

IDENTIFICAÇÃO DE REGIÕES HEMIMETILADAS E A PREDIÇÃO DE
REGIÕES DIFERENCIALMENTE METILADAS IMPRINTADAS NO
GENOMA HUMANO

GRAZIELA DE SÁ MACHADO ARAÚJO

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE - UENF

CAMPOS DOS GOYTACAZES – RJ

OUTUBRO DE 2017

IDENTIFICAÇÃO DE REGIÕES HEMIMETILADAS E A PREDIÇÃO DE
REGIÕES DIFERENCIALMENTE METILADAS IMPRINTADAS NO
GENOMA HUMANO

GRAZIELA DE SÁ MACHADO ARAÚJO

Tese apresentada ao Centro de Biociências
e Biotecnologia – CBB, da Universidade
Estadual do Norte Fluminense, como parte
das exigências para obtenção do título de
Doutor em Biociências e Biotecnologia.

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE - UENF

CAMPOS DOS GOYTACAZES – RJ

OUTUBRO DE 2017

IDENTIFICAÇÃO DE REGIÕES HEMIMETILADAS E A PREDIÇÃO DE
REGIÕES DIFERENCIALMENTE METILADAS IMPRINTADAS NO
GENOMA HUMANO

GRAZIELA DE SÁ MACHADO ARAÚJO

Tese apresentada ao Centro de Biociências e Biotecnologia – CBB, da Universidade Estadual do Norte Fluminense, como parte das exigências para obtenção do título de Doutor em Biociências e Biotecnologia.

Aprovada em ____ de outubro de 2017.

Comissão Examinadora:

Dra. Regina Célia de Souza Campos Fernandes (Doutora em Doenças Infecciosas e Parasitárias) - FMC

Dr. Victor Martin Quintana Flores (Doutor em Biociências e Biotecnologia) - UENF

Dr. Fabrício Moreira Almeida (Doutorado em Biociências e Biotecnologia) - UFRJ

Dr. Filipe Brum Machado (Doutor em Genética) – UENF
(Coorientador)

Dr. Enrique Medina –Acosta (Doutor em parasitologia Médica e Molecular) –UENF.
(Orientador)

DEDICATÓRIA

. . . A Deus, que a todo momento estava presente. . .
. . . A toda minha família que me incentivou para que eu pudesse vencer mais um desafio em minha vida. .
.

À Deus, que me fez em muitos momentos conflitantes ter força para enfrentar.

Aos meus pais, Josely de Sá Machado Araújo e Gustavo Juarez Araújo, que com um amor único me incentivaram e me aconselharam nos momentos difíceis para que eu pudesse superar e alcançar minhas metas.

Aos meus irmãos, Marcelle Sá Machado de Araújo, Gustavo Juarez Araújo Filho e Gustavo Araújo Neto que sempre estiveram na torcida para o meu sucesso.

Aos meus avós Antônio Machado Moço e Marly de Sá Machado por estarem sempre presentes na minha vida.

Ao meu Tio Marquinho e minha madrinha Gizelma minha gratidão por toda atenção e carinho.

AGRADECIMENTOS

Ao **Prof. Dr. Enrique Medina-Acosta**, que ao longo destes anos de convivência, foi mais do que um orientador, foi um grande amigo. Agradeço por todo tempo dispensado ao meu crescimento científico e formação profissional.

Ao **Dr. Filipe Brum Machado**, por toda amizade, coorientação, que foram importantes para a realização deste trabalho.

A **Dra. Thaís Louvain de Souza**, por toda amizade, por todo incentivo, e pela colaboração, que foram fundamentais para à revisão do manuscrito e finalização deste trabalho.

A **Dra. Regina Célia de Souza Campos Fernandes, Dr. Victor Martin Quintana Flores e Dr. Fabrício Moreira Almeida**, por aceitarem prontamente o convite para a participação como membros titulares para a banca de defesa desta tese, e por todas as sugestões e correções que foram essenciais para a finalização deste trabalho.

Aos alunos: **Douglas Terra Machado, Ronaldo da Silva Francisco Junior, Pedro Thyago Mozer Rodrigues, Cristina dos Santos Ferreira, Dra. Cláudia Caixeta Franco Andrade** pela colaboração efetiva nas análises dos experimentos *in silico*, fundamentais para a finalização deste trabalho. Em especial a **Douglas Terra Machado** que em nenhum momento se absteve em me ajudar, não poupando esforços para as edições das figuras sendo um grande incentivador para a finalização deste trabalho.

Aos amigos: **Douglas Terra, Thaís Louvain de Souza, Filipe Brum Machado, Ronaldo da Silva Francisco Junior, Antonio Francisco Alves da Silva, Cláudia Caixeta Franco Andrade, Cristina dos Santos Ferreira, Thiago Barbosa de Souza, Diego Sá, Luísa Mignone Paixão, Juliana Ywasaki, Katrine da Silva Leonardo, Jozimara Teixeira, Cleiton Figueiredo Ozório da Silva, Alan Tardin,**

Pedro Thyago Mozer Rodrigues, Maria Emilce da Rosa Francelino pelo convívio, companheirismo, e aprendizado mútuo ao longo desses anos.

A todo o **corpo docente** da **Universidade Estadual do Norte Fluminense Darcy Ribeiro - UENF**, que direta ou indiretamente participaram efetivamente da minha formação acadêmica.

À **Universidade Estadual do Norte Fluminense Darcy Ribeiro – UENF**, em particular ao **Programa de pós-graduação em Biociências e Biotecnologia**, pela oportunidade concedida.

SUMÁRIO

1. INTRODUÇÃO	15
2. REVISÃO BIBLIOGRÁFICA.....	17
2.1 Imprinting Genômico.....	17
2.2 Regiões Diferencialmente Metiladas (DMRs)	24
2.3 Metilação	27
2.4 Reprogramação do <i>Imprinting</i> nas Gerações	28
2.5 Dissomia Uniparental (UPD).....	29
2.6 Expressão Monoalélica Estocástica	31
2.7 Modificações Pós Traducionais em Proteínas Histonas	32
2.8 RNAs Não Codificantes	33
3. JUSTIFICATIVA.....	34
4. OBJETIVOS.....	35
4.1. Objetivo Geral.....	35
4.2. Objetivos Específicos.....	35
5. MATERIAIS E MÉTODOS.....	36
5.1. Material Biológico e Extração de DNA	36
5.2. Aspectos Éticos da Pesquisa.....	36
5.3. Varredura e desenho de marcadores	36
5.4. Critérios de Seleção para <i>loci</i> Potencialmente Polimórficos	38
5.5. Iniciadores Marcador-Específicos.....	39
5.6. Tipagem dos Marcadores	39
5.6.1 DESM-QFPCR para o fragmento PWWP2AP1-DMR	42
5.6.2 DESM-QFPCR para o fragmento 263 PARD6G-AS1/ADNP2-DMR.....	42
5.6.3 DESM-QFPCR para o fragmento GCSAML-DMR	43
5.6.4 DESM-QFPCR para o fragmento PMF1-DMR.....	43
5.7. Caracterização dos Alelos Amplificados e Sequenciamento	44
5.8. Demonstração do Perfil do Eletroferograma para Caracterização de Predição de <i>Imprinting</i> Genômico.....	44
5.9. Demonstração do Perfil Eletroferograma para Determinação da Origem Parental do <i>Imprinting</i>	45
5.10. RNA-Seq.....	46
6. RESULTADOS	49

6. 1 Seleção de Regiões Diferencialmente Metiladas (DMR)	49
6.2 Região <i>PARD6G-AS1</i>	55
6.3 Região <i>GCSAML</i>	62
6. 4 Região <i>PWWP2AP1</i>	68
6.5 Região <i>PMF1</i>	74
7. Discussão	80
8. CONCLUSÃO	85
8. REFERÊNCIAS	88

SUMÁRIO DE ILUSTRAÇÕES

Figura 1. Assimetria gamética da metilação na região promotora do gene imprintado <i>PEG10</i>	38
Figura 2. O esquema representativo in silico de um SNP do gene do cromossomo 7. Para demonstração com iniciador sentido e anti sentido em verde.	39
Figura 3. Eletroferograma representativo da identificação de locus com perfil de metilação intermediária alélica específica.	45
Figura 4. Eletroferogramas representativos da genotipagem do SNP rs2244352 no gene <i>WRB</i> a fim de se determinar a origem parental da metilação.	46
Figura 5. Assimetria gamética da metilação na região promotora do gene <i>PARD6G-AS1</i>	56
Figura 6. Perfil de expressão do gene <i>PARD6G-AS1</i> em 53 tecidos disponível no GTEX.	57
Figura 7. Caracterização da DMR da região <i>PARD6G-AS1</i> com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho).....	58
Figura 8. Eletroferograma representativo da identificação do locus <i>PARD6G-AS1</i> com perfil de metilação alélica específica.	59
Figura 9. A assimetria gamética na gDMR do gene <i>PARD6G-AS1</i> é de origem materna.	60
Figura 10. Variação na metilação e entre triplicatas em amostras de indivíduos controles utilizando DESM-QF-PCR.	61
Figura 11. Assimetria gamética da metilação na região promotora do gene <i>GCSAML</i>	63
Figura 12. Perfil de expressão do gene <i>GCSAML</i> em 53 tecidos disponível no GTEX.	63
Figura 13. Caracterização da DMR da região <i>GCSAML</i> com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho).....	64
Figura 14. Eletroferograma representativo da identificação do locus <i>GCSAML</i> com perfil de hemimetilação alélica.	65
Figura 15. A assimetria gamética na gDMR do gene <i>GCSAML</i> é de origem materna..	66

Figura 16. Variação na metilação e entre triplicatas em amostras de indivíduos controles utilizando DESM-QF-PCR.	67
Figura 17. Assimetria gamética da metilação na região promotora do gene <i>PWWP2AP1</i>	69
Figura 18. Perfil de expressão do gene <i>PWWP2AP1</i> em 53 tecidos disponível no GTEX.	69
Figura 19. Caracterização da DMR da região <i>PWWP2AP1</i> com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho).....	70
Figura 20. Eletroferograma representativo da identificação do <i>locus PWWP2AP1</i> com perfil de hemimetilação alélica.....	71
Figura 21. Assimetria gamética enviesada na gDMR do gene <i>PWWP2AP1</i>	72
Figura 22. Variação na metilação e entre triplicatas em amostras de indivíduos controles utilizando DESM-QF-PCR.	73
Figura 23. Assimetria gamética da metilação na região promotora do gene <i>PMF1</i> . De cima para baixo, ideograma representado o cromossomo 1, neste a região marcada em vermelho (linha vertical) em que abaixo confere essa região.....	75
Figura 24. Perfil de expressão do gene <i>PMF1</i> em 53 tecidos disponível no GTEX..	76
Figura 25. Caracterização da DMR da região <i>PMF1</i> com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho).....	77
Figura 26. Eletroferograma representativo da identificação do <i>locus PMF1</i> com perfil de hipermetilação alélica.....	78
Figura 27. Simetria gamética na gDMR do gene <i>PMF1</i>	79

TABELAS

Tabela 1. Conteúdo de dados e estatísticas do MetalImprint.....	19
Tabela 2. Genes imprintados detectados neste estudo, com o número de tecidos em cada categoria de expressão de imprintado ou bialélico.....	21
Tabela 3. Visão geral molecular de doenças associadas ao <i>Imprinting</i>	22
Tabela 4. Visão geral clínica das doenças associadas ao <i>Imprinting</i>	23
Tabela 5. Regiões Diferentemente Metiladas.....	26
Tabela 6. Análise integrativa das candidatas à DMRs alvos e desenho dos primers.	41
Tabela 7. Localizações cromossômicas de Regiões Diferentemente Metiladas.	50
Tabela 8. Regiões hemimetiladas escolhidas a partir dos critérios maiores e menores.	54

RESUMO

Em mamíferos placentários ocorre a diferenciação epigenética dos cromossomos parentais durante o desenvolvimento que resulta na expressão monoalélica origem parental-dependente de um discreto número de genes, e esse processo é conhecido como *imprinting* genômico. A metilação do DNA ocorre no contexto do dinucleotídeo CpG. Uma característica do *imprinting* genômico é o estado intermediário de metilação do DNA que pode tanto ser constitutivo em todos tecidos quanto ser tecido-específico, constituindo as chamadas DMRs (do inglês *Differentially Methylated Regions*). A ocorrência das DMRs é geralmente em ilhas CpG ou nas regiões promotoras. Parece haver uma relação entre a ocorrência de DMRs e a expressão monoalélica consistente com *imprinting*, uma vez que os genes imprintados não contêm uma DMR ao longo do seu corpo gênico, e normalmente são regulados à distância em *cis* por outras DMRs. Constatar o *imprinting* de genes candidatos é um desafio porque a expressão pode ser isoforma-específica e/ou tecido-específico, no tanto que no genoma humano foram identificados menos de 125 genes imprintados até o presente estudo. Este trabalho visou à identificação de regiões diferencialmente metiladas e a predição de genes imprintados no genoma humano. A estratégia consistiu na visualização manual dos perfis comparativos de metilação de DNA em dezenas de experimentos de metiloma em repositórios públicos para diversos tecidos e gametas em uma janela integrativa em 40kb em sessões customizadas na plataforma *UCSC Genome Browser*, da universidade da Califórnia em Santa Cruz. Foram identificadas 25 novas regiões diferencialmente metiladas de origem gamética (gDMR) e 61 regiões diferentemente metiladas secundárias (sDMR). Para análise subsequente foram selecionados *loci* segundo os seguintes critérios: (i) ocorrência de pelo menos um polimorfismo de um único nucleotídeo (SNP) com frequência alélica $>0,15$; (ii) sobreposição com ilha CpG; e (iii) não sobreposição com duplicações segmentárias ou repetições simples. Foram desenhados ensaios da PCR para validação experimental dos perfis de metilação e expressão gênica para 4 (quatro) regiões com metilação intermediária nos genes *PWWP2AP1*, *PARD6GAS1*, *GCSAML* e *PMF1*, candidatos ao *imprinting*. Para verificar a predição da origem gamética materna da metilação, desenvolvemos ensaios com base na análise de digestão enzimática sensível a metilação seguida de PCR (DESM-PCR) que permitem interrogar os estados de metilação em sítios

CpG localizados em sequências de reconhecimento da enzima de restrição *HpaII*, que é sensível à metilação. Os SNPs para DNA genômico de células sanguíneas em trios familiares nucleares foram informativos nas análises dos candidatos a iDMRs gaméticas nos *loci PARD6G-AS1* e *PWWP2AP1* e candidato a iDMR secundária no gene *GCSAML*. Para ambas as iDMRs *PARD6G-AS1* e *GCSAML*, os alelos maternos foram consistentemente metilados enquanto os alelos paternos estavam desmetilados. Em contraste, para o candidato a iDMR no pseudogene *PWWP2AP1*, o estado de metilação estava presente nos dois alelos parentais. Já para o *PMF1*, tivemos os dados consistentes com o estado hipermetilado onde os iniciadores foram desenhados. Dado que o estado intermediário de metilação do DNA nas iDMRs não é um critério suficiente para prever com precisão o estado ativo ou silencioso dos *loci*, avaliamos os perfis de expressão transcricional alélica por meio do mapeamento de SNPs heterozigotos mapeados nos genes alvo utilizando experimentos primários e secundários de RNA-Seq do NIH Recurso de banco de dados público Genotype-Tissue Expression (GTEx) e SRA (do inglês *Sequence Read Archive*). Mostramos que os três *loci* são expressos bialelicamente em vários tecidos de vários doadores. Concluímos que as regiões diferencialmente metiladas que estão configuradas para *imprinting* materno no *PARD6G-AS1* e no *GCSAML* estão desacopladas aos efeitos de expressão origem parental-dependentes em múltiplos tecidos humanos.

Palavras-chave: DMR, *imprinting*, *PWWP2AP1*, *PARD6GAS1*, *GCSAML*, *PMF1*.

ABSTRACT

In placental mammals occurs the epigenetic differentiation of the parental chromosomes during the development that results in monoallelic parental-of-origin expression of a discrete number of genes and this process is known as genomic imprinting. The DNA methylation occurs in the context of the dinucleotide CpG. A characteristic of the genomic imprinting is the intermediary methylation state on the DNA that can not only be the constitutive mark in all tissues but can also be tissue-specific, which constitute the called DMRs (Differentially Methylated Regions). The occurrence of DMRs is generally at the CpG islands or at promoter regions. It seems to have a relationship between the occurrence of DMRs and the monoallelic expression consistent with imprinting, once that the imprinted genes don't have a DMR on your gene body, and generally are regulated in cis by others DMRs. Considering the imprinting status of candidate genes is a challenge because the expression can be isoform-specific and/or tissue-specific, that's why the human genome has around 125 genes imprinted until this present study. This work aimed to identify differentially methylated regions and the prediction of imprinted genes on the human genome. The strategy focused for the manual visualization of comparative methylation DNA profiles in lots of methylome experiments from public repositories in diverse tissues and gametes on an integrative window of 40 kb for customized sessions at the UCSC *Genome Browser*, from University of California in Santa Cruz. We identify 25 new gametic differentially methylated regions (gDMRs) and 61 secondary differentially methylated regions (sDMRs). For the subsequent analysis we selected loci based on the following criteria: (i) occurrence of at least one single nucleotide polymorphism (SNP) with allelic frequency >0,15; (ii) overlap with CpG island; and (iii) no overlap with segmental duplications or simple repeats. Primers were designed for PCR with the objective of experimental validation of the methylation profile and genic expression in 4 regions with intermediary methylation of the *PWWP2AP1*, *PARD6G-AS1*, *GCSAML* and *PMF1*, candidate for the imprinting. To verify the prediction of maternal gametic origin of methylation, we developed assays focusing on the analysis of the enzymatic susceptible digestion to methylation at CpG sites mapped in sequences in which there's sites for the *HpaII*, the enzyme sensible for methylation. The SNPs for the genomic DNA on the blood cells in family trios were informative for the analysis of the gametic iDMRs candidates on the *loci*

PARD6G-AS1 and *PWWP2AP1* and the secondary iDMR candidate for the *GCSAML* gene. For both iDMRs of *PARD6G-AS1* and *GCSAML*, the maternal alleles were consistently methylated while the paternal were unmethylated. In contrast, for the iDMR candidate on the pseudogene *PWWP2AP1*, the methylation was present on both parental alleles. However, for the *PMF1*, we had the hypermethylated state in which the primers were designed. Once that the intermediary methylation profile of the iDMRs isn't a sufficient criterion to describe with precision the active and inactive states of the loci, we evaluated the transcriptional allelic expression profile through heterozygote SNPs mapped on the target genes utilizing primary and secondary RNA-Seq experiments from NIH GTEx (Genotype-Tissue Expression) and SRA (Sequence Read Archive). We show that the three loci are biallelic expressed in various tissues from various donors. We conclude that the differentially methylated regions that are configured for the maternal imprinting on the *PARD6G-AS1* and *GCSAML* are decoupled from parent-of-origin expression effects in multiple human tissues.

Keywords: DMR, *imprinting*, *PWWP2AP1*, *PARD6GAS1*, *GCSAML*, *PMF1*.

1. INTRODUÇÃO

Epigenética é um ramo da Genética que estuda as alterações químicas hereditárias que podem implicar na expressão gênica, mas que não alteram a seqüência do DNA. Durante a última década, os pesquisadores mostraram que a regulação epigenética desempenha um papel importante no crescimento celular, diferenciação, doenças autoimunes e câncer. Os principais mecanismos epigenéticos incluem o fenômeno bem entendido de metilação do DNA, modificações pós traducionais em caudas de proteínas histonas e regulação por RNAs não codificantes, um modo de regulação que só foi identificado relativamente recentemente e é uma área de investigação intensiva em curso, porém que é necessária ao processo de inativação do cromossomo X, por exemplo. (Wei, Huang *et al.*, 2017).

Em organismos eucarióticos diplóides, geralmente são derivadas cópias materna e paterna de cada gene e estas são simultâneamente expressas em níveis comparáveis. No entanto, há exceções onde apenas um dos alelos é expresso. Os genes expressos de forma monoalélica podem ser divididos em três classes distintas. (Gimelbrant, Hutchinson *et al.*, 2007).

Uma classe é de genes autossômicos imprintados (tais como *IGF2* e *H19*) cuja expressão monoalélica é regulada de uma forma origem parental-dependente (Reik e Walter, 2001; Gimelbrant, Hutchinson *et al.*, 2007).

Uma segunda classe são genes do cromossomo X inativados, regulados por um processo aleatório em que ao longo do desenvolvimento embrionário, mais especificamente entre o período de 8 e 16 células pós-fecundação, ocorre o processo de inativação do cromossomo X em que metade das células inativa o cromossomo X materno e a outra metade inativa o cromossomo X paterno (Lyon, 1986; Gimelbrant, Hutchinson *et al.*, 2007), em escala celular, vê-se expressão monoalélica dos genes que são inativados, porém em escala tecidual, vê-se expressão bialélica.

Uma terceira classe são genes autossômicos sujeitos a expressão monoalélica estocástica (Pernis, Chiappino *et al.*, 1965; Gimelbrant, Ensminger *et*

al. , 2005; Gimelbrant, Hutchinson *et al.* , 2007). Para estes genes algumas células expressam o alelo materno enquanto outras células expressam o alelo paterno.

Muitos *loci* imprintados foram identificados com distúrbios do desenvolvimento causados pela sua ruptura, particularmente a descoberta de dissomia e outros erros genéticos em distúrbios humanos raros (Yamazawa, Ogata *et al.*, 2010).

O *imprinting* genômico, inativação do cromossomo X e dissomia uniparental não seguem os padrões das heranças das leis de Mendel devido à expressão específica de alelos (ASE). Especificamente, o *imprinting* genômico induz o silenciamento de um dos alelos quando é transmitido por um dos pais, resultando em origem parental-dependente (Surani, Barton *et al.*, 1984; Deng, Berletch *et al.*, 2014). Alguns genes imprintados desempenham um papel decisivo no controle no processo de crescimento e desenvolvimento (Plasschaert e Bartolomei, 2014). A interrupção desses genes causa síndromes genéticas graves, incluindo a síndrome de Prader-Willi (PWS), síndrome de Angelman (AS)(Donlon, 1988), síndrome de Beckwith-Wiedemann (BWS)(Weksberg, Teshima *et al.*, 1993) síndrome de Silver-Russell (SRS)(Abu-Amero, Monk *et al.*, 2008), síndrome de Temple (TS) e a síndrome de Kagami-Ogata (KOS)(Ogata, Kagami *et al.*, 2008) entre outras.

No início da década de 1990, *IGF2* (expresso paternalmente) e *H19* (expresso maternalmente) foram os dois primeiros genes imprintados descobertos em ratos. (Barlow, Stoger *et al.*, 1991; Bartolomei, Zemel *et al.*, 1991). Até a presente data, aproximadamente 150 genes imprintados em humanos foram descobertos (Babak, Deveale *et al.*, 2015; Baran, Subramaniam *et al.*, 2015).

Porém ainda não existe uma ferramenta ou estratégia robusta para identificação de DMRs imprintadas ligadas a genes preditos ao *imprinting* genômico. Alterações epigenéticas ou deleções nesses genes preditos ao *imprinting* genômico estão associados com doenças genéticas e câncer. A busca de novas DMRs imprintadas e genes imprintados e sua confirmação auxiliam na compreensão de doenças epigenéticas bem como nas alterações da célula tumoral podendo servir de subsídios para a geração de produtos tanto para diagnóstico quanto para tratamento.

2. REVISÃO BIBLIOGRÁFICA

2.1 Imprinting Genômico

Em mamíferos placentários ocorre a diferenciação epigenética dos cromossomos parentais durante o desenvolvimento que resulta na expressão monoalélica origem parental-dependente de um discreto número de genes. Essa diferenciação epigenética é conhecida como *imprinting* genômico. (Reik e Walter, 2001; Joshi, Garg *et al.* , 2016). Uma das marcas epigenéticas características nesse processo é a metilação intermediária do DNA que ocorre no contexto do dinucleotídeo CpG, estabelecendo, dessa maneira, regiões diferencialmente metiladas (DMRs, do inglês *Differentially Methylated Region*) (Reik e Walter, 2001). Há uma relação entre a ocorrência de DMRs e a expressão monoalélica imprintada, pois os genes sob esse processo de *imprinting* genômico que não contêm uma DMR são regulados à distância por DMRs em *cis*, ou seja, um ponto de regulação à distância que pode ocorrer na mesma fita do DNA.

No entanto, constatar a presença de um gene imprintado é um desafio uma vez que a expressão monoalélica de um gene imprintado pode ocorrer apenas em uma das várias isoformas, possivelmente, apenas em determinados tecidos, ou apenas em fases específicas do desenvolvimento. Conseqüentemente, o fato de não poder confirmar o *imprinting* em um tecido ou um estágio específico do desenvolvimento não elimina a possibilidade de o gene ser imprintado em qualquer outro tecido ou em alguma outra etapa do desenvolvimento (Morison, Ramsay *et al.*, 2005).

