

Camila Bassi Fernandes da Silva

**Investigação do padrão de metilação em
genes ou regiões candidatas à fissura labial
com ou sem palato não-sindrômica (FL/P NS)**

*Investigation of methylation pattern in candidate
genes or regions to non-syndromic cleft lip with or
without palate (CL/P NS)*

São Paulo

2020

Camila Bassi Fernandes da Silva

Investigação do padrão de metilação em genes
ou regiões candidatas à fissura labial com ou
sem palato não-sindrômica (FL/P NS)

*Investigation of methylation pattern in candidate
genes or regions to non-syndromic cleft lip with or
without palate (CL/P NS)*

Dissertação apresentada ao Instituto
de Biociências da Universidade de
São Paulo, para a obtenção de Título
de Mestre em Ciências, na Área de
Biologia/Genética.

Orientador(a): Prof^a. Dra. Maria Rita
dos Santos e Passos-Bueno

São Paulo

2020

Ficha Catalográfica

Silva, Camila Bassi Fernandes da

Investigação do padrão de metilação em genes candidatos à fissura labial com ou sem palato não-sindrômica (FL/P NS)

101 páginas

Dissertação (Mestrado) - Instituto de Biociências da Universidade de São Paulo. Departamento de Genética e Biologia Evolutiva.

1. Fissuras labiopalatinas 2. Metilação I. Universidade de São Paulo. Instituto de Biociências. Departamento de Genética e Biologia Evolutiva

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a).

Profª. Dra. Maria Rita S. Passos-Bueno

Orientadora

Agradecimentos

Agradeço à minha família, especialmente aos meus pais e minha irmã, por sempre me apoiarem nas minhas decisões e por todo o amor e carinho que sempre me deram. Graças ao suporte deles realizei sonhos que nem imaginava que seria capaz.

Agradeço à Rita por toda a dedicação, orientação e paciência e por essa grande oportunidade de estudo, que me permitiu obter novos aprendizados e experiências de vida.

Agradeço ao Luquinhas, que ontem foi meu parceiro de IC e hoje me ensinou com toda a sua paciência e disponibilidade e me ajudou muito nessa etapa e, principalmente, pela amizade e carinho por todos esses anos.

Agradeço ao “mestre” Luciano, que foi a primeira pessoa a me ensinar praticamente tudo que sei hoje de laboratório, e por sempre estar disposto a me ajudar.

Agradeço aos meus amigos do laboratório Gér, Juninha, Sofi, Letícia, Elisa, Fred, May, Dani, Claudinha, Vanzinha, Cá, Ágatha, Duda, Gabi, Camila G., Diogo, Andressa, Igor, Isa, Naila e Simone, pela companhia nessa etapa e por serem essas pessoas incríveis.

Agradeço aos meus colegas Vinícius e Renan por terem contribuído em uma das etapas desse trabalho e por serem pessoas super prestativas e disponíveis.

Agradeço aos pacientes e familiares que doaram materiais por acreditarem e incentivarem a pesquisa, permitindo a realização deste trabalho.

Agradeço à Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) e ao Centro de Pesquisa, Inovação e Difusão (CEPID) (número do processo FAPESP/CEPID 2013/08028-1) pelo apoio financeiro.

Lista de abreviações

<i>1kGP</i>	<i>1000 Genomes Project</i>
<i>6500ESP</i>	<i>Exome Sequencing Project</i>
<i>AbraOM</i>	Arquivo brasileiro Online de Mutações
<i>CpG</i>	Citosina-fosfato-guanina
<i>EMT</i>	Transição Epitélio-Mesenquimal
<i>ExAC</i>	<i>Exome Aggregation Consortium</i>
<i>eQTL</i>	Loci de características quantitativas de expressão
<i>FLP</i>	Fissura labiopalatina
<i>FL/P</i>	Fissura do lábio com ou sem fissura do palato
<i>FL/P NS</i>	Fissura do lábio com ou sem fissura do palato não-sindrômica
<i>FP</i>	Fissura do palato
<i>GWAS</i>	<i>Genome-wide association studies</i>
<i>MGI</i>	<i>Mouse Genome Informatics</i>
<i>NGS</i>	Sequenciamento de nova geração
<i>PCA</i>	Análise de Componente Principal
<i>PCP</i>	Polaridade celular planar
<i>PCR</i>	Reação em cadeia de polimerase
<i>PFn</i>	Proeminência frontonasal
<i>PMX</i>	Proeminências maxilares
<i>PNL</i>	Proeminências nasais laterais
<i>PNM</i>	Proeminências nasais mediais
<i>PPS</i>	Síndrome do pterígeo poplíteo
<i>SCE</i>	Sequenciamento completo de exoma
<i>SNV</i>	<i>Single Nucleotide Variation</i>
<i>SVW</i>	Síndrome de Van der Woude
<i>UBC</i>	Poliubiquitina-C

Índice

1. Introdução	11
1.1. Fissuras labiopalatinas	11
1.2. Fatores genéticos da fissura do lábio com ou sem fissura do palato (FL/P)	15
1.2.1. FL/Ps síndrômicas.....	15
1.2.2. FL/Ps não-síndrômicas	16
1.3. Epigenética	19
1.4. Fatores Genéticos X Epigenética nas FL/Ps NS.....	21
1.5. Hipóteses do presente trabalho.....	23
2. Objetivos gerais e específicos.....	25
2.1. Objetivos gerais.....	25
2.2. Objetivos específicos.....	25
3. Material e Métodos	27
3.1. Casuística e ética	27
3.2. Abordagens utilizadas para seleção de genes a serem incluídos na análise de metilação	28
3.2.1. Seleção de genes por análise de SCE de famílias inglesas	29
3.2.2. Seleção de genes em famílias brasileiras	31
• Seleção de genes por SCE (Brito, 2016).....	31
• Seleção de genes candidatos de painel de FL/P NS	31
• Reanálise de SCE da família F886	32
• Estudo de ligação da família F886 e reanálise dos dados de SCE	32
3.3. Análise de metilação	34
3.3.1. Análise de metilação para os genes candidatos selecionados por SCE e painel.....	34
3.3.2. Análise de metilação na família F886.....	39
4. Resultados	41
4.1. Genes selecionados para a análise de metilação.....	41
4.1.1. Genes selecionados nas famílias inglesas	41
4.1.2. Genes selecionados nas famílias brasileiras	43
• Genes candidatos de Brito (2016).....	43
• Genes candidatos de painel de FL/P NS.....	46
• Reanálise de SCE da família F886	46
• Estudo de ligação da família F886 e reanálise dos dados de SCE	48
4.2. Análise de metilação	50

4.2.1. Análise de metilação para os genes candidatos selecionados por SCE e painel.....	51
• Análise de metilação na amostra global	52
• Análise de metilação por família	58
4.2.2. Análise de metilação na família F886.....	59
5. Discussão	61
6. Conclusões.....	67
7. Resumo	69
8. <i>Abstract</i>	71
9. Referências Bibliográficas.....	73
10. Anexos	83
Anexo I.....	85
Anexo II.....	89
Anexo III.....	91
Anexo IV	95

1. Introdução

1.1. Fissuras labiopalatinas

As fissuras labiopalatinas (FLP) são malformações congênitas caracterizadas por falhas na formação do lábio superior (com ou sem envolvimento do palato) ou apenas do palato durante o desenvolvimento embrionário (Marazita, 2012; Leslie & Marazita, 2013).

Os pacientes com FLP podem apresentar limitações na fala, na audição e problemas na formação dos dentes. Além disso, é comum apresentarem dificuldades na comunicação e integração social, uma vez que, como a malformação ocorre na face, é notadamente visível pela sociedade, fazendo com que os indivíduos com FLP tenham dificuldades em se relacionar com outras pessoas e, conseqüentemente, podem apresentar problemas psicossociais (Marazita, 2012; Dixon *et al.*, 2011; Leslie & Marazita, 2013). Foi também observado que indivíduos com FLP têm risco de mortalidade no primeiro ano de vida nove vezes maior do que a população em geral em países em desenvolvimento (Carlson *et al.*, 2013), devido principalmente a dificuldades na alimentação. O tratamento desses pacientes é de custo elevado, estimado em US\$200.000,00 (Berk & Marazita, 2002; Leslie & Marazita, 2013), em que incidem procedimentos cirúrgicos e acompanhamento multidisciplinar que envolve uma equipe de profissionais médicos, geneticistas, fonoaudiólogos, dentistas, nutricionistas e psicólogos (Dixon *et al.*, 2011; Leslie & Marazita, 2013; Stuppia *et al.*, 2011; Mossey *et al.*, 2009). Pelas dificuldades que o paciente enfrenta e pelo custo elevado do tratamento, é importante buscar quais fatores e mecanismos levam a essas malformações, com o intuito de desenvolver estratégias de aconselhamento genético e, principalmente, preventivas.

A malformação do lábio e/ou do palato ocorre ao longo do desenvolvimento da face do embrião. O desenvolvimento craniofacial ocorre entre a 3ª e a 10ª semana de gestação e depende de uma sequência de eventos celulares precisamente coordenados como crescimento, proliferação, migração, transição epitélio-mesenquimal, diferenciação e apoptose. Se houver algum distúrbio em algum desses processos, como

falha no crescimento ou na fusão das estruturas, o feto poderá apresentar fissura do lábio superior e/ou do palato (Figuras 1 e 2) (Leslie & Marazita, 2013; Twigg & Wilkie, 2015).

Os indivíduos com FLP apresentam fenótipos variáveis, podendo apenas ter a falha no lábio superior (unilateral ou bilateral), ou também ter afetadas as regiões dento-alveolar e o palato primário (anterior ao forâmen pré-incisivo) e secundário (posterior ao forâmen pré-incisivo) ou apenas o palato, restrito ao palato secundário (forâmen pós-incisivo) (Schutte & Murray, 1999; Taib *et al.*, 2015) (Figura 3). O tipo de fissura mais comum é a fissura labial unilateral do lado esquerdo (Dixon *et al.*, 2011; Gundlach & Maus, 2006).

Tendo em vista que os processos de diferenciação das estruturas que dão origem ao lábio e ao palato ocorrem em momentos diferentes durante a embriogênese, as FLPs foram separadas em duas classes principais: Fissura do lábio com ou sem fissura do palato (FL/P) e Fissura do palato (FP) (Fogh-Andersen; 1942; Fraser, 1955; Mossey *et al.*, 2009). Além disso, observa-se que não há ocorrência dos dois tipos de fissura (FL/P e FP) em uma mesma família, com raras exceções. Dentre essas exceções, destacam-se as famílias com indivíduos com a Síndrome de Van der Woude (SVW), que é uma condição autossômica dominante causada em geral por mutações de perda de função em heterozigose no gene *IRF6* (Dixon *et al.*, 2011). Dessa forma, para estudos genéticos e aconselhamento, tem sido considerado que, para a maioria dos casos, FL/P e FP têm etiologia distinta.

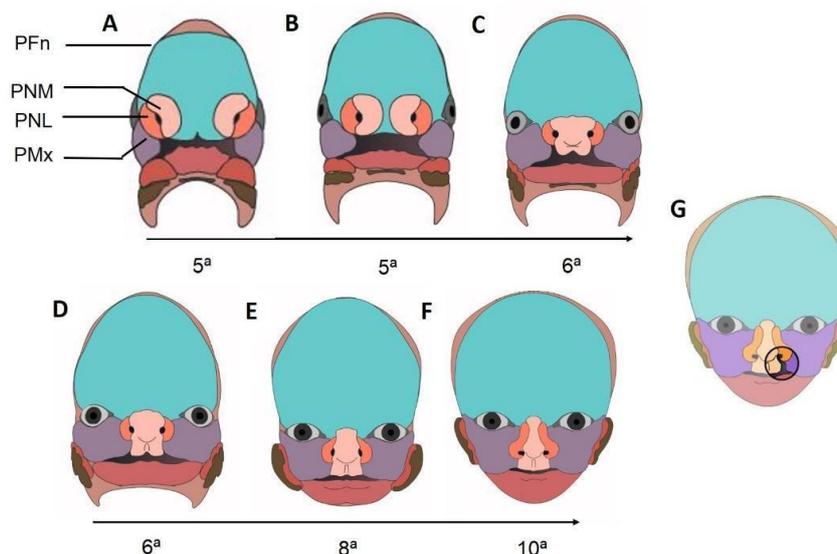


Figura 1 – Formação da face e do lábio superior. Durante este processo, ocorre a transição epitélio-mesenquimal, em que as células da borda da placa neural, a princípio do tipo epiteliais, passam a possuir um fenótipo migratório e mesenquimal, o que permite seu desprendimento do epitélio e migração, formando a proeminência frontonasal e os arcos faríngeos. O primeiro arco faríngeo origina as proeminências maxilares. A proeminência frontonasal origina os pares de proeminências nasais mediais e nasais laterais. A - Na quinta semana de gestação, a proeminência frontonasal (PFn), as proeminências nasais mediais (PNM), proeminências nasais laterais (PNL) e proeminências maxilares (PMx) já estão formadas. B - Essas estruturas crescem e se reorganizam. C - As PNM e PNL se aproximam e se fundem para iniciar a formação do nariz. D - As proeminências nasais já fundidas se aproximam das PMx. E - Inicia-se a fusão das proeminências nasais e das PMx. F - O lábio superior e o palato primário se formam, então, na décima semana de gestação. G - Se houver algum distúrbio, poderá não ocorrer essa fusão, formando a fissura do lábio superior (Adaptado de <http://www.indiana.edu/~anat550/hnanim/face/face.html>) (Jiang *et al.*, 2006).

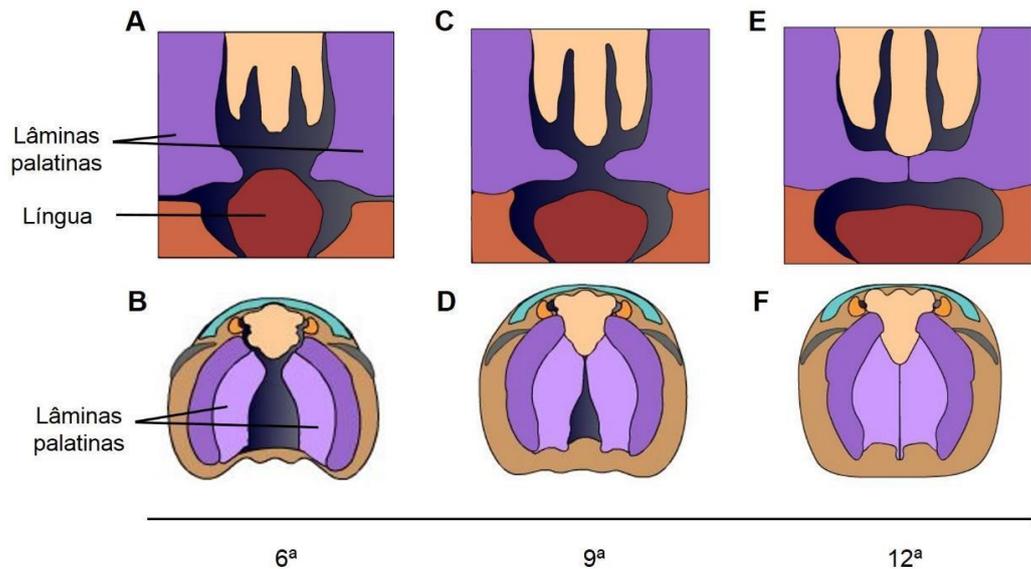


Figura 2 - Formação do palato secundário. A e B - Lâminas palatinas dispostas verticalmente e paralelas, separadas pela língua, na sexta semana de gestação. Corte longitudinal (A) e Corte transversal/Vista ventro-dorsal (B). C e D - As lâminas palatinas crescem e se elevam em uma posição horizontal, enquanto a língua desce. Corte longitudinal (C) e Corte transversal/Vista ventro-dorsal (D). E e F - As lâminas palatinas se fundem, formando o palato secundário, permitindo uma separação entre cavidade oral e nasal. Corte longitudinal (E) e Corte transversal/Vista ventro-dorsal (F) (Adaptado de <http://www.indiana.edu/~anat550/hnanim/face/face.html>) (Gritli-Linde, 2007).

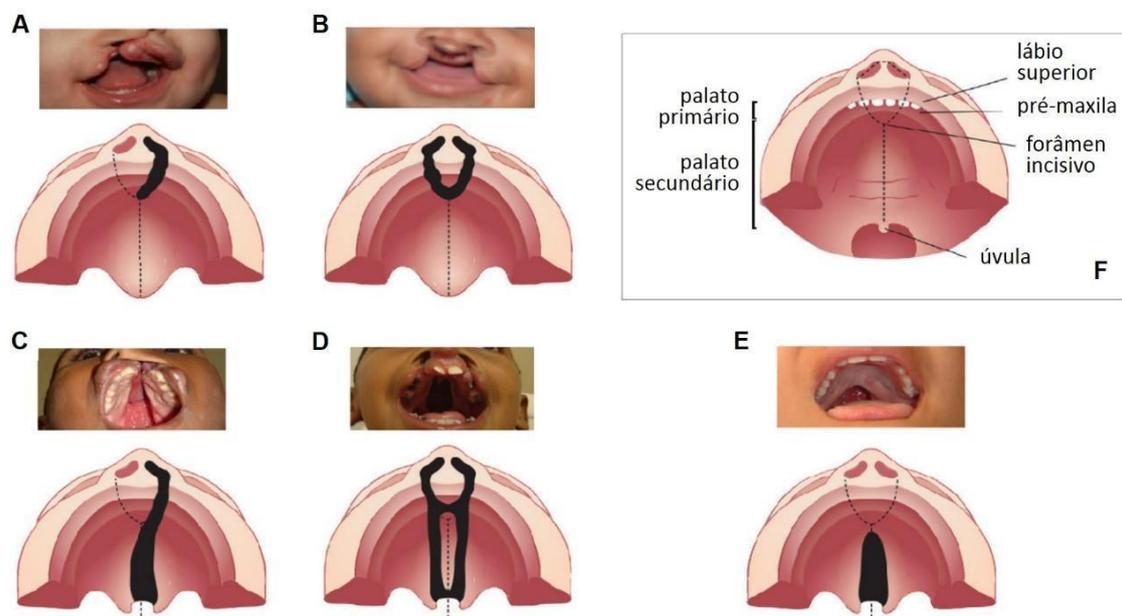


Figura 3 - Variabilidade fenotípica da FLP. A - Fissura labial unilateral. B - Fissura labial bilateral. C - Fissura labial e palatina unilateral. D - Fissura labial e palatina bilateral. E - Fissura palatina. F - Vista ventro-dorsal do interior bucal (Adaptado de Brito *et al.*, 2012b).

A frequência da FL/P varia de acordo com a população. Em africanos, é de cerca de 1 em cada 2.500 nascimentos, sendo a mais baixa do mundo, em europeus é de cerca de 1:1.000 e em Asiáticos e em Ameríndios é de cerca de 2 a cada 1.000 nascimentos, sendo a mais alta (Dixon *et al.*, 2011; Marazita, 2012; Taib *et al.*, 2015). No Brasil, varia de 0.28 a 1.54 a cada 1000 nascimentos (Menegotto & Salzano, 1991; Loffredo *et al.*, 2001, Rodrigues *et al.*, 2009). A FP apresenta uma incidência racial uniforme de 0.5:1000 nascimentos (Taib *et al.*, 2015). Observa-se um aumento da prevalência de FL/P em populações com baixo nível socioeconômico (Dixon *et al.*, 2011; Murray *et al.*, 1997) e diferenças na prevalência de acordo com o sexo, sendo FL/P mais prevalente no sexo masculino e FP, no sexo feminino, em diversos grupos étnicos (Mossey *et al.*, 2009; Murray *et al.*, 1997).

Historicamente, as FL/Ps têm sido maior foco de estudos do que FP, possivelmente devido ao maior número de casos e maior facilidade de identificação (Marazita, 2012; Dixon *et al.*, 2011), sendo também as FL/Ps o foco deste trabalho. Dessa forma, os próximos tópicos serão focados nesse grupo de fissura.

1.2. Fatores genéticos da fissura do lábio com ou sem fissura do palato (FL/P)

A fissura do lábio com ou sem fissura do palato (FL/P) pode ser síndrômica, ou seja, pode estar associada a outras alterações clínicas, ou não-síndrômica, em que a fissura se apresenta como fenótipo isolado.

1.2.1. FL/Ps síndrômicas

As FL/Ps síndrômicas representam 30% dos casos de FL/P, sendo caracterizadas pela presença da fissura do lábio com ou sem palato associada a outras alterações clínicas (Leslie & Marazita, 2013; Jugessur *et al.*, 2009). Podem ocorrer devido a alterações cromossômicas, como deleções, duplicações ou trissomias, ou devido a alterações monogênicas (Stuppia *et al.*, 2011; Mitchell, 2002), como por exemplo a SVW e a Síndrome do pterígeo poplíteo (PPS), ambas causadas por mutações no gene *IRF6* (*Interferon Regulatory Receptor 6*) (Kondo *et al.*, 2002). Tais síndromes apresentam fenótipos adicionais, além da fissura. A SVW é caracterizada pela presença de fossetas no lábio inferior e da fissura do lábio e/ou do palato, enquanto que a PPS é caracterizada

por anomalias cutâneas, musculoesqueléticas e genitais, além da fissura do lábio e/ou do palato (Stuppia *et al.*, 2011).

A identificação de mutações nas formas sindrômicas tem contribuído para a compreensão da etiologia das formas não-sindrômicas (Stanier & Moore, 2004), podendo diferentes variantes no mesmo gene estarem atuando nas duas formas, como por exemplo, variantes genéticas raras (variantes com frequência menor que 1%) sem sentido e do tipo não-sinônima no gene *IRF6* são causativas de SVW e variantes comuns (variantes com frequência maior que 1%) na região promotora desse mesmo gene são de susceptibilidade às FL/P não-sindrômicas (Kondo *et al.*, 2002; Zuccherro *et al.*, 2004; Beaty *et al.*, 2010; Brito *et al.*, 2012c; Paranaíba *et al.*, 2008).

1.2.2. FL/Ps não-sindrômicas

As FL/Ps não-sindrômicas representam os casos de fissura do lábio com ou sem palato, sem outras alterações clínicas associadas, e correspondem a cerca de 70% dos casos (Jugessur *et al.*, 2009; Zuccherro *et al.*, 2004).

A observação de que há pelo menos 2 pacientes com fissura labial com ou sem fissura do palato não sindrômica (FL/P NS) em uma mesma família em 20-30% dos casos (Marazita, 2012; Grosen *et al.*, 2010), a maior concordância de FL/P entre gêmeos monozigóticos (40-60%) do que em dizigóticos (3-5%) na Dinamarca (Jugessur *et al.*, 2009; Christensen & Fogh-Andersen, 1993) e uma estimativa alta da herdabilidade, sendo na população brasileira de 45 a 85% (Brito *et al.*, 2011), são fortes evidências de que a ocorrência da FL/P NS depende de uma contribuição genética.

Com base nessas evidências, tem-se sugerido que, na maioria dos casos, o modelo genético mais provável da FL/P NS é o de herança multifatorial (Grosen *et al.*, 2010; Dixon *et al.*, 2011). O padrão de herança multifatorial depende dos fatores genéticos associados a fatores ambientais, atuando durante a embriogênese, para a manifestação da doença. Como exemplos de fatores ambientais, foi observada uma proporção maior de mães que fizeram uso de álcool, cigarro e/ou tiveram deficiência de ácido fólico durante a gravidez que tiveram filhos afetados do que em mães que não utilizaram essas substâncias (Kelly *et al.*, 2012; Crossan & Duane, 2018; Xu *et al.*, 2018,

Wilcox *et al.*, 2007). Contudo, pouco se sabe sobre o mecanismo de ação de tais fatores e como esses interagem com os fatores genéticos de predisposição.

O modelo de herança multifatorial é o de limiar de predisposição (Falconer, 1965; Mitchell, 2002). Neste modelo, os indivíduos com a doença multifatorial apresentam um acúmulo de alelos de risco, os quais são alelos comuns (variantes presentes em mais que 1-5% da população) e de efeito baixo na determinação do fenótipo. Dessa forma, a manifestação da doença depende de ultrapassar esse limiar. Os indivíduos da população que não atingem esse limiar são clinicamente normais (Falconer, 1965; Manolio *et al.*, 2009).

O método que mais tem sido utilizada para estudar as variantes comuns de risco é o *Genome-Wide Association Study* (GWAS), que parte do princípio que a doença comum é causada por um acúmulo de variantes de pequeno efeito e frequentes na população, isto é, “doença comum, variante comum” (Reich & Lander, 2001; Schork *et al.*, 2009), que correspondem às mesmas variantes no modelo de limiar de predisposição acima citado. O GWAS não necessita de conhecimento prévio dos genes que serão incluídos no estudo (Kruglyak, 2008).

Um exemplo de variante comum que foi identificado por estudos de GWAS e que confere risco em pacientes com FL/P NS é a variante de base única ou SNV (do inglês *Single-Nucleotide Variation*) rs957525 C>A na posição 130015336 da região 8q24.21, que é uma região de deserto gênico (Birnbaum *et al.*, 2009). A associação entre esse SNV e FL/P foi replicada em várias populações, incluindo a brasileira (Brito *et al.*, 2012a).

Além da região 8q24.21, estudos de GWAS detectaram outras regiões associadas à FL/P NS como 1p22.1 e 20q12 nas populações europeia e asiática (Beatty *et al.*, 2010) e 2p24.2, 4p16.2, 5p12, 8p11.23, 17q23, 19q13, entre outras, em estudos multi-étnicos (Leslie *et al.*, 2016; Yu *et al.*, 2017). Já foram confirmadas na população brasileira variantes comuns em 8q24.21, 2p24.2 e 1p22.1 (Brito *et al.*, 2012a; Machado *et al.*, 2018, Bagordakis *et al.*, 2013).

Entretanto, as variantes encontradas até o momento por GWAS não são suficientes para explicar a alta herdabilidade das FL/Ps, o que tem sido chamado de

“herdabilidade perdida”. Uma hipótese é que as variantes que provavelmente não estão sendo detectadas sejam variantes de efeito médio ou grande que, por serem deletérias, são frequentemente removidas pela seleção natural e, conseqüentemente, tornam-se raras na população.

A principal metodologia que tem sido usada para a identificação e estudo das variantes raras é o sequenciamento de nova geração (NGS), dentre elas, a análise de sequenciamento completo de exoma (SCE). Esta técnica busca por mutações na região codificante do genoma, local em que se espera que se tenha uma conservação maior da seqüência de nucleotídeos. Logo, se houver alterações nessa região, espera-se que seja de grande ou médio efeito para o organismo (Tetreault *et al.*, 2015).

Um exemplo bem-sucedido de aplicação de SCE por NGS é a identificação de mutações raras no gene *CDH1* em casos familiares de FL/P NS. O *CDH1* (*Cadherin 1*) codifica caderina epitelial que promove adesão celular entre células epiteliais vizinhas e é superexpresso durante estágios críticos do desenvolvimento do lábio e do palato (Freboung *et al.*, 2006). Brito *et al.* (2015) encontrou variantes do tipo *missense*, *stopgain* e *splicing* em *CDH1* em pacientes brasileiros. Nessas famílias, a segregação da FL/P é compatível com o padrão de herança autossômico dominante com penetrância incompleta, em que as variantes se encontram no estado heterozigoto. A penetrância incompleta, que foi estimada em menos de 50%, sugere que outros fatores genéticos ou ambientais são necessários. Esses casos também podem representar um modelo de herança oligogênica.

