



UNIVERSIDADE ESTADUAL DE CAMPINAS

Faculdade de Ciências Médicas

AMANDA DONATTI

**IDENTIFICAÇÃO DE VARIAÇÕES ESTRUTURAIS (CNVS) E
MODIFICAÇÕES EPIGENÉTICAS EM PACIENTES COM ACIDENTE
VASCULAR CEREBRAL**

Campinas

2016

AMANDA DONATTI

**IDENTIFICAÇÃO DE VARIAÇÕES ESTRUTURAIS (CNVS) E
MODIFICAÇÕES EPIGENÉTICAS EM PACIENTES COM ACIDENTE
VASCULAR CEREBRAL**

Dissertação apresentada à Faculdade de Ciências Médicas da
Universidade Estadual de Campinas como parte dos requisitos
exigidos para a obtenção do título de Mestra em Ciências.

Orientadora: ISCIA TERESINHA LOPES CENDES

Coorientador: RODRIGO SECOLIN

ESTE EXEMPLAR CORRESPONDE À VERSÃO
FINAL DA DISSERTAÇÃO DEFENDIDA PELA
ALUNA AMANDA DONATTI, E ORIENTADA PELA
PROFA. DRA ISCIA TERESINHA LOPES CENDES.

Campinas

2016

Agência(s) de fomento e nº(s) de processo(s): FAPESP, 2013/24932-0

Ficha catalográfica
Universidade Estadual de Campinas
Biblioteca da Faculdade de Ciências Médicas
Ana Paula de Moraes e Oliveira - CRB 8/8985

D715i Donatti, Amanda, 1991-
Identificação de variações estruturais (CNVs) e modificações epigenéticas em pacientes com acidente vascular cerebral / Amanda Donatti. – Campinas, SP : [s.n.], 2016.

Orientador: Iscia Teresinha Lopes-Cendes.
Coorientador: Rodrigo Secolin.
Dissertação (mestrado) – Universidade Estadual de Campinas, Faculdade de Ciências Médicas.

1. Acidente vascular cerebral. 2. Análise de sequência com séries de oligonucleotídeos. 3. Metilação de DNA. I. Lopes-Cendes, Iscia Teresinha, 1964-. II. Secolin, Rodrigo. III. Universidade Estadual de Campinas. Faculdade de Ciências Médicas. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Identification of structural variants (CNVs) and epigenetic changes in patients with stroke

Palavras-chave em inglês:

Stroke

Oligonucleotide array sequence analysis

DNA methylation

Área de concentração: Neurociências

Titulação: Mestra em Ciências

Banca examinadora:

Iscia Teresinha Lopes-Cendes [Orientador]

Wagner Mauad Avelar

Rodrigo Bazan

Data de defesa: 29-01-2016

Programa de Pós-Graduação: Fisiopatologia Médica

BANCA EXAMINADORA DA DEFESA DE MESTRADO

AMANDA DONATTI

ORIENTADORA PROFA. DRA. ISCIA TERESINHA LOPES CENDES

COORIENTADOR PROF. DR. RODRIGO SECOLIN

MEMBROS:

1. PROFA. DRA. ISCIA TERESINHA LOPES CENDES

2. PROF. DR. WAGNER MAUAD AVELAR

3. PROF. DR. RODRIGO BAZAN

Programa de Pós-Graduação em Fisiopatologia Médica da Faculdade de Ciências Médicas da Universidade Estadual de Campinas.

A ata de defesa com as respectivas assinaturas dos membros da banca examinadora encontra-se no processo de vida acadêmica do aluno.

Data: 29/01/2016

Dedicatória

“Dedico este trabalho a minha família, que sempre confiou em mim e sem os quais não conseguiria chegar até aqui”.

Agradecimentos

Agradeço primeiramente a Deus, que me permitiu chegar até aqui e guiou os meus caminhos.

A minha orientadora Dra. Iscia Lopes-Cendes que me deu a oportunidade de fazer meu mestrado sob sua orientação e me acolheu em seu laboratório, pelos ensinamentos e valores que me foram passados e pelo grande exemplo profissional que se tornou para mim.

Ao Dr. Rodrigo Secolin pela co-orientação, ensinamentos, valores, amizade, paciência e todo conhecimento que me foi passado durante esses anos.

ADra. DanyellaDogini, por todo seu tempo, ajuda, paciência na elaboração desse estudo.

A FAPESP pelo apoio financeiro durante a realização deste mestrado.

Ao Laboratório de Biologia Molecular da Universidade Regional de Joinville/SC, pela colaboração e envio das amostras, principalmente aos colaboradores: Leslie Ecker Ferreira, Dr. Paulo França e Dr. Norberto Cabral pelo carinho, atenção e esforço dados durante o projeto, sem o qual este não seria realizado.

Aos Drs. Fabio Rossi Torres e Benilton Carvalho, por toda a ajuda e suporte na realização dos experimentos, análises estatísticas e de bioinformática.

A minha família, principalmente, minha mãe, avó, tios e primos, que sempre acreditaram e investiram em mim, me apoiaram acima de tudo em qualquer decisão tomada por mim e acreditaram em mim e na minha capacidade, sempre me apoiando, chorando e festejando comigo.

A meu namorado Murilo por todo apoio, compreensão, paciência, pelo companheirismo e conselhos durante todo o tempo.

Aos companheiros de laboratório e amigos Alexandre, Amanda, Camila, Isabella, Marcella, Simoni, Vanessa S. e Vanessa Z, que sempre estiveram comigo nos momentos bons e difíceis, e com quem compartilhei muitos momentos nesses últimos anos.

A todos os outros amigos e companheiros de laboratório André, Cristiane, Cynthia, Fernando, Karina, Luciana Bonadia, Luciana Rezende, Marcella, Marilza, Marina, Murilo, Patrícia, Patrícia Aline, Renata, Roby, Saul, Thais, Tânia, Tatiane e Welliton, meu muito obrigado por tudo.

Aos amigos de longa data, que mesmo longe continuam presentes.

Muito Obrigada

Amanda Donatti

Resumo

O acidente vascular cerebral (AVC) é uma condição heterogênea multifatorial, responsável pela interrupção do fluxo sanguíneo para o cérebro. Em geral, o AVC pode ser classificado em dois grandes grupos principais: hemorrágico e isquêmico. Estudos têm demonstrado a presença de fatores genéticos no AVC. Estudos de meta-análise revelaram 19 genes associados ao AVC isquêmico. Contudo, variantes funcionais ainda não foram identificadas em 8 desses genes candidatos (*HDAC9*, *CDKN2B-AS1*, *PDE4D*, *PITX2*, *ZFH3*, *ANGPT1*, *AGTRL1* e *ALOX5AP*). Além disso, em ensaios *in vivo*, foi possível observar que padrões alterados de metilações no DNA podem influenciar o desenvolvimento de AVC ou modificações nas proteínas envolvidas nesse processo, o que pode ser responsável pelo aumento do risco de AVC. Estudos para AVC hemorrágico vêm demonstrando forte influência genética nesses eventos, porém, ainda são menos recorrentes, devido à menor frequência desse evento na população e à sua maior taxa de mortalidade. Variações no número de cópias genômicas (CNVs) são mutações comuns no genoma humano e têm sido apontadas como importantes causas de alterações na expressão gênica. Poucos estudos têm demonstrado a relação entre genes e CNVs encontrados em pacientes com AVC isquêmico e hemorrágico. Os objetivos principais desse trabalho são: identificar fatores epigenéticos na região promotora de genes candidatos em pacientes com AVC isquêmico e identificar variações nos números de cópias (CNVs) em pacientes com AVC hemorrágico. Para isso, foram utilizadas amostras de DNA de pacientes com AVC e indivíduos controle provenientes de Joinville/SC. Através de análises de bioinformática, foram identificadas regiões potencialmente metiláveis (CpGs) nos promotores de genes candidatos para AVC isquêmico. Esses genes tiveram suas regiões promotoras analisadas através da técnica de PCR por metilação específica (MS-PCR). Para análise de CNVs em pacientes com AVC hemorrágico, foram realizados chips de microarranjos de DNA baseados em SNPs (*Genome-Wide Human SNP Array 6.0*; Affymetrix Inc.) e os resultados foram analisados no software Genotyping Console® (Affymetrix Inc.). Os estudos de metilação não geraram resultados significativos devido à não amplificação das ilhas CpGs nas regiões promotoras dos genes candidatos nas

condições técnicas inicialmente testadas nesse estudo. Nos estudos de CNVs foram analisadas 45 amostras de pacientes e 41 amostras de controles, de modo que regiões de ganho e perda genômica de interesse foram identificadas. Através de análises no software METACORE, foi possível identificar 19 genes relacionados ao AVC. Validando esses resultados em 18 pacientes italianos, foi possível selecionar 2 genes em comum as duas coortes: *MACROD2* e *NSF*. O gene *MACROD2* não apresenta variantes descritas na população geral, se tornando forte candidato a influenciar a ocorrência do AVC. Com tais resultados, podemos concluir que variantes genômicas estruturais na forma de CNV sem alguns genes candidato podem ter um papel na predisposição ao AVC hemorrágico. Nossos resultados foram confirmados em duas populações distintas de pacientes com AVC hemorrágico.

Palavras-Chave: AVC; Análise de sequência com séries de oligonucleotídeos; Metilação do DNA

Abstract

Stroke is a heterogeneous and multifactorial condition, responsible for blockage in blood flow to the brain. Stroke can be classified in two main groups: ischemic and hemorrhagic. Many studies demonstrate the presence of genetic factors on stroke. Meta-analysis studies revealed 19 genes related to ischemic stroke. However, functional variants were not identified in 8 candidate genes (*HDAC9*, *CDKN2B-AS1*, *PDE4D*, *PITX2*, *ZFHX3*, *ANGPT1*, *AGTRL1* and *ALOX5AP*). In addition, *in vivo* studies show that abnormal methylation pattern in DNA can influence stroke development or changes in proteins related to this phenotype, which leads to increased risk of stroke. Although some studies demonstrate influence of genetic variants in hemorrhagic stroke, studies with this subtype are less recurrent due to its low population frequency and higher mortality rate when compared to ischemic events. Copy number variations (CNVs) are common mutations which have been identified as significant causes of change in gene expression. Few studies have demonstrated a relation between genes and CNVs for ischemic and hemorrhagic stroke. The main objectives of this study are: to identify epigenetic factor on promoter region of candidate genes to ischemic stroke; and to identify copy number variations in patients with hemorrhagic stroke. We use DNA sample from patients with stroke provided from Joinville/SC. Through bioinformatics analysis, we identified regions potentially methylated (CpGs) on promoter regions of candidate genes from ischemic stroke. These candidate genes had their promoter region analyzed by methyl-specific PCR (MS-PCR). To analyze CNVs in patients with hemorrhagic stroke, we used microarray chips based on SNPs (*Genome-Wide Human SNP Array 6.0*; Affymetrix Inc.), whose results were analyzed by Genotyping Console® Software (Affymetrix Inc.). Our methylation studies did not resulted in any significant findings, since the CpGs islands in the promoter region of candidate genes were not amplified using the conditions initially tested in this study. In the CNVs studies, we analyzed 45 samples from patients and 41 samples from controls and identified 19 genes potentially related to stroke. We validated these results in a group of Italian patients, analyzing 18 patients with hemorrhagic stroke. Overall, we identified 2 genes which were present in both

cohorts: *MACROD2* and *NSF*. Gene *MACROD2* does not have variants described in the normal general population genetic database, which makes it a potential candidate gene to influence hemorrhagic stroke. In conclusion, we have demonstrated for the first time that CNVs in certain candidate genes may influence the risk for hemorrhagic stroke. Our findings were confirmed in two independent samples of patients with hemorrhagic stroke.

Keywords: Stroke; Oligonucleotide array sequence analysis; DNA methylation

Lista de Ilustrações

Figura 1 - Taxa de mortalidade mundial referente a 2012 com base nas 10 principais doenças que acometem a população.	22
Figura 2 - Comparação entre as taxas de mortalidade mundiais entre os anos 2000 e 2012, segundo a OMS.	22
Figura 3 - Taxas de mortalidade mundial em 2012 distribuídas por região econômica.	23
Figura 4 -Esquematização dos principais subtipos de AVC isquêmico.	24
Figura 5 – Exemplos das classificações de AVC hemorrágico de acordo com as localizações das hemorragias.	26
Figura 6 – Conversão do DNA pelo tratamento com bissulfito.	34
Figura 7 - Distribuição das idades das amostras de indivíduos controles, pacientes com AVC isquêmico e AVC hemorrágico provenientes do Biobanco de AVC em Joinville/SC.	44
Figura 8 - Distribuição do gênero entre as populações amostrais de pacientes com AVC isquêmico, AVC hemorrágico e controles provenientes do Biobanco de AVC de Joinville, SC.	45
Figura 9 - Comparação entre as idades das populações amostrais (pacientes com AVC hemorrágico e indivíduos controles), provenientes de colaboração com a Universidade de Pisa/IT.	48
Figura 10 - Descrição do gênero das amostras de indivíduos controles e de pacientes com AVC hemorrágico provenientes de Pisa/IT.	49
Figura 11 - Ilha CpG identificada (em destaque) para o gene AGTRL1, a partir de análise de 1000pb upstream e primeiro éxon.	55
Figura 12 - Identificação de possíveis ilhas CpGs no gene AGTRL1 e os primers desenhados para a região mais estável.	55
Figura 13 - Identificação da segunda região CpG do gene AGTRL1.	57
Figura 14 - Desenho de primers para a segunda região CpG do gene AGTRL1.	57
Figura 15 - Identificação de ilhasCpGs (em destaque) para o gene ALOX5AP, a partir de análise de 1000pb upstream e primeiro éxon.	59
Figura 16 - Identificação de ilhas CpG (em destaque) para o gene ANGPT1, a partir de análise de 1000pb upstream e primeiro éxon.	60

Figura 17 - Ilha CpG identificada (em destaque) para o gene CDKN2B-AS1, a partir de análise de 1000pb upstream e primeiro éxon	61
Figura 18 - Desenho de pares de primers para análise de possível ilha CpG do gene CDKN2b-AS1.	61
Figura 19 - Identificação de Ilhas CpGs (em destaque) para o gene HDAC9, a partir de análise de 1000pb upstream e primeiro éxon.	63
Figura 20 - Desenho de pares de primers para análise das possíveis ilhasCpGs do gene HDAC9.	63
Figura 21 - Identificação de ilhas CpG (em destaque) para o gene PDE4D, a partir de análise de 1000pb upstream e primeiro éxon.	64
Figura 22 - Identificação de ilhas CpG (em destaque) para o gene PDE4D, a partir de análise de 1000pb upstream e primeiro éxon.	65
Figura 23 - Identificação de ilhas CpG (em destaque) para o gene ZFH3, a partir de análise de 1000pb upstream e primeiro éxon.	66
Figura 24 - Desenho de pares de primers para análise das possíveis ilhas CpGs do gene ZFH3.	67
Figura 25 - PCA plot das amostras provenientes de Joinville/SC, comparando indivíduos controles e pacientes com AVC hemorrágico.	71
Figura 26 - Regiões com altas taxas de CNVs (A) por amostras e (B) por região cromossômica.	73
Figura 27 - PCA comparando as amostras de pacientes brasileiros (círculo vermelho), controles brasileiros (triângulo azul), pacientes italianos (quadrado cinza) e controles italianos (losango verde).	77
Figura 28 - PCA comparando as amostras pacientes italianos (quadrado cinza) e controles italianos (losango verde).	78

Lista de Tabelas

Tabela 1 - Genes identificados nos estudos genéticos (Meta-análises e GWAS) para o fenótipo geral de AVC isquêmico.	29
Tabela 2 - Genes identificados através de estudos genéticos e candidatos às alterações relacionadas com os subtipos de AVC isquêmico.	30
Tabela 3- Genes descritos na literatura como relacionados a AVC hemorrágico.	31
Tabela 4 – Comparação entre os fatores de risco que acometem as amostras controle e pacientes com AVC isquêmico.	46
Tabela 5 - Comparação entre os fatores de risco que acometem as amostras controle e pacientes com AVC hemorrágico.	47
Tabela 6–Descrição dos fatores de risco que acometem as amostras de pacientes com AVC hemorrágico provenientes de Pisa/IT.	50
Tabela 7 - Temperaturas de anelamento específicas para cada par de primer obtidos por análises in silico para ilhas CpGs nas regiões promotoras dos genes AGTRL1, CDKN2B-AS1, HDAC9 e ZFH3.	54
Tabela 8 - Descrição dos primers selecionados para a primeira região CpG do gene AGTRL1, dentre os quais os primers F1/R1 (em destaque) foram selecionados.	56
Tabela 9 - Descrição dos primers selecionados para a segunda região CpG do gene AGTRL1, dentre os quais os primers F3/R3 e F4/R4 (em destaque) foram selecionados.	58
Tabela 10 - Descrição dos primers sugeridos para as ilhas CpGs do gene CDKN2B-AS1.	62
Tabela 11 - Descrição dos primers sugeridos para a primeira ilha CPG encontrada no gene HDAC9.	64
Tabela 12 - Descrição dos primers selecionados para as regiões CpGs do gene ZFH3. Os pares de primers selecionados foram F5/R5 e F4/R4 (em destaque), que abrangem a primeira e segunda ilha, respectivamente.	67
Tabela 13- Genes identificados pelo METACORE como potencialmente relacionados com processos neurológicos e vasculares, potencialmente	74

relacionados com AVC.

Tabela 14 - Genes identificados no banco de dados DGV, que não apresentam variantes descritas que englobem toda a sequência gênica. 75

Tabela 15 - Descrição das variantes inéditas de acordo com o banco de dados DGV. 75

Tabela 16 - Genes identificados pelo banco de dados DECIPHER, como relacionados com fenótipos semelhantes ao AVC. 76

Tabela 17 – Descrição das variantes inéditas encontradas no estudo com relação aos fenótipos descritos no banco de dados DECIPHER. 77

Tabela 18 – Genes identificados na coorte italiana e após análise pelo software METACORE na coorte brasileira. 79

Lista de Abreviaturas e Siglas

<i>ACT</i>	Antiquimotripsina alfa 1
<i>AGTRL1</i>	Receptor de Apelina
<i>ALOX5AP</i>	Proteína Araquidonato de Ativação de 5-Lipoxigenase
<i>ANG2</i>	Angiopoitina 2
<i>ANGPT1</i>	Angiopoitina 1
<i>ANXA8</i>	Anexina A8
<i>APL</i>	Leucemia Aguda Promielocítica
<i>APO E</i>	Apolipoproteína E
<i>Apo(a)</i>	Gene Apolipoproteína
<i>AVC</i>	Acidente Vascular Cerebral
<i>BS-PCR</i>	PCR para Sequenciamento por Bissulfito
<i>CADASIL</i>	Leucoencefalopatia na Forma Autossômica Dominante
<i>CARASIL</i>	Leucoencefalopatia na Forma Autossômica Recessiva
<i>CDH13</i>	Caderina 13
<i>CDKN2B-AS1</i>	Inibidor de Quinase Dependente de Ciclina 2B - RNA Antissense 1
<i>Chr</i>	Cromossomo
<i>CNV</i>	Variação no Número de Cópias
<i>COL3A1</i>	Colágeno tipo 3 alfa 1
<i>COL4A1</i>	Colágeno, Tipo IV, Alfa 1
<i>COL4A2</i>	Colágeno, Tipo IV, Alfa 2
<i>COL5A2</i>	Colágeno tipo 5 alfa 2
<i>CpG</i>	Combinação da Estrutura "Citosina + Fosfato + Guanina"
<i>CRP</i>	Gene da Proteína C-reativa
<i>CYP2E1</i>	Citocromo P450 Família 2, Subfamília E, Membro 1
<i>CYP2E9</i>	Cytochrome P450, Família 2, Subfamília C, Polipeptídeo 9
<i>DECIPHER</i>	Banco de Dados de Desequilíbrio Cromossômico e Fenotípico em Seres Humanos utilizando Recursos do Ensembl
<i>DEFB103A</i>	Defensina 103 Alfa
<i>DEFB103B</i>	Defensina 103 Beta
<i>DEFB4B</i>	Defensina 4 Beta
<i>DGV</i>	Banco de Dados de Variantes Genômicas

DNA	Ácido Desoxirribonucleico
DNTP	Desoxirribonucleotídeo Fosfatado
<i>DUPSS22</i>	Fosfatase Dupla Especificidade 22
EDTA	Ácido Etilenodiamino Tetra-Acético
<i>eNOS</i>	Síntese de Óxido Nítrico Endotelial
F(n)/R(n)	Conjunto de Primers Forward e Reverse; sendo (n) o número de cada par
<i>FBXW7</i>	F-box e domínio de repetição WD contendo 7
<i>FGL2</i>	Proteína tipo Fibrinogênio 2
<i>FNTA</i>	Farnesiltransferase alfa
<i>FXIII</i>	Fator de Coagulação XIII
GWAS	Estudos de Associação de Genoma Completo
HAS	Pressão Sistêmica Arterial
<i>HCHWA-D</i>	Amiloidose Tipo Holandês
<i>HCHWA-I</i>	Amiloidose Tipo Islandês
<i>HDAC9</i>	Histona-Desacetilase 9
<i>HGSNAT</i>	Heparano-Alfa-Glucosaminidase N-Acetiltransferase
ID	Identificação
<i>IL-6</i>	Interleucina 6
InDel	Inserção e Deleção Genômica
IT	Itália
kb	Kilo Base
<i>LIMK1</i>	Quinase de Domínio LIM 1
M	Molar
<i>MACROD2</i>	Domínio MACRO contendo 2
MAPK	Proteínas Quinases Ativadas por Mitógenos
Mb	Mega Base
<i>MBTPS1</i>	Fator de Transcrição de Peptidase Ligado à Membrana, sítio 1
MCAO	Oclusão da Artéria Média Cerebral
mg	Miligrama
mL	Mililitro
<i>MLYCD</i>	Malonil-CoA Decarboxilase
mM	Mili Molar

MS-PCR	PCR Metil Específico
MTHFR	Metileno TetrahidrofolatoRedutase
NA	Não respondeu / Desconhece
<i>NAALADL2</i>	N-acetylated alpha-linked acidic dipeptidase-like2
NaCl	Cloreto de Sódio
<i>NBPF1</i>	Ponto de Quebra do Neuroblastoma, Membro 1
ng	Nanograma
<i>NINJ2</i>	Ninjurin 2
nm	Nanômetro
<i>NOMO1</i>	Modulador NODAL 1
<i>NPY4R</i>	Receptor 4 de Neuropeptídeos Y
<i>NSF</i>	Fator de Sensibilidade a N-etilmaleimida
OMS	Organização Mundial da Saúde
<i>OPG</i>	Osteoprotegerina
pb	Par de base
PCA	Análise de Componente Principal
PCR	Reação em Cadeia da Polimerase
<i>PDE4D</i>	Fosfodiesterase 4D
<i>PET112L</i>	RNA transportador-GlutamilAmidotransferase, subunidade B
<i>PFKFB3</i>	6 fosfofructo-2-quinase/frutose-2,6-bifosfatase 3
pg	Pico grama
pH	Potencial Hidrogeniônico/Potencial de Hidrogênio
<i>PITX2</i>	Junção de Homodomínio 2
<i>PKD1-like</i>	Gene tipo 1 da Doença Renal Policística
<i>PMF1</i>	Factor de Modulação da Poliamina-1
<i>PRKCH</i>	Proteína Quinase C
<i>PRODH</i>	Prolina Oxidase 1
<i>RAF1</i>	Proto-oncogene Raf-1, Quinase de Treonina/Serina
<i>RB1CC1</i>	Indução de Espiral Enrolada RB1
<i>RBM17</i>	Proteína de Sítio de Ligação de RNA 17
RNA	Ácido Ribonucleico
rpm	Rotação por minuto
RSB	Tampão de Parada de Reação

SC	Santa Catarina
SDS	Dodecil Sulfato de Sódio
<i>SELE</i>	Selectina E
<i>SLC25A44</i>	Transportador de Sóluto, Família 25, membro 44
SNP	Polimorfismo de Base Única
<i>SNTA</i>	Sintrofina alfa 1
TE	Tris-HCl + EDTA
T _m	Temperatura de Melting
<i>TNF</i>	Fator de Necrose Tumoral
TOAST	"Trial of Org 10172 in Acute Stroke Treatment"
Tris-HCl	TrizmaHidroclorídrica
<i>UCP3</i>	Gene da Proteína de Desacoplamento 3
μL	Micro litro
UNIVILLE	Universidade Regional de Joinville
UTR	Região Não Traduzida
VEGF	Fator de Crescimento do Endotélio Vascular
VS	Versus
<i>ZFX3</i>	Homeobox Indicador de Zinco 3

Sumário

Resumo

Abstract

Lista de Ilustrações

Lista de Tabelas

Lista de Abreviaturas e Siglas

1. Introdução	21
1.1. Acidente Vascular Cerebral	21
1.1.1. Acidente Vascular Cerebral Isquêmico	23
1.1.2. Acidente Vascular Cerebral Hemorrágico	25
1.2. Genética do Acidente Vascular Cerebral	27
1.3. Epigenética	32
1.4. Variação no Número de Cópias Genômicas	35
2. Objetivos	37
Capítulo I – Coorte	38
1. Caracterização Amostral	38
1.1. Biobanco de Joinville/SC	38
1.2. Colaboração com Universidade de Pisa/IT	40
2. Extrações de DNA	41
2.1. Método Fenol/Clorofórmio/Álcool-Isoamílico	41
2.2. Método Salting Out	42
2.3. Qualidade das Amostras	43
3. Resultados	43
3.1. Biobanco de Joinville/SC	43
3.2. Colaboração com Universidade de Pisa/IT	47
Capítulo II - Epigenética	51
1. Material e Métodos	51
1.1. Amostras	51
1.2. Identificação dos Sítios CpGs em Genes Candidatos	51
1.3. Tratamento por Bissulfito das Amostras	52
1.4. PCR Metil-Específico (MS-PCR)	53
2. Resultados e Discussão	54
2.1. Identificação dos Sítios CpGs em Genes Candidatos	54

2.2. Conversão por Bissulfito e MS-PCR	68
Capítulo III – Variação no Número de Cópias	69
1. Métodos	69
1.1. Amostras	69
1.1.1. Amostras provenientes do Biobanco de Joinville/SC	69
1.1.2. Amostras provenientes de colaboração com Universidade de Pisa/IT	69
1.2. Microarranjos de DNA	69
2. Resultados	71
2.1. Variação no Número de Cópias Genômicas de Pacientes com Acidente Vascular Cerebral Hemorrágico	71
2.1.1. Coorte Brasileira	71
2.1.2. Amostras provenientes de Pisa/IT	77
3. Discussão	80
4. Conclusões	89
5. Referências	90
Anexo I – Parecer Comitê de Ética	98

1. Introdução

1.1. Acidente Vascular Cerebral

O acidente vascular cerebral (AVC) é caracterizado por um bloqueio de fluxo sanguíneo para determinada região do cérebro, resultando em perdas de funções neurológicas[1, 2]. Dependendo da região afetada, o paciente pode sofrer danos nas funções motoras ou sensoriais[1]. Tais funções comprometidas variam de acordo com o subtipo de AVC, assim como sua gravidade ou local de ocorrência [1]. Essas perdas de funções podem ser temporárias ou definitivas, com efeitos silenciosos ou devastadores [1].

