

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO

PRISCILA DE SOUSA FERREIRA

**CARACTERIZAÇÃO *IN SILICO* DE REGIÕES
DIFERENCIALMENTE METILADAS EM GENES ASSOCIADOS
AO CÂNCER DE MAMA**

CAMPOS DOS GOYTACAZES – RJ

PRISCILA DE SOUSA FERREIRA

**CARACTERIZAÇÃO *IN SILICO* DE REGIÕES
DIFERENCIALMENTE METILADAS EM GENES ASSOCIADOS
AO CÂNCER DE MAMA**

Dissertação apresentada ao
Centro de Biociências e
Biotecnologia da Universidade
Estadual do Norte Fluminense Darcy
Ribeiro – como parte dos requisitos
para a obtenção do título de Mestre
em Biociências e Biotecnologia

Orientador: Prof. Dr. Milton Masahiko Kanashiro

Coorientador: Prof. Dr. Álvaro Fabrício Lopes Rios

UNIVERSIDADE ESTADUAL DO NORTE FLUMINENSE DARCY RIBEIRO

CAMPOS DOS GOYTACAZES – RJ

FEVEREIRO 2018

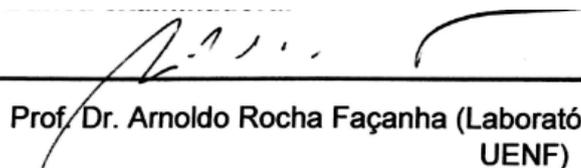
PRISCILA DE SOUSA FERREIRA

**CARACTERIZAÇÃO *IN SILICO* DE REGIÕES
DIFERENCIALMENTE METILADAS EM GENES ASSOCIADOS
AO CÂNCER DE MAMA**

Dissertação apresentada ao Centro de Biociências e Biotecnologia da Universidade Estadual do Norte Fluminense Darcy Ribeiro – como parte dos requisitos para a obtenção do título de Mestre em Biociências e Biotecnologia.

Aprovada em 28 de fevereiro de 2018

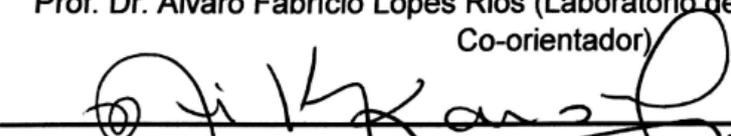
Banca examinadora:


Prof. Dr. Arnaldo Rocha Façanha (Laboratório de Biologia Celular e Tecidual – UENF)


Prof. Dr. Filipe Brum Machado (Departamento de Ciências Biológicas – UEMG)


Prof. Dr. Víctor Martin Quintana Flores (Laboratório de Biotecnologia – UENF)


Prof. Dr. Álvaro Fabrício Lopes Rios (Laboratório de Biotecnologia – UENF – Co-orientador)


Prof. Dr. Milton Masahiko Kanashiro (Laboratório de Biologia do Reconhecer – UENF – Orientador)

AGRADECIMENTOS

À **Deus**, por estar sempre presente, guiando o meu caminho e me dando força. Por me proporcionar a realização de mais um sonho e por me possibilitar crescer e aprender a cada dia.

À minha **família**, pelo apoio e incentivo. Em especial a minha avó **Maria Aparecida**, a minha mãe **Leila** e ao meu irmão **Rodrigo** por sempre estarem ao meu lado, me dando força e acreditando em mim.

À amiga e companheira de laboratório **Paula**, por toda a orientação, pelo carinho, pela ajuda não somente no mestrado mas também na graduação, pelos conselhos, pelas risadas e por sempre estar disposta a ajudar, “quebrando o meu galho”. Você é uma pessoa muito especial!

À amiga e companheira de laboratório **Mariana**, pela amizade, pelos conselhos, pela ajuda e por ser essa pessoa tão querida.

Ao querido **Gian Lucas**, por estar ao meu lado, por me aconselhar, por toda ajuda tanto pessoal quanto na realização desse projeto, pelo incentivo, carinho e força. Você é um presente que Deus me deu.

Ao meu orientador, **Prof. Dr. Milton Masahiko Kanashiro**, pela confiança na realização desse projeto.

Ao meu coorientador, **Prof. Dr. Álvaro Fabrício Lopes Rios**, pela oportunidade de realização desse projeto.

Ao **Prof. Dr. Victor Quintana Flores**, pelos conselhos e amizade.

A todos os amigos que de alguma forma contribuíram para a realização desse projeto. Em especial a **Laura Maciel, dona Marina, Gislane, Laura Marinone, Fernanda Curcio** por todo o carinho e **Wellington** pelo apoio técnico e pela amizade.

À querida e grande amiga, **Gabriela Petroceli** e a sua mãe **Gilza**, pela amizade, pelo carinho dedicado, pelo suporte e por estarem sempre ao meu lado. **Gabi**, você é um presente que Campos me deu.

À **UENF** e **CAPES** pela estrutura, aprovação e apoio financeiro.

ÍNDICE

ABREVIATURAS	viii
LISTA DE FIGURAS	xviii
LISTA DE TABELAS	xviii
LISTA DE GRÁFICOS	xviii
RESUMO	xix
ABSTRACT	xxi
1. INTRODUÇÃO	1
1.1 Epigenética em doenças humanas	1
1.2 Metilação do DNA e sua participação em doenças humanas	7
1.3 Variação epigenética	10
2. OBJETIVOS	17
2.1 Objetivo Geral	17
2.2 Objetivos Específicos	17
3 MATERIAL E MÉTODOS	18
3.1 Repositório de dados públicos	18
3.2 O conjunto de genes escolhidos para a análise	18
3.3 Análise do perfil de metilação do DNA em Ilhas CpGs (CGIs) e shores.	18
3.4 Identificação de elementos altamente conservados associados as cDMRs	20
3.5 Caracterização das cDMRs identificadas quanto a posição em relação ao genes estudados.	20
3.6 Identificação da presença de ligação de fatores de transcrição nas cDMRs identificadas	21
3.7 Análise estatística	22
4 RESULTADOS E DISCUSSÃO	23
4.1 Caracterização <i>in silico</i> dos genes analisados	23
4.2 Caracterização <i>in silico</i> dos perfis de metilação do DNA em genes relacionados a tumores mamários	24
4.3 Padrão de <i>GOM</i> e <i>LOM</i> nas cDMRs – associação em relação a posição nos genes e a CGIs e shores	27
4.4 Números de cDMRs com elementos evolucionariamente conservados e TFBS conservados	33
4.5 Detecção de eventos de ligação de fatores de transcrição (TFs) em CGIs e shores entre perfis hipo e hipermetilados	35
5. CONCLUSÃO	42
6. REFERÊNCIAS BIBLIOGRÁFICAS	43

APÊNDICE 1 – Tabela de genes relacionados a alterações epigenéticas de ganho ou perda de metilação em câncer de mama	62
---	----

ABREVIATURAS

- ❖ 5mC – 5' – metilcitosina
- ❖ 5mC – dif – diferença de metilação
- ❖ *ABCB1* – *ATP binding cassette, sub-family B, member 1*
- ❖ *ABCC1* – *ATP binding cassette, sub-family C, member 1*
- ❖ *ABCG2* – *ATP binding cassette, Sub-family G, member 2*
- ❖ *aDMR* – do inglês, *ageing – specific methylated region* ou *age related DMRs*
- ❖ *aGEN* – expressão de genes associados com a idade (*age associated genes*)
- ❖ ANOVA – análise estatística, do inglês, *Analysis of variance*
- ❖ *APC* – polipose adenomatosa coli (*Adenomatous Polyposis Coli*)
- ❖ *ASM* – do inglês, *allele – specific methylation*
- ❖ *ATM* – ataxia telangiectasia mutada (*Ataxia Telangiectasia Mutated*)
- ❖ *BIN1* – integrador de ponte 1 (*Brifing Integrator 1*)
- ❖ *BRCA1* – câncer de mama 1 (*Breast Cancer 1*)
- ❖ *BRCA2* – câncer de mama 2 (*Breast Cancer 2*)
- ❖ *CAV1* – proteína caveolina 1 (*Caveolin 1, Caveolae Protein*)
- ❖ *CCND1* – ciclina D1 (*Cyclin D1*)
- ❖ *CCND2* – ciclina D2 (*Cyclin D2*)
- ❖ *CDCP1* – domínio CUB contendo proteína 1 (*CUB Domain Containing Protein 1*)
- ❖ *CDH1* – caderina 1, tipo 1, E – caderina (*Cadherin, Type 1, E-cadherin*)

- ❖ *CDH11* – caderina 11 (*Cadherin 11*)
- ❖ *CDKN2A* – inibidor de quinase dependente de ciclina 2A (*Cyclin-Dependent Kinase Inhibitor 2A*)
- ❖ *cDMRs* – do inglês, *cancer differentially methylated region*
- ❖ *CDX2* – homeobox do tipo caudal 2 (*Caudal Type Homeobox 2*)
- ❖ CGI – *shores* – margens das ilhas CpGs
- ❖ *CHD1* - fator de transcrição, do inglês, *chromodomain – helicase - DNA - binding protein 1*
- ❖ *CHD2* - fator de transcrição, do inglês, *chromodomain – helicase - DNA - binding protein 2*
- ❖ *CHEK2* – checkpoint de quinase 2 (*Checkpoint kinase 2*)
- ❖ ChIP-seq – seqüenciamento por DNA derivado de imunoprecipitação de cromatina
- ❖ *C-MYC / MYC* - oncogene
- ❖ COBRA – metodologia para estudo de epigenética, do inglês, *Combined Bisulfite Restriction Analysis*
- ❖ CpG – dinucleotídeo “Citosina – fosfato – Guanina”
- ❖ CpGI ou CGI – do inglês, *CpG Islands*
- ❖ *CTBP2* - fator de transcrição, do inglês, *C – terminal binding protein 2*
- ❖ *CTCF* - fator de transcrição, do inglês, *CCCTC – binding factor*
- ❖ *CXCL1* – ligante de quimiocina 1 - *motif C-X-C* – (*Chemokine – C-X-C motif – Ligand 1*)
- ❖ *DAPK1* – proteína quinase 1 associada a morte (*Death – associated protein kinase 1*)
- ❖ *DCLK1* – quinase tipo doublecortina 1 (*Doublecortin – Like Kinase 1*)

- ❖ *DLEC1* – supressor de câncer pulmonar e esofágico 1 (*Deleted in Lung and Esophagic Cancer 1*)
- ❖ *DMR* – do inglês, *differentially methylated region*
- ❖ *DNM1* – dinamina 1 (*Dynamin 1*)
- ❖ *DNMTs* – enzimas DNA metiltransferases
- ❖ *DPYSL2* – dihidropirimidase tipo 2 (*Dihydropyrimidinase Like 2*)
- ❖ *DPYSL3* – dihidropirimidase tipo 3 (*Dihydropyrimidinase Like 3*)
- ❖ *DPYSL4* – dihidropirimidase tipo 4 (*Dihydropyrimidinase Like 4*)
- ❖ *EGR1* - fator de transcrição, do inglês, *early growth response 1*
- ❖ *EN1* – *engrailed homeobox 1*
- ❖ *ENCODE* – do inglês, *Encyclopedia of DNA elements*
- ❖ *EP300* - fator de transcrição, do inglês, *E1A binding protein p300*
- ❖ *EPBB2* – receptor EPH B2 (*EPH receptor B2*)
- ❖ *EPHB1* – receptor EPH B1 (*EPH Receptor B1*)
- ❖ *eQTLs* – do inglês, *expression quantitative trait locis*
- ❖ *ERBB2* – oncogene homólogo 2 de leucemia eritoblástica viral (*Erythroblastic Leukemia Viral Oncogene Homolog 2*)
- ❖ *ESR1* – receptor de estrogênio 1 (*Estrogen Receptor 1*)
- ❖ *EZH2* – fator de transcrição, do inglês, *Enhancer of zeste homolog 2*
- ❖ *FBXW7* – f-box e WD contendo 7 domínios repetitivos (*F -Box and WD repeat domain containing 7*)
- ❖ *FGF2* – fator de crescimento de fibroblasto 2 (*Fibroblast Growth Fator 2*)
- ❖ *FGFR2* – receptor 2 do fator de crescimento de fibroblasto (*Fibroblast Growth Factor Receptor 2*)

- ❖ *FGFR3* – receptor 3 do fator de crescimento de fibroblasto (*Fibroblast Growth Factor Receptor 3*)
- ❖ *FLT1* – tirosina quinase 1 relacionada a fms (*Fms Related Tyrosine Kinase 1*)
- ❖ *FOXA1* - fator de transcrição, do inglês, *forkhead box protein A1*
- ❖ *FOXD3* – *forkhead Box D3*
- ❖ *FOXJ1* – *forkhead box J1*
- ❖ *GATA 2* - fator de transcrição, do inglês, *GATA binding protein 2*
- ❖ *GDF15* – fator 15 de diferenciação e crescimento (*Growth Differentiation Factor 15*)
- ❖ *GOM* – do inglês, *Gain of Methylation*
- ❖ *GPER1/GPR30* – proteína G acoplada ao receptor de estrogênio 1, do inglês, *G protein coupled estrogen receptor 1*
- ❖ *GRIN2B* – receptor de glutamato, ionotrópico, N – metil – D – aspartato 2B (*Glutamate Receptor, Ionotropic, N – methyl – D – aspartate 2B*)
- ❖ *GSK1* – glicogênio sintase quinase 1 (*Glycogen Synthase Kinase 1*)
- ❖ *GSTM1* – glutationa S – transferase mu 1 (*Glutathione S – transferase mu 1 (muscle)*)
- ❖ *GSTM2* – glutationa S – transferase mu 2 (*Glutathione S – transferase mu 2 (muscle)*)
- ❖ *GSTP1* – glutationa – S – transferase Pi 1 (*Glutathione – S – Transferase Pi 1*)
- ❖ *GWAS* – do inglês, *Genome – Wide Association Studies*
- ❖ H3K27me – metilação da lisina 27 da histona 3
- ❖ H3K27me3 – trimetilação da lisina 27 da histona 3
- ❖ H3K36me – metilação da lisina 36 da histona 3

- ❖ H3K4me ou H3K4me1 – metilação da histona 4 da lisina 3
- ❖ H3K9me – metilação da lisina 9 da histona 3
- ❖ H4K20me – metilação da lisina 20 da histona 4
- ❖ HATs – enzimas histonas acetiltransferases
- ❖ *HCC1954* – linhagem celular relacionada a tumores da glândula mamária
- ❖ *HDAC2* – histona deacetilase 2 (*histone deacetylase 2*)
- ❖ *HER-2/neu* – do inglês, *Human Epidermal Growth Factor Receptor 2*
- ❖ *HMEC* – células do epitélio mamário humano (*Human Mammary Epithelial Cells*)
- ❖ *HORMAD1* – domínio HORMA, do inglês *HORMA Domain Containing 1*
- ❖ *HOTAIR* – transcrito de RNA antisense de HOX (*HOX Transcript Antisense RNA*)
- ❖ *HOXA1* – *homeobox A1*
- ❖ *HOXA10* – *homeobox A10*
- ❖ *HOXA11* – *homeobox A11*
- ❖ *HOXA5* – *homeobox A5*
- ❖ *HOXC11* – *homeobox C11*
- ❖ *HPCAL1* – hipocalcina tipo 1 (*Hippocalcin-like 1*)
- ❖ *HSM* – do inglês, *haplotype – specific methylation*
- ❖ *iDMR* – do inglês, *imprinting – specific differentially methylated region*
- ❖ *IGF2* – fator de crescimento semelhante a insulina 2 (*Insulin – like growth factor 2*)
- ❖ *JUND* - fator de transcrição, do inglês, *jun D proto oncogene*
- ❖ *KDM5B* - fator de transcrição, do inglês, *lysine demethylase 5B*

- ❖ *L3MBTL1* – proteína semelhante a tumor cerebral maligno (3) (*Lethal (3) Malignant Brain Tumor – Like Protein*)
- ❖ *Lods* – segmentos de DNA altamente conservados (*transformed log-odds score*)
- ❖ *LOM* – do inglês, *Loss of Metylation*
- ❖ *MALAT1* – transcrito 1 associado a adenocarcionama pulmonar metastático (*Metastasis Associated Lung Adenocarcinoma Transcript 1*)
- ❖ *MAX* - fator de transcrição, do inglês, *myc associated factor X*
- ❖ *MAZ* - fator de transcrição, do inglês, *myc – associated zinc finger protein*
- ❖ *MCF-7* – linhagem celular de adenocarcinoma mamário
- ❖ *MGMT* – O-6 – metilguanina – DNA metiltransferase (*O-6-methylguanine-DNA methyltransferase*)
- ❖ *MGRN1* –do inglês, *mahogunin ring finger 1*
- ❖ *MPPE1* – metalofosfoesterase 1 (*Metallophosphoesterase 1*)
- ❖ *mQLTs* - do inglês, *methylation quantitative trait locis*
- ❖ *MRE11A* – MRE11 recombinação meiótica 11 homóloga A (*MRE11 Meiotic Recombination 11 Homolog A*)
- ❖ mRNA – RNA mensageiro
- ❖ *MT1G* – metalotionação 1G (*Metallothionein 1G*)
- ❖ *MVP* – do inglês, *methulation variable position*
- ❖ *NBN* – nibrina (*Nibrin*)
- ❖ *NCAM1* – molécula de adesão celular neural 1 (*Neural Cell Adhesion Molecule 1*)
- ❖ ncRNA – RNA não codificante
- ❖ *NEUROD1* – diferenciação neuronal 1 (*Neuronal Differentiation 1*)

- ❖ *NEUROG2* – neurogenina 2 (*Neurogenin 2*)
- ❖ *NSUN2* – família da RNA metiltransferase NOP2/Sun, membro 2 (*NOP2/Sun RNA Methyltransferase Family, Member 2*)
- ❖ *NT5E* – 5' nucleotidase ecto
- ❖ *NTRK1* – receptor tipo 1 neurotrófico de tirosina quinase (*Neurotrophic Tyrosine Kinase, type 1*)
- ❖ *NTRK2* – receptor tipo 2 neurotrófico de tirosina quinase (*Neurotrophic Tyrosine Kinase, type 2*)
- ❖ *NUP155* – nucleoporina 155kDa (*Nucleoporin 155kDa*)
- ❖ *PALLD* – proteína associada ao citoesqueleto (*Palladin, Cytoskeletal Associated Protein*)
- ❖ *PAX3* – do inglês, *paired box 3*
- ❖ *PAX7* – do inglês, *paired box 7*
- ❖ Pb – pares de bases
- ❖ PcG-2/ PCR2 – do inglês, *Polycomb complex repressive 2*
- ❖ *PDE4C* – fosfodiesterase 4C (*Phosphodiesterase 4C*)
- ❖ *PDGFRA* – receptor de fator de crescimento derivado de plaquetas, polipeptídeo alfa (*Platelet Derived Growth Factor Receptor, Alpha Polypeptide*)
- ❖ *PIK3CA* – fosfatidilinositol -4,5 – bifosfato 3 – quinase, subunidade catalítica alfa (*Phosphatidylinositol-4,5-bisphosphate 3-kinase, Catalytic Subunit Alpha*)
- ❖ *PITX2* – *paired - like homeodomain 2*
- ❖ *PKD2* – doença renal policística 2 (*Polycystic Kidney Disease 2*)
- ❖ *POLR2A* – fator de transcrição, do inglês, *RNA Polymerase II Subunit A*
- ❖ *pRB* ou Rb1 – proteína retinoblastoma

- ❖ *PRKCB* – proteína quinase C beta (*Protein Kinase C Beta*)
- ❖ *PRR5- PRR5 – ARHGAP8 readthrough*
- ❖ *PTEN* – fosfatase homóloga a tensina (*Phosphatase and Tensin Homolog*)
- ❖ *PTPRO* – proteína tirosina fosfatase, receptor tipo O (*Protein Tyrosine Phosphatase, Receptor type O*)
- ❖ *PTRF* – polimerase I e fator de transcrição para liberação (*Polymerase I and Transcript Release Factor*)
- ❖ *RAB40C* – membro da família de oncogene RAS (*Member RAS Oncogene Family*)
- ❖ *RAD50 – RAD50 homolog*
- ❖ *RAD51C – RAD51 homolog C*
- ❖ *RARB* – receptor de ácido retinóico, beta (*Retinoic Acid Receptor, beta*)
- ❖ *RASSF1* – associação do domínio Ras – contendo proteína 1 (*Ras Association Domain-Containing Protein 1*)
- ❖ *RASSF2* – associação do domínio Ras - contend proteína 2 (*Ras Association Domain-Containing Protein 2*)
- ❖ *RBBP5* - fator de transcrição, do inglês, *RB binding protein 5*
- ❖ *rDMR* – do inglês, *reprogramming – specific differentially methylated region*
- ❖ *ROR2* – receptor 2 de tirosina quinase (*Receptor Tyrosine Kinase like Orphan Receptor 2*)
- ❖ *RUNX3* – fator de transcrição relacionado com Runt 3 (*Runt – Related Transcription Factor 3*)
- ❖ *SCGB3A1* – secretoglobina, família 3A, membro 1 (*Secretoglobin, Family 3A, Member 1*)
- ❖ *SFRP1* – *secreted frizzled related protein 1*

- ❖ *SFRP4* – secreted frizzled related protein 4
- ❖ *SH3GL2* – domínio SH3 GRB2 tipo 2 (*SH3-Domain GRB2-Like 2*)
- ❖ *SH3GL3* – domínio SH3 GRB2 tipo 3 (*SH3-Domain GRB2-Like 3*)
- ❖ *SIX3* – *SIX Homeobox 3*
- ❖ *SLC6A3* – família de transportadores de soluto 6, membro 3 - neurotransmissor transportador de dopamina (*Solute Carrier Family 6 – Neurotransmitter Transporter, Dopamine), Member 3*)
- ❖ SNP – do inglês, *Single Nucleotide Polymorphism*
- ❖ *SOX1* – região determinante do sexo Y 1 (*Sex Determining Region Y – SRY – box 1*)
- ❖ *SOX17* – região determinante do sexo Y 17 (*Sex Determining Region Y – SRY – box 17*)
- ❖ *SOX18* – região determinante do sexo Y 18 (*Sex Determining Region Y – SRY – box 18*)
- ❖ *SP1* - fator de transcrição Sp1 (*Transcription Factor Sp1*)
- ❖ *SP3* – fator de transcrição Sp3 (*Transcription Factor Sp3*)
- ❖ *SPEG* – *SPEG complex locus*
- ❖ *SRCIN1* – inibidor 1 de sinalização de quinase SRC (*SRC Kinase Signaling Inhibitor 1*)
- ❖ *STK11* – serina/treonina quinase 11 (*Serine/Threonine Kinase 11*)
- ❖ *SUZ12* - fator de transcrição, do inglês, *supressor of zest 12 homolog*
- ❖ *SYNM* – sinemina, proteína de filamento intermediário (*Synemin, Intermediate Filament Protein*)
- ❖ *TAF1* - fator de transcrição (*TATA – box binding protein associated factor 1*)
- ❖ *TAL1* – leucemia linfocítica aguda de célula T 1 (*T-cell Acute Lymphocytic leukemia 1*)