Além de mutações em *CDH1*, por meio da análise de SCE, foram detectadas variantes raras que contribuem para a etiologia das FL/Ps NS, em outros genes, como por exemplo, variantes de perda de função no gene *ARHGAP29* em pacientes filipinos, americanos, indianos, ingleses e brasileiros (Leslie *et al.*, 2012; Chandrasekharan & Ramanathan, 2014; Savastano *et al.*, 2017). Também foram encontradas em amostras multiétnicas, variantes do tipo *missense*, *stopgain*, *nonsense* e de *splicing* no gene *CTNND1* e do tipo *missense* nos genes *PLEKHA5*, *PLEKHA7* e *ESRP2* (Cox *et al.*, 2018) e *frameshift insertion* em *TP63* (Basha *et al.*, 2018).

Identificar variantes raras de efeito médio ou grande bem como aquelas de efeito pequeno é fundamental para a caracterização da arquitetura genética associada às FL/Ps, bem como para o aconselhamento genético. Nos casos de variante rara de efeito médio ou de grande, o risco de novos afetados nessas famílias tende a ser bem maior do que nas famílias em que segregam variantes de efeito pequeno. As variantes raras poderiam explicar a alta recorrência de afetados em algumas famílias. Entretanto, há dificuldades na priorização dessas variantes, pois, no caso de variantes ainda não descritas, não sabemos se elas são, de fato, patogênicas e, quando o gene nunca havia sido antes associado a doença, há a necessidade de replicação de dados e realização de estudos funcionais.

Caso as variantes não sejam detectadas por meio de SCE, um dos motivos pode ser devido à presença de mais de uma variante estar segregando em uma família (heterogeneidade genética). Outro motivo poderia ser por esse tipo de sequenciamento (SCE) cobrir apenas as regiões exônicas e muitas variantes de risco para FL/P podem estar localizadas na região não-codificadora. Um método possível para testar essa hipótese seria o sequenciamento completo do genoma, o que ainda tem um custo relativamente elevado. Um método alternativo no estudo de casos familiares que cobre também as regiões intrônicas é o estudo de ligação, que se baseia na co-segregação do loco genético associado com a doença entre os afetados da família (Marazita *et al.*, 2004). Nesses casos, quanto maior o número de indivíduos da família maior a chance de identificar a região que possa estar ligada a doença (Manolio *et al.*, 2009). O estudo de ligação já foi bastante utilizado na busca de candidatos para FL/P NS em famílias com alta recorrência de afetados antes da técnica de SCE, o qual revelou as seguintes regiões candidatas: 1q32, 2p13, 3q27-28, 9q21, 14q21-24 e 16q24 (Marazita *et al.*, 2004). Entretanto, nenhuma variante foi confirmada ainda nessas regiões.

1.3. Epigenética

Pesquisadores acreditam que uma das formas que os fatores ambientais podem influenciar na expressão gênica dos indivíduos seja por meio de modificações químicas na molécula de DNA ou em histonas, sem a alteração da sequência do DNA, ou seja, alterando o padrão epigenético (Tammen *et al.*, 2013; Jaenish & Bird, 2003).

Um experimento clássico que exemplifica tal afirmação está relacionado com a cor de pelagem da prole de camundongo conforme a alimentação materna. A fêmea gestante pode ter uma prole com a cor de pelagem que pode ser amarela, pseudo-aguti ou mosaico, mas quando a mesma fêmea é exposta a suplementos maternos ou compostos tóxicos, a cor da pelagem de sua prole é, em sua maioria, amarela ou pseudo-aguti, respectivamente. A explicação para esses fenótipos é que há um sítio próximo à região promotora do gene que determina a cor de pelagem que, se estiver hipermetilado (influenciado no experimento pela presença de compostos tóxicos), o fenótipo do camundongo será pseudo-aguti e, se estiver hipometilado (influenciado pela exposição aos suplementos maternos), a cor da pelagem será amarela (Feil & Fraga, 2012).

As alterações epigenéticas são um mecanismo dinâmico e natural da regulação gênica que ocorrem ao longo da vida, a depender do hábito de vida e do ambiente social que o indivíduo vive, ou podem ser herdadas por mitose ou meiose. Um exemplo conhecido de herança epigenética é o processo de inativação do cromossomo X (Feil & Fraga, 2012; Tammen *et al.*, 2013; Calvanese *et al.*, 2009; Skvortsova *et al.*, 2019). Uma vez que cada tecido expressa os genes de diferentes formas e o organismo passa por alterações ao longo da vida, o padrão de metilação é variável para tecidos diferentes de um mesmo organismo (Masser *et al.*, 2013) e de acordo com a idade da pessoa (Tammen *et al.*, 2013; Calvanese *et al.*, 2009).

A molécula de DNA pode apresentar alteração do padrão epigenético por meio da metilação, isto é, pela inserção do grupo metila no carbono 5' do nucleotídeo citosina que estiver localizada em ilhas CpG (citosina-fosfato-guanina). As ilhas CpG são regiões densas em citosinas e guaninas e normalmente estão associadas a regiões promotoras do gene. Devido à sua localização e ao recrutamento de proteínas que se associam às citosinas metiladas, se ocorrer uma hipermetilação nessa região, o promotor em geral se tornará inacessível e, conseqüentemente, não será transcrito. Dessa forma, ocorrerá a inativação do gene (Tammen *et al.*, 2013; Masser *et al.*, 2013; Calvanese *et al.*, 2009; Horsburgh *et al.*, 2015). Também já foi observado que a metilação de DNA pode ativar a transcrição do gene, possivelmente porque, ao se ligar ao promotor, evita a ligação de proteínas repressoras, permitindo a sua transcrição (Halpern *et al.*, 2014).

As modificações químicas também podem ocorrer nas histonas, proteínas que compactam ou descompactam o DNA, encobrendo ou expondo o gene, respectivamente e, conseqüentemente, alterando sua expressão gênica. Alguns exemplos de modificações são a acetilação ou metilação. A acetilação é a inserção do grupo acetila nos aminoácidos lisina e arginina, que leva à ativação do gene pela neutralização de cargas positivas dos aminoácidos, diminuindo a atração entre as histonas, e a metilação é a inserção do grupo metila em aminoácidos que, dependendo da histona metilada, pode levar à ativação ou à inativação do gene pelo recrutamento de proteínas para a cromatina (Horsburgh *et al.*, 2015; Mochizuki *et al.*, 2018).

1.4. Fatores Genéticos X Epigenética nas FL/Ps NS

Mais recentemente, pesquisadores vêm estudando se a alteração epigenética, influenciada por fatores externos, tem alguma participação na alteração do risco de doenças, sendo o câncer a doença mais estudada (Skvortsova *et al.*, 2019).

Foi observado que alguns pacientes com câncer gástrico difuso possuíam mutação em um dos alelos do gene *CDH1*, mesmo gene que apresenta variantes de perda de função causativas para FL/P NS. Existem, inclusive, alguns trabalhos que relatam casos de FL/P e câncer gástrico difuso na mesma família com variantes patogênicas em *CDH1* (Freboureg *et al.*, 2006; Benusiglio *et al.*, 2013). Em alguns casos de câncer gástrico difuso, foi observado que indivíduos heterozigotos apresentavam a doença. Ao verificarem o promotor do *CDH1* nestes pacientes, foi encontrada uma hipermetilação, que levava a um silenciamento do gene. Dessa forma, a explicação proposta foi que o modelo desta doença era de 2 *hits*. O primeiro *hit* seria, então, a variante patogênica na linhagem germinativa em um dos alelos de *CDH1* e o segundo *hit*, a hipermetilação do outro alelo (Grady *et al.*, 2000; Machado *et al.*, 2001).

O mesmo mecanismo foi sugerido e testado em pacientes com FL/P NS com mutação em *CDH1* que, de forma semelhante ao câncer gástrico difuso, tem padrão de herança autossômico dominante com penetrância incompleta (Brito *et al.*, 2015; Alvizi *et al.*, 2017). Foi verificado que os pacientes exibiam um nível de metilação no promotor do gene significativamente superior aos não-penetrantes com variante patogênica em

CDH1 e aos controles (Figura 4). Dessa forma, foi proposto que, além da variante patogênica rara no gene *CDH1*, é necessário um segundo *hit* para que a FL/P se manifeste, como por exemplo, a metilação ou silenciamento do promotor deste gene. A não ocorrência deste silenciamento, que representaria um segundo *hit*, poderia explicar os casos de penetrância incompleta (Brito *et al.*, 2015; Alvizi *et al.*, 2017).

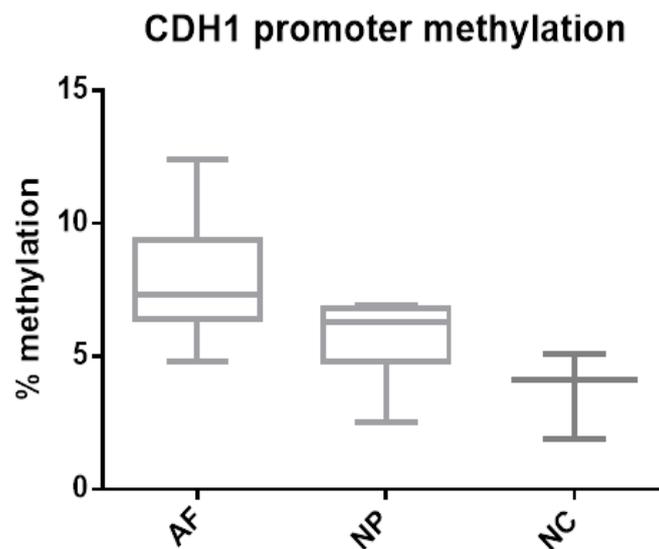


Figura 4 - Porcentagens de metilação de pacientes (AF), familiares não-penetrantes (NP) e controles da população (NC). A variante patogênica em *CDH1* foi observada nos familiares afetados e nos não-penetrantes (Alvizi *et al.*, 2017).

Alvizi *et al.* (2017) também observou, a partir de análise de metiloma, um padrão de metilação distinto e enriquecido em regiões potencialmente regulatórias e ativas no genoma e em regiões candidatas, especialmente em genes da via de transição epitélio-mesenquimal e da via Wnt, em pacientes com FL/P NS comparado aos controles na população brasileira. Além disso, o mesmo trabalho encontrou uma correlação entre os níveis de metilação de tecido de lábio e de sangue, o que indica que o sangue pode ser utilizado para estudo de análise de metilação para FL/P.

Em comparação de níveis de metilação entre pacientes com vários subtipos de FLP (fissura do lábio, fissura do palato e fissura do lábio e do palato), foi observado que o padrão de metilação era similar entre os pacientes com fissura de lábio e fissura de

lábio e do palato (as FL/Ps) e ambos eram diferentes de fissura do palato (FP), o que corrobora com a diferença de etiologia de FL/P e FP (Sharp *et al.*, 2017).

Gonseth *et al.* (2019) observou uma tendência à hipometilação em todo o epigenoma de pacientes com FL/P próximos a regiões ativas e a genes previamente associados a FL/P. O estudo foi realizado em amostras de pacientes anteriores a fortificação obrigatória de folato na dieta nos Estados Unidos. Dessa forma, o trabalho sugeriu que a hipometilação pode ser devido a uma quantidade baixa de doadores do grupo metila, resultante da deficiência de folato, e que alterações na metilação de loci epigenômicos sensíveis à nutrição poderiam explicar o mecanismo pelo qual o risco de FL/P é responsivo a suplementação materna de folato.

Tais trabalhos têm demonstrado a importância do estudo de metilação em pacientes com FL/P NS.

1.5. Hipóteses do presente trabalho

Diante da revisão da bibliografia exposta acima, apesar das variantes raras serem um mecanismo promissor para explicar a arquitetura genética das FL/P NS, uma das limitações importantes é a priorização das variantes de genes candidatos, particularmente aqueles com pouca evidência de poderem estar associadas a essa malformação, ou seja, aquelas variantes em genes candidatos novos, nunca antes sugeridos como associados à FL/P. Uma outra limitação é identificar variantes patogênicas do tipo *missense* e sinônimas. Apesar dos vários programas de predição *in silico*, na maioria das vezes são necessários estudos funcionais, o que é inviável dado o grande número de possíveis variantes patogênicas observadas nos estudos genômicos. Tendo em vista que os casos de FL/P NS com padrão de herança compatível com o autossômico dominante, na sua grande maioria, tem penetrância incompleta, tem-se sugerido que é necessário mais um evento (genético ou epigenético) para a manifestação do quadro clínico, como sugerimos em nosso estudo preliminar (Alvizi *et al.*, 2017), onde observamos que o padrão de metilação da região promotora no *CDH1* dos afetados heterozigotos para variante patogênica em *CDH1* é maior do que nos

indivíduos heterozigotos com a mesma variante, mas sem manifestação clínica, ou nos controles (sem variante patogênica em *CDH1* e clinicamente normais).

Diante desses achados, a hipótese principal desse trabalho é que genes ou regiões candidatas à FL/P NS têm maior probabilidade de terem alteração no padrão de metilação nos afetados do que em não-afetados.

2. Objetivos gerais e específicos

2.1. Objetivos gerais

O objetivo geral do presente trabalho é testar se os genes ou regiões candidatas à FL/P NS selecionados por diferentes metodologias têm maior probabilidade de terem alteração no padrão de metilação nos afetados do que em não-afetados. Foram selecionados alguns genes já conhecidos como causativos de FL/P, como *ARHGAP29*, *TP63*, *SHH* e novos genes candidatos.

2.2. Objetivos específicos

- a) Selecionar genes em famílias inglesas e brasileiras com pelo menos 2 pacientes (casos multiplex) com FL/P NS. Os genes candidatos foram selecionados após análise de sequenciamento completo de exoma, estudo de ligação ou sequenciamento de painel de genes candidatos;
- b) Verificar se há diferença no padrão de metilação entre pacientes e controles para os genes selecionados, comparando:
 - i) amostra total de pacientes e controles para todos os genes analisados e;
 - ii) pacientes de uma mesma família que apresentam variante patogênica para o gene analisado e controles.

3. Material e Métodos

3.1. Casuística e ética

As amostras de DNA foram extraídas a partir de sangue (coletado com tubos EDTA BD Vacutainer; Becton, Dickison and Company, Bergen County, NJ, EUA) ou de saliva (coletado com *Oragene® DNA Collection Kits OG-500 e OG-575*; DNA Genotek, Ottawa, ON, Canada) e purificados de acordo com protocolos de fabricantes.

Esse estudo incluiu dados de sequenciamento das seguintes casuísticas:

a) Dados de SCE de amostras de DNA de pacientes do *Great Ormond Street Hospital for Children*, que foram obtidos em parceria com o Dr. Philip Stanier da *University College London* (Londres, Reino Unido). As amostras foram provenientes de um total de 7 indivíduos, membros de 3 famílias inglesas multiplex (Famílias 281, BC84-85 e 313), sendo 6 pacientes com FL/P NS e 1 familiar não-afetado.

b) Dados de SCE de 9 famílias multiplex e de estudo de ligação (1 família multiplex, a qual foi também incluída no estudo de SCE), atendidos durante missões da Operação Sorriso dos Estados do Ceará, Pará e Rio de Janeiro, no Hospital das Clínicas da Universidade de São Paulo e na Sobrapar (Sociedade Brasileira de Pesquisa e Assistência para Reabilitação Crânio-facial - Campinas). Esses dados estão disponíveis na tese de doutorado de Brito (2016).

Uma vez selecionados os genes candidatos, foi então realizada análise de metilação da região promotora dos mesmos em uma amostra de pacientes e controles; cumpre ressaltar que os pacientes selecionados para essa análise não necessariamente tinham variante genética nos genes candidatos. Diante desse critério, para a análise de metilação, foram estudados 28 pacientes (18 com variantes em genes candidatos analisados nessa dissertação e 10 sem variantes candidatas), 3 heterozigotos para variantes de risco em genes candidatos analisados nessa dissertação e clinicamente normais, 21 não-afetados pertencentes às famílias F1843, F2570, F7614, F8418, F3196 e F2848 e 47 controles. Todos os pacientes e controles são da população brasileira. Apenas

os 18 pacientes com variantes candidatas das famílias F1843, F2570, F7614, F8418, F3196 e F2848 (selecionadas por Brito, 2016) possuíam amostras de familiares, permitindo análises por família. A casuística mais detalhada para essas famílias está explicada no item 3.3.1.

O projeto que resultou na presente dissertação contou com o parecer consubstanciado do Comitê de Ética em Pesquisa do Instituto de Biociências da Universidade de São Paulo (Número 363.876/2013).

3.2. Abordagens utilizadas para seleção de genes a serem incluídos na análise de metilação

Para se verificar se há diferença no padrão de metilação nos genes relacionados à FL/P NS, foi feita uma seleção dos genes candidatos priorizados após SCE, presentes em painel de genes candidatos ou por meio de estudo de ligação. Os passos realizados neste trabalho estão representados na Figura 5. Cada etapa será melhor explicada nos tópicos seguintes.

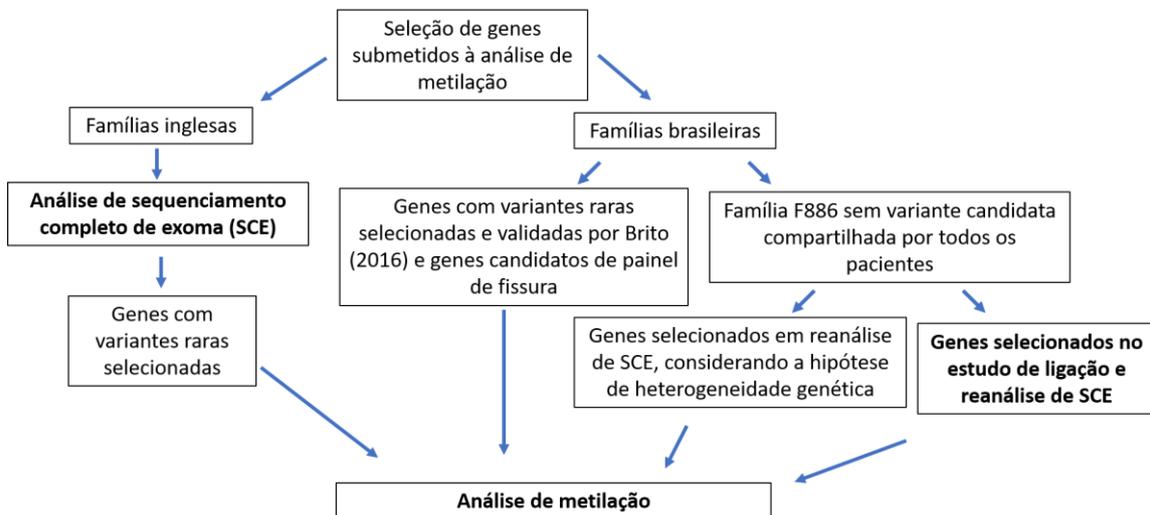


Figura 5 - Esquema simplificado do trabalho realizado. Em negrito estão destacadas as etapas que foram efetivamente realizadas no presente trabalho.

3.2.1. Seleção de genes por análise de SCE de famílias inglesas

Para investigar genes com variantes raras e patogênicas em casos familiares de pacientes com FL/P NS, foi feita uma busca nos dados de SCE de 7 indivíduos de 3 famílias inglesas (Figura 6).

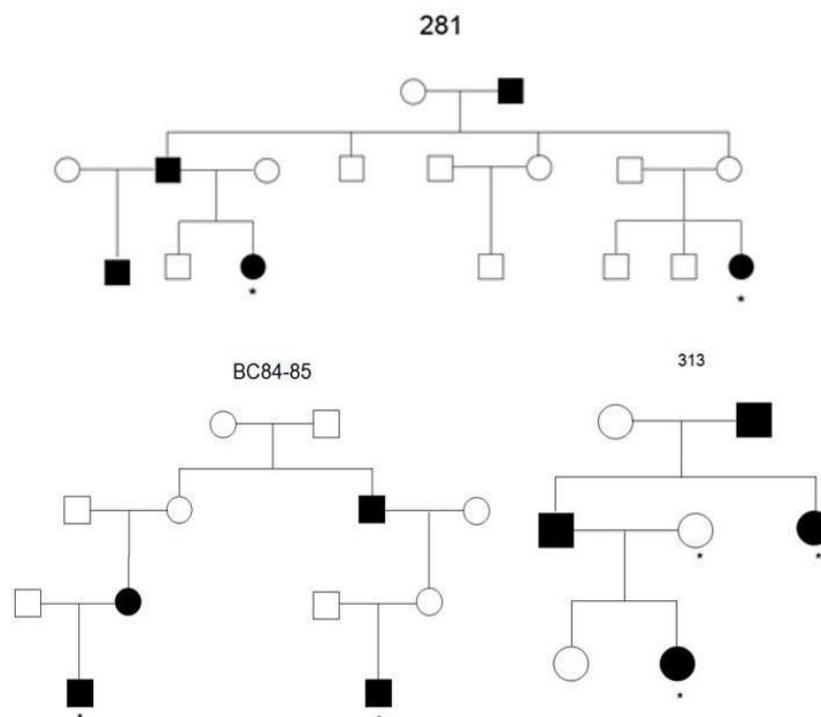


Figura 6 - Representação dos heredogramas das famílias inglesas. Os asteriscos indicam os pacientes e o familiar que foram analisados neste trabalho. Note que o familiar selecionado não pertence à família em que ocorre a manifestação da doença, podendo ser considerado um controle da população, permitindo uma filtragem mais confiável das variantes, pois possivelmente não é portador da variante rara procurada.

O SCE foi realizado pela *University College London* utilizando o *Agilent Exome v4 51 Mb Capture Technology* e, para o preparo da biblioteca, foi utilizado o *Illumina HiSeq2000 (Illumina)*. O alinhamento de sequência, o processamento de dados, as chamadas e as anotações de variantes foram realizados com *Burrows-Wheeler Aligner* (<http://bio-bwa.sourceforge.net/>), *Picard* (<http://broadinstitute.github.io/picard/>), *Genome Analysis Toolkit Package* (<https://www.broadinstitute.org/gatk/>) e *ANNOVAR* (<http://www.openbioinformatics.org/annovar/>), respectivamente.

Para o presente trabalho, a partir dos dados gerados por SCE, foi feita uma investigação de variantes raras por cada família pelo método de filtragem e priorização

de variantes, utilizando a abordagem de acordo com Brito (2016). As famílias foram analisadas de forma individual. Foram excluídas variantes com frequência maior que 1%, utilizando informações de bancos de dados como *1000 Genomes Project* (<http://www.internationalgenome.org/NHLBI/>), *Exome Sequencing Project (ESP6500)* (<http://evs.gs.washington.edu/EVS/>), *Exome Aggregation Consortium (ExAC)* (<http://exac.broadinstitute.org/>) e ABraOM (Naslavsky *et al.*, 2017), e variantes que não apresentassem funções exônicas ou de *splicing*. Foi feita a comparação de paciente com familiar não-afetado, quando havia ambos, esperando-se que a variante rara e patogênica estivesse presente apenas no paciente, e a comparação entre os pacientes da mesma família para buscar variantes presentes em ambos. Foram excluídas as variantes que apresentassem má qualidade de genotipagem e variantes com desproporção alélica, ou seja, se estivessem em heterozigose, o que se esperaria era um número de *reads* semelhante para cada alelo e, se em homozigose, o número de *reads* do alelo presente deveria ser majoritário em relação ao outro. Foram excluídas também as variantes sinônimas e as localizadas em genes com alta variação alélica ou polimorfismos (Fajardo *et al.*, 2012).

Após aplicar os filtros, as variantes que restaram foram melhor estudadas e priorizadas de acordo com critérios como função, expressão e predição de patogenicidade. Esta etapa foi realizada por meio das ferramentas *VarElect* (<http://varelect.genecards.org/>) para identificar os genes que estivessem diretamente ou indiretamente relacionados à FL/PNS, utilizando os termos “*cleft lip*” ou “*cleft lip with or without palate*”, *Mouse Genome Informatics* (<http://www.informatics.jax.org/>) para verificar os fenótipos encontrados em camundongos causados por mutações nos genes estudados e se tais fenótipos estavam relacionados com malformações craniofaciais ou mortes prematuras, *Sysface* (<http://bioinformatics.udel.edu/Research/SysFACE/>) para verificar se ocorre a expressão dos genes durante o desenvolvimento de estruturas craniofaciais, como região frontonasal, mandíbula, maxila e palato em humanos, e *ExAC Browser* (<http://exac.broadinstitute.org/>) para verificar a tolerância a mutações esperada para os genes. Para verificar a conservação e predição de danos em proteínas expressas pelos genes estudados, foram utilizadas as ferramentas *SIFT* (<http://sift.jcvi.org/>) e *Polyphen-2* (<http://genetics.bwh.harvard.edu/pph2/>). Além

destas ferramentas, os genes priorizados pelos critérios supracitados foram também estudados utilizando os bancos de dados *Pubmed* (www.ncbi.nlm.nih.gov/pubmed) e *OMIM* (<http://www.omim.org/>) para se obter um melhor conhecimento de suas funções no organismo, associação desses genes com doenças raras e verificar a existência de publicações anteriores ou não, relacionando o gene em questão à FL/P ou à formação craniofacial.

Os genes selecionados após a filtragem e priorização de variantes foram incluídos na análise de metilação.

3.2.2. Seleção de genes em famílias brasileiras

- *Seleção de genes por SCE (Brito, 2016)*

O SCE de 9 famílias brasileiras multiplex levou à priorização (conforme método de filtragem e priorização de variantes citados no item 3.2.1) de 9 genes candidatos para FL/P NS, nos quais foram encontradas 10 variantes raras potencialmente patogênicas, sendo 8 do tipo *missense*, 1 *stopgain* e 1 *frameshift deletion* em 21 pacientes de 6 famílias brasileiras. Tais genes foram previamente selecionados na tese de Brito (2016) utilizando os mesmos critérios de análise de SCE descritos acima e foram incluídos na análise de metilação.

- *Seleção de genes candidatos de painel de FL/P NS*

Foram selecionados 5 genes presentes em um painel de genes candidatos a FL/P NS definidos pelo nosso grupo de pesquisa (Faria, 2019): dois desses genes estão associados a formas sindrômicas de fissura (*TP63* e *SHH*), um associado a forma não-sindrômica (*ARHGAP29* - Savastano *et al.*, 2017; Leslie *et al.*, 2012; Chandrasekharan & Ramanathan, 2014) e dois outros genes candidatos selecionados a partir de análise de transcriptoma de células-tronco de pacientes com FL/P NS (*BRCA1* - Kobayashi *et al.*, 2013) e de estudo de loci de características quantitativas de expressão (eQTL) (*NOL8* - Masotti *et al.*, 2018).

- *Reanálise de SCE da família F886*

Uma das famílias brasileiras, F886 (Figura 7), estudada por Brito (2016), não apresentou nenhuma variante candidata compartilhada por todos os pacientes na análise de SCE, considerando-se um padrão de herança autossômico dominante.

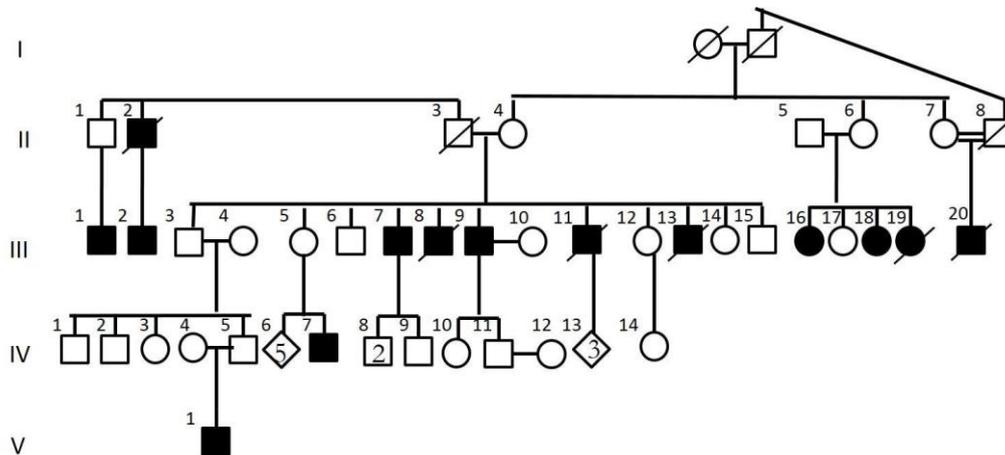


Figura 7 – Heredograma da família F886.

