

**Universidade de Brasília**

**Faculdade de Medicina**

**Programa de Pós-Graduação em Patologia Molecular**

**Variação estrutural no número de cópias e sua implicação na  
expressão de microRNA em humanos**

**Tulio Cesar de Lima Lins**

**2014**

**Tulio Cesar de Lima Lins**

**Varição estrutural no número de cópias e sua implicação na  
expressão de microRNA em humanos**

Tese apresentada ao Programa de Pós-Graduação *Strictu Sensu* em Patologia Molecular da Universidade de Brasília, como requisito para obtenção do Título de Doutor em Patologia Molecular.

Orientador: Prof. Dr. Rinaldo Wellerson Pereira

Brasília

2014

Ficha catalográfica elaborada pela Biblioteca Central da Universidade de  
Brasília. Acervo 1015152.

L759v Lins, Tulio Cesar de Lima.  
Variação estrutural no número de cópias e sua implicação  
na expressão de microRNA em humanos / Tulio Cesar de Lima Lins.  
-- 2014.  
96 f. : il. ; 30 cm.

Tese (doutorado) - Universidade de Brasília, Faculdade  
de Medicina, Programa de Pós-Graduação em Patologia  
Molecular, 2014.

Inclui bibliografia.

Orientação: Rinaldo Wellerson Pereira.

1. Polimorfismo (Genética). 2. Ácido ribonucléico.  
I. Pereira, Rinaldo Wellerson. II. Título.

CDU 575.174.015.3

## **TERMO DE APROVAÇÃO**

Tese apresentada publicamente como requisito para obtenção do Título de Doutor em Patologia Molecular (Genética), defendida em 27 de Fevereiro de 2014 e aprovada pela banca examinadora constituída por:

---

Orientador: Prof. Dr Rinaldo Wellerson Pereira  
Universidade Católica de Brasília e Universidade de Brasília

---

Examinador: Prof. Dr<sup>a</sup>. Aline Pic-Taylor  
Universidade de Brasília

---

Examinador externo: Prof. Dr. Aparecido Divino da Cruz  
Pontifícia Universidade Católica de Goiás

---

Examinador externo: Prof. Dr<sup>a</sup>. Juliana Forte Mazzeu de Araújo  
Universidade Católica de Brasília

---

Examinador externo: Prof. Dr. Robert Pogue  
Universidade Católica de Brasília

---

Suplente: Prof. Dr<sup>a</sup>. Rosangela Vieira de Andrade  
Universidade Católica de Brasília

## **Dedico**

A todos aqueles que me instigaram o espírito de cientista, entre família, amigos,  
profissionais de saúde e professores de todos os níveis.

## **AGRADECIMENTOS**

Primeiramente ao orientador Prof. Dr. Rinaldo Wellerson Pereira, pela parceria, amizade e incentivo, por idealizar, encorajar e acreditar na realização deste projeto;

Aos professores Doutores Aline Pic-Taylor, Aparecido Divino da Cruz (Peixoto), Juliana Forte Mazzeu de Araújo, Robert Pogue, por participarem da banca e avaliarem criticamente esta Tese.

Aos colegas e professores do Programa de Pós-Graduação em Ciências Genômicas e Biotecnologia da Universidade Católica de Brasília, pela amizade, ajuda e parceria; dentre muitos contribuíram em suma importância para a realização desse trabalho: Prof. Dra. Rosângela Vieira de Andrade, Dianny Jimenez, Priscilla Orosco, Nina Jardim, Tainá Macherini, Getulio Oliveira Jr, Clarissa Gomes, Felipe Marques, Marcela Chiabai, Tatiana Marçola, Laureane Pavanelli, Nathalia Gurgel, Larissa Lemos, entre outros não esquecidos.

A todos os professores, funcionários e demais alunos do Programa de Pós-Graduação em Patologia Molecular da UnB e de outros departamentos pelo apoio, incentivo, ensinamentos, ajuda e parceria;

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior do Ministério da Educação (CAPES/MEC) pelo fornecimento da bolsa de Doutorado;

Aos meus pais Tales e Graça, meus irmãos Tales e Taís, demais familiares e a todos os meus queridos amigos pelo acolhimento e carinho quando foi necessário e pelo incentivo em seguir acreditando nos meus sonhos;

Aos demais não mencionados, porém não esquecidos, que em algum momento da minha vida cruzaram meu caminho, sem vocês eu não estaria aqui!

"A vida é como andar de bicicleta. Para manter o equilíbrio, você deve se manter em movimento."

**- Albert Einstein**

## RESUMO

Variações no número de cópias (do inglês *Copy Number Variation* - CNV) são segmentos iguais a ou maiores que 1 kilobase que apresentam variações estruturais como deleções e duplicações. Constituem uma fração de variação genética importante, que interagem, modulam ou direcionam a expressão gênica. MicroRNA são pequenos RNA de aproximadamente 20 nucleotídeos que regulam a expressão gênica pós-transcricional. O objetivo desta Tese foi analisar a expressão de microRNA com relação às variações estruturais do número de cópias dos respectivos genes em humanos, as regiões de CNV-miRNA. Foram selecionadas 11 regiões de provável polimorfismo no número de cópias além de duas regiões invariáveis como controle positivo, que foram avaliadas por triagem molecular em PCR quantitativa em uma amostra de 103 indivíduos da população brasileira. Em seguida, uma segunda amostragem de 30 indivíduos tiveram DNA e RNA coletados de células mononucleadas do sangue (PBMC) e de epitélio bucal e RNA total do plasma para respectiva quantificação da expressão desses microRNA. Para elucidar questões relativas à dosagem gênica, a expressão de microRNA foi comparada com o número de cópias das respectivas regiões gênicas. Das regiões que apresentaram polimorfismo (CNV-miR-570, -miR-499, -miR-126, -miR-150, -miR-338, -miR-1275), apenas uma não teve o microRNA quantificado (CNV-miR-499) por não ter apresentado amplificação nos tecidos e também no plasma. Os microRNA miR-570 e miR-338, tiveram uma expressão relativa crescente de acordo com o número de cópias para frações de PBMC (e plasma somente para o miR-570), no entanto, estatisticamente não significativas. Numa simulação populacional pelo método Monte Carlo ( $n = 1.000$ ), foi identificada diferença significativa para o miR-570, com taxa de mudança progressiva relativa entre ao número de cópias. A hipótese de que o número de cópias de uma região contendo genes miRNA pode afetar a regulação da expressão de genes abre uma perspectiva de novas pesquisas envolvendo a variabilidade fenotípica.

**Palavras-chave:** Polimorfismo populacional, CNV, microRNA, CNV-miRNA, dosagem gênica

## ABSTRACT

Copy number variation (CNV) are DNA segments of 1 kilobase or larger, showing structural variations, such as deletions, insertions or duplications. It constitutes a significant fraction of genetic variation that directly interacts or modulates gene expression. MicroRNA is a class of small RNA of about 20 nucleotides that regulates post-transcriptional gene expression. The aim of this Thesis was to analyze the expression of microRNA regarding the number of copies of their respective genes in humans, the CNV-miRNA regions. Eleven regions of probable polymorphisms in addition to two invariable control regions were selected and evaluated by quantitative PCR molecular screening in a sample of 103 individuals of the Brazilian population. Then, a second sample of 30 subjects had DNA and RNA collected from blood mononuclear cells (PBMC), oral epithelium and total RNA from plasma, quantified the expression of those microRNAs and compared with the number of copies, aiming to elucidate the relative response to gene dosage. From the regions that showed polymorphism (CNV-miR-570, -miR-499, -miR-126, miR-150, -miR-338, -miR-1275) only one was unable to have the microRNA quantified (miR-499) for failing to provide amplification in tissues and plasma. The microRNA miR-570 and miR-338 had an increased relative expression according to the respective number of copies for PBMC fraction (and plasma only for miR-570), though not statistically significant. In a population simulation by Monte Carlo method ( $n = 1,000$ ), a significant difference was identified to miR-570, with progressive rate of relative change between the categories of number of copies. The hypothesis that the copy number variation of a region containing miRNA genes can affect the regulation of gene expression opens a new perspective for research regarding phenotypic variability.

**Key-words:** CNV, population polymorphism, microRNA, gene dosage, CNV-miRNA

## SUMÁRIO

LISTA DE TABELAS.....	1
LISTA DE FIGURAS.....	3
LISTA DE ABREVIATURAS.....	4
INTRODUÇÃO .....	6
HIPÓTESES.....	28
OBJETIVOS .....	29
MATERIAL E MÉTODOS.....	30
RESULTADOS.....	41
DISCUSSÃO.....	65
CONCLUSÕES.....	75
REFERÊNCIAS .....	76
ANEXO I .....	92

## LISTA DE TABELAS

- Tabela 1:** Descrição de 11 microRNA selecionados para o estudo com CNV putativa, a partir do estudo de Marcinkowska et al. (2011). Orientação em relação ao DNA referência; Número de entradas no banco de dados DGV e estudos que depositaram essas entradas, com observações (número de indivíduos) totais da variação com relação a amostra total dos estudos e respectiva frequência relativa. Fonte: o autor.....42
- Tabela 2:** Descrição dos ensaios de microRNA e CNV pelo Genoma de referência NCBI build 37. \* Loco escolhido como controle diploide. † Loco escolhido como controle diploide e como controle endógeno de reação de microRNA. Fonte: o autor.....43
- Tabela 3:** Distribuição da frequência absoluta dos ensaios de miR-CNV de acordo com o número de cópias genômica em 103 indivíduos da população brasileira. Fonte: O autor.....45
- Tabela 4:** Distribuição da frequência relativa dos ensaios de miR-CNV de acordo com o número de cópias genômica em 103 indivíduos da população brasileira. Fonte: o autor.....45
- Tabela 5:** Número absoluto de indivíduos por variação do número cópia das regiões CNV-miR-570 e CNV-miR-499 classificados por sexo. Nível de significância em  $\alpha = 0.05$  para o teste Chi-quadrado: \*  $p = 0.001$ ; †  $p = 0.004$ . Fonte: o autor. ....47
- Tabela 6:** Número absoluto de indivíduos por variação do número cópia classificados por das regiões CNV-miR-570 e CNV-miR-499 Região geográfica. Nível. Nível de significância em  $\alpha = 0.05$  para o teste Chi-quadrado: \*  $p = 0,352$ ; †  $p = 0,001$ . Fonte: o autor.....47
- Tabela 7:** Estimativas de proporção de ancestralidade na amostra da população brasileira (média  $\pm$  desvio padrão) diferenciadas por variação do número de cópia das regiões CNV-miR-570 e CNV-miR-499. Proporções para ancestralidade Europeia (EUR), Africana (AFR) e Nativo-Americana (NAM). Teste ANOVA com significância a 5%. Fonte: o autor. ....47
- Tabela 8:** Comparação das frequências absolutas do CNV-miR-499 entre regiões geográficas da população brasileira de acordo com a média da ancestralidade. Teste de ANOVA multifatorial com significância de 5%,  $p = 0,113$  para EUR;  $p = 0,081$  para AFR e  $p = 0,088$  para NAM. Fonte: o autor.48
- Tabela 9:** Comparação das frequências relativas populacionais das CNV-miRNA entre as amostras do banco de dados DGV (n total = 15.156, máximo e mínimo referente às populações de cada estudo) e a amostra da população brasileira. Fonte: o autor.....49
- Tabela 10:** Comparação das frequências absolutas na amostra de trabalho em 9 regiões de CNV-miR e entre os tecidos do epitélio bucal e PBMC. ....50

<b>Tabela 11:</b> Comparação do genótipo individual por número de cópias em cada amostra. Relação somente das regiões identificadas com diferenças no número de cópias entre células bucais e PBMC. ....	51
<b>Tabela 12:</b> Valores da quantificação relativa dos microRNA entre o epitélio bucal, plasma e PBMC, de acordo com a variação em número de cópias. Fonte: o autor. ....	56
<b>Tabela 13:</b> Teste de Kruskal-Wallis (com Qui-quadrado representando o parâmetro <i>H-statistic</i> ), para comparação de médias da taxa de mudança (FC) dos microRNA de acordo com a variação em número de cópias do epitélio bucal e PBMC. Significância estatística a 5%.....	58
<b>Tabela 13:</b> Dados da simulação Monte Carlo para os microRNA de acordo com a frequência populacional da variação do número de cópias estimada nos 103 indivíduos da população brasileira e expressão diferencial ( $\log_2(2^{-\Delta Ct})$ ) no plasma e PBMC. Teste de Kruskal-Wallis com significância a 5%. * <i>p</i> -valor para PBMC; † <i>p</i> -valor para Plasma.....	62
<b>Tabela 14:</b> Equações previstas para expressão relativa de microRNA ( <i>y</i> ) em função do número de cópias ( <i>x</i> ) para os modelos de simulação em cada situação, com respectivo valor de ajuste ( $R^2$ ) em regressão.....	64

## LISTA DE FIGURAS

- Figura 1:** Tipos de Genes encontrados em regiões de variação do número de cópias classificados de acordo com o ponto de quebra da CNV. Caixas coloridas representam regiões gênicas de *exons* e linhas *introns*. Fonte: adaptado de Woodwark e Bateman (2011). .....16
- Figura 2:** Esquema da via de biogênese e função dos microRNA. Fonte: adaptado de Melo e Melo (2014).....21
- Figura 3:** Representação esquemática da resposta de dosagem gênica ao fenótipo. (a) Relação linear; (b) Retorno decrescente; (c) Efeito estequiométrico; (d) Proteínas com característica de agregação Fonte: Adaptado de Conrad e Antonarakis (2007). .....25
- Figura 4:** Validação da amostra calibrador com cópia diploide para 9 ensaios de CNV, usando uma amostra aleatória (COSP27) e o DNA de referência em dupla análise (DNAT e TDNA). Fonte: O autor. ....44
- Figura 5:** Boxplot dos níveis de cada microRNA por constituinte biológico (PBMC, Plasma e Células do epitélio bucal) e do total agrupado. Linha em negrito da caixa = mediana dos Ct; parte inferior da caixa = primeiro quartil (Q1); parte superior da caixa = terceiro quartil (Q3); barras = limites inferior e superior. Fonte: o autor. ....53
- Figura 6:** Mapa de expressão relativa dos microRNA. Fonte: O autor .....54
- Figura 7:** Expressão relativa entre perda e ganho, com relação ao diploide, dos microRNA: (A) miR-126, (B) miR-570 e (C) miR-338, de acordo o constituinte biológico. Em (A) apenas a comparação de perda e ganho entre células do epitélio bucal, pois PBMC não houve Fonte: O autor.....57
- Figura 8:** Padrão de expressão relativa entre as diferentes classes de número de cópia dos microRNA: (A) miR-570 e (B) miR-338, de acordo o constituinte biológico. Fonte: O autor.....59
- Figura 9:** Distribuição da expressão diferencial ( $\log_2(2^{-\Delta Ct})$ ) gerada por simulação Monte Carlo (n = 1.000) entre as diferentes classes de número de cópia dos microRNA miR-570 (Plasma e PBMC) e miR-338 (PBMC). Fonte: O autor. ....61
- Figura 10:** Modelo simulado para expressão relativa entre as diferentes classes de número de cópia dos microRNA: (A) miR-570 plasma, (B) miR-570 PBMC e (C) miR-126 PBMC. Fonte: O autor.....63

## LISTA DE ABREVIATURAS

DGV – Database of Genomic Variants

CGH – Hibridização Genômica Comparativa

FISH – Hibridização Fluorescente in Situ

BAC - Cromossomos Artificiais De Bactéria

PCR – Reação em Cadeia Da Polimerase

qPCR – Reação Quantitativa em Cadeia Da Polimerase

Ct – Cycle threshold

DNA – Ácido Desoxiribonucleico

RNA – Ácido Ribonucleico

mRNA – RNA mensageiro

ncRNA – RNA não codificadores

rRNA – RNA ribossomal

tRNA – RNA transportador

snRNA – pequenos RNA nucleares

snoRNA – pequenos RNAs nucleolares

miRNA, miR = microRNA

RISC - complexo de silenciamento induzido por RNA

SNP – polimorfismos de única base

LCR – cópia de baixa repetição

CNV – variação (ões) no número de cópias

CNP - polimorfismos no número de cópias

CN - número de cópia

EUR - ancestralidade Europeia

AFR - ancestralidade Africana

NAM - ancestralidade Nativo-Americana

PBMC - células mononucleadas do sangue

FC - taxa de mudança (do inglês *fold change*)

## INTRODUÇÃO

Variações nas sequências do genoma são um dos principais componentes de investigação científica dos fenótipos comuns e patológicos para desvendar os mecanismos das suas origens e desenvolvimento. Na natureza, os fenótipos são complexos na sua etiologia, resultante da interação de componentes genéticos e fatores ambientais. A variação no genoma humano sempre foi alvo dessas investigações, desde os primeiros estudos de ploidia e alterações cariotípicas microscopicamente visíveis, passando pela premissa clássica do dogma central da biologia, chegando até modelos atuais mais complexos, envolvendo controle da expressão gênica por mecanismos genéticos e epigenéticos (LANDER, 2011). Padrões de variação no genoma podem ajudar a identificar as regiões que possuem associações com características biológicas e que, de alguma maneira, atuam no controle desses eventos em populações humanas (LANDER, 2011). O volume e a qualidade de dados gerados pelo sequenciamento massivo de genomas ajudou a ciência na busca pelas explicações de alguns desses mecanismos e suas redes de integração, de forma abrangente, imparcial e livre de hipótese (KELL; OLIVER, 2004).

Desde os primeiros estudos de citogenética, reconheceu-se que perdas ou ganhos de cromossomos inteiros, as aneuploidias, ou alterações estruturais microscopicamente visíveis de fragmentos cromossômicos como as deleções, inserções ou translocações são responsáveis por diversas síndromes genéticas comuns e raras, além de doenças de desenvolvimento intelectual e cognitivo, não associadas a síndromes reconhecidas (LEE, C.; SCHERER, 2010; SHAW et al., 2004). As primeiras descrições de alterações cromossômicas

mais importantes para a medicina foram feitas pelo cientista Francês Lejeune e sua equipe, descrevendo a trissomia do cromossomo 21 como responsável pela Síndrome de Down (LEJEUNE et al., 1959) e a primeira descrição de Síndrome herdada por deleção, a Síndrome Cri-du-chat, envolvendo a deleção parcial do braço curto do cromossomo 5 (LEJEUNE et al., 1963). Uma variação estrutural microscopicamente visível, relacionada a patologias não sindrômicas, descrita a cerca da mesma época, foi a translocação cromossômica recíproca entre os braços longos dos cromossomos 9 e 22,  $t(9;22)$ , cujo produto da fusão dos genes *BCR* e *ABL*, a proteína quimérica BCR-ABL, está presente em diversos casos de leucemia (NOWELL; HUNGERFORD, 1960).

As variações cromossômicas podem ser detectadas e analisadas por metodologias que abrangem todo o genoma (genoma-completo), bem como serem restritas a uma região específica (LEE, J. H.; JEON, 2008). Métodos de investigação usados nas descobertas de alterações cromossômicas ligadas a síndromes genéticas se basearam em análises citogenéticas de genoma-completo, através do cariótipo obtido por coloração. A coloração é uma técnica baseada em princípios químicos usada para originar contraste entre diferentes componentes celulares, entre eles os cromossomos, permitindo sua visualização (DOLAN, 2011). Ao longo do tempo, diversas técnicas de bandeamento cromossômico surgiram, trazendo com elas o aumento da resolução no número de bandas. O bandeamento G, que utiliza o corante Giemsa, é uma das técnicas de rotina mais utilizadas devido à sua simplicidade, sensibilidade e estabilidade. Sua resolução pode gerar de 300 até 850 bandas por genoma haploide, sendo maior a resolução aumenta-se a probabilidade de detecção de alterações cromossômicas de menores tamanhos (DOLAN, 2011).

Outras técnicas de fluorescência química que utilizam sondas específicas foram desenvolvidas para buscar melhor definição e ampliar as precisões metodológicas, como a Hibridização Genômica Comparativa (CGH do inglês *Comparative Genomic Hybridization*), hibridização *in situ* com fluorescência multicolor, e cariotipagem espectral, entre outros métodos de citogenética molecular (MICALÉ; MOHAMED, 2012). Concomitantemente, a hibridização fluorescente foi amplamente utilizada para avaliar a deleção ou duplicação de regiões específicas ou para mapear genes a seus respectivos cromossomos. Como exemplo, a técnica de Hibridização Fluorescente *in Situ* (FISH – do inglês *fluorescence in situ hybridization*) foi desenvolvida e é constantemente utilizada em mapeamentos e validação de regiões com alterações de número de cópias (MICALÉ; MOHAMED, 2012). Métodos com alvos específicos são frequentemente usados na clínica médica para ajudar no diagnóstico de síndromes, pois usando sondas específicas para as regiões de interesse detectam microdeleções ou microduplicações não visualizadas em técnicas de bandeamento (BALLIF et al., 2008; STOFANKO et al., 2013; WEISS et al., 2008).

A reação quantitativa em cadeia da polimerase (qPCR – do inglês *Quantitative Polymerase Chain Reaction*) é outro método que utiliza marcadores fluorescentes em conjunto com a leitura ótica computadorizada em tempo real da reação (D'HAENE et al., 2010). É amplamente usada para quantificar ácidos nucleicos (DNA/RNA) pela amplificação da sequência alvo, permitindo uma maior sensibilidade de detecção. De maneira geral, a qPCR em tempo real é utilizada em um número crescente de aplicações de pesquisa, incluindo desde a discriminação alélica e análise de polimorfismos de única base (SNP do inglês *Single Nucleotide Polymorphisms*), validação de deleção ou duplicação de regiões específicas, testes de detecção de organismos geneticamente modificados, monitoramento e

detecção da carga viral e de outros patógenos (D'HAENE et al., 2010; VANGUILDER et al., 2008). Além disso, é também utilizada extensivamente para quantificação de RNA de diversas classes, por meio de reações de PCR quantitativa por transcrição reversa, ou qRT-PCR (RT – do inglês *Reverse Transcription*), no qual inicialmente o RNA extraído passa por uma reação de transcrição reversa para a quantificação de expressão gênica ou expressão de RNA não codificantes (D'HAENE et al., 2010; VANGUILDER et al., 2008). Por toda essa versatilidade, a qPCR é considerada uma técnica padrão ouro em análises de quantificação de ácidos nucleicos (D'HAENE et al., 2010; VANGUILDER et al., 2008).

No método da qPCR, a quantidade de produto na reação deve, em teoria, duplicar a cada ciclo, resultando em uma amplificação exponencial da região alvo. O fragmento amplificado é monitorado em tempo real, por meio da intensidade de fluorescência emitida, usando uma molécula chamada de repórter. Vários produtos químicos de detecção estão disponíveis comercialmente para medir a quantificação de produto, incluindo as sondas de hidrólise (por exemplo, sondas TaqMan®), ou moléculas de ligação que emitem fluorescência quando ligadas a cadeia dupla do DNA em regiões não-específicas, como o corante SYBR Green® (VANGUILDER et al., 2008).

De maneira geral, nas reações de qPCR, a ausência de uma sequência alvo específica na reação não permite a emissão de fluorescência ou pela presença de uma molécula supressora do sinal da molécula repórter, o quencher (sondas TaqMan®, Applied Biosystems, EUA), ou por não ter incremento de produto para ligação do corante (SYBR Green®, Life Technologies, EUA). Nessas estratégias, a reação de PCR é caracterizada por três fases: a fase exponencial (quando o alvo é amplificado exponencialmente), a fase linear

(quando a quantidade do produto é amplificada a uma taxa linear progressiva); e a fase platô (na qual não há mudança significativa na quantidade de produto amplificado), (D'HAENE et al., 2010; VANGUILDER et al., 2008). Na fase inicial da PCR, os substratos não estão limitados e não há nenhuma degradação dos produtos. À medida que a reação ocorre, o alvo é amplificado de maneira exponencial até atingir uma concentração suficiente para iniciar a emissão de fluorescência em quantidade detectável. O número de ciclos na reação em que esse valor ocorre é definido como limite de transição de fases (Ct do inglês – *Cycle threshold*), ocorrendo no limiar de mudança entre a fase exponencial e a linear (D'HAENE et al., 2010; VANGUILDER et al., 2008). O valor de Ct determinado para cada reação pode ser usado então para gerar uma quantificação relativa usando o método Ct comparativo, onde o valor de Ct encontrado para o alvo é comparado com um gene, ou da expressão gênica, de um controle de reação, geralmente controle endógeno (D'HAENE et al., 2010; VANGUILDER et al., 2008). Nas sondas TaqMan®, quando ocorre a hibridização com a sequência-alvo de interesse, a molécula repórter é liberada emitindo a fluorescência, que é detectada pelo aparelho termociclador. O nível de fluorescência detectado está diretamente relacionado à quantidade de alvo amplificada em cada ciclo de PCR, formando uma curva de amplificação no formato sigmoide (D'HAENE et al., 2010; VANGUILDER et al., 2008).

