



**UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE CIÊNCIAS DA SAÚDE
PROGRAMA DE PÓS-GRADUAÇÃO DE PROCESSOS
INTERATIVOS DOS ÓRGÃOS E SISTEMAS**



YASMIM CRISTINA FERREIRA DE ALMEIDA

FREQUÊNCIA E AVALIAÇÃO FUNCIONAL *IN SILICO* DE VARIANTES GENÉTICAS NO GENE DA ENDOGLINA EM UMA POPULAÇÃO MISCIGENADA

Salvador – Bahia
2019

YASMIM CRISTINA FERREIRA DE ALMEIDA

FREQUÊNCIA E AVALIAÇÃO FUNCIONAL *IN SILICO* DE VARIANTES GENÉTICAS NO GENE DA ENDOGLINA EM UMA POPULAÇÃO MISCIGENADA

Dissertação submetida ao Programa de Pós-graduação de Processos Interativos dos Órgãos e Sistemas do Instituto de Ciências da Saúde da Universidade Federal da Bahia, como requisito parcial para a obtenção do título de Mestre em Processos Interativos dos Órgãos e Sistemas.

Orientadora: Prof. Dra. Camila Alexandrina Viana de Figueirêdo

Coorientadora: Profa. Dra. Keina Maciele Campos Dourado

Salvador – Bahia
2019

Ficha catalográfica elaborada pelo Sistema Universitário de Bibliotecas SIBI/UFBA),
com os dados fornecidos pelo(a) autor(a).

Almeida, Yasmim Cristina Ferreira de
Frequência e Avaliação Funcional In Silico de
Variantes Genéticas no gene da Endoglina em uma
População Miscigenada / Yasmim Cristina Ferreira de
Almeida. -- Salvador-BA, 2019.

70 f. : il

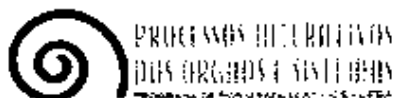
Orientadora: Camila Alexandrina Viana de
Figueirêdo.

Coorientadora: Keina Maciele Campos Dourado.
Dissertação (Mestrado - Programa de Pós Graduação em
Processos Interativos de Órgãos e Sistemas) --
Universidade Federal da Bahia, Instituto de Ciências
da Saúde, 2019.

1. Endoglina. 2. Antígeno CD105. 3. Polimorfismo
Genético. 4. Polimorfismo de Nucleotídeo Único. I.
Figueirêdo, Camila Alexandrina Viana de. II. Dourado,
Keina Maciele Campos. III. Título.

CDU: 575.23

UNIVERSIDADE FEDERAL DA BAHIA
INSTITUTO DE CIÊNCIAS DA SAÚDE



TERMO DE APROVAÇÃO
DEFESA PÚBLICA DE DISSERTAÇÃO

YASMIM CRISTINA FERREIRA DE ALMEIDA

**Frequência e Avaliação Funcional *in silico* de variantes genéticas no gene da Endogлина em
uma população miscigenada
Salvador, Bahia, 11 de dezembro de 2019**

COMISSÃO EXAMINADORA:

PROFA DRA CAMILA ALEXANDRINA VIANA DE FIGUEIREDO (UFBA)
Universidade Federal da Bahia

PROFA. DRA. KEINA MACIELE CAMPOS DOURADO

Uninassau

PROF DR GUSTAVO NUNES DE OLIVEIRA COSTA (UNIFACS)
Universidade Salvador

PROF DR BRUNO DE ALMEIDA LOPES (UFBA)
Universidade Federal da Bahia

Dedico à minha avó

Júlia Pereira Ferreira (*in memoriam*),

Que mesmo ausente, me ensina a cada dia, fazendo morada eterna em meu coração.

AGRADECIMENTOS

Gostaria de registrar e agradecer a todas as pessoas que cruzaram o meu caminho nesta jornada e que foram essenciais para me ensinar algo:

À minha querida Orientadora, Profa. Camila, uma grande inspiração, sempre atenciosa, solícita e compreensível. Obrigada por abrir as portas e pela oportunidade ímpar de trabalhar com você, além de ter me proporcionado enxergar que o caminho do conhecimento é infinito! À Profa. Keina, pelo apoio e por me apresentar a esse grande desafio.

Aos meus colegas do Programa de Pós-Graduação PIOS, em especial aos colegas Thaise e Daniel, pela troca de experiências e amizade, bem como aos queridos secretários do Programa, Célia, Tarcísio e, em especial, ao Sr. Carlos, sempre carismático e disposto a ajudar.

Ao Prof. Maurício Cardeal, por sua grande habilidade em passar o conhecimento, encantando com sua sapiência, desmitificando e rompendo conceitos.

Aos meus colegas do Laboratório IMUNOBIO pela convivência, trocas de experiência, por me permitir errar e aprender, por abrirem este espaço de conhecimento. Às amigas que levarei comigo, Raísa e Ana Paula.

Ao querido Prof. Ryan, por sempre estimular a troca de saberes e por nos encorajar a continuar no caminho.

Às queridas secretárias do PPGIm, Aline e Jusci, por me adotarem e me ajudarem sempre quando necessário, pela simpatia e solicitude. Ao colega Geraldo pela parceria junto a este desafio, e que ainda não acabou.

À UFBA, por abrir este espaço de troca e conhecimento infinito. Aos funcionários e servidores que fazem o melhor que podem, com os recursos à que tem acesso.

Enfim, à minha família, meu porto seguro! Sem os quais, a razão de tudo isso seria em vão. À minha mãe Cristina, meu pai Júlio e aos meus irmãos, Isis e Julinho por todo cuidado e amor incondicional. Vocês são essenciais.

Ao meu companheiro de vida, Eduardo, pela amizade, amor, cumplicidade, compreensão e, além... Obrigada por me estimular a continuar, mesmo quando tudo parecia desandar. Você foi e é um grande parceiro! Só tenho mesmo a agradecer!

À extensão da minha família, Fillipi, Neralci, Zé Luiz, Leo e Sanni pela amizade e amor.

À minha sobrinha e afilhada, a pequenina Maria Luísa, nossa Malu, por ressignificar nossas vidas, nos trazendo luz, amor e esperança!

“Tenha plena consciência das estrelas e do infinito lá no alto. Então a vida parece quase encantada depois de tudo.”

Vincent van Gogh

CONSIDERAÇÕES

Foi de suma importância para este trabalho o acesso aos bancos de dados dos programas Social Change Asthma and Allergy in Latin America (SCAALA), coordenado pelo Prof. Dr. Maurício Barreto e do Programa para o controle da Asma e da Rinite Alérgica na Bahia (ProAR), coordenado pelo Prof. Dr. Álvaro Cruz.

APOIO FINANCEIRO

Fundação de Amparo à Pesquisa do Estado da Bahia – FAPESB

Bolsa de Mestrado

ALMEIDA, Yasmim Cristina Ferreira de. Frequência e análise funcional de variantes genéticas no gene da endoglina em uma população miscigenada. 2019. (Dissertação) Mestrado em Processos Interativos de Órgãos e Sistemas – Instituto de Ciências da Saúde, Universidade Federal da Bahia.

RESUMO

Introdução: a endoglina (ENG, CD105) é um correceptor integrante da família do fator de crescimento transformante-beta e participa da regulação de processos celulares como proliferação, diferenciação, migração e apoptose. ENG é mais conhecida por sua expressão em células endoteliais, desempenhando papel importante na angiogênese e na vasculogênese, porém sua expressão diferencial já foi associada a desfechos patogênicos, inclusive devido a mutações no gene *ENG*. **Objetivos:** descrever a frequência de variantes genéticas em *ENG*, analisar a distribuição da frequência alélica de populações pseudoancestrais, analisar funcionalmente *in silico* as variantes genéticas em *ENG* e descrever as variantes genéticas que possam estar envolvidas em processos patogênicos descritos em outras populações. **Metodologia:** utilizamos os bancos de dados genômicos dos projetos SCAALA e ProAR, para o presente estudo descritivo. A informação genética foi extraída e as análises genéticas descritivas e funcional *in silico* foram realizadas através de programas livres e plataformas on-line. **Resultados:** foram analisadas 85 variantes genéticas, dessas, 50 variantes apresentaram alelo de menor frequência (MAF) entre 0% e 50% e 32 variantes foram monomórficas. As frequências das variantes genéticas, no geral, foram intermediárias entre as frequências de populações africanas e europeias. As análises funcionais *in silico* indicaram que, entre as 33 variantes *missense* extraídas, 10 apresentaram escore prejudicial. **Conclusão:** variantes genéticas em *ENG*, descritas no presente trabalho, foram previamente associadas com os alelos mais frequentes em outras populações ao fenótipo de malformação arteriovenosa cerebral esporádica, telangiectasia hereditária hemorrágica tipo 1 com apresentação de malformação arteriovenosa pulmonar e à pré-eclâmpsia. Estudos adicionais sobre os mecanismos funcionais desse gene são necessários, uma vez que a ENG participa de uma série de processos celulares importantes e mais esforços devem ser feitos para estudos genéticos na população brasileira, considerando a miscigenação dessa população.

Descritores: Endoglina. Antígeno CD105. Polimorfismo genético. Polimorfismo de nucleotídeo único.

ALMEIDA, Yasmim Cristina Ferreira de. *Frequency and functional analysis in silico of genetic variants in the endoglin gene in an admixture population*. 2019. (Dissertação) Mestrado em Processos Interativos de Órgãos e Sistemas – Instituto de Ciências da Saúde, Universidade Federal da Bahia.

ABSTRACT

Introduction: endoglin (ENG, CD105) is a correceptor of the transforming growth factor-beta family and it participates in the cellular processes regulation such as proliferation, differentiation, migration and apoptosis. ENG is most known for its expression in endothelial cells, playing an important role in angiogenesis and vasculogenesis. Furthermore, its differential expression has already been associated with pathogenic outcomes, some of them associated with mutations in the ENG gene. **Objectives:** to describe the genetic variants frequency in the ENG gene in the Brazil's population, to analyze the allele frequency distribution of pseudoancestral populations, to perform in silico analysis with genetic variants found in the endoglin gene and to describe genetic variants that may be involved in pathogenic processes described in other populations. **Methodology:** we used the SCAALA and ProAR genomic database for the present descriptive study. The DNA was genotyped, and functional analysis in silico was performed using free programs and online platforms. **Results:** 85 genetic variants were analyzed, 50 SNVs presented minor allele frequency between 0% and 49% and 32 variants were monomorphic. Functional analysis in silico indicate that from the 33 missense variants, 10 had damaging escore. **Conclusions:** variants in endoglin that have been previously associated with sporadic cerebral arteriovenous malformations, hereditary hemorrhagic telangiectasia type 1 phenotype with pulmonary arteriovenous malformations, and preeclampsia. Further investigation regarding the functional mechanisms of this gene are necessary, since the endoglin participates in a range of important cellular processes and more efforts should be made for genetic studies in the Brazilian population, considering the admixture phenotype.

Keywords: Endoglin. CD105 antigen. Genetic polimorfism. Single nucleotide polimorfism.

LISTA DE FIGURAS

Figura 1 - Representação esquemática da estrutura da endogлина homodimérica.....	24
Figura 2 - Modelo teórico do complexo ternário ENG-BMP9-ALK1.....	25
Figura 3 - Modulação da via de sinalização de TGF- β e BMP9 por endogлина em células endoteliais.....	30
Figura 4 - Gene e isoformas de endogлина.....	32
Figura 5 - Quantidade absoluta das variantes genéticas do gene da endogлина encontradas no banco de dados internacional ARUP.....	35
Figura 6 - Distribuição das variantes genéticas em ENG na população deste estudo...	41
Figura 7 - Análise dos efeitos estruturais de variantes missense consideradas prejudiciais utilizando a plataforma HOPE.....	48
Figura 8 - Arquitetura esquemática dos domínios de ENG e mapa de variantes genéticas com consequência missense na população estudada.	48
Figura 9 - Desequilíbrio de ligação entre as variantes genéticas extraídas do banco de dados SCAALA.....	49
Figura 10 - Desequilíbrio de ligação entre as variantes genéticas extraídas do banco de dados ProAR.....	50

LISTA DE TABELAS

Tabela 1 - Tipos celulares que expressam ENG em tecidos humanos normais.....	22
Tabela 2 - Sumário de proteínas ligantes para endogлина.....	26
Tabela 3 - Sumário das variantes genéticas em ENG extraídas do banco de dados SCAALA.....	39
Tabela 4 - Sumário das variantes genéticas em ENG extraídas do banco de dados ProAR.....	39
Tabela 5 - Diferencial de frequências alélicas entre três populações diferentes do Projeto 1000 Genomas para os marcadores genéticos do banco de dados SCAALA.....	42
Tabela 6 - Diferencial de frequências alélicas entre três populações diferentes do Projeto 1000 Genomas para os marcadores genéticos do banco de dados ProAR.....	43
Tabela 7 - Variantes missense encontradas em ENG na população de estudo e sua classificação segundo algoritmos computacionais (SIFT, PolyPhen-2, MutationTaster2 e PMut).....	46

LISTA DE ABREVIATURAS E SIGLAS

AIMs	marcadores genéticos informativos de ancestralidade (<i>Ancestry Informative Markers</i>)
ALK	Quinase semelhante ao receptor Activina (<i>Activin receptor-Like Kinase</i>)
BAVMs	malformações arteriovenosas cerebrais
BMPs	Proteínas morfogênicas Ósseas (<i>Bone Morphogenetic Proteins</i>)
CD105	expressão imuno-histoquímica de endoglina
CEU	residentes de Utah, EUA, com ancestralidade da europeia setentrional e ocidental
CHB	Chineses da etnia Han de Pequim, China
CT	citoplasmático
DL	Desequilíbrio de Ligação
EC	célula endotelial (<i>Endothelial Cell</i>)
ECTO	ectodomínio
ENG	Endoglina (<i>Endoglin</i>)
HHT	Telangiectasia Hemorrágica Hereditária (<i>Hereditary Hemorrhagic Telangiectasia</i>)
PAH	Hipertensão Arterial Pulmonar
PE	preeclâmpsia
JPS	Síndrome da polipose juvenil (<i>Juvenile Polyposis Syndrome</i>)
kDa	Kilodalton
L-endoglina	Endoglina longa (<i>Large Endoglin</i>)
mAb	Anticorpo monoclonal (<i>Monoclonal antibody</i>)
MAF	Alelo de Menor Frequência (<i>Minor Allele Frequency</i>)
MAV	Malformações arteriovenosas
MMP	Metaloproteinase de matriz
OR	domínio órfão (<i>Orphan Domain</i>)
OR	Razão de chances (<i>Odds ratio</i>)
ProAR	Programa para o controle da asma e da rinite alérgica na Bahia (<i>Asthma and Allergic Rhinitis Control Program in Bahia</i>)
RGD	Ácido tripeptídico arginina-glicina-aspártico
RNA _m	Ácido Ribonucleico mensageiro

SCAALA	Mudanças Sociais, Asma e Alergia na América Latina (Social Changes Asthma and Allergy in Latin America)
sENG	Endoglina solúvel (<i>soluble Endoglin</i>)
S-endoglina	Endoglina curta (<i>Small Endoglin</i>)
SNV	variação de nucleotídeo único (<i>Single Nucleotide Variation</i>)
SP	peptídeo sinal (<i>Signal peptide</i>)
TGF- β	Fator de crescimento transformante beta (<i>Transforming Growth Factor Beta</i>)
T β RII	receptor do fator de crescimento transformante beta tipo II
tSNV	tagging single nucleotide variation
YRI	Iorubas de Ibadan, Nigéria
ZP	zona pelúcida

SUMÁRIO

1 INTRODUÇÃO	18
2 JUSTIFICATIVA	20
3 OBJETIVOS	21
3.1 Objetivo Geral.....	21
3.2 Objetivos Específicos	21
4 REFERENCIAL TEÓRICO	22
4.1 Endogлина	22
4.1.1 Expressão, descoberta e função	22
4.1.1.1 Estrutura de endogлина	23
4.1.1.2 Endogлина e família TGF-β	26
4.1.1.3 Endogлина, sinalização e principal papel fisiológico	27
4.1.2 Gene da endogлина e variantes genéticas	30
5 MATERIAIS E MÉTODOS	35
5.1 População do estudo.....	35
5.2 Extração de DNA e genotipagem.....	36
5.3 Marcadores de ancestralidade	36
5.4 Análise funcional <i>in silico</i>	37
6 RESULTADOS	38
6.1 Descrição da frequência de variantes genéticas no gene da endogлина na população estudada.....	38
6.2 Distribuição da frequência de variantes genéticas no gene da endogлина em populações pseudoancestrais.....	41
6.3 Análise funcional <i>in silico</i> de variantes genéticas no gene da endogлина.....	44
7 DISCUSSÃO	51
8 CONCLUSÃO.....	57
REFERÊNCIAS BIBLIOGRÁFICAS	58
ANEXO.....	70

1 INTRODUÇÃO

A endogлина, também conhecida como ENG ou CD105, é um correceptor que vem sendo amplamente descrito como constitutivamente expresso em células endoteliais vasculares. A ENG exerce papel crucial no processo de angiogênese, no remodelamento vascular e na homeostase, participando do controle de diversos processos celulares, incluindo diferenciação, proliferação, inflamação e cicatrização de feridas, por meio da sua atuação como integrante da família do fator de crescimento transformante-beta (TGF- β) (1–3).

Apesar de ser mais conhecida por sua expressão em células endoteliais ativas, a ENG é encontrada em diferentes tipos celulares, tendo sua expressão diferencial associada a processos patológicos, principalmente envolvendo o endotélio. Sua expressão aumentada em células endoteliais vasculares está relacionada à embriogênese (4,5), sendo também descrita em alguns processos patológicos. Os níveis plasmáticos aumentados da ENG solúvel também são relatados em condições como pré-eclâmpsia (6,7), câncer, metástase (8,9), doenças relacionadas à inflamação, como aterosclerose, diabetes mellitus tipo II (10), psoríase, artrite reumatoide (11) e, mais recentemente, em leucemias agudas (12), representando um potencial biomarcador para essas patologias.

A importância da ENG na fisiologia do organismo foi demonstrada através da inibição do gene *eng* em camundongos, em que embriões homocigotos (ENG^{-/-}) revelam um estágio embrionário precoce e letal devido a malformações cardiovasculares, com ruptura de vasos e falta de transição endotelial-mesenquimal (13–15). Adicionalmente, foi demonstrado que mutações germinativas levam à diminuição de função da ENG, configurando um distúrbio autossômico dominante denominado telangiectasia hemorrágica hereditária tipo 1 (HHT1) (3,15–17). Diversos estudos são conduzidos para a investigação da influência da base genética na patologia e seus diferentes fenótipos de acordo com a apresentação de severidade, bem como a condução de estudos funcionais da molécula, possuindo mais de 500 variantes genéticas descritas em *ENG* associadas ao HHT1 em base de dados internacionais.

Em contrapartida, foi identificado que a superexpressão da ENG no endotélio vascular está relacionada à progressão tumoral, com presença de angiogênese induzida

por tumores sólidos (1), desta forma, os anticorpos anti-endogлина são utilizados como alvo principal para terapia anti-angiogênica (1,2,18). Além disso, *ENG* vem sendo um dos genes candidatos das vias da angiogênese para os estudos de associação das variantes genéticas com pré-eclâmpsia (19–22) devido à presença elevada da *ENG* circulante no soro de mulheres com pré-eclâmpsia, sendo maior e mais significativa antes do início dos sintomas clínicos da condição (6,7).

Considerando que a endogлина apresenta papel fundamental para a homeostase e fisiologia do organismo e que variantes genéticas podem implicar em desfechos fisiopatológicos, o presente estudo propôs descrever a frequência de variantes genéticas em uma amostra populacional brasileira, comparar com populações ancestrais envolvidas na composição demográfica dessa população, analisar funcionalmente através de método *in silico* as variantes genéticas que apresentem consequência *missense*, bem como descrever as variantes genéticas presentes na população brasileira já relatadas na literatura que podem estar potencialmente envolvidas em processos patogênicos em outras populações.