A análise dos afetados dessa família sugere que os indivíduos II-3 e II-4 (Figura 7) sejam provenientes de famílias distintas e ambos possuem familiares com FL/P NS. Por este motivo, é possível que haja heterogeneidade genética nessa família.

Dessa forma, ao longo das realizações deste trabalho, o pós-doutorando Dr. Luciano Abreu Brito reanalisou o SCE de 6 pacientes da família F886 (III-7, III-9, III-18, III-19, III-20 e V-1), utilizando o método de filtragem e de priorização de variantes descrito na Seção 3.2.1, com exceção de que, nesta reanálise, não foram analisados todos os pacientes simultaneamente para procurar alguma variante que estivesse presente em todos. A análise foi feita por pares de indivíduos, supondo que ao menos dois pacientes da família apresentassem a mesma variante rara e os demais apresentassem outra variante distinta. Os genes das variantes priorizadas nesta nova análise também foram incluídos na análise de metilação.

- *Estudo de ligação da família F886 e reanálise dos dados de SCE*

Uma vez que a família F886 apresenta um grande número de afetados, também foi realizado um estudo de ligação com esta família na tentativa de buscar uma região

cromossômica em comum (haplótipo) que esteja segregando com o fenótipo de FL/P NS nessa família.

Para tal análise, realizou-se *SNP-array* de 6 pacientes (III-7, III-9, III-18, III-19, III-20 e V-1) e 9 familiares não-afetados (III-3, III-4, IV-1, IV-2, IV-4, IV-5, IV-9, IV-10 e IV-11) por meio da lâmina Axiom_KP_UCSF_LAT da Plataforma *Affymetrix* para 818.154 SNPs (hg19). Para a análise de ligação, utilizou-se o valor da penetrância de 0.46 para esta família, o qual foi calculado com a colaboração do Professor Dr. Paulo Otto (IB-USP).

Para a análise de ligação, foi utilizado o software *Im_linkage* do pacote Morgan, que apresentou valores de LOD score para interpretação de quais regiões cromossômicas apresentam chances de estarem associadas ao fenótipo, representados através de gráficos usando o *software* R versão 3.4.2.

No presente trabalho, foram consideradas significativas quaisquer regiões que apresentassem LOD Score acima ou próximo de 2.0 (probabilidade de haver ligação 100x maior do que a probabilidade de não haver ligação) que representa o valor máximo de LOD Score para essa família e dos indivíduos que dispomos de material genético.

O preparo dos arquivos para análise do estudo de ligação foi realizado pelo Me. Vinícius Magalhães Borges e pelo Dr. Renan Barbosa Lemes.

Após delimitar as regiões com LOD Score acima ou próximo de 2.0, os genes presentes nessas regiões foram pesquisados no *software* *BioMart* (<https://www.ensembl.org/biomart/martview/1753578ecd8ff9300111955ed8abd48f>) e os genes encontrados foram estudados e avaliados, a partir de critérios de função (bancos de dados *Pubmed* e *OMIM*) e de expressão (*softwares* *MGI* e *Sysface*) relacionados a formação craniofacial ou FLP.

Os genes selecionados no estudo de ligação foram pesquisados dentre as variantes raras compartilhadas por pelo menos dois indivíduos na reanálise da família F886.

As variantes encontradas foram validadas por sequenciamento pelo método de Sanger, que foi realizado com o *BigDye Terminator Cycle Sequencing Kit* (Thermo Fisher). Os produtos de PCR de interesse foram submetidos à eletroforese capilar no sequenciador ABI3730 DNA Analyzer (*Applied Biosystems*) do Centro de Estudos do Genoma Humano. A análise das sequências foi realizada no programa Sequencher 5.2 (*Gene Codes*).

A validação das variantes candidatas foi feita para 6 pacientes da família F886 (III-7, III-9, III-18, III-19, III-20 e V-1), 2 familiares que podem ser portadores da variante de interesse, pois tem descendentes afetados e são clinicamente normais (III-3 e IV-5), 6 familiares clinicamente normais sem descendentes (que também poderiam ser portadores da variante (IV-1, IV-2, IV-3, IV-9, IV-10 e IV-11)) e 4 familiares agregados à família (familiares que podem ser considerados como controles da população, uma vez que são externos à família e não receberam os alelos que estão segregando entre os pacientes (III-4, III-10, IV-4 e IV-12)).

3.3. Análise de metilação

Os genes selecionados pelos critérios acima expostos foram submetidos à análise de metilação para verificar se há diferença entre pacientes e controles.

3.3.1. Análise de metilação para os genes candidatos selecionados por SCE e painel

Foi feita a análise de metilação em pacientes com FL/P NS, familiares e controles. A identificação dos indivíduos por grupos (pacientes, controles, familiares, portadores não-afetados), sexo, grupos de idade ou tipo de tecido (sangue ou saliva) está relacionada no Anexo I.

As famílias que possuíamos amostras tanto de pacientes quanto de familiares estão representadas na Figura 8. Tais famílias são as mesmas em que foram selecionados os 9 genes candidatos por Brito (2016). Dessa forma, os pacientes (N=18) destas famílias, apresentam alguma variante para esses genes, assim como os heterozigotos clinicamente normais (N=3). Também pertencem a essas famílias os familiares selecionados (N=21). Os pacientes (N=10) que não apresentam variante para os genes

incluídos na análise de metilação pertencem a outras famílias e não temos DNA dos familiares destes indivíduos.

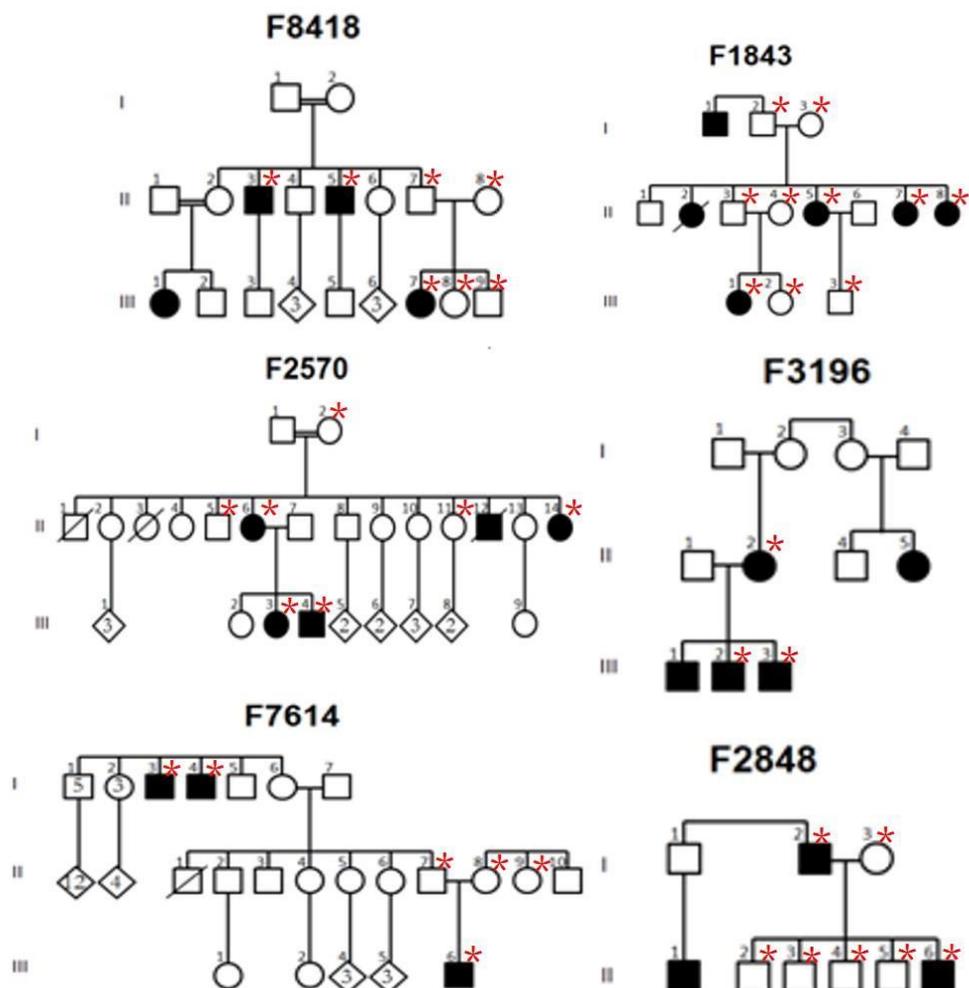


Figura 8 - Famílias brasileiras multiplex com pelo menos 2 pacientes com FL/P NS estudadas no presente trabalho. Os asteriscos indicam os indivíduos em que foram testados o padrão de metilação. As variantes descritas por Brito (2016) foram encontradas nestas famílias e todos os pacientes representados pelo asterisco apresentam variante em algum dos genes selecionados para análise de metilação.

Para esta etapa, foi utilizado o método *Bisulfite Amplicon Sequencing* (Masser *et al.*, 2013), que depende de PCR bissulfito, preparo de biblioteca e sequenciamento.

Primeiramente, utilizando a ferramenta online do *UCSC Genome Browser* (<https://genome.ucsc.edu>), foram selecionadas, no promotor dos genes, regiões contendo sítios CpGs agrupados para cada gene. As regiões selecionadas foram discriminadas pelo número CpG presente no *UCSC*, pois alguns genes contêm mais que uma região de sítios CpG. Caso não houvesse nenhum sítio CpG descrito no *UCSC*, a região era selecionada a partir da sequência de DNA da região promotora do gene. A

partir dessas regiões, foram desenhados *primers* com a conversão das citosinas para timinas em regiões que não são passíveis de serem metiladas, ou seja, onde não há citosinas próximas a guaninas, utilizando a ferramenta online *MethPrimer* (<http://www.urogene.org/methprimer/>) e o *software Methyl Primer Express* (*Applied Biosciences*, USA). Os *primers* desenhados podem ser visualizados no Anexo II. Foram adicionados adaptadores *Illumina* CS1 e CS2 aos *primers Forward* e *Reverse* com a sequência de nucleotídeos ACACTGACGACATGGTTCTACA e TACGGTAGCAGAGACTTGGTCT, respectivamente.

As amostras de DNA utilizadas foram submetidas a um processo de conversão de bissulfito, em que as citosinas não-metiladas foram convertidas em uracilas e as citosinas metiladas permaneciam como citosinas, permitindo a diferenciação entre elas. A conversão foi feita utilizando a técnica de hidroquinona e metabissulfito de sódio saturado (Método *in-house*; Patterson *et al.*, 2011). A partir do DNA convertido, foi feita reação em cadeia de polimerase (PCR) bissulfito-específica para cada par de *primers* desenhados, utilizando o protocolo padrão Platinum PCR Supermix da Invitrogen. Os produtos foram checados por eletroforese em gel de agarose 1,5%.

Uma vez amplificados, foi feita uma mistura de todos os produtos (*pool*) para cada indivíduo e os tamanhos dos produtos foram checados utilizando *BioAnalyzer* para o preparo da biblioteca. Cada amostra de *pool* foi indexada utilizando-se *index* específicos para cada amostra (*Access Array Barcode Library, Fluidigm*, com 1uL de *index*) em reação de PCR e a indexagem foi checada pelo *BioAnalyzer*. Após indexagem, um *pool* com todos os indivíduos-amplicons foi realizado para gerar a biblioteca. A biblioteca foi purificada por *Ampure XP Beads* em coluna magnética e, posteriormente, por purificação em gel, utilizando o *GFX PCR DNA and Gel Band Purification Kit*, para eliminação de bandas com tamanhos menores do que os esperados para os produtos de PCR, e foi checada novamente no *BioAnalyzer* para visualização de picos. A quantificação foi feita por meio do *KAPA Library Quantification kit* (*KAPA Biosystems*, Wilmington, MA) e a corrida foi realizada por *PCR Real Time da Roche (LightCycler 480)*. As bibliotecas foram submetidas ao sequenciamento utilizando a corrida *MiSeq Reagent V2 Kit 250bp pair-ended* no *MiSeq Sequencer (Illumina)*.

Após o sequenciamento, o primeiro tratamento de bioinformática foi a individualização das amostras e a remoção das sequências *index* e adaptadores utilizando o programa *FASTX Barcode Splitter* do pacote *FastX Toolkit R* (http://hannonlab.cshl.edu/fastx_toolkit). Foi feito um filtro de qualidade nos *reads*, selecionando os que apresentavam pelo menos 50% de bases com $Q > 30$, utilizando o programa *FASTQ Quality Filter* do pacote *FastX Toolkit R*. Os arquivos foram convertidos em FASTA com o programa *FASTQ-to-FASTA*, gerando um arquivo para cada região CpG de cada indivíduo, que foi utilizado para a análise de metilação.

Para a quantificação do nível de metilação nas ilhas CpG dos genes foi utilizado o *software BiQ Analyzer HT*, que alinha as sequências genômicas de referência aos *reads* das amostras e calcula o nível de metilação (de 0, completamente não-metilada, a 1, completamente metilada) para cada região analisada de cada amostra. Tal análise foi realizada com o filtro de identidade mínima de sequência de 80%. A partir dos dados obtidos, foi feito um filtro em que se excluiu da análise, para cada gene, os indivíduos que apresentassem número menor que 10 *reads*. As sequências genômicas de referência utilizadas estão representadas no Anexo III.

Com os dados obtidos, foi realizado o teste de Análise de Componente Principal (PCA), utilizando o *software R* versão 3.4.2 e os pacotes “*ggbiplot*” e “*devtools*”, para se verificar como as amostras se agrupavam de acordo com os valores de metilação de todas as regiões CpG.

Na sequência da análise de metilação, foi feito o teste ANOVA, utilizando o *software R* versão 3.4.2, para se verificar a contribuição de cada variável (idade, sexo e tecido) sobre os valores de metilação.

Posteriormente, foi feito um teste de Comparação múltipla das médias (ANCOVA), que compara os valores de metilação entre pacientes e controles, eliminando os efeitos das variáveis “Idade”, “Sexo” e “Tecido”, com base em seus valores de contribuição calculados no teste ANOVA. Foram utilizados os pacotes “*car*” e “*multcomp*” para estas análises no *software R*. Também foram calculadas as diferenças das médias de níveis de metilação de pacientes e controles para se verificar se estava

ocorrendo um aumento ou uma diminuição em seus valores. Para a análise de metilação, foram considerados os grupos "portadores não-afetados" e "familiares" como integrantes do grupo "controles".

A análise foi realizada para cada região ou agrupamento de ilhas CpGs (denominado aqui como "análise por região") e para cada sítio CpG analisado isoladamente (denominado aqui como "análise por sítio CpG"), conforme ilustrado na Figura 9.

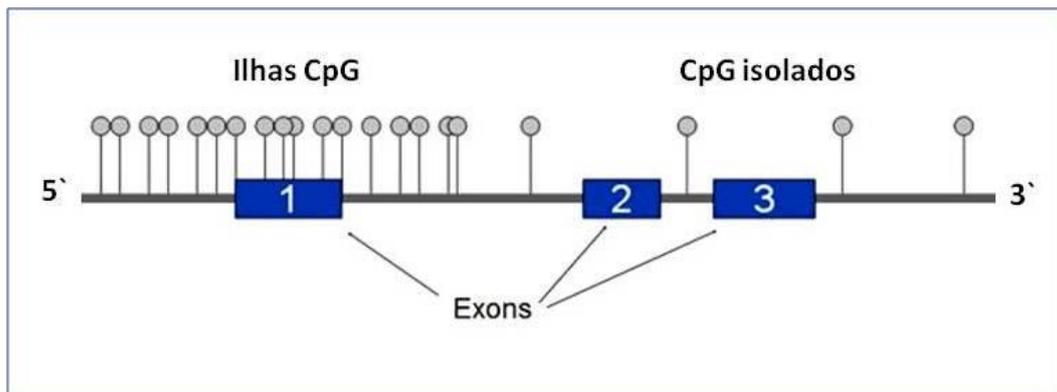


Figura 9 – A análise de metilação pode ser feita para cada sítio CpG isolado (representado pelo círculo cinza), mesmo que esteja presente em uma ilha CpG, sendo possível observar o efeito de cada unidade de CpG, ou para toda a ilha CpG (agrupamento de CpGs), observando o efeito simultâneo de vários CpGs agrupados (Imagem: <http://lifesequencing.blogspot.com.br/p/analise-de-metilacao-por-sequenciamento.html>).

As famílias selecionadas para esta análise apresentavam algumas das variantes para os genes analisados (Brito, 2016). Dessa forma, a análise foi realizada por região e por sítio, das seguintes formas:

- a) para todos os genes candidatos com a amostra total de pacientes *versus* controles (denominada aqui como "amostra global") e;
- b) para os genes em que os pacientes de uma mesma família possuíam variante rara apenas com esses pacientes *versus* todos os controles (aqui denominado como "análise por família").

A significância estatística para a análise por região foi considerada para p-valores < 0,05, após correção de múltiplos testes de *Tukey*. Para a análise por sítio, foram selecionados os genes com pelo menos 3 sítios significativos, com p-valores abaixo

de 0,05 (após correção de múltiplos testes de *Tukey*) e com um mínimo de 4% de diferença das médias entre pacientes e controles.

Na análise por sítio, foi feito um gráfico *Heatmap* para melhor visualização dos p-valores obtidos para cada sítio dentro de cada região CpG, utilizando o *software* R com os pacotes *reshape2* e *ggplot2*.

No estudo da amostra global, também foi realizada a análise de redes de interações proteína-proteína para os genes em que foram previamente encontradas variantes nas famílias brasileiras analisadas (Brito, 2016) e que foram submetidas à análise de metilação e genes que apresentaram p-valores $< 0,05$ nas análises de metilação, utilizando o *software* online *Network Analyst* versão 3.0 (<https://www.networkanalyst.ca/NetworkAnalyst/faces/Secure/network/ListAnalOverview.xhtml>), por meio do banco de dados IMEx Interactome (Breuer *et al.*, 2013). O objetivo desta análise é visualizar se existe interação por intermédio de proteínas entre os genes metilados e os genes selecionados por SCE que apresentam variantes potencialmente patogênicas que segregam nestas famílias brasileiras.

A análise de metilação foi realizada em colaboração com o pós-doutorando Dr. Lucas Alvizi Cruz.

3.3.2. Análise de metilação na família F886

Foi realizada pelo Dr. Lucas Alvizi Cruz a reanálise do metiloma, previamente publicada no trabalho de Alvizi *et al.* (2017), de 5 pacientes da família F886 (III-7, III-9, III-19, III-20 e V-1 da Figura 7) e 59 controles da população brasileira.

Para tal análise, foi utilizado o *pipeline* RnBeads (pacote *Bioconductor* do *software* R), em que foram excluídas as sondas nos cromossomos sexuais e as sondas com p-valor de detecção maior que 0,05. Além disso, na análise de metilação diferencial, foram selecionadas as sondas nas regiões candidatas (dos picos encontrados no estudo de ligação) com p-valor $< 0,05$ e diferença de metilação maior ou igual a 5%.

4. Resultados

4.1. Genes selecionados para a análise de metilação

Nos tópicos abaixo, estão relacionados os genes que foram selecionados a partir das análises de sequenciamento de exoma completo (SCE), de painel de genes candidatos e de estudo de ligação, os quais foram subsequentemente submetidos à análise de metilação.

4.1.1. Genes selecionados nas famílias inglesas

Por meio da análise de SCE, foram obtidas 433.340 variantes para cada família. Após os filtros iniciais, este número reduziu para 16.166 variantes. Destas, na comparação pacientes-familiares ou pacientes-pacientes, obtivemos 54 variantes para a família 281, 75 para a família 313 e 31 para a família BC84.85. A partir dessas variantes, utilizando os critérios de função, expressão e predição de patogenicidade, foram selecionadas três variantes *missense*, uma em cada família, descritas na Tabela 1. Todos os genes encontrados para as famílias inglesas foram submetidos à análise de metilação.

Tabela 1 – Genes e variantes candidatas selecionadas a partir de SCE das famílias inglesas.

Família	Gene	Região	Posição (hg19)	Variante	Tipo	Predição <i>in silico</i> *	Frequência	Informações relevantes (Função do gene, Sysface, MGI e OMIM)
						SIFT / Polyphen-2 HD;HV / ExAC	ExAC / 1kGP / ESP6500 / ABraOM	
281	VCL	10q22.2	75843264	c.1015C>T:p.R339C (NM_003373)	missense	0.02 D / 1 D ; 1 D / z= 2.71 (T)	0,0244 / 0 / 0 / 0	Codifica para proteína citoesquelética que se associa à face citoplasmática de junções do tipo aderente ¹ Anormalidades no fechamento do tubo neural e na adesão celular em camundongos Publicações anteriores relacionadas à FL/P: Jugessur <i>et al.</i> , 2011 Associado a cardiomiopatia hipertrófica (OMIM#613255) (Herança autossômica dominante)
BC84-85	SMEK1	14q32.12	91925172	c.1716A>T:p.E572D (NM_001284281)	missense	0.95 T / 0.0,B; 0.001,B / z= 3.64 (I)	0 / 0 / 0.000384 / 0.00082	Papel importante em organização de microtúbulos, apoptose, parada do ciclo celular e controle de danos no DNA ²
313	RECQL4	8q24.3	145741187	c.1219G>A:p.E407K (NM_004260)	missense	na / na; na / z= -4.22 (T)	0,08178 / 0.000199 / 0 / 0	Proteína de RecQ helicase ativa em vários tipos de células antes e após o nascimento. Acredita-se ser importante nas células do desenvolvimento de ossos e peles ³ Expresso em mandíbula Fissura do palato em camundongos Associado às síndromes Baller-Gerold (OMIM#218600), Rapadilino (OMIM#266280) e Rothmund-Thomson (OMIM#268400) (Herança autossômica recessiva)

1-Weller *et al.*, 1990; 2-Kim *et al.*, 2017; 3-Bernstein *et al.*, 2010.

HD: Polyphen-2 HumDiv; **HV:** Polyphen-2 HumVar; **1kGP:** 1000 Genomes Project; **ExAC:** Exome Aggregation Consortium; **ESP6500:** Exome Sequencing Project database; **ABraOM:** banco de dados do Centro de Estudos do Genoma Humano e Células-Tronco; **na:** não disponível. **D:** patogênico; **P:** possivelmente patogênico; **B:** benigno; **T:** tolerado (SIFT), tolerante (ExAC); **I:** intolerante; **AD:** Herança autossômica dominante; **AR:** Herança autossômica recessiva.

* Valores de *score* para predição de variantes: SIFT patogênico se ≤ 0.05 ; Polyphen-2, provavelmente patogênico se ≥ 0.909 (HumVar) e ≥ 0.957 (HumDiv). ExAC classifica o gene como tolerante (T) ou intolerante (I) a mutações *missense* (z score, intolerante se $z \geq 3$) ou LoF (pLI score, intolerante se $pLI \geq 0.9$).

4.1.2. Genes selecionados nas famílias brasileiras

- *Genes candidatos de Brito (2016)*

A partir da análise do SCE de famílias multiplex brasileiras (Brito, 2016), foram selecionadas e validadas as variantes raras e possivelmente patogênicas representadas na Tabela 2. Os 9 genes que apresentavam essas variantes raras nessas famílias foram selecionados para serem submetidos à análise de metilação.

Tabela 2 - Genes selecionados a partir de SCE em famílias brasileiras (Brito, 2016).

Família	Gene*	Região	Posição (hg19)	Variante	Tipo	Predição <i>in silico</i> **	Frequência**	Informações relevantes (Função do gene, Sysface, MGI e OMIM)
						SIFT / Polyphen-2 HD;HV / ExAC	ExAC / 1kGP / ESP6500 / ABraOM	
F1843	<i>PPM1F</i>	22q11	22300402	c.19C>T;p.Q7X (NM_014634)	<i>Stopgain</i>	0.02 D / na;na / pLI=0 (T)	0 / 0 / 0 / 0	Transporte de N-caderina ¹ Expressão em maxila
F2570	<i>PAX8</i>	2q24	113999636	c.550G>C;p.G184R (NM_003466)	<i>Missense</i>	0.36 T / 1 D ; 1 D / z=0.5 (T)	0 / 0 / 0 / 0	Organogênese e determinação de linhagem ² Associado a hipotireoidismo congênito (OMIM#218700)(AD)
	<i>KIF20B</i>	10q23	91522558	c.4835A>G;p.K1612R	<i>missense</i>	0.19 T / 0.06 B; 0.03 B / z=-3.5 (T)	0.0004 / 0.001 / 0.001 / 0	Citocinese de células neuroepiteliais ³ Fissura de palato e morfologia craniofacial anormal em camundongos
F2848	<i>ROR2</i>	9q22	94487187	c.1589G>A;p.R530Q (NM_004560)	<i>missense</i>	0.05 D / 0.36 B / 0.01 B / z=0.5 (T)	0.002 / 0.001 / 0.002 / 0.001	Participação na via PCP ⁴ Fissura de palato em camundongos Associado a Síndrome de Robinow (OMIM#268310) (AR) e braquidactilia (OMIM#113000) (AD)
	<i>ZEB1</i>	10p11	31809536	c.1213A>G;p.I405V (NM_001174093)	<i>missense</i>	0.09 T / 0.76 P / 0.25 B / z=0.7 (T)	0 / 0 / 0 / 0	Adesão celular ⁵ Fissura do lábio superior em camundongos Associado a distrofia corneal (OMIM#613270)
F7614	<i>KIFAP3</i>	1q24	169961306	c.1142G>C;p.C381S (NM_001204516)	<i>missense</i>	0 D / 0.12 B; 0.25 B / z=1.2 (T)	0 / 0 / 0 / 0	Adesão celular ⁶
F8418	<i>CDK1</i>	10q21	62544513	c.88G>A;p.V30I (NM_001170406)	<i>missense</i>	0.09 T / 0.97 D ; 0.77 D / z=2.6 (T)	0 / 0 / 0 / 0	Divisão celular e reparo de dado de DNA ⁷

F3196	<i>IGF2R</i>	6q25	160468235	c.2096C>T:p.S699L (NM_000876)	<i>missense</i>	0.001 D / 1 D ; 1 D / z=1.4 (T)	0 / 0 / 0 / 0	Regulação da proliferação celular, crescimento, migração e diferenciação ⁸ Expressão em maxila e palato Morfologia craniofacial anormal em camundongos Associado a carcinoma hepatocelular (OMIM#114550)
	<i>PRICKLE1</i>	12q12	42853958	c.2149delA:p.S717fs (NM_001144881)	<i>frameshift deletion</i>	na / na; na / pLI=1 (I)	0 / 0 / 0 / 0	Participação na via PCP ⁴ Expressão em maxila, frontonasal e palato Fissura de palato em camundongos Associado a epilepsia progressiva mioclônica (OMIM#612437) (AR)

1-Phang *et al.*, 2014; 2-Blake & Ziman, 2014; 3-Janisch *et al.*, 2013; 4-Sebbagh & Borg, 2014; De Calisto *et al.*, 2005; 5-Peinado *et al.*, 2007; 6-Tanuma *et al.*, 2009; 7-Blake & Ziman, 2014; 8-Bergman *et al.*, 2013.

HD: Polyphen-2 HumDiv; **HV:** Polyphen-2 HumVar; **1kGP:** 1000 Genomes Project; **ExAC:** Exome Aggregation Consortium; **ESP6500:** Exome Sequencing Project database; **ABraOM:** banco de dados do Centro de Estudos do Genoma Humano e Células-Tronco; **na:** não disponível. **D:** patogênico; **P:** possivelmente patogênico; **B:** benigno; **T:** tolerado (SIFT), tolerante (ExAC); **I:** intolerante; **AD:** Herança autossômica dominante; **AR:** Herança autossômica recessiva.