Esse tipo de expressão em mamíferos fornece o exemplo clássico de memória epigenética origem parental-dependente ao longo da vida . Cerca de cento e cinquenta (150) genes imprintados foram identificados em mamíferos (Williamson Cm, Blake A *et al.*, 2013) embora alguns relatos sugerem que mais de 100 genes mostram evidência de *imprinting* em seres humanos. Recentemente pesquisas de metilação de DNA em todo o genoma e expressão de genes conseguiram confirmar apenas cerca de metade deste número (Joshi, Garg *et al.* , 2016).

O *imprinting* genômico depende de um mecanismo comum que os marca diferencialmente na linha germinativa masculina e feminina (Kelsey e Feil, 2013). Muitos genes imprintados são encontrados em grupos e a expressão imprintada correta de todos os genes agrupados depende da existência da marcação apropriada de um elemento genético, a região controladora de *imprinting* (ICRs, do inglês *Imprinting Control Regions*) (Wan e Bartolomei, 2008; Ferguson-Smith, A. C. , 2011). As propriedades da ICR são importantes para estudos epigenéticos em destaque no espermatozóide e oócito e para verificar como a metilação é mantida após a fertilização, por exemplo. (Proudhon, Duffie *et al.* , 2012).

Metilação do DNA nas regiões de controle de *imprinting* é uma modificação epigenética amplamente estudada. Esse mecanismo requer a reposição da metilação do DNA na linhagem germinativa e sua subsequente resistência ao apagamento durante a transição da célula germinativa para desenvolvimento embrionário (Messerschmidt, De Vries *et al.* , 2012; Seisenberger, Peat, Hore *et al.* , 2013). Enquanto a metilação das ICRs é permanente, os efeitos sobre metilação do DNA e expressão de genes são dependentes de outros fatores, tal como fase de desenvolvimento (Smallwood, Tomizawa *et al.* , 2011) em que podem ocorrer outras marcas epigenéticas.

Os genes imprintados são essenciais para o desenvolvimento embrionário. Surani e colaboradores em 1984 citam que uma célula germinativa que tenha somente material genético da mãe (ginogenético) terá o desenvolvimento do embrião, porém terá um comprometimento do desenvolvimento da placenta. Enquanto que Reik e colaboradores em 1987 descreveram que se houver somente material genético do pai (androgenético), a placenta se desenvolverá, mas o embrião terá comprometimento no seu desenvolvimento. Isso condiz que há expressão monoalélica origem parental-dependente que ocasiona o desenvolvimento completo do embrião e da placenta, onde ambos os genomas parentais se completam para que isso ocorra (Surani, Barton *et al.* , 1984; Reik, Collick *et al.* , 1987).

A alteração na expressão do gene contribui para várias doenças humanas (Luedi, Dietrich *et al.* , 2007). A perda do *imprinting* genômico pode levar a uma variedade de distúrbios, por exemplo, Prader-Willi e Algemean causadas por deleção na região 15q11-13 ou dissomia do cromossomo 15. Se houver deleção dessa

região na cópia herdada do pai, haverá fenótipo clínico com característica de Prader-Willi, enquanto que se houver deleção na cópia herdada da mãe, haverá fenótipo clínico característico de Angelman. O mesmo equivale a dissomia uniparental com duas cópias do cromossomo 15 herdadas do pai, caracterizando Angelman, enquanto que dissomia materna para o mesmo cromossomo caracteriza Prader-Willi. Apesar de um grande esforço de investigação estar sendo realizada para examinar vários mecanismos de *imprinting*, pouco se sabe sobre a criação e regulamentação dos genes imprintados. Existem vários mecanismos epigenéticos que têm relevância nesse processo e que podem ter impacto sobre o desenvolvimento embrionário e crescimento (Biliya e Bulla, 2010).

Os genes imprintados estão envolvidos na etiologia de síndromes raras e têm sido associados a doenças comuns como diabetes e câncer (Santoni, Stamoulis *et al.*, 2017). Portanto há um rastreamento para identificar regiões hemimetiladas no genoma candidatas à iDMR (DMR imprintada) e genes imprintados.

É importante ressaltar que genes imprintados desempenham importantes papéis no crescimento e desenvolvimento (Reik e Walter, 2001). Dados estatísticos mostram na Tabela 1 os genes imprintados, validados, preditos, que perdem *imprinting* e que não são imprintados (Wei, Su *et al.*, 2014). Esses dados disponíveis na literatura demonstram que apesar de muitos genes preditos ao *imprinting*, apenas 32,5% são validados como imprintados. A dificuldade de encontrar genes sob esse processo consiste em muitas vezes ser em estágio específico do desenvolvimento, ser isoforma específico ou tecido específico, como mencionado anteriormente.

Tabela 1. Conteúdo de dados e estatísticas do MetalImprint.

Conteúdo dos dados	Humano
Genes Imprintados	
Total	317
Experimentos validados	65
Preditos	200
Perda de <i>Imprinting</i>	21
Não Imprintado	31

Fonte: Modificado de (Wei, Su *et al.*, 2014).

Alguns dos genes validados, os referidos na tabela 2, têm sido associados com doenças genéticas como, por exemplo, o gene *PEG10* que tem grande impacto em estudos de carcinomas (Bang, Ha *et al.* , 2014; Gao, Zhang *et al.* , 2014). Em tecidos adultos com caracterização sistemática, há variação de *imprinting* entre genes, indivíduos e tecidos. É importante ressaltar que dos 42 genes imprintados citados por Baran, inclui genes novos e previamente identificados. No gene *PEG10*, de 30 tecidos testados apenas 13 tecidos foram imprintados segundo seus estudos. Isso demonstra que apesar do gene ser imprintado, ele pode ser tecido específico (Baran, Subramaniam *et al.*, 2015).

Por meio da determinação do estado de metilação e da identificação de polimorfismos de um único nucleotídeo (SNPs, do inglês *Single Nucleotide Polymorphism*) em trios, ou seja, filho (a), pai e mãe, informativos é possível estabelecer os perfis (monoalélico ou bialélico) de expressão transcricional e o padrão de herança epigenética (Morcos, Ge *et al.* , 2011).

Em termos gerais, a expressão monoalélica dependente da origem parental, isto é sob *imprinting*, está correlacionada com o estado hemimetilado de metilação nas regiões promotoras ou controladoras de *imprinting* (ICRs)(Hackett e Surani, 2013).

Tabela 2. Genes imprintados detectados neste estudo, com o número de tecidos em cada categoria de expressão de imprintado ou bialélico.

	Imprinted	Consistent with imprinted	Biallelic	Consistent with biallelic	Ambiguous	Total tissues with data
CPA4	2	1	8	4	2	17
CST1	1	2	0	3	0	6
DIRAS3	2	15	0	7	3	27
DLK1	7	3	0	1	1	12
FAM50B	11	25	0	1	0	37
GRB10	1	0	28	3	1	33
H19	30	4	0	1	0	35
IGF2-AS	1	7	0	4	0	12
IGF2	21	4	3	2	3	33
INPP5F_V2	3	0	23	10	0	36
KCNQ1	10	4	13	3	6	36
KIF25	1	0	0	7	0	8
L3MBTL1	20	6	0	1	4	31
LPAR6	3	0	6	25	0	34
MAGEL2	1	5	0	1	0	7
MAGI2	2	1	20	9	0	32
MEG3	30	3	0	0	0	33
MEG8	4	7	0	0	0	11
MEG9	5	20	0	0	2	27
MEST	7	12	2	8	4	33
NAP1L5	19	14	0	0	0	33
NDN	4	25	0	0	0	29
NTM	1	0	11	11	0	23
PEG10	13	13	0	1	3	30
PEG3	24	7	0	1	0	32
PLAGL1	28	3	3	0	0	34
PPIEL	4	20	1	3	7	35
PRSS50	1	3	2	16	0	22
PWRN1	3	6	0	1	0	10
RP11-7F17.7	1	4	2	2	2	11
SGK2	1	8	7	4	3	23
SNHG14	29	7	0	1	0	37
SNRPN	31	5	0	0	1	37
SNURF	25	8	0	1	0	34
SYCE1	1	9	0	10	1	21
THEGL	1	3	1	0	0	5
UBE3A	1	1	11	18	0	31
UGT2B4	1	1	0	1	1	4
UTS2	2	2	0	11	1	16
ZDBF2	21	11	0	2	0	34
ZNF331	13	1	7	13	2	36
ZNF597	4	23	1	1	0	29

Até agora, quatro diferentes tipos de mudanças moleculares foram associados a transtornos de *imprinting*: variações numéricas das cópias (CNVs: deleções, duplicações) da região imprintada, dissomia uniparental (UPD), marcas de metilação aberrante ("epimutações"), e mutações pontuais em genes (imprintados) (Soellner, Begemann *et al.*, 2017). Apesar de já terem sido descobertos uma

quantidade considerável de genes imprintados, como os descritos na tabela 2 de Baran (42 genes), apenas doze doenças (tabela 3 e 4) foram descritas associadas ao *imprinting*. (Soellner, Begemann *et al.*, 2017). Na Tabela 3 estão descritos os fatores moleculares de doenças associadas ao Imprinting. Na Tabela 4 estão descritos os fatores clínicos associados às doenças associadas ao *Imprinting*.

Tabela 3. Visão geral molecular de doenças associadas ao *Imprinting*.

Doença	Cromossomo/Região	Alterações cromossômicas
Diabetes Mellitus Transitório Neonatal	6q24	UPD(6)pat dup(6q) PLAGL1:alt- TSS-DMR: LOM
	7	upd(7)mat CNVs (dup7p), del7q
Síndrome Silver–Russell	11p15. 5	upd(11)mat dup(11p15)mat <i>H19/IGF2:IG-DMR</i>
Retardo mental Birk–Barel	8q24. 3	Mutações no gene <i>KCNK9</i>
Síndrome Beckwith–Wiedemann	11p15. 5	upd(11)pat
Síndrome Temple	14q32	upd(14)mat
Síndrome Kagami–Ogata	14q32	upd(14)pat del(14q32)mat
Síndrome de Angelman	15q11q13	upd(15)pat del(15q11q13)mat Mutações no gene <i>UBE3A</i>
Síndrome de Prader–Willi	15q11q13	upd(15)mat
Puberdade precoce	15q11. 2	Mutações no gene <i>MKRN3</i>
Síndrome Schaaf–Yang		Mutações no gene <i>MAGEL2</i>
Pseudohipoparatiroidismo esporádico	20q13	upd(20)pat
upd(20)mat	20	upd(20)mat

Tabela adaptada. (Soellner, Begemann *et al.*, 2017).

Tabela 4. Visão geral clínica das doenças associadas ao *Imprinting*.

Doenças	Características Clínicas
Diabetes Mellitus Transitório Neonatal	Diabetes transitória, hiperglicemia sem cetoacidose.
Síndrome de Silver–Russell	Face frontal triangular, dificuldades de alimentação, assimetria esquelética.
Retardo mental Birk–Barel	Deficiência intelectual, hiperatividade, dificuldades para se alimentar, hipotonia, rosto alongado.
Síndrome de Beckwith–Wiedemann	Crescimento pré e pós-natal excessivo, organomegalia, hipoglicemia neonatal, hemihipertrofia, aumento do risco de tumor.
Síndrome de Temple	Hipotonia, dificuldade de alimentação na infância, obesidade, puberdade precoce.
Síndrome de Kagami–Ogata	Defeitos na parede abdominal e torácica.
Síndrome de Angelman	Retardo mental, microcefalia.
Síndrome de Prader–Willi	Retardo de crescimento pós-natal Atraso mental neonatal Hipotonia, hipogenitalismo hipopigmentação, obesidade.
Puberdade precoce	Puberdade precoce
Síndrome de Schaaf–Yang	Hipotonia neonatal, Problemas de alimentação na infância, atraso no desenvolvimento.
Pseudoparatiroidismo esporádico	Deficiência para paratormônio e outros hormônios, Osteodistrofia, Anomalias de comportamento, crescimento anormal
upd(20)mat	Retardo de crescimento intra-uterino, retardo de crescimento pós-natal

Tabela adaptada. (Soellner, Begemann *et al.*, 2017).

2.2 Regiões Diferencialmente Metiladas (DMRs)

A principal marca epigenética do silenciamento é a modificação do DNA por metilação no contexto do dinucleotídeo CpG em domínios de *imprinting* conhecidos como regiões diferencialmente metiladas (DMR) (Hannula-Jouppi, Muurinen *et al.* , 2014).

Os *loci* imprintados geralmente contêm uma DMR caracterizada por marcas de metilação do DNA herdada do gameta masculino ou feminino (Schulz, Mccole *et al.* , 2009). A metilação do DNA é importante componente no controle da expressão e ocorre em citosinas 5' em dinucleotídeos CpG. A metilação das ilhas CpG é encontrada em genes em que a expressão está permanentemente reprimida, como por genes imprintados, genes no cromossomo X inativo e genes expressos exclusivamente em células germinativas. Portanto quando as ilhas CpG não estão metiladas os genes estão ativos com sítios de início de transcrição (Li, Beard *et al.* , 1993; Bourc'his, Xu *et al.* , 2001; Kaneda, Okano *et al.* , 2004; Arnaud, 2010; Jones, 2012).

Geralmente dois ou mais genes imprintados residem em grupos que têm um centro de controle de *imprinting* (ICC) como DMRs em locais intragênicos ou intergênicos que controlam a expressão dos genes (Wood e Oakey, 2006).

Regiões em gametas têm perfis de metilação opostos que são amplamente resolvidos durante a reprogramação epigenética pós-fertilização, como já mencionado anteriormente.

Contudo, algumas sequências específicas associadas à *loci* imprintados sobrevivem a este processo de desmetilação. Com a exceção de algumas regiões germinais paternalmente metiladas, DMRs associadas a domínios imprintados conhecidos, foi demonstrado que a metilação derivada de espermatozoides é reprogramada pelo estágio de desenvolvimento de blastocisto. Em contraste, um grande número de gDMR derivadas de oócitos sobrevivem ao estágio de blastocisto e persistem exclusivamente como DMRs transientemente metilados apenas na placenta. (Sanchez-Delgado, Court *et al.* , 2016).

A maioria dos genes imprintados possui regiões diferentemente metiladas (DMRs) e a metilação alélica depende da origem parental (Reik e Walter, 2001).

As DMRs estabelecidas através da linhagem germinativa são chamadas de DMRs gaméticas ou primárias. Estas muitas vezes coincidem com as regiões de

controle de *imprinting* (ICRs). Portanto regulam a expressão gênica e outras modificações epigenéticas (Delaval e Feil, 2004; Henckel e Arnaud, 2010). O seu estado de metilação é mantido em todas as linhagens somáticas, uma vez adquiridas.

Outras

As DMRs podem ser "somáticas" ou "secundárias", e são estabelecidas após a fertilização, podendo ser tecido específico (Mann, 2001; Henckel e Arnaud, 2010).

É notório que o perfil das DMRs gaméticas se distinguem das DMRs secundárias, pois nas DMRs gaméticas (gDMR) o metiloma do ovócito e espermatozóide se opõe, um se encontrando hipermetilado e outro hipometilado e os tecidos se encontram com perfil hemimetilado ou apenas um dos tecidos se encontra hemimetilado (tecido específico). Já nas DMRs secundárias (sDMR) tanto o ovócito quanto o espermatozóide, ambos se encontram hipometilado ou hipermetilado e os tecidos se encontram com perfil hemimetilado ou em apenas um tecido está hemimetilado (tecido específico). As DMR imprintadas (iDMR) são de origem parental dependente (Court, Tayama *et al.*, 2014).

É importante ressaltar que genes imprintados desempenham importantes papéis no crescimento e desenvolvimento (Reik e Walter, 2001). Geralmente dois ou mais genes imprintados residem em grupos que tem um centro de controle de *imprinting* (ICC) como regiões diferencialmente metiladas em locais intragênicos ou intergênicos que controlam a expressão dos genes (Wood e Oakey, 2006).

Abaixo estão descritas as DMRs, que podem estar controlando *imprinting* de genes perto ou a longa distância no genoma humano, na Tabela 5. (Docherty, Rezwan *et al.*, 2014) (Court, Tayama *et al.*, 2014).

Tabela 5. Regiões Diferentemente Metiladas

Gene	Chr	Posição Física	Metilação^a
<i>PPIEL</i>	1	chr1:40024626-40025540	M
<i>DIRAS3</i>	1	chr1:68512539-68517273	M
<i>DIRAS3Ex2</i>	1	chr1:68512505-68513486	M
<i>ZDBF2</i>	2	chr2:207114583-207136544	P
<i>NAP1L5</i>	4	chr4:89618184-89619237	M
<i>FAM50B</i>	6	chr6:3849082-3850359	M
<i>PLAGL1</i>	6	chr6:144328078-144329888	M
<i>IGF2R</i>	6	chr6:160426558-160427561	M
<i>WDR27</i>	6	chr6:170054504-170055618	M
<i>GRB10</i>	7	chr7:50848726-50851312	M
<i>PEG10</i>	7	chr7:94285537-94287960	M
<i>MEST</i>	7	chr7:130130122-130134388	M
<i>HTR5A</i>	7	chr7:154862719-154863382	M
<i>CXORF56 psg/ERLIN2</i>	8	chr8:37604992-37606088	M
<i>TRAPPC9</i>	8	chr8:141108147-141111081	M
<i>INPP5F</i>	10	chr10:121578046-121578727	M
<i>H19</i>	11	chr11:2018812-2024740	P
<i>IGF2 DMR2</i>	11	chr11:2153991-2155112	P
<i>IGF2 DMR0</i>	11	chr11:2168333-2169768	P
<i>KvDMR1</i>	11	chr11:2719948-2722259	M
<i>RB1</i>	13	chr13:48892341-48895763	M
<i>IG-DMR</i>	14	chr14:101275427-101278058	P
<i>MEG3</i>	14	chr14:101290524-101293978	P
<i>MEG8</i>	14	chr14:101370741-101371419	M
<i>MKRN3/MIR4508</i>	15	chr15:23807086-23812495	M
<i>MAGEL2</i>	15	chr15:23892425-23894029	M
<i>NDN</i>	15	chr15:23931451-23932759	M
<i>SNRPN</i>	15	chr15:25068564-25069481	M
<i>SNRPN</i>	15	chr15:25093008-25093829	M
<i>SNRPN</i>	15	chr15:25123027-25123905	M
<i>SNURF</i>	15	chr15:25200004-25201976	M
<i>IGF1R</i>	15	chr15:99408496-99409650	M
<i>ZNF597</i>	16	chr16:3481801-3482388	M
<i>ZNF597/NAA60</i>	16	chr16:3492828-3494463	P
<i>ZNF331</i>	19	chr19:54040510-54042212	M
<i>ZNF331</i>	19	chr19:54057086-54058425	M
<i>PEG3</i>	19	chr19:57348493-57353271	M
<i>MCTS2P/HM13</i>	20	chr20:30134663-30135933	M
<i>BLCAP/NNAT</i>	20	chr20:36148604-36150528	M
<i>L3MBTL</i>	20	chr20:42142365-42144040	M

Continuação da tabela 5.

Gene	Chr	Posição Física	Metilação^a
GNAS	20	chr20:57414039-57418612	P
NESP-AS/GNAS-AS1	20	chr20:57425649-57428033	M
GNASXL	20	chr20:57428905-57431463	M
GNASEx1A	20	chr20:57463265-57465201	M
WRB	21	chr21:40757510-40758276	M
NHP2L1	22	chr22:42077774-42078873	M

^a Origem parental da metilação

Outros artigos utilizaram o gene *PEG10*, como controle positivo para *imprinting* genômico (*imprinting* materno). Dentre as DMRs encontradas e citadas no artigo, o gene predito *Gli3* foi sugestivo ao *Imprinting* (Hannula-Jouppi, Muurinen *et al.*, 2014). No entanto é uma região onde não tem hemimetilação no gene nem próximo, assim como outros citados nesse mesmo artigo tendo uma grande evidência que não é uma DMR e nem um gene imprintado.

2. 3 Metilação

A metilação do DNA ocorre em aproximadamente 70% de resíduos de citosina nos genomas de seres humanos e outros mamíferos, quase sempre na sequência 5'-CpG-3'. Conhecemos duas consequências genéticas da sua presença e ambas têm importantes implicações para a saúde humana. Em primeiro lugar, os sítios CpG são *hotspots* para mutação, uma vez que a 5-metilcitosina (5^{me}C) frequentemente muda para T (Bird, 2008).

Mutabilidade do 5^{me}C é sem dúvida a causa mais importante de transtornos genéticos humanos e câncer. Em segundo lugar, a metilação de CpG em promotores de genes pode conduzir a um desligamento transcricional estável. O silenciamento gênico aberrante por metilação de promotores densos de CpG é uma característica da síndrome do X frágil e de muitos cânceres (Bird, 2008).

Os padrões de expressão gênica são determinados por fatores de transcrição ligados ao DNA onde genes de ativação ou repressão da transcrição, reconhecendo a sequência de bases de DNA em regiões promotoras, fazem a interação destes fatores com suas sequências, o que desencadeia uma série de eventos, envolvendo

mudanças na estrutura da cromatina, que leva a um conjunto complexo que ativa transcrição (Bird, 2002).

A característica mais marcante da metilação do DNA dos vertebrados é a presença de ilhas CpG, isto é, regiões ricas em CG não metiladas que possuem densidade de CpG e geralmente estão posicionados nas extremidades 5' de muitos genes humanos (Bird, 2002). A análise computacional da sequência do genoma humano prediz 29.000 ilhas CpG (Lander, Linton *et al.*, 2001; Venter, Adams *et al.*, 2001). Estudos anteriores estimaram que aproximadamente 60% dos genes estão associados às ilhas CpG (Bird, 2002).

2.4 Reprogramação do *Imprinting* nas Gerações

O *imprinting* é reprogramado a cada geração e as marcas existentes na região regulatória são apagadas, e em seguida, um novo conjunto de *imprinting* de acordo com o desenvolvimento das células germinativas é adquirido subsequentemente e marcas imprintadas do zigoto permanecem pós-eventos mitóticos para todos tecidos somáticos, por causa do papel crucial da metilação do DNA pela enzima DNMT2 (DNA-metil transferase 2), que tem papel na manutenção da metilação pós-mitose, por exemplo, (Obata e Kono, 2002).

As células germinativas de ambos os sexos podem sofrer uma reprogramação epigenética necessária para restaurar a totipotência (Seki, Hayashi *et al.*, 2005; Hajkova, Ancelin *et al.*, 2008; Arnaud, 2010).

Novas marcas de metilação são adquiridas mais tarde em estágios de desenvolvimento de células germinativas, em tempos diferentes nos dois sexos. Em células germinativas masculinas os genes que estão sob *imprinting* paterno, ou seja, com expressão materna, são adquiridos em mitose, o *imprinting* começa antes do nascimento, em pró-espermatogônias, e é mantido através das fases meióticas e haploides (Davis, Yang *et al.*, 2000; Kato, Kaneda *et al.*, 2007).

No entanto, as marcas de metilação maternas são totalmente adquiridas na fase da metáfase da meiose II. Sendo assim o *imprinting* materno é adquirido após o nascimento, durante o crescimento do oócito (Lucifero, Mann *et al.*, 2004; Hiura, Obata *et al.*, 2006).

A reprogramação de metilomas parentais é essencial para o desenvolvimento embrionário. Em mamíferos, 5-metilcitosinas paternas (5mCs) foram propostas para

serem ativamente convertidas em bases oxidadas. Acredita-se que estas bases oxidadas paternas e 5mCs maternas sejam passivamente diluídas por divisões celulares.

Portanto, o metiloma paterno e pelo menos uma proporção significativa de metiloma materno passam por desmetilação ativa durante o desenvolvimento embrionário. No entanto, o metiloma paterno e pelo menos uma proporção significativa de metiloma materno passam por desmetilação ativa durante o desenvolvimento embrionário. Adicionalmente, todas as regiões de controle de *imprinting* conhecidas (ICRs) foram classificadas em ICR de linha germinal ou somáticas. (Wang, Zhang *et al.* , 2014).

Os epigenomas de esperma e oócito são drasticamente diferentes. A reprogramação de epigenomas parentais é essencial para a compatibilidade da totipotência durante o desenvolvimento do embrião (Hackett e Surani, 2013). Em mamíferos, ocorrem duas ondas de desmetilação de DNA do genoma: durante o desenvolvimento de células germinativas primordiais (PGC) e embriogênese precoce (Seisenberger, Peat e Reik, 2013; Wu e Zhang, 2014).

É importante ressaltar que se o *imprinting* for materno, sempre vai passar o cromossomo materno imprintado para o filho, mesmo que o cromossomo seja proveniente do avô materno. E se o *imprinting* for paterno que é mais raro (como *H19* e *MEG3*) sempre vai passar o cromossomo paterno imprintado para o filho.

Na formação do zigoto pós-fecundação e na formação das células germinativas (gametas) predomina o estado hipometilado. Nesse momento que ocorre a metilação *De novo*. Há uma diferença do tempo de desmetilação das células primordiais germinativas masculinas e femininas.

