	Shiftless antiviral inhibitor of ribosomal frameshifting protein GN=SHFL SV=2	-	Exclusiva	
P01782	Immunoglobulin heavy variable 3-9 GN=IGHV3-9 SV=2	-	Exclusiva	
Q92686	Neurogranin GN=NRGN SV=1	-	Exclusiva	
Q96BZ4	5'-3' exonuclease PLD4 GN=PLD4 PE=2 SV=2	-	Exclusiva	
Q9H1E3	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1 GN=NUCKS1 SV=1	-	Exclusiva	
P26440	Isovaleryl-CoA dehydrogenase, mitochondrial GN=IVD SV=2	-	Exclusiva	
Q92843-2	Isoform 3 of Bcl-2-like protein 2 GN=BCL2L2	-	Exclusiva	
Q13595	Transformer-2 protein homolog alpha GN=TRA2A SV=1	-	Exclusiva	
P18510-3	Isoform 3 of Interleukin-1 receptor antagonist protein GN=IL1RN	-	Exclusiva	
P30049	ATP synthase subunit delta, mitochondrial GN=ATP5F1D SV=2	-	Exclusiva	

A6NMS3	Olfactory receptor 5K4 GN=OR5K4 PE=3 SV=1	-	Exclusiva
P17096	High mobility group protein HMG-I/HMG-Y GN=HMGA1 SV=3	-	Exclusiva
Q16658	Fascin GN=FSCN1 SV=3	-	Exclusiva
Q9UQN3	Charged multivesicular body protein 2b GN=CHMP2B SV=1	-	Exclusiva
Q9C0E8-4	Isoform 4 of Endoplasmic reticulum junction formation protein lunapark GN=LNPk	-	Exclusiva
P28907	ADP-ribosyl cyclase/cyclic ADP-ribose hydrolase 1 GN=CD38 SV=2	-	Exclusiva
Q8WVN8-3	Isoform 3 of Ubiquitin-conjugating enzyme E2 Q2 GN=UBE2Q2	-	Exclusiva
A5A3E0	POTE ankyrin domain family member F GN=POTEF SV=2	-	Exclusiva
Q96S19	Spermatid perinuclear RNA-binding protein GN=STRBP SV=1	-	Exclusiva
Q9Y315	Deoxyribose-phosphate aldolase GN=DERA SV=2	-	Exclusiva
P07602-3	Isoform Sap-mu-9 of Prosaposin GN=PSAP	-	Exclusiva
O43660	Pleiotropic regulator 1 GN=PLRG1 SV=1	-	Exclusiva
P84090	Enhancer of rudimentary homolog GN=ERH SV=1	-	Exclusiva
Q01459	Di-N-acetylchitobiase GN=CTBS SV=1	-	Exclusiva
O14556	Glyceraldehyde-3-phosphate dehydrogenase, testis-specific GN=GAPDHS SV=2	-	Exclusiva
Q9Y296	Trafficking protein particle complex subunit 4 GN=TRAPPC4 SV=1	-	Exclusiva
P04066	Tissue alpha-L-fucosidase GN=FUCA1 SV=4	-	Exclusiva
P03973	Antileukoprotease GN=SLPI SV=2	-	Exclusiva
P98161	Polycystin-1 GN=PKD1 SV=3	-	Exclusiva
Q2M3G0	ATP-binding cassette sub-family B member 5 GN=ABCB5 SV=4	-	Exclusiva
P01824	Immunoglobulin heavy variable 4-39 GN=IGHV4-39 SV=2	-	Exclusiva
P29966	Myristoylated alanine-rich C-kinase substrate GN=MARCKS SV=4	-	Exclusiva
P40199	Carcinoembryonic antigen-related cell adhesion molecule 6 GN=CEACAM6 SV=3	-	Exclusiva
P01593	Immunoglobulin kappa variable 1D-33 GN=IGKV1D-33 SV=2	-	Exclusiva
Q9NZV6	Methionine-R-sulfoxide reductase B1 GN=MSRB1 SV=3	-	Exclusiva
P01040	Cystatin-A GN=CSTA SV=1	-	Exclusiva
Q9P2Y5	UV radiation resistance-associated gene protein GN=UVRAG SV=1	-	Exclusiva
B9A064	Immunoglobulin lambda-like polypeptide 5 GN=IGLL5 PE=2 SV=2	-	Exclusiva
P51668	Ubiquitin-conjugating enzyme E2 D1 GN=UBE2D1 SV=1	-	Exclusiva
P46527	Cyclin-dependent kinase inhibitor 1B GN=CDKN1B SV=1	-	Exclusiva
P62888	60S ribosomal protein L30 GN=RPL30 SV=2	-	Exclusiva
Q9Y530	ADP-ribose glycohydrolase OARD1 GN=OARD1 SV=2	-	Exclusiva
Q16836-2	Isoform 2 of Hydroxyacyl-coenzyme A dehydrogenase, mitochondrial GN=HADH	-	Exclusiva
Q9HBW1	Leucine-rich repeat-containing protein 4 GN=LRRC4 SV=2	-	Exclusiva
P62879	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2 GN=GNB2 SV=3	-	Exclusiva
P23141-2	Isoform 2 of Liver carboxylesterase 1 GN=CES1	-	Exclusiva

P07311	Acylphosphatase-1 GN=ACYP1 SV=2	-	Exclusiva
P00747	Plasminogen GN=PLG SV=2	-	Exclusiva
P23527	Histone H2B type 1-O GN=H2BC17 SV=3	-	Exclusiva
P00734	Prothrombin GN=F2 SV=2	-	Exclusiva
P02790	Hemopexin GN=HPX SV=2	-	Exclusiva
Q03591	Complement factor H-related protein 1 GN=CFHR1 SV=2	-	Exclusiva
P00736	Complement C1r subcomponent GN=C1R SV=2	-	Exclusiva
P20742	Pregnancy zone protein GN=PZP SV=4	-	Exclusiva
P01019	Angiotensinogen GN=AGT SV=1	-	Exclusiva
Q9BPX5	Actin-related protein 2/3 complex subunit 5-like protein GN=ARPC5L SV=1	-	Exclusiva
Q16773	Kynurenine--oxoglutarate transaminase 1 GN=KYAT1 SV=1	-	Exclusiva
Q8TEQ6	Gem-associated protein 5 GN=GEMIN5 SV=3	-	Exclusiva
AOA0B4J2H0	Immunoglobulin heavy variable 1-69D GN=IGHV1-69D SV=1	-	Exclusiva
P02792	Ferritin light chain GN=FTL SV=2	-	Exclusiva
Q8IWT0	Protein archease GN=ZBTB8OS SV=2	-	Exclusiva
P35030	Trypsin-3 GN=PRSS3 SV=2	-	Exclusiva
O43598	2'-deoxynucleoside 5'-phosphate N-hydrolase 1 GN=DNPH1 SV=1	-	Exclusiva
P05546	Heparin cofactor 2 GN=SERPIND1 SV=3	-	Exclusiva
Q92506	(3R)-3-hydroxyacyl-CoA dehydrogenase GN=HSD17B8 SV=2	-	Exclusiva
O43451	Maltase-glucoamylase, intestinal GN=MGAM SV=5	-	Exclusiva
P21980	Protein-glutamine gamma-glutamyltransferase 2 GN=TGM2 SV=2	-	Exclusiva
P04839	Cytochrome b-245 heavy chain GN=CYBB SV=2	-	Exclusiva
Q14520	Hyaluronan-binding protein 2 GN=HABP2 SV=1	-	Exclusiva
Q8IW45-2	Isoform 2 of ATP-dependent (S)-NAD(P)H-hydrate dehydratase GN=NAXD	-	Exclusiva
Q9BZE9-2	Isoform 2 of Tether containing UBX domain for GLUT4 GN=ASPSCR1	-	Exclusiva
Q9Y6H1	Coiled-coil-helix-coiled-coil-helix domain-containing protein 2 GN=CHCHD2 SV=1	-	Exclusiva
P50876	E3 ubiquitin-protein ligase RNF144A GN=RNF144A SV=2	-	Exclusiva
Q96B97-2	Isoform 2 of SH3 domain-containing kinase-binding protein 1 GN=SH3KBP1	-	Exclusiva
Q9UBV8	Peflin GN=PEF1 SV=1	-	Exclusiva
P20290	Transcription factor BTF3 GN=BTF3 SV=1	-	Exclusiva
P51124	Granzyme M GN=GZMM SV=2	-	Exclusiva
Q16774-3	Isoform 3 of Guanylate kinase GN=GUK1	-	Exclusiva
Q9NVE7	4'-phosphopantetheine phosphatase GN=PANK4 SV=1	-	Exclusiva
O95372	Acyl-protein thioesterase 2 GN=LYPLA2 SV=1	-	Exclusiva
Q96K17	Transcription factor BTF3 homolog 4 GN=BTF3L4 SV=1	-	Exclusiva
Q6K0P9	Pyrin and HIN domain-containing protein 1 GN=PYHIN1 SV=1	-	Exclusiva
Q86WQ0-2	Isoform 2 of Nuclear receptor 2C2-associated protein GN=NR2C2AP	-	Exclusiva