- ❖ *tDMR* ou T-MDRs – do inglês, *tissue – specific differentially methylated region*
- ❖ TF – fator de transcrição
- ❖ *TFBBS* – do inglês, *transcription factor binding sites*
- ❖ *TGFB1* – fator transformador de crescimento, beta 1 (*Transforming Growth Factor, Beta 1*)
- ❖ *TLX1* – do inglês, *T – cell Leukemia Homeobox 1*
- ❖ *TMEFF2* – proteína transmembrana com EGF e dois domínios semelhantes a follistatina 2 (*Transmembrane Protein with EGF-like and Two Follistatin-like Domains 2*)
- ❖ *TP53* – proteína tumoral p53 (*Tumor Protein p53*)
- ❖ *TWIST1* – gene que codifica um fator de transcrição com motivo hélice – loop – hélice 1 (*Twist Basic Helix-loop-helix Transcription Factor 1*)
- ❖ USCS – University of california, Santa Cruz
- ❖ *VMR* – do inglês, *variably methylated region*
- ❖ *WGBS* – do inglês, *whole – genome bisulfite sequencing*
- ❖ *WNT1* – *wingless-type MMTV integration site family*
- ❖ *YY1* - fator de transcrição, do inglês, *Yin Yang 1*
- ❖ *ZEB1* – “dedo de zinco” E - box ligando homeobox 1 (zinc finger E-box binding homeobox 1)
- ❖ *ZNF217* – proteína de “dedo de zinco” 217 (*Zinc Finger Protein 217*)
- ❖ *ZNF584* – proteína de “dedo de zinco” 584 (*Zinc Finger Protein 584*)

LISTA DE FIGURAS

Figura 1: Representação de resolução para estudo da metilação do DNA dependendo do contexto biológico.....	9
--	---

LISTA DE TABELAS

Tabela 1: Genes relacionados a alterações epigenéticas de ganho ou perda de metilação em câncer de mama.....	62
--	----

Tabela 2: Número de eventos de GOM e LOM nas CDMRs.....	28
---	----

LISTA DE GRÁFICOS

Gráfico 1: Níveis de metilação do DNA em relação a posição dos genes analisados (<i>upstream</i> , <i>intron</i> , <i>downstream</i>) entre as três linhagens investigadas.....	29
---	----

Gráfico 2: Níveis de metilação do DNA nas três linhagens investigadas em regiões a mais de 2kb das CGIs estudadas.....	30
--	----

Gráfico 3: Distribuição dos valores de dif 5mC% entre CGI e <i>shore</i> entre as linhagens <i>HCC1954</i> e <i>MCF-7</i>	32
---	----

Gráfico 4: Fatores de transcrição (TFs) que se ligam as CGIs tanto hipometiladas quanto hipermetiladas.....	36
---	----

Gráfico 5: Fatores de transcrição (TFs) que se ligam as <i>shores</i> tanto hipometiladas quanto hipermetiladas.....	37
--	----

RESUMO

Apesar dos avanços em protocolos de biologia molecular e análise genômica para identificação de sítios de alterações epigenéticas (epimutações) associados à etiologia de doenças humanas, muitos estudos voltam sua atenção para variações nas regiões promotoras, especialmente àquelas associadas as ilhas CpGs (CGIs). No entanto, uma vasta área do genoma, além dos promotores gênicos, pode abrigar alterações epigenéticas que podem estar associadas ao surgimento de doenças como o câncer de mama. Regiões intergênicas, introns e mesmos as regiões adjacentes as CGIs (as *shores*) podem abrigar sítios de alterações epigenéticas associados a elementos regulatórios como acentuadores, silenciadores e insuladores. Com a disponibilidade dos bancos de dados públicos de epigenomas, como os metilomas de DNA, a identificação e caracterização genômica de novas regiões associadas com a metilação diferencial entre células tumorais e saudáveis pode ser melhor explorada. A identificação de regiões diferencialmente metiladas (DMRs) em câncer de mama pode auxiliar na caracterização epigenética desses tumores e gerar uma nova fonte de alvos moleculares para utilização em terapias futuras. O presente estudo teve como objetivo a identificação e caracterização de DMRs em um conjunto de genes com relatos de alterações epigenéticas observadas em tumores mamários humanos. Foram investigados dados de metilação do DNA e características genômicas locais, como sobreposição nas CGIs e *shores*, presença de fatores de transcrição e conservação de sequências entre diferentes espécies nas DMRs identificadas. Foi realizada uma revisão bibliográfica de genes que apresentavam o perfil de metilação alterado em carcinomas mamários, a qual resultou em um total de 113 genes. Foram analisadas as CGIs e suas respectivas *shores* em cada um dos 113 genes. As análises *in silico* foram realizadas utilizando dados obtidos pelo UCSC *genome browser*. Foram associados às DMRs, dados de *Whole Genome Bisulfite Sequencing* (WGBS), dados de conservação de sequências entre espécies, associação nas CGIs e *shores* e, a presença de ligação de fatores de transcrição utilizando também dados disponíveis na plataforma *online* supracitada. O cruzamento entre dados genômicos e as coordenadas das DMRS foi realizado a partir da utilização da ferramenta *Data Integrator*

(UCSC). Foram obtidos dados quantitativos de metilação do DNA de duas linhagens de carcinoma mamário (*MCF-7* e *HCC1954*) e de tecido saudável (*HMEC*). Os resultados demonstraram que a maioria das DMRs encontradas estão localizadas em *shores*. Os perfis de metilação encontrados nas linhagens tumorais demonstraram uma prevalência de hipermetilação das DMRs. Quando observada a variação em termos de eventos de ganho e perda de metilação do DNA, as *shores* apresentaram uma diferença significativa em termos de eventos de alteração da metilação do DNA. Esses dados corroboram os dados da literatura para a ocorrência de um perfil de hipermetilação de DMRs em células tumorais. Grande parte dessas regiões apresentaram ocorrência de sequências evolucionariamente conservadas. Os dados apresentados também estão de acordo com relatos mais recentes, apontando as *shores* como elementos onde grande parte da variação epigenética em tumores é encontrada. Foram encontrados padrões diferenciais para ocupação por fatores de transcrição (TFs) entre CGI e *shores* e, entre regiões hiper e hipometiladas. Os resultados do presente estudo apontam as *shores* como as regiões de alteração epigenética prevalentes em linhagens de tumores de mama. Os diversos fatores de transcrição associados às regiões genômicas, são um indicativo de uma assinatura molecular que pode estar ligada a natureza das alterações epigenéticas em cada DMR.

Palavras – chaves: Epigenética, câncer de mama, região diferencialmente metilada (DMRs)

ABSTRACT

Despite the advance of molecular biology protocols and genomic analysis to identify the regions of epigenetic mutations (epimutations) related to etiology of human diseases, many studies are still focused in variations that happen in promoters regions, specially that ones related to the CpGS (CGIs) islands. However, a great area of the genoma, besides gene promoters, can contain epigenetic changes that can be associated with the rising of some diseases as breast cancer. Intergenic zones, introns and even the zones bordering the CGIs (as shores) can shelter places of epigenetic changes that associated with regulatory elements as enhancers, silencers and insulators. The identification and genomic description of new regions associated with differential methylation between tumor cells and healthy cells can be better explored because nowadays the public databases of epigenome are available, as the methyloma of DNA for example. The identification of differentially methylated regions (DMRs) in breast cancer can help the process of epigenetic description of these tumors and, create a new source of molecular targets that could be used in future therapies. The objective of this study was the identification and description of the DMRs jointly some genes in which was observed some epigenetic modifications in human breast tumors. The DNA methylation and genomic local characteristics were investigated as the overlap of CGI and shores, presence of transcription factors and conservation of sequences between different species in the DMRs identified. It was made a biographical revision of some genes that were expressing their methylation profile modified in breast carcinomas, in which has result the 113 genes. It were analysed the CGIs and their respective shores in which one of 113 gene. The analysis *in silico* were made using data extracted from UCSC genome browser. Some data from the *Whole Genome Bisulfite Sequencing* (WGBS), from the sequence conservation between species, the association between CGIs and their shores, and the presence of connection between transcription factors, that also used data extracted from the web browser before mentioned, were associated with the DMRs. The cross between the genomic data and the coordinates of the DMRs was made using the *Data Integrator* (UCSC) tool. Quantitative DNA

methylation data were obtained from two mammary carcinoma lines (MCF-7 and HCC1954) and from healthy tissue (HMEC). The results showed that most of the DMRs were founded in shores. The methylation profile found in the tumor lines showed a prevalence of hypermethylation of the DMRs. When observed the variation in terms of gain and loss of DNA methylation, the shores presented a significant difference in terms of DNA methylation alteration. These data corroborate with the literature data for the occurrence of a hypermethylation profile of DMRs in tumor cells. Most of these regions had evolutionarily conserved sequences. The data presented are also in agreement with more recent reports, pointing the shores as elements where much of the epigenetic variation in tumors is found. Differential patterns were found for occupancy by transcription factors (TFs) between CGI and shores, and between hyper and hypomethylated regions. The results of the present study point to shores as the regions of epigenetic alteration prevalent in lineages of breast tumors. The various transcription factors associated with genomic regions are indicative of a molecular signature may be linked to the nature of epigenetic changes in each DMR.

Key words: Epigenetics, breast cancer, differentially methylated region (DMRs)

1. INTRODUÇÃO

1.1 Epigenética em doenças humanas

A epigenética é um ramo da genética que investiga mudanças na atividade dos genes, as quais podem ser herdáveis e não são atribuídas a alterações na sequência do DNA (BROOKES & SHI, 2014). Os mecanismos epigenéticos possuem um papel central no desenvolvimento de um indivíduo, uma vez que atuam como fatores de interação entre a informação genética contida no genoma e o ambiente em que o mesmo está inserido. Dessa forma, a variação em características fenotípicas de um indivíduo é mediada pela ação do ambiente sobre os mecanismos epigenéticos (MOORE, 2016)

Atualmente, os mecanismos epigenéticos mais estudados são os RNAs não codificadores de proteínas (ncRNAs), as modificações covalentes de histonas e a metilação do DNA (BROOKES & SHI, 2014; GU et al., 2016; ZHANG et al., 2017). Os mecanismos epigenéticos podem atuar tanto de forma direta quanto indireta na regulação da expressão gênica, modulando a ação de fatores de regulação da transcrição (KOFINK et al., 2013). Sua influência no controle da transcrição gênica consiste principalmente na regulação do *status* de condensação da cromatina, no qual os genes encontram-se inseridos (GIBNEY & NOLAN, 2010; HAMILTON, 2011). Alterações nesses mecanismos podem dessa forma, contribuir para o surgimento de diversas doenças humanas (BAKUSIC et al., 2016; BROOKES & SHI, 2014).

Novos protocolos de sequenciamento em larga escala e a identificação a partir dos mesmos de uma ampla gama de elementos regulatórios tem

auxiliado na elucidação da etiologia de várias doenças genéticas multifatoriais (RAKYAN et al., 2011). Por exemplo, a análise genômica utilizando metodologias de GWAS (*genome – wide association studies*) tem permitido a identificação de centenas de SNPs associados a doenças humanas (HINDORFF et al., 2009). No entanto, as bases genéticas de muitas patologias ainda não estão bem elucidadas, e mesmo técnicas como o sequenciamento completo do genoma ou somente de exomas ainda não explicam a origem de muitas doenças (FEINBERG & IRIZARRY, 2010; KULIS & ESTELLER, 2010; PETRONIS, 2010). Nesse cenário, o entendimento da participação dos fatores epigenéticos pode contribuir para a compreensão dos mecanismos moleculares associados à etiologia de muitas patologias, onde apenas a base genética não é suficiente (RAKYAN et al., 2011).

Os ncRNAs são moléculas que estão amplamente distribuídas no transcriptoma humano. Esses RNAs atuam em diversos mecanismos biológicos, como a inativação do cromossomo X em mamíferos, no mecanismo de compensação de doses em insetos, no *imprinting* genômico e no silenciamento mediado por proteínas do complexo *polycomb* (MATTICK & MAKUNIN, 2006; ZARATIEGUI et al., 2007). A ação dos ncRNAs pode ocorrer através do recrutamento da maquinaria de modificação epigenética aos genes (JANOWSKI et al., 2006; KIM et al., 2006). Um exemplo da ação de recrutamento das maquinarias protéicas de regulação gênica desses RNAs vem da descoberta de diversos ncRNAs encontrados expressos em regiões a montante ou sobrepostos a região 5' de genes que codificam proteínas (CORE et al., 2008; FEJES-TOTH et al., 2009; KAPRANOV et al., 2007; PREKER et al., 2008; SEILA et al., 2008).

As modificações covalentes de histonas são depositadas nas caudas N e C terminais dessas proteínas e são mediadas por diversas enzimas que catalisam diferentes tipos de modificações em diferentes tipos de resíduos de aminoácidos (KOUZARIDES, 2007). As histonas constituem um grupo de nucleoproteínas compostas por domínios globulares e caudas N- terminais com formato irregular, sendo responsáveis pela compactação do DNA e remodelamento da cromatina (BARSKI et al., 2007). Entre as modificações mais estudadas encontram-se a metilação, a acetilação, a ubiquitinação e a fosforilação de histonas (HARR et al., 2016)

A acetilação de histonas ocorre devido à transferência de um grupo acetil por enzimas histonas acetiltransferases (HATs) nos resíduos de lisina nas caudas N-terminais das histonas. Essa modificação altera suas características físico-químicas alterando a capacidade de se ligarem lateralmente aos grupamentos fosfato da molécula de DNA, neutralizando a carga positiva da lisina (ANDREOLI & DEL RIO, 2014). A acetilação de histonas geralmente está associada com a ativação da transcrição gênica. Assim, a acetilação das lisinas 9 e 14 da histona H3 e das lisinas 5, 8, 12 e 16 da histona H4 estão associadas a ativação pela formação de eucromatina. Antagonicamente, a de-acetilação dos resíduos de lisina leva a condensação da cromatina e inativação da transcrição gênica (WEI et al., 2017). A acetilação de histonas também interage com a metilação do DNA (REIK, 2007; TCHURIKOV, 2005).

A metilação de histonas, que frequentemente ocorre nos resíduos de lisina ou arginina em H3 e H4, também contribui para regular o *status* de

condensação da cromatina. A metilação de lisina nas histonas pode levar tanto a ativação quanto ao silenciamento gênico, dependendo do contexto biológico. A metilação da lisina 9 da histona 3 (H3K9me), da lisina 27 da histona 3 (H3K27me) e da lisina 20 da histona 4 (H4K20me) são conhecidas por silenciar diferentes regiões do genoma através do recrutamento do complexo *polycomb* 2 (PCR2), formando heterocromatina (FERRARI et al., 2014; MOREY & HELIN, 2010). No entanto, a metilação da lisina 4 da histona 3 (H3K4me) e lisina 36 da histona 3 (H3K36me) são consideradas marcas de ativação (MAUPETIT-MÉHOUAS et al., 2016).

A metilação do DNA é o mecanismo epigenético mais investigado até o momento, quando se deseja avaliar padrões epigenéticos diferenciais em indivíduos saudáveis ou acometidos por diferentes tipos de doenças. Em mamíferos, a metilação do DNA ocorre preferencialmente em dinucleotídeos CpGs (5mC), sendo regulada por um grupo de enzimas denominadas DNA metiltransferases (DNMTs). Essas enzimas catalisam a transferência de um grupo metil a um carbono 5' de um anel de citosina (VAN OTTERDIJK & MICHELS, 2016). A presença dessa marca em promotores gênicos ou elementos regulatórios é quase sempre um indicativo de formação de heterocromatina, que está associada ao silenciamento gênico local (MOORE, 2016).

Nas últimas duas décadas, um crescente número de publicações tem correlacionado a presença de alterações nos perfis de marcas epigenéticas à etiologia de diversas doenças humanas (ROBERTSON, 2005). Resultados provenientes de diversos estudos vêm demonstrando o papel da epigenética

no desenvolvimento de doenças associadas a alterações metabólicas como o diabetes tipo 2, a obesidade, as doenças cardiovasculares, as doenças imunológicas, os transtornos psiquiátricos e no desenvolvimento de câncer (BAUER et al., 2016; RACITI et al., 2017; RAKYAN et al., 2011; SHRIVASTAVA & DESOUSA, 2016; SMOLLER, 2016).

BAYLIN & JONES (2011) demonstraram que as alterações na metilação do DNA podem cooperar com eventos genéticos no desenvolvimento de diversos carcinomas humanos. Têm sido identificados diversos padrões de metilação alterados em adenocarcinomas pulmonares, gástricos, pancreáticos, câncer de cólon, câncer de ovários e câncer de mama (CHAE et al., 2016; FUJIMOTO et al., 2005; GUO et al., 2016; HANSMANN et al., 2012; HOLM et al., 2016; IRIZARRY et al., 2009a; TIAN et al., 2016; YAN et al., 2017). Segundo IRIZARRY et al., (2009a) existe uma tendência dos oncogenes apresentarem hipometilação enquanto que os supressores tumorais se encontram hipermetilados.

Estudos em carcinomas mamários vêm demonstrando a importância da epigenética, com foco especial na metilação do DNA, no desenvolvimento desses tipos de tumores (HON et al., 2012; MANJEGOWDA et al., 2017; PARK et al., 2012; TERRY et al., 2016). CHAE et al., (2016) demonstraram a importância do processo de metilação em diversos subtipos de câncer de mama. Segundo os autores, o processo de metilação do DNA é crucial para a regulação gênica desses subtipos tumorais.

O câncer de mama é um dos tipos de tumores cuja ocorrência de novos casos vem aumentando em todo o mundo. Esse tumor é o segundo mais

frequente e o tipo mais comum em mulheres, tendo em média uma estimativa de 1,6 milhões de novos casos por ano mundialmente (JEMAL et al., 2011; MURRAY et al., 2012). No Brasil e no mundo a taxa de mortalidade por essa doença é alta, pois o diagnóstico normalmente é realizado em estágios avançados (DEY, 2014; FELIX et al., 2011)

A maioria dos tumores de mama se origina de células epiteliais e apresentam mutações em diferentes genes. Segundo OZBUN & BUTEL, (1995) pelo menos metade dos tumores possui mutação no gene *TP53*, um supressor tumoral. Estudos remotos também acharam mutações comuns em carcinomas mamários que desregulam a via da proteína retinoblastoma (pRB), incluindo perda da expressão de *RB1* ou *CDKN2A (P16^{INK1a})* (VARLEY et al., 1989, BRENNER et al., 1996) ou superexpressão da proteína Ciclina D1 (GILLETT et al., 1994).

Células de câncer de mama também podem apresentar alteração na via de sinalização Ras (CLARK & DER, 1995), decorrentes de diversos mecanismos, sendo que o mais comum destas é a superexpressão do gene *HER-2/neu* (SLAMON et al., 1989). A amplificação de *C-MYC* também foi observada (ESCOT et al., 1986). Mesmo que muitas dessas mutações já tenham sido descritas para o fenótipo tumoral de carcinoma mamário, nenhuma delas está envolvida universalmente em todos os casos de câncer de mama humano. A quantidade de genes mutados que existem no genoma e, que contribuem para o fenótipo tumoral ainda é desconhecida (ELENBAAS et al., 2000).

Enquanto grande número de mutações genéticas está relacionado ao surgimento do câncer de mama, a atuação de fatores epigenéticos é cada vez mais evidente (HON et al., 2012). As alterações nos perfis de metilação do DNA em diferentes *loci* têm sido associadas ao início do desenvolvimento tumoral (KOHONEN-CORISH et al., 2007). Alguns genes que frequentemente se encontram hipermetilados em câncer de mama estão envolvidos em mecanismos celulares como a apoptose (*RASSF1*, *HOXA5*, *TWIST1*), o controle da senescência celular (*CCND2*, *CDKN2A*), enquanto outros regulam o reparo do DNA (*BRCA1*), o crescimento celular (*ESR1*, *PGR*) e a invasão de tecidual (*CDH1*) (DWORKIN et al., 2009; JOVANOVIC et al., 2010).

1.2 Metilação do DNA e sua participação em doenças humanas

A metilação do DNA é uma modificação covalente herdável, crucial para diversos processos biológicos tais como a regulação da transcrição gênica, diferenciação celular e também no que tange o desenvolvimento de diversas doenças (TIRADO-MAGALLANES et al., 2017).

Em um contexto *loci* específico, a ação de silenciamento da metilação do DNA tem sido estudada principalmente em regiões ricas em dinucleotídeos CpGs como por exemplo, nos promotores gênicos associados a ilhas CpGs (*CpG islands* - CGIs), pois em mamíferos 80% dos promotores estão associados a CGIs, (TIRADO-MAGALLANES et al., 2017). No entanto, alterações de metilação do DNA em regiões adjacentes as CGIs (como as *CGI-shores*) têm sido encontradas em diferentes doenças e relacionadas à regulação de genes próximos (IRIZARRY et al., 2009b).

Durante alguns anos, devido a limitações tecnológicas e ao estudo da metilação do DNA apenas na função de silenciamento gênico, pesquisas alvejaram somente promotores gênicos (particularmente dos promotores associados á CGIs).

Com o advento das novas tecnologias de sequenciamento global do genoma, em especial a *whole-genome bisulfite sequencing* (WGBS), o estudo da metilação do DNA tem alcançado novas faixas de resolução e expansão além das CGIs, abrindo a possibilidade para diversas funções da metilação do DNA (ADUSUMALLI et al., 2014). Essas novas descobertas possibilitaram o desenvolvimento de estratégias computacionais que não limitam seus estudos apenas à metilação no contexto de associação da expressão gênica nos promotores (ADUSUMALLI et al., 2014). Novas metodologias utilizadas em estudos recentes revelaram a necessidade de se buscar diferentes escalas genômicas, de uma forma que seja possível correlacionar a metilação do DNA com as características da cromatina (TIRADO-MAGALLANES et al., 2017) (Figura 1).

Apesar de classicamente a metilação do DNA estar relacionada à função repressiva da transcrição gênica, dados derivados de sequenciamento global de protocolos baseados em modificação do DNA por bissulfito de sódio, indicam que a função de repressão é apenas uma das faces dessa marca epigenética. A metilação do DNA também já foi associada à ativação gênica, a regulação do *splicing* de mRNAs, ao posicionamento de nucleossomos e ao recrutamento de fatores de transcrição (BAHAR et al., 2014; BELL et al., 2011; BREILING & LYKO, 2015; CHODAVARAPU et al., 2010; FUJIMOTO et al., 2005; GIBBS et al., 2010; GUTIERREZ-ARCELUS et al., 2013; HU et al., 2013;

JIMENEZ-USECHE et al., 2013; KELLY et al., 2012; SHUKLA et al., 2011; ZHU et al., 2003).

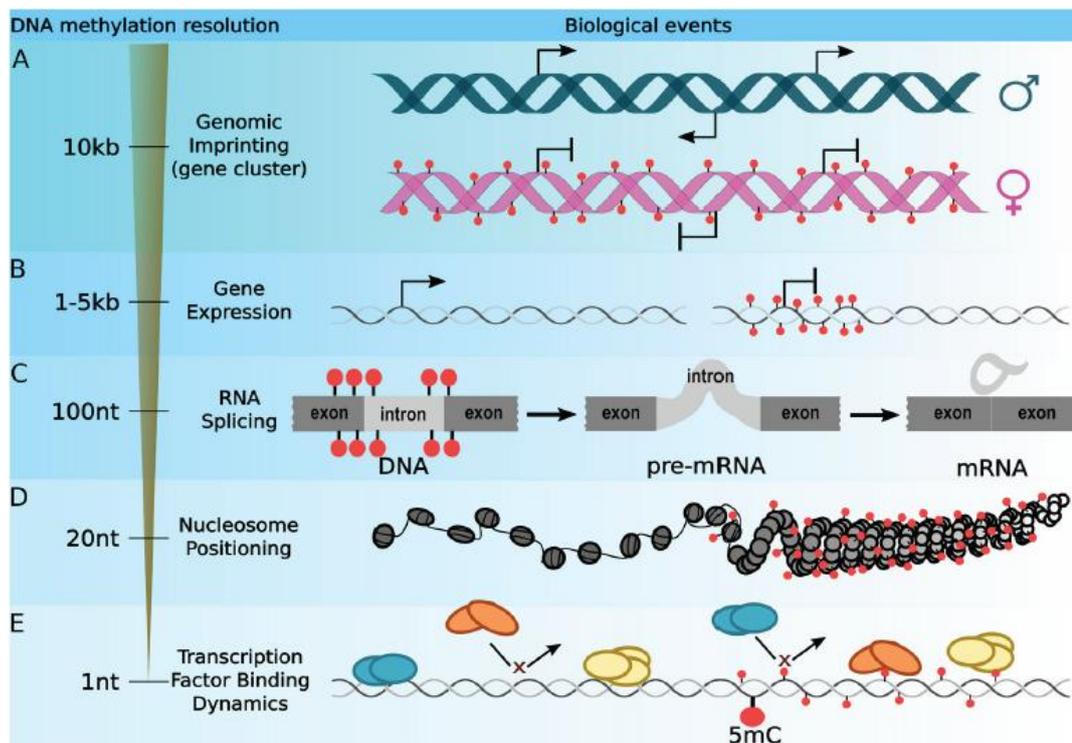


Figura 1: **Representação de resoluções para estudo da metilação do DNA dependendo do contexto biológico.** (A) 1 a 10kb - Estudo de imprinting genômico (B) 1 – 5kb - Estudo da metilação correlacionada a expressão de genes. (C) (100 pb) Para o estudo do processo de metilação correlacionado a outros mecanismos como splicing (D) 20 pb - Estudo da correlação entre a metilação do DNA e o posicionamento de nucleossomos. (E) 1 pb – Estudos visando a análise de recrutamento de fatores de transcrição (TIRADO-MAGALLANES et al., 2017).