O *imprinting* genômico em mamíferos é importante para o desenvolvimento embrionário (Surani, Kothary *et al.*, 1990). A perda de *imprinting* está associada com muitas doenças humanas (Lalande, 1996).

2.5 Dissomia Uniparental (UPD)

Genes imprintados expressam apenas o alelo materno ou paterno e o desequilíbrio dos genes imprintados pode ser por dissomia uniparental (UPD), onde a herança de ambos os cromossomos pode ser apenas da mãe ou apenas do pai

(Engel, 1980; Joshi, Garg *et al.* , 2016). Esse evento resulta clinicamente em síndromes como UPD no cromossomo 15 (Materna/ Paterna) Prader-Willi, Angelman, UPD no cromossomo 11 (Paterna) Beckwith-wiedemann ou UPD no cromossomo 6 (Paterno), Diabetes Mellitus neonatal transitório. (Joshi, Garg *et al.* , 2016).

A perda de metilação em um *locus* imprintado é relatada para a dissomia uniparental materna do cromossomo 7, que é encontrada em cerca de 10% dos pacientes com síndrome de Silver-Russell (Gicquel, Rossignol *et al.* , 2005; Bruce, Hannula-Jouppi *et al.* , 2009). Essa síndrome é caracterizada por restrição de crescimento pré e pós-natal grave, assimetria esquelética e um rosto triangular (Price, Stanhope *et al.*, 1999). Além dessas características típicas tem retardo do crescimento, mas nenhum gene tem sido conclusivamente envolvido nessa síndrome (Engel, 1980; Buiting, 2010; Hannula-Jouppi, Muurinen *et al.* , 2014).

O *imprinting* ocorre através de marcas epigenéticas nos dois alelos parentais, com a maioria dos *loci* imprintados marcados pela presença de DMRs. Para identificar sítios de viés epigenético parental, existem estudos perfilando os padrões de metilação do DNA em indivíduos com UPD em vários cromossomos (Joshi, Garg *et al.* , 2016).

Portanto definindo DMRs imprintadas foram observados modestos níveis de metilação do DNA em cada CpG na síndrome de Prader-Willi e Angelman na região 15q11-13, como mencionada anteriormente, demonstrando que a influência do *imprinting* não se limita a elementos reguladores individuais, como as ilhas CpG, podendo se estender por domínios cromossômicos inteiros. Foi analisado que no gene estudado tem uma DMR com microdeleção na região 15q11. 2, sugerindo potenciais efeitos a esse distúrbio genômico (Joshi, Garg *et al.* , 2016).

Além disso, foram utilizados dados de RNA-Seq em que detectaram assinaturas consistentes com o *imprinting* associado com várias DMRs. Utilizando vários metilomas de sangue em uma amostra populacional, definiram padrões de variação epigenética (Joshi, Garg *et al.* , 2016).

Em DMRs identificaram indivíduos raros com ganho global ou perda de metilação através de múltiplos *loci* imprintados. Esses dados forneceram um mapa de viés epigenético parental no genoma humano (Joshi, Garg *et al.* , 2016).

É importante ressaltar que cada gene é avaliado quanto à consistência com a expressão monoalélica. Os dados consideram três categorias de expressão gênica: monoalélico, bialélico, desvio do bialélico (Joshi, Garg *et al.* , 2016).

Os *loci* imprintados nem sempre têm como marcas epigenéticas elementos reguladores ilhas CpG ou promotores à região Prader- Willi/ Angelman e sim pode estender-se sobre domínios cromossômicos inteiros. Em segundo lugar, nem todos os locais metilados com genes imprintados tem um padrão com perfil de metilação intermediária, com um alelo parental hipermetilado e outro alelo hipometilado na mesma região. Foram identificados também raros indivíduos que exibiram consistentemente elevados ou baixos níveis de metilação em dezenas de *loci* imprintados (Joshi, Garg *et al.* , 2016).

Embora a perda de *imprinting* seja uma causa conhecida de síndrome congênita, as consequências funcionais das perturbações em outros *loci* imprintados permanecem desconhecidas (Joshi, Garg *et al.* , 2016).

2.6 Expressão Monoalélica Estocástica

Interessantemente, há outro mecanismo de expressão monoalélica além do caracterizado por origem parental-dependente. Este, na literatura, é chamado de estocástico (ou randômico) (Eckersley-Maslin e Spector, 2014).

No caso desse tipo expressão monoalélica estocástica, a transcrição ocorre a partir de apenas um dos dois alelos homólogos em cada célula diplóide. Curiosamente, entre 5% e 15% dos genes autossomos exibem expressão monoalélica randômica, em que as diferentes células expressam apenas um alelo de forma independente da sequência genômica de base, de uma forma específica do tipo celular. A expressão gênica monoalélica é provavelmente gerada através da regulação independente estocástica dos dois alelos na diferenciação, e tem variadas implicações para a célula e para o organismo, em particular no que diz respeito à doença. Convém mencionar que a expressão monoalélica aleatória aumenta após a diferenciação e que ocorre em uma variedade de tipos celulares humanos (Eckersley-Maslin e Spector, 2014).

2.7 Modificações Pós Traducionais em Proteínas Histonas

A primeira ordem de compactação do DNA é dada pela estrutura dos nucleossomos, a unidade funcional da cromatina, formado pela junção de proteínas histonas, responsáveis pelo remodelamento da cromatina. Em sua estrutura as histonas classificam-se como um complexo de oito proteínas (H2A, H2B, H3 e H4) aos pares mais a histona H1 (Luger, Mader *et al.*, 1997; Kouzarides, 2007).

Histonas são compostas por domínios funcionais e caudas N-terminais. Diversas modificações pós-traducionais acometem as sequências das caudas das histonas. Geralmente, os resíduos de aminoácidos que mais sofrem modificações são as lisinas e argininas (Barski, Cuddapah *et al.*, 2007; Kouzarides, 2007). Sendo as modificações mais comuns: acetilação, metilação, fosforilação e ubiquitinação, influenciando de formas distintas a estrutura da cromatina e a expressão gênica (Kouzarides, 2007).

A regulação da expressão gênica durante os diferentes estágios de vida da célula leva a cromatina a sofrer modificações drásticas em seu estado estrutural, isso por sua vez pode interferir no perfil alélico de muitos genes. A estrutura aberta da eucromatina facilita o acesso de fatores em *trans* para início da transcrição (Grewal e Moazed, 2003), durante o qual, as histonas exercem um papel fundamental para organização da cromatina.

Dependendo do tipo de modificação do presente na cauda da histona a expressão do gene pode ou não acontecer, por exemplo, a transferência de um grupamento acetil por enzimas histonas acetiltransferases (HATs) nas lisinas das caudas N-terminais de histonas altera suas propriedades físico-químicas, levando-as a perda da capacidade de se ligarem lateralmente aos grupamentos fosfatos da molécula de DNA por neutralizarem a carga positiva das lisinas, isso leva a expressão gênica (Li, Carey *et al.*, 2007) (Andreoli e Del Rio, 2014).

Além da sua função na regulação da transcrição, acetilação de histonas participa no processo de replicação do DNA, agindo eficientemente sobre a ligação DNA-nucleossomo, quebra da dupla fita de DNA e facilitando o acesso da maquinaria de reparo do DNA (Butler, 2009; Unnikrishnan, Gafken *et al.*, 2010).

As proteínas histonas envolvem o DNA para formar nucleossomos, que compactam ainda mais na estrutura da cromatina. Além das histonas anteriormente

mencionadas, existem histonas variantes que têm frequentemente papéis fundamentais na regulação da dinâmica da cromatina e na acessibilidade do DNA subjacente (Shi, Wen *et al.* , 2016).

Além disso, modificações como metilação de histonas podem levar a expressão ou silenciamento de alguns loci, apresentando um importante papel para o desenvolvimento e regulação de genes. Por exemplo, no contexto do *Imprinting* genômico a presença de marcas diferenciais em alelos diferencialmente marcados constituem um panorama para mecanismos epigenéticos regulatórios de genes imprintados (Bartolomei e Tilghman, 1997; Delaval, Wagschal *et al.* , 2006).

Portanto, *imprinting* é um mecanismo epigenético envolvendo tanto metilação do DNA quanto modificações em caudas de histonas nos alelos parentais, com a maioria dos genes imprintados marcados por CpG ricos em DMRs, caracterizando as chamadas ilhas CpG (iCpG) (Joshi, Garg *et al.* , 2016).

2.8 RNAs Não Codificantes

A maioria dos transcritos humanos não são traduzidos, mas um grande número deles ainda tem funções vitais. Os RNAs não codificantes (ncRNAs) são um agrupamento de RNAs que não codificam proteínas funcionais e foram originalmente considerados como meramente reguladores da expressão genética a nível pós-transcricional. Contudo, existe um número crescente de evidências de que os RNAs não codificantes desempenham um papel importante no controle epigenético, como por exemplo, o gene *XIST* (*X-Inactive Specific Transcript*), que quando expresso ocorre a da inativação do cromossomo X, recruta complexos de repressão que têm fundamental importância na formação do corpúsculo de Barr (Wei, Huang *et al.* , 2017).

Portanto, esses ncRNAs destacam o papel proeminente do RNA na regulação da expressão gênica. Há desenvolvimentos de pesquisas recentes com o propósito de chegar a uma melhor compreensão de RNAs não codificantes e seus mecanismos de ação em células, ganhando assim um entendimento preliminar de que os RNAs não codificadores retornam a uma rede reguladora epigenética (Wei, Huang *et al.* , 2017).

3. JUSTIFICATIVA

O desenvolvimento de um sistema robusto para a identificação computacional de novas regiões hemimetiladas no genoma humano será um valioso instrumento para validar *in silico* DMR e genes sujeitos ao *imprinting*. A identificação de novas DMRs imprintadas pode ser um subsídio para o estudo de genes envolvidos em diversas patologias associadas às alterações epigenéticas e de expressão gênica consistente com *imprinting* genômico.

4. OBJETIVOS

4. 1. Objetivo Geral

Identificar computacionalmente novas regiões hemimetiladas no genoma humano e validar a predição de DMR e genes imprintados no genoma humano através da caracterização da metilação monoalélica dependente da origem parental e confirmação da expressão monoalélica da iDMR através de RNA-Seq.

4. 2. Objetivos Específicos

- Identificar computacionalmente regiões hemimetiladas no genoma humano;
- Validar a predição de DMR e genes imprintados no genoma humano;
- Investigar metilação de origem parental utilizando polimorfismos de único nucleotídeo;
- Identificar computacionalmente através de RNA-Seq novos SNPs com expressão monoalélica e bialélica a fim de validar a predição de genes imprintados.

5. MATERIAIS E MÉTODOS

5. 1. Material Biológico e Extração de DNA

Sangue periférico de voluntários saudáveis foram submetidas ao método de extração do DNA genômico utilizando o kit de extração e purificação de DNA genômico *Illustra blood genomicPrep*, GE Healthcare, UK, conforme a especificação do fabricante. Todas as amostras extraídas foram quantificadas para que, o que foi quantificado pelo Thermo Scientific NanoDrop 2000c (Thermo Fisher Scientific) e estocadas a -20°C. As amostras foram armazenadas no banco de DNA do Núcleo de Diagnóstico de Investigação Molecular (NUDIM).

5. 2. Aspectos Éticos da Pesquisa

As amostras utilizadas fazem parte de um banco repositório de DNA do NUDIM que possui autorização do Comitê de Ética em pesquisa envolvendo seres humanos da Faculdade de Medicina de Campos, RJ e do CONEP sob CAAE no. 35385714. 0. 0000. 5244 (Conselho Nacional de Saúde do Brasil - CONEP – Ministério da Saúde, Plataforma Brasil). A inclusão neste projeto e disponibilização do DNA no Banco de DNA foi voluntária por meio de assinatura de Termo de Consentimento Livre e Esclarecido.

5. 3. Varredura e desenho de marcadores

Domínios DMR supostamente imprintados, foram identificados computacionalmente por comparação de dados de perfis de metilação em 120 projetos metilomas disponíveis em plataformas públicas. A visualização dos dados de metilomas foi feita na plataforma *Genome Browser* da UCSC, customizando pistas (*tracks*) dos perfis de metilação em centenas de linhagens celulares e tecidos humanos incluindo oócito e espermatozóide. O perfil alvo está representado na figura 1, que demonstra o perfil de um gene conhecido na literatura (*PEG10*) do cromossomo 7, onde as células do oócito são hipermetiladas e as células dos espermatozoides são hipometiladas sendo um exemplo de perfil característico de *imprinting* materno, com expressão paterna. O método de rastreamento incorporou tanto DMR primária (representada na figura 1) quanto os genes com

perfil de DMR secundária. A janela de visualização foi de 40 Kb, previamente estabelecida pelo nosso grupo de pesquisa.

A metodologia padrão-ouro para a identificação dos resíduos de C metilados no contexto CpG é o sequenciamento de DNA tratado com bissulfito de sódio, o qual converte as citosinas não metiladas em uracilas através da perda do grupo amina dessas bases e posteriormente na reação de sequenciamento para resíduos T. As citosinas metiladas, por outro lado, não são convertidas pelo bissulfito de sódio (Allen S. Yang e Issa., 2004). A proporção de número de sequências lidas (*reads*) contendo resíduos T ou C em cada sítio CpG de referência determina o estado de metilação nesse ponto. Por exemplo, para uma cobertura de 20 reads, em hipótese, o estado hipometilado corresponde a 20 reads com resíduos T. O estado hipermetilado, a 20 reads com resíduos C e o hemimetilado a 10 reads com T e 10 reads com C. Nessa análise, os níveis de metilação são expressos em porcentagem de 0-100. Esses dados são então esquematizados e armazenados em plataformas online de pesquisa genômica, como, por exemplo, o *Genome Browser*, da Universidade da Califórnia em Santa Cruz.

Utilizando essa ferramenta os critérios de inclusão para predição de um gene imprintado foram:

- Critério Maior

- Perfil hemimetilado entre (0,35 a 0,65) em todos os tecidos disponíveis no track.
- Ilha CpG na região da hemimetilação ou regiões próximas.
- Presença de AMR (região de metilação alélica) na região da hemimetilação.

- Critério Menor

- Analisar se na região onde tem hemimetilação em todos os tecidos há região promotora do gene, pois é descrito na literatura que a metilação regula a região promotora não sendo algo normal esta hemimetilação constitutiva.

- Analisar se na região da hemimetilação tem retrotransposon. Pois esta assimetria da metilação pode estar sendo proveniente de outro cromossomo dependendo da porcentagem de homologia.

É notório que observar os perfis dos metilomas dos genes conhecidamente imprintados ou regiões próximas a estes genes imprintados, para definir algoritmos, para encontrar novas DMRs no genoma e analisar se é imprintada e se pertence a um gene imprintado é de enorme valia. Em genes conhecidamente imprintados, por exemplo, como *PEG10* a DMR é gamética, o ovócito (hipermetilado) e o espermatozóide (hipometilado), nos tecidos tem perfil hemimetilação na maioria dos tecidos (metilomas) e a expressão é monoalélica origem parental dependente, sendo um gene de *imprinting* materno como é mostrado na figura 1 abaixo.



Figura 1. Assimetria gamética da metilação na região promotora do gene imprintado *PEG10*. Perfis hipermetilado em células do oócito e hipometilado em células do espermatozoide e metilação intermediária nas demais linhagens celulares é um exemplo de perfil característico de *imprinting* materno.

5. 4. Critérios de Seleção para *loci* Potencialmente Polimórficos

Para selecionar novos marcadores potencialmente polimórficos de DNA para estudo de metilação origem parental dependente no genoma humano, foram escolhidos SNPs intrônicos e exônicos na área hemimetilada em genes preditos com frequência mínima alélica (MAF) maior que 0,15. Na escolha dos SNPs foi verificado se este se encontra em regiões de repetição *LINE*, *SINE*, pois estas regiões não são as melhores regiões para análise por serem regiões de transposons, repetidas. Adicionalmente para estudo de *Imprinting* genômico no genoma humano através de

estudo *in silico* com RNA-Seq foram selecionados SNPs exônicos (MAF>0,2) em genes preditos e com metilação origem parental dependente confirmada.

5.5. Iniciadores Marcador-Específicos

Para o desenho de iniciadores específicos para os ensaios de Digestão Enzimática Sensível a Metilação Associada à Reação em Cadeia da Polimerase Quantitativa por Fluorescência (DESM-QFPCR) e sequenciamento por extensão de um único nucleotídeo foram utilizados os programas livres e online *OligoPerfect™ Designer* da *Invitrogen™* (<http://www.invitrogen.com/>) e o *OligoCalc Oligonucleotide Properties Calculator* (<http://www.basic.northwestern.edu/biotoools/oligoalc.html>). Para o ensaio de DESM-QFPCR os iniciadores utilizados na amplificação dos alvos foram modificados com fluorocromos 6-FAM. O produto de amplificação não tinha mais que 400 pares de base e os marcadores não apresentavam complementaridade (Figura 2).

chr7:42005486-42005755

```
ACCTCCTCATCTTTTCCCTGCCTTCCATCCTCCTGTACCAATTGATGCCAGACATCATGAGGGCCGTTACCATT  
ACGATCCATCTCCGATTCCCTCCATTGCATATGACTTCCGCCTTATCTAGTAGCCCTACGTATCCGGACCTGCCCT  
TCATTAGGATCTCCCCACACCGGAACCCCACTGCTGCTTCCGAGTCTCCCTTCAGCCCTCCACATCCCTACATT  
AATCCCTACATGGACTATATCCGCTCCTTGACACAGCAGCCCATCGCTCTCCATGATCTCA GCAACCCGTGGGCTG  
AGCCCTACAGATG
```

```
ACCTCCTCATCTTTTCCCTGCCTTCCATCCTCCTGTACCAATTGATGCCAGACATCATGAGGGCCGTTACCATT  
ACGATCCATCTCCGATTCCCTCCATTGCATATGACTTCCGCCTTATCTAGTAGCCCTACGTATCCGGACCTGCCCT  
TCATTAGGATCTCCCCACACCGGAACCCCGCTGCTGCTTCCGAGTCTCCCTTCAGCCCTCCACATCCCTACATT  
AATCCCTACATGGACTATATCCGCTCCTTGACACAGCAGCCCATCGCTCTCCATGATCTCA GCAACCCGTGGGCTG  
AGCCCTACAGATG
```

Figura 2. O esquema representativo *in silico* de um SNP do gene do cromossomo 7. Para demonstração com iniciador sentido e anti sentido em verde. O produto de amplificação possui 203pb em que o SNP rs846266 se encontra na posição de 99pb.

5.6. Tipagem dos Marcadores

Os SNPs foram tipados pela metodologia de SNaPshot®, utilizando o kit comercial ABI PRISM® SNaPshot® Multiplex e os iniciadores desenhados para este estudo. Para padronização dos ensaios foram realizados gradientes com as seguintes variáveis: temperatura de anelamento, concentração de iniciadores,

concentração de $MgCl_2$ e concentração de DNA alvo. A reação de amplificação foi preparada em volume final de 12,5 μ L realizada da mesma forma que a utilizada para o ensaio DESM-QFPCR para cada *locus*. Para a confirmação da especificidade da reação, os produtos da PCR alvos do SNaPshot foram analisados em gel de agarose 2% para cada marcador de DNA genômico.

O ensaio DESM-QFPCR determina o percentual de metilação no *locus* alvo. A quantificação é realizada em unidades relativas de fluorescência em relação ao *locus* Controle Negativo (sem sítio de restrição, CN), pela razão entre os valores de área de cada fragmento nos ensaios teste e referência. Juntamente com o *locus* alvo e o CN, foi utilizado um Controle Positivo (com sítios de restrição a enzimas sensíveis a metilação, CP). Cada amostra de DNA foi estratificada em dois tubos: amostras tratadas (ensaio teste) e não tratadas (ensaio referência) com enzimas de restrição sensíveis a metilação (DESM). Como controle positivo foi utilizado um amplicon alvo de 248pb no gene *ESCO2* ou 150pb no gene *PPP2R3B*, sabidamente desmetilado, contendo quatro sítios *HhaI*, um sítio *BstUI*, e um sítio *HpaII* no amplicon de 248pb. Como controle negativo foi utilizada uma região de 270pb no gene *WRB* (21q22. 2) sem sítios de restrição para as enzimas *HhaI*, *BstUI* e *HpaII*. No gene *PWWP2AP1* foi investigado o estado de metilação de um sítio CpG-*HpaII* em um fragmento de 261pb. No gene *PARD6G-AS1/ADNP2* foi investigado o estado de metilação de um sítio CpG-*HpaII* em um fragmento de 263pb. No gene *GCSAML* foi investigado o estado de metilação de sítios CpG-*HpaII* em um fragmento de 264pb. No gene *PMF1* foi investigado o estado de metilação de um sítio CpG-*HpaII* em um fragmento de 263pb.

Os iniciadores desenvolvidos bem como o número de sítios para cada alvo foram desenhados especialmente para esse projeto. Em comum, essas quatro regiões possuem uma extensão maior de hemimetilação comparada às demais e SNPs passíveis de serem utilizados para a análise experimental quanto à metilação e a origem parental. Também para serem testados experimentalmente foram desenhados o controle positivo (*PPP2R3B*) e controle negativo (*WRB*) (tabela 6).

Tabela 6. Análise integrativa das candidatas a DMRs alvos e desenho dos iniciadores.

Gene	SNP	MAF	Sentido	Antissentido	Pb	Enzimas
<i>PWWP2AP1</i>	rs1176323	G=0.4397/2202	ACGGAGGACGACAAGATAAGC	CTGGAGGCTGAGGGGAAG	261	<i>HhaI, HpaII</i>
<i>PARD6G-AS1</i>	rs11281142	CTGTGGTGC=0.3335/1670	CTCGGCTGCCTCGTCCTA	TTCAACTCGATAAACCTGGTG	263	<i>BstUI, HhaI, HpaII</i>
<i>GCSAML</i>	rs6700954	C=0.3876/1941	CAGTCACCGCTGCCTCTG	GGTTGCAGGCAGTTGAAG	264	<i>BstUI, HhaI, HpaII</i>
<i>PMF1</i>	rs4998248	G=0.3834/1920	GTGTCAGGCGTTCTCGTCTC	GTTAGAATATGCGGTGTTTGGTT	263	<i>HpaII</i>
<i>PPP2R3B</i>			GCTCAGGGCAGCTTCAAA	AAAGCCTCCTGCAAAAGCTC	160	<i>HhaI, HpaII</i>
<i>WRB</i>			TTCAGGGTCTGGAGCTTCCT	AATGTTAATCCAAGACCCACGTA	269-271	-

5.6.1 DESM-QFPCR para o fragmento PWWP2AP1-DMR

Reações DESM foram preparadas em volume final de 10µL, contendo 400ng de DNA, tampão CutSmart® 1X, e 6U de Hhal CutSmart® (*New England Biolabs*), as reações de digestão enzimática foram realizadas no termociclador Veriti® 96-Well VeriFlex (*Applied Biosystems*™) sob a seguinte condição térmica: 37°C por 120 minutos e 65°C por 20 minutos. Após digestão, uma alíquota de 1,5µL do DNA digerido foi submetida à amplificação por QFPCR. A reação de amplificação foi preparada em volume final de 12,5µL sob a seguinte formulação dos iniciadores descritos no apêndice A: 0,8 µM do par de iniciadores ESCO2 ou PPP2R3B (CP), 0,12 µM do par de iniciadores WRB (CN), 0,2 µM do par de iniciadores PWWP2AP1, 2,5M de betaína (*SIGMA-ALDRICH*®), 0,2mM de dNTPs, 15mM de Tris-HCl, 50mM KCl, 2mM de MgCl₂, 1,5U de DNA polimerase Taq Sinapse® (*Sinapse*). Para cada ensaio teste, foi preparado um ensaio referência, onde não foi adicionado a enzima de restrição HpaII. As reações foram realizadas no termociclador Veriti® 96-Well VeriFlex (*Applied Biosystems*™), a reação de amplificação foi realizada sob o seguinte ciclo térmico: (i) 95°C por 11 minutos, (ii) 28 ciclos à 94°C por um minuto, 62°C por um minuto, 72°C por um minuto, e (iii) 60°C por 60 minutos.