Q96BW5	Phosphotriesterase-related protein	GN=PTER SV=1	-	Exclusiva
P20701-2	Isoform 2 of Integrin alpha-L	GN=ITGAL	-	Exclusiva
P51688	N-sulphoglucosamine sulphohydrolase	GN=SGSH SV=1	-	Exclusiva
Q96EY5	Multivesicular body subunit 12A	GN=MVB12A SV=1	-	Exclusiva
Q99471	Prefoldin subunit 5	GN=PFDN5 SV=2	-	Exclusiva
Q96PP8	Guanylate-binding protein 5	GN=GBP5 SV=1	-	Exclusiva
Q9Y2S2	Lambda-crystallin homolog	GN=CRYL1 SV=3	-	Exclusiva
Q9UL26	Ras-related protein Rab-22A	GN=RAB22A SV=2	-	Exclusiva
P62841	40S ribosomal protein S15	GN=RPS15 SV=2	-	Exclusiva
Q9BXR0	Queuine tRNA-ribosyltransferase catalytic subunit 1	GN=QTRT1 SV=3	-	Exclusiva
P61289-2	Isoform 2 of Proteasome activator complex subunit 3	GN=PSME3	-	Exclusiva
Q96PE3	Inositol polyphosphate-4-phosphatase type I A	GN=INPP4A SV=1	-	Exclusiva
P09326	CD48 antigen	GN=CD48 SV=2	-	Exclusiva
O75884	Serine hydrolase RBBP9	GN=RBBP9 SV=2	-	Exclusiva
P55327-3	Isoform 3 of Tumor protein D52	GN=TPD52	-	Exclusiva
Q96J01	THO complex subunit 3	GN=THOC3 SV=1	-	Exclusiva
Q8WX92-2	Isoform 2 of Negative elongation factor B	GN=NELFB	-	Exclusiva
Q9Y211	Nischarin	GN=NISCH SV=3	-	Exclusiva
Q14166	Tubulin--tyrosine ligase-like protein 12	GN=TLL12 SV=2	-	Exclusiva
P20591	Interferon-induced GTP-binding protein Mx1	GN=MX1 SV=4	-	Exclusiva
P04003	C4b-binding protein alpha chain	GN=C4BPA SV=2	-	Exclusiva
P49354	Protein farnesyltransferase/geranylgeranyltransferase type-1 subunit alpha	GN=FNTA SV=1	-	Exclusiva
P02585	Troponin C, skeletal muscle	GN=TNNC2 SV=2	-	Exclusiva
Q13496	Myotubularin	GN=MTM1 SV=2	-	Exclusiva
O43747-2	Isoform 2 of AP-1 complex subunit gamma-1	GN=AP1G1	-	Exclusiva
P30048	Thioredoxin-dependent peroxide reductase, mitochondrial	GN=PRDX3 SV=3	-	Exclusiva
Q6PCE3	Glucose 1,6-bisphosphate synthase	GN=PGM2L1 SV=3	-	Exclusiva
Q8IVE3	Pleckstrin homology domain-containing family H member 2	GN=PLEKHH2 SV=2	-	Exclusiva
P49643	DNA primase large subunit	GN=PRIM2 SV=2	-	Exclusiva
Q5VWW2	GTPase-activating Rap/Ran-GAP domain-like protein 3	GN=GARNL3 PE=2 SV=2	-	Exclusiva
Q14980	Nuclear mitotic apparatus protein 1	GN=NUMA1 SV=2	-	Exclusiva
P34897	Serine hydroxymethyltransferase, mitochondrial	GN=SHMT2 SV=3	-	Exclusiva
Q96LW4	DNA-directed primase/polymerase protein	GN=PRIMPOL SV=3	-	Exclusiva
P43652	Afamin	GN=AFM SV=1	-	Exclusiva
O43290	U4/U6.U5 tri-snRNP-associated protein 1	GN=SART1 SV=1	-	Exclusiva
Q93073	Selenocysteine insertion sequence-binding protein 2-like	GN=SECISBP2L SV=3	-	Exclusiva

Q7Z7G8	Vacuolar protein sorting-associated protein 13B SV=2	GN=VPS13B	-	Exclusiva
O75264	Small integral membrane protein 24	GN=SMIM24 PE=2 SV=2	-	Exclusiva
Q9H3H3-3	Isoform 3 of UPF0696 protein C11orf68	GN=C11orf68	-	Exclusiva
Q8WUD4	Coiled-coil domain-containing protein 12	GN=CCDC12 SV=1	-	Exclusiva
P05186	Alkaline phosphatase, tissue-nonspecific isozyme	GN=ALPL SV=4	-	Exclusiva
P68431	Histone H3.1	GN=H3C1 SV=2	3,61	1,78
Q9BRF8	Serine/threonine-protein phosphatase	CPPED1 GN=CPPED1 SV=3	3,58	2,07
P0CG39	POTE ankyrin domain family member J	GN=POTEJ PE=3 SV=1	3,48	1,73
P13727	Bone marrow proteoglycan	GN=PRG2 SV=2	2,98	1,47
Q99417	c-Myc-binding protein	GN=MYCBP SV=3	2,87	1,41
O95466-3	Isoform 3 of Formin-like protein 1	GN=FMNL1	2,86	1,45
Q9NTM9	Copper homeostasis protein cutC homolog	GN=CUTC SV=1	2,74	2,24
P01859	Immunoglobulin heavy constant gamma 2	GN=IGHG2 SV=2	2,69	2,24
P12955	Xaa-Pro dipeptidase	GN=PEPD SV=3	2,67	2,37
P01860	Immunoglobulin heavy constant gamma 3	GN=IGHG3 SV=2	2,64	1,35
Q08623-4	Isoform 4 of Pseudouridine-5'-phosphatase	GN=PUDP	2,50	2,06
Q01085-2	Isoform 2 of Nucleolysin TIAR	GN=TIAL1	2,05	1,86
P04632	Calpain small subunit 1	GN=CAPNS1 SV=1	1,98	2,01
P04004	Vitronectin	GN=VTN SV=1	1,98	1,38
P40306	Proteasome subunit beta type-10	GN=PSMB10 SV=1	1,93	1,59
O00757	Fructose-1,6-bisphosphatase isozyme 2	GN=FBP2 SV=2	1,84	1,43
P18887	DNA repair protein XRCC1	GN=XRCC1 SV=2	1,77	2,26
Q15637-5	Isoform 5 of Splicing factor 1	GN=SF1	1,56	1,92
Q969H8	Myeloid-derived growth factor	GN=MYDGF SV=1	1,55	1,39
Q9NZT2	Opioid growth factor receptor	GN=OGFR SV=3	1,53	1,41
O00160	Unconventional myosin-I α	GN=MYO1F SV=3	1,48	1,30
Q9H008	Phospholysine phosphohistidine inorganic pyrophosphate phosphatase	GN=LHPP SV=2	1,45	1,95
Q96GX9	Methylthioribulose-1-phosphate dehydratase	GN=APIP SV=1	1,40	1,87
P23588	Eukaryotic translation initiation factor 4B	GN=EIF4B SV=2	1,38	1,99
Q6IA69	Glutamine-dependent NAD(+) synthetase	GN=NADSYN1 SV=3	1,36	1,88
Q15185	Prostaglandin E synthase 3	GN=PTGES3 SV=1	1,36	1,69
P16930	Fumarylacetoacetase	GN=FAH SV=2	1,30	1,47
Q96C23	Galactose mutarotase	GN=GALM SV=1	1,01	1,69
Proteínas reguladas negativamente				
P41439	Folate receptor gamma	GN=FOLR3 SV=2	-	Exclusiva
P07947	Tyrosine-protein kinase Yes	GN=YES1 SV=3	-	Exclusiva
Q13310-3	Isoform 3 of Polyadenylate-binding protein 4	GN=PABPC4	-	Exclusiva
Q9NY65	Tubulin alpha-8 chain	GN=TUBA8 SV=1	-	Exclusiva
O75688	Protein phosphatase 1B	GN=PPM1B SV=1	-	Exclusiva
Q06187-2	Isoform BTK-C of Tyrosine-protein kinase BTK	GN=BTK	-	Exclusiva

Q9Y608-4	Isoform 4 of Leucine-rich repeat flightless-interacting protein 2 GN=LRRFIP2	-	Exclusiva
Q7L576	Cytoplasmic FMR1-interacting protein 1 GN=CYFIP1 SV=1	-	Exclusiva
P50851-2	Isoform 2 of Lipopolysaccharide-responsive and beige-like anchor protein GN=LRBA	-	Exclusiva
Q9NRW1	Ras-related protein Rab-6B GN=RAB6B SV=1	-	Exclusiva
P57772	Selenocysteine-specific elongation factor GN=EEFSEC SV=4	-	Exclusiva
O14672	Disintegrin and metalloproteinase domain-containing protein 10 GN=ADAM10 SV=1	-	Exclusiva
Q9NWW8	BRISC and BRCA1-A complex member 1 GN=BABAM1 SV=1	-	Exclusiva
P05388	60S acidic ribosomal protein P0 GN=RPLP0 SV=1	-	Exclusiva
Q9C0I1	Myotubularin-related protein 12 GN=MTMR12 SV=2	-	Exclusiva
P41227	N-alpha-acetyltransferase 10 GN=NAA10 SV=1	-	Exclusiva
Q9Y6Y8	SEC23-interacting protein GN=SEC23IP SV=1	-	Exclusiva
Q9UNE7	E3 ubiquitin-protein ligase CHIP GN=STUB1 SV=2	-	Exclusiva
Q9C0C9	(E3-independent) E2 ubiquitin-conjugating enzyme GN=UBE2O SV=3	-	Exclusiva
Q9P107	GEM-interacting protein GN=GMIP SV=2	-	Exclusiva
Q15642	Cdc42-interacting protein 4 GN=TRIP10 SV=3	-	Exclusiva
Q5T447	E3 ubiquitin-protein ligase HECTD3 GN=HECTD3 SV=1	-	Exclusiva
Q9UPN3	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5 GN=MACF1 SV=4	-	Exclusiva
Q6ZVM7	TOM1-like protein 2 GN=TOM1L2 SV=1	-	Exclusiva
Q14112	Nidogen-2 GN=NID2 SV=3	-	Exclusiva
Q12874	Splicing factor 3A subunit 3 GN=SF3A3 SV=1	-	Exclusiva
Q9BW04	Specifically androgen-regulated gene protein GN=SARG SV=2	-	Exclusiva
Q6FI81	Anamorsin GN=CIAPIN1 SV=2	-	Exclusiva
Q9Y570-4	Isoform 4 of Protein phosphatase methylesterase 1 GN=PPME1	-	Exclusiva
O00193	Small acidic protein GN=SMAP SV=1	-	Exclusiva
Q9BZL4	Protein phosphatase 1 regulatory subunit 12C GN=PPP1R12C SV=1	-	Exclusiva
Q9BZZ5	Apoptosis inhibitor 5 GN=API5 SV=3	-	Exclusiva
O43633	Charged multivesicular body protein 2a GN=CHMP2A SV=1	-	Exclusiva
Q9UI12	V-type proton ATPase subunit H GN=ATP6V1H SV=1	-	Exclusiva
Q86YP4-3	Isoform 3 of Transcriptional repressor p66-alpha GN=GATAD2A	-	Exclusiva
Q13263	Transcription intermediary factor 1-beta GN=TRIM28 SV=5	-	Exclusiva
C9JLW8	Mapk-regulated corepressor-interacting protein 1 GN=MCRIP1 SV=1	-	Exclusiva
P32189	Glycerol kinase GN=GK SV=3	-	Exclusiva
Q05048	Cleavage stimulation factor subunit 1 GN=CSTF1 SV=1	-	Exclusiva
Q8TDX7	Serine/threonine-protein kinase Nek7 GN=NEK7 SV=1	-	Exclusiva
Q9UBE0	SUMO-activating enzyme subunit 1 GN=SAE1 SV=1	-	Exclusiva
P62195	26S proteasome regulatory subunit 8 GN=PSMC5 SV=1	-	Exclusiva
P46060	Ran GTPase-activating protein 1 GN=RANGAP1 SV=1	-	Exclusiva
Q9BX10	GTP-binding protein 2 GN=GTPBP2 SV=1	-	Exclusiva

