

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE BIOCÊNCIAS
BACHARELADO EM BIOMEDICINA**

AMANDA KELLY DO NASCIMENTO MENDONÇA

**ABORDAGEM COMPUTACIONAL PARA DETECÇÃO DE ANEUPLOIDIAS
FETAIS**

**NATAL - RN
2016**

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO NORTE
CENTRO DE BIOCÊNCIAS
BACHARELADO EM BIOMEDICINA**

AMANDA KELLY DO NASCIMENTO MENDONÇA

**ABORDAGEM COMPUTACIONAL PARA DETECÇÃO DE ANEUPLOIDIAS
FETAIS**

Monografia apresentada à coordenação do curso de biomedicina da Universidade Federal do Rio Grande do Norte, como requisito parcial à obtenção do título de bacharel em Biomedicina.

Prof. Dr. Sandro José de Souza

**NATAL - RN
2016**

AMANDA KELLY DO NASCIMENTO MENDONÇA

**ABORDAGEM COMPUTACIONAL PARA DETECÇÃO DE ANEUPLOIDIAS
FETAIS**

A monografia aprovada por todos os membros da Banca examinadora foi aceita pelo Curso de Biomedicina e homologada pelos membros da banca, como requisito parcial à obtenção do título de bacharel em Biomedicina.

Natal, 13 de Dezembro de 2016.

BANCA EXAMINADORA



Prof. Dr. Sandro José de Souza
Instituto do Cérebro - ICe, UFRN.



Prof. Dr. César Rennó Costa
Instituto Metr pole Digital – IMD, UFRN.



Prof. Dr. Gustavo Ant nio de Souza
Instituto do C rebro - ICe, UFRN.

Universidade Federal do Rio Grande do Norte - UFRN
Sistema de Bibliotecas - SISBI

Catálogo de Publicação na Fonte. UFRN - Biblioteca Setorial do Centro de Biociências - CB

Mendonça, Amanda Kelly do Nascimento.

Abordagem computacional para detecção de aneuploidias fetais
/ Amanda Kelly do Nascimento Mendonça. - Natal, 2016.
34 f.: il.

Monografia (Graduação) - Universidade Federal do Rio Grande
do Norte. Centro de Biociências. Bacharelado em Biomedicina.
Orientador: Prof. Dr. Sandro José de Souza.

1. Feto - Monografia. 2. NIPT - Monografia. 3. CAADy -
Monografia. 4. Aneuploidias - Monografia. I. Souza, Sandro José
de. II. Universidade Federal do Rio Grande do Norte. III.
Título.

RN/UF/BSE-CB

CDU 612.647

SUMÁRIO

1. INTRODUÇÃO

- 1.1. Do Material Genômico às Trissomias
- 1.2. DNA Fetal no Plasma Materno
- 1.3. Teste Pré-natal Não-invasivo
- 1.4. Sequenciamento Paralelo Massivo do Genoma
- 1.5. Metodologias de Detecção para Aneuploidias Fetais
- 1.6. Abordagem Computacional para Detecção de Aneuploidias Fetais (CAADy)

2. REFERÊNCIAS

3. NORMAS DO PERIÓDICO

- 3.1. Preparo da submissão
- 3.2. Informações básicas
- 3.3. Padrão de seções
- 3.4. Formato de referências
- 3.5. Referenciando tabelas e figuras
- 3.6. Tipos de arquivo
- 3.7. Estilo de considerações

4. ARTIGO

ÍNDICE

LISTA DE ABREVIATURAS.....	5
LISTA DE TABELAS.....	6
Tabela 1. Cariótipo fetal.....	23
Tabela 2. Comparação entre os métodos de z-score IR (desvio-padrão) e z-score IR modificado (MAD) para detecção dos casos de trissomias dos cromossomos 13, 18 e 21.....	26
LISTA DE FIGURAS.....	7
Figura 1. Representação esquemática da metodologia de detecção abordada pela CAADy.....	25
Figura 2. Seleção do chr1R baseada na proximidade do conteúdo CG com o cromossomo da trissomia.....	26
Figura 3. Trissomias detectadas pelo CAADy.....	28.
RESUMO.....	8
ABSTRACT.....	9
INTRODUÇÃO.....	10
1.1. Do Material Genômico às Trissomias.....	
1.2. DNA Fetal no Plasma Materno.....	10
1.3. Teste Pré-natal Não-invasivo.....	10
1.4. Sequenciamento Paralelo Massivo do Genoma.....	11
1.5. Metodologias de Detecção para Aneuploidias Fetais.....	12
1.6. Abordagem Computacional para Detecção de Aneuploidias Fetais (CAADy)	
REFERÊNCIAS.....	14
NORMAS DO PERIÓDICO.....	16
3.1. Preparo da submissão.....	16
3.2. Informações básicas.....	16
3.3. Padrão de seções.....	16
3.4. Formato de referências.....	17
3.5. Referenciando tabelas e figuras.....	18
3.6. Tipos de arquivo.....	18
3.7. Estilo de considerações.....	19
ARTIGO.....	21

LISTA DE ABREVIATURAS

CAADy – Abordagem computacional para detecção de aneuploidias fetais

cffDNA – cell free fetal DNA

CG – citosina e guanina

chr – cromossomo

IR – Referência interna

MAD – desvio absoluto mediano

MPGS - Massively Parallel Sequencing

mRNA – Acido ribonucleico mensageiro

NGS - *Next-Generation Sequencing*

NIFTY – Noninvasive fetal trisomy test

NIPT – Teste pré-natal não-invasivo

PCR – Reação em cadeia da polimerase

SRA - Sequence Read Archive

LISTA DE TABELAS

Tabela 1. Cariótipo fetal.....	23
Tabela 2. Comparação entre os métodos de z-score IR (desvio-padrão) e z-score IR modificado (MAD) para detecção dos casos de trissomias dos cromossomos 13, 18 e 21.....	26

LISTA DE FIGURAS

Figura 1. Esquematização da metodologia de detecção abordada pela CAADy....	25
Figura 2. Seleção do chrIR baseada na proximidade do conteúdo CG com o cromossomo da trissomia.....	26
Figura 3. Trissomias detectadas pelo CAADy.....	28

RESUMO

A descoberta da presença do DNA fetal (*cffDNA*) no plasma e soro materno favoreceu novas abordagens ao diagnóstico não invasivo de desordens fetais (NIPT) promovendo uma melhor assistência ao pré-natal. A rápida detecção de *cffDNA* já é possível em torno da sétima semana gestacional com aumento de sua concentração ao longo do período. Assim, tais características possibilitam o reconhecimento antecipado de afecções fetais e a ação de medidas paliativas. No contexto de uso de *Next-Generation Sequencing* (NGS), diversos métodos têm sido descritos na literatura afirmando uma consistente sensibilidade para detecção de casos de aneuploidias como as síndromes de Down, Edward, Klinefelter, Patau e Turner. Nós propomos um método *in silico*, CAADy (Abordagem Computacional para Detecção de Aneuploidias Fetais), uma estratégia para detecção de aneuploidias fetais. A metodologia é fundamentada na remoção de amostras *outliers* provenientes de diferentes tecnologias de sequenciamento e permitirá a identificação de enfermidades genéticas de origem cromossômica. Neste estudo foi usado o método de *z-score* com interferência interna (IR) calculado pelo desvio absoluto mediano (MAD) para identificar os casos de trissomia dos cromossomos 13, 18 e 21 (2, 12 e 16 casos, respectivamente) em 903 amostras de *cffDNA* disponíveis no portal SRA (<http://www.ncbi.nlm.nih.gov/sra/SRA047257>). Nós detectamos todos os casos de trissomias com 100% de sensibilidade e especificidade. A confirmação dos resultados foi baseada no cariótipo fetal. Assim, nossa metodologia apresentou grande potencial para detectar trissomias.

Palavras-chave: *cffDNA*, NIPT, aneuploidias, trissomias, CAADy.