5.6.2 DESM-QFPCR para o fragmento 263 PARD6G-AS1/ADNP2-DMR

Reações DESM foram preparadas em volume final de 10µL, contendo 400ng de DNA, tampão CutSmart® 1X, e 6U de Hhal e HpaII CutSmart® (*New England Biolabs*), as reações de digestão enzimática foram realizadas no termociclador Veriti® 96-Well VeriFlex (*Applied Biosystems*™) sob a seguinte condição térmica: 37°C por 120 minutos e 65°C por 20 minutos. Após digestão, uma alíquota de 1,5µL do DNA digerido foi submetida à amplificação por QFPCR. A reação de amplificação foi preparada em volume final de 12,5µL sob a seguinte formulação dos iniciadores descritos no apêndice A: 2 µM do par de iniciadores ESCO2 ou PPP2R3B (CP), 0,08 µM do par de iniciadores WRB (CN), 0,38 µM do par de iniciadores PARD6G-AS1/ADNP2, 2,5M de betaína (*SIGMA-ALDRICH*®), 0,2mM de dNTPs, 15mM de Tris-HCl, 50mM KCl, 2mM de MgCl₂, 1,5U de DNA polimerase Taq Sinapse® (*Sinapse*). Para cada ensaio teste, foi preparado um ensaio referência, onde não foi adicionado a enzima de restrição HpaII. As reações foram realizadas no termociclador Veriti® 96-Well VeriFlex (*Applied Biosystems*™), a reação de

amplificação foi realizada sob o seguinte ciclo térmico: (i) 95°C por 11 minutos, (ii) 28 ciclos à 94°C por um minuto, 62°C por um minuto, 72°C por um minuto, e (iii) 60°C por 60 minutos.

5.6.3 DESM-QFPCR para o fragmento GCSAML-DMR

Reações DESM foram preparadas em volume final de 10µL, contendo 400ng de DNA, tampão CutSmart® 1X, e 6U de *HhaI* CutSmart® (*New England Biolabs*), as reações de digestão enzimática foram realizadas no termociclador *Veriti*® 96-Well *VeriFlex* (*Applied Biosystems*™) sob a seguinte condição térmica: 37°C por 120 minutos e 65°C por 20 minutos. Após digestão, uma alíquota de 1,5µL do DNA digerido foi submetida à amplificação por QFPCR. A reação de amplificação foi preparada em volume final de 12,5µL sob a seguinte formulação dos iniciadores descritos no apêndice A: 1,6 µM do par de iniciadores *ESCO2* ou *PPP2R3B* (CP), 0,12 µM do par de iniciadores *WRB* (CN), 0,16 µM do par de iniciadores *GCSAML*, 2,5M de betaína (*SIGMA-ALDRICH*®), 0,2mM de dNTPs, 15mM de Tris-HCl, 50mM KCl, 2mM de MgCl₂, 1,5U de DNA polimerase Taq *Sinapse*® (*Sinapse*). Para cada ensaio teste, foi preparado um ensaio referência, onde não foi adicionado a enzima de restrição *HpaI*. As reações foram realizadas no termociclador *Veriti*® 96-Well *VeriFlex* (*Applied Biosystems*™), a reação de amplificação foi realizada sob o seguinte ciclo térmico: (i) 95°C por 11 minutos, (ii) 28 ciclos à 94°C por um minuto, 62°C por um minuto, 72°C por um minuto, e (iii) 60°C por 60 minutos.

5.6.4 DESM-QFPCR para o fragmento PMF1-DMR

Reações DESM foram preparadas em volume final de 10µL, contendo 400ng de DNA, tampão CutSmart® 1X, e 6U de *HhaI* CutSmart® (*New England Biolabs*), as reações de digestão enzimática foram realizadas no termociclador *Veriti*® 96-Well *VeriFlex* (*Applied Biosystems*™) sob a seguinte condição térmica: 37°C por 120 minutos e 65°C por 20 minutos. Após digestão, uma alíquota de 1,5µL do DNA digerido foi submetida à amplificação por QFPCR. A reação de amplificação foi preparada em volume final de 12,5µL sob a seguinte formulação dos iniciadores descritos no apêndice A: 1,2 µM do par de iniciadores *ESCO2* ou *PPP2R3B* (CP), 0,09 µM do par de iniciadores *WRB* (CN), 0,4 µM do par de iniciadores *PMF1*, 2,5M de betaína (*SIGMA-ALDRICH*®), 0,2mM de dNTPs, 15mM de Tris-HCl, 50mM KCl,

2mM de MgCl₂, 1,5U de DNA polimerase Taq Sinapse[®] (*Sinapse*). Para cada ensaio teste, foi preparado um ensaio referência, onde não foi adicionado a enzima de restrição HpaII. As reações foram realizadas no termociclador *Veriti[®] 96-Well VeriFlex* (*Applied Biosystems*[™]), a reação de amplificação foi realizada sob o seguinte ciclo térmico: (i) 95°C por 11 minutos, (ii) 28 ciclos à 94°C por um minuto, 62°C por um minuto, 72°C por um minuto, e (iii) 60°C por 60 minutos.

5. 7. Caracterização dos Alelos Amplificados e Sequenciamento

Aos produtos da extensão de um único nucleotídeo (SNaPshot[®]) foram adicionados formamida (Hi-Di Formamida, Applied Biosystems) e o padrão de massa molecular GeneScan LIZ 120 marcado com o fluorocromo LIZ[™] (fluorescência laranja) (Applied Biosystems).

Aos produtos da DESM-QF-PCR foram adicionados formamida (Hi-Di Formamida, Applied Biosystems) e o padrão de massa molecular GeneScan LIZ 500 marcado com o fluorocromo LIZ[™] (fluorescência laranja) (Applied Biosystems).

Os produtos de amplificação foram separados por eletroforese capilar na plataforma ABI Prism[™] 310 Genetic Analyzer, utilizando o polímero POP4 (contendo 4-6% de acrilamida). Os perfis eletroforéticos foram analisados utilizando os programas GeneScan[™] e Genotyper[™] (Applied Biosystems).

5.8. Demonstração do Perfil do Eletroferograma para Caracterização de Predição de *Imprinting* Genômico

Na Figura 3 há uma demonstração de como foram os ensaios e visualização dos eletroferogramas antes e depois da digestão com enzima sensível a metilação (que digere região desmetilada) para verificação do estado de metilação do *locus*.

É importante ressaltar que as enzimas (*Bst*UI, *Hha*I, *Hpa*II) digerem região desmetilada e que o controle positivo desaparece depois da digestão, já o controle negativo continua com mesmo perfil após a digestão, justamente por não ter sítios para essas enzimas. Se a amostra alvo ficar, após a digestão, com a metade da área, configura um perfil de metilação intermediária que sugere *imprinting* genômico.

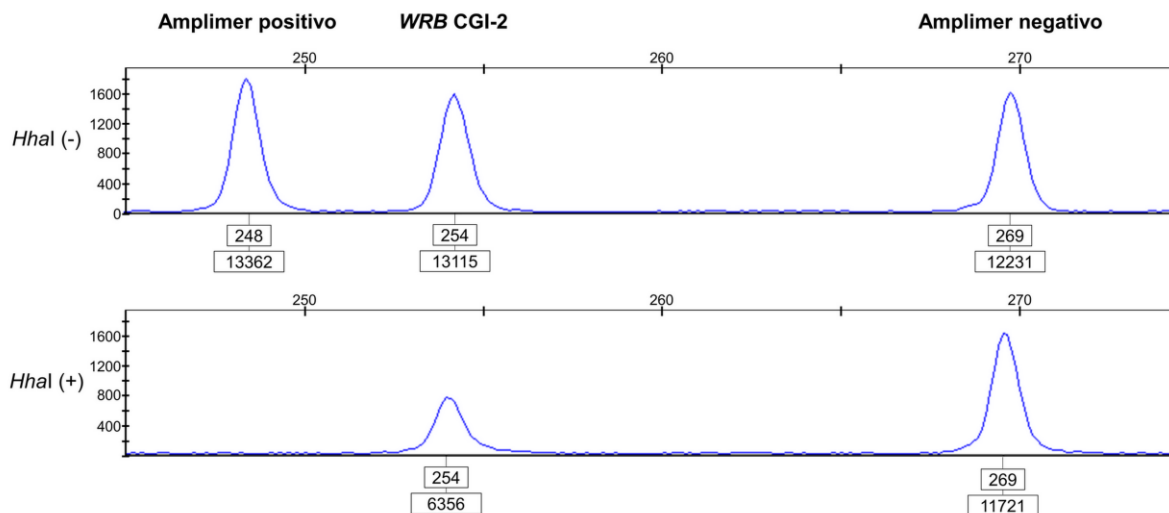


Figura 3. Eletroferograma representativo da identificação de locus com perfil de metilação intermediária alélica específica. À esquerda com 248pb temos o controle positivo que depois da digestão com a enzima *HhaI* desaparece a região desmetilada. Ao meio com 254pb temos o alvo que depois da digestão, apresentou metade da área (perfil hemimetilado) e do lado direito com 269pb o controle negativo sem sítio de digestão, permanecendo sem modificação depois da digestão. Demonstrando, no entanto um exemplo característico de *imprinting* genômico. Adaptado de (Alves Da Silva, Machado *et al.* , 2016).

5.9. Demonstração do Perfil Eletroferograma para Determinação da Origem Parental do *Imprinting*

Para determinar a origem parental do “*Imprinting*” é necessário em uma DMR desenhar iniciadores com sítio da enzima de restrição (*BstUI*, *HhaI* ou *HpaII*) e SNP de preferência com MAF maior que 0,2. Na figura 4 está uma demonstração de como determinar a origem parental do “*Imprinting*” utilizando amostras de pai, mãe e filho. Na figura verifica-se que após a digestão o alelo A de origem paterna que foi digerido (região desmetilada) e o filho permaneceu apenas com o alelo materno C caracterizando “*Imprinting*” Materno.

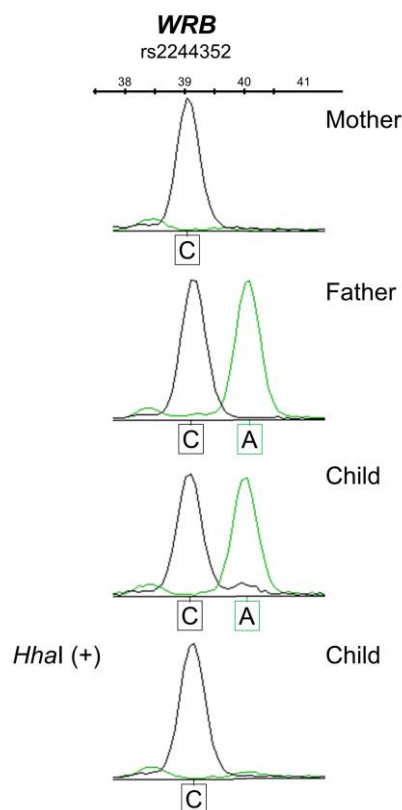


Figura 4. Eletroferogramas representativos da genotipagem do SNP rs2244352 no gene *WRB* a fim de se determinar a origem parental da metilação. A metilação específica para o alelo da DMR foi avaliada experimentalmente através do SNP rs2244352 no DNA genômico de uma amostra controle e seus pais. A permanência do alelo rs2244352*C de origem materna demonstra que o alelo materno foi consistentemente metilado (refratários à digestão com *HhaI*). Enquanto que o alelo paterno rs2244352*A é não metilado (susceptível à digestão com *HhaI*). Assim a base genética subjacente para os perfis de hemimetilação observados no *WRB* DMR é inequivocamente devido à herança de impressões maternas alelo- específicos. Adaptado de (Alves Da Silva, Machado *et al.* , 2016).

5.10. RNA-Seq

RNA-Seq é uma técnica da biologia molecular utilizada para investigar a expressão gênica a partir da presença ou ausência do cDNA. Nós utilizamos a técnica em uma metodologia *in silico* a qual investigamos a expressão dos genes de interesse em um campo amostral oriundo de bancos de dados públicos da plataforma SRA, no NCBI.

Com auxílio da plataforma *Genome Browser*, da Universidade da Califórnia em Santa Cruz, foram selecionados polimorfismos de nucleotídeo único (SNPs) exônicos nos genes de interesse. Esses SNPs foram obtidos e analisados no banco de dados database SNPs do NCBI quanto ao seu MAF global ($>0,2$ e $< 0,5$), classe de variação (selecione apenas SNP), classe funcional (3' UTR, 5' UTR, éxon) e estado de validação (pelo projeto 1000 genomas).

As sequências foram organizadas de modo a ter 14 nucleotídeos antes da variação alélica e 14 nucleotídeos após a variação. Dessa forma, junto à variação alélica, que caracteriza o SNP, foram obtidos 29 nucleotídeos totais.

Cada SNP foi editado com relação ao seu código IUPAC. Por exemplo, se o SNP envolve C/T, o código IUPAC para essa variação é Y, e assim foi feito aos demais SNPs.

Amostras públicas de tecido adiposo, adrenal, artéria, cérebro, cólon, coração, esôfago, estômago, fígado, mama, músculo, ovário, pâncreas, pele e sangue total, pituitária, placenta, próstata, pulmão, rim, testículo, tireóide, trompas e útero foram obtidas com auxílio da plataforma SRA, do NCBI e filtradas anteriormente por nosso grupo de pesquisa de modo que existisse apenas amostras únicas, ou seja, sem misturas, e saudáveis.

As sequências dos SNPs com IUPAC, alelo 1 e alelo 2 foram filtradas em cada amostra e os *reads*, as leituras de cada SNP em determinadas amostras, foram obtidos.

Posteriormente, foi realizado o cálculo da fração alélica dividindo o número de *reads* obtido pelo alelo 1 sobre o número de *reads* obtido pelo total (IUPAC), após a mesma aplicação foi realizada para o alelo 2.

Com base nos resultados da fração alélica, esses dados foram caracterizados como expressão bialélica, desvio do bialélico e expressão monoalélica.

Para expressão bialélica, a fração alélica de ambos alelos deve ser entre 0,35 e 0,65; para desvio do bialélico, a fração de um dos alelos deve ser entre 0,15 e 0,35 enquanto que o outro deve ser entre 0,65 e 0,85; e para caracterização de expressão monoalélica, a fração de um alelo deve ser entre 0,15 e 0 e o outro entre 0,85 e 1,00.

A expressão monoalélica seguida de FLIP, que é uma mudança da expressão alélica de um alelo para o outro alelo no campo amostral de experimentos, é indicativa de possível *imprinting* genômico desse gene.

6. RESULTADOS

6. 1 Seleção de Regiões Diferencialmente Metiladas (DMR)

A visualização dos dados de metilomas foi feita no *Genome Browser* da UCSC, entretanto com uma janela de visualização de 40Kb. Para garantir o sucesso do rastreamento/anotação manual, cada janela de visualização foi analisada independentemente por dois observadores, em colaboração com a pós-doutoranda Cláudia Caixeta Franco.

Utilizando dessas novas análises de rastreamento no genoma encontramos 123 regiões diferentemente metiladas em todo o genoma (Tabela 7). Algumas das regiões identificadas mapeiam em genes conhecidamente imprintados, mas nem todos os genes sabidamente imprintados possuem uma região diferentemente metilada ou a região diferentemente metilada exibe um perfil hemimetilado em todos os tecidos. As novas 70 DMRs identificadas constituem regiões preditas ao *imprinting* genômico e estão destacadas em cor cinza na tabela 7.

Foram identificadas 25 novas regiões diferentemente metiladas gaméticas (gDMR) e 61 regiões diferentemente metiladas secundárias (sDMR). Para análise subsequente foram selecionados *loci* por (i) ocorrência de pelo menos um polimorfismo de único nucleotídeo com frequência alélica $>0,15$, (ii) sobreposição com ilha CpG e (iii) não sobreposição com duplicações segmentárias ou repetições simples. Foram desenhados ensaios da PCR para validação experimental dos perfis de metilação e expressão gênica para 4 DMRs nos genes *PWWP2AP1*, *PARD6GAS1*, *GCSAML*, *PMF1*, candidatos ao *imprinting*.

Os perfis de hemimetilações encontradas, foram comparados em diferentes estudos de metiloma de genes conhecidamente imprintado como o *PEG10*. Portanto, para verificar a origem parental da metilação das preditas DMRs desenvolvemos ensaios de DESM-PCR (Digestão Enzimática Sensível a Metilação seguida da PCR) que permitem interrogar os estados de metilação em sítios CpG localizados em sequências de reconhecimento da enzima de restrição *HpaII*, susceptível a metilação. Foram feitas as padronizações dos demais ensaios da DESM-PCR. Os perfis de metilação foram determinados em amostras de DNA extraído de sangue periférico.

Tabela 7. Localizações cromossômicas de Regiões Diferentemente Metiladas.

Localização	DMR	CpG	Predição A ¹	Predição B ²	Gene
chr1:40,020,707-40,065,175	Primária	28	X		<i>PPIEL</i>
chr1:40,139,179-40,179,254	Secundária	62		X	<i>HPCAL4</i>
chr1:41,281,768-41,322,068	Secundária	76		X	<i>KCNQ4</i>
chr1:68,491,424-68,531,612	Primária	22	X		<i>GNG12-AS1</i>
chr1:147,682,144-147,722,428	Secundária	70		X	<i>NBPF19/ LOC101927468</i>
chr1:147,795,178-147,835,462	Secundária		X		<i>NBPF19</i>
chr1:148,846,473-148,886,630	Primária	223	X		<i>NBPF25P - DMR1</i>
chr1:149,004,001-149,044,207	Secundária		X		<i>NBPF25P - DMR2</i>
chr1:149,004,001-149,044,207	Secundária	53	X		<i>PMF1/ PMF1-BGLAP</i>
chr1:196,849,571-196,890,015	Secundária			X	<i>CFHR4</i>
chr1:240,636,958-240,676,993	Secundária	51	X		<i>GREM2/MIR1273E</i>
chr1:247,664,882-247,705,084	Secundária	74	X		<i>GCSAML</i>
chr2:25,353,591-25,393,666	Secundária	85		X	<i>PONC</i>
chr2:25,414,233-25,454,421	Secundária	57	X		
chr2:95,531,245-95,571,276	Primária	55	X		<i>LOC442028/TEKT4</i>
chr2:118,594,961-118,637,153	Secundária	155	X		
chr2:132,988,904-133,029,306	Secundária	194		X	<i>ANKRD3OBL</i>
chr2:207,103,606-207,144,068	Secundária	22	X		<i>GPR1AS</i>
chr3:570,338-610,413	Secundária	42		X	<i>AK126307</i>
chr3:42,956,725-42,996,744	Primária	24	X		<i>KRBOX1</i>
chr3:75,692,191-75,732,885	Secundária	265		X	<i>LINC00960</i>
chr3:130,442,066-130,482,125	Primária			X	<i>PIK3R4</i>
chr4:39,738-80,151	Secundária	67	X		<i>ZNF595/ZNF718</i>
chr4:6,087,718-6,128,206	Primária	68		X	<i>JAKMIP1</i>
chr4:74,844,746-74,885,080	Secundária	22		X	<i>PF4</i>
chr4:89,614,385-89,654,436	Primário	31	X		<i>HERC3/NAP1L5</i>
chr4:174,420,492-174,460,637	Secundária	34		X	
chr4:190,824,798-190,864,943	Secundária	91	X		<i>FRG1</i>
chr4:190,920,304-190,960,638	Secundária	621	X		
chr5:1,717,546-1,757,612	Secundária	23	X		
chr5:8,421,514-8,461,956	Secundária	88	X		<i>LOC729506/ LOC100505738</i>
chr5:37,193,591-37,233,751	Primária	28		X	<i>C5orf42</i>
chr5:135,385,673-135,426,229	Primária	24	X		<i>TGFBI (DOWNSTREAM)</i>
chr5:140,165,431-140,205,607	Secundária	120	X		<i>PCDHA1</i>
chr5:140,165,431-140,205,607	Secundária	125	X		<i>PCDHA2</i>
chr5:140,165,431-140,205,607	Secundária	121	X		<i>PCDHA3</i>
chr5:140,165,431-140,205,607	Secundária	111	X		<i>PCDHA4</i>
chr5:140,165,431-140,205,607	Secundária	118	X		<i>PCDHA5</i>
chr6:280,244-320,394	Secundária	120	X		<i>DUSP22</i>

Continuação da tabela 7.

Localização	DMR	CpG	Predição A ¹	Predição B ²	Gene
chr6:3,810,632-3,851,551	Primária	148	X		<i>FAM50B</i>
chr6:144,300,437-144,340,541	Primária	118	X		<i>PLAGL1/ HYMAI</i>
chr6:160409343-160449573	Primária	99	X		<i>IGF2R/ AIRN</i>
chr6:170,038,789-170,078,940	Primária	23	X		<i>WDR27</i>
chr7:22,852,576-22,892,663	Secundária	26		X	<i>TOMM7</i>
chr7:23515602-23555878	Primária		X		<i>RPS2P32/TRA2A-DOWNSTREAM</i>
chr7:32,334,202-32,374,480	Secundária		X		<i>PDE1C</i>
chr7:32,783,595-32,823,723	Secundária	40		X	<i>LINC00997</i>
chr7:42,876,538-42,916,742	Primária	18	X		<i>C7orf25</i>
chr7:50,433,908-50,474,564	Secundária	79	X		<i>IKZF1</i>
chr7:50,838,071-50,878,275	Primária	131	X		<i>GRB10</i>
chr7:55,505,661-55,546,204	Secundária	47	X		<i>VOPP1-DOWNSTREAM</i>
chr7:56,223,484-56,263,914	Secundária	97		X	
chr7:56,493,213-56,533,417	Secundária	38	X		<i>LOC650226</i>
chr7:56,536,996-56,577,314	Secundária	18		X	<i>LOC100240728</i>
chr7:57,254,933-57,295,137	Secundária	54		X	<i>LOC105375297-DOWNSTREAM</i>
chr7:57,913,331-57,953,632	Secundária	153		X	
chr7:64,384,806-64,424,959	Secundária			X	<i>ZNF273-DOWNSTREAM</i>
chr7:94,273,580-94,313,962	Primária	114	X		<i>SGCE/PEG10</i>
chr7:101,932,175-101,972,858	Secundária	98		X	
chr7:127,870,995-127,911,394	Secundária	67	X		<i>LEP</i>
chr7:130,109,211-130,149,687	Primária	184	X		<i>MEST/MESTIT1/MIR335/COPG2</i>
chr7:138,336,645-138,376,893	Primária	74	X		<i>SVOPL</i>
chr7:150,646,733-150,686,981	Primária	82		X	<i>KCNH2</i>
chr8:1,493,767-1,534,677	Primária	134		X	<i>DLGAP2</i>
chr8:95,116,378-95,156,802	Primária			X	<i>CDH 17-DOWNSTREAM</i>
chr8:141,078,342-141,118,766	Primária	210	X		<i>TRAPPC9</i>
chr9:68,378,203-68,418,975	Secundária	53	X		
chr9:79,180,283-79,220,655	Secundária	36	X		
chr9:98,065,201-98,105,229	Primária	28	X		<i>FANCC</i>
chr10:27,674,766-27,714,869	Primária	73	X		<i>PTCHD3</i>
chr10:121,557,354-121,598,161	Primária	84	X		<i>INPP5F/MCMBP</i>
chr11:1,999,505-2,039,980	Primária	65	X		<i>H19</i>
chr11:2,144,878-2,177,699	Secundária	28		X	<i>IGF2/INS-IGF2</i>
chr11:2,688,898-2,729,048	Primária	172	X		<i>KCNQ1/KCNQ10T1</i>
chr11:13,493,615-13,534,090	Secundária			x	<i>PTH-DOWNSTREAM</i>
chr11:49,057,839-49,097,838	Secundária	84		X	<i>TRIM64C-DOWNSTREAM</i>
chr11:49,841,328-49,881,554	Secundária	90		X	
chr11:49,896,275-49,936,576	Secundária	86	X		

Continuação da tabela 7.

Localização	DMR	CpG	Predição A ¹	Predição B ²	Gene
chr11:75,122,783-75,162,971	Primária	32	X		<i>KLHL35</i>
chr11:76,841,424-76,881,952	Primária	60		X	<i>MYO7A</i>
chr11:94,751,004-94,791,270	Secundária	85		X	<i>KDMHE-DOWNSTREAM</i>
chr12:64,192,683-64,232,787	Secundária	45		X	<i>TMEM5-DOWNSTREAM</i>
chr12:133,015,103-133,055,347	Secundária	35		X	
chr12:133,782,189-133,822,193	Secundária	98		X	<i>ANHX</i>
chr13:30,059,591-30,099,823	Primária	23		X	<i>MYUS2</i>
chr13:48,877,281-48,917,513	Primária	85		X	<i>RB1</i>
chr13:60,829,701-60,870,412	Primária		X		<i>LINC00434</i>
chr13:81,212,596-81,252,977	Primária	31	X		<i>PWWP2AP1</i>
chr14:101,268,588-101,308,660	Secundária	86	X		<i>MEG3</i>
chr15:23,879,034-23,919,149	Secundária			X	<i>MAGEL2</i>
chr15:23,903,561-23,943,831	Secundária	104	X		<i>NDN</i>
chr15:24,993,368-25,033,658	Secundária	64	X		
chr15:25,183,088-25,223,378	Primária	77	X		<i>SNRPN/SNURF</i>
chr15:99,395,302-99,435,647	Primária		X		<i>IGF1R</i>
chr16:3,460,803-3,502,393	Secundária	41	X		<i>ZNF597 / NAA60</i>
chr17:21,196,690-21,236,810	Secundária	18	X		<i>MAP2K3-DOWNSTREAM</i>
chr17:31,132,480-31,172,722	Secundária	79	X		<i>MYO1D</i>
chr17:33,474,836-33,516,364	Secundária	28		X	<i>UNC45B</i>
chr18:14,985,285-15,025,479	Secundária	305		X	
chr18:77,366,131-77,406,383	Primária	185		X	
chr18:77,893,133-77,934,131	Primária	28	X		<i>PARD6G-AS1</i>
chr19:54,038,922-54,079,345	Primária	100	X		<i>ZNF331</i>
chr19:54,038,922-54,079,345	Primária	83	X		<i>ZNF331</i>
chr19:57,341,094-57,381,188	Primária	83	X		<i>ZIM2/PEG3</i>
chr20:25,821,829-25,861,932	Secundária	49		X	<i>LOC101926935</i>
chr20:26,174,015-26,214,118	Secundária	218		X	<i>MIR663AHG</i>
chr20:30,119,419-30,159,522	Primária	25	X		<i>MCTS2P</i>
chr20:36,125,169-36,165,272	Primária	120		X	<i>BLCAP/NNAP</i>
chr20:42,116,537-42,156,875	Primária	35	X		<i>L3MBTL1</i>
chr20:57,394,633-57,434,737	Secundária	177	X		<i>GNAS</i>
chr20:57,394,633-57,434,737	Primária	190	X		<i>GNAS</i>
chr20:57,441,538-57,481,641	Secundária	499	X		<i>GNAS</i>
chr21:9,814,459-9,854,459	Secundária	335		X	
chr21:9,941,098-9,981,123	Secundária	30		X	<i>TEKT4P2</i>
chr21:9,797,239-9,837,364	Secundária	335		X	
chr21:15,338,433-15,378,615	Secundária	143		X	<i>ANKRD20A11P</i>
chr21:30,445,982-30,486,164	Secundária	41		X	<i>MAP3K7CL</i>

Continuação da tabela 7.

