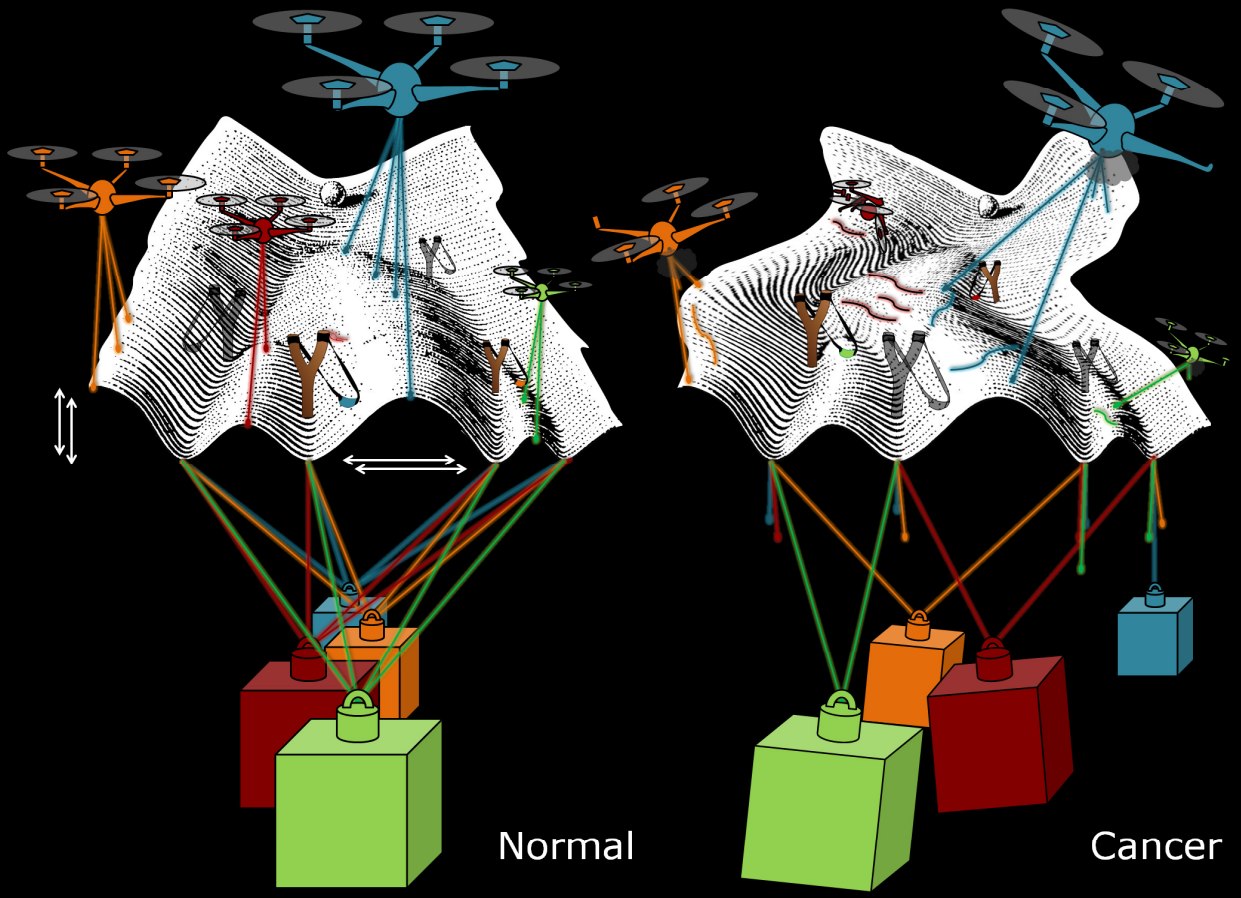


Humberto Jorge Gomes Ferreira

Epigenetic regulation of non-coding RNAs in cancer

Tese de doutoramento em Ciências Farmacêuticas, especialidade Biologia Celular e Molecular, orientada pelo Professor Doutor Manel Esteller Badosa e pela Professora Doutora Maria Celeste Fernandes Lopes e apresentada à Faculdade de Farmácia da Universidade de Coimbra

Julho 2016

· U C ·

Universidade de Coimbra

Humberto Jorge Gomes Ferreira, PharmD, MSc

Epigenetic regulation of non-coding RNAs in cancer

Thesis in Pharmaceutical Sciences, specialty of Cellular and Molecular Biology, presented to the Faculty of Pharmacy of the University of Coimbra as a requirement for the PhD degree.

Thesis Supervisors:

Manel Esteller Badosa, M.D., Ph.D.

Maria Celeste Fernandes Lopes, Ph.D.

Coimbra, July 2016

Cover Illustration Credits:

"*Conrad Hal Waddington, The strategy of the genes, 1957*"

Adapted by Humberto Jorge Gomes Ferreira

On the cover:

Metaphor explaining the disruption of the genetic (weights) and epigenetic (drones) equilibrium in cancer, leading consequently to an altered expression of non-coding RNAs (slingshots) that, in turn, contribute to the reprogramming of malignant cells.

UNIVERSIDADE DE COIMBRA

Tese de doutoramento em Ciências Farmacêuticas, especialidade Biologia Celular e Molecular, orientada pelo Professor Doutor Manel Esteller Badosa do Instituto de Investigación Biomédica de Bellvitge (Barcelona, Espanha) e pela Professora Doutora Maria Celeste Fernandes Lopes da Faculdade de Farmácia da Universidade de Coimbra.

Title (English)	Epigenetic regulation of non-coding RNAs in cancer

---

Título (Português)	Regulação Epigenética de RNAs não codificantes em cancro

---

Author (Autor)	Humberto Jorge Gomes Ferreira
Supervisor (Orientador)	Prof. Doutor Manel Esteller Badosa Instituto de Investigación Biomédica de Bellvitge Barcelona, Espanha
Co-Supervisor (Co-orientador)	Prof. Doutora Maria Celeste Fernandes Lopes Faculdade de Farmácia, Universidade de Coimbra, Coimbra, Portugal

Year (Ano)	2016

The work presented in this doctoral thesis was performed at the Cancer Epigenetics Group, from the Cancer Epigenetics and Biology Program, at the Bellvitge Biomedical Research Institute, supervised by the Professor Doctor Manel Esteller Badosa. The doctoral studies were performed according to the PhD Programme in Experimental Biology and Biomedicine, from the Centre for Neurosciences and Cell Biology, at the University of Coimbra, Portugal, under the supervision of the Professor Doctor Maria Celeste Fernandes Lopes from the Faculty of Pharmacy of the University of Coimbra.

Financial support was obtained from the Portuguese Foundation for Science and Technology (FCT) in the form of a PhD fellowship addressed to the author (SFRH/BD/33887/2009).

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

À minha mãe,

“*Eles não sabem, nem sonham,*

que o sonho comanda a vida.

Que sempre que um homem sonha

o mundo pula e avança

como bola colorida

entre as mãos de uma criança.”

António Gedeão

Acknowledgements

One day someone said that "whenever a man dreams, the world jumps and moves, as a colored ball in the hands of a child". Moved by their curiosity children look for challenges every day. Over the last years I was challenged by the complexity of a simple cell, aiming unconsciously to cure cancer. Instead of satisfy my curiosity and help “others”, curing their cancer, I became even more curious and “others” helped me over my first steps in science. And here I am, after a long walk where many people have contributed to my personal and professional enrichment, becoming part of who I am and to those I would like to dedicate some lines.

First of all, I would like to thank the PDBEB PhD Programme and Fundação para a Ciência e a Tecnologia (FCT), for trusting and giving me this opportunity.

I would like to thank **Manel**, my PhD thesis director, who trusted me from the beginning, giving me the great opportunity to form part of his team, presenting me some challenges and giving me freedom to think and learn. À **Professora Celeste**, a minha co-directora de tese, o meu sincero muito obrigado, por toda a sua simpatia e apoio que me foi dando, transmitindo-me sempre motivação e força para continuar em frente.

To my unofficial supervisor **Holger**, who introduced me the world of Epigenetics and non-coding RNAs, thanks for being always there, giving me a lot of opportunities to learn and explore this unknown world. I really appreciated not only to work and discuss about science with you, but also to have fun outside of the lab.

Going back on time, many other people importantly contributed to enter into this challenging world. Inicialmente gostaria de agradecer ao **Professor Luís de Almeida** que foi quem decisivamente me cativou para este mundo, dando-me sempre o seu apoio. Je voudrais aussi remercier au **Professeur Patrick Couvreur** et au **Dr. Harivardhan**, avec qui « j'ai guéri le cancer ». Au moins, c'est ce que je pensais, au début. Outra oportunidade para conhecer este mundo foi-me dada pelo **Professor João Pinto** ao qual também gostaria de agradecer.

Obrigado a todos os meus amigos mais próximos da Faculdade que sempre me deram força para continuar, apesar da distância. Especialmente, obrigado **Sandra** por sempre teres acreditado em mim, me teres ajudado, apoiado e compreendido.

Al caer de paracaídas en el PEBC sin conocer a nadie, me ha recibido un grupo de amigos que han hecho que mi experiencia aquí pudiera ser inolvidable y muy enriquecedora. Muchas de estas personas, aunque ya esparcidas por todo el mundo, afortunadamente seguirán siendo las mismas de siempre. La única barrera serán las distancias físicas, pero nos iremos encontrando por el mundo. **Javi** y **José**, echo mucho de menos que los gatos, monos y elefantes se escapasen del zoo en los finales de tarde aburridos. Me habéis ayudado muchísimo y seréis siempre referencias para mí a todos los niveles. **Paolo** y **Jordi**, después de las pausas para reírnos un poco, todo es más animado y sale mejor. Pena que os habéis ido muy rápido. "Pequeña" **Alexiaaaaa**, aunque te cambies de continente para no molestarte, sabes que me encanta por eso creo que no te librarás de mí así tan fácil, muchas gracias por todo. Y **Julica**, ya sabes que me has dejado marcado… menudos sustos. **Carmencita**, gracias por tu amistad y por te preocupares conmigo queriendo sacarme del lab... yo también me preocupo por ti y tengo que avisarte que los vegetales están preparando una revolución contra ti, así que cuídate. **Vanesa**, contigo casi me enfado porque llegas y te vas y yo sigo en el mismo sitio. No olvidaré tus consejos prácticos y que *lo perfecto es lo contrario de lo bueno*. Sé que debería hacerte más caso. Gracias por todos los momentos que hemos compartido. **Laia**, gracias por tu sonrisa, por hacerme sonreír y por todo lo que hemos compartido.

Montse y **Pere**, mis fieles amigos de la gula, muchas gracias por no fallarme nunca en nuestros merecidos momentos de reposición de energías. Montse, gracias por ser tan atenta con el "pegotito" en el paki. Pere, ya sabes que me encantan nuestras teorías frikis, hay que seguir con eso…puede que alguna fructifique. Aunque estoy seguro que cambiarás ese tipo de teorías por tus grandes descubrimientos. **Rute** Marlene, eu "olho para a direita" e tu acabas de chegar, eu "olho para a esquerda" e tu acabas de ir, como têm de ser as coisas. Obrigado pelos conselhos e sábias verdades que sempre dizes.

Miguel, **Paula**, **Fer**, tengo que deciros que por mejores que hayan sido nuestras cenas en el lab, me gusta más cenar con vosotros "en la calle". Sois el mejor ejemplo del buen trabajo de equipo, haciendo el posible y el imposible… no fuera la barba caerme después, hasta pensaría en repetirlo... Miguel, eres un crack, un buen amigo y disfruté muchísimo trabajando y discutiendo ciencia contigo. Gracias por haber confiado en mi trabajo y compartido conmigo tus experiencias dentro y fuera del lab. Eco-Paulinha,

comparto tu visión ante la mayoría de las (des)igualdades. Gracias por estar siempre lista para ayudar a los demás. GRANDE Fer, eres una de las mejores personas que conozco y soy un afortunado por haberte conocido. Aparte llevas dentro unas cuantas enciclopedias y aprendí mucho contigo. Muchas gracias por todo.

Martita K, me acuerdo lo bien que me has enseñado a extraer RNA después de que hubieras estado pintando flores en el lab, luego te busqué para que me ayudases con la cuanti y ya no estabas. **Paloma**, tu igual, me has dejado tirado en cultivos y ya no tengo a quien molestar... Muchas gracias por todos los momentos, chicas.

Anna y **Sebas**, la verdad es que una tesis sin momentos de relax es imposible. Me habéis enseñado un mundo donde no hay estrés y que cada mundo tiene su proprio ritmo, nos vamos a bucear? Gracias por todo. To **Laia P**, **Janina** and **Melania**… Girls, I hope I was a good "teacher" and thanks a lot for all the moments we shared.

Amigos y colegas del PEBC, gracias por todos los momentos que hemos compartido, dentro y fuera del lab, y por toda la fuerza que siempre me habéis transmitido.

Aida y **Cris**, qué buena idea lo del Geocaching e ir pa'Tossa. Aunque con tanta fiesta poco me dejabais dormir, aun así esa y todas las otras veces ha merecido realmente mucho la pena. Habéis conseguido muy fácil algo que parecía casi imposible, sacarme del lab… y como lo disfruté, "madre mía, madre mía". Muchas gracias por todos eses momentos, nos vamos de fiesta?

Cátia, muito obrigado pela tua ajuda e preocupação ilimitada, tanto a nível profissional como pessoal, bem como por teres sempre pronto um abraço amigo e uma visão cor-de-rosa e optimista de tudo, mesmo nas coisas com um aspecto mais sombrio.

À minha **família** que desde sempre me deu um imenso apoio em todas as minhas opções mesmo sabendo que escolhia os caminhos mais difíceis, mas que seriam os caminhos necessários às minhas conquistas pessoais mais desejadas.

E ao meu grande amigo **café**, obrigado por me manteres acordado depois de tantas noites mal dormidas ou demasiado curtas. Mas sei que tens outro sabor... em Portugal!

Humberto

List of Contents

Index of Figures

Chapter V

Chapter VI

Index of tables

Abbreviations

#	
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine

A	
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
APC	Adenomatous polyposis coli
ASXL1	Additional sex combs like 1
ASTN2	Astrotactin 2

B	
BRAF	B-Raf proto-oncogene, serine/threonine kinase
BRCA1	Breast cancer 1, early onset
BS	Bisulfite

C	
CAV1	Caveolin-1
CGI	CpG Island
ChIA-PET	Chromatin interaction analysis by paired-end tag
ChIP	Chromatin immunoprecipitation
CI	Confidence interval
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
c-KIT	v-kit Hardy-Zuckerman 4 feline sarcoma viral oncogene homolog
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemias
CNV	Copy number variation
CpG	Cytosine-phosphate-guanine
CREB2	Cyclic AMP-responsive element-binding protein 2
CREBBP	cAMP response element-binding protein
CTNNB1	Catenin (cadherin-associated protein), beta 1

D	
DCC	Deleted in colorectal carcinoma
DKC1	Dyskerin
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
Dnmt1	DNA methyltransferase 1 (*mus musculus*)
DNMT3A	DNA methyltransferase 3A

E

EFNB2	Ephrin-B2
EGC	Embryonic germ cell
EGFR	Epidermal growth factor receptor
EHMT2	Euchromatic histone-lysine N-methyltransferase 2
ENCODE	Encyclopedia of DNA Elements
EP300	E1A binding protein p300
eRNA	Enhancer RNA
EZH2	Enhancer of zeste homolog 2

F

FAB	French-American-British
FAP	Familial adenomatous polyposis
FDR	False discovery rate
FOXQ1	Forkhead box Q1

G

GAS5	Growth arrest specific 5
GCNIS	Germ cell neoplasia in situ
GEO	Gene Expression Omnibus
GO	Gene Ontology
GPRC5A	G protein-coupled receptor, class C, group 5, member A
GRB7	Growth factor receptor-bound protein 7

H

H_xK_yac	Histone(x) lysine(y) acethylated
$H_xK_yme_n$	Histone(x) lysine(y) (n)methylated
H3R2	Histone 3 arginine 2
HAT	Histone acetyltransferase
HCC	Hepatocellular carcinoma
HDAC	Histone deacetylase
HDAC1	Histone deacetylase 1
HDM	Histone demethylase
HK2	Hexokinase 2
HMR	Hypomethylated region
HMT	Histone methyltransferase
HNPCC	Hereditary non-polyposis colon cancer
HOTAIR	HOX transcript antisense RNA

I

IAP	Intracisternal A-particle
IARC	International Agency for Research on Cancer
IDH1	Isocitrate dehydrogenase 1
IGF2	Insulin-like growth factor-2

J

JARID1	Jumonji, AT-rich interactive domain 1
JMJD2C	Jumonji domain containing protein 2C

K

KAT6B	K(Lysine) acetyltransferase 6B
KDM2B	Lysine (K)-specific demethylase 2B
KRAS	Kirsten rat sarcoma viral oncogene homolog

L

lncRNA	Long ncRNA
LOH	Loss of heterozygosity
LSD1	Lysine (K)-specific demethylase 1
LYL1	Lymphoblastic leukemia-associated hematopoiesis regulator 1

M

MBD	Methyl CpG-binding domain
MBD1	Methyl-CpG binding domain protein 1
MeCP2	Methyl CpG binding protein 2
MEF2B	Myocyte enhancer factor 2B
MIRLET7BHG	MIRLET7B host gene
MGMT	O-6-methylguanine-DNA methyltransferase
miRNA	MicroRNA
MLH	MutL homolog
MLL	Mixed lineage leukemia
MMP9	Matrix metallopeptidase 9
mRNA	Messenger RNA
MSI	Microsatellite instability
MUTYH	MutY DNA glycosylase
MYC	v-Myc avian myelocytomatosis viral oncogene homolog

N

ncRNA	Non-coding RNA
NRAS	Neuroblastoma RAS viral (V-Ras) oncogene homolog
NSD1	Nuclear receptor binding SET domain protein 1
NSE	Non-seminoma

O

OR	Odds ratio
OS	Overall survival
ox-BS	Oxidative bisulfite

P

PGC	Primordial germ cell

PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
piRNA	Piwi-interacting RNA
PIWIL1	Piwi like RNA-mediated gene silencing 1
PML-RARA	Promyelocytic leukemia-retinoic acid receptor alpha fusion oncoprotein
PMS2	PMS1 homolog 2

R

RARβ2	Retinoic acid receptor β2 gene
RB1	Retinoblastoma 1 gene
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
RNF43	Ring finger protein 43
R-RAS	Related RAS viral oncogene homolog
rRNA	Ribosomal RNA

S

S100A4	S100 calcium binding protein A4
SCLC	Small cell lung cancer
SE	Seminoma (*except Chapter VI*)
SE	Super-Enhancer (*only Chapter VI*)
SEER	Surveillance, Epidemiology, and End Results
SEMA5A	Semaphorin 5A
SETDB1	SET domain, bifurcated 1
shRNA	Small hairpin RNA
SIRT1	Sirtuin 1
SLC47A1	Solute carrier family 47 member 1
SMAD2	SMAD family member 2
SMYD3	SET and MYND domain containing 3
sncRNA	Small non-coding RNA
snoRNA	Small nucleolar RNA
SNP	Single nucleotide polymorphism
snRNA	Small nuclear RNA
SRA	SET and RING-associated

T

TBC1D16	TBC1 domain family, member 16
TCGA	The Cancer Genome Atlas
TDRD1	Tudor domain containing protein 1
TERRA	Telomeric repeat containing RNA
TET2	Ten-eleven translocation oncogene family member 2
TF	Transcription factor
TFBS	Transcription factor binding site
TGCC	Testicular Germ Cell Cancer
TGFβ	Transforming growth factor beta

TP53	Tumor protein P53
tRNA	Transfer RNA
TSS	Transcription start site
TTD	Tandem tudor domain
T-UCR	Transcribed ultraconserved region

U

UHRF1	Ubiquitin-like with PHD and ring finger domains1

V

VASH1	Vasohibin 1

W

WGBS	Whole genome bisulfite sequencing
WHO	World Health Organization
WNT5A	Wingless-type MMTV integration site family, member 5A

Z

ZBTB4	Zinc finger and BTB domain containing 4
ZFAS1	ZNFX1 antisense RNA 1

Note: When more than one gene/protein from the same family is mentioned, the list contains only the first mention.

Abstract

Tumoral evolutionary process occurs through the sequential accumulation of mutations and epimutations that are responsible for cell heterogeneity and sub-clonal selection, as well as for drug resistance and patients associated mortality. Recently, diverse classes of non-coding RNAs (ncRNAs) were described to be implicated in the regulation of key players of carcinogenesis. By standard and high-throughput methods, we analyzed the epigenetic landscape of different types of cancer, uncovering cancer-related pathways, emphasizing those related to the regulation of ncRNAs.

Small nucleolar RNAs (snoRNAs) guide post-transcriptional modifications of spliceosomal and ribosomal RNAs. Some members of this class of RNAs are disrupted in cancer, where modifications in ribosome biogenesis have also been implicated. We verified that *SNORD123*, *ACA59B* and *U70C* are transcriptionally silenced by DNA hypermethylation of the CpG Island that overlaps the promoter region of their host gene. Of particular interest, *SNORD123* and *ACA59B* are conserved across vertebrates but they do not have a known target (*orphan* snoRNAs). Taking into account that these snoRNAs are expressed in normal colon and are epigenetically repressed in some colorectal cancer cell lines, we suggested that they can have a potential contribution to carcinogenesis. Moreover, we described the DNA hypermethylation of these three snoRNAs in leukemia samples.

Piwi-interacting RNAs (piRNAs) are mainly expressed in germline cells, playing a key role in the epigenetic silence of transposons or guiding their cleavage. We reported the epigenetic transcriptional inactivation of the genes encoding the piRNA-related proteins, *PIWIL1*, *PIWIL2*, *PIWIL4* and *TDRD1*, in both seminomas and non-seminomas. These epigenetic lesions occur in a context of piRNA downregulation and loss of DNA methylation at LINE-1 loci. Importantly, recent studies had shown a similar epigenetic transcriptional disruption in other cancer types; and in non-genetic infertility syndromes, that are epidemiologically linked with testicular cancer.

To better characterize the epigenetic landscape of a cancer cell, we interrogated the entire methylome in several cancer and normal samples. We first established the methylome of two acute myeloid leukemia (AML) cell lines, OCI-AML5 and OCI-AML3, the latter harboring a missense mutation in *DNMT3A*, present in ~20% of the

AML patients. By comparison with the methylation profile of AML samples, we identified a set of twelve differentially methylated candidate target loci for DNMT3A in AML, validating their transcriptional reactivation in our cell line model. Thus, the leukemogenic gene *MEIS1* was actively expressed in OCI-AML3. By screening the highest-ranked differentially methylated regions that potentially regulate non-protein coding genes we described a signature of four hypomethylation-associated transcriptional reactivated ncRNAs in the *DNMT3A* mutant cell line, namely *ENST00000413346*, *LOC100506585*, *ENST00000443490* and *MIRLET7B host gene* (*MIRLET7BHG*). We also suggested that some of the DNMT3A potential target genes could be linked to worst prognosis observed in AML patients harboring the *DNMT3A* mutation, particularly *MEIS1* and the host gene that carries both *let-7a-3* and *let-7b*. These two microRNAs were previously described to be overexpressed in AML.

Based on the loss of differentiation in cancer cells and expanding our study to other tissues, we interrogated the methylation profile of genomic regions known to be responsible for cell identity, namely super-enhancers. We established a correlation among tumor-related hypermethylation of super-enhancers and transcriptional silencing of the corresponding related genes. Our results showed that their methylation profile is also associated with specific cancer types. However, the methylation of the super-enhancer that regulates the host gene of *let-7a-3* and *let-7b* tumor suppressors was linked to their silencing in both lung and breast epithelial cancers. In colorectal cancer, we described tumor-related super-enhancers undergoing hypomethylation-related transcriptional activation of the related genes, such as *MYC* and *RNF43* oncogenes. We hypothesized that the impaired expression and binding of transcription factors could establish novel super-enhancers. We identified FOXQ1 as a probable transcription factor responsible for the loss of DNA methylation at colorectal cancer-specific super-enhancers that control *MYC* and *RNF43*.

DNA methylomes highlight the epigenetic landscape that regulates the expression of key players in cancer biology. Some of these players are non-coding RNAs that should be exploited as biomarkers for cancer diagnosis, or as targets in personalized therapeutic approaches to control tumor progression and/or metastasis.

Keywords: Epigenetics, DNA methylation, Non-coding RNAs, Small-nucleolar RNAs, Piwi-interacting RNAs, Super-enhancers

Resumo

O processo evolutivo de um tumor é feito através da acumulação sequencial de mutações e epimutações, sendo estas responsáveis pela heterogeneidade celular e por uma seleção sub-clonal, bem como pela resistência dos doentes a fármacos e mortalidade associada. Recentemente, diversas classes de RNAs não-codificantes (ncRNAs) foram implicadas na regulação de elementos-chave da carcinogénese. Através de métodos *standard* e de *high-throughput*, analisámos o perfil epigenético de diferentes tipos de cancro, encontrando vias transcripcionais alteradas, dando especial ênfase às vias relacionadas com a regulação dos ncRNAs.

Os *small nucleolar RNAs* (snoRNAs) dirigem modificações pós-transcripcionais de RNAs spliceossomais e ribossomais. Alguns membros desta classe de RNAs estão desregulados em cancro, onde modificações na biogénese do ribossoma também têm sido implicadas. Verificámos que os snoRNAs *SNORD123*, *ACA59B* e *U70C* estão silenciados transcripcionalmente por um aumento da metilação do DNA da *ilha CpG* sobreposta à região promotora do seu gene hospedeiro. É de destacar que os *SNORD123* e *ACA59B* estão conservadas em vertebrados, mas não têm um alvo conhecido (*snoRNAs órfãos*). Tendo em conta que estes são expressos em cólon saudável e que estão epigeneticamente silenciados em algumas linhas celulares de cancro colorrectal, sugerimos que também podem contribuir para o processo de carcinogénese. Além disso, o facto de termos detectado um aumento de metilação do DNA correspondente a estes três *snoRNAs* em amostras de leucemia, reforça a nossa teoria.

Os *piwi-interacting RNAs* (piRNAs) são expressos principalmente em células germinativas, desempenhando um papel fundamental no silenciamento epigenético de transposons ou dirigindo a sua clivagem. Os nossos resultados demonstram o silenciamento epigenético da transcripção dos genes que codificam para as proteínas relacionadas com os *piRNAs*, nomeadamente *PIWIL1*, *PIWIL2*, *PIWIL4* e *TDRD1*, tanto em seminomas como em não-seminomas. Estas lesões epigenéticas ocorrem num contexto de baixos níveis de *piRNA*s e de perda de metilação do DNA nas regiões correspondentes ao LINE-1. De forma semelhante, estudos recentes descrevem o silenciamento epigenético da transcrição noutros tipos de cancro; e em síndromes não genéticos de infertilidade, neste caso epidemiologicamente associados ao cancro de testículo.