2 JUSTIFICATIVA

A endoglin medeia diversas funções celulares, compreendendo um importante papel na regulação da angiogênese e remodelamento vascular. Apesar de ser mais conhecida por sua expressão em células endoteliais, a endoglin já foi encontrada expressa em diferentes tipos celulares, tendo sua expressão diferencial associada a desfechos patogênicos. Desta forma, conhecer as variantes genéticas no gene da endoglin que estão presentes na população brasileira, considerando que esta seja miscigenada, tri-híbrida (descendente de africanos, europeus e ameríndios), comparando com as populações ancestrais, com predomínio africano e europeu, possibilitará descrever e analisar as variantes genéticas presentes na população brasileira e que podem estar potencialmente envolvidas em processos patogênicos já descritos em outras populações.

3 OBJETIVOS

3.1 Objetivo Geral

- Descrever a frequência de variantes genéticas no gene da endoglina em uma população brasileira.

3.2 Objetivos Específicos

- Analisar a distribuição da frequência alélica de populações pseudoancestrais;
- Descrever através de análises *in silico* as características funcionais e regulatórias de variantes genéticas no gene da endoglina;
- Descrever variantes genéticas no gene da endoglina que possam estar envolvidas em processos patogênicos descritas em outras populações.

4 REFERENCIAL TEÓRICO

4.1 Endoglina

4.1.1 Expressão, descoberta e função

A endoglina, também conhecida como ENG ou CD105, foi identificada e descrita como um marcador de superfície em uma linhagem de célula leucêmica pré-B, denominada HOON, através do desenvolvimento de anticorpo monoclonal murino (mAb) 44G4 (23). Com o avanço das pesquisas acerca do mAb 44G4 e seu uso em imunohistoquímica de cortes histológicos diversos (linfonodo, amígdala, baço, timo, rim, pulmão e fígado), verificou-se que o mAb 44G4 reage fortemente com as células endoteliais (ECs) de todos os tipos de vasos sanguíneos, como capilares, arteríolas, veias, vênulas e vênulas endoteliais altas, inclusive no cordão umbilical (23,24). Desde então, a ENG vem sendo amplamente descrita como constitutivamente expressa em ECs vasculares, exercendo papel crucial no processo de angiogênese, remodelamento vascular e homeostase, participando do controle de diversos processos celulares, incluindo diferenciação, proliferação, inflamação e cicatrização de feridas (1–3,18,25).

Outros tecidos e tipos celulares para além das ECs são relacionados à expressão da ENG, como sinciciotrofoblastos de placenta a termo (26), células do estroma (27) e certas células hematopoiéticas de linhagens linfóide e mieloide, como monócitos e macrófagos ativados (28), bem como células tronco hematopoiéticas (29,30) (Tabela 1).

Tabela 1 - Tipos celulares que expressam ENG em tecidos humanos normais.

Localização	Tipo celular
Sistema cardiovascular	Células endoteliais Musculatura lisa das células endoteliais
Sistema hematológico	Fibroblastos estromais de medula óssea Células progenitoras B Precusores eritroides Células CD34+ circulantes
Sistema reticuloendotelial	Macrófagos intersticiais da polpa vermelha do baço Monócitos ativados Macrófagos diferenciados Células dendríticas foliculares em órgãos linfóides
Sistema genitourinário	Células intersticiais da região extraglomerular do rim Células mesangiais da região glomerular do rim Células basais dos túbulos seminíferos
Sistema tegumentar	Melanócitos
Embrião	Sinciciotrofoblastos de placenta Células mesenquimais embrionárias do coração

Fonte: Adaptado de Dallas *et al.*, 2008 (31).

A expressão do gene da endoglina (*ENG*) foi catalogada pelo Genotype-Tissue Expression Project (GTEx Project, <https://www.gtexportal.org/>) em diferentes tecidos, apresentando maior expressão em tecidos do sistema cardiovascular (artérias e coração), seguido de pulmão e ovário, e menor expressão em sangue periférico, sistema nervoso e em linfócitos transformados pelo EBV (Anexo). No entanto, sua função e estrutura continuam sendo explorados para elucidação, visto que a expressão diferencial de *ENG* implica em processos patológicos envolvendo, principalmente, a vasculatura.

4.1.1.1 Estrutura de endoglina

A endoglina é uma glicoproteína transmembranar do tipo I, expressa como um homodímero de 180 kDa, contendo duas subunidades ligadas por ponte dissulfeto. A estrutura da *ENG* humana (Figura 1) é composta por 658 resíduos de aminoácidos dispostos em um grande domínio extracelular (576 aminoácidos), uma região transmembranar (25 aminoácidos) e uma pequena cauda citoplasmática (47 aminoácidos) (32,33).

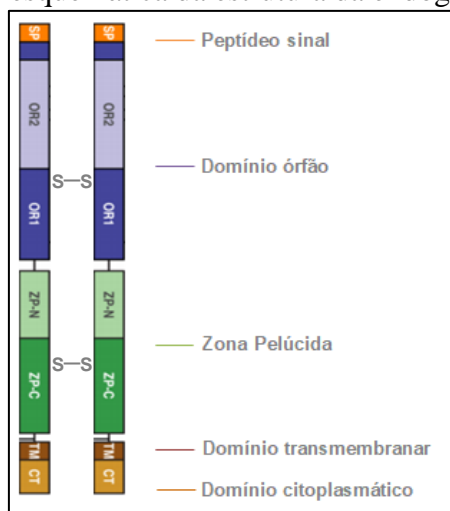
A molécula *ENG* inicia com um peptídeo sinal (resíduos 1-25), seguida do domínio extracelular, denominado ectodomínio (ECTO), que se apresenta como uma cúpula feita de monômeros orientados antiparalelamente, compreendido em domínios/módulos: domínio órfão (OR, do inglês *orphan domain*; resíduos 26-330), assim denominado por não possuir estrutura com homologia descrita, e dois módulos de Zona Pelúcida (ZP) (resíduos 349-576) (33,34).

Estruturalmente, a *ENG* pertence à família ZP, que corresponde a uma estrutura bipartida (ZP-N e ZP-C) e compreende uma importante região para a polimerização e a interação de proteínas, contendo oito resíduos de cisteínas (Cys) conservadas e distribuídas pelos módulos ZP, permitindo ligações dissulfeto intramoleculares (34,35), implicando na homodimerização e na estabilidade da molécula (3,33). O domínio ZP-N contém uma sequência que compreende uma alça exposta de um ácido tripeptídeo arginina-glicina-aspártico (RGD) nas posições 399-401, constituindo prováveis locais de glicosilação ligados a N (33,34,36) e apresenta, também, potenciais sítios de glicosilação ligados a O: Asn88, Asn102, Asn121, Asn134 e Asn307 (32). Dessa forma, a *ENG* é altamente glicosilada, contribuindo para o seu papel na adesão celular mediada por integrinas (32,37) e permitindo sua conformação multimodular (7,32,34,38). A estrutura de ZP-C compreende cerca de 113 aminoácidos

e contém uma única ligação dissulfeto intramolecular (33), incluindo os resíduos Arg437-Gly558, um importante local para interação com os receptores da família TGF- β : ALK1, ALK5 e T β R2 (discutidos na próxima seção), sendo que ALK1 também se liga a ZP-N na região que engloba os resíduos Glu26-Arg437 (34,36).

A estrutura de OR abrange os resíduos 26-330 e consiste em dois domínios entrelaçados (OR1 e OR2, vide Figura 1) que compartilham uma dobra entre si, sendo OR2 alinhado por sua porção C-terminal com N-terminal de OR1. Ambos possuem em torno de 18% de similaridade entre suas sequências de aminoácidos e contêm uma ligação dissulfeto conservada envolvendo os resíduos C30-C207 em OR1 e C53-C182 em OR2. Acredita-se que OR1 tenha sido gerado através da duplicação da sequência de OR2 com posterior permutação circular (33). Os resíduos de cisteína (Cys30, Cys53, Cys182, Cys207, Cys242, Cys330 e Cys350) encontrados na porção OR são alvos potenciais no envolvimento de ligações intermoleculares da ENG (34).

Figura 1 – Representação esquemática da estrutura da endogлина homodimérica.



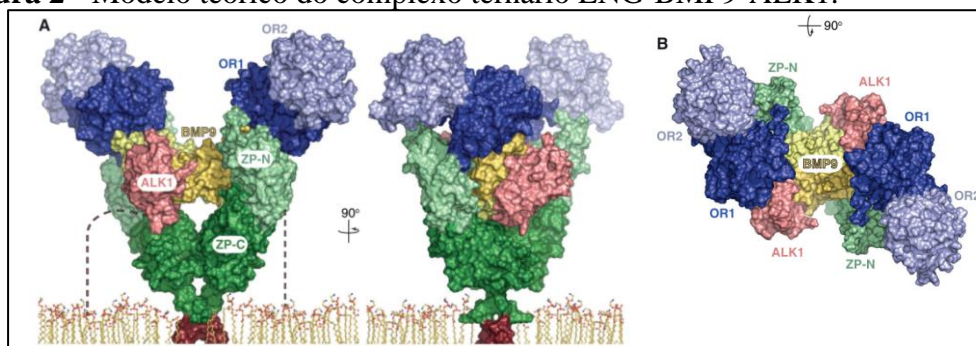
Nota: SP, peptídeo sinal; OR, região órfão; ZP, zona pelúcida; TM, domínio transmembranar; CT, domínio citoplasmático.

Fonte: Adaptado de Saito *et al.*, 2017 (33).

O módulo OR possui alta afinidade para o ligante BMP9 (também conhecido como fator de diferenciação de crescimento 2, GDF2) e ALK1 em ECs (39–41). Um estudo *in vitro* recente evidenciou que ENG faz ligação direta ao BMP9 envolvendo os resíduos N268-I271, S278 e F282, pressupondo que BMP9 seja retido na superfície plasmática devido à bivalência de ENG, o que possibilita a ligação de ALK1 em sítio diferente de BMP9, configurando um complexo ternário (Figura 2). Neste aspecto, a ZP

parece intervir através da sua dimerização, com uma disposição estrutural semelhante a um anticorpo (Figura 2A), permitindo uma maior estabilidade e avidéz do ECTO de ENG e aumentando a retenção de ligantes, possibilitando dois locais de ligação (ver Figura 2B). Outros componentes importantes que se ligam a porção OR em ENG e que sobrepõe ao sítio de ligação a BMP9, são mAb anti-endogлина, como o TRC105 com epítipo envolvendo os resíduos P231-E276 (33) e mAb 44G4 e P4A4 nos resíduos Y277-G331 (42), desta forma, inibindo a sinalização dependente de BMP9 e a cadeia de sinalização intermediada por ALK1 que estão envolvidas na proliferação celular (discutidas na seção 4.1.1.3).

Figura 2 - Modelo teórico do complexo ternário ENG-BMP9-ALK1.



Notas: (A) Possível arquitetura do complexo ternário ENG-BMP9-ALK1 na membrana plasmática; (B) Vista superior do modelo ternário ENG-BMP9-ALK1, sendo BMP9 em amarelo e ALK1 em rosa.

Fonte: Adaptado de Saito *et al.*, 2017 (33).

O domínio transmembranar de ENG possui 25 resíduos hidrofóbicos (Leu587-Trp611) (34) seguido pelo domínio citoplasmático (CT). O domínio CT (resíduos 612-658) não possui atividade enzimática conhecida, mas é rica em resíduos de serina / treonina (Ser / Tre) (32,38), funcionando similarmente a quinase, pois esses resíduos servem de substrato para fosforilação por ALK5, ALK1 e, constitutivamente, por T β RII (36). A região CT possui um domínio de ligação PDZ na porção C-terminal, composto pelos resíduos Ser-Ser-Met-Ala (1,43). Em outras espécies esse domínio é ausente, como em cães, murinos e suínos (25). Um estudo de mutação dirigida eliminando o domínio PDZ por deleção evidenciou que os resíduos de serina *em* PDZ parecem regular a fosforilação dos resíduos adjacentes de Ser / Tre, bem como a localização e glicosilação de ENG (44). Além disso, o domínio CT interage com proteínas Tctex2 β (45), β -arrestina2 (46), Zyxin e ZRP-1 (44) (Tabela 2) e supõe-se que através dessas ligações, a ENG esteja envolvida na organização do citoesqueleto, podendo mediar a

adesão e a migração celular, o transporte de proteína via vesículas endocíticas e os rearranjos de actina F (43,44).

Tabela 2 - Sumário de proteínas ligantes para endogлина.

Domínio da Endogлина	Ligantes da família TGF- β	T β RI	T β RII	Outras proteínas
Ectodomínio	TGF- β 1			
	TGF- β 3	ALK1		
	Activina A	ALK2	T β RII	
	BMP-2	ALK3	ActRII	
	BMP-7	ALK5	BMPRII	
	BMP-9 BMP-10	ALK6		
Domínio citoplasmático		ALK1 ALK5	T β RII	Tctex2 β β -Arrestin2 Zyxin ZRP-1

Fonte: Adaptado de Bernabeu *et al.*, 2009 (1) e Koleva *et al.*, 2006 (44).

A ENG possui uma forma solúvel (sENG ou sCD105) circulante de ECTO (resíduos 1-586). O derramamento de sENG ocorre sob certas condições fisiológicas como inflamação e lesão endotelial, bem como na presença de fator de necrose tumoral- α (TNF- α) e anticorpos anti-endogлина (25,47,48), através da clivagem pela metaloproteinase de matriz 14 (MMP-14 ou MT1-MMP) (8) nos resíduos G586-L587 (33). Recentemente, evidências apontaram a MMP-12 como uma importante protease responsável pela liberação de sENG em macrófagos pró-inflamatórios e ECs em modelos *in vivo* e *in vitro* (48). Após clivagem, a sENG pode formar complexos com receptores de sinalização ligados a ligantes da família TGF- β (41) atuando negativamente, ou seja, com efeito antiangiogênico e apresentando papel importante na atividade pró-inflamatória (6,47–49). Os níveis plasmáticos aumentados de sENG foram relatados em condições fisiopatológicos envolvendo o endotélio como na pré-eclâmpsia (6,7), câncer e metástase (8,9), doenças relacionadas à inflamação como aterosclerose, diabetes mellitus tipo II (10), psoríase, artrite reumatóide (11), e mais recentemente em leucemias agudas (12), representando um potencial biomarcador para essas patologias.

4.1.1.2 Endogлина e família TGF- β

O papel de ENG está relacionado, principalmente, à sua atuação como integrante da família do fator de crescimento transformante-beta (TGF- β). A família de genes de citocinas humanas do TGF- β codifica para mais de 30 membros, formando um conjunto de citocinas multifuncionais, entre elas as isoformas de TGF- β (TGF- β 1, TGF- β 2 e TGF- β 3), as proteínas morfogenéticas ósseas (BMPs), ativinas, inibinas, nodal e

hormônios anti-Mülleriano (50–53). As moléculas da família TGF- β são diméricas e geralmente estabilizadas por ligações de dissulfeto, sendo sintetizadas como grandes precursoras e liberadas como moléculas ativas quando clivadas no terminal carboxila (54), permitindo que possam se ligar aos receptores heteroméricos da família TGF- β (55).

Os membros de TGF- β formam complexos heteroméricos através da ligação de dois diferentes tipos de receptores quinases: tipo I e tipo II. Os receptores tipo I, denominados também como quinases do tipo activina (ALK), são um grupo constituído por sete membros (ALK1 a ALK7) compostos por serina/treonina quinases que são fosforiladas e ativadas por receptores tipo II. Os receptores tipo II são serina/treonina quinases (54) constitutivamente ativas dentro da célula (56) e compreendem cinco membros, sendo eles: receptor TGF- β de tipo II (T β RII), os receptores Activina e BMP (ActRII, ActRIIB e BMPRII) e o receptor de tipo II da substância inibitória Mülleriana (54).

O complexo de receptores TGF- β abrange, ainda, os correceptores, ditos auxiliares ou acessórios de sinalização: são os receptores tipo III – endogлина e β -glicano (40,41,57,58). Ambos possuem cerca de 71% de semelhança entre suas porções transmembrana e citoplasmática (59) e agem regulando o acesso do ligante TGF- β ao complexo heteromérico de receptores I e II, modulando as respostas celulares (36,60,61).

Foi postulado, inicialmente, que ENG se liga ao TGF- β 1 e ao TGF- β 3 com elevada afinidade em ECs humanas (57), devido a sua capacidade de interação com os receptores de sinalização tipo I (ALK5 e ALK1) e tipo II (36,61), desta forma, não se ligaria ao ligante por si só, exigindo um agente de reticulação. No entanto, já foi descrito que a ENG pode se ligar a receptores mesmo na ausência do ligante, interagindo direta e indiretamente com os diferentes membros de TGF- β (34,36,40,41). Estudos *in vitro* comprovaram que a ENG se liga também a activina-A, BMP-2 e BMP-7 através de receptores de activina tipo II, mas não se liga ao TGF- β 2 (40,41,57).

4.1.1.3 Endogлина, sinalização e principal papel fisiológico

A formação e a remodelação dos vasos sanguíneos são processos essenciais para a manutenção da homeostase e a função dos tecidos. Esses processos se referem à vasculogênese e à angiogênese. A vasculogênese está relacionada à diferenciação primária *in situ* de precursores endoteliais da mesoderme e a sua organização no plexo capilar primário, já a angiogênese se refere à formação *de novo* por um processo de expansão a partir de vasos pré-existentes (62,63). É relatado que a endogлина é altamente expressa durante a embriogênese em ECs vasculares em locais de angiogênese ativa (4) e atua positivamente para a função do saco vitelino (14). *In vitro*, os níveis de endogлина agem diferencialmente em células endoteliais da veia umbilical humana (HUVEC). Enquanto que a sobreexpressão de ENG está relacionada à proliferação celular, níveis baixos estão relacionados à quiescência de HUVEC (64).

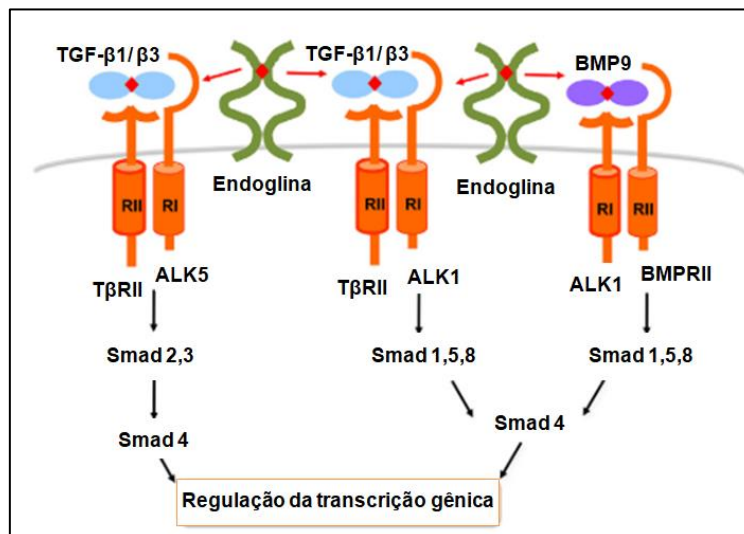
Acredita-se que os estímulos para o aumento de expressão de ENG em vasos ativados sejam provenientes de mecanismos multifatoriais, como a hipóxia, a lesão vascular e a ação de citocinas (25). O microambiente com baixa biodisponibilidade de oxigênio acarreta o acúmulo do fator de transcrição induzível por hipóxia 1 (HIF-1) que induz a expressão de endogлина nos níveis de RNA mensageiro (RNAm) (65), resultando em uma nova rede vascular que fornece sangue para o tecido hipóxico. Por sua vez, o tumor exige suprimento contínuo de oxigênio e nutrientes para manter a proliferação celular, sendo assim, a angiogênese representa um processo crucial para o crescimento do tumor, e os fatores relacionados a este processo compreendem os principais alvos para o tratamento de tumores (25). Foi revelado por imunohistoquímica que diversos tumores sólidos e medula óssea leucêmica submetidos à angiogênese ativa apresentam forte marcação para endogлина no endotélio dos vasos tumorais (66). Assim, ENG representa um alvo importante para anti-angiogênese, visto que a expressão de endogлина é um marcador mais forte de ECs do que outros marcadores, como o receptor 2 do fator de crescimento endotelial vascular (VEGFR2), o fator de von Willebrand (vWF) ou as moléculas de adesão (ICAM, PECAM, VCAM) (67).