### **Variações no número de cópias**

O advento das técnicas de biologia molecular foram fundamentais para o surgimento de ferramentas de alta tecnologia para o rápido desenvolvimento científico. Durante o projeto de sequenciamento do genoma humano na transição do século XX para o XXI, a

aplicação de um conjunto de clones em vetores de cromossomos artificiais de bactéria (BAC, do inglês *Bacterial Artificial Chromosome*) permitiu que o processo de sequenciamento ocorresse em escala massiva. O microarranjo baseado em BAC para hibridização genômica comparativa, o CGH-BAC-Array, possibilitou a identificação de variações submicroscópicas em todo genoma por meio de genotipagem de pacientes-controle saudáveis (IAFRATE et al., 2004; SEBAT et al., 2004; TUZUN et al., 2005). Embora o termo “saudável” tenha sido utilizado originalmente para identificar indivíduos livres de determinadas doenças, a amplitude de fenótipos comuns variáveis fez com que uma diversidade de polimorfismos genéticos possivelmente associados fosse identificada em indivíduos sem manifestações patológicas. Com isso, o termo “grupo controle” ou mesmo os dados de sequenciamento de um único indivíduo tomado como controle, podem gerar questões de ordem prática na análise populacional de variantes genéticas, criando uma suposição hipotética onde os termos saudáveis ou controle são usados como definição *a priori*, que, na realidade, por si é multifatorial e complexo de estruturar.

A variabilidade fenotípica em populações humanas é determinada em grande parte por dois fatores complementares: meio ambiente e informação genética. A variação genética é causada por diferentes variantes presentes no genoma humano, sejam polimorfismos comuns ou mutações menos frequentes. O grande número de variações estruturais entre indivíduos constituiu-se como uma importante fonte de diversidade e variabilidade genética populacional. Além disso, as variações estruturais foram incluídas em estudos de grande porte que buscavam entender seu papel na variação fenotípica comum e na susceptibilidade a doenças (IAFRATE et al., 2004; SEBAT et al., 2004; TUZUN et al., 2005).

Até recentemente, pensava-se que a maioria destas variantes fosse mudanças pontuais de um nucleotídeo, SNP, que estão presentes no genoma humano em torno de aproximadamente 10 milhões, separadas em categorias de polimorfismos frequentes (frequência de alelos acima de 1% em populações) ou variantes raras. Por definição, variantes raras são definidas como aquelas abaixo de 1% (FRAZER et al., 2009). No entanto, de maneira menos estrita, essas definições de percentual podem variar em frequência populacional entre 0,1% e 3% dependendo do tipo de polimorfismo e a relação genética com outras variantes, podendo chegar até 5% caso não apresentem desequilíbrio de ligação (FRAZER et al., 2009). Abaixo desse limiar podem ser caracterizadas como variantes novas (FRAZER et al., 2009). Os polimorfismos do genoma humano foram descritos a partir dos resultados de dois grandes projetos derivados do genoma humano: *The International HapMap Project* (GIBBS et al., 2003) e seu seguimento, o *One Thousand Genome Project* (KAISER, 2008).

À medida que foram surgindo novas ferramentas de análise, grandes duplicações e deleções foram identificadas também como fonte de variação genética sem estarem aparentemente relacionadas com nenhuma doença (BARBER et al., 1998; IAFRATE et al., 2004; SEBAT et al., 2004). O termo mais abrangente para essas variações, variação estrutural de segmentos genômicos, inclui em seu escopo o polimorfismos que envolvem ganho ou perda de porções a partir de 1000 pares de bases (ou 1 kilobase, kb) de extensão, ou maior (FEUK et al., 2006; SCHERER et al., 2007). Além disso, rearranjos de posição, como as translocações, ou de orientação, como as inversões, e outros mais complexos também complementam o conjunto de variações definidas pelo termo. As duplicações segmentares, também chamadas de cópia de baixa repetição (LCR, do inglês *Low Copy Repeat*) ou

Duplicons, são definidas como segmentos de DNA que ocorrem em duas ou mais cópias por genoma haplóide, com as diferentes cópias partilhando a partir de 90% de identidade de sequência (BAILEY et al., 2001).

Dentro desse escopo, as variações no número de cópias (CNV – do inglês *Copy Number Variation*) são definidas como alterações genômicas com pelo menos 1 kb de extensão, envolvendo somente ganhos ou perdas em comparação com um genoma de referência (IAFRATE et al., 2004; LEVY et al., 2007; SCHERER et al., 2007; SEBAT et al., 2004). Devido a limitações técnicas, tais rearranjos no genoma humano eram assumidos como limitados em escala. No entanto, desde 2004, o desenvolvimento de tecnologias de microarranjos e a disponibilidade da sequência completa do genoma humano tornaram possíveis e eficazes a obtenção de mapas individuais de genoma completo com os locais, precisão de cópias e frequências de CNV, revelando-as como uma das maiores fontes de variação genética humana (IAFRATE et al., 2004; REDON et al., 2006; SEBAT et al., 2004; TUZUN et al., 2005).

É inegável que os SNP tenham aparecido como a maior fonte de variação genética em termos de número (LANDER, 2011). No entanto, o nível sem precedência de diversidade genética conferida por CNV abriu um novo tópico no entendimento da variação fenotípica, evolução humana e susceptibilidade a doenças. As variações no número de cópias cobrem individualmente aproximadamente 12% do genoma humano (REDON et al., 2006), com uma média de 12 variantes por indivíduo (FEUK et al., 2006; ITSARA et al., 2009; ITSARA et al., 2010), podendo chegar entre 20 até 30 CNV, sendo esses números dependentes dos limites de detecção definidos pela densidade dos *arrays* por genoma haploide e densidade

das sondas no mapa genômico, ou seja, regiões com alta ou baixa densidade de sondas (JAKOBSSON et al., 2008; LEVY et al., 2007).

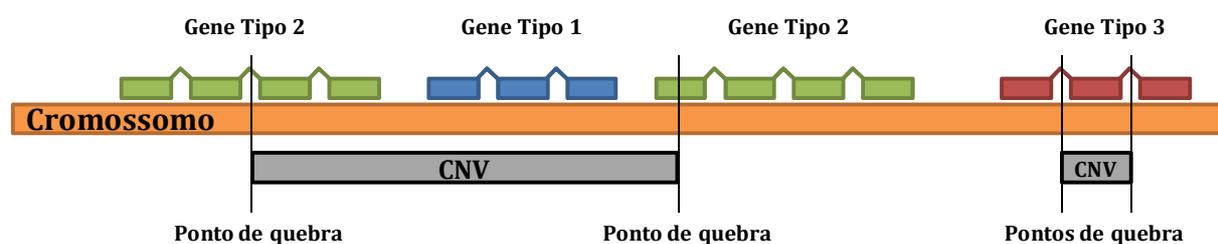
Muito embora essas variações sejam pouco numerosas quando comparado aos SNP, elas podem se estender por abrangência desde 1 kb por evento, podendo atingir regiões que codificam proteínas (SEBAT et al., 2004) ou regiões regulatórias da expressão gênica (STRANGER et al., 2007), resultando em uma fração significativa do genoma, tanto em extensão quanto em função (WOODWARK; BATEMAN, 2011). Em termos de número, a estimativa de variação individual pode chegar a mais de 350 milhões de pares de base (REDON et al., 2006), o que coloca definitivamente as CNV como uma importante fonte da variação genética.

As mutações somáticas estão presentes em diversos tipos celulares e em tecidos aparentemente sem alteração morfológica ou funcional, sendo crucial para o entendimento do desenvolvimento embrionário e do processo de envelhecimento (DE, 2011). Durante o ciclo de vida de um indivíduo, desde a fecundação de um óvulo até o envelhecimento, as células somáticas podem acumular diversas alterações genéticas, de maneira que a mesma linhagem vinda de um único óvulo fertilizado pode dar origem a células com diferenças genéticas relevantes entre tecidos distintos, ou mesmo dentro do mesmo tecido (YOUSSOUFIAN; PYERITZ, 2002). A existência de células com genótipos distintos em um mesmo organismo é denominada de mosaicismo somático (DE, 2011; YOUSSOUFIAN; PYERITZ, 2002).

O mosaicismo comumente surge como resultado de diferentes alterações genéticas como SNP, CNV, rearranjos estruturais, ganhos e perdas de cromossomos, por diversos

mecanismos, tanto induzidos, como por exemplo, exposição à radiação ultravioleta e o uso de drogas lícitas e/ou ilícitas; quanto espontâneos, por exemplo, elementos transponíveis, erros na replicação de DNA ou mecanismos defeituosos de reparo do DNA (DE, 2011; YOUSSEFIAN; PYERITZ, 2002). Assim, as CNV aparecem potencialmente como candidatas ao mosaicismos, podendo ocorrer em células germinativas, em células somáticas que geram agregados teciduais dispersos ou mesmo diferentes mosaicos e eventos únicos em células individuais (MCCONNELL et al., 2013).

A sua extensão e localização de pontos de quebra permite que diversos tipos de interferência genômica possam ocorrer. Assim, Woodwark e Bateman (2011) propuseram uma nomenclatura para diferentes tipos de genes afetados pelas CNV (Figura 01). Os genes que variam em número de cópia e que estão posicionados inteiramente dentro de CNV representam a primeira classe, ou tipo 1. A segunda classe, ou tipo 2, são genes em que pelo menos um dos pontos de quebra da CNV localiza-se dentro dos seus limites de extensão levando a fusão ou truncamentos. A terceira classe de genes afetados por CNV, ou tipo 3, são aqueles cuja CNV se encontra inteiramente dentro dele, geralmente esses genes podem ser interrompidos dependendo se o limite da CNV se localize em porções que atinjam os exons (WOODWARK; BATEMAN, 2011).



**Figura 1:** Tipos de Genes encontrados em regiões de variação do número de cópias classificados de acordo com o ponto de quebra da CNV. Caixas coloridas representam regiões gênicas de *exons* e linhas *introns*. Fonte: adaptado de Woodwark e Bateman (2011).

Por representarem estruturas genômicas que podem estar associadas a fenótipos diversos, e que variam de ocorrência e frequência entre indivíduos e entre populações, as CNV podem ser consideradas como modelos alélicos de variação genômica, atribuindo a referência de alelo ao número de cópias presentes (KATO et al., 2009). Em populações diversas, CNV com frequências populacionais acima de 1% são caracterizadas como polimorfismos no número de cópias (CNP, do inglês *Copy Number Polymorphism*). Uma variante de menor frequência é definida por aquela menos comum em uma determinada população, no entanto o uso deste termo se complica quando uma região é multialélica, pois a frequência de um alelo, ou o número de cópia neste caso, pode variar de acordo com a população estudada e com a técnica utilizada para identificar as cópias (SCHERER et al., 2007). Portanto, o termo frequência do número de cópia representaria melhor a realidade desse tipo de polimorfismo (LOCKE et al., 2006).

As CNV abriram precedentes também para estudos evolutivos e populacionais. Diferenças significativas na frequência entre CNV foram descritas entre diferentes grupos

étnicos populacionais (CONRAD, D. F.; HURLES, 2007; JAKOBSSON et al., 2008; LI et al., 2009; MCELROY et al., 2009; REDON et al., 2006; WINEINGER et al., 2011; XU et al., 2011; ZHANG, Y. B. et al., 2012), inclusive com relatos de CNV observadas exclusivamente em determinadas populações (ARMENGOL et al., 2009; PRONOLD et al., 2012; XU et al., 2011). Isso sugere um papel importante das CNV nos modelos de estratificação de populações mundiais, implicando em uma semelhança na história evolutiva de marcadores genéticos entre CNV em comparação com outros tipos de polimorfismos (JAKOBSSON et al., 2008).

Diversos exemplos de polimorfismos de número de cópias em diferentes populações são relatados na literatura, com famílias de genes que podem atingir até centenas cópias (CAMPBELL et al., 2011; SUDMANT et al., 2010). O gene da Sulfotransferase Citosólica Humana (SULT1A1) possui distribuição do número de cópias que varia entre 1 e 5, e sua frequência para ganho (3 cópias ou mais) é relatada em torno de 26% em populações de origem Europeia e 63% para populações de origem Africana, constituindo um exemplo de CNV polimórfica em populações humanas (HEBBRING et al., 2007). Outro exemplo é o gene da amilase salivar, AMY1, cujo polimorfismo varia entre 2 a 15 cópias, e é relatado em diversas populações com histórico contrastante em relação ao padrão de consumo dietético de amido (PERRY et al., 2007).

Um importante papel das CNV na evolução pode ser inferido a partir da análise das categorias funcionais dos genes que mais frequentemente se apresentam nesta classe de variante estrutural. Uma das características mais importantes para os estudos funcionais de CNV é sua localização predominantemente fora de sequências codificantes e elementos reguladores ultraconservados (REDON et al., 2006). Estes elementos são sequências de,

pelo menos, 200 pares de bases altamente conservadas entre várias espécies, incluindo humanos, ratos e camundongos (BEJERANO et al., 2004). As categorias funcionais com maior enriquecimento para CNV foram para genes envolvidos na adesão celular, na percepção sensorial olfativa e de respostas a estímulos químicos (NGUYEN et al., 2006; REDON et al., 2006). Além disso, as categorias que foram sub-representadas incluiu genes envolvidos na sinalização e proliferação celular, bem como os genes envolvidos na regulação de fosforilação de proteínas (NGUYEN et al., 2006; REDON et al., 2006).

Entretanto, as variantes estruturais também podem influenciar os genes em *cis* ou em *trans* ao longo de grandes distâncias genômicas ou mesmo em cromossomos distintos, respectivamente. Assim, as CNV podem afetar a expressão gênica em centenas de quilobases de distância do local do rearranjo, atuando em elementos regulatórios de transcrição e tradução, e também por meio do controle de RNA mensageiro (mRNA) e taxas de degradação de proteínas (WEISCHENFELDT et al., 2013). Uma classe de moléculas envolvidas principalmente na regulação pós-transcricional da expressão gênica são o RNA não codificantes, entre eles os microRNA.

## **MicroRNA**

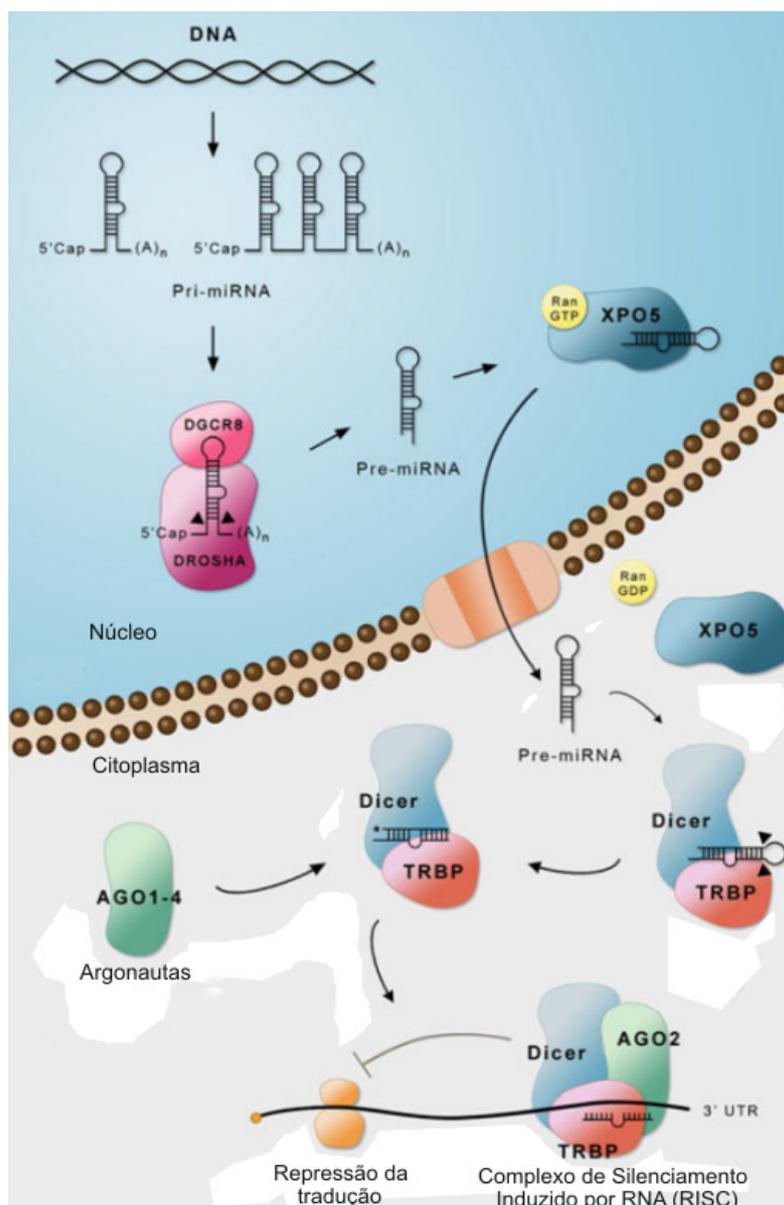
O entendimento do controle da expressão gênica foi constantemente modificado nas duas últimas décadas pela descrição extensiva de RNA não codificadores (ncRNA). A maioria desempenha funções específicas, tais como o RNA ribossomal (rRNA) e o RNA transportador (tRNA), envolvidos na tradução do mRNA. No entanto, foi identificada nos

genomas de diversos organismos, uma classe abundante de pequenos RNA com prováveis funções de regulação pós-transcricional de determinados processos biológicos, tais como desenvolvimento, crescimento, diferenciação e metabolismo celular (AMBROS et al., 2003; LAU et al., 2001; REINHART et al., 2002). Entre os ncRNA, os pequenos RNA nucleares (snRNA) e os pequenos RNAs nucleolares (snoRNA) participam, respectivamente, do processamento de mRNA (ou em inglês, *splicing*) e na modificação do rRNA. Além deles, os microRNA notoriamente ganharam espaço no meio científico devido a sua função de regular a expressão gênica por meio da clivagem ou inibição da tradução de mRNA alvo (AMBROS et al., 2003; LAU et al., 2001; REINHART et al., 2002).

Os microRNA são curtas moléculas de RNA de cadeia simples que são complementares a um ou mais RNA mensageiros e variam entre 21 a 30 nucleotídeos de tamanho, de acordo com mecanismos distintos de biogênese (FARAZI et al., 2008; MELO; MELO, 2014). A biogênese dos miRNA (Figura 02) é iniciada primeiramente pela transcrição da sua unidade transcricional, genes codificantes de RNA, por meio da enzima RNA polimerase II, gerando um transcrito longo de RNA chamado de miRNA primário (pri-miRNA), contendo um ou vários transcritos (MELO; MELO, 2014). O transcrito primário é modificado da mesma maneira que os transcritos de mRNA, ou seja, recebe a marcação da 7-metilguanossina na extremidade 5' (5'cap) e a cauda de poli-adenosina (poli-A) na extremidade 3' (AMBROS et al., 2003). Então são processados no interior do núcleo em miRNA em forma precursora de grampo em fita dupla (pre-miRNA, de aproximadamente 70 pb) pela enzima RNase III Drosha (RNase III endonuclease) juntamente com o cofator *DiGeorge syndrome critical region 8* (DGCR8), (DAVIS-DUSENBERY; HATA, 2010; LEE, Y. et al., 2003). Uma vez processado, o microRNA é exportado do núcleo para o citosol pela

Exportina-5 (LUND et al., 2004). O complexo enzimático Dicer, um membro da superfamília RNase III, juntamente com a proteína de ligação TAR (TRBP), medeia o processamento citoplasmático do pre-miRNA em miRNA maduros, gerando um miRNA de fita dupla de 19 a 24 pb (LUND et al., 2004). Uma destas fitas, chamada de fita passageira, é degradada por RNases, enquanto a fita guia, ou fita madura, seletivamente se associa ao complexo de silenciamento induzido por RNA (RISC – do inglês *RNA-induced silencing complex*) e interage com os mRNA alvos por complementaridade na região 3'UTR, para degradá-los ou reprimir sua tradução (SCHWARZ et al., 2003).

O reconhecimento acontece com um grau de complementaridade imperfeito, com falta de complementaridade principalmente na parte central do miRNA. A complementaridade perfeita promove a clivagem do mRNA catalisada pela associação da proteína ARGONAUTA ao complexo miRISC (CARTHEW; SONTHEIMER, 2009; INUI et al., 2010). A única proteína Argonauta com atividade de clivagem do mRNA em humanos é a Ago2. No entanto, todas as demais quatro proteínas Ago se associam com miRNA e atuam no silenciamento gênico (CARTHEW; SONTHEIMER, 2009; INUI et al., 2010). A repressão da tradução do mRNA ocorre quando a complementaridade é imperfeita e, por mecanismos ainda não elucidados, sugere-se que a repressão aconteça durante a elongação e o miRNA promova a dissociação entre os ribossomos e o mRNA (CARTHEW; SONTHEIMER, 2009; INUI et al., 2010).



**Figura 2:** Esquema da via de biogênese e função dos microRNA. Fonte: adaptado de Melo e Melo (2014)

As transcrições primárias dos microRNA, em seres humanos, são processadas no núcleo e clivadas pelo complexo Drosha-DGCR8 que reconhece a estrutura de RNA de cadeia dupla do pri-miRNA e libera um precursor (pre-miRNA) de 60 a 70 nucleotídeos. Este pré-miRNA é exportado para o citoplasma via exportina-5, onde é novamente processado em miRNA/miRNA\* pelo complexo DICER-TRBP, e em seguida associada a proteína Argonauta (Ago2 em humanos), formando o complexo de silenciamento (RISC). A fita simples miRNA serve como guia para o reconhecimento do alvo, enquanto a fita passageira miRNA\* é clivada pela Ago2.

As sequências de DNA que codificam para miRNA, os genes de microRNA, podem ser encontradas em diversas regiões genômicas, como as intergênicas e intrônicas (INUI et al., 2010). Estima-se que existam milhares de miRNA identificados em várias espécies de organismos, dos quais 1.872 precursores e 2.578 maduros estão descritos em seres humanos (dados de mirbase.org em Janeiro de 2014). Evidências sugerem que os miRNA estão diretamente envolvidos na biogênese de diversas características fenotípicas, como por exemplo, a regulação de secreção de insulina (POY et al., 2004), uma vez que cerca de um terço dos genes humanos pode ser controlado por miRNA (BEREZIKOV et al., 2005). Comparado com genes não regulados por miRNA, no entanto, os genes regulados por miRNA em geral têm uma variação maior na sua expressão entre diferentes grupos étnicos ou populacionais (LU; CLARK, 2012), além da expressão diferenciada dos próprios miRNA entre grupos distintos (HUANG et al., 2011).

### **Implicações funcionais do número de cópias**

A variação no número de cópias de regiões genômicas possui algumas características evolutivas que podem ser funcionais ou meramente estruturais, incluindo sequências intrônicas ou exônicas, regiões promotoras com sítios de ligação para fator de transcrição ou regiões intergênicas (KONRAD et al., 2011). Cópias múltiplas de um gene podem servir de alternativa de segurança para genes essenciais. Por exemplo, para um gene que perde sua função em decorrência de uma mutação, sua cópia pode assumir o papel funcional, evitando possíveis efeitos deletérios (KONRAD et al., 2011). Esse modelo foi proposto como o mecanismo primário para a preservação de cópias de genes duplicados

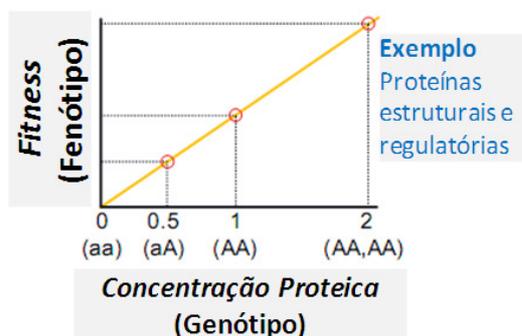
onde a duplicação é um evento neutro à pressão seletiva. Em contraste, o modelo de divergência assume que uma duplicação pode ser positivamente selecionada porque doses mais elevadas de proteína fornecem uma vantagem seletiva imediata, em casos de genes sensíveis ao efeito de dosagem (SOSKINE; TAWFIK, 2010).