Ainda que o mecanismo de metilação do DNA seja estudado há vários anos, todas as facetas do processo de metilação ainda não estão completamente elucidadas. BRENET et al., (2011) demonstraram que quando o processo de metilação ocorre no primeiro exon de um gene o resultado é o silenciamento gênico, assim como quando a metilação ocorre na região

promotora de um gene (ARAI et al., 2006; PENG et al., 2006; SUZUKI & BIRD, 2008; TIRADO-MAGALLANES et al., 2017). Contrariamente à função nas regiões 5' dos genes, quando a metilação do DNA ocorre no corpo gênico, está associada positivamente a transcrição do gene (FENG et al., 2010; JONES, 2012; WU et al., 2010; YU et al., 2013; ZEMACH et al., 2010; ZHANG et al., 2006; ZILBERMAN et al., 2007).

1.3 Variação epigenética

Alguns aspectos que devem ser considerados nos estudos associados alterações nos perfis de metilação do DNA são: a extensão da alteração; o perfil célula ou tecido específico e; a possível origem dos fatores que causam variação. A metilação do DNA pode variar em decorrência de diversos fatores, sendo estes genéticos ou ambientais. No contexto genômico a extensão da variação pode ser limitada a um único nucleotídeo, a regiões de tamanhos variados e mesmo ao genoma como um todo. Segundo revisado por RAKYAN et al., (2011) a variação da metilação do DNA pode ocorrer em diferentes contextos *loci* específicos em:

- Posições variáveis de metilação (*Methylation variable position - MVP*). Quando um único sítio CpG apresenta metilação diferencial (IRIZARRY et al., 2009a);
- Em regiões diferencialmente metiladas (*Differentially methylated region - DMR*). Uma região do genoma onde múltiplos CG adjacentes mostram padrão diferencial de metilação. As DMRs podem ocorrer em diferentes contextos, como por exemplo: em gene regulados por *imprinting*

genômico (iDMR — *imprinting-specific differentially methylated region*) (HANNA & KELSEY, 2017) de maneira tecido específico (tDMR — *tissue-specific differentially methylated region*) (IRIZARRY et al., 2009a); durante eventos de reprogramação celular (rDMR — *reprogramming-specific differentially methylated region*) (MANTSOKI et al., 2015); em diferentes tipos de câncer (cDMR — *cancer-specific differentially methylated region*) (CHAE et al., 2016; HEYN et al., 2016) e; durante o envelhecimento (*aDMR — ageing-specific differentially methylated region*) (LIN et al., 2016; ZHANG et al., 2017);

- Regiões de metilação variável (*Variably methylated region - VMR*). Regiões que são definidas pelo aumento da variabilidade de metilação, mais que o ganho ou perda específico dessa marca epigenética (FEINBERG & IRIZARRY, 2010)
- *Metilação alelo específica (Allele-specific methylation - ASM)*. Posições ou regiões do genoma cuja metilação do DNA está associada preferencialmente a um dos alelos, dependendo da origem parental, da ação de polimorfismos genéticos ou de eventos estocásticos (RAKYAN et al., 2011);
- *Metilação haplótipo específica (Haplotype-specific methylation - HSM)*. Consiste em uma região diferencialmente metilada definida pela herança conjunta de um grupo de SNPs (Haplótipo) (KINDT et al., 2017);
- *Ilhas CpGs (CpG islands - CGIs)*. Regiões enriquecidas em dinucleotídeos CpGs, não metiladas em diversos tecidos, mas que podem apresentar ganho de metilação em doenças como o câncer

(GOPALAKRISHNAN et al., 2008; GUO et al., 2016; HEYN et al., 2016; YAN et al., 2017); e

- *As margens das ilhas CpGs (CGI - shores)*. São regiões imediatamente adjacentes as CGIs (2Kb) que apresentam alta variação na metilação do DNA apesar de baixa densidade de sítios CpGs (CHAE et al., 2016; IRIZARRY et al., 2009b).

Mesmo que o mecanismo de metilação do DNA seja estudado por alguns anos, o modelo ainda não está totalmente esclarecido, necessitando de investigação mais profunda, devido à complexidade do tema. Como citado anteriormente, grande parte dos estudos limitam-se apenas a função clássica atribuída ao processo de metilação, o silenciamento gênico. No entanto, como já fora mencionado, essa marca epigenética possui outras funções que necessitam de maiores investigações (JONES, 2012).

Como revisado por JONES (2012) a posição no genoma em que o mecanismo de metilação do DNA ocorre influencia a expressão gênica, pois dependendo do local (corpo gênico ou região promotora, shore ou CGI) metilado o efeito na expressão dos genes pode ser de ativação ou de repressão. Assim, o mecanismo de metilação do DNA possui diversas variantes, não sendo tão previsível como se acreditava.

Análises de pirosequenciamento por bissulfito demonstraram que a metilação do DNA tecido-específica não ocorre somente em CGIs como se acreditava, mas também nas CGI-*shores*. Esses dados abrem novos campos para investigações de como o direcionamento da metilação do DNA para essas regiões afeta a regulação da expressão gênica (IRIZARRY et al., 2009b). As investigações de IRIZARRY et al., (2009b) apontam o papel das CGI-*shores* na

expressão gênica no contexto de metilação diferencial tecido–específica e em câncer.

A ação da metilação do DNA sobre a ligação sítio específica dos fatores de transcrição (TF) ao DNA é outro aspecto regulatório importante dessa marca epigenética. A informação da sequência do DNA que é lida por fatores de transcrição (TFs) e outras proteínas que interagem com sequências específicas do DNA podem ser reguladas pela ação da metilação (AHSENDORF et al., 2017).

A metilação do DNA pode aumentar ou diminuir a afinidade de interação do DNA com proteínas (FUJIMOTO et al., 2005; HU et al., 2013; ZHU et al., 2003), que por sua vez modula o recrutamento de fatores de transcrição, influenciando assim a ativação ou repressão de genes, dependendo dos elementos regulatórios (promotores, insuladores, potenciadores e silenciadores) e o/ou fatores de transcrição (repressores ou ativadores) (JONES, 1999; LAI et al., 2010; TATETSU et al., 2007). Um estudo realizado por HU et al., (2013) analisou o efeito da metilação de uma única citosina de CGIs em 154 TFs que se ligavam ao DNA. Esses autores verificaram que a interação dos sítios de ligação metilados no DNA com TFs não são unidirecionais como havia sido proposto. Dependendo do TF, a interação desses elementos com o DNA poderia ser melhorada ou dificultada pela metilação. Além disso, alguns TFs mostraram a capacidade de se ligarem a diferentes motivos dependendo do *status* de metilação do DNA. Estudos anteriores também corroboram os dados de HU et al., (2013) ao demonstrarem que a metilação do DNA fora de CGIs, mas próximas a TFBSs podem alterar a força de ligação de TFs com o DNA (FUJIMOTO et al., 2005).

O mecanismo de metilação do DNA também pode ocorrer em QTLs (*quantitative trait locus*), levando a alterações epigenéticas como modificações na cromatina e na expressão gênica e muitas vezes aumentando o risco de desenvolvimento de doenças. Inicialmente parece haver uma relação positiva entre QTLs de expressão (eQTLs) e de metilação (mQTLs), onde a metilação do DNA em elementos regulatórios mais distais parece ter um efeito ativador na expressão gênica. Dessa forma, há indícios que TFBSs vinculados a mQTLs desempenham um papel na alteração dos níveis de metilação nessa região (BANOVIČH et al., 2014).

MOAREFI & CHÉDIN (2011) identificaram que a metilação em regiões centroméricas constitui outro evento relevante, uma vez que promove a estabilidade cromossômica, ajudando a manter o funcionamento correto de processos biológicos essenciais para a vida, como por exemplo, a segregação de cromossomos no processo de mitose celular. Estes autores também encontraram evidências de que o processo de metilação é importante para a estabilidade do genoma, pois é susceptível a suprimir a expressão de elementos transponíveis.

Considerando essas variações na ação da metilação do DNA em diferentes etapas da regulação gênica, estudos mais detalhados correlacionando a região do genoma e/ou elementos regulatórios onde a metilação ocorre são necessários para desvendar o potencial regulatório total dessa marca.

Atualmente, a caracterização molecular da heterogeneidade celular em tecido tumoral representa um dos maiores desafios das análises epigenômicas (TIRADO-MAGALLANES et al., 2017). A identificação de câncer DMRs

(cDMRs), a partir de bancos de dados públicos, é crucial para a análise da metilação derivada de populações celulares heterogêneas (HE et al., 2013; XIE et al., 2011a; ZHENG et al., 2014). Amostras tumorais possuem uma difícil identificação de sinais de metilação devido à complexidade da amostra, pois são compostas por vários subtipos celulares diferentes, que se desenvolvem e diferenciam-se de uma forma não sincrônica (ZHENG et al., 2014).

Apesar da dificuldade de identificação de padrões de metilação em amostras de câncer devido a sua complexidade, estudos conseguiram identificar que as CGIs em amostras tumorais tendem a estar com um perfil de metilação aberrante (PORTELA & ESTELLER, 2010; ROBERTSON & WOLFFE, 2000). No entanto, pesquisas também demonstraram que não apenas as CGIs, mas as CGIs – *shores* são responsáveis pelo desenvolvimento de diversas patologias estando estas diretamente relacionadas com a metilação diferencial tecido-específica em câncer (IRIZARRY et al., 2009b).

Estudos mais recentes verificaram que no contexto das DMRs parece que apenas poucos dinucleotídeos, ou até mesmo um, podem ser responsáveis pelo *status* de metilação de uma região (NILE et al., 2008; TSUBOI et al., 2017).

Alterações epigenéticas de DMRs em regiões diferentes das CGIs como as *shores*, parecem abrigar uma ampla gama de variações epigenéticas e a metilação em um ou poucos nucleotídeos parece estar relacionada a mudanças na expressão gênica. Dessa forma, há extrema necessidade de se caracterizar com maior precisão as DMRs em amostras tumorais a fim de se obter informações mais relevantes para a área clínica, e para posterior

identificação dessas regiões no desenvolvimento e progressão do câncer de mama.

2. OBJETIVOS

2.1 Objetivo Geral

Identificar e caracterizar regiões com metilação diferencial do DNA entre linhagens de câncer de mama e uma linhagem não tumoral em genes associados à etiologia do carcinoma mamário humano, utilizando dados de WGBS presentes em banco de dados públicos.

2.2 Objetivos Específicos

2.2.1 Obter e analisar os perfis de metilação do DNA das linhagens celulares humanas de carcinoma mamário provenientes do UCSC *genome browser* em um conjunto de 113 genes com relatos prévios de alteração de metilação do DNA em tumores de mama;

2.2.2 Caracterizar as cDMRs quanto a localização no *locus* gênico, presença de sítios de fatores de transcrição, elementos conservados e perfis de metilação; e

2.2.3 Analisar o conjunto de dados obtidos e integrá-los de modo que possam determinar assinaturas epigenéticas em diversos genes relacionados ao câncer de mama.

3 MATERIAL E MÉTODOS

3.1 Repositório de dados públicos

Os dados de WGBS (*Whole Genome Bisulfite Sequencing*) utilizados nesse trabalho foram obtidos a partir de bancos de dados públicos, disponíveis no UCSC *Genome Browser*, (*human genome* GRCh37/hg19) (<http://genome.usc.edu>). Foram utilizados dados de WGBS de linhagens *MCF7* (GRIMMER et al., 2014), *HCC1954* e *HMEC* (HON et al., 2012).

3.2 O conjunto de genes escolhidos para a análise

A fim de verificar o perfil de metilação de linhagens tumorais de diferentes genes relacionados a etiologia do câncer de mama, foi realizado um levantamento bibliográfico dos genes que apresentavam perfil de metilação alterado neste tipo de tumor (CHAE et al., 2016; HOLM et al., 2016; JOHNSON et al., 2017; KHAKPOUR et al., 2017; PARK et al., 2012; SPITZWIESER et al., 2016; TERRY et al., 2016; TIAN et al., 2016; ZHANG et al., 2017)

3.3 Análise do perfil de metilação do DNA em Ilhas CpGs (CGIs) e *shores*.

Foram analisadas CGI e suas respectivas *shores* (2Kb a montante e 2Kb a jusante de cada CGI) descritas no UCSC *genome browser* (*CpG island track*) em cada um dos 113 genes selecionados para o presente estudo. O perfil de metilação das linhagens de carcinoma mamário e tecido saudável

foram analisados em planilha do *Windows Excel* 2016. Para cada um dos 113 genes foi criada uma planilha que apresentava dados quantitativos de metilação das três linhagens celulares (*MCF-7*, *HCC1954* e *HMEC*).

A partir dos dados de metilação do DNA em dinucleotídeos CpGs únicos (5mC) foi realizado o cálculo da diferença de metilação entre a linhagem tumoral e as linhagens celular de epitélio mamário humano utilizando a fórmula $5mC\ dif = 5mC^{HMEC} - 5mC^{Linhagen_tumoral}$. Os resultados de 5mC-dif foram utilizados para avaliar eventos de ganho (*GOM – gain of methylation*) ou perda (*LOM – loss of methylation*) de metilação em cada dinucleotídeo CpG.

Foram classificadas como DMRs regiões com no mínimo 100 pb (pares de bases) e com 5mC-dif média de -10% (*LOM*) ou + 10% (*GOM*) entre a linhagem *HMEC* e ambas as linhagens tumorais (modificado de JÜHLING et al., 2016).

Após essa etapa, foi criada uma nova planilha onde o principal objetivo foi apresentar por meio de gráficos como se comportava o perfil de metilação de cada uma das três linhagens celulares. Nessa planilha foram colocadas outras informações como por exemplo: (1) as coordenadas gênicas da região em que está ocorrendo *GOM* ou *LOM*; (2) a região em que o fenômeno ocorre (CGI, *shore*, ou CGI-*shore*); (3) a posição de cada DMR em relação ao gene (2Kb a montante, dentro de íntrons, 2Kb a jusante e regiões que não se enquadravam nas três anteriores); (4) e a estatística descritiva da metilação em cada região.

3.4 Identificação de elementos altamente conservados associados as cDMRs

As coordenadas de cada cDMR encontradas foram utilizadas como um *custom track* no UCSC *genome browser* e utilizadas na ferramenta *Data Integrator* nesse mesmo *web browser* para identificação de regiões altamente conservadas e sítios de ligação de fatores de transcrição (TFBS - *transcription factor binding sites*) conservados entre diferentes espécies. Para a verificação foram adicionados ao *Data integrator* os seguintes tracks: foi adicionado um *track* com as coordenadas genômicas a serem analisadas, *TFBS conserved* e *Conservation – Element (phastCons em 100 vertebrados)*. Os segmentos conservados identificados foram representados pela sigla *lod* (essa sigla deriva de *transformed log-odds score*; com valores que variam de 0-1000), e representa parte dos cálculos de predição de segmentos conservados entre 100 espécies de vertebrados utilizados pelo *PhastCons*.

3.5 Caracterização das cDMRs identificadas quanto a posição em relação ao genes estudados.

As coordenadas de posições a montante, a jusante e de introns do genoma humano (GRCh37/hg19) foram obtidas em banco de dados disponíveis no UCSC *genome browser*. Foi utilizada a ferramenta *Table browser* com as seguintes especificações: *clade=mammal > genome=human> assembly=Feb. 2009 (GRCh37/hg19) > group=Gene and Gene prediction >*

region=genome > output file=custom track > get output= "Upstream by 2000 bp" / "Intron" / "Downstream by 2000 bp".

Esses *custom tracks* foram integrados com *custom track* das coordenadas genômicas da cDMRs identificadas pela análise do padrão diferencial de metilação do DNA entre as linhagens de câncer de mama (HCC1954/MCF7 vs HMEC). Para essa integração de coordenadas, foi utilizada a ferramenta *Data integrator*, disponível no UCSC *genome browser*. Essa integração permitiu correlacionar cada cDMR a sua posição em relação aos genes estudados.

3.6 Identificação da presença de ligação de fatores de transcrição nas cDMRs identificadas

Para auxiliar na elucidação da possível função regulatória das DMRs identificadas no presente estudo, foi investigada a presença de ligação de fatores de transcrição (TFs) nas sequências identificadas a partir de dados de ChIP-seq de bancos de dados do ENCODE (*Encyclopedia of DNA elements*). Foram utilizados dados da base de dados *Factorbook* obtidos a partir do UCSC *genome browser* (https://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=653189759_FhgBQzA6A0uC47jmu77AZj4nVEaA&c=chr1&g=wgEncodeRegTfbsClusteredV3).

O banco de dados *Factorbook Motifs* contém dados de 837 experimentos contemplando 167 fatores de transcrição. Para averiguação da ligação de TFs as cDMRs as coordenadas genômicas das mesmas foram integradas aos dados do *Factorbook* utilizando a ferramenta *Data Integrator*

(UCSC genome browser > tools). As coordenadas das cDMRs foram divididas em CGI hipermetiladas e hipometiladas (CGI-hiper e hipo) e *shores* hipermetiladas e hipometiladas (*shore*-hiper e hipo). Regiões de ligação entre CGI e *shores* (CGI-*shores*) não foram incluídas por não haver como distinguir de onde a metilação sofre mais influência (CGI ou *shore*).

3.7 Análise estatística

Os dados referentes à estatística descritiva e a análise de variância foram obtidos no *software* GraphPad Prism 5.0.

4 RESULTADOS E DISCUSSÃO

4.1 Caracterização *in silico* dos genes analisados

Foi encontrado um total de 113 genes, listados a seguir: *BRCA1*, *HORMAD1*, *ERBB2*, *PIKA3CA*, *TP53*, *CDH1*, *BRCA2*, *APC*, *RASSF1*, *HPCAL1*, *NEUROG2*, *SLC6A3*, *RAB40C*, *ZNF584*, *FOXD3*, *PDE4C*, *MPPE1*, *PRR5*, *RAD51C*, *ATM*, *NEUROD1*, *TAL1*, *SOX1*, *FGF2*, *ESR1*, *GSTM2*, *SIX3*, *EN1*, *PAX3*, *GSK1*, *PKD2*, *CCND1*, *TWIST1*, *CXCL1*, *CDCP1*, *SOX17*, *DLEC1*, *GRIN2B*, *GSTP1*, *HOXA1*, *HOXA10*, *IGF2*, *MT1G*, *RARB*, *RUNX3*, *SCGB3A1*, *SFRP1*, *SFRP4*, *TMEFF2*, *CCND2*, *CDKN2A*, *DAPK1*, *MGMT*, *PITX2*, *NT5E*, *ZNF217*, *PTPRO*, *HOTAIR*, *L3MBTL1*, *GSTM1*, *SYNM*, *FBXW7*, *NUP155*, *STK11*, *PTEN*, *CHEK2*, *MRE11A*, *RAD50*, *NBN*, *FGFR2*, *NTRK1*, *NTRK2*, *PDGFRA*, *PAX7*, *CDH11*, *TLX1*, *FGFR3*, *CDX2*, *NCAM1*, *DPYSL3*, *DPYSL2*, *DPYSL4*, *SH3GL3*, *EPHB2*, *DNM1*, *BIN1*, *SH3GL2*, *FLT1*, *ROR2*, *SPEG*, *PRKCB*, *DCLK1*, *EPHB1*, *ZEB1*, *SRCIN1*, *MALAT1*, *MGRN1*, *NSUN2*, *FOXJ1*, *HDAC2*, *WNT1*, *SOX18*, *SP3*, *HOXA11*, *HOXC11*, *CAV1*, *PTRF*, *TGFB1*, *GDF15*, *PALLD*, *ABCB1*, *ABCC1* e *ABCG2*.

Para os 113 genes escolhidos para análise no presente estudo, foi realizada a caracterização *in silico* para identificação de alterações de ganho ou perda de metilação do DNA (CHAE et al., 2016; HOLM et al., 2016; JOHNSON et al., 2017; PARK et al., 2012; SPITZWIESER et al., 2016; TERRY et al., 2016; TIAN et al., 2016; TIRADO-MAGALLANES et al., 2017; ZHANG et al., 2017), o número de ilhas CpGs que cada gene possuía e a posição dessas CGIs em relação ao gene, conforme demonstrado na tabela 1 do apêndice.

4.2 Caracterização *in silico* dos perfis de metilação do DNA em genes relacionados a tumores mamários

Como apresentado pela tabela 1, cada um dos 113 genes analisados pode possuir uma ou mais CpGI. Para verificação de *GOM* ou *LOM* em cada uma dessas ilhas CpGs e suas respectivas *shores* foram utilizados os critérios estabelecidos em materiais e métodos (3.3). Dessa forma cada gene apresentou diferentes perfis de metilação, referentes as diferentes CGIs e *shores* presentes nos mesmos. Seguindo esse critério, dos 113 genes analisados foram encontradas 986 regiões com alteração de *GOM* ou *LOM* em relação a linhagem controle *HMEC*, sendo que a média de cDMRs encontradas por gene foi de aproximadamente 9.

Os dados demonstraram que dessas 986 regiões que apresentaram alterações nos perfis de 5mC%, 697 ocorreram em *shores* (aproximadamente 70% dos eventos), 179 em CpGI (aproximadamente 20% dos eventos) e 110 na interseção *shore/CpGI* (aproximadamente 10% dos eventos). A grande maioria das alterações encontradas nas linhagens celulares *HCC1954* e *MCF-7* (aproximadamente 85,7%) foram de ganho de metilação (*GOM*).

A posição de cada DMR em relação ao gene estudado foi analisada, no intuito de caracterizar quantas regiões eram definidas como *upstream* (a montante), *intron* e *downstream* (a jusante) e, dentro dessa divisão quantas correspondiam a CGI, *shores* e CGI-*shore*. Dessa forma, das 986 regiões analisadas foram classificadas apenas 621 que apresentaram esse critério sem sobreposição de posições por isoformas ou de sobreposição a outros genes.

Dessas 621 cDMRs, 45,2% estão localizadas *upstream* dos genes, sendo que essa porcentagem é composta por 111 *shores*, 28 CGIs, e 18 CGI-*shores*. As regiões intrônicas correspondem a 63,9% sendo composta de 306 *shores*, 75 CGIs e 40 CGI – *shores*. Já as regiões *downstream* correspondem a 31,61% (43 cDMRs) com 31 *shores*, 6 CGIs e 6 CGI – *shores*. Foram analisadas somente regiões que não apresentavam sobreposição quanto as características de posição em relação ao gene.

Os dados supracitados corroboram o estudo realizado por IRIZARRY et al., (2009b), o qual analisou amostras de tecidos de três linhagens embrionárias distintas e identificou DMRs tecido específicas (T-DMRs) em CGI e suas regiões circundantes 2Kb *upstream* e *downstream* – as *shores*. Os dados revelaram que apenas 6% das T-DMRs estão nas ilhas CpGs, enquanto que 76% se localizam nas *shores* e 18% se encontram em regiões acima de 2kb das respectivas ilhas (IRIZARRY et al., 2009b)

Ao analisar os dados WGBS de 30 linhagens celulares de câncer de mama, que representam subtipos desse tipo de tumor e que são funcionalmente distintos, CHAE et al., (2016) encontraram diferenças significativas de metilação entre CGIs e *shores*.