De maneira a caracterizar melhor o perfil epigenético de uma célula cancerígena, estudámos o metiloma de vários tipos de cancro e de amostras de tecido normal. Em primeiro lugar, estabelecemos o metiloma de duas linhas celulares de leucemia mieloide aguda (AML), OCI-AML5 e OCI-AML3, a última das quais tem uma mutação *missense* no gene *DNMT3A*, estando presente em ~ 20% dos doentes com AML. Comparando o perfil de metilação do DNA de amostras de AML, identificámos doze regiões diferencialmente metiladas candidatas à ação da DNMT3A, validando a reativação da sua transcrição no modelo de linhas celulares. Deste modo, vimos que o gene leucemogénico *MEIS1* se expressa ativamente em OCI-AML3. Analisando as regiões com maiores diferenças a nível de metilação de DNA e que pudessem potencialmente regular genes não codificantes para proteínas, detetámos a existência de quatro ncRNAs, *ENST00000413346*, *LOC100506585*, *ENST00000443490* e *MIRLET7BHG*, associados a uma reativação transcripcional devido à perda de metilação do DNA na linha celular portadora da mutação no gene *DNMT3A*. Sugerimos que alguns dos potenciais alvos da DNMT3A, poderiam estar relacionados com um pior prognóstico em pacientes com AML que são portadores da mutação no gene *DNMT3A*, especialmente *MEIS1* e o gene hospedeiro que alberga *let-7a-3* e *let-7b*, tendo já sido descrito que estes microRNAs exibem maior expressão em AML.

Tendo presente a perda de diferenciação das células cancerígenas e querendo expandir o nosso estudo a outros tecidos, investigámos o perfil de metilação de regiões descritas como responsáveis pela identidade celular, os *super-enhancers*. No contexto tumoral encontrámos uma correlação entre o aumento de metilação do DNA dos *super-enhancers* e o silenciamento transcripcional dos genes correspondentes. Apesar do perfil de metilação dos *super-enhancers* ser específico do tipo de cancro, a metilação do *super-enhancer* que regula o gene hospedeiro dos supressores tumorais *let-7a-3* e *let-7b* foi associada ao seu silenciamento, tanto em cancro de pulmão como em cancro de mama, ambos epiteliais. No caso do cancro colorrectal, descrevemos *super-enhancers* submetidos a uma perda de metilação associada à ativação transcripcional dos oncogenes *MYC* e *RNF43*. Estabelecemos ainda a teoria de que a expressão e ligação desreguladas de fatores de transcripção poderiam promover a formação de novos *super-enhancers*. Identificámos FOXQ1 como um provável fator de transcripção, responsável pela perda de metilação dos *super-enhancers* específicos de cancro colorrectal e que controlam *MYC* e *RNF43*.

Os metilomas de DNA destacam o perfil epigenético que regula a expressão de elementos-chave na biologia do cancro. Alguns destes elementos são ncRNAs que deveriam ser explorados como biomarcadores para diagnóstico de cancro e também como alvos em estratégias de terapia personalizada, com vista ao controlo da progressão tumoral e/ou das metástases.

Palavras-chave: Epigenética, Metilação do DNA, RNAs não-codificantes, *small nucleolar RNAs*, *piwi-interacting RNAs*, *super-enhancers*

CHAPTER I

General Introduction

1. Cancer

The generic term *cancer* is related to a large group of pathological conditions associated to the abnormal and uncontrolled growth of cells. Also called *malignant tumor* or *neoplasm*, it is due to the active transformation of normal cells into highly malignant successors. They disturb the homeostasis between the different types of cells among any tissue or organ[1-3]. The uncontrolled ability to divide leads to the emergence of large populations of cells which no longer follow the standard principles that regulate the normal tissue construction and maintenance, such as cell differentiation, growth and programmed death. A hallmark of this malignant disease is the ability to invade adjoining tissues. Cancer cells, outside the limits of the normal tissue from which they derive, can enter into the circulation and spread to distant organs and eventually establish secondary tumors. These secondary tumors, called metastases, are the main cause of death from cancer[1].

1.1 Historic perspective

Cancer is a disease as old as man, haunting our lives from the antiquity. The actual increase in life expectancy accentuates even more this man's oldest foe that was aptly considered a *"monster more insatiable than the guillotine"*[4]. Its definition comes from earliest civilizations and was continuously updated, in the light of new knowledge derived from observational studies and from the tireless look for new effective treatments.

The origin of the word *cancer* is attributed to Hippocrates (460-375BC), the often called *father of medicine*. He used the terms *carcinos* and *carcinoma* to distinguish between malignant non-ulcerating and malignant ulcerating tumors, respectively. *Carcinos* means *crab* in Greek, whereas the equivalent Latin word is *cancer*. At that time, Hippocrates described and drew externally visible tumors, since it was not allowed to open the body, according to Greek traditions. Breast cancer was not treatable then, but it was one of the few types of tumor that could be seen outwardly. Probably he used the term *crab* to highlight the similarities found between the continuous growth of this tumor, accompanied by swollen finger-like spreading veins, and the silhouette of a crab[5]. The concept *scirrhos* (*scirrus* in Latin) was also introduced by this physician to

indicate tumor rigidity on palpation and he divided this pathology into benign and malign. He also discriminated between superficial and deep carcinomas, considering that they should be treated as singular entities. In terms of therapy, at that time it was suggested that tumors that were not treatable by medicine should be cured by "*iron*" (knife extraction). Likewise, those that couldn't be cured by iron should be cured by "*fire*" (cautery). Those not cured by any option referred before were considered untreatable. He also introduced the concept of palliative care and indicated that any treatment of occult or deep-seated tumors could decrease the survival period of a patient[6].

Aulus Celsus (25BC-50AD), a roman physician translated the Greek term *carcinos* into *cancer*. He described a variety of superficial cancers and identified the differences between cancer and non-malignant cellular proliferations[7]. He highlighted the involvement of the axillary glands in breast cancer and suggested the potential dissemination of primary tumors[8]. Later on, the Greek physician, Claudius Galen (130-200AD), attributed the word *oncos* (the Greek term for *swelling*) to describe tumors (origin of the term Oncology) and introduced the term *sarcoma* for tumors of raw meat (in Greek *sarkos*)[8]. Breast cancer started to be treated by total mastectomy, being reported by Aetius (527-565); and Lanfranc (1252-1315) described the differences between benign and malignant breast tumors[6-9].

Historically, the practice of autopsy, the illustration of tumors and the introduction of microscopy were essential for the understanding of cancer. Concomitantly, new theories have emerged: the cellular theory by Robert Hooke (1635-1703), the Blastema Theory by Johannes Müller (1801-1858) and the Chronic Irritation Theory by Rudolf Virchow (1821-1902)[8,10-13].

The term metastasis introduced by Joseph Recamier (1774-1852) gained a very important relevance after the confirmation that malignant cells could migrate from the original tumor and metastize (Karl Thiersch, 1822-1895). This process was thought to take place through lymphatic vessels except for sarcomas, where malignant cells should enter the bloodstream (Theodor Billroth, 1829-1894). Meanwhile, John Birkett (1815-1904) described the process by which malignant epithelial cells initiate the metastatic process, calling it *microinvasion*[11,14,15].

The impressive technological breakthrough of the 20th century allowed an extraordinary jump in the scientific knowledge. In oncology, this exponential development allowed scientists to growth tumor cells artificially (Alexis Carrel, 1873-1944) and Theodor Boveri (1862-1915) suggested that "*malignant tumors might be the result of a certain abnormal condition of the chromosomes, which may arise from multipolar mitosis"*, setting the basis for the Chromosomal Theory of Cancer[16,17]. In 1953, the discovery of the deoxyribonucleic acid (DNA) structure by James Watson and Francis Crick[18] trigger a huge impetus in the molecular biology and genetics worlds. This impulse originated two of the biggest discoveries in cancer field. The discovery of oncogenes, highlighting their malignant potential[19], and of tumor suppressor genes which were commonly downregulated in tumors[20], gave rise to a new era in the oncologic field. Oncogenes and tumor suppressors, as well as a plethora of associated mutations driving the malignant pathway, supported the postulation of the **"***Somatic Mutation Theory"* for cancer origin. Accordingly, a single somatic cell accumulates multiple genomic mutations often affecting genes that control cell cycle and proliferation, inducing an uncontrolled hyperproliferation, initiating the tumorigenic process[21].

Importantly, in the last 30 years of cancer research, the amount of information published is such huge that overpass our capacity of integration. Therefore, cancer research area is becoming more specialized in subfields that need to be somehow connected. Nowadays, the integration of all the acquired knowledge must be addressed by multidisciplinary teams to join the different pieces of this complex puzzle.

1.2 Cancer Epidemiology

Cancer comprises currently more than 100 distinct types, with their respective subtypes of tumors[22]. Its overall health impact, prognostic outcome and limited treatments, are since centuries ago a huge concern.

In 2012, cancer figured as a leading cause of mortality and morbidity, with 14,1 million new cases and 8,2 million cancer associated deaths worldwide (13% of all deaths)[1,23]. In order of frequency, the six most mortality-associated cancers in men were lung, liver, stomach, colorectal, prostate and esophagus. In women, this list is composed by breast, lung, colorectal, cervical, stomach and liver **(Fig. 1.1)**[23,24]. It is important to emphasize

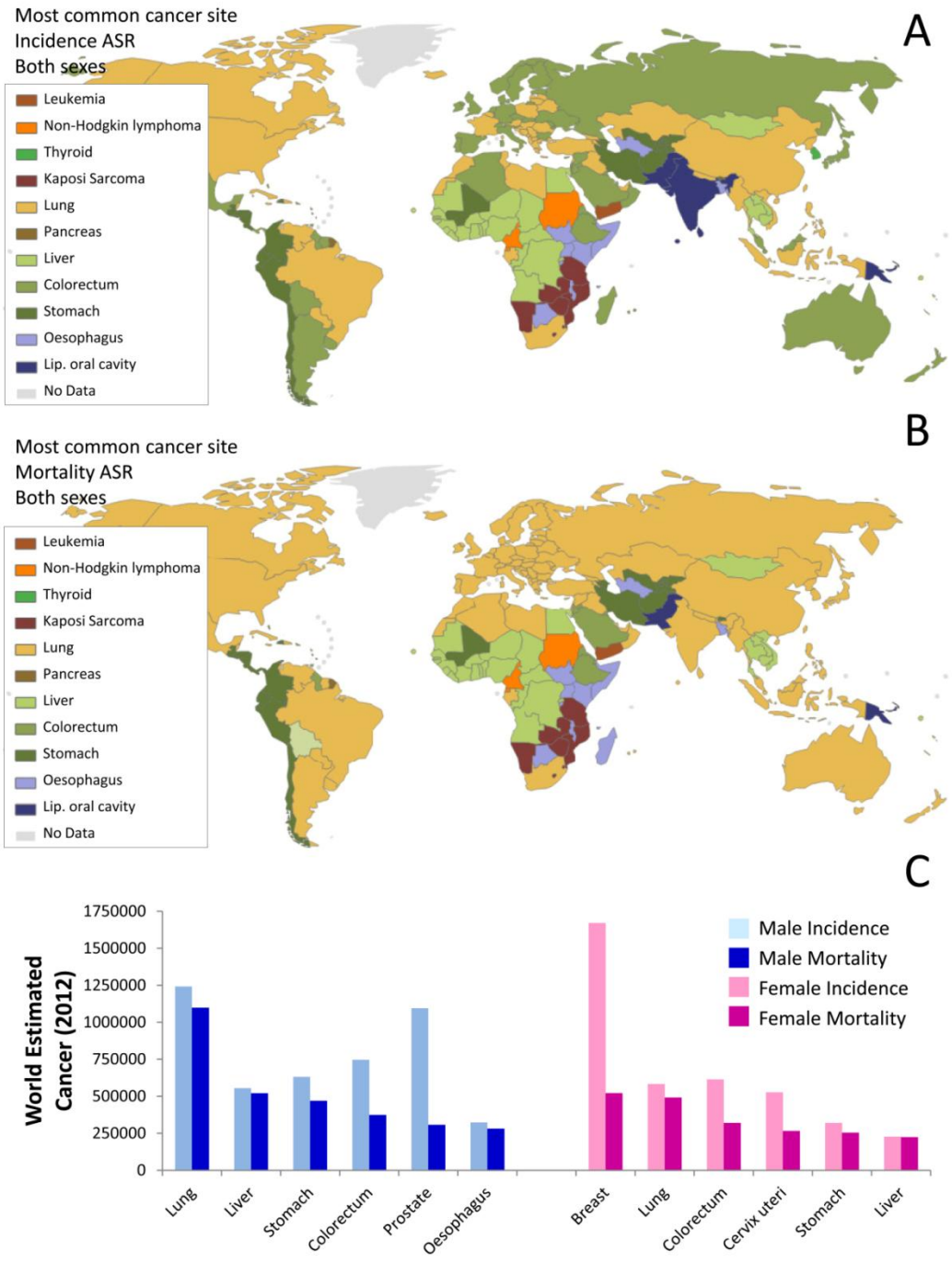


Figure 1.1. Global Cancer Incidence and Mortality. A and B: Map of the distribution of cancer incidence and mortality worldwide. **C**: Graph representing the world estimated cancer incidence and mortality cases in 2012, distributed by sex. This graph only shows the 6 types of cancer with higher mortality among men and women. *Adapted from GLOBOCAN, 2012*[23].

that cervical cancer is however the most prevalent and lethal type of cancer in some developing countries[1].

Importantly, there is a clear difference on the incidence of certain types of cancers across countries that present different economic growth. Low- and middle-income countries register about 70% of cancer deaths worldwide. They present significant differences in terms of cancer etiology, where viral and microorganism infections assume the top positions, raising the risk for cancer development[1].

1.3 Cancer Etiologies

Aiming to discover the origin of cancer, researchers prompted the establishment of several correlations between lifestyles and cancer development **(Fig. 1.2)**. Historically, the first recognized correlation was the fact that celibate nuns had almost no cervical cancer; nevertheless they had a normal incidence of breast cancer. Accordingly, Bernardino Ramazzini (1633-1714) hypothesized that female hormonal cycles and sexual hormones could predispose them to this second malignancy[25]. Nowadays the International Agency for Research on Cancer (IARC) is responsible for the release of periodical monographs evaluating the possible causes of cancer. A wide range of causes has been categorized, generally with a very specific exposed subpopulation. Chemicals (e.g. formaldehyde), personal habits (e.g. tobacco smoking) or viral infections (e.g. hepatitis B virus, HBV) are some of these examples.

1.3.1 Genetics

The incidence of particular types of cancer within families led to the establishment of some genetic correlations. For instance, the hereditary predisposition for gastric, breast and some other types of cancer was hypothesized and later confirmed with the discovery of oncogenes and tumor suppressor genes[15]. In the case of retinoblastoma, the presence of an inherited mutation was described to predispose for a second one, ultimately leading to the onset of the tumor[20].

1.3.2 Environmental and Lifestyle

In developed countries, tobacco and diet are associated to almost 50% of cancer incidence. Nevertheless, it is important to refer some almost exclusive factors that

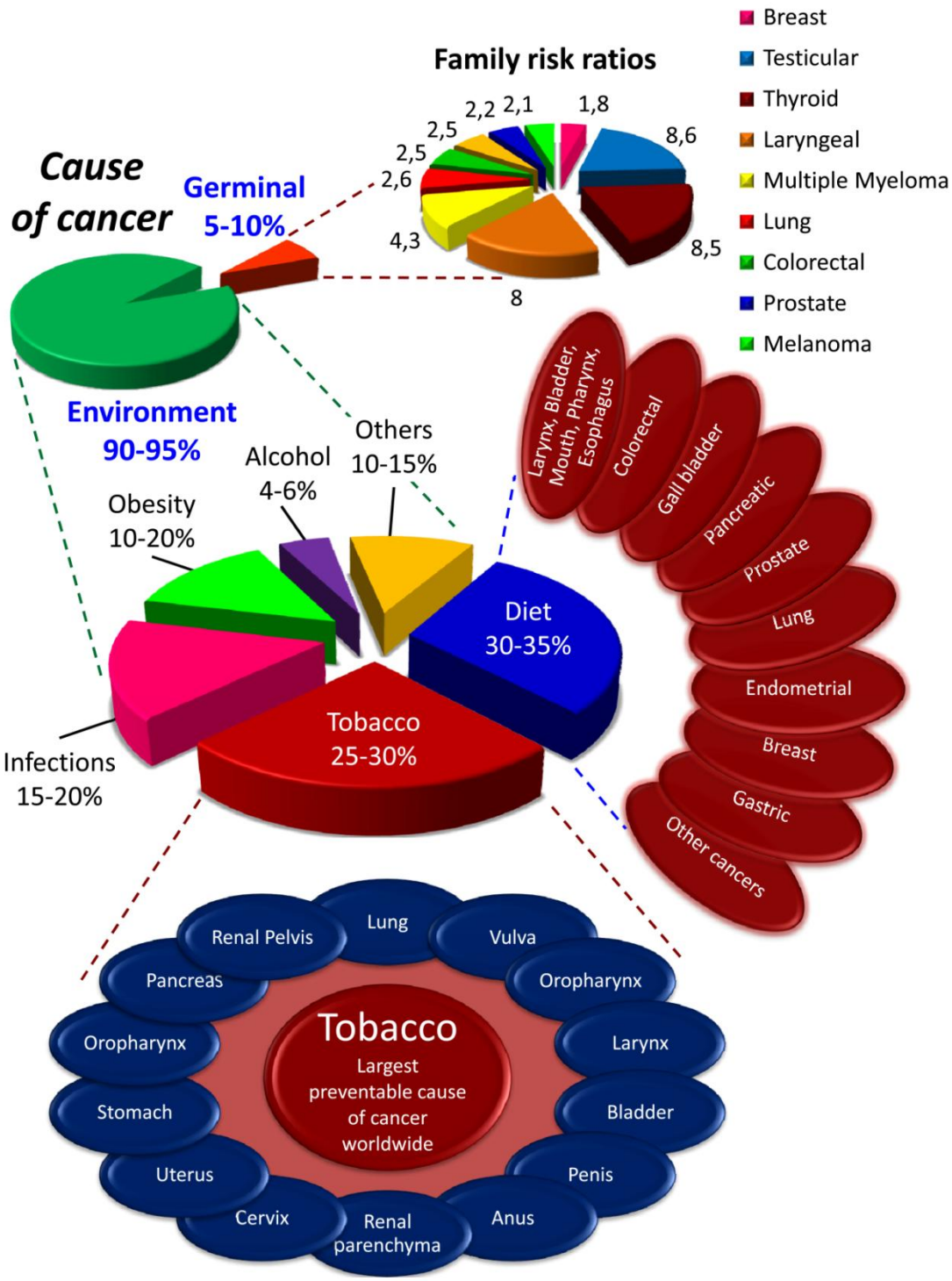


Figure 1.2. The Genetic and Environmental Etiologies of Cancer. Familiar history of cancer contributes only with 5-10% to cancer risk (upper left circular graph). The upper right graph represents the familial risk ratios (compared to all population) for different types of cancer. Environmental contribution accounts for 90-95% of cancer risk, being distributed in several factors. Diet (30-35%) and tobacco (25-30%) occupy the top positions in the list. Interestingly, they are preventable factors that should be taken into consideration to reduce the incidence of the most prevalent cancer types. Middle graph represents the percentage of each environmental cause of cancer and concomitant graphical projection of the cancers that have been etiologically associated to smoking and diet. Adapted from *Anand, P. et al., 2008*[24].

increase cancer risk in these countries, namely the contribution of pollution, food additives and use of pharmaceuticals.

In the late 20th century, tobacco smoke was classified as the most significant human carcinogen, being responsible for 30% of cancer deaths in the United States[26]. Etiologic studies demonstrated that lung cancer risk increase 22 times in male and 12 times in female smokers, respectively[27]. Importantly, tobacco consumption is the major cause of cancer that can be prevented[1].

Alcohol consumption is another avoidable cause of cancer and it has been associated to 3,6% of all cancers[28]. Its uptake increases the risk of liver, colorectal, oral cavity, larynx, pharynx, esophagus and also female breast cancers[29]. The incidence is dose-dependent, with higher cancer occurrence for moderate/high alcohol intakes, compared to light drinkers[30,31].

Physical inactivity, dietary factors, obesity, environmental pollution, occupational carcinogens, and radiation are other main cancer risk factors worldwide. Relevantly, the ultraviolet radiation exposure is carcinogenic and the major cause of skin cancer[1].

Recently, covering a very broad population, the consumption of red meat and processed meat was evaluated in terms of carcinogenicity as "*probably carcinogenic to humans*". Epidemiological data and mechanistic evidences showed a positive correlation between consumption of red meat (cancer risk could increase by 17% for every 100 gram portion) and processed meat (18% for every 50 gram portion, eaten daily) with colorectal cancer. Additionally, the IARC pointed out for the existence of a positive association between consumption of red meat and pancreatic and prostate cancer; and processed meat and gastric cancer[32].

In terms of pharmaceuticals, female oral contraceptives were associated with an increased risk of breast cancer and a decreased risk of ovarian and endometrial cancers; however, tamoxifen increases the latter one. The risk of vaginal cancer was associated with the employ of the estrogen diethylstilbestrol. Conversely, post-menopausal hormone substitution therapy decreases the risk of colon cancer but increases the risk of endometrial and breast ones. The risk of colorectal cancer is also decreased by the use of non-steroidal anti-inflammatory drugs[33].

1.3.3 Infections

Over the last decades, parasitic infections of *Schistosoma haematobium* in the urinary system, *Clonorchis sinensis* in the liver, and *Helicobacter pylori* in the stomach, were strongly correlated to bladder carcinoma, cholangiocarcinoma of the bile ducts, and gastric cancer, respectively[14,34-37]. Chronic infections account for one fifth of all cancers worldwide. Clearly contributing to these numbers are the human papillomavirus (HPV) and the hepatitis B virus (HBV), associated to cervical and liver cancer, respectively[1].

1.3.4 Cancer Burden Reduction

According to World Health Organization (WHO), the avoidance of certain risk factors, specially tobacco use, could reduce in more than 30% the total number of cancer deaths[1]. In terms of tobacco consumption, a comprehensive ban on tobacco sponsorship, promotion, advertising and national smoke-free laws could decrease significantly its consumption[1]. Changes in lifestyle; familiarity with major risk factors, like diet, sun exposure and occupational exposure to carcinogens; and knowledge of cancer familiar history would be also important in terms of prevention[38].

Personalized and effective treatment, comprising chemotherapy, surgery and radiotherapy, as well as a more tailored diagnosis and early detection, could raise considerably cancer survival rates. In developing countries, there is a special require for cancer control plans in order to improve cancer prevention and care[1,38]. On the other side of the coin, developed countries started to prevent the man's oldest foe, sometimes with extreme measures. In women with or without previous breast cancer diagnosis, the contralateral prophylactic mastectomy (CPM) and the bilateral prophylactic mastectomy (BPM) have increased over the last years. The indication of these invasive chirurgical interventions are based on high risk factors that include mutations in *breast cancer 1, early onset (*BRCA1), BRCA2 or other breast cancer predisposition genes, familiar history, diagnosis at young age and breast histology. The difficulties of the decision-making process have important psychological and ethical components. However efficient and innovative reconstructive approaches turned these procedures more attractive and eligible[39]. A study published in 2008 evaluated the negative and positive expectations pre- and postoperatively, body image perception, sexual activity, health-related quality of life, anxiety and depression after bilateral prophylactic mastectomy

followed by reconstruction. In this group of women, the follow-up indicated negative impact on body image and sexuality, but no effects on quality of life, anxiety or depression[40]. Partial or total prophylactic surgical resection of ovary (oophorectomy), testis (orchiopexy), colon (colectomy), thyroid (thyroiectomy) and stomach (gastrectomy) can also prevent the onset of their related cancers. Recently, a prophylactic prostatectomy (prostate) was also conducted; however the associated risks are nowadays excessively severe to generalize this type of intervention in a near future. Some of the risks comprise urinary incontinence and impotence[41].

1.4 The Hallmarks of Cancer

Despite the enormous variety and complexity throughout cancer types, there is a set of common physiopathological implications shared by the majority of cancers. Tumor cells enter in a process of mitotic immortality, supported by a persistent proliferative signaling, cell death resistance selection, escaping from growth suppressors and enabling angiogenesis, acquiring a more aggressive phenotype when they undergo under an invasion process and cell survival in an ectopic environment[22]. The hallmarks of cancer were recently upgraded by two new concepts. Cancer cells experience a metabolic reprogramming to ensure a continuous cell growth and proliferation; and they adopt a set of features that allow them to escape from the immune system **(Fig. 1.3)**[42,43]. Under a selective pressure, altered cells find out several ways to be undetected. They thrive in a chronically inflamed microenvironment, evading the immune system recognition and suppressing the immune reactivity[42]. Tumor inflammation is one of the processes that contribute to the acquisition of some of the hallmarks of cancer. Malignant lesions are usually enriched in activated inflammatory cells, sources of pro-angiogenic and growth factors. Concomitantly, tumor microenvironment is also supplied with reactive oxygen species, inducing DNA damage and genomic instability[44].