O efeito de dosagem gênica corresponde à quantidade de produto do gene (mRNA ou proteína expressos) que irá aumentar de acordo com cada cópia adicional do gene, até um limite superior ser atingido, assumindo que não existam redes de retroregulação que irão suprimir a expressão de genes (INNAN; KONDRASHOV, 2009). Fenótipos comuns têm sido relacionados positivamente com nível de expressão de proteínas, em uma maneira na qual o conceito de equilíbrio genético sensível à dosagem pode ser totalmente aplicável. O caso mais notável envolvendo variação fenotípica comum é o gene da amilase salivar (AMY1), detectados em uma frequência diferenciada populacionalmente entre ganhos e perdas cromossômicas (IAFRATE et al., 2004). Especificamente, a enzima amilase salivar foi identificada com níveis de expressão aumentados em função do aumento no número de cópias da região genômica que é transcrita, tendo sido ainda positivamente correlacionada com populações que possuem tradicionalmente dietas ricas em amido. Ou seja, em média, quanto maior o consumo de amido, mais cópias do gene AMY1 e maior concentração da enzima amilase salivar estarão presentes na população (PERRY et al., 2007). As implicações funcionais dessas alterações são, portanto, evidentes no impacto que a expressão gênica pode exercer sobre o fenótipo, de forma abrangente. Por exemplo, a deleção de um gene pode gerar a depleção de seu(s) produto(s), especialmente quando a região abriga genes onde a expressão de uma única cópia não é capaz de expressar o fenótipo associado, como no caso dos genes haploinsuficientes (CONRAD, B.; ANTONARAKIS, 2007; DANG et al.,

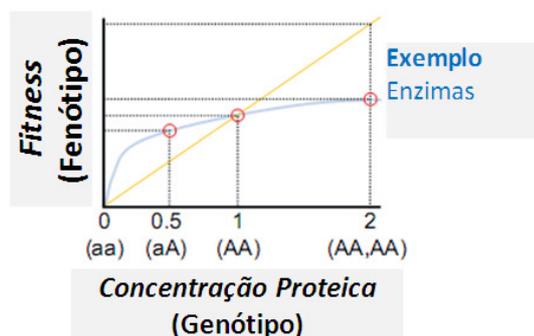
2008), ou ainda, cópias extras de genes podem causar a superexpressão de seus produtos (CONRAD, B.; ANTONARAKIS, 2007; PERRY et al., 2007).

Apesar de ser uma interação complexa e repleta de exceções (GOUT et al., 2010), existem propostas de modelos alternativamente simples para explicar o comportamento de dosagem gênica com relação aos fenótipos (CONRAD, B.; ANTONARAKIS, 2007). Entre elas podem-se destacar quatro (Figura 3): (a) Relação linear: diretamente proporcional entre a concentração de produto e o genótipo, sendo normalmente encontrada em proteínas estruturais e regulatórias; (b) Retorno decrescente: a concentração de produto mostra pouca variação ao longo de grandes intervalos de dose a partir de uma dosagem funcional mais adaptada, envolve caso de enzimas que funcionam em concentrações limitadas; (c) Efeito estequiométrico: essa alternativa corresponde à diminuição do fenótipo (ou *fitness* genético) tanto para incremento quanto para supressão de dosagem gênica. Certas categorias de genes funcionais enriquecidos para os reguladores de transcrição e moléculas sinalizadoras fazem parte dessa categoria. Geralmente envolvem complexos proteicos com subunidades em que, se uma das unidades monoméricas estiver presente em uma quantidade muito abaixo ou muito acima da concentração de formação dos intermediários, o complexo não se forma na quantidade preconizada (VEITIA et al., 2008); (d) Proteínas com característica de agregação: nesse modelo proteínas tendem a se agregar caso o limiar de expressão diploide tenha sido ultrapassado, como é o caso da proteína precursora da  $\beta$ -amilóide, uma das responsáveis, em parte, pela etiologia da doença de Alzheimer. Conrad e Antonarakis (2007) propuseram esses modelos baseados em diversos dados publicados em relação ao *fitness* genético e modelos de genótipos de haploinsuficiência até duplicação do alelo selvagem (indicados na figura 3 pelos símbolos “a” e “A”).

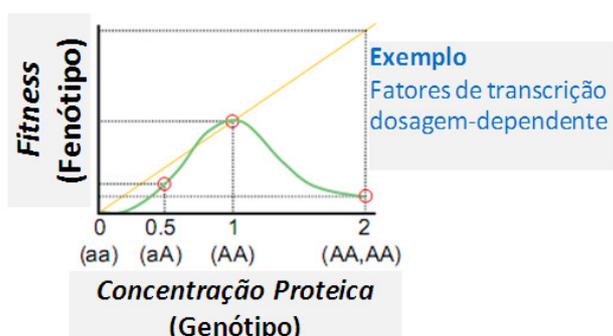
### a Função Linear



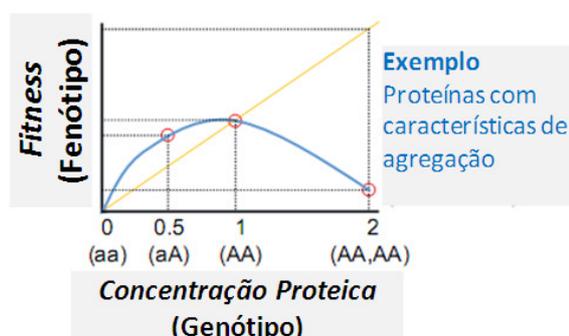
### b Retorno decrescente



### c Relação Estequiométrica



### d Agregação



**Figura 3:** Representação esquemática da resposta de dosagem gênica ao fenótipo. (a) Relação linear; (b) Retorno decrescente; (c) Efeito estequiométrico; (d) Proteínas com característica de agregação. Fonte: Adaptado de Conrad e Antonarakis (2007).

No entanto, algumas destas variantes existem também nos genomas de indivíduos fenotipicamente normais, logo, elas podem não ser uma causa direta somente de doenças genéticas, mas a sua presença pode levar a rearranjos cromossômicos que podem influenciar a expressão de genes específicos dando origem a variações fenotípicas comuns (GIGLIO et al., 2001; HOLLOX et al., 2003; IAFRATE et al., 2004; OSBORNE et al., 2001; SHAW; LUPSKI, 2004). Dentre essas possibilidades, se enquadram as CNV de regiões não

codificantes e as que regulam a expressão de genes próximos ou distantes pela alteração dos seus potenciadores, ou do inglês, *enhancers* (KLOPOCKI; MUNDLOS, 2011).

Os genes de microRNA, por estarem representados em regiões pequenas do genoma, estão mais propensos a se enquadrarem em genes com CNV do tipo 1 e, assim, estariam afetados diretamente no quantitativo de seu produto (MARCINKOWSKA et al., 2011). Pelo fato das CNV participarem de 18% das variações na expressão de genes, existe uma possível correlação em que a expressão dos miRNA pode ser modificada pelo número de cópias genômicas (MARCINKOWSKA et al., 2011).

Em casos de polissomia, como a trissomia do cromossomo 21, em cinco dos miRNA derivados do cromossomo 21 humano, três foram superexpressos em um dos experimentos, o miR-99a, let-7c e miR-155 (KUHN et al., 2008), e um deles, o miR-155, foi diferencialmente expresso em fibroblastos isolados a partir do indivíduo com Síndrome de Down quando comparado com o irmão gêmeo não-afetado (SETHUPATHY et al., 2007). Foi verificado que a superexpressão de microRNA do cromossomo 21 resultou na diminuição específica da expressão de proteínas alvo que contribuem, em parte, para o fenótipo da Síndrome de Down (ELTON et al., 2010). Esses experimentos demonstraram que, de certa maneira, a expressão de microRNA pode ser modulada de acordo com o número de cópias no genoma, interferindo assim no processo traducional e conseqüentemente na variação fenotípica (ELTON et al., 2010). A expressão diferenciada de microRNA tem sido observada também em regiões recorrentes de alteração do número de cópias, tanto perda como ganho, em casos de câncer, como os de leucemia linfoblástica aguda (NINOMIYA et al.,

2012), leucemia linfocítica crônica (CALIN et al., 2002) e adenoma hipofisário (BOTTONI et al., 2005).

Esse fenômeno abre perspectivas para estudos de fenótipos comuns e patológicos envolvendo a variação do número de cópias em genomas humanos e sua expressão diferenciada. Estudos recentes de bioinformática identificaram genes para os pré-microRNA em sítios gênicos que estão inseridos em regiões de variações estruturais recorrentes e polimórficas (DUAN, S. et al., 2009; FELEKKIS et al., 2011; MARCINKOWSKA et al., 2011). Sugere-se que a variação do número de cópias, tanto de regiões de microRNA quanto a de seus alvos na qual eles agem, podem ter modulado os microRNA a atuar como equilibradores da expressão gênica durante a evolução, numa tentativa de regular a expressão gênica e aumentar a tolerância à plasticidade do genoma (FELEKKIS et al., 2011). Porém, o estudo da variabilidade comum e patológica entre variação do número de cópias e expressão dos respectivos microRNA ainda necessita de investigações a fim de elucidar se sua resposta pode ser classificada como sensível à dosagem.

## **HIPÓTESES**

As regiões genômicas de microRNA descritas como polimórficas por estudos de bioinformática, apresentam variação no número de cópias em uma amostra da população brasileira.

A expressão de microRNA é afetada pelo número de cópias genômicas das respectivas regiões.

## **OBJETIVOS**

### **Objetivo Principal**

Analisar a expressão de microRNA com relação às variações estruturais do número de cópias dos respectivos genes em humanos.

### **Objetivos Específicos**

- Identificar regiões polimórficas no número de cópias contendo genes de microRNA;
- Caracterizar as regiões de CNV como polimórficas em uma amostra da população Brasileira;
- Quantificar o número de cópias das CNV em indivíduos sem alterações patológicas;
- Quantificar a expressão de microRNA em indivíduos sem alterações patológicas;
- Correlacionar o número de cópias das CNV com a expressão de microRNA
- Estabelecer modelos de interação entre o número de cópias e expressão de microRNA.

## **MATERIAL E MÉTODOS**

### **Amostras biológicas**

A amostragem utilizada para triagem molecular de CNV foi constituída por 103 indivíduos não relacionados da população Brasileira (Masculino n= 52 e Feminino n= 51), distribuídos entre as cinco regiões geopolíticas na seguinte configuração: Centro-Oeste n=21; Nordeste n=20; Norte n= 22; Sudeste n=22; Sul n=18. Essa amostragem aborda a caracterização de variabilidade genética em termos populacionais, portanto os indivíduos que compuseram a amostragem não passaram por critérios de seleção para inclusão e exclusão de acordo com características fenotípicas. Para fins de cálculos probabilísticos, essa amostragem apresenta um poder de 90% para uma frequência alélica mínima de 0,01 (B-RAO, 2001)

Para a avaliação da relação entre o número de cópias e a expressão de microRNA foram coletados, ao mesmo momento, o sangue periférico (origem embrionária mesodérmica mesenquimal) e células epiteliais da mucosa oral (origem embrionária ectodérmica) de 30 indivíduos, que tiveram o DNA e o RNA extraídos desses dois tecidos, além de RNA total extraído do plasma, fluido corporal com quantidades significativas de microRNA circulante (WEBER et al., 2010). O projeto foi aprovado pelo Comitê de Ética em Pesquisa com Seres Humanos da Universidade Católica de Brasília, via Plataforma Brasil CAAE: 13575813.1.0000.0029 - número do parecer: 218.781. Todos os voluntários

assinaram um termo de consentimento livre e esclarecido autorizando o uso de seu material biológico para esta pesquisa (Anexo I, página 96).

A extração do DNA de sangue periférico foi realizada por coleta venosa em tubo Vacutainer® (BD Diagnostics, EUA) contendo EDTA como anticoagulante, e estocado em 4°C de temperatura para posterior extração pelo método de alta concentração salina (do inglês “*Salting out*”). Para a extração de DNA de células do epitélio da mucosa oral foi coletado um esfregaço bucal com escova cervical durante um minuto e inserido em um tubo de microcentrífuga de 1,5 mL, que prosseguiu para extração com o kit comercial GFX Genomic DNA Purification Kit (GE Healthcare, EUA).

A extração de RNA das células bucais foi realizada com modificações do protocolo descrito por Spira et al. (2004). Um esfregaço das células bucais foi realizado com uma escova cervical por um minuto e, em seguida, foi inserido em um tubo de microcentrífuga de 1,5 mL contendo 0,75 mL de solução RNeasy (Qiagen, EUA) e incubado em baixa temperatura para conservação (4°C por até 72h ou -20°C por até 15 dias). Após o armazenamento os tubos foram centrifugados a 8000 g para formar pellet de células. Descartou-se totalmente o RNeasy, e, se necessário, com auxílio de micropipeta retirou-se o restante de modo a não deixar solução no tubo. Em seguida, as frações de RNA total foram extraídas segundo o protocolo de extração pelo reagente miRNeasy (Qiagen, EUA).

Para o isolamento de células mononucleadas do sangue (PBMC – do inglês *peripheral blood mononuclear cells*) foi coletado aproximadamente 4,0 mL de sangue periférico em tubo Vacutainer® (BD Diagnostics, EUA) contendo EDTA como anticoagulante, estocados em temperatura ambiente e processados em até 2 horas. A separação das frações do sangue

total foi realizada pelo método de gradiente em Ficoll-hypaque densidade 1.077 g/L, (Pharmacia Biotech, Brasil). O método consiste na adição vagarosa de 3,0 mL de sangue recém coletado, com ajuda de uma micropipeta de 1000 µL em um tubo de centrifuga de 8,0 mL contendo 2,0 mL de Ficoll-hypaque em temperatura ambiente. Em seguida foi centrifugado durante 20 minutos a 1500-1800 RFC (aproximadamente 5000 rpm em rotor de 10 cm de raio) e frenagem ajustada para o nível mínimo (se possível sem frenagem). Após a centrifugação, as células mononucleares e plaquetas constituíram uma camada esbranquiçada entre a camada de Ficoll-hypaque (acima da camada de hemácias) e a de plasma. Foi aspirada a maior quantidade possível de plasma sem perturbar a camada de células, que em seguida foi imediatamente coletada com ajuda de uma micropipeta de 200 µL e foi transferida para um tubo de microcentrifuga de 1,5 mL contendo 0,75 mL de solução RNAlater™ (Qiagen, EUA). Posteriormente, as frações de RNA total foram extraídas das células mononucleadas isoladas e do plasma sanguíneo segundo o protocolo do reagente miRNeasy (Qiagen, EUA).

O protocolo de extração de RNA total pelo kit miRNeasy (Qiagen, EUA), consistiu em primeiramente isolar as células ou fluido de interesse lisadas com 500 µL de Trizol® (Invitrogen Life Technologies, EUA) ou Qiazoll® (Qiagen, EUA) sobre vortex constante por 1 minuto. Deixou-se descansar em temperatura ambiente por 5 minutos e adicionou-se 140 µL de clorofórmio, fechando o tubo e agitando vigorosamente sobre vortex constante por 15 segundos. Deixou-se descansar em temperatura ambiente por 2 a 3 minutos. Centrifugou-se a 12.000 rpm a 4°C por 15 minutos e transferiu-se o sobrenadante (fase aquosa) para novo microtubo (do kit). Adicionou-se etanol 100% em 1,5 vezes o volume do transferido e misturou-se vagarosamente. Sem parar, pipetou-se 700 µL da amostra

(incluindo qualquer precipitado que tenha sido formado) em uma coluna do Kit miRNeasy. Centrifugou-se a 10.000 rpm por 15 segundos em temperatura ambiente. Descartou-se a fase aquosa repetindo o passo anterior caso tenha algum restante da amostra. Adicionou-se 700  $\mu\text{L}$  do Buffer RWT na coluna e centrifugou-se a 10.000 rpm por 15 segundos em temperatura ambiente. Descartou-se a fase aquosa. Em seguida, pipetou-se 500  $\mu\text{L}$  de Buffer RPE na coluna e centrifugou-se a 10.000 rpm por 15 segundos em temperatura ambiente. Descartou-se a fase aquosa e repetiu-se esse passo (500  $\mu\text{L}$  de Buffer RPE) com centrifugação por 2 minutos. Descartou-se a fase aquosa e repetiu-se a centrifugação por mais 1 minuto. Colocou-se a coluna em um novo microtubo de 1.5  $\mu\text{L}$  e pipetou-se 25  $\mu\text{L}$  de água RNase free diretamente na membrana da coluna. Por fim, centrifugou-se a 10.000 rpm por 1 minuto em temperatura ambiente e repetiu-se a adição de 25  $\mu\text{L}$  de água RNase free e centrifugação. Essa última fração de 50  $\mu\text{L}$  representa o RNA total extraído da amostra, que foi devidamente estocado em freezer a  $-80^{\circ}\text{C}$ .

As amostras de RNA e DNA foram submetidas à eletroforese em gel de agarose (1%), utilizando como agente intercalante fluorescente o brometo de etídio (10 $\mu\text{g}/\text{mL}$ ) e padrões de DNA de fago  $\lambda$  com concentrações conhecidas. A visualização das bandas foi realizada no transiluminador de ultra-violeta DyNA Light (Labnet, EUA) com registro em fotodocumentador (Lab Trade, Brasil). Caso obtido sucesso de extração, as amostras foram quantificadas utilizando fluorímetro Qubit® 1.0 (Qiagen, EUA) conforme as recomendações do fabricante. A partir das concentrações originais, as amostras de DNA foram diluídas para a concentração de trabalho de 10 ng/ $\mu\text{L}$ . Devido a baixa quantidade de miRNA circulantes no plasma, não foi possível quantificar o RNA extraído em todas as frações e constituintes biológicos. Para normalizar a quantidade de RNA na reação de transcrição reversa, foi

utilizado o volume fixo constante (5 $\mu$ L) em reações conjugadas em multiplex para todas as amostras (MIZUNO et al., 2011).

### **Regiões Alvo**

As regiões alvo de CNV-miRNA foram exploradas na literatura com ênfase no trabalho de Marcinkowska e colaboradores (2011) que classificaram por análise de bioinformática regiões de alta probabilidade de polimorfismo. Deste trabalho foram selecionados 05 locos com possível variação no número de cópias que possuísem ensaios com sondas para PCR quantitativa TaqMan® (Applied Biosystems, EUA), disponíveis para avaliação do número de cópias e quantificação de microRNA. Além disso, outras 08 regiões foram escolhidas com base em estudos de microRNA em câncer (SLATTERY et al., 2011) e averiguação das regiões de CNV pelo banco de dados do *Database of Genomic Variants (DGV* - [www.projects.tcag.ca/variation](http://www.projects.tcag.ca/variation)), estabelecendo o mesmo critério anterior para aquisição de sondas. Como controle diploide foi escolhida uma região sem variações registradas no DGV e que fosse igualmente utilizada como controle positivo das reações de microRNA.

### **PCR quantitativa- Quantificação de CNV**

As CNV de interesse foram quantificadas utilizando ensaios customizados TaqMan® (Applied Biosystems, EUA) para as regiões específicas. O gene Ribonuclease P RNA component H1 (*RPPH1*) localizado no cromossomo 14 citobanda 14q11.2, descrito

comumente como RNaseP, conhecido por apresentar sempre duas cópias em um genoma diploide, foi empregado como controle interno positivo para calcular o número de cópias do alvo de interesse e para confirmar que a reação foi amplificada com êxito.

As reações de qPCR foram realizadas seguindo o protocolo descrito por Mayo e colaboradores (MAYO et al., 2010) contendo 20ng de DNA, 5 µL TaqMan® Universal PCR Master Mix (Applied Biosystems, EUA), 0,5 µL de sondas e iniciadores das região específicas e de controle ( $20\times$  *working stock*), e, por fim, água ultrapura livre de Nuclease, completando o volume final total para 10 µL. A amplificação foi realizada num sistema de PCR em tempo real StepOnePlus™ Real-Time PCR System (Applied Biosystems, EUA) com um sistema de ciclagem realizado com temperatura inicial de 95°C durante 10 minutos, seguido por 40 ciclos de 95°C durante 15 segundos, 60°C durante 1 minuto.

Os valores de Ct (do inglês *Cycle threshold*) foram gerados a partir de um limiar pré-estabelecido e  $\Delta$ Ct (delta-Ct) foram calculados a partir da diferença dos valores de Ct do gene controle do número de cópia e da região alvo utilizando o software CopyCaller™ (Applied BioSystems, EUA), utilizando o método  $2^{-\Delta\Delta Cq}$  (WEAVER et al., 2010). O número de cópias calculado (CCN) de uma amostra foi obtido mediante da multiplicação da quantidade relativa ( $RQ = 2^{-\Delta\Delta Cq}$ ) estabelecida na reação e o número de cópias de um calibrador, ou seja, uma amostra com número de cópias conhecido. Uma vez obtido o número de cópias calculado foi estabelecido o número de cópias predito, que seria o número inteiro aproximado a partir do valor calculado, levando em conta a RQ e os desvios padrões de cada corrida. Em cada placa foi utilizado um controle negativo de amplificação (água ultrapura no lugar do DNA) e também uma amostra calibrador doada de um voluntário,

como controle positivo. Essa amostra foi submetida a testes de dupla análise (prova/contraprova) em qPCR usando um DNA de referência controle diploide, Affymetrix GeneChip® reference Genomic DNA, para determinação do número exato de cópias da amostra calibrador definitiva (CalD). Todas as reações em todas as placas foram feitas em triplicata.

### **Reação de transcrição reversa de RNA total**

A reação de transcrição reversa (RTR) foi realizada a partir do RNA total extraído para obter o DNA complementar (cDNA) de cada amostra. Foi realizada com o Kit Taq-Man miRNA Reverse Transcription (Applied BioSystems, EUA), juntamente com os primers específicos para os miRNA. Para tal, foi utilizado um protocolo de reação em multiplex sugerido pelo fabricante (user bulletin, Publication Part Number 4465407), com modificações (página 92). Primeiramente foram preparados dois conjuntos multiplex (MPX), contendo 5 iniciadores RTR cada. Os conjuntos de iniciadores foram avaliados para verificar as possíveis formações de dímeros de iniciadores para PCR multiplex, usando como parâmetro de estimativa a energia livre de Gibbs a -10 kcal/mol, no programa online PriDimerCheking release 0.1.0 (SHEN et al., 2010). Para os conjuntos multiplex foram misturados os seguintes iniciadores de RTR em quantidades equimolares - MPX1: hsa-mir-570, hsa-mir-126, hsa-mir-338, hsa-mir-650; RNU6B; MPX2: hsa-mir-449, hsa-mir-1275, hsa-mir-150, hsa-mir-16, RPL21. Para preparar a mistura de reação foi misturada para cada tubo: 10 mM dNTPs (1,0 µL), enzima Multiscribe Reverse Transcriptase (1,2 µL), Reverse-Transcription Buffer (1,5 µL), Inibidor de RNase (0,2 µL), RTR multiplex (2,5 µL total - 0,5

$\mu\text{L}$  de cada iniciador), água livre de RNases (3,6  $\mu\text{L}$ ) e RNA total da amostra (5  $\mu\text{L}$ ), somando um total de 15  $\mu\text{L}$ . A reação foi realizada em tubos de polipropileno livre de Nucleases 0,2 mL, e em seguida foi levada para o termociclador Veriti™ (Applied BioSystems, EUA) nas seguintes condições: 16°C por 30 min, 42°C por 30 min, 85°C por 5 min.

### **PCR quantitativa – Quantificação de microRNA**

As reações de RTR PCR foram diluídas na proporção de 1:5 (60  $\mu\text{L}$  água livre de RNases para 15  $\mu\text{L}$  de reação) e as reações de PCR em tempo real foram feitas em triplicata com volume total diminuído para 10  $\mu\text{L}$  (MITCHELL et al., 2008), com master mix contendo água livre de nucleases (1,5  $\mu\text{L}$ ), 2X TaqMan® Universal PCR Master Mix (4,0  $\mu\text{L}$ ) e 20X TaqMan® miRNA Assay (0,5  $\mu\text{L}$ ). Foram distribuídos 6,0  $\mu\text{L}$  de master mix em uma placa de 96 poços (*MicroAmp Fast 96-well Reaction plate*, Applied Biosystems, EUA) e, então, acrescentados 4,0  $\mu\text{L}$  de produto de RTR, obtido previamente. Em seguida, a PCR em tempo real foi feita no StepOnePlus™ Real-Time PCR System (Applied BioSystems, EUA) nas condições: 95°C por 10 min, seguido por 40 ciclos de 95°C por 15 seg e 60°C por 1 min.