Outros estudos demonstram a importância da ENG para a homeostasia do organismo através da inibição do gene da endogлина em camundongos, em que embriões homocigotos *null* para endogлина (ENG^{-/-}) revelam um estágio embrionário precoce e letal devido a malformações cardiovasculares, com ruptura de vasos e falta de transição endotelial-mesenquimal (13–15). Foi evidenciado que mutações germinativas que levam a diminuição de função de ENG, configuram um distúrbio autossômico

dominante denominado telangiectasia hemorrágica hereditária tipo 1 ou síndrome de Osler-Weber-Rendu (HHT1, MIM 187300), devido a haploinsuficiência de endogлина ($ENG^{+/-}$) que reduz significativamente os níveis de ENG em 50%. A HHT1 é caracterizada por displasia vascular com frequentes sangramentos e afeta a mucosa nasal, trato gastrointestinal e a pele, além da apresentação de quadros mais graves devido a malformações vasculares (MAV) em pulmão, cérebro e fígado (3,15–17). Provavelmente a formação de lesões nesses tecidos se dá quando ENG haploinsuficiente é ativada na presença de algum estímulo externo, como inflamação, infecção, lesão vascular, isquemia ou trauma (25). Por outro lado, os camundongos heterozigotos para $ENG^{+/-}$ parecem ter um efeito protetor para o desenvolvimento do tumor, apresentando taxas de crescimento tumoral diminuídas e densidades capilares mais baixas em comparação com camundongos do tipo selvagem (68), o que reforça ainda mais o envolvimento de endogлина no crescimento do tumor e seu papel na angiogênese.

Demonstrou-se que a ENG é regulada pela fosforilação através de receptores TGF- β tipo I e II. Os efeitos exercidos pelo TGF- β resultam em um delicado equilíbrio entre as vias de sinalização que se iniciam por ALK5 e ALK1, este último sendo predominantemente endotelial (69). Após o reconhecimento do ligante intermediado pela endogлина, o receptor do tipo II transfosforila o receptor do tipo I, que por sua vez propaga o sinal através de proteína Smad regulada por receptor (R-Smad, Smad1, 2, 3, 5, 8) e, posteriormente, formam complexos com co-Smad4, translocando-se para o núcleo e regulando a transcrição de genes alvo (2,7) (Figura 3). É amplamente aceito que ocorre a inibição da proliferação e migração celular via TGF- β / ALK5 através da fosforilação de Smad 2 / 3 e promoção dessas respostas via TGF- β / ALK1 com a fosforilação de Smad 1 / 5 / 8 (2). Evidências apontam que a atividade regulada por ALK1 é enriquecida pela fosforilação de ALK5, sendo capaz de recrutar e ativar ALK1 (70). É importante ressaltar que ALK5 é o mediador predominante da sinalização de TGF- β em células endoteliais quiescentes, enquanto ALK1 é preferencialmente ativado durante a angiogênese (71). Por sua vez, a atuação de ENG é importante para a sinalização eficiente de TGF- β / ALK1, promovendo o estado de ativação das ECs durante a angiogênese e inibe indiretamente a via TGF- β / ALK5. Portanto, ENG mostra um papel crucial para a proliferação de ECs, visto que sua ausência anula a sinalização de ALK1, enquanto que ALK5 estaria aumentada (1,2) (Figura 3).

Figura 3 - Modulação da via de sinalização de TGF- β e BMP9 por endogлина em células endoteliais.



Fonte: Adaptado de Gregory *et al.*, 2014 (7).

Mais tarde foram sugeridos novos ligantes para a ENG, já compreendidos como ligantes de ALK1: BMP-9 e BMP-10 (39,72). Posteriormente foi mostrado que ENG e ALK1 se ligam com maior afinidade ao ligante BMP9 do que ao TGF β -1 na sinalização celular. Dessa maneira, a sinalização desencadeada por ENG / BMP9 / ALK1 leva a fosforilação de Smads 1 / 5 / 8 levando a expressão de genes alvo envolvidos na proliferação e migração celular (61,73) (Figura 3).

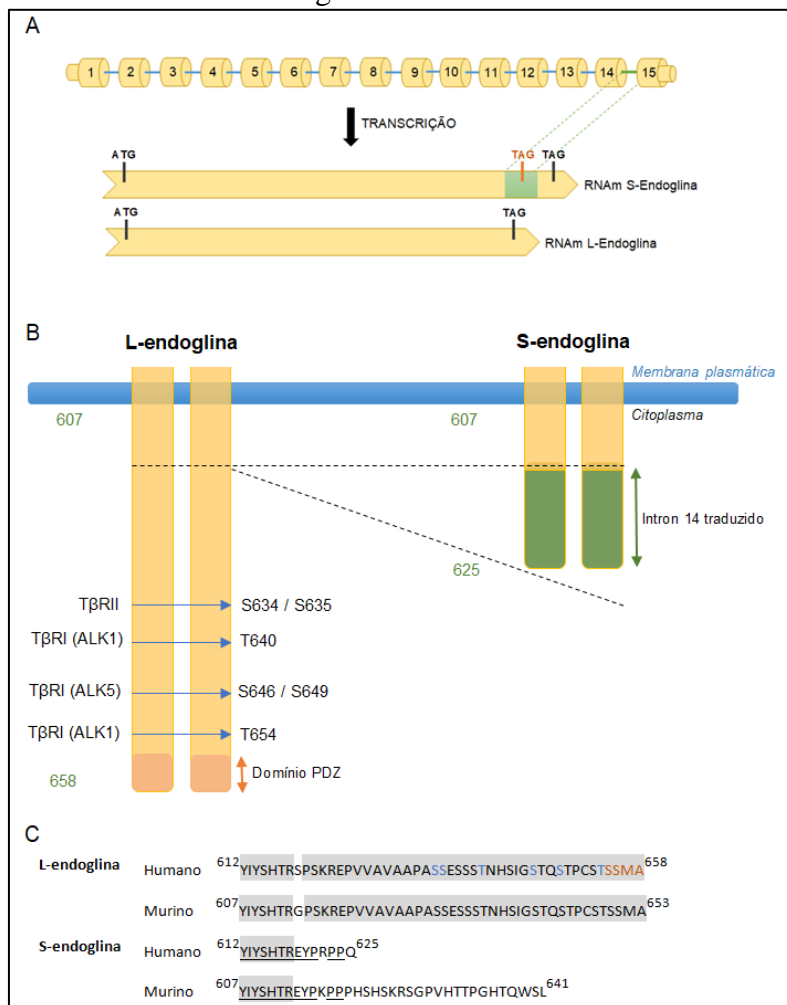
Achados recentes apontam a importância de ENG no contexto de processos angiogênicos além da sua atividade na perspectiva da família TGF- β , podendo impactar na regulação do comportamento das células formadoras de colônias endoteliais (ECFCs) através de sua ação como molécula envolvida na adesão celular (37).

4.1.2 Gene da endogлина e variantes genéticas

O gene que codifica a endogлина humana (*ENG*) fica localizado na região 34 do braço longo do cromossomo 9 (9q34) (74), tem cerca de 40kb e contém 15 *éxons* (regiões expressas, do inglês *expressed regions*), dos quais 13 codificam o domínio extracelular (16,75). Um estudo primário investigando a região promotora de *ENG* identificou que sua transcrição é conduzida por duas regiões ricas em GC (sequência consenso GGGCGG) permitindo a ligação a fatores de transcrição e apresentando diferentes locais de início de transcrição, como domínios para Sp1, ets, GATA, AP-2, NF- κ B e *Mad* (75).

A existência de isoformas de ENG em humanos (38) e em camundongos (43) foi evidenciada pelo estudo de DNA complementar (cDNA) de *ENG*, reveladas pela presença de uma estrutura longa e mais predominante, a L-endogлина, e uma variante curta, S-endogлина. Supõe-se que a S-endogлина seja gerada por *splicing* alternativo através da retenção do *intron* (sequência interveniente, do inglês *intragenic region*) 14 resultando em um RNAm alterado, apresentando diferente sequência de aminoácidos e a inclusão de um códon de parada dando origem a uma proteína mais curta (Figura 4A e Figura 4B). Assim, as isoformas se diferenciam em relação ao comprimento da região CT que, em humanos, é composta por 47 aminoácidos em L-endogлина e 14 aminoácidos em S-endogлина (38).

A S-endogлина não possui a sequência PDZ, bem como resíduos de Ser / Tre que servem de substrato para fosforilação por T β RI (ALK1 e ALK5) e T β RII (Figura 4B). Ambas isoformas humanas se mostraram expressas na membrana plasmática como homodímeros em células mieloides, endoteliais e placentárias (38). Além disso, foi mostrado, em estudo por transfecção, que as isoformas podem se associar formando uma molécula heterodimérica (L-S), o que evidencia a complexidade funcional de ENG (76). Modelos murinos transgênicos para superexpressão de S-endogлина exibiram um perfil angiogênico deficiente devido ao retardo do crescimento tumoral quando comparados a camundongos selvagens para endogлина (ENG^{+/+}) (76), semelhante aos efeitos observados em camundongos sem L-endogлина (77,78). Dessa forma, a molécula S-endogлина parece apresentar um comportamento antiangiogênico, diferindo da ação proangiogênica da L-endogлина, e essa produção de S-endogлина parece ser induzida durante a senescência endotelial, o que pode culminar em patologia vascular dependente da idade (25).

Figura 4 - Gene e isoformas de endogлина.

Nota: (A) Representação esquemática do gene da endogлина e os RNAm das duas isoformas ilustrando a retenção do intron 14 na S-endogлина. (B) Representação esquemática do domínio citoplasmático de isoformas de endogлина humana, L-endogлина e S-endogлина, ilustrando resíduos suscetíveis a fosforilação por receptores TGF-β em L-endogлина. (C) Sequência peptídica do domínio citoplasmático de isoformas de endogлина em humano e murino; em destaque na cor cinza estão representados os resíduos conservados, em azul os resíduos fosforilados, em laranja os resíduos do domínio PDZ e em sublinhado os aminoácidos em comum em S-endogлина nas espécies de humano e murino.

Fonte: Adaptado de Pérez Gomez *et al.* (43), López-Novoa *et al.* (25), Koleva *et al.* (44), Núñez-Gomez *et al.* (69).

Foi relatado, anteriormente, que mutações na ENG estão associados à HHT1. Essa patologia afeta 1 indivíduo a cada 5.000 a 10.000 indivíduos e esta estimativa é ainda superior em locais geograficamente isolados (3,17). Mais de 500 variantes genéticas em *ENG* associadas ao HHT1 são descritas em base de dados do ARUP Laboratories do Departamento de Patologia da Universidade de Utah (http://arup.utah.edu/database/ENG/ENG_welcome.php) e ENG database (<https://databases.lovd.nl/shared/variants/ENG>). A ocorrência de mutações germinativas em *ENG* também é investigada numa provável suscetibilidade a um distúrbio hereditário do câncer, a síndrome da polipose juvenil (JPS, MIM 174900) e ocorre

concomitantemente em menos de 2% dos pacientes com HHT (79), mas até então a sugestão de que mutações em *ENG* estejam associadas ao desfecho não foi confirmada e o papel de *ENG* ainda é questionado em relação à fisiopatologia da doença. Além de mutações em *ENG*, tanto a HHT como JPS são relacionadas a mutações nos genes que codificam proteínas envolvidas na sinalização da família TGF- β , como *ACVRL1* (1,16), *SMAD4* (16,79), *BMPRA1* e *PTEN* (80), sendo esses genes investigados concomitantemente nesses desfechos.

Em contrapartida, a superexpressão de *ENG* no endotélio vascular está relacionada à progressão tumoral, com presença de angiogênese induzida por tumores sólidos (1). Nesses casos, os anticorpos anti-endogлина são utilizados como alvo principal para terapia anti-angiogênica através do bloqueio funcional da via *ENG* / *BMP9* / *ALK1* (1,2,9). Por sua vez, a presença de mutações em epítomos reconhecidos por mAbs (42) (citados na seção 4.1.1.1) bem como o estado de ativação das ECs (1), podem gerar falta de reconhecimento por mAb anti-endogлина (42) e, consequentemente, uma possível falha terapêutica, devido a *ENG* estar relacionada a progressão maligna tumoral (1).

Em outro sentido, o gene da endogлина vem sendo um dos genes candidatos das vias da angiogênese para os estudos de associação das variantes genéticas com preeclâmpsia (19,22), devido à presença elevada de *ENG* no soro de mulheres com preeclâmpsia, sendo maior e mais significativa antes do início dos sintomas clínicos da condição (6,7). Entretanto, ainda há poucos estudos com esta abordagem direcionada para a endogлина.

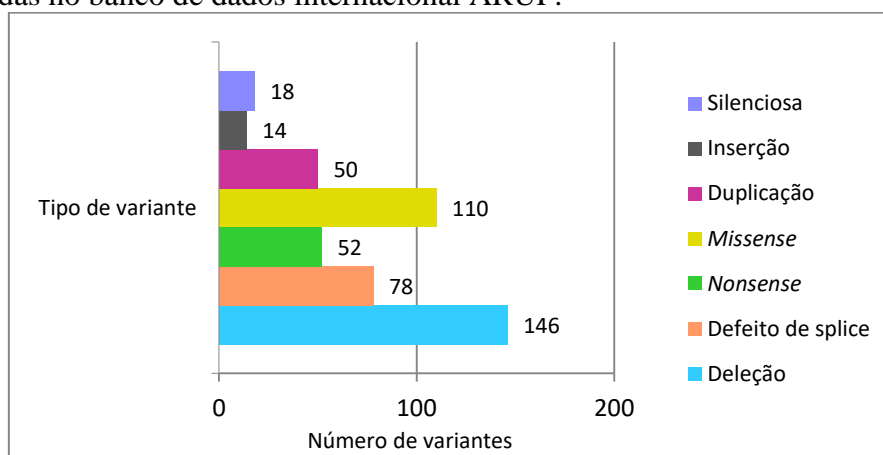
No contexto da HHT1, diversos estudos investigam a causa relacionada às mutações em *ENG* e acredita-se que as mutações patogênicas sejam decorrentes de alelos não funcionais, uma vez que *ENG* mutante é retida intracelularmente devido ao dobramento prejudicado da molécula (3,33,81). Diante disso, as variantes genéticas conhecidas como variação de nucleotídeo único (do inglês *SNV*, *single nucleotide variation*), uma variação na sequência de DNA que afeta somente uma base de nucleotídeo, particularmente aquelas que podem levar à alteração única de aminoácido, chamadas mutações de sentido trocado ou de sentido inverso (em inglês: *missense*), são alvos para a compreensão do impacto funcional na molécula de *ENG* (3,81). As variantes *missense* constituem mais de 50% das mutações que conhecidamente estão

envolvidas em doenças hereditárias (82). De acordo com o banco de dados internacional de mutação em *ENG* da rede ARUP Laboratories, 23,50% das mutações são *missense* (Figura 5) e dessas, mais da metade tem diagnóstico indefinido ou incompleto (http://arup.utah.edu/database/ENG/ENG_welcome.php).

A grande maioria das mutações *missense* em *ENG* associadas ao HHT1 estão presentes no ectodomínio da molécula (3,33) e isto é atribuível devido ao seu tamanho muito maior em relação ao domínio intracelular (83), ocorrendo entre os éxons 1 e 9, enquanto que as mutações investigadas em JPS ocorrem nos éxons 11 e 12 (1). No entanto, mais de 15% dos casos de HHT não apresentam mutações em regiões codificadoras ou nas junções éxon/intron nos genes relacionados à doença (84). As mutações em regiões não codificadoras, conhecidas como introns, e em regiões reguladoras, já foram descritas e associadas ao HHT1. Nesses casos, a regulação transcricional pode ser afetada e impedir a ligação de fatores de transcrição, podendo interferir na expressão gênica e molecular da proteína, resultando no desfecho patogênico (85,86).

Considerando que a endogлина apresenta papel fundamental para a homeostase e fisiologia do organismo e que variantes genéticas podem implicar em desfechos fisiopatológicos, o presente estudo propôs descrever a frequência de variantes genéticas em uma amostra populacional brasileira e analisar a frequência de populações pseudoancestrais, descrever as características funcionais e regulatórias de variantes genéticas presentes na população brasileira através de análise funcional *in silico*, bem como descrever as variantes genéticas encontradas na população brasileira que já foram relatadas na literatura e que podem estar potencialmente envolvidas em processos patogênicos em outras populações.

Figura 5 - Quantidade absoluta das variantes genéticas do gene da endogлина encontradas no banco de dados internacional ARUP.



Fonte: Adaptado de ARUP Laboratories (http://arup.utah.edu/database/ENG/ENG_welcome.php).

5 MATERIAIS E MÉTODOS

5.1 População do estudo

Este estudo foi realizado com o uso de bancos de dados genômicos de duas coortes desenvolvidas no Estado da Bahia, localizado no Nordeste do Brasil: Social Changes, Asthma and Allergy in Latin America Programme (SCAALA) e Programa para o Controle da Asma e Rinite Alérgica no Estado da Bahia (ProAR). O Programa do SCAALA compreendeu 1309 crianças residentes da cidade de Salvador, com idade entre 4 e 11 anos, sem relação sanguínea e foi aprovado pelo Comitê de Ética do Instituto Nacional de Ciências da Saúde da Universidade Federal da Bahia (registro 003-05) e pela Comissão Nacional de Ética em Pesquisa (Resolução No. 15895/2011).

O desenho do estudo foi relatado em outros trabalhos (87,88). O termo de consentimento livre e esclarecido foi obtido junto ao responsável legal ou aos pais de cada criança. O programa ProAR incluiu 1170 indivíduos com idade superior a 18 anos, sem relação sanguínea, e foi aprovado pelo Comitê de Ética da Maternidade Climério de Oliveira (Resolução No. 099/2009) e pela Comissão Nacional de Ética em Pesquisa (Resolução nº 15782 / 2010), tendo os termos de consentimento aplicados a cada indivíduo incluído no estudo.

5.2 Extração de DNA e genotipagem

A extração e a genotipagem de DNA foram realizadas anteriormente nos diferentes projetos citados. A extração de DNA foi realizada a partir de amostras de sangue através do Gentra Puregene Blood Kit (Qiagen; Hilden, Alemanha) de acordo com as instruções do fabricante. O programa SCAALA realizou a genotipagem das amostras através do chip *Illumina BeadChip HumanOmni 2.5-8 Kit BeadChip*, uma ampla plataforma de varredura de genoma com 2,5 milhões de marcadores genotipados pelo projeto EPIGEN-Brazil Consortium (<https://epigen.grude.ufmg.br>) (89). O programa ProAR utilizou o chip *Illumina Multi-Ethnic AMR / AFR-8*. Para este estudo, a informação genética foi extraída do cromossomo 9, localização genômica NC_000009.12 (130577291-130617052) (GRCh37.p13).

Após a extração da informação genética, foi realizado um controle de qualidade através do Software Plink 1.90 (90) para excesso de dados genotípicos faltantes ($\text{mind} > 0,01$ e $\text{geno} > 0,01$). O coeficiente de Desequilíbrio de Ligação (LD, *linkage disequilibrium*) foi calculado utilizando o software Haploview 4.2 (91), com base no valor do quadrado da correlação (r^2) entre as variantes genéticas em diferentes loci (92), utilizando o valor de $r^2 \geq 0,8$ para alta correlação (93).