All together the previous hallmarks allow cancer cells to survive, divide and colonize neighboring and distant vital tissues. This cellular malignancy is based on the acquirement and accumulation of genetic and epigenetic changes, hereditarily maintained across cancer cell divisions[45]. Genomic instability affects both tumor suppressor genes and oncogenes. Generally, tumor suppressor genes, after an initial loss of heterozygosity (LOH), can be silenced by a genetic mutation; or epigenetically, by

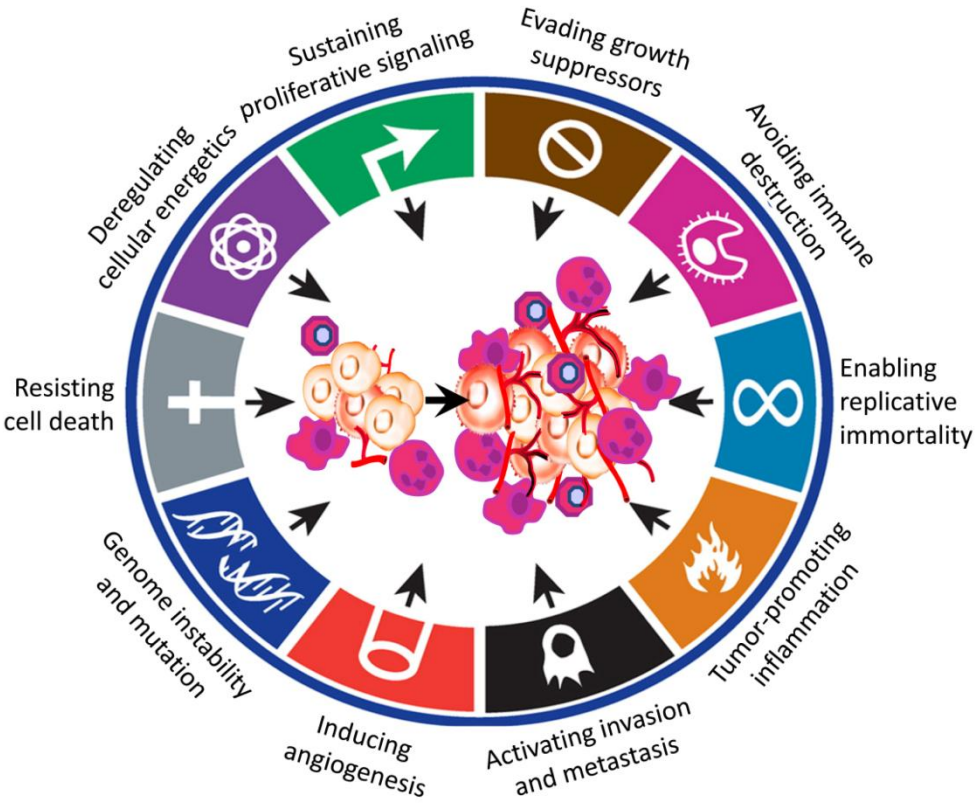


Figure 1.3. Hallmarks of Cancer. An accumulation of genetic and epigenetic modifications cooperate to the acquirement of a set of hallmarks that characterize malignant lesions. Schematic representation of the ten recognized hallmarks in cancer. *Adapted from Hanahan, D., 2011*[43].

DNA methylation or histone modifications[46-51]. Fundamentally, tumor suppressor genes need to be inactivated in both alleles (loss of function), whereas the gain-of-function of oncogenes can be mediated by the activation of a unique allele[22]. The mechanisms of activation of oncogenes include gain-of-function mutations, DNA copy number amplifications and chromosomal rearrangements[22,51,52].

1.5 Colorectal Cancer

1.5.1 Epidemiology and Etiology

In 2012, according to the GLOBOCAN database[23], the estimated number of new colorectal cancer cases worldwide accounted to around 1,4 millions. This type of cancer is the second most common cancer in women (614.000 new cases; 9,2%) and the third in men (746.000 new cases; 10%). There is a clear geographical disproportion in its worldwide incidence, where more than half part of the new cases occurs in the

developed countries. Nevertheless, in terms of mortality, the less developed countries showed 52% of total number of deaths by colorectal cancer with 361.000 cases (both sexes)[23].

In terms of etiology, the attributable lifestyle-related risk factors are diet (specially rich in red meats, processed meats and meats cooked at very high temperatures), physical inactivity, high stress, obesity, smoking and high consumption of alcohol[53]. Other risk factors not related to lifestyle comprise: age (increased risk in older people); type 2 diabetes; history of ulcerative colitis, Crohn's disease, or certain kinds of polyps; and race or ethnic background, for instance being African American or Ashkenazi[53].

1.5.2 Genetic Inheritance

Colorectal tumors occur sporadically in the majority of cases. Nevertheless, there are three recognized inherited syndromes linked to a higher incidence, being responsible for 5-10% of the cases. Germ-line mutations in tumor suppressor genes, including *adenomatous polyposis coli* (*APC*) gene, linked with Familial Adenomatous Polyposis (FAP); *mutL homolog 1* and *mutL homolog 2* (*MLH1* or *MSH2*), linked with Lynch syndrome (also called hereditary non-polyposis colon cancer, HNPCC); and *mutY DNA glycosylase* (*MUTYH*), linked with MAP syndrome (*MUTYH* associated polyposis), lead to a higher risk for the development of the disease. During colonic oncogenesis through the last syndrome, *APC* gene is secondarily mutated, whereas it is congenital mutated in the FAP syndrome, or somatically in sporadic colorectal cancers[54-56].

1.5.3 Molecular Alterations

A genetic model for colorectal tumorigenesis was defined by an ordered set of pathological events. At molecular level, premalignant entities experience a sequence of specific genetic alterations that are translated into well-characterized clinical morphological alterations[57].

According to this model, colorectal cancer starts with adenomatous polyps that arise from the normal colonic mucosa, being considered the precursor lesions. Consequently, they initiate a multistep evolution leading to an adenoma-carcinoma status, that is very well characterized **(Fig. 1.4)**[58,59]. Clear evidences point out that the majority of colorectal cancers start with an abnormal activation of the Wnt / β-Catenin Signaling

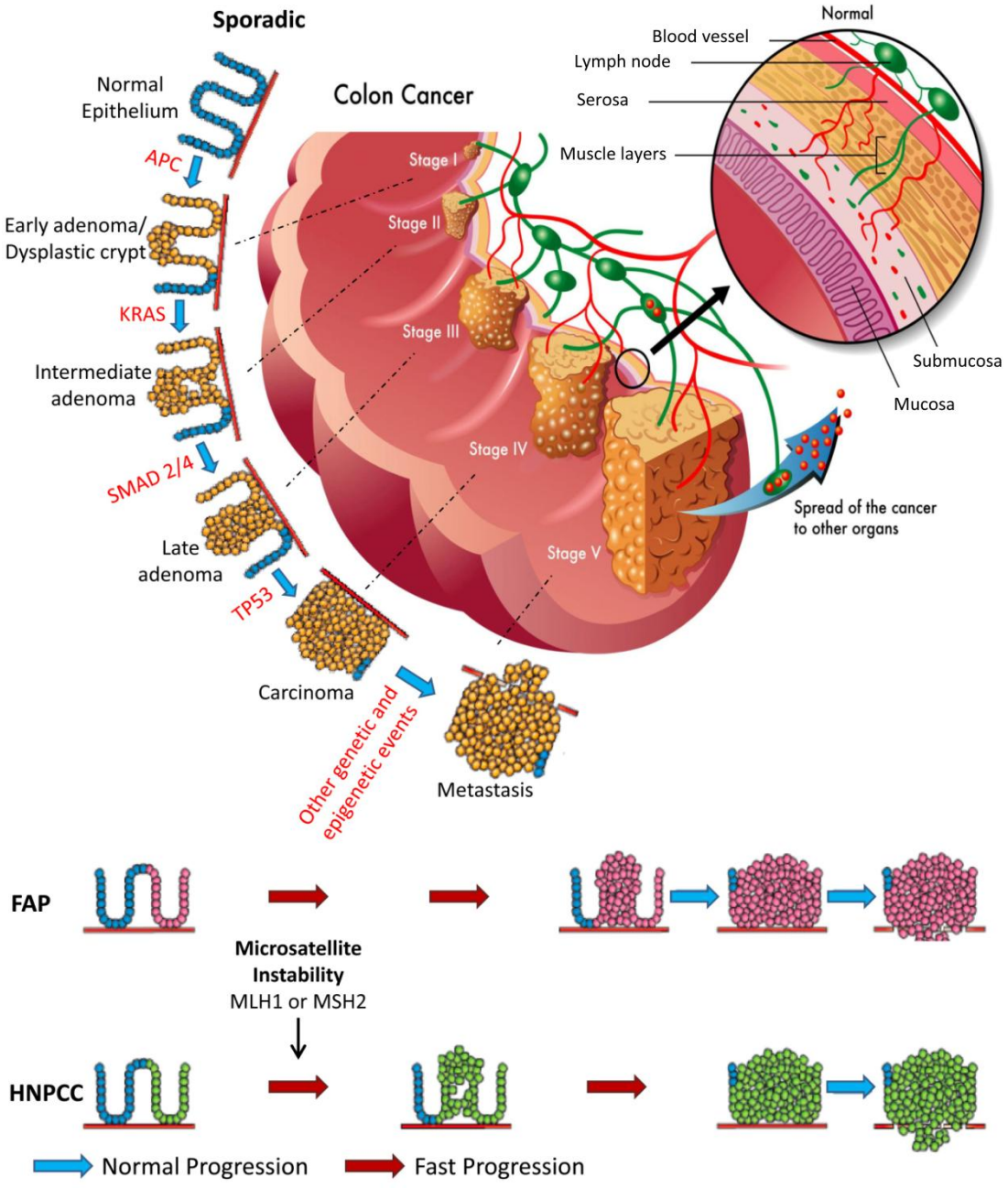


Figure 1.4. Colorectal Tumor Progression. There are several key biological alterations that take place during colorectal malignancy. Wnt pathway activation (*APC*), EGFR signaling activation (*KRAS*), TGFβ response inactivation (*SMAD 2/4*), loss of TP53 function (*TP53*) and other genetic and epigenenetic events cooperate to promote the malignant process. The large majority of the colorectal cancers are sporadic. However one patient out of twenty develops colon or rectal cancer due to genetic predisposition, namely FAP and HNPCC. Their increased risk is accompanied by a fast progression in the early stages of the tumoral development. *Adapted from Davies, R.J et al. 2005*[61] *and http://www.singaporemedicalclinic.com/.*

Pathway (*APC* mutations), followed by the activation of the epidermal growth factor receptor (EGFR) signaling network by *kirsten rat sarcoma viral oncogene homolog* (*KRAS*) mutations. At late stages, transforming growth factor beta (TGFβ) response inactivation (SMAD family member 2/4, SMAD2/4) and loss of tumor protein P53 (TP53) function cooperate to support the malignant process[60,61]. Despite the genetic model of colorectal cancer, progression presumes of the accumulation of other genetic and non genetic defects[62]. Epigenetic events in cancer-related genes and non-coding RNAs (ncRNAs) involvement is being considered more and more as essential to uncover the tip of this iceberg[63,64]. Cancer cells, that initially hold a limited set of genetic and epigenetic alterations, start to accumulate and to combine cellular modifications, resulting in subpopulations of cells that activate distinct pathways, getting selective advantages in terms of cellular maintenance and proliferation[43,63,65].

There are three canonical molecular pathways implicated in colorectal cancer progression: chromosomal instability (CIN); microsatellite instability (MSI); and Cytosine-phosphate-Guanine (CpG) island methylator phenotype (CIMP)[66]. The molecular profile of CIN positive tumors includes gene copy number variation (CNV), chromosomal rearrangements, aneuploidy and a frequent LOH[67]. It is implicated in the inactivation of tumor suppressor genes, namely *APC*, *TP53*, *SMAD4/SMAD2* and *deleted in colorectal carcinoma* (*DCC*); and activation of oncogenes, namely *KRAS*, *phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit alpha* (*PIK3CA*) and *Catenin (Cadherin-Associated Protein) Beta 1* (*CTNNB1*), within both sporadic and inherited colorectal tumors[68-71]. MSI is a hypermutable phenotype caused by the loss of DNA mismatch repair genes (*MLH1*; *MSH2*; *MSH6*; and *PMS1 Homolog 2, PMS2*)[66]. Around 15% of all colorectal cancers harbor this feature, having a tendency to arise in the proximal colon with a mucinous or signet ring appearance. They are poorly differentiated, tend to present lymphocytic infiltrates, respond differentially to chemotherapeutics and have a slight better prognosis. The inactivation of *MLH1* can be attributed to the acquired hypermethylation of its promoter, through CIMP pathogenic pathway. CIMP refers to the extensive hypermethylation of CpG islands in the promoter regions of several loci, namely those regulating the expression of tumor suppressor genes[66].

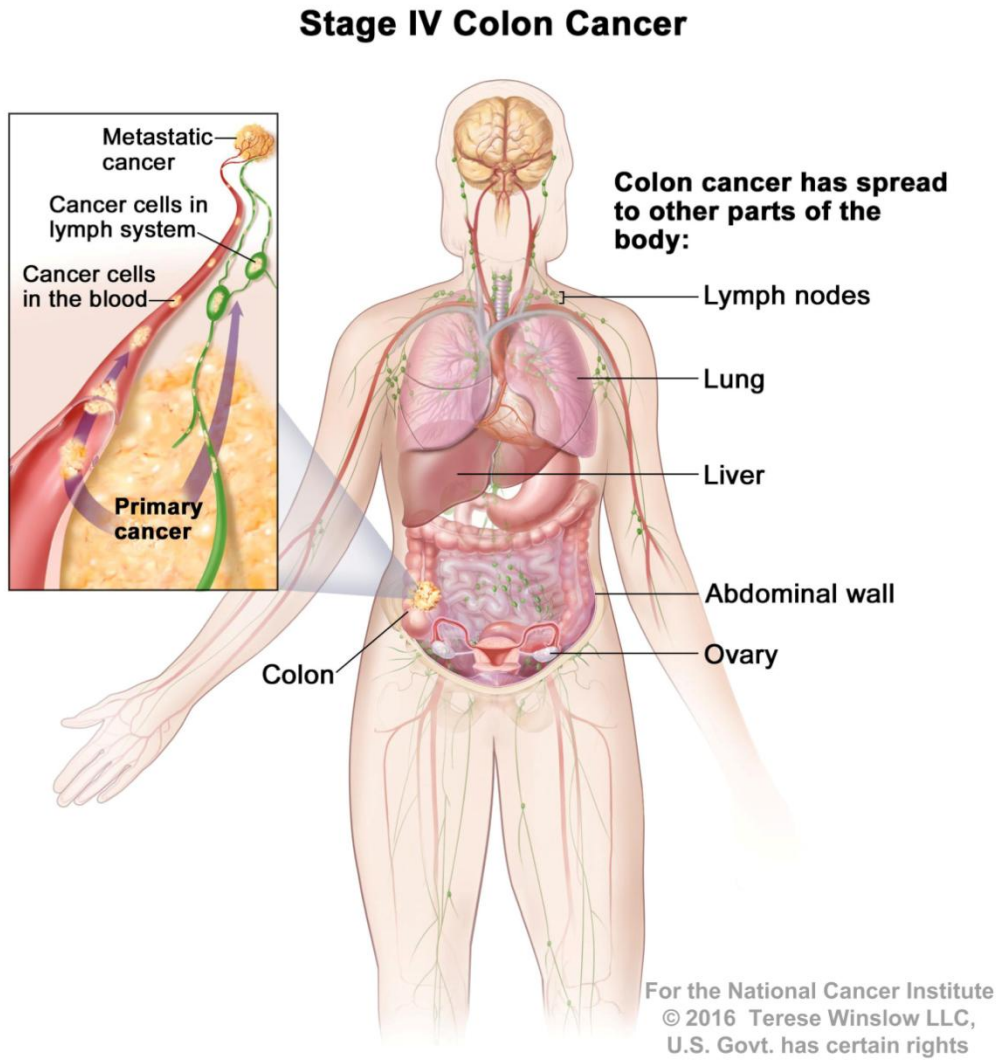


Figure 1.5. Colon Cancer Metastasis. Individuals with Stage IV Colon Cancer have metastasis that arises by the spreading of cancer cells through the blood vessels and lymph system, to other parts of the body. Regional lymph nodes and liver are the most affected organs, but colon cancer can also metastasize to lung, brain, bone, peritoneum, ovary and distant lymph nodes. *Illustration by Terese Winslow LLC (with permission).*

1.5.4 Stages of Disease

Throughout tumor progression there are well-defined stages. The one at which cancer is diagnosed determines treatment options, being crucial for the prognosis. At early stages, when confined to bowel mucosa, colorectal cancer is usually curable with a 5-year survival rate of more than 90%, according to the Surveillance, Epidemiology, and End Results (SEER) Program[72]. Commonly the primary colorectal tumor evolutes to metastases, mainly into the liver, the most common metastatic target site (50-60% of all cases)[73,74]. Patients with colorectal cancer liver metastasis (stage IV) have an overall median survival following liver resection of 3,6 years[75]. Colorectal cancer can also metastasize to lung[76], brain[77], bone[78], peritoneum[79], ovary[80,81] and distant lymph nodes **(Fig. 1.5)**[82,83].

1.6 Testicular Germ Cell Cancer

Germ cell cancers comprise a vast group of tumors. More than 95% of the patients present the primary tumor in the testis, whereas only a small fraction is found extragonadally, namely in the retroperitoneum or in the mediastinum **(Table 1.1)**[84-86]. Testicular germ cell cancer (TGCC) is a relatively uncommon malignancy, accounting for only 1% of all cancer in males. However, it is the most frequent solid malignancy of Caucasian males aged from 15 to 35 years and the main cause of cancer-related mortality and morbidity of young patients, thus at their peak of productivity[23,86-88]. In addition, in the last four decades, the worldwide incidence strikingly increased, making imperative the understanding of the etiology and molecular events that took place in the development of TGCC[88,89].

1.6.1 Epidemiology and Etiology

According to the database GLOBOCAN (2012), there are more than 50.000 new cases of TGCC globally, per year. The incidence of this cancer is very unbalanced worldwide with North and West Europe presenting the higher rates (8,8 and 7,4 new cases per 100.000 per year, respectively) compared to Asia and Africa showing the lowest ones (0,7 and 0,3 new cases per 100.000 per year, respectively). The rate of incidence increases according to the regional human development, ranging from 0,4 in *low human development* regions, to 5,6 new cases per 100.000 per year in *very high human development* areas[23].

Norway, Denmark and Switzerland ranked in the first positions among the countries with higher cumulative risk, defined by the probability of getting this malignancy before the age of 75 years old[23]. Statistics points out easily that in terms of etiopathogenesis of testicular cancer, there is a clear association between the incidence rates and ethnicity. Caucasian males, particularly in northern Europe, and black males in Africa, show the highest and lowest incidence rates, respectively. This difference should be derived from the combination between genetic factors and environmental exposure to deleterious agents. Importantly, some of these substances have an estrogen-like effect that reiterates the estrogen excess theory, the most valued hypothesis for the pathogenesis of testicular cancer. According to this theory, the process of oncogenesis can be initiated by a relative excess of estrogens in the mother's early pregnancy, concomitantly to the gestational development of the gonads[90,91]. These malignant cells are nevertheless kept silenced until the endocrine stimuli of puberty. This hypothesis is supported by case control studies reporting an increase in the incidence of testicular cancer in children, after exogenous estrogen exposure during early gestation[92,93]. Curiously, the low incidence of testicular cancer in black African men could be explained by the high serum androgen (testosterone) and lower relative estrogen levels in black African women during their pregnancy[90,94]. Other identified risk factors include cryptorchidism[95]; germ cell tumor in the contralateral testis[96]; familial testis cancer history[97]; gonadal dysgenesis[98]; infertility[99], Klinefelter syndrome[100,101], high levels of maternal Epstein-Bar IgG antibodies[102], decreased androgen levels in puberty and early adulthood[103], and environment[104].

1.6.2 Development

During germ-cell development, healthy primordial germ cells (PGCs), presenting biparental pattern of genomic imprinting, originate either an oocyte in females or spermatozoa in males. The differentiation process comprises a uniparental genomic imprinting establishment in detriment of the biparental one that is removed. From embryo to adult age, the process of differentiation can be compromised by intrinsic and environmental factors, resulting in a diversity of different germ cell tumors, according to the developmental stage of the PGCs, age and sex of the patient[105,106].

In a first stage, PGCs instead of differentiate into gonocytes can be reprogrammed to pluripotent embryonic germ cells (EGCs), originating infantile germ cell tumors,

namely teratomas or yolk-sac tumors. In a second instance, arrest of differentiation of gonocytes, followed by a microenvironment adaptation (in gonads, thymus, or pineal gland/ hypothalamus/ pituitary area) and proliferation, can originate a seminoma (SE)/ dysgerminoma/ germinoma. The reprogramming of a seminomatous tumor cell into a pluripotent embryonal carcinoma cell originates a non-seminoma (NSE). Both tumors are originated by a germ cell neoplasia *in situ* (GCNIS)[105,106]. Less frequently, a SE cell might originate a NSE[107], explaining the common discovery of non-seminomatous metastasis in patients diagnosed with SEs, during autopsies **(Fig. 1.6)**[108].

In females, germ cells go into meiosis in intra-uterine development, while in males the onset of meiosis takes place at puberty. Consequently, the amount of target PGCs/ gonocytes that can fail maturation, experiencing a malignant transformation, is visibly lower in females than in males, having a lower and higher incidence, respectively. On the other hand, the number of target cells susceptible to originate teratomas is higher in ovary than in testis[105,106]. Pure teratomas represent 95% of ovarian germ cell tumors, but only 4% of testicular germ cell tumors[109-111].

1.6.3 Types

Testicular germ cell tumors comprise three types of different tumors (I, II and III), according to their anatomical site, phenotype and origin **(Table 1.1)**[105,112]. Testicular type II germ cell tumors comprise SE and NSE and have a median age of incidence of 35 and 25 years old, respectively. They are derived from PGCs or gonocytes, where genomic imprinting is erased and characterized by some chromosomal abnormalities[113]. After reprogramming from either a testicular GCNIS or a SE, NSE can be represented by different histological elements, being classified in embryonic cell carcinoma (malignant equivalent of embryonic germ cells), choriocarcinoma (correspondent to the extra-embryonic differentiation), yolk sac tumor and teratoma (representing somatic differentiation)[105]. The aggressiveness of NSE elucidates the reason by which they appear at younger ages compared to slothful SE, to which is assigned a "*loss of stem cell capacity"*, appearing in older patients[107,114]. In the case of teratomas, they exhibit an embryonic differentiation, resembling organ structures of all germ layers. The lower or higher histological grade of immaturity (predominantly neuroepithelial) categorize them

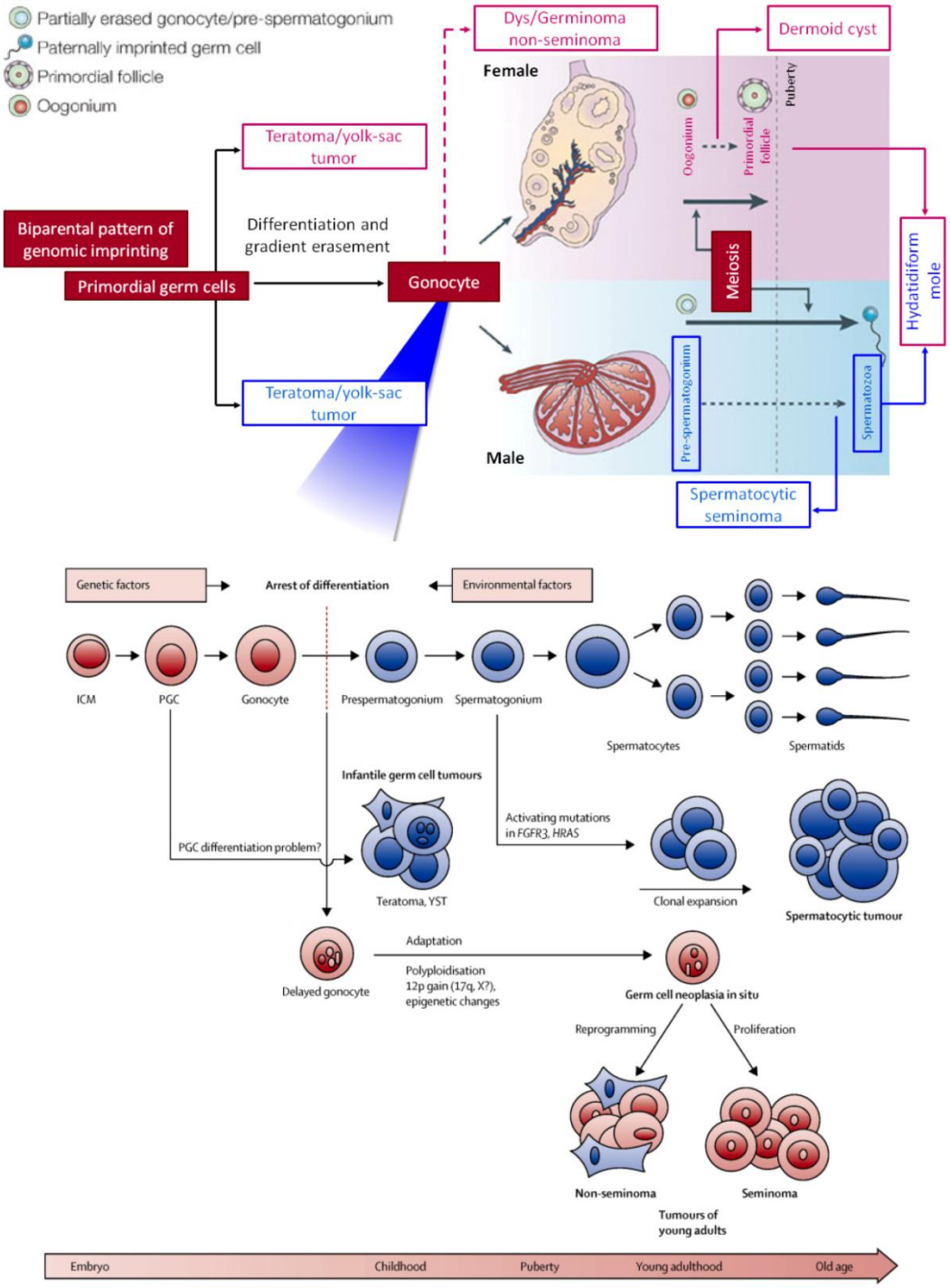


Figure 1.6. Germ Cell Tumors: Histological Origin of Seminomas and Non-Seminomas. PGCs differentiate to oocytes in female or to spermatozoa in males. The maturation can be compromised by intrinsic and environmental factors in different developmental stages, giving rise to different germ cell tumors. Firstly, PGCs can be malignantly reprogrammed into pluripotent EGCs, originating teratomas or yolk-sac tumors. Secondly, during the differentiation process of PGCs, they can originate SEs/ dysgerminomas / germinomas. NSEs are formed by reprogramming of a GCNIS, whereas the SEs by their proliferation. *Adapted from Oosterhuis and Looijenga, 2005[105]; Rajpert-De Meyts, McGlynn et al., 2016[106].*

Table 1.1. Classification of Germ Cell Tumors.