Q6YHK3	CD109 antigen GN=CD109 SV=2	-	Exclusiva
Q9H6T3	RNA polymerase II-associated protein 3 GN=RPAP3 SV=2	-	Exclusiva
Q9NXV6	CDKN2A-interacting protein GN=CDKN2AIP SV=3	-	Exclusiva
Q9BX68	Histidine triad nucleotide-binding protein 2, mitochondrial GN=HINT2 SV=1	-	Exclusiva
Q15029	116 kDa U5 small nuclear ribonucleoprotein component GN=EFTUD2 SV=1	-	Exclusiva
Q9Y230	RuvB-like 2 GN=RUVBL2 SV=3	-	Exclusiva
Q6UX04	Spliceosome-associated protein CWC27 homolog GN=CWC27 SV=1	-	Exclusiva
Q15418-2	Isoform 2 of Ribosomal protein S6 kinase alpha-1 GN=RPS6KA1	-	Exclusiva
Q13492	Phosphatidylinositol-binding clathrin assembly protein GN=PICALM SV=2	-	Exclusiva
Q07020	60S ribosomal protein L18 GN=RPL18 SV=2	-	Exclusiva
Q32P44	Echinoderm microtubule-associated protein-like 3 GN=EML3 SV=1	-	Exclusiva
P61962	DDB1- and CUL4-associated factor 7 GN=DCAF7 SV=1	-	Exclusiva
Q14C86-6	Isoform 6 of GTPase-activating protein and VPS9 domain-containing protein 1 GN=GAPVD1	-	Exclusiva
O43852-3	Isoform 3 of Calumenin GN=CALU	-	Exclusiva
Q9GZP0	Platelet-derived growth factor D GN=PDGFD SV=1	-	Exclusiva
P40429	60S ribosomal protein L13a GN=RPL13A SV=2	-	Exclusiva
O00178	GTP-binding protein 1 GN=GTPBP1 SV=3	-	Exclusiva
Q9NRF8	CTP synthase 2 GN=CTPS2 SV=1	-	Exclusiva
Q86X27	Ras-specific guanine nucleotide-releasing factor RalGPS2 GN=RALGPS2 SV=1	-	Exclusiva
P05121	Plasminogen activator inhibitor 1 GN=SERPINE1 SV=1	-	Exclusiva
Q7LDG7-2	Isoform 2 of RAS guanyl-releasing protein 2 GN=RASGRP2	-	Exclusiva
Q9Y6K1	DNA (cytosine-5)-methyltransferase 3A GN=DNMT3A SV=4	-	Exclusiva
O95248-4	Isoform 4 of Myotubularin-related protein 5 GN=SBF1	-	Exclusiva
Q8TC07	TBC1 domain family member 15 GN=TBC1D15 SV=2	-	Exclusiva
P52948	Nuclear pore complex protein Nup98-Nup96 GN=NUP98 SV=4	-	Exclusiva
Q8IVB4	Sodium/hydrogen exchanger 9 GN=SLC9A9 SV=1	-	Exclusiva
Q12851	Mitogen-activated protein kinase kinase kinase kinase 2 GN=MAP4K2 SV=2	-	Exclusiva
Q9UBP0	Spastin GN=SPAST SV=1	-	Exclusiva
P19086	Guanine nucleotide-binding protein G(z) subunit alpha GN=GNAZ SV=3	-	Exclusiva
Q9NR19-2	Isoform 2 of Acetyl-coenzyme A synthetase, cytoplasmic GN=ACSS2	-	Exclusiva
Q96B26	Exosome complex component RRP43 GN=EXOSC8 SV=1	-	Exclusiva
Q9BUR4	Telomerase Cajal body protein 1 GN=WRAP53 SV=1	-	Exclusiva
Q9H7J1	Protein phosphatase 1 regulatory subunit 3E GN=PPP1R3E SV=2	-	Exclusiva
P55036	26S proteasome non-ATPase regulatory subunit 4 GN=PSMD4 SV=1	-	Exclusiva

Q99747	Gamma-soluble NSF attachment protein GN=NAPG SV=1	-	Exclusiva
Q7RTV0	PHD finger-like domain-containing protein 5A GN=PHF5A SV=1	-	Exclusiva
Q8N1G2	Cap-specific mRNA (nucleoside-2'-O-)-methyltransferase 1 GN=CMTR1 SV=1	-	Exclusiva
Q15836	Vesicle-associated membrane protein 3 GN=VAMP3 SV=3	-	Exclusiva
O95793	Double-stranded RNA-binding protein Staufen homolog 1 GN=STAU1 SV=2	-	Exclusiva
Q6GMV3	Putative peptidyl-tRNA hydrolase PTRHD1 GN=PTRHD1 SV=1	-	Exclusiva
Q8NHV4-3	Isoform 3 of Protein NEDD1 GN=NEDD1	-	Exclusiva
Q9UGP4	LIM domain-containing protein 1 GN=LIMD1 SV=1	-	Exclusiva
Q9Y385	Ubiquitin-conjugating enzyme E2 J1 GN=UBE2J1 SV=2	-	Exclusiva
Q8IYB7	DIS3-like exonuclease 2 GN=DIS3L2 SV=4	-	Exclusiva
Q96FX7	tRNA (adenine(58)-N(1))-methyltransferase catalytic subunit TRMT61A GN=TRMT61A SV=1	-	Exclusiva
P61966	AP-1 complex subunit sigma-1A GN=AP1S1 SV=1	-	Exclusiva
Q8IYB5	Stromal membrane-associated protein 1 GN=SMAP1 SV=2	-	Exclusiva
Q9UJA5	tRNA (adenine(58)-N(1))-methyltransferase non-catalytic subunit TRM6 GN=TRMT6 SV=1	-	Exclusiva
P19388	DNA-directed RNA polymerases I, II, and III subunit RPABC1 GN=POLR2E SV=4	-	Exclusiva
Q96EM0	Trans-3-hydroxy-L-proline dehydratase GN=L3HYPDH SV=2	-	Exclusiva
Q8WXE9-3	Isoform 2 of Stonin-2 GN=STON2	-	Exclusiva
Q92572	AP-3 complex subunit sigma-1 GN=AP3S1 SV=1	-	Exclusiva
Q8N3F0	Maturin GN=MTURN SV=2	-	Exclusiva
O95866-2	Isoform A of Megakaryocyte and platelet inhibitory receptor G6b GN=MPIG6B	-	Exclusiva
P49789	Bis(5'-adenosyl)-triphosphatase GN=FHIT SV=3	-	Exclusiva
Q9H5N1	Rab GTPase-binding effector protein 2 GN=RABEP2 SV=2	-	Exclusiva
Q92796	Disks large homolog 3 GN=DLG3 SV=2	-	Exclusiva
P53814-6	Isoform B3 of Smoothelin GN=SMTN	-	Exclusiva
Q8N8N7	Prostaglandin reductase 2 GN=PTGR2 SV=1	-	Exclusiva
O15198	Mothers against decapentaplegic homolog 9 GN=SMAD9 SV=1	-	Exclusiva
Q9Y3X0	Coiled-coil domain-containing protein 9 GN=CCDC9 SV=1	-	Exclusiva
Q13427	Peptidyl-prolyl cis-trans isomerase G GN=PPIG SV=2	-	Exclusiva
P18440	Arylamine N-acetyltransferase 1 GN=NAT1 SV=2	-	Exclusiva
Q15427	Splicing factor 3B subunit 4 GN=SF3B4 SV=1	-	Exclusiva
Q7Z6Z7	E3 ubiquitin-protein ligase HUWE1 GN=HUWE1 SV=3	-	Exclusiva
O14964	Hepatocyte growth factor-regulated tyrosine kinase substrate GN=HGS SV=1	-	Exclusiva
Q9Y6K9-2	Isoform 2 of NF-kappa-B essential modulator GN=IKBKG	-	Exclusiva
P36915	Guanine nucleotide-binding protein-like 1 GN=GNL1 SV=2	-	Exclusiva
P62910	60S ribosomal protein L32 GN=RPL32 SV=2	-	Exclusiva
O14908	PDZ domain-containing protein GIPC1 GN=GIPC1 SV=2	-	Exclusiva
Q969M7	NEDD8-conjugating enzyme UBE2F GN=UBE2F SV=1	-	Exclusiva
Q3MHD2-2	Isoform 2 of Protein LSM12 homolog GN=LSM12	-	Exclusiva