5.3 Marcadores de ancestralidade

Para analisar a variação da frequência alélica entre as populações que deram origem à população brasileira, foi empregada uma análise através do diferencial de frequências alélicas (δ) que permitem a identificação de marcadores genéticos informativos de ancestralidade (do inglês *Ancestry Informative Markers*, AIMs). Os AIMs se referem a grandes variações de diferencial de frequências alélicas ($\delta \geq 0,3$) entre duas populações de origens geográficas distintas (97,98). A implicação dos AIMs é importante para a estimativa da ancestralidade biogeográfica, pois fornece dados de proporções do genoma individual que foram herdadas de cada uma das populações ancestrais envolvidas no processo de miscigenação (99). Para obter a frequência de alelos das populações de origem, foi utilizada a plataforma de dados do International Genome Sample Resource (IGSR) (<http://www.internationalgenome.org>), através de dados do 1000 genomas no Ensembl, um navegador de genomas, utilizando a referência de genoma anterior (GRCh37p.13).

Os principais grupos étnicos que deram origem à população brasileira foram africanos, europeus e americanos nativos (ameríndios). Dessa forma, foram utilizados dados de descendentes contemporâneos de populações de origens aproximadas, ditos pseudoancestrais, como, indivíduos residentes de Utah nos Estados Unidos, que possuem ancestralidade europeia setentrional e ocidental (CEU), indivíduos Iorubas de Ibadan, Nigéria (YRI) e Chineses da etnia Han de Pequim, China (CHB) (94). Tendo em vista os indícios do envolvimento da população ancestral de CHB na formação da população ameríndia (95), bem como os dados que indicam semelhança das frequências alélicas entre essas populações (96), a amostra CHB foi inserida como referência para os ameríndios.

5.4 Análise funcional *in silico*

Análise funcional *in silico* foram conduzidas para cada variante genética, no caso variação de nucleotídeo único (do inglês *Single Nucleotide Variation*, SNV) presentes em *ENG*. Todas as variantes genéticas usadas nesse estudo estão catalogadas em bancos de dados públicos, como o repositório de polimorfismos dbSNP (<http://www.ncbi.nlm.nih.gov/SNP/>).

O banco de dados RegulomeDB (<http://www.regulomedb.org>) foi utilizado para a interpretação de funções regulatórias de regiões intergênicas no genoma humano. Este banco utiliza um conjunto de dados do ENCODE (enciclopédia de elementos de DNA) e outras fontes. RegulomeDB possibilita identificar um suposto potencial regulatório e variantes funcionais através de um sistema de pontuação, variando de 1 a 6. As variantes com escores de 1a a 1f indicam que provavelmente podem afetar a ligação e a expressão de um alvo genético, já escores de 2a a 2c indicam que provavelmente afeta a ligação, enquanto que 3a e 3b indicam menor probabilidade de afetar a ligação, e as pontuações de 4, 5 e 6 indicam evidências mínimas de ligação ou nenhum dado disponível (100). Assim, pontuações mais baixas estão associadas ao aumento da evidência de que a variante esteja localizada em uma região funcional.

Softwares abertos foram utilizados para a predição do impacto da substituição de aminoácidos na estabilidade e função das proteínas: Polimorphism Phenotyping v2 (PolyPhen-2, <http://genetics.bwh.harvard.edu/pph2/>) caracteriza uma substituição de aminoácidos como "Benigno", "possivelmente prejudicial" ou "provavelmente prejudicial" (101); Sort Intolerant from Tolerant (SIFT, <https://sift.bii.a-star.edu.sg/>)

prevê se uma alteração é "Tolerado" ou "prejudicial" (102); MutationTaster2 software (<http://www.mutationtaster.org/>) prevê uma alteração como "causador de doença" (ou seja, provavelmente deletério), "causador de doença automática" (ou seja, sabidamente deletério), "polimorfismo" (ou seja, provavelmente inofensivo), "Polimorfismo automático" (ou seja, conhecido por ser inofensivo) (103); PMut (<http://mmb.irbbarcelona.org/PMut/analyses/new/>) fornece uma previsão binária para "neutro" ou "patológico" (104). As mutações foram classificadas como patogênicas se dois ou mais softwares concordassem entre si, sendo adotado um escore de 1 a 4, método descrito por Pousada *et al.* 2016 (105). Variantes polimórficas com pontuação igual ou superior a dois foram analisadas através do software HOPE (<https://www3.cmbi.umcn.nl/hope/>). O HOPE analisa os efeitos estruturais de uma mutação pontual em uma sequência de proteínas, fornecendo uma análise 3D da estrutura da proteína.

6 RESULTADOS

6.1 Descrição da frequência de variantes genéticas no gene da endogлина na população estudada

Após a extração da informação genética, foi aplicado o filtro de controle de qualidade, tendo duas variantes removidas pelo HWE e duas variantes pelo excesso de dados genotípicos faltantes, totalizando 85 variantes genéticas incluídas neste estudo. A Tabela 3 e a Tabela 4 detalham as variantes genéticas. Um total de 50 variantes polimórficas possuem MAF entre 0% e 50% e apresentam como consequência: 39 intrônicos, 7 *missense* e 4 sinônimos (Figura 6A). As variantes monomórficas, aquelas que apresentam 100% do alelo referência ou ancestral, correspondem a 32 variantes, evidenciando como consequência: 5 intrônicos, 25 *missense*, 1 sinônimo e 1 acceptor de splice (Figura 6B). Ambos os bancos de dados exibem variantes genéticas comuns (rs10987746, rs10121110, rs11545664, rs12379490, rs17557600, rs41421145, rs116330805), marcadas em negrito na Tabela 3 e Tabela 4.

A maioria das variações que ocorrem estão localizadas no intron 1, seguido pelo intron 2 e intron 12, enquanto as variantes encontradas nos éxons apresentaram uma distribuição de 1 a 3 SNVs por éxon, tendo o éxon 12 a maior quantidade, incluindo uma variante sinônima (Figura 6C).

Tabela 3 - Sumário das variantes genéticas em ENG extraídas do banco de dados SCAALA.

rs ID	Mudança de nucleotídeo	Mudança de aminoácido	Consequência da variante	Localização	Alelos ^a	MAF ^b	RegulomeDB escore ^c
rs10987746	c.1686+306A>G		Intron	Intron 12	C/T	0,48	4
rs1998923	c.67+4663C>T		Intron	Intron 1	A/G	0,46	4
rs10121110	c.219+2965C>T		Intron	Intron 2	G/A	0,44	5
rs10760503	c.1742-72T>C		Intron	Intron 13	A/G	0,44	4
rs11789185	c.68-1994A>C		Intron	Intron 1	G/T	0,42	4
rs4837192	c.67+1656C>T		Intron	Intron 1	A/G	0,38	3a
rs11794565	c.219+4096G>A		Intron	Intron 2	C/T	0,31	4
rs10819309	c.1311+178T>C		Intron	Intron 10	G/A	0,30	4
rs41492548	c.68-1835C>T		Intron	Intron 1	G/A	0,27	2b
rs41432051	c.219+3143C>T		Intron	Intron 2	G/A	0,24	5
rs1330684	c.1687-117G>A		Intron	Intron 12	C/T	0,21	5
rs16930129 (rs11545664)	c.207G>A	p.Leu69=	Sinônimo	Éxon 2	C/T	0,21	
rs12551892	c.68-2188A>G		Intron	Intron 1	T/C	0,20	2b
rs12379490	c.67+4350T>G		Intron	Intron 1	A/C	0,18	4
rs12001427	c.1134+1552A>G		Intron	Intron 8	T/C	0,17	6
rs11792480	c.220-6019C>T		Intron	Intron 2	G/A	0,16	4
rs41429144	c.1742-293C>T		Intron	Intron 13	G/A	0,10	4
rs17557600	c.220-5117A>G		Intron	Intron 2	T/C	0,09	6
rs5031024	c.1687-422T>C		Intron	Intron 12	T/C	0,08	5
rs36092484	c.1060C>T	p.Leu354=	Sinônimo	Éxon 8	G/A	0,07	
rs41421145	c.67+4109T>C		Intron	Intron 1	A/G	0,07	5
rs11789912	c.68-215T>C		Intron	Intron 1	A/G	0,06	5
rs41459245	c.220-6194T>C		Intron	Intron 2	A/G	0,05	4
rs3739817	c.1029C>T	p.Thr343=	Sinônimo	Éxon 8	G/A	0,04	
rs41358947	c.1794T>C	p.Gly598=	Sinônimo	Éxon 14	A/G	0,03	
rs41505655	c.68-256A>G		Intron	Intron 1	T/C	0,02	
rs12553394	c.360+1134C>T		Intron	Intron 3	G/A	0,02	2b
rs41527247	c.67+1350C>T		Intron	Intron 1	G/A	0,02	3a
rs2900256	c.68-4921A>G		Intron	Intron 1	C/T	0,02	5
rs116397503	c.68-4033G>A		Intron	Intron 1	C/T	0,01	4
rs41336944	c.219+5693A>C		Intron	Intron 2	T/G	0,008	5
rs115450389	c.1452C>T	p.Ser484=	Sinônimo	Éxon 12	G/A	0,008	
rs77946642	c.361-1485G>A		Intron	Intron 3	C/T	0,008	2b
rs41316968	c.1686+317A>G		Intron	Intron 12	T/C	0,006	4
rs116330805	c.1510G>A	p.Val504Met	Missense	Éxon 12	C/T	<0,001	
rs199919128	c.360+435A>G		Intron	Intron 3	T	0	5
rs117415442	c.220-3421C>T		Intron	Intron 2	G	0	
rs117253420	c.67+305C>T		Intron	Intron 1	G	0	4

Nota: ^aO primeiro é o alelo de referência e o segundo é o menor alelo; ^bAlelo de menor frequência; ^cA pontuação do RegulomeDB é calculada para variantes em regiões não codificantes. Em negrito estão as variantes genéticas em comum entre os bancos de dados. Os espaços em branco referem-se à falta de informações de dados.

Tabela 4 - Sumário das variantes genéticas em ENG extraídas do banco de dados ProAR.

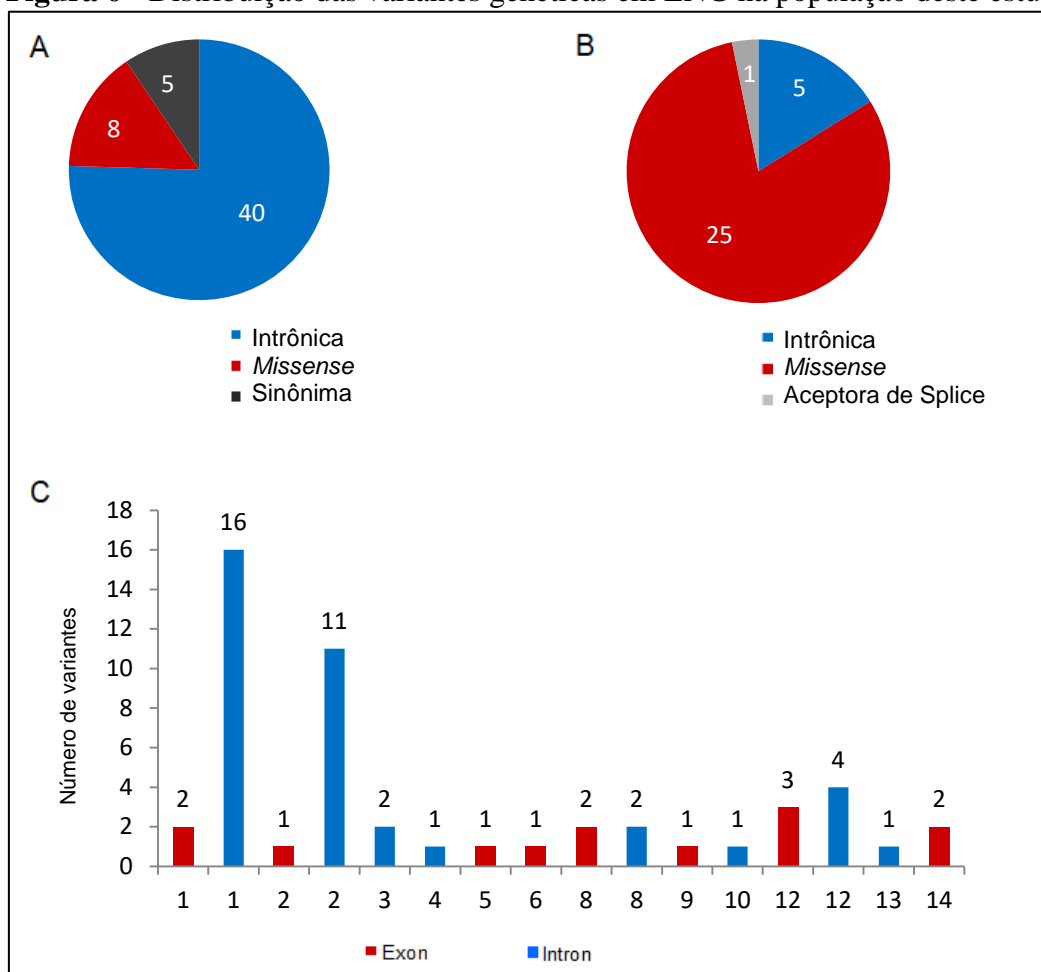
rs ID	Mudança de nucleotídeo	Mudança de aminoácido	Consequência da variante	Localização	Alelos ^a	MAF ^b	RegulomeDB escore ^c
rs10987746	c.1686+306A>G		Intron	Intron 12	C/T	0,50	4
rs10121110	c.219+2965C>T		Intron	Intron 2	G/A	0,43	5
rs4836585	c.67+4185G>T		Intron	Intron 1	A/C	0,43	5
rs10115613	c.67+817C>T		Intron	Intron 1	A/G	0,39	5
rs2417056	c.219+6449C>T		Intron	Intron 2	G/A	0,32	5

Continuação da tabela 4:

rs34116890	c.1135-858T>C		Intron	Intron 8	A/G	0,23	3a
rs16930129 (rs11545664)	c.207G>A	p.Leu69=	Sinônimo	Éxon 2	C/T	0,21	
rs12379490	c.67+4350T>G		Intron	Intron 1	A/C	0,18	4
rs17557600	c.220-5117A>G		Intron	Intron 2	T/C	0,09	6
rs35400405	c.14C>T	p.Thr5Met	<i>Missense</i>	Éxon 1	G/A	0,09	
rs41421145	c.67+4109T>G		Intron	Intron 1	A/G	0,08	5
rs41331249	c.1742-490C>T		Intron	Intron 13	G/A	0,05	4
rs141664673	c.68-1473G>A		Intron	Intron 1	C/T	0,04	5
rs41512447	c.219+1221C>T		Intron	Intron 2	G/A	0,03	4
rs41409546	c.524-30G>A		Intron	Intron 4	C/T	0,03	4
rs116472935	c.67+3286G>A		Intron	Intron 1	C/T	0,02	5
rs139334561	c.7C>G	p. Arg3Cys	<i>Missense</i>	Éxon 1	G/A	0,006	
rs7847860	c.219+25G>T		Intron	Intron 2	C/A	0,002	5
rs116330805	c.1510G>A	p.Val504Met	<i>Missense</i>	Éxon 12	C/T	<0,001	
rs563407740	c.219+863A>G		Intron	Intron 2	A/G	<0,001	
rs143724056	c.1258A>G	p.Met420Val	<i>Missense</i>	Éxon 9	A/G	<0,001	
rs148002300	c.1844C>T	p.Ser615Leu	<i>Missense</i>	Éxon 14	G/A	<0,001	
rs200372420	c.694C>T	p.Arg232Trp	<i>Missense</i>	Éxon 6	G/A	<0,001	
rs201393380	c.617G>C	p.Gly206Ala	<i>Missense</i>	Éxon 5	C/G	<0,001	
rs141330288	c.1447G>A	p.Val483Ile	<i>Missense</i>	Éxon 12	G/A	<0,001	
rs118179851	c.1135-945G>C		Intron	Intron 8	G	0	3a
rs138799379	c.1712G>A	p.Arg571His	<i>Missense</i>	Éxon 13	G	0	
rs139398993	c.392C>T	p.Pro131Leu	<i>Missense</i>	Éxon 4	G	0	
rs140124489	c.220-4568C>T		Intron	Intron 2	G	0	6
rs142896669	c.1633G>A	p.Gly545Ser	<i>Missense</i>	Éxon 12	G	0	
rs143054595	c.1133C>T	p.Ala378Val	<i>Missense</i>	Éxon 8	G	0	
rs147792018	c.565G>T	p.Asp189Tyr	<i>Missense</i>	Éxon 5	G	0	
rs150932144	c.640G>A	p.Gly214Ser	<i>Missense</i>	Éxon 5	G	0	
rs1800956	c.1096G>C	p.Asp366His	<i>Missense</i>	Éxon 8	C	0	
rs199840979	c.388C>T	p.Pro130Ser	<i>Missense</i>	Éxon 4	G	0	
rs200960408	c.218C>T	p.Thr73Met	<i>Missense</i>	Éxon 2	G	0	
rs201768056	c.1762G>C	p.Val588Leu	<i>Missense</i>	Éxon 14	G	0	
rs202048202	c.69T>C	p.Ser23=	Sinônimo	Éxon 2	A	0	
rs369997021	c.1274C>T	p.Ala425Val	<i>Missense</i>	Éxon 10	G	0	
rs373842615	c.1273-2A>G		Aceptor de splice	Intron 9	A	0	2b
rs41322046	c.572G>A	p.Gly191Asp	<i>Missense</i>	Éxon 5	G	0	
rs548424658	c.596G>C	p.Arg199Pro	<i>Missense</i>	Éxon 5	G	0	
rs548690138	c.1093G>A	p.Asp365Asn	<i>Missense</i>	Éxon 8	G	0	
rs756604495	c.602C>A	p.Pro201Gln	<i>Missense</i>	Éxon 5	C	0	
rs762209698	c.613C>T	p.Arg205Trp	<i>Missense</i>	Éxon 5	G	0	
rs763475207	c.1469T>C	p.Leu490Ser	<i>Missense</i>	Éxon 12	A	0	
rs765564489	c.1538A>G	p.Lys513Arg	<i>Missense</i>	Éxon 12	A	0	
rs768491175	c.1655C>T	p..Ala552Val	<i>Missense</i>	Éxon 12	G	0	
rs781374732	c.224C>T	p.Pro75Leu	<i>Missense</i>	Éxon 3	G	0	
rs866067907	c.1914C>A	p.Ser638Arg	<i>Missense</i>	Éxon 15	C	0	
rs866646256	c.1973C>A	p.Ala658Glu	<i>Missense</i>	Éxon 15	C	0	
rs868296721	c.1781G>A	p.Gly594Asp	<i>Missense</i>	Éxon 14	G	0	
rs868572062	c.1972G>C	p.Ala658Pro	<i>Missense</i>	Éxon 15	G	0	
rs868730895	c.1058A>G	p.Glu353Gly	<i>Missense</i>	Éxon 8	A	0	

Nota: ^aO primeiro é o alelo de referência e o segundo é o menor alelo; ^bAlelo de menor frequência; ^cA pontuação do RegulomeDB é calculada para variantes em regiões não codificantes. Em negrito estão as variantes genéticas em comum entre os bancos de dados. Os espaços em branco referem-se à falta de informações de dados.

Figura 6 - Distribuição das variantes genéticas em ENG na população deste estudo.



Nota: (A) Consequência de variantes genéticas polimórficas. (B) Consequência de variantes genéticas monomórficas. (C) Variantes genéticas polimórficas distribuídas em *ENG* de acordo com suas localizações em íntrons e éxons.

6.2 Distribuição da frequência de variantes genéticas no gene da endogлина em populações pseudoancestrais

Os marcadores genéticos encontrados na população brasileira no gene da *ENG* foram analisados para o diferencial entre as frequências alélicas (δ) de populações pseudoancestrais de africanos (YRI), europeus (CEU) e ameríndios (CHB), possibilitando identificar AIMs. Os dados foram apresentados na Tabela 5 e na Tabela 6. Das 85 variantes genéticas extraídas dos bancos SCAALA e ProAR, somente 13 se caracterizam como AIMs, dentre esses, 1 AIM distingue para as três populações pseudoancestrais (rs10115613), 8 distinguem YRI para dois pseudoancestrais, CEU ou CHB (rs1998923, rs10121110, rs11789185, rs4837192, rs11794565, rs1330684, rs4836585 e rs2417056), 2 distinguem CEU de YRI ou CHB (rs11792480 e

rs34116890), 1 distingue CHB de CEU (rs10819309) e 1 distingue CHB de YRI (rs10760503).