Type	Anatomical site	Phenotype	Age	Originating cell	Genomic imprinting	Genotype
I	Testis/ovary/ sacral region/ retroperitoneum/ mediastinum/ neck/midline brain/other rare sites	(Immature) teratoma/ yolk-sac tumour	Neonates and children	Early PGC/ gonocyte	Biparental, partially erased	Diploid (teratoma). Aneuploid (yolk-sac tumour): gain of 1q, 12(p13) and 20q, and loss of 1p,4 and 6q
II	Testis	Seminoma/ non-seminoma	>15 years (median age 35 and 25 years)	PGC/gonocyte	Erased	Aneuploid (+/– triploid): gain of X, 7, 8, 12p and 21; and loss of Y, 1p, 11, 13 and 18
	Ovary	Dysgerminoma/ non-seminoma	>4 years	PGC/gonocyte	Erased	Aneuploid
	Dysgenetic gonad	Dysgerminoma/ non-seminoma	Congenital	PGC/gonocyte	Erased	Diploid/tetraploid
	Anterior mediastinum (thymus)	Seminoma/ non-seminoma	Adolescents	PGC/gonocyte	Erased	Diploid/tri-tetraploid
	Midline brain (pineal gland/ hypothalamus)	Germinoma/ non-seminoma	Children (median age 13 years)	PGC/gonocyte	Erased	Diploid/tri-tetraploid
III	Testis	Spermatocytic seminoma	>50 years	Spermatogonium/ spermatocyte	Partially complete paternal	Aneuploid: gain of 9
IV	Ovary	Dermoid cyst	Children/adults	Oogonia/oocyte	Partially complete maternal	(Near) diploid, diploid/tetraploid, peritriploid (gain of X, 7, 12 and 15)
V	Placenta/uterus	Hydatidiform mole	Fertile period	Empty ovum/ spermatozoa	Completely paternal	Diploid (XX and XY)

PGC, primordial germ cell.

Adapted from Oosterhuis and Looijenga, 2005[105].

as mature or immature, respectively[115,116]. Generally, testicular germ cell tumors contain a mixture of multiple histological patterns, rarely than a unique one[117-119].

The treatment of TGCC is based on the aggressiveness of the tumor. In these terms, tumors are divided in two unique groups, the first one encompassing pure SE and the second covering all the others NSE. This last group includes those tumors that comprise a mixture of non- and seminomatous tumors, since NSEs are more aggressive and invasive[86,120].

1.6.4 Molecular Alterations

Aneuploidy and complex karyotypes are the most common cellular features of testicular germ cell tumors of adolescent and young adult men, with loss of chromosomes Y, 1p, 11, 13 and 18 and gain of X, 7, 8, 12p and 21[86,105]. DNA ploidy pattern in NSEs is more heterogeneous than in SEs, with hyperdiploidy to hypertripoidy in the former in comparison with triploidy and tetraploidy in the lattest[121]. Particularly, gain of chromosome arm 12p, most usually as an isochromosome in more than 80% of the cases, i(12p), is the most frequent aberration and it is a genetic hallmark assigned to the

development of an invasive TGCC, by an unclear mechanism **(Table 1.1)**[113,122-124]. Moreover i(12p) isochromosome copies have a tendency to be higher in NSE compared to SE. Tumors without i(12p) show sometimes other structural alterations of chromosome 12[121]. The Y chromosome deletion gr/gr (1.6 Mb) is associated with infertility and confers susceptibility to testicular germ cell tumors, being more associated with SEs than NSEs[125-127]. Activating mutations and/or increased expression levels of *v-Kit hardy-zuckerman 4 feline sarcoma viral oncogene homolog* (*c-KIT*); g*rowth factor receptor-bound protein 7* (*GRB7*); *KRAS, neuroblastoma RAS viral (V-Ras) oncogene homolog* (*NRAS*); and *B-Raf proto-oncogene, serine/threonine kinase* (*BRAF*) and a maternal inherited 2,7 Mb locus on Xq27 have been also associated with TGCC[128-131].

The onset, progression and metastatic behavior of testicular germ cell tumors have been endorsed to the contribution of several microRNAs (miRNAs). They are potential serum-based biomarkers for diagnosis, prognosis and for drug response and resistance mechanisms, comprising miR-371, miR-372, miR-373 and miR-367[132-135]. Hypermethylation of some tumor suppressor gene promoter regions as of Alu sequences is observed in NSEs, while a global loss of methylation of imprinted genes and of LINE-1 sequences are noted in either SEs and NSEs[106].

1.7 Leukemia

Leukemia is a clonal proliferation of hematopoietic stem cells that replace the normal bone marrow with malignant blood cells **(Fig. 1.7)**[136]. As a consequence, a substantial decrease in erythrocytes, platelets and normal leukocytes originate some symptoms that include fatigue and breathless (anemic)[137-139]; easy bruising and bleeding[140-142]; and increased risk of infection[143-146], respectively. The main causes of death in adults with acute leukemia are infections, hemorrhage and organ failure[147,148].

Leukemias can be characterized in several types and subtypes, according to the rate of progression and type of white blood cells affected. Therefore, the most frequent subtypes of leukemia are acute lymphoblastic leukemia (ALL), frequently detected in children; and acute myeloid leukemia (AML), chronic lymphocytic leukemia (CLL), and chronic myeloid leukemias (CML), more often diagnosed in adults[149].

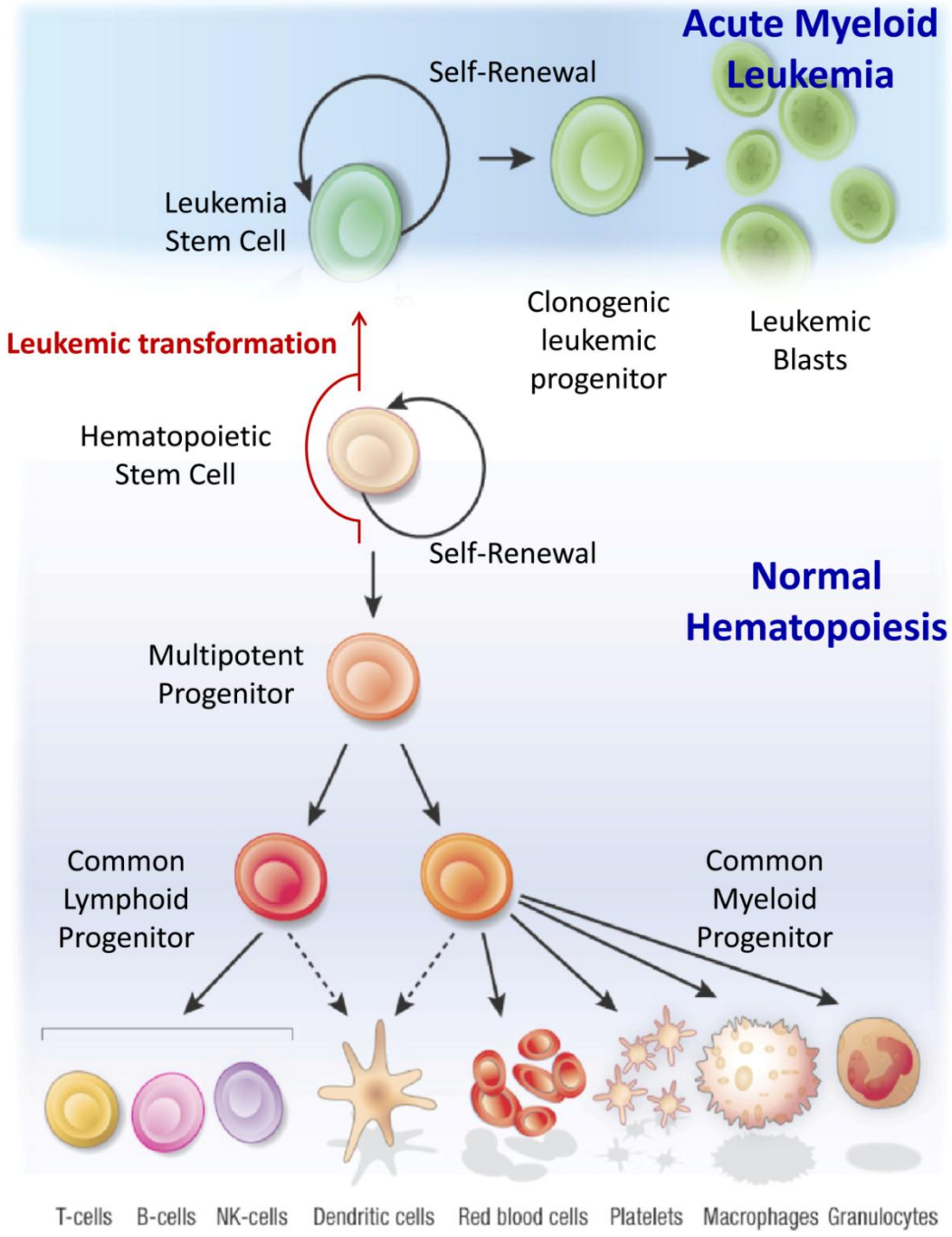


Figure 1.7. Normal Hematopoiesis and Leukemic Transformation. The origin of the hematopoietic system cells is based on the self-renewal of normal hematopoietic stem cells that originate progressively lineage-restricted progenitors with gradually less ability to divide. In AML, it is hypothesized that leukemic stem cells arise from immature myeloid progenitors or hematopoietic stem cells, by the accumulation of mutations, chromosomal rearrangements and epigenetic changes. *Adapted from Tan, BT., et al., 2006*[136].

1.7.1 Epidemiology and Etiology

According to the data published by GLOBOCAN in 2012, the estimated number of new leukemia cases worldwide accounted to around 350.000 new cases with an unbalanced distribution by sex, with 200.000 new cases in men and 150.000 new cases in women. In the same year the estimated number of leukemia-related deaths was around 265 000, proportionally distributed by sex, taking into account the correspondent incidence[23]. Thirty per cent of all leukemias correspond to AMLs, with 18.300 new cases in Europe, per year, around 0,6% of all cancers[150].

In terms of etiology, it was observed that obesity is associated with an increase risk of leukemia. Researchers suggested that endocrinologic, metabolic, immunologic and inflammatory-like alterations could induce the malignant transformation, by disturbing DNA repair mechanisms, gene function or by epigenetic alterations. Other possibility lies in the fact that obesity can create a selective environment where dormant pre-existing clones can emerge[151]. Epidemiologic studies suggested that individuals who have a history of hematologic malignancy or that have been exposed to ionizing radiation have higher risk for the development of leukemias. The latter case includes persons who were submitted to two or three computed tomography scans (especially if they were young), cancer patients who received radiotherapy, medical radiation workers (before 1950) and atomic bomb survivors. Furthermore, exposure to benzene is associated to a higher the risk of AML in adults and exposure to household pesticides *in utero* or in the first three years of life was associated also to a higher the risk of ALL in childhood. Genetically, Down syndrome and neurofibromatosis are also associated with a higher risk of childhood AML and ALL[149]. In terms of therapy outcomes, a recent study identified statin therapy and tobacco consumption as factors impacting positively or negatively the remission after chemotherapy, respectively[152].

1.7.2 Acute Myeloid Leukemia

The AML is also known as acute non-lymphoblastic leukemia or acute myelogenous leukemia. It comprises a group of different non-solid malignancies characterized by the accelerated proliferation of abnormal white blood cells, with concomitant colonization and accumulation of immature blood-forming cells, also called myeloblasts, in bone marrow and blood. This results in an excessive number of immature white cells and

reduced number of red blood cells or platelets[153-155]. This malignancy affects mostly persons at older ages, with 54% of the patients over 65, and 33% over 75 years. The median age at diagnosis is 66 years[156]. Depending on blast count in the peripheral blood and on the presence of complications such as marrow failure, tissue infiltration or hyperuricemia, untreated patients typically die over a period of days or weeks by sepsis, hemorrhage or pulmonary or cerebral leucostasis[150].

1.7.2.1 Classification

The determination of the stage (extension) of a specific cancer type is essential in terms of diagnosis, prognosis and therapeutic clinical decisions. In terms of solid tumors, the classification is generally based on the size and tumor spreading. Not usually forming tumors, AMLs are confined to bone marrow, spreading sometimes to liver and spleen. Thus, according to the French-American-British (FAB) classification, they are classified based on the cell type of origin and on their maturity **(Table 1.2)**[157]. The recently established WHO classification takes into account new research discoveries that affect the prognosis of a patient, having a more exhaustive categorization, for instance the presence of known chromosome abnormalities such as translocations **(Table 1.3)**[158].

Table 1.2. The FAB Classification of Acute Myeloid Leukemia.

FAB Subtype	Name
M0	Undifferentiated AML
M1	AML without maturation (poorly differentiated)
M2	AML with maturation (more differentiated)
M3	Acute promyelocytic leukemia (APML)
M4	Acute myelomonocytic leukemia (AMML) Subtype M4 eos: Acute myelomonocytic leukemia with >5% eosinophils
M5	Acute monocytic leukemia Subtypes: M5a: Acute monoblastic leukemia - poorly differentiated M5b: Acute monocytic leukemia - more differentiated
M6	Acute erythroblastic leukemia
M7	Acute megakaryoblastic leukemia

Adapted from Bennett, J.M. et al.,1976[157].

Table 1.3. The WHO Classification of Acute Myeloid Leukemia.

Acute Myeloid Leukemia and Related Neoplasms
Acute myeloid leukemia with recurrent genetic abnormalities
AML with t(8;21)(q22;q22); *RUNX1-RUNX1T1*
AML with inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
APL with t(15;17)(q22;q12); *PML-RARA*
AML with t(9;11)(p22;q23); *MLLT3-MLL*
AML with t(6;9)(p23;q34); *DEK-NUP214*
AML with inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1*
AML (megakaryoblastic) with t(1;22)(p13;q13); *RBM15-MKL1*
Provisional entity: AML with mutated NPM1
Provisional entity: AML with mutated CEBPA
Acute myeloid leukemia with myelodysplasia-related changes
Therapy-related myeloid neoplasms
Acute myeloid leukemia, not otherwise specified
AML with minimal differentiation
AML without maturation
AML with maturation
Acute myelomonocytic leukemia
Acute monoblastic/monocytic leukemia
Acute erythroid leukemia
Pure erythroid leukemia
Erythroleukemia, erythroid/myeloid
Acute megakaryoblastic leukemia
Acute basophilic leukemia
Acute panmyelosis with myelofibrosis
Myeloid sarcoma
Myeloid proliferations related to Down syndrome
Transient abnormal myelopoiesis
Myeloid leukemia associated with Down syndrome
Blastic plasmacytoid dendritic cell neoplasm

Adapted from Vardiman, JW. et al., 2009[158].

1.7.2.1 Molecular Alterations

The AML is a heterogeneous malignancy characterized by abundant recurrent molecular alterations and chromosomal abnormalities that have been and continue to be identified **(Fig. 1.7)**[136]. Karyotypic changes and chromosomal rearrangements, comprising deletions, duplications, inversions and translocations, characterize 50-70% of the patients with AML. For instance, in terms of karyotype, trisomy 8 is the most frequently abnormality observed, among many others[150,159].

Many genes derived from these chromosomal abnormalities were studied aiming their targeting through the establishment of specific therapies to destroy leukemic blasts. One of these examples is the presence of the t(15;17)(q22;q21) translocation that was demonstrated in more than 95% of the patients harboring an AML subtype M3, resulting in the expression of the PML-RARα oncofusion protein that acts as transcriptional repressor affecting gene expression programs, interfering in differentiation, apoptosis and proliferation. Nevertheless, the presence of this oncofusion protein predicts a beneficial response to all-trans retinoic acid and arsenic trioxide treatments[159-162]. The karyotypic abnormalities are important prognostic factors in AML, being used as predictors of clinical outcome and in the clinical managing of AML patients[163-166]. Moreover, the percentage of patients with cytogenetic abnormalities account to around 58% in both child and adult patients, decreasing with age (0-14 years , 73,2%; 15-34 years, 61,6%; more than 35 years, 49,2%)[164]. This issue was approached by European LeukemiaNet with the creation of a standardized reporting system to correlate the most common cytogenetic and mutational abnormalities (for the genes NPM1, CEBPA, and FLT3) found in adult AML patients with their predictable clinical outcome **(Table 1.4)**[167].

Among several other mutations in AML, the discovery of somatic mutations in the epigenetic modifiers *DNA Methyltransferase 3A* (*DNMT3A*), *ten-eleven translocation oncogene family member 2* (*TET2*), *mixed lineage leukemia* (*MLL*), *enhancer of zeste homolog 2* (*EZH2*) and *isocitrate dehydrogenase 1* (*IDH1*), *isocitrate dehydrogenase 2* (*IDH2*) and *additional sex combs like 1* (*ASXL1*) genes, suggests the implication of epigenetic machinery in the progression of this malignancy[168,169].

Table 1.4. Standardized Reporting System for Genetic Abnormalities in Acute Myeloid Leukemias. This systems correlates genetic findings with the clinical outcome of the patients.

Genetic Group	Subsets
Favorable	t(8;21)(q22;q22); *RUNX1-RUNX1T1*
	inv(16)(p13.1q22) or t(16;16)(p13.1;q22); *CBFB-MYH11*
	Mutated *NPM1* without *FLT3*-ITD (normal karyotype)
	Mutated *CEBPA* (normal karyotype)
Intermediate-I	Mutated *NPM1* and *FLT3*-ITD (normal karyotype)
	Wild-type *NPM1* and *FLT3*-ITD (normal karyotype)
	Wild-type *NPM1* without *FLT3*-ITD (normal karyotype)
Intermediate-II	t(9;11)(p22;q23); *MLLT3-MLL*
	Cytogenetic abnormalities not classified as favorable or adverse
Adverse	inv(3)(q21q26.2) or t(3;3)(q21;q26.2); *RPN1-EVI1*
	t(6;9)(p23;q34); *DEK-NUP214*
	t(v;11)(v;q23); *MLL* rearranged
	−5 or del(5q); −7; abnl(17p); complex karyotype

Adapted from Dohner, H. et al., 2010[167].

2. Epigenetics

Epigenetics from the Greek term *epi-*, means "*upon, over, above*"; and *genetiko*: means *"genitive, origin, genesis".* The first concept ("*epigenotype"*) was introduced by Conrad Waddington in 1942 to describe the molecular mechanism by which the genetic information is translated into a specific trait or phenotype, making reference to the fruit-fly *Drosophila melanogaster*[170,171]. The concept of epigenetic inheritance was broadly mentioned to explain phenomena not elucidated by genetics. In a broad-spectrum idea, epigenetics was defined as the mechanism by which the environment is able to modify the phenotype, without genetic variations[172]. Nutrition[173,174], behavior[175], tobacco[176], lifestyle[177,178] and aging[179] are some factors able to modulate the epigenetic landscape.

This field of study comprises different molecular layers, being the most important ones the histone modifications and the DNA methylation. In response to exogenous stimuli, they modify the chromatin condensation degree, modulating the accessibility to DNA-based processes, namely transcription, replication, recombination and DNA repair[63]. In terms of transcriptional activity, the status of the chromatin controls the accessibility to transcription factors. Consequently, it delineates the multiplicity of expression output through cellular differentiation and development[180-182], defining lineage patterns and specific cell-type identities[183], as well as adult cell renewal ability **(Fig 1.8)**. The preference for an alternative transcription start site (TSS) and splicing processes can also be attributed to specific chromatin profiles[184].

Both histone trimethylation and DNA methylation regulate negatively telomere length and the latter inhibits telomere recombination[185]. Interestingly, successive cell divisions result in telomere shortening to critical lengths, resulting in the decrease of the epigenetic heterochromatin marks and increase in histone acetylation. The open chromatin status is associated to the appearance of age-related pathologies[185-187]. Overall, the interplay between telomere-length and epigenetic status highlight their importance in cancer progression and aging[188-191].

In a healthy cell, transposable elements and endogenous retroviruses are examples of chromosomal loci epigenetically repressed by DNA methylation and histone modifications, resulting in high levels of chromatin compaction, in order to prevent deleterious translocations to other genomic loci, insertional mutagenesis and to ensure

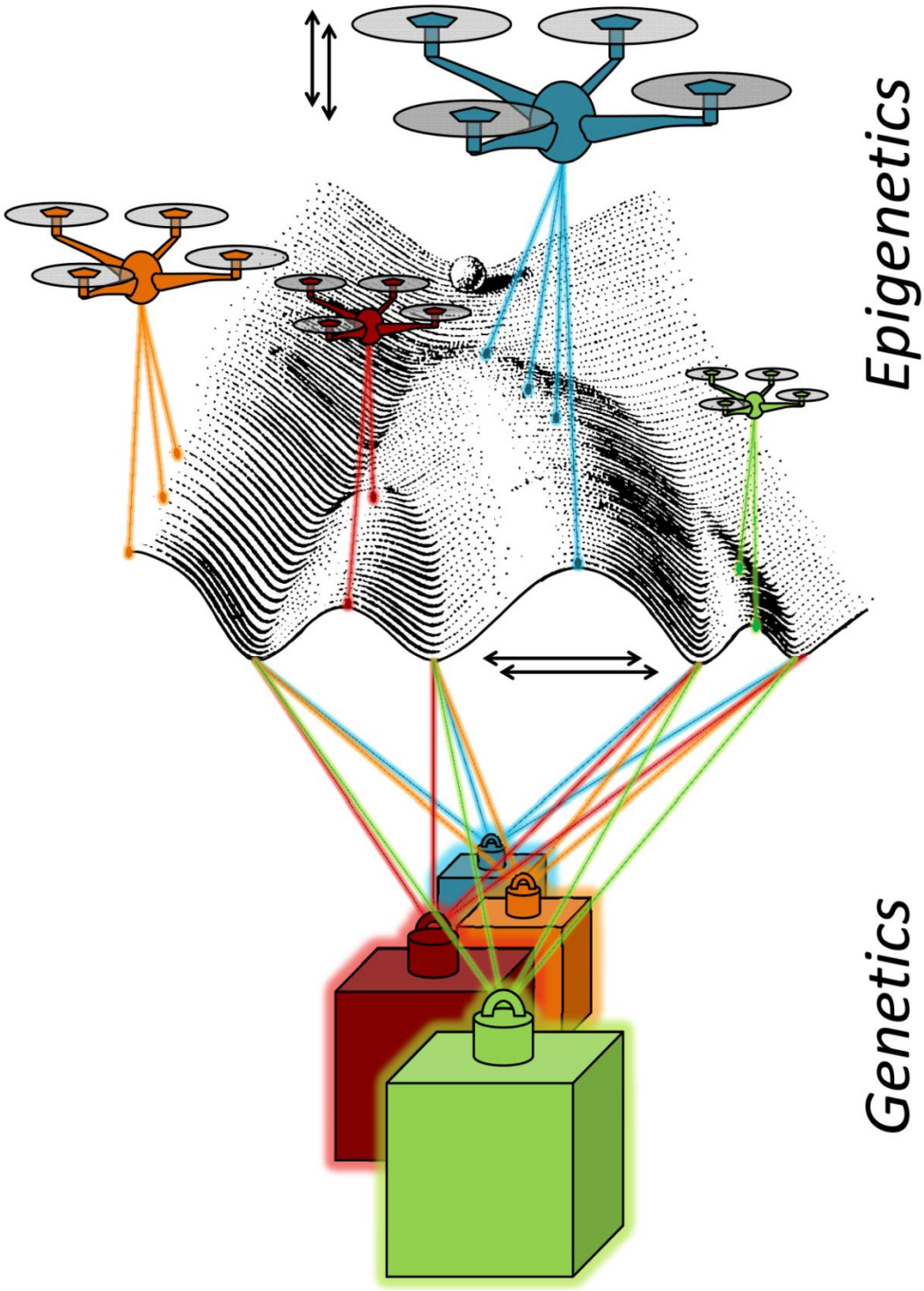


Figure 1.8. Where Epigenetics Meets Genetics and Cell Identity. The epigenetic landscape is a metaphor to explain the way toward differentiation of a particular cell taking into account the different developmental pathways that a cell might take. These pathways are represented by dynamic and flexible ridges and valleys controlled by the gene regulation derived from both genetic and epigenetic pressures (ropes). Genetic (weights) and epigenetics (drones) patterns (movements) establish diverse routes, cooperating to delineate the way from where a cell rolls down towards differentiation. *Adapted from Conrad Hal Waddington, The strategy of the genes,1957*[171].

genomic stability[192-196]. Other physiologic processes importantly regulated by epigenetics comprise genomic imprinting[197,198] and X-chromosome inactivation[199,200]. Epigenetics can therefore be understood as an essential regulatory mechanism layer, above genetics, able to modulate the majority of the genomic and transcriptomic processes, as well as those related to the maintenance of genomic stability. So, it "decides" what to make with the genetic information of a particular cell[201].

During the 20th century, a European subpopulation was submitted to a severe restriction of caloric intake by tragic historical reasons. This represents probably the best human model to study the trans-generational impact of the external environment, namely the relationship between diet, epigenetics and gene expression. The *Hongerwinter* (hunger winter) started in 1944, after the Second World War by limited food supplying to some Nazi-occupied regions in Holland due to the cut of vital supply routes in a rigorous winter. As a consequence, until May 1945 when the country was liberated, a severe restriction of caloric intake affected those populations, including pregnant women and their *in utero* offspring at different stages of gestation[202]. Children that experienced *in utero* famine were smaller, underweight and prone to have glucose intolerance. The offspring presented during their life increased susceptibility to diabetes, atherogenic lipid profile, obesity, coronary heart disease, disturbed blood coagulation, renal disease and increased stress responsiveness. Women submitted to early gestational famine exposure were documented to have higher risk of breast cancer[202-204]. Sixty years later, the imprinted *insulin-like growth factor-2* (*IGF2*) gene was found to present less DNA methylation in these individuals (offspring) in contrast to their same-sex siblings, not exposed to this nutritional privation during gestation. This precise periconceptional exposure reinforce that the establishment and maintaining of the epigenome at very early mammalian development is essential[205]. Transgenerational epigenetic studies suggested that epigenetically adapted phenotypes do not disappear suddenly between two consecutive generations. Instead of that, this dynamic epigenetic phenomenon should be understood as an erosion or slowly enhancement over multiple generations[206].

Some researchers consider that Epigenetics also interrogates the nucleosome positioning and ncRNAs, contributing to delineate the phenotype of a particular cell. Nucleosome positioning is a regulatory layer of gene expression that is able to block the accessibility of DNA to activators and to transcription factors, or to prevent the elongation of the

transcripts by RNA polymerases. Gene transcription is then regulated according to DNA packaging into nucleosomes and to their exact positioning around the TSSs. A disturbance in the position of the nucleosomes can change the activity of the RNA polymerases[207]. For instance, the loss of a nucleosome activates downstream promoters being associated with gene activation, whereas the overlapping between a nucleosome and a TSS is correlated with gene silencing[208-210]. Interestingly, this epigenetic mechanism is also involved in sculpting the DNA methylation landscape[211].