* Genes reportados em Brito (2016). Heredogramas das famílias representados na Figura 8.

** Valores de *score* para predição de variantes: SIFT patogênico se ≤ 0.05 ; Polyphen-2, provavelmente patogênico se ≥ 0.909 (HumVar) e ≥ 0.957 (HumDiv). ExAC classifica o gene como tolerante (T) ou intolerante (I) a mutações *missense* (z score, intolerante se $z \geq 3$) ou LoF (pLI score, intolerante se $pLI \geq 0.9$).

- *Genes candidatos de painel de FL/P NS*

Foram selecionados 5 genes presentes em painel de genes candidatos de pacientes com FL/P NS. Os genes estão relacionados na Tabela 3.

Tabela 3 - Genes selecionados a partir de painel de FL/P NS.

Gene	Critério de inclusão
<i>ARHGAP29</i>	Apresenta variantes encontradas em pacientes com FL/P NS, inclusive na população brasileira ¹
<i>BRCA1</i>	Expressão em primórdio orofacial embrionário e expressão desregulada em células tronco de polpa de dente de pacientes com FL/P em relação a controles ²
<i>NOL8</i>	eQTLs identificados em músculo <i>oris orbicularis</i> de pacientes com FL/P NS ³
<i>TP63</i>	Associado a FL/P Síndrômica – ectrodactilia ⁴
<i>SHH</i>	Associado a FL/P Síndrômica – holoprosencefalia ⁵

1-Leslie *et al.*, 2012; Chandrasekharan & Ramanathan, 2014; Savastano *et al.*, 2017; 2-Kobayashi *et al.*, 2013; 3-Masotti *et al.*, 2018; 4-van Bokhoven *et al.*, 2001; Sripathomsawat *et al.*, 2011; 5-Verlinsky *et al.*, 2003.

- *Reanálise de SCE da família F886*

Na reanálise de SCE dos 6 pacientes da família F886 (III-7, III-9, III-18, III-19, III-20 e V-1), considerando a hipótese de heterogeneidade genética, foram escolhidas 3 variantes do tipo *stopgain* ou *frameshift insertion* em possíveis genes candidatos pelo pós-doutorando Dr. Luciano Abreu Brito. Os genes estão representados na Tabela 4.

Tabela 4 – Genes e variantes selecionados por SCE na família F886 (Brito, 2016), considerando a hipótese de heterogeneidade genética.

Gene	Região	Posição (hg19)	Variante	Pacientes que carregam a variante	Predição <i>in silico</i> SIFT / Polyphen-2 HD;HV / ExAC	Frequência ExAC / 1kGP / ESP6500 / ABraOM	Informações relevantes (Função do gene, Sysface, MGI e OMIM)
<i>NUMA1</i>	11q13.4	71726283	c.2266G>T:p.E756X (NM_006185) <i>Stopgain</i>	III-18, III-19 e III-20	0.04 D / na; na / pLI = 1 (I)	0 / 0 / 0 / 0	Proteína de matriz nuclear, regulador de ciclo celular, recruta proteínas do centrômero, formação e manutenção do fuso mitótico Expresso em mandíbula Associado a leucemia
<i>STAT5A</i>	17q21.2	40452264	c.708_709insGC:p.G236fs (NM_001288719) <i>frameshift insertion</i>	III-9, III-19, III-20 e V-I	na / na; na / pLI = 1 (I)	0 / 0 / 0 / 0	Resposta a interleucinas
<i>THAP11</i>	16q22.1	67876755	c.299dupC:p.A100fs (NM_020457) <i>frameshift insertion</i>	III-9, III-19, III-20 e V-I	na / na; na / pLI = 0.47 (T)	0 / 0 / 0 / 0	Letalidade em camundongos

HD: Polyphen-2 HumDiv; **HV:** Polyphen-2 HumVar; **1kGP:** 1000 Genomes Project; **ExAC:** Exome Aggregation Consortium; **ESP6500:** Exome Sequencing Project database; **ABraOM:** banco de dados do Centro de Estudos do Genoma Humano e Células-Tronco; **na:** não disponível. **D:** patogênico; **P:** possivelmente patogênico; **B:** benigno; **T:** tolerado (SIFT), tolerante (ExAC); **I:** intolerante; **AD:** Herança autossômica dominante; **AR:** Herança autossômica recessiva.

* Valores de *score* para predição de variantes: SIFT patogênico se ≤ 0.05 ; Polyphen-2, provavelmente patogênico se ≥ 0.909 (HumVar) e ≥ 0.957 (HumDiv). ExAC classifica o gene como tolerante (T) ou intolerante (I) a mutações *missense* (*z score*, intolerante se $z \geq 3$) ou LoF (*pLI score*, intolerante se $pLI \geq 0.9$).

- *Estudo de ligação da família F886 e reanálise dos dados de SCE*

Os gráficos com valores de LOD Score acima ou próximo de 2.0, que demonstram as regiões cromossômicas com maiores chances de estarem associadas ao fenótipo, obtidos no estudo de ligação da família F886 estão representados na Figura 10.

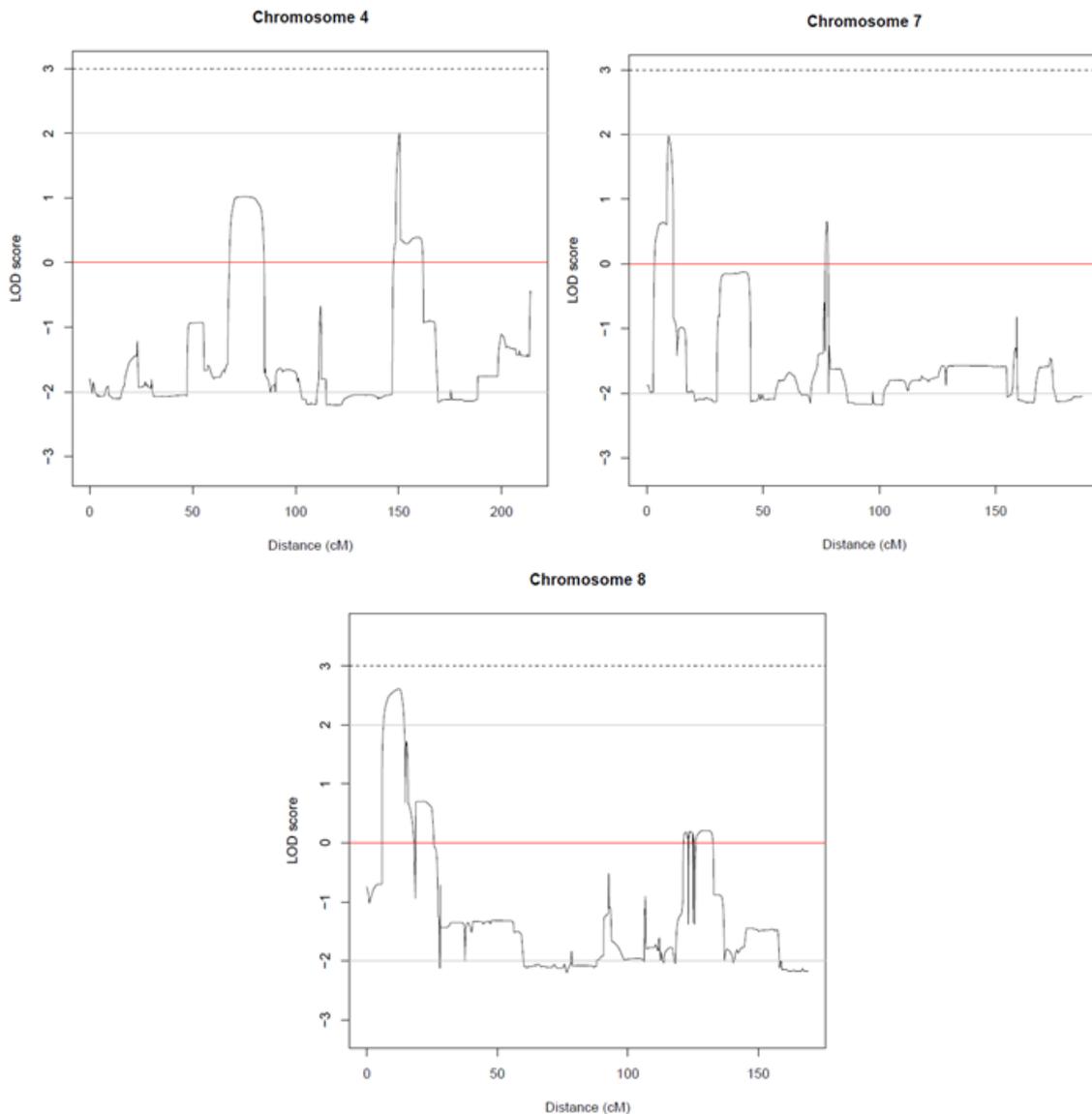


Figura 10 – Gráficos com valores de LOD Score acima de 2.0 obtidos no estudo de ligação da família F886.

Em seguida, baseando-se nos dados de LOD Score, foram definidas as posições físicas dos cromossomos 4, 7 e 8 (Tabela 5) com maior probabilidade de conterem variantes de risco para a FL/P nessas famílias.

Tabela 5 – Posições genética e física selecionadas com LOD Score acima ou próximo de 2.0.

Cromossomo	Posição genética inicial	Posição genética final	Posição física inicial	Posição física final	LOD Score máximo
4	147.554	161.683	145537850	160349432	2.0148
7	3.448	11.371	2815140	7256440	1.9764
8	5.787	25.636	2964003	12911318	2.6074

Foram observados 370 genes nas três regiões delimitadas. Os genes foram avaliados e estudados de acordo com suas funções e expressão. Foram escolhidos 5 genes e, dentre esses genes, foi selecionado o gene *DCHS2* no cromossomo 4, pois além de ter participação na via PCP, também observamos variantes raras para esse gene em 4 pacientes na reanálise do SCE desta família.

Foram encontradas 2 variantes em *DCHS2*, sendo uma *missense* (e25; NM_017639) e uma intrônica (i20). Os dados das variantes estão representados na Tabela 6.

Tabela 6 – Variantes raras em *DCHS2* na família F886.

Gene	Região	Posição (hg19)	Variante	Predição <i>in silico</i>	Frequência	Informações relevantes (Função do gene, Sysface, MGI e OMIM)
				SIFT / Polyphen-2 HD;HV / ExAC	ExAC / 1kGP / ESP6500 / ABraOM	
<i>DCHS2</i>	4q31.3	155158029	c.6410G>T:p.G2137V (NM_017639) <i>missense</i> (e25)	na / 1 D; 0.999 D / z = -0.25 (T)	na / 0 / 0 / 0.002	Participação da via Fat/Dchs2
		155180652	Intrônico (i20)	na / na; na / na	na / 0 / 0 / 0.003	

HD: Polyphen-2 HumDiv; **HV:** Polyphen-2 HumVar; **1kGP:** 1000 Genomes Project; **ExAC:** Exome Aggregation Consortium; **ESP6500:** Exome Sequencing Project database; **ABraOM:** banco de dados do Centro de Estudos do Genoma Humano e Células-Tronco; **na:** não disponível. **D:** patogênico; **P:** possivelmente patogênico; **B:** benigno; **T:** tolerado (SIFT), tolerante (ExAC); **I:** intolerante.

* Valores de *score* para predição de variantes: SIFT patogênico se ≤ 0.05 ; Polyphen-2, provavelmente patogênico se ≥ 0.909 (HumVar) e ≥ 0.957 (HumDiv). ExAC classifica o gene como tolerante (T) ou intolerante (I) a mutações *missense* (z score, intolerante se $z \geq 3$) ou LoF (pLI score, intolerante se $pLI \geq 0.9$).

Ambas as variantes (e25 e i20) se encontravam em heterozigose para os pacientes III-7, III-9, III-19 e III-20, e em homozigose para o alelo raro em III-18. O paciente V-I não apresentava a variante.

Os resultados das variantes encontradas em *DCHS2* foram validados pela análise de sequenciamento pelo método Sanger. Na Figura 11 estão relacionados os genótipos encontrados na validação das variantes e25 e i20 na família F886.

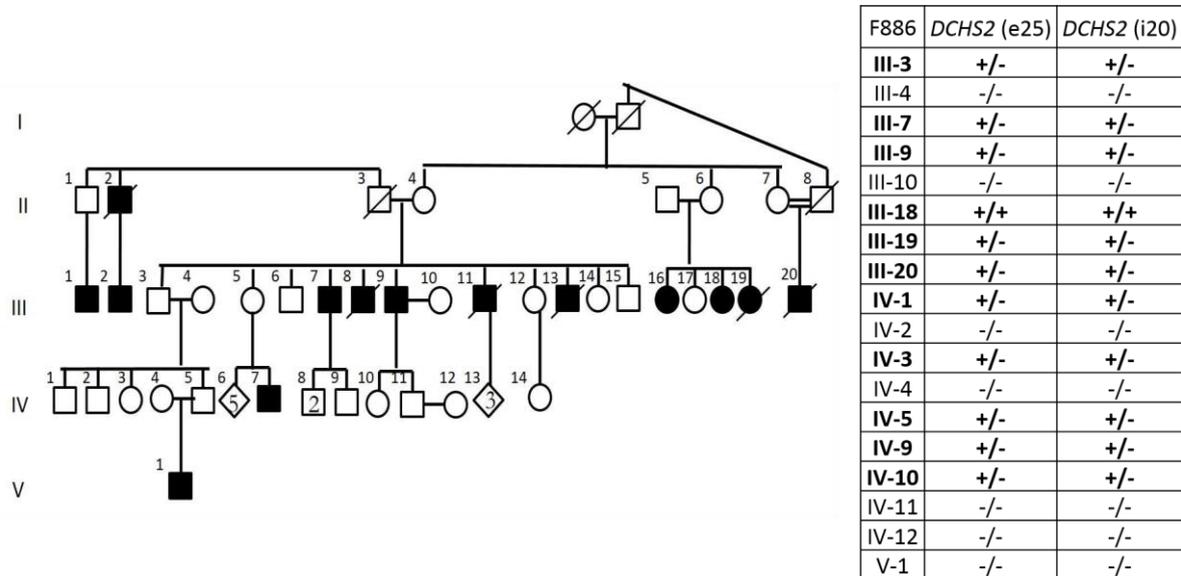


Figura 11 – Heredograma e genótipos das variantes e25 e i20 para cada indivíduo da família F886.

Foi observado que os controles da família (III-4, III-10, IV-4 e IV-12) não apresentavam as variantes e, além dos pacientes (III-7, III-9, III-18, III-19 e III-20), os familiares não-afetados que possuem descendentes afetados (III-3 e IV-5) apresentavam as variantes de *DCHS2*. Além disso, foi observado que as duas variantes apresentam o mesmo padrão de genótipo em cada indivíduo (ambas heterozigotas ou homozigotas para o mesmo alelo).

4.2. Análise de metilação

Para esta etapa, foram incluídos os genes selecionados nas análises de SCE, de painel de genes candidatos e de estudo de ligação, o que resultou em um total de 21 genes candidatos a FL/P NS (*ARHGAP29*, *BRCA1*, *CDK1*, *DCHS2*, *IGF2R*, *KIFAP3*, *KIF20B*, *NOL8*, *NUMA1*, *PAX8*, *PPM1F*, *PRICKLE1*, *RECQL4*, *ROR2*, *SHH*, *SMEK1*, *STAT5A*, *THAP11*, *TP63*, *VCL* e *ZEB1*), dentre eles, 1 gene já associado à forma não-sindrômica (*ARHGAP29*), três genes associados a formas síndrômicas (*SHH*, *TP63*, *ROR2*) e um gene (*DCHS2*),

localizado em uma região cromossômica candidata por estudos de ligação combinado com análise de SCE.

4.2.1. Análise de metilação para os genes candidatos selecionados por SCE e painel

As regiões de ilhas CpGs que foram submetidas à análise de metilação para os genes candidatos selecionados por SCE de Brito (2016), na reanálise da família F886 (com exceção de *DCSH2* - Seção 4.2.2) e do painel estão indicadas na Tabela 7, assim como os tamanhos dos produtos de PCR esperados somados aos adaptadores e *index* e quantidade de sítios CpGs de cada gene. A localização cromossômica das regiões CpG selecionadas estão representadas no Anexo III.

Tabela 7 – Número da região CpG (UCSC hg19) selecionada para cada gene, tamanhos dos produtos esperados e quantidade de sítios CpG.

Gene	Número da Ilha CpG	Tamanho do produto (pb)	Quantidade de sítios CpG	Gene	Número da Ilha CpG	Tamanho do produto (pb)	Quantidade de sítios CpG
<i>ARHGAP29</i>	73	309	21	<i>RECQL4</i>	47	341	13
<i>BRCA1</i>	24	374	17	<i>RECQL4</i>	164	348	32
<i>CDK1</i>	69	376	22	<i>ROR2</i>	21	343	17
<i>IGF2R</i>	124	390	39	<i>ROR2</i>	210	424	26
<i>KIF20B</i>	51	224	6	<i>SHH</i>	32	335	25
<i>KIFAP3</i>	30	348	17	<i>SHH</i>	542	378	22
<i>NOL8</i>	64	277	11	<i>SMEK1</i>	166	378	23
<i>NUMA1</i>	51	314	13	<i>STAT5A</i>	85	310	26
<i>PAX8</i>	25	427	26	<i>THAP11</i>	92	387	34
<i>PAX8</i>	27	260	10	<i>TP63</i>	Promotor 1	349	6
<i>PAX8</i>	131	406	20	<i>TP63</i>	Promotor 2	289	4
<i>PPM1F</i>	21	419	25	<i>TP63</i>	Promotor 3	448	4
<i>PRICKLE1</i>	23	374	15	<i>VCL</i>	87	445	38
<i>PRICKLE1</i>	107	248	21	<i>ZEB1</i>	238	431	25
<i>PRICKLE1</i>	165	303	23				

Os valores de metilação obtidos para cada indivíduo para as ilhas CpGs analisadas (por região) estão representados no Anexo IV.

- *Análise de metilação na amostra global*

Essa análise tem como objetivo verificar se o padrão de metilação dos genes candidatos difere entre pacientes e controles.

O agrupamento das amostras de acordo com os valores de metilação de todas as regiões CpG obtido a partir da Análise de Componente Principal (PCA) está representado na Figura 12.

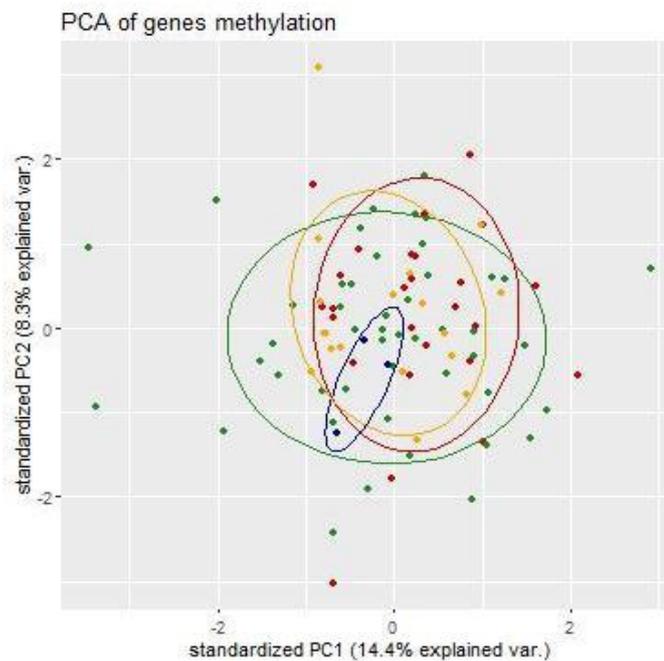


Figura 12 - Análise de Componente Principal. Cada ponto representa um indivíduo e o círculo, como eles se distribuem. Verde - controles, azul - portadores não-afetados, vermelho - pacientes, amarelo - familiares.

Foi observado que os dados brutos de nível de metilação de controles, pacientes, familiares e portadores não-afetados se distribuem uniformemente, sem alguma tendência clara em um primeiro momento.

Posteriormente, verificamos a contribuição das variáveis “idade”, “sexo” e “tecido” sobre a metilação de cada região CpG. Para isso, foi feito o teste ANOVA para cada gene na amostra global.

A contribuição das variáveis “idade”, “sexo” e “tecido” está representada na Tabela 8.

Tabela 8 - Contribuição das variáveis “idade”, “sexo” e “tecido” sobre o nível de metilação para cada região CpG. O asterisco (*) indica p-valor < 0,05.

	Idade	Sexo	Tecido
<i>ARHGAP29</i>	0.0007579***	0.528949668	0.267855252
<i>BRCA1</i>	0.22799645	0.866758887	-
<i>CDK1</i>	0.69439111	0.197179568	0.0944274
<i>KIF20B</i>	0.641686842	0.380420164	0.931936368
<i>KIFAP3</i>	0.37672051	0.336777713	0.445845983
<i>NOL8</i>	0.039291426 *	0.119225221	0.663337504
<i>NUMA1</i>	0.1067807	0.123055127	0.141743615
<i>PAX8_CpG27</i>	0.066441605	0.659708218	9.5697110 ***
<i>PAX8_CpG131</i>	0.902770957	0.160058139	0.0164202 *
<i>PRICKLE1_CpG23</i>	-	-	-
<i>PRICKLE1_CpG107</i>	0.0023664**	0.171932639	7.1199e-05 ***
<i>PRICKLE1_CpG165</i>	0.037953329*	0.0924194 .	1.3026e-07 ***
<i>RECQL4_CpG47</i>	0.011935431*	0.837674347	0.223571478
<i>ROR2_CpG21</i>	0.084019024	0.333345855	0.773611889
<i>ROR2_CpG210</i>	0.308246556	0.827318506	0.325721004
<i>SHH_CpG542</i>	0.181158108	0.285746386	0.321381568
<i>SHPRH</i>	0.420196517	0.815539084	0.987207846
<i>SMEK1</i>	0.450173419	0.974571952	0.265563906
<i>STAT5A</i>	0.469409444	0.515388784	0.966728532
<i>TP63_Promoter1</i>	0.774726439	0.95126156	0.788765331
<i>TP63_Promoter2</i>	0.038433733*	0.478554901	5.3422e-14 ***
<i>ZEB1</i>	0.215322099	0.761537006	0.088857
<i>IGF2R</i>	0.834297138	0.869430077	0.277043318
<i>PAX8_CpG25</i>	0.455182965	0.107380213	0.989163857
<i>PPM1F_CpG21</i>	0.716470616	0.120955744	0.6954164
<i>RECQL4_CpG164</i>	0.017690759*	0.681284771	0.158057469
<i>SHH_CpG32</i>	0.177025812	0.26809579	0.465282041
<i>THAP11</i>	0.145223824	0.996721242	0.152246605
<i>TP63_Promoter3</i>	0.335688724	0.295960151	0.0006142 ***
<i>VCL</i>	0.918064825	0.340349004	0.486140456

Foi observado que as variáveis “idade” e “tecido” tem contribuição sobre os valores de metilação, enquanto que a variável “sexo” não apresentou contribuição. Dessa forma, os dados de metilação foram corrigidos a partir desses valores de contribuição pelo teste ANCOVA. As Tabelas 9 e 10 representam as médias do nível de metilação de pacientes e controles, a diferença entre as médias, o p-valor obtido após correção por ANCOVA e por *Tukey* e o número de pacientes e controles que permaneceram após os filtros realizados, na análise por região e na análise por sítio CpG (apenas os p-valores abaixo de 0,1), respectivamente.

Tabela 9 – Análise por região. Médias e diferença das médias do nível de metilação para pacientes e controles, resultado do teste ANCOVA (valor corrigido por Tukey) e número de pacientes e controles. O destaque em negrito representa o gene com p-valor < 0,05.

Gene	Média A (Pacientes)	Média B (Controles)	Diferença (Média A - Média B)	P-valor	Número Pacientes	Número Controles
ARHGAP29	0.0301923	0.0333928	-0.0032005	0.9838367	26	56
BRCA1	0.5676666	0.6138333	-0.0461666	0.394987	3	6
CDK1	0.0243333	0.0301702	-0.0058368	0.16439	18	47
KIF20B	0.0016153	0.0017164	-0.000101	0.6639452	26	67
KIFAP3	0.0046956	0.0050615	-0.0003658	0.7359305	23	65
NOL8	0.0036538	0.0046507	-0.0009969	0.4060469	26	63
NUMA1	0.0037307	0.0042537	-0.0005229	0.4136060	26	67
PAX8_CpG27	0.0053461	0.0065	-0.0011538	0.2383281	26	68
PAX8_CpG131	0.046	0.0401111	0.0058888	0.4207049	6	18
PRICKLE1_CpG23	0.013	0.0168	-0.0038	0.7951616	1	5
PRICKLE1_CpG107	0.0028	0.0033134	-0.0005134	0.9814776	25	67
PRICKLE1_CpG165	0.0034	0.0037164	-0.0003164	0.6154517	25	67
RECQL4_CpG47	0.9171666	0.9163692	0.00079743	0.5684752	24	65
ROR2_CpG21	0.4780769	0.5158378	-0.03776091	0.0694659	13	37
ROR2_CpG210	0.0104444	0.0119756	-0.00153116	0.7577127	18	41
SHH_CpG542	0.007	0.0086666	-0.00166666	0.3036001	2	9
SMEK1	0.0053846	0.0066428	-0.00125824	0.4225145	13	42
STAT5A	0.0026875	0.0027	-1.2499999e-05	0.8389751	16	60
TP63_Promotor1	0.8262380	0.7863863	0.0398517	0.0251912*	21	44
TP63_Promotor2	0.88764	0.8579848	0.0296551	0.1052293	25	66
ZEB1	0.00635	0.0067647	-0.0004147	0.98309	20	51
IGF2R	0.0033636	0.0028863	0.0004772	0.6863161	11	44
PAX8_CpG25	0.1285833	0.1597692	-0.0311858	0.1666101	12	26
PPM1F_CpG21	0.0214444	0.0208409	0.0006035	0.4383608	18	44
RECQL4_CpG164	0.0013809	0.0017959	-0.0004149	0.5986029	21	49
SHH_CpG32	0.0153333	0.0320666	-0.0167333	0.8961271	3	15
THAP11	0.00204	0.0018421	0.0001978	0.9107911	25	57
TP63_Promotor3	0.9325	0.87425	0.05825	0.4325151	2	12
VCL	0.002875	0.006	-0.003125	0.2587041	8	17

Tabela 10 – Análise por sítio CpG. Médias e diferença das médias do nível de metilação para pacientes e controles, resultado do teste ANCOVA (valor corrigido por *Tukey*; relacionados apenas os resultados com p-valores abaixo de 0,1) e número de pacientes e controles. Os destaques em negrito representam as regiões CpG com pelo menos 3 sítios com p-valor < 0,05 e diferença de 4%.