Q03111	Protein ENL GN=MLLT1 SV=2	-	Exclusiva
Q9BRJ6	Uncharacterized protein C7orf50 GN=C7orf50 SV=1	-	Exclusiva
Q8N9U0	Tandem C2 domains nuclear protein GN=TC2N SV=2	-	Exclusiva
Q5T0N5	Formin-binding protein 1-like GN=FNBP1L SV=3	-	Exclusiva
P62829	60S ribosomal protein L23 GN=RPL23 SV=1	-	Exclusiva
Q9UNP9-3	Isoform 3 of Peptidyl-prolyl cis-trans isomerase E GN=PPIE	-	Exclusiva
Q15287	RNA-binding protein with serine-rich domain 1 GN=RNPS1 SV=1	-	Exclusiva
Q9NWX4	Histone PARylation factor 1 GN=HPF1 SV=2	-	Exclusiva
Q9NY27	Serine/threonine-protein phosphatase 4 regulatory subunit 2 GN=PPP4R2 SV=3	-	Exclusiva
O15160	DNA-directed RNA polymerases I and III subunit RPAC1 GN=POLR1C SV=1	-	Exclusiva
Q96EK6	Glucosamine 6-phosphate N-acetyltransferase GN=GPNAT1 SV=1	-	Exclusiva
Q9H267	Vacuolar protein sorting-associated protein 33B GN=VPS33B SV=2	-	Exclusiva
P23919	Thymidylate kinase GN=DTYMK SV=4	-	Exclusiva
Q6ZSR9	Uncharacterized protein FLJ45252 PE=2 SV=2	-	Exclusiva
Q16512-2	Isoform 2 of Serine/threonine-protein kinase N1 GN=PKN1	-	Exclusiva
Q92963-3	Isoform 3 of GTP-binding protein Rit1 GN=RIT1	-	Exclusiva
P20036	HLA class II histocompatibility antigen, DP alpha 1 chain GN=HLA-DPA1 SV=1	-	Exclusiva
Q8IV48	3'-5' exoribonuclease 1 GN=ERI1 SV=3	-	Exclusiva
O60884	DnaJ homolog subfamily A member 2 GN=DNAJA2 SV=1	-	Exclusiva
Q14116	Interleukin-18 GN=IL18 SV=1	-	Exclusiva
Q13459	Unconventional myosin-IXb GN=MYO9B SV=3	-	Exclusiva
Q9BZF9	Uveal autoantigen with coiled-coil domains and ankyrin repeats GN=UACA SV=2	-	Exclusiva
Q14353	Guanidinoacetate N-methyltransferase GN=GAMT SV=1	-	Exclusiva
Q9HB19	Pleckstrin homology domain-containing family A member 2 GN=PLEKHA2 SV=2	-	Exclusiva
Q9BSD7	Cancer-related nucleoside-triphosphatase GN=NTPCR SV=1	-	Exclusiva
O43567	E3 ubiquitin-protein ligase RNF13 GN=RNF13 SV=1	-	Exclusiva
Q9NPE3	H/ACA ribonucleoprotein complex subunit 3 GN=NOP10 SV=1	-	Exclusiva
Q9BV57	1,2-dihydroxy-3-keto-5-methylthiopentene dioxygenase GN=ADI1 SV=1	-	Exclusiva
Q6PGP7	Tetratricopeptide repeat protein 37 GN=TTC37 SV=1	-	Exclusiva
P45985-2	Isoform 2 of Dual specificity mitogen-activated protein kinase kinase 4 GN=MAP2K4	-	Exclusiva
Q9NQC7	Ubiquitin carboxyl-terminal hydrolase CYLD GN=CYLD SV=1	-	Exclusiva
Q06787	Synaptic functional regulator FMR1 GN=FMR1 SV=1	-	Exclusiva
Q9NXH9	tRNA (guanine(26)-N(2))-dimethyltransferase GN=TRMT1 SV=1	-	Exclusiva
Q86S22	Trafficking protein particle complex subunit 6B GN=TRAPPC6B SV=1	-	Exclusiva

Q9NPQ8	Synembryn-A GN=RIC8A SV=3	-	Exclusiva
Q9Y487	V-type proton ATPase 116 kDa subunit a2 GN=ATP6V0A2 SV=2	-	Exclusiva
Q92974	Rho guanine nucleotide exchange factor 2 GN=ARHGEF2 SV=4	-	Exclusiva
Q9HAU5	Regulator of nonsense transcripts 2 GN=UPF2 SV=1	-	Exclusiva
Q15404-2	Isoform 2 of Ras suppressor protein 1 GN=RSU1	-	Exclusiva
P08519	Apolipoprotein(a) GN=LPA SV=1	-	Exclusiva
P00746	Complement factor D GN=CFD SV=5	-	Exclusiva
P57764	Gasdermin-D GN=GSDMD SV=1	-	Exclusiva
Q969Q5	Ras-related protein Rab-24 GN=RAB24 SV=1	-	Exclusiva
P35658-5	Isoform 5 of Nuclear pore complex protein Nup214 GN=NUP214	-	Exclusiva
Q96C11-3	Isoform 4 of FGGY carbohydrate kinase domain-containing protein GN=FGGY	-	Exclusiva
O00629	Importin subunit alpha-3 GN=KPNA4 SV=1	-	Exclusiva
Q96B45	BLOC-1-related complex subunit 7 GN=BORCS7 PE=3 SV=2	-	Exclusiva
Q08043	Alpha-actinin-3 GN=ACTN3 SV=2	-	Exclusiva
Q9H1P3	Oxysterol-binding protein-related protein 2 GN=OSBPL2 SV=1	-	Exclusiva
Q9H074	Polyadenylate-binding protein-interacting protein 1 GN=PAIP1 SV=1	-	Exclusiva
P20340	Ras-related protein Rab-6A GN=RAB6A SV=3	-	Exclusiva
P0DJ18	Serum amyloid A-1 protein GN=SAA1 SV=1	-	Exclusiva
Q14444	Caprin-1 GN=CAPRIN1 SV=2	-	Exclusiva
Q9H3P7	Golgi resident protein GCP60 GN=ACBD3 SV=4	-	Exclusiva
P62917	60S ribosomal protein L8 GN=RPL8 SV=2	-	Exclusiva
P23921	Ribonucleoside-diphosphate reductase large subunit GN=RRM1 SV=1	-	Exclusiva
P55273	Cyclin-dependent kinase 4 inhibitor D GN=CDKN2D SV=1	-	Exclusiva
Q9UBF2	Coatomer subunit gamma-2 GN=COPG2 SV=1	-	Exclusiva
P52788	Spermine synthase GN=SMS SV=2	-	Exclusiva
P52292	Importin subunit alpha-1 GN=KPNA2 SV=1	-	Exclusiva
Q96JH7	Deubiquitinating protein VCIPI1 GN=VCIPI1 SV=2	-	Exclusiva
P51114	Fragile X mental retardation syndrome-related protein 1 GN=FXR1 SV=3	-	Exclusiva
Q99816	Tumor susceptibility gene 101 protein GN=TSG101 SV=2	-	Exclusiva
Q8N6T3-2	Isoform 2 of ADP-ribosylation factor GTPase-activating protein 1 GN=ARFGAP1	-	Exclusiva
Q9NZQ3	NCK-interacting protein with SH3 domain GN=NCKIPSD SV=1	-	Exclusiva
P54619-3	Isoform 3 of 5'-AMP-activated protein kinase subunit gamma-1 GN=PRKAG1	-	Exclusiva
Q9P1Z2	Calcium-binding and coiled-coil domain-containing protein 1 GN=CALCOCO1 SV=2	-	Exclusiva
P53618	Coatomer subunit beta GN=COPB1 SV=3	-	Exclusiva
Q15003	Condensin complex subunit 2 GN=NCAPH SV=3	-	Exclusiva
Q00341	Vigilin GN=HDLBP SV=2	-	Exclusiva

Q92841	Probable ATP-dependent RNA helicase DDX17 SV=2	GN=DDX17	-1,01	1,46
Q96G03	Phosphoglucomutase-2	GN=PGM2 SV=4	-1,07	1,53
Q13303-2	Isoform 2 of Voltage-gated potassium channel subunit beta-2 GN=KCNAB2		-1,08	1,46
Q01970	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-3	GN=PLCB3 SV=2	-1,10	1,46
P61978-2	Isoform 2 of Heterogeneous nuclear ribonucleoprotein K GN=HNRNPK		-1,11	1,70
P23526	Adenosylhomocysteinase	GN=AHCY SV=4	-1,11	1,46
Q07955-2	Isoform ASF-2 of Serine/arginine-rich splicing factor 1 GN=SRSF1		-1,16	1,31
Q16555	Dihydropyrimidinase-related protein 2	GN=DPYSL2 SV=1	-1,16	1,49
P31153	S-adenosylmethionine synthase isoform type-2 SV=1	GN=MAT2A	-1,17	1,66
P13798	Acylamino-acid-releasing enzyme	GN=APEH SV=4	-1,17	1,39
P0DMV9	Heat shock 70 kDa protein 1B	GN=HSPA1B SV=1	-1,18	1,51
P00338-3	Isoform 3 of L-lactate dehydrogenase A chain	GN=LDHA	-1,19	1,30
P22314	Ubiquitin-like modifier-activating enzyme 1	GN=UBA1 SV=3	-1,20	2,01
Q9BWD1	Acetyl-CoA acetyltransferase, cytosolic	GN=ACAT2 SV=2	-1,23	1,80
P55884-2	Isoform 2 of Eukaryotic translation initiation factor 3 subunit B GN=EIF3B		-1,23	1,38
P50395	Rab GDP dissociation inhibitor beta	GN=GDI2 SV=2	-1,23	1,43
Q9NPP4	NLR family CARD domain-containing protein 4	GN=NLRC4 SV=2	-1,23	1,47
P07195	L-lactate dehydrogenase B chain	GN=LDHB SV=2	-1,25	1,86
Q9Y2J8	Protein-arginine deiminase type-2	GN=PADI2 SV=2	-1,25	1,72
O43396	Thioredoxin-like protein 1	GN=TXNL1 SV=3	-1,26	1,88
P61224	Ras-related protein Rap-1b	GN=RAP1B SV=1	-1,26	1,49
P13489	Ribonuclease inhibitor	GN=RNH1 SV=2	-1,27	1,36
P42331-4	Isoform 4 of Rho GTPase-activating protein 25	GN=ARHGAP25	-1,28	1,64
O15117-3	Isoform 3 of FYN-binding protein 1	GN=FYB1	-1,28	1,36
Q9NZL9	Methionine adenosyltransferase 2 subunit beta SV=1	GN=MAT2B	-1,28	2,11
P07237	Protein disulfide-isomerase	GN=P4HB SV=3	-1,29	2,35
O43148-2	Isoform 2 of mRNA cap guanine-N7 methyltransferase GN=RNMT		-1,32	2,18
P05067	Amyloid-beta precursor protein	GN=APP SV=3	-1,32	1,41
Q06830	Peroxiredoxin-1	GN=PRDX1 SV=1	-1,34	2,03
P48739-3	Isoform 3 of Phosphatidylinositol transfer protein beta isoform GN=PITPNB		-1,34	1,91
Q8NBS9	Thioredoxin domain-containing protein 5	GN=TXNDC5 SV=2	-1,36	1,52
Q08945	FACT complex subunit SSRP1	GN=SSRP1 SV=1	-1,36	1,42
P48595	Serpin B10	GN=SERPINB10 SV=1	-1,36	1,53
P17655	Calpain-2 catalytic subunit	GN=CAPN2 SV=6	-1,37	1,44
P29144	Tripeptidyl-peptidase 2	GN=TPP2 SV=4	-1,37	4,05
Q01469	Fatty acid-binding protein 5	GN=FABP5 SV=3	-1,37	1,42