ABSTRACT

The discovery of the cell free fetal DNA (cffDNA) in maternal plasma and serum allowed the development of new test for fetal disorders in a non-invasive way (NIPT). The detection of the cffDNA is possible around the seventh week, and its concentration increases along the gestational period. Those characteristics allow the early identification of many fetal issues and the palliative actions. Regarding the Next-Generation Sequencing (NGS) tech, several methods has been described on literature asserting a consistent sensibility to detection of aneuploidy cases as Down, Edward and Patau syndromes. We propose a method in *silico*, CAADy (Computational Approach for Detection of Fetal Aneuploidies), a strategy to detection of fetal aneuploidies. That method removes outliers which can come from different technologies of sequencing and it will allow the identification of genetic diseases of chromosomal origin. In this study was used the method of z-score with internal reference calculated by median absolute deviation (MAD) for identify the cases of trisomy of chromosomes 13, 18 and 21 (2, 12 and 16 cases, respectively) included in the 903 samples of cffDNA availables in the SRA portal (<http://www.ncbi.nlm.nih.gov/sra/SRA047257>). We detected all cases of trisomies and with 100% of sensitivity and specificity. The confirmation of results observed in *silico* were based on the karyotype. Therefore, our methodology was presented as great for the detection of trisomies.

Keywords: cffDNA, NIPT, aneuploidies, trisomies, CAADy.

1. INTRODUÇÃO

1.1. Do Material Genômico às Trissomias

O ácido desoxirribonucleico (DNA) é o componente principal que contém as informações necessárias para o perfeito funcionamento celular. Sua estrutura é formada por duas fitas em alfa-hélice compostas de nucleotídeos conectados por ligações de hidrogênio. Os nucleotídeos possuem em sua unidade, um fosfato, uma desoxipentose e uma base nitrogenada. Esta última pode ser uma timina, guanina, citosina ou adenina. Além disso, o DNA pode ser empacotado à nível de cromossomo. Sendo que em nossas células há 23 pares de cromossomos onde, 2 são heterossomos e 44 autossomos (Griffiths *et al*, 2013).

Os cromossomos são herdados durante a divisão celular chamada de meiose. Ela é responsável pela formação dos gametas masculino e feminino. Uma meiose correta divide a célula com 46 cromossomos, em quatro células-filhas com 23 cromossomos cada. Isso porque a meiose ocorre em duas etapas de divisão. Porém, durante esse processo pode haver uma distribuição desigual dos cromossomos. Em consequência disso, o funcionamento celular será comprometido. Se, tais células defeituosas forem transmitidas à geração, na qual, provavelmente, será desenvolvida uma síndrome (Pierce, 2016).

As síndromes cromossomais podem ser resultantes da presença de um cromossomo a mais e são chamadas de trissomias. As trissomias mais comuns na população mundial ocorrem nos cromossomos 13, 18 e 21, conhecidas por síndrome de Patau, Edwards e Down, respectivamente. Diversos sinais e sintomas estão associados a estas síndromes, dentre eles, os distúrbios cardíacos são os mais preocupantes (Schaefer & Thompson, 2015).

Logo, a detecção antecipada destas afecções é um fator-chave para a melhoria do bem-estar na vida desses pacientes. Em geral, isto é feito pela triagem pré-natal que contempla exames capazes de identificar problemas fetais. A ultrassonografia de translucência nugal, por exemplo, é capaz de detectar malformações no feto, principalmente, a síndrome de Down, a partir da 11^a até 14^a semanas gestacionais. Caso houver indicativo de problemas, a amniocentese é recomendada para investigar a situação. No entanto, as principais desvantagens desta técnica são o risco de aborto por ser um procedimento invasivo. Assim, novas técnicas não-invasivas são necessárias para substituir a amniocentese. Uma das alternativas é o teste não-invasivo que utiliza o DNA fetal (cell-free fetal DNA,

cffDNA) presente no plasma materno para quantificar altas ou baixas representações genômicas (Qi *et al.*, 2016).

1.2. DNA Fetal no Plasma Materno

Em 1997, a presença do *cffDNA* no plasma materno foi demonstrada pela primeira vez por Lo *et al* (Lo *et al.*, 1997). Embora, diversas hipóteses busquem uma explicação para o fenômeno, ainda não há evidências concretas que o esclareça. Há fundamentos condizentes de que a placenta é a maior fonte de *cffDNA* e o aumento da idade gestacional exacerba a sua quantidade (Illanes *et al.*, 2007; Bischoff *et al.*, 2012). Uma acentuada elevação foi demonstrada nas últimas 8 semanas gestacionais, indicando uma provável quebra da barreira mãe-feto (Bischoff *et al.*, 2012). Ademais, alguns grupos conseguiram detectá-lo após 35 dias de concepção, suportando uma origem trofoblástica, antes de estabelecer a circulação fetomaternal (Illanes *et al.*, 2005).

A fração de *cffDNA* no plasma materno tem sido calculada por 3 metodologias. Quando o feto é masculino, podem ser facilmente estimados os fragmentos do cromossomo Y usando PCR (reação em cadeia da polimerase). Ademais, há ainda o uso comparativo do padrão de metilação e polimorfismos fetal e materno (Chan *et al.*, 2006; Fan *et al.*, 2012). Contudo, as desvantagens correspondem a restrição de uso para fetos masculinos e o requerimento de informações genéticas parenterais (Kang *et al.*, 2016).

1.3. Teste Pré-natal Não-invasivo

A triagem pré-natal é feita pela combinação de testes sorológicos e ultrassonográficos. Secundariamente, após a identificação positiva na triagem, o *cffDNA* tem sido aplicado como opção adicional (Palomaki, Lambert-Messerlian & Haddow, 2016). A aplicação do DNA derivado do plasma ou soro materno tem sido uma proposta interessante para o diagnóstico molecular, pois minimiza a necessidade de que técnicas invasivas sejam realizadas durante o acompanhamento pré-natal. Tal como, o uso em outras indicações clínicas, providenciando assim, o acesso não-invasivo ao genoma fetal antecipadamente e, o manejo de gestações de risco (Illanes *et al.*, 2007; Bischoff *et al.*, 2012; Lehmann-Werman *et al.*, 2016).

O teste pré-natal não-invasivo (NIPT) é a nova tecnologia usada para detectar trissomias dos cromossomos 13, 18 e 21, através da análise de *cffDNA* na circulação materna (Qi *et al.*, 2016). A fração de *cffDNA* no plasma materno é cerca de 5-20% sendo esta uma concentração crítica para a acurácia deste teste (Kang *et al.*, 2016). A quantidade de *cffDNA* é elevada quando o feto é aneuploide e influenciada pelo crescimento fetal (Bischoff *et al.*, 2012).

Alta sensibilidade ($97\pm 100\%$) e especificidade (100%) é exibida pelo NIPT para detecção de trissomias autossômicas. Após sua implementação na prática clínica, a preferência por testes de alto risco diminuiu e a taxa de detecção de aneuploidias aumenta sem a necessidade de numerosos procedimentos invasivos (Qi *et al.*, 2016). Entretanto, o principal obstáculo para o rápido desenvolvimento e baixo custo diagnóstico do NIPT é quando há baixa fração de *cffDNA* (<4%) disponível, principalmente, durante o início da gestação, onde a quantidade de DNA fetal livre é proporcional a idade gestacional (Norton *et al.*, 2012; Wang *et al.*, 2013).

1.4. Sequenciamento Paralelo Massivo do Genoma (MPGS)

A exploração dos fragmentos de *cffDNA* na circulação materna outorgou o emprego do sequenciamento de nova geração (*next-generation sequencing*, NGS) como uma ferramenta tecnológica auxiliadora na identificação de trissomias fetais e de outras desordens genéticas, contrapondo o uso de técnicas invasivas como a amniocentese (Chen *et al.*, 2011; Bianchi *et al.*, 2012). Concomitante com a implementação futura de sequenciamento de baixo custo, o *cffDNA* pode ser uma alternativa de primeira escolha na triagem neonatal (Palomaki *et al.*, 2016). Apesar do seu lançamento comercial em 2011, há pouco conhecimento acerca da sua performance na atual prática clínica (Valderramos *et al.*, 2016).

Uma maneira de quantificar o *cffDNA* é submeter as amostras de plasma ou soro ao *Massively Parallel Sequencing* (MPGS), seguido pelo mapeamento e enumeração dos fragmentos por cromossomo de origem. A partir do número de sequências alinhadas para cada cromossomo é possível mensurar altas representações quando há aneuploidia fetal. Assim, esta metodologia de detecção de trissomias não requer a diferenciação do DNA fetal do materno. Além disso, pode revelar diversas características dos ácidos nucleicos presentes no plasma, como a

distribuição de *mRNA*, epigenética, metilação do DNA e modificação de histonas, por exemplo, entre outras informações (Fan *et al.*, 2008).