Gene	Sítio CpG	Média A (Pacientes)	Média B (Controles)	Diferença (Média A - Média B)	P-valor	Número Pacientes	Número Controles
<i>ARHGAP29</i>	Site.5	0.016257	0.033324	-0.017066	0.044864*	26	56
	Site.17	0.010887	0.006325	0.004563	0.00596**	26	56
	Site.18	0.008877	0.007662	0.001215	0.068439	26	56
<i>CDK1</i>	Site.1	0	0.002569	-0.002570	0.046067*	18	47
	Site.6	0.032481	0.064970	-0.032490	0.027518*	18	47
	Site.18	0.038218	0.073515	-0.035297	0.046038*	18	47
<i>IGF2R</i>	Site.26	0.007142	0.001292	0.005850	0.022128*	11	44
<i>NOL8</i>	Site.4	0.001191	0.000546	0.000645	0.030858*	26	63
<i>PAX8_CpG25</i>	Site.2	0.163227	0.237366	-0.074138	0.095988	12	26
	Site.6	0.091402	0.141478	-0.050076	0.024945*	12	26
	Site.9	0.130512	0.206307	-0.075795	0.064273	12	26
	Site.26	0.058333	0.000000	0.058333	0.006530*	12	26
<i>PAX8_CpG27</i>	Site.7	0.004154	0.005437	-0.001283	0.048643*	26	68
	Site.9	0.002629	0.003712	-0.001083	0.043864*	26	68
<i>PAX8_CpG131</i>	Site.5	0.023473	0.011149	0.012324	0.087995	6	18
	Site.9	0.025613	0.006256	0.019356	0.081185	6	18
<i>PRICKLE1_CpG107</i>	Site.3	0.001607	0.003641	-0.002034	0.047036*	25	67
<i>PRICKLE1_CpG165</i>	Site.1	0.003665	0.002582	0.001083	0.0994051	25	67
	Site.8	0.00277	0.001288	0.001482	0.00169**	25	67
	Site.11	0.001351	0.000606	0.000745	0.0525517	25	67
	Site.15	0.002811	0.002087	0.000724	0.0979320	25	67
	Site.17	0.003385	0.001913	0.001472	0.022820*	25	67
<i>RECQL4_CpG164</i>	Site.27	0.004337	0.000333	0.004004	0.034304*	21	49
<i>ROR2_CpG21</i>	Site.1	0.282216	0.327922	-0.045707	0.045095*	13	37
	Site.2	0.669174	0.733388	-0.064213	0.077941	13	37
	Site.3	0.324359	0.381226	-0.056867	0.00840**	13	37
	Site.10	0.561164	0.611593	-0.050429	0.043323*	13	37
	Site.13	0.197469	0.275953	-0.078484	0.084978	13	37
	Site.14	0.370476	0.428206	-0.057730	0.090944	13	37
	Site.15	0.136073	0.270900	-0.134827	0.00146**	13	37
<i>SHH_CpG542</i>	Site.22	0.103899	0.023419	0.080481	0.072011	2	9
<i>STAT5A</i>	Site.26	0.001119	0.003764	-0.002645	0.090081	16	60
<i>THAP11</i>	Site.1	0.001293	0.000373	0.000920	0.067454	25	57
	Site.22	0.004245	0.000693	0.003552	0.09518	25	57
	Site.33	0.001888	0.000269	0.001619	0.045140*	25	57

<i>TP63_Promotor1</i>	Site.3	0.903355	0.861318	0.042037	0.08048	21	44
	Site.4	0.792832	0.732257	0.060575	0.027518*	21	44
	Site.5	0.773872	0.712482	0.061390	0.063064	21	44
	Site.6	0.536802	0.435773	0.101029	0.064092	21	44
<i>TP63_Promotor2</i>	Site.1	0.876670	0.839990	0.036680	0.098217	25	66
	Site.4	0.825681	0.771010	0.054672	0.027398*	25	66
<i>TP63_Promotor3</i>	Site.3	0.533333	0.753920	-0.220587	0.065716	2	12
<i>VCL</i>	Site.7	0.000000	0.005348	-0.005348	0.085975	8	17
	Site.8	0.000000	0.000000	0.000000	0.085975	8	17

A Figura 13 representa todos os p-valores obtidos na análise por sítio após correção de *Tukey* para cada região CpG/gene, incluindo os p-valores indicados na Tabela 10.

Seguindo os critérios de seleção expostos na Seção 3.3.1, foi considerado como significativo o gene *TP63* promotor 1 (p-valor=0.0251912) na análise por região. Na análise por sítio, apesar de não terem sido observados 3 sítios significativos no gene *TP63*, observa-se uma variação de 4 a 10% de diferença de metilação em 4 sítios consecutivos e apenas uma região significativa (sítio 4). Assim sendo, tanto para a análise por região, quanto para a análise por sítio CpG, foi observada uma tendência à hipermetilação para *TP63* nos pacientes.

Não foram observados resultados significativos na análise por região do gene *ROR2*. Entretanto, foi considerada como significativa a ilha CpG número 21 do mesmo gene (4 sítios significativos) na análise por sítio, sendo observada uma tendência à hipometilação nos pacientes em *ROR2*.

Dentre os resultados acima expostos, destaca-se o gene *TP63* que foi estatisticamente significativo na análise por região e apresenta valores altos de diferença de metilação na análise por sítio. Diante disso, realizamos a análise de redes para o gene *TP63* e os genes que apresentam variantes validadas nas famílias brasileiras (Famílias representadas na Figura 8 e Genes da Tabela 2 - *PPM1F*, *ROR2*, *IGF2R*, *PAX8*, *PRICKLE1*, *KIF20B*, *ZEB1*, *KIFAP3* e *CDK1*). O objetivo dessa análise foi verificar se o gene *TP63* atua por intermédio de proteínas com os genes validados nessas famílias. A Figura 14 representa a análise de rede de interação proteína-proteína realizada para tais genes. A

proteína central, poliubiquitina-C (UBC), interage com os genes *TP63*, *ROR2*, *CDK1*, *PRICKLE1*, *KIFAP3*, *KIF20B*, *PPM1F* e *IGF2R*.

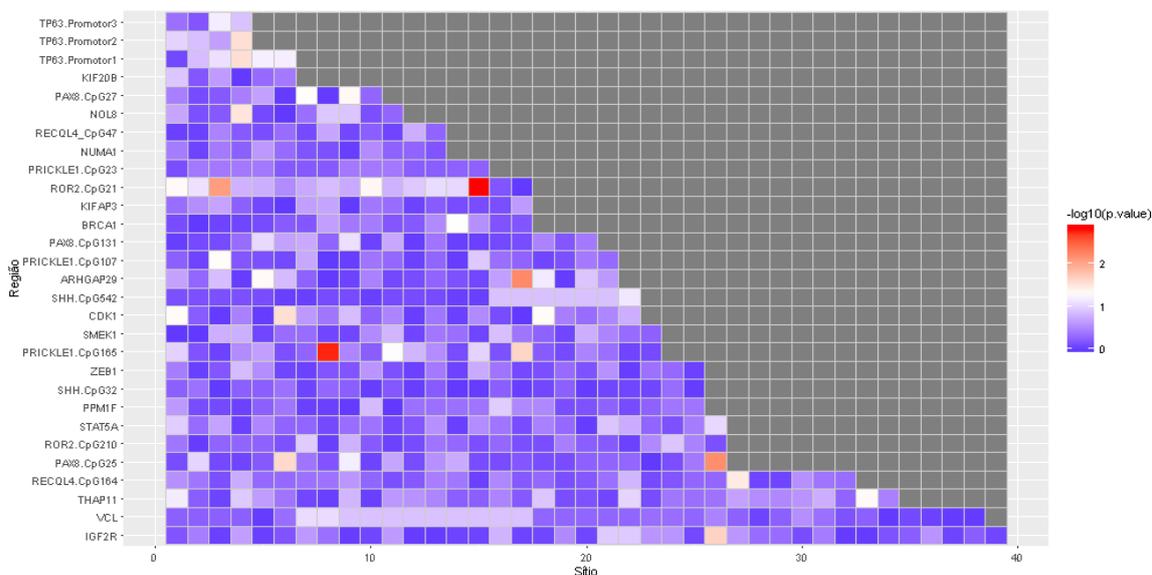


Figura 13 – P-valores obtidos na análise por sítio CpG para cada Região CpG/Gene. Os p-valores estão representados em escala logarítmica, em que, acima do limite em branco até o vermelho estão os p-valores abaixo de 0,05 e do branco ao azul, os p-valores acima de 0,05.

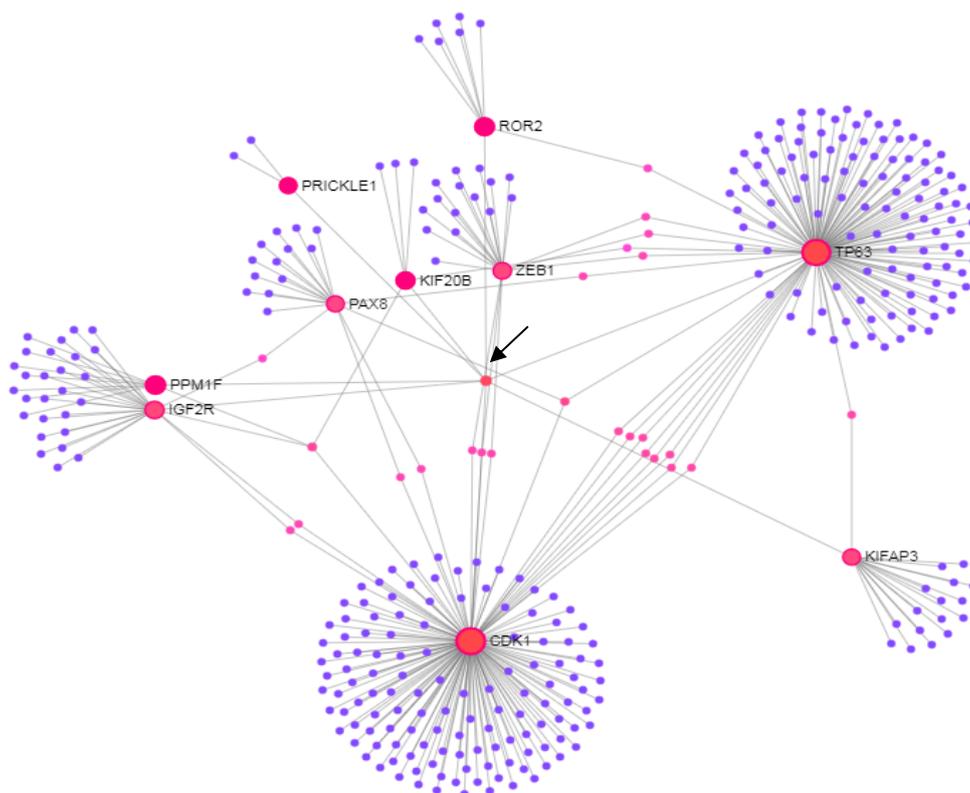


Figura 14 – Análise de redes de interação proteína-proteína dos genes selecionados nas famílias brasileiras. A seta indica a proteína UBC, que interage com a maioria dos genes.

- *Análise de metilação por família*

Foi feita a análise de dados de metilação, comparando-se os níveis de metilação gene-específico para pacientes que possuíam a variante rara para o gene analisado (conforme tabela 2 - o gene *PRICKLE1* e *IGF2R* na família F3196, *PPM1F* em F1843, *KIF20B* e *PAX8* em F2570, *KIFAP3* em F7614, *ROR2* e *ZEB1* em F2848 e *CDK1* em F8418) com os controles. As médias do nível de metilação de pacientes e controles, a diferença entre as médias, o p-valor obtido após correção no teste ANCOVA e no teste de *Tukey* e o número de pacientes e controles que permaneceram após os filtros realizados, nas análises por família por região e por sítio CpG estão representados nas tabelas 11 e 12, respectivamente. Os genes *ROR2* e *IGF2R* não estão presentes na análise por não apresentarem N amostral suficiente de pacientes após os filtros pré-análise (filtro de identidade mínima, número de *reads*).

Tabela 11 – Análise por região por família. Médias e diferença das médias do nível de metilação para pacientes e controles, resultado do teste ANCOVA (valor corrigido por *Tukey*) e número de pacientes com a variante rara e controles.

Família	Gene	Média A (Pacientes)	Média B (Controles)	Diferença (Média A - Média B)	P-valor	Número Pacientes	Número Controles
F8418	<i>CDK1</i>	0.028	0.030152	-0.002152	0.888115	1	46
F2570	<i>KIF20B</i>	0.0015	0.001703	-0.000203	0.588037	4	64
F7614	<i>KIFAP3</i>	0.005	0.00514	-0.000140	0.776606	2	64
F2570	<i>PAX8_CpG25</i>	0.156	0.158375	-0.002374	0.857449	1	24
F2570	<i>PAX8_CpG27</i>	0.0055	0.00647	-0.000976	0.154649	4	65
F2570	<i>PAX8_CpG131</i>	0.03766	0.04011	-0.002444	0.845158	3	18
F1843	<i>PPM1F_CpG21</i>	0.02	0.02068	-0.000682	0.482970	3	41
F3196	<i>PRICKLE1_CpG107</i>	0.003	0.00331	-0.000313	0.343963	3	67
F3196	<i>PRICKLE1_CpG165</i>	0.00266	0.00371	-0.001049	0.074022	3	67
F2848	<i>ZEB1</i>	0.009	0.00682	0.0021702	0.871247	1	47

Tabela 12 – Análise por sítio CpG por família. Médias e diferença das médias do nível de metilação para pacientes e controles, resultado do teste ANCOVA (valor corrigido por *Tukey*; apenas resultados com p-valores abaixo de 0,1) e número de pacientes com a variante rara e controles.

Gene	Sítio	Média A (Pacientes)	Média B (Controles)	Diferença (Média A - Média B)	P-valor	Número Pacientes	Número Controles
<i>CDK1</i>	Site.13	0.052381	0.011085	0.041296	0.026422*	1	46
<i>PAX8_CpG25</i>	Site.6	0.005917	0.139860	-0.133942	0.063084	1	24
<i>PAX8_CpG27</i>	Site.5	0.005230	0.008364	-0.003134	0.092894	4	65
	Site.7	0.003435	0.005410	-0.001975	0.07549	4	65
<i>PAX8_CpG131</i>	Site.4	0.000000	0.029619	-0.029619	0.022668*	2	18
	Site.7	0.080000	0.026749	0.053251	0.036220*	2	18
	Site.9	0.054233	0.006257	0.047976	0.00047**	2	18
<i>PPM1F_CpG21</i>	Site.16	0.023810	0.000000	0.023810	0.00033**	3	42
<i>PRICKLE1_CpG107</i>	Site.10	0.000000	0.002848	-0.002848	0.086125	3	67
<i>ZEB1</i>	Site.18	0.045455	0.009790	0.035665	0.085764	1	47
	Site.21	0.125000	0.008994	0.116006	3.1e-10 ***	1	47

Para a análise de metilação por região (Tabela 11), não foram encontrados valores significativos dentro dos parâmetros de análise.

Quanto à análise por sítio CpG (Tabela 12), foram encontrados valores significativos para o gene *PAX8* CpG 131 em 3 sítios, porém um deles apresenta menos que 4% de diferença entre paciente e controle e não há um mesmo padrão observado nos 3 sítios, uma vez que dois estão hipermetilados e um está hipometilado. Dessa forma, tais dados não foram considerados como relevantes no presente trabalho.

Os genes *CDK1* e *ZEB1* apresentaram 1 sítio significativo com diferenças de 4,1% e 11,6% respectivamente, entretanto, não foram encontrados 3 sítios significativos, conforme parâmetros de análise e tais genes não foram considerados significativos.

4.2.2. Análise de metilação na família F886

Tendo em vista que os pacientes da família F886 haviam sido incluídos na análise de metiloma (Alvizi *et al.*, 2016), estes dados foram reanalisados com foco nas regiões encontradas no estudo de ligação e reanálise de SCE dessa família e, portanto, foram observados os valores de metilação em *DCHS2*. Foram observadas duas regiões significativas do gene *DCHS2* com 11% de diferença de metilação, sendo uma delas

exônica com p-valor de 0,0049 e a outra em região intergênica a 200kb de *DCHS2*, com p-valor de 0,039 (Tabela 13). Foi observada uma tendência à hipermetilação nos pacientes nas duas posições.

Tabela 13 – Dados de metiloma para os pacientes da família F886 *versus* controles

Posição (hg19)	Tipo	Média A (Pacientes)	Média B (Controles)	Diferença (Média A – Média B)	P-valor
chr4:155411603	Exônica	0.169820029	0.059258869	0.11056116	0.00490665
chr4:155662375	Intrônica	0.209916599	0.096369018	0.113547581	0.039813565

5. Discussão

O modelo mais provável para FL/P NS é o de herança multifatorial (Grosen *et al.*, 2010; Dixon *et al.*, 2011). Entretanto, apenas as variantes comuns não explicam a alta herdabilidade da doença e foi sugerido que variantes raras de efeito médio ou grande estejam contribuindo para algumas doenças complexas (Manolio *et al.*, 2009; Pritchard *et al.*, 2002).

Por meio do sequenciamento completo de exoma (SCE), alguns trabalhos têm encontrado variantes raras em famílias multiplex com FL/P NS, como por exemplo, no gene *CDH1* (Brito *et al.*, 2015). Entretanto, há limitações na busca por variantes raras. Devido à baixa frequência das variantes (menor que 1%), há dificuldades na replicação dos resultados em outras famílias e/ou populações. Além disso, não se sabe quantos e quais genes estão de fato envolvidos na etiologia da FL/P NS, o que dificulta na decisão da priorização das variantes.

Tendo em vista que as FL/Ps tem padrão de herança multifatorial, pesquisadores têm estudado se a presença ou não do fenótipo pode estar sendo influenciada pela alteração do padrão epigenético nos genes envolvidos nessas doenças. Alguns trabalhos têm demonstrado importância do estudo da epigenética nas FL/Ps (Alvizi *et al.*, 2017; Sharp *et al.*, 2017; Gonseth *et al.*, 2019). Com base nisso, avaliamos se existe uma alteração de metilação mais frequente em genes selecionados como candidatos à FL/P NS ao compararmos pacientes e controles, o que poderia contribuir para maior evidência de que esses genes candidatos estejam de fato relacionados ao fenótipo e que a epigenética pode estar influenciando no quadro clínico.

Para tal análise, foram selecionados 21 genes a partir de análise de SCE, painel e estudo de ligação em famílias inglesas e brasileiras. Dentre os 21 genes, *ARHGAP29* já foi associado à FL/P NS (Leslie *et al.*, 2012; Chandrasekharan & Ramanathan, 2014; Savastano *et al.*, 2017), 3 genes já foram associados à formas sindrômicas de fissura, o *TP63*, que foi associado a ectrodactilia (doença congênita que afeta a formação das mãos e dos pés (OMIM#604292)), o *SHH*, que foi associado a holoprosencefalia (doença congênita caracterizada pela clivagem incompleta do prosencéfalo (OMIM#142945)) e

ROR2, que foi associado a Síndrome de Robinow (encurtamento dos membros e anomalias do crânio (OMIM#268310)) e braquidactilia (encurtamento dos dedos (OMIM#113000)). Os demais genes nunca haviam sido associados à FL/P, de forma que as evidências que essas variantes sejam causativas do quadro clínico são ainda bastante preliminares.

Na nossa casuística, a análise de metilação das ilhas CpG da região promotora dos 20 genes testados (com exceção de *DCHS2*, que será descrito posteriormente) revelou como relevantes os genes *TP63* e *ROR2*.

No caso de *TP63*, encontramos um resultado consistente em relação à tendência à hipermetilação desse gene, que foi observada tanto na análise por região, quanto na análise por sítio CpG, na amostra global de pacientes. O presente trabalho encontrou níveis de metilação mais altos na análise por região (p-valor=0.0251912) e na análise por sítio CpG (de 4 a 10% de metilação em 4 sítios consecutivos) em pacientes com FL/P NS no gene *TP63*, indicando que o promotor desse gene pode estar hipermetilado e sua expressão diminuída nos pacientes, mesmo efeito encontrado no gene *CDH1* (Alvizi *et al.*, 2017). A análise por sítio pode não ter sido significativa devido ao número amostral analisado.

Na análise de redes de interação proteína-proteína, foi observado que o gene *TP63* interage com os genes *CDK1*, *PRICKLE1*, *ROR2*, *KIFAP3*, *KIF20B*, *PPM1F* e *IGF2R*, por meio de uma proteína, a poliubiquitina-C (UBC). A poliubiquitina está envolvida em diferentes processos, como reparo de DNA, regulação do ciclo celular e degradação de proteínas por meio de proteossoma (Hung *et al.*, 2006; Komander, 2009). Vale ressaltar que, dentre os 28 pacientes analisados, 18 apresentam alguma variante para os 7 genes que interagem com *TP63*. Apesar de não se ter conhecimento sobre como ocorre essa interação por meio da UBC, uma possível explicação é que os níveis de *TP63* estejam desregulados em uma fase crítica do desenvolvimento craniofacial em função da alteração de metilação, o que pode ter influência sobre os genes que apresentam variantes nessas famílias, ou seja, além da variante, a hipermetilação de *TP63* pode ser um evento adicional para o quadro clínico.

O gene *TP63* (*Tumor protein p63*) codifica a proteína p63, que age como um fator de transcrição, exercendo regulação transcricional de vários genes e micro-RNAs, responsáveis pela adesão celular, apoptose, proliferação e diferenciação (Papakonstantinou *et al.*, 2019; Su *et al.*, 2013). Foi observada hipometilação e superexpressão de *TP63* em pacientes com leucemia crônica e acredita-se que mudanças sutis em sua expressão podem ter papel decisivo no destino da célula (Papakonstantinou *et al.*, 2019). O *TP63* participa de diversas vias de sinalização, como Notch e Wnt (Yoh & Prywes, 2015).

A via Notch é importante para o controle da diferenciação epidermal. A ativação desta via pode reprimir ou ativar p63, dependendo da célula (Yalcin-Ozuysal *et al.*, 2010; Ross & Kadesch, 2004).

Em relação à via Wnt, um grupo de trabalho sugeriu que a rede Wnt-p63-*IRF6*, dependente de proteínas Pbx, controla o desenvolvimento da ectoderme do terço médio da face e que a desregulação dessa rede leva à formação anormal do lábio e do palato primário em camundongos (Ferreti *et al.*, 2011). Além disso, foi observado que *IRF6* (gene com variantes já descritas para FL/P; Kondo *et al.*, 2002; Zucchero *et al.*, 2004; Beaty *et al.*, 2010) e *TP63* interagem epistaticamente em camundongos com fissura do palato heterozigotos para ambos genes (Thomason *et al.*, 2010). A proteína p63 ativa transcionalmente o gene *IRF6* que, quando ativado, induz a regulação negativa de p63 mediada por proteossoma, limitando o potencial proliferativo de queratinócitos (Botti *et al.*, 2011; Rotondo *et al.*, 2016).

Dessa forma, a redução da expressão do *TP63* e, conseqüentemente, da proteína p63, pode estar influenciando nas sinalizações das vias Wnt e Notch, assim como na expressão de *IRF6*.

Além do gene *TP63*, julgamos também relevante destacar os resultados obtidos para *ROR2*, uma vez que foram observados valores estatisticamente significativos para *ROR2* na análise de metilação por sítio CpG na amostra global. Foi observada uma tendência à hipometilação nos pacientes na análise por sítio CpG. Uma vez que não

foram observados resultados significativos na análise por região, é necessário um estudo com maior tamanho amostral.

O gene *ROR2* (*Receptor Tyrosine Kinase-Like Orphan Receptor 2*) tem participação nas vias de sinalização Wnt canônica e não-canônica. A via de sinalização Wnt regula diversos eventos durante o desenvolvimento embrionário, como organogênese, diferenciação, polarização e migração celular (Komiya *et al.*, 2008). Dentre as vias não-canônicas está a via PCP, que inicia eventos da via de polaridade celular planar (PCP). A via PCP tem papel na polaridade celular e tecidual, controlando eventos como a migração de células da crista neural e promovendo direções que controlam e coordenam a integração celular que levam à construção do indivíduo (Sebbagh & Borg, 2014; De Calisto *et al.*, 2005). Estudos demonstram que a desregulação na sinalização da via Wnt implica na malformação de estruturas craniofaciais em humanos (Chiquet *et al.*, 2008; Menezes *et al.*, 2010). Alvizi *et al.* (2017) observou um padrão de metilação distinto e enriquecido em regiões potencialmente regulatórias e ativas no genoma, incluindo em genes que participam da via Wnt.

Variantes comuns de *ROR2* já foram associadas à fissura do palato em humanos (Wang *et al.*, 2012). Além disso, *ROR2* é expresso durante a formação craniofacial de camundongos (Topczewski *et al.*, 2011).

A expressão do gene *ROR2* é regulada por mudança de metilação do promotor, sendo seu promotor progressivamente metilado durante a diferenciação das células-tronco mesenquimais em osteoblastos (Tarfiei *et al.*, 2011).

Tais evidências demonstram que a regulação correta do nível de metilação no promotor de *ROR2* tem importância no seu funcionamento normal nos tipos celulares estudados.

Entretanto, o gene *ROR2* apresentou valores de metilação inferiores para os pacientes aqui estudados. Pesquisadores sugerem que a hipometilação do promotor do gene produz o efeito contrário da hipermetilação, ou seja, a superexpressão do gene, em células tumorais (Husni *et al.*, 2019; Shen *et al.*, 2019). É provável que a hipometilação de *ROR2* esteja aumentando sua expressão, quando em uma situação normal, estaria

diminuída, afetando sua regulação e, conseqüentemente, passos da sinalização da via Wnt.

Em resumo, esses resultados sugerem que alteração de metilação em *TP63* e *ROR2* pode estar contribuindo para a FL/P NS. É possível que essa alteração de metilação represente um evento adicional além da mutação encontrada nos genes candidatos aqui selecionados ou outros fatores genéticos ainda não identificados.

Para que se analise o efeito da metilação de DNA sobre a expressão dos genes, é necessário, para os próximos passos, a realização de estudos que correlacionem os níveis de RNA mensageiro com a metilação de DNA para *TP63* e *ROR2*.

Em uma segunda análise, visamos verificar se havia alteração de metilação nos indivíduos portadores com mutação patogênica nos genes candidatos selecionados. Para isso, foi feita a análise com os dados de metilação relacionando as famílias e os genes selecionados para elas na análise de SCE (Tabela 2). Tendo em vista que não foram encontrados no mínimo 3 sítios significativos ou a diferença da média entre pacientes e controles era menor que 4%, consideramos que não foi encontrado nenhum gene significativo nesta análise.

Para os 20 genes testados, com exceção de *TP63* e *ROR2*, não foram encontrados resultados significativos na análise de metilação, o que sugere que não há alteração no padrão de metilação no promotor desses genes nos pacientes testados. Dessa forma, tais genes não apresentaram evidências adicionais de estarem envolvidos com a predisposição de FL/Ps na casuística do presente trabalho. Tais resultados podem ser devido ao tamanho amostral estudado.

Por fim, a família F886, estudada no presente trabalho, é proveniente de Cidade de Cruz (Ceará) e se destaca pelo número de afetados e presença de consanguinidade entre alguns dos pais de pacientes. Uma das possíveis dificuldades na detecção do gene desta família é de que pode haver mais de um gene atuando nesta família, uma vez que II-3 e II-4 são aparentemente de famílias distintas, não consanguíneas. Ainda, como pais e avós de V-1 não têm FL/P e essa malformação é relativamente comum, deve-se também levar em conta que a etiologia de FL/P em V-1 possa ser diferente da maioria

dos outros casos na família. O estudo de ligação na família F886, que foi realizado na tentativa de melhor dissecarmos a arquitetura genética dessa família, revelou a região contendo o gene *DCHS2* como a mais provável de ter o loco de risco para a FL/P nessa família. Além disso, encontramos variantes para o gene *DCHS2* na reanálise de SCE. A distância entre as variantes é de 22 kb e foram observados os mesmos genótipos para as duas variantes em todos os indivíduos, o que indica que elas poderiam estar segregando no mesmo haplótipo. Exceto pelo paciente V-I, esse haplótipo está segregando com a doença. O resultado observado em V-I pode ser decorrente de um evento de recombinação entre o loco da doença e esses marcadores ou que a FL/P nesse indivíduo tenha outra causa, conforme discutido acima.

Interessantemente, além de observarmos que há um haplótipo na região contendo o *DCHS2* (4q31.3; chr4:155411603-155662375) segregando com o fenótipo, observou-se hipermetilação próxima da região promotora do *DCHS2* nos afetados desta família quando comparados com controles da população brasileira. *DCHS2* (*Dachsous Cadherin-Related 2*) codifica uma caderina e participa da via Fat/Dchs2, que regula a via PCP em *Drosophila* (Le Pabic *et al.*, 2014). Como dito anteriormente, a via PCP tem importância na formação craniofacial e, portanto, a desregulação de expressão desse gene pode estar contribuindo para o fenótipo de FL/P nessa família. Dessa forma, consideramos promissora essa região como candidata à FL/P NS.