Atualmente, de acordo com dados da Organização Mundial da Saúde (OMS), as doenças cardiovasculares, principalmente isquemia cardíaca e AVC, abrangem as maiores causas de mortalidade mundial (Figura 1). Comparando-se as taxas de mortalidade mundiais entre 2000 e 2012, as doenças cardiovasculares apresentaram um aumento em sua ocorrência (Figura 2). Ao avaliar as taxas de mortalidade por grupos econômicos, o AVC é a principal causa de morte dentre os países com renda econômica média-alta, grupo no qual o Brasil está incluso, e entre as principais causas de morte nos grupos com renda econômica média-baixa e alta (Figura 3). Além disso, o AVC caracteriza-se como a principal causa de invalidez em adultos e um importante fator de risco para desenvolvimento de demência e problemas cognitivos relacionados à idade [3-5]. O tratamento depende do tipo e da região afetada, de modo que as causas do AVC e as regiões danificadas devem ser precisamente determinadas [1]. O AVC pode ser classificado em dois tipos principais: hemorrágico (OMIM #614519) e isquêmico (OMIM #60137) [1].

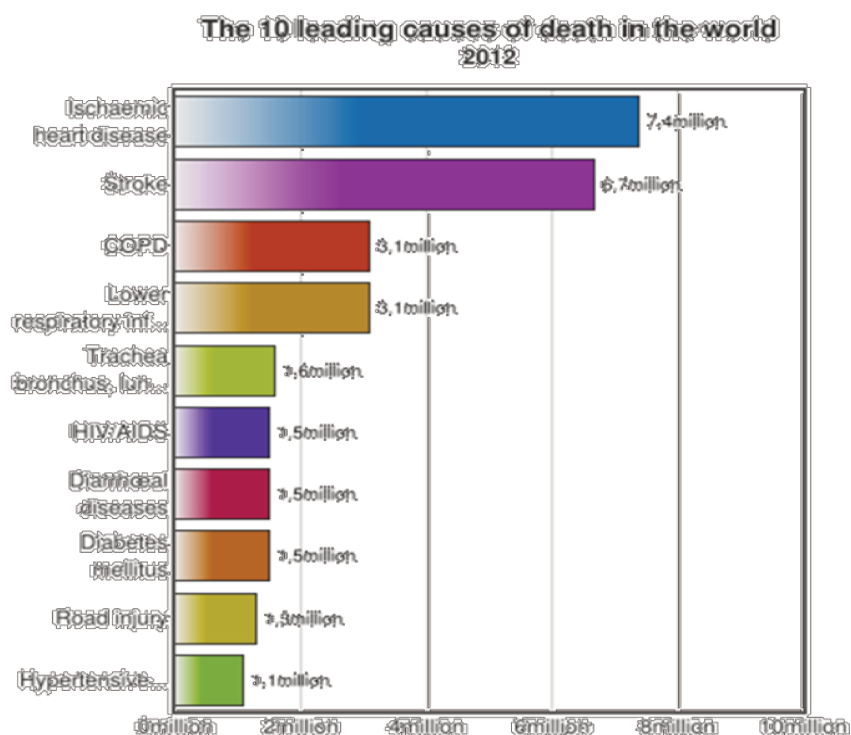


Figura 1 - Taxa de mortalidade mundial referente a 2012 com base nas 10 principais doenças que acometem a população. Isquemia cardíaca e AVC são as duas principais causas, acometendo 7,4 e 6,7 milhões de habitantes (OMS. Disponível em: <http://www.who.int/mediacentre/factsheets/fs310/en/>).

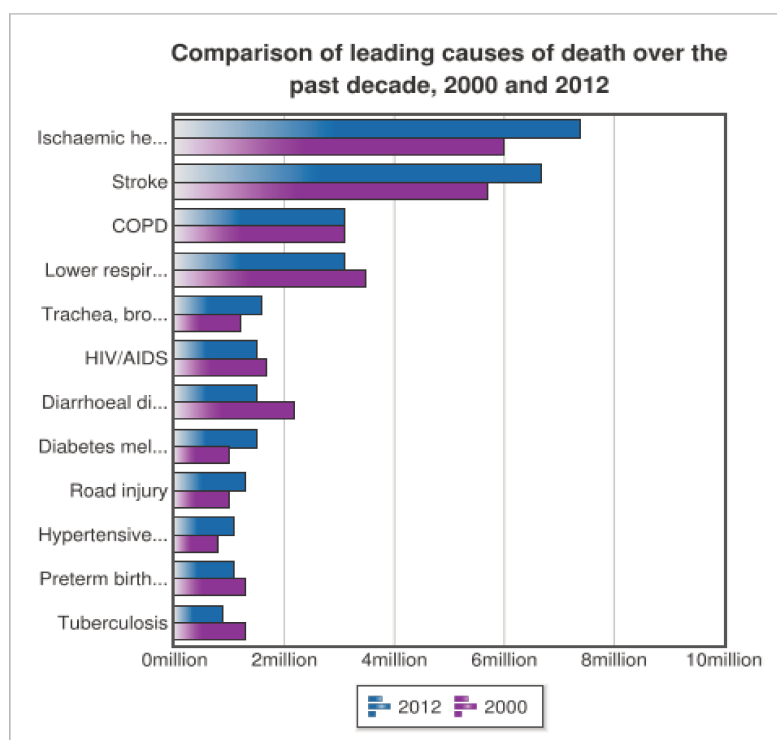


Figura 3 – Comparação entre as taxas de mortalidade mundiais entre os anos 2000 e 2012, segundo a OMS. O índice de letalidade de doenças cardiovasculares aumentou nesse período. (OMS. Disponível em: <http://www.who.int/mediacentre/factsheets/fs310/en/>)

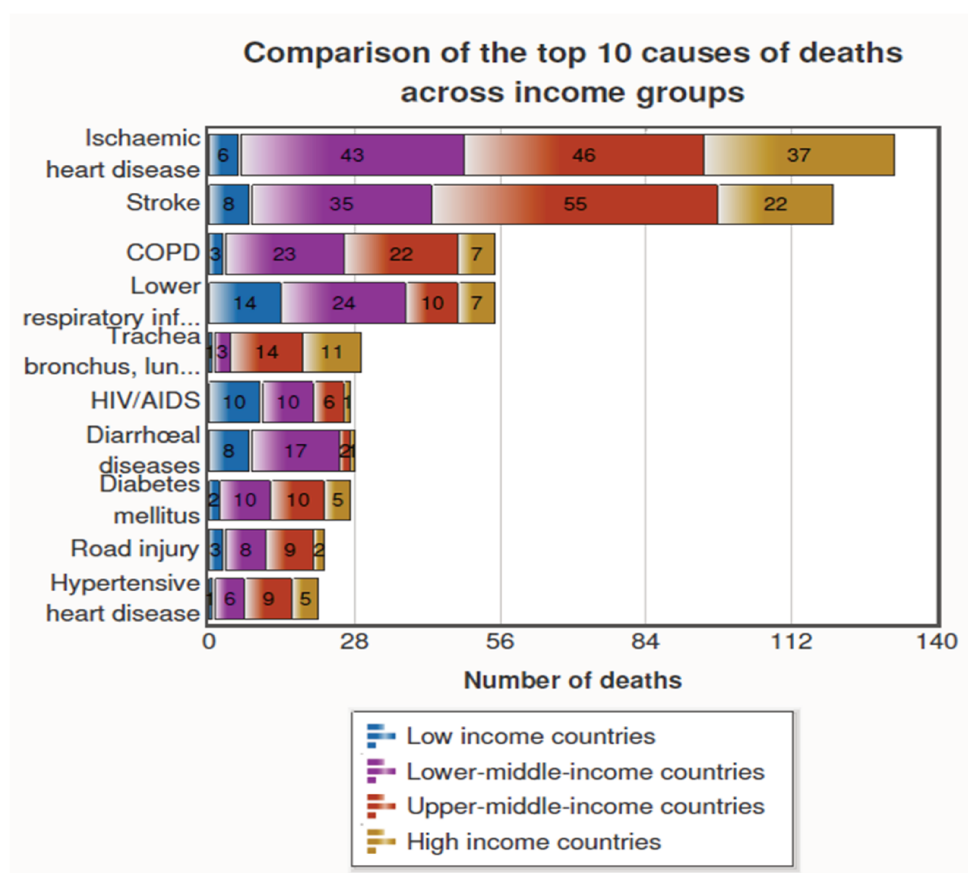


Figura 4 - Taxas de mortalidade mundial em 2012 distribuídas por região econômica: Baixa renda, renda média-baixa, renda média-alta e alta renda. É possível verificar que isquemia cardíaca e AVC são as principais causas de morte nos países com maior renda, sendo que o AVC acomete a maior parte da população dos países com renda média-alta, incluindo o Brasil. (OMS. Disponível em: <http://www.who.int/mediacentre/factsheets/fs310/en/index1.html#>)

1.1.1. Acidente Vascular Cerebral Isquêmico

O AVC isquêmico ocorre em 80 a 85% dos casos e é caracterizado pela oclusão das artérias cerebrais, impedindo o fluxo sanguíneo e o suprimento de oxigênio para a área afetada[1, 2,5]. O AVC isquêmico pode ser classificado em cinco subtipos, de acordo com o mecanismo de dano aos vasos sanguíneos ou causa da redução de fluxo sanguíneo para o cérebro: cardioembolismo, aterosclerose dos grandes vasos, oclusão dos pequenos vasos, outras causas e criptogênico[6].

O embolismo é caracterizado pela formação de um coágulo em uma artéria em uma região distante do cérebro e segue com a corrente sanguínea

para a região cerebral, levando à oclusão dos vasos [1, 2]. Esse coágulo pode ter origem em três regiões principais: coração (denominado cardioembólico), aorta ou nas principais artérias do pescoço que se encontram na mesma rota da circulação sanguínea (Figura 4)[1]. A aterosclerose dos grandes vasos leva à estenose e formação de coágulos nos vasos cerebrais, de modo que o aumento destes coágulos preenche a artéria, levando-a a obstrução (Figura 4)[1]. O terceiro subtipo é causado pela oclusão dos pequenos vasos cerebrais, ou lacunas, sendo que não há evidências de disfunção cortical[1]. Esses vasos podem ser qualquer estrutura vascular, sejam pequenas artérias, arteríolas e capilares, que estejam localizadas no parênquima cerebral ou no espaço subaracnóide[7]. Essa definição auxilia na diferenciação entre as doenças de grandes e pequenos vasos sanguíneos [7]. O AVC classificado como “outras causas” se refere a eventos causados por outros fenótipos que não sejam os fatores de risco principais, podendo ter sido causado por doenças e síndromes genéticas não relacionadas diretamente ao AVC, por exemplo[6]. Quando a classificação se dá por indeterminado, ou criptogênico, é devido a não identificação de fatores causais do evento nos pacientes[6].

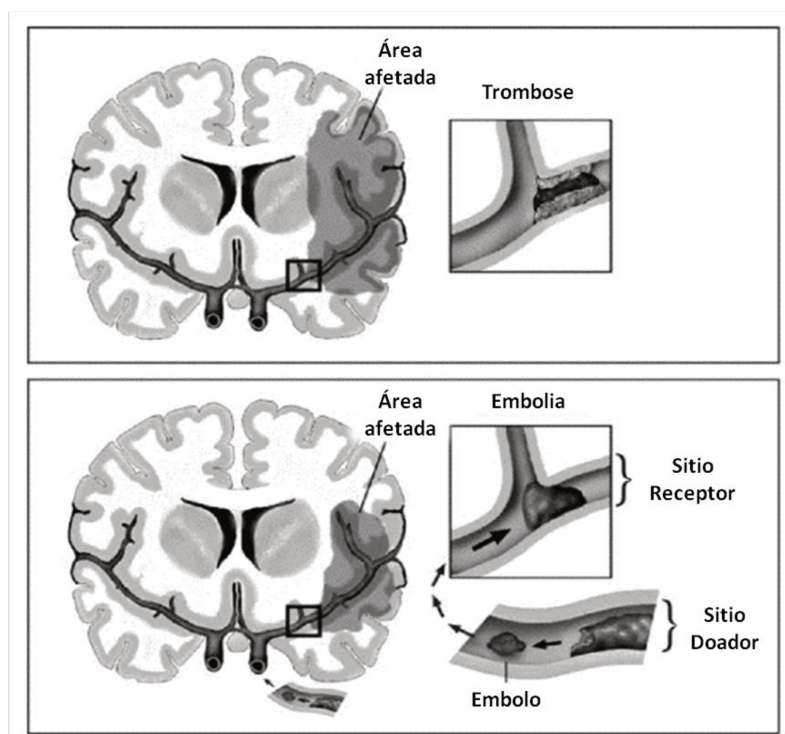


Figura 4-Esquematização dos principais subtipos de AVC isquêmico. Na imagem superior, há a formação de trombos nas artérias cerebrais, causando falta de oxigenação em uma maior área do cérebro, denominada arteriosclerose. Na imagem inferior, há a formação de coágulos em outras partes do corpo, principalmente coração,

que seguem pela corrente sanguínea até os vasos cerebrais, aonde se estabelecem, obstruindo-os e impedindo a oxigenação de determinada área. Esse segundo exemplo refere-se à cardioembolia[1].

O AVC isquêmico apresenta relação com alguns fatores de risco, que podem estar associados à alterações vasculares ou ao seu mecanismo de ação[1]. Os principais fatores de risco englobam hipertensão arterial sistêmica, tabagismo, obesidade, diabetes, arteriosclerose, cardioembolismo e doença dos pequenos vasos, sendo que estes fatores podem ser modificados pelo paciente [4-6, 8]. Entretanto, há riscos não controláveis que podem influenciar o surgimento desse evento, como idade, sexo e histórico familiar [3, 8]. Em relação ao fator etnia, poucos estudos demonstraram um risco aumentado de AVC isquêmico em indivíduos negros e hispânicos caribenhos, em comparação com indivíduos brancos, a partir da análise de uma amostra de indivíduos residentes na mesma comunidade urbana no norte de Manhattan [9].

1.1.2. Acidente Vascular Cerebral Hemorrágico

O AVC hemorrágico atinge de 15 a 20% dos casos e é caracterizado pela ruptura dos vasos sanguíneos cerebrais, com consequente extravasamento de sangue e interrupção do suprimento de oxigênio para a área afetada[5]. A classificação dos subtipos de AVC hemorrágico se dá pela região onde ocorre a ruptura dos vasos[1]. Quando o sangramento dos vasos ocorre internamente à pia-máter do cérebro, é denominada hemorragia intracerebral (Figura 5); quando a hemorragia ocorre entre a pia-máter e a aracnóide, o AVC é denominado subaracnóide (Figura 5)[1]. Rupturas acima da região aracnóide, porém dentro da dura-máter, são conhecidas por hemorragias subdurais (Figura 5)[1]. Por fim, quando o sangramento ocorre acima da dura-máter, porém internamente à caixa craniana, é conhecida por hemorragia epidural (Figura 5)[1].

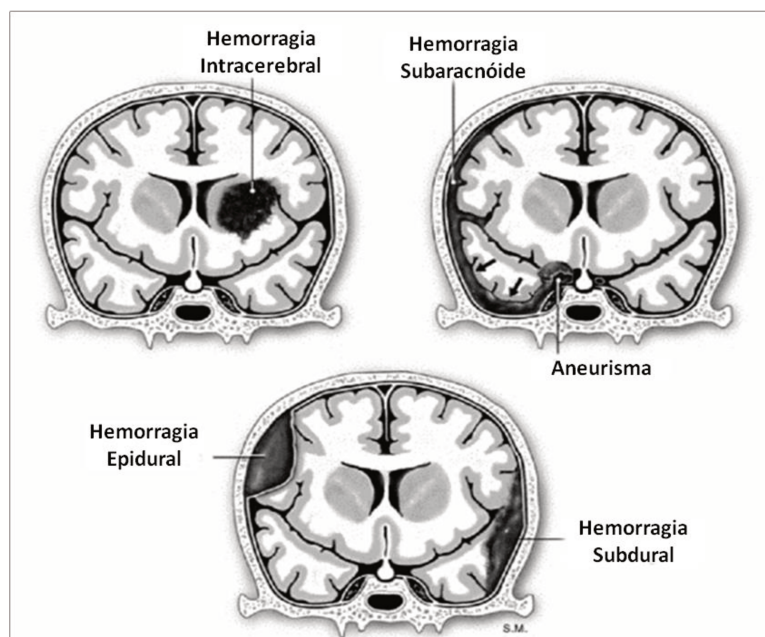


Figura 5 – Exemplos das classificações de AVC hemorrágico de acordo com as localizações das hemorragias. Na imagem superior à esquerda, identifica-se a região afetada em uma hemorragia intracerebral. Na imagem superior direita, identificam-se as regiões afetadas durante um aneurisma e uma hemorragia subaracnóide. Na imagem inferior, é possível identificar as hemorragias epidurais e subdurais[1].

Fatores de risco tendem a influenciar a ocorrência de alguns subtipos de AVC hemorrágico[1]. As hemorragias intracerebrais são comumente causadas por hipertensão, malformações vasculares, traumas, hemofilia, baixo número de plaquetas, podendo ocorrer também, com prescrição de anticoagulantes e após uso de drogas trombolíticas [1]. As hemorragias subaracnóides podem ser causadas por aneurismas arteriais, malformações vasculares próximos à superfície cerebral, tendências hemorrágicas e lesões na cabeça [1]. As hemorragias subdurais e epidurais, ocorrem, basicamente, através de traumas [1]. Esses fatores auxiliam a fragilizar a parede dos vasos sanguíneos, permitindo seu rompimento [1-4].

As hemorragias intracerebrais são as mais frequentes dentre os AVCs hemorrágicos e sua incidência tende a aumentar ao longo dos anos, devido ao aumento da expectativa de vida e usos de tratamentos antitrombóticos, especialmente em idosos [10].

Histopatologicamente, a fundamentação do AVC hemorrágico intracerebral ocorre baseada no local da lesão no cérebro, classificando-o em lobar e não-lobar[10]. Hemorragias lobares originam-se no córtex ou nas

junções corticais-subcorticais do cérebro e, em geral, estão associadas com angiopatias [10]. Hemorragias não-lobares, por sua vez, originam-se em estruturas mais profundas dos hemisférios cerebrais, cerebelo ou tronco cerebral, e tendem a ser associadas com vasculopatias[10].

1.2. Genética do Acidente Vascular Cerebral

Os fatores de risco já conhecidos para AVC, apesar de apresentarem forte influência sobre o fenótipo, explicam apenas 60% das ocorrências [11]. Deste modo, é possível que haja outros fatores não identificados que influenciem este evento [11].

Algumas condições mendelianas monogênicas podem resultar em AVC isquêmico ou hemorrágico, como arteriopatia cerebral com infartos subcorticais e leucoencefalopatia na forma autossômica dominante (CADASIL) e autossômica recessiva (CARASIL), doença de Fabry, anemia falciforme, entre outras[5]. Contudo, essas desordens monogênicas são importantes para pacientes específicos, apresentando pouco risco para a população e se tornando causas raras do AVC [5].

Estudos em gêmeos, famílias e modelos animais evidenciam a presença de fatores genéticos relacionados ao desenvolvimento do AVC [8, 11]. Estudos recentes demonstram que a genética envolvida no AVC tende a ser complexa, apresentando variantes genéticas de susceptibilidade para o surgimento de fenótipos precipitantes, tais como arteriosclerose. Muitos estudos têm investigado fatores genéticos de risco para AVC, por meio de estudos de associação de genes candidatos e avaliação de todo o genoma (*Genome-wide association studies* - GWAS) [5, 8, 11].

Ambos os estudos apresentam importância para o entendimento da genética de doenças complexas [8]. Os estudos de GWAS se referem à genotipagem de milhares de polimorfismos de base única (SNPs) ao longo do genoma, que são testados para associações com o fenótipo de interesse[12]. Esses estudos proporcionam uma poderosa ferramenta para a descoberta de genes que afetam o risco individual para o desenvolvimento de doenças neurológicas, psiquiátricas, câncer, cardiovasculares, entre outras [12, 13]. Essas metodologias, entretanto, são dependentes de grande quantidade

amostral, em torno de 1000 pacientes ou mais[8]. Essa dificuldade, associada com diferenças nas características das amostras e no desenho experimental, pode explicar grande parte da baixa reprodutibilidade de resultados entre diferentes estudos[8].

Considerando que o AVC isquêmico é o mais frequente na população, estudos genéticos relacionados a essa patologia tendem a ser mais amplos, quando comparados com os realizados para AVC hemorrágico.

Para AVC isquêmico, já foram realizados estudos de Meta-análises incluindo GWAS e estudos de genes candidatos adicionais [3, 4,14-16]. Por meio destes estudos, foram identificados um total de 23 polimorfismos de base única (SNPs) e 2 inserções ou deleções (InDels) em 19 genes (Tabela 1; Tabela 2).

Utilizando o fenótipo total de AVC isquêmico, sem diferenciação dos subtipos, foram encontrados 16 SNPs localizados em regiões codificantes e são mutações não-sinônimas, que podem estar diretamente relacionadas com o desenvolvimento do AVC[3, 4,14-16]. Entretanto, quatro SNPs estão localizados em íntrons, formando um haplótipo, quatro SNPs estão localizados em regiões intergênicas, e um SNP apresenta-se na região 3' UTR do gene *ANGPT1* (Tabela 1)[3, 4,14-16]. Os SNPs intrônicos para o fenótipo total estão localizados no gene *ALOX5AP* e os SNPs intergênicos para esse fenótipo localizam-se nos genes: *AGTRL1*, *HDAC9*, *PITX2* e *NINJ2*. Este último, por sua vez, não foi replicado nas meta-análises realizadas[3, 4,14-16].

Tabela 1 - Genes identificados nos estudos genéticos (Meta-análises e GWAS) para o fenótipo geral de AVC isquêmico[3, 4, 14-16].

Gene	Polimorfismo	Tipo Polimorfismo	Local
<i>AGTRL1</i>	rs994358	SNP	intergênico
<i>HDAC9</i>	rs2107595	SNP	intergênico
<i>PITX2</i>	rs6843082	SNP	intergênico
<i>NINJ2*</i>	rs11833579	SNP	intergênico
<i>CELSR1</i>	rs6007897	SNP	Éxon
	rs4044210	SNP	Éxon
<i>ALOX5AP</i>	rs17222814 rs10507391 rs4769874 rs9551963	SNP-haplótipo	Íntron
<i>F5</i>	p.R506Q	SNP	Éxon
<i>F2</i>	p.G20210A	SNP	Éxon
<i>ITGB3</i>	p.L33P	SNP	Éxon
<i>ACE</i>	I/D	InDel	Íntron
<i>PAI-1</i>	4G/5G	InDel	Éxon
<i>AGT</i>	p.M235T	SNP	Éxon
<i>MTHFR</i>	p.C677T	SNP	Éxon
<i>ANGPT1</i>	rs2507800	SNP	3'UTR

*Não replicados em meta-análises

Analisando os dados em termos de subtipos, observa-se que o AVC isquêmico causado por arteriosclerose apresentou SNPsem regiões não codificantes relacionados a 2 genes: *HDAC9* e *CDKN2B-AS1*. O AVC isquêmico causado por doença nos pequenos vasos (lacunar) apresentou um SNP em um éxon do gene *PRKCH*. Na isquemia causada por cardioembolismo, identificou-se SNPsem regiões não codificantes nos genes: *PITX2*, *ZFHX3*, *PDE4D* (Tabela 2).

Tabela 2 - Genes identificados através de estudos genéticos e candidatos às alterações relacionadas com os subtipos de AVC isquêmico[3, 4, 14-16].

Subtipo	Gene	Polimorfismo	Tipo Polimorfismo	Local
Arterioesclerose	<i>HDAC9</i> **	rs2107595	SNP	intergênico
	<i>CDKN2B-AS1</i>	rs2383207	SNP	Íntron
Lacunar	<i>PRKCH</i> *	p.V1425I	SNP	éxon
Cardioembolia	<i>PITX2</i> **	rs6843082	SNP	intergênico
	<i>ZFHX3</i>	rs879324	SNP	íntron
	<i>PDE4D</i>	rs1396476	SNP	íntron
		rs2910829	SNP	Íntron
		rs966221	SNP	Íntron
		rs152312	SNP	Íntron

*Não replicado nas meta-análises; **Aparece também no fenótipo total

Estes estudos genômicos demonstraram variantes que podem estar relacionadas à genética do AVC. Contudo, oito genes descritos tanto para o fenótipo total, quanto para as sub-classificações de AVC isquêmico, não apresentaram possíveis variantes funcionais (*HDAC9*; *CDKN2B-AS1*; *PDE4D*; *PITX2*; *ZFHX3*; *AGTRL1*, *ALOX5AP* e *ANGPT1*), devido à associação de polimorfismos intergênicos ou intrônicos próximos a estes genes[3, 4, 14-16].

Íntrons e regiões intergênicas são potenciais áreas regulatórias de genes [17]. Essas regiões podem ser transcritas, mas não são traduzidas (com pequenas exceções, como tradução de introns em *splicing* alternativo), sendo que não há evidências de que alterações nessas áreas apresentem alto poder funcional. Entretanto, há uma probabilidade de que polimorfismos localizados nessas regiões influenciem a expressão gênica de genes próximos[17].

Neste caso, e também considerando que o AVC sofre uma forte influência ambiental, uma das hipóteses é que estes genes aumentem o risco de AVC isquêmico devido à influência de fatores regulatórios nessas regiões, como alterações por metilações, sítios de microRNAs ou sítios de *splicing* alternativos.

Por outro lado, devido à sua menor frequência na população e à alta taxa de mortalidade desses eventos, estudos genéticos sobre AVC

hemorrágico são menos recorrentes. Até o momento, foram descritos na literatura 19 genes relacionados a essa patologia (Tabela 3) [11, 15,16, 18-20].

Os genes descritos, em geral, relacionam-se com os processos de coagulação, acúmulo e degradação de lipídeos, vascularidade, dentre outros [11, 15,16, 18-20]. Contudo, a presença de uma variante funcional, potencialmente causal e diretamente relacionada ao AVC hemorrágico ainda não foi identificada.

Considerando a abrangência da patologia e a baixa frequência de estudos relacionados, ainda são necessárias novas pesquisas para melhor compreensão dos mecanismos do AVC hemorrágico.