Os resultados obtidos foram analisados primeiramente com o auxílio do *software* SDS versão 2.2 (Applied BioSystems, EUA). A normalização dos dados foi feita pela média dos valores de Ct dos genes constitutivos endógenos (RNU6B e RPL21) utilizados como controles. Em seguida, obteve-se os valores de  $\Delta\text{Ct}$  subtraindo-se dos valores de Ct de cada microRNA os valores da média dos Ct dos normalizadores ( $\Delta\text{Ct} = \text{Ct alvo} - \text{Ct constitutivo}$ ), para cada amostra (SCHMITTGEN; LIVAK, 2008). O passo seguinte foi utilizar o cálculo de

expressão relativa para os microRNA levando em consideração somente os normalizadores, sendo definida como a taxa de expressão relativa aos genes constitutivos, e expressa matematicamente por  $2^{-\Delta Ct}$ , sendo que a aplicação do logaritmo na base 2 gera a expressão relativa entre genes ( $ER = \log_2 2^{(-\Delta Ct)}$ ).

Em seguida, fez-se a comparação entre grupos de referência e grupos de interesse pelo cálculo de  $\Delta\Delta Ct$  ( $\Delta\Delta Ct = \Delta Ct \text{ amostra} - \Delta Ct \text{ referência}$ ). Porém, como o modelo experimental não envolve tratamentos ou momentos diferentes em uma mesma amostra, os grupos foram tomados a partir do número de cópias predito para as regiões genômicas, sendo que uma ou nenhuma cópia representou o grupo “Perda”, duas cópias representou o grupo “diploide” e três ou mais cópias o grupo “Ganho”. Desta forma, o grupo “diploide” foi assumido como grupo de referência e os demais como grupos de comparação. Como as amostras, nesse caso, não são pareadas, a média dos grupos foi utilizada como determinação de variação entre as proporções. Expressão matemática que representa a taxa de mudança (FC do inglês *Fold Change*) foi dada por  $2^{-\Delta\Delta Ct}$  e a aplicação do logaritmo na base 2 do FC gerou a expressão relativa entre grupos ( $ER = \log_2 2^{(-\Delta\Delta Ct)}$ ). No mesmo sentido, Os cálculos de normalização e de  $\Delta Ct$  foram realizados no software DataAssist™ (Applied BioSystems, EUA) que forneceu uma planilha para exportação no Excel, no qual foram concluídos os cálculos.

A eficiência de todos os iniciadores TaqMan® (CNV e microRNA) foi confirmada usando uma diluição seriada de quatro a seis pontos, para se obter uma curva padrão utilizando as mesmas especificações de corrida estabelecidas no StepOnePlus™ Real-Time PCR System (Applied BioSystems, EUA). Com o valor da sua inclinação, foi calculada a

eficiência, cujos valores aceitos deveriam estar entre 90 e 110% e os valores de regressão ( $R^2$ ) aceitos como significativos a partir de 0,995.

### **Análise estatística**

Para verificar a significância estatística dos dados gerados, o programa *SPSS statistics v.13* foi utilizado como ferramenta para a análise. Para testar a dependência de variáveis categóricas e contínuas com o número de cópias foram utilizados, respectivamente, os testes de Qui-quadrado, teste-t e One-way ANOVA. A expressão diferencial dos microRNA foi analisada para significância de distribuição normal através do teste de Kolmogorov-Smirnov e comparada por teste-t entre os diferentes constituintes biológicos, epitélio bucal, células mononucleadas e plasma sanguíneo. Foi elaborado um mapa de expressão relativa para determinar de maneira geral quais microRNA estavam acima (*up-regulated*) ou abaixo da expressão dos controles endógenos (*down-regulated*), baseado em distâncias de correlação de Pearson e agrupadas pelo método de ligação completa utilizando o *software* DataAssist™ (Applied BioSystems, EUA).

Para determinar os microRNA que parecem ter diferentes distribuições entre os diferentes subgrupos de CNV, foi usado o teste não paramétrico de Kruskal-Wallis para comparar mais de duas populações ( $k > 3$ ) em um delineamento experimental inteiramente casualizado, classificando as CNV de acordo com o número de cópias, ou também como grupos de perda ( $CN \leq 1$ ), diploide ( $CN = 2$ ) ou ganho ( $CN \geq 3$ ). Para avaliar os microRNA do plasma, a análise foi relativa ao número de cópias das células mononucleadas do sangue,

visto que existe forte correlação entre eles (PRITCHARD et al., 2011). A partir dos dados obtidos da expressão relativa aos genes constitutivos ( $\log_2 2^{-\Delta Ct}$ ) foi gerada uma simulação usando o modelo Monte-Carlo para predição de comportamento das variáveis aleatórias assumindo a distribuição normal em uma população gerada de 1.000 a 10.000 indivíduos. A simulação Monte Carlo gera números estocásticos das variáveis contínuas (no caso a expressão relativa dos microRNA) com base nos dados reais da distribuição de expressão (média e desvio padrão) e designa esses valores para a população gerada para cada número de cópias baseada na frequência de ocorrência da variação do número de cópias medida na população brasileira (GINSBERG et al., 2002). O teste de Kruskal-Wallis foi utilizado para comparar a diferença das médias entre os grupos. Também foram estabelecidos modelos preditivos de expressão para número de cópias não existentes na população, baseados em dois tipos de equação: (a) equação polinomial de primeira ordem ( $y = ax + b$ ) e (b) equação logarítmica ( $y = a + b \ln x$ ). O ajuste da curva foi feito por regressão ( $R^2$ ). Os modelos foram plotados e calculados no programa *GraphMathica for Windows* versão 2.2e. Para todos os testes estatísticos, o grau de significância válido foi de  $p < 0,05$ , ou seja, 95% de confiabilidade.

Para fins de compreensão, nesta Tese as regiões genômicas de maneira geral, independente de polimorfismo, foram denominadas como CNV-miRNA ou CNV-miR precedendo o número do microRNA. Por exemplo: CNV-miR-1 para representar a região genômica que contém o gene do microRNA hsa-miR-1. Desta forma espera-se evitar a confusão entre os termos de região genômica e do seu transcrito, o próprio microRNA.

## RESULTADOS

A busca por regiões de CNV contendo genes de microRNA na literatura rendeu um total de 209 CNV-miRNA, alguns dos quais com dois ou mais genes localizados mutuamente em uma mesma região de CNV (MARCINKOWSKA et al., 2011). Destes, foram selecionados 11 (Tabela 01) que apresentaram estudos prévios de CNV com entradas depositadas no banco de dados DGV em constantes atualizações (ABECASIS et al., 2012; AHN et al., 2009; BENTLEY et al., 2008; CAMPBELL et al., 2011; CONRAD, D. F. et al., 2006; COOPER et al., 2008; DE SMITH et al., 2007; GUSEV et al., 2009; IAFRATE et al., 2004; ITSARA et al., 2009; JAKOBSSON et al., 2008; JU et al., 2010; KIDD et al., 2008; KIDD et al., 2010; KIM, J. I. et al., 2009; KORBEL et al., 2007; LEVY et al., 2007; LOCKE et al., 2006; MCCARROLL et al., 2006; MCCARROLL et al., 2008; MCKERNAN et al., 2009; MILLS et al., 2006; PANG et al., 2010; PARK et al., 2010; PERRY et al., 2008; PINTO et al., 2007; REDON et al., 2006; SEBAT et al., 2004; SHAIKH et al., 2009; SHARP et al., 2005; SIMON-SANCHEZ et al., 2007; TEAGUE et al., 2010; TUZUN et al., 2005; WANG et al., 2007; WHEELER et al., 2008; WONG, K. K. et al., 2007; WONG, L. P. et al., 2012; XU et al., 2011; ZOGOPOULOS et al., 2007). Para os locos selecionados foi anotado o total de observações de perda e ganho de sequência na amostra geral, permitindo assim o cálculo médio das frequências de ocorrência de CNV (Tabela 01).

Além dos selecionados, foram escolhidos no DGV outros dois locos como controle invariável, ou seja, potencialmente com duas cópias (hsa-miR-16-1 e RNU-6B). A tabela 02 sumariza todas as regiões com a descrição de localização cromossômica dos locos e dos ensaios TaqMan® para microRNA e CNV e os códigos utilizados para identificação dos ensaios no Laboratório, no caso de CNV os três últimos números do fornecedor.

**Tabela 1:** Descrição de 11 microRNA selecionados para o estudo com CNV putativa, a partir do estudo de Marcinkowska et al. (2011). Orientação em relação ao DNA referência; Número de entradas no banco de dados DGV e estudos que depositaram essas entradas, com observações (número de indivíduos) totais da variação com relação a amostra total dos estudos e respectiva frequência relativa. Fonte: o autor.

miRNA	Orientação	DGV	Número de estudos	Observações totais (Ganho/Perda)	Amostra total	Frequência (%)	Referências
CNV-miR-650	+	44	20	624	13704	4,55%	Cooper et al. (2008); de Smith et al. (2007); Itsara et al. (2009); Kim et al. (2009); Locke et al. (2006); McCarroll et al. (2006); Sebat et al. (2004); Shaikh et al. (2009); Sharp et al. (2005); Tuzun et al. (2005); Wang et al. (2007); Wong et al. (2007); Xu et al. (2011); Abecasis et al. (2012); Park et al. (2010); Wong et al. (2012); Ju et al. (2010); Kidd et al. (2010); Pang et al. (2010); Campbell et al. (2011). Xu et al. (2011); Abecasis et al. (2012); Redon et al. (2006); Locke et al. (2006); Wong et al. (2012); Conrad et al. (2006); de Smith et al. (2007); Sharp et al. (2005); Perry et al. (2008); Park et al. (2010); Kim et al. (2009); McKernan et al. (2009); Bentley et al. (2008); Ju et al. (2010).
CNV-miR-570	+	44	17	176	8542	2,06%	Xu et al. (2011); Itsara et al. (2009); Jakobsson et al. (2008); Redon et al. (2006); Simon-Sanchez et al. (2007); de Smith et al. (2007); Perry et al. (2008); Teague et al. (2010).
CNV-miR-126	+	27	8	108	7514	1,44%	Wong et al. (2007); Jakobsson et al. (2008); Park et al. (2010); Conrad et al. (2009); Xu et al. (2011); Shaikh et al. (2009).
CNV-miR-338	-	19	6	127	9579	1,33%	Perry et al. (2008); Wong et al. (2007); Xu et al. (2011).
CNV-miR-150	-	7	3	52	6659	0,78%	Conrad et al. (2009); Itsara et al. (2009); Shaikh et al. (2009).
CNV-miR-183	-	3	3	15	2008	0,75%	Simon-Sanchez et al. (2007); Jakobsson et al. (2008); Redon et al. (2006); Itsara et al. (2009); Xu et al. (2011).
CNV-miR-1-1	+	12	5	47	8985	0,52%	Park et al. (2010); Jakobsson et al. (2008); Shaikh et al. (2009); Wong et al. (2007).
CNV-let-7g	-	7	4	15	9128	0,16%	Abecasis et al. (2012); Conrad et al. (2009); Xu et al. (2011).
CNV-miR-1275	+	4	3	8	8135	0,10%	Wong et al. (2007); Kidd et al. (2008); Xu et al. (2011).
CNV-miR-499	+	3	3	5	6637	0,08%	Redon et al. (2006); Xu et al. (2011).
CNV-miR-31	-	3	2	3	6804	0,04%	

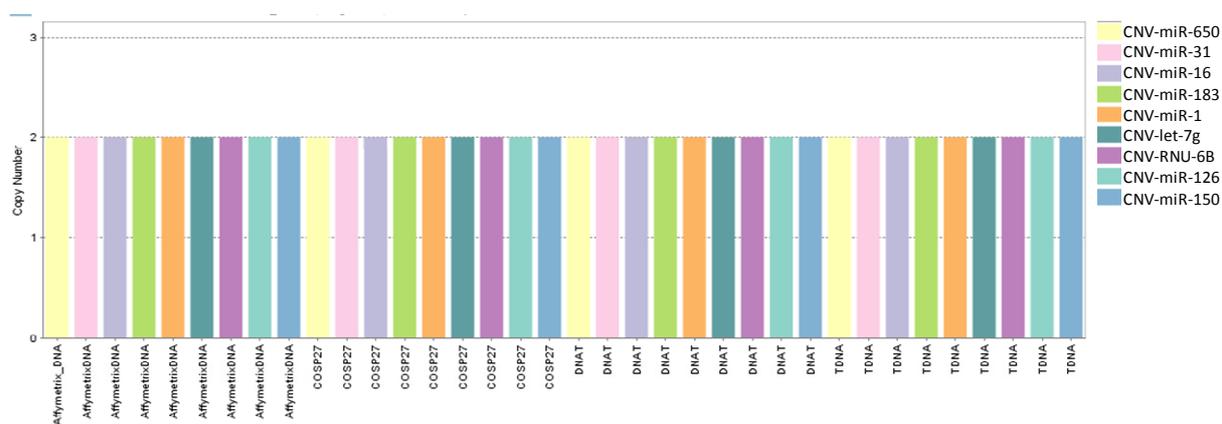
**Tabela 2:** Descrição dos ensaios de microRNA e CNV pelo Genoma de referência NCBI build 37. \* Loco escolhido como controle diploide. †Loco escolhido como controle diploide e como controle endógeno de reação de microRNA. Fonte: o autor.

microRNA	Ensaio miR	Chr	miR-Início	miR-Fim	Ensaio CNV	Posição da sonda
hsa-miR-570	2347	3	195426272	195426368	Hs03487806_cn	195425793
hsa-let-7g	2282	3	52302294	52302377	Hs03459881_cn	52302140
hsa-miR-1275	2840	6	33967749	33967828	Hs06754732_cn	33968165
hsa-miR-183	2269	7	129414745	129414854	Hs04967643_cn	129413527
hsa-miR-126	2228	9	139565054	139565138	Hs02940913_cn	139565397
hsa-miR-31	2279	9	21512114	21512184	Hs03281785_cn	21512259
hsa-miR-338-3p	2252	17	79099683	79099749	Hs03984183_cn	79100616
hsa-miR-150	473	19	50004042	50004125	Hs04012189_cn	50003831
hsa-miR-499b-3p	464911_mat	20	33578203	33578275	Hs01950141_cn	33578082
hsa-miR-1-1	2222	20	61151513	61151583	Hs00892069_cn	61150760
hsa-miR-650	1603	22	23165270	23165365	Hs03322330_cn	23166219
*hsa-miR-16-1	000391	13	50623109	50623197	Hs03846573_cn	50622625
†RNU-6B	001093	10	13259331	13259375	Hs05225739_cn	13259207

### Triagem molecular das regiões genômicas

Para a triagem molecular das regiões selecionadas foi realizada primeiramente a validação dos iniciadores, determinada pela curva padrão de eficiência e respectivos valores de regressão ( $R^2$ ). Uma amostra de DNA reunido de 10 indivíduos com uma concentração inicial de 500ng/ $\mu$ L foi utilizada em diluições seriadas de 1:4. Os valores de eficiência variaram de 97,19% a 109,75% e  $R^2$  entre 0,993 a 0,996.

O ajuste da amostra calibrador definitiva foi realizado com a análise dos dados de RQ, juntamente com o controle endógeno de reação RNaseP, usando a amostra de referência “*Affymetrix GeneChip® reference Genomic DNA*”, com duas cópias dos genes controle e alvos. A análise validou o calibrador como duas cópias para todas as regiões investigadas (Figura 04).



**Figura 4:** Validação da amostra calibrador com cópia diploide para 9 ensaios de CNV, usando uma amostra aleatória (COSP27) e o DNA de referência em dupla análise (DNAT e TDNA). Fonte: O autor.

A triagem molecular das 13 regiões CNV-miRNA foi realizada com sondas de qPCR em 103 indivíduos da população brasileira. Sete das regiões CNV-miRNA, incluindo as putativas regiões monomórficas CNV-miR-16-1 e CNV-RNU-6B, não apresentaram variantes polimórficas, isto é, 100% da amostra com duas cópias (Tabela 3 para frequência absoluta e Tabela 04 para frequência relativa). Enquanto as outras seis regiões apresentaram polimorfismo com a frequência de polimorfismo entre 1,95% (CNV-miR-1275) e máximo de 37,9% (CNV-miR-570) em 103 indivíduos da população Brasileira. As regiões de CNV-miR-570 e CNV-miR-499b-3p foram as mais polimórficas, variando de 1 a 4 cópias, e zero a 6 cópias, respectivamente (Tabela 3 e Tabela 4).

**Tabela 3:** Distribuição da frequência absoluta dos ensaios de miR-CNV de acordo com o número de cópias genômica em 103 indivíduos da população brasileira. Fonte: O autor.

CNV-miRNA	Número de cópias						
	0	1	2	3	4	5	6
CNV-miR-570	0	12	64	25	2	0	0
CNV-miR-499	2	6	81	3	5	4	2
CNV-miR-126	0	2	98	3	0	0	0
CNV-miR-150	0	2	99	2	0	0	0
CNV-miR-338	0	1	99	2	1	0	0
CNV-miR-1275	0	1	101	1	0	0	0
CNV-miR-650	0	0	103	0	0	0	0
CNV-miR-183	0	0	103	0	0	0	0
CNV-let-7g	0	0	103	0	0	0	0
CNV-miR-31	0	0	103	0	0	0	0
CNV-miR-1	0	0	103	0	0	0	0
CNV-miR-16	0	0	103	0	0	0	0
CNV-RNU-6B	0	0	103	0	0	0	0

**Tabela 4:** Distribuição da frequência relativa dos ensaios de miR-CNV de acordo com o número de cópias genômica em 103 indivíduos da população brasileira. Fonte: o autor.

CNV-miRNA	Número de cópias						
	0	1	2	3	4	5	6
CNV-miR-570	0,0	0,117	0,621	0,243	0,019	0,00	0,00
CNV-miR-499	0,019	0,058	0,786	0,029	0,049	0,039	0,019
CNV-miR-126	0,0	0,019	0,951	0,029	0,0	0,0	0,0
CNV-miR-150	0,0	0,019	0,961	0,019	0,0	0,0	0,0
CNV-miR-338	0,0	0,010	0,961	0,019	0,010	0,0	0,0
CNV-miR-1275	0,0	0,010	0,980	0,010	0,0	0,0	0,0
CNV-miR-650	0,0	0,0	1,00	0,0	0,0	0,0	0,0
CNV-miR-183	0,0	0,0	1,00	0,0	0,0	0,0	0,0
CNV-let-7g	0,0	0,0	1,00	0,0	0,0	0,0	0,0
CNV-miR-31	0,0	0,0	1,00	0,0	0,0	0,0	0,0
CNV-miR-1	0,0	0,0	1,00	0,0	0,0	0,0	0,0
CNV-miR-16	0,0	0,0	1,00	0,0	0,0	0,0	0,0
CNV-RNU-6B	0,0	0,0	1,00	0,0	0,0	0,0	0,0

Foram avaliados os dados disponíveis para essa amostragem com fim de detectar alguma interação com o número de cópias. As variáveis disponíveis eram: Gênero, região geográfica e estimativa de ancestralidade genética. Foi identificada uma diferença significativa com relação ao gênero, em que uma maior proporção de mulheres possuiu perda do número de cópias para a CNV-miR-570 ( $p = 0,014$ ) e uma maior proporção de homens possuíram ganho para a CNV-miR-499 ( $p = 0,033$ ) (Tabela 5). Para as regiões geográficas, não foram encontradas diferenças para a CNV-miR-570 ( $p = 0,352$ ). No entanto, para a CNV-miR-499 foi encontrada uma diferença entre as regiões ( $p = 0,01$ ), notoriamente identificado pela região sudeste (Tabela 6). Com dados do perfil de ancestralidade genética da amostra foi correlacionado com número de cópia. Apesar de apresentar pequena dissimilaridade nas estimativas de ancestralidade entre os grupos, não houve diferença estatisticamente significativa (Tabela 7), mesmo quando avaliado com ANOVA multifatorial (Tabela 8) ( $p = 0,113$  para EUR;  $p = 0,081$  para AFR e  $p = 0,088$  para NAM).

Um detalhamento comparativo com as amostras de banco de dados DGV foi realizado para cada uma das regiões genômicas em questão (Tabela 9), mostrando as diferenças entre a frequência total de polimorfismo em todos os estudos ( $n = 15.156$ ) e as frequências de polimorfismo máximo e mínimo, considerando cada estudo individualmente.

**Tabela 5:** Número absoluto de indivíduos por variação do número cópia das regiões CNV-miR-570 e CNV-miR-499 classificados por sexo. Nível de significância em  $\alpha = 0.05$  para o teste Chi-quadrado: \*  $p = 0.001$ ; †  $p = 0.004$ . Fonte: o autor.

Número de cópia	CNV-miR-570 *		CNV-miR-499 †	
	Feminino	Masculino	Feminino	Masculino
Perda	11	1	4	4
Diplóide	29	35	45	36
Ganho	11	16	2	12

**Tabela 6:** Número absoluto de indivíduos por variação do número cópia classificados por das regiões CNV-miR-570 e CNV-miR-499 Região geográfica. Nível. Nível de significância em  $\alpha = 0.05$  para o teste Chi-quadrado: \*  $p = 0,352$ ; †  $p = 0,001$ . Fonte: o autor.

Região	CNV-miR-570 *			CNV-miR-499†		
	Perda	Diplóide	Ganho	Perda	Diplóide	Ganho
Centro-Oeste (n=21)	0	15	6	1	20	0
Nordeste (n=20)	4	9	7	2	18	0
Norte (n=22)	4	15	3	2	19	1
Sudeste (n=22)	3	14	5	0	12	10
Sul (n=18)	1	11	6	3	12	3

**Tabela 7:** Estimativas de proporção de ancestralidade na amostra da população brasileira (média  $\pm$  desvio padrão) diferenciadas por variação do número de cópia das regiões CNV-miR-570 e CNV-miR-499. Proporções para ancestralidade Europeia (EUR), Africana (AFR) e Nativo-Americana (NAM). Teste ANOVA com significância a 5%. Fonte: o autor.

Região	EUR	AFR	NAM
<b>CNV-miR-570</b>	$p = 0,226$	$p = 0,439$	$p = 0,263$
Perda (n=12)	$0.677 \pm 0.256$	$0.225 \pm 0.172$	$0.099 \pm 0.129$
Diplóide (n=64)	$0.735 \pm 0.200$	$0.172 \pm 0.151$	$0.093 \pm 0.116$
Ganho (n=27)	$0.646 \pm 0.284$	$0.207 \pm 0.179$	$0.147 \pm 0.200$
<b>CNV-miR-499</b>	$p = 0,428$	$p = 0,122$	$p = 0,560$
Perda (n=8)	$0.702 \pm 0.297$	$0.137 \pm 0.147$	$0.161 \pm 0.271$
Diplóide (n=81)	$0.718 \pm 0.225$	$0.179 \pm 0.162$	$0.103 \pm 0.129$
Ganho (n=14)	$0.630 \pm 0.236$	$0.264 \pm 0.145$	$0.106 \pm 0.137$
Total (n=103)	$0.705 \pm 0.232$	$0.187 \pm 0.161$	$0.108 \pm 0.144$

**Tabela 8:** Comparação das frequências absolutas do CNV-miR-499 entre regiões geográficas da população brasileira de acordo com a média da ancestralidade. Teste de ANOVA multifatorial com significância de 5%,  $p = 0,113$  para EUR;  $p = 0,081$  para AFR e  $p = 0,088$  para NAM. Fonte: o autor.

<b>Região</b>	<b>CNV-miR-499</b>	<b>n</b>	<b>EUR</b>	<b>AFR</b>	<b>AMR</b>
Centro-Oeste	Perda	1	0,363	0,419	0,218
	Diploide	20	0,617	0,226	0,157
	Ganho	0	.	.	.
Nordeste	Perda	2	0,967	0,016	0,017
	Diploide	18	0,719	0,179	0,102
	Ganho	0	.	.	.
Norte	Perda	2	0,727	0,193	0,080
	Diploide	19	0,666	0,217	0,118
	Ganho	1	0,499	0,273	0,228
Sudeste	Perda	0	.	.	.
	Diploide	12	0,803	0,142	0,055
	Ganho	10	0,596	0,283	0,121
Sul	Perda	3	0,622	0,086	0,292
	Diploide	12	0,886	0,076	0,037
	Ganho	3	0,788	0,198	0,014

**Tabela 9:** Comparação das frequências relativas populacionais das CNV-miRNA entre as amostras do banco de dados DGV (n total = 15.156, máximo e mínimo referente às populações de cada estudo) e a amostra da população brasileira. Fonte: o autor.