Localização	DMR	CpG	Predição A ¹	Predição B ²	Gene
chr21:30,522,328-30,562,510	Primária			X	MAP3K7CL
chr21:40,731,595-40,771,777	Secundária	27	X		WRB
chr21:48,063,854-48,103,886	Primária	89		X	PRMT2
chr22:16,984,599-17,023,684	Secundária	113		X	
chr22:42,053,487-42,093,590	Primária	42		X	NHP2L1

¹ Região predita ao *imprinting* nesse estudo por exibir padrão diferencialmente metilado comparável a aquele exibido nas DMRs conhecidas.

² Região predita ao *imprinting* nesse estudo por exibir padrão irregular de hemimetilação não comparável a aquele exibido nas DMRs conhecidas.

Após essa ampla análise, foi elaborado um “*track*” incluindo a informação da metilação em oócito e espermatozóide, que foi primordial para discriminar uma região diferentemente metilada primária de uma região diferentemente metilada secundária. Das 70 novas DMRs preditas, 20% são gaméticas.

Portanto, para encontrar DMR imprintadas associadas a genes imprintados, rastreamos o perfil dos genes conhecidamente imprintados, como por exemplo o gene PEG10, que é utilizado como controle positivo de *Imprinting* em várias linhas de pesquisas (Hannula-Jouppi, Muurinen *et al.*, 2014) ao observar sua DMR na plataforma *UCSC Genome Browser* tem metilação intermediária constitutiva nos metilomas disponíveis no *Track*. Ao buscar novos alvos procuramos um perfil comparável a este encontrado na maior parte dos genes conhecidamente imprintados.

De acordo com as conclusões obtidas no estudo de metilação, os genes imprintados que não tem hemimetilação no tecido, podem estar sendo controladas por Regiões Diferentemente Metiladas próximas, com posição física de no máximo 2,5 milhões de nucleotídeos de distância. Se o gene imprintado não tiver uma DMR a essa distância, seu controle de *Imprinting* é devido a outros fatores.

A visualização dos dados de metilomas feitas no *Genome Browser* da UCSC, utilizando os critérios maiores e menores permitiram através da análise dos perfis de metilação, escolher inicialmente 15 regiões hemimetiladas (Tabela 8). Essas regiões tinham todos os critérios maiores e pelo menos um dos critérios menores. Essas novas regiões hemimetiladas encontradas constituem regiões promissoras, pois

ainda não foram listadas na literatura. Dessas 7/15 (46,6%) são de DMR primária e 8/15 (53,3%) não possuem DMR na região promotora do gene. Não existe diferença significativa entre o tipo de DMR primária e secundária e a prevalência de retrotransposon ($p=1,000$, Teste Fisher) nem com a prevalência de DMR na região promotora ($p=1,000$, Teste Fisher). Dessas 15 regiões, 4 (*PWWP2AP1*, *PARD6G-AS1*, *GCSAML* e *PMF1*) foram escolhidas para serem testadas experimentalmente.

Tabela 8. Regiões hemimetiladas escolhidas a partir dos critérios maiores e menores.

Gene	Local ^a	DMR	Retrotransposon	DMR na região promotora ^b	Hemimetilação no blastocisto	Número de CG na Ilha CpG
<i>PMF1</i>	1	Secundária	Ausente	Ausente	Hipometilado	47
<i>GCSAML</i>	1	Secundária	Ausente	Presente	Hipometilado	65
<i>C5orf42</i>	5	Primária	Presente	Ausente	Hemimetilado	28
<i>LOC100505738</i>	5	Secundária	Ausente	Presente	Hipometilado	88
<i>VTRNA2-1</i>	5	Primária	Ausente	Presente	Hemimetilado	24
<i>PCDHA3</i>	5	Secundária	Ausente	Presente	Hipometilado	121
<i>TMEM14C</i>	6	Secundária	Ausente	Ausente	Hipometilado	64
<i>DLGAP2</i>	8	Primária	Ausente	Ausente	Hemimetilado	118
<i>FANCC</i>	9	Primária	Ausente	Ausente	Hemimetilado	26
<i>ANHX</i>	12	Secundária	Presente	Presente	Hipometilado	79
<i>LINC00434</i>	13	Primária	Ausente	Presente	Hemimetilado	
<i>PWWP2AP1</i>	13	Primária	Ausente	Presente	Hemimetilado	31
<i>MYO1D</i>	17	Secundária	Ausente	Ausente	Hemimetilado	53
<i>PARD6G-AS1</i>	18	Primária	Ausente	Presente	Hemimetilado	28
<i>PRMT2</i>	21	Primária	Ausente	Presente	Hemimetilado	89

^a Cromossomo em que o gene está localizado.

^b Presença ou ausência de DMR na região promotora.

6.2 Região *PARD6G-AS1*

PARD6G-AS1 é um gene não codificante, em que cinco transcritos são sintetizados por esse gene com 3333, 1058, 647, 584 e 525 nucleotídeos. A DMR da região *PARD6G-AS1* é primária e possui uma ilha CpG com 28 resíduos “CG”. A DMR está localizada na região promotora do gene *PARD6G-AS1* (Figura 5), localizado na região 18q12 do cromossomo 18. O tecido que possui maior expressão para esse gene de acordo com o GTEx é o testículo (Figura 6).

Não há ainda funcionalidade descrita para esse gene, entretanto um artigo é encontrado no banco de dados “PubMed” utilizando a palavra chave “*PARD6G-AS1*”. Curiosamente esse artigo descreve uma associação epigenética entre cg06092953 (chr18; 77,905,699, gene mais perto = *PARD6G-AS1*) e agressividade ($p = 9.0 \times 10^{-7}$, FDR = 0.18) em um estudo de coorte com gêmeos holandeses (Van Dongen, Nivard *et al.*, 2015).

Além deste estudo, outro artigo recentemente divulgado foi uma ampla análise do genoma de metilação do DNA em pacientes com Pseudohipoparatiroidismo (PHP) onde verificou-se metilação anormal para algumas das DMR em humanos, e isso incluiu hipometilação significativa para *PARD6G-AS1*. Essa análise foi feita com Illumina Infinium Human Methylation 450K Bead-Chip (HM450k) que foi usado para determinar o genoma em toda a largura de perfis de metilação do DNA em pacientes com (PHP) e controles saudáveis compatíveis com idade e gênero (Roctus, Martin-Trujillo *et al.*, 2016).

Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA methylation, um gráfico de calor foi produzido para DMR da região *PARD6G-AS1* (figura 7). Todas as regiões apresentam a mesma configuração: extremidades da DMR estão hipermetiladas enquanto a região central está hemimetilada. Apenas a região central da DMR do gene *PMF1* se apresenta hipometilada para as células mononucleares de sangue periférico. Quanto à metilação, o estado de metilação do DNA na região *PARD6G-AS1* foi determinado para 19 amostras de DNA de voluntários através do ensaio DESM-QF-PCR (Figura 8). A média de metilação foi baixa, com $24,88\% \pm 21,15$ para as 19 amostras controles. A região de desenho dos iniciadores compreendia um polimorfismo do tipo In/Del (-/CTGTGGTGC) de 9 pares de base (SNP rs11281142). A taxa de heterozigose observada para as amostras controles foi de 0,554 enquanto que a heterozigose esperada foi de 0,480.

De acordo com a análise DESM-QF-PCR em três trios informativos, a origem parental da metilação do DNA para essa região é materna (Figura 9). Ou seja, as evidências através da análise epigenética de metilação indicam que o *imprinting* é materno. Para a confirmação desse achado foi realizado uma investigação da expressão transcricional dos genes alvos por mineração de dados de RNA-Seq. Concomitantemente foi feita a identificação computacional de 21 SNPs selecionados a partir de uma análise em *megaregions* a partir dos genes alvo com pelo menos 80 *reads*. A validação da predição foi feita para alguns genes para tipagem de DNA e cDNA.



Figura 5. Assimetria gamética da metilação na região promotora do gene *PARD6G-AS1*. De cima para baixo, ideograma representado o cromossomo 18, neste a região marcada em vermelho (linha vertical) em que abaixo confere essa região. Abaixo, as isoformas do gene *PARD6G-AS1*, CTCF, POL2, ilhas CpG e metilação nos gametas, tecidos somáticos e fetais.

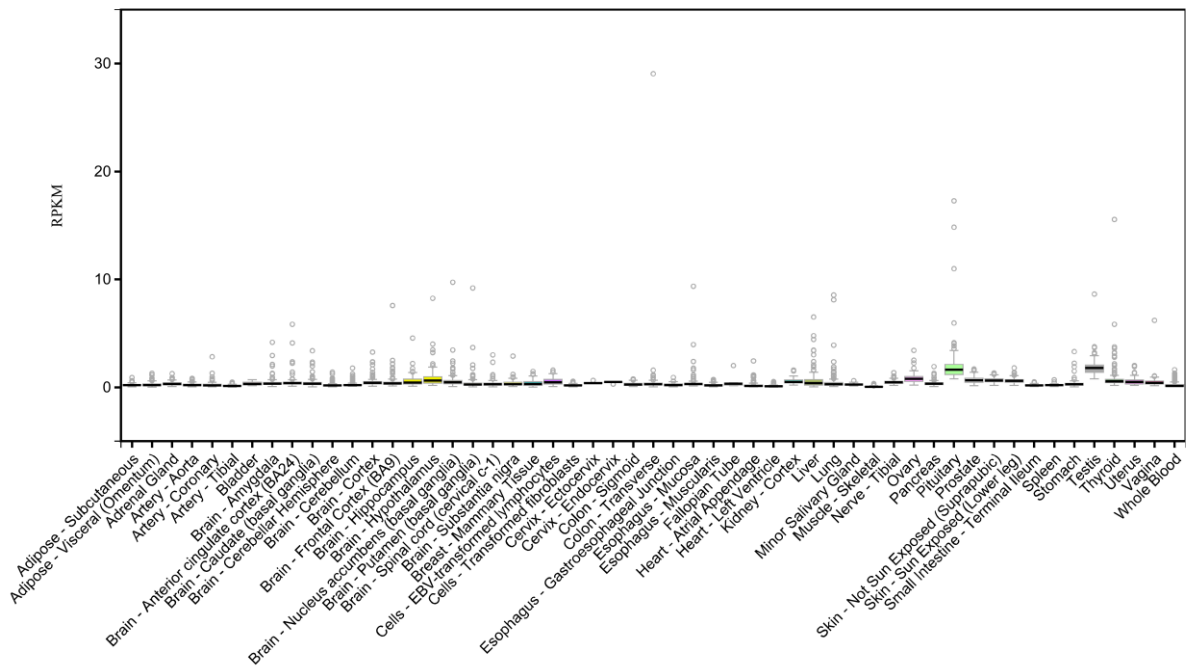


Figura 6. Perfil de expressão do gene *PARD6G-AS1* em 53 tecidos disponível no GTEX. O gene *PARD6G-AS1* possui baixa expressão em todos os tecidos, os tecidos com maior expressão são o testículo e pituitária.

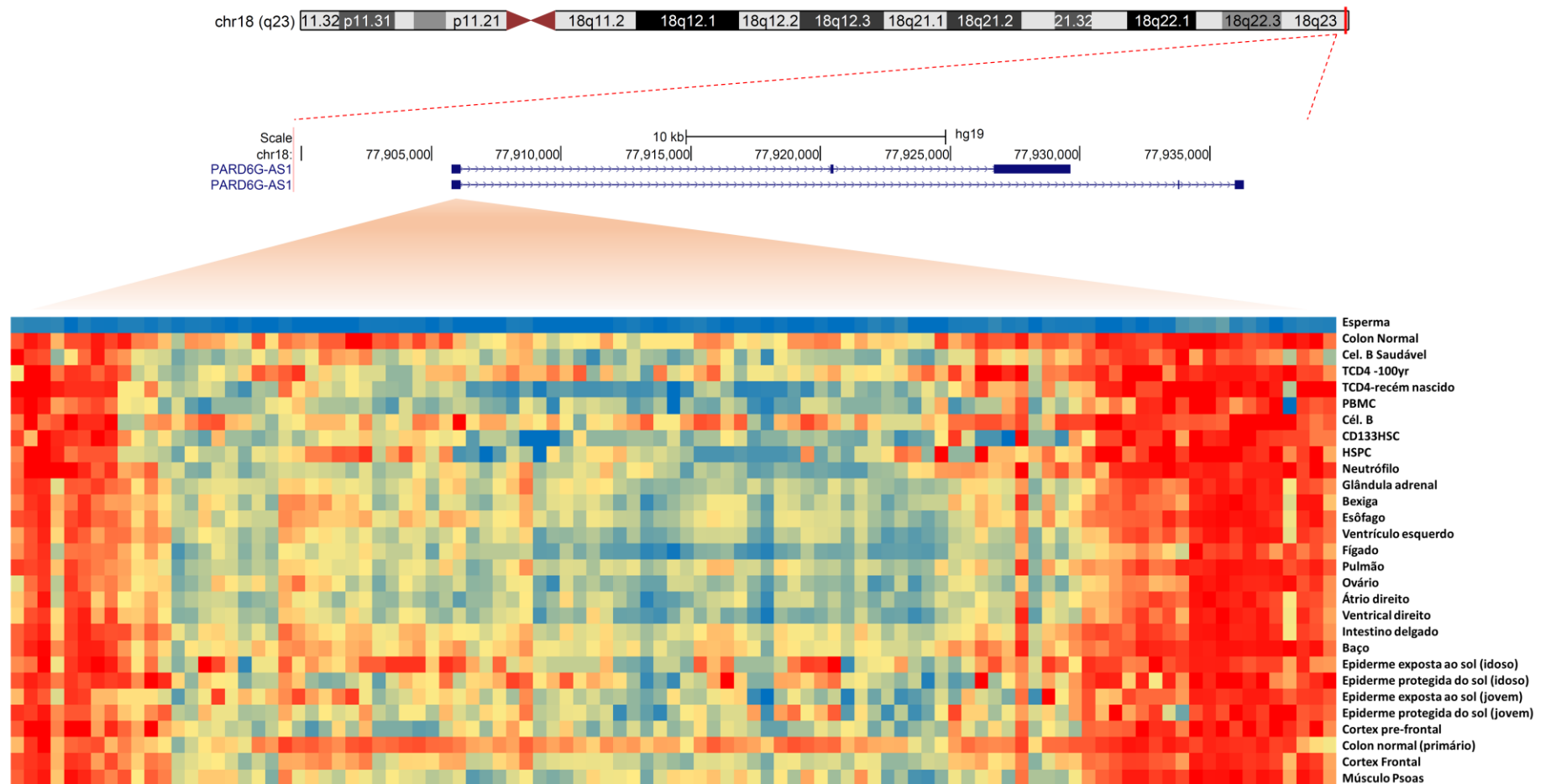


Figura 7. Caracterização da DMR da região *PARD6G-AS1* com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA methylation. A posição física referente a essa região de hemimetilação, em amarelo, é chr18:77904635-77907548.

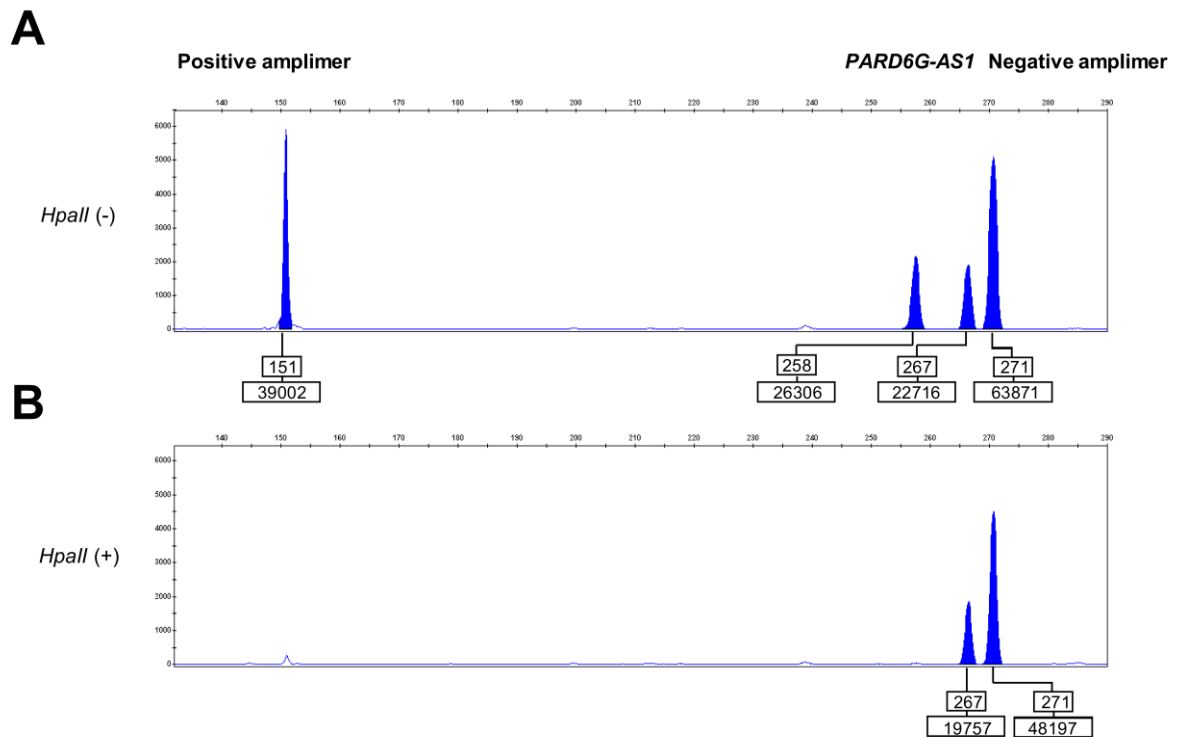


Figura 8. Eletroferograma representativo da identificação do locus *PARD6G-AS1* com perfil de metilação alélica específica. À esquerda com 151pb temos o controle positivo que depois da digestão com a enzima *HpaII* desaparece por ser uma região desmetilada. Ao meio os dois alelos 258 (rs11281142⁻) e 267 (rs11281142^{*}CTGTGGTGC) que depois da digestão, apenas o alelo 267 se apresentou e do lado direito com 271pb o controle negativo sem sítio de digestão, permanecendo sem modificação depois da digestão.

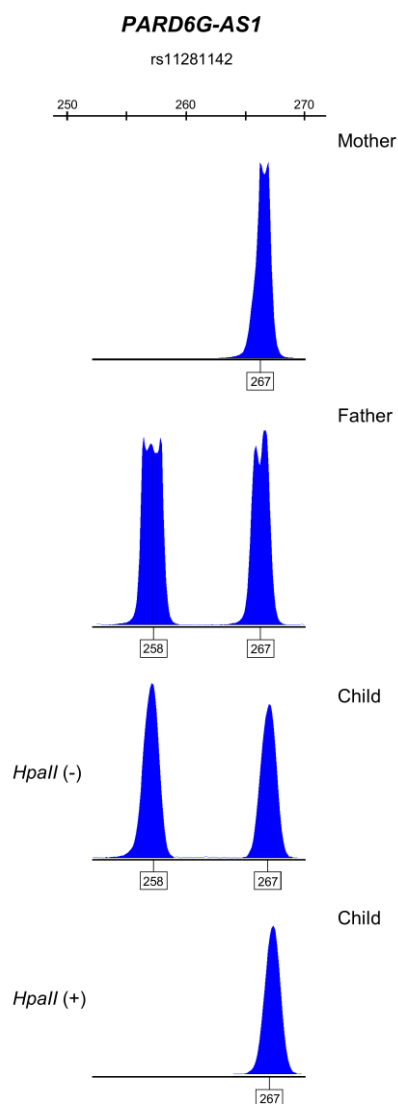


Figura 9. A assimetria gamética na gDMR do gene *PARD6G-AS1* é de origem materna. Eletroferograma representativo do perfil de metilação pontual na gDMR alvo em amostras de DNA da família 1634 (mãe, pai, filho) para um loco In/Del ligado ao sítio CpG alvo da interrogação do estado de metilação dos alelos parentais. A mãe (M) é homocigota para o alelo 265bp, enquanto que o pai (P) e o filho (F) são heterocigotos para o In/Del (DNA não digerido. ND; alelos 256pb e 265pb). No filho, o alelo paterno obrigatório (256bp) é susceptível à digestão com a enzima *HpaII*, enquanto o alelo materno, refratário à digestão, é metilado. Os demais produtos da PCR correspondem aos controles positivo e negativo da digestão.

A média da metilação da DMR do gene *PARD6G-AS1* determinada pela DESM-QF-PCR em 16 amostras de sangue foi de 31,50%. O desvio padrão encontrado variava de acordo com a amostra (**Figura 10**).

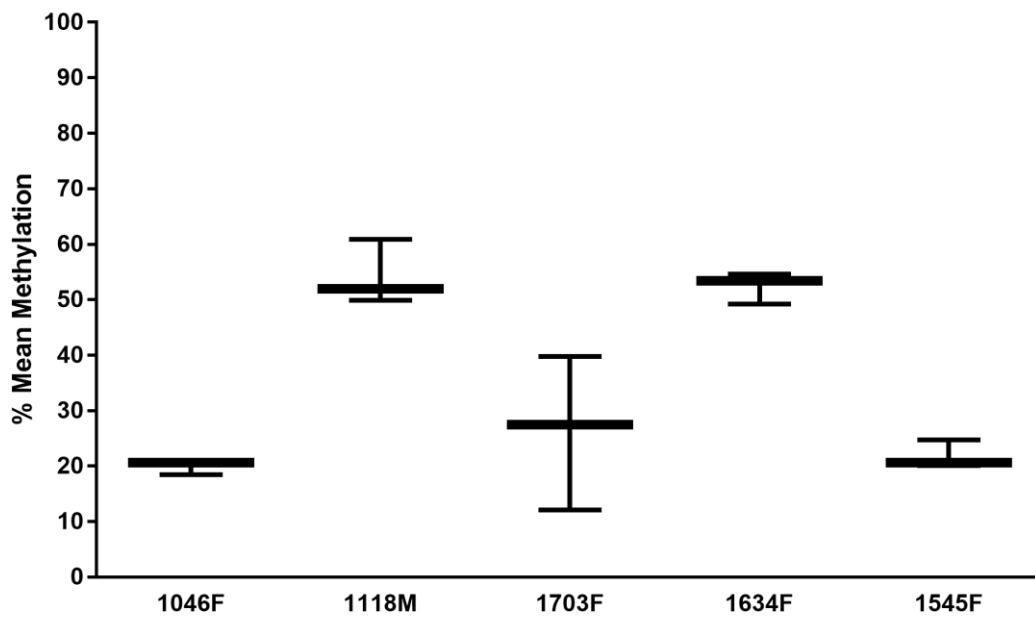


Figura 10. Variação na metilação e entre triplicatas em amostras de indivíduos controles utilizando DESM-QF-PCR. A metilação do gene *PARD6G-AS1* foi determinada por DESM-QF-PCR nas amostras dos doadores 1046F, 1118M, 1703F, 1634F e 1545F, a amostra 1703F apresentou uma maior variação da metilação aferida na triplicata.