Concerning ncRNAs, these are epigenetic regulators that silence or target coding messenger RNAs (mRNAs), being also responsible for chromatin remodeling[212,213]. All these layers of epigenetic control have an additive role and act in symphony to control genome stability and cell homeostasis **(Fig. 1.9)**[214-217].

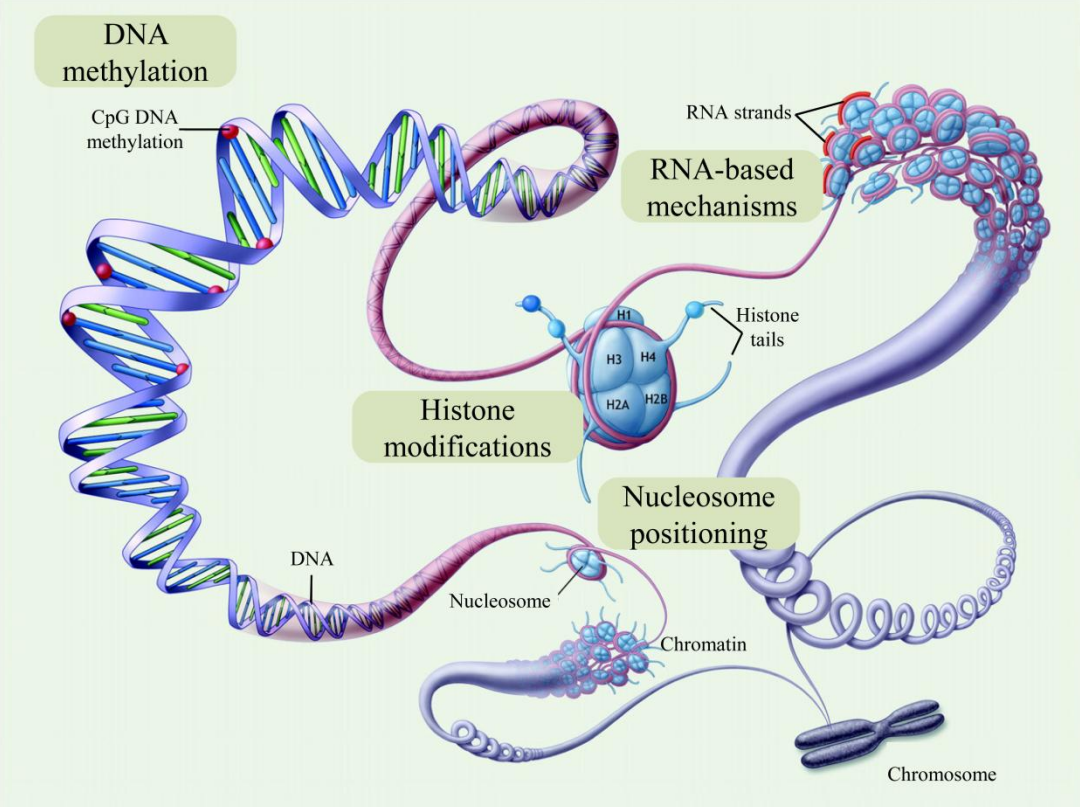


Figure 1.9. Epigenetic Mechanisms of Gene Regulation. In the nucleus, the double strand of DNA is packaged by an octamer of histones, forming a complex called chromatin. Undergoing further higher level of condensation, the chromatin give rise to chromosomes. The main epigenetic mechanisms able to modulate the chromatin structure and consequently gene expression are DNA methylation, histone post-transcriptional modifications, ncRNAS and nucleosome positioning. *Adapted from Matouk and Marsden, 2008*[214].

2.1 Histone Modifications

In 1884, Albrecht Kossel isolated by the first time histones and at that time suggested they could have a key role in transcriptional gene regulation[218]. Almost one century later, Kornberg and Thomas proposed the existence of a particle that consisted in an octamer of histones (H2A, H2B, H3 and H4) structurally organized, with approximately 147 base pairs of DNA wrapped in superhelical turns. The nucleosome was defined as the unit of the chromatin, allowing chromatin packaging and chromosome formation[219,220]. N-terminal histone tails can experience up to 16 classes of post-transcriptional covalent modifications, modulating nucleosome dynamics and chromatin structure by altering noncovalent connections within and among nucleosomes. These modifications comprised methylation, acetylation, phosphorylation and ubiquitination **(Fig 1.10)**[221,222]. Specifically recognized and mediated by particular proteins, these modifications become effective by recruitment of remodeling enzymes and chromatin modifiers[222]. Accordingly, the blockade or allowance of transcriptional activity of a specific genomic locus depends on the combination of the associated covalent histone modifications that are coupled into different levels of chromatin packaging[222-225].

The euchromatin, a relaxed chromatin state with an open conformation, is associated to an active transcription and is characterized typically by high levels of acetylation of histone 3 and histone 4 and di- or trimethylation of histone 3 lysine 4 (H3K4me2 or H3K4me3). By contrast, the heterochromatin, a condensed chromatin state with a close conformation, shows high levels of H3K9me and H3K27me, among other histone covalent modifications. The role of chromatin during gene transcription is assured by nucleosomal positioning and distribution of these post-translational histone modifications throughout the genes, minutely defined according to the relative location of the open reading frame (ORF), core promoter and upstream region[208,226,227]. The transcription and repression levels are determined by the accessibility of those regions controlled by the nucleosome positioning patterns. In summary, transcriptionally active and inactive promoter regions have low and high nucleosome occupancy, respectively[228,229].

All biological processes based on the genomic sequence, such as DNA repair, mitotic replication, meiotic recombination and transcription, are regulated by histone-modifying

enzymes in combination with DNA methyltransferases (DNMTs) and methyl-binding proteins (MBPs)[50].

The histone code changes dynamically according to the cellular requirements at certain moment, for instance, in order to assist or block gene transcription[230]. The histone marks belong to a language that can be interpreted, erased or modified by additional histone modifications. The proteins that recognize, catalyze and remove these specific chemical modifications are known as readers, writers and erasers, respectively[231]. The chromatin-interacting protein families include among others, histone methyltransferases (HMTs), histone acetyltransferases (HATs), histone demethylases (HDMs), histone deacetylases (HDACs) and histone kinases and phosphatases[232].

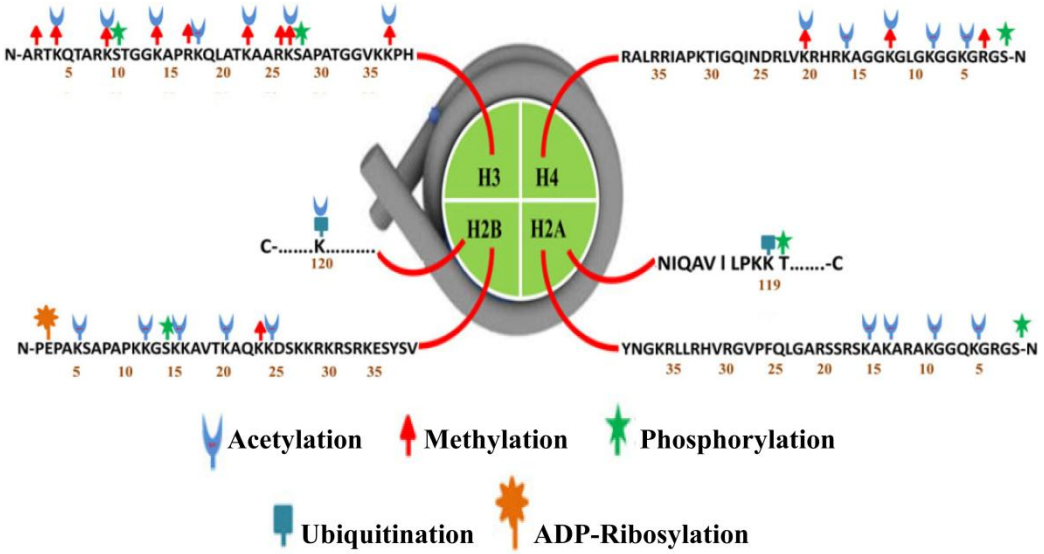


Figure 1.10. Histone Tail Post-Translational Modifications. In a context-dependent manner, histones undergo reversible covalent modifications, such as acetylation, methylation, phosphorylation, ubiquitination and ADP-ribosylation. *Adapted from Azad and Tomar, 2014[221].*

2.2 DNA methylation

The DNA methylation is a widely studied epigenetic modification in mammals[233]. In 1975, both Riggs and Holiday suggested that this DNA modification could affect gene expression[234,235], being nowadays associated to silencing of both coding and non-coding genes[236,237]. CpG dinucleotides are the most common targets of DNA methylation. The covalent inherited addition of a methyl group from S-adenosyl-methionine occurs in the 5' carbon of the cytosine ring (5-methylcytosine, 5mC)[238].

The global CpG dinucleotide under-representation, across mammalian genome, is contrasted exceptionally by loci of large repetitive sequences (endoparasitic sequences, centromeres and telomeres) and short and dense extensions of CpGs, frequently associated to gene promoters, called CpG Islands (CGIs). The ascription of these loci to CGIs are based on prediction algorithms with ad hoc thresholds, being the prediction accuracy variable[239]. The most used algorithm defines a CGI as a genomic region longer than 200 bases, presenting at least a content of 50% of guanines and cytosines and an observed-to-expected CpG ratio greater than 0,6[238,240,241]. According to these parameters, despite 72% of all mammalian gene promoters have a high CpG concentration[242], only 50% is considered to be linked with at least one CGI[243].

Mammalian developmental studies have shown the importance of dynamic alterations in DNA methylation[244,245]. The human genome holds approximately 29.000 CGIs[246], corresponding to 5% of all CGs and 1% of the entire genome[247]. They are spread in a non-random pattern, preferentially close to TSSs in the promoter regions or in the first exons[246,248,249]. Despite the large majority of the epigenetic studies about the transcriptional effect of DNA methylation are focused in CGIs, their upstream regions up to 2Kb, termed *CGI shores*, are the most susceptible to hypermethylation and hypomethylation events and correlate strongly with gene expression[250-252].Throughout mammalian genomes, cytosines of around 75% of CpG dinucleotides are methylated. The exception resides in those clustered in CGIs that are generally loci protected from DNA methylation[239,247,253]. The methylation status of cytosines at CpGs in CGIs is associated to stable and inherited patterns of activity regulation of downstream genes of these promoters. DNA methylation of CGIs is normally connected to gene silencing[248,254].

In normal cells, heavy DNA methylation of repetitive sequences prevents translocations, chromosome instability and gene disruption, by the reactivation of parasitic sequences[255-257]. Curiously, the acquisition of transposable elements through the eukaryotic genome overlapped co-evolutionarily with epigenetic inhibitory mechanisms able to block the mobility and silence their expression[258]. Nevertheless, they may have also evolved to participate in tissue-specific regulatory networks, lacking DNA methylation in those tissues and exhibiting enhancer activity[259]. Genomic imprinting is described by the transcriptional silencing associated to the

hypermethylation of one of the parental alleles[260,261]. Evolutionarily, it is though that X-chromosome inactivation and imprinting could have evolved together[262]. In 1983, Ehlrich and colleagues described the existence of tissue-specific differences at DNA methylation levels across mammals[263]. In terms of cell differentiation, about 5% of the CGIs are differentially methylated in a tissue-specific manner and some of them, namely those within promoter regions, are associated with tissue-specific expression profiles being crucial during embryonic development[264]. Evolutionarily, the atypical methylcytosine-thymine mutation and consequent depletion, could explain the genome-wide decrease of cytosines and guanines to about 40% of all nucleotides. Accordingly, the genomic frequency of CpG dinucleotides is around 25% of their predictable occurrence[265,266].

There is a non-random profile of DNA methylation across the genome. Classical studies consider the individuality of each CpG dinucleotide and the corresponding DNA methylation status without any relationship among them. Novel insights about CpG distribution suggested the existence of a strong association between DNA methylation patterns and the density and number of CpGs. According to this bimodality of cluster methylation, CpGs distributed in large and dense clusters are generally hypomethylated, whereas clusters with a sparse distribution or less CpGs are mostly hypermethylated. Accordingly, a certain methylated or unmethylated CpG exerts a positive feedback at nearby CpGs, by long-range methylation or short-range demethylation, respectively[267].

Mammalian DNA methylation is established and maintained by the functional combination of DNMTs. Thus, this epigenetic mark is inherited over the mitotic or meiotic cell divisions[238] through three main enzymes called DNMT1, DNMT3A, and DNMT3B. Their functional overlapping is responsible for mammalian development and cellular differentiation[268]. DNMT1 methyltransferase is responsible for the maintenance of the DNA methylation. It binds preferentially to hemimethylated double-strands and restores the entire methylated CpG dinucleotides after DNA replication, copying the methylation profile from the old strands to the new ones **(Fig. 1.11)**[269,270]. By contrast, DNMT3A and DNMT3B are the DNA methyltransferases responsible for the *de novo* methylation, being crucial during mammalian development at very early stages[271,272]. In combination, all these three DNMTs have an indispensable functionality, since their lacking result in embryonic lethality or impaired embryonic development[272,273].

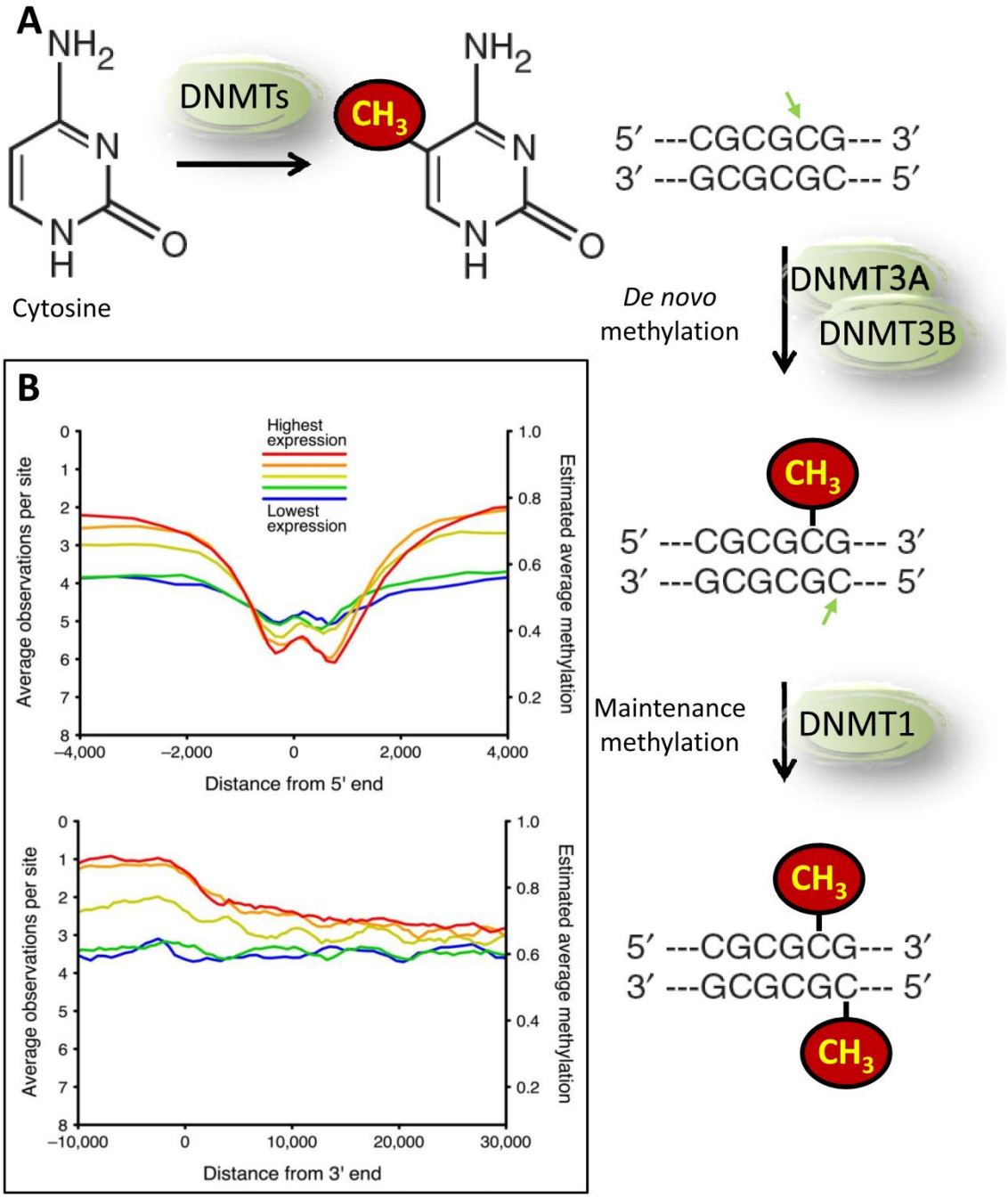


Figure 1.11. DNA Methylation Mechanisms. (A) In a CpG dinucleotide context, DNA methylation occurs at cytosine bases when a methyl group is added at the 5' position on the pyrimidine ring. DNMT3A and DNMT3B are involved in the *de novo* methylation and DNMT1 is responsible for the subsequent methylation of the hemi-methylated DNA at the complementary strand. (B) Genome-wide correlation between methylation and transcription. Highly expressed genes have low DNA methylated promoters and high DNA methylated gene bodies, maintaining a high DNA methylation until the 3' end. *Adapted from Ball et al., 2009[248] and Day and Sweatt, 2010[271].*

There is a genome-wide correlation between promoter CGIs hypermethylation and transcriptional silencing in different cell types[274-276]. This repression is based on the spatial impediment for the accessibility of transcription factors to their binding sites[277] and on the decrease on the affinity of DNA to enroll into nucleosomes, changing nucleosome positioning around TSSs[278]. This model of transcriptional repression is supported by a methyl CpG-binding domain (MBD)-containing family that attaches to methylated CpGs altering the nucleosomal architecture, translating the information given by the DNA methylation into a function or activity[279]. This family of proteins is composed by the methyl CpG binding protein 2 (MeCP2), methyl-CpG binding domain protein 1 (MBD1), MBD2, MBD3, MBD4, MBD5 and MBD6[280]. MBD3, MBD5 and MBD6 despite present a conserved methyl binding domain, *in vitro* they are unable to bind efficiently to methylated DNA. By contrast, in a DNA sequence–independent way, MeCP2, MBD2 and ubiquitin-like with PHD and ring finger domains 1 protein (UHRF1) (referred latter) bind strongly to methylated CpG dinucleotides. Additionally, two other enzymatic families recognize these DNA modifications, namely the SET- and RING-associated domain and the KAISO families. The first one comprises UHRF1 and UHRF2 and the second one is formed by Kaiso, zinc finger and BTB domain containing 4 (ZBTB4) and ZBTB38 proteins, acting through the zinc-finger domain C2H2[281,282]. Kaiso family proteins bind efficiently to methylated DNA and act in a sequence-dependent manner[280,283,284].

The discovery of 5-hydroxymethylcytosine (5hmC) modification changed in some way the epigenetic perspective around DNA methylation that was considered as a relatively invariable modification. This modification is quite abundant in specific tissues and the enzymes responsible for this dynamic modification, from the 5mC to 5hmC, are the 2-oxoglutarate- and Fe(II)-dependent oxygenases TET1[285], TET2[286] and TET3[287,288]. The 5hmC is regarded as an intermediary modification in DNA demethylation or a signal for transcription factors[289]. This class of TET proteins was first depicted in human myeloid malignancies[285-287] and the reversibility of the DNA methylation brought new issues to cell differentiation, embryonic development and cancer fields[290]. TET1 and TET2 are considered central players in pluripotency, while TET3, not existing in embryonic stem cells, possibly directs hydroxymethylation in differentiated cells, in addiction with TET1 or TET2[290,291].

In mammals, the DNA methylation was the first epigenetic mark described in dinucleotides CpG. The knowledge about their location is exponentially increasing at single-base resolution, bringing new pieces to complete the puzzle about the dynamic balance between genomic DNA methylation and its regulation[292]. Unknown pieces were recently revealed by the discovery of DNA methylation in a non-CpG context. CpHs (H=Adenine/Cytosine/Thymine) were identified in embryonic stem cells, disappearing upon induced differentiation and restored in induced pluripotent stem cells[293]. Moreover, another study showed that CpH methylation involves DNMT3A, is established *de novo* during neuronal maturation and suggests its possible importance in the nervous system as a key epigenetic modification[294].

2.3 Linking DNA Methylation and Histone Modifications

Evolutionarily, several proteins have more than one recognition domain, allowing the concomitant detection of different marks, arising from different spatial locations, promoting finally the adequate function according to the neighboring input they receive.

UHRF1 is a protein regulator of the chromatin structure and transcriptional activity, bridging two layers of epigenetic marks, specifically DNA methylation and histone-tail covalent modifications[295]. It comprises different structural motifs, namely an ubiquitin-like (Ubl) domain, a tandem tudor domain (TTD), a PHD domain, a SET and RING-associated (SRA) domain and a RING domain[296]. The UHRF1 SRA domain promotes the recruitment of DNMT1 to the hemimethylated DNA at the replication forks, promoting specifically the methylation of hemimethylated CpG dinucleotides[297,298], while the RING domain with E3 ubiquitin ligase regulates its stability[299,300]. Through its TTD, UHRF1 identify the H3K9me3 repressive mark and associates with unmethylated H3K4, directing the maintenance of DNA methylation[295,301]; and through its PHD motif that recognizes the unmodified Histone 3 Arginine 2 (H3R2) histone mark[302,303]. H3R2 methylation significantly abrogates PHD-mediated binding[304]. Briefly, through its different domains, UHRF1 catalyzes commonly the DNA methylation in a nucleosome context, establishing a relationship between DNA methylation and histone tails[305].

The role of MeCP2 and MBD2 in transcriptional repression is due to the recruitment of HDACs and co-occurrence within protein complexes associated to gene silencing[306,307]. The interaction of MeCP2 with the co-repressor complexes Sin3-HDAC and N-CoR–

SMRT endorse transcriptional repression, by association with histone deacetylase 1 (HDAC1) and HDAC3, respectively[308-310]. For instance, in the first case, while MeCP2 identify the DNA methylation, HDAC1 deacetylates the histone tails to promote transcriptional repression[308,310]. The existence of a link between DNA methylation and transcriptional silencing through specific histone modifications prompted to think about a higher layer of gene regulation, namely the derived nucleosomal conformation. The nucleosome-remodeling and histone deacetylase (NuRD) complex have both HDAC and ATP-dependent nucleosome disruption activities. This complex helps the transcriptional repression by assisting repressors to contact to chromatin and interacting with MBD2 and MBD3 in a reciprocally exclusive mode[311,312]. The histone core deacetylation catalysis converts the open and transcriptionally active chromatin into a closed structure, isolated from the transcriptional machinery. The link between promoter CGI methylation, MBDs, HDACs and a chromatin remodeling machinery fulfill the set of causes and consequences that lead to the transcriptionally inactive chromatin state, being finally translated in gene repression in cis[280,313].

2.4 Epigenetic Modifications in Cancer

The onset and progression of cancer is characterized by the cooperation between genetic and epigenetic changes that create variability in cell populations. Through a continuous cellular Darwinian evolutionary process, cells with advantageous features are selected according to their genome and epigenome[314,315]. Differentially DNA methylation in specific regions, histone tail modifications and altered gene regulation of chromatin-modifying enzymes characterize the epigenetic landscape of a cancer cell[62,207,316].

2.4.1 Histone Modifications

In cancer, aberrant patterns of histone modifications collaborate in the promotion of the tumorigenic process, varying the cellular epigenetic landscape[317]. Not always contributing for the malignancy event, these chemical alterations in histone tails are deregulated in cancer, being one of its hallmarks[50,318]. The enzymes responsible for the interpretation, addition and deletion of such posttranslational modifications of histone tails are often deregulated in terms of copy number, mutated or involved in translocations **(Fig. 1.12)**[319]. Post-translational modifications in terms of histone

acetylation, methylation, deacetylation and demethylation have been reported during tumorigenesis[231].

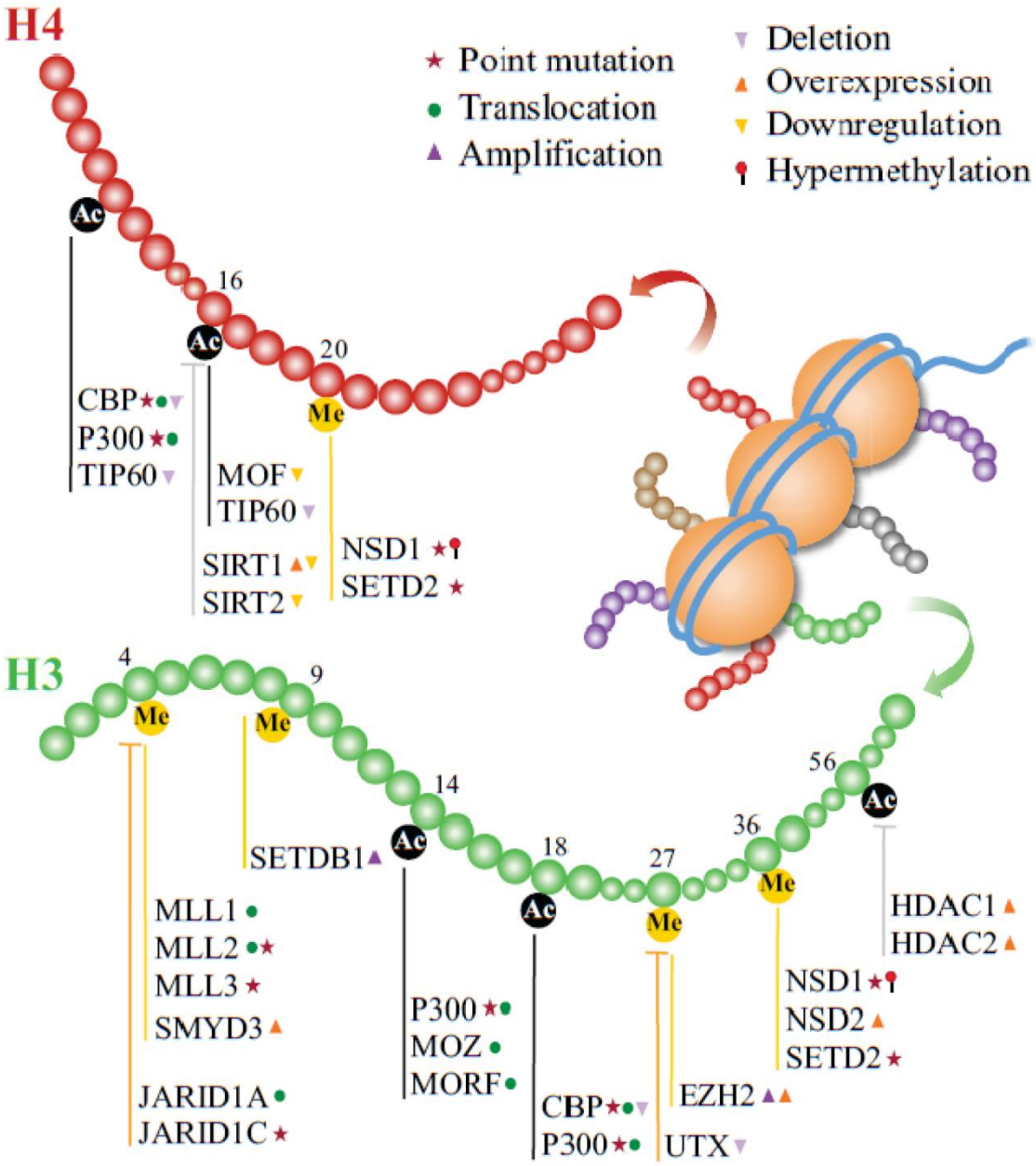


Figure 1.12. Histone Modifications in Cancer. The transcriptionally active or inactive chromatin state is synchronized by enzymes that control DNA methylation, histone modifications and ATP-dependent chromatin remodeling complexes. In cancer, the aberrant epigenetic landscape comprises abnormal patterns of histone modifications due to the deregulation of some of their regulators, such as HMTs, HATs, HDMs and HDACs. The figure summarize some of these epigenetic players and the affected post-transcriptional modifications. *Adapted from Simo-Riudalbas and Esteller, 2015*[231].