CNV-miRNA	DGV				BRA	
	Frequência total	Máximo	Referência	Mínimo		Referência
CNV-miR-570	1,16%	60,78%	de Smith et al. (2007)	0,23%	Xu et al. (2011)	37,86%
CNV-miR-499	0,03%	11,11%	Kidd et al. (2008)	0,02%	Xu et al. (2011)	21,36%
CNV-miR-126	0,71%	5,64%	Jakobsson et al. (2008)	0,32%	Itsara et al. (2009)	4,85%
CNV-miR-150	0,34%	64,52%	Perry et al. (2008)	0,06%	Xu et al. (2011)	3,88%
CNV-miR-338	0,84%	58,06%	Park et al. (2010)	0,05%	Shaikh et al. (2009)	3,88%
CNV-miR-1275	0,05%	0,67%	Conrad et al. (2009)	0,03%	Xu et al. (2011)	1,94%
CNV-miR-1-1	0,31%	4,51%	Jakobsson et al. (2008)	0,13%	Itsara et al. (2009)	0,00%
CNV-let-7g	0,10%	4,21%	Wong et al. (2007)	0,05%	Xu et al. (2011)	0,00%
CNV-miR-183	0,10%	0,59%	Itsara et al. (2009)	0,06%	Shaikh et al. (2009)	0,00%
CNV-miR-31	0,02%	0,37%	Redon et al. (2006)	0,03%	Xu et al. (2011)	0,00%
CNV-miR-650	4,12%	100%	Park et al. (2010)	0,03%	Campbell et al. (2011)	0,00%

Dados de outras 30 amostras com DNA extraído de células do epitélio bucal e de células mononucleadas do sangue (PBMC) foram gerados para 9 das 13 CNV-miRNA investigadas em primeiro momento, devido a parte delas terem apresentado caráter monomórfico na triagem inicial. Ainda assim optou-se por usar os controles invariáveis CNV-miR-16 e CNV-RNU6B. Foi observado que houve discrepância entre o número de cópias de diferentes tecidos celulares (Tabela 10) em um mesmo indivíduo, com eventos que foram de 20% de ocorrência (CNV-miRNA-499 e CNV-miRNA-338 com 6 casos), seguidos por 16,6% (CNV-miRNA-126 com 5 casos), 13,3% (CNV-miRNA-150 com 4 casos) e 10% (CNV-miRNA-570 com 3 casos) (Tabela 11).

**Tabela 10:** Comparação das frequências absolutas na amostra de trabalho em 9 regiões de CNV-miR e entre os tecidos do epitélio bucal e PBMC.

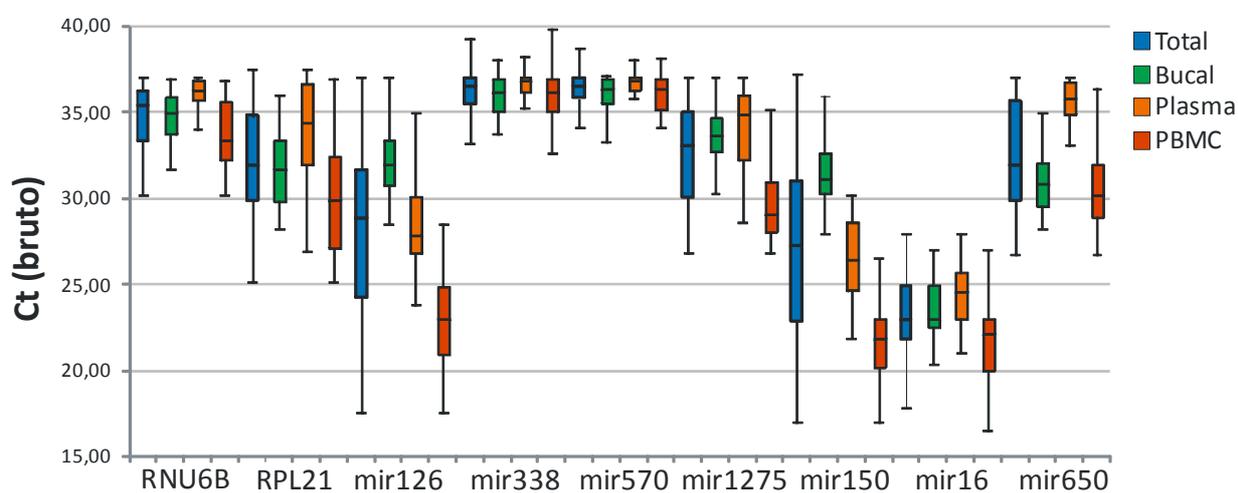
CNV-miR	Tipo celular	Número de cópias					Alterações		Total de alterações
		0	1	2	3	4	Perda	Ganho	
CNV-miR-126	Bucal	0	1	26	3	0	1	3	4
	PBMC	0	3	27	0	0	3	0	3
CNV-miR-338	Bucal	0	4	19	4	3	4	7	11
	PBMC	0	5	21	4	0	5	4	9
CNV-miR-570	Bucal	0	3	17	9	1	3	10	13
	PBMC	0	1	17	11	1	1	12	13
CNV-miR-499	Bucal	0	4	25	1	0	4	1	5
	PBMC	0	3	27	0	0	3	0	3
CNV-miR-1275	Bucal	0	2	28	0	0	2	0	2
	PBMC	0	2	28	0	0	2	0	2
CNV-miR-150	Bucal	0	2	28	0	0	2	0	2
	PBMC	0	2	28	0	0	2	0	2
CNV-miR-650	Bucal	0	0	30	0	0	0	0	0
	PBMC	0	0	30	0	0	0	0	0
CNV-miR-16	Bucal	0	0	30	0	0	0	0	0
	PBMC	0	0	30	0	0	0	0	0
CNV-RNU6B	Bucal	0	0	30	0	0	0	0	0
	PBMC	0	0	30	0	0	0	0	0



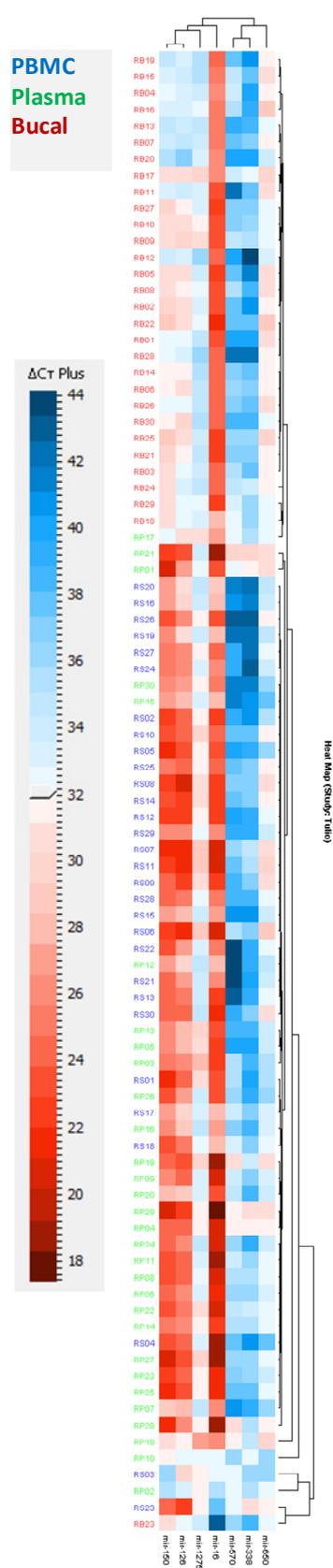
## **Expressão dos microRNA**

A expressão dos microRNA foi avaliada em três constituintes biológicas diferentes: células do epitélio bucal, células mononucleadas do sangue e por fim, do plasma sanguíneo. Para as três amostragens foram realizadas o cálculo de expressão diferencial a partir dos dados brutos (Figura 5). Inicialmente o hsa-miR-499 foi considerado como potencial candidato para avaliar a expressão diferencial, no entanto sua amplificação não foi observada, inviabilizando sua utilização nos tratamentos estatísticos. A diferença de expressão para cada microRNA entre os três constituintes biológicos foi significativa para a maioria das comparações, exceto para bucal x PBMC nos miR-338 ( $p=0,879$ ), miR-570 ( $p=0,826$ ), miR-650 ( $p=0,481$ ). Apesar da diferença identificada nos dados de Ct brutos entre as amostras biológicas, a avaliação dos controles endógenos para a normalização foi aceita dentro dos parâmetros de análise do software, uma vez que os normalizadores são calculados pelo desvio-padrão dos valores de  $\Delta Ct$  entre os controles contra a média dos demais microRNA para gerar uma pontuação de estabilidade.

A verificação de expressão diferencial foi evidenciada no mapa gerado pelos cálculos de agrupamento a partir do  $\Delta Ct$  de cada amostra em cada microRNA. Observa-se no mapa um dendrograma gerado pela interação entre as amostras e a expressão relativa dos microRNA, demonstrando variabilidade entre as amostras (Figura 6). Foi observada diferença entre o epitélio bucal em relação ao plasma e PBMC. Esses dois últimos formaram um único grupo mais homogêneo, porém com uma tendência das amostras biológicas estarem agrupadas com seus pares. O mapa também evidenciou diferença entre os microRNA expressos, em média, acima (miR-338, miR-570 e miR-650) ou abaixo dos controles endógenos (miR-16, miR-1275, miR-126 e miR-150).



**Figura 5:** Boxplot dos níveis de cada microRNA por constituinte biológico (PBMC, Plasma e Células do epitélio bucal) e do total agrupado. Linha em negrito da caixa = mediana dos Ct; parte inferior da caixa = primeiro quartil (Q1); parte superior da caixa = terceiro quartil (Q3); barras = limites inferior e superior. Fonte: o autor.



**Figura 6:** Mapa de expressão relativa dos microRNA. Fonte: O autor

Cada amostra está representada por uma linha horizontal, sendo o grupo biológico identificado pelas cores vermelho (epitélio bucal), verde (plasma) e azul (PBMC), e cada coluna representando um microRNA diferente. A variação de cores do azul para o vermelho representa a variação de Ct com relação aos controles endógenos indo do positivo (azul) para o negativo (vermelho).

### **Correlação microRNA com CNV**

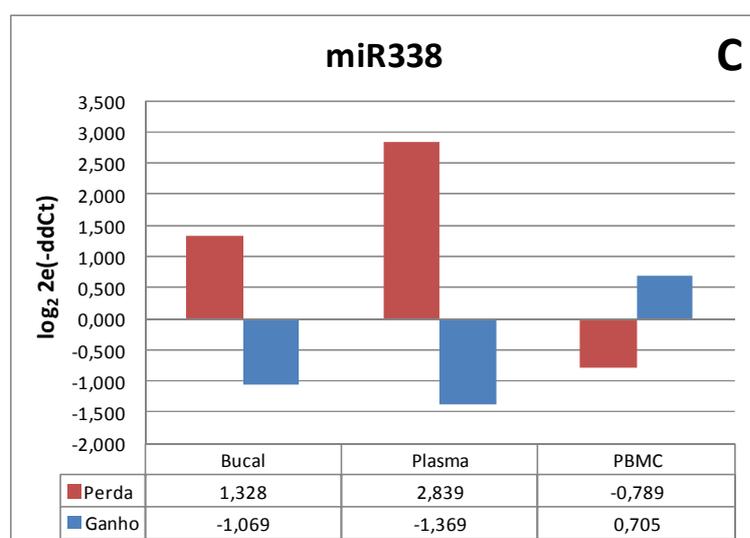
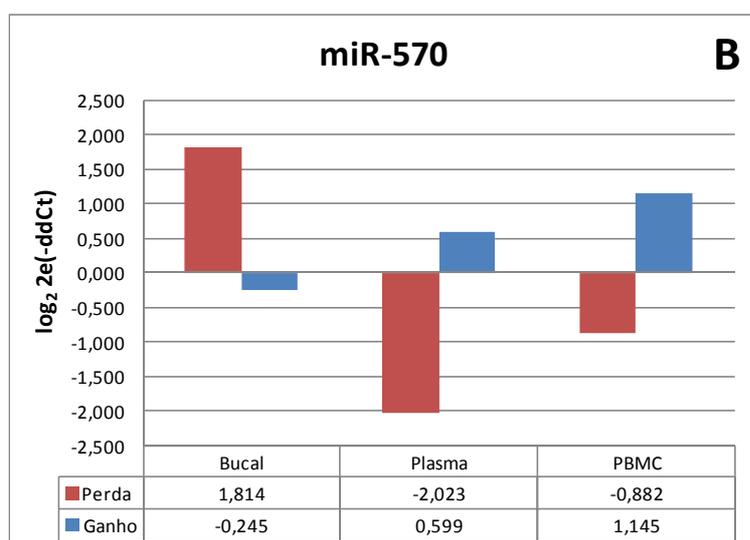
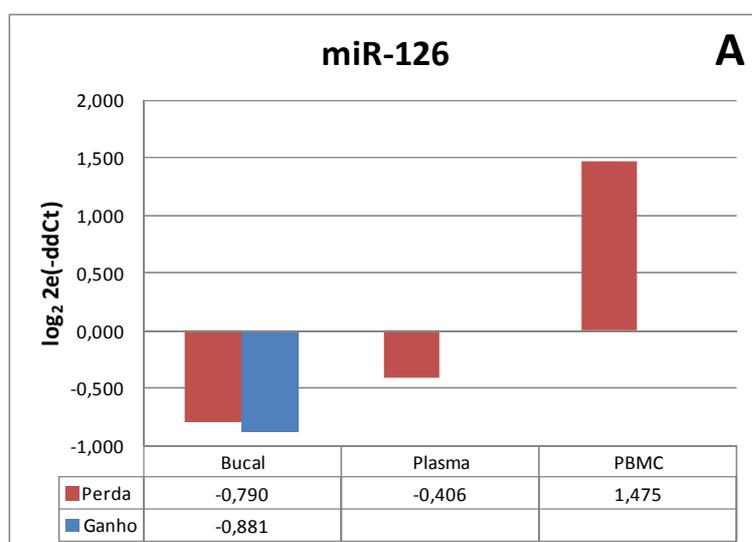
Antes de realizar qualquer teste de associação, os dados foram verificados para o tipo de distribuição de acordo com o teste de probabilidade de distribuição normal de Kolmogorov-Smirnov. Todos os dados contínuos, como os valores de  $\Delta Ct$ , apresentaram distribuição normal ( $p > 0,05$ ), enquanto os discretos, como o número de cópias e as respectivas classes, não assumiram a distribuição normal ( $p < 0,05$ ). Assim, subsequentemente foram aplicados testes não paramétricos para verificar as associações entre CNV e microRNA.

As quantificações relativas tiveram diferentes respostas com relação ao número de cópias e também com ao constituinte biológico (Tabela 12). A situação hipotética de resposta por dosagem foi observada para as células mononucleadas do sangue em todos os microRNA onde houve variação de ganho e perda de região cromossomal (miR-338 e miR-570). Somente para o miR-570 essa situação se repetiu para a fração plasmática do sangue (Figura 07). Teste não paramétrico de Kruskal-Wallis, no entanto, não detectou nenhuma diferença significativa entre a comparação de médias da taxa de mudança (FC) desses microRNA nesses constituintes biológicos (Tabela 13).

Os microRNA que apresentaram somente variação de perda, os miR-150 e miR-1275, tiveram avaliados como mudança de expressão relativa apenas a relação FC entre perda e diploide, sendo considerado uma taxa com incremento de expressão aquela em que o valor de  $\Delta\Delta Ct$  assume valor positivo maior que zero. Este fato ocorreu apenas para a fração plasmática dos dois microRNA (Tabela 12).

**Tabela 12:** Valores da quantificação relativa dos microRNA entre o epitélio bucal, plasma e PBMC, de acordo com a variação em número de cópias. Fonte: o autor.

microRNA	$\Delta Ct$			$\Delta\Delta Ct$		$2^{-\Delta\Delta Ct}$ (FC)		$\log_2(2^{-\Delta\Delta Ct})$	
	Perda	Diploide	Ganho	Perda	Ganho	Perda	Ganho	Perda	Ganho
<b>miR-126 bucal</b>	-0,923	-1,713	-0,833	0,790	0,881	0,578	0,543	-0,790	-0,881
<b>miR-126 plasma</b>	-6,779	-7,185		0,406		0,755		-0,406	
<b>miR-126 pbmc</b>	-10,122	-8,647		-1,475		2,779		1,475	
<b>miR-570 bucal</b>	0,874	2,688	2,933	-1,814	0,245	3,516	0,844	1,814	-0,245
<b>miR-570 plasma</b>	2,030	0,007	-0,592	2,023	-0,599	0,246	1,515	-2,023	0,599
<b>miR-570 pbmc</b>	5,401	4,519	3,374	0,882	-1,145	0,543	2,212	-0,882	1,145
<b>miR-338 bucal</b>	2,709	4,037	5,105	-1,328	1,069	2,510	0,477	1,328	-1,069
<b>miR-338 plasma</b>	-0,332	2,507	3,877	-2,839	1,369	7,155	0,387	2,839	-1,369
<b>miR-338 pbmc</b>	5,666	4,877	4,172	0,789	-0,705	0,579	1,630	-0,789	0,705
<b>miR-150 bucal</b>	-2,423	-2,118	-	-0,305	-	1,235	-	0,305	-
<b>miR-150 plasma</b>	-7,967	-9,008	-	1,041	-	0,486	-	-1,041	-
<b>miR-150 pbmc</b>	-12,050	-9,745	-	-2,305	-	4,942	-	2,305	-
<b>miR-1275 bucal</b>	-0,947	-0,063	-	-0,884	-	1,845	-	0,884	-
<b>miR-1275 plasma</b>	-0,262	-2,139	-	1,877	-	0,272	-	-1,877	-
<b>miR-1275 pbmc</b>	-4,351	-1,765	-	-2,586	-	6,003	-	2,586	-

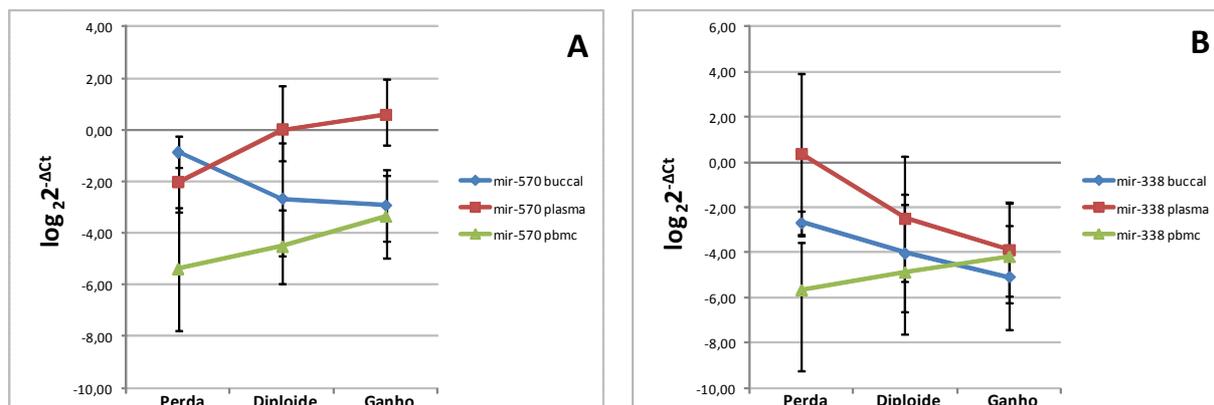


**Figura 7:** Expressão relativa entre perda e ganho, com relação ao diploide, dos microRNA: (A) miR-126, (B) miR-570 e (C) miR-338, de acordo o constituinte biológico. Em (A) apenas a comparação de perda e ganho entre células do epitélio bucal, pois PBMC não houve Fonte: O autor.

**Tabela 13:** Teste de Kruskal-Wallis (com Qui-quadrado representando o parâmetro *H-statistic*), para comparação de médias da taxa de mudança (FC) dos microRNA de acordo com a variação em número de cópias do epitélio bucal e PBMC. Significância estatística a 5%.

microRNA	Bucal_CN		PBMC_CN	
	Qui-quadrado	<i>p</i> -valor	Qui-quadrado	<i>p</i> -valor
<b>miR-570 bucal</b>	3,947	0,139	1,819	0,403
<b>miR-570 plasma</b>	0,366	0,833	2,585	0,275
<b>miR-570 PBMC</b>	1,598	0,450	1,897	0,387
<b>miR-126 bucal</b>	1,819	0,403	0,747	0,388
<b>miR-126 plasma</b>	2,585	0,275	0,345	0,557
<b>miR-126 PBMC</b>	1,897	0,387	1,301	0,254
<b>miR-338 bucal</b>	3,875	0,144	0,028	0,868
<b>miR-338 plasma</b>	4,414	0,110	0,062	0,803
<b>miR-338 PBMC</b>	1,124	0,570	3,657	0,056

Para entender melhor o comportamento da expressão de acordo com o número de cópias, foi traçado um gráfico com a taxa de mudança em relação aos genes constituintes, o equivalente ao  $\log_2 2^{(-\Delta Ct)}$  (Figura 08). Sugere-se que a inclinação da curva seja condizente com a taxa de mudança de expressão.



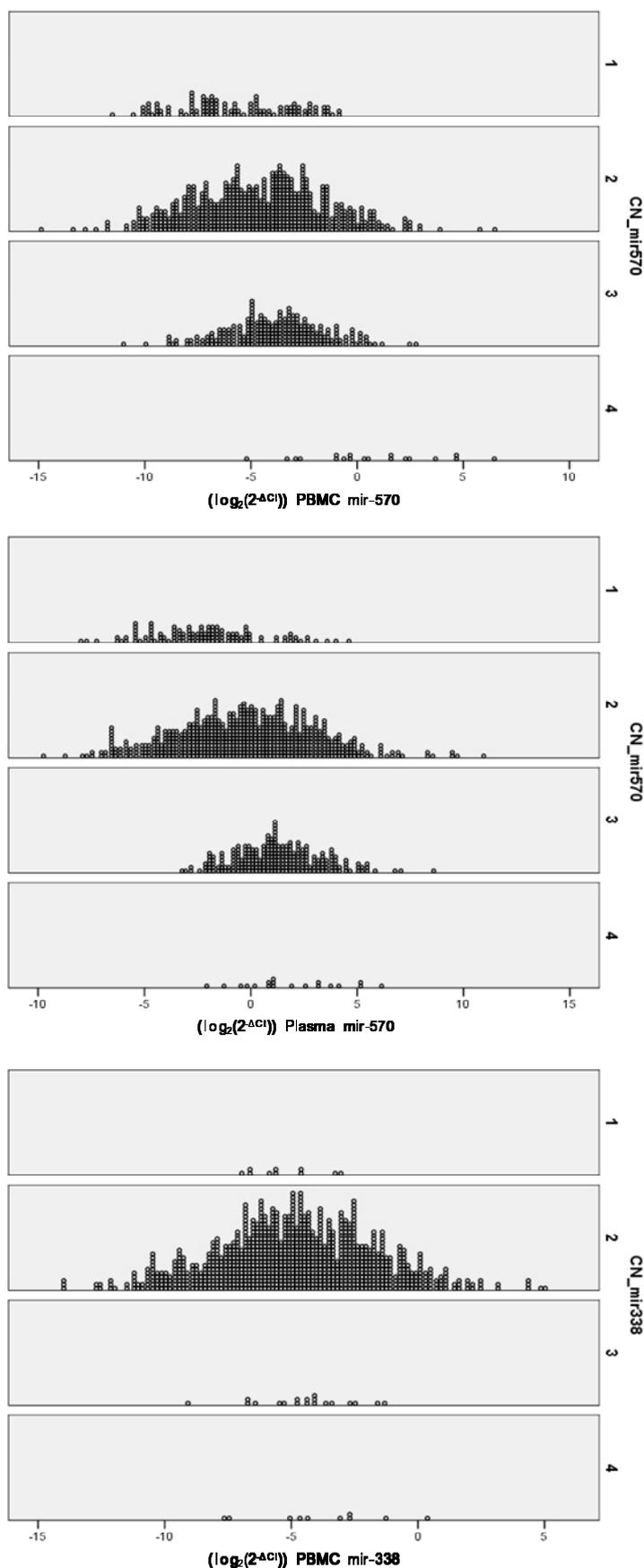
**Figura 8:** Padrão de expressão relativa entre as diferentes classes de número de cópia dos microRNA: (A) miR-570 e (B) miR-338, de acordo o constituinte biológico. Fonte: O autor.

### Simulação Monte Carlo

O modelo de simulação estocástica de Monte Carlo foi estabelecido com base nos dados reais de expressão relativa entre as diferentes classes de número de cópia dos microRNA miR-570 e miR-338. Os dados gerados para 1.000 e 10.000 mostraram não ter diferença significativa entre eles ( $p > 0,05$ ), portanto os dados foram gerados e analisados com o menor valor possível até atingir significância estatística.

Os valores atribuídos na simulação foram distribuídos discretamente assumindo uma função normal (Figura 09), com medidas de distribuição da expressão relativa dos microRNA gerados para 1.000 indivíduos (Tabela 13). Para interação com  $n = 1.000$  foram encontradas diferenças significativas no microRNA miR-570 (plasma e PBMC). Enquanto para o miR-338, que precisou de sucessivas avaliações com número maior de interações, porém todas não significativas:  $n = 10.000$  ( $p = 0,572$ ),  $n = 100.000$  ( $p = 0,561$ ) e  $n = 1.000.000$  ( $p = 0,563$ ).