6.3 Região *GCSAML*

A DMR da região *GCSAML* é secundária e possui uma ilha CpG com 65 resíduos “CG”. A DMR está localizada na região promotora do gene *GCSAML* (Figura 11), localizado na região 1q44 do cromossomo 1. *GCSAML* é um gene não codificante, em que onze transcritos são sintetizados por esse gene com 4130, 4063, 4030, 3959, 3820, 1822, 1308, 1094, 562, 522 e 558 nucleotídeos. Sendo que apenas o último transcrito não codifica proteína. A proteína codificada por esse gene, com 135 e/ou 115 aminoácidos é transcrita em células presentes nos centros germinativos, e parece ter importância na maturação e diferenciação dos linfócitos B. Um artigo é encontrado no banco de dados “PubMed” utilizando a palavra chave “*GCSAML*”. A expressão de transcrito e da proteína *GCSAML* é aumentada após auto enxerto em uma espécie de esponja marinha (Blumbach, Diehl-Seifert *et al.* , 1999).

Os tecidos que possuem maior expressão quantitativa para esse gene de acordo com o GTEX são cérebro, próstata e testículo (Figura 12). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA methylation, um gráfico de calor foi produzido para DMR da região *GCSAML* (Figura 13). Todas as regiões apresentam a mesma configuração: extremidades da DMR estão hipermetiladas enquanto a região central está hemimetilada. Quanto à metilação, o estado de metilação do DNA na região *GCSAML* foi determinada para 19 amostras de DNA de voluntários através do ensaio DESM-QF-PCR (Figura 14). A média de metilação foi 24,88% +/- 21,15 para as amostras controles. A região de desenho dos iniciadores compreendia um SNP (rs6700954). A taxa de heterozigose observada foi 0,554 enquanto que a heterozigose esperada foi 0,480.

De acordo com a análise DESM-QF-PCR em três trios informativos, a origem parental da metilação do DNA para essa região é materna (Figura 15). Ou seja, as evidências através da análise epigenética de metilação indicam que o *imprinting* é materno. Para a confirmação desse achado foi realizado uma investigação da expressão transcricional dos genes alvos por mineração de dados de RNA-Seq. Concomitantemente foi feita a identificação computacional de 21 SNPs selecionados a partir de uma análise em *megaregions* a partir dos genes alvo com pelo menos 80 *reads*. A validação da predição foi feita para alguns genes para tipagem de DNA e cDNA.

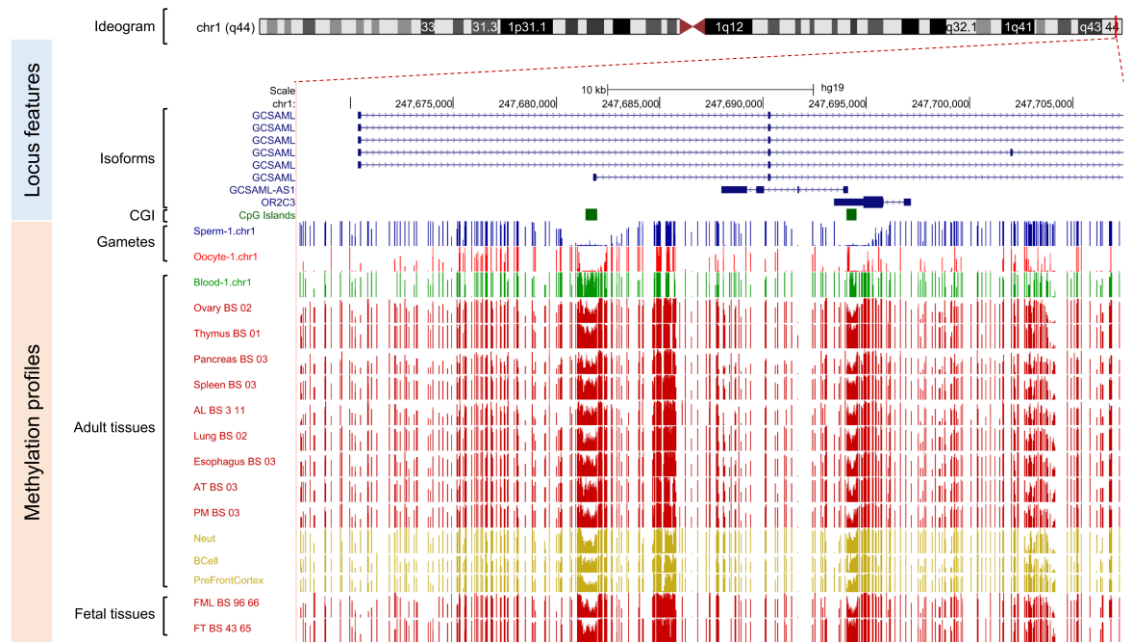


Figura 11. Assimetria gamética da metilação na região promotora do gene *GCSAML*. De cima para baixo, ideograma representado o cromossomo 1, neste a região marcada em vermelho (linha vertical) em que abaixo confere essa região. Abaixo, as isoformas do gene *GCSAML*, CTCF, POL2, ilhas CpG e metilação nos gametas, tecidos somáticos e fetais.

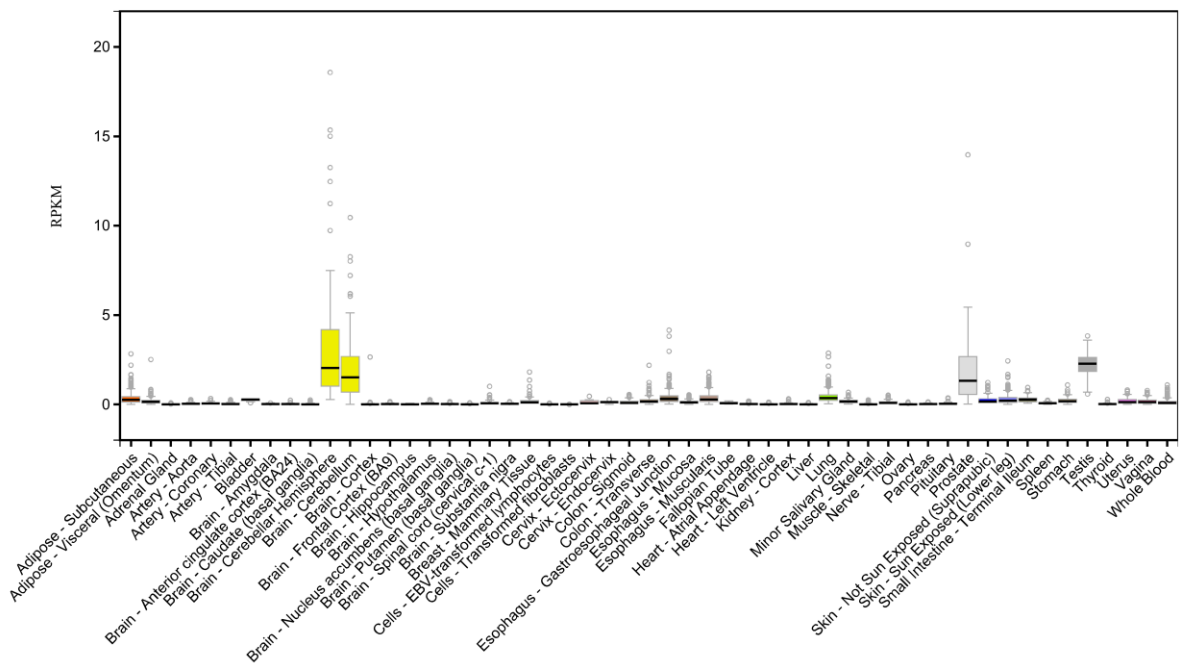


Figura 12. Perfil de expressão do gene *GCSAML* em 53 tecidos disponível no GTEx. O gene *GCSAML* possui baixa expressão em grande parte dos tecidos, entretanto é possível observar que o cérebro, próstata e testículo são os tecidos com maior expressão desse gene.

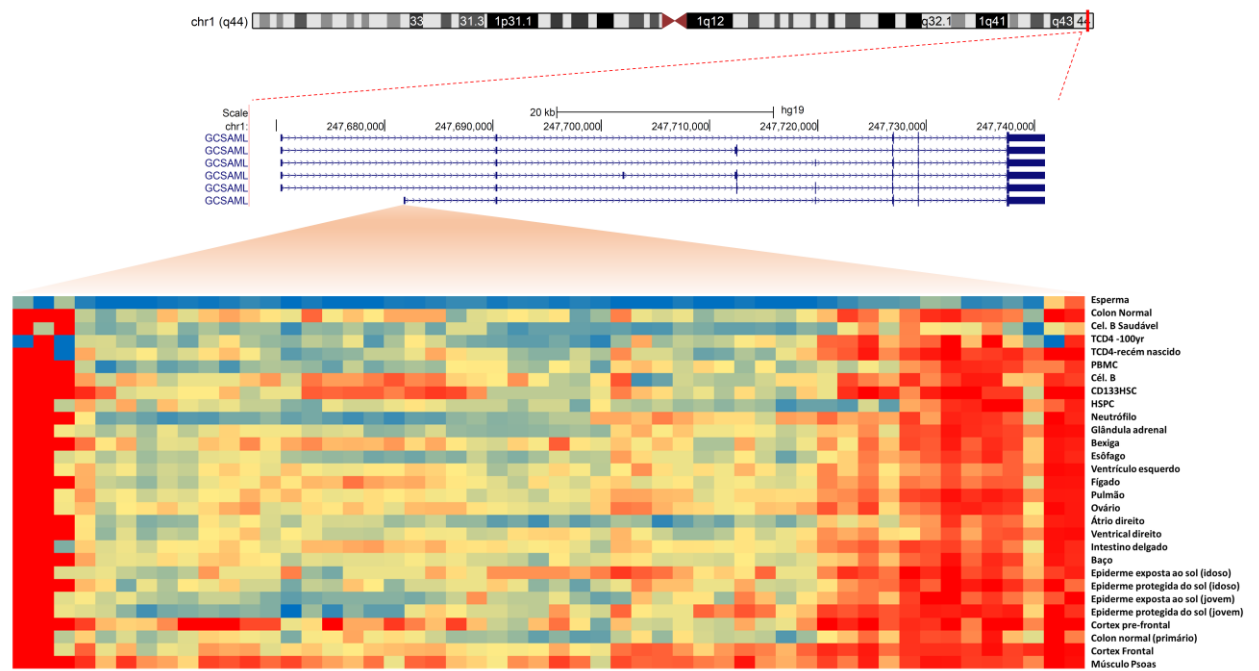


Figura 13. Caracterização da DMR da região *GCSAML* com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA methylation. A posição física referente a essa região de hemimetilação, em amarelo, é chr1:247680935-247682391.

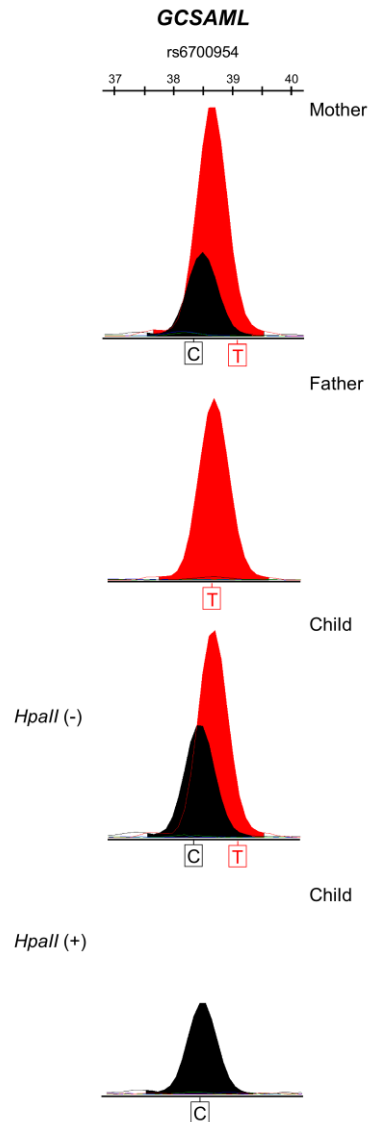


Figura 15. A assimetria gamética na gDMR do gene *GCSAML* é de origem materna. Eletroferograma representativo do perfil de metilação pontual na gDMR alvo em amostras de DNA de da família 1655 (mãe, pai, filho) informativo o SNP rs6700954 ligado ao sítio CpG alvo da interrogação do estado de metilação dos alelos parentais. O pai (P) é homocigoto rs6700954*T, enquanto que a mãe (M) e o filho (F) são heterocigotos rs6700954*T/C. No filho, o alelo paterno obrigatório rs6700954*T é suscetível à digestão com a enzima *HpaII*, enquanto o alelo materno, refratário à digestão, é metilado.

A média da metilação da DMR do gene *GCSAML* determinada pela DESM-QF-PCR em 8 amostras de sangue foi de 50,30%. O desvio padrão encontrado variava de acordo com a amostra (**Figura 16**).

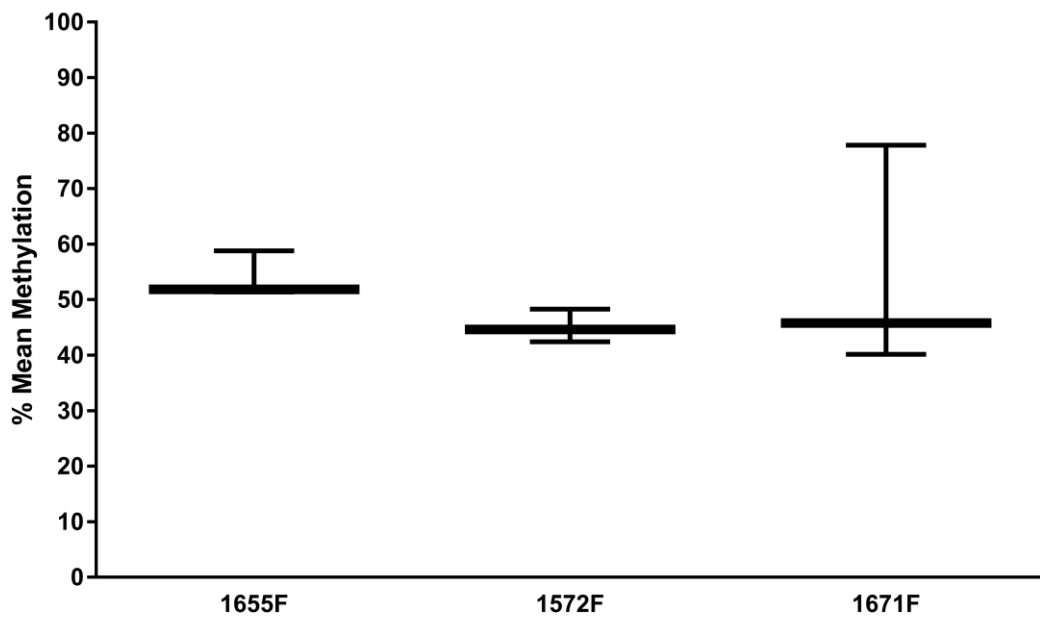


Figura 16. Variação na metilação e entre triplicatas em amostras de indivíduos controles utilizando DESM-QF-PCR. A metilação do gene *GCSAML* foi determinada por DESM-QF-PCR nas amostras dos doadores 1655F, 1572F e 1671F, a amostra 1671F apresentou uma maior variação da metilação aferida na triplicata.

6.4 Região *PWWP2AP1*

PWWP2AP1 é um pseudogene não codificante, em que apenas um transcrito é sintetizado por esse gene com 2273 nucleotídeos. Não há ainda funcionalidade descrita para esse gene, e nenhum artigo é encontrado no banco de dados “PubMed” utilizando a palavra chave “*PWWP2AP1*”. Além disso, nenhum SNP é citado no EXAC, revelando que a cobertura desse gene é baixa nos testes de Exomas. A DMR da região *PWWP2AP1* é primária e possui uma ilha CpG com 31 resíduos “CG”. A DMR está localizada na região promotora do pseudogene *PWWP2AP1* (Figura 17), localizado na região 13q31. 1 do cromossomo 13.

Os tecidos que possuem maior expressão quantitativamente falando para esse gene de acordo com o GTEX é o testículo (Figura 18). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA Methylation, um gráfico de calor foi produzido para DMR da região *PWWP2AP1* (Figura 19). Todas as regiões apresentam a mesma configuração: extremidades da DMR estão hipermetiladas enquanto a região central está hemimetilada.

Quanto à metilação, o estado de metilação do DNA na região *PWWP2AP1* foi determinado para 17 amostras de DNA de voluntários através do ensaio DESM-QF-PCR (Figura 20). A média de metilação foi 24,88% +/- 21,15 para as 17 amostras controles. O SNP rs1176323, possui uma taxa de heterozigose observada nas amostras controles de 0,563 e esperada de 0,482.

De acordo com a análise DESM-QF-PCR em três trios informativos para o SNP rs1176323, a origem parental da metilação do DNA para essa região é materna (Figura 21). Entretanto um resquício do alelo paterno é observado. A proporção entre o alelo C/T antes e após a digestão com *HpaII* cai de 20% para 11,1%. As análises utilizando DESM-QF-PCR indicam que a metilação não é parental específica, pois ambos os alelos permanecem metilados, apesar da metilação paterna ser menor que a materna. Para a confirmação desse achado foi realizado uma investigação da expressão transcricional dos genes alvos por mineração de dados de RNA-Seq. Concomitantemente foi feita a identificação computacional de 21 SNPs selecionados a partir de uma análise em *megaregions* a partir do gene *PWWP2AP1* com pelo menos 80 reads.

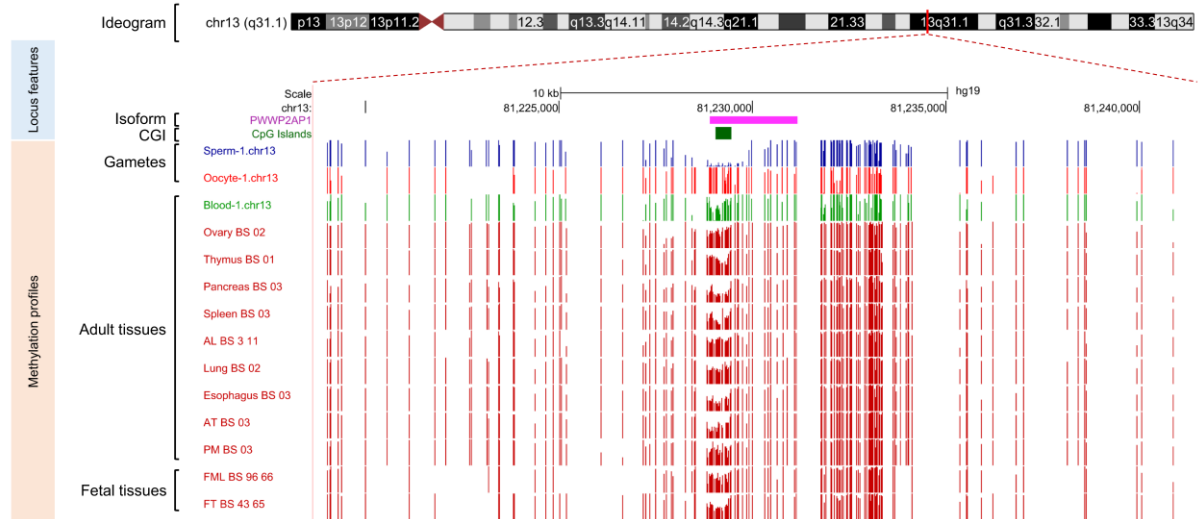


Figura 17. Assimetria gamética da metilação na região promotora do gene *PWWP2AP1*. De cima para baixo, ideograma representado o cromossomo 13, neste a região marcada em vermelho (linha vertical) em que abaixo confere essa região. Abaixo, a isoforma do gene *PWWP2AP1*, ilhas CpG e metilação nos gametas, tecidos somáticos e fetais.

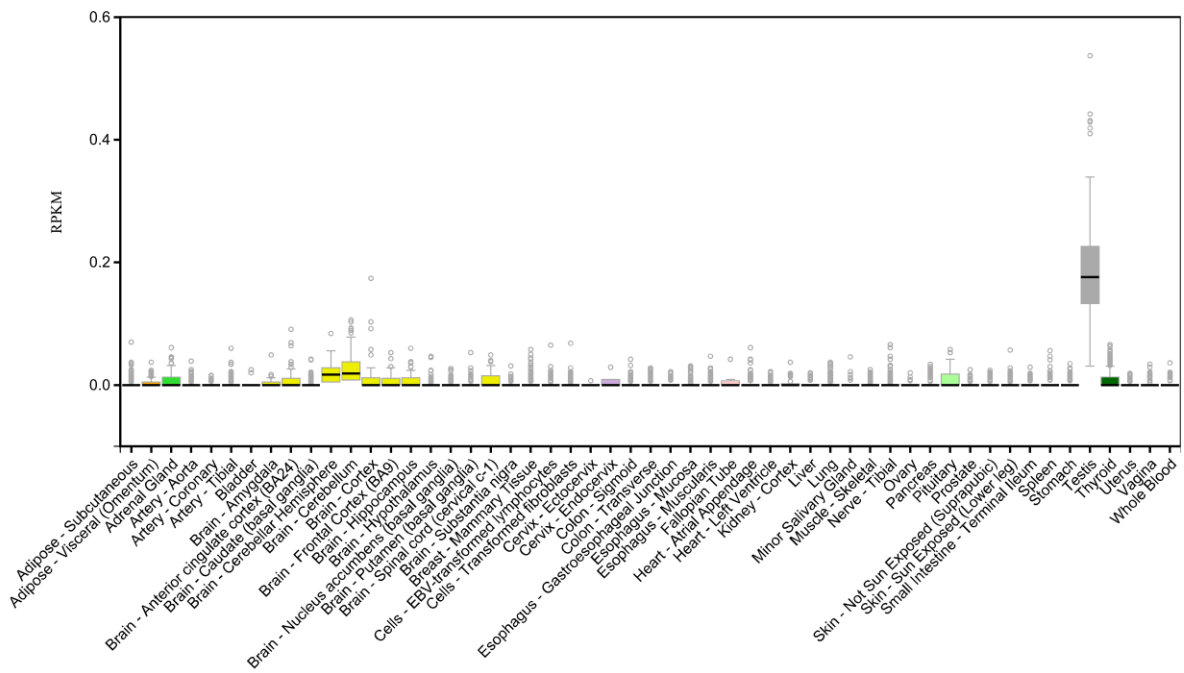


Figura 18. Perfil de expressão do gene *PWWP2AP1* em 53 tecidos disponíveis no GTEx. O gene *PWWP2AP1* possui baixa expressão em todos os tecidos, entretanto é possível observar que o testículo é o tecido com maior expressão desse gene.

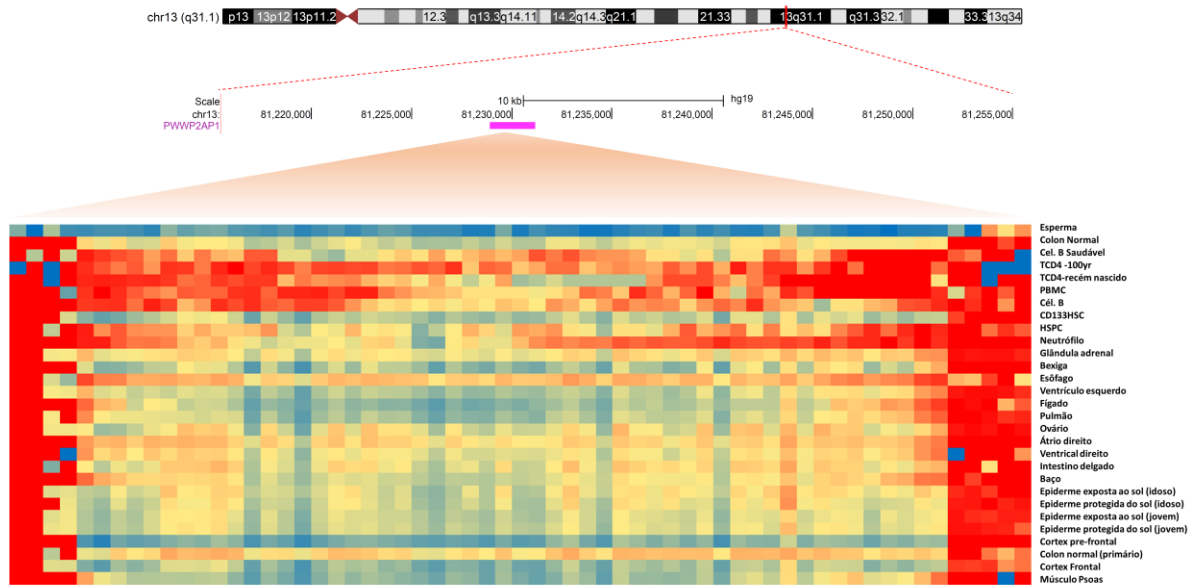


Figura 19. Caracterização da DMR da região *PWWP2AP1* com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA methylation. A posição física referente a essa região de hemimetilação, em amarelo, é chr13:81229156-81229416.