A sensibilidade da tecnologia *MPGS* é potencialmente mais sensível e com menor risco de contaminação do que a PCR em tempo real (Chiu *et al.*, 2008). Teoricamente, não há um viés biológico ou analítico do *MPGS* e espera-se que haja o sequenciamento uniforme do genoma completo. No entanto, tem sido reportado o contrário (Kozarewa *et al.*, 2009). Chiu *et al* (2008), por exemplo, demonstrou que o conjunto de sequências de DNA por amostra pode variar de corrida para corrida de sequenciamento. Alguns cromossomos apresentam grande variação no número de fragmentos sequenciados de amostra para amostra. Este problema está relacionado ao conteúdo CG dos cromossomos que interfere na sensibilidade de detecção das aneuploidias. Todavia, as aneuploidias humanas mais comuns (13, 18 e 21, por exemplo) possuem uma baixa variação no conteúdo CG e portanto, apresentam alta sensibilidade de detecção (Fan *et al.*, 2008). Ademais, a mensuração do representativo genômico dos cromossomos 13 e 18 é menos precisa do que para o cromossomo 21. Como a variação do conteúdo CG contribui para tal acontecimento, uma correção quantitativa para minimizá-la pode permitir o diagnóstico acurado destas trissomias (Kozarewa *et al.*, 2009).

1.5. Metodologias de Detecção para Aneuploidias Fetais

As abordagens de detecção de trissomias consistem basicamente na representação percentual da quantidade de sequências únicas para o cromossomo de interesse dentro do conjunto de dados, seguido do cálculo de *z-score*. Desta maneira, o *z-score* nos demonstrará o número de desvios-padrão em relação à média do conjunto amostral normal. Todavia, cada metodologia separadamente apresenta seu diferencial. A aplicação do *z-score* padrão, sem nenhuma remoção do viés GC, foi aplicado por Chiu *et al* (2008) para detectar a síndrome de Down. Em contrapartida, Chen *et al* (2011) descreveu uma abordagem de *z-score* com a estratégia de remover tal interferência. Essa abordagem também aplicada por Lau *et al* (2011) consiste em utilizar cromossomos referência para excluir o viés GC. Diferente das abordagens anteriores, o *NIFTY* (Jiang *et al.*, 2012) compõe-se da combinação de um teste-t de *Student* otimizado com regressão polinomial linear ponderada.

O grupo desenvolvedor do *NIFTY* fez uma avaliação de desempenho das quatro abordagens usando o coeficiente de variação. O maior coeficiente de variação encontrado foi no *z-score* padrão quando comparado às demais metodologias. Logo, ela possui baixa sensibilidade de detecção para trissomias. Já, a performance do *z-score* com correção GC apresentou 99% de sensibilidade e especificidade para trissomias 13, 18 e 21. No entanto, os resultados não foram precisos para aneuploidias sexuais, devido à dificuldade de identificação do gênero fetal. Assim como o *z-score* padrão, o uso do chrIR para correção do viés GC também apresentou alto coeficiente de variação e alto risco de falso negativo para detectar aneuploidias XXY e XYY. Em contraste, o *NIFTY* exibiu alta acurácia na detecção de aneuploidias sexuais, tais como, XO, XXY e XYY (Jiang *et al.*, 2012).

1.6. Abordagem Computacional para Detecção de Aneuploidias Fetais (CAADy).

As metodologias de *z-score* padrão calculadas pelo desvio-padrão e média, em geral, apresentam ainda problemas com falso-positivos e falso-negativos nos testes. Para resolver isso, propomos um novo cálculo de *z-score* e mais acurado para detectar com maior sensibilidade e especificidade tais casos. A nossa abordagem é baseada no cálculo de *z-score* modificado, que utiliza o desvio absoluto mediano (MAD) e a mediana como medidas mais sensíveis à presença de *outliers* nas amostras. Assim, altas ou baixas representações genômicas podem ser testadas e identificadas no conjunto amostral.

2. REFERÊNCIAS

1. Bianchi DW, *et al.* CARE Study Group: DNA sequencing versus standard pre-natal aneuploidy screening. *N Engl J Med.* 2012; 370(9):799–808.
2. Bischoff FZ, *et al.* Cell-free fetal DNA and intact fetal cells in maternal blood circulation: implications for first and second trimester non-invasive prenatal diagnosis. *Hum Reprod Update.* 2012; 8, 493–500.
3. Chan KC, *et al.* Hypermethylated RASSF1A in maternal plasma: a universal fetal DNA marker that improves the reliability of noninvasive prenatal diagnosis. *Clin Chem.* 2006; 52:2211–2218.
4. Chen EZ, *et al.* Noninvasive prenatal diagnosis of fetal trisomy 18 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One.* 2011; 6:e21791.
5. Chiu RW, *et al.* Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci USA.* 2008; 105(51):20458–20463.
6. Fan HC, *et al.* Noninvasive diagnosis of fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad.* 2008;105:16266-71.
7. Fan HC, *et al.* Non-invasive prenatal measurement of the fetal genome. *Nature.* 2012; 487(7407):320–324.
8. Griffiths A J F, *et al.* *Introdução à Genética.* 10ª edição. Rio de Janeiro: Ed. Guanabara Koogan, 2013.
9. Illanes S, Avent N, Soothill PW. Cell-free fetal DNA in maternal plasma: an important advance to link fetal genetics to obstetric ultrasound. *Ultrasound Obstet Gynecol.* 2005; 25(4):317–22.
10. Illanes S, *et al.* Early detection of cell-free fetal DNA in maternal plasma. *Early Hum Dev.* 2007; 83:563-566.
11. Jiang F, *et al.* Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis methodology for fetal autosomal and sex chromosomal aneuploidies. *BMC Med Genomics.* 2012; 5:57.
12. Kang X, *et al.* An Advanced Model to Precisely Estimate the Cell-Free Fetal DNA Concentration in Maternal Plasma. *PLoS One.* 2016; 11(9):e0161928.

13. Kozarewa I, *et al.* Amplification-free Illumina sequencing-library preparation facilitates improved mapping and assembly of (G+C)-biased genomes. *Nat Methods.* 2009; 6: 291-295.
14. Lau TK, *et al.* Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal plasma DNA sequencing. *J Matern Fetal Neonatal Med.* 2011; 25(8):1370–1374.
15. Lehmann-Werman R, *et al.* Identification of tissue-specific cell death using methylation patterns of circulating DNA. *Proc. Natl Acad. Sci.* 2016; 113(13), E1826–E1834.
16. Lo YMD, *et al.* Presence of fetal DNA in maternal plasma and serum. *Lancet.* 1997; 350:485–7.
17. Norton ME, *et al.* Non-Invasive Chromosomal Evaluation (NICE) Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and trisomy 18. *Am J Obstet Gynecol.* 2012; 207(2):137 e131-138.
18. Palomaki GE, Lambert-Messerlian GM, Haddow JE. Serum versus cfDNA screening: Where have all the trisomies gone? *Am J Obstet Gynecol.* 2016; 215:583-7.
19. Pierce, BA. *Genética: um enfoque conceitual.* 5 edição. Rio de Janeiro: Guanabara Koogan, 2016.
20. Qi G, *et al.* Noninvasive prenatal testing in routine clinical practice for a high-risk population: Experience from a center. *Medicine (Baltimore).* 2016; 95(41):e5126.
21. Schaefer GB, Thompson JN. *Genética médica: uma abordagem integrada.* Porto Alegre: AMGH, 2015.
22. Valderramos SG, *et al.* Cell-free DNA screening in clinical practice: Abnormal autosomal aneuploidy and microdeletion results. *Am J Obstet Gynecol.* 2016; pii:S0002-9378(16)30384-2.
23. Wang E, *et al.* Gestational age and maternal weight effects on fetal cell-free DNA in maternal plasma. *Prenat Diagn.* 2013; 33(7):662–666.

3. NORMAS DO PERIÓDICO

O presente estudo será submetido a revista *PeerJ* que publica conhecimentos científicos mundialmente com o acesso aberto ao público.