Tabela 3- Genes descritos na literatura como relacionados a AVC hemorrágico[11, 15, 16, 18-20].

Sigla	Gene	Variante
<i>ACT</i>	Alpha-1 antichymotrypsin	A/T polymorfism
<i>APO E</i>	Apolipoprotein E	ε2/ε3/ε4
<i>Apo(a)</i>	Apolipoprotein(a) gene	TTTTA repeat
<i>COL4A1</i>	Collagen, type IV, alpha 1	Pro352Leu
		Arg538Gly
<i>COL4A2</i>	Collagen, type IV, alpha 2	3448C>A
		5068G>A
		3368A>G
<i>CRP</i>	C-reactive protein gene	H5 haplotype (A-T-C)
<i>CYP2E9</i>	Cytochrome P450, family 2, subfamily C, polypeptide 9	Ile359Leu
		Cys144Arg
<i>eNOS</i>	Endothelial nitric oxide synthase	786T>C
<i>FXIII</i>	Coagulation Factor XIII	Val34Leu
<i>HCHWA-D</i>	Amyloidosis Dutch type	Gln693Glu
<i>HCHWA-I</i>	Amyloidosis Icelandic type	Cys68Gln
<i>LIMK1</i>	LIM domain kinase 1	rs710968
<i>IL-6</i>	Interleukin-6	572G>C
<i>OPG</i>	Osteoprotegerin	1181C>C
<i>PKD1-like</i>	Polycystic kidney disease 1-like gene	Gly243Asp
<i>PMF1</i>	Polyamine-modulated factor 1	rs2984613
<i>SLC25A44</i>	Solute carrier family 25, member 44	
<i>TNF</i>	Tumor necrosis factor	863C>A
<i>SELE</i>	Selectin E	Ser128Arg
<i>UCP3</i>	Uncoupling protein 3 gene	55C>T

O desenvolvimento de novas técnicas de análise tem ampliado a identificação de variantes não-codificantes. Contudo, as funções dessas variantes ainda são fatores limitantes para a compreensão de diversas doenças[21].

A expressão gênica é altamente influenciada por elementos regulatórios, como alterações em regiões promotoras; elementos silenciadores, e potenciadores; fatores epigenéticos, como metilações e acetilações; pequenas sequências de RNAs não codificantes (ncRNAs), alterações nas sequências cromossômicas, entre outros [21]. A validação dessas regiões regulatórias tende a permitir melhor compreensão de diversos mecanismos patogênicos de variantes associadas a doenças, sejam estas, mendelianas ou complexas, com variantes germinativas ou somáticas [21].

1.3. Epigenética

A epigenética pode ser descrito como o estudo das alterações na expressão gênica devido a modificações covalentes no DNA as quais não estão relacionadas com alterações na sequência de nucleotídeos [22]. Estas modificações incluem a metilação de citosinas no DNA e nas histonas, remodelagem dos nucleossomos, formação de heterocromatina e regulação por RNAs não codificantes[23]. O mecanismo epigenético se refere a processos moleculares inter-relacionados e complexos que modulam a expressão gênica, incluindo transcrição, processamentos pós-transcricionais do RNA e traduções[23].

O mecanismo de metilação, tanto das bases nitrogenadas das sequências de DNA, quando das histonas, é o mais frequente e estudado. As metilações do DNA são estabilizadas e mantidas enzimaticamente pela DNA metiltransferase, que transfere grupos metil da S-adenosilmethionina para o 5º resíduo de carbono nas citosinas em várias regiões genômicas, conhecidas como sítios CpG[22-24]. Citosinas presentes nessas regiões, geralmente, são altamente metiladas, podendo chegar a 60-80% de metilações em mamíferos [24]. Em geral, a metilação do DNA inibe o processo de transcrição irreversivelmente [23]. Os alvos epigenéticos representam uma área promissora, devido a seus sítios de interface entre influências ambientais e facilidades de adaptação a essas alterações.

Sítios CpGs são dinucleotídeos de citosina e guanina separados apenas por um fosfodiéster[25]. Quando houver alta frequência desses sítios

em uma região genômica, essas são chamadas de ilhas CpGs[25]. Para se caracterizar como uma ilha, a sequência deve conter uma frequência de (G+C) igual ou maior a 0,50; uma relação entre a frequência observada pela esperada igual ou maior que 0,60; e ambos ocorrendo dentro de uma janela de 200 pares de bases ou mais[26].

Essas ilhas estão, geralmente, localizadas em regiões promotoras dos genes, e realizam significativo controle da expressão gênica [25]. A metilação dessas ilhas CpGs é um dos fatores epigenéticos mais comuns, já que a presença dessas sequências nas regiões promotoras é encontrada em aproximadamente 50% dos genes expressos constitutivamente (*"housekeeping"*) no genoma [25, 26]. A metilação dessas ilhas pode ter dois efeitos sobre o mecanismo de transcrição: deslocar fatores de transcrição, que normalmente se ligam ao DNA; e atrair proteínas de metilação, que estão associadas ao silenciamento gênico e compactação da cromatina[26]. Em tecidos normais, as ilhas CpGs estão comumente não metiladas, de modo que, ao serem metiladas, inibem a transcrição gênica em questão [25].

Atualmente, há várias técnicas utilizadas para a análise epigenética no genoma, seja em pequenas regiões genômicas, específicas ou não, ou no genoma completo. Essas técnicas abrangem desde imunoprecipitação do DNA, uso de enzimas de restrição e ligação de proteínas para sítios específicos de adição de metil, além da conversão por bissulfito e sequenciamentos de nova geração [24]. A conversão por bissulfito é uma das técnicas mais conhecidas e utilizadas para esses estudos. Neste método, ocorre um tratamento do DNA com bissulfito de sódio, no qual as citosinas que não apresentam metilação são transformadas em uracila, enquanto as metiladas não sofrem alterações [24]. Ao amplificar essas sequências, a enzima DNA polimerase transformará as uracilas em timinas, gerando uma sequência dupla fita de DNA (Figura 6)[24]. Assim, comparando as sequências tratadas e originais, é possível inferir o padrão inicial de metilação do DNA pela quantificação das citosinas [24].

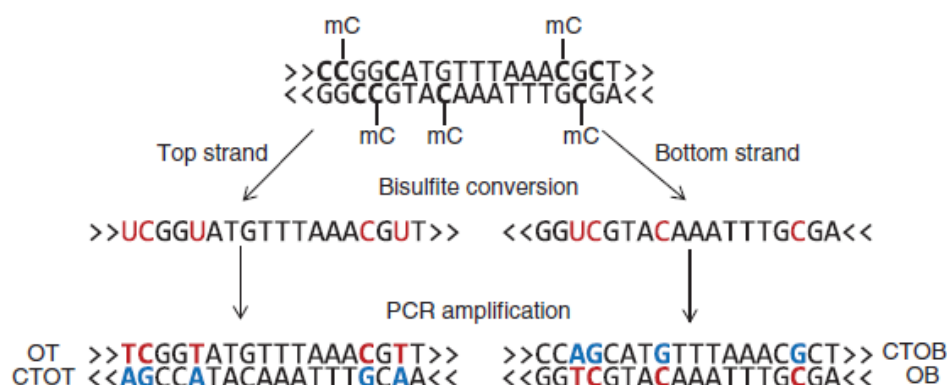


Figura 6 – Conversão do DNA pelo tratamento com bissulfito. As citosinas não metiladas são convertidas em uracila, enquanto as metiladas são mantidas na sequência. Durante a amplificação, as uracilas são transformadas em timina, gerando uma dupla fita de DNA. mC: metil-citosina; OT: fita com sequência original forward (*Original Top Strand*); CTOT: fita complementar à original forward; OB: fita com sequência original reverse (*Original Bottom Strand*); CTOB: fita complementar à original reverse[24].

No sistema nervoso central, os mecanismos epigenéticos agem como reguladores de desenvolvimento, homeostase e plasticidade[23]. No caso do AVC, esses processos estão envolvidos nos mecanismos celulares e moleculares que fundamentam seu desenvolvimento e recuperação, incluindo as respostas de estresse que modulam viabilidade celular e promovem reparação e funcional reorganização dos tecidos[23].

No AVC isquêmico, a função da metilação de DNA apresenta vários pequenos aspectos, com efeitos sobre todo o genoma e sobre genes específicos que influenciam a vulnerabilidade do sistema nervoso central[23]. Estudos em animais revelam que em algumas doenças, como oclusão da artéria média cerebral (MCAO), a metilação do DNA em cérebros isquêmicos é aumentada, podendo ser responsável por promover a morte celular[23]. De fato, testes demonstram importante atividade da enzima DNA metiltransferase[23].

Verificou-se, também nesse estudo, que a deficiência em outro importante fator para a metilação, a metilenetetraidrofolato redutase (MTHFR), causa hiperhomocisteinemia e resulta em um aumento no risco de AVC e doenças cardiovasculares[23]. Esse fator está relacionado com o metabolismo do folato e formação de reservas celulares de grupos doadores de metil [23]. Um polimorfismo, *MTHFR 677C>T*, uma variante gênica presente no gene do

fator MTHFR, também manifestou associação com risco aumentado de AVC, doenças cardiovasculares e desordens de desenvolvimento neural[23].

Portanto, pode-se observar que metilações em genes relacionados com o sistema nervoso central apresentam importantes efeitos no mecanismo de AVC isquêmico.

1.4. Estudos de Variações nos Números de Cópias no DNA

Variações no número de cópias (*copy number variation* - CNVs) são deleções ou duplicações maiores que 1kb de DNA[27-30]. Esse valor tende a ser empírico, já que alguns estudos determinam a existência dessas variantes com tamanhos a partir de 50kb[31]. Variantes menores que esse tamanho tendem a não ser precisas, já que ainda não há métodos moleculares curados para estudar esses pequenos rearranjos na escala genômica [32].

Essas variações podem ser estruturalmente mais simples, como duplicações de uma mesma região, ou apresentarem estruturas mais complexas, como ganhos de sequências homólogas ao longo do genoma[29]. Alterações genômicas por recombinação são frequentes ao longo do genoma, podendo abranger 13% do mesmo [32]. Estima-se que 5% do genoma seja formado por regiões de duplicações [33]. As CNVs podem gerar alterações significativas na expressão gênica devido à variação no número de cópias de um gene ou ruptura da sequência gênica [28, 29].

As CNVs têm sido reconhecidas como importante fonte de variações genéticas entre mutações normais e patogênicas e podem ou não conter total ou parcialmente genes em sua extensão[30]. Essas alterações podem ser herdadas ou esporádicas, de modo que CNVs *de novo* são mais propícias a serem causais para uma doença [32]. Contudo, os efeitos fenotípicos dessas alterações são muitas vezes complexos e altamente dependentes das sequências afetadas [32].

Em geral, CNVs influenciam os riscos a doenças através de alguns mecanismos, como impacto direto na expressão gênica, interrupção na transcrição, dosagem ou regulação gênica[34]. A identificação de CNVs que causam o aumento ou diminuição do risco de doenças em pacientes em

estudos de coorte vem sendo facilitado através das técnicas para análise dessas variações [30].

Técnicas mais recentes, como chips de microarranjos e sequenciamentos de alto desempenho, permitem análises mais abrangentes dos CNVs do que técnicas citogenéticas tradicionais, que permitem a detecção de anomalias de, no mínimo, 5Mb de tamanho [35].

Os microarranjos baseados em SNPs providenciam informações genóticas adicionadas às relações de intensidade, permitindo que essa plataforma tenha o potencial para detectar CNVs em uma resolução significativamente maior que outras de primeira geração, baseadas em oligonucleotídeos e clonagens [35]. Desse modo, os dados gerados por essa técnica tendem a ser maiores e mais complexos, ampliando a importância da análise e interpretação desses dados. Uma correta interpretação funcional dos dados dos CNVs deve ser baseada em conjuntos de confiança das regiões candidatas, detectando também, CNVs raros e menores. Em geral, essa interpretação combina os dados de intensidade e genótipo [35].

CNVs têm sido catalogadas em bancos de dados públicos, como o Banco de Dados de Variantes Genéticas de Toronto (DGV) e o Banco de Dados de Desequilíbrio Cromossômico e Fenotípico em Seres Humanos utilizando Recursos do Ensembl (DECIPHER) [32].

Estudos de CNVs em pacientes com AVC isquêmico demonstram a existência de variantes raras que possam estar associadas com a ocorrência desse evento [36-38]. Genes foram identificados nesses CNVs a partir da comparação entre amostras casos e controles, e relacionados com as vias de organização extracelular, organização de fibrilas de colágeno, e receptores de fatores de crescimento [36]. Dentre esses genes, pode-se citar: *COL3A1*, *COL5A2*, *SNTA1* e *DUPSS22* [36]. Entretanto, em alguns estudos realizados com base em arranjos de SNPs, é possível que CNVs raros ou menores não tenham sido identificadas devido à baixa abrangência de polimorfismos descritos a determinada região genômica [37].

Um estudo realizado para análise de CNVs em pacientes com AVC hemorrágico subaracnóide na população japonesa demonstrou leve associação de duas variantes: chr4:153210505–153212191 (deleção em 4q31.3; $p=0,04$) e chr10:6265006–6267388 (duplicação em 10p15.1; $p=0,05$) [30]. Neste estudo

foram identificados alguns genes potencialmente relacionados a hemorragias subaracnóides, como: *PET112L*, *FBXW7*, *RBM17* e *PFKFB3*[30]. O gene *PET112L* é um homólogo do gene *PET112*, presente em *Saccharomyces cerevisiae*, estando expresso em músculos e tecidos cardíacos, nos quais realiza forte oxidação fosforativa[30]. O gene *FBXW7* está relacionado à regulação da apoptose neuronal, enquanto os genes *RBM17* e *PFKFB3* são relacionados com diabetes, obesidade e doenças endócrinas [30]. Novos estudos são necessários, ainda, para elucidar os mecanismos de ação dessa doença, a fim de melhorar a qualidade das profilaxias, dos tratamentos e de vida dos pacientes.

2. Objetivos

Identificar variantes genéticas, potencialmente funcionais, em pacientes com Acidente Vascular Cerebral

- i. Identificar modificações epigenéticas na região promotora de genes candidatos em pacientes com AVC isquêmico.
- ii. Identificar variantes no número de cópias em pacientes com AVC hemorrágico

CAPÍTULO I – Coorte

1. Caracterização Amostral

1.1. Biobanco de Joinville/SC

Para esse estudo foram utilizadas amostras de pacientes e controles da região sul do país, obtidas a partir de colaboração com o biobanco de AVC da Universidade Regional de Joinville – UNIVILLE, de Santa Catarina. O biobanco tem por objetivo a análise prospectiva de pacientes com AVC, independente da classificação, a partir de estudo epidemiológico na região de Joinville, SC. O sangue periférico dos pacientes foi extraído pela técnica de fenol-clorofórmio/álcool-isoamílico[39], armazenado na UNIVILLE, sendo que uma alíquota foi enviada para nosso laboratório para os estudos genéticos. Considerando a predominante ancestralidade européia dessa região, a análise genética das amostras provenientes de Joinville, SC, proporcionou a vantagem de apresentar mínima estratificação da população [40, 41].

Este biobanco é constituído por 1480 pacientes com AVC isquêmico (54,6% homens; média de idade 64 ± 14 anos); 90 pacientes com AVC hemorrágico (57,7% homens; média de idade 64 ± 12 anos); e 2459 controles (52,7% homens; média de idade de 65 ± 12 anos). Os indivíduos controles coletados devem, preferencialmente ter idade semelhante à dos pacientes, não possuir consanguinidade com os mesmos e não ter sido diagnosticado com AVC. As amostras dos pacientes são coletadas durante internação hospitalar dos mesmos. As amostras foram obtidas em três hospitais de Joinville: na Unidade de AVC presente no Hospital Municipal São José, no Centro Hospitalar Unimed e no Hospital Dona Helena.

O diagnóstico final de AVC isquêmico ocorreu de acordo com o *Trial of Org 10172 in Acute Stroke Treatment* (TOAST) [6]. A classificação ocorre após exames clínicos e avaliação médica desses pacientes. Assim, o biobanco contém, para AVC isquêmico: 273 pacientes com arteriosclerose, 332 com cardioembolismo, 289 com lacunar, 227 com ataque isquêmico transitório, 170 com outras causas de AVC e 99 com causa indeterminada. O diagnóstico de AVC hemorrágico foi realizado com base na localização da hemorragia. Desse modo, a amostra é constituída por: 14 pacientes com hemorragia lobar, 42 com não-lobar e 34 com localização ainda indeterminada. Essas informações passaram a ser obtidas pela equipe recentemente, motivo pelo qual as amostras ainda não apresentam essa classificação.

As amostras de pacientes foram caracterizadas em relação à idade, sexo, histórico familiar de ocorrência de AVC dos indivíduos, presença de hipertensão arterial sistêmica (HAS), diabetes, etilismo e tabagismo.

Foi realizada uma análise descritiva, comparando os principais fatores de risco com a população amostral. Para o fator de risco tabagismo, a classificação ocorreu de acordo com descrição pela Organização Mundial da Saúde (OMS). Para a estatística, as classificações “fumantes” e “ex-fumantes” foram agrupadas como fator de risco positivo. Para análise do histórico familiar, o paciente foi questionado sobre a ocorrência de AVC nos pais e parentes próximos até a 2ª geração. O etilismo foi avaliado por questionamento durante a consulta clínica e foi considerado positivo para indivíduos que ingerem bebidas alcoólicas eventualmente/socialmente ou são considerados elitistas. Essa análise ocorreu baseada na quantidade de doses que esse indivíduo ingere por mês. Para isso a quantidade de doses que o indivíduo ingere por semana é somada, totalizando a quantidade mensal. Os critérios para dosagem são: caso homem, ingestão de duas latas ou uma garrafa de cerveja, uma dose de destilado ou dois cálices de vinho ao dia; caso mulher, ingestão de uma lata ou meia garrafa de cerveja, meia dose de destilado ou um cálice de vinho ao dia. Embriaguez uma vez na semana também é considerada um critério para essa classificação. Para ser classificado como ingestão de bebidas alcoólicas eventualmente/socialmente, a somatória dessas doses deve ser inferior a 21 doses/mês, para homens, e 14 doses/mês, para mulheres. Acima desses valores, o indivíduo é considerado elitista.

As informações referentes à frequência de atividade física não foram obtidas de todos os indivíduos. Nesta classificação, foram selecionados como positivos os indivíduos que informaram realizar atividades físicas leves e/ou moderadas. São considerados inativos, os indivíduos que não realizam atividade física voluntária, paga ou recreativa. Os ativos leves são considerados indivíduos aposentados ou com empregos sedentários que realizam atividade física por meia hora por dia ou que possuam emprego sedentário, porém, sem trabalho sentado e sem atividade física. Os indivíduos ativos moderados são os que possuem trabalhos sedentários com meia a uma hora de atividade física por dia, trabalhos sedentários com meia hora de

atividade física por dia ou trabalho físico, mas sem atividade física recreativa. A frequência de atividade física foi questionada apenas a uma parte da amostra.

A fim de avaliar se os dados clínicos não apresentam diferenças significativas entre casos e controles, um teste exato de Fisher foi realizado para as variáveis categóricas. As variáveis numéricas foram avaliadas, inicialmente, pelo teste de normalidade de Shapiro-Wilk. O teste t de Student foi realizado caso as variáveis sigam uma distribuição normal e, caso contrário, o teste não paramétrico Wilcoxon-Mann-Witney foi utilizado para avaliação das diferenças entre casos e controles. Visto que as análises são feitas comparando os grupos AVC isquêmico vs. controle e AVC hemorrágico vs. controle, foi adotado um nível de significância $\alpha = 0,025$ ($0,05/2$), ajustado para múltiplos testes por meio da correção de Bonferroni.

Esse projeto foi aprovado pelos Comitês de Ética da FCM-UNICAMP (nº 257.020) e da Univille (nº 090/2013), e todos os indivíduos assinaram um termo de consentimento livre esclarecido antes de participar do estudo.

1.2. Colaboração com Universidade de Pisa/IT

A partir de colaboração com o Dr. Michelangelo Mancuso, da Universidade de Pisa/IT, foram encaminhadas ao nosso laboratório amostras de 18 pacientes com AVC hemorrágico e de 30 indivíduos controles. As amostras foram extraídas por utilizando a técnica de *salting out*[42].

As amostras de pacientes foram caracterizadas para idade, sexo, ano de ocorrência e número de eventos, histórico familiar de ocorrência de AVC, presença de diabetes, hipertensão, tabagismo, etilismo e hipercolesterolemia. Para as amostras de controles, foram encaminhados apenas dados de sexo, idade e etnia. Tanto pacientes quanto controles foram descritos como caucasianos. Em relação ao histórico familiar dos pacientes, foi questionado sobre a ocorrência de AVC até a primeira geração, apenas.

2. Extrações de DNA

2.1. Método do Fenol-Clorofórmio/Álcool-Isoamílico

O método do fenol/clorofórmio/álcool-isoamílico é o mais antigo utilizado para extração de DNA a partir de sangue completo, ou seja, sem prévio processamento. Nessa técnica, com o uso de enzimas e solventes orgânicos, ocorre a lise celular, degradação de proteínas e moléculas de RNA e, por fim, a precipitação do DNA[43]. É importante ressaltar que as amostras utilizadas foram frescas, a fim de evitar hemólise das células e baixa qualidade final das amostras.

Inicialmente, centrifugar as amostras de sangue a 2500rpm por 10 minutos a fim de separar o plasma dos leucócitos e hemácias. Para a extração de DNA, o plasma é descartado, de modo que o processo continua com as outras células precipitadas. Transferir, então, o aglomerado celular para um tubo falcon de 15mL e adicionar 6mL de tampão RSB 1x e 6 gotas de NONIDET®P40 Substitute (Amresco LLC, OH, Estados Unidos). Esses reagentes auxiliarão na lise celular. Agitar por 10 minutos e centrifugar novamente a 2500rpm por 10 minutos. Descartar o sobrenadante. Adicionar 0,5mL de tampão RSB 1x, 3mL de detergente SDS 10% e 80µL de proteinase K 20mg/mL. O tampão irá estabilizar as moléculas e o pH da reação, enquanto o detergente irá realizar a lise de eritrócitos e da membrana celular, ao mesmo tempo que a enzima cliva as proteínas. Deixar a reação overnight à 65°C, temperatura ótima para atividade da proteinase K.

A segunda etapa é caracterizada pela purificação da molécula de DNA. Para isso, adiciona-se compostos orgânicos que auxiliam a degradação de proteínas e organelas[43].

Portanto, após o período de incubação da amostra, adicionar 3mL de fenol 0,1M, pré-tratado com Tris-HCl para equilibrar a solução em pH=8. O fenol separa a fase orgânica da fase aquosa, mantendo na primeira fase os contaminantes da amostra. Agitar por 10 minutos e centrifugar a 2500rpm por 10 minutos. Nesta etapa, as centrifugações irão separar a fase orgânica (que contém as soluções orgânicas de purificação) da fase aquosa (contendo as moléculas solúveis em água, como ácidos nucleicos). Transferir a fase superior para outro tubo, adicionar 1,5mL de fenol 0,1M e 1,5mL de clorofórmio-álcool-

isoamílico, solução preparado proporção 24:1. O clorofórmio precipita proteínas, enquanto o álcool-isoamílico é um aditivo que evita a formação de espumas e inativa a enzima RNase. Agitar novamente por 10 minutos e centrifugar a 2500rpm por 10 minutos. Retirar a fase inferior da solução (fase orgânica). Adicionar 3mL de clorofórmio-álcool-isoamílico (proporção de 24:1), agitar por 10 minutos e centrifugar a 2500rpm por 10 minutos. Descartar completamente a fase inferior da solução. Adicionar 9mL de etanol 100% e gelado e agitar levemente até a precipitação do DNA. Ressuspender essa molécula em tampão TE (Tris-HCl10mM e EDTA 1mM) e armazenar a -20°C até uso.

Esse método de extração tem bom rendimento e com boa qualidade final[43]. Entretanto, uma desvantagem é o uso de compostos tóxicos e extenso trabalho experimental[43]. Outro ponto, é que se os compostos, principalmente fenol, não forem completamente eliminados da amostra, podem influenciar na baixa qualidade do DNA[43].

2.2. Método Salting Out

Esse método, por sua vez, envolve a retirada de proteínas celulares pela desidratação e precipitação das mesmas com solução saturada de cloreto de sódio (NaCl) [42, 43]. Além desse composto, essa técnica utiliza outros compostos atóxicos, como proteinase K eRNase[42, 43]. Apresenta, também, bom rendimento e pureza das amostras, sendo relativamente rápido e barato [43]. Uma restrição dessa técnica é o uso de diferentes enzimas, já que cada uma apresenta um padrão de afinidade a salinidade diferente. O protocolo abaixo foi obtido de acordo com Miller et al, 1987 [42].

Para essa extração deve-se: ressuspender o sangue não coagulado com 3mL de tampão de lise (RSB). Manter as células em lise overnight para digestão a 37°C em solução com 0,2mL de SDS 10% e 0,5mL de proteinase K. Após a digestão, adicionar 1mL de NaCl saturado 6M e agitar vigorosamente por 15 segundos. Centrifugar a 2500rpm por 15 minutos. As proteínas tratadas com o cloreto precipitam, mantendo o DNA em solução. Transferir o sobrenadante para outro tubo falcon de 15mL. Adicionar 2 volumes de etanol

absoluto e inverter várias vezes até a precipitação do DNA. Transferir o DNA para tubo criogênico contendo 100-200µL de solução tampão TE.

2.3. Qualidade das amostras de DNA

Após a extração, as amostras de DNA foram submetidas à análise de quantidade e grau de pureza ($\lambda_{260/280}$), as quais foram realizadas no Laboratório de Genética Molecular, Departamento de Genética Médica - FCM, utilizando o equipamento *Epoch* (BioTek Instruments, Inc. Winooski, VT, Estados Unidos). Os DNAs usados foram escolhidos com base em sua pureza (razão entre os valores das absorbâncias 260/280 nm), sendo que esta deveria estar entre 1,8 e 2,0, quantificação e integridade. A integridade dos DNAs foi testada por eletroforese em gel de agarose 1,5%. Foram separadas alíquotas desses DNAs e diluídas de acordo com a técnica a ser utilizada. Para técnicas de microarranjos de DNA, a quantificação ideal das diluições foi 50ng/µL, podendo variar de 45ng/µL a 60ng/µL, como descrito pelo protocolo do fabricante.

3. Resultados

3.1. Biobanco de Joinville/SC

Até o momento, foram enviados à UNICAMP 246 amostras de DNA de indivíduos controles, 175 de pacientes com AVC isquêmico aterotrombótico, e 60 de pacientes com AVC hemorrágico. Em conjunto, foram encaminhadas as avaliações clínicas. Os controles coletados incluíram os cônjuges dos pacientes.