Os dados plotados em gráfico permitiram estimar uma equação geral da reta em função da distribuição dos pontos (Figura 10). Os ajustes da curva por regressão ( $R^2$ ) não permitiram uma aproximação matemática homogênea entre os microRNA e os respectivos constituintes biológicos para determinar o comportamento geral de expressão de acordo com o número de cópias. Esse fato ocorreu devido a algumas curvas apresentarem melhor ajuste para o modelo linear (miR-338 PBMC), outra para o logarítmico (miR-570 Plasma), enquanto uma (miR-570 PBMC) não teve ajuste suficientemente adequado para os modelos propostos, com  $R^2$  menor que 0,95 (Tabela 14).

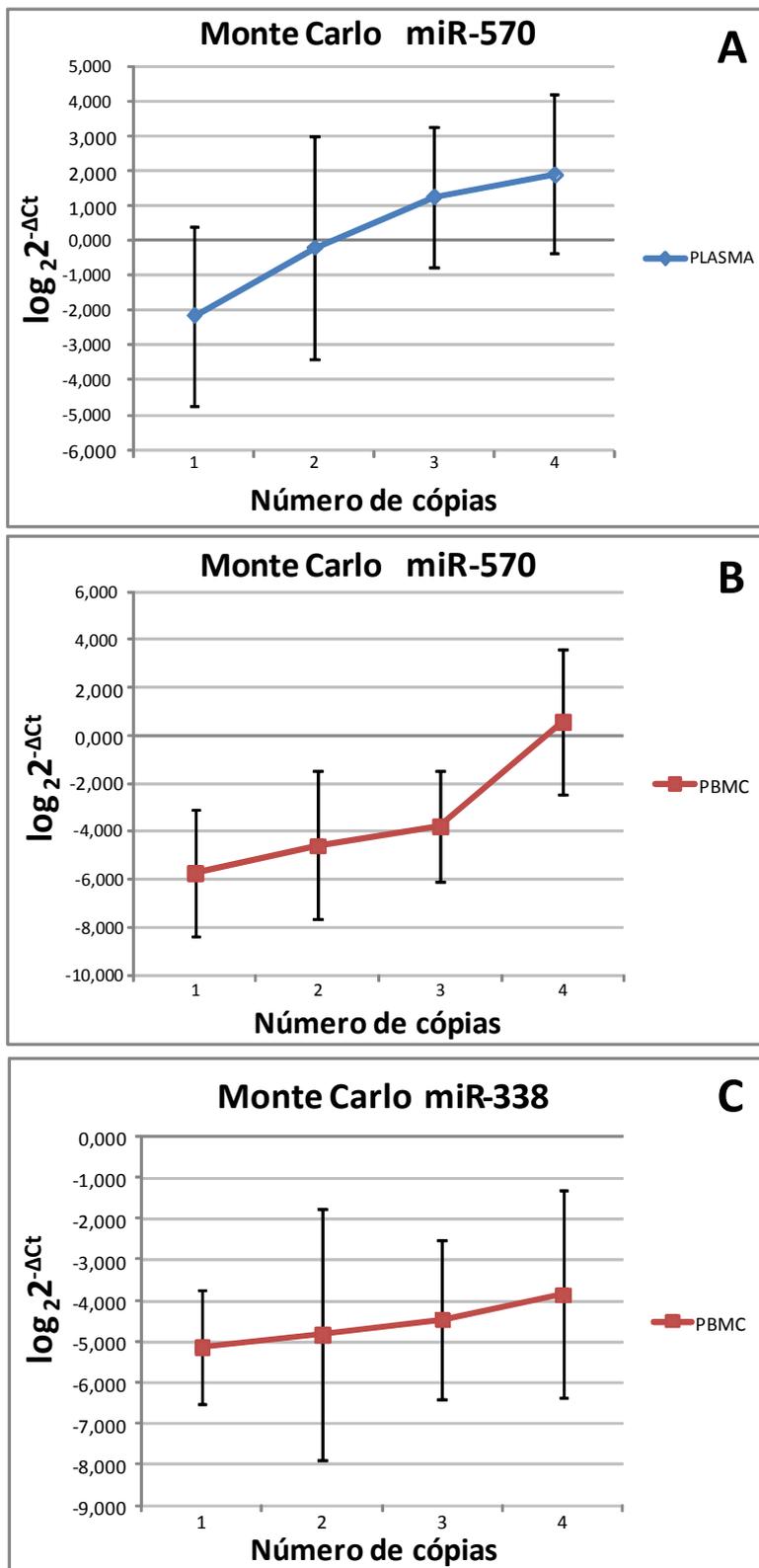


**Figura 9:** Distribuição da expressão diferencial ( $\log_2(2^{-\Delta Ct})$ ) gerada por simulação Monte Carlo ( $n = 1.000$ ) entre as diferentes classes de número de cópia dos microRNA miR-570 (Plasma e PBMC) e miR-338 (PBMC). Fonte: O autor.

Cada ponto representa valores da simulação de expressão diferencial de microRNA de um indivíduo para o número de cópia atribuído nas respectivas caixas.

**Tabela 14:** Dados da simulação Monte Carlo para os microRNA de acordo com a frequência populacional da variação do número de cópias estimada nos 103 indivíduos da população brasileira e expressão diferencial ( $\log_2(2^{-\Delta Ct})$ ) no plasma e PBMC. Teste de Kruskal-Wallis com significância a 5%. \**p*-valor para PBMC; †*p*-valor para Plasma.

CNV-miR	Número de cópia	Frequência populacional	n (total = 1000)	$\log_2(2^{-\Delta Ct})$ PBMC	$\log_2(2^{-\Delta Ct})$ Plasma	p-valor
CNV-miR-570	1	0,1165	117	-2,142 ± 2,573	-5,718 ± 2,655	< 0,0001*
	2	0,6214	621	-0,191 ± 3,188	-4,592 ± 3,092	
	3	0,2427	243	1,277 ± 2,010	-3,793 ± 2,275	< 0,0001†
	4	0,0194	19	1,912 ± 2,292	0,577 ± 3,033	
CNV-miR-338	1	0,0097	10	-5,140 ± 1,373	-	0,556
	2	0,9612	961	-4,833 ± 3,059	-	
	3	0,0194	19	-4,470 ± 1,952	-	
	4	0,0097	10	-3,850 ± 2,531	-	



**Figura 10:** Modelo simulado para expressão relativa entre as diferentes classes de número de cópia dos microRNA: (A) miR-570 plasma, (B) miR-570 PBMC e (C) miR-126 PBMC. Fonte: O autor.

**Tabela 15:** Equações previstas para expressão relativa de microRNA ( $y$ ) em função do número de cópias ( $x$ ) para os modelos de simulação em cada situação, com respectivo valor de ajuste ( $R^2$ ) em regressão.

<b>microRNA</b>	<b>Equação linear</b>	<b>Ajuste (<math>R^2</math>)</b>	<b>Equação logarítmica</b>	<b>Ajuste (<math>R^2</math>)</b>
miR-570 Plasma	$y = 1,363x - 3,1935$	0,9548	$y = 2,9894\ln(x) - 2,161$	0,9959
miR-570 PBMC	$y = 1,9684x - 8,3024$	0,8510	$y = 3,8706\ln(x) - 6,4566$	0,7135
miR-338 PBMC	$y = 0,4233x - 5,6313$	0,9711	$y = 0,8631\ln(x) - 5,2589$	0,8757

## DISCUSSÃO

A investigação de microRNA na modulação de fenômenos biológicos tem adquirido grandes interesses da pesquisas de fenômenos patológicos e não patológicos. A regulação da sua biogênese e atividade, no entanto, ainda não enfoca totalmente na análise molecular da organização cromossomal ou variações estruturais, mas têm concentrado seus esforços em mutações específicas, tais como SNP localizado nas regiões genômicas de pri-miRNA (DUAN, R. et al., 2007; DUAN, S. et al., 2009), e dentro das regiões não regulatórias de genes alvo, como a 3'UTR, alterando a atividade dos microRNA no reconhecimento dos seus sítios de ligação (SANDBERG et al., 2008). Esta Tese centrou-se nas variações estruturais do DNA em locos de microRNA que variam em número de cópias. Os resultados aqui apresentados validaram algumas CNV-miRNA tomadas *a priori* como polimórficas, além de confirmar cópias inalteradas de regiões monomórficas putativas na população brasileira e também avaliou o comportamento da expressão de acordo com os polimorfismos observados.

A partir de 11 regiões putativas, 6 foram confirmadas como polimórfica, das quais 2 mostraram relativa variabilidade. Uma análise prévia de bioinformática de regiões de CNV-miRNA, utilizando o banco de dados DGV (MARCINKOWSKA et al., 2011) demonstrou que a fração de locos polimórficos foi duas vezes menor do que o esperado, sugerindo que pelo menos as CNV altamente polimórficas podem estar sobre seleção negativa, e que uma parte significativa das que estão depositadas no DGV são raras, ou podem representar falsos positivos (ITSARA et al., 2009; MARCINKOWSKA et al., 2011). O fato de algumas regiões terem sido identificadas como monomórficas na amostra estudada da população brasileira

não indica necessariamente que estas sejam regiões invariáveis, uma vez que a frequência de ocorrência pode chegar a 1 em 5000 ou mais, como foi o caso de algumas CNV descritas nos bancos de dados (Tabela 9). Desta maneira, é importante que estudos possam validar essas regiões utilizando diferentes populações a fim de averiguar a sua utilidade em estudos genéticos.

Por ocasião da elaboração do projeto referente a este estudo, inicialmente existiam somente 14 estudos depositados no DGV (MARCINKOWSKA et al., 2011). Subsequentemente, com o aumento da casuística, novas variantes foram identificadas e descritas, alterando a frequência das CNV previamente relatadas. Este cenário tem se mostrado promissor para os estudos de genética a fim de esclarecer o papel das variações genômicas na composição de fenótipos e na regulação dos fenômenos biológicos. É importante mencionar, que muitas das CNV-miRNA identificadas como altamente polimórficas por análise bioinformática (MARCINKOWSKA et al., 2011), não possuíam ensaios TaqMan® disponíveis para análise de CNV. Portanto, ainda existe um contingente significativo de CNV que constituem potenciais locos polimórficos para serem investigados.

Diferenças na frequência do número de cópias das regiões polimórficas do genoma são importantes para estudos evolutivos envolvendo modelos de seleção genética entre diferentes populações (CAMPBELL et al., 2011; ITSARA et al., 2009; SEBAT et al., 2004; SUDMANT et al., 2010), com respectiva procura de possíveis candidatos com impacto funcional na expressão gênica (CONRAD, D. F. et al., 2010; MANDEL et al., 2010; PERRY et al., 2007; STRANGER et al., 2007). No entanto, até o momento pouco se sabe sobre os efeitos da CNV na expressão dos miRNA. Além disso, a variação das CNV entre populações

de diferentes origens geográficas, especialmente aquelas com processo de miscigenação ainda é pouco explorada. Até o momento, somente um CNV-miRNA foi descrito, o miR-720 (PRONOLD et al., 2012) localizado em 3q26.1 com variações depositadas no DVG (XU et al., 2011). As medidas quantitativas de ancestralidade não sugeriram nenhuma correlação com o número de cópias a partir da análise de dados em CNV-miRNA na amostra brasileira.

Por outro lado, uma diferença estatística foi observada para o gênero, quando considerada como variável independente. Recentemente, foi identificado um viés sexual na variação do número de cópias relacionado a ancestralidade africana em amostras europeias (SHADRAVAN, 2013), sugerindo um papel evolutivo funcional para o número de cópias de genes da família de receptores olfativos. No caso da população brasileira, os testes estatísticos não mostraram evidências consistentes de diferença por estratificação pela ancestralidade. No entanto, foram observadas diferenças com relação ao gênero e também a região geográfica. Dado que a população brasileira possui uma intensa história de miscigenação cultural e de populações (LINS et al., 2010), e que esse fenômeno não foi observado na reprodução com as demais 30 amostras, seria necessário investigar em futuros estudos se essas regiões são propensas a efeitos de seleção positiva ou se os resultados observados são simplesmente efeitos de amostragem.

Outra observação que chama atenção para diferentes resultados foi o fato da CNV-miR-650 ter sido identificada como altamente polimórfica em diversos estudos e não ter sido identificado polimorfismo em 133 indivíduos da população brasileira. Uma possível explicação está na posição da sonda TaqMan® utilizada na análise, que, distante cerca de 850 pb do gene, poderia se localizar fora do ponto de quebra da CNV, não detectando as

variações. No entanto, a diversidade de variantes descritas no banco de dados com relação às suas extensões (entre 100 e 800 kb para o miR-650) sugere que a situação pode ser outra, como a taxa de concordância entre genomas, o uso indiscriminado de padrões de referência e diferentes métodos de bioinformática usados nas análises.

Por definição, CNV é considerada uma variação estrutural em um indivíduo em comparação a um genoma de referência (IAFRATE et al., 2004; LEVY et al., 2007; SCHERER et al., 2007; SEBAT et al., 2004). A sequência padrão de DNA de referência do genoma humano passa por constantes e sucessivas atualizações em curadoria de um consórcio próprio. Porém, é derivada, na sua maior parte, de um único doador, com o resto da sequência proveniente de um mosaico de outras fontes. Esse modelo original de um único indivíduo não é suficiente para representar a complexidade de regiões com diversidade alélica e estrutural, incluindo as lacunas de complexas deleções, duplicações e repetições em tandem (SCHERER et al., 2007). A definição de grupos controles pelo fenótipo é essencial para evitar possíveis vieses, necessitando cada vez mais métodos criteriosos de fenotipagem.

Desta maneira, ter uma taxa de concordância entre dois genomas pode ser confuso como, por exemplo, na distinção entre a perda em uma amostra de teste de um ganho em uma amostra de referência (SCHERER et al., 2007). Além disso, não existe uma padronização para uso de uma única amostra de referência devido principalmente às limitações técnicas de cada tipo de caracterização de CNV, pois em alguns estudos foram usados conjuntos de várias amostras (IAFRATE et al., 2004; SEBAT et al., 2004), já outros utilizaram apenas uma (LOCKE et al., 2006; REDON et al., 2006; SHARP et al., 2005; WONG,

K. K. et al., 2007), enquanto o mais apropriado seria usar a sequência de referência do genoma humano (HINDS et al., 2006). Essas limitações técnicas também são de suma importância para identificar variantes em uma mesma amostra. Usando diferentes plataformas experimentais e algoritmos de análise pesquisadores chegaram a valores de até 80% de concordância para uma mesma amostra repetida no mesmo equipamento (KOIKE et al., 2011). Portanto, é necessário confirmar através de outras técnicas, e em outras amostras, a particularidade da falta de polimorfismo nessa região da CNV-miR-650.

Nesse mesmo sentido, na tentativa de estimar a taxa de falsos-positivos entre CNV, tem sido notada a variação em diversas situações como em irmãos monozigóticos (BRUDER et al., 2008) e também em diferentes tecidos de um mesmo indivíduo (PIOTROWSKI et al., 2008). O polimorfismo intraindividual, ou mosaicismosomático, pode levar a graves consequências em um indivíduo, caso a célula ou grupo de células, não sofra apoptose e propague a mutação implicando em várias mudanças na fisiologia celular. Em casos específicos, o mosaicismosomático de regiões genômicas de grandes extensões podem modificar processos de controle de transcrição gênica ou da sinalização celular, acarretando um espectro de fenótipos implicados na fisiologia, desde alterações no neurodesenvolvimento normal e patológico, até o processo de envelhecimento (PIOTROWSKI et al., 2008; PODURI et al., 2013).

As alterações entre células mononucleadas do sangue e do epitélio bucal neste estudo foram aceitas como verdadeiras, mas serão reanalisadas utilizando outras tecnologias, como os arranjos baseados em SNP, como parte de projetos em andamento. Ainda que a diferença tenha sido pequena, o epitélio bucal apresentou um maior número de

alterações com relação às células mononucleadas (Tabelas 10 e 11), e a comparação de CNV com a expressão de microRNA foi realizada pelo respectivo tecido. Muito embora o mosaicismo somático seja frequentemente associado a doenças (BIESECKER; SPINNER, 2013; BRUDER et al., 2008; PODURI et al., 2013), é cada vez mais comum o relato em células somáticas de indivíduos sem processos patológicos (DE, 2011). Essa constatação, no entanto possui um fator que pode ser crucial para a análise de CNV. Em geral, estima-se que aproximadamente 30% das células de fibroblastos possuem variações somáticas nos seus genomas (ABYZOV et al., 2012). Além disso, o mosaicismo pode surgir como um evento em um aglomerado de células em regiões específicas de um tecido ou mesmo em eventos únicos em células individuais (MCCONNELL et al., 2013).

Sendo a qPCR um método analítico sensível às concentrações iniciais de material biológico, a presença não homogênea de mosaicismo somático pode gerar ruídos quantitativos, indicando resultados falsos positivos e/ou negativos (WEAVER et al., 2010). Esse fato tem uma extrema implicação em estudos que utilizam DNA de um tecido (preferencialmente sangue e epitélio bucal, cuja extração é simples e eficiente), mas avaliam características fenotípicas ou de expressão gênica de outro, prática essa que extremamente comum na pesquisa científica devido ao difícil acesso aos tecidos de interesse e dificuldades (éticas, legais e técnicas) de realizar biopsias. Como a expressão de microRNA é diferenciada entre tecidos (LIANG et al., 2007) esse impacto deve ser levado em consideração, podendo refletir incongruência nos resultados de estudos de associação genética a fenômenos patológicos.

Uma característica observada na presente casuística foi o fato de uma região polimórfica, a CNV-miR-499, não ter apresentado expressão do seu microRNA nos materiais biológicos pesquisados. O perfil da expressão de microRNA em diferentes tecidos e constituintes biológicos, como os fluidos corporais, já foi descrito em diversos organismos (KIM, V. N.; NAM, 2006), inclusive em humanos (LIANG et al., 2007; WEBER et al., 2010), sendo alguns expressos exclusivamente em determinados tecidos (LIANG et al., 2007). O hsa-mir-499 tem expressão exclusiva no tecido muscular cardíaco, e não pode ser detectado como microRNA circulante em plasma, exceto em indivíduos com infarto do miocárdio. Em estudos prévios, demonstrou-se que a expressão do hsa-miR-499 era elevada em infarto agudo do miocárdio, miocardite viral e insuficiência cardíaca aguda (CHENG et al., 2014; CORSTEN et al., 2010). Assim como o hsa-miR-499 pode ser utilizado como biomarcador sob tais condições patológicas, é necessário investigar indivíduos com histórico pessoal e familiar de infarto e doenças coronarianas a fim de elucidar o papel do polimorfismo de número de cópias e a possível correlação com a expressão do hsa-miR-499, colocando a CNV como potencial biomarcador genético. Ainda, a associação detectada entre gênero e a frequência do número de cópias da CNV-miR-499 na triagem amostral da população brasileira é também sugestiva de futuras pesquisas, uma vez que mulheres são mais propensas de apresentar episódios de infarto do miocárdio sem dor no peito e com maior mortalidade do que os homens {Canto, 2012 #598}.

A consequência fenotípica da duplicação de grandes regiões ou mesmo cromossomos inteiros é, pelo menos em parte, determinada pela extensão de desequilíbrio regulatório podendo oscilar a dosagem de vias de transdução de sinal, fluxo metabólico, ou processos de comunicação celular (VEITIA et al., 2008). Na teoria clássica de balanço de

dosagem, é esperado que um gene duplicado (que apresente três cópias por genoma diploide) resulte em um aumento de 1,5 vezes da expressão de mRNA. No entanto, experimentalmente observa-se que existem diversos mecanismos que contradizem a hipótese clássica, seja pelo efeito estequiométrico de concentração de produto, seja por haploinsuficiência ou por outros mecanismos compensatórios de regulação (CONRAD, B.; ANTONARAKIS, 2007; VEITIA et al., 2008). Portanto, o efeito do desequilíbrio de dosagem gênica é influenciado pela taxa de síntese e degradação dos produtos envolvidos, sejam proteínas ou mRNA, em mecanismos complexos diversos que não atendem a um único modelo (CONRAD, B.; ANTONARAKIS, 2007; GOUT et al., 2010; KONDRASHOV; KOONIN, 2004).

De maneira geral, especificamente em genes que estão localizados dentro de CNV, foi observado previamente que a expressão média em deleções era menor em comparação com duplicações, com pequenas diferenças em ordem de magnitude (SCHUSTER-BOCKLER et al., 2010). Além disso, existiram casos em que as deleções aparentemente resultaram no aumento da expressão e vice-versa (SCHUSTER-BOCKLER et al., 2010). Nesse sentido, corroborando as observações de Schuster-Bockler e colaboradores, (2010), a expressão de microRNA não seguiu uma regularidade entre os diferentes casos, mas utilizando algoritmos de simulação Monte Carlo, a taxa de produto entre as diferentes classes de número de cópia ficou entre 0,42 e 2,54 para o modelo linear e entre 0,86 e 4,74 para o logarítmico.

A compensação de dosagem possui diversos mecanismos, incluindo a regulação da expressão gênica pela organização da estrutura tridimensional da cromatina (GIRTON;

JOHANSEN, 2008). Os microRNA possuem semelhanças com genes codificadores de proteínas na sua regulação da transcrição. Evidências de estudos biológicos demonstram desde a transcrição pela RNA Polimerase II (LEE, Y. et al., 2004), além de um conjunto de fatores de transcrição associados à RNAPII regulando a expressão de miRNA (CARTHEW; SONTHEIMER, 2009). Além disso, a regulação da transcrição de miRNA também envolve promotores e rede de interações espaciais, cobrindo quase um terço de todo miRNAoma. Os ncRNA longos e curtos estão correlacionados com o complexo de recrutamento de cromatina para determinados locos do genoma para alterar a sua estrutura e promover a expressão gênica (CHEN et al., 2013; ZHANG, R. et al., 2011). Ainda se sabe pouco sobre esse efeito nos microRNA e se as CNV contendo microRNA interferem de alguma maneira nessas interações espaciais da cromatina e nas redes de regulação por microRNA (CHEN et al., 2013).

O único estudo que mediu de forma prática a quantidade de microRNA com relação ao número de cópias genômicas envolveu o estudo de indivíduos com trissomia do cromossomo 21 (KUHN et al., 2008), no qual as taxas de expressão relativa foram entre 0,5 a 2 vezes maiores entre portadores da síndrome e controle em amostras de hipocampo e do coração. Não existem na literatura outros estudos que descrevem amostras não patológicas com o mesmo perfil de avaliação entre CNV e microRNA. Como exemplo, estudos de indivíduos com deficiência intelectual (QIAO et al., 2013) e com autismo (VAISHNAVI et al., 2013) apenas se limitaram a regiões de CNV, aprofundando os resultados em bioinformática para possíveis microRNA que estariam presentes nessas regiões, mas sem apurar a quantificação dos mesmos. Entretanto, em câncer, estudos apontam para os mesmos indicativos, no qual duplicações de regiões contendo genes de microRNA também

expressam respectivamente maiores quantidades de seu produto (AN et al., 2013; CALIN et al., 2002; FANCIULLI et al., 2010; NINOMIYA et al., 2012; SLATTERY et al., 2011). A expressão de microRNA com CNV também foi relacionada em pacientes portadores de doença cardíaca congênita, no qual os microRNA hsa-miR-570 e hsa-miR-650, ambos avaliados nesta Tese, se mostraram candidatos promissores na rede de regulação pós-transcricional de proteínas associadas à doença (XING et al., 2013).

Pela complexidade da temática abordada nesta Tese, diversos aspectos devem ser levados em consideração para a interpretação desses dados. Tópicos como flutuações de expressão intraindividual e entre diferentes tecidos, a funcionalidade de microRNA na regulação da expressão gênica e respectivo efeito no controle de ciclos celulares são apenas alguns fatores que destacam a importância e necessidade deste trabalho precursor. As evidências levantadas nesta Tese sugerem que o número de cópias de uma região contendo genes miRNA influencia positivamente na sua respectiva expressão, conforme a hipótese inicial. Entretanto, por se tratar de um fenômeno biológico pesquisado *in vivo*, os modelos utilizados tiveram respostas variadas. Portanto, serão necessários futuros estudos para corroborar esses achados com números mais expressivos de microRNA e regiões polimórficas, assim como no número amostral e outros tecidos e fluidos corporais. Ainda, experimentos controlados, usando células-tronco em cultivo celular, por exemplo, com a finalidade de aumentar as evidências e consolidar essa afirmação, abrindo uma perspectiva de novas pesquisas envolvendo o controle da expressão gênica sobre a variabilidade fenotípica comum e patológica.