Já dados de GUO et al., (2016) demonstraram que a metilação aberrante nas *shores* em promotores gênicos, pode ser correlacionada com um mecanismo de inativação de genes supressores tumorais em adenocarcinomas de cárdia gástrica. A metilação do DNA em regiões que não são CGIs pode levar a inativação de genes e contribuir para o estabelecimento

de padrões de metilação específicos do câncer (ENDOHO et al., 2005; GUO et al., 2016; SATOH et al., 2002).

Os resultados encontrados no presente estudo somados àqueles descritos na literatura científica supracitada apontam as *shores* como uma nova classe de DMRs em diferentes tipos de câncer. A função regulatória dessas regiões ainda deve ser alvo de estudos futuros, uma vez que a proporção de DMRs associadas a essas regiões parece superar a presença em DMRs em CGIs (CHAE et al., 2016; IRIZARRY et al., 2009b). Essas alterações de metilação tumor específicas abrem um leque de novas regiões que podem estar associadas a etiologia de diferentes tipos de câncer.

MANJEGOWDA et al., (2017), ao analisar por ensaios de COBRA e sequenciamento por bissulfito, a metilação do DNA no gene *GPER1/GPR30*, demonstraram que as *shores* são regiões importantes na análise de DMRs dessa proteína e, que inclusive podem influenciar a regulação de *GPER1/GPR30* nos tumores de mama.

RAO et al., (2013) em um estudo de expressão e metilação do gene da *CAVEOLINA-1 (CAV1)* em câncer de mama, demonstraram que as *shores* foram as regiões que apresentaram uma diferença de metilação significativa nas análises de DMRs. Esses autores destacaram inclusive que a agressividade do tumor e as taxas de sobrevivência de pacientes podem estar relacionadas aos níveis de metilação nessas regiões. Os autores sugerem que o estudo de metilação nas *shores* em *CAV1* pode ser usado como um novo marcador prognóstico para câncer de mama do tipo basal.

Os resultados do presente estudo ratificam os dados encontrados na literatura, ao relatar que as *shores* são regiões que concentram grande variação epigenética em termos de metilação do DNA, maior ainda que aquela encontrada em CGIs.

Essa metilação nas *shores* e em outras regiões fora de CGIs reforçam a inclusão das mesmas em protocolos e investigações moleculares do câncer para estudos de caracterização de mecanismos de desenvolvimento desse grupo de doenças. Se tratando de carcinomas, os trabalhos aqui mencionados não só constataram a importância das *shores* na progressão da doença como relacionaram que os níveis de metilação nessas regiões estão associados a subtipos tumorais, agressividade da doença e sobrevivência de pacientes (no caso de câncer de mama) e que a hipermetilação nessas regiões podem estar silenciando supressores tumorais, levando ao desenvolvimento de tumores.

Considerando os resultados apresentados para os 113 genes, é importante ressaltar que essas cDMRs alteradas, e com um perfil semelhante (*GOM* ou *LOM*) nas duas linhagens de câncer de mama estudadas, podem representar novos alvos tanto para o diagnóstico quanto para o tratamento de carcinomas mamários.

4.3 Padrão de *GOM* e *LOM* nas cDMRs – associação em relação a posição nos genes e a CGIs e shores

Analisando as cDMRs utilizando os parâmetros de padrão de ganho (*GOM*) e perda de metilação (*LOM*) pode se notar uma clara prevalência de *GOM* nas cDMRs identificadas (*LOM* n = 141 sítios; *GOM* n = 845 sítios).

Quanto a distribuição desses perfis epigenéticos as cDMRs que apresentaram *LOM* são predominantemente *shores* (CGI=5,67%; *SHORE*=92,19%; CGI-*SHORE*=2,12%). As cDMRs associadas a *GOM* apresentaram uma distribuição menos desigual quanto a sobreposição a CGIs e *shores* (CGI=20,23%; *SHORE*=67,21%; CGI-*SHORES*=12,54%) (tabela 2).

Tabela 1: **Número de eventos de GOM e LOM nas cDMRs.** Número de eventos e porcentagem em relação ao total de cDMRS identificadas de eventos de *GOM* e *LOM* analisados entre as CGIs, *shores*, CGI – *shores*.

	Nº eventos	Porcentagem	<i>Shores</i>	CGI	CGI- <i>shores</i>
<i>LOM</i>	141	12.3%	130	8	3
<i>GOM</i>	845	87.7%	568	171	106

O gráfico 1 apresenta o perfil de metilação de *HMEC*, *HCC1954* e *MCF-7* nas regiões *downstream*, *upstream* e introns. O perfil de metilação encontrado nas linhagens tumorais demonstrou uma prevalência de hipermetilação das DMRs, independente de sua posição no genoma. No entanto, analisando *HMEC* vemos que ela apresenta um perfil de hipermetilação nas regiões que correspondem aos *introns* e *downstream* e diferencialmente metilada na região *upstream*.

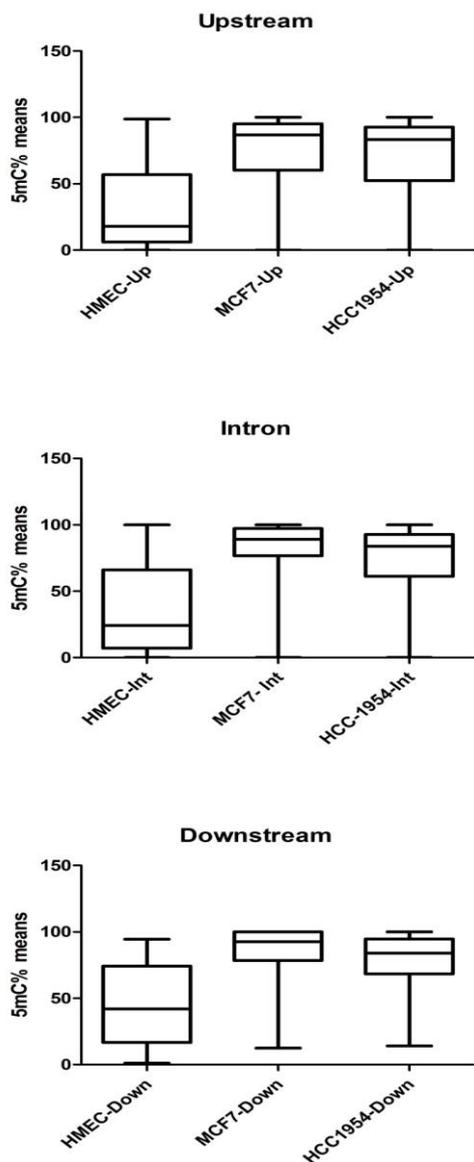


Gráfico 1: Níveis de metilação do DNA em relação à posição dos genes analisados (*upstream*, *intron*, *downstream*) entre as três linhagens investigadas. Box plots mostrando a mediana, interquartil e valores máximo e mínimo das médias de 5mC%.

Analisando o perfil de metilação das cDMRs que se encontram a mais de 2kb das ilhas CpGs analisadas, fora de *shores* e CGIs, (Gráfico 2) foi possível verificar que as linhagens tumorais também tendem a estar

hipermetiladas de maneira semelhante as demais regiões investigadas quando comparadas a linhagem controle HMEC. Os resultados da análise de variância (ANOVA – Bonferroni's post test, $\alpha = 0.01$) demonstraram diferenças estatísticas significantes quando comparados os valores de média de 5mC% das cDMRs (para regiões *upstream*, *downstream*, intron e regiões distantes, mais de 2Kb das CGIs) entre as linhagens *MCF7* e *HCC1954* em comparação a linhagem controle *HMEC*.

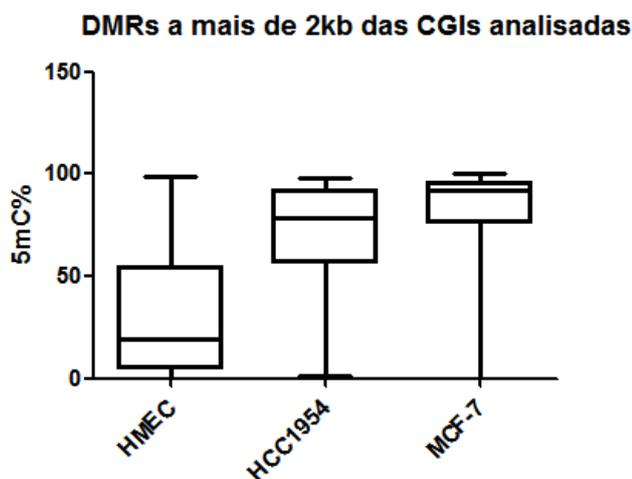


Gráfico 2: **Níveis de metilação do DNA das três linhagens investigadas em regiões a mais de 2kb das CGIs estudadas.** Box plots mostrando a mediana, interquartis e valores máximo e mínimo das médias de 5mC%.

Esses dados reafirmam os dados da literatura para a ocorrência de um perfil de hipermetilação em cDMRs em células de câncer de cólon comparados a tecidos saudáveis (IRIZARRY et al., 2009b).

Estudos realizados em outros tipos de tumores, como câncer gástrico também demonstram um perfil de hipermetilação nas linhagens tumorais (ENDOH et al., 2005). Nesse estudo, foram analisados os perfis de metilação em múltiplas regiões do gene *RASSF2*, incluindo sítios CGI e regiões associadas a sítios de ligação de fatores de transcrição em 10 linhagens celulares de câncer gástrico, bem como 78 tecidos de câncer gástrico primário e epitélio gástrico não neoplásico correspondente. Foi encontrado um perfil de hipermetilação de *RASSF2* em pelo menos uma das regiões analisadas, sendo detectado em 7 das 10 linhagens celulares estudadas. Duas linhagens celulares exibiram o perfil de hipermetilação em todas as regiões analisadas, incluindo os sítios de início de transcrição (o que resultou na perda da expressão de mRNA de *RASSF2*).

Em oposição aos perfis de hipermetilação, tem sido identificados perfis de hipometilação em *shores* em estudos com diferentes tipos de câncer de mama (CHAE et al., 2016). Em estudos onde *shores* foram analisadas em adenocarcinomas de cárdia gástrica, foram encontrados perfis de hipermetilação dessas regiões (GUO et al., 2016). Apesar de relatos frequentes de hipermetilação em CGI, relatos de hipometilação constituem um indicativo que as *shores* podem representar um nova classe de elementos regulatórios com maior heterogeneidade de alterações epigenéticas.

Quando comparados o perfil de variação dos valores de perda e ganho de metilação entre CGI e *shores* para *MCF7* e *HCC1954* ($5mC \text{ dif} = \text{média } 5mC \text{ linhagem tumoral} - \text{média } 5mC \text{ Controle}$) utilizando ANOVA, foram encontradas diferenças significativas entre *shores* e CGIs (Gráfico 3).

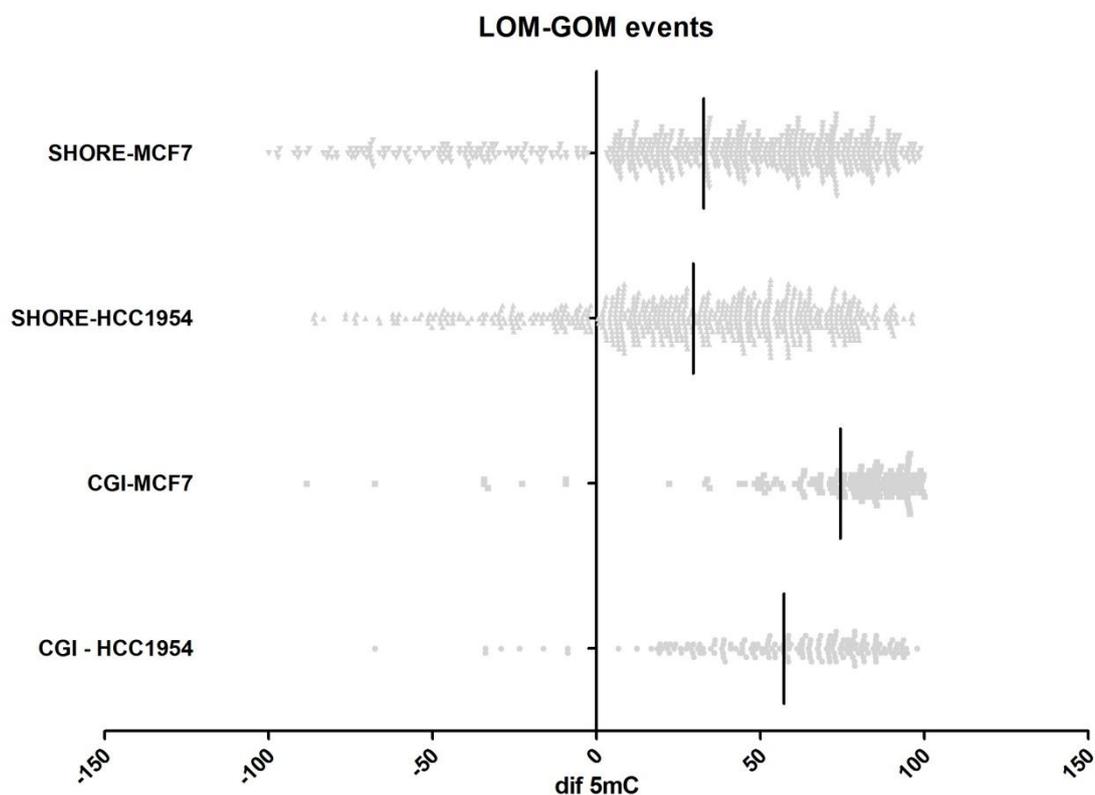


Gráfico 3: **Distribuição dos valores de dif 5mC% entre CGI e shores entre as linhagens HCC1954 e MCF7.** A figura demonstra os eventos de ganho e perda de metilação nas linhagens tumorais, pela região analisada.

O gráfico 3 demonstra que há uma predominância de eventos de *GOM* em linhagens tumorais, sendo que a maioria desses eventos ocorre em *shores*. No entanto, ao analisar o número de eventos em cada uma das regiões, nota-se que as *shores*, além de apresentarem *GOM* também apresentam maiores proporções de eventos de *LOM*. A proporção mais equilibrada de eventos

GOM/LOM reforça a hipótese de as *shores* constituem uma classe de elementos onde a heterogeneidade de alterações epigenéticas é maior.

A análise de estatística descritiva demonstra que analisando apenas as *shores*, os valores das medianas para as porcentagens de metilação do DNA em dinucleotídeos CpGs (5mC%) para *MCF7* foram de 42.77% enquanto que para *HCC1954* foi igual a 33.50%. Os valores das médias de 5mC% para *MCF7* e *HCC1954* foram 32.69 ± 1.74 e 29.57 ± 1.42 , respectivamente. Ao analisarmos os mesmos parâmetros, apenas em CGIs temos que a mediana de *MCF7* foi 82.68 e *HCC1954*= 62.80. Os valores das médias apresentados para *MCF-7*= 74.46 ± 2.18 e em *HCC1954*= 57.25 ± 2.13 . Os resultados da análise de variância (ANOVA – Bonferroni's post test, $\alpha = 0.01$) demonstraram diferenças estatísticas significantes em um número de eventos de *GOM* e *LOM* entre *shores* e CGIs em ambas as linhagens, *MCF7* e *HCC1954*.

Dessa forma, o presente estudo demonstra que as cDMRs das linhagens celulares de carcinoma mamário tendem a ganhar metilação nas linhagens tumorais. A maior frequência de eventos de *LOM* detectados nas *shores* demonstram a maior variabilidade epigenética dessas regiões e, reforçam a investigação das mesmas em estudos de câncer de mama.

4.4 Números de cDMRs com elementos evolucionariamente conservados e TFBS conservados

A variabilidade de alterações nos perfis de 5mC% nas cDMRs são um indicativo de que essas regiões (*shores* ou CGIs) possam apresentar

elementos regulatórios associados aos genes investigados. No entanto, apenas os perfis de metilação não são suficientes para garantir esse tipo de afirmação.

A presença de segmentos de DNA evolucionariamente conservados constitui um forte indicador da presença de elementos regulatórios, especialmente quando associados a presença de sítios de ligação de fatores de transcrição também conservados entre diferentes espécies (VAQUERIZAS et al., 2009; WINGENDER et al., 2015; XIE et al., 2011b).

Ao confrontarmos os dados genômicos (presença de segmentos de DNA e sítios de ligação de fatores de transcrição conservados evolucionariamente) com as coordenadas das cDMRs identificadas, através da ferramenta *Data Integrator* (disponível no UCSC), foi verificada a presença de *segmentos de DNA (lods)* e TFBS evolucionariamente conservados, obtendo-se um total de 585 regiões. Desse total, 398 regiões continham tanto *lods* quanto TFBS.

Aproximadamente 30% dessas regiões estavam localizadas em CGIs enquanto que 56% dessas estavam localizadas em *shores*. Como parâmetros foram usados o alinhamento de 100 espécies diferentes de vertebrados e medidas de conservação evolutiva utilizando dois métodos (phastCons e phyloP). Também nesse critério, as *shores* foram predominantes em porcentagem, isso reforça a hipótese dessas regiões como uma nova classe de elementos regulatórios ainda pouco explorada.

As identificação de sequências conservadas entre espécies é uma abordagem muito utilizada para busca por elementos regulatórios que possam estar envolvidos na regulação gênica local ou em genes distantes (BOFFELLI

et al., 2003; KELLIS et al., 2003; MARGULIES et al., 2003; WOOLFE et al., 2005). No entanto, segundo SIEPEL et al., (2005) uma mesma sequência pode estar vinculada a mais de um fator de transcrição (TF) podendo fornecer exemplo de conservação, devido a múltiplas funções vinculadas.

A utilização de estratégias que vinculem a identificação de cDMRs a elementos regulatórios conservados entre espécies representa uma forma de identificação de segmentos genômicos com maior probabilidade de estarem associados a etiologia de diversos tumores.

Dessa forma, o presente trabalho identificou sequências de cDMRs apresentando conservação de TFBS entre espécies, demonstrando assim que essas regiões podem apresentar um caráter regulatório associado aos genes onde estão localizadas.

4.5 Detecção de eventos de ligação de fatores de transcrição (TFs) em CGIs e *shores* entre perfis hipo e hipermetilados.

Ação dos TFs está associada ao estabelecimento de estados de condensação diferencial da cromatina, por meio da interação com sítos de ligação no DNA. Os TFs podem recrutar co-fatores e enzimas que modificam histonas e o DNA, controlando assim a condensação da cromatina e consequentemente a transcrição gênica (GAL-YAM et al., 2006; ZARET & CARROLL, 2011).

Após a investigação sobre elemento altamente conservados (aqui referidos como *lods*) e TFBS das cDMRs identificadas, foi utilizado o banco de

dados *Factorbook Motifs* para a averiguar a ocorrência real de ligação de TFs a cDMRs. As coordenadas das cDMRs foram divididas em CGI hipermetiladas e hipometiladas (CGI-hiper e hipo) e *shores* hipermetiladas e hipometilada (*shore*-hiper e hipo). Os resultados da ocorrência de ligação de TFs encontram-se representados nos gráficos 4 e 5.

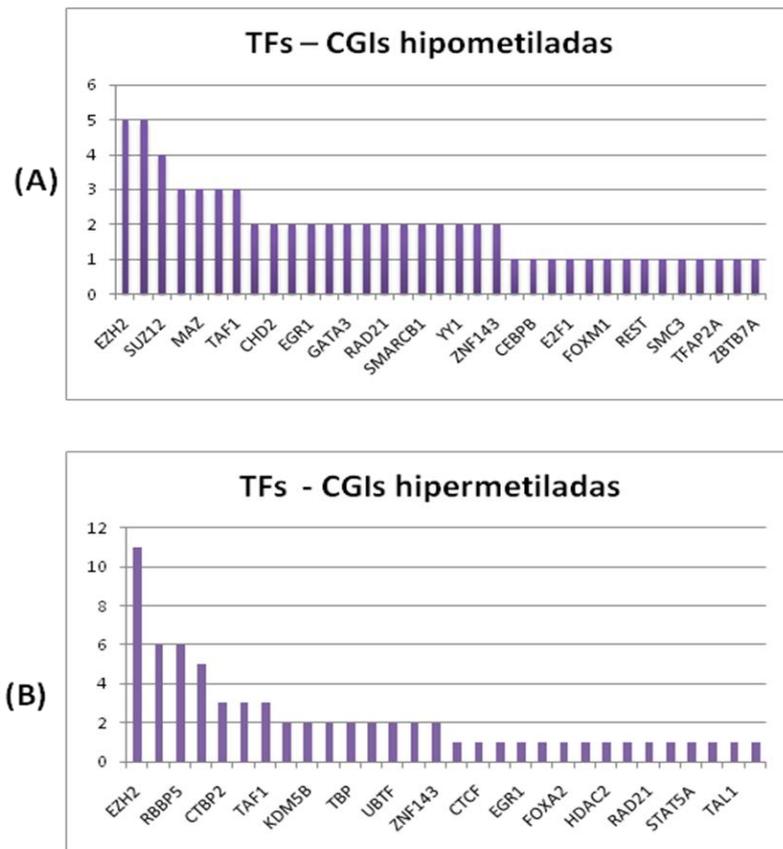


Gráfico 4: **Fatores de transcrição (TFs) que se ligam as CGIs tanto hipometiladas quanto hipermetiladas.** (A) Fatores de Transcrição (TFs) presentes nas regiões das DMRs CGI hipometiladas. (B) Fatores de Transcrição (TFs) presentes nas regiões das DMRs CGI hipermetiladas

Nas CGIs hipometiladas, os fatores de transcrição mais abundantes foram: EZH2, POLR2A, SUZ12, FOXA1, MAZ, MYC, TAF1, CHD1, CHD2 e CTCF. Já as CGIs hipermetiladas apresentaram a prevalência dos seguintes TFs: EZH2, CHD1, RBBP5, SUZ12, CTBP2, POLR2A, TAF1, GATA2, KDM5B e SP1. Vale ressaltar que EZH2 e SUZ12 são componentes de PcG-2.

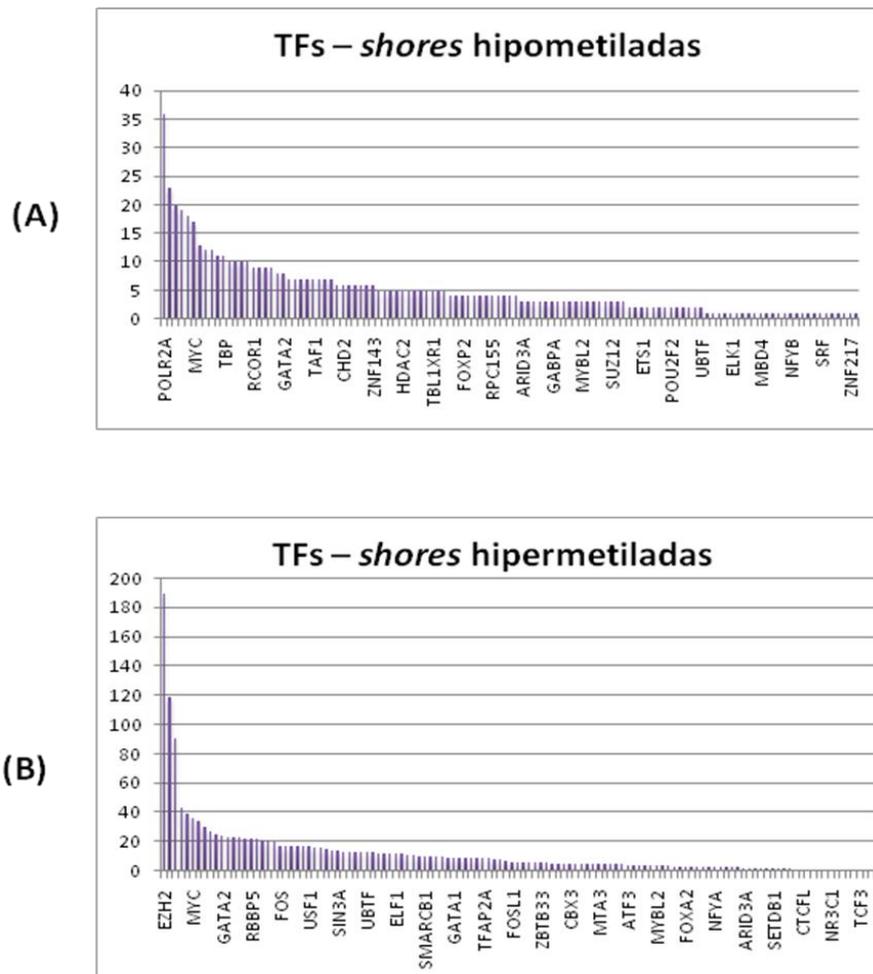


Gráfico 5: Fatores de transcrição (TFs) que se ligam as shores tanto hipometiladas quanto hipermetiladas. (A) Fatores de Transcrição (TFs) presentes

nas regiões das DMRs shores hipometiladas. (B) Fatores de Transcrição (TFs) presentes nas regiões das DMRs shores hipermetiladas.