A média de idade dos indivíduos controles e dos pacientes de AVC isquêmico e hemorrágico foi de 65 anos (desvio padrão = 12 anos), 66 anos (desvio padrão = 11 anos) e 62 anos (desvio padrão = 14 anos), respectivamente (Figura 7). As amostras de pacientes com AVC isquêmico e hemorrágico apresentaram distribuição normal em relação a idade (Shapiro-Wilk p-value=0,543 e 0,592, respectivamente), enquanto a amostra controle não (Shapiro-Wilk p-value=0,0068). Através de testes estatísticos, pode-se concluir que não houve diferenças estatísticas entre os grupos (AVC isquêmico

VS. Controles p-value=0,449; AVC hemorrágico VS. Controles p-value=0,175). É possível verificar que as amostras apresentam a maioria dos indivíduos do sexo masculino (61% para AVC isquêmico e 55% para grupo controle), com exceção do grupo de pacientes com AVC hemorrágico que apresenta similaridade entre ambos os gêneros (48% de indivíduos do sexo masculino) (Figura 8). Os testes estatísticos demonstram que não há diferenças significativas entre os grupos (AVC isquêmico VS. Controles p-value=0,21; AVC hemorrágico VS. Controles p-value=0,98).

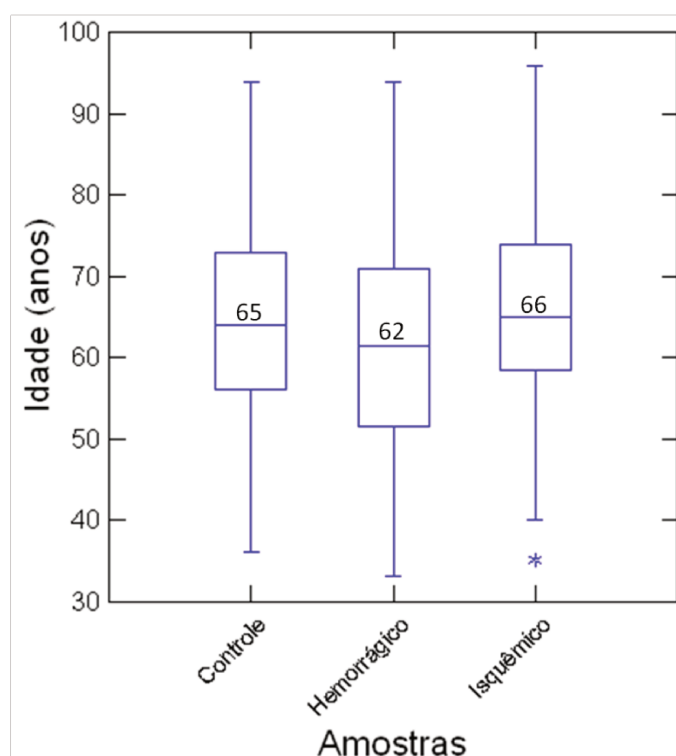


Figura 7 - Gráfico da distribuição das idades das amostras de indivíduos controles, pacientes com AVC isquêmico e AVC hemorrágico provenientes do Biobanco de AVC em Joinville/SC (MyStat Software).

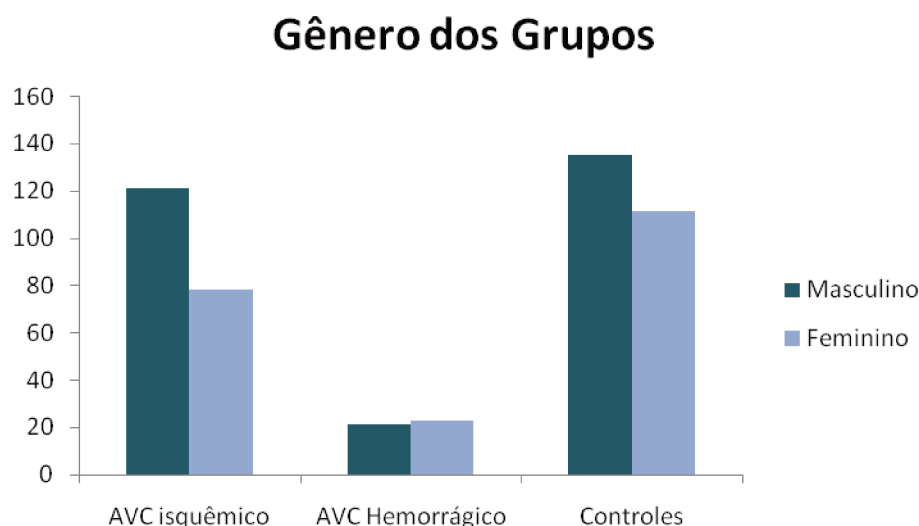


Figura 8 - Distribuição do gênero entre as populações amostrais de pacientes com AVC isquêmico, AVC hemorrágico e controles provenientes do Biobanco de AVC de Joinville, SC.

A caracterização das amostras para fatores de risco ocorreu baseada nas seguintes informações: histórico familiar, hipertensão arterial, diabetes, tabagismo, etilismo e frequência de atividade física.

De acordo com a Tabela 4, dos fatores de risco, o mais frequente na amostra controles é o tabagismo (46%), seguido pela hipertensão arterial (43%) e histórico familiar (21%). Destes fatores de risco, podemos perceber que os mais frequentes na amostra de AVC isquêmico são: hipertensão arterial (76%), tabagismo (62%) e histórico familiar (50%). Os mesmos critérios da classificação dos controles foram utilizados para as amostras de pacientes.

Através de cálculos estatísticos, baseando-se em uma distribuição com 95% de confiabilidade, pode-se verificar que as amostras são estatisticamente diferentes quando comparadas para cada fator de risco entre controles e pacientes com AVC isquêmico. Pode-se concluir que hipertensão arterial, tabagismo e histórico familiar são os três principais fatores de risco que acometem tanto pacientes com AVC isquêmico, quanto indivíduos controles. Não foram realizadas comparações estatísticas para a classe “Etilismo”, porque essa informação não foi obtida para o grupo controle.

Tabela 4 –Comparação entre os fatores de risco que acometem as amostras controlee pacientes com AVC isquêmico.

Fatores de Risco	Controles			AVC isquêmico			P-valor
	Sim	Não	NA	Sim	Não	NA	
Histórico	52	194	0	100	92	7	1,72 ⁻¹²
Familiar	(21%)	(79%)	(0%)	(50%)	(46%)	(4%)	
HAS	107	135	4	152	44	3	1,52 x10 ⁻¹¹
	(43%)	(55%)	(2%)	(76%)	(22%)	(2%)	
Diabetes	49	187	10	72	124	3	<1 x10 ⁻¹⁵
	(20%)	(76%)	(4%)	(36%)	(62%)	(2%)	
Tabagismo	114	131	1	123	76	0	3,86 x10 ⁻³
	(46%)	(53%)	(0,4%)	(62%)	(38%)	(0%)	
Etilismo	-	-	-	67	132	0	-
				(34%)	(66%)	(0%)	
Atividade Física	14	46	186	52	143	1	<1 x10 ⁻¹⁵
	(6%)	(19%)	(76%)	(26%)	(73%)	(1%)	
TOTAL	336	693	201	566	813	14	

NA = Desconhece/Não respondeu; HAS = hipertensão arterial sistêmica

Para as amostras de AVC hemorrágico, os fatores de risco mais frequentes são: hipertensão arterial (78%), tabagismo (48%) e etilismo (37%). (Tabela 5). No caso de AVC hemorrágico, o etilismo está entre os 3 principais fatores de risco que acometem esses pacientes. Dados para etilismo não foram coletados para o grupo controle pelo Biobanco de Joinville, SC. Dentre os outros fatores, para a categoria tabagismo, ambos os grupos AVC hemorrágico e controle apresentam semelhança estatística. Não foram realizadas comparações estatísticas para a classe “Etilismo”, porque essa informação não foi obtida para o grupo controle.

Tabela 5 - Comparação entre os fatores de risco que acometem as amostras controle e pacientes com AVC hemorrágico.

Fator de Risco	Controles			AVC hemorrágico			P-valor
	Sim (%)	Não (%)	NA (%)	Sim (%)	Não (%)	NA (%)	
Histórico Familiar	52 (21%)	194 (79%)	0 (0%)	18 (30%)	37 (62%)	5 (8%)	6×10^{-7}
HAS	107 (43%)	135 (55%)	4 (2%)	47 (78%)	12 (20%)	1 (2%)	$6,63 \times 10^{-6}$
Diabetes	49 (20%)	187 (76%)	10 (4%)	4 (7%)	52 (87%)	4 (7%)	$1,55 \times 10^{-6}$
Tabagismo	114 (46%)	131 (53%)	1 (0,4%)	29 (48%)	31 (52%)	0	$8,57 \times 10^{-1}$
Etilismo	-	-	-	22 (37%)	38 (63%)	0	-
Atividade Física	14 (6%)	46 (19%)	186 (76%)	14 (23%)	46 (77%)	0	$<1 \times 10^{-15}$
TOTAL	336	693	201	134	216	10	

NA = Desconhece/Não respondeu; HAS = hipertensão arterial sistêmica

3.2. Amostras provenientes da Universidade de Pisa/IT

As amostras provenientes da Itália foram encaminhadas ao laboratório, com DNA periférico já extraído, a fim de colaboração entre as universidades. As médias de idade entre os grupos são: 63 anos (desvio padrão = 11 anos) para o grupo de pacientes e 70 anos (desvio padrão = 7 anos) para o grupo de controles (Figura 9). De acordo com análises estatísticas, ambas as coortes apresentam distribuição normal dos dados “idade” (Shapiro-Wilk_{pacientes} p-value=0,96 e Shapiro-Wilk_{controles} p-value=0,75).

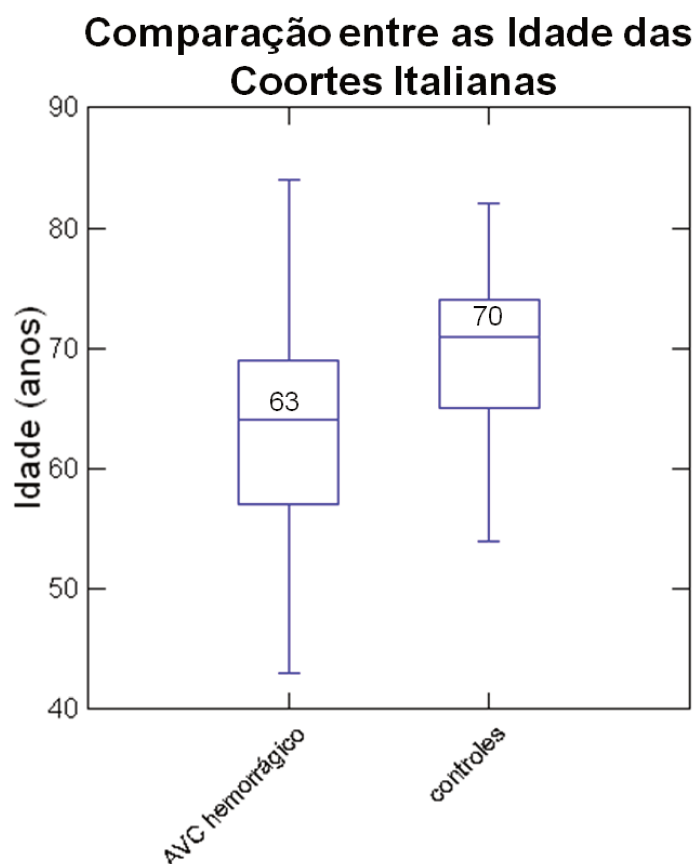


Figura 9 - Comparação entre as idades das populações amostrais (pacientes com AVC hemorrágico e indivíduos controles), provenientes de colaboração com a Universidade de Pisa/IT.

Em relação ao gênero dos indivíduos, é possível verificar que há mais indivíduos do sexo masculino para o grupo de pacientes e mais indivíduos do sexo feminino para o grupo de controles (Figura 10). Contudo, de acordo com testes estatísticos, não há diferença significativa entre os grupos (Pacientes_{Itália} VS. Controles_{Itália} p -value=0,23).

O grupo controle é formado de indivíduos que não possuem consanguinidade com os pacientes e, em geral, são coletados nas consultas hospitalares. Desse modo, a frequência de indivíduos femininos nesse grupo é uma sugestão de que foram coletadas as esposas dos pacientes. O fato de o grupo controle ter maior frequência de mulheres, também pode ser uma explicação para a maior média de idade dessas amostras, já que a expectativa de vida de indivíduos do sexo feminino é consideravelmente maior na

população. Essa alta expectativa de vida, também tende a influenciar o risco de ocorrência de AVC nesses indivíduos [44].

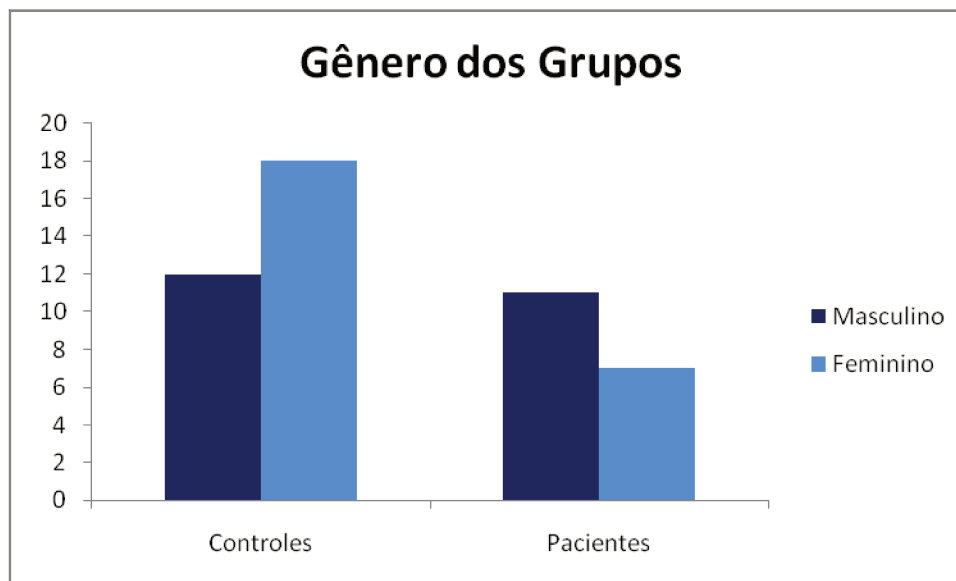


Figura 10 - Descrição do gênero das amostras de indivíduos controles e de pacientes com AVC hemorrágico provenientes de Pisa/IT.

As amostras também foram caracterizadas em relação a fatores de risco. Para as amostras controles, entretanto, somente foram obtidas as informações referentes a sexo, idade e etnia. Para as amostras de pacientes, foram obtidas informações sobre hipertensão, diabetes, hipercolesterolemia, tabagismo, etilismo e histórico familiar. Para essa amostra, o fator de risco mais frequente é a hipertensão arterial (83%), seguido de hipercolesterolemia (50%) e diabetes (33%) (Tabela 6). Contudo, a informação “desconhecida” ou “sem resposta” é muito frequente para os fatores tabagismo e etilismo, dois dos principais fatores de risco para o AVC hemorrágico (Tabela 6).

Tabela 6 –Descrição dos fatores de risco que acometem as amostras de pacientes com AVC hemorrágico provenientes de Pisa/IT.

Fatores de Risco	AVC hemorrágico		
	Sim	Não	NA
Histórico Familiar	-	18 (100%)	-
HAS	15 (83%)	3 (17%)	-
Diabetes	6 (33%)	12 (67%)	-
Tabagismo	-	2 (11%)	16 (89%)
Etilismo	-	5 (28%)	13 (72%)
Hipercolesterolemia	9 (50%)	9 (50%)	-
TOTAL	30	49	29

NA = Desconhece/Não respondeu; HAS = hipertensão arterial sistêmica

CAPÍTULO II –Epigenética

1. Material e Métodos

1.1. Amostras

As amostras utilizadas para essa análise foram as de AVC isquêmico provenientes do biobanco de Joinville/SC.

1.2. Identificação dos sítios CpGs nos Genes Candidatos

Baseado nos estudos de meta análises e GWAS, os oito genes que não apresentaram variantes funcionais (*HDAC9*; *CDKN2B-AS1*; *PDE4D*; *PITX2*; *ZFHX3*; *AGTRL1*, *ALOX5AP* e *ANGPT1*), foram selecionados para identificação de ilhas CpGs. Visto que a região promotora apresenta, em geral, vários sítios de regulação da expressão gênica, ela é o principal alvo de estudos de metilação e identificação de ilhas CpGs, comumente regiões potencialmente metiladas. Desse modo, foram realizadas buscas por essas regiões de metilação nas regiões promotoras desses genes.

Para cada gene, as regiões promotoras foram localizadas pelo banco de dados Ensembl(www.ensembl.org). Foram selecionados o primeiro éxon do gene e os 1000pb *upstream*. A região de 1000pb *upstream* do início da transcrição do gene foi selecionada a fim de avaliar toda a região promotora (normalmente localizada a 300pb antes do começo da transcrição) em busca de ilhas CpG e o primeiro éxon foi analisado visando encontrar alguma ilha que se estendesse até o início do gene, podendo alterar sua expressão. Essa sequência foi analisada no software EMBOSS CpGPlot(www.ebi.ac.uk/Tools/seqstats/emboss_cpgplot), a fim de identificar regiões que pudessem ser potenciais sítios de metilação. Essa identificação é baseada na frequência decitosinase guaninas observadas (porcentagem mínima de 50%); na razão de CpGs observadas por esperadas por região gênica (mínimo de 0,60); e no tamanho mínimo da ilha CpG (parâmetro padrão igual a 200pb)[45].

As potenciais regiões CpG encontradas foram analisadas no programa *MethPrimer*(<http://www.urogene.org/cgi-bin/methprimer/methprimer.cgi>), com a finalidade de desenho de *primers* específicos para metilações, baseando-se nos DNAs tratados com bissulfito[46]. O programa pode desenhar *primers* para dois tipos de PCR: “PCR específico de metilação” (*Methylation-Specific PCR - MS-PCR*) e “PCR para sequenciamento de bissulfito” (*Bissulfite-Sequencing PCR – BS-PCR*)[46]. Além disso, o *MethPrimer* também prediz ilhas CpGs nas sequências de DNA[46]. Para adicionar sequências nesse programa, não há necessidade de qualquer edição no formato[46]. Foram utilizadas sequências no formato FASTA. Ao inserir a sequência, é possível alterar os parâmetros de interesse para os *primers*, como tamanho do produto, temperatura de melting (T_m), tamanho e quantidade de CpGs englobadas pelos *primers*[46]. Dentre as opções testadas, foram utilizadas as configurações padrão do programa: Tamanho do produto de 100-300pb, T_m do *primer* entre 50-60°C, tamanho do *primer* entre 20-30pb, quantidade de CpGs igual ou maior a 4 [46]. Os *primers* foram desenhados para regiões convertidas e não convertidas por bissulfito.

1.3. Tratamento por Bissulfito das Amostras

Nesse método, o DNA é convertido pela técnica de bissulfito, na qual as citosinas não metiladas serão modificadas para uracila enquanto as citosinas metiladas não serão alteradas. Para isso, utiliza-se o kit para conversão de bissulfito *EZ DNA Methylation-Lighting Kit* (ZymoResearch, Irvine, CA, Estados Unidos).

Nesse método, a quantidade inicial do DNA deve ser entre 100pg a 2µg, sendo a concentração ideal entre 200ng a 500ng. Utilizou-se a concentração de 20ng/µL de DNA, a fim de obter uma quantidade inicial de 400ng de DNA. Adicionar 20µL de DNA em um tubo de microcentrífuga de 200µL. Adicionar 130µL do reagente “*Lightning Conversion Reagent*”. Centrifugar rapidamente para garantir que não há gotas nas paredes do tubo. Colocar no termociclador seguindo o ciclo: 98°C por 8 minutos; 54°C por 60 minutos; 4°C por até 20 horas. Preparar a coluna de purificação “*Zymo-SpinTM IC Column*” colocando-a

no tubos de coleta “*Collection Tube*”. Transferir a amostra para a coluna de purificação. Adicionar 600µL de tampão “*M-Binding Buffer*” na coluna contendo a amostra. Misturar por inversões. Centrifugar a 13000g por 30 segundos. Descartar o sobrenadante. Adicionar 200µL de tampão “*L-Desulphonation Buffer*” à coluna e armazenar de 15 a 20 minutos à temperatura ambiente. Centrifugar a 13000g por 30 segundos. Adicionar 200µL de tampão de lavagem “*M-Wash Buffer*” à coluna e centrifugar a 13000g por 30 segundos. Repetir essa etapa de lavagem. Colocar a coluna e tubo de microcentrífuga 1,5mL. Adicionar 10µL de tampão de eluição “*M-Elution Buffer*” diretamente na matriz da coluna. Centrifugar a 13000g por 30 segundos para eluição do DNA. O DNA pode ser analisado imediatamente ou armazenado a -20°C para uso posterior.

As amostras de DNA foram submetidas à análise de quantidade e grau de pureza ($\lambda_{260/280}$), as quais foram realizadas utilizando o equipamento *Epoch* (BioTek Instruments, Inc. Winooski, VT, Estados Unidos).

1.4. PCR Metil-Específico (MS-PCR)

A técnica PCR Metil-Específico (*methyl-specific PCR*; MS-PCR) é sensível e específica para metilações nos sítios CpG[47]. Após a conversão das amostras, é necessário a amplificação das sequências através da técnica de PCR a partir dos *primers* desenhados especificamente para as ilhas CpGs nas regiões promotoras dos genes. Esses *primers* são obtidos para as sequências metiladas e não metiladas.

A reação de PCR consiste na amplificação das sequências gênicas a partir de uma sequência homóloga à região de interesse. Nessa técnica, apenas a região alvo será amplificada para maiores análises. Portanto, para essa reação, deve-se adicionar a um tubo de microcentrífuga de 200µL: 6,2µL de água de injeção, 1µL de tampão 10x, 1µL de solução de dinucleotídeos (DNTP) a 1,25mM, 0,35µL de solução de cloreto de magnésio a 50mM, 0,25µL de *primer* forward diluído a 10mM, 0,25µL de *primer* reverse diluído a 10mM, 0,3µL de enzima Taq DNA polimerase e 0,5µL de DNA. A reação deve ser adicionada a um termociclador para a amplificação. A amplificação deve

ocorrer em 35 ciclos, seguindo as etapas: 30 segundos a 95°C, 30 segundos na temperatura de anelamento (Tabela 7) e 30 segundos a 72°C. A temperatura de anelamento é específica para cada par de *primer*, baseando-se na temperatura de melting de cada um (Tabela 7).

Tabela 7 - Temperaturas de anelamento específicas para cada par de *primer* obtidos por análises *in silico* para ilhas CpGs nas regiões promotoras dos genes *AGTRL1*, *CDKN2B-AS1*, *HDAC9* e *ZFHX3*.

Gene	Temperatura de anelamento (°C)
<i>AGTRL1</i>	67
<i>CDKN2B-AS1</i>	66
<i>HDAC9</i>	64
<i>ZFHX3</i>	74

Os resultados são analisados em gel de agarose 1,5%, demonstrando a presença ou ausência de metilação na região.

2. Resultados Discussão

2.1. Identificação de SítiosCpGs em Genes Candidatos

Entre os oito genes selecionados, somente os genes *AGTRL1*, *HDAC9*, *CDKN2B-AS1* e *ZFHX3* apresentaram prováveis sítios CpGs, de acordo com análises pelos softwares EMBOSS CpGPLOT e MethPrimer.

a) *AGTRL1*

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot identificou uma região CpG ao final da sequência (Figura 11). Entretanto, ao analisarmos essa mesma sequência no software MethPrimer, foram identificadas duas potenciais ilhas CpGs (Figura 12). A primeira ilha apresenta 140pb, localizando-se entre 716 – 855pb, enquanto a segunda, possui 351pb de tamanho, localizada entre 968 – 1318pb.

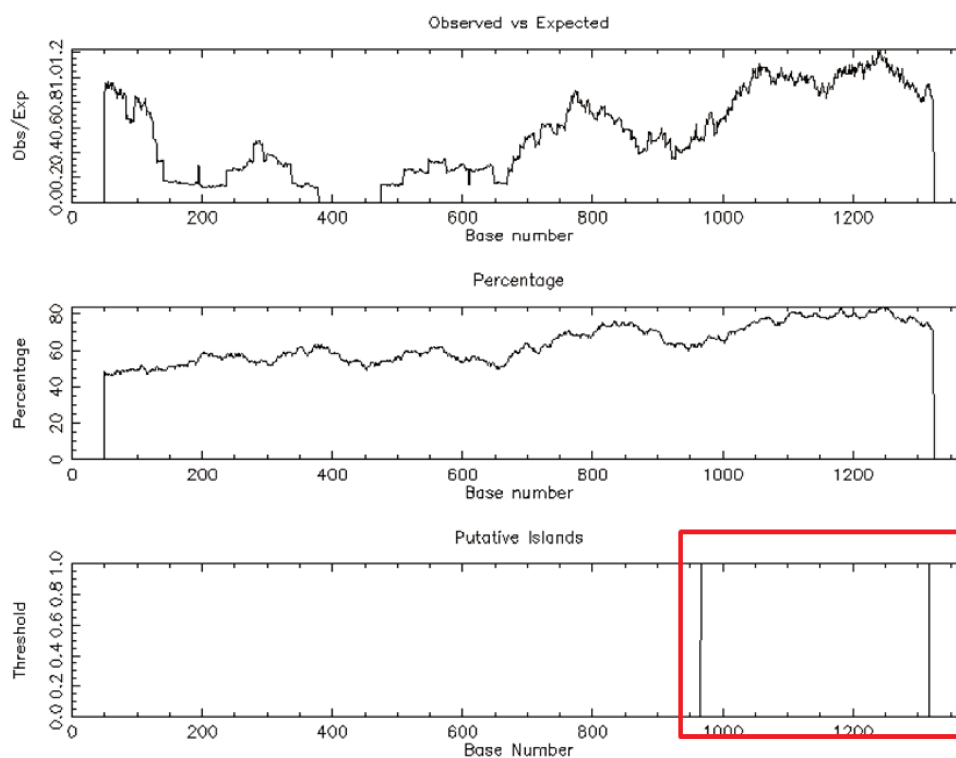


Figura 11 - Ilha CpG identificada (em destaque) para o gene *AGTRL1*, a partir de análise de 1000pb *upstream* e primeiro éxon (EMBOSS CpGPlot).