Tabela 5 – Diferencial de frequências alélicas entre três populações diferentes do Projeto 1000 Genomas para os marcadores genéticos do banco de dados SCAALA.

rs ID	Alelos ^a	Alelo referência ^a	Frequência do alelo referência ^a			δ^b		
			YRI	CEU	CHB	YRI/CEU	YRI/CHB	CEU/CHB
rs10987746	T/C	C	0.431	0.439	0.665	0.008	0.234	0.226
rs1998923	G/A	G	0.764	0.126	0.155	0.638	0.609	0.029
rs10121110	G/A	G	0.704	0.399	0.107	0.305	0.597	0.292
rs10760503	A/G	G	0.333	0.419	0.665	0.086	0.332	0.246
rs11789185	T/G	T	0.722	0.106	0.087	0.616	0.635	0.019
rs4837192	G/A	G	0.616	0.121	0.16	0.495	0.456	0.039
rs11794565	C/T	C	0.852	0.525	0.408	0.327	0.444	0.117
rs10819309	A/G	G	0.819	0.54	0.937	0.279	0.118	0.397
rs41492548	G/A	G	0.616	0.808	0.641	0.192	0.025	0.167
rs41432051	G/A	G	0.648	0.874	0.922	0.226	0.274	0.048
rs1330684	C/T	C	0.991	0.652	0.447	0.339	0.544	0.205
rs16930129 (rs11545664)	C/T	C	0.759	0.889	0.922	0.130	0.163	0.033
rs12551892	T/C	T	0.759	0.909	0.922	0.150	0.163	0.013
rs12379490	A/C	A	0.782	0.899	1	0.117	0.218	0.101
rs12001427	T/C	C	0.245	0.051	0.097	0.194	0.148	0.046
rs11792480	G/A	G	0.995	0.662	0.990	0.333	0.005	0.328
rs41429144	G/A	G	0.838	1	1	0.162	0.162	0
rs17557600	T/C	T	0.852	0.944	0.665	0.092	0.187	0.279
rs5031024	A/G	A	0.935	0.990	0.903	0.055	0.032	0.087
rs36092484	G/A	G	0.940	0.985	1	0.045	0.060	0.015
rs41421145	A/G	A	0.847	0.985	1	0.138	0.153	0.015
rs11789912	A/G	A	0.944	0.985	0.990	0.041	0.046	0.005
rs41459245	A/G	G	0.097	0	0	0.097	0.097	0
rs3739817	G/A	G	0.944	0.934	0.981	0.010	0.037	0.047
rs41358947	A/G	A	0.944	1	1	0.056	0.056	0
rs41505655	T/C	T	0.921	1	1	0.079	0.079	0
rs12553394	G/A	G	0.986	0.960	1	0.026	0.014	0.040
rs41527247	G/A	G	0.935	1	1	0.065	0.065	0
rs2900256	T/C	C	0.944	1	1	0.056	0.056	0
rs116397503	C/T	C	0.949	1	1	0.051	0.051	0
rs41336944	T/G	T	0.991	1	1	0.009		
rs115450389	G/A	G	0.968	1	1	0.032	0.032	0
rs77946642	C/T	C	0.968	1	1	0.032	0.032	0
rs41316968	T/C	T	1	0.960	1	0.040	0	0.040
rs116330805	C/T	C	0.968	1	1	0.032	0.032	0
rs199919128								
rs117415442	G/A	G	1	0.995	0.937	0.005	0.063	0.058
rs117253420	G/A	G	1	1	0.917	0	0.083	0.083

Nota: ^aDados provenientes da base de dados do Projeto 1000 Genomas; ^bDiferença entre frequências alélicas de populações da base de dados do Projeto 1000 Genomas. Em negrito estão as variantes genéticas em comum entre os bancos de dados do SCAALA e do ProAR. Os espaços em branco referem-se à falta de informações de dados.

Tabela 6 - Diferencial de frequências alélicas entre três populações diferentes do Projeto 1000 Genomas para os marcadores genéticos do banco de dados ProAR.

rs ID	Alelos ^a	Alelo referência ^a	Frequência do alelo referência ^a			δ^b		
			YRI	CEU	CHB	YRI/CEU	YRI/CHB	CEU/CHB
rs10987746	T/C	C	0.431	0.439	0.665	0.008	0.234	0.226
rs10121110	G/A	G	0.704	0.399	0.107	0.305	0.597	0.292
rs4836585	C/A	C	0.699	0.121	0.087	0.578	0.612	0.034
rs10115613	G/A	G	0.639	0.121	1	0.518	0.361	0.879
rs2417056	G/A	A	0.481	0.111	0.029	0.37	0.452	0.082
rs34116890	A/G	A	0.991	0.525	0.932	0.466	0.059	0.407
rs16930129 (rs11545664)	C/T	C	0.759	0.889	0.922	0.130	0.163	0.033
rs12379490	A/C	A	0.782	0.899	1	0.117	0.218	0.101
rs17557600	T/C	T	0.852	0.944	0.665	0.092	0.187	0.279
rs35400405	G/A	G	0.856	0.985	1	0.129	0.144	0.015
rs41421145	A/G	A	0.847	0.985	1	0.138	0.153	0.015
rs41331249	G/A	G	0.944	1	1	0.056	0.056	0
rs141664673	C/T	C	0.955	0.939	0.995	0.016	0.04	0.056
rs41512447	G/A	G	0.907	1	1	0.093	0.093	0
rs41409546	C/T	C	0.944	1	1	0.056	0.056	0
rs116472935	C/T	C	0.977	1	1	0.023	0.023	0
rs139334561	G/A	G	-	-				
rs7847860	C/A	C	0.931	0.924	1	0.007	0.069	0.076
rs116330805	C/T	C	0.968	1	1	0.032	0.032	0.000
rs563407740	T/C	T	0.972	0.985	0.990	0.013	0.018	0.005
rs143724056	T/C	T	1	1	1	0	0	0
rs148002300								
rs200372420	G/A/C	G	1	1	1	0	0	0
rs201393380								
rs141330288	C/T	C	1	1	1	0	1	0
rs118179851	C/G	C	1	1	0.976	0	0.024	0.024
rs138799379	C/T	C	1	1	1	0	0	0
rs139398993	G/A	G	1	0.995	1	0.005	0	0.005
rs140124489	G/A	G	1	1	0.966	0	0.034	0.034
rs142896669	C/T	C	1	1	1	0	0	0
rs143054595								
rs147792018								
rs150932144	C/T	C	1	1	1	0	0	0
rs1800956	C/G	C	1	0.99	0.903	0.01	0.097	
rs199840979								
rs200960408	G/A	G	1	1	1	0	0	0
rs201768056								
rs202048202	A/G	A	1	1	0.995	0	0.005	0.005

Continuação da tabela 6:

rs369997021								
rs373842615								
rs41322046	C/T	C	1	1	1	0	0	0
rs548424658	C/G/T	C	1	1	1	0	0	0
rs548690138	C/T	C	1	1	1	0	0	0
rs756604495								
rs762209698								
rs763475207								
rs765564489								
rs768491175								
rs781374732								
rs866067907								
rs866646256								
rs868296721								
rs868572062								
rs868730895								

Nota: ^aDados provenientes da base de dados do Projeto 1000 Genomas; ^bDiferença entre frequências alélicas de populações da base de dados do Projeto 1000 Genomas. Em negrito estão as variantes genéticas em comum entre os bancos de dados do SCAALA e do ProAR. Os espaços em branco referem-se à falta de informações de dados.

6.3 Análise funcional *in silico* de variantes genéticas no gene da endoglin

As pontuações de variantes relacionadas ao RegulomeDB em regiões não codificantes são apresentadas na Tabela 2 e na Tabela 3. As pontuações mais baixas, 2b (rs41492548, rs12551892, rs12553394, rs77946642, rs373842615) e 3a (rs4837192, rs34116890, rs41527247, rs118179851), indicam as possíveis variantes que podem estar envolvidas em regiões reguladoras e funcionais, enquanto valores mais altos assumem que as variantes não estão diretamente envolvidas nos processos funcionais.

As variantes *missense* foram analisadas através de diferentes *softwares* para predição da patogenicidade e estão apresentadas na Tabela 7. Das 33 variantes analisadas, 10 têm um escore que indica patogenicidade, e dessas, apenas 2 variantes são polimórficas (rs201393380 e rs200372420, ambas possuem MAF entre 0% e 1%). Essas duas variantes, rs201393380 e rs200372420, foram analisadas usando a plataforma HOPE para prever o impacto estrutural na proteína (mostrado na Figura 7). O rs201393380 corresponde à alteração de aminoácidos Gly206Ala (G206A). O software HOPE prevê que o resíduo mutante é maior e mais hidrofóbico que o resíduo do tipo selvagem, localizado em uma região (OR1, mostrado na Figura 8) necessária

para a interação com o ligante BMP9, e tal alteração pode abolir a flexibilidade conferida pela glicina, podendo perturbar a estrutura deste domínio. O rs200372420 corresponde à alteração de aminoácidos Arg232Trp (R232W). A previsão para esta troca é de que o resíduo mutante é maior e tem carga positiva, enquanto o resíduo do tipo selvagem tem carga neutra. Essa alteração também ocorre em uma região (OR1, mostrada na Figura 8) necessária para a interação com o ligante BMP9, a variante fica localizada na superfície da proteína, podendo interromper as interações com outras moléculas ou outras partes da proteína.

Em relação ao LD, as seguintes variantes extraídas do banco SCAALA apresentam valores que indicam alta correlação entre si: rs115450389 e rs77946642 ($r^2 = 1,0$), rs11545664 e rs12551892 ($r^2 = 0,92$), rs11545664 e rs12379490 ($r^2 = 0,83$). As variantes rs12551892 e rs12379490 ($r^2 = 0,77$), rs11789185 e rs1998923 ($r^2 = 0,76$), rs1998923 e rs4837192 ($r^2 = 0,70$), rs10760503 e rs10987746 ($r^2 = 0,70$), apresentam média correlação (Figura 9). As variantes extraídas do banco ProAR, rs4836585 e rs10115613 apresentam alta correlação ($r^2 = 0,80$), já as variantes rs11545664 e rs12379490 apresentam média correlação ($r^2 = 0,78$) (Figura 10).

Tabela 7 - Variantes missense encontradas em ENG na população de estudo e sua classificação segundo algoritmos computacionais (SIFT, PolyPhen-2, MutationTaster2 e PMut).

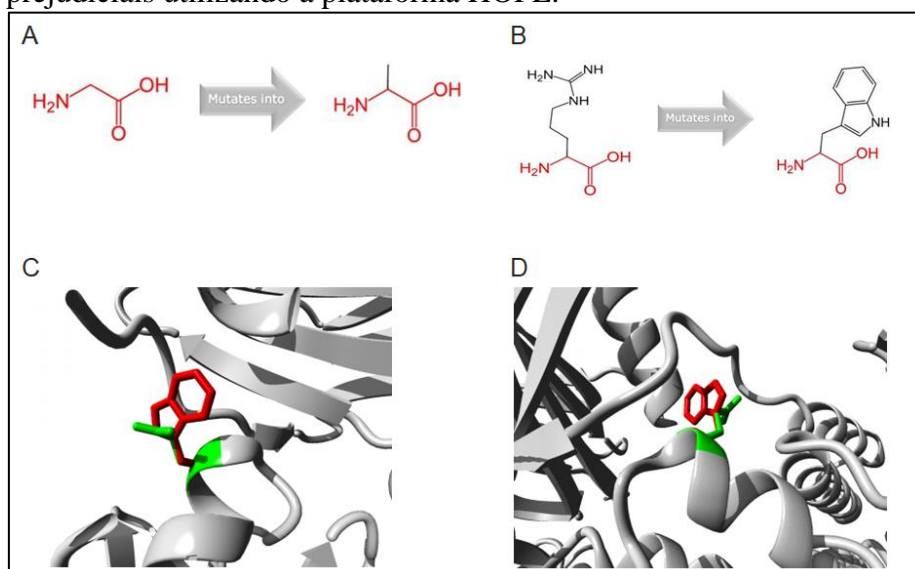
rs ID	Éxon	Mudança de nucleotídeo	Mudança de aminoácido	Alelos ^a	MAF ^b	Domínio da proteína	SIFT	PolyPhen2	MutationTaster2	PMut	Score ^c
rs35400405	1	c.14C>T	p.Thr5Met	G/A	0,09	SP	Tolerado	Benigno	Polimorfismo	Neutro	0
rs139334561	1	c.7C>G	p.Arg3Cys	G/A	0,006	SP	Prejudicial	Benigno	Polimorfismo	Neutro	1
rs201393380	5	c.617G>C	p.Gly206Ala	C/G	<0,001	OR1	Prejudicial	Benigno	Polimorfismo	Patológico	2
rs200372420	6	c.694C>T	p.Arg232Trp	G/A	<0,001	OR1	Prejudicial	Possivelmente Prejudicial	Polimorfismo	Neutro	2
rs143724056	9	c.1258A>G	p.Met420Val	A/G	<0,001	ZP-N	Tolerado	Benigno	Polimorfismo	Neutro	0
rs141330288	12	c.1447G>A	p.Val483Ile	G/A	<0,001	ZP-C	Tolerado	Benigno	Polimorfismo	Neutro	0
rs116330805	12	c.1510G>A	p.Val504Met	C/T	<0,001	ZP-C	Tolerado	Possivelmente Prejudicial	Polimorfismo	Neutro	1
rs148002300	14	c.1844C>T	p.Ser615Leu	G/A	<0,001	CT	Tolerado	Benigno	Causador de doença	Neutro	1
rs200960408	2	c.218C>T	p.Thr73Met	G	0	OR2	Tolerado	Benigno	Polimorfismo	Neutro	0
rs781374732	3	c.224C>T	p.Pro75Leu	G	0	OR2		Benigno	Polimorfismo	Neutro	0
rs199840979	4	c.388C>T	p.Pro130Ser	G	0	OR2	Tolerado	Benigno	Polimorfismo	Neutro	0
rs139398993	4	c.392C>T	p.Pro131Leu	G	0	OR2	Tolerado	Benigno	Polimorfismo	Neutro	0
rs147792018	5	c.565G>T	p.Asp189Tyr	G	0	OR2	Tolerado	Possivelmente Prejudicial	Causador de doença	Neutro	2
rs41322046	5	c.572G>A	p.Gly191Asp	G	0	OR2	Prejudicial	Possivelmente Prejudicial	Causador de doença	Neutro	3
rs548424658	5	c.596G>C	p.Arg199Pro	G	0	OR2		Benigno	Polimorfismo	Neutro	0
rs756604495	5	c.602C>A	p.Pro201Gln	C	0	OR1		Benigno	Polimorfismo	Neutro	0
rs762209698	5	c.613C>T	p.Arg205Trp	G	0	OR1		Benigno	Polimorfismo	Neutro	0
rs150932144	5	c.640G>A	p.Gly214Ser	G	0	OR1	Tolerado	Possivelmente Prejudicial	Causador de doença	Neutro	2
rs868730895	8	c.1058A>G	p.Glu353Gly	A	0	ZP-N		Benigno	Polimorfismo	Neutro	0
rs548690138	8	c.1093G>A	p.Asp365Asn	G	0	ZP-N		Benigno	Polimorfismo	Neutro	0
rs1800956	8	c.1096G>C	p.Asp366His	C	0	ZP-N	Prejudicial	Possivelmente Prejudicial	Polimorfismo	Neutro	2
rs143054595	8	c.1133C>T	p.Ala378Val	G	0	ZP-N	Tolerado	Benigno	Causador de doença	Neutro	1
rs369997021	10	c.1274C>T	p.Ala425Val	G	0	ZP-N	Tolerado	Benigno	Polimorfismo	Neutro	0
rs763475207	12	c.1469T>C	p.Leu490Ser	A	0	ZP-C		Benigno	Polimorfismo	Patológico	1
rs765564489	12	c.1538A>G	p.Lys513Arg	A	0	ZP-C		Possivelmente Prejudicial	Polimorfismo	Neutro	1
rs142896669	12	c.1633G>A	p.Gly545Ser	G	0	ZP-C	Prejudicial	Possivelmente Prejudicial	Causador de doença	Neutro	3

Continuação da Tabela 7:

rs768491175	12	c.1655C>T	p..Ala552Val	G	0	ZP-C		Benigno	Polimorfismo	Neutro	0
rs138799379	13	c.1712G>A	p.Arg571His	G	0	ZP-C	Prejudicial	Benigno	Polimorfismo	Neutro	1
rs201768056	14	c.1762G>C	p.Val588Leu	G	0	TM	Tolerado	Benigno	Polimorfismo	Neutro	0
rs868296721	14	c.1781G>A	p.Gly594Asp	G	0	TM		Possivelmente Prejudicial	Causador de doença	Patológico	3
rs866067907	15	c.1914C>A	p.Ser638Arg	C	0	CT		Possivelmente Prejudicial		Patológico	2
rs868572062	15	c.1972G>C	p.Ala658Pro	G	0	CT		Benigno		Patológico	1
rs866646256	15	c.1973C>A	p.Ala658Glu	C	0	CT		Possivelmente Prejudicial	Polimorfismo	Patológico	2

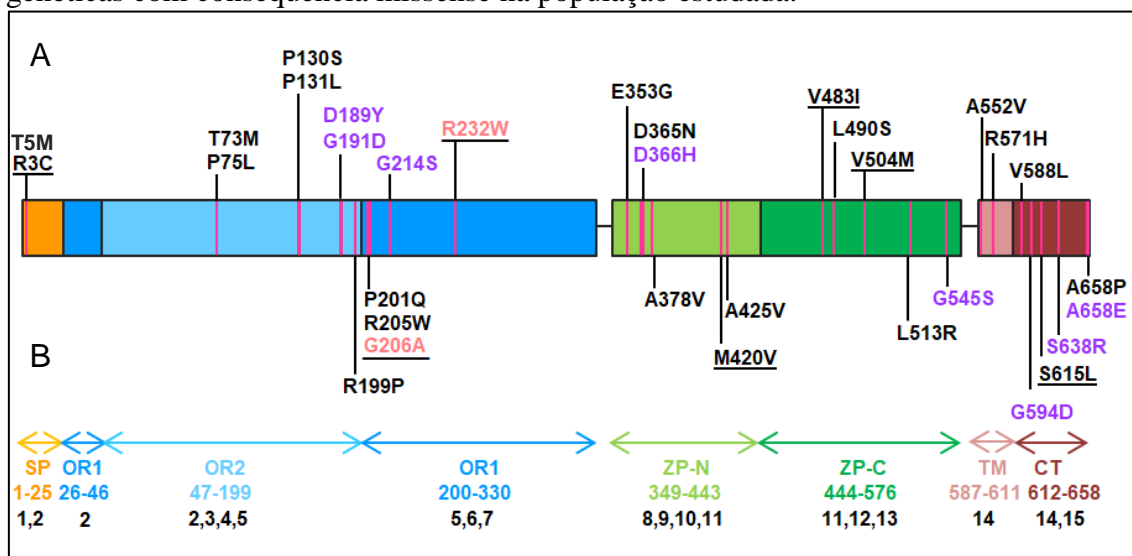
Nota: ^aO primeiro é o alelo de referência e o segundo é o menor alelo; ^balelo de menor frequência; ^cestes resultados são considerados prejudiciais se a pontuação for igual ou superior a dois; espaços em branco referem-se à falta de informações de dados.

Figura 7 - Análise dos efeitos estruturais de variantes missense consideradas prejudiciais utilizando a plataforma HOPE.