P40925-3	Isoform 3 of Malate dehydrogenase, cytoplasmic	GN=MDH1	-1,38	1,69
P29692-2	Isoform 2 of Elongation factor 1-delta	GN=EEF1D	-1,39	1,52
P61106	Ras-related protein Rab-14	GN=RAB14 SV=4	-1,39	1,55
Q06323	Proteasome activator complex subunit 1	GN=PSME1 SV=1	-1,39	1,51
P11021	Endoplasmic reticulum chaperone BiP	GN=HSPA5 SV=2	-1,40	2,41
O75832	26S proteasome non-ATPase regulatory subunit 10	GN=PSMD10 SV=1	-1,40	1,56
Q92609-2	Isoform 2 of TBC1 domain family member 5	GN=TBC1D5	-1,40	1,52
O75874	Isocitrate dehydrogenase [NADP] cytoplasmic	GN=IDH1 SV=2	-1,40	3,12
P33176	Kinesin-1 heavy chain	GN=KIF5B SV=1	-1,42	1,49
O60547	GDP-mannose 4,6 dehydratase	GN=GMDS SV=1	-1,42	2,48
P61086	Ubiquitin-conjugating enzyme E2 K	GN=UBE2K SV=3	-1,43	1,60
P27797	Calreticulin	GN=CALR SV=1	-1,43	2,48
P50995	Annexin A11	GN=ANXA11 SV=1	-1,44	1,81
P23284	Peptidyl-prolyl cis-trans isomerase B	GN=PPIB SV=2	-1,44	1,56
P00492	Hypoxanthine-guanine phosphoribosyltransferase	GN=HPRT1 SV=2	-1,44	1,46
Q7KZF4	Staphylococcal nuclease domain-containing protein 1	GN=SND1 SV=1	-1,45	2,18
P14324-2	Isoform 2 of Farnesyl pyrophosphate synthase	GN=FDPS	-1,46	2,06
Q13043	Serine/threonine-protein kinase 4	GN=STK4 SV=2	-1,46	1,39
P28838	Cytosol aminopeptidase	GN=LAP3 SV=3	-1,47	1,63
O60256	Phosphoribosyl pyrophosphate synthase-associated protein 2	GN=PRPSAP2 SV=1	-1,47	1,60
Q9UNZ2-5	Isoform 3 of NSFL1 cofactor p47	GN=NSFL1C	-1,48	3,17
Q52LJ0	Protein FAM98B	GN=FAM98B SV=2	-1,49	1,70
P63279	SUMO-conjugating enzyme UBC9	GN=UBE2I SV=1	-1,49	1,47
P40197	Platelet glycoprotein V	GN=GP5 SV=1	-1,50	1,71
P58546	Myotrophin	GN=MTPN SV=2	-1,50	1,49
P19367	Hexokinase-1	GN=HK1 SV=3	-1,51	1,71
O75367	Core histone macro-H2A.1	GN=MACROH2A1 SV=4	-1,52	2,21
P26640	Valine--tRNA ligase	GN=VAR5 SV=4	-1,53	1,76
P60709	Actin, cytoplasmic 1	GN=ACTB SV=1	-1,53	1,31
P16885	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase gamma-2	GN=PLCG2 SV=4	-1,53	1,53
P49593	Protein phosphatase 1F	GN=PPM1F SV=3	-1,55	2,32
P48637	Glutathione synthetase	GN=GSS SV=1	-1,55	1,58
P21333	Filamin-A	GN=FLNA SV=4	-1,55	2,70
Q8WUM4	Programmed cell death 6-interacting protein	GN=PDCD6IP SV=1	-1,55	2,41
Q15080	Neutrophil cytosol factor 4	GN=NCF4 SV=2	-1,56	1,71
Q8IUI8-2	Isoform 2 of Cytokine receptor-like factor 3	GN=CRLF3	-1,56	2,11
P25098	Beta-adrenergic receptor kinase 1	GN=GRK2 SV=2	-1,56	2,26
Q9NYU2	UDP-glucose:glycoprotein glucosyltransferase 1	GN=UGGT1 SV=3	-1,57	1,47

P07900-2	Isoform 2 of Heat shock protein HSP 90-alpha	GN=HSP90AA1	-1,57	1,46
P26641-2	Isoform 2 of Elongation factor 1-gamma	GN=EEF1G	-1,57	3,20
Q9NUQ9	CYFIP-related Rac1 interactor B	GN=CYRIB SV=1	-1,58	1,35
O75347	Tubulin-specific chaperone A	GN=TBCA SV=3	-1,59	2,04
P25788	Proteasome subunit alpha type-3	GN=PSMA3 SV=2	-1,59	2,23
Q16181	Septin-7	GN=SEPTIN7 SV=2	-1,59	2,10
P61081	NEDD8-conjugating enzyme Ubc12	GN=UBE2M SV=1	-1,60	1,53
P08238	Heat shock protein HSP 90-beta	GN=HSP90AB1 SV=4	-1,61	2,41
P49721	Proteasome subunit beta type-2	GN=PSMB2 SV=1	-1,61	1,89
P46940	Ras GTPase-activating-like protein IQGAP1	GN=IQGAP1 SV=1	-1,62	2,00
P34932	Heat shock 70 kDa protein 4	GN=HSPA4 SV=4	-1,62	2,05
P15153	Ras-related C3 botulinum toxin substrate 2	GN=RAC2 SV=1	-1,62	1,69
Q96Q11	CCA tRNA nucleotidyltransferase 1, mitochondrial	GN=TRNT1 SV=2	-1,62	2,36
P25786-2	Isoform Long of Proteasome subunit alpha type-1	GN=PSMA1	-1,63	1,40
P29350-3	Isoform 2 of Tyrosine-protein phosphatase non-receptor type 6	GN=PTPN6	-1,63	1,80
P14618-2	Isoform M1 of Pyruvate kinase PKM	GN=PKM	-1,64	1,38
P12429	Annexin A3	GN=ANXA3 SV=3	-1,66	1,54
Q01082	Spectrin beta chain, non-erythrocytic 1	GN=SPTBN1 SV=2	-1,67	1,93
Q9Y2Q3	Glutathione S-transferase kappa 1	GN=GSTK1 SV=3	-1,67	1,37
P41250-2	Isoform 2 of Glycine--tRNA ligase	GN=GARS1	-1,68	1,53
O95352	Ubiquitin-like modifier-activating enzyme ATG7	GN=ATG7 SV=1	-1,69	1,35
P00387-3	Isoform 3 of NADH-cytochrome b5 reductase 3	GN=CYB5R3	-1,70	1,49
P11310-2	Isoform 2 of Medium-chain specific acyl-CoA dehydrogenase, mitochondrial	GN=ACADM	-1,70	1,76
P14625	Endoplasmic reticulum chaperone	GN=HSP90B1 SV=1	-1,70	1,93
Q9UQE7	Structural maintenance of chromosomes protein 3	GN=SMC3 SV=2	-1,70	1,90
P48507	Glutamate--cysteine ligase regulatory subunit	GN=GCLM SV=1	-1,71	1,80
O14818	Proteasome subunit alpha type-7	GN=PSMA7 SV=1	-1,74	1,95
P06703	Protein S100-A6	GN=S100A6 SV=1	-1,74	2,81
P40121	Macrophage-capping protein	GN=CAPG SV=2	-1,74	1,46
P49327	Fatty acid synthase	GN=FASN SV=3	-1,75	1,30
O95292	Vesicle-associated membrane protein-associated protein B/C	GN=VAPB SV=3	-1,75	1,40
B5ME19	Eukaryotic translation initiation factor 3 subunit C-like protein	GN=EIF3CL SV=1	-1,76	1,66
P00491	Purine nucleoside phosphorylase	GN=PNP SV=2	-1,77	1,96
P11142	Heat shock cognate 71 kDa protein	GN=HSPA8 SV=1	-1,77	5,03
P52566	Rho GDP-dissociation inhibitor 2	GN=ARHGDI2 SV=3	-1,78	2,12
P23381	Tryptophan--tRNA ligase, cytoplasmic	GN=WARS1 SV=2	-1,78	2,32
Q15181	Inorganic pyrophosphatase	GN=PPA1 SV=2	-1,78	3,53
P63104	14-3-3 protein zeta/delta	GN=YWHAZ SV=1	-1,79	2,04