Os fatores de transcrição mais abundantes nas *shores* hipometiladas foram: POLR2A, MAX, CTCF, YY1, EP300, MYC, JUND, MAS, SP1 e EZH2, enquanto que nas *shores* hipermetiladas encontramos EZH2, POLR2A, SUZ12, CTCF, CTBP2, MYC, MAX, CHD1, EGR1 e TAF1.

Consirando os quatro tipos de cDMRs estudadas, fica clara a ligação diferencial de fatores de transcrição a cada uma delas. Essa ligação diferencial de TFs sugere que os mecanismos que levam a eventos de *GOM* e *LOM* podem ser distintitos, inclusive entre CGIs e *shores*.

Dentre os TFs encontrados há componentes do PcG2 (EZH2 e SUZ12). Esses fatores estão associados a formação de heterocromatina e indiretamente a metilação do DNA (TIWARI et al., 2008). Também foram encontrados relatos de ligação de proteínas CTCF e YY1. O YY1 é um fator de transcrição cuja função pode estar associada a perda ou ganho de metilação do DNA dependendo da região alvejada e, possivelmente, da ação conjunta com outros fatores (KIM et al., 2009).

O fator CTCF é uma proteína insuladora que também apresenta função repressora, sua não ligação está associada ao ganho de metilação em DMRs de genes regulados por *imprinting genômico* e a consequente desregulação gênica local (KIM et al., 2009)

Os promotores associados a CGIs podem ser reprimidos por vários mecanismos, inclusive pela ação de proteínas *polycomb*. Estudos de análise da

metilação em células tumorais demonstraram que a hipermetilação do DNA juntamente com a ação do complexo *polycomb* podem reprogramar o epigenoma de células tumorais, reduzindo a sua plasticidade regulatória na expressão de genes (GAL-YAM et al., 2008; OHM et al., 2007; SCHLESINGER et al., 2007; WIDSCHWENDTER et al., 2007).

DOZMOROV (2015) ao tentar correlacionar a expressão de genes associadas com o envelhecimento (aGEN – *age associated genes*), assim como a metilação (aDMRs – *age related DMRs*) encontrou assinaturas comuns entre essas regiões, que incluem a ligação em EZH2, SUZ12 e CTCF a ocorrência simultânea de marcas de repressão (H3K27me3) quanto de ativação (H3K4me1), características de domínios bivalentes (BERNSTEIN et al., 2006)

Os domínios bivalentes geralmente tendem a assumir um padrão de marcas epigenéticas de ativação ou repressão, devido a estímulos durante a diferenciação (BERNSTEIN et al., 2006). Apesar do *status* de bivalência não ter sido investigado no presente estudo, a possível existência do mesmo, devido a presença de componentes do complexo PcG-2, poderia ser um fator associado a ocorrência de alterações epigenéticas nas cDMRs.

Já o estudo de SIEBOLD et al., (2010), ratifica essas possibilidades, ao sugerir que o PcG-2 desempenha um papel direto na modulação da longevidade e resistência ao estresse em *Drosophilas*. Um dado interessante é que os resultados encontrados por Dozmorov também incluem uma associação de PcG-2 com aDMRs hipermetiladas, sugerindo que a regulação realizada por

PcG-2 pode estar fortemente relacionada ao aumento da metilação com a idade (aDMRs).

Outros autores já demonstraram que a hipermetilação de alguns sítios e a assinatura de PcG-2 estão correlacionadas (BEERMAN et al., 2013; JUNG & PFEIFER, 2015) e que alternativamente, PcG-2 pode recrutar histona metiltransferases e subsequentemente DNA metiltransferases, o que resultaria na hipermetilação da região (CEDAR & BERGMAN, 2012; VIRÉ et al., 2006).

Apesar de o foco desse estudo ser as cDMRs, é de grande relevância destacar resultados semelhantes, mesmo que estejam direcionados a outras perguntas. Fatores de transcrição relacionados a PcG-2, parecem exercer um papel funcional nas alterações epigenéticas encontradas nas cDMRs, devendo o mesmo ser melhor explorado.

Outras fontes significativas demonstraram ainda uma forte tendência de hipermetilação do DNA em regiões que estão próximas a TFs, sugerindo que haja uma perda na expressão de alguns genes, principalmente supressores tumorais, levando ao desenvolvimento de tumores (ENDOHO et al., 2005; GUO et al., 2016; JOHNSON et al., 2017).

Um estudo realizado por JIN et al., (2016) investigou os efeitos da metilação nas interações DNA/proteínas, escolhendo três fatores de transcrição humanos. Os resultados do trabalho demonstraram que a metilação do DNA (principalmente em regiões que não são CGIs) podem fornecer um mecanismo para regular a expressão de genes, através da influência direta de fatores de transcrição.

No entanto, a regulação de genes é um evento altamente complexo, podendo ser regulada por pelo menos três aspectos da metilação do DNA: (1) Os fatores de transcrição podem ser bloqueados devido as interações da metilação do DNA nas regiões promotoras; (2) Quando os TFs conseguem se ligar ao seu alvo, a estabilidade da ligação pode ser modificada diretamente por regiões metiladas do DNA; e (3) Logo após a transcrição de um gene ser iniciada, o alongamento pode ser bloqueado por algumas proteínas que estão associadas ao processo de metilação no corpo gênico.

Dessa forma, sugere-se que as modificações epigenéticas existentes nas DMRs estudadas podem estar associadas aos TFs, principalmente aos do PcG-2, que por sua vez pode está correlacionada com o perfil hipermetilado das linhagens celulares de câncer de mama encontradas no presente trabalho.

5. CONCLUSÃO

- ❖ As linhagens tumorais tendem a ganhar metilação, se tornando hipermetiladas, como demonstrado na literatura. Dessa forma, sugerimos que a metilação do DNA pode ser usada como um epimarcador em tumores mamários;
- ❖ Apesar de muitos estudos focarem nas CGIs, os dados apresentados demonstram que a maior variabilidade de epimutações encontra-se nas *shores* e que essas regiões podem constituir uma nova classe de elementos regulatórios em câncer;
- ❖ O cruzamento de dados genômicos com as DMRs identificadas mostrou a presença de *sequências* e TFBS *evolucionariamente conservados*, tanto em *shores* quanto em CGIs. No entanto, a grande predominância desses elementos encontra-se nas *shores* (56%), o que ratifica mais uma vez a importância dessas regiões no estudo de cDMRs;
- ❖ Os diversos fatores de transcrição associados às regiões genômicas, são um indicativo de que uma assinatura molecular pode estar ligada a natureza das alterações epigenéticas em cada DMR e que elementos do complexo repressor *polycomb* II (PcG-2) podem estar relacionados a hipermetilação nessas regiões.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ADUSUMALLI, S.; MOHD, F.; MOHD, O.; SOONG, R.; BENOUKRAF, T. Methodological aspects of whole-genome bisulfite sequencing analysis. **Briefings in Bioinformatics**, v. 16, p. 369–379, 2014.
- AHSENDORF, T.; MÜLLER, F.J.; TOPKAR, V.; GUNAWARDENA, V.; EILS, R. Transcription factors, coregulators, and epigenetic marks are linearly correlated and highly redundant. **Plos One**, v. 12, p. 1–25, 2017.
- ANDREOLI, F.; DEL RIO, A. Physicochemical modifications of histones and their impact on epigenomics. **Drug Discovery Today**, p. 1–8, 2014.
- ARAI, E.; KANAI, Y.; USHIJIMA, S.; FUJIMOTO, H.; MUKAI, K.; HORORASHI, S. Regional DNA hypermethylation and DNA methyltransferase (DNMT) 1 protein overexpression in both renal tumors and corresponding nontumorous renal tissues. **International Journal of Cancer**, v. 119, p. 288–296, 2006.
- BAHAR HALPERN, K.; VANA, T.; WALKER, M. D. Paradoxical role of DNA methylation in activation of FoxA2 gene expression during endoderm development. **Journal of Biological Chemistry**, v. 289, p. 23882–23892, 2014.
- BAKUSIC, J.; SCHAUFELI, W.; CLAES, S.; GODDERIS, L. Stress, burnout and depression: A systematic review on DNA methylation mechanisms. **Journal of Psychosomatic Research**, p. 1–41, 2016.
- BANOVICH, N. E.; LAN, X.; MCVICKER, G.; VAN DE GEIJN, B.; DEGNER, F. J.; BLISCHAK, J.D.; ROUX, J. PRITCHARD, J.K.; GILAD, Y. Methylation QTLs Are Associated with Coordinated Changes in Transcription Factor Binding, Histone Modifications, and Gene Expression Levels. **Plos Genetics**, v. 10, p. 1–12, 2014.
- BARSKI, A.; CUDDAPAH, S.; CUI, K., ROH, T-Y.; SCHONES, D.E.; WANG, Z.; WEI, G; CHEPELEV, I.; ZHAO, K. High-Resolution Profiling of Histone Methylations in the Human Genome. **Cell**, v. 129, p. 823–837, 2007.

- BAUER, M.; FINK, B.; THÜRMAN, L.; ESZLINGER, M.; HERBERTH, G., LEHMANN, I Tobacco smoking differently influences cell types of the innate and adaptive immune system—indications from CpG site methylation. **Clinical Epigenetics**, v. 8, p. 1–12, 2016.
- BAYLIN, S. B.; JONES, P. A. A decade of exploring the cancer epigenome — biological and translational implications. **Nature Reviews Cancer**, v. 11, n. 10, p. 726–734, 2011.
- BEERMAN, I; BOCK, C.; GARRISON, B.S.; SMITH, Z.D.; GU, H.; MEISSNER, A.; ROSSI, D. Proliferation-dependent alterations of the DNA methylation landscape underlie hematopoietic stem cell aging. **Cell Stem Cell**, v. 12, p. 413–425, 2013.
- BELL, J. T.; PAI, A. A.; PICKRELL, J. K.; GAFFNEY, D. J.; PIQUE-REGI, R.; DEGNER, G. F.; GILAD, Y.; PRITCHARD, J. K. DNA methylation patterns associate with genetic and gene expression variation in HapMap cell lines. **Genome Biology**, v. 12, p. 1–13, 2011.
- BERNSTEIN, B. E.; MIKKELSEN, T. S.; XIE, X.; KAMAL, M.; HUEBERT, D. J.; CUFF, J.; FRY, B.; MEISSNER, A.; WERNIG, M.; PLATH, K.; JAENISCH, R.; WAGSCHAL, A.; FEIL, R.; SCHEREIBER, S. L.; LANDER, E. S. A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. **Cell**, v. 125, p. 315–326, 2006.
- BOFFELLI, D.; MCAULIFFE, J.; OVCHARENKO, D.; LEWIS, K. D.; OVCHARENKO, I.; PATCHER, L.; RUBIN, E. M. Phylogenetic shadowing of primate sequences to find functional regions of the human genome. **Science**, v. 299, p. 1391–1394, 2003.
- BREILING, A.; LYKO, F. Epigenetic regulatory functions of DNA modifications: 5-methylcytosine and beyond. **Epigenetics and Chromatin**, v. 8, p. 1–9, 2015.
- BRENET, F.; MOH, M.; FUNK, P.; FEIERSTEIN, E.; VIALE, A. J.; SOCCI, N. D.; SCANDURA, J. M. DNA methylation of the first exon is tightly linked to transcriptional silencing. **Plos One**, v. 6, p. 1–12, 2011.

- BRENNER, A. J.; PALADUGU, A.; WANG, H.; OLOPADE, O. I.; DREYLING, M. H.; ALDAZ, C. M. Preferential loss of expression of p16(INK4a) rather than p19(ARF) in breast cancer. **Clinical Cancer Research**. v. 2, p. 1993–1998, 1996.
- BROOKES, E.; SHI, Y. Diverse epigenetic mechanisms of human disease. **Annual Review of Genetic**, v. 15, p. 237-268, 2014
- CEDAR, H.; BERGMAN, Y. Programming of DNA methylation patterns. **Annual Review of Biochemistry**, v. 81, p. 97–117, 2012.
- CHAE, H.; LEE, S.; NEPHEW, K. P.; KIM, S. Subtype-specific CpG island shore methylation and mutation patterns in 30 breast cancer cell lines. **BioMed Central Systems Biology**, v. 10, p. 434–548, 2016.
- CHODAVARAPU, R. A.; FENG, S.; BERNATAVICHUTE, Y. V.; CHEN, P-Y.; STROUD, H.; YU, Y.; HETZEL, J.; KUO, F.; KIM, J.; COKUS, S. J.; CASERO, D.; BERNAL, M.; HUIJSER, P.; CLARK, A. T.; KRÄMER, U.; MERCHANT, S. S.; ZHANG, X.; JACOBSEN, S. E.; PELLEGRINI, M. Relationship between nucleosome positioning and DNA methylation. **Nature**, v. 466, p. 388–392, 2010.
- CLARK, G. J.; DER, C. J. Aberrant function of the Ras signal transduction pathway in human breast cancer. **Breast Cancer Research Treatment**. v.35, p. 133–144, 1995.
- CORE, L. J.; WATERFALL, J. J.; LIS, J. T. Nascent RNA Sequencing Reveals Widespread Pausing and Divergent Initiation at Human Promoters. **Science**, v. 322, p. 1845–1848, 2008.
- DEY, S. Preventing breast cancer in LMICs via screening and/or early detection: The real and the surreal. **World Journal of Clinical Oncology**, v. 5, p. 509–519, 2014.
- DOZMOROV, M. G. Polycomb repressive complex 2 epigenomic signature defines age-associated hypermethylation and gene expression changes. **Epigenetics**, v. 10, p. 484–495, 2015.

- DWORKIN, A. M.; HUANG, T. H.-M.; TOLAND, A. E. Epigenetic alterations in the breast: Implications for breast cancer detection, prognosis and treatment. **Seminars in cancer biology**, v. 19, p. 165–171, 2009.
- ELENBAAS, B.; SPIRIO, S.; KOERNER, F.; FLEMING, M. D.; ZIMONJIC, D. B.; DONAHER, J. L.; POPESCU, N. C.; HAHN, W. C.; WEINBER, R. A. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. **Genes & development**, v. 15, p. 50–65, 2000.
- ENDO, M.; TAMURA, G.; HOMMA, N.; TERASHIMA, M.; NISHIZUKA, S.; MOTOYAMA, T. RASSF2, a potential tumour suppressor, is silenced by CpG island hypermethylation in gastric cancer. **British Journal of Cancer**, v. 93, p. 1395–1399, 2005.
- ESCOT, C.; THEILLE, C.; LINDEREAU, R.; SPYRATOS, F.; CHAMPEME, M.H.; GEST, J.; CALLAHAN, R. Genetic alteration of the c-myc protooncogene (MYC) in human primary breast carcinomas. **Proceedings of the National Academy of Sciences**. v. 83, p. 4834–4838, 1986
- FEINBERG, A. P.; IRIZARRY, R. A. Stochastic epigenetic variation as a driving force of development, evolutionary adaptation, and disease. **Proceedings of the National Academy of Sciences**, v. 107, p. 1757–1764, 2010.
- FEJES-TOTH, K.; SOTIROVA, V.; SACHIDANANDAM, R.; ASSAF, G.; HANNON, G. J.; KAPRANOV, P.; FOISSAC, S.; WILLINGHAM, A. T.; DUTAGGUPTA, R.; DUMAIS, E.; GINGERAS, T. T. Post-transcriptional processing generates a diversity of 5'- modified long and short RNAs. **Nature**, v. 457, p. 1028–1032, 2009.
- FELIX, JD., CASTRO, DS., AMORIM, MHC., ZANDONADE, E. Tendência de mortalidade por câncer de mama em mulheres no estado do Espírito Santo, no período de 1980 a 2007. **Revista Brasileira de Cancerologia**. v. 57, p. 159-166, 2011.
- FENG, S.; COKUS, S. J.; ZHANG, X.; CHEN, P-Y.; BOSTICK, M.; GOLL, M.

G.; HETZEL, J.; JAIN, J.; STRAUSS, S. H.; HALPERN, M. E.; UKOMADU, C.; SADLER, K. C.; PELLEGRINI, M.; JACOBSEN, S. E Conservation and divergence of methylation patterning in plants and animals. **Proceedings of the National Academy of Sciences**, v. 107, p. 8689–8694, 2010.

FERRARI, K. J.; SCELFO, A.; JAMMULA, S.; CUOMO, A.; BAROZZI, I.; STÜTZER, F.; BONALDI, T.; PASINI, D. Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. **Molecular Cell**, v. 53, p. 49–62, 2014.

FUJIMOTO, M.; KITAZAWA, R.; MAEDA, S.; KITAZAWA, S. Methylation adjacent to negatively regulating AP-1 site reactivates TrkA gene expression during cancer progression. **Oncogene**, v. 24, p. 5108–5118, 2005.

GAL-YAM, E. N.; JEONG, S.; TANAY, A.; EGGER, G.; LEE, A. S.; JONES, P. A. Constitutive nucleosome depletion and ordered factor assembly at the GRP78 promoter revealed by single molecule footprinting. **Plos Genetics**, v. 2, p. 1451–1463, 2006.

GAL-YAM, E. N.; EGGER, G.; INIGUEZ, L.; HOLSTER, H.; EINARSSON, S.; ZHANG, X.; LIN, J. C.; LIANG, G.; JONES, P. A.; TANAY, A. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. **Proceedings of the National Academy of Sciences of the United States of America**, v. 105, p. 12979–12984, 2008.

GIBBS, J. R.; VAN DER BRUG, M. P.; HERNANDEZ, D. G., TRAYNOR, B. J.; NALLS, M. A.; LAI, S-L.; AREPALLI, S.; DILMAN, A.; RAFFERTY, I. P.; TRONCOSO, J.; JOHNSON, R.; ZIELKE, H. R.; FERRUCCI, L.; LONGO, D. L.; COOKSON, M. R.; SINGLETON, A. B. Abundant quantitative trait loci exist for DNA methylation and gene expression in Human Brain. **Plos Genetics**, v. 6, p. 1–13, 2010.

GIBNEY, E. R.; NOLAN, C. M. Epigenetics and gene expression. **Heredity**, v. 105, p. 4–13, 2010.

- GILLETT, C.; FANTL, V.; SMITH, R.; FISHER, C.; BARTEK, J.; DICKSON, C.; BARNES, D.; PETERS, G. Amplification and overexpression of cyclin D1 in breast cancer detected by immunohistochemical staining. **Cancer Research**. v. 54, p.1812–1817, 1994
- GOPALAKRISHNAN, S.; EMBURGH, B. O. VAN; ROBERTSON, K. D. DNA methylation in development and human disease. **Mutation Research - Genetic Toxicology and Environmental Mutagenesis**, v. 647, p. 1–17, 2008.
- GRIMMER, M. R.; STOLZENBURG, S.; FORD, E.; LISTER, R.; BLANCAFORT, P.; FARNHAM, P. J. Analysis of an artificial zinc finger epigenetic modulator: Widespread binding but limited regulation. **Nucleic Acids Research**, v. 42, p. 10856–10868, 2014.
- GU, X.; BOLDRUP, L.; COATES, P. J.; FAHRAEUS, R.; NYLANDER, E.; LOIZOU, C.; OLOFSSON, K.; NORBERG - SPAAK, L.; GÄRSKOG, O.; NYLANDER, K. Epigenetic regulation of OAS2 shows disease-specific DNA methylation profiles at individual CpG sites. **Scientific Reports**, v. 6, p. 1–8, 2016.
- GUO, Y-L.; ZHU, T-N.; GUO, W.; DONG, Z-M.; ZHOU, Z.; CUI, Y-J.; ZHAO, R-J. Aberrant CpG Island Shore Region Methylation of CAV1 Is Associated with Tumor Progression and Poor Prognosis in Gastric Cardia Adenocarcinoma. **Archives of Medical Research**, v. 47, p. 460–470, 2016.
- GUTIERREZ-ARCELUS, M.; LAPPALAINEN, T.; MONTGOMERY, S. B.; BUIL, A.; ONGEN, H.; YUROVSKY, A.; BRYOIS, J.; GIGER, T.; ROMANO, L.; PLANCHON, A.; FALCONNET, E.; BIELSER, D.; GAGNEBIN, M.; PADIOLEAU, I.; BOREL, C.; LETOURNEAU, A.; MAKRYTHANASIS, P.; GUIPPONI, M.; GEHRIG, C.; ANTONARAKIS, S. E.; DERMITZAKIS, E. T. Passive and active DNA methylation and the interplay with genetic variation in gene regulation. **eLife**, p. 1–18, 2013.
- HAMILTON, J. P. Epigenetics: Principles and practice. **Genetic Basis of**

Digestive Diseases, v. 29, p. 130–135, 2011.

HANNA, C. W.; KELSEY, G. Genomic imprinting beyond DNA methylation: A role for maternal histones. **Genome Biology**, v. 18, p. 1–4, 2017.

HANSMANN, T.; PLIUSHCH, G.; LEUBNER, M.; KROLL, P.; ENDT, D.; GEHRIG, A.; PREISLER - ADAMS, S.; WIEACKER, P.; HAAF, T. Constitutive promoter methylation of BRCA1 and RAD51C in patients with familial ovarian cancer and early-onset sporadic breast cancer. **Human Molecular Genetics**, v. 21, p. 4669–4679, 2012.

HARR, J. C.; GONZALEZ-SANDOVAL, A.; GASSER, S. M. Histones and histone modifications in perinuclear chromatin anchoring: from yeast to man. **EMBO reports**, v. 17, p. 139–155, 2016.

HE, J.; SUN, X.; SHAO, X.; LIANG, L.; XIE, H. DMEAS: DNA methylation entropy analysis software. **Bioinformatics**, v. 29, p. 2044–2045, 2013.

HEYN, H.; VIDAL, E.; FERREIRA, H. J.; VIZOSO, M.; SAYOLS, S.; GOMEZ, A.; MORAN, S.; BOQUE - SATRE, R.; GUIL, S.; MARTINEZ - CARDUS, A.; LIN, C. Y.; ROYO, R.; SANCHEZ - MUT, J. V.; MARTINEZ, R.; GUT, M.; TORRENTS, D.; OROZCO, M.; GUT, I.; YOUNG, R. A.; ESTELLER, M. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. **Genome Biology**, v. 17, p. 1–16, 2016.

HINDORFF, L. A.; SETHUPATHY, P.; JUNKINS, H. A.; RAMOS, E. M.; MEHTA, J. P.; COLLINS, F. S.; MANOLIO, T. A. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. **Proceedings of the National Academy of Sciences**, v. 106, p. 9362–9367, 2009.