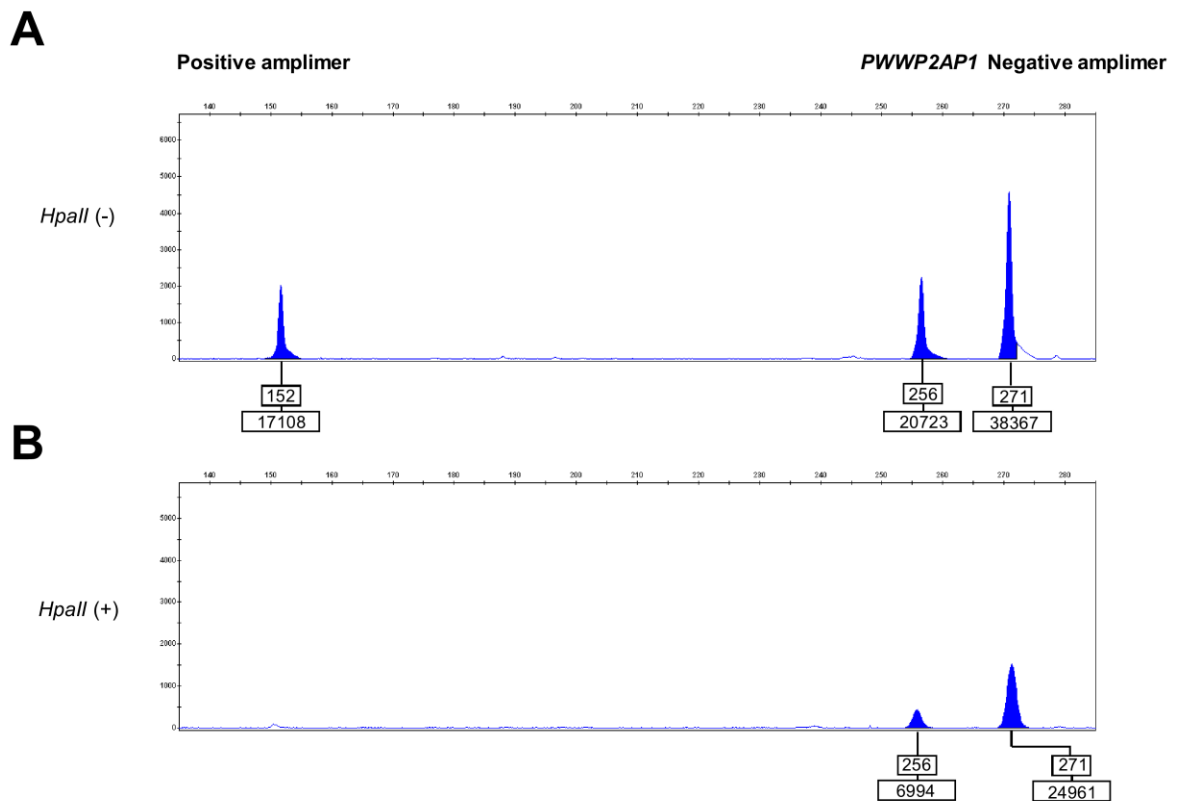


Figura 20. Eletroferograma representativo da identificação do locus *PWWP2AP1* com perfil de hemimetilação alélica. À esquerda com 152pb temos o controle positivo que depois da digestão com a enzima *HpaII* desaparece por ser uma região desmetilada. Ao meio com 256pb temos o alvo que depois da digestão, apresentou metade da área (perfil hemimetilado) e do lado direito com 271pb o controle negativo sem sítio de digestão, permanecendo sem modificação depois da digestão.

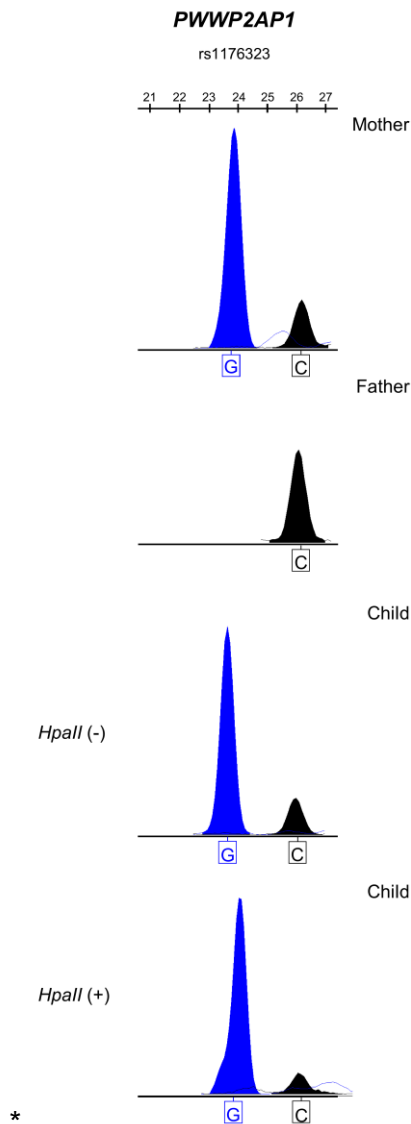


Figura 21. Assimetria gamética enviesada na gDMR do gene *PWWP2AP1*. Eletroferograma representativo do perfil de metilação pontual na gDMR alvo na família 1572 (mãe, pai, filho) informativo para um SNP rs1176323 ligado ao sítio CpG alvo da interrogação do estado de metilação dos alelos parentais. O pai (P) é homocigoto para o alelo C, enquanto que a mãe (M) e o filho (F) são heterocigotos rs1176323*C/T. No filho, o alelo materno é mais refratário à digestão por *HpaII*, enquanto que o alelo paterno obrigatório rs1176323*C é mais susceptível à digestão com a enzima *HpaII*. Entretanto, o filho ainda apresenta o alelo paterno rs1176323*C apresentando perfil enviesado de metilação parental.

A média da metilação da DMR do gene *PWWP2AP1* determinada pela DESM-QF-PCR em 7 amostras de sangue foi de 59,85%. O desvio padrão encontrado variava de acordo com a amostra (**Figura 22**).

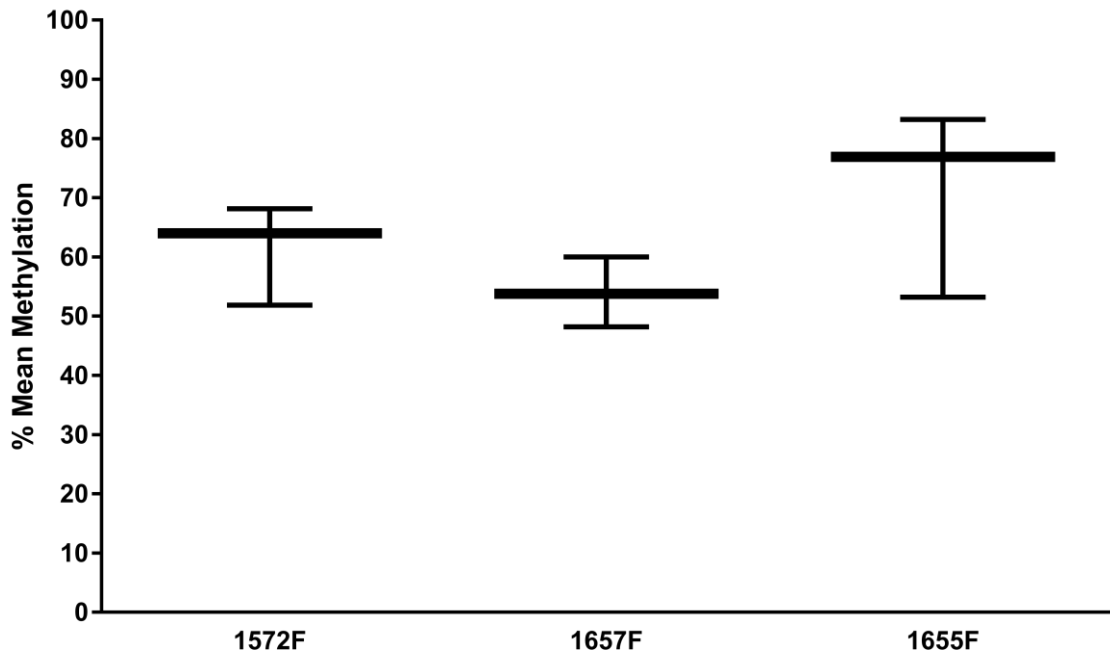


Figura 22. Variação na metilação e entre triplicatas em amostras de indivíduos controles utilizando DESM-QF-PCR. A metilação do gene *PWWP2AP1* foi determinada por DESM-QF-PCR nas amostras dos doadores 1572F, 1657F e 1655F, a amostra 1655F apresentou uma maior variação da metilação aferida na triplicata.

6.5 Região *PMF1*

As poliaminas são importantes reguladores do crescimento celular e da apoptose. O factor-1 modulado com poliamina (PMF-1) está envolvido na homeostase de poliamina. A DMR da região *PMF1* é secundária e possui uma ilha CpG com 47 resíduos “CG”. A DMR está localizada na região promotora do gene *PMF1* (*Fator modulador da Poliamina 1*; Figura 23), localizado na região 1q22 do cromossomo 1. Após a identificação de uma ilha CpG enriquecida que abrange o promotor *PMF1*, um estudo descreveu que a metilação de *PMF1* em amostras urinárias distinguiu pacientes com câncer de bexiga dos controles.

PMF1 é um gene com seis transcritos sintetizados por esse gene com 1068, 1019, 884, 933, 837 e 1275. Sendo que os três últimos transcritos não codificam proteína. A proteína codificada por esse gene, com 205 e/ou 207 aminoácidos pode estar envolvida na anáfase e metáfase mitótica. A busca no banco de dados “PubMed” utilizando a palavra chave “PMF1”, resultou em 36 artigos. Entre eles, estudos de associação com o AVC hemorrágico (Carpenter, Singh *et al.* , 2016; Chauhan e Debette, 2016) e com aneuploidias em células hematopoiéticas em homens (Wright, Day *et al.* , 2017).

O tecido que possui maior expressão quantitativamente falando para esse gene de acordo com o GTEX é a tireóide, sendo que os demais também estão sendo expressos (Figura 24). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA Methylation, um gráfico de calor foi produzido para DMR da região *PMF1* (figura 25). Todas as regiões apresentam a mesma configuração: extremidades da DMR estão hipermetiladas enquanto a região central está hemimetilada.

Quanto à metilação, o estado de metilação do DNA na região *PMF1* foi determinado para 19 amostras de DNA de voluntários através do ensaio DESM-QF-PCR (Figura 26). A média de metilação foi 24,88% +/- 21,15 para as amostras controles. A região de desenho dos iniciadores compreendia um SNP (rs4998248). A taxa de heterozigose observada foi 0,500 enquanto que a heterozigose esperada foi 0,420.

De acordo com a análise DESM-QF-PCR, a região é hipermetilada (Figura 26), e com a análise de cinco trios não foi possível observar metilação do DNA origem parental específica (Figura 27). Ou seja, as evidências através da análise

epigenética de metilação indicam que esse gene não é imprintado. Para a confirmação desse achado foi realizado uma investigação da expressão transcricional dos genes alvos por mineração de dados de RNA-Seq. Concomitantemente foi feita a identificação computacional de 21 SNPs selecionados a partir de uma análise em *megaregions* a partir do gene *PMF1* com pelo menos 80 *reads*.

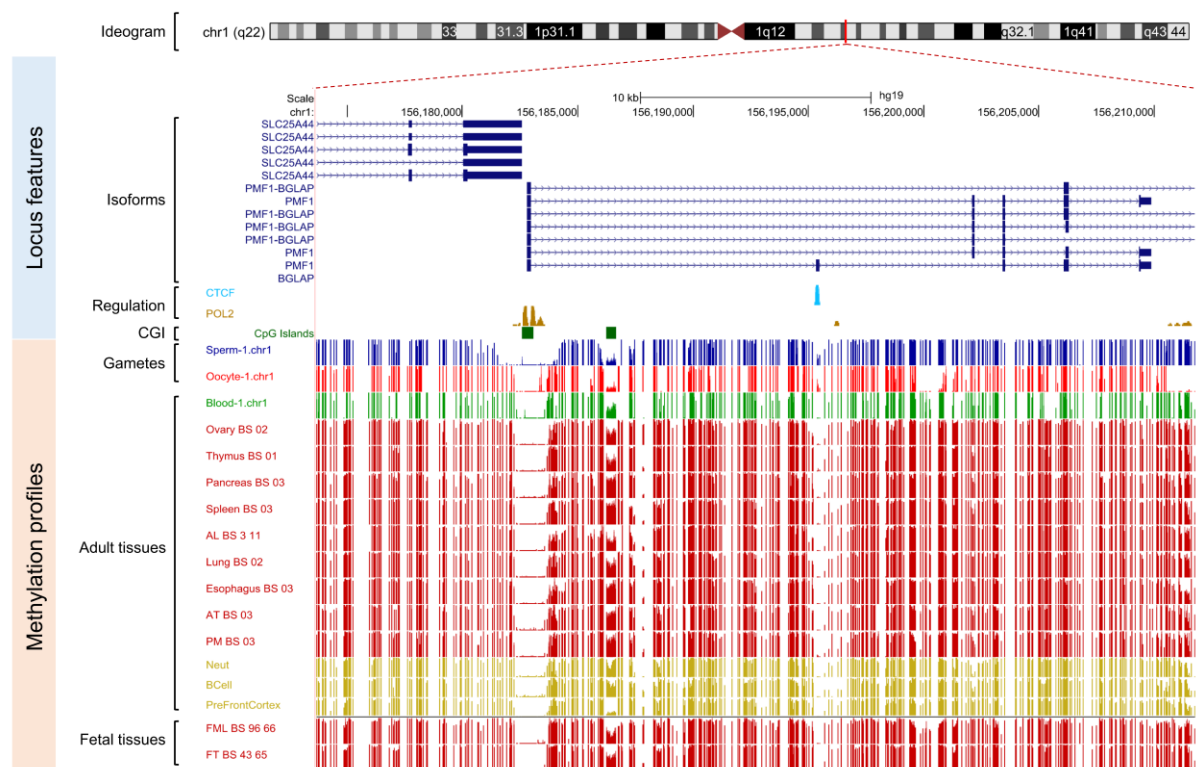


Figura 23. Assimetria gamética da metilação na região promotora do gene *PMF1*. De cima para baixo, ideograma representado o cromossomo 1, neste a região marcada em vermelho (linha vertical) em que abaixo confere essa região. Abaixo, as isoformas do gene *PMF1*, ilhas CpG e metilação nos gametas, tecidos somáticos e fetais.

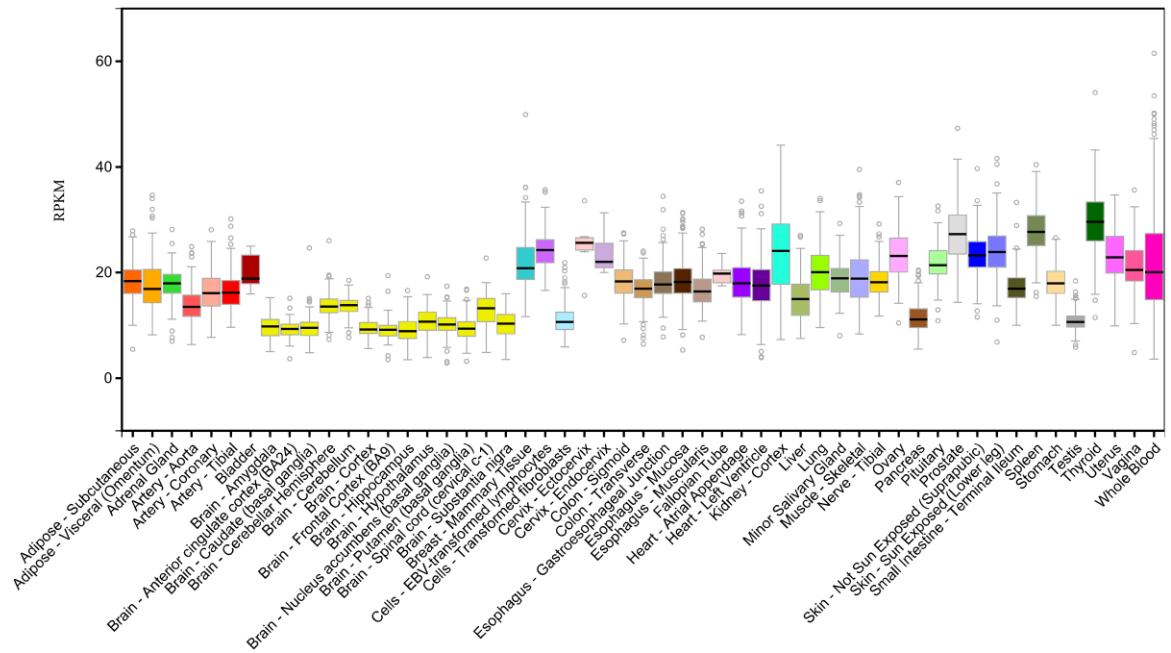


Figura 24. Perfil de expressão do gene *PMF1* em 53 tecidos disponíveis no GTEX. O gene *PMF1* possui alta expressão em grande parte dos tecidos, entretanto é possível observar que a tireoide é o tecido com maior expressão desse gene.

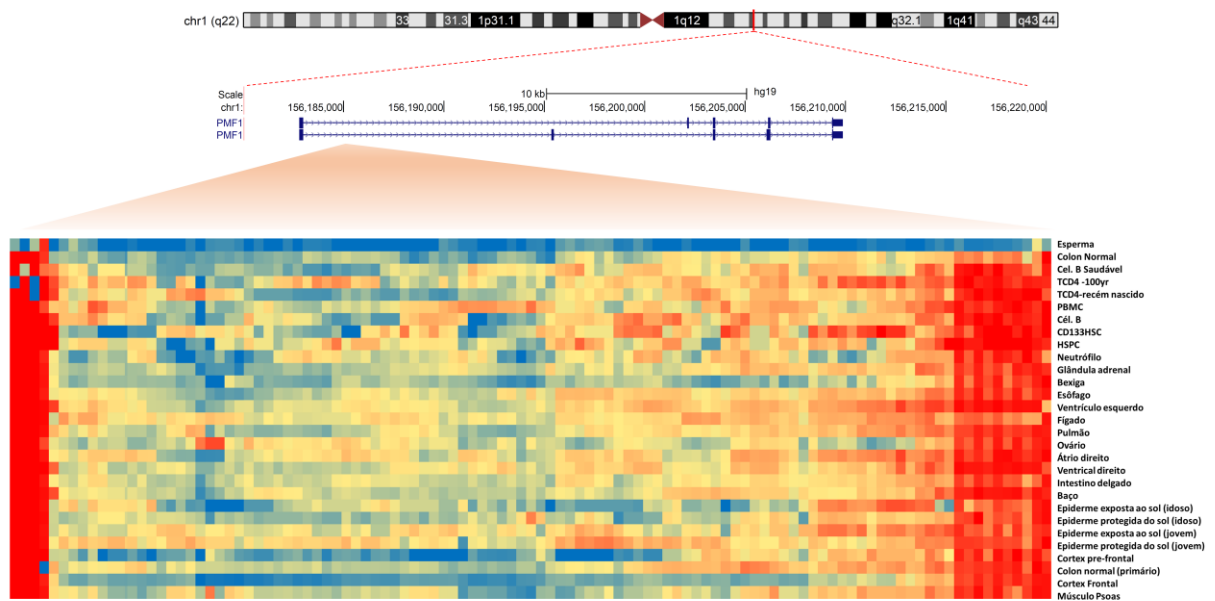


Figura 25. Caracterização da DMR da região *PMF1* com um gráfico de calor representativo dos perfis de hipometilação (em azul), metilação intermediária (em amarelo) e hipermetilação (em vermelho). Com dados de metiloma de 129 tecidos de 25 estudos no PublicHub DNA methylation. A posição física referente a essa região de hemimetilação, em amarelo, é chr1:156186540-156186802.

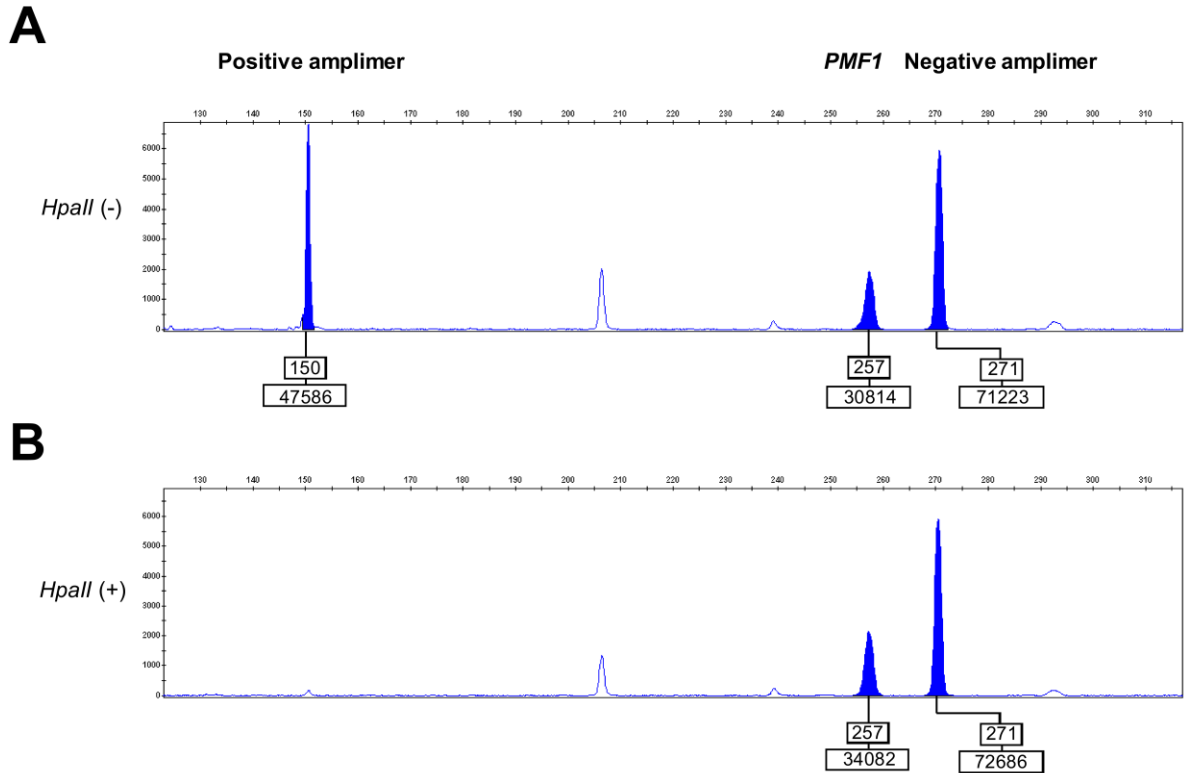


Figura 26. Eletroferograma representativo da identificação do *locus PMF1* com perfil de hipermetilação alélica. À esquerda com 150pb temos o controle positivo que depois da digestão com a enzima *HpaII* desaparece por ser uma região desmetilada. Ao meio com 257pb temos o alvo que depois da digestão, apresentou uma área aumentada (perfil hipermetilação) e do lado direito com 271pb o controle negativo sem sítio de digestão, permanecendo sem modificação depois da digestão.

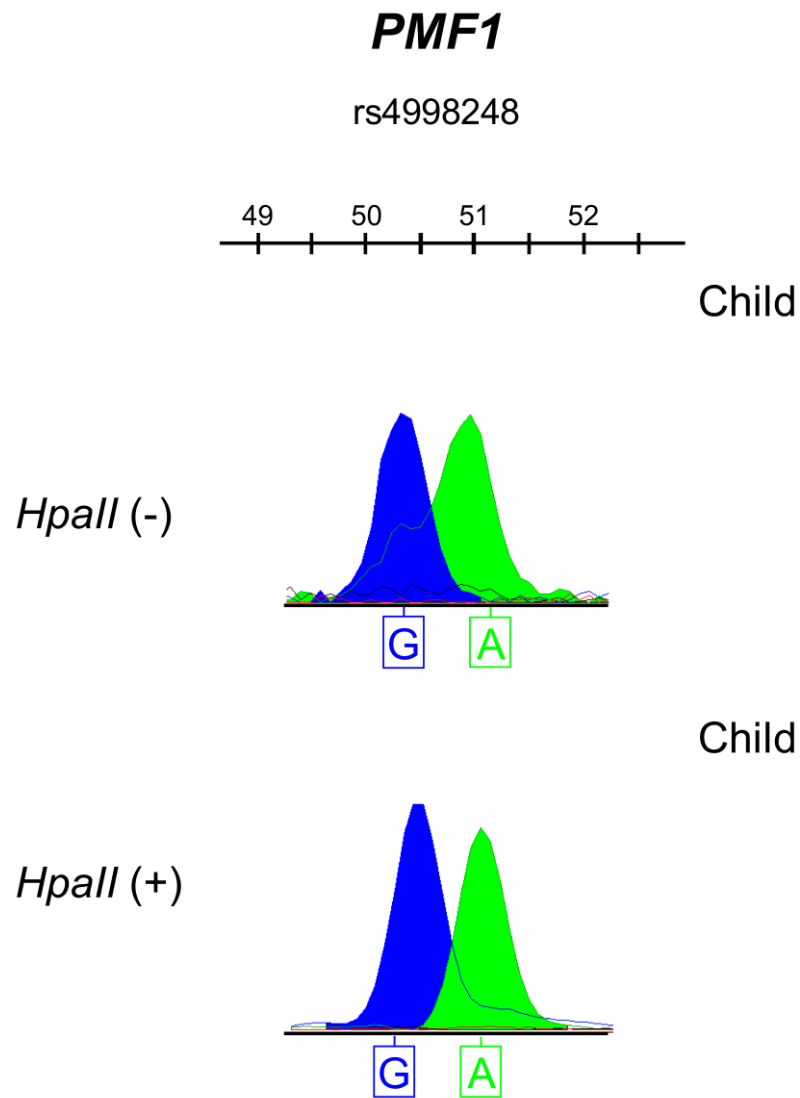


Figura 27. Simetria gamética na gDMR do gene *PMF1*. Eletroferograma representativo do perfil de metilação pontual na gDMR alvo em amostras de DNA de um indivíduo controle 1545F informativo para o SNP rs4998248 ligado ao sítio CpG alvo da interrogação do estado de metilação dos alelos parentais. Ambos os alelos são refratários à digestão.