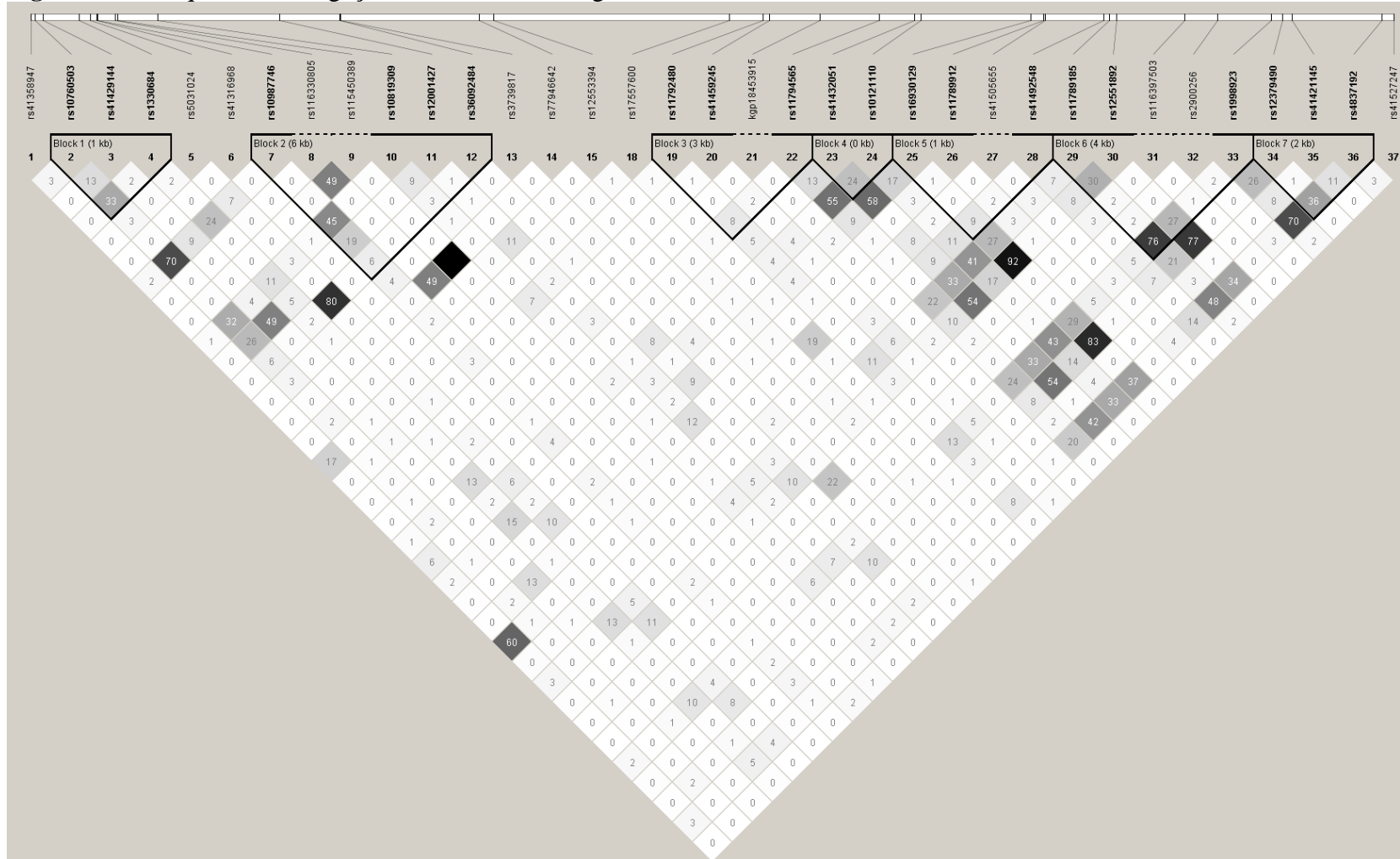
Nota: (A) e (B) se referem às estruturas esquemáticas do aminoácido original (à esquerda) e do mutante (à direita). (A) Alteração de aminoácidos p.Gly206Ala (rs201393380): o resíduo mutante é maior e mais hidrofóbico que o resíduo de tipo selvagem, localizado em uma região necessária para a interação com BMP9, essa alteração pode abolir a flexibilidade conferida pela glicina e pode perturbar a estrutura central desse domínio. (B) Alteração de aminoácidos p.Arg232Trp (rs200372420): o resíduo mutante é maior e apresenta carga positiva, enquanto o resíduo do tipo selvagem tem carga neutra, a mudança ocorre em uma região necessária para a interação com BMP9, localizada na superfície da proteína, podendo interferir nas interações com outras moléculas ou outras partes da proteína. (C) e (D) referem-se a estruturas esquemáticas 3D em diferentes ângulos da alteração de aminoácido p.Arg232Trp, o aminoácido original (em verde) e o mutante (em vermelho).

Figura 8 - Arquitetura esquemática dos domínios de ENG e mapa de variantes genéticas com consequência missense na população estudada.



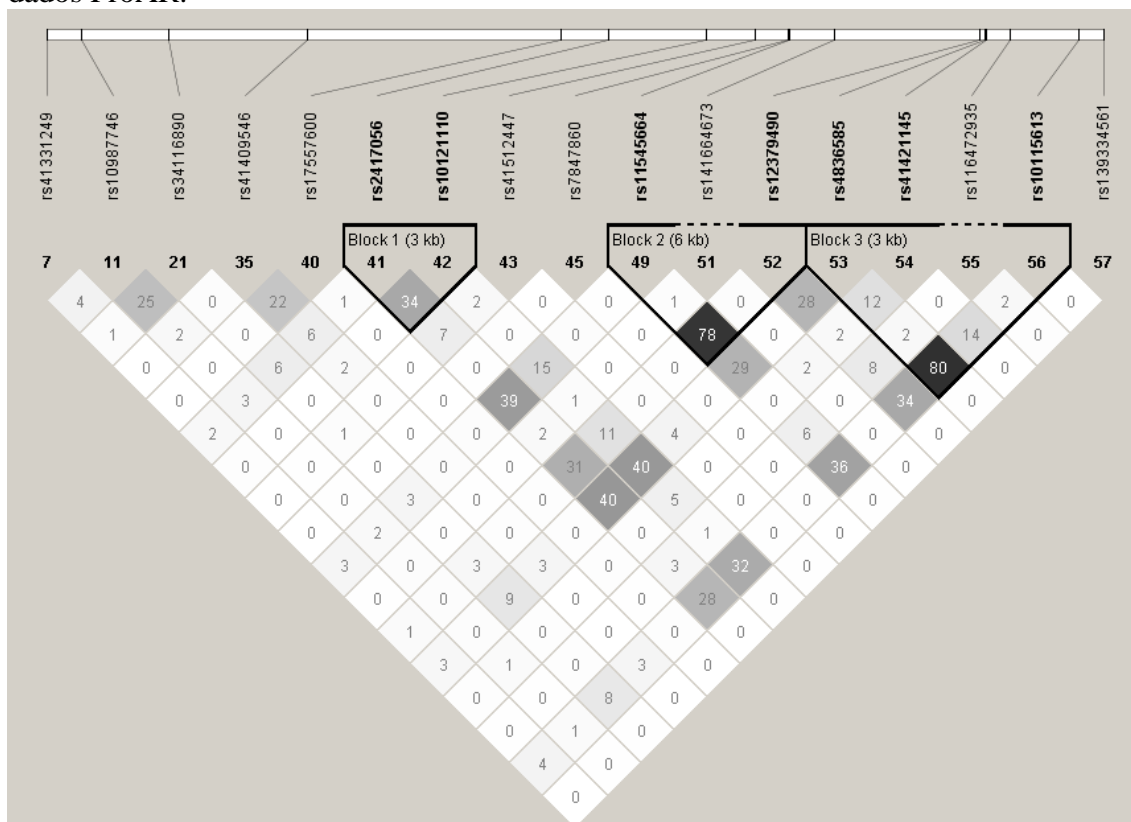
Nota: (A) Estrutura esquemática de Endoglina e mapa de variantes genéticas com consequência *missense*. As mutações destacadas em roxo foram aquelas com pontuação ≥ 2 mostradas na Tabela 7, bem como as mutações destacadas em rosa, analisadas na Figura 7; as mutações sublinhadas referem-se a variantes com MAF > 0 ; as outras variantes apresentam MAF = 0. (B) Indicação dos domínios da proteína e respectivos intervalos de aminoácidos (1-658 aminoácidos); os éxons traduzidos para cada módulo são indicados em preto (1-15 éxons). Fonte: Dados do estudo e adaptado de Saito *et al.* (33), Béllon *et al.* (38).

Figura 9 - Desequilíbrio de ligação entre as variantes genéticas extraídas do banco de dados SCAALA.



Nota: a cor dos quadrados ilustra a força dos valores de r^2 em pares em uma escala em que quadrados sólidos em preto ($r^2 \geq 0,8$) indicam alta correlação, tons de cinza indicam correlação intermediária e em branco indicam baixa correlação. O valor de r^2 também é indicado dentro de cada quadrado.

Figura 10 - Desequilíbrio de ligação entre as variantes genéticas extraídas do banco de dados ProAR.



Nota: a cor dos quadrados ilustra a força dos valores de r^2 em pares em uma escala em que quadrados sólidos em preto ($r^2 \geq 0,8$) indicam alta correlação, tons de cinza indicam correlação intermediária e em branco indicam baixa correlação. O valor de r^2 também é indicado dentro de cada quadrado.

7 DISCUSSÃO

Este estudo teve como objetivo descrever a frequência de variantes genéticas no gene *ENG* em uma população do Nordeste do Brasil, analisando a distribuição de frequências de populações pseudoancestrais, uma vez que é sabido que as variantes nesse gene podem estar envolvidas em processos patogênicos descritos anteriormente em outras populações.

Mutações na linha germinativa que levam à diminuição da função de *ENG* configuram a HHT1 devido à haploinsuficiência de $ENG^{(+/-)}$. Essa doença culmina em displasia vascular com sangramentos frequentes e afeta a mucosa nasal, o trato gastrointestinal e a pele. Além disso, essas mutações podem levar a condições graves devido a malformações arteriovenosas (MAV) no pulmão, cérebro e fígado (15–17).

Estudos genéticos de rastreamento de mutações em famílias com histórico de HHT, identificaram a ocorrência de SNVs em *ENG* como rs3739817 (frequência de 5% (106) e 4% (107)), rs11545664 (frequência de 11% (106) e 1% (107)), rs1800956 (107) em 32 famílias da Europa Setentrional (106) e em 4 famílias da Itália (107), rs3739817, rs116330805, rs11545664, rs41322046 em 160 pacientes de famílias não relacionadas da França (108), rs116330805, rs11545664, rs41322046 e rs41358947 em 13 famílias Canadenses (109), mas nenhuma análise de associação dessas variantes foi realizada para inferir um possível risco ou proteção para o desfecho. Em nosso estudo, essas alterações ocorreram em nossa população, exceto para o rs41322046 que é monomórfico para o alelo G. O alelo T-rs11545664 possui frequência de 21%, superior ao relatado nos estudos (106,107), o alelo A-rs3739817 possui frequência de 0,04 neste estudo, frequência menor do que nos estudos em referência (106,107), alelo G-rs41358947 ocorre em 3% e T-rs116330805 ocorre <0,1% da população (Tabela 3).

Estudos posteriores exploraram o impacto de algumas mutações *missense* em *ENG*, mostrando que estas raramente podem ser detectadas e expressas na superfície celular, sendo retidas intracelularmente (17,106,110). Por outro lado, um estudo mais recente também analisou variantes *missense* em *ENG*, incluindo rs13933456 (R3C), rs116330805 (V504M), rs148002300 (S615L), rs201393380 (G206A) e estes foram encontrados na superfície celular, como ocorre em *ENG* de tipo selvagem, sendo consideradas, portanto, mutações silenciosas (81). Vale destacar que foi esperado que o mutante S615L pudesse provocar alguma alteração na fosforilação, visto que o resíduo

selvagem se trata de uma serina, mas não foram evidenciados defeitos de sinalização na ocorrência dessa variante (81). Mallet e colaboradores (2015) também apontam que os mutantes rs13933456 (R3C), rs116330805 (V504M) e rs148002300 (S615L) em *ENG* foram relatados no banco de dados internacional associados ao fenótipo HHT benigno e incerto, enquanto o mutante rs201393380 (G206A) foi descrito em pacientes com HHT improvável ou possível, enfatizando a importância da realização de estudos funcionais (81). As análises *in silico* do presente estudo indicam que as variantes rs13933456 (R3C), rs116330805 (V504M) e rs148002300 (S615L) não são prejudiciais (Tabela 5), o que é consistente com os resultados de Mallet e colaboradores (2015). A análise *in silico* não é totalmente confiável, uma vez que as diferentes plataformas apresentam previsões conflitantes (105), portanto, para esta pesquisa utilizamos diferentes softwares e foi adotado o uso de escore (≥ 2) para indicar patogenicidade. O mutante rs201392280 (G206A) foi considerado patogênico de acordo com o escore (2) (Tabela 5) e o alelo G-rs201393380 apresentou frequência $< 0,001$, sendo raro na população brasileira (Tabela 4). A análise *in silico* subsequente, utilizando o software HOPE foi realizada para prever os possíveis efeitos estruturais para essa variante, rs201393380 (Figura 7A), e indica que a alteração de glicina pela alanina pode abolir a flexibilidade conferida pela glicina e afetar a estrutura central do domínio órfão (OR1, mostrado na Figura 8), importante para a interação com BMP9, um ligante crucial para a cascata de sinalização que ativa a angiogênese (72). A interação ENG-BMP9-ALK1 é amplamente investigada no contexto da HHT, pois a resposta à formação ou ao reparo de vasos sanguíneos depende da sinalização e do reconhecimento de ligantes específicos para ENG, como BMP9. Sendo assim, alterações na estrutura da molécula de endoglin podem alterar a sua avidéz na interação intermolecular e, conseqüentemente, afetar sua atividade que é crucial para mediar funções celulares que implicam na homeostase do organismo.

A variante sinônima rs11545664 (L69L) foi investigada na suscetibilidade para malformações arteriovenosas cerebrais (BAVMs) e para o risco de hemorragia intracraniana (HIC) em um estudo de caso-controle no Sul do Brasil. Esse estudo foi composto por 63 casos (80,6% caucasianos, 6,5% afro-brasileiros, 12,9% misturados) e 96 controles (85,4%, 6,3% e 8,3%, respectivamente). A análise mostrou apenas uma tendência de associação, mas sem significância, entre rs11545664 (T/C) e a ocorrência de BAVM em um genótipo completo (111).

Um estudo realizado na Universidade da Califórnia, em São Francisco, envolveu 177 pacientes brancos com BAVM esporádico, ou seja, sem associação com HHT ou outra patologia cerebrovascular, e 129 controles. O rs11545664 mostrou associação com o desfecho para o genótipo GG (odds ratio, $OR = 1,88$, [95% CI, 1,04 a 3,42], $p = 0,002$). A associação encontrada foi aumentada (OR , 2,68, [95% CI, 1,64 a 4,39], $p < 0,001$) quando combinada com a variante rs2071219 do gene *ACVRL1* (gene que codifica ALK1, membro da família do TGF- β), ambas variantes estão em alta correlação para o LD ($r^2 = 0,76$) (112). O rs11545664 também foi investigado em relação à associação com BAVMs esporádicas ou relacionadas ao HHT em pacientes holandeses, mas nenhuma associação significativa foi encontrada e, em uma tentativa de validação, a combinação das variantes rs11545664 e rs2071219-*ACVRL1* não mostrou OR significativa para qualquer combinação de alelos (113). A BAVM ocorre em pacientes com HHT (16), e a ocorrência familiar de BAVM esporádica é rara (114). O rs11545664 reduz a ligação prevista para uma proteína de *splice* (SRp40) (115). Estudos recentes mostram que essa consequência pode interferir no gene e na expressão molecular da proteína, resultando no desfecho patogênico (HHT1) (85,116). Variantes sinônimas, como rs11545664, não apenas representam variabilidade na composição genética de um indivíduo ou um evento espúrio como já foi muito difundido (117). A variante sinônima pode alterar a cinética de tradução do RNAm devido à alteração da sequência dos ácidos nucleicos, compondo um códon sinônimo, dessa maneira, pode afetar o dobramento da proteína co-traducional e pode repercutir na produção e função da proteína, influenciando o risco e o espectro dos sintomas da doença, bem como podendo afetar a resposta terapêutica (117).

Um estudo realizado na Holanda, composto por uma população mista, mostrou que, entre 649 holandeses caucasianos (mais predominantes) e 72 negros das Antilhas Holandesas, o alelo C-rs10987746 foi significativamente associado ao aumento do risco de MAV pulmonar em pacientes com HHT1 [risco relativo = 1,3 (1,0018–1,7424)], apresentando maior risco em indivíduos homocigotos ($p = 0,004$). O alelo C ocorreu em 56,5% dos pacientes com HHT1 com MAV pulmonar e 43,5% entre os pacientes com HHT1 sem MAV pulmonar. Em um modelo dominante, o mesmo alelo foi associado a baixos níveis de expressão de RNAm de ENG ($p = 0,004$) em um painel de linhas celulares de linfoblastos humanos (118). Em nossa população, o alelo C-rs10987746 foi 52% frequente (Tabela 3 e Tabela 4).

As variantes *missense* rs142896669, rs41322046, sinônima rs11545664 e intrônica rs7847860 foram relatadas em 57 pacientes europeus com hipertensão arterial pulmonar (PAH) e 55 controles (105). A variante rs41322046 (G191D) foi encontrada em 10 pacientes italianos com PAH e foi mais frequente na PAH idiopática do que na PAH associada a outras condições ($p = 0,040$). A frequência do alelo A- rs41322046 foi de 10% nos casos e não foi detectada nos controles, os quais foram associadas a uma idade precoce para o diagnóstico de PAH (média de 10 anos antes, $p = 0,035$) (105). Em nosso estudo, o rs41322046 é monomórfico para o alelo G. Essa variante foi considerada patogênica de acordo com o escore (3) com base nos resultados de análises *in silico* (Tabela 7) e de acordo com as análises realizadas por Pousada *et al.* (2016) (105). A variante provavelmente é prejudicial à proteína porque o aminoácido glicina é um resíduo altamente conservado (105). Outros estudos descrevem rs41322046 como um polimorfismo ou uma variante rara (108,109) e todas essas variantes descritas anteriormente, rs11545664, rs7847860, rs3739817, rs142896669, incluindo rs41322046, são classificadas como variantes benignas pelo banco de dados internacional ARUP. Devido às divergências encontradas em relação à patogenicidade da variante rs41322046 e por estar presente em baixa frequência na população geral, mas estando presente em pacientes com PAH, Pousada *et al.* (2016) realizaram um estudo funcional com a construção de um minigene a partir do éxon 5, no qual é encontrada a variante rs41322046, expandindo para 200 pb nas junções com íntrons, mas não foram encontradas variantes ligadas aos processos de splicing (105).