2.4.1.1 Histone Acetylation

Cancer progression is associated with HATs deregulation. The histone acetylase E1A binding protein p300 (EP300)[320] controls transcription through chromatin remodeling, playing a central role in cell differentiation and proliferation. Missense and nonsense mutations were reported in gastric, colorectal, breast, pancreatic and small cell lung (SCLC) cancers[321-323]. It was also shown that the *EP300* locus undergo LOH in colon, breast and ovarian cancers[324]. The ability of E1A adenovirus to induce replication of human quiescent cells is based on the E1A-p300/CBP interaction, promoting a global hypoacetylation of H3K18 and restricting H3K18ac to a limited set of genes. As a result, cell cycle is stimulated, promoting cell growth and division[325,326]. In SCLC, different mutations were found in the HAT *cAMP response element-binding protein* (*CREBBP*)[323]. Moreover, genomic loss of the *K(Lysine) Acetyltransferase 6B* (*KAT6B*) tumor suppressor was associated with decrease in H3K23ac[327]. Mutations in the transcriptional activator *Myocyte Enhancer Factor 2B* (*MEF2B*), reported to interact with HDACs, were found in non-Hodgkin lymphomas[328]. Mono-allelic loss of *KAT5* in lymphomas, head-and-neck and breast cancer decrease its expression levels, reducing its modulation in DNA-damage response and increasing malignancy[329]. Chromosomal translocations result frequently in aberrant fusion proteins. Often they affect some subtypes of leukemia and involve HATs like *EP300*[330,331], *MOZ* (*MYST3/KAT6A*)[332], *MORF* (*MYST4/KAT6B*)[333] and *CREBBP* (*CBP*)[332,334].

At histone level, one of the major alterations experienced in cancer cells is the overall loss of H4K16ac[318]. Several studies pointed out the presence of genetic mutations or disrupted expression of HDACs leading ultimately to the impaired expression of genes involved in the tumorigenic process, namely those controlling cell-cycle regulation, cell proliferation and apoptosis[335].

2.4.1.2 Histone Deacetylation

The HDACs are another class of histone modifiers with aberrant expression in cancer[335,336]. For instance, *HDAC1* is overexpressed in colon[337], breast[338,339], gastric[340], prostate[341], pancreatic[342], ovarian[343], cervical[344] and liver[345] carcinomas. An increased expression of *HDAC2* is commonly observed in breast[338], cervical[344], gastric[340], colorectal[337,346] and hepatocellular[345] cancers. The same HDAC up-regulation is

observed for multiple tumor types, such as *HDAC3* in colon[346,347] and breast[338]; *HDAC4* in colon[348], stomach[349] and prostate[350]; *HDAC5* in sarcoma[351], breast[352], melanoma[353] and colon[346]; *HDAC6* in melanoma[353], liver[354] and glioblastoma[355]; *HDAC7* in pancreas[356] and colon[346]; and *HDAC8* in urethra[357] and neuroblastoma tumors[358]. There is a controversy about the tumorigenic function of *HDAC2*. Despite being overexpressed in some colorectal cancer patients[337,346], this HDAC was also found to bear an inactivating mutation in some sporadic colorectal carcinomas with MSI. In clinics, this duality should be taken into account, as it has a potential significance for the best pharmacological treatment selection[359-361].

An increased expression of *sirtuin 1* (*SIRT1*) was observed in colorectal[362], hepatocellular[363], ovarian[364] and breast[365] carcinomas. Controversially, *SIRT1* frameshift mutations were found in colorectal and gastric carcinomas with MSI[366]. *SIRT2* overexpression was associated with uveal melanoma[367] and hepatocellular carcinoma (HCC) metastasis, mediating the epithelial to mesenchymal transition in the later one[368]. Highly expression of *SIRT3* was observed in papillary thyroid carcinomas[369] while repression of *SIRT4* expression point out its tumor suppressive function in colorectal[370] and breast[371] cancers. Finally, *SIRT7* up-regulation was demonstrated to have an important role in the human colorectal development and progression[372].

2.4.1.3 Histone Methylation

The tumorigenic development and progression is characterized by an unusual pattern of histone methyl marks genome-wide. This abnormality is instigated by the abnormal activity of HMTs and HDMs, either by an altered gene expression regulation, or by CNV and chromosomal translocations[194,373,374]. According to these changes, tumors are found to have increased H3K9me3[375,376] and H3K27me3[377,378] repressive marks; and decreased H3K4me3 active mark[318] and H4K20me3 repressive mark[379].

The *MLL1* H3K4 methyltransferase gene harbors one of the biggest promiscuous recombination hot spots of the human genome. In cancer, recombination events occur with more than 50 *MLL* fusion partners, contributing to its partial tandem duplication and failure of its methyltransferase activity[380,381]. Accordingly, *MLL* translocations are found in more than 70% of pediatric leukemias and in about 10% of AMLs in adults[381]. Apart of the leukaemogenic *MLL* translocations, the disruption of *MLL* genes also occur

in solid tumors such as prostate[382], gastric[383], head and neck[384], bladder[385], hepatocellular[386], squamous cell lung[387] and SCLC[323] tumors. Gene silencing mediated by the overexpression of *EZH2* H3K27-methyltransferase has been observed in solid malignancies such as testis, larynx, bladder, stomach and breast cancers, among others[388]. In epithelial ovarian cancer, EZH2-mediated methylation and silencing of *Vasohibin 1* (*VASH1*) gene promotes tumor angiogenesis[389]. Nevertheless, in Kaposi sarcoma tumors, the stimulation of angiogenesis is achieved by EZH2 acting as a transcriptional gene activator, inducing the proangiogenic factor *Ephrin-B2* (*EFNB2*)[390]. This dual role in cancer is supported by the fact that EZH2 is susceptible to inactivating mutations in B-cell lymphomas of germinal-center origin[328] and myelodysplastic syndromes[391]. The hypertrimethylation of H3K27 was described in human B-cell lymphoma, by contrast, as a result of a EZH2 gain-of-function mutation[392].

During tumorigenesis, the epigenetic regulation of the *SET and MYND domain containing 3* (*SMYD3*) H3K4 methyltransferase increases the level of cancer promoting-genes such as *matrix metallopeptidase 9* (*MMP9*), associated with metastases[393]. Accordingly, overexpression of *SMYD3* was observed in colorectal carcinoma[394], hepatocellular cancer[394], breast cancer[395], pancreatic ductal adenocarcinoma[396] and lung adenocarcinoma[396]. Recently, it was described that this HMT oncogene is required for both liver and colon chemically induced carcinogenesis in mice[397].

Other HMTs were described to play a key role in other types of cancer. For instance, *SET Domain, Bifurcated 1* (*SETDB1*) H3K9 methyltransferase was shown to be recurrently amplified in liver cancer[398], melanoma[399] and in NSCLC and SCLC[400]. Another study showed that the inhibition of euchromatic histone-lysine N-methyltransferase 2 (EHMT2) in neuroblastoma, specifically reduces global H3K9me2, decreasing growth and inducing apoptosis[401]. Conversely, translocation-mediated silencing of *nuclear receptor binding SET domain protein 1* (*NSD1*) H3K36 and H4K20 HMT or its transcriptional silencing associated with CGI-promoter hypermethylation was observed in AMLs[402,403], and gliomas and neuroblastomas[404], respectively.

2.4.1.4 Histone Demethylation

Several enzymes of the HDMs family are deregulated in cancer, contributing to specific histone tail patterns inherent to an advantageous epigenetic landscape. For instance,

lysine (K)-specific demethylase 1 (LSD1) H3K4 demethylase represses transcription by exercising its function in H3K4me1 and H3K4me2, which are histone marks characteristic of an active transcriptional state[405-407]. Its overexpression and consequent transcriptional repressive function, was observed in colorectal cancer, regulating genes related to proliferation and metastasis[408]. The up-regulation of *LSD1* and concomitant support of human carcinogenesis, through chromatin regulation, supporting mechanisms such as proliferation, migration and invasion[409], was also confirmed in breast[410], prostate[411], bladder and NSCLC[412] tumors, among others. Several studies suggested that despite its function as a transcriptional repressor, the LSD1 interaction with androgen or estrogen nuclear receptors could change its histone modification target from H3K4me1/me2 to H3K9me1/me2, changing its function to a transcriptional activator[413]. A cooperative demethylation activity is achieved by the LSD1 (H3K9me1/me2 demethylase) interaction with jumonji domain containing protein 2C (JMJD2C) (H3K9me3 demethylase), promoting androgen receptor-dependent gene expression[414]. Accordingly, members of the JMJD2C H3K9me3 demethylase subfamily are over-expressed by an increase in their genomic copy number in breast cancer[415], esophageal squamous cell carcinoma cell lines[416], medulloblastoma[417,418], Hodgkin lymphomas[419] and primary mediastinal B cell lymphomas (PMBL)[419]. Jumonji, AT-rich interactive domain 1 (JARID1) family and lysine (K)-specific demethylase 2B (KDM2B, also known as FBXL10) are in charge of the demethylation of H3K4me2/me3; and H3K36me1/me2 and H3K4me3, respectively[420]. They are up-regulated in breast[421], NSCL[422], SCL[422] and bladder[422] tumors; and in leukemias[423,424], SEs[425] and pancreatic ductal adenocarcinomas[426], correspondingly. Lastly, *KDM6A* H3K27 demethylase undergo inactivating mutations in some types of malignancies[427-430].

2.4.2 DNA Methylation

The knowledge of an increasing number of genes showing epigenetic alterations in cancer emphasizes their significance in this disease. During tumor progression in cancer, the specific DNA methylation profile of a cancer cell changes progressively according to the original tissue and tumor stage. These genome-wide events are nowadays easy to characterize, being suggested as a new generation of biomarkers for diagnosis, prognosis and as predictors of drug response[431].

2.4.2.1 DNA Hypomethylation

One of the first epigenetic abnormalities discovered in cancer was the global DNA hypomethylation[432]. More than 30 years ago, Feinberg and colleagues were able to distinguish human tumors from their correspondent normal tissue based on DNA methylation, namely hypomethylation[432]. Twenty years later, a global hypomethylation in all normal tissues in mice was achieved by generating mice carrying a hypomorphic *DNA methyltransferase 1* (*Dnmt1*) allele. Importantly, it was demonstrated that DNA hypomethylation induced CIN and tumor formation[433,434].

Hypomethylation events in cancer are known to occur in repetitive sequences, CpG-poor promoters, CpG-rich promoters and introns[50,435]. For instance, hypomethylation of LINE-1 and latent viral sequences incorporated in the genome, usually repressed through DNA methylation, can be activated and contribute to cancer progression[436-439]. Retrotransposon activation can cause genomic instability by mitotic recombination, through deletions, translocations and chromosomal rearrangements[433]. An altered gene expression can be achieved by mutagenic insertion in non-coding regulatory regions (indirectly contributing to transcriptional regulation) or within a gene itself , by creating new exons, by altering the alternative splicing, or by creating new polyadenylation sites[258,440]. So, during cancer progression the global hypomethylation reactivates transposable elements, leading finally to different patterns of gene expression, LOH and aneuploidy **(Fig. 1.13)**[441-445].

The balance of active and inactive transcribed genes can be disrupted by a direct regulation of DNA methylation. Promoter DNA hypomethylation can lead to the reactivation of genes or specific isoforms generally silenced in normal cells, contributing to cancer phenotype. Hypomethylation-dependent overexpression of several coding and non-coding genes can be observed in several cancer types in both primary tumors and metastasis. Some of the oncogenes undergoing this epigenetic re-expression include r*elated RAS viral oncogene homolog* (*R-RAS*) in gastric cancer[446], *wingless-type MMTV integration site family, member 5A* (*WNT5A*) in prostate cancer[447], *S100 calcium binding protein A4* (*S100A4*) in colon cancer and pancreatic ductal adenocarcinoma[448,449], *miR-191* in HCC[450] and *miR-128a* in ALL[451,452]. A recent study demonstrated that a hypomethylation event reactivates a short isoform of *TBC1 domain family, member 16* (*TBC1D16*) that exacerbates melanoma growth and metastasis, both

in vitro and *in vivo*, being associated with a poor clinical outcome in melanoma[453]. Genomic imprinting could also be affected by DNA hypomethylation, activating the transcription of maternal or paternal imprinted loci. For instance, loss of imprinting in cancer can affect the *IGF2* gene in prostate[454], breast[455], colorectal[456] and Wilms tumors[457].

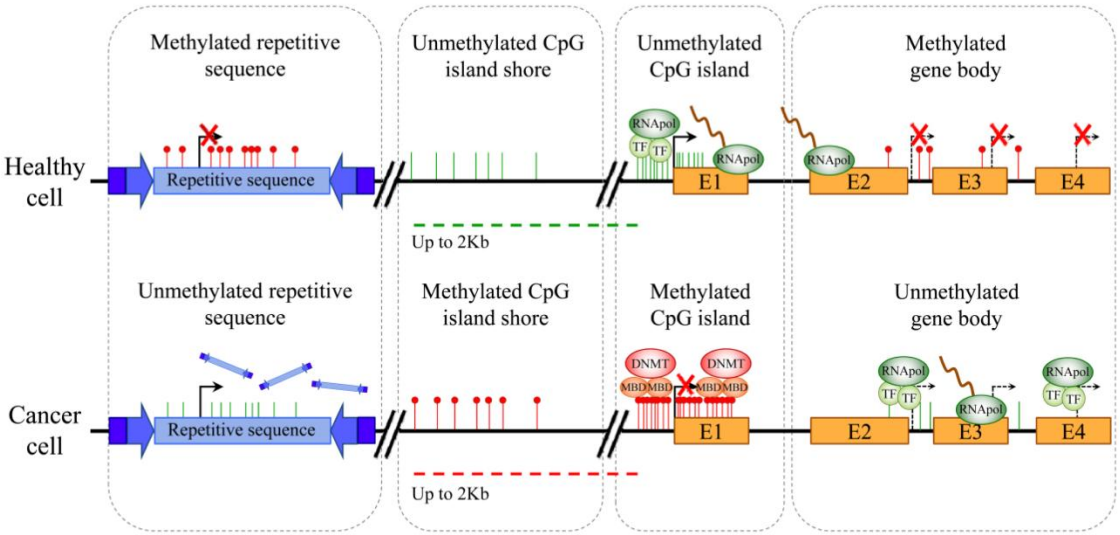


Figure 1.13. DNA Methylation Patterns in Cancer. The global hypomethylation is a general event in cancer that leads to chromosomal instability, translocations and gene disruption through the activation of repetitive sequences normally repressed by hypermethylation. At gene bodies, the decrease in DNA methylation levels allows the transcription to be initiated at several incorrect sites. In contrast, the hypermethylation at CGIs and shores leads to the repression of several tumor suppressor genes. *Adapted from Portela and Esteller, 2010*[207].

Overlapping with tissue-specific differentially methylated regions, CGI shores are associated with epigenetic reprogramming and cancer. The hypomethylation of CGI shores, initially hypermethylated in normal cells, have been associated with transcriptional activation of c*aveolin-1* (*CAV1*) in highly aggressive breast cancer and *hexokinase 2* (*HK2*) in HCC. Curiously, in the later case, the CGI itself that is unmethylated in normal cells can experience methylation through a CIMP, repressing the expression of HK2[458,459].

2.4.2.2 DNA Hypermethylation

The hypermethylation fingerprint during cancer progression is not random. In cancer, CGI promoter hypermethylation is involved in the inactivation of tumor suppressor

genes **(Fig. 1.13)**. This disruption leads to the failure of main cellular pathways such as DNA repair, cell cycle control, apoptosis and cell adhesion. As a result, the inactivation of specific genes confers thereby a proliferative advantage, resulting in clonal selection[316]. According to the Knudson's two-hit hypothesis, this event is functionally equivalent to a genetic mutation with a similar or even higher frequency[460,461]. DNA hypermethylation could affect one or both alleles of a tumor suppressor gene, depending on the mutation (inactivation) status of each allele prior to this event[460]. The first described tumor suppressor gene[20] inactivated by DNA hypermethylation was the *retinoblastoma 1 gene* (*RB1*)[462]. Soon after, similar hypermethylation events were found in several genomic loci, sustaining the mechanistic hypothesis of epigenetic transcriptional silencing of tumor suppressor genes by DNA methylation in cancer[63].

In cancer, a very large and increasing number of tumor suppressor genes are known to be inactivated by promoter hypermethylation, such as *BRCA1* in breast and ovarian cancer[463]; *O-6-methylguanine-DNA methyltransferase* (*MGMT*) in colorectal cancer and gliomas[464]; and *MLH1* in colon[465], endometrial[466] and gastric[467] cancers. The hypermethylation of normally unmethylated CGI shores was also observed in pancreatic adenocarcinomas[468] and bladder cancer[469].

DNMTs are responsible to sculpt the epigenetic landscape in terms of DNA methylation, having an abnormal activity in several malignancies. For instance, highly recurrent somatic missense mutations in *DNMT3A* have been observed in around 20,5% of the patients with AML, predicting their prognostic and therapeutic response to chemotherapy[470-472]. The mutation of *DNMT3A* decreases its enzymatic activity[472], by dominantly blocking the capacity of wild-type DNMT3A to form active tetramers[473]. Several cancer types are characterized by an increase in the expression of *DNMT1*, *DNMT3A* and *DNMT3B* levels and an associated hypermethylation of CGIs[474]. Polymorphisms in *DNMT3B* gene promoter region lead to its increased activity and are linked to an earlier onset of HNPCC[475] and an increased risk of lung[476], breast[477] and prostate[478] cancers. By other hand, changes in the cellular context during tumor progression can change not only the activity of DNMTs but also their recruitment to specific target genes. In leukemias, the expression of the promyelocytic leukemia-retinoic acid receptor alpha fusion oncoprotein (PML-RARA) and interaction with

DNMT1 and DNMT3A lead to the promoter DNA methylation-dependent silencing of the *retinoic acid receptor β2 gene* (*RARβ2*)[479].

Other DNA hypermethylation-related abnormalities have been shown in cancer, namely related to MBD and TET families. A study conducted in pre and postmenopausal women, by genotyping two SNPs within the *MBD2* gene, one intronic and one in the non-coding exon, has shown an altered breast cancer risk dependent on menopausal status[480]. Impaired TET2 activity, concomitant to an inefficient conversion of 5mC to 5hmC was observed in hematological malignancies[481,482], where this gene was also found to be homozygous mutated[483,484]. Genome-wide redistribution and loss in the hydroxymethylation levels was also observed in several solid tumors including prostate cancer, esophageal squamous cell carcinoma, HCC and cholangiocarcinoma[485-487].

In a genetic point of view, DNA methylation can also drive non-transcriptional events. Methylated cytosines (at CpGs) could be considered as endogenous mutagens and carcinogens in humans, increasing the potential for cytosine to thymine transition mutations (at least 10 times higher). Methylated CpGs are probably responsible for more than 30% of all known disease-related point mutations[488,489].

In summary, in terms of the epigenome, cancer cells undergo a genome wide DNA hypomethylation simultaneous to focal hypermethylation events in promoter regions, being selected in terms of the advantageous acquired epimutations[490,491]. During tumor progression, hypomethylation and *de novo* methylation could be functionally analogous to gain-of-function and loss-of-function genetic mutations, respectively. These changes in transcriptional activity affect mechanistically different pathways such as those related to apoptosis, cell adhesion, angiogenesis, cell-cycle regulation and DNA repair[62,274].

3. Non-Coding RNAs

3.1 Background

According to the central dogma of Biology by Francis Crick in 1958, ribonucleic acid (RNA) molecules were only considered as an intermediary between genes and proteins: "*the coded genetic information hard-wired into DNA is transcribed into individual transportable cassettes, composed of messenger RNA (mRNA); each mRNA cassette contains the program for synthesis of a particular protein (or small number of*

proteins)"[492]. Historically, few types of ncRNAs were considered functional, namely the ribosomal (rRNA) and the transfer (tRNA) RNAs, as they could explain the protein synthesis processes[493,494].

Intriguingly, in human genetics, one of the biggest mysteries was the fact that less than 2% of our genome is translated into proteins, with more than 98% having no coding potential, being considered as *junk* accumulated during the evolutionary process[495]. However, the establishment of a stronger positive correlation between the genomic proportion of non-protein-coding regions and the complexity of the different organisms, in comparison to protein coding genes, highlighted their importance. Evolutionarily, this *junk* part of the genome was conserved and enlarged according to the developmental complexity, suggesting that non-protein-coding regions underwent an advantageous selection. Curiously, genes holding large amounts of intronic sequences are considerably more expressed in the nervous system and downregulated in cancers[496]. This could suggest that these genes hold intronic regulatory sequences, probably related to cell division and tissue-specific gene expression.

Recently, the Encyclopedia of DNA Elements (ENCODE) Consortium have assigned biochemical functions for 80% of the genome, mapping regions of transcription, transcription factor association, chromatin structure and histone modifications[497]. Although the function of RNA as a regulatory molecule has been overlooked until a recent past, the *central dogma* was challenged by observing that 74,7% of the entire genome is actively transcribed under a given physiological context[498]. Thus, the information contained in a particular sequence of the genome can be converted into a RNA transcript which in turn can either be translated into a protein or influence the transcription or translation of other genes.

3.2 Molecular Functions and Epigenetics

The classical definition of epigenetics comprises two main layers of gene regulation: DNA methylation and histone post-translational modifications. Recently, several records in the molecular biology field have unmasked a new complex and dynamic layer of gene regulation abreast of the previous ones. This new conceptual revision is aware of the direct or indirect interaction of ncRNA molecules with proteins, DNA and other RNAs, molding the epigenetic landscape[499]. Accordingly, ncRNAs can bind and

recruit histone modifying complexes and/or modulate the activity of DNA methyltransferases, affecting the genome organization and expression of both protein coding and/or non-coding RNA molecules. In turn, the first RNA molecules can be transcriptionally or translationally regulated through the interaction with the second ones **(Fig. 1.14)**[499,500]. Importantly, ncRNAs not only affect gene expression of other RNAs, but they are regulated by epigenetic mechanisms themselves.

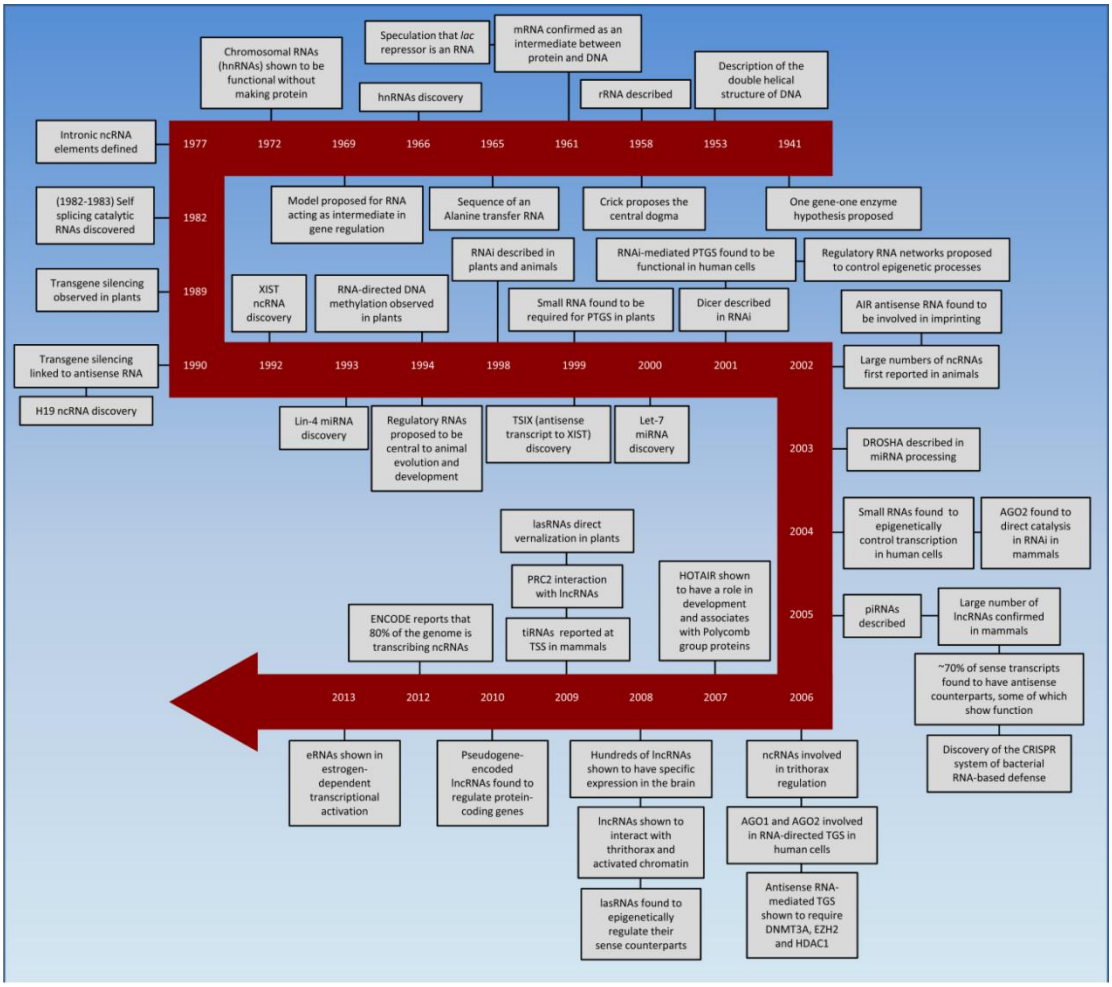


Figure 1.14. Timeline of Non-coding RNAs and their Regulatory Functions. AGO, Argonaute; *AIR*, also known as *AIRN* (*antisense of IGF2R non-protein coding RNA*); CRISPR, clustered regularly interspaced short palindromic repeat; DROSHA, Drosha, Ribonuclease Type III; *H19*, *H19 imprinted maternally expressed transcript*; hnRNA, heterogeneous nuclear RNA; *HOTAIR*, *HOX transcript antisense RNA*; PRC2, Polycomb repressive complex 2; PTGS, post-transcriptional gene silencing; RNAi, RNA interference; TGS, transcriptional gene silencing; tiRNA, transcription initiation RNA; *XIST*, *X inactive specific transcript*. *Adapted from Morris and Mattick, 2014*[500].