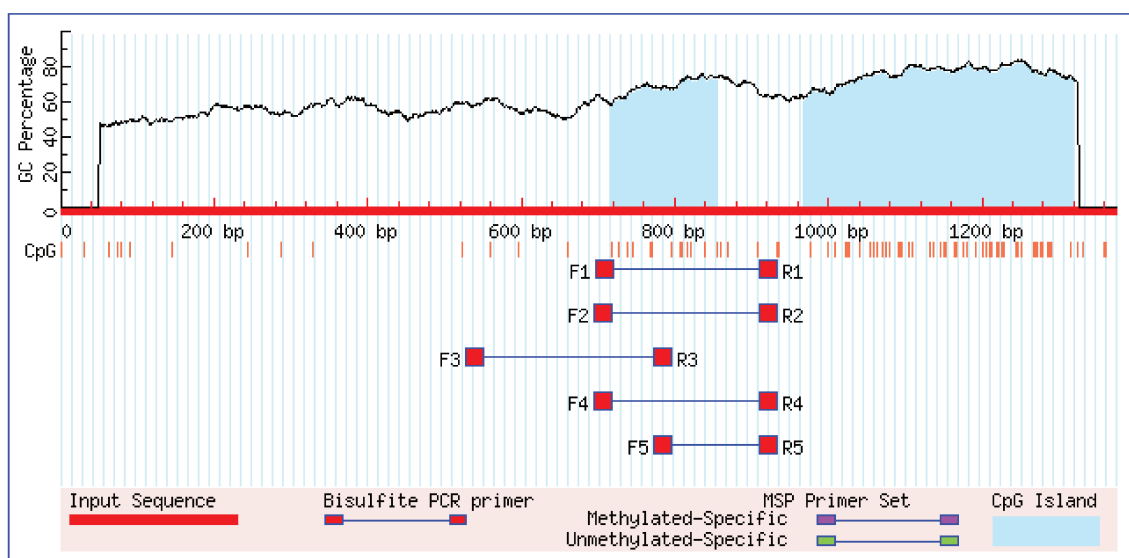


Figura 12 - Identificação de possíveis ilhas CpGs no gene *AGTRL1* e os *primers* desenhados para a região mais estável (MethPrimer).

Dentre os conjuntos de *primers* desenhados para a primeira região, o par que apresentou maior especificidade foi o F1/R1 (Tabela 8). Essa especificidade se baseia no tamanho e na localização dos *primers*, maior número de citosinas e ilhas CpGs englobadas pela região do *primer*, e mais próximas temperaturas de melting (T_m) dos *primers* *forward* e *reverse*. A discrepância entre as ilhas identificadas pelo EMBOSS e MethPrimer, já que este segundo identificou uma ilha *upstream* a identificada pelo primeiro, pode ser devido as diferenças nas configurações basais entre os dois softwares, principalmente quando relacionado ao tamanho das ilhas.

Tabela 8 - Descrição dos *primers* selecionados para a primeira região CpG do gene *AGTRL1*, dentre os quais os *primers* F1/R1 (em destaque) foram selecionados.

	Primer	Start	Size	T_m	GC%	"C"s	Sequence
1	Left Primer	697	23	53,89	65,22	11	GTTTTGTAAATTTTTGGTTTT
	Right Primer	934	24	53,39	54,17	6	AACCCACTTAATCATATAATCATC
	Product size: 238, T_m : 70.0, CpGs in product: 16						
2	Left Primer	695	25	57,84	64,00	11	AGGTTTTGTAAATTTTTGGTTTT
	Right Primer	934	24	53,39	54,17	6	AACCCACTTAATCATATAATCATC
	Product size: 240, T_m : 69.9, CpGs in product: 16						
3	Left Primer	527	25	57,21	64,00	5	GTAAGGGTAAGGATTAGGGTTAGAG
	Right Primer	795	24	57,63	62,50	7	AAATAACCCCAACAACAAACTAC
	Product size: 269, T_m : 70.5, CpGs in product: 9						
4	Left Primer	696	24	57,00	66,67	11	GGTTTTGTAAATTTTTGGTTTT
	Right Primer	934	24	53,39	54,17	6	AACCCACTTAATCATATAATCATC
	Product size: 239, T_m : 69.9, CpGs in product: 16						
5	Left Primer	772	24	57,63	62,50	7	GTAGTTTTGTTGTGGGGTTATTT
	Right Primer	934	24	53,39	54,17	6	AACCCACTTAATCATATAATCATC
	Product size: 163, T_m : 68.0, CpGs in product: 10						

Para a análise da segunda ilha CpG encontrada, foram simulados novos *primers* a partir do primeiro éxon, seguido por 1000pb *downstream*, gerando uma extensa região potencialmente metilável (Figura 13). Deste modo, foram sugeridas algumas opções de *primers*, partilhando a região de interesse. Destes *primers*, foram selecionados, com base nos mesmos princípios do anterior, os pares F3/R3 e F4/R4 (Figura 14; Tabela 9).

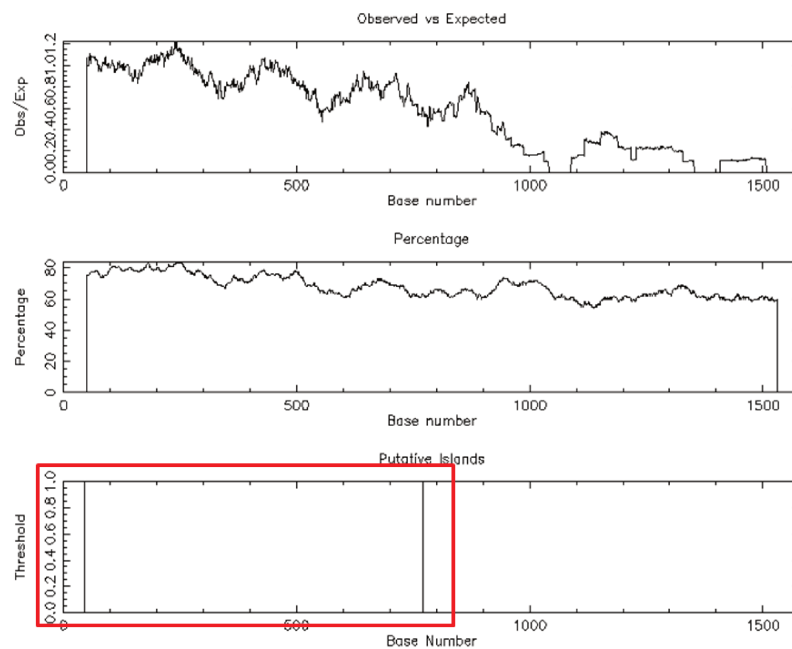


Figura 13 - Identificação da segunda região CpG do gene *AGTRL1*.

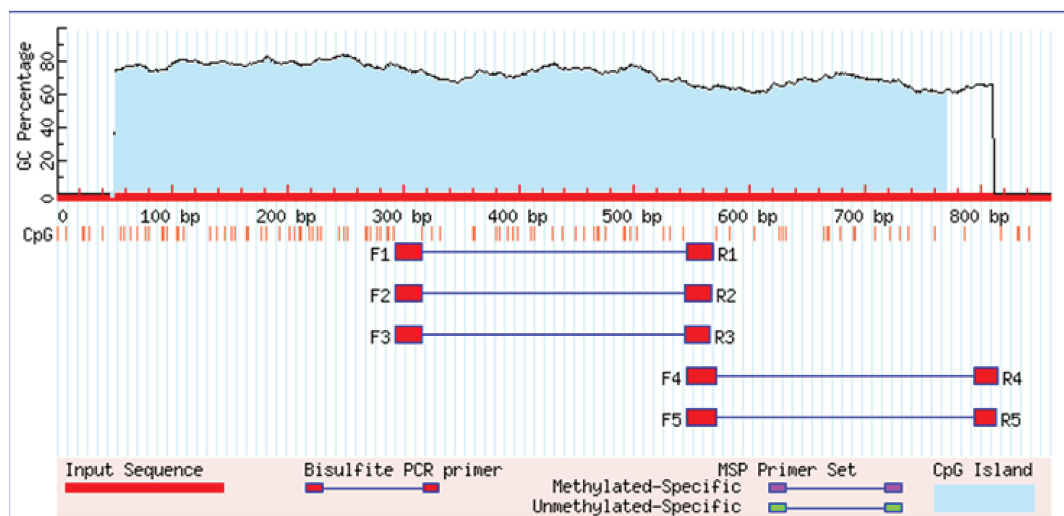


Figura 14 - Desenho de *primers* para a segunda região CpG do gene *AGTRL1*. Como a região é muito extensa, os *primers* foram desenhados por seções (*MethPrimer*). *Primers* para a primeira e segunda seção desta ilha CpG, na qual foram selecionados os pares F3/R3, para a primeira, e F4/R4, para a segunda.

Tabela 9 - Descrição dos *primers* selecionados para a segunda região CpG do gene *AGTRL1*, dentre os quais os *primers* F3/R3 e F4/R4 (em destaque) foram selecionados.

	Primer	Start	Size	Tm	GC%	"C"s	Sequence
1	Left Primer	293	24	53,46	58,33	6	GGTTTAAGGAGTAGTATGAATTTG
	Right Primer	568	24	56,88	58,33	7	AAAAACAATCCTTATAACCAACC
	Product size: 276, Tm: 79.6, CpGs in product: 27						
2	Left Primer	293	24	53,46	58,33	6	GGTTTAAGGAGTAGTATGAATTTG
	Right Primer	567	23	55,59	60,87	7	AAAAACAATCCTTATAACCAACC
	Product size: 275, Tm: 79.6, CpGs in product: 27						
3	Left Primer	293	24	53,46	58,33	6	GGTTTAAGGAGTAGTATGAATTTG
	Right Primer	566	22	54,18	63,64	7	AAAACAATCCTTATAACCAACC
	Product size: 274, Tm: 79.6, CpGs in product: 27						
4	Left Primer	545	26	59,15	53,85	7	GGTTGGTTATAAGGATTGTTTTTTT
	Right Primer	815	21	59,36	71,43	6	AACTCCCAACCTTCAAAAACC
	Product size: 271, Tm: 79.5, CpGs in product: 18						
5	Left Primer	545	26	59,15	53,85	7	GGTTGGTTATAAGGATTGTTTTTTT
	Right Primer	814	20	58,00	70,00	5	ACTCCCAACCTTCAAAAACC
	Product size: 270, Tm: 79.4, CpGs in product: 18						

b) ALOX5AP

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot não identificou regiões CpG na sequência (Figura 15).

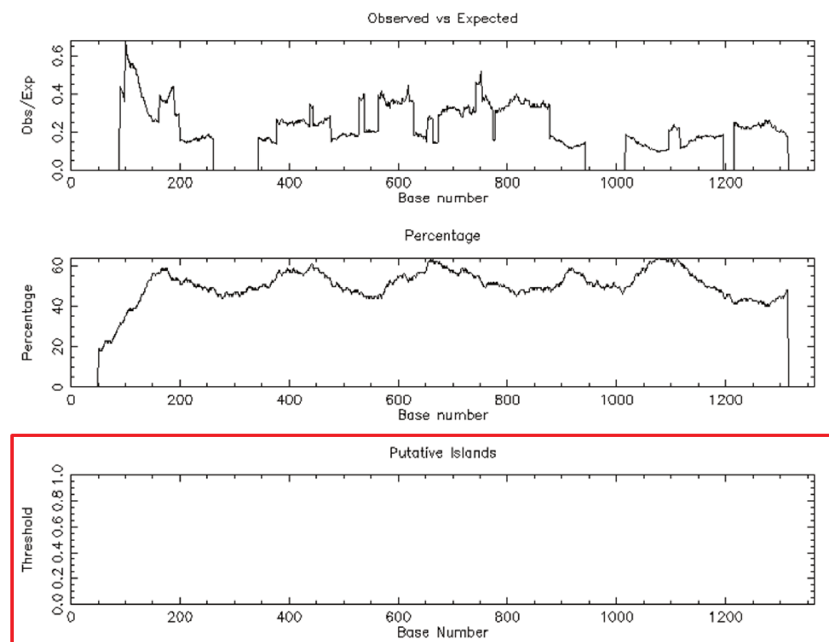


Figura 15 - Identificação de ilhasCpG (em destaque) para o gene *ALOX5AP*, a partir de análise de 1000pb *upstream* e primeiro éxon. Não foram encontradas regiões potencialmente metiláveis (EMBOSS CpGPlot).

c) ANGPT1

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot não identificou regiões CpG na sequência (Figura 16).

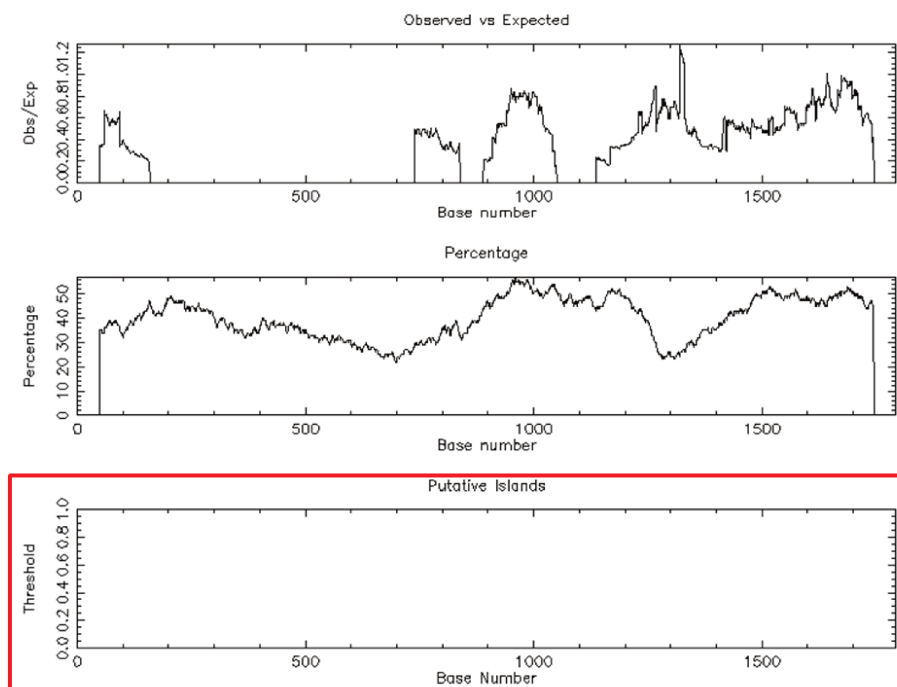


Figura 16 - Identificação de ilhas CpG (em destaque) para o gene *ANGPT1*, a partir de análise de 1000pb *upstream* e primeiro éxon. Não foram encontradas regiões potencialmente metiláveis (EMBOSS CpGPlot).

d) *CDKN2B-AS1*

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot identificou uma região CpG na sequência (Figura 17). Entretanto, como esta é uma região extensa, os *primers* desenhados no MethPrimer não a abrangem completamente, de modo que são sugeridos *primers* que partilham a região selecionada em duas (Figura 18). A região CpG apresenta 955pb de extensão, localizando-se entre 162pb e 1116pb da sequência. Os pares de *primers* selecionados para esse gene foram F1/R1 e F2/R2, devido aos melhores parâmetros apresentados (Figura 18; Tabela 10).

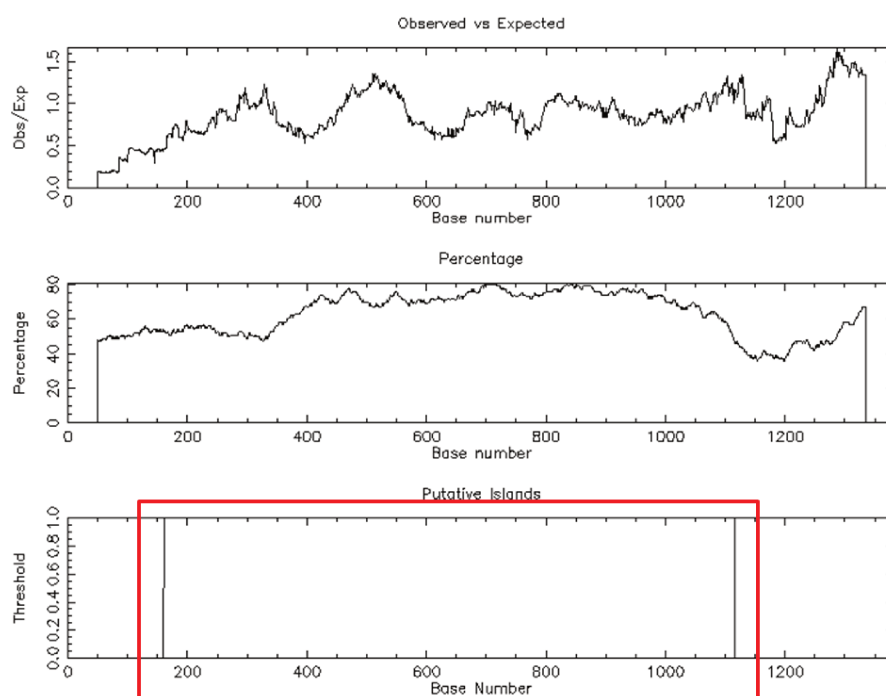


Figura 17 5- Ilha CpG identificada (em destaque) para o gene *CDKN2B-AS1*, a partir de análise de 1000pb *upstream* e primeiro éxon (EMBOSS CpGPlot).

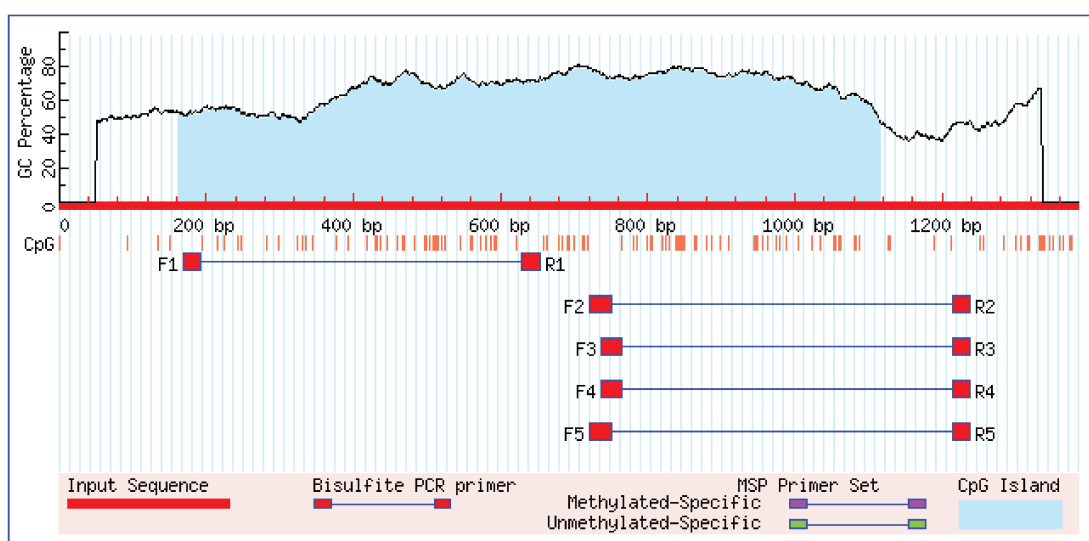


Figura 18 - Desenho de pares de *primers* para análise de possível ilha CpG do gene *CDKN2b-AS1*. Devido à extensão da ilha, foram selecionados pares que cobrissem toda a região em duas partes (F1/R1; F2/R2) (MethPrimer).

Tabela 10 - Descrição dos *primers* sugeridos para as ilhas CpGs do gene *CDKN2B-AS1*. Os *primers* selecionados foram os pares F1/R1 e F2/R2 (em destaque), para a primeira e segunda seção, respectivamente, da ilha.

	Primer	Start	Size	Tm	GC%	"C"s	Sequence
1	Left Primer	170	25	58,79	52,00	4	AGGGAGGTTTAGGTTTAATGGTATT
	Right Primer	654	26	59,31	61,54	10	TCCAATCTACAATTAACAAACAAA
	Product size: 485, Tm: 72.4, CpGs in product: 40						
2	Left Primer	720	30	54,25	80,00	23	GTTTTTATTTTATTTTATTTTATTTT
	Right Primer	1237	25	54,36	52,00	5	ACTAACATTCAACCTCCTAATTAAC
	Product size: 518, Tm: 73.1, CpGs in product: 44						
3	Left Primer	736	29	51,98	79,31	23	TTTATTTTATTTTATTTTATTTTATT
	Right Primer	1237	25	54,36	51,00	5	ACTAACATTCAACCTCCTAATTAAC
	Product size: 502, Tm: 72.4, CpGs in product: 44						
4	Left Primer	737	28	50,80	78,57	22	TTATTTTATTTTATTTTATTTTATT
	Right Primer	1237	25	54,36	52,00	5	ACTAACATTCAACCTCCTAATTAAC
	Product size: 501, Tm: 72.4, CpGs in product: 44						
5	Left Primer	721	30	53,74	76,67	23	TTTTTATTTTATTTTATTTTATTTTA
	Right Primer	1237	25	54,36	52,00	5	ACTAACATTCAACCTCCTAATTAAC
	Product size: 517, Tm: 73.0, CpGs in product: 44						

e) HDAC9

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot identificou duas regiões CpGs na sequência (Figura 19). A primeira localiza-se de 290 a 542pb, enquanto a segunda, entre 583 a 1143pb. A análise através o MethPrimer, entretanto, gerou apenas *primers* para a primeira região (Figura 20). Desse modo, foi selecionado o par de *primers* F2/R2 (Tabela 11).

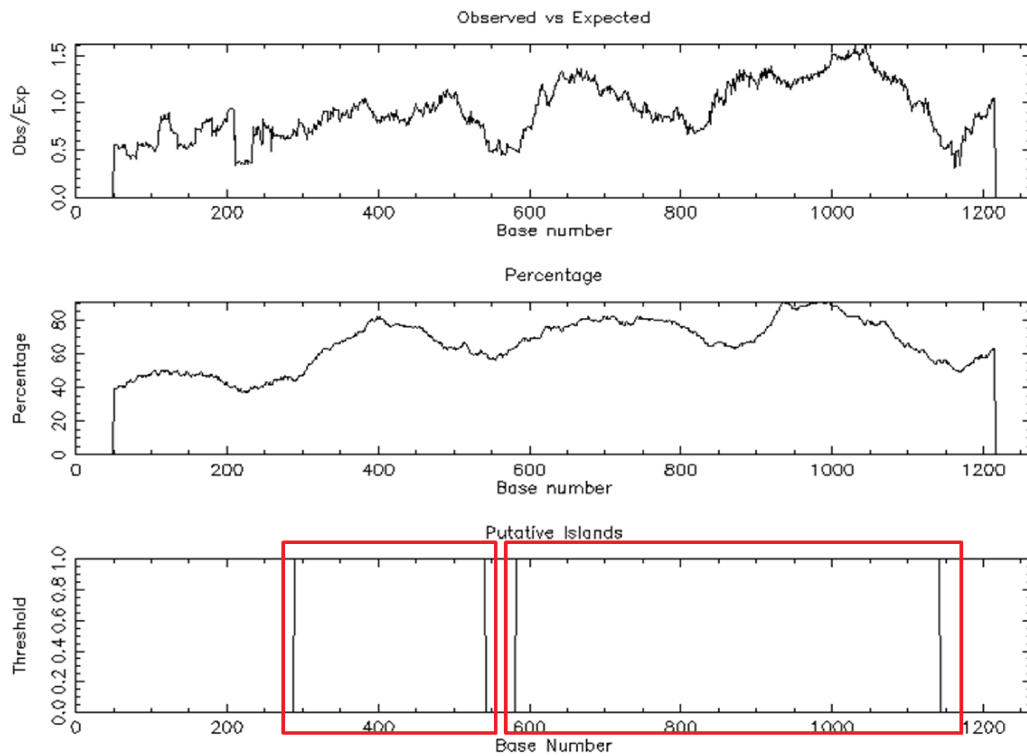


Figura 19 - Identificação de Ilhas CpGs (em destaque) para o gene *HDAC9*, a partir de análise de 1000pb *upstream* e primeiro éxon (EMBOSS CpGPlot).

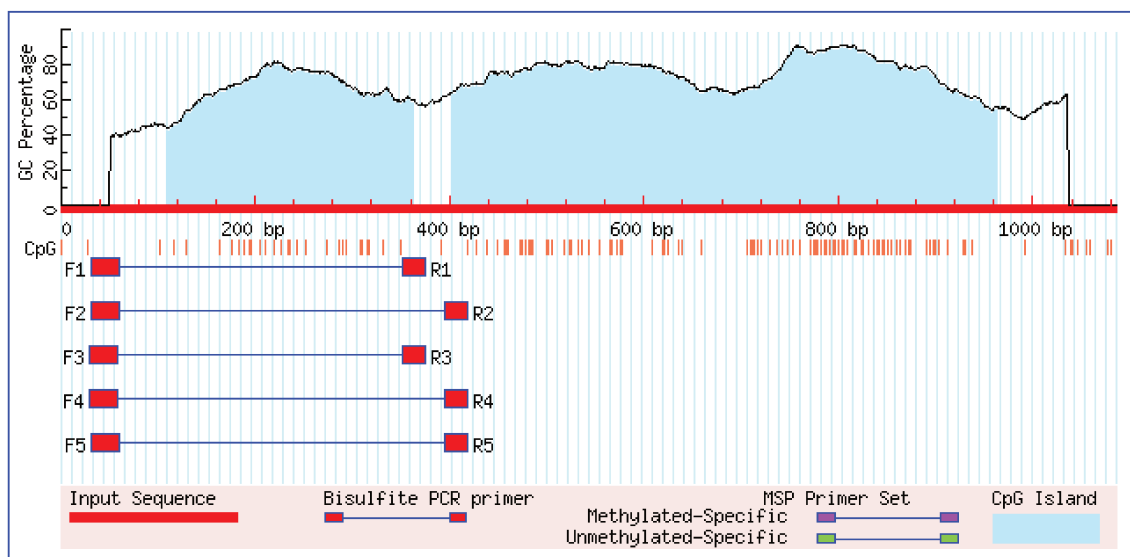


Figura 20 - Desenho de pares de *primers* para análise das possíveis ilhasCpGs do gene *HDAC9*. Foram sugeridos *primers* apenas para a primeira região identificada, de modo que o par F2/R2 foi o selecionado (MethPrimer).

Tabela 11 - Descrição dos *primers* sugeridos para a primeira ilha CPG encontrada no gene *HDAC9*. O par F2/R2 (em destaque) foi selecionado.