As variantes intrônicas rs11792480 e rs10121110, apresentadas como tagging SNV (tSNV) marcando diferentes haploblocos ($r^2 = 0,29$, separados por cerca de 4.000 bases), foram independentemente associadas à pré-eclâmpsia (PE) em mulheres norte-americanas brancas, mas não em negras, em um estudo de caso-controle, composto por 181 casos e 174 controles (19). O alelo G-rs11792480 ocorreu em 71,7% das mulheres com PE e 63% nos controles ($p = 0,01$) e o alelo A-rs10121110 ocorreu em 66% e 55,3% ($p = 0,004$), respectivamente. Em nosso estudo, a correlação entre essas variantes foi ainda inferior ($r^2 = 0,04$) (Figura 9) e ambas não apresentam forte correlação para nenhuma outra variante (Figura 9 e Figura 10), o alelo G-rs11792480 teve uma frequência de 84% (Tabela 3) e o alelo A-rs10121110 apresentou frequência de cerca de 44% (Tabela 3 e Tabela 4). Ainda em se tratando do estudo em referência (19), os autores observam que mulheres brancas que herdaram o alelo A-rs10121110 em

homozigose tiveram um risco 2,22 vezes maior de desenvolver PE em comparação com mulheres com o genótipo GG ($p = 0,008$, [99% CI: 1,02 a 5,13]) (19). Estudos adicionais procuraram validar esses achados, mas não foram encontradas associações com as variantes rs11792480 e rs10121110 em *ENG* e PE nas seguintes coortes: 77 casos e 63 controles em noruegueses caucasianos, 69 casos e 106 controles em caucasianos americanos (21), 130 casos e 316 controles em chineses (20) e 75 casos e 50 controles em egípcios (22). Devido a essas diferenças, alguns argumentos são sugeridos, como: uma possível falta de robustez na análise de associação devido ao tamanho da coorte, os haplótipos marcados pelo tSNV selecionado podem diferir por ancestralidade ou devido a diferenças na definição do fenótipo de PE (21). Além disso, ainda não está claro se os diferentes achados entre mulheres brancas e negras no estudo de Bell *et al.* (2013) são atribuídos devido à heterogeneidade da população ou à diferença no tamanho da amostra (pequena para o grupo de mulheres negras) (20). Outro estudo na perspectiva da PE com uma amostra negra norte-americana (184 casos e 305 controles) muito maior que a amostra de mulheres brancas (32 casos e 85 controles), bem como maior que as amostras dos estudos citados anteriormente, não mostra uma associação de SNVs em *ENG* com o desfecho de PE (119). Nas mulheres negras, a variante rs17557600 foi investigada, mas não apresentou associação com PE para o alelo T ($p = 0,06$; *OR* 1,46 [IC 95%: 0,98 a 2,17]), e apresenta frequência de 90% (119). No presente estudo, o alelo T-rs17557600 teve uma frequência de 91% (Tabela 3 e Tabela 4).

Devido à sua atuação como mediadora na sinalização da angiogênese, a *ENG* é relacionada à progressão tumoral maligna, sendo alvo para terapia antiangiogênica com o uso de anticorpos monoclonais através do bloqueio funcional da via *ENG*-BMP9-*ALK1* (1,2,9). Mas no contexto do câncer as variantes genéticas em *ENG* são pouco descritas e exploradas. Um estudo recente realizado nos Estados Unidos incluiu *ENG* como um dos genes candidatos à quimiorresistência no câncer epitelial de ovário seroso (EOC) em estágio avançado, devido à *ENG* estar relacionada à regulação ovariana de células-tronco cancerígenas (120). Porém, nenhuma associação foi encontrada para 130 variantes em *ENG* com uma resposta incomum à terapia entre pacientes com EOC (máxima verossimilhança da mistura, $AML = 0,184$). Entre os pacientes, o alelo T-rs41432051 teve uma frequência de 15%, A-rs2417056 com 17% e A-rs11545664 com 13% (120). Outro estudo identificou o rs11545664 como um componente em um

modelo de predição para a toxicidade por esofagite aguda pós-quimiorradioterapia em pacientes com câncer de pulmão (121). Combinando dados clínicos, parâmetros de tratamento e informações genóticas, a variante sinônima rs11545664 em *ENG* e outras variantes, rs2302535 (*EGFR*), rs1131877 (*TRAF3*) e rs2230528 (*ITGB2*), foram consideradas neste modelo de previsão. A frequência do alelo T-rs11545664 foi de 12,4% para todos os pacientes, diminuindo para 7,75% naqueles com esofagite aguda (121).

Uma pesquisa recente do nosso grupo mostrou que a expressão de *ENG* desempenha um papel importante no desenvolvimento e no tratamento da leucemia aguda (12). Diferentes fatores podem estar diretamente associados à etiologia das neoplasias hematopoiéticas, incluindo fatores genéticos, imunológicos e ambientais (122). Nesse sentido, *ENG* prova ser um forte gene candidato e nosso grupo já vem desenvolvendo um projeto com a finalidade de analisar as variantes genéticas na leucemia.

8 CONCLUSÃO

A frequência de variantes genéticas em *ENG* na população brasileira se mostrou, em geral, intermediária entre os padrões africano e europeu. O fenótipo HHT1 com malformações arteriovenosas pulmonares, AVM cerebral esporádica e pré-eclâmpsia têm sido associadas em outros estudos com os alelos mais frequentes no gene *ENG*. Os SNVs associados nos estudos trazidos foram mais frequentes em nossa população, com exceção do alelo A-rs10121110 em relação à pré-eclâmpsia. É importante destacar a variante sinônima rs11545664 (L69L), localizada no éxon 2, que foi investigada em diferentes desfechos. No contexto do câncer, variantes genéticas em *ENG* são pouco descritas e exploradas, nesse sentido, nosso grupo já vem desenvolvendo um projeto com a finalidade de analisar as variantes genéticas na leucemia.

Este trabalho, visa ainda, desenvolver uma análise de componentes principais para a estimação da ancestralidade biogeográfica da população do estudo, inclusive utilizando dados de AIMs. Posteriormente pretendemos explorar *ENG* como gene candidato para leucemia em estudo de associação entre genótipo e fenótipo, do tipo caso-controle, bem como analisar funcionalmente as variantes genéticas nesse contexto.

Mais pesquisas sobre os mecanismos funcionais desse gene são necessárias, pois ele participa de uma série de processos celulares importantes e mais esforços devem ser feitos para estudos genéticos na população brasileira considerando a miscigenação, visto que a coloração da pele é um preditor não confiável de ancestralidade étnica do indivíduo.

REFERÊNCIAS BIBLIOGRÁFICAS

1. Bernabeu C, Lopez-Novoa JM, Quintanilla M. The emerging role of TGF- β superfamily coreceptors in cancer. *Biochim Biophys Acta - Mol Basis Dis.* 2009;1792(10):954–73. Available from: <http://dx.doi.org/10.1016/j.bbadis.2009.07.003>
2. Lebrin F, Goumans M-J, Jonker L, Carvalho RLC, Valdimarsdottir G, Thorikay M, et al. Endoglin promotes endothelial cell proliferation and TGF-beta/ALK1 signal transduction. *EMBO J.* 2004 Oct 13;23(20):4018–28. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/15385967>
3. Abdalla SA, Letarte M. Hereditary haemorrhagic telangiectasia: Current views on genetics and mechanisms of disease. *J Med Genet.* 2006;43(2):97–110.
4. Jonker L, Arthur H. Endoglin expression in early development is associated with vasculogenesis and angiogenesis. *Mech Dev.* 2002;(110):193–196.
5. Perlingeiro RCR. Endoglin is required for hemangioblast and early hematopoietic development. *Development.* 2007;134(16):3041–8.
6. Venkatesha S, Toporsian M, Lam C, Hanai J, Mammoto T, Kim YM, et al. Soluble endoglin contributes to the pathogenesis of preeclampsia. *Nat Med.* 2006 Jun 4;12(6):642–9. Available from: <http://www.nature.com/articles/nm1429>
7. Gregory AL, Xu G, Sotov V, Letarte M. Review: The enigmatic role of endoglin in the placenta. *Placenta.* 2014;35(SUPPL):S93-99.
8. Hawinkels LJAC, Kuiper P, Wiercinska E, Verspaget HW, Liu Z, Pardali E, et al. Matrix metalloproteinase-14 (MT1-MMP)-mediated endoglin shedding inhibits tumor angiogenesis. *Cancer Res.* 2010;70(10):4141–50.
9. Li C, Guo B, Bernabeu C, Kumar S. Plasma levels of soluble CD105 correlate with metastasis in patients with breast cancer. *Int J cancer.* 2000 Mar 20;89(2):122–6. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10754488>
101. Blázquez-Medela AM, García-Ortiz L, Gómez-Marcos MA, Recio-Rodríguez JI, Sánchez-Rodríguez A, López-Novoa JM, et al. Increased plasma soluble endoglin levels as an indicator of cardiovascular alterations in hypertensive and diabetic patients. *BMC Med.* 2010;8(1):86. Available from: <http://www.biomedcentral.com/1741-7015/8/86>
11. Honsawek S, Tanavalee A, Yuktanandana P. Elevated Circulating and Synovial Fluid Endoglin Are Associated with Primary Knee Osteoarthritis Severity. *Arch*

- Med Res. 2009;40(7):590–4. Available from: <http://dx.doi.org/10.1016/j.arcmed.2009.07.010>
12. Dourado KMC, Baik J, Oliveira VKP, Beltrame M, Yamamoto A, Theuer CP, et al. Endoglin: a novel target for therapeutic intervention in acute leukemias revealed in xenograft mouse models. *Blood*. 2017;129(18):2526–36.
 13. Li DY, Sorensen LK, Brooke BS, Urness LD, Davis EC, Taylor DG, et al. Defective angiogenesis in mice lacking endoglin. *Science*. 1999;284(5419):1534–7.
 14. Arthur HM, Ure J, Smith AJH, Renforth G, Wilson DI, Torsney E, et al. Endoglin, an ancillary TGF β receptor, is required for extraembryonic angiogenesis and plays a key role in heart development. *Dev Biol*. 2000;217(1):42–53.
 15. Bourdeau A, Dumont DJ, Letarte M. A murine model of hereditary hemorrhagic telangiectasia. *J Clin Invest*. 1999 Nov 15;104(10):1343–51. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/10562296>
 16. McAllister K a, Grogg KM, Johnson DW, Gallione CJ, Baldwin M a, Jackson CE, et al. Endoglin, a TGF-beta binding protein of endothelial cells, is the gene for hereditary haemorrhagic telangiectasia type 1. *Nat Genet*. 1994;8(4):345–51.
 17. Pece N, Vera S, Cymerman U, White RI, Wrana JL, Letarte M. Mutant endoglin in hereditary hemorrhagic telangiectasia type 1 is transiently expressed intracellularly and is not a dominant negative. *J Clin Invest J Clin Invest*. 1997;100(10):2568–79. Available from: <https://doi.org/10.1172/JCII19800>.
 18. Li C, Guo B, Bernabeu C, Kumar S. Angiogenesis in breast cancer: the role of transforming growth factor beta and CD105. *Microsc Res Tech*. 2001;52:437–49.
 19. Bell MJ, Roberts JM, Founds SA, Jeyabalan A, Terhorst L, Conley YP. Variation in endoglin pathway genes is associated with preeclampsia: a case – control candidate gene association study. *BioMed Cent Ltd*. 2013;13(82):1–9.
 20. Li X, Tan H, Chen M, Zhou S. Transforming growth factor beta 1 related gene polymorphisms in gestational hypertension and preeclampsia: A case-control candidate gene association study. *Pregnancy Hypertens*. 2017;12:155–60. Available from: <https://doi.org/10.1016/j.preghy.2017.11.010>
 21. Schmella MJ, Roberts JM, Conley YP, Ren D, Storvold GL, Ingles SA, et al. Endoglin pathway genetic variation in preeclampsia: A validation study in Norwegian and Latina cohorts. *Pregnancy Hypertens*. 2018;12(March

- 2017):144–9.
22. ElMonier AA, El-Boghdady NA, Abdelaziz MA, Shaheen AA. Association between endoglin/transforming growth factor beta receptors 1, 2 gene polymorphisms and the level of soluble endoglin with preeclampsia in Egyptian women. *Arch Biochem Biophys*. 2019;662(October 2018):7–14. Available from: <https://doi.org/10.1016/j.abb.2018.11.022>
 23. Quackenbush EJ, Letarte M. Identification of several cell surface proteins of non-T, non-B acute lymphoblastic leukemia by using monoclonal antibodies. *J Immunol*. 1985;134(2):1276–85. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/3155538>
 24. Gougos, A. und Letarte M. Identification of a human endothelial cell antigen with monoclonal antibody 44G4 produced against a pre-B leukemic cell line. *J Immunol*. 1988;141(6):1925–33.
 25. López-Novoa JM, Bernabeu C. The physiological role of endoglin in the cardiovascular system. *Am J Physiol Circ Physiol*. 2010;299(4):H959–74.
 26. Gougos A, St Jacques S, Greaves A, O’Connell P, D’Apice A, Bühring H, et al. Identification of distinct epitopes of endoglin, an RGD-containing glycoprotein of endothelial cells, leukemic cells, and syncytiotrophoblasts. *Int Immunol*. 1992;4(1):83–92.
 27. St-Jacques S, Cymerman U, Pece N, Letarte M. Molecular characterization and in situ localization of murine endoglin reveal that it is a transforming growth factor-beta binding protein of endothelial and stromal cells. *Endocrinology*. 1994;134:2645–57.
 28. Lastres P, Bellon C, Cabanas F, Sanchez-Madrir F, Acevedo A, Gougos A, et al. Regulated expression on human macrophages of endoglin, an Arg-Gly-Asp-containing surface antigen. *Eur J Immunol*. 1992;22:393–7.
 29. Chen C-Z, Li M, de Graaf D, Monti S, Gottgens B, Sanchez M-J, et al. Nonlinear partial differential equations and applications: Identification of endoglin as a functional marker that defines long-term repopulating hematopoietic stem cells. *Proc Natl Acad Sci*. 2002 Nov 26;99(24):15468–73. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12438646>
 30. Perlingeiro RCR. Endoglin is required for hemangioblast and early hematopoietic development. *Dev Dis*. 2007;134:3041–7.
 31. Dallas NA, Samuel S, Xia L, Fan F, Gray MJ, Lim SJ, et al. Endoglin (CD105):

- A marker of tumor vasculature and potential target for therapy. *Clin Cancer Res.* 2008;14(7):1931–7.
32. Gougos A, Letarte M. Communication Primary Structure of Endoglin , an Glycoprotein of. *J Biol Chem.* 1990;265(15):8361–4.
 33. Saito T, Bokhove M, Croci R, Zamora-Caballero S, Han L, Letarte M, et al. Structural Basis of the Human Endoglin-BMP9 Interaction: Insights into BMP Signaling and HHT1. *Cell Rep.* 2017;19(9):1917–28. Available from: <http://dx.doi.org/10.1016/j.celrep.2017.05.011>
 34. Llorca O, Trujillo A, Blanco FJ, Bernabeu C. Structural Model of Human Endoglin, a Transmembrane Receptor Responsible for Hereditary Hemorrhagic Telangiectasia. *J Mol Biol.* 2007;365(3):694–705.
 35. Jovine L, Darie CC, Litscher ES, Wassarman PM. Zona Pellucida Domain Proteins. *Annu Rev Biochem.* 2005;74(1):83–114.
 36. Guerrero-Esteo M, Sá Nchez-Elsner T, Letamendia A, Bernabé C. Extracellular and Cytoplasmic Domains of Endoglin Interact with the Transforming Growth Factor- β Receptors I and II*. *J Biol Chem.* 2002;277(9):29197–209.
 37. Rossi E, Smadja DM, Boscolo E, Langa C, Arevalo MA, Pericacho M, et al. Endoglin regulates mural cell adhesion in the circulatory system. *Cell Mol Life Sci.* 2016;73(8):1715–39.
 38. Bellón T, Corbi A, Lastres P, Calés C, Cebrián M, Vera S, et al. Identification and expression of two forms of the human transforming growth factor- β -binding protein endoglin with distinct cytoplasmic regions. *Eur J Immunol.* 1993;23(9):2340–5.
 39. Alt A, Miguel-Romero L, Donderis J, Aristorena M, Blanco FJ, Round A, et al. Structural and functional insights into endoglin ligand recognition and binding. *PLoS One.* 2012;7(2).
 40. Barbara NP, Wrana JL, Letarte M, Barbara NP, Wrana JL, Letarte M. Endoglin Is an Accessory Protein That Interacts with the Signaling Receptor Complex of Multiple Members of the Transforming Growth Factor- β Endoglin Is an Accessory Protein That Interacts with the Signaling Receptor Complex. *J Biol Chem.* 1999;274(2):584–94.
 41. Letamendia A, Lastres P, Botella L, Raab U, Langa C, Velasco B, et al. Role of endoglin in cellular responses to transforming growth factor-b. A comparative study with betaglycan. *J Biol Chem.* 1998;273:33011–9.

42. Pichuantes S, Vera S, Bourdeau A, Pece N, Kumar S, Wayner E, et al. Mapping epitopes to distinct regions of the extracellular domain of endoglin using bacterially expressed recombinant fragments. *Tissue Antigens*. 1997;50(3):265–76.
43. Pérez-Gómez E, del Castillo G, Santibáñez JF, Lêpez-Novoa JM, Bernabéu C, Quintanilla M. The Role of the TGF- β Coreceptor Endoglin in Cancer. *Sci World J*. 2010 Dec 14;10:2367–84. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21170488>
44. Koleva RI, Conley BA, Romero D, Riley KS, Marto JA, Lux A, et al. Endoglin structure and function: Determinants of endoglin phosphorylation by transforming growth factor- β receptors. *J Biol Chem*. 2006;281(35):25110–23.
45. Meng QJ, Lux A, Holloschi A, Li J, Hughes JMX, Foerg T, et al. Identification of Tctex2 β , a novel dynein light chain family member that interacts with different transforming growth factor- β receptors. *J Biol Chem*. 2006;281(48):37069–80.
46. Lee NY, Blobel GC. The interaction of endoglin with β -arrestin2 regulates transforming growth factor- β -mediated ERK activation and migration in endothelial cells. *J Biol Chem*. 2007;282(29):21507–17.
47. Gallardo-Vara E, Tual-Chalot S, Botella LM, Arthur HM, Bernabeu C. Soluble endoglin regulates expression of angiogenesis-related proteins and induction of arteriovenous malformations in a mouse model of hereditary hemorrhagic telangiectasia. *Dis Model Mech*. 2018;11(9):dmm034397.
48. Aristorena M, Gallardo-vara E, Vicen M, Casas-engel MD Las, Ojeda-fernandez L, Blanco FJ, et al. MMP-12 , Secreted by Pro-Inflammatory Macrophages , Targets Endoglin in Human Macrophages and Endothelial Cells. *Int J Mol Sci*. 2019;20(3107):1–19.
49. Varejckova M, Gallardo-Vara E, Vicen M, Vitverova B, Fikrova P, Dolezelova E, et al. Soluble endoglin modulates the pro-inflammatory mediators NF- κ B and IL-6 in cultured human endothelial cells. Vol. 175, *Life Sciences*. Elsevier Inc; 2017. 52–60 p. Available from: <http://dx.doi.org/10.1016/j.lfs.2017.03.014>
50. Derynck R, Miyazono K. The TGF- β family. *Cold Spring Harb*. 2008;1114 p.
51. Moustakas A, Heldin C. Non-Smad TGF- β signals. *J Cell Sci*. 2005;118:3573–3584.
52. Massagué J. TGF β signalling in context. *Nat Rev Mol Cell Biol*. 2012;13:616–30.