The functionality of ncRNAs has been approached by different methods. Depending on the cellular context of a cell, different molecular functions have been attributed to ncRNAs, participating crucially in epigenetic processes that control cell differentiation, development and cellular physiology. Their expression and direct or indirect targeting of proteins can be disrupted by several genetic and epigenetic factors in tumorigenesis, and other human diseases, such as neurological, cardiovascular and developmental diseases[501-503]. Particularly, the involvement of distinct classes of ncRNAs in cancer, is being translated by their annotation as good biomarkers for diagnosis, prognosis and predictors for drug response[504-509].

3.3 Classification

A large part of the genome is transcribed into RNAs, but only a small fraction encodes proteins[495]. The discovery of a complex layer of gene regulation above the *central dogma* mediated by several epigenetic mechanisms encouraged the characterization of their different entities. Among this network, ncRNAs assume a relevant importance by a range of different molecular mechanisms from the transcriptional regulation of coding genes until their conversion into proteins[510]. The later process is assisted by rRNAs and tRNAs which mediate the translation of a particular mRNA into a protein, by forming part of the ribosomal components and by the specific transport of amino acids directed by a three-nucleotide sequence (codon), respectively. Excluding the rRNAs and tRNAs, currently the transcriptome comprise two main groups of ncRNA molecules based on their length: small non-coding RNAs (sncRNAs) and long non-coding RNAs (lncRNAs), depending if they have less or more than 200 nucleotides, respectively (with some exceptions) **(Table 1.5)**[511].

sncRNAs are divided in different subclasses, being characterized by different lengths, cellular location, function and pathway of biogenesis. Among others, they comprise the following subclasses: small nuclear RNAs (snRNAs), piwi-interacting RNAs (piRNAs), small nucleolar RNAs (snoRNAs) and the most widely studied class of ncRNAs, the miRNAs. The later ones are ncRNAs of ~22 nucleotides that mediate transcriptional gene repression by limiting the translation of mRNA into proteins by different mechanisms, guided by sequence complementarity between the small RNA and the target mRNAs. It is estimated that more than 60% of protein-coding genes are regulated by miRNAs **(Fig. 1.15)**[512].

Table 1.5. Types of ncRNAs. Description of some of the most studied ncRNAs.

Name	Size	Location	Number in humans	Functions	Illustrative examples
Short ncRNAs					
miRNAs	19-24 nt	Encoded at widespread locations	> 1 424	Targeting of mRNAs and many others	miR-15/16, miR-124a, miR-34b/c, miR-200
piRNAs	26-31 nt	Clusters, intragenic	23 439	Transposon repression, DNA methylation	piRNAs targeting RASGRF1 and LINE-1 and IAP elements
tiRNAs	17-18 nt	Downstream of TSSs	> 5000	Regulation of transcription?	Associated with the CAP1gene
Mid-size ncRNAs					
snoRNAs	60-300 nt	Intronic	> 300	rRNA modifications	U50, SNORD
PASRs	22-200 nt	5′regions of protein-coding genes	> 10 000	Unknown	Half of protein-coding genes
TSSa-RNAs	20-90 nt	-250 bp and +50bp of TSSs	> 10 000	Maintenance of transcription?	Associated with RNF12 and CCDC52 genes
PROMPTs	<200 nt	-0,5 kb and -2,5kb of TSSs	unknown	Activation of transcription?	Associated with EXT1 and RBM39 genes
Long ncRNAs					
lincRNAs	>200 nt	Widespread loci	> 1000	Examples include scaffold DNA-chromatin complexes	HOTAIR, HOTTIP, lincRNA-p21
T-UCRs	>200 nt	Widespread loci	> 350	Regulation of miRNA and mRNA levels?	Uc.283+, uc.338, uc160+
Other lncRNAs	>200 nt	Widespread loci	> 3000	Examples include X-chromosome inactivation, telomere regulation, imprinting	XIST, TSIX, TERRAs, p15AS, H19, HYMAI

CAP1, *CAP, adenylate cyclase-associated protein 1*; *CCDC52*, *coiled-coil domain containing 52* (also known as *SPICE1*); *EXT1*, *exostosin 1*; *HOTTIP*, *HOXA distal transcript antisense RNA*; *HYMAI*, *hydatidiform mole associated and imprinted*; IAP, intracisternal A-particle; PASRs, promoter-associated small RNAs; PROMPTs, promoter upstream transcripts; *RASGRF1*, *RAS-protein-specific guanine nucleotide-releasing factor 1*; *RBM39*, *RNA - binding motif protein 39*; *RNF12*, *ring finger protein 12* (also known as *RLIM*); TERRAs, telomeric repeat containing RNAs; tiRNAs, transcription initiation RNAs; TSSa-RNAs, TSS-associated RNAs; *TSIX*, *XIST antisense transcript*. *Adapted from Esteller, 2011*[503].

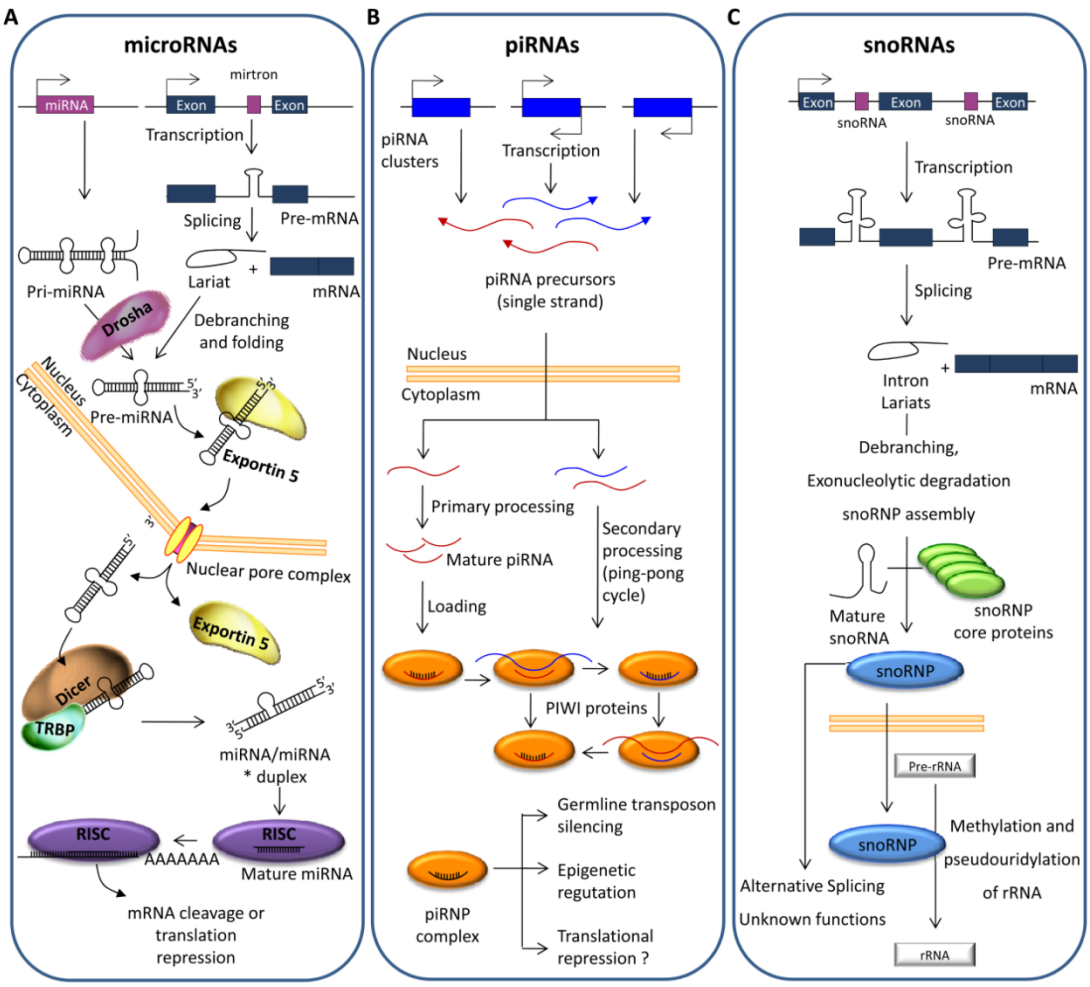


Figure 1.15. Biogenesis and functions of ncRNAs. (A) miRNAs are transcribed as individual transcripts (pri-miRNAs) or encoded in introns of host genes (mirtrons). These transcripts undergo further processing by the Drosha complex or the lariat-debranching enzyme, respectively, giving rise to precursor miRNAs (pre-miRNAs). In turn, following their export to the nucleus and further processing by Dicer and TAR RNA-binding protein 2 (TARBP2), the mature miRNAs generated are loaded into the RNA-induced silencing complex (RISC) and participate in targeted mRNA cleavage or translation repression. (B) piRNAs are generally encoded in mono-or bidirectional clusters. Following a PIWI-protein-catalysed amplification loop (called the "ping-pong cycle"), via sense and antisense intermediates, additional piRNAs are produced and subsequently loaded into PIWI ribonucleoprotein (piRNP) complexes, participating in germline transposon silencing and epigenetic regulation. (C) SnoRNAs are mainly located in introns of both coding an non-coding genes. Following splicing, debranching and trimming, snoRNAs function as small nucleolar ribonucleoproteins (snoRNPs), guiding methylation and pseudouridylation of rRNA. They also participate in alternative splicing and some had unknown functions ("orphan snoRNAs"). *Adapted from Esteller, 2011*[503].

Recently, lncRNAs have been recognized as transcriptional and translational key modulators of gene expression programs, as well as controllers of mRNA stability, being crucial in mammalian cell differentiation and development[500,513,514]. lncRNAs comprise different subclasses such as long intergenic noncoding RNAs (lincRNAs), natural antisense transcripts (NATs), transcribed ultraconserved regions (T-UCRs), telomeric repeat containing RNA (TERRA) and enhancer RNAs (eRNAs). lncRNAs are generally located in the chromatin and nucleus being their expression positively correlated with the expression of antisense protein-coding genes. Less expressed than coding genes, lncRNAs show a tissue-specific expression profile and there is a fraction known to be processed into small RNAs[511].

CHAPTER II

Aims and Thesis Outlines

1. Aims

In the last decades of cancer research scientists contributed to a huge amount of discoveries, bringing new players to an unknown landscape that is being more and more explored.

Since the discovery of DNA, the scientific research in cancer was based mainly on *Genetics* and in the *central dogma* through chromosomal abnormalities and the study of protein-coding oncogenes and tumor suppressors, throughout gain-of-function and inactivating mutations, respectively. Later, the attempts to look for possible explanations that support the existence of different phenotypes for the same genotype led to the discovery of two new layers of gene regulation. Here, the actual scenario of cancer biology comprising both *protein-coding genes* and *Genetics* was challenged by their cooperativeness with *ncRNAs* and *Epigenetics*. Based on the fact that members of other classes of ncRNAs are epigenetically silenced in cancer, we hypothesized that members of snoRNA and piRNA classes could be also repressed directly or indirectly by hypermethylation (**Chapter III** and **IV**).

While genetic changes are irreversible, the reversibility of epigenetic events opens a therapeutic opportunity to target key regulators of the epigenetic landscape and, at the same time, a challenge due to their rather unspecific genome-wide effects. Moreover, the epigenetic background is tissue-dependent, meaning that the same epigenetic regulator can behave as a tumor-suppressor or as an oncogene in a cell context dependent manner. For instance, *DNMT3A* that is recurrently mutated in leukemias[471], is overexpressed in several solid tumors[474]. Traditionally, epigenetic studies are based on the CIMP pathogenic pathway, looking specifically at DNA hypermethylation-dependent silencing of tumor suppressor genes. In order to fill this gap, we hypothesized that leukemia cells harboring a mutation in *DNMT3A* that decreases its enzymatic activity[472], should present a distinct methylome characterized by the specific hypomethylation at several loci that encode both coding and non-coding RNAs, which become re-expressed (**Chapter V**).

Classically, tumor progression studies focus their attention in epigenetic changes at proximal promoter regions, while epigenetic alterations at other regulatory loci are ignored. Nevertheless, the expression of genes that define cell identity is associated to

the existence of super-enhancers, cis-acting gene regulatory sequences[515]. In cancer cells, they are able to transcriptionally control oncogenes and other transforming genes[516-519]. We hypothesized that both protein-coding and ncRNAs could be epigenetically deregulated by aberrant profiles of DNA methylation at distal regulatory sequences such as super-enhancers (**Chapter VI**).

2. Thesis Outlines

In the first project entitled "**CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer**" (**Chapter III**) we aim to identify snoRNAs whose associated 5′-CpG islands become hypermethylated in cancer. Typically, they guide the modification of specific sites in ribosomal and spliceosomal RNAs. Nevertheless, there is a subset of snoRNAs with unknown functions ("*orphan snoRNAs*"). In this study we will focus our attention in several snoRNAs, giving more attention to snoRNAs with unknown targets and interrogating their possible role in cancer.

In a second project entitled "**Epigenetic loss of the PIWI/piRNA machinery in human testicular tumorigenesis**" (**Chapter IV**), we aim to investigate whether the disruption of PIWI proteins through their promoter region DNA hypermethylation affects piRNAs expression. Moreover we hypothesize that their silencing in testicular germ cell tumor cell lines and primary SE and NSE tumors would be accompanied by piRNA downregulation and concomitant LINE-1 loss of DNA methylation.

The third project entitled "**DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia**" (**Chapter V**) aims to identify specific transcripts undergoing re-activation in AML cell lines and patients harboring *DNMT3A* mutations, being our first goal, the identification of ncRNAs undergoing hypomethylation-mediated re-expression.

In our fourth project entitled "**Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer**" (**Chapter VI**) we aim to establish a relationship between the DNA methylation beyond those of proximal promoter gene regions and associated transcriptional regulation of downstream coding and non-coding genes.

CHAPTER III

CpG island hypermethylation-associated silencing of small nucleolar RNAs in human cancer

RNA Biol. 2012 Jun;9(6):881-90

Humberto J. Ferreira

Holger Heyn

Cátia Moutinho

Manel Esteller

1. Abstract

Much effort in cancer research has focused on the tiny part of our genome that codes for mRNA. However, it has recently been recognized that microRNAs also contribute decisively to tumorigenesis. Studies have also shown that epigenetic silencing by CpG island hypermethylation of microRNAs with tumor suppressor activities is a common feature of human cancer. The importance of other classes of non-coding RNAs, such as long intergenic ncRNAs (lincRNAs) and transcribed ultraconserved regions (T-UCRs) as altered elements in neoplasia, is also gaining recognition. Thus, we wondered whether there were other ncRNAs undergoing CpG island hypermethylation-associated inactivation in cancer cells. We focused on the small nucleolar RNAs (snoRNAs), a subset of ncRNA with a wide variety of cellular functions, such as chemical modification of RNA, pre-RNA processing and control of alternative splicing. By data mining snoRNA databases and the scientific literature, we selected 49 snoRNAs that had a CpG island within ≤ 2 Kb or that were processed from a host gene with a 5′-CpG island. Bisulfite genomic sequencing of multiple clones in normal colon mucosa and the colorectal cancer cell line HCT-116 showed that 46 snoRNAs were equally methylated in the two samples: completely unmethylated (n = 26) or fully methylated (n = 20). Most interestingly, the host gene-associated 5′-CpG islands of the snoRNAs *SNORD123*, *U70C* and *ACA59B* were hypermethylated in the cancer cells but not in the corresponding normal tissue. CpG island hypermethylation was associated with the transcriptional silencing of the respective snoRNAs. Results of a DNA methylation microarray platform in a comprehensive collection of normal tissues, cancer cell lines and primary malignancies demonstrated that the observed hypermethylation of snoRNAs was a common feature of various tumor types, particularly in leukemias. Overall, our findings indicate the existence of a new subclass of ncRNAs, snoRNAs, that are targeted by epigenetic inactivation in human cancer.

2. Introduction

Coding exons account for only 1.5% of the genome[520], despite being the focus of most of the current biomedical research. A large proportion of the genome is made up of non-protein coding regions that might have critical biological important, containing gene regulatory regions (transcriptional and splicing types), matrix attachment sites, origins of replication, other functional elements and non-coding RNAs (ncRNAs)[521,522]. The physiological and pathological importance of this functional part of the non-protein-coding genome is particularly apparent in a large class of small non-coding RNAs (sncRNAs) known as microRNAs[523]. These are about 22 nucleotides long and repress gene expression in a variety of eukaryotic organisms[523]. In human cancer, miRNA expression profiles differ between normal tissues and derived tumors and between tumor types[524]. miRNAs can act as oncogenes or tumor suppressors, with a key role in tumorigenesis[525-527]. Defects in miRNA function have been associated with a failure of miRNA post-transcriptional regulation[528], miRNA transcriptional repression by oncogenic factors[529], loss-of-function genetic alterations in the genes involved in miRNA-processing pathways[530-532] and transcriptional silencing associated with hypermethylation of CpG island promoters[533]. Thus, as occurs with miRNAs, it is likely that other types of ncRNAs are also involved in human tumorigenesis and are characterized by epigenetic and genetic defects in this disease[522]. In this context, Ultraconserved Regions (UCRs), a subset of conserved sequences that are located in intragenic and intergenic regions[534,535], are altered at the transcriptional level in human tumorigenesis[536]. Interestingly, transcribed UCRs (T-UCRs) undergo DNA methylation-associated silencing in cancer cells[537].

Another important class of ncRNAs that are potentially altered in human cancer are the small nucleolar RNAs (snoRNAs), which are localized in the nucleolus[538]. They are responsible for methylation[539,540] and pseudouridylation of rRNA (rRNA) at about 50–100 sites per eukaryotic ribosome. However there is increasing evidence of the targeting of other classes of RNAs, such as mRNAs[541]. snoRNAs are divided into two main classes: box C/D and box H/ACA[542], on the basis of their conserved secondary structural characteristics and associated modification reactions[538,543]. The C/D box snoRNAs guide the position-specific 2′-O-methylation and are associated with a core of four proteins: fibrillarin (the methyltransferase), NOP56, NOP5/NOP58 and NHPX.

The H/ACA snoRNAs direct RNA pseudouridylation of rRNA and are associated with dyskerin (the pseudouridine synthase), GAR1, NHP2 and NOP10[541,543,544]. Mutations in the human *dyskerin* gene, *NHP215* and *NOP1016* gene are associated with the X-linked genetic disorder, dyskeratosis congenita (DC), where malfunction of rRNA and shortening of telomeres have been observed[545-547]. As mutations in the *dyskerin* gene have also been associated with cancer susceptibility[545-547], it was suggested that snoRNAs were involved in the onset and progression of cancer. One of the first studies to address this possibility reported a snoRNA to be highly expressed in normal brain, but significantly reduced in meningioma[548]. Other studies showed that *GAS5*, a snoRNA-host gene, controls cell survival by inducing or sensitizing cells to apoptosis[549,550]. A substantial decrease of *GAS5* in breast cancer samples compared with adjacent unaffected normal breast epithelial tissues also suggests that it has a role as a tumor suppressor gene[550].

The association between snoRNAs and cancer was underlined by other studies showing that a homozygous 2-bp (TT) deletion of the snoRNA *U50* is strongly associated with prostate cancer[551] and that *U50* undergoes frequent genomic heterozygous deletion and transcriptional downregulation in breast cancer[552]. *U50* overexpression reduces colony-forming ability in prostate and breast cancer cells[551,552]. Taken together, these studies suggest that non-coding snoRNA *U50* is important in the development and progression of breast and prostate cancers[551,552]. More recently, it was shown that a diversity of snoRNAs are differentially expressed in non-small-cell lung cancer with respect to the corresponding matched tissue[553], encouraging investigation into the possible role of snoRNAs in oncogenesis. Another study has linked at least one snoRNA to the post-transcriptional processing of a protein-coding gene[554]. There is also evidence that other snoRNAs might be involved in the regulation of gene expression by giving rise to other regulatory RNA species, such as miRNAs, linking snoRNAs to RNA silencing[555]. It would therefore be very interesting to identify the function and mechanisms of the orphan snoRNAs.

The downregulation of tumor suppressor protein-coding genes (e.g., *hMLH1*, *BRCA1*, *VHL* and $p16^{INK4a}$)[49] and ncRNAs with growth inhibitory functions, such as miRNAs[533] has been closely linked to the presence of CpG island promoter hypermethylation. Thus, we wondered whether the same mechanism could play a role in the loss of adequate

snoRNA expression in tumors. Usually, snoRNAs are genomically found in the introns of protein-coding or non-protein-coding host genes, with each intron carrying only one single snoRNA and their transcription being synchronized with that of the host gene. After splicing they are generally processed by debranching and exonucleolytic trimming of the 5′- and 3′-ends[541,556-558], and assembled with specific core proteins that are essential for their localization and correct enzymatic activity, and for preventing their degradation[544]. However, intergenic snoRNAs are independently transcribed by RNA polII as independent units[541], and some human intron-encoded snoRNAs may have their own independent promoters[559]. Herein, we present a double candidate and genomic approach to unmask snoRNA-associated CpG islands that undergo cancer-specific hypermethylation-associated transcriptional silencing, such as *SNORD123*, *U70C* and *ACA59B*. These findings support a model in which epigenetic disruption of emerging new classes of ncRNAs, such as snoRNAs, is a common feature of human tumorigenesis.

3. Materials and Methods

Cell lines, culture conditions and primary study samples

The human cancer cell lines examined in this study were obtained from the American Type Culture Collection (ATCC). HCT-116 and DKO cells were a generous gift from Dr. Bert Vogelstein (Johns Hopkins Kimmel Comprehensive Cancer Center). Cells were maintained in appropriate media and treated with 1 μM 5-aza-2´-deoxycytidine (Sigma) for 48h to achieve demethylation[537]. DNA samples from normal tissues and primary leukemias were obtained at the time of the clinically indicated surgical procedures. All patients gave written consent to participate in the study and the Ethics Committee of each hospital approved the study protocol.

DNA methylation analyses

The CpG Island Searcher Program[560] was used to determine which snoRNAs were located within 2 Kb of a CpG island. DNA methylation status was established by PCR analysis of bisulfite-modified genomic DNA, which induces chemical conversion of unmethylated, but not methylated, cytosine to uracil. Two procedures were used. First, methylation status was established by bisulfite genomic sequencing of the

corresponding CpG islands. Eight independent clones were analyzed. The second analysis used methylation-specific PCR with primers specific for either the methylated or modified unmethylated DNA. The primers used are described in **Table 3.1**.

Quantification of snoRNAs with real-time PCR

Quantitative real-time PCR was performed to quantify the level of snoRNAs, as described previously[537]. Briefly, to quantify *SEMA5A* and the snoRNA-host genes 1 μg of purified and DNase-treated (TURBO DNA-free, Ambion) total RNA was reverse-transcribed using Thermoscript RT and random primer hexamers. cDNA was amplified by real-time PCR using SYBR (Applied Biosystems) green detection. Reverse transcription using a custom-designed TaqMan microRNA Reverse Transcription Kit (Applied Biosystems) was used to quantify the *SNORD123* and *U70C*, providing specificity for the mature RNA target. Reactions were performed in an Applied Biosystems 7900HT Fast Real-Time PCR system in 384-well plates. Expression values of *ASTN2, LOC100505806, SLC47A1* and *SEMA5A* were normalized to *HPRT1* and expression values of *U70C* and *SNORD123* to *RNU19*, respectively. Total RNA was extracted from three independent experiments and real-time PCR reactions were performed in triplicate. The primers used are described in **Table 3.1.**

Northern-blot

Fifteen micrograms of total RNA were loaded in a 15% denaturating polyacrylamide gel containing 8M Urea in 0.5XTBE buffer system. Decade Marker (Ambion) was prepared according to manufacturer's protocol, using [γ-^{32}P] ATP (PerkinElmer) and simultaneously loaded into the gel. Both RNA and marker were resolved by gel electrophoresis and transferred onto Hybond-N+ membrane (Amersham) in 0.5XTBE, followed by UV-cross linking (1200 Jules). Both *SNORD123* and *5.8S* rRNA probes were radiolabeled with 25 μCie using T4 kinase (Invitrogen) and purified with Nucaway Spin columns (Ambion). The membrane was prehybridized in hybridization buffer for 1h and hybridized overnight in the same solution at 45°C containing the *SNORD123* probe previously heated at 95°C for 2 min. The membrane was washed at low stringency followed by film exposure. The membrane was then hybridized with the *5.8S* rRNA probe using the same conditions followed by quantification using phosphorimager technology. All the probes used are described in **Table 3.1.**

Table 3.1 (Part 1). Oligonucleotides used in the study.