Q8NBF2	NHL repeat-containing protein 2 GN=NHLRC2 SV=1	-1,79	2,00
Q9NQW7-3	Isoform 3 of Xaa-Pro aminopeptidase 1 GN=XPNPEP1	-1,80	2,23
P35236-3	Isoform 3 of Tyrosine-protein phosphatase non-receptor type 7 GN=PTPN7	-1,81	1,37
P28066	Proteasome subunit alpha type-5 GN=PSMA5 SV=3	-1,81	1,71
P08754	Guanine nucleotide-binding protein G(i) subunit alpha-3 GN=GNAI3 SV=3	-1,81	1,64
P31150	Rab GDP dissociation inhibitor alpha GN=GDI1 SV=2	-1,84	3,70
P09960	Leukotriene A-4 hydrolase GN=LTA4H SV=2	-1,85	1,33
Q9Y237-2	Isoform 2 of Peptidyl-prolyl cis-trans isomerase NIMA-interacting 4 GN=PIN4	-1,85	2,04
P50990	T-complex protein 1 subunit theta GN=CCT8 SV=4	-1,86	2,97
Q14141	Septin-6 GN=SEPTIN6 SV=4	-1,86	1,92
P09493-5	Isoform 5 of Tropomyosin alpha-1 chain GN=TPM1	-1,86	2,26
P61160	Actin-related protein 2 GN=ACTR2 SV=1	-1,88	2,42
Q14696	LRP chaperone MESD GN=MESD SV=2	-1,88	1,67
P07948-2	Isoform 2 of Tyrosine-protein kinase Lyn GN=LYN	-1,89	1,54
P00367	Glutamate dehydrogenase 1, mitochondrial GN=GLUD1 SV=2	-1,89	1,56
P50991	T-complex protein 1 subunit delta GN=CCT4 SV=4	-1,90	1,36
Q13347	Eukaryotic translation initiation factor 3 subunit I GN=EIF3I SV=1	-1,90	1,54
Q14697	Neutral alpha-glucosidase AB GN=GANAB SV=3	-1,91	1,38
O00194	Ras-related protein Rab-27B GN=RAB27B SV=4	-1,91	2,14
P62314	Small nuclear ribonucleoprotein Sm D1 GN=SNRPD1 SV=1	-1,92	1,78
Q99832	T-complex protein 1 subunit eta GN=CCT7 SV=2	-1,93	2,75
Q9H3U1	Protein unc-45 homolog A GN=UNC45A SV=1	-1,93	1,48
P26599-3	Isoform 3 of Polypyrimidine tract-binding protein 1 GN=PTBP1	-1,93	1,47
P06753	Tropomyosin alpha-3 chain GN=TPM3 SV=2	-1,93	1,36
Q9NR12	PDZ and LIM domain protein 7 GN=PDLIM7 SV=1	-1,94	2,56
P55209	Nucleosome assembly protein 1-like 1 GN=NAP1L1 SV=1	-1,95	1,98
Q15027	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 1 GN=ACAP1 SV=1	-1,95	1,68
O75475	PC4 and SFRS1-interacting protein GN=PSIP1 SV=1	-1,96	1,67
Q9H939	Proline-serine-threonine phosphatase-interacting protein 2 GN=PSTPIP2 SV=4	-1,97	2,14
P10644	cAMP-dependent protein kinase type I-alpha regulatory subunit GN=PRKAR1A SV=1	-1,98	2,32
P50552	Vasodilator-stimulated phosphoprotein GN=VASP SV=3	-1,98	2,50
P62140	Serine/threonine-protein phosphatase PP1-beta catalytic subunit GN=PPP1CB SV=3	-1,99	1,69
P27105	Stomatin GN=STOM SV=3	-2,00	2,06
P13797	Plastin-3 GN=PLS3 SV=4	-2,00	1,36
Q12904-2	Isoform 2 of Aminoacyl tRNA synthase complex-interacting multifunctional protein 1 GN=AIMP1	-2,00	2,08
Q5TFE4	5'-nucleotidase domain-containing protein 1 GN=NT5DC1 SV=1	-2,00	1,56

P15927-3	Isoform 3 of Replication protein A 32 kDa subunit	GN=RPA2	-2,01	2,39
P55072	Transitional endoplasmic reticulum ATPase	GN=VCP SV=4	-2,01	1,86
P37802-2	Isoform 2 of Transgelin-2	GN=TAGLN2	-2,02	1,53
P61981	14-3-3 protein gamma	GN=YWHAG SV=2	-2,02	2,97
Q9H0P0	Cytosolic 5'-nucleotidase 3A	GN=NT5C3A SV=3	-2,02	1,66
P09417	Dihydropteridine reductase	GN=QDPR SV=2	-2,03	3,74
P27348	14-3-3 protein theta	GN=YWHAQ SV=1	-2,03	1,53
Q13564-4	Isoform 4 of NEDD8-activating enzyme E1 regulatory subunit	GN=NAE1	-2,03	1,62
O00567	Nucleolar protein 56	GN=NOP56 SV=4	-2,03	1,47
Q9NYL9	Tropomodulin-3	GN=TMOD3 SV=1	-2,04	2,36
Q99439	Calponin-2	GN=CNN2 SV=4	-2,04	1,93
O00161	Synaptosomal-associated protein 23	GN=SNAP23 SV=1	-2,08	1,32
O75351	Vacuolar protein sorting-associated protein 4B	GN=VPS4B SV=2	-2,09	1,55
P18206	Vinculin	GN=VCL SV=4	-2,10	3,34
P61088	Ubiquitin-conjugating enzyme E2 N	GN=UBE2N SV=1	-2,11	1,87
O00429	Dynamamin-1-like protein	GN=DNM1L SV=2	-2,11	2,75
P52565	Rho GDP-dissociation inhibitor 1	GN=ARHGDI1 SV=3	-2,11	1,50
P07203	Glutathione peroxidase 1	GN=GPX1 SV=4	-2,12	2,64
Q92530	Proteasome inhibitor PI31 subunit	GN=PSMF1 SV=2	-2,14	1,58
Q96IU4	Protein ABHD14B	GN=ABHD14B SV=1	-2,14	2,09
P35579	Myosin-9	GN=MYH9 SV=4	-2,14	1,98
P62942	Peptidyl-prolyl cis-trans isomerase FKBP1A	GN=FKBP1A SV=2	-2,14	2,49
P08567	Pleckstrin	GN=PLEK SV=3	-2,14	1,62
Q15833	Syntaxin-binding protein 2	GN=STXBP2 SV=2	-2,15	1,64
Q12905	Interleukin enhancer-binding factor 2	GN=ILF2 SV=2	-2,16	1,40
P13010	X-ray repair cross-complementing protein 5	GN=XRCC5 SV=3	-2,17	1,65
O43707	Alpha-actinin-4	GN=ACTN4 SV=2	-2,17	2,73
P51452	Dual specificity protein phosphatase 3	GN=DUSP3 SV=1	-2,19	2,48
P53396	ATP-citrate synthase	GN=ACLY SV=3	-2,20	2,18
P14543	Nidogen-1	GN=NID1 SV=3	-2,21	2,31
P09486	SPARC	GN=SPARC SV=1	-2,21	2,87
P20061	Transcobalamin-1	GN=TCN1 SV=2	-2,22	1,31
P11940	Polyadenylate-binding protein 1	GN=PABPC1 SV=2	-2,22	1,90
Q9NVS9	Pyridoxine-5'-phosphate oxidase	GN=PNPO SV=1	-2,24	2,43
Q9Y6E0	Serine/threonine-protein kinase 24	GN=STK24 SV=1	-2,27	1,38
Q96S44	EKC/KEOPS complex subunit TP53RK	GN=TP53RK SV=2	-2,27	1,83
P07225	Vitamin K-dependent protein S	GN=PROS1 SV=1	-2,29	1,78
P49368	T-complex protein 1 subunit gamma	GN=CCT3 SV=4	-2,30	1,90
Q9H4A3-7	Isoform 6 of Serine/threonine-protein kinase WNK1	GN=WNK1	-2,32	2,92
P38606	V-type proton ATPase catalytic subunit A	GN=ATP6V1A SV=2	-2,32	2,75
P24844	Myosin regulatory light polypeptide 9	GN=MYL9 SV=4	-2,33	1,59
Q96AZ6	Interferon-stimulated gene 20 kDa protein	GN=ISG20 SV=2	-2,34	1,36

P49903	Selenide, water dikinase 1 GN=SEPHS1 SV=2	-2,34	1,48
P09874	Poly [ADP-ribose] polymerase 1 GN=PARP1 SV=4	-2,34	2,03
Q9Y5X3	Sorting nexin-5 GN=SNX5 SV=1	-2,36	1,83
O75563	Src kinase-associated phosphoprotein 2 GN=SKAP2 SV=1	-2,37	1,56
P06753-2	Isoform 2 of Tropomyosin alpha-3 chain GN=TPM3	-2,37	2,08
Q15843	NEDD8 GN=NEDD8 SV=1	-2,37	2,33
Q00013	55 kDa erythrocyte membrane protein GN=MPP1 SV=2	-2,40	1,76
P01137	Transforming growth factor beta-1 proprotein GN=TGFB1 SV=2	-2,41	2,06
P62495	Eukaryotic peptide chain release factor subunit 1 GN=ETF1 SV=3	-2,42	2,16
Q9Y613	FH1/FH2 domain-containing protein 1 GN=FHOD1 SV=3	-2,42	2,39
O60506	Heterogeneous nuclear ribonucleoprotein Q GN=SYNCRIP SV=2	-2,42	1,84
Q14974	Importin subunit beta-1 GN=KPNB1 SV=2	-2,43	1,86
Q14315	Filamin-C GN=FLNC SV=3	-2,43	1,91
P07996	Thrombospondin-1 GN=THBS1 SV=2	-2,44	2,35
Q15084-2	Isoform 2 of Protein disulfide-isomerase A6 GN=PDIA6	-2,44	2,16
P08514	Integrin alpha-IIb GN=ITGA2B SV=3	-2,46	1,83
P54577	Tyrosine--tRNA ligase, cytoplasmic GN=YARS1 SV=4	-2,47	2,58
P02675	Fibrinogen beta chain GN=FGB SV=2	-2,48	2,36
Q9H0E2	Toll-interacting protein GN=TOLLIP SV=1	-2,49	1,48
Q70J99-3	Isoform 3 of Protein unc-13 homolog D GN=UNC13D	-2,50	2,94
P17987	T-complex protein 1 subunit alpha GN=TCP1 SV=1	-2,51	1,68
P48643	T-complex protein 1 subunit epsilon GN=CCT5 SV=1	-2,55	1,93
Q9NRN7	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase GN=AASDHPPT SV=2	-2,57	2,22
P40227	T-complex protein 1 subunit zeta GN=CCT6A SV=3	-2,63	2,06
Q13148	TAR DNA-binding protein 43 GN=TARDBP SV=1	-2,65	2,50
P07384	Calpain-1 catalytic subunit GN=CAPN1 SV=1	-2,65	3,09
O60664	Perilipin-3 GN=PLIN3 SV=3	-2,68	1,93
P05771-2	Isoform Beta-II of Protein kinase C beta type GN=PRKCB	-2,72	2,73
Q86UX7	Fermitin family homolog 3 GN=FERMT3 SV=1	-2,73	1,79
P55957-2	Isoform 2 of BH3-interacting domain death agonist GN=BID	-2,74	1,34
O14979	Heterogeneous nuclear ribonucleoprotein D-like GN=HNRNPDL SV=3	-2,77	1,75
Q00577	Transcriptional activator protein Pur-alpha GN=PURA SV=2	-2,78	1,59
Q9Y490	Talin-1 GN=TLN1 SV=3	-2,79	3,47
P30419	Glycylpeptide N-tetradecanoyltransferase 1 GN=NMT1 SV=2	-2,79	1,95
Q96C90	Protein phosphatase 1 regulatory subunit 14B GN=PPP1R14B SV=3	-2,83	2,38
P60660-2	Isoform Smooth muscle of Myosin light polypeptide 6 GN=MYL6	-2,83	2,56
Q14498	RNA-binding protein 39 GN=RBM39 SV=2	-2,88	1,77
Q13418	Integrin-linked protein kinase GN=ILK SV=2	-2,88	1,50
P13224-2	Isoform 2 of Platelet glycoprotein Ib beta chain GN=GP1BB	-2,89	1,49