HOLM, K.; STAAF, J.; LAUSS, M.; AINE, M.; LINDGREN, D.; BENDAHL, P-O.; VALLON - CHRISTERSSON, J.; BARKARDOTTIR, R. B.; HÖUGLUND, M.; BORG, A.; JÖNSSON, J.; RINGNÉR, M. An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. **Breast Cancer Research**, v. 18, p. 1–20, 2016.

- HON, G. C.; HAWKINS, R. D.; CABALLERO, O. L.; LO, C.; LISTER, L.; PELIZZOLA, M.; VALSESIA, A.; YE, Z.; KUAN, S.; EDSALL, L. E.; CAMARGO, A. A.; STEVENSON, B. J.; ECKER, J. R.; BAFNA, V.; STRAUSBERG, R. L.; SIMPSON, A. J.; REN, B. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. **Genome Research**, v. 22, p. 246–258, 2012.
- HU, S.; WAN, J.; SU, Y.; SONG, Q.; ZENG, Y.; NGUYEN, H. N.; SHIN, J.; COX, E.; RHO, H. S.; WOODARD, C.; XIA, S.; LIU, S.; LYU, H.; MING, G-L.; WADE, H.; SONG, H.; QIAN, J., ZHU, H. DNA methylation presents distinct binding sites for human transcription factors. **eLife**, v. 2, p. 1–16, 2013.
- IRIZARRY, R. A.; LADD-ACOSTA, C.; WEN, B.; WU, Z.; MONTANO, C.; ONYANGO, P.; CUI, H.; GABO, K.; RONGIONE, M.; WEBSTER, M.; JI, H.; POTASH, J. SABUNCIYAN, S.; FEINBERG, A. P. Genome-wide methylation analysis of human colon cancer reveals similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. **Nature Genetics**, v. 41, p. 1–25, 2009a.
- IRIZARRY, R. A.; LADD-ACOSTA, C.; WEN, B.; WU, Z.; MONTANO, C.; ONYANGO, P.; CUI, H.; GABO, K.; RONGIONE, M.; WEBSTER, M.; JI, H.; POTASH, J. B.; SABUNCIYAN, S.; FEINBERG, A. P. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. **Nature Genetics**, v. 41, p. 178–186, 2009b.
- JANOWSKI, B.; KUFFMAN, K. E.; SCHWARTZ, J. C.; RAM, R.; NORDSELL, R.; SHAMES, D. S.; MINNA, J. D.; COREY, D. R. Involvement of AGO1 and AGO2 in mammalian transcriptional silencing. **Nature Structural & Molecular Biology**, v. 13, p. 787–792, 2006.
- JEMAL, A.; BRAY, F.; CENTER, M. M.; FERLAY, J.; WARD, E.; FORMAN, D. Global Cancer Statistics: 2011. **A Cancer Journal for Clinicians**, v. 61, p. 69–90, 2011.

- JIMENEZ-USECHE, I.; KE, J.; TIAN, Y.; SHIM, D.; HOWELL, S. C.; QIU, X.; YUAN, C. DNA methylation regulated nucleosome dynamics. **Scientific Reports**, v. 3, p. 1–5, 2013.
- JIN, J.; LIAN, T.; GU, C.; YU, K.; GAO, Y. Q.; SU, X-D. The effects of cytosine methylation on general transcription factors. **Scientific Reports**, v. 6, p. 1–13, 2016.
- JOHNSON, K. C.; HOUSEMAN, E. A.; KING, J. E.; CHRISTENSEN, B. C. Normal breast tissue DNA methylation differences at regulatory elements are associated with the cancer risk factor age. **Breast Cancer Research**, v. 19, p. 1–11, 2017.
- JONES, P. A. The DNA methylation paradox. **Trends in Genetics**, v. 15, p. 34–37, 1999.
- JONES, P. A. Functions of DNA methylation: Islands, start sites, gene bodies and beyond. **Nature Reviews Genetics**, v. 13, p. 484–492, 2012.
- JOVANOVIC, J.; RONNEBERG, J. A.; TOST, J.; KRISTENSEN, V. The epigenetics of breast cancer. **Molecular Oncology**, v. 4, p. 242–254, 2010.
- JÜHLING, F.; KRETZMER, H.; BERNHART, S. H.; OTTO, C.; STADLER, P. F.; HOFFMANN, S. Metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data. **Genome Research**, v. 26, p. 256–262, 2016.
- JUNG, M.; PFEIFER, G. P. Aging and DNA methylation. **BMC biology**, v. 13, p. 1–8, 2015.
- KAPRANOV, P.; CHENG, J.; DIKE, S.; NIX, D. A.; DUTTAGUPTA, R.; WILLINGHAM, A. T.; STADLER, P. F.; HERTEL, J.; HACKERMÜLLER, J.; HOFACKER, I. L.; BELL, I.; CHEUNG, E.; DRENKOW, J.; DUMAIS, E.; PATEL, S.; HELT, G.; GANESH, M.; GHOSH, S.; PICCOLBONI, A.; SEMENTCHENKO, V.; TAMMANA, H.; GINGERAS, T. R. RNA maps reveal new RNA classes and a possible function for pervasive transcription. **Science**, v. 316, p. 1484–1488, 2007.

- KELLIS, M.; PATTERSON, N.; ENDRIZZI, M.; BIRREN, B.; LANDER, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. **Nature**, v. 423, p. 241–254, 2003.
- KELLY, T. K.; LIU, Y.; LAY, F. D.; LIANG, G.; BERMAN, B. P.; JONES, P. A. Genome-wide mapping of nucleosome positioning and DNA methylation within individual DNA molecules. **Genome Research**, v. 22, p. 2497–2506, 2012.
- KHAKPOUR, G.; NORUZINIA, M.; IZADI, P.; KARAMI, F.; AHMADVAND, M.; HESHMAT, R.; AMOLI, M. M.; TAVAKKOLY - BAZZAZ, J. Methyloomics of breast cancer: Seeking epimarkers in peripheral blood of young subjects. **Tumor Biology**, p. 1–11, 2017.
- KIM, D.; VILLENEUVE, L. M.; MORRIS, K. V.; ROSSI, J. J Argonaute-1 directs siRNA-mediated transcriptional gene silencing in human cells. **Nature Structural & Molecular Biology**, v. 13, p. 793–797, 2006.
- KIM, J. DO; KANG, K.; KIM, J. YY1's role in DNA methylation of Peg3 and Xist. **Nucleic Acids Research**, v. 37, p. 5656–5664, 2009.
- KINDT, A. S. D.; FUERST, R. W.; KNOOP, J.; LAIMIGHOFER, M.; TELIEPS, T.; KIPPICH, M.; WOERHEIDE, M. A.; WAHL, S.; WILSON, R.; SEDLMEIER, E-M.; HOMMEL, A.; TODD, J. A.; KRUMISIEK, J.; ZIEGLER, A - G.; BONIFACIO, E. Allele-specific methylation of type 1 diabetes susceptibility genes. **Journal of Autoimmunity**, p. 1–12, 2017.
- KOFINK, D.; BOKIS, M. P. M.; TIMMERS, H. T. M.; KAS, M. J. Epigenetic dynamics in psychiatric disorders: Environmental programming of neurodevelopmental processes. **Neuroscience and Biobehavioral Reviews**, v. 37, p. 831–845, 2013.
- KOHONEN-CORISH, M. R. J.; SIGGLEKOW, N. D.; SUSANTO, J.; CHAPUIS, P. H.; BOKEY, E. L.; DENT, O. F.; CHAN, C.; LIN, B. P. C.; SENG, T. J.; LAIRD, P. W.; YOUNG, J.; LEGGETT, B. A.; JASS, J. R.; SHUTHERLAND, R. L. Promoter methylation of the mutated in colorectal cancer gene is a frequent early event in colorectal cancer. **Oncogene**, v.

26, p. 4435–4441, 2007.

KOUZARIDES, T. Chromatin modifications and their function. **Cell**. v. 128, p. 693–705, 2007.

KULIS, M.; ESTELLER, M. DNA methylation and cancer. **Advances in Genetics**, v. 70, p. 27–56, 2010.

LAI, A. Y.; FATEMI, M.; DHASARATHY, A.; MALONE, C.; SOBOL, S. E.; GEIGERMAN, C.; JAYE, D. L.; MAV, D.; SHAH, R.; LI, L.; WADE, P. A. DNA methylation prevents CTCF-mediated silencing of the oncogene BCL6 in B cell lymphomas. **The Journal of Experimental Medicine**, v. 207, p. 1939–1950, 2010.

LEIGHT, J. L.; DRAIN, A.P.; WEAVER, V. M. Extracellular matrix remodeling and stiffening modulate tumor phenotype and treatment response. **Annual Review of Cancer Biology**, v 1, p. 313-334, 2017.

LIN, Q.; WEIDNER, C. I.; COSTA, I. G.; MARIONI, R. E.; FERREIRA, M. R. P.; DEARY, I. J.; WAGNER, W. DNA methylation levels at individual age-associated CpG sites can be indicative for life expectancy. **Aging**, v. 8, p. 394–401, 2016.

MANJEGOWDA, M. C.; GUPTA, P. S.; LIMAYE, A. M. Hyper-methylation of the upstream CpG island shore is a likely mechanism of GPER1 silencing in breast cancer cells. **Gene**, p. 1–37, 2017.

MANTSOKI, A.; DEVAILLY, G.; JOSHI, A. CpG island erosion, polycomb occupancy and sequence motif enrichment at bivalent promoters in mammalian embryonic stem cells. **Scientific Reports**, v. 5, p. 1–15, 2015.

MARGULIES, E. H.; BLANCHETTE, M.; NISC, C. S. P.; HAUSSLER, D.; GREEN, E. D. Identification and characterization of multi-species conserved sequences. **Genome Research**, v. 13, p. 2507–2518, 2003.

MATTICK, J. S.; MAKUNIN, I. V. Non-coding RNA. **Human molecular genetics**, v. 15, p. 17–29, 2006.

- MAUPETIT-MÉHOUS, S.; MONTIBUS, B.; NURY, D.; TAYAMA, C.; WASSEF, M.; KOTA, S. K.; FOGLI, A.; CAMPOS, F. C.; HATA, K.; FEIL, R.; MARGUERON, R.; NAKABAYASHI, K.; COURT, F.; ARNAUD, P. Imprinting control regions (ICRs) are marked by mono-allelic bivalent chromatin when transcriptionally inactive. **Nucleic Acids Research**, v. 44, p. 621–635, 2016.
- MOAREFI, A. H.; CHÉDIN, F. ICF syndrome mutations cause a broad spectrum of biochemical defects in DNMT3B-mediated de novo DNA methylation. **Journal of Molecular Biology**, v. 409, p. 758–772, 2011.
- MOORE, D. S. Behavioral epigenetics. **Wiley Interdisciplinary Reviews: Systems Biology and Medicine**, p. 1–8, 2016.
- MOREY, L.; HELIN, K. Polycomb group protein-mediated repression of transcription. **Trends in Biochemical Sciences**, v. 35, p. 323–332, 2010.
- MURRAY, S.; BRIASOULIS, E.; LINARDOU, H.; BAFALOUKOS, D.; PAPADIMITRIOU, C. Taxane resistance in breast cancer: Mechanisms, predictive biomarkers and circumvention strategies. **Cancer Treatment Reviews**, v. 38, p. 890–903, 2012.
- NILE, C. J.; READ, R. C.; AKIL, M.; DUFF, G. W.; WILSON, A. G. Methylation status of a single CpG site in the IL6 promoter is related to IL6 messenger RNA levels and rheumatoid arthritis. **Arthritis and Rheumatism**, v. 58, p. 2686–2693, 2008.
- OHM, J. E.; MCGARVEY, K. M.; YU, X.; CHENG, L.; SCHUEBEL, K. E.; COPE, L.; MOHAMMAD, H. L.; CHEN, W.; DANIEL, V. C.; YU, W.; BERMAN, D. M.; JENUWEIN, T.; PRUITT, K.; SHARKIS, S. J.; WATKINS, D. N.; HERMAN, J. G.; BAYLIN, S. B. A Stem Cell-Like Chromatin Pattern May Predispose Tumor Suppressor Genes to DNA Hypermethylation and Silencing in Adult Cancers. **Nature genetics**, v. 39, p. 237–242, 2007.
- OZBUN, M. A.; BUTEL, J. S. Tumor suppressor p53 mutations and breast cancer: A critical analysis. **Advances in Cancer Research**, v. 66, p. 71–141, 1995.

- PARK, S. Y.; KNON, H. J.; CHOI, Y., LEE, H. E.; KIM, S-W.; KIM, J. H.; KIM, I. A.; JUNG, N.; CHO, N-Y.; KANG, G. W. Distinct patterns of promoter CpG island methylation of breast cancer subtypes are associated with stem cell phenotypes. **Modern Pathology**, v. 25, p. 185–196, 2012.
- PENG, D. F.; KANAI, Y.; SAWADA, M.; HIRAOKA, N.; KITAZAWA, S.; HIROHASHI, S. DNA methylation of multiple tumor-related genes in association with overexpression of DNA methyltransferase 1 (DNMT1) during multistage carcinogenesis of the pancreas. **Carcinogenesis**, v. 27, n. 6, p. 1160–1168, 2006.
- PETRONIS, A. Epigenetics as a unifying principle in the aetiology of complex traits and diseases. **Nature**, v. 465, p. 721–727, 2010.
- PORTELA, A.; ESTELLER, M. Epigenetic modifications and human disease. **Nature Biotechnology**, v. 28, p. 1057–1068, 2010.
- PREKER, P.; NIELSEN, J.; HAMMLER, S.; LYKKE - ANDERSEN, S.; CHRISTENSEN, M. S.; MAPENDANO, C. K.; SCHIERUP, M. H.; JENSEN, T. H. RNA Exosome Depletion Reveals Transcription Upstream of Active Human Promoters. **Science**, v. 322, p. 1851–1854, 2008.
- RACITI, G. A.. SPINELLI, R.; DESIDERIO, A.; LONGO, M.; PARILLO, L.; NIGRO, C.; D'ESPOSITO, V.; MIRRA, P.; FIORY, F.; PILONE, V.; FORESTIERI, P.; FORMISANO, P.; PASTAN, I.; MIELE, C.; BEGUINOT, F Specific CpG hyper-methylation leads to Ankrd26 gene down-regulation in white adipose tissue of a mouse model of diet-induced obesity. **Scientific Reports**, v. 7, p. 1–13, 2017.
- RAKYAN, V. K.; DOWN, T. A.; BALDING, D. J.; BECK, S. Epigenome-wide association studies for common human diseases. **Nature Reviews Genetics**, v. 12, p. 529–541, 2011.
- RAO, X.; EVANS, J.; CHAE, H.; PILROSE, J.; KIM, S.; YAN, P.; HUANG, R-L.; LAI, H-C.; LIN, H.; LIU, Y.; MILLER, D.; RHEE, J-K.;HUANG, Y-W.; GU, F.; GRAY, J. W.; HUANG, T-H. M.; NEPHEW, K. P CpG island shore

methylation regulates caveolin-1 expression in breast cancer. **Oncogene**, v. 32, p. 4519–4528, 2013.

REIK, W. Stability and flexibility of epigenetic gene regulation in mammalian development. **Nature**, v. 447, p. 425–432, 2007.

ROBERTSON, K. D. DNA methylation and human disease. **Nature Reviews Genetics**, v. 6, p. 597–610, 2005.

ROBERTSON, K. D.; WOLFFE, A. P. DNA methylation in health and disease. **Nature Reviews Genetics**, v. 1, p. 11–19, 2000.

SATOH, A.; TOYOTA, M.; ITOH, F.; KIKUCHI, T.; OBATA, T.; SASAKI, Y.; SUZUKI, H.; YAWATA, A.; KUSANO, M.; FUJITA, M.; HOSOKAWA, M.; YANAGIHARA, K.; TOKINO, T.; IMAI, K. DNA methylation and histone deacetylation associated with silencing DAP kinase gene expression in colorectal and gastric cancers. **British Journal of Cancer**, v. 86, p. 1817–1823, 2002.

SCHLESINGER, Y.; STRAUSSMAN, S.; KESHET, I.; FARKASH, S.; HECHT, M.; ZIMMERMAN, J.; EDEN, E.; YAKHINI, Z.; BEN - SHUSHAN, E.; REUBINOFF, B. E.; BERGMAN, H.; SIMON, I.; CEDAR, H. Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. **Nature Genetics**, v. 39, p. 232–236, 2007.

SEILA, A. C.; CALABRESE, M.; LEVINE, S. S.; YEO, G. W.; RAHL, P. B.; FLYNN, R. A.; YOUNG, R. A.; SHARP, P. A. Divergent transcription from active promoters. **Science**, v. 322, p. 1849–1851, 2008.

SHRIVASTAVA, A.; DESOUSA, A. Resilience: A psychobiological construct for psychiatric disorders. **Indian Journal of Psychiatry**, v. 58, p. 38–43, 2016.

SHUKLA, S.; KAVAK, E.; GREGORY, M.; IMASHIMIZU, M.; SHUTINOSKI, B.; KASHLEV, M.; OBERDOERFFER, P.; SANDBERG, T.; OBERDOERFFER, S. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. **Nature**, v. 479, p. 74–81, 2011.

SIEBOLD, A. P.; BENERJEE, R.; TIE, F.; KISS, D. L.; MOSKOWITZ, J.;

HARTE, P. J. Polycomb Repressive Complex 2 and Trithorax modulate *Drosophila* longevity and stress resistance. **Proceedings of the National Academy of Sciences**, v. 107, p. 169–174, 2010.

SIEPEL, A. BEJERANO, G.; PEDERSEN, J. S.; HINRICHS, A. S.; HOU, M.; ROSENBLOOM, K.; CLAWSON, H.; SPIETH, J.; HILLIER, L. W.; RICHARDS, S.; WEINSTOCK, G. M.; WILSON, R. K.; GIBBS, R. A.; KENT, W. J.; MILLER, W.; HAUSSLER, D. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. **Genome Research**, v. 15, p. 1034–1050, 2005.

SLAMON, D. J.; GODOLPHIN, W.; JONES, L. A.; HOLT, J. A.; WONG, S. G.; KEITH, D. E.; LEVIN, W. J.; STUART, S. G., UDOVE, J.; ULLRICH, A., et al., Studies of the HER-2/neu proto-oncogene in human breast and ovarian cancer. **Science**. v. 244, p. 707–712, 1989.

SMOLLER, J. W. The Genetics of Stress-Related Disorders: PTSD, Depression, and Anxiety Disorders. **Neuropsychopharmacology**, v. 41, p. 297–319, 2016.

SPITZWIESER, M.; PIRKER, C.; KOBLMÜLLER, B.; PFEILER, G.; HACKER, S.; BERGER, W.; HEFFETER, P.; CICHNA - MARKL, M. Promoter methylation patterns of ABCB1, ABCC1 and ABCG2 in human cancer cell lines, multidrug-resistant cell models and tumor, tumor-adjacent and tumor-distant tissues from breast cancer patients. **Oncotarget**, v. 7, p. 73347–73369, 2016.

SUZUKI, M. M.; BIRD, A. DNA methylation landscapes: Provocative insights from epigenomics. **Nature Reviews Genetics**, v. 9, p. 465–476, 2008.

TATETSU, H.; UENO, S.; HATA, H.; YAMADA, Y.; TAKEYA, M.; MITSUYA, H.; TENEN, D. G.; OKUNO, T. Down-regulation of PU.1 by methylation of distal regulatory elements and the promoter is required for myeloma cell growth. **Cancer Research**, v. 67, p. 5328–5336, 2007.

- TCHURIKOV, N. A. Molecular mechanisms of epigenetics. **Biochemistry (Mosc)**, v. 70, p. 406-423, 2005.
- TERRY, M. B.; MCDONALD, J. A.; WU, H. C.; ENG, S.; SANTELLA, R. M. Epigenetic Biomarkers of Breast Cancer Risk: Across the Breast Cancer Prevention Continuum. **Advances in Experimental Medicine and Biology**, v. 882, p. 33–68, 2016.
- TIAN, S.; BERTELSMANN, K.; YU, L.; SUN, S. DNA Methylation Heterogeneity Patterns in Breast Cancer Cell Lines. **Cancer Informatics**, v. 15, p. 1–9, 2016.
- TIRADO-MAGALLANES, R.; REBBANI, K.; LIM, R.; PRADHAN, S.; BENOUKRAF, T. Whole genome DNA methylation: beyond genes silencing. **Oncotarget**, v. 8, p. 5629–5637, 2017.
- TIWARI, V. K.; MCGARVEY, K. M.; LICCHESI, J. D. F.; OHM, J. E.; HERMAN, J. G.; SCHÜBELER, D.; BAYLIN, S. B. PcG Proteins, DNA Methylation, and Gene Repression by Chromatin Looping. **Plos Biology**, v. 6, p. 2911–2927, 2008.
- TSUBOI, K.; NAGAMOTO, T.; GOHNO, T.; HIGUSHI, T.; SASAKI, S.; FUJIKI, N.; KUROSUMI, M.; TAKEI, H.; YAMAGUSHI, Y.; NIWA, T.; HAYASHI, S-I. Single CpG site methylation controls estrogen receptor gene transcription and correlates with hormone therapy resistance. **Journal of Steroid Biochemistry and Molecular Biology**, v. 171, p. 209–217, 2017.
- VAN OTTERDIJK, S. D.; MICHELS, K. B. Transgenerational epigenetic inheritance in mammals: How good is the evidence? **FASEB Journal**, v. 30, p. 2457–2465, 2016.
- VAQUERIZAS, J. M.; KUMMERFELD, S. K.; TEICHMANN, S. A.; LUSCOMBE, N. M. A census of human transcription factors: Function, expression and evolution. **Nature Reviews Genetics**, v. 10, p. 252–263, 2009.
- VARLEY, J. M.; ARMOUR, J.; SWALLOW, J. E.; JEFFREYS, A. J.; PONDER, B. A.; T'ANG, A.; FUNG, Y. K.; BRAMMAR, W. J.; WALKER, R. A. The

retinoblastoma gene is frequently altered leading to loss of expression in primary breast tumours. **Oncogene**, v. 4, p. 725–729, 1989.

VIRÉ, E.; BRENNER, C.; DEPLUS, R.; BLANCHON, L.; FRAGA, M.; DIDELOT, C.; MOREY, L., VAN EYNDE, A.; BERNARD, D.; VANDERWINDEN, J. M.; BOLLEN, M., ESTELLER, M.; DI CROCE, L.; DE LAUNOIT, Y.; FUKS, F The Polycomb group protein EZH2 directly controls DNA methylation. **Nature**, v. 439, p. 871–874, 2006.

WEI, J. W.; HUANG, K.; YANG, C.; KANG, C-S Non-coding RNAs as regulators in epigenetics (Review). **Oncology Reports**, v. 37, p. 3–9, 2017.

WIDSCHWENDTER, M.; FIELG, H.; EGLE, D.; MUELLER - HOLZNER, E.; SPIZZO, G.; MARTH, C.; WEISENBERG, D. J.; CAMPAN, M.; YOUNG, J.; JACOBS, I.; LAIRD, P. W. Epigenetic stem cell signature in cancer. **Nature Genetics**, v. 39, p. 157–158, 2007.

WINGENDER, E.; SCHOEPS, T.; HAUBROCK, M.; DÖNITZ, J. TFClass: A classification of human transcription factors and their rodent orthologs. **Nucleic Acids Research**, v. 43, p. 97–102, 2015.

WOOLFE, A.; GOODSON, M.; GOODE, D. K.; SNELL, P.; MCEWEN, G. K.; VAVOURI, T.; SMITH, S. F.; NORTH, P.; CALLAWAY, H.; KELLY, K.; WALTER, K.; ABNIZOVA, I.; GILKS, W.; EDWARDS, T. J. K.; COOKE, J. E.; ELGAR, G. Highly conserved non-coding sequences are associated with vertebrate development. **Plos Biology**, v. 3, p. 116–130, 2005.