6. Conclusões

Com base em tais resultados, podemos concluir que:

- a) *TP63* e *ROR2* estão mais frequentemente com alteração de metilação nos pacientes do que nos controles. Esses resultados sugerem pela 1ª vez, que a metilação nesses genes podem ser um fator de risco para FL/P NS. Entretanto, não sabemos se esse padrão de metilação está associado aos fatores de risco genético sugeridos para essas famílias.
- b) Os genes candidatos selecionados, incluindo o gene *ARGHAP29*, que já é associado à FL/P, não apresentaram alteração do padrão de metilação. Esses resultados podem ser decorrentes do pequeno tamanho amostral e/ou que, exceto pelas variantes em *ARGHAP29*, as variantes selecionadas como de risco representam falsos positivos, ou ainda, que a alteração de metilação em genes de risco para FL/P não seja comum.
- c) O estudo de ligação combinado com estudo de metilação em uma das famílias sugeriu uma nova região de risco para FL/P NS, que inclui o gene *DCHS2*.

7. Resumo

As fissuras labiopalatinas não sindrômicas (FL/P NS) são malformações congênitas que afetam o lábio e/ou o palato. O padrão de herança sugerido para as FL/P NS é o de herança multifatorial. Um dos modelos genéticos mais testados é aquele em que os pacientes apresentam um acúmulo de variantes de baixo risco e frequentes na população. Entretanto, apenas essas variantes não explicam a alta herdabilidade da doença. Foi sugerido, então, que variantes raras de efeito médio e grande estejam atuando na predisposição à FL/P NS, entretanto, a identificação dessas variantes representam um grande desafio. Uma vez que apenas a contribuição genética não explica o fenótipo, tem-se sugerido que a alteração do padrão epigenético possa estar influenciando no quadro clínico dos pacientes. Alguns trabalhos observaram que há alteração de metilação no promotor de genes candidatos em pacientes com FL/P NS, dentre eles, o gene *CDH1*. Dessa forma, o objetivo do presente trabalho foi verificar se os genes selecionados como possíveis causas para FL/P NS têm maior probabilidade de terem alteração no padrão de metilação nos afetados de famílias multiplex do que em não-afetados, o que poderia adicionar evidências de que os genes estejam contribuindo para o fenótipo. Foram estudados o padrão de metilação da região promotora (região global e por sítio) de 16 novos genes candidatos e de 4 genes já associados às FL/Ps (*ARHGAP29*, *SHH*, *TP63* e *ROR2*) em uma casuística brasileira de 28 afetados e 71 controles. Foi observada hipermetilação no gene *TP63* (p-valor=0.025 na análise por região) e hipometilação em *ROR2* (p-valores=0.045, 0.008, 0.043 e 0,0015 em 4 sítios dentro da região CpG). Ambos os genes estão envolvidos na sinalização da via Wnt que, quando desregulada, pode ser um fator de risco para FL/P. Também foi encontrada no presente trabalho uma nova região de risco candidata para FL/P NS, baseado em estudo de ligação, genômica e análise epigenética. *DCHS2*, o melhor candidato dessa região, tem participação na via PCP, o que torna essa região ainda mais promissora. Tais resultados sugerem que a alteração do padrão epigenético nesses genes pode estar influenciando no quadro clínico nas FL/Ps; porém esse estudo não contribuiu para uma melhor classificação da maioria dos novos genes candidatos, o que pode ter sido por limitações do tamanho amostral e metodologia.

Palavras-chave: Fissura labial com ou sem fissura do palato não-sindrômica, FL/P NS, epigenética, metilação, CpG, *TP63*, *ROR2*, *DCHS2*.

8. Abstract

Non-syndromic cleft lip with or without cleft palate (NSCL/P) are congenital malformations that affect the lip and/or palate. Multifactorial is the most likely inheritance pattern for NSCL/P. Accumulation of low risk and frequent variants has been the most tested model in genomic studies. However, just the presence of these variants does not explain the high heritability of the disease. Rare variants of medium and large effects have been suggested as another possible model to be acting on NSCL/P predisposition, however, identification of these variants represents a great challenge. Changes in the epigenetic pattern may be influencing the clinical condition of patients in addition to genetic alterations. In this regard, methylation alterations in the promoter region of candidate genes in patients with NSCL/P, including *CDH1*, have been shown by some authors. Thus, the aim of the present study was to verify if novel candidate genes for NSCL/P selected from genomic studies are more likely to have changes in the methylation pattern in affected individuals from multiplex families than in non-affected ones, which could add more evidences that these genes are good FL/P candidates. We studied the methylation pattern of the promoter region (global region and by site) of 16 novel candidate genes and 4 genes that are already associated with CL/P (*ARHGAP29*, *SHH*, *TP63* e *ROR2*) on a Brazilian cohort of 28 affected individuals and 71 controls. It was observed hypermethylation on *TP63* gene (p-value=0.025 on analysis by region) and hypomethylation on *ROR2* (p-values=0.045, 0.008, 0.043 and 0.0015 on 4 sites within the CpG region). Both genes are involved in Wnt signalization pathway, which when disrupted, can be a risk factor for CL/P. A new candidate risk region for NSCL/P, based on linkage, genomic and epigenetic analysis, was also found in this study. *DCHS2*, the best candidate gene within this region, participates in PCP pathway, which makes this candidate region even more promising. In conclusion, such results suggest that the alteration of the epigenetic pattern in these genes may be influencing the clinical condition observed in NSCL/Ps; however, our study did not contribute in adding more evidence about the pathogenicity status of the prioritized novel candidate variants and genes, which can be due to our limited sample size and methodology.

Keywords: Nonsyndromic cleft lip with or without cleft palate, NSCL/P, epigenetics, methylation, CpG, *TP63*, *ROR2*, *DCHS2*.

9. Referências Bibliográficas

Alvizi L, Ke X, Brito LA, *et al.*, 2017. Differential methylation is associated with non-syndromic cleft lip and palate and contributes to penetrance effects. *Scientific Reports*, 7(1):2441.

Bagordakis E, Paranaiba LMR, Brito LA, *et al.*, 2013. Polymorphisms at regions 1p22.1 (rs560426) and 8q24 (rs1530300) are risk markers for nonsyndromic cleft lip and/or palate in the Brazilian population. *American Journal of Medical Genetics. Part A*, 161A(5):1177–1180.

Bahar Halpern K, Vana T & Walker MD, 2014. Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development. *The Journal of Biological Chemistry*, 289(34):23882–23892.

Basha M, Demeer B, Revencu N, *et al.*, 2018. Whole exome sequencing identifies mutations in 10% of patients with familial non-syndromic cleft lip and/or palate in genes mutated in well-known syndromes. *Journal of Medical Genetics*, 55(7):449–458.

Beaty TH, Murray JC, Marazita ML, *et al.*, 2010. A genome-wide association study of cleft lip with and without cleft palate identifies risk variants near MAFB and ABCA4. *Nature Genetics*, 42(6):525–529.

Benusiglio PR, Caron O, Consolino E, *et al.*, 2013. Cleft lip, cleft palate, hereditary diffuse gastric cancer and germline mutations in CDH1. *International Journal of Cancer*, 132(10):2470.

Bergman D, Halje M, Nordin M & Engström W, 2013. Insulin-like growth factor 2 in development and disease: a mini-review. *Gerontology*, 59(3):240–249.

Berk, NW, Marazita, ML, 2002. The Costs of Cleft Lip and Palate: Personal and Societal Implications. In: Wyszynski, DF., editor. *Cleft Lip and Palate: From Origin to Treatment*. Oxford: Oxford University Press.

Bernstein KA, Gangloff S & Rothstein R, 2010. The RecQ DNA helicases in DNA repair. *Annual Review of Genetics*, 44:393–417.

Birnbaum S, Ludwig KU, Reutter H, *et al.*, 2009. Key susceptibility locus for nonsyndromic cleft lip with or without cleft palate on chromosome 8q24. *Nature Genetics*, 41(4):473–477.

Blake JA & Ziman MR, 2014. Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development*, 141(4):737–751.

Botti E, Spallone G, Moretti F, *et al.*, 2011. Developmental factor IRF6 exhibits tumor suppressor activity in squamous cell carcinomas. *Proceedings of the National Academy of Sciences of the United States of America*, 108(33):13710–13715.

Breuer K, Foroushani AK, Laird MR, *et al.*, 2013. InnateDB: systems biology of innate immunity and beyond--recent updates and continuing curation. *Nucleic Acids Research*, 41(Database issue):D1228-33.

Brito LA, 2016. Exome analysis in multiplex families reveals novel candidate genes for nonsyndromic cleft lip / palate. Tese de doutorado. 29-60.

Brito LA, Cruz LA, Rocha KM, *et al.*, 2011. Genetic contribution for non-syndromic cleft lip with or without cleft palate (NS CL/P) in different regions of Brazil and implications for association studies. *American Journal of Medical Genetics. Part A*, 155A(7):1581–1587.

Brito LA, Paranaíba LMR, Bassi CFS, *et al.*, 2012a. Region 8q24 is a susceptibility locus for nonsyndromic oral clefting in Brazil. *American Journal of Medical Genetics. Part A*, 158A(9):2170–2175.

Brito LA, Meira JGC, Kobayashi GS & Passos-Bueno MR, 2012b. Genetics and management of the patient with orofacial cleft. *Plastic surgery international*, 2012:782821.

Brito LA, Bassi CF, Massotti C, *et al.*, 2012c. *IRF6* is a risk factor for nonsyndromic cleft lip in the Brazilian Population. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 94(6):464–468.

Brito LA, Yamamoto GL, Melo S, *et al.*, 2015. Rare Variants in the Epithelial Cadherin Gene Underlying the Genetic Etiology of Nonsyndromic Cleft Lip with or without Cleft Palate. *Human Mutation*, 36(11):1029–1033.

Calvanese V, Lara E, Kahn A & Fraga MF, 2009. The role of epigenetics in aging and age-related diseases. *Ageing Research Reviews*, 8(4):268–276.

Carlson L, Hatcher KW & Vander Burg R, 2013. Elevated infant mortality rates among oral cleft and isolated oral cleft cases: a meta-analysis of studies from 1943 to 2010. *The Cleft Palate-Craniofacial Journal*, 50(1):2–12.

Chandrasekharan D & Ramanathan A, 2014. Identification of a novel heterozygous truncation mutation in exon 1 of ARHGAP29 in an Indian subject with nonsyndromic cleft lip with cleft palate. *European journal of dentistry*, 8(4):528–532.

Chiquet BT, Blanton SH, Burt A, *et al.*, 2008. Variation in WNT genes is associated with non-syndromic cleft lip with or without cleft palate. *Human Molecular Genetics*, 17(14):2212–2218.

Christensen K & Fogh-Andersen P, 1993. Cleft lip (+/- cleft palate) in Danish twins, 1970-1990. *American Journal of Medical Genetics*, 47(6):910–916.

Cox LL, Cox TC, Moreno Uribe LM, *et al.*, 2018. Mutations in the Epithelial Cadherin-p120-Catenin Complex Cause Mendelian Non-Syndromic Cleft Lip with or without Cleft Palate. *American Journal of Human Genetics*, 102(6):1143–1157.

Crossan E & Duane B, 2018. Is there an association between maternal smoking and oral clefts? *Evidence-based dentistry*, 19(1):24–25.

De Calisto J, Araya C, Marchant L, Riaz CF & Mayor R, 2005. Essential role of non-canonical Wnt signalling in neural crest migration. *Development*, 132(11):2587–2597.

Dixon MJ, Marazita ML, Beaty TH & Murray JC, 2011. Cleft lip and palate: understanding genetic and environmental influences. *Nature Reviews. Genetics*, 12(3):167–178.

Falconer DS, 1965. The inheritance of liability to certain diseases, estimated from the incidence among relatives. *Annals of Human Genetics*, 29(1):51–76.

Faria LA, 2019. Non-syndromic orofacial clefts: enrichment of rare loss-of-functions variants in genes under selective constraint. Tese de doutorado. 21-62.

Feil R & Fraga MF, 2012. Epigenetics and the environment: emerging patterns and implications. *Nature Reviews. Genetics*, 13(2):97–109.

Ferretti E, Li B, Zewdu R, *et al.*, 2011. A conserved Pbx-Wnt-p63-Irf6 regulatory module controls face morphogenesis by promoting epithelial apoptosis. *Developmental Cell*, 21(4):627–641.

Fogh-Andersen P, 1942. *Inheritance of Harelip and Cleft Palate*. Munksgaard.

Fraser FC, 1955. Thoughts on the etiology of clefts of the palate and lip. *Acta genetica et statistica medica*, 5(4):358–369.

Frebourg T, Oliveira C, Hochain P, *et al.*, 2006. Cleft lip/palate and CDH1/E-cadherin mutations in families with hereditary diffuse gastric cancer. *Journal of Medical Genetics*, 43(2):138–142.

Fuentes Fajardo KV, Adams D, NISC Comparative Sequencing Program, *et al.*, 2012. Detecting false-positive signals in exome sequencing. *Human Mutation*, 33(4):609–613.

Gonseth S, Shaw GM, Roy R, *et al.*, 2019. Epigenomic profiling of newborns with isolated orofacial clefts reveals widespread DNA methylation changes and implicates metastable epiallele regions in disease risk. *Epigenetics*, 14(2):198–213.

Grady WM, Willis J, Guilford PJ, *et al.*, 2000. Methylation of the CDH1 promoter as the second genetic hit in hereditary diffuse gastric cancer. *Nature Genetics*, 26(1):16–17.

Gritli-Linde A, 2007. Molecular control of secondary palate development. *Developmental Biology*, 301(2):309–326.

Grosen D, Chevrier C, Skytthe A, *et al.*, 2010. A cohort study of recurrence patterns among more than 54,000 relatives of oral cleft cases in Denmark: support for the multifactorial threshold model of inheritance. *Journal of Medical Genetics*, 47(3):162–168.

Gundlach KKH & Maus C, 2006. Epidemiological studies on the frequency of clefts in Europe and world-wide. *Journal of cranio-maxillo-facial surgery : official publication of the European Association for Cranio-Maxillo-Facial Surgery*, 34 Suppl 2:1–2.

Horsburgh S, Robson-Ansley P, Adams R & Smith C, 2015. Exercise and inflammation-related epigenetic modifications: focus on DNA methylation. *Exercise Immunology Review*, 21:26–41.

Howe LJ, Richardson TG, Arathimos R, *et al.*, 2019. Evidence for DNA methylation mediating genetic liability to non-syndromic cleft lip/palate. *Epigenomics*, 11(2):133–145.

Huang F, Kirkpatrick D, Jiang X, Gygi S & Sorkin A, 2006. Differential regulation of EGF receptor internalization and degradation by multiubiquitination within the kinase domain. *Molecular Cell*, 21(6):737–748.

Husni RE, Shiba-Ishii A, Nakagawa T, *et al.*, 2019. DNA hypomethylation-related overexpression of SFN, GORASP2 and ZYG11A is a novel prognostic biomarker for early stage lung adenocarcinoma. *Oncotarget*, 10(17):1625–1636.

Ira G, Pellicoli A, Balijja A, *et al.*, 2004. DNA end resection, homologous recombination and DNA damage checkpoint activation require CDK1. *Nature*, 431(7011):1011–1017.

Jaenisch R & Bird A, 2003. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genetics*, 33 Suppl:245–254.

Janisch KM, Vock VM, Fleming MS, *et al.*, 2013. The vertebrate-specific Kinesin-6, Kif20b, is required for normal cytokinesis of polarized cortical stem cells and cerebral cortex size. *Development*, 140(23):4672–4682.

Jiang R, Bush JO & Lidral AC, 2006. Development of the upper lip: morphogenetic and molecular mechanisms. *Developmental Dynamics*, 235(5):1152–1166.

Jugessur A, Farlie PG & Kilpatrick N, 2009. The genetics of isolated orofacial clefts: from genotypes to subphenotypes. *Oral Diseases*, 15(7):437–453.

Jugessur A, Shi M, Gjessing HK, *et al.*, 2011. Fetal genetic risk of isolated cleft lip only versus isolated cleft lip and palate: a subphenotype analysis using two population-based studies of orofacial clefts in Scandinavia. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 91(2):85–92.

Kelly D, O’Dowd T & Reulbach U, 2012. Use of folic acid supplements and risk of cleft lip and palate in infants: a population-based cohort study. *The British Journal of General Practice*, 62(600):e466-72.

Kim B-R, Kwon Y & Rho SB, 2017. BMI-1 interacts with sMEK1 and inactivates sMEK1-induced apoptotic cell death. *Oncology Reports*, 37(1):579–586.

Kobayashi GS, Alvizi L, Sunaga DY, *et al.*, 2013. Susceptibility to DNA damage as a molecular mechanism for non-syndromic cleft lip and palate. *Plos One*, 8(6):e65677.

Komander D, 2009. The emerging complexity of protein ubiquitination. *Biochemical Society Transactions*, 37(Pt 5):937–953.

Komiya Y & Habas R, 2008. Wnt signal transduction pathways. *Organogenesis*, 4(2):68–75.

Kondo S, Schutte BC, Richardson RJ, *et al.*, 2002. Mutations in IRF6 cause Van der Woude and popliteal pterygium syndromes. *Nature Genetics*, 32(2):285–289.

Kruglyak L, 2008. The road to genome-wide association studies. *Nature Reviews. Genetics*, 9(4):314–318.

Le Pabic P, Ng C & Schilling TF, 2014. Fat-Dachsous signaling coordinates cartilage differentiation and polarity during craniofacial development. *PLoS Genetics*, 10(10):e1004726.

Leslie EJ, Carlson JC, Shaffer JR, *et al.*, 2016. A multi-ethnic genome-wide association study identifies novel loci for non-syndromic cleft lip with or without cleft palate on 2p24.2, 17q23 and 19q13. *Human Molecular Genetics*, 25(13):2862–2872.

Leslie EJ, Mansilla MA, Biggs LC, *et al.*, 2012. Expression and mutation analyses implicate ARHGAP29 as the etiologic gene for the cleft lip with or without cleft palate locus identified by genome-wide association on chromosome 1p22. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 94(11):934–942.

Leslie EJ & Marazita ML, 2013. Genetics of cleft lip and cleft palate. *American Journal of Medical Genetics. Part C, Seminars in Medical Genetics*, 163C(4):246–258.

Loffredo LC, Souza JM, Freitas JA & Mossey PA, 2001. Oral clefts and vitamin supplementation. *The Cleft Palate-Craniofacial Journal*, 38(1):76–83.

Machado JC, Oliveira C, Carvalho R, *et al.*, 2001. E-cadherin gene (CDH1) promoter methylation as the second hit in sporadic diffuse gastric carcinoma. *Oncogene*, 20(12):1525–1528.

Machado RA, Nogueira EN, Martelli-Júnior H, *et al.*, 2018. 2p24.2 (rs7552) is a susceptibility locus for nonsyndromic cleft lip with or without cleft palate in the Brazilian population. *Clinical Genetics*, 93(6):1199–1204.

Malumbres M, 2014. Cyclin-dependent kinases. *Genome Biology*, 15(6):122.

Manolio TA, Collins FS, Cox NJ, *et al.*, 2009. Finding the missing heritability of complex diseases. *Nature*, 461(7265):747–753.

Marazita ML, 2012. The evolution of human genetic studies of cleft lip and cleft palate. *Annual Review of Genomics and Human Genetics*, 13:263–283.

Marazita ML, Murray JC, Lidral AC, *et al.*, 2004. Meta-analysis of 13 genome scans reveals multiple cleft lip/palate genes with novel loci on 9q21 and 2q32-35. *American Journal of Human Genetics*, 75(2):161–173.

Masotti C, Brito LA, Nica AC, *et al.*, 2018. MRPL53, a New Candidate Gene for Orofacial Clefting, Identified Using an eQTL Approach. *Journal of Dental Research*, 97(1):33–40.

Masser DR, Berg AS & Freeman WM, 2013. Focused, high accuracy 5-methylcytosine quantitation with base resolution by benchtop next-generation sequencing. *Epigenetics & Chromatin*, 6(1):33.

Menegotto BG & Salzano FM, 1991. Epidemiology of oral clefts in a large South American sample. *The Cleft Palate-Craniofacial Journal*, 28(4):373–6; discussion 376.

Menezes R, Letra A, Kim AH, *et al.*, 2010. Studies with Wnt genes and nonsyndromic cleft lip and palate. *Birth Defects Research. Part A, Clinical and Molecular Teratology*, 88(11):995–1000.

Mitchell LE, 2002. Mode of Inheritance of Oral Clefts. In DF Wyszynski (ed.), *Cleft lip and palate: From origin to treatment*. . Oxford: Oxford University Press, p. 234–239.

Mochizuki K, Hariya N & Kubota T, 2018. Novel models of epigenetic gene regulation in the nutritional environment. *Advances in Experimental Medicine and Biology*, 1012:11–18.

Mossey PA, Little J, Munger RG, Dixon MJ & Shaw WC, 2009. Cleft lip and palate. *The Lancet*, 374(9703):1773–1785.

Murray JC, Daack-Hirsch S, Buetow KH, *et al.*, 1997. Clinical and epidemiologic studies of cleft lip and palate in the Philippines. *The Cleft Palate-Craniofacial Journal*, 34(1):7–10.

Naslavsky MS, Yamamoto GL, De Almeida TF, *et al.*, 2017. Exomic variants of an elderly cohort of Brazilians in the ABraOM database. *Human Mutation*, 38(7):751–763.

Papakonstantinou N, Ntoufa S, Tsagiopoulou M, *et al.*, 2019. Integrated epigenomic and transcriptomic analysis reveals TP63 as a novel player in clinically aggressive chronic lymphocytic leukemia. *International Journal of Cancer*, 144(11):2695–2706.

Paranaíba, LM, Martelli-Júnior H, Oliveira Swerts MS, *et al.*, 2008. Novel mutations in the IRF6 gene in Brazilian families with Van der Woude syndrome. *International Journal of Molecular Medicine*, 22 (4):507–511.

Patterson K, Molloy L, Qu W & Clark S, 2011. DNA methylation: bisulphite modification and analysis. *Journal of Visualized Experiments*, (56).

Peinado H, Olmeda D & Cano A, 2007. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nature Reviews. Cancer*, 7(6):415–428.

Phang H-Q, Hoon J-L, Lai S-K, *et al.*, 2014. POPX2 phosphatase regulates the KIF3 kinesin motor complex. *Journal of Cell Science*, 127(Pt 4):727–739.

Pritchard JK & Cox NJ, 2002. The allelic architecture of human disease genes: common disease-common variant...or not? *Human Molecular Genetics*, 11(20):2417–2423.

Reich DE & Lander ES, 2001. On the allelic spectrum of human disease. *Trends in Genetics*, 17(9):502–510.

Rodrigues K, Sena MF De, Roncalli AG & Ferreira MAF, 2009. Prevalence of orofacial clefts and social factors in Brazil. *Brazilian oral research*, 23(1):38–42.

Ross DA & Kadesch T, 2004. Consequences of Notch-mediated induction of Jagged1. *Experimental Cell Research*, 296(2):173–182.

Rotondo JC, Borghi A, Selvatici R, *et al.*, 2016. Hypermethylation-Induced Inactivation of the IRF6 Gene as a Possible Early Event in Progression of Vulvar Squamous Cell Carcinoma Associated With Lichen Sclerosus. *JAMA dermatology*, 152(8):928–933.

Savastano CP, Brito LA, Faria AC, *et al.*, 2017. Impact of rare variants in ARHGAP29 to the etiology of oral clefts: role of loss-of-function vs missense variants. *Clinical Genetics*, 91(5):683–689.

Schork NJ, Murray SS, Frazer KA & Topol EJ, 2009. Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development*, 19(3):212–219.

Schutte BC & Murray JC, 1999. The many faces and factors of orofacial clefts. *Human Molecular Genetics*, 8(10):1853–1859.

Sebbagh M & Borg J-P, 2014. Insight into planar cell polarity. *Experimental Cell Research*, 328(2):284–295.

Sharp GC, Ho K, Davies A, *et al.*, 2017. Distinct DNA methylation profiles in subtypes of orofacial cleft. *Clinical epigenetics*, 9:63.

Shen Z, Zhu L, Zhang C, Cui X & Lu J, 2019. Overexpression of BHLHE41, correlated with DNA hypomethylation in 3'UTR region, promotes the growth of human clear cell renal cell carcinoma. *Oncology Reports*, 41(4):2137–2147.

Skvortsova K, Stirzaker C & Taberlay P, 2019. The DNA methylation landscape in cancer. *Essays in biochemistry*,

Sripathomsawat W, Tanpaiboon P, Heering J, *et al.*, 2011. Phenotypic analysis of Arg227 mutations of TP63 with emphasis on dental phenotype and micturition difficulties in EEC syndrome. *American Journal of Medical Genetics. Part A*, 155A(1):228–232.

Stanier P & Moore GE, 2004. Genetics of cleft lip and palate: syndromic genes contribute to the incidence of non-syndromic clefts. *Human Molecular Genetics*, 13 Spec No 1:R73-81.

Stuppia L, Capogreco M, Marzo G, *et al.*, 2011. Genetics of syndromic and nonsyndromic cleft lip and palate. *The Journal of Craniofacial Surgery*, 22(5):1722–1726.

Su X, Chakravarti D & Flores ER, 2013. p63 steps into the limelight: crucial roles in the suppression of tumorigenesis and metastasis. *Nature Reviews. Cancer*, 13(2):136–143.

Taib BG, Taib AG, Swift AC & Van Eeden S, 2015. Cleft lip and palate: diagnosis and management. *British Journal of Hospital Medicine*, 76(10):584–5, 588.

Tammen SA, Friso S & Choi S-W, 2013. Epigenetics: the link between nature and nurture. *Molecular aspects of medicine*, 34(4):753–764.

Tanuma N, Nomura M, Ikeda M, *et al.*, 2009. Protein phosphatase Dusp26 associates with KIF3 motor and promotes N-cadherin-mediated cell-cell adhesion. *Oncogene*, 28(5):752–761.

Tarfiei G, Noruzinia M, Soleimani M, *et al.*, 2011. ROR2 promoter methylation change in osteoblastic differentiation of mesenchymal stem cells. *Cell journal*, 13(1):11–15.

Tetreault M, Bareke E, Nadaf J, Alirezaie N & Majewski J, 2015. Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. *Expert Review of Molecular Diagnostics*, 15(6):749–760.

Thomason HA, Zhou H, Kouwenhoven EN, *et al.*, 2010. Cooperation between the transcription factors p63 and IRF6 is essential to prevent cleft palate in mice. *The Journal of Clinical Investigation*, 120(5):1561–1569.

Topczewski J, Dale RM & Sisson BE, 2011. Planar cell polarity signaling in craniofacial development. *Organogenesis*, 7(4):255–259.

Twigg SRF & Wilkie AOM, 2015. New insights into craniofacial malformations. *Human Molecular Genetics*, 24(R1):R50-9.

Van Bokhoven H, Hamel BC, Bamshad M, *et al.*, 2001. p63 Gene mutations in eec syndrome, limb-mammary syndrome, and isolated split hand-split foot malformation suggest a genotype-phenotype correlation. *American Journal of Human Genetics*, 69(3):481–492.

Verlinsky Y, Rechitsky S, Verlinsky O, *et al.*, 2003. Preimplantation diagnosis for sonic hedgehog mutation causing familial holoprosencephaly. *The New England Journal of Medicine*, 348(15):1449–1454.

Wang H, Hetmanski JB, Ruczinski I, *et al.*, 2012. ROR2 gene is associated with risk of non-syndromic cleft palate in an Asian population. *Chinese Medical Journal*, 125(3):476–480.