Q9UMX5	Neudesin GN=NENF SV=1	-2,89	1,68
P07951-2	Isoform 2 of Tropomyosin beta chain GN=TPM2	-2,94	1,56
O60229	Kalirin GN=KALRN SV=3	-2,98	1,70
P63000-2	Isoform B of Ras-related C3 botulinum toxin substrate 1 GN=RAC1	-3,04	3,44
O14950	Myosin regulatory light chain 12B GN=MYL12B SV=2	-3,16	2,36
P12814	Alpha-actinin-1 GN=ACTN1 SV=2	-3,24	3,60
P62258	14-3-3 protein epsilon GN=YWHAE SV=1	-3,32	3,62
P08708	40S ribosomal protein S17 GN=RPS17 SV=2	-3,34	1,54
P67936	Tropomyosin alpha-4 chain GN=TPM4 SV=3	-3,34	3,38
Q04917	14-3-3 protein eta GN=YWHAH SV=4	-3,37	2,83
Q96B97	SH3 domain-containing kinase-binding protein 1 GN=SH3KBP1 SV=2	-3,44	1,93
P05106	Integrin beta-3 GN=ITGB3 SV=2	-3,54	2,62
P02671	Fibrinogen alpha chain GN=FGA SV=2	-3,56	2,72
Q9NZN3	EH domain-containing protein 3 GN=EHD3 SV=2	-3,57	1,95
P01033	Metalloproteinase inhibitor 1 GN=TIMP1 SV=1	-3,64	2,90
P61204	ADP-ribosylation factor 3 GN=ARF3 SV=2	-3,97	2,76
Q14677-3	Isoform 3 of Clathrin interactor 1 GN=CLINT1	-4,49	2,06

Tabela anexo 11.6. Proteínas diferencialmente expressas em leucócitos de pacientes com CHH em relação aos controles saudáveis. A tabela abaixo apresenta o número de acesso da proteína (Homo sapiens no banco de dados UniProt, o nome da proteína e a descrição; diferença de expressão/fold difference (razão pacientes/controles). Todas as proteínas são referentes à espécie Homo sapiens e foram ordenadas de forma decrescente em relação às razões paciente/controle. GN = Nome do gene (do inglês, Gene Name); SV = número da versão da sequência (do inglês, SequenceVersion)

11.7 Proteínas diferencialmente expressas em leucócitos de pacientes com CHH do presente relacionadas às vias de sinalização

Via de sinalização	ID uniprot	Nome do gene	Nome da proteína
WNT	Positivamente regulada		
	P62805	H4C1	Histone H4 GN=H4C1 SV=2
	P62879	GNB2	Guanine nucleotide-binding protein G(I)/G(S)/G(T) subunit beta-2 GN=GNB2 SV=3
	P68431	H3C1	Histone H3.1 GN=H3C1 SV=2
	P04908	H2AC4	Histone H2A type 1-B/E GN=H2AC4 SV=2
	P23527	H2BC17	Histone H2B type 1-O GN=H2BC17 SV=3
	P40306	PSMB10	Proteasome subunit beta type-10 GN=PSMB10 SV=1
	Negativamente regulada		
	P28066	PSMA5	Proteasome subunit alpha type-5 GN=PSMA5 SV=3
	Q06323	PSME1	Proteasome activator complex subunit 1 GN=PSME1 SV=1
	O75832	PSMD10	26S proteasome non-ATPase regulatory subunit 10 GN=PSMD10 SV=1
	P25788	PSMA3	Proteasome subunit alpha type-3 GN=PSMA3 SV=2
	P15153	RAC2	Ras-related C3 botulinum toxin substrate 2 GN=RAC2 SV=1
	P62195	PSMC5	26S proteasome regulatory subunit 8 GN=PSMC5 SV=1
	Q01970	PLCB3	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-3 GN=PLCB3 SV=2
	P63104	YWHAZ	14-3-3 protein zeta/delta GN=YWHAZ SV=1
	Q92530	PSMF1	Proteasome inhibitor PI31 subunit GN=PSMF1 SV=2
	P55036	PSMD4	26S proteasome non-ATPase regulatory subunit 4 GN=PSMD4 SV=1
	P49721	PSMB2	Proteasome subunit beta type-2 GN=PSMB2 SV=1
O14818	PSMA7	Proteasome subunit alpha type-7 GN=PSMA7 SV=1	
TGFB	Positivamente regulada		
	P51668	UBE2D1	Ubiquitin-conjugating enzyme E2 D1 GN=UBE2D1 SV=1
	Negativamente regulada		
	P05121	SERPINE1	Plasminogen activator inhibitor 1 GN=SERPINE1 SV=1
	O15198	SMAD9	Mothers against decapentaplegic homolog 9 GN=SMAD9 SV=1
	P01137	TGFB1	Transforming growth factor beta-1 proprotein GN=TGFB1 SV=2
	Q15843	NEDD8	NEDD8 GN=NEDD8 SV=1
	P62942	FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A GN=FKBP1A SV=2
	P09874	PARP1	Poly [ADP-ribose] polymerase 1 GN=PARP1 SV=4
	Q9UNE7	STUB1	E3 ubiquitin-protein ligase CHIP GN=STUB1 SV=2
	P61081	UBE2M	NEDD8-conjugating enzyme Ubc12 GN=UBE2M SV=1
	P62140	PPP1CB	Serine/threonine-protein phosphatase PP1-beta catalytic subunit GN=PPP1CB SV=3
P05106	ITGB3	Integrin beta-3 GN=ITGB3 SV=2	
Rho GTPase	Positivamente regulada		
P62805	H4C1	Histone H4 GN=H4C1 SV=2	

Q9Y2I1	NISCH	Nischarin	GN=NISCH SV=3
P68431	H3C1	Histone H3.1	GN=H3C1 SV=2
P34897	SHMT2	Serine hydroxymethyltransferase, mitochondrial	GN=SHMT2 SV=3
P04908	H2AC4	Histone H2A type 1-B/E	GN=H2AC4 SV=2
P23527	H2BC17	Histone H2B type 1-O	GN=H2BC17 SV=3
P46527	CDKN1B	Cyclin-dependent kinase inhibitor 1B	GN=CDKN1B SV=1
P04839	CYBB	Cytochrome b-245 heavy chain	GN=CYBB SV=2
Negativamente regulada			
P08238	HSP90AB1	Heat shock protein HSP 90-beta	GN=HSP90AB1 SV=4
P46060	RANGAP1	Ran GTPase-activating protein 1	GN=RANGAP1 SV=1
P35579	MYH9	Myosin-9	GN=MYH9 SV=4
P62140	PPP1CB	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	GN=PPP1CB SV=3
P55072	VCP	Transitional endoplasmic reticulum ATPase	GN=VCP SV=4
P61981	YWHAG	14-3-3 protein gamma	GN=YWHAG SV=2
P15153	RAC2	Ras-related C3 botulinum toxin substrate 2	GN=RAC2 SV=1
Q9BZF9	UACA	Uveal autoantigen with coiled-coil domains and ankyrin repeats	GN=UACA SV=2
Q7L576	CYFIP1	Cytoplasmic FMR1-interacting protein 1	GN=CYFIP1 SV=1
Q5T0N5	FNBP1L	Formin-binding protein 1-like	GN=FNBP1L SV=3
P06753	TPM3	Tropomyosin alpha-3 chain	GN=TPM3 SV=2
P60709	ACTB	Actin, cytoplasmic 1	GN=ACTB SV=1
Q15836	VAMP3	Vesicle-associated membrane protein 3	GN=VAMP3 SV=3
O95292	VAPB	Vesicle-associated membrane protein-associated protein B/C	GN=VAPB SV=3
Q9P107	GMIP	GEM-interacting protein	GN=GMIP SV=2
P61160	ACTR2	Actin-related protein 2	GN=ACTR2 SV=1
P40227	CCT6A	T-complex protein 1 subunit zeta	GN=CCT6A SV=3
P63104	YWHAZ	14-3-3 protein zeta/delta	GN=YWHAZ SV=1
P62258	YWHAE	14-3-3 protein epsilon	GN=YWHAE SV=1
Q9BX68	HINT2	Histidine triad nucleotide-binding protein 2, mitochondrial	GN=HINT2 SV=1
O60229	KALRN	Kalirin	GN=KALRN SV=3
O14964	HGS	Hepatocyte growth factor-regulated tyrosine kinase substrate	GN=HGS SV=1
Q13459	MYO9B	Unconventional myosin-IXb	GN=MYO9B SV=3
Q9NZQ3	NCKIPSD	NCK-interacting protein with SH3 domain	GN=NCKIPSD SV=1
Q92974	ARHGEF2	Rho guanine nucleotide exchange factor 2	GN=ARHGEF2 SV=4
Q9NY65	TUBA8	Tubulin alpha-8 chain	GN=TUBA8 SV=1
P52565	ARHGDIA	Rho GDP-dissociation inhibitor 1	GN=ARHGDIA SV=3
Q04917	YWHAH	14-3-3 protein eta	GN=YWHAH SV=4
P52566	ARHGDIB	Rho GDP-dissociation inhibitor 2	GN=ARHGDIB SV=3
Q15080	NCF4	Neutrophil cytosol factor 4	GN=NCF4 SV=2
P24844	MYL9	Myosin regulatory light polypeptide 9	GN=MYL9 SV=4
O00161	SNAP23	Synaptosomal-associated protein 23	GN=SNAP23 SV=1