	Primer	Start	Size	Tm	GC%	"C"s	Sequence
1	Left Primer	3	30	54,62	40,00	9	TTTTTTATTGATAAATTTTAAATGTAGTA
	Right Primer	375	24	56,44	58,33	9	AAAAAAAACCAAAAACCTAAAC
	Product size: 345, Tm: 71.8, CpGs in product: 27						
2	Left Primer	31	30	54,62	40,00	9	TTTTTTATTGATAAATTTTAAATGTAGTA
	Right Primer	418	25	54,61	64,00	9	AAAAAAAACCTAAACTACACCAAC
	Product size: 388, Tm: 72.0, CpGs in product: 28						
3	Left Primer	30	30	55,17	43,33	9	GTTTTTTATTGATAAATTTTAAATGTAGT
	Right Primer	375	24	56,44	58,33	9	AAAAAAAACCAAAAACCTAAAC
	Product size: 346, Tm: 71.8, CpGs in product: 27						
4	Left Primer	30	30	55,17	43,33	9	GTTTTTTATTGATAAATTTTAAATGTAGT
	Right Primer	418	25	54,61	64,00	9	AAAAAAAACCTAAACTACACCAAC
	Product size: 389, Tm: 72.0, CpGs in product: 28						
5	Left Primer	32	30	53,91	40,00	9	TTTTTATTGATAAATTTTAAATGTAGTAT
	Right Primer	418	25	54,61	64,00	9	AAAAAAAACCTAAACTACACCAAC
	Product size: 387, Tm: 72.0, CpGs in product: 28						

f) *PDE4D*

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot não identificou regiões CpG na sequência (Figura 21).

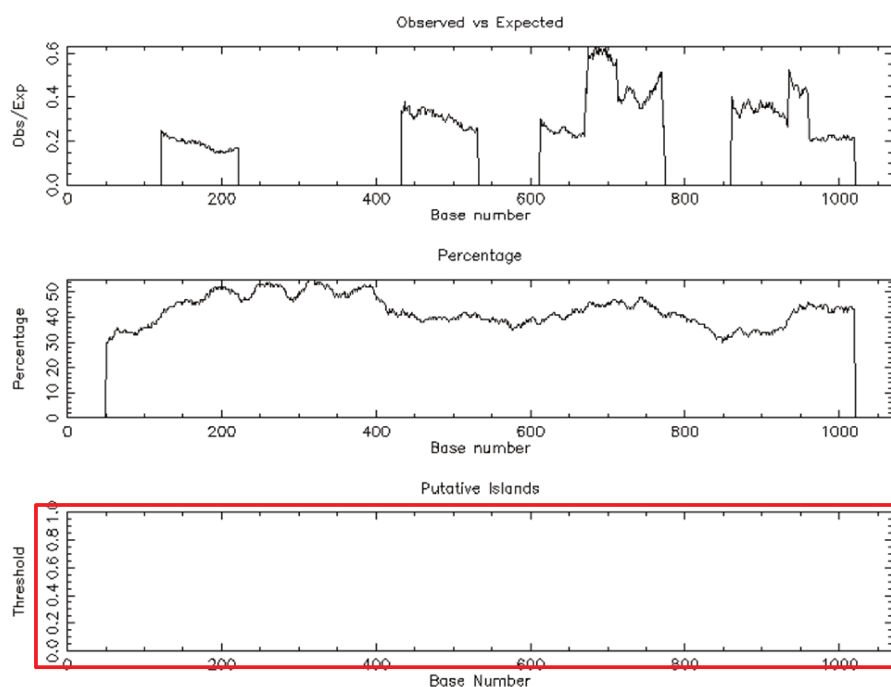


Figura 216 - Identificação de ilhas CpG (em destaque) para o gene *PDE4D*, a partir de análise de 1000pb *upstream* e primeiro éxon. Não foram encontradas regiões potencialmente metiláveis (EMBOSS CpGPlot).

g) PITX2

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot não identificou regiões CpG na sequência (Figura 22).

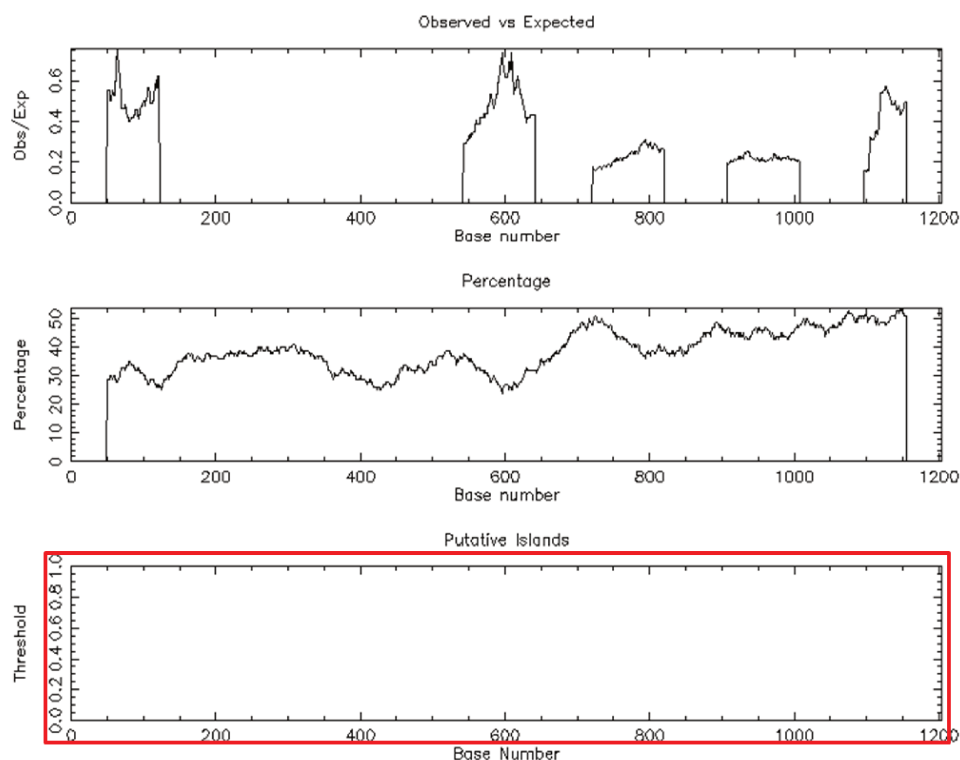


Figura 22 - Identificação de ilhas CpG (em destaque) para o gene *PDE4D*, a partir de análise de 1000pb *upstream* e primeiro éxon. Não foram encontradas regiões potencialmente metiláveis (EMBOSS CpGPlot).

h) ZFX3

Ao analisar os 1000pb *upstream* mais o primeiro éxon do gene, o software EMBOSS CpGPlot identificou uma região CpG na região exônica da sequência (Figura 23). Entretanto, ao analisar essa região no software MethPrimer, foi identificada uma segunda e menor ilha *upstream*. A discrepância entre as ilhas identificadas pelo EMBOSS e MethPrimer, já que este segundo identificou uma ilha *upstream* a identificada pelo primeiro, pode ser devido as diferenças nas configurações basais entre os dois softwares, principalmente quando relacionado ao tamanho das ilhas. Assim, as ilhas

identificadas foram: a primeira entre 855pbe 955pb; e a segunda entre 1248pbe 1664pb (Figura 24). Deste modo, foram selecionados dois pares de *primers* que abrangeriam ambas as ilhas encontradas, F4/R4 e F5/R5 (Figura 24; Tabela 12)

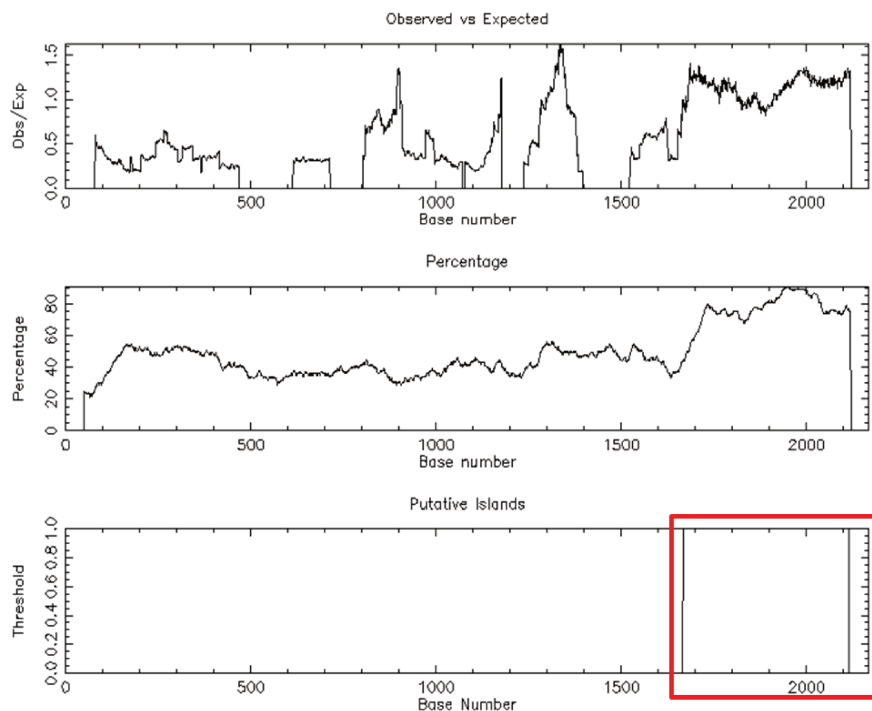


Figura 23-Identificação de ilhas CpG (em destaque) para o gene *ZFHX3*, a partir de análise de 1000pb *upstream* e primeiro éxon. Foi identificada uma ilha potencial ao final da sequência input (EMBOSS CpGPlot).

Tabela 12 - Descrição dos *primers* seleccionados para as regiões CpGs do gene *ZFX3*. Os pares de *primers* seleccionados foram F5/R5e F4/R4 (em destaque), que abrangem a primeira e segunda ilha, respectivamente.

	Primer	Start	Size	Tm	GC%	"C"s	Sequence
1	Left Primer	1224	23	53,89	28,00	4	TAGAATAAAGATTTAAAGATTTTTT
	Right Primer	1643	20	54,96	65,00	6	AAACCCATCAAAACAACAAC
	Product size: 420, Tm: 64.4, CpGs in product: 59						
2	Left Primer	1223	26	51,02	30,77	4	GTAGAATAAAGATTTAAAGATTTTTT
	Right Primer	1643	20	54,96	65,00	6	AAACCCATCAAAACAACAAC
	Product size: 421, Tm: 64.4, CpGs in product: 59						
3	Left Primer	1222	27	52,03	29,63	4	AGTAGAATAAAGATTTAAAGATTTTTT
	Right Primer	1643	20	54,96	65,00	6	AAACCCATCAAAACAACAAC
	Product size: 422, Tm: 64.4, CpGs in product: 59						
4	Left Primer	1157	23	53,18	52,17	4	GTGTTTGGGTAATGAAGTATATG
	Right Primer	1643	20	54,96	65,00	6	AAACCCATCAAAACAACAAC
	Product size: 487, Tm: 64.7, CpGs in product: 60						
5	Left Primer	842	26	53,18	42,31	8	TTGTTGTTTTTATGTATTTTATTTTT
	Right Primer	1179	23	53,18	52,17	4	CATATACTTCATTACCCAAACAC
	Product size: 338, Tm: 65.0, CpGs in product: 9						

2.2. Conversão por Bissulfito e MS-PCR

As amostras foram convertidas pela técnica de bissulfito utilizando o kit *EzDna-Methylation- Lighting Kit* da ZymoResearch. A quantificação média das amostras foi 15ng/μL, de modo que a recuperação média das amostras de DNA inicial/convertido foi de 77%. Essa conversão está abaixo do proposto pelo kit, cuja recuperação é dita como maior que 80%.

Contudo, ao analisarmos as amostras pela técnica de PCR, verificamos que não houve amplificação das mesmas nas condições inicialmente testadas nesse trabalho. Estudos adicionais testando outras condições e técnicas podem ser realizados no futuro.

CAPÍTULO III – Microarranjos de DNA e CNVs

1. Métodos

1.1. Amostras

1.1.1. Amostras provenientes do Biobanco de Joinville/SC

Para este estudo foram utilizadas as amostras de AVC hemorrágico e indivíduos controles provenientes do Biobanco de AVC de Joinville/SC. Neste estudo, foram avaliadas 86 amostras, sendo 45 controles e 41 pacientes. A diferença entre esses números é devido à qualidade das amostras. Infelizmente, 15 amostras de pacientes e 19 amostras de controles não apresentaram razão 260/280 e absorbância a 320nm significativas para a análise pela técnica de microarranjos.

1.1.2. Amostras provenientes de colaboração com Universidade de Pisa/IT

As amostras encaminhadas em colaboração internacional com o grupo de pesquisa do Dr. Michelangelo Mancuso, da Universidade de Pisa – Itália, foram analisadas de acordo com o método de microarranjos de DNA. Foram encaminhadas 18 amostras de pacientes com AVC hemorrágico e 30 controles, entretanto, devido a baixa qualidade das amostras, de ambos os grupos, apenas amostras de 12 pacientes foram analisadas.

1.2. Microarranjos de DNA

As amostras de ambas as coortes foram analisadas utilizando a mesma técnica e análise. Assim, elas foram submetidas à análise de CNV através da técnica de microarranjos de DNA, por meio do chip *Genome-WideHuman SNP Array 6.0* (Affymetrix Inc.). Esses chips possuem mais de 1.8 milhões de marcadores genéticos, incluindo 906600 polimorfismos de base única (SNPs) e 946000 sondas para detecção de variações nos números de cópias (CNVs). Os dados foram analisados no software Genotyping Console® (Affymetrix Inc.) A identificação dos genótipos e das CNVs foi realizada por meio dos algoritmos de Modelo Linear Robusto Bayesiano usando Mahalanobis (BRLMM) e *Canary*, respectivamente. Uma possível análise a ser realizada com as CNVs identificadas, é classificá-las de acordo com a quantidade de

eventos por regiões cromossômicas ou por amostra. Essa análise é conhecida por CNVR e é realizada pelo software JMCNV (<http://www.lmb.cnptia.embrapa.br/tools/JMCNV/JMCNVUpload>).

Considerando que esse é um estudo descritivo, utilizamos a estratégia de comparação com base em uma população referência. Essa população foi criada a partir das amostras controles encaminhadas de Joinville/SC. As amostras de pacientes foram, portanto, comparadas uma a uma, com toda a população referência criada, visando a identificação de variantes que estejam presentes apenas nas amostras de pacientes que e possam estar vinculadas com algum efeito sobre o AVC (variantes causais).

Os parâmetros utilizados para a identificação das CNVs se basearam no número de marcadores por região: para ser considerado um ganho genético, as variantes deveriam conter, no mínimo, 50 marcadores; para ser considerada uma região de perda genética, as variantes deveriam conter, no mínimo, 25 marcadores. A partir dos genes encontrados, uma rede gênica foi construída no software METACORE®(Thomas-Reuter) para confirmar possíveis interações dos genes encontrados e suas variantes.

Após a identificação de genes potencialmente relacionados ao AVC, as variantes foram comparadas em dois bancos de dados de variantes: o *Database of Genomic Variants* (DGV) (<http://dgv.tcag.ca>) e o DECIPHER(<http://decipher.sanger.ac.uk/>). O primeiro é um banco de dados de variantes estruturais maiores que 50kb, presente no genoma humano. Esse banco de dados é uma relevante fonte de variantes normais, permitindo sua comparação com variantes patogênicas. O segundo banco de dados, por sua vez, descreve variantes relevantes clinicamente, utilizando variantes que já foram identificadas com descrições fenotípicas patogênicas, se tornando uma significativa fonte para identificação de mutações potencialmente relacionadas a doenças.

A fim de confirmar se os grupos de pacientes e controles são geneticamente semelhantes para comparação, foi realizada uma análise de estratificação da população por meio da análise de componentes principais (*principal component analysis* - PCA) utilizando o programa R [48]. Esse gráfico se baseia na comparação dos componentes principais 1 e 2, obtidos a partir de dados da genotipagem dessas amostras.

A análise do PCA foi realizada utilizando um programa desenvolvido em nosso laboratório (*mahalanobisDist-package*), o qual determina a distância de *Mahalanobis* entre os casos e controles, estima a distribuição nula por meio de permutação, e compara os dois valores, resultando em um valor de p, sendo que valores de $p < 0,05$ indicam evidência de estratificação da população. Para a permutação, foram utilizadas 1.000.000 replicações.

2. RESULTADOS

2.1. Variação no Número de Cópias Genômicas de Pacientes com Acidente Vascular Cerebral Hemorrágico

2.1.1. Coorte Brasileira

A análise das amostras de pacientes ocorreu confrontando-as com uma população referência, criada a partir das amostras controles, provenientes da mesma região. Isso permite que haja menor variação entre os grupos. Para confirmar essa teoria, foi realizado um gráfico de PCA entre esses dois grupos estudados. Os resultados de PCA indicam que não há evidência de estratificação entre casos e controles (valor de $p = 0,2185$) para a população de Joinville (Figura 25).

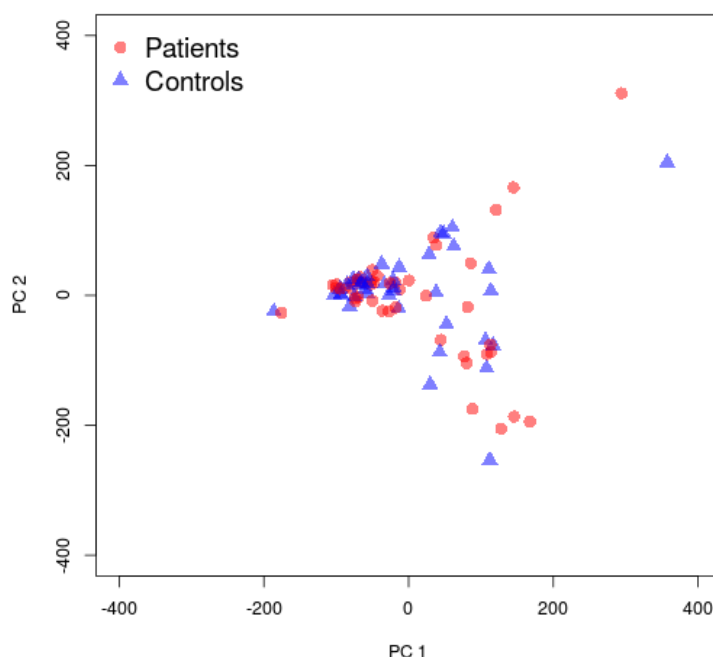


Figura 25 - PCA plot das amostras provenientes de Joinville/SC, comparando indivíduos controles e pacientes com AVC hemorrágico (R Software).

Os experimentos sempre foram executados com amostras controles e de pacientes concomitantemente e distribuídas de forma aleatória, a fim de reduzir erros devido ao *batch effect* dos experimentos.

De acordo com os resultados, foram identificadas 384 regiões CNVs, que englobavam 191 genes. Foram identificadas 24 CNVs inéditas de acordo com análise do Genotyping Console, que se localizavam, principalmente, nos cromossomos X e Y, de modo que apenas sete CNVs localizam-se nos cromossomos 2, 3, 4, 8, 9 e 10. As variantes germinativas não foram analisadas nesse projeto, devido à alta taxa de recombinações desses cromossomos. As variantes inéditas somáticas também foram comparadas com os bancos de dados DGV e DECIPHER. Contudo, maiores análises são necessárias para maior compreensão da função dessas CNVs.

Através da identificação das CNVs, é possível classificá-las e analisá-las com base na sua quantidade por amostra ou por região cromossômica (CNVRs). Assim, foram identificadas 3 potenciais regiões de CNVRs, sendo que a presente no cromossomo 15q11.2 engloba vários pseudogenes e regiões regulatórias, que não causam efeito imediato sobre o mecanismo do AVC (Figura 24). As outras duas regiões selecionadas são as do cromossomo 8p23.1 e 17q21.31, que englobam os genes *DEFB4B* e *NSF*, relacionados com processos imunológicos e reparos após danos vasculares, respectivamente (Figura 26).

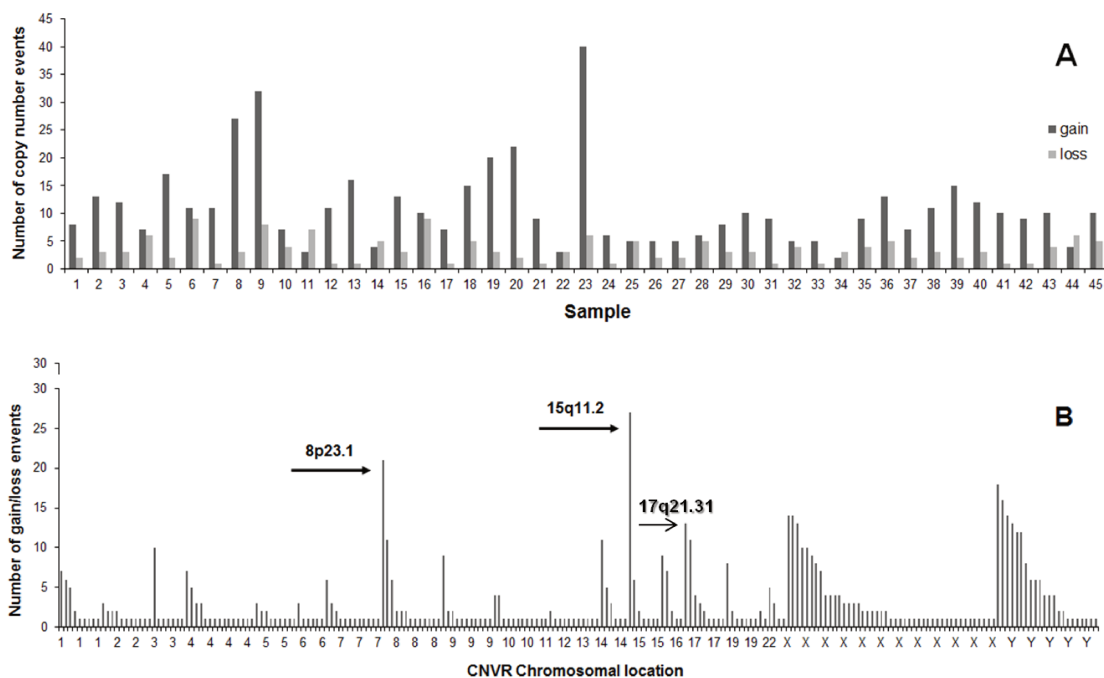


Figura 26 - Regiões com altas taxas de CNVs (A) por amostras e (B) por região cromossômica

Após a análise das CNVs, os genes identificados nas variantes foram analisados pelo software METACORE. Para essa análise foram excluídos apenas os pseudogenes identificados para as amostras. Assim sendo, foi possível selecionar 19 genes potencialmente relacionados a fisiopatologia do AVC (Tabela 13). Essas variantes são tanto duplicações como deleções, conforme Tabela 13, e englobavam as sequências completas desses genes.

Tabela 13 - Genes identificados pelo METACORE como potencialmente relacionados com processos neurológicos e vasculares, potencialmente relacionados com AVC.

Genes Candidatos	CNV	Banda Cromossômica	Tamanho (kb)	Número de Marcadores	Número de Cópias	Número de Indivíduos
ANXA8	Duplicação	10q11.22	483 (325 - 641)	160 (69 - 251)	3	2
	Deleção	10q11.22	186	67	1	1
CDH13	Duplicação	16q23.3	368	378	3	1
CYP2E1	Duplicação	10q26.3	133	108	3	1
DEFB103A	Duplicação	8p23.1	552 (429 - 612)	124 (84 - 147)	3	8
	Deleção	8p23.1	534 (521 - 548)	121 (117 - 125)	1	2
DEFB103B	Duplicação	8p23.1	552 (429 - 612)	124 (84 - 147)	3	8
	Deleção	8p23.1	534 (521 - 548)	121 (117 - 125)	1	2
DEFB4B	Duplicação	8p23.1	552 (429 - 612)	124 (84 - 147)	3	8
	Deleção	8p23.1	534 (521 - 548)	121 (117 - 125)	1	2
FGL2	Duplicação	7q11.23	424	270	3	1
FNTA	Duplicação	8p11.21	887	241	3	1
HGSNAT	Duplicação	8p11.21	887	241	3	1
MACROD2	Deleção	20p12.1	199	122	1	1
MBTPS1	Duplicação	16q23.3	368	378	3	1
MLYCD	Duplicação	16q23.3	368	378	3	1
NAALADL2	Deleção	3q26.31	381	252	1	1
NBPF1	Duplicação	1p36.13	388	121	3	1
NOMO1	Duplicação	16p13.11	251	58	3	1
NPY4R	Duplicação	10q11.22	483 (325 - 641)	160 (69 - 251)	3	2
	Deleção	10q11.22	186	67	1	1
NSF	Duplicação	17q21.31	502 (407 - 620)	112 (55 - 173)	3	8
	Deleção	17q21.31	344 (277 - 390)	39 (27 - 48)	1	10
PRODH	Duplicação	22q11.21	122	92	3	1
	Deleção	22q11.21	131	95	1	2
RAF1	Duplicação	3p25.2	176	206	3	1
RB1CC1	Duplicação	8q11.23	196	131	3	1

kb = kilobases

Ao analisarmos as variantes encontradas na nossa coorte no banco de dados DGV, pudemos verificar que quase todas estavam presentes na população geral. Entretanto, 4 delas não apresentam descrições de variantes que englobem a sequência gênica completa (Tabela 14).

Tabela 14 - Genes identificados no banco de dados DGV, que não apresentam variantes descritas que englobem toda a sequência gênica.

Genes Candidatos	CNV	Banda Cromossômica	Tamanho (kb)	Número de Marcadores	Número de Cópias	Número de Indivíduos
<i>CDH13</i>	Duplicação	16q23.3	368	378	3	1
<i>FNTA</i>	Duplicação	8p11.21	887	241	3	1
<i>MACROD2</i>	Deleção	20p12.1	199	122	1	1
<i>NAALADL2</i>	Deleção	3q26.31	381	252	1	1

Ao comparar as variantes inéditas identificadas pelo *Genotyping Console* com o banco de dados DGV, três variantes (presentes nos cromossomos 3, 8 e 10) não apresentaram descrição no banco de dados de variantes controles (Tabela 15). Assim, uma hipótese é que essas variantes possam influenciar a genética e a ocorrência do AVC. Contudo, suas funções genéticas ainda não foram previstas devido à falta de elementos funcionais descritos.

Tabela 15 - Descrição das variantes inéditas de acordo com o banco de dados DGV. As variantes dos cromossomos 3, 8 e 10 não apresentaram variantes descritas no banco de dados que abrangessem toda a região CNV identificada no estudo. As variantes presentes no cromossomo 2, 4 e 9 foram descritas no banco de dados com mesmo tamanho e variação do encontrado no estudo.