## CONCLUSÕES

Até a presente data, este é o primeiro estudo a investigar a relação direta entre a expressão de microRNA em regiões genômicas de variação no número de cópias em humanos sem características patológicas.

Essas regiões foram previamente assinaladas por bioinformática e por estudos de microarranjos de SNP e CGH, mas somente em alguns casos foram validadas como polimórficas na população brasileira. Para as demais regiões consideradas monomórficas, pode ser necessário uma maior amostragem para confirmar se as CNV detectadas podem ser raras ou mesmo representar falsos positivos de outros estudos.

Segundo os dados relatados aqui, existem indícios da diferença na expressão de microRNA em regiões que variam em número de cópias. Essa expressão não parece ocorrer da mesma maneira, nem com a mesma magnitude, em diferentes tecidos ou constituintes biológicos. Porém, os dados devem ser olhados com cautela, uma vez que a amostragem de casos foi pequena e, portanto, requer uma investigação de maior porte, tanto com uma amostragem mais significativa em relação ao número de cópias, quanto como utilizando tecnologias de sequenciamento em alta escala de genoma completo e também de sequenciamento global de microRNA.

Ainda são necessários estudos funcionais para identificação da interação desses microRNA com a regulação da expressão dos seus genes alvos e assim constituir uma rede integrativa entre genoma, transcriptoma e proteoma.

## REFERÊNCIAS

ABECASIS, G. R.; AUTON, A.; BROOKS, L. D.; DEPRISTO, M. A.; DURBIN, R. M.; HANDSAKER, R. E.; KANG, H. M.; MARTH, G. T.; MCVEAN, G. A. An integrated map of genetic variation from 1,092 human genomes. **Nature**, v.491, n.7422, p.56-65, Nov 1, 2012.

ABYZOV, A.; MARIANI, J.; PALEJEV, D. et al. Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells. **Nature**, v.492, n.7429, p.438-42, Dec 20, 2012.

AHN, S.; WANG, R. T.; PARK, C. C.; LIN, A.; LEAHY, R. M.; LANGE, K.; SMITH, D. J. Directed mammalian gene regulatory networks using expression and comparative genomic hybridization microarray data from radiation hybrids. **PLoS Comput Biol**, v.5, n.6, p.e1000407, Jun, 2009.

AMBROS, V.; BARTEL, B.; BARTEL, D. P. et al. A uniform system for microRNA annotation. **RNA**, v.9, n.3, p.277-9, Mar, 2003.

AN, J.; PAN, Y.; YAN, Z.; LI, W.; CUI, J.; YUAN, J.; TIAN, L.; XING, R.; LU, Y. MiR-23a in amplified 19p13.13 loci targets metallothionein 2A and promotes growth in gastric cancer cells. **J Cell Biochem**, v.114, n.9, p.2160-9, Sep, 2013.

ARMENGOL, L.; VILLATORO, S.; GONZALEZ, J. R.; PANTANO, L.; GARCIA-ARAGONES, M.; RABIONET, R.; CACERES, M.; ESTIVILL, X. Identification of copy number variants defining genomic differences among major human groups. **PLoS One**, v.4, n.9, p.e7230, 2009.

B-RAO, C. Sample size considerations in genetic polymorphism studies. **Hum Hered**, v.52, n.4, p.191-200, 2001.

BAILEY, J. A.; YAVOR, A. M.; MASSA, H. F.; TRASK, B. J.; EICHLER, E. E. Segmental duplications: organization and impact within the current human genome project assembly. **Genome Res**, v.11, n.6, p.1005-17, Jun, 2001.

BALLIF, B. C.; THEISEN, A.; COPPINGER, J. et al. Expanding the clinical phenotype of the 3q29 microdeletion syndrome and characterization of the reciprocal microduplication. **Mol Cytogenet**, v.1, p.8, 2008.

BARBER, J. C.; JOYCE, C. A.; COLLINSON, M. N. et al. Duplication of 8p23.1: a cytogenetic anomaly with no established clinical significance. **J Med Genet**, v.35, n.6, p.491-6, Jun, 1998.

BEJERANO, G.; PHEASANT, M.; MAKUNIN, I.; STEPHEN, S.; KENT, W. J.; MATTICK, J. S.; HAUSSLER, D. Ultraconserved elements in the human genome. **Science**, v.304, n.5675, p.1321-5, May 28, 2004.

BENTLEY, D. R.; BALASUBRAMANIAN, S.; SWERDLOW, H. P. et al. Accurate whole human genome sequencing using reversible terminator chemistry. **Nature**, v.456, n.7218, p.53-9, Nov 6, 2008.

BEREZIKOV, E.; GURYEV, V.; VAN DE BELT, J.; WIENHOLDS, E.; PLASTERK, R. H.; CUPPEN, E. Phylogenetic shadowing and computational identification of human microRNA genes. **Cell**, v.120, n.1, p.21-4, Jan 14, 2005.

BIESECKER, L. G.; SPINNER, N. B. A genomic view of mosaicism and human disease. **Nat Rev Genet**, v.14, n.5, p.307-20, May, 2013.

BOTTONI, A.; PICCIN, D.; TAGLIATI, F.; LUCHIN, A.; ZATELLI, M. C.; DEGLI UBERTI, E. C. miR-15a and miR-16-1 down-regulation in pituitary adenomas. **J Cell Physiol**, v.204, n.1, p.280-5, Jul, 2005.

BRUDER, C. E.; PIOTROWSKI, A.; GIJSBERS, A. A. et al. Phenotypically concordant and discordant monozygotic twins display different DNA copy-number-variation profiles. **Am J Hum Genet**, v.82, n.3, p.763-71, Mar, 2008.

CALIN, G. A.; DUMITRU, C. D.; SHIMIZU, M. et al. Frequent deletions and down-regulation of micro-RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. **Proc Natl Acad Sci U S A**, v.99, n.24, p.15524-9, Nov 26, 2002.

CAMPBELL, C. D.; SAMPAS, N.; TSALENKO, A. et al. Population-genetic properties of differentiated human copy-number polymorphisms. **Am J Hum Genet**, v.88, n.3, p.317-32, Mar 11, 2011.

CARTHEW, R. W.; SONTHEIMER, E. J. Origins and Mechanisms of miRNAs and siRNAs. **Cell**, v.136, n.4, p.642-55, Feb 20, 2009.

CHEN, D.; FU, L. Y.; ZHANG, Z. et al. Dissecting the chromatin interactome of microRNA genes. **Nucleic Acids Res**, Dec 18, 2013.

CHENG, C.; WANG, Q.; YOU, W.; CHEN, M.; XIA, J. MiRNAs as Biomarkers of Myocardial Infarction: A Meta-Analysis. **PLoS One**, v.9, n.2, p.e88566, 2014.

CONRAD, B.; ANTONARAKIS, S. E. Gene duplication: a drive for phenotypic diversity and cause of human disease. **Annu Rev Genomics Hum Genet**, v.8, p.17-35, 2007.

CONRAD, D. F.; ANDREWS, T. D.; CARTER, N. P.; HURLES, M. E.; PRITCHARD, J. K. A high-resolution survey of deletion polymorphism in the human genome. **Nat Genet**, v.38, n.1, p.75-81, Jan, 2006.

CONRAD, D. F.; HURLES, M. E. The population genetics of structural variation. **Nat Genet**, v.39, n.7 Suppl, p.S30-6, Jul, 2007.

CONRAD, D. F.; PINTO, D.; REDON, R. et al. Origins and functional impact of copy number variation in the human genome. **Nature**, v.464, n.7289, p.704-12, Apr 1, 2010.

COOPER, G. M.; ZERR, T.; KIDD, J. M.; EICHLER, E. E.; NICKERSON, D. A. Systematic assessment of copy number variant detection via genome-wide SNP genotyping. **Nat Genet**, v.40, n.10, p.1199-203, Oct, 2008.

CORSTEN, M. F.; DENNERT, R.; JOCHEMS, S. et al. Circulating MicroRNA-208b and MicroRNA-499 reflect myocardial damage in cardiovascular disease. **Circ Cardiovasc Genet**, v.3, n.6, p.499-506, Dec, 2010.

D'HAENE, B.; VANDESOMPELE, J.; HELLEMANS, J. Accurate and objective copy number profiling using real-time quantitative PCR. **Methods**, v.50, n.4, p.262-70, Apr, 2010.

DANG, V. T.; KASSAHN, K. S.; MARCOS, A. E.; RAGAN, M. A. Identification of human haploinsufficient genes and their genomic proximity to segmental duplications. **Eur J Hum Genet**, v.16, n.11, p.1350-7, Nov, 2008.

DAVIS-DUSENBERY, B. N.; HATA, A. Mechanisms of control of microRNA biogenesis. **J Biochem**, v.148, n.4, p.381-92, Oct, 2010.

DE, S. Somatic mosaicism in healthy human tissues. **Trends Genet**, v.27, n.6, p.217-23, Jun, 2011.

DE SMITH, A. J.; TSALENKO, A.; SAMPAS, N. et al. Array CGH analysis of copy number variation identifies 1284 new genes variant in healthy white males: implications for association studies of complex diseases. **Hum Mol Genet**, v.16, n.23, p.2783-94, Dec 1, 2007.

DOLAN, M. The role of the Giemsa stain in cytogenetics. **Biotech Histochem**, v.86, n.2, p.94-7, Apr, 2011.

DUAN, R.; PAK, C.; JIN, P. Single nucleotide polymorphism associated with mature miR-125a alters the processing of pri-miRNA. **Hum Mol Genet**, v.16, n.9, p.1124-31, May 1, 2007.

DUAN, S.; MI, S.; ZHANG, W.; DOLAN, M. E. Comprehensive analysis of the impact of SNPs and CNVs on human microRNAs and their regulatory genes. **RNA Biol**, v.6, n.4, p.412-25, Sep-Oct, 2009.

ELTON, T. S.; SANSOM, S. E.; MARTIN, M. M. Trisomy-21 gene dosage overexpression of miRNAs results in the haploinsufficiency of specific target proteins. **RNA biology**, v.7, n.5, p.540, 2010.

FANCIULLI, M.; PETRETTO, E.; AITMAN, T. J. Gene copy number variation and common human disease. **Clin Genet**, v.77, n.3, p.201-13, Mar, 2010.

FARAZI, T. A.; JURANEK, S. A.; TUSCHL, T. The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. **Development**, v.135, n.7, p.1201-14, Apr, 2008.

FELEKKIS, K.; VOSKARIDES, K.; DWEEP, H.; STICHT, C.; GRETZ, N.; DELTAS, C. Increased number of microRNA target sites in genes encoded in CNV regions. Evidence for an evolutionary genomic interaction. **Mol Biol Evol**, v.28, n.9, p.2421-4, Sep, 2011.

FEUK, L.; MARSHALL, C. R.; WINTLE, R. F.; SCHERER, S. W. Structural variants: changing the landscape of chromosomes and design of disease studies. **Hum Mol Genet**, v.15 Spec No 1, p.R57-66, Apr 15, 2006.

FRAZER, K. A.; MURRAY, S. S.; SCHORK, N. J.; TOPOL, E. J. Human genetic variation and its contribution to complex traits. **Nat Rev Genet**, v.10, n.4, p.241-51, Apr, 2009.

GIBBS, R. A.; BELMONT, J. W.; HARDENBOL, P. et al. The international HapMap project. **Nature**, v.426, n.6968, p.789-796, 2003.

GIGLIO, S.; BROMAN, K. W.; MATSUMOTO, N. et al. Olfactory receptor-gene clusters, genomic-inversion polymorphisms, and common chromosome rearrangements. **Am J Hum Genet**, v.68, n.4, p.874-83, Apr, 2001.

GINSBERG, G.; SMOLENSKI, S.; HATTIS, D.; SONAWANE, B. Population distribution of aldehyde dehydrogenase-2 genetic polymorphism: implications for risk assessment. **Regul Toxicol Pharmacol**, v.36, n.3, p.297-309, Dec, 2002.

GIRTON, J. R.; JOHANSEN, K. M. Chromatin structure and the regulation of gene expression: the lessons of PEV in Drosophila. **Adv Genet**, v.61, p.1-43, 2008.

GOUT, J. F.; KAHN, D.; DURET, L. The relationship among gene expression, the evolution of gene dosage, and the rate of protein evolution. **PLoS Genet**, v.6, n.5, p.e1000944, May, 2010.

GUSEV, A.; LOWE, J. K.; STOFFEL, M.; DALY, M. J.; ALTSHULER, D.; BRESLOW, J. L.; FRIEDMAN, J. M.; PE'ER, I. Whole population, genome-wide mapping of hidden relatedness. **Genome Res**, v.19, n.2, p.318-26, Feb, 2009.

HEBBRING, S. J.; ADJEI, A. A.; BAER, J. L.; JENKINS, G. D.; ZHANG, J.; CUNNINGHAM, J. M.; SCHAID, D. J.; WEINSHILBOUM, R. M.; THIBODEAU, S. N. Human SULT1A1 gene: copy number differences and functional implications. **Hum Mol Genet**, v.16, n.5, p.463-70, Mar 1, 2007.

HINDS, D. A.; KLOEK, A. P.; JEN, M.; CHEN, X.; FRAZER, K. A. Common deletions and SNPs are in linkage disequilibrium in the human genome. **Nat Genet**, v.38, n.1, p.82-5, Jan, 2006.

HOLLOX, E. J.; ARMOUR, J. A.; BARBER, J. C. Extensive normal copy number variation of a beta-defensin antimicrobial-gene cluster. **Am J Hum Genet**, v.73, n.3, p.591-600, Sep, 2003.

HUANG, R. S.; GAMAZON, E. R.; ZILIAK, D. et al. Population differences in microRNA expression and biological implications. **RNA Biol**, v.8, n.4, p.692-701, Jul 1, 2011.

IAFRATE, A. J.; FEUK, L.; RIVERA, M. N.; LISTEWNIK, M. L.; DONAHOE, P. K.; QI, Y.; SCHERER, S. W.; LEE, C. Detection of large-scale variation in the human genome. **Nat Genet**, v.36, n.9, p.949-51, Sep, 2004.

INNAN, H.; KONDRASHOV, F. The evolution of gene duplications: classifying and distinguishing between models. **Nat Rev Genet**, v.11, n.2, p.97-108, Feb, 2009.

INUI, M.; MARTELLO, G.; PICCOLO, S. MicroRNA control of signal transduction. **Nat Rev Mol Cell Biol**, v.11, n.4, p.252-63, Apr, 2010.

ITSARA, A.; COOPER, G. M.; BAKER, C. et al. Population analysis of large copy number variants and hotspots of human genetic disease. **Am J Hum Genet**, v.84, n.2, p.148-61, Feb, 2009.

ITSARA, A.; WU, H.; SMITH, J. D.; NICKERSON, D. A.; ROMIEU, I.; LONDON, S. J.; EICHLER, E. E. De novo rates and selection of large copy number variation. **Genome Res**, v.20, n.11, p.1469-81, Nov, 2010.

JAKOBSSON, M.; SCHOLZ, S. W.; SCHEET, P. et al. Genotype, haplotype and copy-number variation in worldwide human populations. **Nature**, v.451, n.7181, p.998-1003, Feb 21, 2008.

JU, Y. S.; HONG, D.; KIM, S.; PARK, S. S.; LEE, S.; PARK, H.; KIM, J. I.; SEO, J. S. Reference-unbiased copy number variant analysis using CGH microarrays. **Nucleic Acids Res**, v.38, n.20, p.e190, Nov, 2010.

KAISER, J. A plan to capture human diversity in 1000 genomes. **Science**, v.319, n.5862, p.395-395, 2008.

KATO, M.; KAWAGUCHI, T.; ISHIKAWA, S. et al. Population-genetic nature of copy number variations in the human genome. **Hum Mol Genet**, v.19, n.5, p.761-73, Mar 1, 2009.

KELL, D.; OLIVER, S. Here is the evidence, now what is the hypothesis? The complementary roles of inductive and hypothesis-driven science in the post-genomic era. **BioEssays: news and reviews in molecular, cellular and developmental biology**, v.26, n.1, p.99, 2004.

KIDD, J. M.; COOPER, G. M.; DONAHUE, W. F. et al. Mapping and sequencing of structural variation from eight human genomes. **Nature**, v.453, n.7191, p.56-64, May 1, 2008.

KIDD, J. M.; GRAVES, T.; NEWMAN, T. L. et al. A human genome structural variation sequencing resource reveals insights into mutational mechanisms. **Cell**, v.143, n.5, p.837-47, Nov 24, 2010.

KIM, J. I.; JU, Y. S.; PARK, H. et al. A highly annotated whole-genome sequence of a Korean individual. **Nature**, v.460, n.7258, p.1011-5, Aug 20, 2009.

KIM, V. N.; NAM, J. W. Genomics of microRNA. **Trends Genet**, v.22, n.3, p.165-73, Mar, 2006.

KLOPOCKI, E.; MUNDLOS, S. Copy-number variations, noncoding sequences, and human phenotypes. **Annu Rev Genomics Hum Genet**, v.12, p.53-72, Sep 22, 2011.

KOIKE, A.; NISHIDA, N.; YAMASHITA, D.; TOKUNAGA, K. Comparative analysis of copy number variation detection methods and database construction. **BMC Genet**, v.12, p.29, 2011.

KONDRASHOV, F. A.; KOONIN, E. V. A common framework for understanding the origin of genetic dominance and evolutionary fates of gene duplications. **Trends Genet**, v.20, n.7, p.287-90, Jul, 2004.

KONRAD, A.; TEUFEL, A. I.; GRAHNEN, J. A.; LIBERLES, D. A. Toward a general model for the evolutionary dynamics of gene duplicates. **Genome Biol Evol**, v.3, p.1197-209, 2011.

KORBEL, J. O.; URBAN, A. E.; GRUBERT, F. et al. Systematic prediction and validation of breakpoints associated with copy-number variants in the human genome. **Proc Natl Acad Sci U S A**, v.104, n.24, p.10110-5, Jun 12, 2007.

KUHN, D. E.; NUOVO, G. J.; MARTIN, M. M. et al. Human chromosome 21-derived miRNAs are overexpressed in down syndrome brains and hearts. **Biochem Biophys Res Commun**, v.370, n.3, p.473-7, Jun 6, 2008.

LANDER, E. S. Initial impact of the sequencing of the human genome. **Nature**, v.470, n.7333, p.187-97, Feb 10, 2011.

LAU, N. C.; LIM, L. P.; WEINSTEIN, E. G.; BARTEL, D. P. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. **Science**, v.294, n.5543, p.858-62, Oct 26, 2001.

LEE, C.; SCHERER, S. W. The clinical context of copy number variation in the human genome. **Expert Rev Mol Med**, v.12, p.e8, 2010.

LEE, J. H.; JEON, J. T. Methods to detect and analyze copy number variations at the genome-wide and locus-specific levels. **Cytogenet Genome Res**, v.123, n.1-4, p.333-42, 2008.

LEE, Y.; AHN, C.; HAN, J. et al. The nuclear RNase III Drosha initiates microRNA processing. **Nature**, v.425, n.6956, p.415-9, Sep 25, 2003.

LEE, Y.; KIM, M.; HAN, J.; YEOM, K. H.; LEE, S.; BAEK, S. H.; KIM, V. N. MicroRNA genes are transcribed by RNA polymerase II. **EMBO J**, v.23, n.20, p.4051-60, Oct 13, 2004.

LEJEUNE, J.; GAUTIER, M.; TURPIN, R. Etude des chromosomes somatiques de neuf enfants mongoliens. [Study of somatic chromosomes from 9 mongoloid children]. **C R Hebd Seances Acad Sci**, v.248, n.11, p.1721-2, Mar 16, 1959.

LEJEUNE, J.; LAFOURCADE, J.; BERGER, R.; VIALATTE, J.; BOESWILLWALD, M.; SERINGE, P.; TURPIN, R. Trois cas de d'el'etion partielle du bras court d'un chromosome 5. [3 Cases of Partial Deletion of the Short Arm of a 5 Chromosome]. **C R Hebd Seances Acad Sci**, v.257, p.3098-102, Nov 18, 1963.

LEVY, S.; SUTTON, G.; NG, P. C. et al. The diploid genome sequence of an individual human. **PLoS Biol**, v.5, n.10, p.e254, Sep 4, 2007.

LI, J.; YANG, T.; WANG, L. et al. Whole genome distribution and ethnic differentiation of copy number variation in Caucasian and Asian populations. **PLoS One**, v.4, n.11, p.e7958, 2009.

LIANG, Y.; RIDZON, D.; WONG, L.; CHEN, C. Characterization of microRNA expression profiles in normal human tissues. **BMC Genomics**, v.8, p.166, 2007.

LINS, T. C.; VIEIRA, R. G.; ABREU, B. S.; GRATTAPAGLIA, D.; PEREIRA, R. W. Genetic composition of Brazilian population samples based on a set of twenty-eight ancestry informative SNPs. **Am J Hum Biol**, v.22, n.2, p.187-92, Mar-Apr, 2010.

LOCKE, D. P.; SHARP, A. J.; MCCARROLL, S. A. et al. Linkage disequilibrium and heritability of copy-number polymorphisms within duplicated regions of the human genome. **Am J Hum Genet**, v.79, n.2, p.275-90, Aug, 2006.

LU, J.; CLARK, A. G. Impact of microRNA regulation on variation in human gene expression. **Genome Res**, Mar 28, 2012.

LUND, E.; GUTTINGER, S.; CALADO, A.; DAHLBERG, J. E.; KUTAY, U. Nuclear export of microRNA precursors. **Science**, v.303, n.5654, p.95-8, Jan 2, 2004.

MANDEL, A. L.; PEYROT DES GACHONS, C.; PLANK, K. L.; ALARCON, S.; BRESLIN, P. A. Individual differences in AMY1 gene copy number, salivary alpha-amylase levels, and the perception of oral starch. **PLoS One**, v.5, n.10, p.e13352, 2010.

MARCINKOWSKA, M.; SZYMANSKI, M.; KRZYZOSIAK, W. J.; KOZLOWSKI, P. Copy number variation of microRNA genes in the human genome. **BMC Genomics**, v.12, p.183, 2011.

MAYO, P.; HARTSHORNE, T.; LI, K.; MCMUNN-GIBSON, C.; SPENCER, K.; SCHNETZ-BOUTAUD, N. CNV Analysis Using TaqMan Copy Number Assays. **Current Protocols in Human Genetics**, v.2, p.1-2.13, 2010.

MCCARROLL, S. A.; HADNOTT, T. N.; PERRY, G. H. et al. Common deletion polymorphisms in the human genome. **Nat Genet**, v.38, n.1, p.86-92, Jan, 2006.

MCCARROLL, S. A.; KURUVILLA, F. G.; KORN, J. M. et al. Integrated detection and population-genetic analysis of SNPs and copy number variation. **Nat Genet**, v.40, n.10, p.1166-74, Oct, 2008.

MCCONNELL, M. J.; LINDBERG, M. R.; BRENNAND, K. J. et al. Mosaic copy number variation in human neurons. **Science**, v.342, n.6158, p.632-7, Nov 1, 2013.

MCELROY, J. P.; NELSON, M. R.; CAILLIER, S. J.; OKSENBERG, J. R. Copy number variation in African Americans. **BMC Genet**, v.10, p.15, 2009.

MCKERNAN, K. J.; PECKHAM, H. E.; COSTA, G. L. et al. Sequence and structural variation in a human genome uncovered by short-read, massively parallel ligation sequencing using two-base encoding. **Genome Res**, v.19, n.9, p.1527-41, Sep, 2009.

MELO, C. A.; MELO, S. A. Biogenesis and Physiology of MicroRNAs. In: (Ed.). **Non-coding RNAs and Cancer**: Springer, 2014. Biogenesis and Physiology of MicroRNAs, p.5-24

MICALE, M. A.; MOHAMED, A. N. 16 Molecular Cytogenetic (FISH) Analysis of Hematolymphoid Disorders. **Laboratory Hematology Practice**, p.181, 2012.

MILLS, R. E.; LUTTIG, C. T.; LARKINS, C. E.; BEAUCHAMP, A.; TSUI, C.; PITTARD, W. S.; DEVINE, S. E. An initial map of insertion and deletion (INDEL) variation in the human genome. **Genome Res**, v.16, n.9, p.1182-90, Sep, 2006.

MITCHELL, P. S.; PARKIN, R. K.; KROH, E. M. et al. Circulating microRNAs as stable blood-based markers for cancer detection. **Proc Natl Acad Sci U S A**, v.105, n.30, p.10513-8, Jul 29, 2008.