3.1. Preparo da submissão

- Não é necessário enviar *cover letter* para aceitação de pré-submissões e submissões.
- Incluir no texto o número de linhas; margem de 2.5 cm em todas as bordas; fonte *times new roman*, tamanho 12.
- Arquivos normalmente esperados: texto manuscrito, algumas figuras, tabelas, código ou dado-bruto e informações suplementares.
- Os dados devem ser disponibilizados através de *link* e no arquivo suplementar. Em alguns casos, não será obrigatório carregá-los no arquivo suplementar. As exceções são situações nas quais os dados não estão disponíveis para publicações ou são muito extensos. Nestas circunstâncias, deverão ser disponibilizados em repositório online.
- As submissões podem ser enviadas em arquivo único contendo figuras e tabelas no texto ou em múltiplos arquivos, sendo estas separadas.
- Durante a submissão devem ser preenchidos os seguintes campos no formulário online: título, *abstract*, autores, assunto, palavras-chave, editores/revisores, declarações e arquivos.

3.2. Informações básicas

As informações dos autores e suas afiliações podem ser inseridas na primeira página do manuscrito, abaixo do título.

3.3. Padrão de Seções

Autores

Abstract (máx. de 500 palavras ou 3.000 caracteres contendo: contexto geral, métodos, resultados e discussão).

Introdução; Materiais e Métodos; Resultados; Discussão; Conclusão; Agradecimentos; Referências

3.4. Formato de referências

Citações no texto:

- Para três ou menos autores, listar todos os nomes. Para mais de 4 autores, abreviar com o primeiro autor *et al.*
- Múltiplas referências sobre o mesmo item devem ser separadas com ponto e vírgula (;) e obedecer a ordem cronológica.
- Referências do mesmo autor no mesmo ano podem ser diferenciadas por letras (Smith, 2001a; Smith, 2001b).
- Artigos citados que estão em processo de aceitação, incluir na lista de referências e providenciar uma cópia nas informações suplementares.
- Citações de trabalhos não publicados, em preparação ou sob revisão com dados não publicados, usar as iniciais do autores e sobrenomes apenas no texto e não incluir na seção de referências.
- Referências pessoais devem ser evitadas, mas, se absolutamente necessário, podem ser referenciadas como “pers. comm.”, incluindo nomes individuais e ano.

Seção de referências:

- Cada referência deve ser listada usando o seguinte formato: lista de autores com iniciais. Ano da publicação. Título do artigo. Nome da revista, volume: páginas. DOI (quando disponível). Exemplo: Smith JL, Jones P, Wang X. 2004. Investigating ecological destruction in the Amazon. *Journal of the Amazon Rainforest* 112:368-374. DOI: 10.1234/amazon.15886.
- Referências para documentos em geral como patentes, relatórios técnicos de agências do governo ou grupos de pesquisa científica. Incluir nome dos autores, título da página ou artigo, data de publicação, nome do editor, sempre que possível, URL, acesso de datas e número de identificação como número de patentes. Exemplo: Boettiger C. 2013. knitcitations: citations for knitr markdown files. Available at <https://github.com/cboettig/knitcitations> (accessed 10 July 2012)
- Exemplo de referências para livros: James FY. 2010. *Understanding corn and wheat*. Oxford: Oxford University Press.

- Exemplo de teses: Blair HJ. 1989. Structural modifications of the fern genus *Lecanopteris*(Polypodiaceae). D. Phil. Thesis, Cambridge University.
- Exemplo de webpage: Johnson S. 2010. Italian Plants. Available at <http://www.italianplants.com>(accessed 22 March 2011).
- Exemplo de abstract: Thomas D, Scharfenecker U, Schiltmeyer B. 2006. Low potential for drug-drug interaction of lacosamide [abstract no. 2.235]. *Epilepsia* 47 (Suppl. 4): 200
- A seção de referências podem ser ordenadas por autor, ano e título. Todas as citações no texto devem aparecer na lista de referências e todas as referências listadas devem ser citadas no manuscrito.

3.5. Referenciando tabelas e figuras

As figuras e tabelas protegidas por direitos autorais devem apresentar autorização, assim como, deve ser incluída na legenda da tabela os créditos de permissão.

3.6. Tipos de arquivo

Manuscrito

- Para o texto manuscrito, a PeerJ aceita os formatos .DOC (MS Word), .DOCX (MS Word), .TeX, .ODT (OpenDocument) e .RTF.
- O texto deve ser justificado para margem esquerda.

Figuras

- Indicar o número da figura no seu nome de acordo com a ordem de aparecimento no texto.
- A citação das figuras deve seguir a ordem de aparecimento no texto.
- O título é uma sentença única.
- A legenda é uma descrição.
- Os formatos são EPS, JPEG, PDF ou PNG.
- Espaços em branco devem ser eliminados em torno das figuras.
- O tipo e tamanho da fonte devem ser legíveis.
- Tamanho mínimo: 900x900 pixels.

Tabelas

- Indicar o número da tabela no seu nome de acordo com a ordem de aparecimento no texto.
- A citação das tabelas deve seguir a ordem de aparecimento no texto.
- O título é uma sentença única.
- A legenda é uma descrição.
- Os formatos são .DOC (MS Word), .DOCX (MS Word), .ODT (Open Doc).
- As tabelas devem caber numa única página 21.6 x 28 cm com 2.5 de margem.
- Inserir notas de rodapé abaixo da tabela para explicar abreviaturas.

3.7. Estilo de considerações

Unidades, símbolos matemáticos e abreviações

- Usar o Sistema Internacional de Unidades.

Texto manuscrito

- As submissões que resultem em um manuscrito final composto de 45 páginas ou mais podem incorrer em uma taxa de serviço adicional para cobrir os custos de produção.
- Os artigos devem ser escritos em inglês claro e inequívoco para uma audiência internacional.
- As normas estabelecidas de escrita acadêmica dentro de seu campo devem ser seguidas.
- A nomenclatura biológica e médica deve respeitar as diretrizes reconhecidas estabelecidas pelos regulamentos de comitês internacionais ou órgãos autorizados para áreas específicas, conforme aplicável.

Figuras

- As figuras devem ser minimamente processadas ou manipuladas para adicionar rótulos, setas ou mudar contraste ou brilho se aplicado a toda a imagem, bem como aos controles.
- Todas as figuras devem ser citadas no texto. Usar a abreviatura “Fig.” ou “Figura” quando iniciando a frase.
- As legendas devem aparecer no texto manuscrito, ao lado da figura, descrevendo claramente o conteúdo desta.

1 **4. ARTIGO**

2

3 **ABORDAGEM COMPUTACIONAL PARA DETECÇÃO DE ANEUPLOIDIAS**

4 **FETAIS**

5 A. K. N. Mendonça¹, J. E. Kroll¹, S. J. Souza^{1*}

6 ¹ Instituto do Cérebro, Universidade Federal do Rio Grande do Norte, Natal 59064-560, RN,
7 Brasil

8 *Autor correspondente, e-mail: sandro@neuro.ufrn.br

9 **RESUMO**

10 A descoberta da presença do DNA fetal (*cffDNA*) no plasma e soro materno favoreceu novas
11 abordagens ao diagnóstico não invasivo de desordens fetais (NIPT) promovendo uma melhor
12 assistência ao pré-natal. A rápida detecção de *cffDNA* já é possível em torno da sétima semana
13 gestacional com aumento de sua concentração ao longo do período. Assim, tais características
14 possibilitam o reconhecimento antecipado de afecções fetais e a ação de medidas paliativas.
15 No contexto de uso de *Next-Generation Sequencing* (NGS), diversos métodos têm sido
16 descritos na literatura afirmando uma consistente sensibilidade para detecção de casos de
17 aneuploidias como as síndromes de *Down*, *Edward*, *Klinefelter*, *Patau* e *Turner*. Nós
18 propomos um método *in silico*, CAADy (Abordagem Computacional para Detecção de
19 Aneuploidias fetais), uma estratégia para detecção de aneuploidias fetais. A metodologia é
20 fundamentada na remoção de amostras *outliers* provenientes de diferentes tecnologias de
21 sequenciamento e permitirá a identificação de enfermidades genéticas de origem
22 cromossômica. Neste estudo foi usado o método de z-score com interferência interna (IR)
23 calculado pelo desvio absoluto mediano (MAD) para identificar os casos de trissomia dos
24 cromossomos 13, 18 e 21 (2, 12 e 16 casos, respectivamente) em 903 amostras de *cffDNA*
25 disponíveis no portal SRA (<http://www.ncbi.nlm.nih.gov/sra/SRA047257>). Nós detectamos
26 todos os casos de trissomias com 100% de sensibilidade e especificidade. A confirmação dos
27 resultados foi baseada no cariótipo fetal. Assim, nossa metodologia apresentou grande
28 potencial para detectar trissomias.