Weller PA, Ogryzko EP, Corben EB, *et al.*, 1990. Complete sequence of human vinculin and assignment of the gene to chromosome 10. *Proceedings of the National Academy of Sciences of the United States of America*, 87(15):5667–5671.

Westendorf JJ, Kahler RA & Schroeder TM, 2004. Wnt signaling in osteoblasts and bone diseases. *Gene*, 341:19–39.

Wilcox AJ, Lie RT, Solvoll K, *et al.*, 2007. Folic acid supplements and risk of facial clefts: national population based case-control study. *BMJ (Clinical Research Ed.)*, 334(7591):464.

Xu D-P, Qu W-D, Sun C, *et al.*, 2018. A Study on Environmental Factors for Nonsyndromic Cleft Lip and/or Palate. *The Journal of Craniofacial Surgery*, 29(2):364–367.

Yalcin-Ozuysal O, Fiche M, Guitierrez M, *et al.*, 2010. Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. *Cell Death and Differentiation*, 17(10):1600–1612.

Yoh K & Prywes R, 2015. Pathway Regulation of p63, a Director of Epithelial Cell Fate. *Frontiers in endocrinology*, 6:51.

Yu Y, Zuo X, He M, *et al.*, 2017. Genome-wide analyses of non-syndromic cleft lip with palate identify 14 novel loci and genetic heterogeneity. *Nature Communications*, 8:14364.

Zucchero TM, Cooper ME, Maher BS, *et al.*, 2004. Interferon regulatory factor 6 (IRF6) gene variants and the risk of isolated cleft lip or palate. *The New England Journal of Medicine*, 351(8):769–780.

10. Anexos

Anexo I – Amostras de controles, pacientes e familiares selecionados para o presente trabalho e seus dados. Código dos heredogramas: Código dos indivíduos correspondentes aos heredogramas da Figura 8. Tecido: 1-Sangue, 2-Mucosa Oral; Idade: 1-De 0 a 10 anos, 2-De 11 a 20 anos, 3-De 21 a 30 anos, 4-De 31 a 40 anos, 5-De 41 a 50 anos, 6-De 51 a 60 anos, 7-Acima de 60 anos; Sexo: 1-Feminino, 2-Masculino.

Indivíduo	Código dos heredogramas	Grupo	Tecido	Idade	Sexo
c.26039	-	Controle	1	5	1
c.26053	-	Controle	1	5	1
c.26054	-	Controle	1	7	1
c.26078	-	Controle	1	7	1
c.26101	-	Controle	1	5	1
c.26177	-	Controle	1	5	1
c.26219	-	Controle	1	6	1
c.26232	-	Controle	1	6	2
c.26390	-	Controle	1	7	1
c.27368	-	Controle	1	6	2
c.27705	-	Controle	1	6	2
c.27769	-	Controle	1	5	1
c.27848	-	Controle	1	5	1
c.27851	-	Controle	1	6	2
c.27931	-	Controle	1	6	1
c.28039	-	Controle	1	6	1
c107	-	Controle	1	7	2
c111	-	Controle	1	6	1
c112	-	Controle	1	7	2
c130	-	Controle	1	6	1
c131	-	Controle	1	2	1
c132	-	Controle	1	5	1
c137	-	Controle	1	5	2
c142	-	Controle	1	5	1
c156	-	Controle	1	7	2
c18	-	Controle	1	1	1
c3	-	Controle	1	1	2
c38	-	Controle	1	1	2
c39	-	Controle	1	1	2
c40	-	Controle	1	1	2
c41	-	Controle	1	1	2
c43	-	Controle	1	1	2
c44	-	Controle	1	1	2
c46	-	Controle	1	1	1
c48	-	Controle	1	1	2
c50	-	Controle	1	1	1
F9350-1	-	Controle	2	6	1
F9351-1	-	Controle	2	4	1
F9357-1	-	Controle	2	5	2

F9365-1	-	Controle	2	6	1
F9399-1	-	Controle	2	5	1
F9400-1	-	Controle	2	5	2
F9405-1	-	Controle	2	7	2
F9412-1	-	Controle	2	7	2
F9413-1	-	Controle	2	5	2
F9414-1	-	Controle	2	7	2
F9434-1	-	Controle	2	6	2
F9467-1	-	Controle	2	7	1
F1843-7	II-3	Portador não-afetado	1	4	2
F7614-2	II-7	Portador não-afetado	1	3	2
F8418-3	II-7	Portador não-afetado	1	5	2
F10318-1	-	Paciente	2	1	2
F1843-1	II-8	Paciente	1	2	1
F1843-2	II-5	Paciente	1	4	1
F1843-5	III-1	Paciente	1	2	1
F185-1	-	Paciente	1	2	2
F2549-1	-	Paciente	2	2	2
F2570-1	II-14	Paciente	1	4	1
F2570-4	III-4	Paciente	1	2	2
F2570-5	III-3	Paciente	1	2	1
F2570-6	II-6	Paciente	1	5	1
F2848-1	II-6	Paciente	2	1	2
F2848-3	I-2	Paciente	2	5	2
F3196-1	III-2	Paciente	1	3	2
F3196-2	II-2	Paciente	1	6	1
F3196-3	III-3	Paciente	1	3	2
F6543-1	-	Paciente	2	3	1
F7391-1	-	Paciente	NA	6	1
F7614-1	III-6	Paciente	1	1	2
F7614-4	I-4	Paciente	1	5	2
F7614-5	I-3	Paciente	1	4	2
F7867-1	-	Paciente	2	3	2
F8104-1	-	Paciente	2	2	1
F8418-1	III-7	Paciente	1	2	1
F8418-4	II-5	Paciente	1	4	2
F8418-7	II-3	Paciente	2	NA	2
F8422-1	-	Paciente	2	2	2
F9112-1	-	Paciente	2	1	1
F9113-1	-	Paciente	2	1	2
F1843-10	I-3	Familiar	1	6	1
F1843-4	III-3	Familiar	1	1	2
F1843-6	III-2	Familiar	1	1	1
F1843-8	II-4	Familiar	1	4	1
F1843-9	I-2	Familiar	1	6	2

F2570-2	I-2	Familiar	1	7	1
F2570-3	II-11	Familiar	1	4	1
F2570-7	II-5	Familiar	1	5	2
F2848-2	I-3	Familiar	2	4	1
F2848-4	II-5	Familiar	2	2	2
F2848-5	II-4	Familiar	2	NA	2
F2848-6	II-3	Familiar	2	NA	2
F7614-3	II-8	Familiar	1	2	1
F7614-6	II-9	Familiar	1	3	1
F7614-7	II-10	Familiar	1	2	2
F8418-2	II-8	Familiar	1	4	1
F8418-5	III-8	Familiar	1	2	1
F8418-6	III-9	Familiar	1	2	2

Anexo II - *Primers* desenhados utilizando a ferramenta online *Methprimer* ou o software *Methyl Primer Express*.

Gene	Sequência F	Sequência R
ARHGAP29	ATATTAAGAGTTAAATGTGTAAGGT	TATCCTCRAACCAAACTAAA
BRCA1	TTTTAGTTTAGTGTGTTATTTTT	CCAACTACTTCCTTACCACTTC
CDK1	GTAGTTGGGTTTGATTGGTTGT	ACCTTCTAAAAAACCC
IGF2R	AGGAGTTTTGGGGTTTTAAGT	ACTAAAATCCTACCTCACTTTC
KIF20B	GATTGGGAAGTGGTTAGT	TTCTATCCTTAAACAATAAAAATAT
KIFAP3	TTTTATAGTTTTGTGTTTAAATATTG	CTTCTATTCCCATCCTCAAAAA
NOL8	TTAGTTTTGTTTTTAGTGAGGTA	AAAACCTAAATCCTCCACTTCC
NUMA1	TATTGGGAAAGAGAAGGGTATTAT	CCTCAAATTACTCAAACAACCTCC
PAX8 CpG131	GATTTGGTTTAGGTTTTGAGG	AAAAATCAATATTTCAAATAAATCTCC
PAX8 CpG25	TGGTTTTATTTGTATGGGATTTTT	AATCAAAAAACAACATCTTCTCTCT
PAX8 CpG27	TTTGATAAATATTTGTAGGAAATG	TTTACTCATTATCTTAAAAACTCC
PPM1F	ATGYGAGGTATTTGATAGA	AAAACAACATCCACAATAACA
PRICKLE1 CpG107	GAATTATTTAGAGTTTAGAAAGTTTT	CCTACAACCAACTACCAAC
PRICKLE1 CpG165	GGTTGTTATTTAGTTTTGGGATTTAGATA	AACACCAACCCCTCTATTCTCT
PRICKLE1 CpG23	AGGGTGTTTATTTGAGTTTTTTTT	ATCCCCCACACTAAATTTTCATAA
RECQL4 CpG164	GGATAAGTAGTTAATGGGAGG	AACTCCAAATACAACCACTACCT
RECQL4 CpG47	AAGTTGGGGTAGTTTTAGATGTAG	ACCTTAAACACACATAATCCCATC
ROR2 CpG21	GGTTGAGAGGTATAGTAGGTAGT	AACACTCTTCCAAAAACTTATC
ROR2 CpG210	TTTAGGTAGTTTTAAGTTGATTTTT	CATCCCTTTTATAATCCCTAAAC
SHH CpG32	GTTTATGGAATTGATGATTTT	TACCATTCCAACCCCTATCTAATA
SHH CpG542	GTTGTTTATTTGGGAGGAGGTTT	ACACTATTAATAATCAAACACACTTATA
SMEK1	TTTTTTTTATTTGGTTTTGTTAT	CTACAACACTCCCTAAATAACCCC
STAT5A	TAGGTAGTTGATTTTTGAGGAGGAA	ACAATAAAATATCCCCTCTC
THAP11	GTTGGGAAAGAATTTGTTTTT	CCTTAAATAATATCCCACTACC
TP63 Promotor 1	AATTTTATTGGTTAAATTAAGAAA	AAAACAATACTATAAAATAACACACC
TP63 Promotor 2	TTGTGATTTAGTATGATAAAATAATATTAG	TCCACATTTACAAAATAAAAAATC
TP63 Promotor 3	GTTTATAGTTGGGTATATATTAGGAAATTT	TATATACAATTTACAAAACATTACATTA
VCL	TTTTTTTTAGGAATGGGGGAG	CCAAATAAAAAATCTACTATACCACC
ZEB1	AATTTTATTGAAGTTATTTTTATTT	AATTCTCCCTATACCCTATACCCTC

Anexo III - Sequências genômicas de referência e localizações cromossômicas (hg19) para cada gene ou região de ilhas CpG utilizadas para alinhamento no *software* *Biq Analyzer HT*. chr: cromossomo.

Gene (Localização cromossômica)	Sequência referência
<i>ARHGAP29</i> (chr1:94,702,611-94,702,816)	ACATTAAGAGTAAATGTGCAAAGGCTTCAGAACAGCAGGCGGACCCGTTG TACGCGCTGTGTGCTAACTTAATTACGATCGTCACCGTCTGCTCCCGGCCACCG CTGGCGTCTCCAGCTCGTCCGCGCTCCCTCGCCGGGCGCGCCCTAGCCCC GGCCGGACCCCTCCCGTGGACTCCAGTCTGGCTCGAGGACA
<i>BRCA1</i> (chr17:41,278,215-41,278,485)	CCTCAGCCCCAGTGTGTTATTTTTCGGGTTCAGCTTGCTTTTGCCCCGTCTCCGT CGACGCAATCGCCACAGTCAATGGGGTGGTCTTTTGAGGGACAAGTGGTAA GAGCCAATCTTCTGGCGAAAACGCGGAGAAACGGGACTAGTTACTGTCTTTGT CCGCCATGTTAGATTACCCACAGAGATAGCGGCAGAGCTGGCAGCGGACGGT CTTTGCATTGCCGCTCCCGAGGGGGCGGGAAGCTGGTAAGGAAGCAGCCTGG
<i>CDK1</i> (chr10:62,538,144-62,538,416)	GTAGCTGGGCTCTGATTGGCTGCTTTGAAAGTCTACGGGCTACCCGATTGGTGA ATCCGGGGCCCTTTAGCGCGGTGAGTTTAAAAGTCTCGCAGTCTGGCTCAAAGC TGGCTCTTGGAAATTGAGCGGAGAGCGACGCGGTTGTTGTAGCTGCCGCTCGCG CCGCCGCGGAATAATAAGCCGGGTACAGTGGCTGGGGTCAAGGTCGTGTCTAG GGGACGCGCCGAGGGCTCGGAGGGCGAGTATTGAGGAACGGGGTCTCTAAG AAGGC
<i>IGF2R</i> (chr6:160,390,506-160,390,792)	AGGAGCTCCTGGGGTCTCAAGTCGCTCCGTTCCGCGCCGGGGTCCGCCCGT TCCCGAGGGAAAGTTGCTGCGTGGTGTCCGCGAGTCCCGCCGGTCCCGTGGC CAGGCGGGGGCGCGGGGGCACACGCGTCTGCGGCTCTGGAGCCCGCG TCCGGCCACTTGTGGCTGTGCTGAGTCCGGGGTGGCTTTGTCCGGAGCGGTG CCCGGCCCTTGGGGTCCGCGCCTGCCCTCGCGCGCCGGGGCTCCGGGAAAG TTGGAGGCAGGAAGTCCAGT
<i>KIF20B</i> (chr10:91,460,960-91,461,080)	GACTGGGAAGTGGCTAGCCCGCCATTCTCGGCCCGTCTTAGGGGAGTCCGAG GACTAAGTGCAGTGGCAGTGGCCTCACCTCGGGCCTTCGCCACATCTTTATTGCC TCAAGGACAGAA
<i>KIFAP3</i> (chr1:170,043,778-170,044,022)	CTCCACAGCTTCTGTGCCCAAACACTGGAGCGGCCAGACCCGCCAGAGTC GCCTAAGCCGGGCCGTACGACGCATGCGCTTGTCTGACCTCTGTATGCGCA GGCTCAGTCTGAGGCGGTGGCCTCAGAGTTTACGACCCGGGGAGCCGAGCAG GGAAGTTTTGGTGCAGTTTTGAGCGTCCGTGTATAAGGAGTGGCTCCGTAGTCA CCGCCCTTGGAGTGGGAATAGAAAG
<i>NOL8</i> (chr9:95,087,925-95,088,098)	CCAGCCCTGCCTTTAGTGTAGGCACGCGCCGGGGCGGCTCCAAGGCTGCCGGC TAGGGTGTAGCAGCGAGCGCTTGGGCTTGGAGTAGCAAAGAAGCGCTCTCCA GAGAAAGTATAGCCACTGCTTAGACAGCCAGGGAAACGTGTGCGGGGAAGTGG AGGACTCAGGCTCT
<i>NUMA1</i> (chr11:71,791,684-71,791,894)	CACTGGGAAAGAGAAGGGGTATCACCGCTTCCGGACCCCGGCTGTACTTGAAG GCAAAGAGAAGTACAAATCCCAGCGTACCCGCGGCCCTGAAGCCCCGCCCTG ACAAACTGAAGTCCCGGTAAGCATCGCGTCACTTATGGCGCTGCCGGT TGTGGTGACGAAAGCAGTTGCCATGGAGTTGCTCTGAGTAACCTGAGG
<i>PAX8 CpG25</i> (chr2:114,048,778-114,049,101)	TGGCTTCATTTGCATGGGATCCTTCTTTTATGCAGCCTCAGTATCAGAACGAGC CAGGAGGTTAACGAGCGTCGTCGTGCAGCGTCCCGGGACACCAGGGCCAGC GCGCGCTCTGCTTCCCGCGCTGCCCTCTGGGACTCCGACAGCGCGGCTGTCA CGGGGCTGGAGCTGTCAGTTTCCAGCCCTCCGCCCGCAGGGCTGGGCTCC GGCGCAGGAGAGGCGCAGCAGTTCCCATCCTCCCTGGCAAAGCGGTTCA CAGCCCGCTCTCATGGGCTGAACGAAGGAGGAAGACAGCTGTTTTCTGATC
<i>PAX8 CpG27</i> (chr2:114,082,644-114,082,800)	TCTGACAACTATTTGTAGGAAAATGAGATTTGGTCCGCGCTGGATGGCTCTCT GGGAGGCAGGCGTGGGGGTGGGGTGTGCGAGGCCCGTGCGCCCCCAACCC CCACCAGAGCCACGGCGCTCACGGAGCTTTAAGATAAATGAGCAA

<p><i>PAX8 CpG131</i> (chr2:114,034,747-114,035,049)</p>	<p>GATTCTGGCCAGGTCTCTGAGGGGAATGGGGGCCGGGAGGGAGCATTCTGCT GCAGCTCTAGAGAGAACCCGAGTGAGAAATCTTCCGGCGCCCCAGGGGTTCCC GCTCCCCTCTAGTTGGGCCATGGCCCTGGGACTGGAGACTGCGCCGAGAGCC CGGGCCCCCGCAGTGCCTTGGCTAAATAGAGATCTCAGCCTCGTGCAGCC GATCCGACCCTCTTATTATAACAAATGTCCCGTTTGAGAGCGGGACATCATAT TCGCGGAGAATTCTTCTGAAATATTGATTCC</p>
<p><i>PPM1F</i> (chr22:22,292,439-22,292,754)</p>	<p>ATGCGAGGCACTCTGACAGACGGGGAAGCAAGGGCAGGCGAGACTTCTTCTA GCTGCGCCCTCCCTCTTCCAGGCCTGCAGGAGCCCGAGCGTAGGACTGGGC TGGTGCCGGCGCAGCAACGCTCTTCTCACGGCAGGTGGCGCTGTCTCCCGC TGGACGGACAGAGACCCATCAGGCGGGGAGGGCCTGGGCGGTGGCGGGCC CCTTGGTGGGCGCCAGAGCGGAGGAGCAGCAGCTGCCACCGCCGGAAGCCT GCTGGCTCGGCCGAGAACAATGGCCGCTGTCACTGTGGACAGCTGTCTC</p>
<p><i>PRICKLE1 CpG23</i> (chr12:42,984,012-42,984,282)</p>	<p>AGGGGTGCCTATTTGAGTTCCTTCTCCTTCGGGGGTGTGTGAAGGGGCGAGT AATTAACACTTAGATTCTTCCGGCGCGCTTGCCGAGCTGGGGACCCGTTTGG ACTCGATCCTGTCTCGTGTAAATCCACAGGCTGGTCTGAGTCTTGCCCGGAAC CGGAGGAGTGATCTTTCAGAGCAAGAAGCCAAAGTTACGCTCTAGTCCGTTCC CTTCCCGCCTCGACGGCCTGGTCTCTCCATGAAACTCAGTGTGGGGGAT</p>
<p><i>PRICKLE1 CpG107</i> (chr12:42,983,396-42,983,540)</p>	<p>GAACCATCCAGAGCCAGAAAGTCTCCGTCCGACCGCGGCGCTCTCGGGGA AGTCAGCGCAGGCTGCGGAGGCTGGCATGTTCCGGGCGCGCTGTCGGGCGGC GGCGGCCGCGGAGACGGCGCTGGCAGCTGGGCTGCAGG</p>
<p><i>PRICKLE1 CpG165</i> (chr12:42,877,138-42,877,337)</p>	<p>GGTGCCATCCAGCCTGGGACCCAGACAGCGGCTCCACCTGCGGCCCGGACA GCGGCCGAGCGCGCCTGCAGCCCTCGCGGCGCCGGCCCTGAGCCCTCGGC GCGCACTAGCCGCCCGCCGCTCTGCGTCTGAAGGAGGCTGGCGGCCCGGAGC GGGAGCGGAGGACTCCCAAGGAACAGAGGGCTGGGTGCC</p>
<p><i>RECQL4 CpG47</i> (chr8:145,738,680-145,738,918)</p>	<p>AAGCTTGGGGGAGCCAGATGCAGCACAGCCCGCACATCTGGCCGGTCCAGC CCCATCCCAAAGGCCACCGTGGCCACCACCACCCGCAACTGGCCCTGCATGAAG GCTCGCTGTACCGCCCGCTTCCCGGCTGCACATGCCGCGTGGTAGCCTCG GCTGTGGTTTTGGGGGACGACCTTTGGGGAAGACAGGAGATGGTCACTGGG ATGGGACCATGTGTGCCAAGGT</p>
<p><i>RECQL4 CpG164</i> (chr8:145,743,313-145,743,557)</p>	<p>GGACAAGCAGCAATGGGAGGCGTCAACGTCTATCGCGGGGCGCCGCGGCGCCG GGCGGGGAGCGGCGGACGGTCTAGGCGGGGCTGGAGGCGGTGGCTGCGGTTG CGGGACCGGTGAGCGCGGGCGGCGGGGGGGTGTCTGAGCGGGGCGGCGCAC GCCCTGGCCCTGCCTGGCCTCAAGGGGCGGGGACCGGTGCTCGGGGCGGC CTGGCCGAGGGAGGCAAGTGGCTGCATTTGGGAGCC</p>
<p><i>ROR2 CpG21</i> (chr9:94,662,646-94,662,885)</p>	<p>GGCTGAGAGGCACAGCAGGACCGACCTGGCGCAGGCCCCCCACGCAGGG AACGGTGGCGGCGCGGGGCAATCCTCCGGCTGCCTGGAGCAAAGAGGCTGGC AGGGACACGAGGGCGCGGGCCTGCACGTGGGACCCGAGGAAACAAGAAAT GAGGAAGGCACCGCTGGCATCCAGAACATTCAAAGCAGCAAGGAGGCGGG AAAACGACAAGCTTCTGGAAGAGAGTGT</p>
<p><i>ROR2 CpG210</i> (chr9:94,711,230-94,711,550)</p>	<p>CCTAGGCAGGTCCCAAGCTGCACTTCCGAACCTCCCGCCGTTCTCTGCGCCC GGATCCAAGCAGCCGAAACCGCGCAGGGACTCGGGGGAAGTGGGGCCCCGGG AGGAAACCACTAGGAGCCGGGCCCCGCCCTCCCGCCGAGCCTCTACAG GTCCCGGGGTCCCCCGCGCGGCCAGAAGGCGGCCGCCCCAGCCGCTCA GGGACCCGCGATCCCAAGTCAAAACCAATTCATCCAGATCACACCACTTCGG CCACCCAGAGCGGTGGGGGAGAGGAGTTTACGGGACCAACAAAAGGGATG</p>
<p><i>SHH CpG32</i> (chr7:155,604,820-155,605,051)</p>	<p>GCCCATGGAAGTATGACTTCCGAGCTGTCCCGTGGGGTCCGTGCGCGAGTG CGCGGGCGGGTGTGTGCGTGTGCGCTCTCTTGGCTTTCCCTTCTCGCTCC GGCTCGCCGCTCGCTCTCCCTCGCTGGCTGCCTCGCTTTTCTTCTATATA ACCTTGCCCGCCGCGGTGCGGGGACTCTCACCGCTCCCAACCCAGACAGGG GCTGGAATGGCA</p>

<p><i>SHH CpG542</i> (chr7:155,597,818-155,598,092)</p>	<p>GCTGCTCACTTGGGAGGAGGCCTCCGGAGACCCCCGCTAGGGGGACCCCGCGG GCAGGGAGCCGGGGAGGAGGTTTTGCGCAGGGGGCGCCCTGGGCCAGGGGC CGCCGCCGAGCCCGGAGGTGGGGGGCGCCCTCTGCCTGCCTGCAGCCGGG AGCTCTCCGGGCCGACACCCGCCCTGCCGCCCGCTCCACGCCTTTGCTTTCT GTGGGGACGAGAAGCCACTCTAGCCTTTACCTACAAGTGTGCCTTGACACCTTAA CAGTGT</p>
<p><i>SMEK1</i> (chr14:91,976,407-91,976,681)</p>	<p>CCTTCCTTATACCTGGCCCTGCCACCGCGGCCAGTGGCTCCCTCCGCTCCCCCTG TTTTAAACGTCAGAAGCCGACAGGCGCCTGATCCTGCTGCAACTGCTACCACTGC CCGGGAGACTCTGAGGGCCCCGGCCGCCAAGCGACCCGGACTAGTGGAAGT CGGTTTCCCTCCGCCCGCTCTCTCCGTCTTCTCACACAGCCAAAACCGCCGCC ATCTTGTTTCGCCCTCAAAGCGGGCGCGGGGCCACCCAGGGAGCAGCTGCA G</p>
<p><i>STAT5A</i> (chr17:40,440,466-40,440,672)</p>	<p>CAGGCAGCTGACCTTTGAGGAGGAAATCGCTGCTCTCCGCTCCTTCTGTAGTAA CAGCCGCCGCTGCCGCCGCCAGGAACCCCGCCGGGAGCGAGAGCCGCGG GGCGCAGAGCCGCCCGGCTGCCGGACGTTGCGGCCACCAGGTGGGTGACC CGGTGGCGCTCCTCGCGCGCGCCGAGAGGGGACACTTACTGC</p>
<p><i>THAP11</i> (chr16:67,876,071-67,876,354)</p>	<p>GCTGGGGAAAGAATCTGCTCTTCGGAGGCTATCGCAGTGCCTTGAAGGCGGC AGGAATCCGCACACACCCGCCCGGAAGTGAGCCGCTTCTGGCGCGGGCATTG TGGGGGGCGTAGTCTTTCTCAGGCGGTCTTCCGGCGTCCCGGGGGCCAC TCCCGAGCGCAGGCGGGCAGCCAGGCGGGCGCGCGCGGGCCGGCAGGA AGCGTATTCTGGGCACGGGGCGCCGGGCGGGCCGGCTGCGCCGAGCGGCAGT GGTGGGATACCACCAAGG</p>
<p><i>TP63 Promotor 1</i> (chr3:189,349,104-189,349,349)</p>	<p>AATCTCATTGGCTAAAATCAAGAAACGCTCCGCCTTTTGCAAATATGTATGAAG GAGAGAAGTGCCTAAACTTCTATGTCTGATAGCATTGACCCTATTGCTTTTAGCC TCCCGGCTTATATCTATATATACACAGGTATATGTGTATATTTATATAATTGTT TCCGTTGTTGATATCAAAGACAGTTGAAGGAAATGAATTTGAAACTTACGGT GTGCCACCCTACGTAAGTCCCT</p>
<p><i>TP63 Promotor 2</i> (chr3:189,306,517-189,306,702)</p>	<p>CTGTGACTCAGCATGATAAAACAACATCAGCCGGCACTATGCATTTTCAGTTCTC CCGAATCTTGAATTGAGAGAGAAACAGCTAATTCTCTCCATAGGTTTTCGGTTA TGAATTCTGTATCCAAGAAAGCTAGATCTGGAAAGATCCTTAGAGATCATCGATC CTCTATTTTGTAAATGTGGA</p>
<p><i>TP63 Promotor 3</i> (chr3:189,507,407-189,507,751)</p>	<p>GCCTATAGTTGGGTATATATTAGGAAACCTTAAATTATGTACAGAGAGAGAAAG AGAGAGAGGGACTTGAGTTCTGTTATCTTCTTAAAGTAGATTATATTGTAAGGGT CTCGGGTGGGGGGTGGCAAATCCTGGAGCCAGAAGAAAGGACAGCAGC ATTGATCAATCTTACAGCTAACATGTTGTACTGGAAACAATGCCAGACTCAA TTTAGTGAGGTAAGGATTTTAGATTTTAGCACTCCATTTAGAGATGCTTTTTAATT TTTATTTTGTAAAAAACTTACGTAATTTGCGGTTCTCGTCCACCAATGTAATGT TTTGAAATTGTATATA</p>
<p><i>VCL</i> (chr10:75,757,694-75,758,035)</p>	<p>CCTCCCTTAGGAATGGGGGAGGGGCGGGTTTTGGCGCGCGGTGGCTGGGGC GGAGCATCTCGAAAAGGGACCAGTAGGAGTGGCGGCCAGGTCTGAGCTGCGC TGCCATTGGGCGCGGCAGCAGAGGGGGCGGGGAGCCCGTGGAGCTGGTTAC GCCGAGGGAAGCCCGACTCCGTAAGTCTGCTGCACAGTCTGCTCTTCCGCCGTT CCGGCCCCGTGGATCCTACTTCTGTGCGCCCGCGGTTCCGCCCCCGCTCGCCG CCGCGATGCCAGTGTTCATACGCGCAGATCGAGAGCATCCTGGAGCCGGTGG CACAGCAGATCTCCACCTGG</p>
<p><i>ZEB1</i> (chr10:31,607,415-31,607,742)</p>	<p>AATCTCATTGAAGTCACTTCCCATCCCGTTCTGTTGGGGGAAACCAGGCGTCC CTGGAAGGGAAGGGAAGGGAGTCCGGGCTGCGCGGGTCAAGTAGCCTCTCTCC GGTCGCCGCTGTCTCGCCGTCCCAAACTGCCCTTCCCCTCATCAAGGGAAC TCCCCGGGAGTCCAGACCGCGATCCCTTCTTGGCCCCGGGGTGGCGGGGGC GGACACGCGAGGCGTGGGACTGATGGTAGCCCTGCTCCAGGAAGCAGGACAGG CGGGGACCTCTGGGCGCTCGGAGGGGACGCTCCGAGGGCACAGGGTACAGGG AGAATC</p>

Anexo IV - Valores de metilação das regiões CpGs selecionadas para cada região CpG/gene estudado para cada indivíduo. NA indica valores que foram removidos após filtros.