WU, H.; COSKUN, V.; TAO, J.; XIE, W.; GE, W.; YOSHIKAWA, K.; LI, E.; ZHANG, Y.; SUN, Y. E. Dnmt3a-dependent nonpromoter DNA methylation facilitates transcription of neurogenic genes. **Science**, v. 329, p. 444–448, 2010.

XIE, H.; WANG, M.; DE ANDRADE, A.; BONALDO, M. F., GALAT, V.; ARNDT, K.; RAJARAM, V.; GOLDMAN, S.; TOMITA, T.; SOARES, M. B. Genome-wide quantitative assessment of variation in DNA methylation patterns. **Nucleic Acids Research**, v. 39, p. 4099–4108, 2011a.

- XIE, Z.; HU, S.; QIAN, J.; BLACKSHAW, S.; ZHU, H. Systematic characterization of protein-DNA interactions. **Nucleic Acids Research**, v. 39, p. 4099–4108, 2011b.
- YAN, H.; GUAN, Q.; HE, J.; LIN, Y.; ZHANG, J.; LI, H.; LIU, H.; GU, H.; GUO, Z.; HE, F. Individualized analysis reveals CpG sites with methylation aberrations in almost all lung adenocarcinoma tissues. **Journal of Translational Medicine**, v. 15, p. 1–9, 2017.
- YU, D.-H.; WARE, C.; WATERLAND, R. A.; ZHANG, J.; CHEN, M.-H.; GADKARI, M.; KUNDE-RAMAMOORTHY, G.; NOSAVANH, L. K.; SHEN, L. Developmentally Programmed 3' CpG Island Methylation Confers Tissue- and Cell-Type-Specific Transcriptional Activation. **Molecular and Cellular Biology**, v. 33, p. 1845–1858, 2013.
- ZARATIEGUI, M.; IRVINE, D. V.; MARTIENSSEN, R. A. Noncoding RNAs and Gene Silencing. **Cell**, v. 128, p. 763–776, 2007.
- ZARET, K. S.; CARROLL, J. S. Pioneer transcription factors: establishing competence for gene expression. **Genes and Development**, v. 25, p. 2227–2241, 2011.
- ZEMACH, A.; MCDANIEL, Y. E.; SILVA, P.; ZILBERMAN, D. Genome-wide evolutionary analysis of eukaryotic DNA methylation. **Science**, v. 328, p. 916–919, 2010.
- ZHANG, C.; WANG, X.; LI, X.; ZHAO, N.; WANG, Y.; HAN, X.; CI, C.; LI, M.; ZHANG, Y. The landscape of DNA methylation-mediated regulation of long non-coding RNAs in breast cancer. **Oncotarget**, 2017.
- ZHANG, X.; YAZAKI, J.; SUNDARESAN, A.; COKUS, S.; CHAN, S. W.-L.; CHEN, H.; HEDERSON, I. R.; SHINN, P.; PELLEGRINI, M.; JACOBSEN, S. E.; ECKER, J. R. Genome-wide High-Resolution Mapping and Functional Analysis of DNA Methylation in Arabidopsis. **Cell**, v. 126, p. 1189–1201, 2006.
- ZHENG, X.; ZHAO, Q.; WU, H.-J.; LI, W.; WANG, H.; MEYER, C. A.; QIN, Q. A.; XU, H.; ZANG, C.; JIANG, P.; LI, F.; HOU, Y.; HE, J.; WANG, J.; WANG,

J.; ZHANG, P.; ZHANG, Y.; LIU, X. S. MethylPurify: tumor purity deconvolution and differential methylation detection from single tumor DNA methylomes. **Genome biology**, v. 15, p. 1–13, 2014.

ZHU, W.; SRINIVASAN, K.; DAI, Z.; DUAN, W.; DRUHAN, L. J.; DING, H; YEE, L.; VILLALONA - CALERO, M. A.; PLASS, C.; OTTERSON, G. A. Methylation of Adjacent CpG Sites Affects Sp1 / Sp3 Binding and Activity in the p21 Cip1 Promoter. **Molecular and Cellular Biology**, v. 23, p. 4056–4065, 2003.

ZILBERMAN, D.; GHERING, M.; TRAN, R. K.; BALLINGER, T.; HENIKOFF, S. Genome-wide analysis of Arabidopsis thaliana DNA methylation uncovers an interdependence between methylation and transcription. **Nature Genetics**, v. 39, n. 1, p. 61–69, 2007.

APÊNDICE 1 – Tabela de genes relacionados a alterações epigenéticas de ganho ou perda de metilação em câncer de mama

Tabela 2: **Genes relacionados a alterações epigenéticas de ganho ou perda de metilação em câncer de mama.** Revisão bibliográfica sobre os genes que apresentam alterações epigenéticas em tumores mamários, contendo os autores utilizados, os genes e algumas características genômicas. Todas as informações sobre função biológica dos genes apresentadas na tabela foram retiradas do banco de dados National Center Biotechnology Information (<https://www.ncbi.nlm.nih.gov/>)

Genes	Referência	CGI	Posição da CGI	Função Biológica do Gene
<i>BRCA1</i>	Holm <i>et al.</i> , 2016	1	Intergênica	Supressor tumoral. Relacionado a câncer de mama e de ovários. Desempenha um importante papel no reparo do DNA.
<i>HORMAD1</i>	Holm <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Marcador tumoral. Família de genes envolvidos na ligação com a cromatina e na regulação do ciclo celular.
<i>ERBB2</i>	Holm <i>et al.</i> , 2016	3	2 intergênicas e 1 intergênica	Codifica um receptor de tirosina quinase. Sua superexpressão está relacionada a

				vários tipos de câncer, dentre eles mama e ovário.
PIK3CA	Holm <i>et al.</i> , 2016	1	Intragênica	Comumente relacionado a câncer cervical. Codifica a subunidade α catalítica da proteína fosfatidilinositol 4,5-bifosfato-3-quinase que usa ATP para fosforilar PtdIns, Ptdln4P, PtdIns(4,5) P2.
TP53	Holm <i>et al.</i> , 2016	2	1 intragênica e a 1 intergênica	Supressor tumoral. Relacionado a regulação do ciclo celular, apoptose, senescência, reparo do DNA e mudanças no metabolismo.
CDH1	Holm <i>et al.</i> , 2016	1	Intragênica	Representante da superfamília das caderinas. Mutações nesse gene estão relacionados a câncer, como ovários,

				mama, colorretal, entre outros. A perda de sua função contribui para a progressão do câncer por aumentar a proliferação, invasão e metástase.
BRCA2	Holm <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Supressor tumoral. Relacionado a câncer de mama e de ovários. Desempenha um importante papel no reparo do DNA.
APC	Johnson <i>et al.</i> , 2017	1	Intragênica	Proteína supressora que atua como antagonista na via de sinalização Wnt. Está envolvido em outros processos como migração, adesão, ativação transcricional e apoptose. Mutações nesse gene geram proteína truncadas que causam polipose adenomatose

				familiar, que geralmente progride para fenótipos oncológicos.
RASSF1	Johnson <i>et al.</i> , 2017	2	2 intragênicas	Proteína similar a proteína efetora Ras. Perda ou alteração na expressão está relacionada a câncer. A proteína interage com proteínas de reparo de DNA.
HPCAL1	Khakpour <i>et al.</i> , 2017	2	2 intragênicas	Codifica um dos membros de uma família de proteínas relacionadas a ligação de cálcio em neurônios específicos, similar a hipocalceína em humanos. Pode estar envolvida na regulação dependente de cálcio ou na fosforilação de rodopsinas. Também pode ser uma importante via de sinalização neural no sistema nervoso

				central.
NEUROG2	Khakpour <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Fator de transcrição hélice- <i>loop</i> -hélice neural. Expresso em células progenitoras neurais no desenvolvimento do sistema nervosos central e periférico.
SLC6A3	Khakpour <i>et al.</i> , 2017	3	2 intragênicas e 1 a <i>upstream</i> (-2Kb)	Transportador de dopamina. Membro da família de neurotransmissores de transporte dependentes de sódio e cloreto. Variações nessa proteína estão associadas a epilepsia idiopática, transtorno de déficit de atenção, hiperatividade, dependência de álcool e cocaína, susceptibilidade a doença de Parkinson e proteção contra dependência

				à nicotina.
RAB40C	Khakpour <i>et al.</i> , 2017	7	1 <i>upstream</i> e 6 intragênica	Membro da família de oncogenes RAS. Provável componente de reconhecimento de substrato de um complexo E3 ubiquitina ligase ECS (Elongin-Cullin-SOCS-box-caixa). Relacionada a ubiquitinação de proteínas alvos e sua subsequente degradação no proteossoma.
ZNF584	Khakpour <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Está envolvida na regulação transcricional. Codifica uma proteína com domínio dedo de zinco.
FOXD3	Khakpour <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Proteína pertencente a família <i>forkhead</i> de fatores de transcrição. Mutações

				nesse gene estão relacionadas a susceptibilidade autoimune.
PDE4C	Khakpour <i>et al.</i> , 2017	4	Intragênicas	Proteína da família das fosfodiesterases de nucleotídeos cíclicos (PDE) e subfamília PDE4. Regula a concentração celular de cAMP, desempenhando um papel fundamental em muitos processo biológicos.
MPPE1	Khakpour <i>et al.</i> , 2017	2	1 intergênica e 1 Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Metaloprotease 1. Necessária para o transporte de GPI do retículo endoplasmático para o complexo de Golgi. Alguns estudos relatam o papel de metaloproteases no remodelamento da matriz extracelular levando ao fenótipo

				oncológico devido a maior capacidade de invasibilidade e metástase (LEIGHT ET AL., 2017).
<i>PRR5</i>	Khakpour <i>et al.</i> , 2017	5	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Codifica uma proteína com o domínio rico em prolina. Supressor tumoral estando envolvido com tumorigênese mamária e colorretal. Há transcritos desse gene relacionados a proteínas ativadoras de GTPases Rho (ARHGAP8).
<i>RAD51C</i>	Khakpour <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família RAD51, sendo conhecida pelo seu envolvimento em recombinação homóloga e reparo do DNA. Sua superexpressão está relacionada a progressão tumoral em

				câncer de mama.
ATM	Khakpour <i>et al.</i> , 2017	1	Intragênica	Codifica proteínas membros da família PI3/PI4 quinase. Relacionada as principais vias do <i>checkpoint</i> do ciclo celular, que são necessárias a resposta celular para danos no DNA e estabilidade genoma.
NEUROD1	Khakpour <i>et al.</i> , 2017	1	Intergênica	Fator de transcrição da família NeuroD. Regula genes envolvidos na expressão da insulina.
TAL1	Khakpour <i>et al.</i> , 2017	2	Intragênica	Fator de transcrição. Codifica uma proteína de leucemia linfocítica aguda.
SOX1	Khakpour <i>et al.</i> ,	2	Inicia a montante (-2Kb) do sítio de início de	Fator de transcrição envolvido na regulação do desenvolvimento

	2017		transcrição e estende para o corpo gênico	embrionário. A proteína codificada pode atuar como um ativador transcricional após formar um complexo protéico com outras proteínas.
FGF2	Khakpour <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Proteína membro da família de fator de crescimento de fibroblasto. Se ligam a heparina e possuem amplas atividades mitogênicas e angiogênicas. Está associada a diversos processos biológicos como desenvolvimento de membros, sistema nervoso, cicatrização de feridas e crescimento tumoral.
ESR1	Khakpour <i>et al.</i> , 2017	1	Intragênica	Fator de transcrição. Codifica um receptor de estrogênio, sendo composto

				<p>por vários domínios de ligação de hormônios, DNA e fator de transcrição. Os receptores de estrogênio estão relacionados a tumores de mama e endométrio.</p>
GSTM2	Khakpour <i>et al.</i> , 2017	1	Intragênica	<p>Glutathione S-transferase que pertence a classe mu. A classe mu das enzimas atuam na desintoxicação de compostos eletrofílicos, incluindo carcinógenos, drogas terapêuticas, toxinas ambientais e produtos do estresse oxidativo. Variações genéticas nessa proteína podem alterar a susceptibilidade de um indivíduo a carcinógenos e toxinas, além de afetar a</p>

				toxicidade e a eficácia de certas drogas.
SIX3	Khakpour <i>et al.</i> , 2017	1	Intragênica	Membro da família de fatores de transcrição <i>homeobox sine oculis</i> . Desempenha um papel no desenvolvimento de olhos e mutações nesse gene têm sido associadas a holoprosencefalia tipo 2.
EM1	Khakpour <i>et al.</i> , 2017	6	3 intergênicas, 2 intragênicas e 1 intragênica se estendendo depois da terminação do gene	Acredita-se que o gene <i>homeobox</i> tenha um papel no controle do desenvolvimento. Mutações nos homólogos em camundongos produzem diferentes efeitos no desenvolvimento que são letais. Os homólogos humanos 1 e 2 estão envolvidos na formação de

				padrões durante o desenvolvimento do sistema nervoso central.
PAX3	Khakpour <i>et al.</i> , 2017	5	4 intragênicas e 1 intergênica	Membro da família de fatores de transcrição (PAX). Desempenham papéis críticos durante o desenvolvimento fetal. Mutações estão relacionadas a síndrome de Waardenburg e rabdomyosarcoma alveolar.
GSK1	Khakpour <i>et al.</i> , 2017	1	Intragênica	Proteína glicogênio sintase quinase. Envolvidas na via Wnt de câncer de cólon de útero. Desregulação desta proteína está envolvida em carcinomas hepatocelular, câncer gástrico e neuroblastoma.

PKD2	Khakpour <i>et al.</i> , 2017	1	Intragênica	Proteína membro da família das policistinas. Está relacionada ao transporte e sinalização de cálcio em células epiteliais renais.
CCND1	Khakpour <i>et al.</i> , 2017	3	2 intragênicas e 1 Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Proteínas pertencentes a família das ciclinas. A ciclina codificada por esta proteína funciona como uma subunidade reguladora de CDK4 e CDK6, cuja a atividade é necessária para a transição de G1 para S no ciclo celular. Esta proteína interage com a proteína supressora tumoral Rb. Mutações, amplificação e superexpressão desse gene são observadas frequentemente em diversos tumores e podem contribuir para

				o desenvolvimento tumoral.
<i>TWIST1</i>	Khakpour <i>et al.</i> , 2017	1	Intragênica se estendendo depois da terminação do gene	Fator de transcrição com domínio hélice – <i>loop</i> – hélice que desempenha um importante papel no desenvolvimento embrionário. Regula a transcrição de genes envolvidos no fechamento da sutura craniana durante o desenvolvimento do crânio. Também possui funções no fechamento neural, desenvolvimento de membros e metabolismo da gordura marrom. Sua superexpressão está relacionada a múltiplos tumores.
<i>CXCL1</i>	Khakpour <i>et al.</i> ,	1	Inicia a montante (-2Kb) do sítio de início de	Membro da subfamília CXC de

	2017		transcrição e estende para o corpo gênico	quimiocinas. Desempenha papel na inflamação e como um quimioatrativo de neutrófilos. Superexpressão dessa proteína está associada ao crescimento e progressão de certos tumores.
CDCP1	Khakpour <i>et al.</i> , 2017	1	Intragênica	Proteína transmembrana que atua como substrato para as proteínas da família das quinases Src. Desempenha um papel na regulação dependente da fosforilação de tirosina em eventos celulares que estão envolvidos na invasão tumoral e metástase.
SOX17	Khakpour <i>et al.</i> , 2017	1	Intragênica	Membro da família de fatores de transcrição SOX, envolvidos na regulação

				do desenvolvimento embrionário e no fenótipo celular. A proteína codificada pode atuar como um regulador transcricional após formar complexos protéicos com outras proteínas.
DLEC1	Park <i>et al.</i> , 2012	1	Intragênica	A regulação negativa desse gene já foi observada em vários tumores, incluindo câncer de pulmão, esofágico, renal e carcinomas de células escamosas de cabeça e pescoço.
GRIN2B	Park <i>et al.</i> , 2012	2	Intragênica	Membro da família de receptores N-metil-D-aspartato (NMDA). Expressão precoce desse gene sugere um papel no desenvolvimento do cérebro, na

				<p>formação de circuitos, na plasticidade sináptica, na migração e diferenciação celular. Mutações que ocorrem naturalmente nesses genes estão associadas a distúrbios do neurodesenvolvimento, incluindo transtorno do espectro de autismo, transtorno do déficit de atenção com hiperatividade, epilepsia e esquizofrenia.</p>
GSTP1	Park <i>et al.</i> , 2012	1	Intragênica	<p>Glutionina-S-transferase (GSTs). Desempenham um papel importante na desintoxicação catalisando a conjugação de muitos compostos hidrofóbicos e eletrofilicos com glutathione reduzida. Possui atividade no metabolismo de</p>

				xenobióticos e desempenha um papel na susceptibilidade ao câncer e outras doenças.
HOXA1	Park <i>et al.</i> , 2012	2	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico e 1 intragênica	Fatores de transcrição denominados genes <i>homeobox</i> . Sua expressão é espacial e regulada durante o desenvolvimento embrionário. Está envolvida na regulação da expressão gênica, morfogênese e diferenciação.
HOXA10	Park <i>et al.</i> , 2012	2	Intragênica	Fatores de transcrição denominados genes <i>homeobox</i> . Sua expressão é regulada durante o desenvolvimento embrionário. A expressão dessa proteína está correlacionada com fertilidade,

				viabilidade de embriões e regulação do comprometimento da linhagem hematopoiética.
IGF2	Park <i>et al.</i> , 2012	4	Intragênica	Membro da família da insulina de fatores de crescimento que estão envolvidos no desenvolvimento e crescimento. Gene <i>imprintado</i> com expressão do alelo paterno. Alterações epigenéticas nesse gene estão associadas a síndrome de Beckwith-Wiedemann, síndrome de Silver – Russell e tumor de Wilms.
MT1G	Park <i>et al.</i> , 2012	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Supressor tumoral. Está associado a diversos tipos de tumores como hepatoblastoma, câncer colorretal e

				câncer de tireóide através da modulação da via de sinalização celular PI3K/Akt.
RARB	Park <i>et al.</i> , 2012	1	Intergênica	Receptor de ácido retinóico beta. Membro da superfamília de receptores do hormônio esteróide - tireóide. Liga-se ao ácido retinóico que está envolvido com a sinalização celular na morfogênese embrionária, crescimento e diferenciação celular. O gene foi identificado pela primeira vez em carcinomas hepatocelulares.
RUNX3	Park <i>et al.</i> , 2012	3	Intragênicas	Fator de transcrição. Interage com vários outros fatores de transcrição. É um supressor tumoral, sendo deletados em

				carcinomas ou silenciados.
SCGB3A1	Park <i>et al.</i> , 2012	1	Intragênica se estendendo depois da terminação do gene	Membros da família de secreção de globinas. Expresso em tecidos pulmonares. Já foi relacionada com alguns tipos tumorais, sendo utilizadas como biomarcadores em alguns deles.
SFRP1	Park <i>et al.</i> , 2012	1	Intragênica	Membro da família <i>SFRP</i> , atuando como um modulador da via de sinalização Wnt. O silenciamento epigenético desse gene leva a desregulação da via Wnt, que está associada ao câncer.
SFRP4	Park <i>et al.</i> , 2012	1	Intragênica	Membro da família <i>SFRP</i> , atuando como um modulador da via de sinalização Wnt. A expressão de <i>SFRP4</i> no miocárdio

				ventricular correlaciona-se com genes envolvidos na apoptose.
TMEFF2	Park <i>et al.</i> , 2012	1	Intragênica	Membro da família de proteínas transmembranas tomorregulinas. Esta proteína funciona como um oncogene e um supressor tumoral dependendo do contexto celular e pode regular a invasão de células em câncer de próstata. Em vários tipos de tumores este gene encontra-se hipermetilado.
CCND2	Park <i>et al.</i> , 2012	2	Intergênica e intragênica	Família das ciclinas. A ciclina codificada por esta proteína funciona como uma subunidade reguladora de CDK4 e CDK6, cuja a atividade é necessária para a

				transição de G1 para S no ciclo celular. Esta proteína interage com a proteína supressora tumoral Rb. Altos níveis de expressão desses genes foram observados em tumores ovarianos e testiculares.
CDKN2A	Park <i>et al.</i> , 2012	5	4 intragênicas e 1 intergênica	CDK4. Possui função regulatória de p53, influenciando na progressão do ciclo celular.
DAPK1	Spitzwieser <i>et al.</i> , 2016	3	Intragênica	Proteína serina-treonina quinase 1. Mediador positivo de morte celular programada. Candidato a supressor tumoral.
MGMT	Spitzwieser <i>et al.</i> ,	1	Intragênica	Proteína de reparo de DNA. Está

	2016			envolvida na defesa celular contra mutagênese e toxicidade de agentes alquilantes. Mutações no promotor desse gene têm sido associadas a vários tipos de câncer como colorretal, câncer de pulmão, linfoma e glioblastoma.
PITX2	Terry <i>et al.</i> , 2016	7	Intragênicas	Membro da família <i>homeobox</i> RIEG/PITX. Codifica um fator de transcrição e regula a expressão de lisil hidroxilase do procolágeno. Desempenha papel na diferenciação terminal do fenótipo de células somatotróficas e lactotróficas, estando envolvida no desenvolvimento de órgãos oculares, dentais e abdominais. Está envolvida na

				regulação do hormônio prolactina.
NT5E	Terry <i>et al.</i> , 2016	1	Intragênica	A proteína codificada por esse gene está envolvida na diferenciação de linfócitos. Defeitos neste gene podem levar a calcificação de articulações e artérias.
ZNF217	Terry <i>et al.</i> , 2016	2	Intragênicas	Proteína com domínio dedo de zinco. Sua função está associada a funções apoptóticas. É um alvo terapêutico em câncer de mama e biomarcador prognóstico para a doença.
PTPRO	Terry <i>et al.</i> , 2016	1	Intragênica	Membro da família de receptores de proteínas tirosinas fosfatases. A proteína codificada pode ter muitas isoformas e funções específicas, incluindo a

				regulação e produção de osteoclastos, inibição da proliferação celular e está envolvida no processo apoptótico. É um candidato a supressor tumoral. A diminuição da expressão desse gene está relacionado a diversos tumores.
HOTAIR	Terry <i>et al.</i> , 2016	4	1 intergênica e 3 intragênicas	Promove a repressão epigenética do gene HOX. É bastante expresso em tumores.
L3MBTL1	Terry <i>et al.</i> , 2016	2	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico e 1 intragênica	Representante de um dos membros do complexo <i>Polycomb</i> . Possui atividade na regulação gênica através de modificações na cromatina. A proteína codificada também pode ser necessária

				no processo de mitose.
GSTM1	Terry <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Glutathione-S-transferase que pertence a classe mu. A classe mu das enzimas atuam na desintoxicação de compostos eletrofilicos, incluindo carcinógenos, drogas terapêuticas, toxinas ambientais e produtos do estresse oxidativo. Variações genéticas nessa proteína podem alterar a susceptibilidade de um indivíduo a carcinógenos e toxinas, além de afetar a toxicidade e a eficácia de certos medicamentos. Mutações nulas nesta classe têm sido associadas a vários tipos de tumores, provavelmente devido a maior susceptibilidade a toxinas

				ambientais e carcinógenos.
SYNM	Terry <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família de filamentos intermediários. São proteínas de citoesqueleto que conferem resistência ao estresse mecânico. Essa proteína fornece um importante suporte estrutural aos músculos.
FBXW7	Terry <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família de proteína F-box que constituem o complexo da proteína ubiquitina ligase (SCFs) que funcionam na ubiquitinação dependente de fosforilação. Essas proteínas liga-se diretamente a ciclina E degradando-a por meio de ubiquitinação. Mutações nesses

				genes são encontradas em linhagens celulares de câncer de mama e de ovário, implicando o potencial papel do gene no desenvolvimento do câncer.
NUP155	Terry <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Nucleosporina. A proteína codificada por este gene desempenha um papel na formação da dupla membrana e na fusão do envelope nuclear. Também pode estar envolvida na fisiologia cardíaca e pode estar associada à patogênese da fibrilação arterial.
STK11	Tian <i>et al.</i> , 2016	4	1 Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico, 2 intragênicas e 1 intragênica se estendendo depois	Membro da família serina/treonina quinase que regula a polaridade celular e funciona como um supressor tumoral.

			da terminação do gene	Mutações nesse gene têm sido associadas a síndrome Peutz-Jeghers (distúrbio autossômico dominante caracterizado pelo crescimento de pólipos no trato gastro-intestinal), máculas pigmentadas na pele e na boca e outras neoplasias.
PTEN	Tian <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Supressor tumoral. Há grande incidência de mutações nesse gene em câncer. Codifica a proteína fosfatidilinositol-3,4,5-trifosfato 3-fosfatase. Ao contrário da maioria das proteínas tirosina fosfatases, esta proteína preferencialmente desfosforila substratos de fosfoinosítídeos. Regula negativamente

				os níveis intracelulares de fosfatidilinositol-3,4,5-trifosfato nas células e funciona como um supressor de tumor, regulando negativamente a via de sinalização de AKT / PKB.
CHEK2	Tian <i>et al.</i> , 2016	1	Intergênica	Regulador do ponto de verificação do ciclo celular e supressor tumoral. Esta proteína nuclear é um membro da subfamília CDS1 das proteínas quinases serina/treonina. Quando ativada, a proteína codificada é conhecida por inibir a fosfatase CDC25C, impedindo a entrada na mitose, e mostrou estabilizar a proteína supressora de tumor p53, levando à parada do ciclo celular em G1.