29 Palavras-chave: *cffDNA*, NIPT, aneuploidias, trissomias, CAADy.

30

31 **INTRODUÇÃO**

32 A descoberta do DNA fetal livre no plasma e soro materno (Lo *et al.*, 1997) permitiu o
33 desenvolvimento de novos testes não-invasivos para detecção de desordens fetais (NIPT)

34 (Bianchi *et al.*, 2012; Jiang *et al.*, 2012; Song *et al.*, 2013). A aplicação do DNA derivado do
35 plasma ou soro materno tem sido uma proposta interessante para o diagnóstico molecular,
36 pois minimiza a necessidade de que técnicas invasivas sejam realizadas durante o
37 acompanhamento pré-natal (Bischoff *et al.*, 2002; Illanes *et al.*, 2007). Tal como, o uso em
38 outras indicações clínicas, providenciando assim, o acesso não-invasivo ao genoma fetal
39 antecipadamente e, o manejo de gestações de risco (Lehmann-Werman *et al.*, 2016).

40 A fração de *cffDNA* no plasma materno é cerca de 5-20% sendo esta uma concentração
41 crítica para a acurácia deste teste (Kang *et al.*, 2016). Não obstante, alta sensibilidade
42 (97~100%) e especificidade (100%) tem sido exibida pelo NIPT para detecção de trissomias
43 autossômicas (Qi *et al.*, 2016). Ademais, após a sua implementação na prática clínica, a
44 preferência por testes de alto risco foi diminuindo e a taxa de detecção de aneuploidias foi
45 aumentando sem a necessidade de numerosos procedimentos invasivos (Kang *et al.*, 2016).

46 Uma maneira de quantificar o *cffDNA* é submeter as amostras de plasma ou soro ao
47 sequenciamento *shotgun*, seguido pelo mapeamento e enumeração dos fragmentos por
48 cromossomo de origem. A partir do número de sequências alinhadas para cada cromossomo é
49 possível mensurar altas representações quando há aneuploidia fetal. Assim, esta metodologia
50 de detecção de trissomias não requer a diferenciação do DNA fetal do materno (Fan *et al.*,
51 2008).

52 As abordagens de detecção de trissomias consistem basicamente na representação
53 percentual da quantidade de sequências únicas para o cromossomo de interesse dentro do
54 conjunto de dados, seguido do cálculo de *z-score*. Desta maneira, o *z-score* nos indicará o
55 número de desvios-padrão em relação a média do conjunto amostral normal (Chiu *et al.*,
56 2008; Chen *et al.*, 2011; Lau *et al.*, 2011). Porém, já foi demonstrada a influência na
57 quantidade de sequências geradas no sequenciamento, por regiões ricas em nucleotídeos
58 guanina e citosina, que são distribuídas distintamente nos cromossomos (Dohm *et al.*, 2008;
59 Alkan *et al.*, 2009; Kozarewa *et al.*, 2009; Aird *et al.*, 2011). Por isso, a normalização do
60 conteúdo CG é proposta por alguns autores no cálculo do *z-score* (Chen *et al.*, 2011; Lau *et al.*,
61 2011). A aplicação do cromossomo referência interna (chrIR), com conteúdo CG similar
62 ao cromossomo da trissomia pode corrigir tal viés (Lau *et al.*, 2011). Baseando-se nisso,
63 propomos aqui uma nova e mais robusta estratégia *in silico*, CAADy (*Computational*
64 *Approach for Detection of Fetal Aneuploidies*), que combina o uso do *z-score* com referência
65 interna (IR), porém, calculado a partir do desvio mediano absoluto (MAD) para remover
66 valores extremos (*outliers*) do conjunto amostral e identificar os casos de trissomias.

67

68 MATERIAIS E MÉTODOS

69 Coleção de dados

70 A coleção de amostras foi obtida no portal SRA
71 (<http://www.ncbi.nlm.nih.gov/sra/SRA047257>). Elas são provenientes de sangue venoso
72 periférico de 903 gestantes, e foram sequenciadas seguindo o protocolo modificado da
73 *Illumina*, através das plataformas *Illumina GAIIx* e *HiSeq 2000*, produzindo uma variação em
74 torno de 2-4 milhões de sequências. O grupo de amostras são compostas por gestantes de 20-
75 45 anos de idade e período gestacional entre 10-34 semanas, correspondentes do primeiro ao
76 terceiro trimestre. De acordo com os resultados da cariotipagem completa usando fluido
77 amniótico, 866 são euploides e 37 aneuploides do grupo de 903 fetos. Os casos de
78 aneuploidias incluem: 2 trissomias do cromossomo 13; 12 trissomias do cromossomo 18; 16
79 trissomias do cromossomo 21; 4 casos de síndrome de Turner 45,X (3 casos típicos e 1
80 mosaico); 2 casos de síndrome Klinefelter XXY e 1 caso de síndrome de Jacobs XYY (Tabela
81 1).

82

83 **Tabela 1.** Cariotipo fetal.

Cariótipo	n.	Classificação
46, (XX) / 46, (XY)	866	Euploides
47, (XX/XY), +13	2	Síndrome de Patau
47, (XX/XY), +18	12	Síndrome de Edwards
47, (XX/XY), +21	16	Síndrome de Down
45, X	3	Síndrome de Turner
45,X/46,XX	1	
47, XXY	2	Síndrome de Klinefelter
47, XYY	1	Síndrome de Jacobs
Total	903	

84

85 Processamento de dados amostrais

86 As sequências foram convertidas do formato SRA para FASTQ (*software fastq-dump*). O
87 alinhamento foi feito pelo *Bowtie-1.1.2*, adequado para alinhar sequências de 35 pb
88 produzidas no sequenciamento. Em seguida, a contagem das sequências únicas dos
89 cromossomos foi obtida para cada amostra utilizando *scripts* próprios.

90

91 Conteúdo CG dos cromossomos

92 Para corrigir o viés CG das amostras, foi mensurado o conteúdo CG de todos os cromossomos
 93 do conjunto amostral. O cálculo foi baseado na média percentual de CG das sequências de
 94 cada cromossomo. O chrIR para cada trissomia foi selecionado com base na proximidade do
 95 conteúdo CG e na maior capacidade de detecção, tanto no *z-score* IR, quanto no *z-score* IR
 96 modificado.

$$97 \quad \%CG = \frac{\left| \sum \frac{CG_{chrN}}{Total\ pb} \right|}{n}$$

98

99 **Representação da fração genômica (FGR)**

100 Para cada trissomia foi calculada a razão entre o número de sequências do respectivo
 101 cromossomo e o total presente em cada amostra. Assim, representando a porcentagem relativa
 102 na amostra. Além disso, o FGR também foi calculado para o chrIR escolhido. Em seguida, foi
 103 obtida a razão entre os FGR do cromossomo-alvo e chrIR.

$$104 \quad FGR = \frac{UR_{chrN}}{URT_{amostra}},$$

105 onde UR ChrN são as sequências únicas do cromossomo de interesse e URT samples é o total
 106 de sequências únicas das amostras

$$107 \quad FGR_{relativo} = \frac{\%UR_{chrN}}{\%UR_{chrIR}},$$

108 onde % UR ChrN é a porcentagem de sequências únicas do cromossomo de interesse ou FGR
 109 ChrN e % UR ChrIR é a porcentagem de sequências únicas do chrIR ou FGR ChrIR.