P67936	TPM4	Tropomyosin alpha-4 chain	GN=TPM4 SV=3
P21333	FLNA	Filamin-A	GN=FLNA SV=4
Q01082	SPTBN1	Spectrin beta chain, non-erythrocytic 1	GN=SPTBN1 SV=2
O14950	MYL12B	Myosin regulatory light chain 12B	GN=MYL12B SV=2
Q13492	PICALM	Phosphatidylinositol-binding clathrin assembly protein	GN=PICALM SV=2
Q99832	CCT7	T-complex protein 1 subunit eta	GN=CCT7 SV=2
P12814	ACTN1	Alpha-actinin-1	GN=ACTN1 SV=2
Q15642	TRIP10	Cdc42-interacting protein 4	GN=TRIP10 SV=3
O43396	TXNL1	Thioredoxin-like protein 1	GN=TXNL1 SV=3
P27348	YWHAQ	14-3-3 protein theta	GN=YWHAQ SV=1
P27105	STOM	Stomatin	GN=STOM SV=3
P46940	IQGAP1	Ras GTPase-activating-like protein IQGAP1	GN=IQGAP1 SV=1
P52948	NUP98	Nuclear pore complex protein Nup98-Nup96	GN=NUP98 SV=4
P33176	KIF5B	Kinesin-1 heavy chain	GN=KIF5B SV=1
Q9NYL9	TMOD3	Tropomodulin-3	GN=TMOD3 SV=1

Tabela anexo 11.7. Proteínas diferencialmente expressas em leucócitos de pacientes com CHH do presente relacionadas às vias de sinalização. A tabela abaixo apresenta o número de acesso da proteína no banco de dados UniProt, o símbolo do gene e o nome da proteína. Todas as proteínas são referentes à espécie *Homo sapiens*. GN = Nome do gene (do inglês, Gene Name); SV = número da versão da sequência (do inglês, *SequenceVersion*)

11.8 Proteínas diferencialmente expressas em leucócitos de pacientes com CHH do presente estudo também detectadas em ensaio de transcriptoma da literatura (Rogler *et al.*, 2013).

ID uniprot	Nome do gene	Nome da proteína
Q7Z6P3	RAB44	Ras-related protein Rab-44 GN=RAB44 SV=4
P52565	ARHGDI1	Rho GDP-dissociation inhibitor 1 GN=ARHGDI1 SV=3
Q08043	ACTN3	Alpha-actinin-3 GN=ACTN3 SV=2
P52565	ARHGDI1	Rho GDP-dissociation inhibitor 1 GN=ARHGDI1 SV=3
Q9NZQ3	NCKIPSD	NCK-interacting protein with SH3 domain GN=NCKIPSD SV=1
Q03111	MLLT1	Protein ENL GN=MLLT1 SV=2
Q15181	PPA1	Inorganic pyrophosphatase GN=PPA1 SV=2
P61086	UBE2K	Ubiquitin-conjugating enzyme E2 K GN=UBE2K SV=3
Q13492	PICALM	Phosphatidylinositol-binding clathrin assembly protein GN=PICALM SV=2
Q9H074	PAIP1	Polyadenylate-binding protein-interacting protein 1 GN=PAIP1 SV=1
P07311	ACYP1	Acylphosphatase-1 GN=ACYP1 SV=2
Q8TDX7	NEK7	Serine/threonine-protein kinase Nek7 GN=NEK7 SV=1
P61088	UBE2N	Ubiquitin-conjugating enzyme E2 N GN=UBE2N SV=1
O60884	DNAJA2	DnaJ homolog subfamily A member 2 GN=DNAJA2 SV=1
P52788	SMS	Spermine synthase GN=SMS SV=2
Q9NRN7	AASDHPPT	L-aminoadipate-semialdehyde dehydrogenase-phosphopantetheinyl transferase GN=AASDHPPT SV=2
P63104	YWHAZ	14-3-3 protein zeta/delta GN=YWHAZ SV=1
P08754	GNAI3	Guanine nucleotide-binding protein G(i) subunit alpha-3 GN=GNAI3 SV=3
Q96B26	EXOSC8	Exosome complex component RRP43 GN=EXOSC8 SV=1
O43567	RNF13	E3 ubiquitin-protein ligase RNF13 GN=RNF13 SV=1
P30048	PRDX3	Thioredoxin-dependent peroxide reductase, mitochondrial GN=PRDX3 SV=3
P13797	PLS3	Plastin-3 GN=PLS3 SV=4
P61224	RAP1B	Ras-related protein Rap-1b GN=RAP1B SV=1
Q8NBF2	NHLRC2	NHL repeat-containing protein 2 GN=NHLRC2 SV=1
P61224	RAP1B	Ras-related protein Rap-1b GN=RAP1B SV=1
Q9UBP0	SPAST	Spastin GN=SPAST SV=1
Q5T447	HECTD3	E3 ubiquitin-protein ligase HECTD3 GN=HECTD3 SV=1
Q96C90	PPP1R14B	Protein phosphatase 1 regulatory subunit 14B GN=PPP1R14B SV=3
Q9BZZ5	API5	Apoptosis inhibitor 5 GN=API5 SV=3
Q9Y385	UBE2J1	Ubiquitin-conjugating enzyme E2 J1 GN=UBE2J1 SV=2
P62942	FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A GN=FKBP1A SV=2
O95793	STAU1	Double-stranded RNA-binding protein Staufen homolog 1 GN=STAU1 SV=2
Q86S22	TRAPPC6B	Trafficking protein particle complex subunit 6B GN=TRAPPC6B SV=1
Q9NZL9	MAT2B	Methionine adenosyltransferase 2 subunit beta GN=MAT2B SV=1
Q5T0N5	FNBP1L	Formin-binding protein 1-like GN=FNBP1L SV=3
P62942	FKBP1A	Peptidyl-prolyl cis-trans isomerase FKBP1A GN=FKBP1A SV=2

Q9H1E3	NUCKS1	Nuclear ubiquitous casein and cyclin-dependent kinase substrate 1	GN=NUCKS1 SV=1
P38606	ATP6V1A	V-type proton ATPase catalytic subunit A	GN=ATP6V1A SV=2
P07947	YES1	Tyrosine-protein kinase Yes	GN=YES1 SV=3
Q9NVS9	PNPO	Pyridoxine-5'-phosphate oxidase	GN=PNPO SV=1
O14672	ADAM10	Disintegrin and metalloproteinase domain-containing protein 10	GN=ADAM10 SV=1
O75874	IDH1	Isocitrate dehydrogenase [NADP] cytoplasmic	GN=IDH1 SV=2
Q03111	MLLT1	Protein ENL	GN=MLLT1 SV=2
Q9UNE7	STUB1	E3 ubiquitin-protein ligase CHIP	GN=STUB1 SV=2
O75884	RBBP9	Serine hydrolase RBBP9	GN=RBBP9 SV=2

Tabela anexo 11.8. Proteínas diferencialmente expressas em leucócitos de pacientes com CHH do presente estudo também detectadas em ensaio de transcriptoma da literatura (Rogler *et al.*, 2013). A tabela abaixo apresenta o número de acesso da proteína no banco de dados UniProt, o símbolo do gene e o nome da proteína. Homo sapiens. GN = Nome do gene (do inglês, Gene Name); SV = número da versão da sequência (do inglês, *SequenceVersion*)

11.9 Artigo publicado

Identification of Novel and Recurrent *RMRP* Variants in a Series of Brazilian Patients with Cartilage-Hair Hypoplasia: McKusick Syndrome

Maria E. Gomes^a Luiza Calatrava Paternostro^a Valéria R. Moura^a
Deborah Antunes^b Ernesto R. Caffarena^b Dafne Horovitz^c
Maria T. Sanseverino^d Gabriela Ferraz Leal^e Têmis M. Felix^f
Denise Pontes Cavalcanti^g Juan Clinton Llerena Jr.^{c, h, i} Sayonara Gonzalez^a

^aLaboratório de Medicina Genômica, Centro de Genética Médica Dr. José Carlos Cabral de Almeida & Centro de Referência para Doenças Raras, IFF/FIOCRUZ, Rio de Janeiro, Brazil; ^bLaboratório de Biofísica Computacional e Modelagem Molecular (PROCC), IOC/FIOCRUZ, Rio de Janeiro, Brazil; ^cUnidade de Genética Clínica, Centro de Genética Médica Dr. José Carlos Cabral de Almeida & Centro de Referência para Doenças Raras, IFF/FIOCRUZ, Rio de Janeiro, Brazil; ^dUniversidade Federal do Rio Grande do Sul/UFRGS, Porto Alegre, Brazil; ^eInstituto de Medicina Integral Prof. Fernando Figueira, Recife, Brazil; ^fServiço de Genética Médica Hospital de Clínicas de Porto Alegre (HCPA), Porto Alegre, Brazil; ^gGrupo de Displasias Esqueléticas, Departamento de Genética Médica, FCM-UNICAMP, Campinas, Brazil; ^hINAGEMP – Instituto Nacional de Genética Médica Populacional, Porto Alegre, Brazil; ⁱFaculdade de Medicina Fundação Arthur Sá Earp Jr, Petrópolis, Brazil

Keywords

Cartilage-hair hypoplasia · Lymphoma · McKusick syndrome · Metaphyseal bone dysplasia · *RMRP* · RNase MRP

Abstract

Cartilage-hair hypoplasia syndrome (CHH) is an autosomal recessive disorder caused by pathogenic variants of the *RMRP* gene and characterized by metaphyseal bone dysplasia associated with hypotrichosis, immunodeficiency, and predisposition to malignancy. However, the genotype-phenotype correlation in CHH is not well understood. Here, we report a single country cohort of 23 Brazilian patients with clinical and radiological features consistent with CHH. We found 23 different pathogenic variants in the *RMRP* gene – 12 novel and 11 previously described in the literature. Interestingly, the most frequent Finnish pathogenic variant related to CHH (g.71A>G) was not found in our cohort. In contrast, more than

50% of the patients carried the rare g.196C>T variant suggesting a possible founder effect in the Brazilian population. In silico analysis showed that pathogenic variants occurred either in the regions conserved in mammalian species or within essential domains for the ribonucleoprotein complex. Pathogenicity prediction studies can improve the understanding of how these variants affect RNA.

© 2019 S. Karger AG, Basel

Cartilage-hair hypoplasia syndrome (CHH; 250250) is an autosomal recessive disorder with a broad clinical spectrum characterized by metaphyseal skeletal dysplasia, disproportionate short stature, joint laxity, light-colored fine and sparse hair, gastrointestinal problems, anemia, im-

J.C.L. and S.G. jointly supervised this work.