				<p>Além disso, essa proteína interage e fosforila o BRCA1, permitindo que esse gene restaure a sobrevivência após danos no DNA. Mutações neste gene têm sido associadas à síndrome de Li-Fraumeni, um fenótipo de câncer familiar altamente penetrante, geralmente associado a mutações hereditárias em TP53. Além disso, acredita-se que mutações neste gene conferem uma predisposição para sarcomas, câncer de mama e tumores cerebrais.</p>
MRE11A	Tian <i>et al.</i> , 2016	1	Intragênica se estendendo depois da terminação do gene	<p>Proteína nuclear envolvida na recombinação homóloga, na manutenção do comprimento dos telômeros e no</p>

				<p>reparo de quebra de fita dupla do DNA. A proteína tem atividade de exonuclease e de endonuclease. Ela forma um complexo com o homólogo RAD50; este complexo é necessário para a junção não-homóloga de extremidades de DNA e possui aumento de endonuclease e exonuclease na fita de DNA.</p>
<i>RAD50</i>	Tian <i>et al.</i> , 2016	1	Intragênica	<p>A proteína codificada por este gene é altamente similar a <i>Saccharomyces cerevisiae</i> Rad50, uma proteína envolvida no reparo na quebra da fita dupla de DNA. Essa proteína é importante para o reparo na quebra de fita dupla do DNA, a ativação do ponto de</p>

				verificação do ciclo celular, manutenção do telômero e a recombinação meiótica.
NBN	Tian <i>et al.</i> , 2016	1	Intragênica	Membro do complexo de reparação de ruptura de fita dupla MRE11/RAD50 que consiste em 5 proteínas. Acredita-se que esteja envolvido no reparo de quebra de fita dupla do DNA e na ativação do ponto de verificação induzido por dano ao DNA. Mutações nesse gene estão associadas à síndrome de quebra de Nijmegen, uma síndrome de instabilidade cromossômica autossômica recessiva caracterizada por microcefalia, retardo de crescimento, imunodeficiência e predisposição ao

				câncer
FGFR2	Tian <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família de receptores do fator de crescimento de fibroblastos. Este membro particular da família é um receptor de alta afinidade para o fator de crescimento ácido, básico e/ou queratinócito, dependendo da isoforma. Mutações nesse gene estão associadas à síndrome de Crouzon, síndrome de Pfeiffer, craniossinostose, síndrome de Apert, síndrome de Jackson-Weiss, síndrome de Beare-Stevenson cutis gyrate, síndrome de Saethre-Chotzen e craniossinostose sindrômica.

<i>NTRK1</i>	Tian <i>et al.</i> , 2016	4	Intragênica	<p>Membro da família de receptores de tirosina quinase neurotrófica (NTRK). Esta quinase é um receptor ligado à membrana que, após ligação à neurotrofina, se fosforila e é membro da via da MAPK. A presença desta quinase leva à diferenciação celular e pode desempenhar um papel na especificação dos subtipos de neurônios sensoriais. Mutações nesse gene têm sido associadas à insensibilidade congênita à dor, anidrose, comportamento auto-mutilante, incapacidade cognitiva e câncer.</p>
---------------------	---------------------------	---	-------------	---

<i>NTRK2</i>	Tian <i>et al.</i> , 2016	1	Intragênica	Membro da família de receptores de tirosina quinase neurotrófica (NTRK). Esta quinase é um receptor ligado à membrana que, após ligação à neurotrofina, se fosforila. É membro da via da MAPK. Sinalização através desta quinase leva à diferenciação celular. Mutações nesse gene têm sido associadas com obesidade e transtornos de humor.
<i>PDGFRA</i>	Tian <i>et al.</i> , 2016	7	Intragênica	Receptor da tirosina quinase da superfície celular para membros da família do fator de crescimento derivado de plaquetas. Esses fatores de crescimento são mitógenos para células

				de origem mesenquimal. Estudos sugerem que este gene desempenha um papel no desenvolvimento de órgãos, cicatrização de feridas e progressão tumoral. Mutações neste gene têm sido associadas com síndrome hipereosinofílica idiopática, tumores estromais gastrointestinais somáticos e familiares, e uma variedade de outros cânceres.
PAX7	Tian <i>et al.</i> , 2016	8	Intragênicas	Membro da família de fatores de transcrição (PAX). Esses genes desempenham papéis críticos durante o desenvolvimento fetal e o crescimento do

				câncer.
CDH11	Tian <i>et al.</i> , 2016	2	1 intragênica e 1 intragênica se estendendo depois da terminação do gene	Caderina clássica do tipo II da superfamília caderina. Proteínas integrais de membrana que medeiam a adesão célula-célula dependente de cálcio. A expressão deste tipo de caderina nas linhagens celulares osteoblásticas e a sua regulação positiva durante a diferenciação sugerem uma função específica no desenvolvimento e manutenção óssea.
TLX1	Tian <i>et al.</i> , 2016	3	2 intragênicas e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Fator de transcrição nuclear que pertence à subfamília NK-linked dos genes <i>homeobox</i> . A proteína codificada é

				necessária para o desenvolvimento normal do baço durante a embriogênese. A expressão desse gene está associada a certas leucemias linfoblásticas agudas de células T.
FGFR3	Tian <i>et al.</i> , 2016	7	6 intragênicas e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família do receptor do fator de crescimento de fibroblastos (FGFR). Este membro particular liga-se à hormônios do crescimento de fibroblastos ácidos e básicos e desempenha um papel no desenvolvimento e manutenção óssea. Também está associado a mitogênese e diferenciação celular. Mutações nesse gene levam à craniossinostose e a múltiplos tipos de

				displasia esquelética
CDX2	Tian <i>et al.</i> , 2016	4	2 intragênicas, 2 intergênicas	Membro da família do fator de transcrição <i>homeobox</i> . A proteína codificada é um importante regulador de genes específicos do intestino envolvidos no crescimento celular e diferenciação. Esta proteína também desempenha um papel no desenvolvimento embrionário inicial do trato intestinal. A expressão aberrante desse gene está associada à inflamação intestinal e à tumorigênese.
NCAM1	Tian <i>et al.</i> , 2016	1	Intragênica	Proteína de adesão celular que é um membro da superfamília das imunoglobulinas. A proteína codificada

				<p>está envolvida nas interações célula-célula, bem como nas interações célula-matriz durante o desenvolvimento e a diferenciação. Demonstrou-se que a proteína codificada está envolvida no desenvolvimento do sistema nervoso, nas células envolvidas na expansão de células T e células dendríticas que desempenham um papel importante na vigilância imunológica.</p>
DPYSL3	Tian <i>et al.</i> , 2016	2	<p>1 intragênica e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico</p>	<p>Proteína 3 relacionada a Dihidropirimidinase. Membro da família de proteínas mediadoras da resposta às colapsinas. As proteínas mediadoras da resposta à colapsina formam homo e</p>

				hetero tetrâmeros e facilitam a orientação, o crescimento e a polaridade dos neurônios. Necessária para sinalização por semaforinas de classe 3 e subsequente remodelação do citoesqueleto. Desempenha um papel na orientação dos axônios e migração celular.
DPYSL2	Tian <i>et al.</i> , 2016	2	1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico e 1 intragênica	Proteína 2 relacionada a Dihidropirimidinase. Membro da família de proteínas mediadoras da resposta às colapsinas. As proteínas mediadoras da resposta à colapsina formam homo e hetero tetrâmeros e facilitam a orientação, o crescimento e a polaridade

				<p>dos neurônios. A proteína codificada promove a montagem de microtúbulos e é necessária na papel na sinalização sináptica através de interações com os canais de cálcio. Este gene foi relacionado a múltiplos distúrbios neurológicos e a hiperfosforilação da proteína codificada pode desempenhar um papel fundamental no desenvolvimento da doença de Alzheimer.</p>
<i>DPYSL4</i>	Tian <i>et al.</i> , 2016	3	<p>1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico, 1 intragênica e 1 intergênicas</p>	<p>Proteína 4 relacionada a Dihidropirimidinase. Membro da família de proteínas mediadoras da resposta às colapsinas. As proteínas mediadoras da</p>

				resposta à colapsina formam homo e hetero tetrâmeros e facilitam a orientação, o crescimento e a polaridade dos neurônios. Necessária para sinalização por semaforinas de classe 3 e subsequente remodelação do citoesqueleto. Desempenha um papel na orientação dos axônios e migração celular.
SH3GL3	Tian <i>et al.</i> , 2016	1	Intragênica	Endofilina A3. Proteína associada a endocitose.
EPHB2	Tian <i>et al.</i> , 2016	2	1 intragênica e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família de receptores Eph das glicoproteínas transmembranas do receptor da tirosina quinase. Eles ligam-

				<p>se a efrinas e estão envolvidos em diversos processos celulares, incluindo motilidade, divisão e diferenciação. Uma característica distintiva da sinalização Eph-efrina é que ambos, os receptores e os ligantes são competentes para transduzir uma cascata de sinalização, resultando em sinalização bidirecional. Esta proteína pertence a um subgrupo dos receptores Eph chamado EphB. As proteínas deste subgrupo são distinguidas de outros membros da família por homologia de sequência e afinidade de ligação preferencial para ligantes de efrina B ligados à membrana.</p>
--	--	--	--	--

<i>DNM1</i>	Tian <i>et al.</i> , 2016	4	3 intragênicas e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da subfamília dinamina de proteínas de ligação a GTP. A proteína codificada possui propriedades mecanoquímicas únicas usadas para romper membranas e está envolvida na endocitose mediada por clatrina e em outros processos de tráfico vesicular. Actina e outras proteínas do citoesqueleto atuam como parceiros de ligação para a proteína codificada, que também pode se auto-montar levando à estimulação da atividade da GTPase.
<i>BIN1</i>	Tian <i>et al.</i> , 2016	2	1 intragênica e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo	Proteína adaptadora nucleocitoplasmática, que foi inicialmente identificada como uma proteína que

			<p>gênico</p>	<p>interage com MYC com características de um supressor tumoral. As isoformas que são expressas no sistema nervoso central podem estar envolvidas na endocitose da vesícula sináptica e podem interagir com a dinamina, sinaptojanina, endofilina e clatrina. Isoformas que são expressas em isoformas musculares localizam-se no citoplasma e no núcleo e ativam um processo apoptótico independente de caspases. Estudos em ratos sugerem que este gene desempenha um papel importante no desenvolvimento do músculo cardíaco.</p>
--	--	--	---------------	--

SH3GL2	Tian <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Endofilina A1. Relacionado com a endocitose das vesículas sinápticas. Cooperar com o SH3GL2 para mediar o tráfego endocítico e a sinalização precoce do BDNF-NTRK2 a partir dos endossomas iniciais.
FLT1	Tian <i>et al.</i> , 2016	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família de receptor de crescimento endotelial vascular. Esta proteína se liga ao VEGFR-A, VEGFR-B e ao fator de crescimento placentário. Desempenha um papel importante na angiogênese e vasculogênese. As isoformas incluem variantes solúveis encurtadas do receptor transmembrana. As isoformas solúveis estão associadas

				ao aparecimento de pré-eclâmpsia.
ROR2	Tian <i>et al.</i> , 2016	5	4 intragênicas e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Proteína receptora tirosina quinase e proteína transmembrana do tipo I que pertence a subfamília ROR dos receptores de superfície celular. A proteína pode estar envolvida na formação inicial dos condrócitos e pode ser necessária para o desenvolvimento da cartilagem e no crescimento. Mutações nesse gene podem causar braquidactilia tipo B (distúrbio esquelético caracterizado por hipoplasia/aplasia de falanges distais e unhas). Além disso, mutações nesse gene podem causar a forma autossômica recessiva da

				síndrome de Robinow, que é caracterizada por displasia esquelética com encurtamento generalizado do osso do membro, defeitos segmentares da coluna, braquidactilia e uma aparência facial dismórfica.
SPEG	Tian <i>et al.</i> , 2016	5	4 intragênicas e 1 inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família das quinases de cadeia leve da miosina. Esta família de proteínas é necessária para o desenvolvimento do citoesqueleto miocitário.
PRKCB	Tian <i>et al.</i> , 2016	1	Intragênica	Proteína quinase C. Família de proteínas quinases específicas de serina e treonina que podem ser ativadas pelo cálcio e

pelo segundo mensageiro diacilglicerol. Os membros da família PKC fosforila uma ampla variedade de alvos protéicos e são conhecidos por estarem envolvidos em diversas vias de sinalização celular. Elas também servem como receptores principais de ésteres de forbol, uma classe de promotores tumorais. Cada membro da família PKC tem um perfil de expressão específico e acredita-se que desempenha um papel distinto nas células. Foi relatado que esta proteína quinase está envolvida em muitas funções celulares diferentes, tais como ativação de células B, indução de

				apoptose, proliferação de células endoteliais e absorção de açúcar no intestino.
<i>DCLK1</i>	Tian <i>et al.</i> , 2016	1	Intragênica	Membro da superfamília das proteínas quinases e da família das duplacortinas. A proteína codificada está envolvida em vários processos celulares diferentes, incluindo migração neuronal, transporte retrógrado, apoptose neuronal e neurogênese. Este gene é regulado positivamente pelo fator neurotrófico derivado do cérebro e está associado à memória e às habilidades cognitivas gerais.

<i>EPHB1</i>	Tian <i>et al.</i> , 2016	1	Intragênica	<p>Família de receptores de efrinas. Os receptores de efrina constituem o maior subgrupo da família de receptores de tirosina quinase (RTK). A proteína codificada por este gene é um receptor para os membros da família da efrina-B e medeiam numerosos processos de desenvolvimento, particularmente no sistema nervoso.</p>
<i>ZEB1</i>	Zhang <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	<p>Fator de transcrição com motivo dedo de zinco. A proteína codificada desempenha um papel na repressão transcricional da interleucina 2. Mutações neste gene têm sido associadas à distrofia corneana polimorfa posterior 3 e à distrofia</p>

				corneana endotelial de Fuchs tardia.
<i>SRCIN1</i>	Zhang <i>et al.</i> , 2017	7	6 intragênicas e 1 intergênica	Atua como um regulador negativo da proteína serina-tirosina quinase (SRC), inibindo a sua atividade. Regula a morfologia da espinha dendrítica. Envolvido na exocitose dependente de cálcio. Pode desempenhar um papel na liberação de neurotransmissores ou na manutenção de sinapses. Auxilia na migração celular.
<i>MALAT1</i>	Zhang <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Codifica um ncRNA semelhante a um tRNA. Este transcrito é retido no núcleo onde pode formar estruturas moleculares úteis em complexos de

				ribonucleoproteínas. Pode atuar como um regulador transcricional de vários genes, incluindo genes envolvidos na metástase do câncer, migração celular e na regulação do ciclo celular. Sua alta expressão em múltiplos tecidos cancerosos tem sido associada à proliferação e metástase de células tumorais.
MGRN1	Zhang <i>et al.</i> , 2017	7	Intragênicas	Proteína ubiquitina-quinase. Desempenha um papel na regulação do tráfico de endossomos para lisossomos e atua em várias vias de sinalização.
NSUN2	Zhang <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de	Metiltransferase que atua nos

			transcrição e estende para o corpo gênico	precursores de tRNA contendo íntrons. Esta modificação é necessária para estabilizar o emparelhamento anticódon-códon e traduzir corretamente o mRNA.
FOXJ1	Zhang <i>et al.</i> , 2017	2	1 intragênica e 1 intragênica se estendendo depois da terminação do gene	Membro da família <i>forkhead</i> de fatores de transcrição. Polimorfismos nesse gene estão associados ao lúpus eritematoso sistêmico e à rinite alérgica.
HDAC2	Zhang <i>et al.</i> , 2017	1	Intragênica se estendendo depois da terminação do gene	Histona deacetilase. Esta proteína forma complexos repressores transcricionais por associação com muitas proteínas diferentes, incluindo YY1, um fator de transcrição de dedo de zinco de mamífero. Assim, desempenha um papel

				importante na regulação transcricional, na progressão do ciclo celular e nos eventos de desenvolvimento.
WNT1	Zhang <i>et al.</i> , 2017	1	Inicia a montante (-2Kb) do sítio de início de transcrição e estende para o corpo gênico	Membro da família de genes WNT que codifica genes relacionados a proteínas de sinalização. Estas proteínas foram relacionadas na oncogênese e em vários processos do desenvolvimento, incluindo a regulação da morfologia celular e padronização durante a embriogênese.
SOX18	Zhang <i>et al.</i> , 2017	2	1 intergênica e 1 intragênica	Membro da família de fatores de transcrição SOX, envolvidos na regulação do desenvolvimento embrionário e na diferenciação celular. Pode atuar como

				<p>um regulador transcricional após formar um complexo proteico com outras proteínas. Esta proteína desempenha um papel no desenvolvimento de cabelo, vasos sanguíneos e vasos linfáticos. Mutações nesse gene têm sido associadas com formas recessivas e dominantes da hipotireoidite-linfedema-telangiectasia.</p>
SP3	Zhang <i>et al.</i> , 2017	1	Intragênica	<p>Família de genes relacionados a Sp1. Esta proteína contém um motivo de ligação de dedo de zinco que se liga ao DNA e vários domínios de transativação. Foi relatado como um fator de transcrição bifuncional que estimula ou reprime a</p>

				transcrição de numerosos genes.
HOXA11	Zhang <i>et al.</i> , 2017	3	2 intergênicas, 1 intragênica	Família de genes <i>homeobox</i> . Essa família de fatores de transcrição é altamente conservada e desempenham um papel importante na morfogênese de todos os organismos multicelulares. A expressão dessas proteínas é temporalmente regulada durante o desenvolvimento embrionário. Este gene está envolvido na regulação do desenvolvimento uterino e é necessário para a fertilidade feminina.
HOXC11	Zhang <i>et al.</i> , 2017	1	Intragênica	Família de genes <i>homeobox</i> . Essa família é altamente conservada de fatores de transcrição que desempenham um papel

				importante na morfogênese de todos os organismos multicelulares. O produto deste gene se liga a um elemento promotor da hidrolase lactase-florizina. Também pode desempenhar um papel no desenvolvimento intestinal precoce.
CAV1	Chae <i>et al.</i> , 2016	2	Intragênica	Principal componentes de cavéolas encontradas na maioria das células. O gene é um candidato a supressor de tumor e um regulador negativo da cascata de quinase ativada por mitógeno Ras-p42/44.
PTRF	Chae <i>et al.</i> , 2016	3	Intragênicas	Proteína que regula a transcrição de rRNA promovendo a dissociação de

				<p>complexos de transcrição e o reinício da polimerase I em transcritos de rRNA nascentes. Acredita-se que ela desempenha um papel crítico na formação de cavéolas e na estabilização de caveolinas. Mutações neste gene resultam em um distúrbio caracterizado por lipodistrofia generalizada e distrofia muscular.</p>
TGFB1	Chae <i>et al.</i> , 2016	3	2 intragênica e 1 intragênica se estendendo depois da terminação do gene	<p>Ligante secretado da superfamília das proteínas TFG-beta. Os ligantes dessa família ligam-se a vários receptores de TGF-beta, conduzindo o recrutamento e ativação de fatores de transcrição da família SMA, que regulam a expressão</p>

				<p>gênica. Esta proteína regula a proliferação, diferenciação e crescimento celular, e pode modular a expressão e ativação de outros fatores de crescimento, incluindo interferon gama e fator de necrose tumoral alfa. Este gene é frequentemente regulado em células tumorais, sendo superexpressos e as mutações neste gene resultam na doença de Camurati-Engelmann.</p>
GDF15	Chae <i>et al.</i> , 2016	2	Intragênicas	<p>Ligante secretado da superfamília das proteínas TFG-beta. Os ligantes dessa família ligam-se a vários receptores de TGF-beta, conduzindo o recrutamento e ativação de fatores de transcrição da</p>

				<p>família SMAD, que regulam a expressão gênica. Esta proteína é expressa em vários tipos celulares e atua em resposta ao estresse da célula após alguma lesão. Níveis elevados dessa proteína está associada a hipóxia tecidual, inflamação e estresse oxidativo.</p>
PALLD	Chae <i>et al.</i> , 2016	3	Intragênicas	<p>Proteína do citoesqueleto, necessária na organização do citoesqueleto de actina. A proteína é um componente dos microfilamentos contendo actina, e está envolvida no controle da forma, adesão e contração da célula. Os polimorfismos nesse gene estão associados à suscetibilidade ao câncer de pâncreas</p>

				tipo 1 e também ao risco de infarto do miocárdio.
ABCB1	Spitzwieser <i>et al.</i> , 2016	3	Intragênicas	<p>Proteína associada a membrana. Membro da superfamília de transportadores de cassete de ligação de ATP. Essas proteínas transportam várias moléculas através de membranas intracelulares e extracelulares. Membro da subfamília MDR/TAP, estando envolvida na resistência a múltiplas drogas. A proteína codificada por este gene é uma bomba de efluxo de drogas dependente de ATP para compostos xenobióticos com ampla especificidade ao substrato. É responsável pela</p>

				diminuição do acúmulo de drogas em células multirresistentes e muitas vezes medeia o desenvolvimento de resistência a drogas anticâncer. Esta proteína também funciona como um transportador na barreira hematoencefálica.
ABCC1	Spitzwieser <i>et al.</i> , 2016	1	Intragênica	Proteína da superfamília de transportadores de cassete de ligação de ATP. Essas proteínas transportam várias moléculas através de membranas intracelulares e extracelulares. Este transportador completo é um membro da subfamília MRP que está envolvida na resistência a múltiplas drogas. Esta proteína funciona como um transportador

				<p>multiespecífico de ânions orgânicos, com glutathione oxidada, leucotrienos cisteínicos e aflatoxina B1 ativada. Também transporta glicuronídeos e sulfato conjugados de hormônios esteróides e sais biliares.</p>
ABCG2	Spitzwieser <i>et al.</i> , 2016	1	Intragênica	<p>Proteína associada a membrana. Membro da superfamília de transportadores de cassete de ligação de ATP. Essas proteínas transportam várias moléculas através de membranas intracelulares e extracelulares. Alternativamente referida como uma proteína de resistência ao câncer de mama, ela funciona como um</p>

				<p>transportador xenobiótico que pode desempenhar um papel importante na resistência a múltiplas drogas. Provavelmente serve como um mecanismo de defesa celular em resposta à exposição a mitoxantrona e antraciclina.</p>
--	--	--	--	---