110

111 **Separação das trissomias do conjunto de dados**

112 Os casos de trissomias foram separados do conjunto euploide, através do *z-score* IR
 113 modificado, calculado com base no desvio absoluto mediano (MAD), ao invés do desvio-
 114 padrão. As amostras com *z-score* IR modificado ≥ 3 foram marcadas como aneuploides. A
 115 equação é dada por:

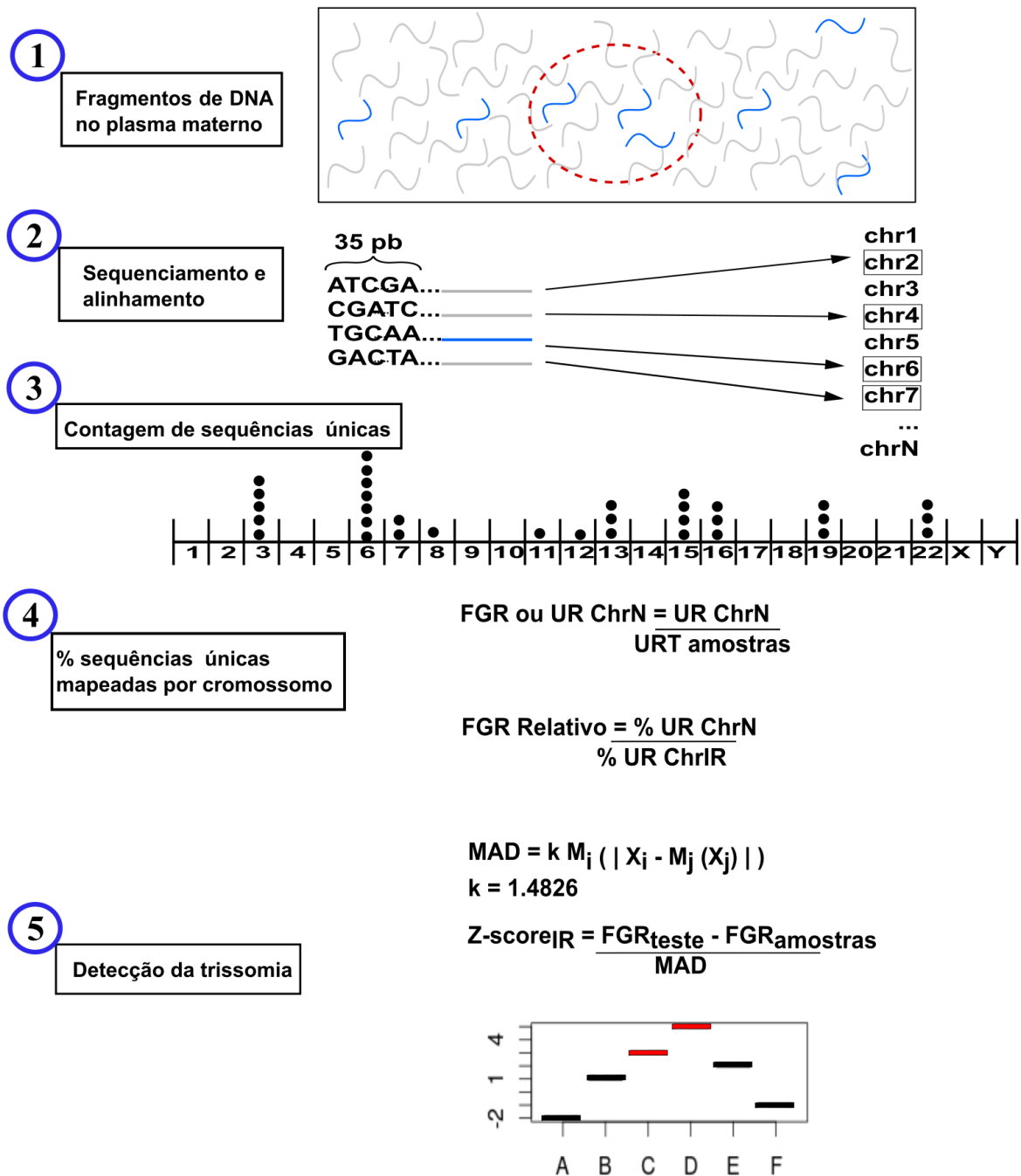
$$116 \quad Z-score_{IR} = \frac{(FGR_{teste} - FGR_{amostras})}{MAD},$$

117 onde o FGR_{teste} é da amostra teste e o $FGR_{amostras}$ é a mediana referente ao conjunto de
 118 amostras, e

119 $MAD = k M_i (|X_i - M_j(X_j)|)$ é o desvio absoluto mediano e

120 $k = 1.4826$ uma constante.

121 A figura 1 resume a metodologia aplicada pelo CAADy para identificar as trissomias.



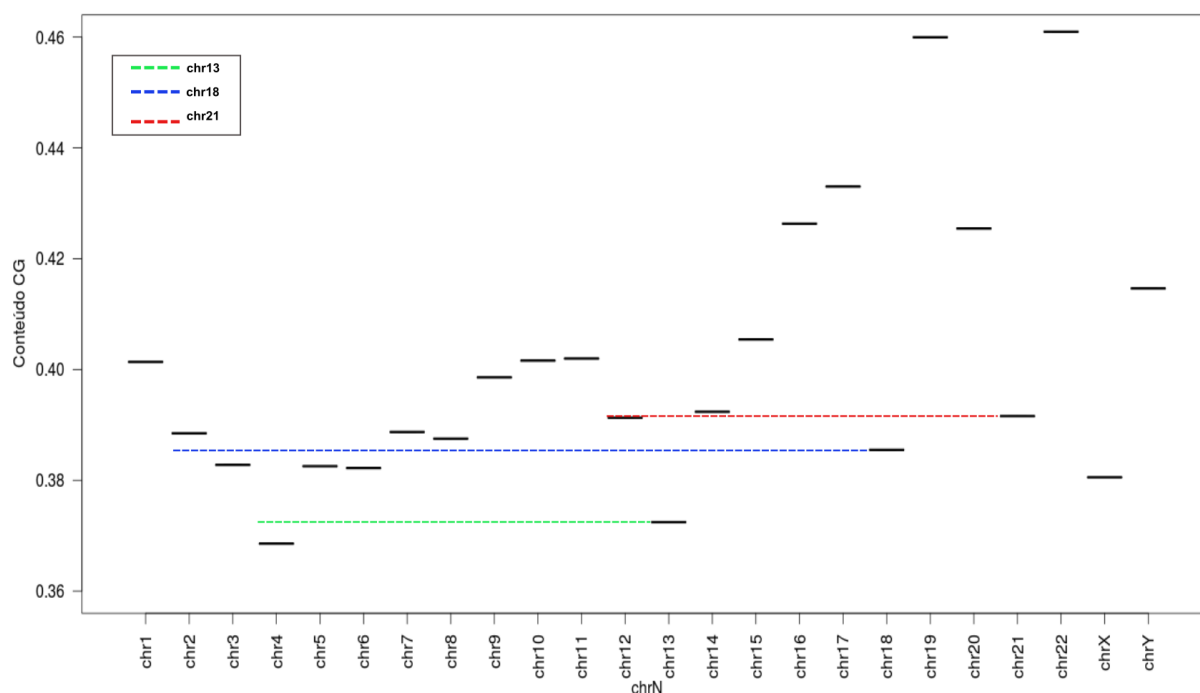
122 **Figura 1.** Esquematização da metodologia de detecção abordada pela CAADy.

123

124 RESULTADOS

125 Viés CG dos Cromossomos

126 O teste para seleção dos cromossomos referência interna foi baseado na proximidade do
 127 conteúdo CG para o cromossomo da trissomia (Fig. 2). O único com CG próximo ao chr13
 128 (0.3724) foi o chr 4 (0.3685). Já para o chr18 foram testados seis cromossomos: chr2
 129 (0.3884), chr3 (0.3828), chr5 (0.3825), chr6 (0.3822), chr7 (0.3887) e chr8(0.3875). Para
 130 trissomia do 21 (0.3916) foram testados os chr12 e chr14 (0.3913 e 0.3923, respectivamente).



132 **Figura 2.** Seleção do chrIR baseada na proximidade do conteúdo CG com o cromossomo da
 133 trissomia.

134

135 **Capacidade de detecção do CAADy**

136 Os chrIR escolhidos constaram naqueles que apresentassem melhor desempenho de
 137 detecção em ambos os testes de z-score, calculados pelo desvio-padrão ou MAD. O melhor
 138 desempenho para todas trissomias foi obtido pelo z-score IR modificado que utiliza o MAD,
 139 ao invés do desvio-padrão. Para a T13, o chrIR selecionado foi o cromossomo 4; já para T18,
 140 o melhor chrIR em ambos os testes foi o cromossomo 6, apresentando 100% de sensibilidade
 141 e especificidade. O chr14 também teve sucesso para T21 nos dois testes de z-score (Tabela 2).