53. Wakefield L, Hill C. Beyond TGF β : Roles of other TGF β superfamily members in cancer. *Nat Rev Cancer*. 2013;13:328–341.
54. Heldin CH, Moustakas A. Signaling receptors for TGF- β family members. *Cold Spring Harb Perspect Biol*. 2016;8(8):1–33.
55. De Crescenzo G, Hinck CS, Shu Z, Zúñiga J, Yang J, Tang Y, et al. Three key residues underlie the differential affinity of the TGF β isoforms for the TGF β type II receptor. *J Mol Biol*. 2006;355(1):47–62.
56. Wrana J, Attisano L, Wieser R, Ventura F, Massagué J. Mechanism of activation of the TGF-beta receptor. *Nature*. 1994;4(370):341–7.
57. Cheifetz S, Bellon T, Cales C, Vera S, Bernabeu C, Massague J, et al. Endoglin is a component of the transforming growth factor-beta receptor system in human endothelial cells. *J Biol Chem*. 1992;267(27):19027–30.
58. Lastres P, Letamendía A, Zhang H, Rius C, Almendro N, Raab U, et al. Endoglin modulates cellular responses to TGF- β 1. *J Cell Biol*. 1996;133(5):1109–21.
59. López-Casillas F, Cheifetz S, Doody J, Andres JL, Lane WS, Massague J. Structure and expression of the membrane proteoglycan betaglycan, a component of the TGF- β receptor system. *Cell*. 1991;67(4):785–95.
60. Wrana J, Attisano L, Cárcamo J, Zentella A, Doody J, Laiho M, et al. TGF beta signals through a heteromeric protein kinase receptor complex. *Cell*. 1992;71(6):1003–14.
61. Bernabeu C, Blanco F, Langa C, Garrido-Martin E, Botella L. Involvement of the TGF-beta superfamily signalling pathway in hereditary haemorrhagic telangiectasia. *J Appl Biomed*. 2010;J Appl Bio(8):169–171.
62. Risau W, Flamme I. Vasculogenesis. *Annu Rev Cell Dev Biol*. 1995;11:73–91.
63. Risau W. Mechanisms of angiogenesis. *Nature*. 1997;386:671–4.
64. Burrows F, Derbyshire E, Tazzari P, Amlot P, Gazdar A, King S, et al. Up-regulation of endoglin on vascular endothelial cells in human solid tumors: implications for diagnosis and therapy. *Clin Cancer Res*. 1995;1:1623–4.
65. Sánchez-Elsner T, Botella LM, Velasco B, Langa C, Bernabéu C. Endoglin Expression Is Regulated by Transcriptional Cooperation between the Hypoxia and Transforming Growth Factor- β Pathways. *J Biol Chem*. 2002 Nov 15;277(46):43799–808. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/12228247>
66. Chakhachiro ZI, Zuo Z, Aladily TN, Kantarjian HM, Cortes JE, Alayed K, et al.

- CD105 (Endoglin) Is Highly Overexpressed in a Subset of Cases of Acute Myeloid Leukemias. *Am J Clin Pathol.* 2013 Sep 1;140(3):370–8. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/23955456>
67. Graulich W, Nettelbeck D, Fischer D, Kissel T, Müller R. Cell type specificity of the human endoglin promoter. *Gene.* 1999;(227):55–62.
 68. Düwel A, Eleno N, Jerkic M, Arevalo M, Bolaños J, Bernabeu C, et al. Reduced tumor growth and angiogenesis in endoglin-haploinsufficient mice. *Tumour Biol.* 2007;28:1–8.
 69. Núñez-Gómez E, Pericacho M, Ollauri-Ibáñez C, Bernabéu C, López-Novoa JM. The role of endoglin in post-ischemic revascularization. *Angiogenesis.* 2017;20(1):1–24.
 70. Lebrin F, Deckers M, Bertolino P, Ten Dijke P. TGF-beta receptor function in the endothelium. *Cardiovasc Res.* 2005;(65):599–608.
 71. Bertolino P, Deckers M, Lebrin F, ten Dijke P. Transforming growth factor-beta signal transduction in angiogenesis and vascular disorders. *Chest.* 2005;128:585S-90S. Available from: [https://linkinghub.elsevier.com/retrieve/pii/S0012-3692\(15\)31007-2](https://linkinghub.elsevier.com/retrieve/pii/S0012-3692(15)31007-2)
 72. Castonguay R, Werner ED, Matthews RG, Presman E, Mulivor AW, Solban N, et al. Soluble endoglin specifically binds bone morphogenetic proteins 9 and 10 via its orphan domain, inhibits blood vessel formation, and suppresses tumor growth. *J Biol Chem.* 2011;286(34):30034–46.
 73. Lebrin F, Mummery C. Endoglin-mediated vascular remodeling: mechanisms underlying hereditary hemorrhagic telangiectasia. *Trends Cardiovasc Med.* 2008;18(1):25–32.
 74. Fernández-Ruiz E, St-Jacques S, Bellón T, Letarte M, Bernabéu C. Assignment of the human endoglin gene (End) to 9q34→qter. *Cytogenet Genome Res.* 1993;64(3–4):204–7.
 75. Ríus C, Smith JD, Almendro N, Langa C, Botella LM, Marchuk DA, et al. Cloning of the Promoter Region of Human Endoglin, the Target Gene for Hereditary Hemorrhagic Telangiectasia Type 1. *Blood.* 1998;92(12):4677–90. Available from: <http://www.bloodjournal.org/content/bloodjournal/92/12/4677.full.pdf>
 76. Pérez-Gómez E, Eleno N, López-Novoa JM, Ramirez JR, Velasco B, Letarte M, et al. Characterization of murine S-endoglin isoform and its effects on tumor

- development. *Oncogene*. 2005;24(27):4450–61.
77. Seghers L, de Vries M, Pardali E, Hoefler I, Hierck B, ten Dijke P, et al. Shear induced collateral artery growth modulated by endoglin but not by ALK1. *J Cell Mol Med*. 2012;16:2440–50.
 78. Quintanilla M, Ramirez J, Perez-Gomez E, Romero D, Velasco B, Letarte M, et al. Expression of the TGF-beta coreceptor endoglin in epidermal keratinocytes and its dual role in multistage mouse skin carcinogenesis. *Oncogene*. 2003;22:5676–85.
 79. Gallione CJ, Repetto GM, Legius E, Rustgi AK, Schelley SL, Tejpar S, et al. A combined syndrome of juvenile polyposis and hereditary haemorrhagic telangiectasia associated with mutations in MADH4 (SMAD4). *Lancet*. 2004;363, 852–859.
 80. Van Hattem WA, Brosens LAA, De Leng WWJ, Morsink FH, Lens S, Carvalho R, et al. Large genomic deletions of SMAD4, BMPR1A and PTEN in juvenile polyposis. *Gut*. 2008;57(5):623–7.
 81. Mallet C, Lamribet K, Giraud S, Dupuis-Girod S, Feige JJ, Bailly S, et al. Functional analysis of endoglin mutations from hereditary hemorrhagic telangiectasia type 1 patients reveals different mechanisms for endoglin loss of function. *Hum Mol Genet*. 2015;24(4):1142–54.
 82. Krawczak M, Ball E, Fenton I, Stenson P, Abeyasinghe S, Thomas N, et al. Human gene mutation database-a biomedical information and research resource. 2000;15:45–51.
 83. Ali BR, Ben-Rebeh I, John A, Akawi NA, Milhem RM, Al-Shehhi NA, et al. Endoplasmic reticulum quality control is involved in the mechanism of endoglin-mediated hereditary haemorrhagic telangiectasia. *PLoS One*. 2011;6(10).
 84. McDonald J, Wooderchak-Donahue W, VanSant Webb C, Whitehead K, Stevenson DA, Bayrak-Toydemir P. Hereditary hemorrhagic telangiectasia: Genetics and molecular diagnostics in a new era. *Front Genet*. 2015;5(JAN):1–9.
 85. Albiñana V, Zafra MP, Colau J, Zarrabeitia R, Recio-Poveda L, Olavarrieta L, et al. Mutation affecting the proximal promoter of Endoglin as the origin of hereditary hemorrhagic telangiectasia type 1. *BMC Med Genet*. 2017;18(1):1–9.
 86. Plumitallo S, Ruiz-Llorente L, Langa C, Morini J, Babini G, Cappelletti D, et al. Functional analysis of a novel ENG variant in a patient with hereditary hemorrhagic telangiectasia (HHT) identifies a new Sp1 binding-site. *Gene*.

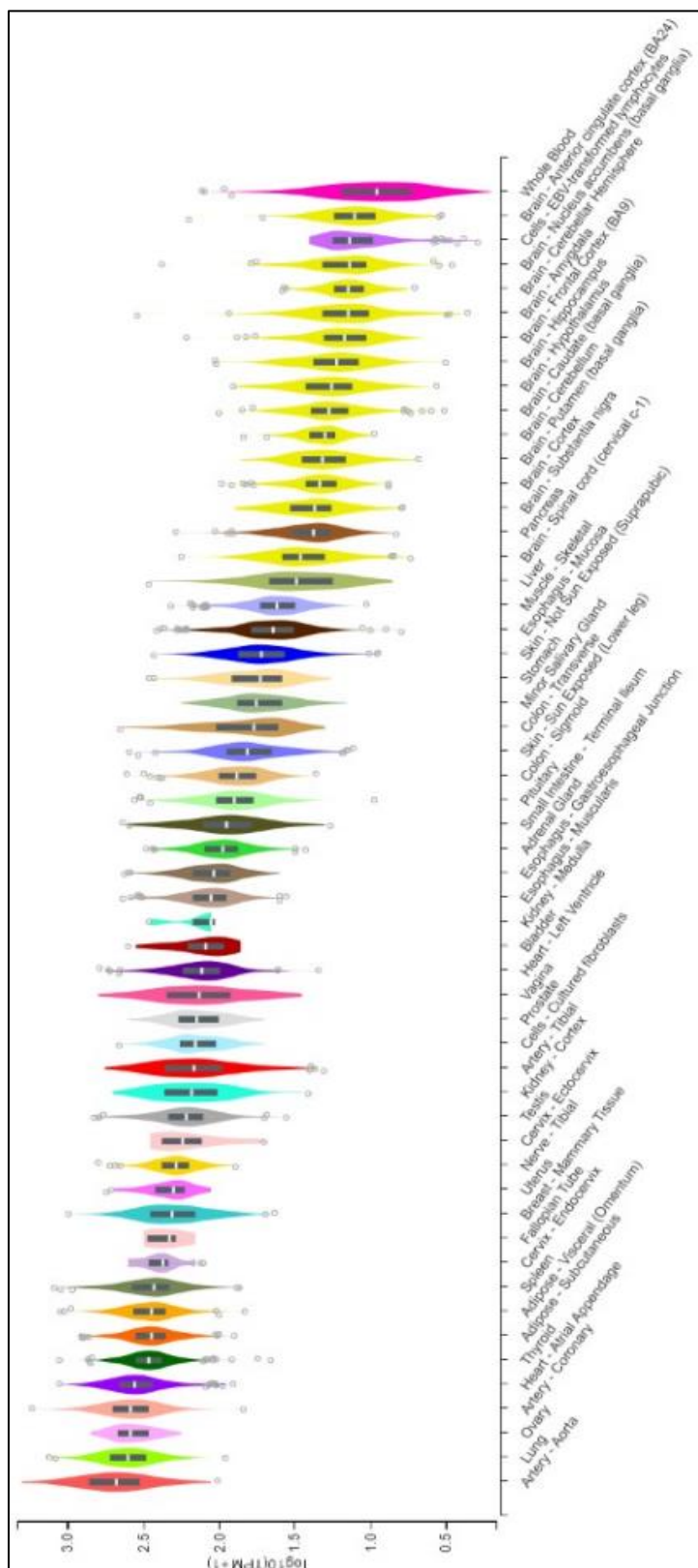
- 2018;647:85–92. Available from: <https://doi.org/10.1016/j.gene.2018.01.007>
87. Barreto ML, Cunha SS, Alcântara-Neves N, Carvalho LP, Cruz AA, Stein RT, et al. Risk factors and immunological pathways for asthma and other allergic diseases in children: background and methodology of a longitudinal study in a large urban center in Northeastern Brazil (Salvador-SCAALA study). *BMC Pulm Med.* 2006 Dec 23;6(1):15. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/16796729>
 88. Figueiredo CA, Alcantara-Neves NM, Amorim LD, Silva NB, Carvalho LCP de, Cooper PJ, et al. Evidence for a modulatory effect of IL-10 on both Th1 and Th2 cytokine production: the role of the environment. *Clin Immunol.* 2011 Apr;139(1):57–64. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/21285005>
 89. Lima-Costa MF, Rodrigues LC, Barreto ML, Gouveia M, Horta BL, Mambrini J, et al. Genomic ancestry and ethnoracial self-classification based on 5,871 community-dwelling Brazilians (The Epigen Initiative). *Sci Rep.* 2015;5:9812. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/25913126>
 90. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *Am J Hum Genet.* 2007;81(3):559–75. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0002929707613524>
 91. Barrett J, Fry B, Maller J, Daly M. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics.* 2005;21(2):263–5.
 92. Neale B. Introduction to Linkage Disequilibrium, the HapMap, and Imputation. *Cold Spring Harbor Protocols.* Cold Spring Harb Protoc. 2010;2010(3):1–4.
 93. Mu W, Zhang W. Molecular Approaches, Models, and Techniques in Pharmacogenomic Research and Development. *Pharmacogenomics.* 2013;273–94.
 94. Magalhães T, Rani S, Nunes G, Costa DO, Figueiredo MA, Melo PS, et al. The correlation between ancestry and color in two cities of Northeast Brazil with contrasting ethnic compositions. *Eur J Hum Genet.* 2015;23(October 2014):984–9.
 95. Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, et al. Genomic evidence for the Pleistocene and recent population history of Native Americans. *Science (80-).* 2015;349(6250):aab3884.

96. Lindenau JD, Salzano FM, Guimarães LSP, Callegari-Jacques SM, Hurtado AM, Hill KR, et al. Distribution patterns of variability for 18 immune system genes in Amerindians - relationship with history and epidemiology. *Tissue Antigens*. 2013;82(3):177–85.
97. Shriver MD, Smith MW, Jin L, Marcini A, Akey JM, Deka R, et al. Ethnic-affiliation estimation by use of population-specific DNA markers. *Am J Hum Genet*. 1997;60(4):957–64.
98. Collins-Schramm HE, Chima B, Morii T, Wah K, Figueroa Y, Criswell LA, et al. Mexican American ancestry-informative markers: examination of population structure and marker characteristics in European Americans, Mexican Americans, Amerindians and Asians. *Hum Genet*. 2004;114:263–71.
99. Royal C, Novembre J, Fullerton S, Goldstein D, Long C, Bamshad M, et al. Inferring genetic ancestry: opportunities, challenges, and implications. *Am J Hum Genet*. 2010;86(5):661–73.
100. Boyle AP, Hong EL, Hariharan M, Cheng Y, Schaub MA, Kasowski M, et al. Annotation of functional variation in personal genomes using RegulomeDB. *Genome Research* 2012, 22(9):1790-1797. PMID: 22955989. *Genome Res*. 2012;22(9):1790–7.
101. Adzhubei I, Jordan D, Sunyaev S. Predicting functional effect of human missense mutations using PolyPhen-2. *Curr Protoc Hum Genet*. 2013;7(7):20.
102. Kumar P, Henikoff S, Ng P. Predicting the effects of coding nonsynonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009;4(7):1073–81.
103. Schwarz J, Cooper D, Schuelke M, Seelow D. MutationTaster2: mutation prediction for the deep-sequencing age. *Nat Methods*. 2014;11(4):361–2.
104. Ferrer-Costa C, Orozco M, de la Cruz X. Sequence-based prediction of pathological mutations. *Proteins*. 2004;57(4):811–9.
105. Pousada G, Balóira A, Fontán D, Núñez M, Valverde D. Mutational and clinical analysis of the ENG gene in patients with pulmonary arterial hypertension. *BMC Genet*. 2016;17(1):7–9. Available from: <http://dx.doi.org/10.1186/s12863-016-0384-3>
106. Shovlin CL, Hughes JMB, Scott J, Seidman CE, Seidman JG. Characterization of Endoglin and Identification of Novel Mutations in Hereditary Hemorrhagic Telangiectasia. *Am J Hum Genet*. 1997;61(1):68–79.

107. Lastella P, Sabbà C, Lenato G, Resta N, Lattanzi W, Gallitelli M, et al. Endoglin gene mutations and polymorphisms in Italian patients with hereditary haemorrhagic telangiectasia. *Clin Genet*. 2003;63(6):536–40.
108. Lesca G, Plauchu H, Coulet F, Lefebvre S, Plessis G, Odent S, et al. Molecular Screening of ALK1/ACVRL1 and ENG Genes in Hereditary Hemorrhagic Telangiectasia in France. *Hum Mutat*. 2004;23(4):289–99.
109. Abdalla SA, Cymerman U, Rushlow D, Chen N, Stoeber GP, Lemire EG, et al. Novel mutations and polymorphisms in genes causing hereditary hemorrhagic telangiectasia. *Hum Mutat*. 2005;25(3):1–8.
110. Paquet M-E. Analysis of several endoglin mutants reveals no endogenous mature or secreted protein capable of interfering with normal endoglin function. *Hum Mol Genet*. 2001;10(13):1347–57.
111. Franciscatto AC, Ludwig FS, Matte US, Mota S, Stefani MA. Replication Study of Polymorphisms Associated With Brain Arteriovenous Malformation in a Population From South of Brazil. *Cureus*. 2016;8(2):1–10.
112. Pawlikowska L, Tran MN, Achrol AS, Ha C, Burchard E, Choudhry S, et al. Polymorphisms in Transforming Growth Factor-Beta-Related Genes ALK1 and ENG Are Associated With Sporadic Brain Arteriovenous Malformations. *Stroke*. 2005 Oct 1;36(10):2278–80. Available from: <http://stroke.ahajournals.org/cgi/doi/10.1161/01.STR.0000182253.91167.f8>
113. Boshuisen K, Brundel M, Kovel CGF De, Tom G, Pawlikowska L, Koeleman BPC, et al. Polymorphisms in ACVRL1 and Endoglin genes are not associated with sporadic and HHT related brain AVMs in Dutch patients. *Natl Institutes Heal*. 2013;4(3):375–8.
114. Kamiryo T, Nelson PK, Bose A, Zalzal P, Jafar JJ. Familial arteriovenous malformations in siblings. *Surg Neurol*. 2000 Mar;53(3):255–9. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0090301999001706>
115. Cartegni L, Wang J, Zhu Z, Zhang MQ, Krainer AR. ESEfinder: A web resource to identify exonic splicing enhancers. *Nucleic Acids Res*. 2003;31(13):3568–71.
116. Plumitallo S, Ruiz-Llorente L, Langa C, Morini J, Babini G, Cappelletti D, et al. Functional analysis of a novel ENG variant in a patient with hereditary hemorrhagic telangiectasia (HHT) identifies a new Sp1 binding-site. *Gene*. 2018 Mar;647:85–92. Available from: <http://linkinghub.elsevier.com/retrieve/pii/S0378111918300076>

117. Rauscher R, Ignatova Z. Timing during translation matters: Synonymous mutations in human pathologies influence protein folding and function. *Biochem Soc Trans.* 2018;46(4):937–44.
118. Letteboer TGW, Benzinou M, Merrick C, Quigley D, Zhau K, Kim IJ, et al. Genetic variation in the functional ENG allele inherited from the non-affected parent associates with presence of pulmonary arteriovenous malformations in hereditary hemorrhagic telangiectasia 1 (HHT1) and may influence expression of PTPN14. *Front Genet.* 2015;5:1–9.
119. Srinivas SK, Morrison AC, Andrela CM, Elovitz MA. Allelic variations in angiogenic pathway genes are associated with preeclampsia. *Am J Obstet Gynecol.* 2010;202(5):445.e1-445.e11. Available from: <http://dx.doi.org/10.1016/j.ajog.2010.01.040>
120. Permuth-Wey J, Fulp WJ, Reid BM, Chen Z, Georgeades C, Cheng JQ, et al. STAT3 polymorphisms may predict an unfavorable response to first-line platinum-based therapy for women with advanced serous epithelial ovarian cancer. *Int J Cancer.* 2016;138(3):612–9.
121. De Ruyck K, Sabbe N, Oberije C, Vandecasteele K, Thas O, De Ruyscher D, et al. Development of a multicomponent prediction model for acute esophagitis in lung cancer patients receiving chemoradiotherapy. *Int J Radiat Oncol Biol Phys.* 2011;81(2):537–44.
122. Wiemels J. Perspectives on the causes of childhood leukemia. *Chem Biol Interact.* 2012 Apr 5;196(3):59–67. Available from: <http://www.ncbi.nlm.nih.gov/pubmed/22326931>

ANEXO - Níveis de expressão do gene da endogлина em diferentes tecidos.



Nota: Os valores de expressão são mostrados em TPM (Transcrições Por Milhão) em escala logarítmica, mostrados em box plot para cada tecido específico, em ordem decrescente do nível de expressão. Fonte: GTEx Analysis Release V8 (GTEx Project, <https://www.gtexportal.org/>).