MIZUNO, H.; NAKAMURA, A.; AOKI, Y.; ITO, N.; KISHI, S.; YAMAMOTO, K.; SEKIGUCHI, M.; TAKEDA, S.; HASHIDO, K. Identification of muscle-specific microRNAs in serum of muscular dystrophy animal models: promising novel blood-based markers for muscular dystrophy. **PLoS One**, v.6, n.3, p.e18388, 2011.

NGUYEN, D. Q.; WEBBER, C.; PONTING, C. P. Bias of selection on human copy-number variants. **PLoS Genet**, v.2, n.2, p.e20, Feb, 2006.

NINOMIYA, S.; TYYBAKINOJA, A.; BORZE, I.; RATY, R.; SAARINEN-PIHKALA, U. M.; USVASALO, A.; ELONEN, E.; KNUUTILA, S. Integrated Analysis of Gene Copy Number, Copy Neutral LOH, and microRNA Profiles in Adult Acute Lymphoblastic Leukemia. **Cytogenet Genome Res**, Mar 28, 2012.

NOWELL, P. C.; HUNGERFORD, D. A. Chromosome studies on normal and leukemic human leukocytes. **J Natl Cancer Inst**, v.25, p.85-109, Jul, 1960.

OSBORNE, L. R.; LI, M.; POBER, B. et al. A 1.5 million-base pair inversion polymorphism in families with Williams-Beuren syndrome. **Nat Genet**, v.29, n.3, p.321-5, Nov, 2001.

PANG, A. W.; MACDONALD, J. R.; PINTO, D. et al. Towards a comprehensive structural variation map of an individual human genome. **Genome Biol**, v.11, n.5, p.R52, 2010.

PARK, H.; KIM, J. I.; JU, Y. S. et al. Discovery of common Asian copy number variants using integrated high-resolution array CGH and massively parallel DNA sequencing. **Nat Genet**, v.42, n.5, p.400-5, May, 2010.

PERRY, G. H.; BEN-DOR, A.; TSALENKO, A. et al. The fine-scale and complex architecture of human copy-number variation. **Am J Hum Genet**, v.82, n.3, p.685-95, Mar, 2008.

PERRY, G. H.; DOMINY, N. J.; CLAW, K. G. et al. Diet and the evolution of human amylase gene copy number variation. **Nat Genet**, v.39, n.10, p.1256-60, Oct, 2007.

PINTO, D.; MARSHALL, C.; FEUK, L.; SCHERER, S. W. Copy-number variation in control population cohorts. **Hum Mol Genet**, v.16 Spec No. 2, p.R168-73, Oct 15, 2007.

PIOTROWSKI, A.; BRUDER, C. E.; ANDERSSON, R. et al. Somatic mosaicism for copy number variation in differentiated human tissues. **Hum Mutat**, v.29, n.9, p.1118-24, Sep, 2008.

PODURI, A.; EVRONY, G. D.; CAI, X.; WALSH, C. A. Somatic mutation, genomic variation, and neurological disease. **Science**, v.341, n.6141, p.1237758, Jul 5, 2013.

POY, M. N.; ELIASSON, L.; KRUTZFELDT, J. et al. A pancreatic islet-specific microRNA regulates insulin secretion. **Nature**, v.432, n.7014, p.226-30, Nov 11, 2004.

PRITCHARD, C. C.; KROH, E.; WOOD, B.; ARROYO, J. D.; DOUGHERTY, K. J.; MIYAJI, M. M.; TAIT, J. F.; TEWARI, M. Blood cell origin of circulating microRNAs: a cautionary note for cancer biomarker studies. **Cancer Prev Res (Phila)**, v.5, n.3, p.492-7, Mar, 2011.

PRONOLD, M.; VALI, M.; PIQUE-REGI, R.; ASGHARZADEH, S. Copy number variation signature to predict human ancestry. **BMC Bioinformatics**, v.13, p.336, 2012.

QIAO, Y.; BADDUKE, C.; MERCIER, E.; LEWIS, S. M.; PAVLIDIS, P.; RAJCAN-SEPAROVIC, E. miRNA and miRNA target genes in copy number variations occurring in individuals with intellectual disability. **BMC Genomics**, v.14, p.544, 2013.

REDON, R.; ISHIKAWA, S.; FITCH, K. R. et al. Global variation in copy number in the human genome. **Nature**, v.444, n.7118, p.444-54, Nov 23, 2006.

REINHART, B. J.; WEINSTEIN, E. G.; RHOADES, M. W.; BARTEL, B.; BARTEL, D. P. MicroRNAs in plants. **Genes Dev**, v.16, n.13, p.1616-26, Jul 1, 2002.

SANDBERG, R.; NEILSON, J. R.; SARMA, A.; SHARP, P. A.; BURGE, C. B. Proliferating cells express mRNAs with shortened 3' untranslated regions and fewer microRNA target sites. **Science**, v.320, n.5883, p.1643-7, Jun 20, 2008.

SCHERER, S. W.; LEE, C.; BIRNEY, E.; ALTSHULER, D. M.; EICHLER, E. E.; CARTER, N. P.; HURLES, M. E.; FEUK, L. Challenges and standards in integrating surveys of structural variation. **Nat Genet**, v.39, n.7 Suppl, p.S7-15, Jul, 2007.

SCHMITTGEN, T. D.; LIVAK, K. J. Analyzing real-time PCR data by the comparative C(T) method. **Nat Protoc**, v.3, n.6, p.1101-8, 2008.

SCHUSTER-BOCKLER, B.; CONRAD, D.; BATEMAN, A. Dosage sensitivity shapes the evolution of copy-number varied regions. **PLoS One**, v.5, n.3, p.e9474, 2010.

SCHWARZ, D. S.; HUTVAGNER, G.; DU, T.; XU, Z.; ARONIN, N.; ZAMORE, P. D. Asymmetry in the assembly of the RNAi enzyme complex. **Cell**, v.115, n.2, p.199-208, Oct 17, 2003.

SEBAT, J.; LAKSHMI, B.; TROGE, J. et al. Large-scale copy number polymorphism in the human genome. **Science**, v.305, n.5683, p.525-8, Jul 23, 2004.

SETHUPATHY, P.; BOREL, C.; GAGNEBIN, M.; GRANT, G. R.; DEUTSCH, S.; ELTON, T. S.; HATZIGEORGIOU, A. G.; ANTONARAKIS, S. E. Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3' untranslated region: a

mechanism for functional single-nucleotide polymorphisms related to phenotypes. **Am J Hum Genet**, v.81, n.2, p.405-13, Aug, 2007.

SHADRAVAN, F. Sex bias in copy number variation of olfactory receptor gene family depends on ethnicity. **Front Genet**, v.4, p.32, 2013.

SHAIKH, T. H.; GAI, X.; PERIN, J. C. et al. High-resolution mapping and analysis of copy number variations in the human genome: a data resource for clinical and research applications. **Genome Res**, v.19, n.9, p.1682-90, Sep, 2009.

SHARP, A. J.; LOCKE, D. P.; MCGRATH, S. D. et al. Segmental duplications and copy-number variation in the human genome. **Am J Hum Genet**, v.77, n.1, p.78-88, Jul, 2005.

SHAW, C. J.; LUPSKI, J. R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. **Hum Mol Genet**, v.13 Spec No 1, p.R57-64, Apr 1, 2004.

SHAW, C. J.; WITHERS, M. A.; LUPSKI, J. R. Uncommon deletions of the Smith-Magenis syndrome region can be recurrent when alternate low-copy repeats act as homologous recombination substrates. **Am J Hum Genet**, v.75, n.1, p.75-81, Jul, 2004.

SHEN, Z.; QU, W.; WANG, W. et al. MPprimer: a program for reliable multiplex PCR primer design. **BMC Bioinformatics**, v.11, p.143, 2010.

SIMON-SANCHEZ, J.; SCHOLZ, S.; FUNG, H. C. et al. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. **Hum Mol Genet**, v.16, n.1, p.1-14, Jan 1, 2007.

SLATTERY, M. L.; WOLFF, E.; HOFFMAN, M. D.; PELLATT, D. F.; MILASH, B.; WOLFF, R. K. MicroRNAs and colon and rectal cancer: differential expression by tumor location and subtype. **Genes Chromosomes Cancer**, v.50, n.3, p.196-206, Mar, 2011.

SOSKINE, M.; TAWFIK, D. S. Mutational effects and the evolution of new protein functions. **Nat Rev Genet**, v.11, n.8, p.572-82, Aug, 2010.

SPIRA, A.; BEANE, J.; SCHEMBRI, F.; LIU, G.; DING, C.; GILMAN, S.; YANG, X.; CANTOR, C.; BRODY, J. S. Noninvasive method for obtaining RNA from buccal mucosa epithelial cells for gene expression profiling. **Biotechniques**, v.36, n.3, p.484-7, Mar, 2004.

STOFANKO, M.; GONCALVES-DORNELAS, H.; CUNHA, P. S.; PENA, H. B.; VIANNA-MORGANTE, A. M.; PENA, S. D. Simple, rapid and inexpensive quantitative fluorescent PCR method for detection of microdeletion and microduplication syndromes. **PLoS One**, v.8, n.4, p.e61328, 2013.

STRANGER, B. E.; FORREST, M. S.; DUNNING, M. et al. Relative impact of nucleotide and copy number variation on gene expression phenotypes. **Science**, v.315, n.5813, p.848-53, Feb 9, 2007.

SUDMANT, P. H.; KITZMAN, J. O.; ANTONACCI, F. et al. Diversity of human copy number variation and multicopy genes. **Science**, v.330, n.6004, p.641-6, Oct 29, 2010.

TEAGUE, B.; WATERMAN, M. S.; GOLDSTEIN, S. et al. High-resolution human genome structure by single-molecule analysis. **Proc Natl Acad Sci U S A**, v.107, n.24, p.10848-53, Jun 15, 2010.

TUZUN, E.; SHARP, A. J.; BAILEY, J. A. et al. Fine-scale structural variation of the human genome. **Nat Genet**, v.37, n.7, p.727-32, Jul, 2005.

VAISHNAVI, V.; MANIKANDAN, M.; TIWARY, B. K.; MUNIRAJAN, A. K. Insights on the functional impact of microRNAs present in autism-associated copy number variants. **PLoS One**, v.8, n.2, p.e56781, 2013.

VANGUILDER, H. D.; VRANA, K. E.; FREEMAN, W. M. Twenty-five years of quantitative PCR for gene expression analysis. **Biotechniques**, v.44, n.5, p.619-26, Apr, 2008.

VEITIA, R. A.; BOTTANI, S.; BIRCHLER, J. A. Cellular reactions to gene dosage imbalance: genomic, transcriptomic and proteomic effects. **Trends Genet**, v.24, n.8, p.390-7, Aug, 2008.

WANG, Y.; MOORHEAD, M.; KARLIN-NEUMANN, G. et al. Analysis of molecular inversion probe performance for allele copy number determination. **Genome Biol**, v.8, n.11, p.R246, 2007.

WEAVER, S.; DUBE, S.; MIR, A.; QIN, J.; SUN, G.; RAMAKRISHNAN, R.; JONES, R. C.; LIVAK, K. J. Taking qPCR to a higher level: Analysis of CNV reveals the power of high throughput qPCR to enhance quantitative resolution. **Methods**, v.50, n.4, p.271-6, Apr, 2010.

WEBER, J. A.; BAXTER, D. H.; ZHANG, S.; HUANG, D. Y.; HUANG, K. H.; LEE, M. J.; GALAS, D. J.; WANG, K. The microRNA spectrum in 12 body fluids. **Clin Chem**, v.56, n.11, p.1733-41, Nov, 2010.

WEISCHENFELDT, J.; SYMMONS, O.; SPITZ, F.; KORBEL, J. O. Phenotypic impact of genomic structural variation: insights from and for human disease. **Nat Rev Genet**, v.14, n.2, p.125-38, Feb, 2013.

WEISS, L. A.; SHEN, Y.; KORN, J. M. et al. Association between microdeletion and microduplication at 16p11.2 and autism. **N Engl J Med**, v.358, n.7, p.667-75, Feb 14, 2008.

WHEELER, D. A.; SRINIVASAN, M.; EGHOLM, M. et al. The complete genome of an individual by massively parallel DNA sequencing. **Nature**, v.452, n.7189, p.872-6, Apr 17, 2008.

WINEINGER, N. E.; PAJEWSKI, N. M.; KENNEDY, R. E. et al. Characterization of autosomal copy-number variation in African Americans: the HyperGEN Study. **Eur J Hum Genet**, v.19, n.12, p.1271-5, Dec, 2011.

WONG, K. K.; DELEEuw, R. J.; DOSANJH, N. S. et al. A comprehensive analysis of common copy-number variations in the human genome. **Am J Hum Genet**, v.80, n.1, p.91-104, Jan, 2007.

WONG, L. P.; ONG, R. T.; POH, W. T. et al. Deep whole-genome sequencing of 100 southeast Asian Malays. **Am J Hum Genet**, v.92, n.1, p.52-66, Jan 10, 2012.

WOODWARK, C.; BATEMAN, A. The characterisation of three types of genes that overlie copy number variable regions. **PLoS One**, v.6, n.5, p.e14814, 2011.

XING, H. J.; LI, Y. J.; MA, Q. M. et al. Identification of microRNAs present in congenital heart disease associated copy number variants. **Eur Rev Med Pharmacol Sci**, v.17, n.15, p.2114-20, Aug, 2013.

XU, H.; POH, W. T.; SIM, X. et al. SgD-CNV, a database for common and rare copy number variants in three Asian populations. **Hum Mutat**, v.32, n.12, p.1341-9, Dec, 2011.

YOUSSOUFIAN, H.; PYERITZ, R. E. Mechanisms and consequences of somatic mosaicism in humans. **Nat Rev Genet**, v.3, n.10, p.748-58, Oct, 2002.

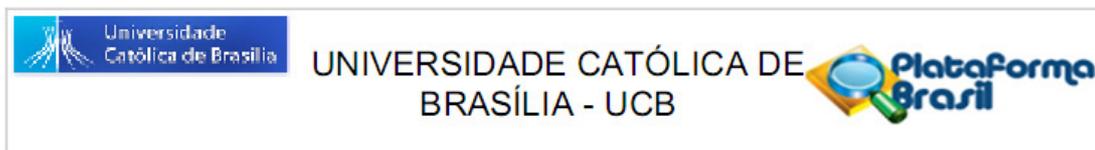
ZHANG, R.; ZHANG, L.; YU, W. Genome-wide expression of non-coding RNA and global chromatin modification. **Acta Biochim Biophys Sin (Shanghai)**, v.44, n.1, p.40-7, Jan, 2011.

ZHANG, Y. B.; LI, X.; ZHANG, F.; WANG, D. M.; YU, J. A preliminary study of copy number variation in Tibetans. **PLoS One**, v.7, n.7, p.e41768, 2012.

ZOGOPOULOS, G.; HA, K. C.; NAQIB, F. et al. Germ-line DNA copy number variation frequencies in a large North American population. **Hum Genet**, v.122, n.3-4, p.345-53, Nov, 2007.

## **ANEXO I**

- Parecer consubstanciado do Comitê de Ética em Pesquisa;
- Termo de consentimento livre e esclarecido;



## PARECER CONSUBSTANCIADO DO CEP

### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Análise de variações no número de cópias gênicas e sua correlação com expressão de microRNAs.

**Pesquisador:** Rinaldo Wellerson Pereira

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

**Versão:** 1

**CAAE:** 13575813.1.0000.0029

**Instituição Proponente:** Universidade Católica de Brasília - UCB

**Patrocinador Principal:** Conselho Nacional de Desenvolvimento Científico e Tecnológico ((CNPq))

### DADOS DO PARECER

**Número do Parecer:** 218.781

**Data da Relatoria:** 11/03/2013

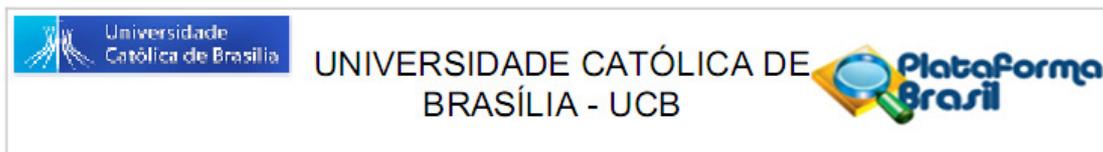
#### **Apresentação do Projeto:**

Projeto de Doutorado do aluno Msc. Tulio Cesar de Lima Lins, sob orientação do Prof. Dr. Rinaldo Wellerson Pereira.

MicroRNA constitui uma classe de pequenos RNAs não codificantes, que possuem entre 18 e 24 pares de base e estão envolvidos em diversos processos biológicos, incluindo metabolismo, desenvolvimento, proliferação celular e apoptose, tendo também papel em doenças humanas como diabetes e câncer. O objetivo de investigar a variabilidade em número de cópias de regiões gênicas com potencial impacto na expressão e ação de miRNAs em pares de tecido de origem no ectoderma (epitélio bucal) e no mesoderma (células da linhagem leucocitária).

Metodologia: será composta por um grupo de brasileiros que representam a maior diversidade de nossa população e também diferentes espectros de miscigenação, recrutados entre a população de estudantes, professores e demais funcionários que circulam pelos Campi da Universidade Católica de Brasília. Os voluntários depreenderão de aproximadamente 10 minutos para a leitura e entendimento da pesquisa, coleta do sangue utilizando seringa estéril e coleta de células bucais, realizados por profissional qualificado. O recrutamento será realizado através de palestras agendadas nos Campi da UCB, visitas em salas

<b>Endereço:</b> QS 07 Lote 01 EPCT - Bloco Central - João Batista de la Salle - Sala - A-005			
<b>Bairro:</b> Taguatinga		<b>CEP:</b> 71.966-700	
<b>UF:</b> DF	<b>Município:</b> BRASILIA		
<b>Telefone:</b> (61)3356-9784	<b>Fax:</b> (61)3356-3010	<b>E-mail:</b> cep@ucb.br	



de aula e distribuição de e-mail pelas listas da Universidade. O critério de inclusão e exclusão será indivíduo maior de 18 anos e assinar o termo de consentimento livre e esclarecido.

A pesquisa será realizada no Laboratório de Ciências Genômicas e Biotecnologia no Programa de Pós Graduação em Ciências Genômicas e Biotecnologia da Universidade Católica de Brasília. A coleta do material biológico será no Campus I e II da Universidade Católica de Brasília.

O início esta previsto para 02 de Abril de 2012 e o término: 02 de Abril de 2014.

Projeto aprovado pelo CNPq.

**Objetivo da Pesquisa:**

O objetivo de investigar a variabilidade em número de cópias de regiões gênicas com potencial impacto na expressão e ação de miRNAs em pares de tecido de origem no ectoderma (epitélio bucal) e no mesoderma (células da linhagem leucocitária).

**Avaliação dos Riscos e Benefícios:**

Os riscos atribuíveis aos participantes neste projeto de pesquisa estão relacionados à coleta de material biológico, os quais são desprezíveis, já que são poucos invasivos e serão realizados por profissional qualificado. Riscos individuais associados à geração dos resultados não existem, já que os dados genéticos utilizados neste estudo não são preditivos de nenhum critério clínico ou médico. No que diz respeito aos riscos envolvidos com a coleta de material genético, serão evitados utilizando profissional capacitado com treinamento para tal e sempre utilizando materiais descartáveis e correta assepsia antes e depois da coleta.

**Comentários e Considerações sobre a Pesquisa:**

A pesquisa esta adequada aos princípios científicos que a justifiquem. O pesquisador descreve o projeto de forma detalhada e ordenada. O desenho esta adequado ao estudo e a população estudada se adequa ao objetivo do estudo. Todo o método foi descrito é necessário apenas descrever o local que será realizada a coleta do sangue dos voluntários (Sugere-se especial atenção dos pesquisadores a este item, para que se adeque as normas de segurança)

**Considerações sobre os Termos de apresentação obrigatória:**

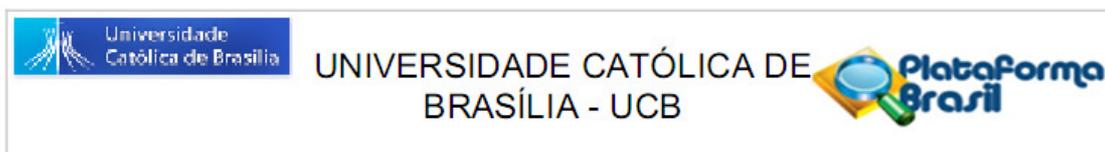
O TCLE, carta de encaminhamento, folha de rosto, orçamento e cronograma estão adequados.

**Recomendações:**

Sugere-se especial atenção dos pesquisadores ao item local de coleta de dados, para que se adeque as normas de segurança.

Sugere-se que o pesquisador investigue também se os voluntários já possuem histórico de doença (cancêr, diabetes) acredita-se que esta informação será útil para o projeto.

**Endereço:** QS 07 Lote 01 EPCT - Bloco Central - João Batista de la Salle - Sala - A-005  
**Bairro:** Taguatinga **CEP:** 71.966-700  
**UF:** DF **Município:** BRASILIA  
**Telefone:** (61)3356-9784 **Fax:** (61)3356-3010 **E-mail:** cep@ucb.br

**Conclusões ou Pendências e Lista de Inadequações:**

O projeto atende às exigências da Resolução 196/96 e recomenda-se sua aprovação pelo Comitê de Ética em Pesquisa.

Após a realização da pesquisa é compromisso dos/das proponentes a entrega de relatório final ou versão final do trabalho.

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

**Considerações Finais a critério do CEP:**

BRASILIA, 14 de Março de 2013

---

**Assinador por:**  
**Yomara Lima Mota**  
**(Coordenador)**

**Endereço:** QS 07 Lote 01 EPCT - Bloco Central - João Batista de la Salle - Sala - A-005  
**Bairro:** Taguatinga **CEP:** 71.966-700  
**UF:** DF **Município:** BRASILIA  
**Telefone:** (61)3356-9784 **Fax:** (61)3356-3010 **E-mail:** cep@ucb.br

## Termo de consentimento livre e esclarecido

**Projeto de pesquisa:** Análise de variações no número de cópias gênicas e sua correlação com expressão de microRNA.

Este projeto tem como justificativa estudar as variações de número de cópias de regiões do genoma que podem alterar a quantidade da expressão gênica em uma população saudável. O objetivo é correlacionar o número de cópias dessas regiões com a quantidade do produto de sua expressão.

Você está sendo convidado a participar deste estudo na qualidade de voluntário. Como critério de inclusão, deverá ser maior de 18 anos, ler e compreender (de forma esclarecida) o presente termo de consentimento livre e esclarecido. Você tem toda a liberdade em decidir se deseja ou não participar do mesmo. Não há nenhuma consequência negativa caso você decida não participar. Caso você decida participar o procedimento será o seguinte:

Um profissional qualificado fará a coleta de amostras biológicas de sangue e células da mucosa oral (interior da bochecha) seguindo normas de segurança para que não ocorram desconfortos e possíveis riscos. Os materiais serão utilizados para extração de DNA e RNA, utilizados para o estudo do número de cópias no genoma e expressão de microRNA.

As amostras biológicas não serão disponibilizadas para outros estudos em qualquer hipótese. Os dados e registros relacionados a este projeto serão mantidos no Programa de Ciências Genômicas e Biotecnologia da Universidade Católica de Brasília, com garantia do sigilo que assegure a sua privacidade quanto aos dados confidenciais envolvidos na pesquisa. Em qualquer momento e por qualquer circunstância você tem toda liberdade de requerer esclarecimentos sobre a metodologia e acompanhamento de resultados durante o curso da pesquisa.

Você tem a total liberdade de se recusar a participar, ou retirar seu consentimento, em qualquer fase da pesquisa, sem penalização alguma e sem prejuízo ao seu cuidado. Não existem benefícios esperados, uma vez que a pesquisa é de caráter experimental e não está associada à testes clínicos. Ao assinar esse termo você autoriza a coleta de amostras biológicas de acordo com os termos descrito acima.

O pesquisador abaixo assinado se responsabiliza com a privacidade das informações realizadas no âmbito da pesquisa.

---

### Assinatura do responsável pela pesquisa

Prof. Dr. Rinaldo Wellerson Pereira

RG: 4208151 SSP-MG

CPF: 844342796-53

Telefone: (61) 3448 7129

Email: rinaldo.pereira@catolica.edu.br

Universidade Católica de Brasília

Data da Coleta: \_\_\_\_/\_\_\_\_/\_\_\_\_

---

### Assinatura do voluntário

Nome Completo:

\_\_\_\_\_

RG: \_\_\_\_\_

CPF: \_\_\_\_\_

Data de Nascimento: \_\_\_\_/\_\_\_\_/\_\_\_\_

Sexo:  Masculino  Feminino