142

		Abordagem					
Aneuploidias	chrIR	Z-score IR modificado			Z-score IR		
		n.	S	E	n.	S	E
T13	chrIR4	2	100%	100%	3	100%	99,88%
T18	chrIR2	12	100%	100%	14	100%	99,77%
	chrIR3	12	100%	100%	14	100%	99,77%
	chrIR5	13	100%	99,88%	16	100%	99,55%
	chrIR6	12	100%	100%	12	100%	100%
	chrIR7	12	100%	100%	14	100%	99,77%

	chrIR8	13	100%	99,88%	14	100%	99,77%
	chrIR12	16	100%	100%	19	100%	99,66%
T21	chrIR14	16	100%	100%	16	100%	100%

143

144 **Tabela 2.** Comparação entre os métodos de z-score IR (desvio-padrão) e z-score IR
145 modificado (MAD) para detecção dos casos de trissomias dos cromossomos 13, 18 e 21 (T13,
146 T18 e T21, respectivamente). ChrIR – cromossomo referência interna; S – sensibilidade; E –
147 especificidade; em negrito, os chrIR selecionados.

148

149 **Trissomias identificadas no conjunto amostral**

150 Todos os casos esperados de trissomias foram detectados pelo CAADy para os cromossomos
151 13, 18 e 21 (2, 12 e 16 casos, respectivamente) representando 100% de especificidade e
152 sensibilidade do método (Fig. 3). Os chrIR utilizados para normalizar o possível viés CG
153 dessas aneuploidias foram chr4, chr6 e chr14, respectivamente. Isto relaciona-se à
154 similaridade do conteúdo CG de cada chrIR com os cromossomos testados das trissomias
155 (Fig. 2).

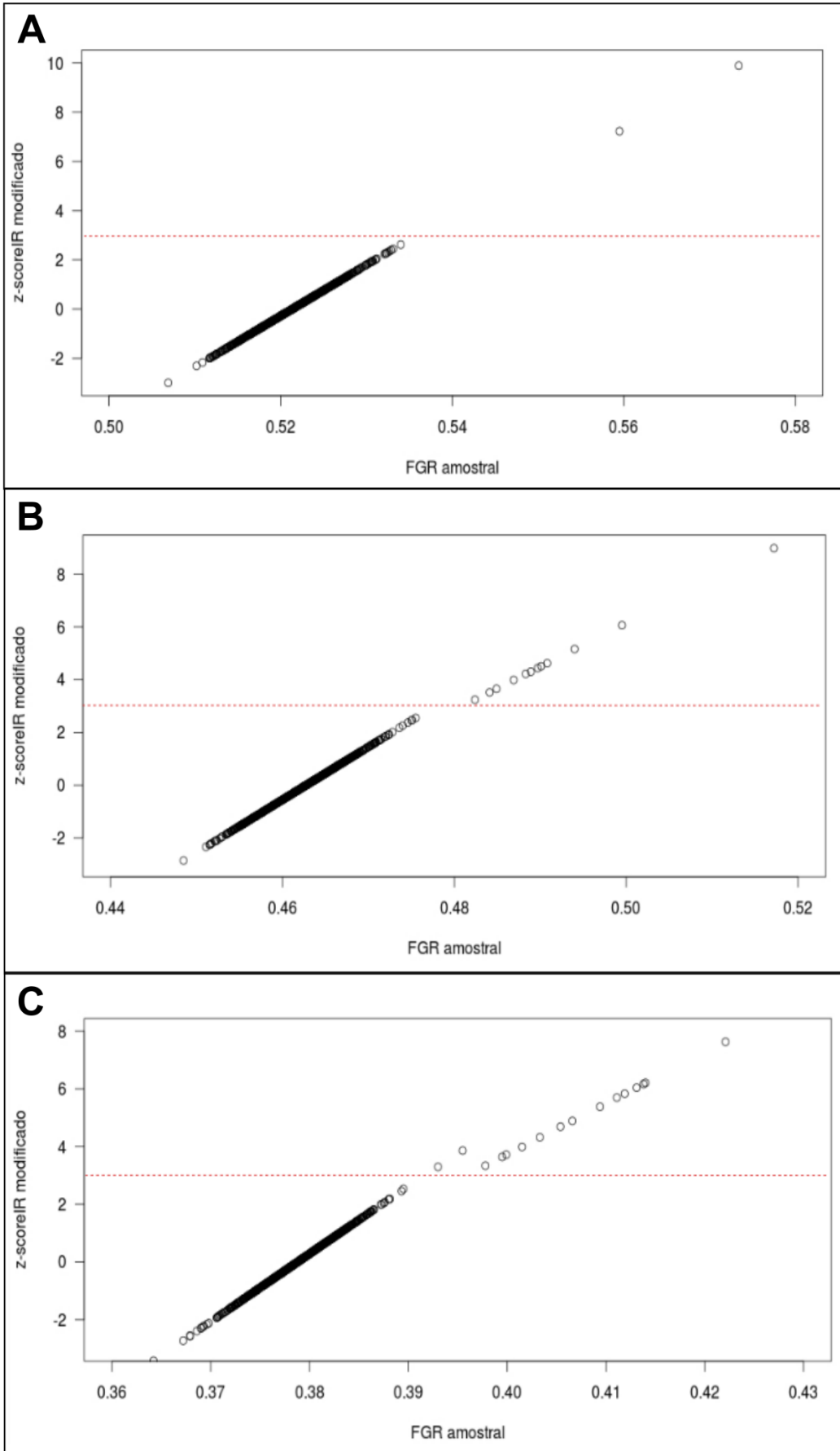
156

157 **Teste de validação com *dataset* independente**

158 Um grupo de amostras sem trissomia (n. 500) foi separado aleatoriamente como referência
159 para testar a validade do método de *z-score* IR modificado. As demais amostras (n. 403)
160 foram testadas cada uma como amostras testes do cálculo. Ao final, foram identificados todos
161 os casos de trissomias conforme a tabela 2.

162

163



165 **Figura 3.** Trissomias detectadas pelo CAADy. A – 2 casos de trissomia chr13; B – 12 casos
166 de trissomia do chr18; C – 16 casos de trissomia do chr21.
167

168 DISCUSSÃO

169 Devido a sua alta sensibilidade e especificidade (Qi *et al.*, 2016), o NIPT tem sido cada
170 vez mais aplicado na prática clínica para substituir o uso de técnicas invasivas durante a
171 triagem pré-natal (Chen *et al.*, 2011; Bianchi *et al.*, 2012). Contudo, algumas barreiras ainda
172 dificultam a sua implementação clínica, como a quantidade de *cffDNA* disponível no plasma
173 ou soro materno; o custo do sequenciamento; interferência do viés GC e a burocracia dos
174 comitês de saúde (Norton *et al.*, 2012; Wang *et al.*, 2013). Apesar disso, o benefício de
175 analisar o genoma fetal ainda no útero materno, sem a necessidade do uso de técnicas que
176 poderiam causar o aborto, possibilita o manejo antecipado das gestações de risco (Illanes *et*
177 *al.*, 2007; Bischoff *et al.*, 2012; Lehmann-Werman *et al.*, 2016).

178 A detecção de trissomias dos cromossomos 13, 18 e 21 utilizando-se do *cffDNA* (Qi *et*
179 *al.*, 2016) é feita a partir do sequenciamento das amostras, mapeamento por cromossomo e
180 enumeração das sequencias (Fan *et al.*, 2008). Em seguida, são submetidas, em geral, ao teste
181 *z-score* para mensurar altas representações quando há aneuploidia fetal. O *z-score* tem
182 demonstrado boa performance para tal finalidade (Chiu *et al.*, 2008; Chen *et al.*, 2011; Lau *et*
183 *al.*, 2011). Porém, o seu cálculo é feito a partir da média e desvio-padrão amostral, o que pode
184 resultar em alguns problemas. Por exemplo, tanto a média, quanto o desvio-padrão são
185 fortemente impactados por *outliers*, que diminui a precisão do método na detecção de casos
186 onde a fração de *cffDNA* referentes as trissomias é baixa (Leys *et al.*, 2013). Assim, nossa
187 proposta de aperfeiçoar o *z-score* com o uso do MAD mostrou-se vantajosa pois aumentou a
188 sensibilidade do método proporcionando uma análise mais refinada e robusta do conjunto de
189 dados, sendo totalmente imune ao tamanho amostral e à presença de *outliers* (Leys *et al.*,
190 2013).

191

192 CONCLUSÃO

193 O presente estudo demonstrou a aplicação de uma abordagem mais robusta e acurada
194 para detecção de trissomias fetais. Além disso, o uso do *z-score* IR modificado é adequado na
195 remoção dos *outliers*, perfazendo-se ideal em casos de altas ou baixas representações

196 cromossômicas aneuploides. De modo, nosso intuito foi alcançado e esperamos contribuir
197 futuramente com mais conhecimentos voltados para o NIPT.