DGV ID	Cromossomo	Posição	Ganho/Perda
nsv584432	2	214914089-215070272	Perda
	3	133822625-133964424	Ganho
esv3569319	4	18802753-18933759	Perda
	8	136497296-136620534	Perda
nsv1051081	9	131235207-131344034	Ganho
nsv1051081	9	131236201-131344034	Ganho
	10	30840200-30945458	Perda

Ao analisar os genes identificados pelo METACORE no banco de dados DECIPHER, foram selecionados 8 genes que pudessem estar relacionados com algum fenótipo semelhante ao AVC (Tabela 16). Contudo, nem sempre as variantes descritas no software englobam apenas fenótipos específicos para esse evento. Por exemplo, hipertensão é um dos principais fatores de risco para o AVC, porém, também está presente em outras doenças.

Sendo que esses fatores ainda não apresentam uma genética bem elucidada e eles podem sofrer influência da variante gênica em questão, qualquer fenótipo potencialmente relacionado ao AVC, sendo específico ou não, foi incluído na análise, mantendo-se uma ressalva para os fenótipos concomitantes apresentados pela síndrome descrita.

Tabela 16 - Genes identificados pelo banco de dados DECIPHER, como relacionados com fenótipos semelhantes ao AVC. Os fenótipos marcados com (*) foram selecionados com ressalvas, já que foram descritos concomitantemente a outros fenótipos, os quais nem sempre se relacionavam a eventos neurovasculares.

DECIPHER ID	Genes Candidatos	Intervalo (Mb)	Principal Fenótipo Descrito no DECIPHER*	Pacientes
250056	CYP2E1	10.25	Hipertensão*	3
875		3.47	Anormalidade no sistema nervoso	1
261038	DEFB4B	5.29	Anormalidade nos vasos sanguíneos*	1
294232	MACROD2	0.176	Cardiomiopatia	1
288510	NOMO1	1.61	Má formação do coração e vasos sanguíneos*	2
274067		15.4	Anormalidade nos vasos sanguíneos*	2
293487	NSF	1.96	Comprometimento neurológico da fala	1
255533	PRODH	3.29	Anormalidade nos vasos sanguíneos*	2
256303		3.16	Anormalidade no coração* Anormalidade nos vasos sanguíneos*	2
280373	RAF1	39.59	Anormalidade no coração*	2
267704	RB1CC1	3.79	Aumento da pressão intracranial*	1

Quando avaliamos as variantes inéditas identificadas pelo estudo e não presentes no banco de dados normais, foi possível identificar alguns fenótipos relacionados com processos vasculares, que podem estar envolvidos na ocorrência do AVC (Tabela 17). A variante identificada no cromossomo 3 apresenta potencial relação com o evento devido à descrição de um único fenótipo para a síndrome em questão: Anormalidade no coração. Seu efeito na população ainda não é conhecido.

Tabela 17 – Descrição das variantes inéditas encontradas no estudo com relação aos fenótipos descritos no banco de dados DECIPHER. Os fenótipos marcados com (*) foram selecionados com ressalvas, já que foram descritos concomitantemente a outros fenótipos, os quais nem sempre se relacionavam a eventos neurovasculares.

DECIPHER ID	Variante	Intervalo (Mb)	Fenótipo	Pacientes
263046	Chr3:125730698-139379982	13.65	Anormalidade no coração	1
296426	Chr8: 47536057-14617403	98.64	Agenesia do corpo caloso*; Defeito do septo ventricular*	2
296426	Chr10: 27133014-37893252	10.76	Regurgitação aórtica*	1

2.1.2. Coorte Italiana

As amostras de pacientes italianos foram comparadas com a população referência de controles brasileiros, proveniente de Joinville/SC. Essa iniciativa ocorreu baseada na ancestralidade européia dessa região, de modo que havia uma hipótese de não estratificação amostral. Entretanto, ambas as coortes apresentam estrutura genética diferente, como demonstrado pelo gráfico de PCA (Figura 27), o que tende a enviesar as análises.

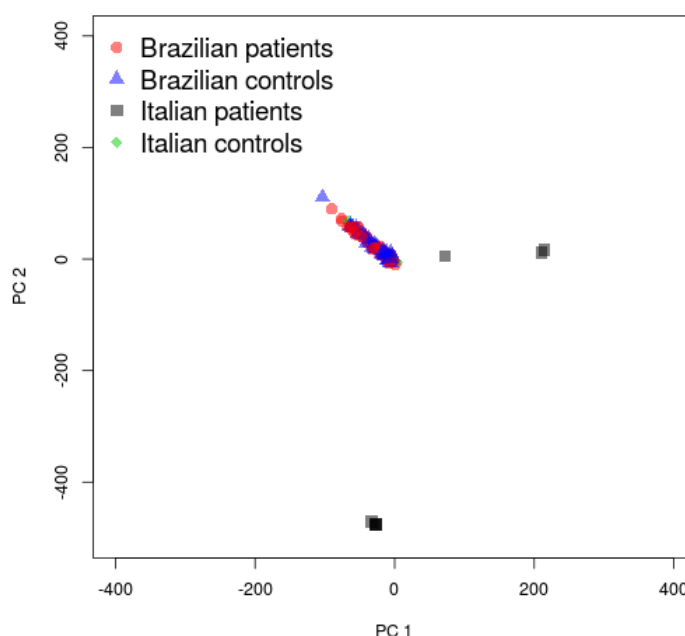


Figura 27 - PCA comparando as amostras de pacientes brasileiros (círculo vermelho), controles brasileiros (triângulo azul), pacientes italianos (quadrado cinza) e controles italianos (losango verde). Visualmente, é possível identificar que há uma estratificação amostral entre a coorte de pacientes italianos com as outras coortes, de modo que análises entre esses grupos tende a estar enviesada (R Software).

Contudo, quando comparamos a coorte de pacientes italianos com alguns controles provenientes da mesma região, a análise também aparece estratificada (Figura 28). Desse modo, ainda é necessário confirmar o motivo da estratificação, que pode estar relacionada a diferentes regiões de origem dos pacientes e controles, a técnica de extração de DNA, a diferenças nas metodologias utilizadas para microarranjos ou a própria estrutura genética da população examinada.

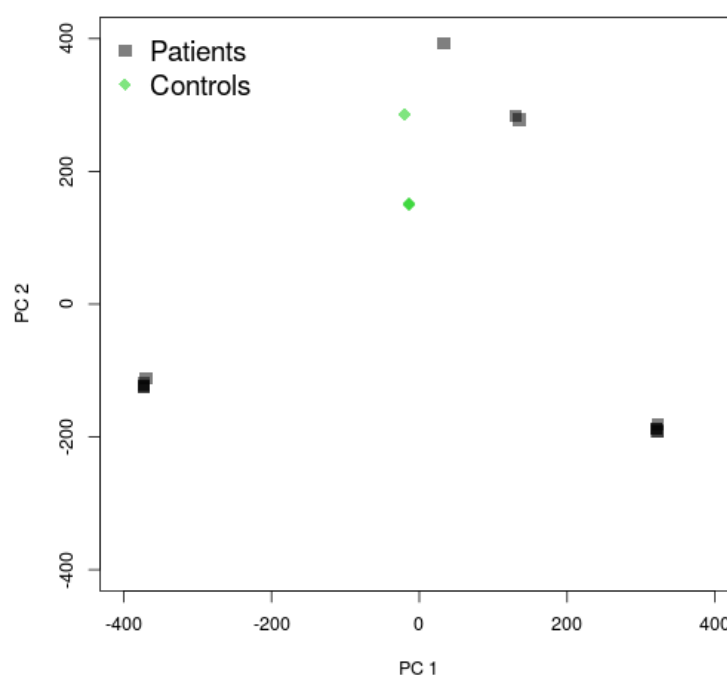


Figura 28-PCA comparando as amostras pacientes italianos (quadrado cinza) e controles italianos (losango verde). Visualmente, é possível identificar que não há um agrupamento das amostras, gerando uma estratificação amostral entre as coortes. Desse modo, as análises entre esses grupos tende a estar enviesada (R Software).

Apesar da ressalva quanto a diferenças na estrutura genética das coortes, a coorte de pacientes italianos foi analisada em comparação com a população referência brasileira, a fim de verificar semelhanças e diferenças entre ambas as populações.

Desse modo, foram identificados 29 genes em comum entre a coorte de pacientes italianos e a de pacientes brasileiros. Quando realizada comparação entre os genes identificados na coorte italiana com os 19 identificados pelo Software METACORE, 2 genes foram identificados em

comum: *MACROD2* e *NSF* (Tabela 18). Entretanto, é importante ressaltar que esse resultado ainda é parcial e pode estar enviesado.

Tabela 18 – Genes identificados na coorte italiana e após análise pelo software METACORE na coorte brasileira.

Genes Candidatos	CNV	Bandas Cromossômicas	Tamanho (kb)	Número de Marcadores	Número de Cópias	Número de Indivíduos
<i>MACROD2</i>	Deleção	20p12.1	143,5 (114 - 173)	163 (138 - 188)	1	2
<i>NSF</i>	Duplicação	17q21.31	134,5 (117 - 152)	75,5 (58 - 93)	3	2
	Deleção	17q21.31	329 (322 - 337)	34,5 (34 - 35)	1	3

3. Discussão

Variações genéticas no genoma humano podem causar significativa influência em várias doenças [28]. Essas variações podem ocorrer de diferentes maneiras: alterações de uma única base na sequência gênica, grandes variações estruturais ou, ainda, modificações covalentes que não influenciam a sequência de nucleotídeos em si, mas a expressão dessas regiões [22, 27, 28]. A expressão gênica é um dos processos mais relevantes nos organismos, tendo várias funções fundamentais, por exemplo, transcrição da sequência gênica, tradução protéica e diferenciação celular [49]. Alterações na expressão gênica são potenciais causas de variabilidade e estão diretamente relacionadas à ocorrência de doenças.

Epigenética é uma modificação covalente altamente responsável por alterações na expressão gênica [22]. Uma forma dessa modificação é a adição de grupos metil (-CH₃) em resíduos de citosinas, comumente seguidos por guaninas, distribuídos ao longo de uma região genômica [22, 23]. Essa região, com alta presença de citosinas e guaninas, é conhecida por ilha CpG [22, 25]. Ilhas CpGs são comumente encontradas nas regiões promotoras dos genes, o que influencia no controle da expressão gênica [25]. Ao analisarmos a presença dessas ilhas nos promotores gênicos, é possível estabelecer um parâmetro de análise para que apenas os genes com potencial influência epigenética sejam avaliados nos experimentos.

No caso deste estudo, apenas 4 genes apresentam potenciais ilhas CpGs nas regiões promotoras. Isso evidencia que há uma probabilidade desses genes terem expressão influenciada por metilações. Estudos mostram que 50% dos genes expressos constitutivamente apresentam alta frequência de ilhas CpGs, porém nem todos sofrem influência de metilações para a regulação da expressão gênica [26]. Contudo, essa comprovação somente é possível com estudos experimentais. Infelizmente, as amplificações dessas regiões promotoras pela técnica tradicional de PCR não foram eficazes. Esse fato pode ser devido a erros no desenho de *primers*, a problemas durante a conversão por bissulfito das amostras ou a técnica em questão. É possível que técnicas mais recentes e sensíveis possam gerar resultados mais satisfatórios

e conclusivos. Um método de interesse seria o sequenciamento dessas sequências promotoras, seguido por análise comparativa das regiões metiladas e não metiladas.

Entretanto, para os genes que, de acordo com análises *in silico*, não apresentaram regiões potencialmente metiláveis nos promotores, não há evidências de que eles não possam sofrer outros tipos de regulação epigenética. Sítios regulatórios, como sítios de ligação de microRNAs, são importantes fatores que podem influenciar a expressão ou função gênica após a transcrição[50, 51]. Inicialmente, pensava-se na presença de sítios de microRNAs apenas na região 3'UTR do gene, de modo que a ligação do microRNA nessa região interferisse na tradução protéica[50]. Contudo, novas descobertas demonstram a interação de microRNAs com diferentes regiões genômicas (3'UTR, 5'UTR, promotores e sequências codificantes), promovendo a regulação da expressão[50]. Contudo, a presença de polimorfismos, principalmente não-sinônimos, na sequência alvo dos microRNAs pode desregular a interação microRNA-alvo, alterando os padrões de expressão do gene [51].

Dentre os polimorfismos identificados por estudos de metanálises para AVC, 14 polimorfismos, presentes em 8 genes, apresentam-se nas regiões não traduzidas, como intrônicas, intergênicas e 3'UTR, o que pode afetar a expressão gênica em doenças complexas [51]. Chen et al (2010) identificou uma interação entre o polimorfismo rs2507800, presente na região 3'UTR do gene *ANGPT1*, e o microRNA *mir-211*[52]. De acordo com esse trabalho, o alelo T do polimorfismo pode aumentar o risco para AVC através do enfraquecimento da ligação microRNA-alvo, gerando uma redução na expressão do gene e, conseqüentemente, aumentando a vulnerabilidade dos vasos sanguíneos [52]. Assim, essa associação pode se tornar um fator de risco independente para o AVC [52].

Erros durante recombinações cromossômicas são outra potencial fonte de alterações na expressão gênica. CNVs são importantes mutações no genoma com significativa influência na variabilidade genética[30]. Essas mutações podem ser associadas com diversidade normal ou patogênica e tem

sido amplamente utilizada para diagnósticos de doenças ou compreensão genética[30]. As técnicas de microarranjos de DNA têm auxiliado na compreensão dessas CNVs, principalmente pela significante resolução na detecção dessas variantes, quando comparado com as técnicas de primeira geração [35]. Assim, utilizando uma interpretação baseada em conjuntos de confiança das regiões candidatas, é possível ampliar a quantidade de variantes identificadas com poder funcional [35].

Ao identificar as variantes presentes na coorte, os genes englobados por elas devem ser avaliados, verificando sua viabilidade como variantes funcionais[32]. Para isso, existem bancos de dados, tanto normais quanto patogênicos, que exibem a frequência e o efeito dessas variantes nas populações[32]. O primeiro deles é o DGV, que exibe variantes presentes na população normal, ou seja, sem descrição de doenças, com base em estudos populacionais (<http://dgv.tcag.ca>). O segundo banco de dados é o DECIPHER, cujo objetivo é catalogar variantes já identificadas na literatura que tenham efeito sobre algum fenótipo (<http://decipher.sanger.ac.uk/>).

Os genes identificados por esse estudo, então, se relacionam, basicamente com processos de angiogênese, coagulação sanguínea, imunologia, desenvolvimento de tecidos, principalmente cerebral e cardíaco, metabolismo de lipídeos e apoptose celular.

Os processos vasculares são relevantes para doenças cardiovasculares, como o AVC, devido, principalmente, a sua capacidade de permitir o crescimento e desenvolvimento dos vasos sanguíneos, assim como sua influência na manutenção da integridade dos mesmos. Essas características englobam ampla complexidade regulatória e genética, o que dificulta a seleção de genes ou processos isolados que possam causar grande efeito fenotípico. A formação de novos vasos (angiogênese) ocorre, principalmente, após danos teciduais e pode estar relacionados a diversos mecanismos [2]. O principal fator angiogênico, fator de crescimento do endotélio vascular (VEGF), é responsável pela permeabilidade vascular, estimulação da proliferação, migração e diferenciação celular em diferentes tipos celulares, como endotélio, glia e neurônios [2, 53]. Esse fator é superexpresso em cérebros após o evento AVC [2]. Nas amostras brasileiras, foi

possível identificar um desequilíbrio no número de cópias dos genes *ANXA8*, *CDH13*, *HGSNAT*, *NAALADL2* e *RAF1* associados com processos angiogênicos e vias do VEGF.

Anexinas são descritas em etapas de tráfico vesicular e fusões de membranas, de modo que esses genes podem estar relacionados a diversos processos celulares, como angiogênese, inflamação, fusão e reparo de membranas, apoptose, proliferação e diferenciação celular [54-57]. Essa família gênica tem a região C-terminal conservada e uma região N-terminal alterada, o que contribui com as diferenças entre os genes e com a especificidade dos alvos de membrana [54, 55, 58, 59]. A Anexina A8 (*ANXA8*) tem baixa frequência de expressão na maioria dos tecidos e, apesar do seu potencial alvo de fosforilação, sua função ainda não foi caracterizada. *ANXA8* é reconhecida como alta e especificamente expressa em células de pacientes com leucemia aguda promielocítica (APL), doença associada com hemorragias pulmonares e intracranianas [55, 60, 61].

ANXA8 apresenta similaridade com outras anexinas, como *ANXA2*, *ANXA3* e *ANXA5*, cujas funções já foram descritas como relacionadas à anticoagulação, reparo e prevenção de extensão de danos na membrana, fator angiogênico induzido por VEGF, vias de diferenciação celular, angiogênese e fisiologia vascular [55-57, 59, 62, 63]. Deste modo, é possível presumir que a família das anexinas apresenta relevante influência em doenças vasculares.

Caderinas são proteínas responsáveis pela adesão celular e manutenção da integridade tecidual [64]. Caderina13 (*CDH13*) é predominantemente expressa em sistemas nervoso e cardiovascular, agindo em processos angiogênicos, remodelagem da parede vascular, desenvolvimento e adesão celular [64, 65]. Orget al (2009) descreve uma interação entre *CDH13* e determinação da pressão sanguínea, que tende a ser relacionada a fisiologia cardiovascular, baixa densidade de lipoproteínas e adiponectinas [65]. Essas últimas podem modular sensibilidade a insulina, homeostase de glicose, metabolismo lipídico e resposta a aterosclerose e a proteção inflamatória [65, 66].

O gene *HGSNAT* é principalmente relacionado a mucopolissacaridose tipo III, uma doença autossômica recessiva associada

com metabolismo de heparina e sulfato de heparano[67]. Proteoglicano de sulfato de heparano é o principal componente da matrix extracelular, que participa de processos celulares, como angiogênese e adesão celular, e pode facilitar o desenvolvimento através da ligação de citocinas, fatores de crescimento, quimiocinas e outros fatores regulatórios [68, 69]. Heparanase, enzima que cataliza a degradação desses proteoglicanos, é mediadora de metástases tumorais e angiogênese, além de ser descrita com alta expressão em cérebros pós-ischêmicos em camundongos[69]. De acordo com Li et al (2011), o aumento de heparanase tende a estar correlacionada com o aumento de angiopoietina-2 (Ang-2), VEGF e fator de crescimento de fibroblastos (FGF-2), relevantes fatores angiogênicos[69]. Além disso, heparina também contribui com a inibição da coagulação sanguínea pela ativação de trombinas e fator X [2, 70]. Essa inibição contribui com o aumento do risco de sangramento, levando a hemorragias cerebrais[70].

N-acetylated alpha-linked acid dipeptidase-like2 (*NAALADL2*) age como uma matrix metalloprotease e altera o ambiente tumoral, sendo alvo de interesses terapêuticos[71, 72]. *NAALADL2* foi identificado como um promotor mais agressivo do câncer de próstata e colo de útero, devido a sua habilidade de promover metástases e angiogênese[71, 72]. *NAALADL2* também está associado com má formação (Síndrome Lange) e com danos na artéria coronária em crianças, que podem se desenvolver em seqüelas cardiovasculares em adultos (Doença de Kawasaki) [71, 72].

RAF1 é uma quinase serina/treonina, relacionada a diferentes vias e doenças [53, 73]. É comumente ativada por fosforilação, metilação, rearranjo de proteínas e ligações lipídicas [53, 74]. *RAF1* é uma das principais moléculas ativadas na via MAPK, uma via de sinalizadora de regulação comumente descrita para desenvolvimento tumoral e processos induzidos por VEGF [53, 74]. *RAF1* também está associado com a Síndrome Noonan, uma doença caracterizada por baixa estatura, defeitos cardíacos congênitos e de coagulação, atrasos no desenvolvimento, displasias linfáticas, dentre outras [73]. Desse modo, há sugestões que alterações nas interações de *RAF1* com vias VEGF e MAPK podem causar fenótipos patogênicos, incluindo defeitos vasculares.

Outro significativo processo vascular é a coagulação, que mantém forte influência sobre várias doenças, incluindo o AVC hemorrágico [75]. Terapias de anticoagulação previnem a formação de trombinas, e consequente coágulos, no cérebro, levando a manutenção de intensos sangramentos [75]. Os genes *FGL2* e *NSF* foram identificados em nossas coortes e relacionados a esse processo.

FGL2 está presente na via da protrombina e tem importante papel na microtrombose pela ativação da coagulação [76, 77]. É expresso em células endoteliais microvasculares quando induzidas por citocinas pró-inflamatórias [76, 77]. *FGL2* pode ser relacionado a doenças procoagulantes, inflamatórias e imunológicas caracterizadas pela deposição de fibrina intravascular [76, 77]. De acordo com Dinget al (2010), *fgl2* foi associado com microtromboses cardíacas e como possível causa de coagulação sanguínea como resposta inflamatória em ratos com diabetes do tipo II [76].

NSF é altamente associado com reparo, regeneração e recuperação celular após danos vasculares [78]. Este gene também está envolvido no tráfico vesicular e fusão de membranas, agindo como chaperonas, regulando a exocitose [78-80]. Exocitose age na obstrução microvascular e inflamação vascular pela agregação dos leucócitos e plaquetas no endotélio, o que pode estar relacionada a processos isquêmicos [78-80].

Há evidências de que infecções e inflamações podem interagir com fatores genéticos para o desenvolvimento de AVC [81]. Defensinas são proteínas envolvidas nos processos imunológicos, principalmente contra patógenos [81]. Algumas defensinas podem ser encontradas nos capilares cerebrais, protegendo-os contra infecções [81]. Essas proteínas são amplamente identificadas em estudos de CNVs para doenças complexas devido a uma região flanqueadora que providencia suporte para frequentes rearranjos no processo de recombinação [82]. Além disso, a proximidade com a região telomérica do cromossomo tende a facilitar esse rearranjo [83].

Os genes das Defensinas 4 (*DEFB4*) e 103 (*DEFB103*) apresentam padrão hereditário de recombinação, o que sugere que esses genes apresentem um equilíbrio de ligação [83]. Tiszlavicz et al (2011) identificou uma baixa expressão de *DEFB4* em pacientes com AVC isquêmico [81]. Baseado nas análises realizadas pelo METACORE, *DEFB4* foi associada com doenças

vasculares e cardiovasculares, o que reforça a ideia que a resposta imune inata esteja relevantemente agindo na fisiopatologia do AVC [81].

Algumas condições vasculares também podem estar associadas com déficits de desenvolvimento, que podem causar diversos problemas durante a vida. Os tecidos cardíaco e cerebral possuem desenvolvimentos simultâneos durante a fase embrionária, de modo que falhas nesse processo podem gerar doenças cardíacas ou neurovasculares, como o AVC [84]. Genes com funções de desenvolvimento tecidual tendem a influenciar a estrutura desses tecidos, podendo ocasionar bases para doenças.

MACROD2 é descrito durante o desenvolvimento cerebral e tem sido associado com doenças neurológicas, como autismo, esquizofrenia e infartos cerebrais, como sugerido por análises *in silico* [13, 85].

NOMO1 é um membro da família de fatores de crescimento (*TGFB*) expresso durante embriogênese e é associado com desenvolvimento tecidual, principalmente cardiogênese [86, 87]. Esse gene é responsável, também, por padrões de mesoderme e endoderme e especificações dos primeiros eixos no desenvolvimento cardíaco [86, 87].

Há estudos que demonstram que os fatores genéticos podem agir tanto sobre mecanismos de ação de diversas doenças, incluindo complexas, como sobre os fatores de risco que influenciam a ocorrência dessas enfermidades, intensificando seus efeitos sobre elas.

Genes da família do Citocromo P450 agem sobre a síntese do colesterol, esteróides e lipídeos e o metabolismo de drogas. *CYP2E1* é descrito com ação sobre o metabolismo do álcool e nicotina, principais fatores de risco para o AVC [8, 88-91]. Vasiliouet al (2006) descreve o efeito da sensibilidade do álcool no cérebro [91]. Outras famílias do Citocromo P450 já foram associadas com regulação da pressão sanguínea, hipertensão, leucemia, risco de complicações hemorrágicas, aumento do risco de doenças nas artérias coronárias e eventos cardiovasculares, assim como para AVC isquêmico e hemorrágico [11, 88-93].

Outro gene relacionado a síntese do colesterol e ácidos graxos é o *MBTPS1*, cuja família é responsável pela marcação de proteínas para suas ativações [94]. Todos os genes dessa família estão relacionados com diferentes estágios da aterogênese [94]. *MBTPS1* ativa a síntese do colesterol e

a transcrição do LDL, além de ter baixa expressão em placas aterotrombóticas[94].

NPY é um neuropeptídeo presente nos neurônios do sistema nervoso central e periférico e tem sido descrito com associação a doenças neuropsiquiátricas e obesidade, em alguns estudos[95, 96].

O receptor 4 do *NPY* (*NPY4R*) é amplamente frequente e apresenta alta variabilidade entre as espécies [95]. Em roedores, *NPY4R* foi detectado no coração, intestino, estômago, colo e próstata, enquanto que em humanos, está presente no músculo esquelético, coração, pulmão e cérebro [96]. Estudos em camundongos descreveram danos nas funções cardiovasculares, como redução da pressão sanguínea e do batimento cardíaco e respostas vasoconstritoras, na ausência de *NPY4R*[96]. Esses sintomas são semelhantes aos encontrados em pacientes com bulimia ou anorexia nervosas em humanos [96]. De acordo com Smith-White et al (2002), essas doenças envolvem diversos centros cerebrais, o que sugere que deleções no receptor *NPY4R* interfiram nas funções de outros receptores *NPY* em outras regiões cerebrais [96]. Outros estudos relatam a redução sinérgica de peso corporal, e consequente obesidade, em humanos na presença dos receptores 2 e 4 do *NPY*[95]. Esses receptores também já foram descritos para as desordens metabólicas acima[95]. Desse modo, é possível que receptores *NPY* mantenham forte interações no cérebro e nos sistemas cardiovasculares, influenciando seus efeitos sobre doenças.

Em análises no Software METACORE, o gene *NPY4R* foi relacionado com doenças e vias associadas com processos neurovasculares, incluindo AVC, o que intensifica a hipótese que esses receptores exerçam efeito sobre essas condições. Contudo, estudos mais complexos ainda são essenciais para avaliar essa teoria.

Uma forma mais simples de comprovar a interação de genes com determinado fenótipo, seria verificar a presença dessas variantes em outra população com mesmo padrão fenotípico. Com base nesse conceito, foi realizada uma colaboração internacional com um grupo italiano, a fim de validar os resultados encontrados em nossa coorte. A Itália seria uma potencial região de interesse para comparação devido a proximidade genética entre ambas as populações. Além do mais, a coorte utilizada nesse estudo é proveniente de

Joinville/SC, uma região com grande ancestralidade européia. Entretanto, ao realizarmos uma análise de estratificação amostral, essas populações demonstram diferentes estruturas genéticas, que podem estar relacionadas a diferentes regiões de origem dos pacientes e controles, a técnica de extração de DNA, a diferenças nas metodologias utilizadas para microarranjos ou a estrutura genética da população. A comparação entre as coortes foi realizada, porém, como não foi possível distinguir qual a fonte de variabilidade entre as coortes, prosseguiu-se com ressalvas.