A metilação do gene *PMF1* foi determinada por DESM-QF-PCR em 5 amostras, entretanto nenhuma foi em triplicata. A média da metilação para essa DMR foi de 100%.

7. Discussão

Existe um número considerável de genes imprintados já relatados na literatura, no entanto apenas 12 doenças são associadas com alterações do *imprinting* genômico (Soellner, Begemann *et al.*, 2017). Estudos em busca desses genes imprintados partem da análise do perfil de metilação em *loci* específicos, e tem como principal estratégia utilizar o Illumina 450k (cerca de 480.000 pontos de citosina) para um rastreamento no genoma com mais de 3 sondas e com baixo desvio padrão por *loci*.

Docherty e colaboradores (2014) e Court e colaboradores (2014), que são grupos especialistas na investigação de DMRs, não são tão atuantes em pesquisas de genes imprintados como Baran e colaboradores (2015). A estratégia desses três grupos é encontrar uma região com metilação intermediária, varrendo o metiloma em busca do fator sugestivo que esta é uma possível DMR imprintada em grande parte dos tecidos, como o *SNURF* (Baran, Subramaniam *et al.*, 2015). Entretanto, encontrar uma nova DMR imprintada e com expressão origem parental dependente como o *SNURF* é uma tarefa árdua. *PEG10*, citado por Baran como gene imprintado, apresenta expressão monoalélica em apenas em 13 de 30 tecidos, e o gene citado *GRB10*, foi validado como imprintado apenas no cérebro, ou seja, como *imprinting* tecido-específico (Baran, Subramaniam *et al.*, 2015).

Além dessa dificuldade intrínseca na busca de DMRs imprintadas, a ferramenta utilizada por esses grupos possui uma distribuição de sondas no genoma não uniforme, para alguns genes conhecidamente imprintados, como o *PEG10*, há 53 sondas, um número considerável (Court, Tayama *et al.*, 2014). Entretanto, o nosso alvo *GCSAML* possui apenas 5 sondas, destas, 3 estão localizadas na região hemimetilada. Isso pode ser um dos motivos de Court e colaboradores (2014) não terem apontado essa DMR como uma possível DMR imprintada.

O nosso método de rastreamento em todo o genoma de regiões hemimetiladas na plataforma de análises UCSC *Genome Browser* em uma janela integrativa de 40 kb com múltiplos estudos de DNA tratados com bissulfito de sódio, demonstrou ser uma estratégia eficiente comparada a uma das mais utilizadas nos artigos, como o próprio Illumina 450k (Dedeurwaerder, Defrance *et al.*, 2011) . As análises deste estudo consistiram no racional da estratégia metodológica desenvolvida e

aprimorada pelos membros do nosso laboratório em analisar QF-PCR-DESM com a intenção de verificar o estado de hemimetilação, seguida pelo ensaio de SNaPshot nos produtos da digestão com enzimas sensíveis à metilação (Alves Da Silva, Machado *et al.*, 2016). Esta metodologia foi extremamente eficaz para determinar se as hemimetilações rastreadas consistiam em uma DMR imprintada ou não. E para analisar se a expressão do *loci* gênico era de fato imprintado, foram realizadas análises de expressão por RNA-Seq. Essa foi uma metodologia bem eficaz para determinar se os alvos escolhidos para a parte experimental *in vitro* era uma DMR imprintada causando um efeito de transcrição origem parental-dependente.

Portanto foi possível, por meio da estratégia de visualização em 40 kb, encontrar regiões candidatas à DMRs ainda não encontrados na literatura como, por exemplo, o *GCSAML*, tendo sua iDMR validada nesse presente estudo. É importante analisar o perfil dos genes imprintados no UCSC *Genome Browser* (40 kb) com intuito de ver se o perfil tem similaridade com os novos alvos. Os genes imprintados geralmente possuem uma região com perfil hemimetilado na maior parte dos metilomas analisados, por isso priorizamos rastrear regiões hemimetiladas constitutivas.

Este trabalho aponta pela primeira vez, que *PARD6G-AS1* e *GCSAML* são iDMRs de origem materna. O *imprinting* do 5^mCpG são derivadas de oócitos (*PARD6G-AS1* iDMR) e somáticas (*GCSAML* iDMR) e sua herança é dissociada com os efeitos de origem parental em pelo menos 52 tecidos adultos somáticos, conforme determinado por perfis RNA-Seq que eram consistentes com expressão bialélica em múltiplos SNP informativos. Essa abordagem não pode excluir a possibilidade de expressão monoalélica em tecidos nos quais os níveis de expressão estão abaixo do limite de detecção de 12 leituras por SNP aceitas no nosso projeto experimental. Da mesma forma, não podemos excluir a possibilidade de que as iDMRs sejam associados à expressão de *loci* gênico que estão distantes 2,3 Mb *upstream* e *downstream* dos nossos alvos, que é a distância física mais alta entre qualquer iDMR conhecida e o gene controlado subjacente (s) (ou seja, o *SNRPN / SNURF* iDMR). É importante ressaltar que assim como os nossos alvos há também várias DMRs imprintadas sem genes imprintados, como por exemplo o *WRB*, que possui iDMRs mas que quando validado transcricionalmente com SNPs para diferenciação dos cromossomos parentais, observou-se expressão bialélica (Alves Da Silva, Machado *et al.*, 2016).

O nosso alvo *PARD6G-AS1* já foi apontado como um novo candidato após uma análise em que foi detectada uma hipometilação em três sondas na região associada com a síndrome de Beckwith-Wiedemann e em nove sondas na região associada a Diabetes Neonatal Transitório (Docherty, Rezwan *et al.*, 2014), um número considerável de sondas. Na nossa análise DESM-QF-PCR para esse alvo encontramos uma média de metilação mais baixa (31,50%), ou seja, houve uma tendência à hipometilação e isso pode ser devido à diferença de metilação das células sanguíneas. Os Neutrófilos de acordo com o Sequenciamento bissulfito em todo o genoma (WGBS) são hemimetilados. Entretanto, na população brasileira a ausência do antígeno Duffy, característica frequente em afrodescendentes, é causadora da neutropenia benigna sem repercussão clínica. Porém, pode alterar a proporção de células desses indivíduos, que possuem cerca de metade de neutrófilos que um indivíduo portador do antígeno Duffy (Horuk, 2015).

Da mesma forma, o estado de metilação no local cg18973878 na iDMR materna *PARD6G-AS1* não codificado foi recentemente identificado como um cis-meQTL, significativamente afetado quando o alelo menor T rs11659843 SNP é transmitido pelo pai (valores de p de todo o genoma $1,95 \times 10^{-7}$ a $5,32 \times 10^{-21}$) (Cuellar Partida, 2017). O SNP que exerce o efeito cis-meQTL é intrônico para o gene *PARD6G* de codificação de proteína, e mapeia 79,598 pb (hg19) longe do iDMR materna *PARD6G-AS1*, o que, evidentemente, não se sobrepõe ao gene *PARD6G*.

Também observamos que, dentro da iDMR materna *PARD6G-AS1*, evidências de locais específicos do alelo em haplótipos dependentes foram relatados em vários metilomas (Do, Lang *et al.*, 2016). Nós propomos que os epi-haplótipos acima sejam dependentes da origem parental e, portanto, eles seriam úteis no estudo da susceptibilidade às doenças imprintadas complexas genéticas (Sanchez-Delgado, Riccio *et al.*, 2016)

Como relevância biológica do *imprinting* genômico, observamos que o estado de hipometilação com caráter de significância anormal foi relatado para certos locais de CpG localizados dentro da iDMR de origem gamética *PARD6G-AS1*. Além dos sítios CpG citados anteriormente como tendo relação a um fenótipo clínico, mais dois locais de CpG abrangentes são impactados em pacientes com pseudo-hipoparatiroidismo com defeitos de *imprinting* de cluster *GNAS* (Roctus, Martin-Trujillo *et al.*, 2016). Interessantemente, um estudo de associação com

comportamento de agressão em gêmeos em todo o epigenoma mostrou que o aumento em 0,1% no sítio cg06092953 localizado dentro da recém-validada iDMR materna *PARD6G-AS1* reflete em um aumento de um desvio padrão no score de comportamento agressivo em gêmeos monozigóticos e dizigóticos da Holanda (Van Dongen, Nivard *et al.*, 2015). Entretanto, como usaram mais de 400 gêmeos idênticos, esses dados não indicam que o comportamento agressivo seja dependente da origem parental epi-haplótipos e, sim, da metilação. Nossa hipótese é que a referida associação epigenética com agressão pode ser influenciada pela metilação e não pelo epi-haplótipo origem parental dependente. E que futuros estudos de coorte e caso controle, devem ser realizados a fim de determinar se o envelhecimento e o uso de fármacos com ação desmetilantes (5'Azacitidina) podem não só alterar a metilação desse ponto, como influenciar no score de comportamento agressivo desses indivíduos.

Quanto ao *GCSAML*, assim como previsto na análise 40k, os resultados DESM-QF-PCR apresentaram uma hemimetilação. Alterações nessa hemimetilação para um estado hipermetilado foi relatada no córtex pré-frontal *post mortem* em um caso de síndrome de Klinefelter com o diagnóstico de massa cerebelar reduzida e esquizofrenia (47, cariótipo XXY) (Viana, Pidsley *et al.*, 2014). E posteriormente confirmado em um segundo estudo usando amostras de sangue venoso de cinco pacientes com síndrome de Klinefelter (Wan, Qiu *et al.*, 2015). É conhecido, que pacientes com Klinefelter têm risco aumentado para Sjogren, Lúpus Eritromatoso Sistêmico e assim como os pacientes com Síndrome de Down, possuem risco reduzido para tumores sólidos, porém risco aumentado para tumores de origem hematológica (Scofield, Bruner *et al.*, 2008; Ji, Zoller *et al.*, 2016). Ou seja, o cromossomo X extra deve assim como já foi provado para o 21, alterar o perfil epigenético *cis* e *trans* em vários sítios (Kerkel, Schupf *et al.*, 2010; Jones, Farre *et al.*, 2013).

O nosso alvo *PWWP2AP1* não é uma iDMR, além, disso possui a expressão alélica origem parental independente. O quarto alvo, *PMF1* não foi confirmado a iDMR devido ao desenho ter focado a QF-DESM-PCR na extremidade dessa região hemimetilada. As bordas dessa região são hipermetiladas, o que foi constatado pelo nosso método e por isso no SNaPshot permaneceram ambos os alelos analisados antes e depois da digestão com enzimas sensíveis à metilação.

Apesar de um dos alvos não ser uma iDMR, com a estratégia 40K conseguimos detectar 2 novas iDMRs maternas de um total de 3 bons desenhos experimentais. Esses três bons desenhos foram realizados a partir de uma lista com 123 alvos. Sendo assim, acreditamos que essa estratégia tem o potencial de detectar pelo menos 80 novas iDMRs com potenciais de estarem associadas a genes imprintados.

|

.

|

8. CONCLUSÃO

- ✓ A estratégia de visualização manual dos perfis comparativos de metilação de DNA em experimentos de metiloma em repositórios públicos para diversos tecidos e gametas em sessões customizadas na plataforma *UCSC Genome Browser* elaborados nesse estudo, apesar de laborioso mostrou-se altamente eficiente no rastreamento de regiões hemimetiladas candidatas a iDMRs e genes imprintados no genoma humano.
- ✓ A metodologia desenvolvida pelo nosso grupo de pesquisa, que consiste na digestão da metilação utilizando a enzima de restrição *HpaII*, sensível à metilação, seguida por sequenciamento por extensão de um único nucleotídeo (SNaPshot), foi capaz de concluir se os genes alvo de fato possuíam uma DMR imprintada e qual foi a origem dos alelos, ou seja, se vieram do pai ou da mãe, nos trios informativos com o filho heterozigoto.
- ✓ As regiões hemimetiladas nesses genes *PARD6G-AS1* e *GCSAML* foram considerados DMR imprintadas de origem materna enquanto que o *PWWP2AP1* foi de origem-parental-independente.
- ✓ Os três *loci* citados (*PARD6G-AS1*, *GCSAML*, *PWWP2AP1*) foram validados computacionalmente quanto à expressão por RNA-Seq e tiveram SNPs com expressão bialélica em diversos doadores para 52 diferentes tecidos.
- ✓ A região na extremidade hipermetilada onde foram desenhado os iniciadores referentes ao gene *PMF1* explica a hipermetilação observada nos ensaios, pois os iniciadores deveriam ter sido desenhados mais para o centro da região hemimetilada.
- ✓ Dado a relevância de uma DMR imprintada e que 2/3 das regiões hemimetiladas testadas foram DMR imprintadas, considera-se que dos 123 *loci* gênicos preditos deverá ter possivelmente 82 DMRs imprintadas a serem

encontradas. A descoberta de novas iDMRs pode ter um grande impacto clínico para o entendimento de doenças genéticas ainda sem muitos estudos elucidados, em futuras pesquisas.

✓ As regiões diferentemente metiladas que estão configuradas para o imprinting materno no *PARD6G-AS1* e no *GCSAML* estão desacopladas aos efeitos de expressão origem parental dependente em múltiplos tecidos humanos.

8. REFERÊNCIAS

- Abu-Amero, S., D. Monk, *et al.* The genetic aetiology of Silver-Russell syndrome. J Med Genet, v.45, n.4, Apr, p.193-9. 2008.
- Allen S. Yang, M. R. H. E., Ketan Doshi, Yutaka Kondo, Eloiza H. Tajara And e J.-P. J. Issa. A simple method for estimating global DNA methylation using bisulfite PCR of repetitive DNA elements. Nucleic Acids Research, v.32. 2004.
- Alves Da Silva, A. F., F. B. Machado, *et al.* Trisomy 21 Alters DNA Methylation in Parent-of-Origin-Dependent and -Independent Manners. PLoS One, v.11, n.4, p.e0154108. 2016.
- Babak, T., B. Deveale, *et al.* Genetic conflict reflected in tissue-specific maps of genomic imprinting in human and mouse. Nat Genet, v.47, n.5, May, p.544-9. 2015.
- Baran, Y., M. Subramaniam, *et al.* The landscape of genomic imprinting across diverse adult human tissues. Genome Res, v.25, n.7, Jul, p.927-36. 2015.
- Barlow, D. P., R. Stoger, *et al.* The mouse insulin-like growth factor type-2 receptor is imprinted and closely linked to the Tme locus. Nature, v.349, n.6304, Jan 03, p.84-7. 1991.
- Bartolomei, M. S., S. Zemel, *et al.* Parental imprinting of the mouse H19 gene. Nature, v.351, n.6322, May 09, p.153-5. 1991.
- Biliya, S. e L. A. Bulla, Jr. Genomic imprinting: the influence of differential methylation in the two sexes. Exp Biol Med (Maywood), v.235, n.2, Feb, p.139-47. 2010.
- Bird, A. DNA methylation patterns and epigenetic memory. Genes Dev, v.16, n.1, Jan 01, p.6-21. 2002.
- _____. The methyl-CpG-binding protein MeCP2 and neurological disease. Biochem Soc Trans, v.36, n.Pt 4, Aug, p.575-83. 2008.
- Butler, M. G. Genomic imprinting disorders in humans: a mini-review. J Assist Reprod Genet, v.26, n.9-10, Sep-Oct, p.477-86. 2009.
- Court, F., C. Tayama, *et al.* Genome-wide parent-of-origin DNA methylation analysis reveals the intricacies of human imprinting and suggests a germline methylation-independent mechanism of establishment. Genome Res, v.24, n.4, Apr, p.554-69. 2014.

Cuellar Partida, G., Laurin, C., Ring, S., Gaunt, T.R., Relton, C.L., Davey Smith, G., and Evans, D. (2017). Imprinted loci may be more widespread in humans than previously appreciated and enable limited assignment of parental allelic transmissions in unrelated individuals. bioRxiv. 2017.

Dedeurwaerder, S., M. Defrance, *et al.* Evaluation of the Infinium Methylation 450K technology. Epigenomics, v.3, n.6, Dec, p.771-84. 2011.

Delaval, K. e R. Feil. Epigenetic regulation of mammalian genomic imprinting. Curr Opin Genet Dev, v.14, n.2, Apr, p.188-95. 2004.

Deng, X., J. B. Berletch, *et al.* X chromosome regulation: diverse patterns in development, tissues and disease. Nat Rev Genet, v.15, n.6, Jun, p.367-78. 2014.

Do, C., C. F. Lang, *et al.* Mechanisms and Disease Associations of Haplotype-Dependent Allele-Specific DNA Methylation. Am J Hum Genet, v.98, n.5, May 05, p.934-955. 2016.

Docherty, L. E., F. I. Rezwan, *et al.* Genome-wide DNA methylation analysis of patients with imprinting disorders identifies differentially methylated regions associated with novel candidate imprinted genes. J Med Genet, v.51, n.4, Apr, p.229-38. 2014.

Donlon, T. A. Similar molecular deletions on chromosome 15q11.2 are encountered in both the Prader-Willi and Angelman syndromes. Hum Genet, v.80, n.4, Dec, p.322-8. 1988.

Eckersley-Maslin, M. A. e D. L. Spector. Random monoallelic expression: regulating gene expression one allele at a time. Trends Genet, v.30, n.6, Jun, p.237-44. 2014.

Gimelbrant, A., J. N. Hutchinson, *et al.* Widespread monoallelic expression on human autosomes. Science, v.318, n.5853, Nov 16, p.1136-40. 2007.

Grewal, S. I. e D. Moazed. Heterochromatin and epigenetic control of gene expression. Science, v.301, n.5634, Aug 08, p.798-802. 2003.

Hackett, J. A. e M. A. Surani. Beyond DNA: programming and inheritance of parental methylomes. Cell, v.153, n.4, May 09, p.737-9. 2013.

Hannula-Jouppi, K., M. Muurinen, *et al.* Differentially methylated regions in maternal and paternal uniparental disomy for chromosome 7. Epigenetics, v.9, n.3, Mar, p.351-65. 2014.

Henckel, A. e P. Arnaud. Genome-wide identification of new imprinted genes. Brief Funct Genomics, v.9, n.4, Jul, p.304-14. 2010.

Horuk, R. The Duffy Antigen Receptor for Chemokines DARC/ACKR1. Front Immunol, v.6, p.279. 2015.

- Ji, J., B. Zoller, *et al.* Risk of solid tumors and hematological malignancy in persons with Turner and Klinefelter syndromes: A national cohort study. Int J Cancer, v.139, n.4, Aug 15, p.754-8. 2016.
- Jones, M. J., P. Farre, *et al.* Distinct DNA methylation patterns of cognitive impairment and trisomy 21 in Down syndrome. BMC Med Genomics, v.6, Dec 27, p.58. 2013.
- Kelsey, G. e R. Feil. New insights into establishment and maintenance of DNA methylation imprints in mammals. Philos Trans R Soc Lond B Biol Sci, v.368, n.1609, Jan 5, p.20110336. 2013.
- Kerkel, K., N. Schupf, *et al.* Altered DNA methylation in leukocytes with trisomy 21. PLoS Genet, v.6, n.11, Nov 18, p.e1001212. 2010.
- Kouzarides, T. Chromatin modifications and their function. Cell, v.128, n.4, Feb 23, p.693-705. 2007.
- Lalande, M. Parental imprinting and human disease. Annu Rev Genet, v.30, p.173-95. 1996.
- Li, B., M. Carey, *et al.* The role of chromatin during transcription. Cell, v.128, n.4, Feb 23, p.707-19. 2007.
- Luger, K., A. W. Mader, *et al.* Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature, v.389, n.6648, Sep 18, p.251-60. 1997.
- Mann, J. R. Imprinting in the germ line. Stem Cells, v.19, n.4, p.287-94. 2001.
- Morison, I. M., J. P. Ramsay, *et al.* A census of mammalian imprinting. Trends Genet, v.21, n.8, Aug, p.457-65. 2005.
- Obata, Y. e T. Kono. Maternal primary imprinting is established at a specific time for each gene throughout oocyte growth. J Biol Chem, v.277, n.7, Feb 15, p.5285-9. 2002.
- Ogata, T., M. Kagami, *et al.* Molecular mechanisms regulating phenotypic outcome in paternal and maternal uniparental disomy for chromosome 14. Epigenetics, v.3, n.4, Jul-Aug, p.181-7. 2008.
- Plasschaert, R. N. e M. S. Bartolomei. Genomic imprinting in development, growth, behavior and stem cells. Development, v.141, n.9, May, p.1805-13. 2014.
- Price, S. M., R. Stanhope, *et al.* The spectrum of Silver-Russell syndrome: a clinical and molecular genetic study and new diagnostic criteria. J Med Genet, v.36, n.11, Nov, p.837-42. 1999.
- Reik, W. e J. Walter. Genomic imprinting: parental influence on the genome. Nat Rev Genet, v.2, n.1, Jan, p.21-32. 2001.

- Rochtus, A., A. Martin-Trujillo, *et al.* Genome-wide DNA methylation analysis of pseudohypoparathyroidism patients with GNAS imprinting defects. Clin Epigenetics, v.8, p.10. 2016.
- Sanchez-Delgado, M., A. Riccio, *et al.* Causes and Consequences of Multi-Locus Imprinting Disturbances in Humans. Trends Genet, v.32, n.7, Jul, p.444-55. 2016.
- Santoni, F. A., G. Stamoulis, *et al.* Detection of Imprinted Genes by Single-Cell Allele-Specific Gene Expression. Am J Hum Genet, v.100, n.3, Mar 02, p.444-453. 2017.
- Scofield, R. H., G. R. Bruner, *et al.* Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: support for the notion of a gene-dose effect from the X chromosome. Arthritis Rheum, v.58, n.8, Aug, p.2511-7. 2008.
- Soellner, L., M. Begemann, *et al.* Recent Advances in Imprinting Disorders. Clin Genet, v.91, n.1, Jan, p.3-13. 2017.
- Surani, M. A., S. C. Barton, *et al.* Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. Nature, v.308, n.5959, Apr 5-11, p.548-50. 1984.
- Surani, M. A., R. Kothary, *et al.* Genome imprinting and development in the mouse. Dev Suppl, p.89-98. 1990.
- Unnikrishnan, A., P. R. Gafken, *et al.* Dynamic changes in histone acetylation regulate origins of DNA replication. Nat Struct Mol Biol, v.17, n.4, Apr, p.430-7. 2010.
- Van Dongen, J., M. G. Nivard, *et al.* Epigenome-Wide Association Study of Aggressive Behavior. Twin Res Hum Genet, v.18, n.6, Dec, p.686-98. 2015.
- Viana, J., R. Pidsley, *et al.* Epigenomic and transcriptomic signatures of a Klinefelter syndrome (47,XXY) karyotype in the brain. Epigenetics, v.9, n.4, Apr, p.587-99. 2014.
- Wan, E. S., W. Qiu, *et al.* Genome-wide site-specific differential methylation in the blood of individuals with Klinefelter syndrome. Mol Reprod Dev, v.82, n.5, May, p.377-86. 2015.
- Weksberg, R., I. Teshima, *et al.* Molecular characterization of cytogenetic alterations associated with the Beckwith-Wiedemann syndrome (BWS) phenotype refines the localization and suggests the gene for BWS is imprinted. Hum Mol Genet, v.2, n.5, May, p.549-56. 1993.
- Williamson Cm, Blake A, *et al.* Mouse Imprinting Data and References. Nature. 2013.
- Wood, A. J. e R. J. Oakey. Genomic imprinting in mammals: emerging themes and established theories. PLoS Genet, v.2, n.11, Nov 24, p.e147. 2006.

Yamazawa, K., T. Ogata, *et al.* Uniparental disomy and human disease: an overview.
Am J Med Genet C Semin Med Genet, v.154C, n.3, Aug 15, p.329-34. 2010.