198 **AGRADECIMENTOS**

199 Este projeto é suportado pelo Laboratório de Bioinformática da Universidade Federal do
200 Rio Grande do Norte, UFRN. Agradecemos a disponibilidade dos dados amostrais que foram
201 sequenciados pelo BGI-Shenzhen.

202 **REFERÊNCIAS**

203

204 Aird D, Ross MG, Chen WS, Danielsson M, Fennell T, Russ C, Jaffe DB, Nusbaum C, Gnirke
205 A. 2011. Analyzing and minimizing PCR amplification bias in Illumina sequencing libraries.
206 *Genome Biol*, 12 (2): R18-10.1186/gb-2011-12-2-r18. DOI: 10.1186/gb-2011-12-2-r18.

207 Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO,
208 Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE. 2009. Personalized copy
209 number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41:
210 1061–1067. DOI: 10.1038/ng.437.

211 Bianchi DW, Platt LD, Goldberg JD, Abuhamad AZ, Sehnert AJ, Rava RP. 2012. Genome-
212 wide fetal aneuploidy detection by maternal plasma DNA sequencing. *Obstet Gynecol*, 119:
213 890–901. DOI: 10.1097/AOG.0b013e31824fb482.

214 Bischoff FZ, Sinacori MK, Dang DD, Marquez-Do D, Horne C, Lewis DE, Simpson JL.
215 2002. Cell-free fetal DNA and intact fetal cells in maternal blood circulation: implications for
216 first and second trimester non-invasive prenatal diagnosis. *Hum Reprod Update* 8, 493–500.
217 DOI: 10.1093/humupd/8.6.493.

218 Chen EZ, Chiu RWK, Sun H, Akolekar R, Chan KCA, Leung TY, Jiang P, Zheng YWL, Lun
219 FMF, Chan LYS, Jin Y, Go ATJI, Lau ET, To WWK, Leung WC, Tang RYK, Au-Yeung SKC,
220 Lam H, Kung YY, Zhang X, van Vugt JMG, Minekawa R, Tang MHY, Wang J, Oudejans
221 CBM, Lau TK, Nicolaides KH, Lo YMD. Noninvasive prenatal diagnosis of fetal trisomy 18
222 and trisomy 13 by maternal plasma DNA sequencing. *PLoS One*, 6:e21791. DOI:
223 10.1371/journal.pone.0021791.

224 Chiu RW, Chan KCA, Gao Y, Lau VYM, Zheng W, Leung TY, Foo CHF, Xie B, Tsui NBY,
225 Lun FMF, Zee BCY, Lau TK, Cantor CR, Lo YMD. 2008. Noninvasive prenatal diagnosis of

226 fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal
227 plasma. *Proc Natl Acad Sci USA*, 105:20458-63. DOI: 10.1073/pnas.0810641105.

228 Dohm JC, Lottaz C, Borodina T, Himmelbauer H. 2008. Substantial biases in ultra-short read
229 data sets from high-throughput DNA sequencing. *Nucleic Acids Res*, 36: e105. DOI:
230 10.1093/nar/gkn425.

231 Fan HC, Blumenfeld YJ, Chitkara U, Hudgins L, Quake SR. 2008. Noninvasive diagnosis of
232 fetal aneuploidy by shotgun sequencing DNA from maternal blood. *Proc Natl Acad Sci USA*,
233 105:16266-71. DOI: 10.1073/pnas.0808319105.

234 Illanes S, Denbow M, Kailasam C, Finning K, Soothill PW. 2007. Early detection of cell-free
235 fetal DNA in maternal plasma. *Early Hum Dev*, 83:563-566. DOI:
236 10.1016/j.earlhumdev.2006.11.001.

237 Jiang F, Ren J, Chen F, Zhou Y, Xie J, Dan S, Su Y, Yin B, Su W, Zhang H, Wang W, Chai X,
238 Lin L, Guo H, Li Q, Li P, Yuan Y, Pan X, Li Y, Liu L, Chen H, Xuan Z, Chen S, Zhang C,
239 Zhang H, Tian Z, Zhang Z, Jiang H, Zhao L, Zheng W, Li S, Li Y, Wang J, Wang J, Zhang X.
240 2012. Noninvasive Fetal Trisomy (NIFTY) test: an advanced noninvasive prenatal diagnosis
241 methodology for fetal autosomal and sex chromosomal aneuploidies. *BMC Med Genomics*, 5:
242 57. DOI: 10.1186/1755-8794-5-57.

243 Kang X, Xia J, Wang Y, Xu H, Jiang H, Xie W, Chen F, Zeng P, Li X, Xie Y, Liu H, Huang G,
244 Chen D, Liu P, Jiang H, Zhang X. 2016. An Advanced Model to Precisely Estimate the Cell-
245 Free Fetal DNA Concentration in Maternal Plasma. *PLoS One*, 11(9):e0161928. DOI:
246 10.1371/journal.pone.0161928.

247 Kozarewa I, Ning Z, Quail MA, Sanders MJ, Berriman M, Turner DJ. 2009. Amplification-
248 free Illumina sequencing-library preparation facilitates improved mapping and assembly of
249 (G+C)-biased genomes. *Nat Methods*, 6: 291-295. 10.1038/nmeth.1311. DOI:
250 10.1038/nmeth.1311.

251 Lau TK, Chen F, Pan X, Pooh RK, Jiang F, Li Y, Jiang H, Li X, Chen S, Zhang X. 2011.
252 Noninvasive prenatal diagnosis of common fetal chromosomal aneuploidies by maternal
253 plasma DNA sequencing. *J Matern Fetal Neonatal Med*, 25(8):1370–1374. DOI:
254 10.3109/14767058.2011.635730.

255 Lehmann-Werman R, Neiman D, Zemmour H, Moss J, Magenheimer J, Vaknin-Dembinsky A,
256 Rubertsson S, Nellgård B, Blennow K, Zetterberg H, Spalding K, Haller MJ, Wasserfall CH,

257 Schatz DA, Greenbaum CJ, Dorrell C, Grompe M, Zick A, Hubert A, Maoz M, Fendrich V,
258 Bartsch DK, Golan T, Ben Sasson SA, Zamir G, Razin A, Cedar H, Shapiro AM, Glaser B,
259 Shemer R, Dor Y. 2016. Identification of tissue-specific cell death using methylation patterns
260 of circulating DNA. *Proc. Natl Acad. Sci. USA* 113(13), E1826–E1834. DOI:
261 10.1073/pnas.1519286113.

262 Leys C, Ley C, Klein O, Bernard P, Licata, L. 2013. Detecting outliers: Do not use standard
263 deviation around the mean, use absolute deviation around the median. *Journal of*
264 *Experimental Social Psychology*, 49(4), 764–766. DOI: 10.1016/j.jesp.2013.03.013.

265 Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, et al. 1997. Presence
266 of fetal DNA in maternal plasma and serum. *Lancet*, 350:485–7. DOI: 10.1016/S0140-
267 6736(97)02174-0.

268 Norton ME, Brar H, Weiss J, Karimi A, Laurent LC, Caughey AB, Rodriguez MH, Williams
269 J, Mitchell ME, Adair CD, Lee H, Jacobsson B, Tomlinson MW, Oepkes D, Hollemon D,
270 Sparks AB, Oliphant A, Song K. 2012. Non-Invasive Chromosomal Evaluation (NICE)
271 Study: results of a multicenter prospective cohort study for detection of fetal trisomy 21 and
272 trisomy 18. *Am J Obstet Gynecol*, 207(2):137 e131-138. DOI: 10.1016/j.ajog.2012.05.021

273 Qi G, Yi J, Han B, Liu H, Guo W, Shi C, Yin L. 2016. Noninvasive prenatal testing in routine
274 clinical practice for a high-risk population: Experience from a center. *Medicine (Baltimore)*,
275 95(41):e5126. DOI: 10.1097/MD.00000000000005126.

276 Song Y, Liu C, Qi H, Zhang Y, Bian X, Liu J. 2013. Noninvasive prenatal testing of fetal
277 aneuploidies by massively parallel sequencing in a prospective Chinese population. *Prenat*
278 *Diagn*, 33: 700–706. DOI: 10.1002/pd.4160.

279 Wang E, Batey A, Struble C, Musci T, Song K, Oliphant A. 2013. Gestational age and
280 maternal weight effects on fetal cell-free DNA in maternal plasma. *Prenat Diagn*, 33(7):662–
281 666. DOI: 10.1002/pd.4119.