Foram identificados 2 genes (*MACROD2* e *NSF*) presentes concomitantemente na coorte italiana e brasileira, após análise desta última pelo METACORE. Esses genes relacionam-se principalmente com vias de coagulação, regeneração vascular e desenvolvimento tecidual de cérebro e coração, potencializando a relevância dessas vias no mecanismo do AVC [13, 78-80,85].

Esses genes foram relatados no banco de dados DECIPHER como relacionados a cardiomiopatia e comprometimento neurológico da fala, respectivamente, o que amplia a probabilidade de alguma interação funcional desses genes com o fenótipo em questão. O gene *MACROD2*, além de apresentar uma descrição para cardiomiopatias no DECIPHER, também não apresenta variantes que englobem toda a sua sequência gênica no banco de dados DGV, de modo que não possua variantes descritas na população geral.

Assim, uma hipótese é que variantes no gene *MACROD2* possam ter efeitos sobre o desenvolvimento cardíaco e neuronal durante embriogênese, influenciando futuros fatores de risco para eventos neurovasculares.

Contudo, devido a complexidade do AVC, não podemos ignorar os outros genes identificados nesse estudo como forte candidatos a apresentarem algum efeito sobre o fenótipo estudado.

4. Conclusões

Neste estudo, apenas 50% dos genes candidatos para AVC isquêmico apresentaram ilhas CpGs na região promotora através de análises *in silico* das regiões. Nos estudos experimentais, essas regiões promotoras não foram amplificadas com sucesso, de modo que outras técnicas devem ser utilizadas para validação dos resultados *in silico*.

Foram observadas alterações no número de cópias em 19 genes relacionados com angiogênese, coagulação, adesão celular, imunologia, metabolismo de drogas e desenvolvimento de tecidos cardíaco e cerebral. Através da validação em população italiana, foi possível identificar 2 genes, que continham CNVs presentes em ambas as coortes: *MACROD2* e *NSF*.

5. Referências

1. Caplan, L.R., *Stroke*. 1 ed, ed. A.J. Sumner. Vol. 1. 2006, New York: Demos Medical Publishing. 255.
2. Mohr, J.P.W., P. A.; Grotta, J. C.; Moskowitz, M. A.; Mayberg, M.; von Kummer, R, *Stroke: Pathophysiology, Diagnosis and Management*. 5 ed. 2011, Philadelphia, PA: Elsevier Saunders.
3. Traylor, M., et al., *Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies*. Lancet Neurol, 2012. **11**(11): p. 951-62.
4. Sharma, P., S. Yadav, and J.F. Meschia, *Genetics of ischaemic stroke*. J Neurol Neurosurg Psychiatry, 2013. **84**(12): p. 1302-8.
5. Markus, H.S., *Stroke genetics: prospects for personalized medicine*. BMC Med, 2012. **10**: p. 113.
6. Adams, H.P., Jr., et al., *Classification of subtype of acute ischemic stroke. Definitions for use in a multicenter clinical trial. TOAST. Trial of Org 10172 in Acute Stroke Treatment*. Stroke, 1993. **24**(1): p. 35-41.
7. Pantoni, L., *Cerebral small vessel disease: from pathogenesis and clinical characteristics to therapeutic challenges*. Lancet Neurol, 2010. **9**(7): p. 689-701.
8. Dichgans, M., *Genetics of ischaemic stroke*. Lancet Neurol, 2007. **6**(2): p. 149-61.
9. Sacco, R.L., et al., *Race-ethnic disparities in the impact of stroke risk factors: the northern Manhattan stroke study*. Stroke, 2001. **32**(8): p. 1725-31.
10. Woo, D., et al., *Meta-analysis of genome-wide association studies identifies 1q22 as a susceptibility locus for intracerebral hemorrhage*. Am J Hum Genet, 2014. **94**(4): p. 511-21.
11. Guo, J.M., A.J. Liu, and D.F. Su, *Genetics of stroke*. Acta Pharmacol Sin, 2010. **31**(9): p. 1055-64.
12. Zeng, P., et al., *Statistical analysis for genome-wide association study*. J Biomed Res, 2015. **29**(4): p. 285-97.

13. Jahanshad, N., et al., *Genome-wide scan of healthy human connectome discovers SPON1 gene variant influencing dementia severity*. Proc Natl Acad Sci U S A, 2013. **110**(12): p. 4768-73.
14. Carty, C.L., et al., *Meta-Analysis of Genome-Wide Association Studies Identifies Genetic Risk Factors for Stroke in African Americans*. Stroke, 2015. **46**(8): p. 2063-8.
15. Yamada, Y., et al., *Genetic risk for ischemic and hemorrhagic stroke*. Arterioscler Thromb Vasc Biol, 2006. **26**(8): p. 1920-5.
16. Wang, Q., et al., *C-reactive protein polymorphisms and genetic susceptibility to ischemic stroke and hemorrhagic stroke in the Chinese Han population*. Acta Pharmacol Sin, 2009. **30**(3): p. 291-8.
17. Frankel, N., *Multiple layers of complexity in cis-regulatory regions of developmental genes*. Dev Dyn, 2012. **241**(12): p. 1857-66.
18. Sun, L., et al., *Pentanucleotide TTTTA repeat polymorphism of apolipoprotein(a) gene and plasma lipoprotein(a) are associated with ischemic and hemorrhagic stroke in Chinese: a multicenter case-control study in China*. Stroke, 2003. **34**(7): p. 1617-22.
19. Strand, M., et al., *Polymorphisms at the osteoprotegerin and interleukin-6 genes in relation to first-ever stroke*. Cerebrovasc Dis, 2007. **24**(5): p. 418-25.
20. Mascalchi, M., et al., *Transthyretin amyloidosis and superficial siderosis of the CNS*. Neurology, 1999. **53**(7): p. 1498-503.
21. Ma, M., et al., *Disease-associated variants in different categories of disease located in distinct regulatory elements*. BMC Genomics, 2015. **16 Suppl 8**: p. S3.
22. Feil, R. and M.F. Fraga, *Epigenetics and the environment: emerging patterns and implications*. Nat Rev Genet, 2011. **13**(2): p. 97-109.
23. Qureshi, I.A. and M.F. Mehler, *Emerging role of epigenetics in stroke: part 1: DNA methylation and chromatin modifications*. Arch Neurol, 2010. **67**(11): p. 1316-22.
24. Krueger, F., et al., *DNA methylome analysis using short bisulfite sequencing data*. Nat Methods, 2012. **9**(2): p. 145-51.

25. Rauch, T. and G.P. Pfeifer, *Methylated-CpG island recovery assay: a new technique for the rapid detection of methylated-CpG islands in cancer*. Lab Invest, 2005. **85**(9): p. 1172-80.
26. Fazzari, M.J. and J.M. Greally, *Epigenomics: beyond CpG islands*. Nat Rev Genet, 2004. **5**(6): p. 446-55.
27. Chen, L., et al., *Genome architecture and its roles in human copy number variation*. Genomics Inform, 2014. **12**(4): p. 136-44.
28. Redon, R., et al., *Global variation in copy number in the human genome*. Nature, 2006. **444**(7118): p. 444-54.
29. Mefford, H.C., *CNVs in Epilepsy*. Curr Genet Med Rep, 2014. **2**: p. 162-167.
30. Bae, J.S., et al., *Genome-wide association analysis of copy number variations in subarachnoid aneurysmal hemorrhage*. J Hum Genet, 2010. **55**(11): p. 726-30.
31. Alkan, C., B.P. Coe, and E.E. Eichler, *Genome structural variation discovery and genotyping*. Nat Rev Genet, 2011. **12**(5): p. 363-76.
32. Stankiewicz, P. and J.R. Lupski, *Structural variation in the human genome and its role in disease*. Annu Rev Med, 2010. **61**: p. 437-55.
33. Bailey, J.A., et al., *Human-specific duplication and mosaic transcripts: the recent paralogous structure of chromosome 22*. Am J Hum Genet, 2002. **70**(1): p. 83-100.
34. Talseth-Palmer, B.A., et al., *Continuing difficulties in interpreting CNV data: lessons from a genome-wide CNV association study of Australian HNPCC/lynch syndrome patients*. BMC Med Genomics, 2013. **6**: p. 10.
35. Vandeweyer, G., et al., *CNV-WebStore: online CNV analysis, storage and interpretation*. BMC Bioinformatics, 2011. **12**: p. 4.
36. Grond-Ginsbach, C., et al., *Copy number variation in patients with cervical artery dissection*. Eur J Hum Genet, 2012. **20**(12): p. 1295-9.
37. Matarin, M., et al., *Structural genomic variation in ischemic stroke*. Neurogenetics, 2008. **9**(2): p. 101-8.
38. Norskov, M.S., et al., *Copy number variation in glutathione S-transferases M1 and T1 and ischemic vascular disease: four studies and meta-analyses*. Circ Cardiovasc Genet, 2011. **4**(4): p. 418-28.

39. Sambrook, J.F., E.F.; Maniatis, T., *Molecular Cloning: A Laboratory Manual*. 2 ed. 1989, Cold Spring Harbor, New York: Cold Spring Harbor Laboratory Press.
40. Pena, S.D., et al., *The genomic ancestry of individuals from different geographical regions of Brazil is more uniform than expected*. PLoS One, 2011. **6**(2): p. e17063.
41. Pritchard, J.K.R., N. A., *Use of unlinked genetic markers to detect population stratification in association studies*. Am J Hum Genet, 1999. **65**: p. 8.
42. Miller, S.A., D.D. Dykes, and H.F. Polesky, *A simple salting out procedure for extracting DNA from human nucleated cells*. Nucleic Acids Res, 1988. **16**(3): p. 1215.
43. Santella, R.M., *Approaches to DNA/RNA Extraction and whole genome amplification*. Cancer Epidemiol Biomarkers Prev, 2006. **15**(9): p. 1585-7.
44. Seshadri, S., et al., *The lifetime risk of stroke: estimates from the Framingham Study*. Stroke, 2006. **37**(2): p. 345-50.
45. Rice, P., I. Longden, and A. Bleasby, *EMBOSS: the European Molecular Biology Open Software Suite*. Trends Genet, 2000. **16**(6): p. 276-7.
46. Li, L.C. and R. Dahiya, *MethPrimer: designing primers for methylation PCRs*. Bioinformatics, 2002. **18**(11): p. 1427-31.
47. Herman, J.G., et al., *Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands*. Proc Natl Acad Sci U S A, 1996. **93**(18): p. 9821-6.
48. Team, R.D.C., *R: A language and environment for statistical computing*. 2011, R Foundation for Statistical Computing: Vienna.
49. Naumova, O.Y., et al., *Gene expression in the human brain: the current state of the study of specificity and spatiotemporal dynamics*. Child Dev, 2013. **84**(1): p. 76-88.
50. Godnic, I., et al., *Genome-wide and species-wide in silico screening for intragenic MicroRNAs in human, mouse and chicken*. PLoS One, 2013. **8**(6): p. e65165.
51. Yu, Z., et al., *Aberrant allele frequencies of the SNPs located in microRNA target sites are potentially associated with human cancers*. Nucleic Acids Res, 2007. **35**(13): p. 4535-41.

52. Chen, J., et al., *A functional variant in the 3'-UTR of angiopoietin-1 might reduce stroke risk by interfering with the binding efficiency of microRNA 211*. Hum Mol Genet, 2010. **19**(12): p. 2524-33.
53. Sulpice, E., et al., *Platelet factor 4 disrupts the intracellular signalling cascade induced by vascular endothelial growth factor by both KDR dependent and independent mechanisms*. Eur J Biochem, 2004. **271**(16): p. 3310-8.
54. Poeter, M., et al., *Annexin A8 controls leukocyte recruitment to activated endothelial cells via cell surface delivery of CD63*. Nat Commun, 2014. **5**: p. 3738.
55. Rety, S., et al., *The crystal structure of annexin A8 is similar to that of annexin A3*. J Mol Biol, 2005. **345**(5): p. 1131-9.
56. Park, J.E., et al., *Annexin A3 is a potential angiogenic mediator*. Biochem Biophys Res Commun, 2005. **337**(4): p. 1283-7.
57. Carmeille, R., et al., *Annexin-A5 promotes membrane resealing in human trophoblasts*. Biochim Biophys Acta, 2015. **1853**(9): p. 2033-44.
58. Junker, H., et al., *Proteomic identification of an upregulated isoform of annexin A3 in the rat brain following reversible cerebral ischemia*. Glia, 2007. **55**(16): p. 1630-7.
59. Goebeler, V., et al., *Annexin A8 displays unique phospholipid and F-actin binding properties*. FEBS Lett, 2006. **580**(10): p. 2430-4.
60. Stein, E., et al., *The coagulopathy of acute promyelocytic leukaemia revisited*. Best Pract Res Clin Haematol, 2009. **22**(1): p. 153-63.
61. Chang, K.S., et al., *Specific expression of the annexin VIII gene in acute promyelocytic leukemia*. Blood, 1992. **79**(7): p. 1802-10.
62. Meadows, S.M. and O. Cleaver, *Annexin A3 Regulates Early Blood Vessel Formation*. PLoS One, 2015. **10**(7): p. e0132580.
63. Flood, E.C. and K.A. Hajjar, *The annexin A2 system and vascular homeostasis*. Vascu Pharmacol, 2011. **54**(3-6): p. 59-67.
64. Hebbard, L.W., et al., *T-cadherin supports angiogenesis and adiponectin association with the vasculature in a mouse mammary tumor model*. Cancer Res, 2008. **68**(5): p. 1407-16.

65. Org, E., et al., *Genome-wide scan identifies CDH13 as a novel susceptibility locus contributing to blood pressure determination in two European populations*. Hum Mol Genet, 2009. **18**(12): p. 2288-96.
66. Chung, C.M., et al., *A genome-wide association study reveals a quantitative trait locus of adiponectin on CDH13 that predicts cardiometabolic outcomes*. Diabetes, 2011. **60**(9): p. 2417-23.
67. Fan, X., et al., *Characterization of the biosynthesis, processing and kinetic mechanism of action of the enzyme deficient in mucopolysaccharidosis IIIC*. PLoS One, 2011. **6**(9): p. e24951.
68. Sarrazin, S., W.C. Lamanna, and J.D. Esko, *Heparan sulfate proteoglycans*. Cold Spring Harb Perspect Biol, 2011. **3**(7).
69. Li, J., et al., *Expression of heparanase in vascular cells and astrocytes of the mouse brain after focal cerebral ischemia*. Brain Res, 2012. **1433**: p. 137-44.
70. Pham, M., et al., *Blood coagulation factor XII--a neglected player in stroke pathophysiology*. J Mol Med (Berl), 2012. **90**(2): p. 119-26.
71. Whitaker, H.C., et al., *N-acetyl-L-aspartyl-L-glutamate peptidase-like 2 is overexpressed in cancer and promotes a pro-migratory and pro-metastatic phenotype*. Oncogene, 2014. **33**(45): p. 5274-87.
72. Burgner, D., et al., *A genome-wide association study identifies novel and functionally related susceptibility Loci for Kawasaki disease*. PLoS Genet, 2009. **5**(1): p. e1000319.
73. Allanson, J.E. and A.E. Roberts, *Noonan Syndrome*, in *GeneReviews(R)*, R.A. Pagon, et al., Editors. 1993: Seattle (WA).
74. Houslay, M.D. and W. Kolch, *Cell-type specific integration of cross-talk between extracellular signal-regulated kinase and cAMP signaling*. Mol Pharmacol, 2000. **58**(4): p. 659-68.
75. Emiru, T., et al., *Intracerebral hemorrhage: a review of coagulation function*. Clin Appl Thromb Hemost, 2013. **19**(6): p. 652-62.
76. Ding, Y., et al., *Expression and significance of fgl2 prothrombinase in cardiac microvascular endothelial cells of rats with type 2 diabetes*. J Huazhong Univ Sci Technolog Med Sci, 2010. **30**(5): p. 575-81.
77. Zhang, P., et al., *flg2 as a potential biomarker of acute cerebral ischemic-reperfusion injury*. Microvascular Research, 2015. **99**: p. 36-42.

78. Liu, C. and B. Hu, *Alterations of N-ethylmaleimide-sensitive atpase following transient cerebral ischemia*. Neuroscience, 2004. **128**(4): p. 767-74.
79. Calvert, J.W., et al., *Inhibition of N-ethylmaleimide-sensitive factor protects against myocardial ischemia/reperfusion injury*. Circ Res, 2007. **101**(12): p. 1247-54.
80. Yamakuchi, M., et al., *Exocytosis of endothelial cells is regulated by N-ethylmaleimide-sensitive factor*. Methods Mol Biol, 2008. **440**: p. 203-15.
81. Tiszlavicz, Z., et al., *Genetic polymorphisms of human beta-defensins in patients with ischemic stroke*. Acta Neurol Scand, 2012. **126**(2): p. 109-15.
82. Cantsilieris, S. and S.J. White, *Correlating multiallelic copy number polymorphisms with disease susceptibility*. Hum Mutat, 2013. **34**(1): p. 1-13.
83. Linzmeier, R.M. and T. Ganz, *Human defensin gene copy number polymorphisms: comprehensive analysis of independent variation in alpha- and beta-defensin regions at 8p22-p23*. Genomics, 2005. **86**(4): p. 423-30.
84. McQuillen, P.S. and S.P. Miller, *Congenital heart disease and brain development*. Ann N Y Acad Sci, 2010. **1184**: p. 68-86.
85. Kohanim, O., et al., *Discovery and Replication of Gene Influences on Brain Structure Using LASSO Regression*. Front Neurosci, 2012. **6**: p. 115.
86. Gao, W.L., et al., *The imprinted H19 gene regulates human placental trophoblast cell proliferation via encoding miR-675 that targets Nodal Modulator 1 (NOMO1)*. RNA Biol, 2012. **9**(7): p. 1002-10.
87. Zhang, H., et al., *Silencing of nodal modulator 1 inhibits the differentiation of P19 cells into cardiomyocytes*. Exp Cell Res, 2015. **331**(2): p. 369-76.
88. Choi, J.Y., et al., *CYP2E1 and NQO1 genotypes, smoking and bladder cancer*. Pharmacogenetics, 2003. **13**(6): p. 349-55.
89. Howard, L.A., et al., *CYP2E1*1D regulatory polymorphism: association with alcohol and nicotine dependence*. Pharmacogenetics, 2003. **13**(6): p. 321-8.

90. Kim, S.K., S.V. Yim, and B.C. Lee, *Association between cytochrome P450 promoter polymorphisms and ischemic stroke*. *Exp Ther Med*, 2012. **3**(2): p. 261-268.
91. Vasiliou, V., et al., *CYP2E1 and catalase influence ethanol sensitivity in the central nervous system*. *Pharmacogenet Genomics*, 2006. **16**(1): p. 51-8.
92. Ding, H., et al., *Association of common variants of CYP4A11 and CYP4F2 with stroke in the Han Chinese population*. *Pharmacogenet Genomics*, 2010. **20**(3): p. 187-94.
93. Kitzmiller, J.P., et al., *CYP3A4*22 and CYP3A5*3 are associated with increased levels of plasma simvastatin concentrations in the cholesterol and pharmacogenetics study cohort*. *Pharmacogenet Genomics*, 2014. **24**(10): p. 486-91.
94. Turpeinen, H., et al., *Proprotein convertases in human atherosclerotic plaques: the overexpression of *FURIN* and its substrate cytokines *BAFF* and *APRIL**. *Atherosclerosis*, 2011. **219**(2): p. 799-806.
95. Lee, N.J., et al., *Synergistic attenuation of obesity by Y2- and Y4-receptor double knockout in ob/ob mice*. *Nutrition*, 2008. **24**(9): p. 892-9.
96. Smith-White, M.A., H. Herzog, and E.K. Potter, *Cardiac function in neuropeptide Y Y4 receptor-knockout mice*. *Regul Pept*, 2002. **110**(1): p. 47-54.

Anexos

1. Parecer do Comitê de Ética em Pesquisa – UNICAMP

FACULDADE DE CIÊNCIAS
MÉDICAS - UNICAMP
(CAMPUS CAMPINAS)



PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: BIORREPOSITÓRIO
ESTUDOS DE GENÉTICA MOLECULAR EM DOENÇAS NEUROPSIQUIÁTRICAS
FASE I

Pesquisador: Iscia Teresinha Lopes Cendes

Área Temática: Área 1. Genética Humana.
(Trata-se de pesquisa envolvendo genética humana não contemplada acima.);

Versão: 2

CAAE: 12112913.3.0000.5404

Instituição Proponente: Hospital de Clínicas da UNICAMP

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 257.020

Data da Relatoria: 12/04/2013

Apresentação do Projeto:

Trata-se de um projeto para implantação de biorepositório de doenças neuro-psiquiátricas e casos-controle. O estudo prevê recrutamento e coleta de 700 pacientes e 300 indivíduos controle.

O presente projeto pretende estudar os aspectos moleculares das seguintes doenças: epilepsias, malformações do desenvolvimento cortical, coreias, ataxias, paraparesias espásticas, distonias, transtorno afetivo bipolar, esquizofrenia, doenças musculares, doenças mitocondriais, doença de Parkinson, acidente vascular cerebral e demências. O projeto está dividido em sub-projetos, com a descrição detalhada das estratégias que serão utilizadas para cada uma dessas doenças.

Serão utilizadas diversas técnicas de biologia molecular para identificação de mutações conhecidas ou novas, como PCR, sequenciamento convencional e de terceira geração e análises de bioinformática. No Subprojeto 1 (Epilepsias e Malformações do Desenvolvimento Cortical), serão avaliadas mutações através da implantação da tecnologia de sequenciamento, baseada em equipamento de terceira geração e um novo sistema de detecção eletroônico, sistema Ion Torrent®. No subprojeto 2 será realizada a captura e o sequenciamento do exoma em amostras de DNA de

Endereço: Rua Tessália Vieira de Camargo, 126

Bairro: Barão Geraldo

CEP: 13.063-887

UF: SP

Município: CAMPINAS

Telefone: (19)3521-8936

Fax: (19)3521-7187

E-mail: cep@fcm.unicamp.br

FACULDADE DE CIÊNCIAS
MÉDICAS - UNICAMP
(CAMPUS CAMPINAS)



indivíduos afetados por formas familiares de epilepsia mioclônica juvenil (EMJ), epilepsia do lobo temporal (ELT) e benigna rolaúndica epilepsia (BRE) e Malformações do desenvolvimento cortical (MCD) nas quais as mutações já descritas não foram identificadas. Nessa investigação, serão utilizados kits de captura e enriquecimento Illumina TruSeq®, além do sequenciamento 11 usando a Illumina Hi-seq 2.000. Após o sequenciamento, será realizada análise de bioinformática nos dados obtidos a fim de identificar variantes potencialmente patogênicas.

Os pacientes serão recrutados nos ambulatórios do HC-UNICAMP (Neurogenética, Epilepsia, Epilepsia de difícil controle, Epilepsia infantil, Distúrbio do Movimento, Doenças Neuromusculares, Distonias, Toxina Botulínica, Demeúncias, Neurovascular ou Psiquiatria). Será coletados de 20-30 ml de sangue periférico para o estudo, além do exame clínico e anamnese. Será coletados dados do prontuário médico.

Objetivo da Pesquisa:

Identificar as mutações responsáveis pelos respectivos fenótipos, em um grupo de doenças neuropsiquiátricas, incluindo: epilepsias, malformações do desenvolvimento cortical, coreias, ataxias, paraparesias espásticas, distonias, transtorno afetivo bipolar, esquizofrenia, doenças musculares, doenças mitocondriais, acidente vascular cerebral, doença de Parkinson e demências.

Avaliação dos Riscos e Benefícios:

Os procedimentos a serem realizados são de pequeno risco para o indivíduo, sendo o principal deles a coleta de sangue por punção venosa.

Como benefícios, em alguns casos será possível a identificação das mutações associadas às doenças dos pacientes incluídos no estudo, o que implica na possibilidade de algoritmos mais eficientes de diagnóstico e tratamento. Os autores alertam que, muitas, a da mutação não resulta uma mudança no tratamento, o que é compreensível quando se trata de estudos genéticos.

Informações geradas durante o projeto e que possam ter implicações na confirmação diagnóstica de indivíduos sintomáticos serão comunicadas aos profissionais responsáveis pelo acompanhamento destes pacientes, nas sessões de orientação e aconselhamento genético dos ambulatórios de Neurogenética, cuja pesquisadora principal é a responsável.

Comentários e Considerações sobre a Pesquisa:

O projeto está bem escrito e detalhado, assim como os procedimentos que serão aplicados aos voluntários saudáveis e portadores de doença. apresenta relevância científica e social, a médio e longo prazos. O estudo será patrocinado pelo próprio pesquisador. Foram acrescentadas ao projeto principal as informações sobre o local onde serão recrutados e onde serão coletados os exames do grupo controle.

Considerações sobre os Termos de apresentação obrigatória:

O TCLE está bem redigido, é de fácil entendimento para o voluntário e prevê armazenamento de

Endereço: Rua Tessália Vieira de Camargo, 126
Bairro: Barão Geraldo CEP: 13.083-887
UF: SP Município: CAMPINAS
Telefone: (19)3521-8936 Fax: (19)3521-7187 E-mail: cep@fcm.unicamp.br

**FACULDADE DE CIENCIAS
MEDICAS - UNICAMP
(CAMPUS CAMPINAS)**



material biológico. Estão presentes todos os termos de apresentação obrigatória previstos pela Resolução 196/96 e complementares, assim como o "Regulamento do Biorepositório de Doenças Neuropsiquiátricas".

Recomendações:

Nada a declarar.

Conclusões ou Pendências e Lista de Inadequações:

Foram acrescentadas ao projeto principal as informações sobre o recrutamento dos voluntários do grupo controle, com priorização inicial para membros da família de pacientes, porém não portadores das doenças. As amostras serão coletadas nos ambulatórios de Genética e Neurologia HC-Unicamp e Hemocentro-UNICAMP.

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Considerações Finais a critério do CEP:

Aprovado em reunião do colegiado, em 23 de Abril de 2013.

CAMPINAS, 26 de Abril de 2013

Assinador por:

**Fátima Aparecida Bottcher Luiz
(Coordenador)**

Endereço: Rua Tessália Vieira de Camargo, 126

Bairro: Barão Geraldo

CEP: 13.083-887

UF: SP **Município:** CAMPINAS

Telefone: (19)3521-8936

Fax: (19)3521-7187

E-mail: cep@fcm.unicamp.br