Amostra	ARHGAP29	BRCA1	CDK1	IGF2R	KIF20B	KIFAP3	NOL8	NUMA1	PAX8 CpG25	PAX8 CpG27	PAX8 CpG131	PPM1F	PRICKLE1 CpG23	PRICKLE1 CpG107	PRICKLE1 CpG165
c.26039	0,034	NA	0,03	0.001	0,002	0,004	0,001	0,002	0.191	0,006	0,019	0.025	NA	0,003	0,003
c.26053	0,038	NA	0,02	0	0,001	0,005	0,003	0,003	NA	0,007	NA	NA	NA	0,004	0,003
c.26054	0,026	NA	0,022	0.002	0,002	0,001	0,005	0,005	NA	0,006	NA	0.037	NA	0,004	0,004
c.26078	0,037	NA	0,052	0.003	0,002	0,005	0,005	0,006	NA	0,01	NA	0.021	NA	0,005	0,005
c.26101	NA	NA	NA	0.003	0	0,004	0,002	0,002	NA	0,006	NA	0.017	NA	0,002	0,003
c.26177	0,034	NA	NA	0	0,001	0,002	0,005	0,002	NA	0,006	NA	0.027	NA	0,003	0,004
c.26219	0,039	NA	NA	NA	0,001	0,007	0,005	0,003	NA	0,005	0,041	NA	NA	0,004	0,003
c.26232	0,036	NA	0,009	NA	0,001	0,002	0,003	0,003	NA	0,005	0,038	NA	NA	0,004	0,003
c.26390	0,061	NA	0,032	0.001	0,003	0,005	0,005	0,004	NA	0,017	NA	NA	NA	0,006	0,009
c.27368	0,036	NA	0,049	0.003	0,002	0,005	0,009	0,005	0.168	0,007	0,032	0.035	NA	0,005	0,004
c.27705	0,031	NA	0,018	NA	0,001	0,009	NA	0,005	NA	0,006	NA	NA	NA	0,004	0,004
c.27769	NA	NA	0,036	0.013	0,002	0,006	0,004	0,004	NA	0,009	NA	NA	NA	0,004	0,005
c.27848	0,047	NA	0,026	0.002	0,002	0,008	0,004	0,004	0.166	0,007	NA	0.032	NA	0,004	0,006
c.27851	0,032	NA	0,021	NA	0,001	0,009	0,002	0,003	NA	0,006	0,045	0.015	NA	0,003	0,004
c.27931	0,051	0,358	0,036	0.009	0,002	0,004	0,003	0,004	0.154	0,01	0,033	NA	0	0,006	0,009
c.28039	0,039	NA	0,013	NA	0,002	0,004	0,009	0,003	NA	0,006	0,026	0.033	0,005	0,003	0,003
c107	NA	NA	NA	NA	NA	NA	NA	NA	NA	0,007	NA	NA	NA	NA	0,003
c111	NA	NA	0,067	NA	0,003	0,005	0,01	0,005	NA	0,009	0,052	NA	NA	0,004	0,008
c112	0,077	NA	0,047	0.005	0,003	0,005	0,006	0,007	0.224	0,01	0,063	0.009	NA	0,004	0,005
c130	0,097	NA	0,028	0.002	0,002	0,004	0,005	0,008	0.109	0,008	0,031	0.021	NA	0,004	0,005
c132	NA	NA	NA	NA	0,001	0,014	0	0,004	NA	0,008	NA	NA	NA	0,005	0,004
c137	0,005	NA	0,021	0.004	0,001	0,003	0,003	0,004	0.166	0,008	0,05	0.021	NA	0,002	0,004
c142	NA	NA	0,029	0.003	0,002	0,003	0,006	0,005	0.253	0,008	NA	0.018	NA	0,004	0,004
c156	0,041	NA	NA	NA	0	0	0,017	0,006	NA	0,014	NA	NA	NA	0,008	0,005
c18	0,028	NA	NA	0	0,001	0,007	0,004	0,003	0.117	0,005	NA	0.017	NA	0,004	0,003

c3	0,03	NA	0,037	0,005	0,003	0,004	0,006	0,004	NA	0,005	NA	0,02	NA	0,002	0,003
c38	0,028	NA	0,03	0,001	0,002	0,004	0,004	0,005	NA	0,008	NA	0,027	NA	0,003	0,002
c39	0,028	NA	0,011	0,001	0,002	0,002	0,005	0,002	0,086	0,006	NA	0,016	NA	0,002	0,003
c40	0,034	NA	0,062	0,001	0,005	0,003	0,003	0,005	NA	0,006	NA	0,013	NA	0,002	0,004
c41	0,026	NA	NA	0,005	0,003	0,003	0,003	0,004	NA	0,006	NA	NA	NA	0,004	0,002
c43	0,017	NA	0,038	0,002	0,002	0,005	0,005	0,006	NA	0,011	NA	0,02	NA	0,004	0,003
c44	0,028	0,886	0,019	0,001	0,002	0,006	0,006	0,004	0,227	0,009	NA	0,008	NA	0,005	0,005
c46	0,021	NA	NA	0,004	0,001	NA	0	0,004	NA	0,007	NA	0,017	NA	0	NA
c48	0,027	NA	0,011	0,005	0,002	0,004	0,004	0,004	0,132	0,007	NA	0,013	NA	0,003	0,004
c50	0,034	NA	0,012	0,004	0,002	0,002	0,003	0,003	NA	0,008	NA	0,032	NA	0,003	0,004
F10318-1	0,019	NA	NA	NA	0,002	0	0	0,008	NA	0,004	NA	0,031	NA	0	0
F1843-1	0,028	NA	0,021	NA	0,002	NA	0,005	0,002	0,214	0,01	0,039	0,02	NA	0,003	0,005
F1843-10	0,031	NA	NA	NA	0,002	0,009	0,002	0,001	0,098	0,008	NA	0,026	NA	0,004	0,006
F1843-2	0,028	0,523	0,027	NA	0,002	0,003	0,002	0,003	NA	0,01	NA	0,013	0,013	0,004	0,005
F1843-4	0,031	0,504	0,036	NA	0,001	0,008	0,005	0,004	0,128	0,005	0,038	NA	0,03	0,003	0,004
F1843-5	0,036	NA	0,024	NA	0,002	0,004	0,003	0,005	0,21	0,005	0,032	0,027	NA	0,003	0,004
F1843-6	0,029	0,701	0,04	0	0,003	0,005	0,004	0,003	0,147	0,005	0,031	0,016	0,042	0,003	0,004
F1843-7	0,027	NA	0,056	0,004	0,004	0,004	0,005	0,005	NA	0,005	0,049	0,027	NA	0,004	0,005
F1843-8	0,034	0,486	0,026	0	0,002	0,009	0,005	0,006	0,176	0,006	0,036	NA	0,007	0,004	0,004
F1843-9	0,038	NA	NA	NA	0,001	0,006	0,001	0,003	0,119	0,009	NA	NA	NA	0,006	0,005
F185-1	0,038	NA	0,012	0,003	0,002	0,009	0,002	0,002	0,136	0,006	NA	0,023	NA	0,004	0,008
F2570-1	0,031	NA	0,017	NA	0,001	0	0,004	0,001	NA	0,006	NA	0,009	NA	0,002	0,002
F2570-2	0,031	NA	0,034	NA	0,002	0,006	0,004	0,005	0,197	0,008	NA	0,013	NA	0,004	0,005
F2570-3	0,026	NA	0,038	NA	0,002	0,003	0,002	0,004	0,156	0,004	NA	0,018	NA	0,002	0,003
F2570-4	0,027	NA	0,022	NA	0,002	0,004	0,002	0,003	NA	0,005	0,036	0,026	NA	0,003	0,003
F2570-5	0,023	NA	0,011	0,001	0,001	0,005	0,004	0,002	NA	0,005	0,045	0,023	NA	0,002	0,003
F2570-6	0,029	0,605	0,051	0,004	0,002	0,01	0,004	0,004	0,156	0,006	0,032	NA	NA	0,002	0,005
F2570-7	0,039	NA	0,019	NA	0,002	0,002	0,011	0,002	NA	0,009	NA	NA	NA	0,004	0,005

F2848-1	0,023	NA	NA	NA	0,001	0,003	0,004	0,002	0.121	0,003	NA	NA	NA	0,001	0,001
F2848-2	0,029	NA	0,028	NA	0,006	0,001	0,003	0,003	0.283	0,008	NA	0.02	NA	0,004	0,004
F2848-3	0,016	NA	NA	NA	0,003	0,002	0,005	0,007	NA	0,003	NA	NA	NA	0,003	0,002
F2848-4	0,009	NA	NA	0.012	0	0,004	0,004	0,001	NA	0,004	NA	NA	NA	0,003	0,002
F2848-5	0,023	NA	0,033	0	0,001	0	0,004	0,003	NA	0,004	NA	0.019	NA	0,001	0,002
F2848-6	0,021	0,748	0,041	0.002	0,001	0,012	0,003	0,004	0.066	0,004	NA	0.027	NA	0,004	0,002
F3196-1	0,029	NA	0,025	NA	0,003	0,006	0,001	0,002	0.09	0,004	NA	0.019	NA	0,003	0,003
F3196-2	0,033	NA	0,051	NA	0,001	0,003	0,004	0,003	NA	0,006	NA	NA	NA	0,003	0,003
F3196-3	0,027	NA	NA	NA	0,001	0,004	0,005	0,002	NA	0,004	NA	NA	NA	0,003	0,002
F6543-1	0,036	NA	0,022	0.006	0,001	0,009	0,003	0,004	NA	0,005	NA	0.032	NA	0,003	0,003
F7391-1	0,028	0,575	0,052	0.008	0,001	0,008	0,004	0,003	NA	0,008	NA	NA	NA	0,003	0,005
F7614-1	0,03	NA	0,01	NA	0	0,004	0,007	0,002	NA	0,005	NA	0.015	NA	0,003	0,004
F7614-2	0,032	NA	0,018	0.006	0,001	0	0,004	0,006	NA	0,007	NA	0.026	NA	0,004	0,004
F7614-3	0,027	NA	0,08	NA	0	0,007	0,006	0,006	NA	0,005	0,053	0.024	NA	0,002	0,003
F7614-4	0,033	NA	0,002	0	0,002	0,006	0,004	0,007	0.107	0,006	NA	0.021	NA	0,004	0,005
F7614-5	0,045	NA	NA	NA	0	NA	0,008	0	NA	0,003	NA	0.022	NA	NA	NA
F7614-6	0,027	NA	NA	NA	0,001	NA	0	0,005	0.124	0,005	NA	0.018	NA	0,001	0,004
F7614-7	0,031	NA	0,023	NA	0,001	0,006	0,008	0,004	0.122	0,006	NA	0.016	NA	0,002	0,004
F7867-1	0,035	NA	0,017	0.003	0,001	0,002	0,004	0,005	NA	0,006	NA	0.029	NA	0,002	0,004
F8104-1	0,031	NA	0,033	0.006	0,005	0,008	0,003	0,004	0.091	0,005	0,092	0.025	NA	0,004	0,003
F8418-1	0,037	NA	NA	NA	0,001	0	0,007	0,001	0.115	0,006	NA	NA	NA	0,002	0,004
F8418-2	0,028	NA	NA	NA	0,001	0,005	0,007	0,004	NA	0,006	NA	0.031	NA	0,002	0,003
F8418-3	0,029	NA	0,031	0.003	0,002	0,006	0,003	0,005	NA	0,007	NA	0.024	NA	0,002	0,003
F8418-4	0,03	NA	0,028	0	0,004	0,008	0,003	0,012	NA	0,007	NA	0.021	NA	0,005	0,004
F8418-5	0,029	NA	NA	0.001	0,001	0,005	0,003	0,004	0.221	0,008	NA	0.023	NA	0,005	0,006
F8418-6	0,034	NA	NA	0.001	0,002	0	NA	0,005	NA	0,008	NA	NA	NA	0,006	0,004
F8418-7	0,029	NA	NA	NA	0,002	NA	0	0,003	0.149	0,005	NA	0.019	NA	0,001	0,002
F8422-1	0,039	NA	0,013	0.002	0	0,006	0,001	0,006	0.098	0,003	NA	0.011	NA	0,004	0,003

F9113-1	0,025	NA	NA	0.004	0	0,004	0,006	0,004	0.056	0,003	NA	NA	NA	0,003	0,002
F9350-1	NA	NA	NA	0.004	0,005	0,006	0,008	0,007	NA	0,004	NA	NA	NA	0,001	0,004
F9351-1	0,041	NA	0,005	0	0,001	0,008	0,006	0,005	NA	0,002	0,05	0.02	NA	0,002	0,002
F9357-1	0,034	NA	0,02	NA	0,001	0,006	0,004	0,005	NA	0,004	NA	NA	NA	0,002	0
F9365-1	NA	NA	NA	NA	0,002	0,002	NA	0,007	NA	0,002	NA	NA	NA	0,003	0,002
F9399-1	0,022	NA	0,004	0.005	0,001	0,012	0,005	0,004	0.17	0,005	0,035	NA	NA	0,002	0,003
F9400-1	0,031	NA	0,026	0.002	0,001	0,006	0,004	0,003	NA	0,003	NA	NA	NA	0,002	0,001
F9405-1	NA	NA	NA	NA	0,001	0,006	0,004	0,006	NA	0,003	NA	NA	NA	0,002	0,002
F9412-1	NA	NA	0,025	0.003	0,001	0,002	0,006	0,005	NA	0,004	NA	0.012	NA	0,002	0,002
F9413-1	0,045	NA	NA	0.002	0,001	0	0,006	0,003	0.154	0,004	NA	0.009	NA	0,002	0,001
F9414-1	0,03	NA	0,026	0.002	0,001	0,009	0,005	0,005	NA	0,005	NA	0.016	NA	0,003	0,002
F9434-1	NA	NA	NA	NA	0,001	0,013	NA	0,008	NA	0,003	NA	NA	NA	0,002	0,003
F9467-1	NA	NA	0,033	0	0,001	0,007	0,006	0,003	NA	0,003	NA	0.019	NA	0,002	0,002

Amostra	RECQL4 CpG47	RECQL4 CpG164	ROR2 CpG21	ROR2 CpG210	SHH CpG32	SHH CpG542	SMEK1	STAT5A	THAP11	TP63_promotor1	TP63_promotor2	TP63_Promotor3	VCL	ZEB1
c.26039	0,909	0.002	NA	0,018	NA	NA	0,009	0,003	0.003	0,876	0,932	NA	NA	NA
c.26053	0,912	NA	NA	0,021	NA	NA	NA	0,001	0.001	NA	0,935	NA	NA	0,005
c.26054	0,916	0.001	0,463	0,005	NA	0,012	NA	0,002	0.002	NA	0,896	NA	NA	0,005
c.26078	0,912	0.004	0,547	NA	NA	NA	0	0,002	0.002	NA	0,918	NA	NA	0,004
c.26101	0,918	NA	0,529	0,006	NA	NA	NA	0,001	0	0,64	0,896	NA	NA	0,006
c.26177	0,918	0.003	0,576	NA	0.012	0,004	0	0,002	0.002	0,735	0,942	NA	NA	0,005
c.26219	0,915	NA	NA	NA	NA	NA	0,001	0,002	NA	0,776	0,941	NA	NA	NA
c.26232	0,893	NA	0,278	0,013	NA	NA	0,004	0,002	NA	0,798	0,91	NA	NA	NA
c.26390	0,923	0	0,605	0,01	0.161	0,004	0,006	0,002	0.002	NA	0,902	NA	NA	0,012
c.27368	0,918	0.009	0,478	0,004	NA	NA	0,002	0,002	0.002	0,809	0,925	0.889	NA	0,01
c.27705	0,898	NA	NA	0,008	NA	NA	0,004	NA	NA	0,799	0,935	0.935	NA	0,006
c.27769	0,927	NA	0,418	0,009	NA	0,01	0,015	0,003	NA	NA	0,857	NA	NA	0,003

c.27848	0,923	0.001	0,345	NA	NA	0,006	0	0,004	0.006	NA	0,924	NA	NA	0,005
c.27851	0,908	0.001	NA	0,007	NA	NA	0,009	0,005	NA	0,8	0,942	NA	NA	0,007
c.27931	0,924	0	0,672	0,013	0.107	0,005	0,007	0,003	NA	0,721	0,952	0.953	0.003	0,006
c.28039	0,921	NA	0,476	0,008	NA	NA	0,004	0,003	NA	NA	0,946	NA	NA	0,008
c107	NA													
c111	0,906	0.001	NA	0,02	NA	NA	0,003	0,005	0	0,671	0,94	0.865	NA	0,008
c112	0,919	0.001	NA	0,02	0.01	0,014	0,008	0,004	0.001	0,865	0,914	NA	NA	0,004
c130	0,907	0.004	0,602	0,001	NA	NA	0,005	0,002	0	0,775	0,907	NA	NA	0,004
c132	0,953	NA	NA	NA	NA	NA	0	0,004	NA	0,731	0,936	NA	NA	0,017
c137	0,903	0.001	0,357	0,022	0	NA	0,011	0	0.001	0,714	0,933	1	NA	0,006
c142	0,917	0.002	NA	0,016	NA	NA	0,005	0,002	0.001	NA	0,918	0.86	0.029	0,005
c156	0,89	0.002	NA	NA	NA	NA	0,013	0,003	0.001	0,802	0,92	NA	NA	NA
c18	0,925	0	NA	0,015	0.021	NA	0,005	0,001	0.002	NA	0,947	NA	NA	NA
c3	0,926	0	0,596	NA	NA	NA	NA	0,002	0.001	NA	0,951	NA	NA	0,015
c38	0,933	0	NA	NA	0.035	NA	NA	0,004	0.003	NA	0,956	NA	0.012	NA
c39	0,934	0.002	NA	0,012	NA	NA	NA	0,002	0	NA	0,93	NA	NA	0,003
c40	0,917	0	NA	0	NA	NA	NA	0,002	0.001	0,758	0,912	NA	0.004	0,006
c41	0,938	NA	0,793	NA	NA	NA	NA	0,002	0	NA	0,97	NA	NA	NA
c43	0,941	0.001	0,709	0,017	NA	NA	NA	0,004	0.001	NA	0,917	NA	NA	0,006
c44	0,937	0.001	0,437	NA	0.023	NA	NA	0,004	0.005	NA	0,941	NA	0.007	0,006
c46	NA	0	NA	NA	NA	NA	0,011	0	0.003	0,872	0,942	NA	NA	NA
c48	0,95	0	NA	NA	0.027	NA	NA	0,005	0.003	NA	0,952	NA	NA	NA
c50	0,913	NA	NA	NA	0.014	NA	NA	0,001	0.002	NA	0,93	NA	NA	0
F10318-1	0,882	NA	0.006	0,758	0,558	NA	NA	NA						
F1843-1	0,914	0.001	0,509	0,01	0.039	NA	0,001	NA	0.002	0,816	0,93	NA	0	0,006
F1843-10	0,93	0.001	NA	NA	NA	NA	0,006	NA	0	0,764	0,954	0.979	0.005	NA
F1843-2	0,907	0.001	0,354	0,024	NA	NA	0,013	0,003	0.002	0,796	0,918	NA	0	0,007
F1843-4	0,917	0.001	0,505	0,013	NA	NA	0,017	0,004	0.001	0,762	0,922	NA	0	0,007

F1843-5	0,922	0.002	0,456	0,008	NA	NA	0,014	0,006	0.003	0,796	0,91	NA	0.004	0,005
F1843-6	0,933	0.002	0,463	0,007	NA	0,002	0,005	0,004	0.001	0,877	0,941	0.833	0.004	0,004
F1843-7	0,917	0	0,341	NA	0.017	NA	0	0	0.005	0,802	0,922	NA	0.002	NA
F1843-8	0,914	0	0,595	0,014	NA	NA	0,007	0,008	0.004	0,836	0,937	NA	NA	0,007
F1843-9	0,884	0.003	0,297	0,011	NA	NA	NA	0,002	0.001	0,738	0,95	NA	NA	0,007
F185-1	0,92	0.002	NA	NA	NA	NA	NA	0	0.002	0,84	0,951	NA	NA	0,007
F2570-1	0,876	0	NA	0,01	NA	NA	NA	NA	0.002	0,845	0,925	1	NA	0,005
F2570-2	0,92	0.006	0,488	0,018	NA	NA	NA	NA	0.002	0,835	0,907	NA	0.008	0,009
F2570-3	0,906	0	0,591	0,014	0.004	NA	NA	0,002	0.001	0,795	0,909	0.909	NA	0,006
F2570-4	0,925	0.002	0,446	0,018	0.004	NA	0	0,003	0.003	0,837	0,932	NA	NA	0,003
F2570-5	0,906	0	0,46	0,008	NA	NA	0,019	0,002	0.002	0,798	0,925	NA	NA	0,007
F2570-6	0,913	0.004	0,376	0,029	NA	0,006	0	0,003	0	0,803	0,954	0.865	0.005	0,005
F2570-7	0,905	0.002	0,357	0,039	NA	NA	0,015	0,003	0.001	0,908	0,96	NA	0	0,004
F2848-1	NA	0.002	NA	NA	NA	NA	NA	NA	0.003	0,911	0,794	NA	NA	NA
F2848-2	0,915	0.007	NA	NA	NA	NA	NA	0,001	0	0,785	0,932	NA	0.006	0,005
F2848-3	0,929	NA	NA	NA	NA	NA	NA	0	NA	0,813	0,6	NA	NA	0,009
F2848-4	0,901	NA	NA	NA	NA	NA	0	0,007	0.003	0,788	NA	NA	NA	0,005
F2848-5	0,932	NA	NA	NA	NA	NA	NA	0,003	0	0,855	0,765	NA	NA	0,007
F2848-6	0,908	NA	0,526	0,007	NA	NA	0,011	0,001	0.003	0,785	0,913	NA	0.005	0,007
F3196-1	0,917	0.001	NA	NA	NA	NA	NA	NA	0	0,767	0,932	NA	0	NA
F3196-2	0,915	0.003	NA	0,006	NA	NA	NA	NA	0.002	0,865	0,94	NA	NA	NA
F3196-3	0,909	0	NA	0,008	0.003	NA	NA	NA	0.001	0,831	0,924	NA	NA	0,008
F6543-1	0,939	NA	0,556	0,006	NA	NA	NA	0,001	0.001	NA	0,966	NA	NA	0,006
F7391-1	0,919	0.003	0,617	0,013	NA	NA	0,002	0,002	0.002	NA	0,908	NA	NA	0,002
F7614-1	0,943	0.001	NA	0,005	NA	NA	0	0,004	0.002	0,905	0,948	NA	NA	0,005
F7614-2	0,933	0.001	0,692	0,004	NA	NA	NA	0,005	0.002	0,723	0,91	0.9	NA	0,004
F7614-3	0,926	0.003	0,697	0,006	NA	NA	0,013	0,003	0.003	0,792	0,926	NA	0.002	0,004
F7614-4	0,932	0.001	0,606	0,01	NA	NA	0,002	0,002	0.001	0,802	0,96	NA	0.002	0,007

F7614-5	NA	0	NA	NA	NA	NA	NA	NA	0	NA	NA	NA	NA	NA
F7614-6	NA	0.004	NA	NA	0.015	NA	NA	NA	0.002	0,828	0,922	NA	0	NA
F7614-7	0,917	0.001	NA	0,013	NA	NA	0,003	NA	0.001	0,754	0,927	NA	NA	0,005
F7867-1	0,919	0.003	0,58	0,019	NA	0,008	0,002	0,004	0.001	0,891	0,904	NA	0	0,007
F8104-1	0,92	0.001	0,318	0,011	NA	NA	0,009	0,001	0.004	NA	0,933	NA	NA	0,009
F8418-1	0,914	0	NA	0	NA	NA	NA	0,003	0.004	0,906	0,907	NA	NA	0,004
F8418-2	0,911	0.003	NA	0,013	NA	NA	NA	0,001	0.001	0,801	0,931	NA	NA	0,007
F8418-3	0,92	0.002	NA	0,021	NA	NA	0	0,004	0.001	0,747	0,922	NA	0	0,007
F8418-4	0,93	0	0,56	0	NA	NA	0	0,001	0.001	0,896	0,942	NA	NA	0,009
F8418-5	0,912	0.002	0,578	0,011	NA	NA	0,003	0,003	0	0,697	0,94	NA	0.015	0,011
F8418-6	0,907	0	NA	NA	NA	NA	0,008	0,003	0.005	0,779	0,945	NA	NA	0,008
F8418-7	0,922	0.002	NA	NA	NA	NA	NA	NA	0.002	0,738	0,878	NA	NA	NA
F8422-1	0,921	NA	0,377	0,003	NA	NA	0,008	0,008	0.002	NA	0,878	NA	0.012	0,007
F9113-1	0,918	NA	0.003	0,737	0,774	NA	NA	0,009						
F9350-1	0,914	NA	0,442	NA	NA	NA	0,009	0,002	0.003	0,856	0,426	NA	NA	0,014
F9351-1	0,884	0.001	NA	0,008	0.014	NA	0,005	0,001	0.001	0,725	0,816	NA	NA	0,01
F9357-1	0,894	NA	0,417	0,004	NA	NA	0,01	0,001	NA	0,767	0,734	NA	NA	0,008
F9365-1	0,894	NA	0,752	NA	NA	NA								
F9399-1	0,912	0	0,517	0,012	NA	NA	0,005	0,003	0	0,867	0,621	0.638	NA	0,006
F9400-1	0,908	NA	0,489	0,007	NA	0,021	0,01	0,003	0.006	0,941	0,291	NA	NA	0,005
F9405-1	0,919	NA	NA	NA	NA	NA	NA	0	0.002	NA	0,282	NA	NA	0,01
F9412-1	0,925	0.003	0,487	NA	NA	NA	NA	0,004	0.002	NA	0,728	NA	NA	NA
F9413-1	0,927	0.003	0,53	0,016	NA	NA	0,019	0,001	0	NA	0,559	0.73	NA	NA
F9414-1	0,913	0.002	0,516	0,008	NA	NA	0,011	0,004	0.004	0,652	0,769	NA	NA	0,007
F9434-1	0,907	NA	0,546	NA	NA	NA								
F9467-1	0,915	0.003	0,677	NA	NA	NA	NA	0,005	0.001	NA	0,099	NA	NA	0,009