Bisulfite Sequencing		
Gene	**Forward 5' → 3'**	**Reverse 5' → 3'**
ACA28	AAGTTTGGAAAAGTTTATAGGGT	CTCACTCTATCACCCAAACTAA
ACA2b	TGTAATTTTAATATTTTGGGAGG	ACCAAAAAAACRAAAAAACTTAA
ACA49	TTTGTTTTTTAGGGTGGTTTTT	AACACCAATAACTAACCCATCA
ACA54	TTTAGATTGGAGTGTAATGGTG	CTAAAAATTCAAATCTTCCCAC
ACA55	AAGTGGTTTTTTAAGGGTGGT	ACCCCAAAATTTTTCTTATTCA
ACA58	TTGGTTTAAAAAGTAGTTTTTTTG	ATTCTTTTTTTTAAAACRAAATCT
HBII-180A and HBII-180B	TAAGGYGAGGAATATTTTTATT	CTACAAACTCCACCTCCCAAAT
HBII-316	GTATTGGAAAAATGAATAGAATTAAT	AATTTTAAAAAAAACCTTATTTCRACTA
HBII-336	GAAGTTTTYGTTAGGGTTTTT	CCTCCAAACCTTTAAATTATCTAACTA
HBII-52-1	ATGATTATTTTTATAATGGTTTATGG	TTTCTTTTTTATCAATTTTTCTATTTA
HBII-82	GTGTTGTGATTTTGGTTTATTGA	ACAAAACAAAACACCATCTTAA
HBII-85-3	TATGTGTGTTGTTTTGTGGAAG	CTCATTTTATTCAACTTTTCCAAA
hTR	GTAGTGGGTGTTTTYGGAG	CTCAAATTAATTTCACCTTTAAAAAAA
mgU12-22/U4-8 and U91	GGAGTTAAGGGTTGAGGAATTT	CATCACCTAATTCCCAAAACTC
mgU2-25/61 and HBII-382	GTTYGGTGGTTTTTTTTAGAAT	AACCRTACCCAAATCAAAAT
SNORD116.1	AGGGTTTTGATTTAGTTGGTTT	ATCAAACAATATCACCCTAAAAAA
U104 and ACA62	GTTGGTTTGTTTTTTTAAAAGAA	CTCTAACRCTACAAATTAACCC
U105B	ATTAGTTATTAGGGAGGTTAATAYGGT	AAACCAACCTAAACAACACRA
U13	TTGAAGATTAAAAAGAAAAAAAATT	AAAAATAACACATCTCACACAA
U19-2	GGGTTTTTATTGTGTTGTTTAGATT	ATAAACATTTACCTTCTCCCCA
U37	TTTGGTTAAAGTGAAGAAATGATT	CAATAAATCCCCCTAAAACAC
U38A and U38B	GATTTTTTAAAGTGTTGGGGTT	CAACCCCTACAAAAAAATAACA
U60	TGTAGTTTTTATAYGAGGTGTTAGA	TCACRAACTCTACATAAAAAAAA
U63	AGTAGAGATAGGGTTTTAGTATGTTG	ATCTTCTAAATAAAAATTTTTTTCCT
U88	AAGGTGAGTAGAAGTTTAAATTTTT	AAACACACAACTTTCTTCTCTT
U3 Subset (U3-2, U3-2B, U3-4, U3, U3-3)	AGATTTYGTTTGGGGGTG	ACCATTAAAACTATACTTTCAAAAATC
ASTN2 (U70C)	GGGTAGTGTGGATATYGTGA	ACCACTAAAACTTACACRAAACC
C7orf40 (ACA9)	TGTAGTGAAGGTTTGGTTTTTA	CTCCAACCTTCTAATCACTAAAA
PPP2R5A (U98b)	TTTTTATTTAYGATTGGTTGTG	CTACCTAAAACCTCCTCTCAAA
RPL13A (U32A, U33, U34 and U35A)	AATAAAGTAAGTTYGGATTGAGTG	AAACCCATACCTACACCTCC
SEMA5A/LOC100505806 (SNORD123)	GGTGGTAAAGTTGGGTGTTTA	AAAACCCCCAAAAATCAAC
SLC47A1 (ACA59B)	TTTGTTTTTTAAGTGTAGGAGTTT	TACCACTAAAACCACCTTACTC
SNHG5 (U50 and U50B)	TTATTGGATATGGAAGTAGATAGGG	TCTACCAACCAACTACCCTTTC
SNHG8 (ACA24)	GYGGGGAGTGAAATAGTATAGGT	AACAAATCCCAAAAATACATAAAAA
TCP1 (ACA20 and ACA29)	TTTTTTTAGTGTTGYGGTTA	CCCCAACTCTAACTTAAAAAAA

Table 3.1 (Part 2). Oligonucleotides used in the study.

Northern Blot Probes		
Transcript	**Oligonucleotide**	
SNORD123	GAATCAGCGCCCCAGAATTCATCATTTTCACCAAGTGTTTT	
5.8S rRNA	GCCCCGGGAGGAACCCGGGGCCGCAAGTGCGTTCGAAGTGTCGATGAT	
qRT-PCR		
Gene	**Forward 5' → 3'**	**Reverse 5' → 3'**
ASTN2	TCGCCGCCGCAGCAAAGGTTT	TGGCCGGGTTTAACCAGTCGGA
HPRT1	TGACACTGGCAAAACAATGCA	GGTCCTTTTCACCAGCAAGCT
LOC100505806	CTCCCTCTCTGCCGGCCTCG	GGTCACTCATGGCCCACAGTCA
SLC47A1	TCACGCTGGCAATCGCGGTT	AGAGGAGCAGGACGAGCGCA
TaqMan Assays (Applied Biosystems)		
Transcript	**Assay ID**	
SNORD123	CS89JBX	
U70C	CSKAJX3	
SEMA5A	Hs01549381_m1	
RNU19	1003	
Methyl-Specific PCR		
Gene	**Forward 5' → 3'**	**Reverse 5' → 3'**
SEMA5A/LOC100505806	U:GTTTGAGGTTGTGAAGTTTATTT;	U: CACACCAATCTACTTACCAAAA;
(SNORD123)	M: TGAGGTTGCGAAGTTTATTC	M: GCCGATCTACTTACCGAAA

Key to symbols: R=A+G; Y=C+T; **U**=Unmethylated; **M**=Methylated

Infinium 450K DNA methylation array

The DNA methylation levels at 10 CpG sites encompassing the *SNORD123*-associated CpG island were determined using the Infinium 450K DNA methylation microarray, as previously described[561]. Briefly, DNA was quantified by Quant-iT™ PicoGreen dsDNA Reagent (Invitrogen) and the integrity was analyzed in a 1.3% agarose gel. Bisulfite conversion of 600 ng of each sample was done according to the manufacturer's recommendations for the Illumina Infinium Assay. Effective bisulfite conversion was checked for three controls that were converted simultaneously with the samples. Four µl of bisulfite-converted DNA were used to hybridize on an Infinium HumanMethylation 450 BeadChip, following the Illumina Infinium HD Methylation protocol. The chip was

analyzed using an Illumina HiScan SQ fluorescent scanner. The intensities of the images were measured using GenomeStudio (2010.3) Methylation module (1.8.5) software. The methylation score of each CpG is represented as a β value.

4. Results and Discussion

4.1 snoRNA CpG island DNA methylation analyses

To identify snoRNAs with putative DNA methylation-related inactivation in human tumors, we data-mined the scientific literature on snoRNAs published during 2000–2011, as made available by PubMed.gov, the human genome browser at UCSC[562] and the snoRNA-LBME-db database[541]. The CpG Island Searcher Program[560] was used to determine which snoRNAs were located within ≤2 Kb of a CpG island, since it has been estimated that more than 90% of the human promoters of another type of ncRNA, the microRNAs, are located 1 Kb upstream of the mature transcript[563,564]. The DNA methylation status of CpG islands within 2 Kb are also important for regulating the expression of a second type of ncRNA, the T-UCRs[537]. We also included snoRNAs that were processed from a host gene RNA containing a 5′-CpG island. **Figure 3.1** illustrates both categories of snoRNA-related CpG islands. Using the described conditions, we selected 49 snoRNAs that had a CpG island within a distance of ≤ 2Kb (15 intergenic independent snoRNAs and 24 within an intron of a host gene) or that were processed from the transcript of a host gene with a 5′-CpG island (n = 10). The characteristics of the 49 selected snoRNAs and the summarized results are shown in **Table 3.2**.

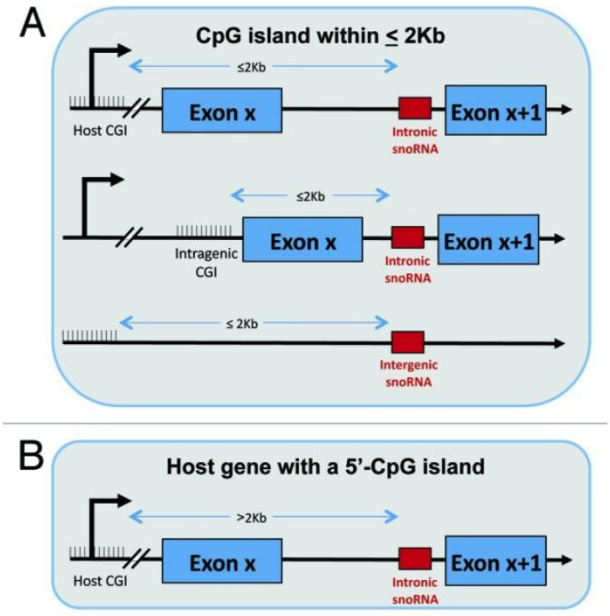


Figure 3.1. Types of CpG islands associated with snoRNAs in this study. A, Upstream CpG islands of a snoRNA located within 2 Kb. It includes intergenic independent snoRNAs and snoRNAs within an intron of a host gene. B, 5′-CpG islands of host genes for which RNA processing generates the expression of an intronic resident snoRNA.

Table 3.2. DNA methylation profile of CpG islands associated with snoRNAs. Green and red rectangles represent unmethylated and methylated CpG islands, respectively.

Name	Class	Chr.	Strand	Length (bp)	Target RNA	Host gene	Analyzed CpG Island	Normal Colon	HCT116
ACA24	HAcaBox	4	+	131	18S rRNA U863 and 18S rRNA U609	SNHG8	CpG island within ≤ 2Kb		
U50	CDBox	6	-	74	28S rRNA C2848 and 28S rRNA G2863	SNHG5	CpG island within ≤ 2Kb		
U50B	CDBox	6	-	71	Unknown	SNHG5	CpG island within ≤ 2Kb		
ACA9	HAcaBox	7	-	133	28S rRNA U1670 and 28S rRNA U1769	C7orf40	CpG island within ≤ 2Kb		
HBII-336	CDBox	7	+	74	18S rRNA A576	Independent	CpG island within ≤ 2Kb		
hTR	scaRna	3	-	548	Unknown	Independent	CpG island within ≤ 2Kb		
U91	scaRna	18	+	83	U4 snRNA C8	Independent	CpG island within ≤ 2Kb		
mgU12-22/U4-8	scaRna	18	+	421	U4 snRNA C8 and U12 snRNA G22	Independent	CpG island within ≤ 2Kb		
HBII-382	scaRna	1	+	82	U2 snRNA C61 and U2 snRNA G11	Independent	CpG island within ≤ 2Kb		
mgU2-25/61	scaRna	1	+	420	U2 snRNA G25 and U2 snRNA C61	Independent	CpG island within ≤ 2Kb		
U13	CDBox	8	+	104	Unknown	Independent	CpG island within ≤ 2Kb		
ACA62	HAcaBox	17	+	133	18S rRNA U34 and 18S rRNA U105	Independent	CpG island within ≤ 2Kb		
U104	CDBox	17	+	80	28S rRNA C1327	Independent	CpG island within ≤ 2Kb		
U60	CDBox	16	-	83	28S rRNA G4340	Independent	CpG island within ≤ 2Kb		
U3	CDBox	17	+	217	Unknown	Independent	CpG island within ≤ 2Kb		
U3-2	CDBox	17	+	217	Unknown	Independent	CpG island within ≤ 2Kb		
U3-2B	CDBox	17	-	217	Unknown	Independent	CpG island within ≤ 2Kb		
U3-3	CDBox	17	-	217	Unknown	Independent	CpG island within ≤ 2Kb		
U3-4	CDBox	17	-	217	Unknown	Independent	CpG island within ≤ 2Kb		
U98b	HAcaBox	1	+	133	Unknown	PPP2R5A	Host gene with a 5'-CpG island		
ACA20	HAcaBox	6	-	132	18S rRNA U651	TCP1	Host gene with a 5'-CpG island		
ACA29	HAcaBox	6	-	140	Unknown	TCP1	Host gene with a 5'-CpG island		
U32A	CDBox	19	+	77	18S rRNA G1328 and 28S rRNA A1511	RPL13A	Host gene with a 5'-CpG island		
U33	CDBox	19	+	83	18S rRNA U1326	RPL13A	Host gene with a 5'-CpG island		
U34	CDBox	19	+	66	28S rRNA U2824	RPL13A	Host gene with a 5'-CpG island		
U35A	CDBox	19	+	86	28S rRNA C4506	RPL13A	Host gene with a 5'-CpG island		
SNORD123	CDBox	5	+	70	Unknown	LOC100505806	Host gene with a 5'-CpG island		
U70C	HAcaBox	9	-	136	18S rRNA U1692	ASTN2	Host gene with a 5'-CpG island		
ACA59B	HAcaBox	17	+	152	Unknown	SLC47A1	Host gene with a 5'-CpG island		
U38A	CDBox	1	+	71	28S rRNA A1858	RPS8	CpG island within ≤ 2Kb		
U38B	CDBox	1	+	69	28S rRNA A1858	RPS8	CpG island within ≤ 2Kb		
HBII-52-1	CDBox	15	+	82	serotonin receptor 5HT-2C mRNA?	SNURF-SNRNP-UBE3A antisense	CpG island within ≤ 2Kb		
HBII-85-3	CDBox	15	+	97	Unknown	SNURF-SNRNP-UBE3A antisense	CpG island within ≤ 2Kb		
SNORD116.1	CDBox	1	-	92	Unknown	USH2A	CpG island within ≤ 2Kb		
ACA55	HAcaBox	1	-	137	18S U36	PABPC4	CpG island within ≤ 2Kb		
U88	scaRna	2	+	266	U5 snRNA U41	ATG16L1	CpG island within ≤ 2Kb		
HBII-316	CDBox	2	+	89	28S rRNA A3846	WDR43	CpG island within ≤ 2Kb		
ACA58	HAcaBox	3	-	137	28S rRNA U3823	MRPL3	CpG island within ≤ 2Kb		
U19-2	HAcaBox	5	+	204	28S rRNA U3741 and 28S rRNA U3743	ATP6V0E	CpG island within ≤ 2Kb		
U63	CDBox	5	-	68	28S rRNA A4541	HSPA9	CpG island within ≤ 2Kb		
ACA54	HAcaBox	11	-	123	28S rRNA U3801 and 28S rRNA U4539	NAP1L4	CpG island within ≤ 2Kb		
ACA49	HAcaBox	12	+	136	Unknown	EP400	CpG island within ≤ 2Kb		
ACA2b	HAcaBox	12	-	137	28S rRNA U4263 and 28S rRNA U4282	C12orf41	CpG island within ≤ 2Kb		
ACA28	HAcaBox	14	+	127	18S rRNA U815 and 18S rRNA U866	EIF5	CpG island within ≤ 2Kb		
HBII-82	CDBox	16	+	94	28S rRNA G3923	SF3B3	CpG island within ≤ 2Kb		
U105B	CDBox	19	+	79	18S rRNA U799	PPAN	CpG island within ≤ 2Kb		
U37	CDBox	19	-	66	28S rRNA A3697	EEF2	CpG island within ≤ 2Kb		
HBII-180A	CDBox	19	-	97	28S rRNA C3680	C19orf48	CpG island within ≤ 2Kb		
HBII-180B	CDBox	19	-	97	28S rRNA C3680	C19orf48	CpG island within ≤ 2Kb		

We performed bisulfite genomic sequencing of multiple clones using primers encompassing the 49 snoRNA-associated CpG islands to determine the DNA methylation patterns in normal colon mucosa and the colorectal cancer cell line HCT-116. We observed a completely unmethylated status for 23 snoRNA-related CpG islands in normal tissue and colon cancer cells (**Table 3.2**). Examples of the bisulfite sequencing analyses are shown in **Figure 3.2**. We also found a dense DNA methylation profile for 20 snoRNA-associated CpG islands in normal colon mucosa and HCT-116 colorectal cancer cells (**Table 3.2**). Examples of the bisulfite sequencing analyses are shown in **Figure 3.3**. Most importantly, we found a cancer-specific hypermethylation event for the snoRNAs *SNORD123*, *U70C* and *ACA59B*. Their associated CpG islands were completely unmethylated in normal colon mucosa and heavily hypermethylated in HCT-116 colorectal cancer cells (**Table 3.2** and **Fig. 3.4**). For all three cases, the CpG islands studied were in the 5′-region of the host gene where the snoRNA is located: the long non-coding gene *LOC100505806* (*SNORD123*), a*strotactin 2* (*U70C*) and s*olute carrier family 47 member 1* (*ACA59B*). The DNA methylation results were also confirmed using methylation-specific PCR (**Fig. 3.5**).

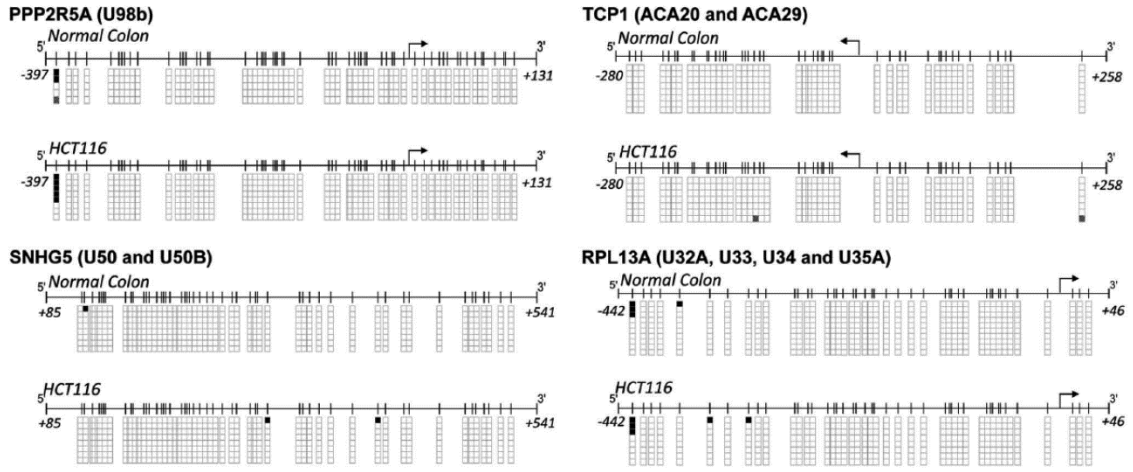


Figure 3.2. Bisulfite genomic sequencing of the CpG islands associated with the snoRNAs *U98b* (host gene *PPP2R5A*), *ACA20*/*ACA29* (host gene *TCP1*), *U50*/*U50B* (host gene *SNHG5*) and *U32A*/*U33*/*U34*/*U35A* (host gene *RPL13A*) in normal colon and the colorectal cancer cell line HCT-116. CpG dinucleotides are represented as short vertical lines. Eight single clones are represented for each sample. Presence of a methylated or unmethylated cytosine is indicated by a black or white square, respectively. Transcription start sites are represented by vertical black arrows.

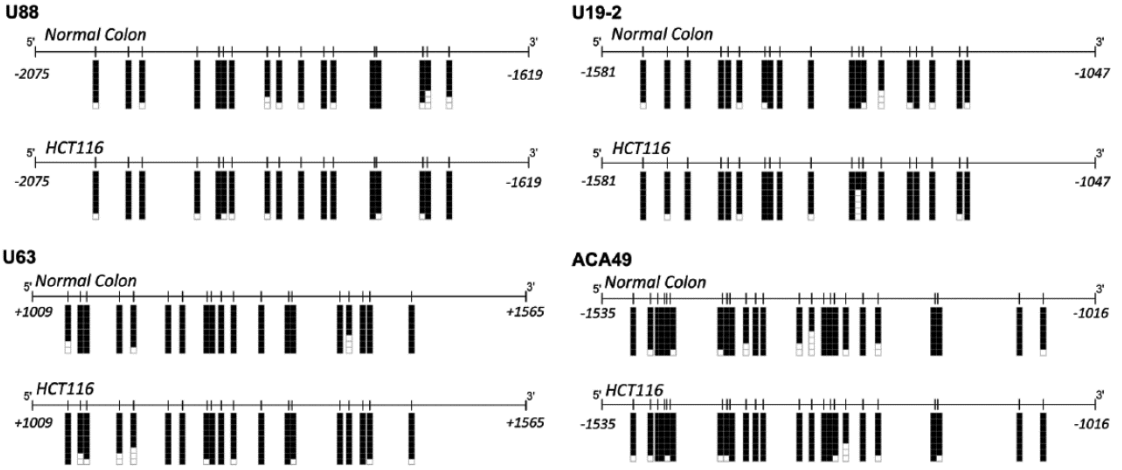


Figure 3.3. Bisulfite genomic sequencing of the CpG islands associated with the snoRNAs *U88*, *U19–2*, *U63* and *ACA49* in normal colon and the colorectal cancer cell line HCT-116. CpG dinucleotides are represented as short vertical lines. Eight single clones are represented for each sample. Presence of a methylated or unmethylated cytosine is indicated by a black or white square, respectively. Transcription start sites are represented by vertical black arrows.

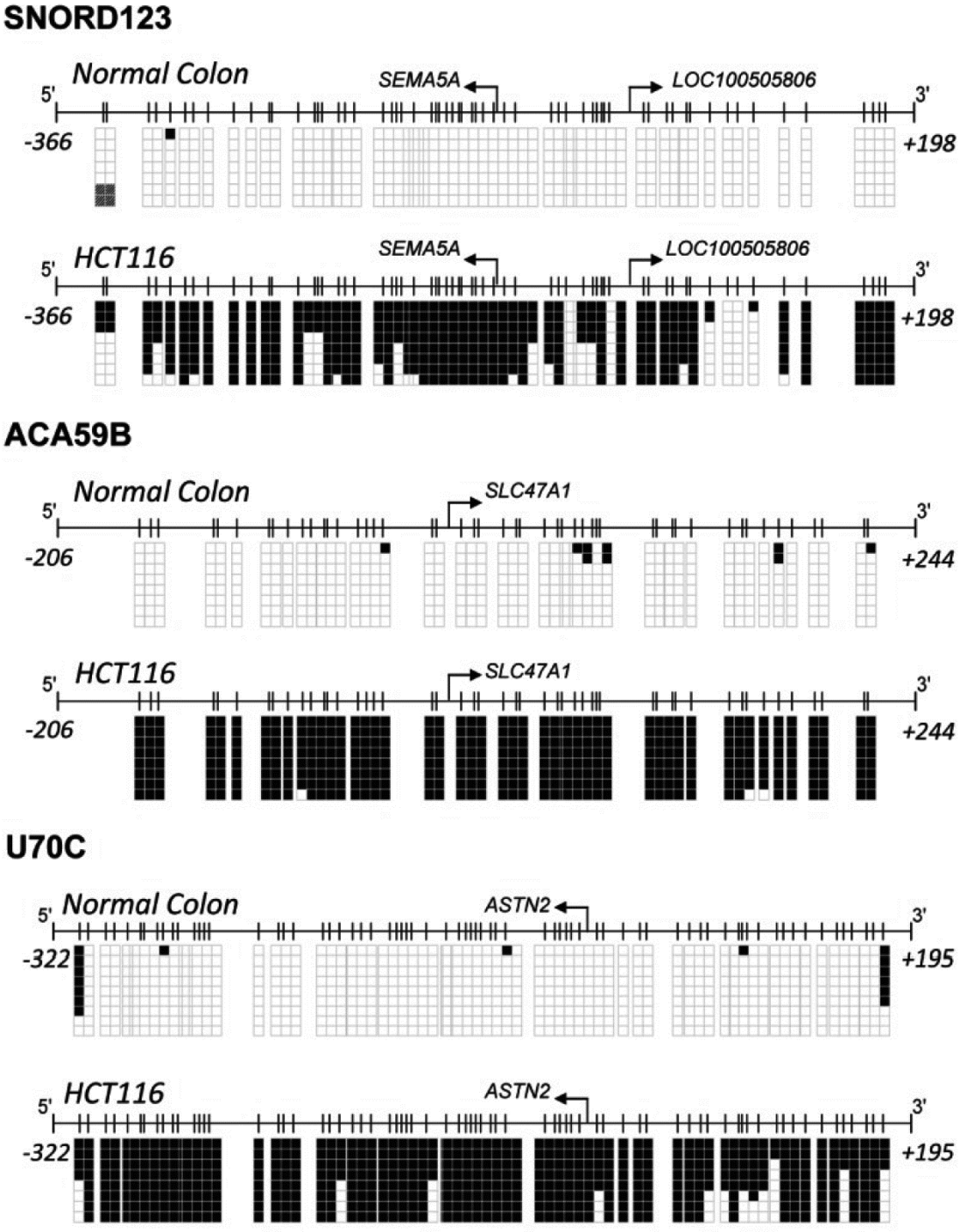


Figure 3.4. Bisulfite genomic sequencing of the CpG islands associated with the snoRNAs *SNORD123* (host gene is the lncRNA *LOC100505806*), ACA59B (host gene *SLC47A1*) and *U70C* (host gene *ASTN2*) in normal colon and the colorectal cancer cell line HCT-116. CpG dinucleotides are represented as short vertical lines. Eight single clones are represented for each sample. Presence of a methylated or unmethylated cytosine is indicated by a black or white square, respectively. Transcription start sites are represented by vertical black arrows.

4.2 Hypermethylation of snoRNA-related CpG islands is associated with transcriptional silencing

To demonstrate transcriptional silencing of these snoRNAs in cancer cells in association with the presence of CpG island hypermethylation, we measured transcript levels by quantitative RT-PCR. For the *SNORD123*, we analyzed the expression of the snoRNA itself, the long non-coding RNA *LOC100505806* from which the snoRNA is processed and the mRNA of the s*emaphorin 5A* (*SEMA5A*) that it is transcribed in the opposite direction from the same CpG island (**Fig. 3.4**). No expression of the *SNORD123*, *LOC100505806* and *SEMA5A* transcripts could be detected in HCT-116 cells in which the shared CpG island was methylated (**Fig. 3.6A**). Normal colon mucosa expressed the three transcripts (**Fig. 3.6A**). Methylation-specific PCR analyses of two additional colon cancer cell lines identified a hypermethylated and unmethylated CpG island in SW48 and DLD1 cells, respectively (**Fig. 3.5**). Loss of the *SNORD123*, *LOC100505806* and *SEMA5A* transcripts was observed in the hypermethylated SW48 cells and expression of the three transcripts was found in the unmethylated DLD1 cells (**Fig. 3.6A**). The absence of the *SNORD123* transcript in HCT-116 and SW48 cells and its presence in DLD1 cells was also validated by Northern-blot analyses (**Fig. 3.6B**). For the snoRNAs *U70C* and *ACA59B*, no expression of the snoRNA *U70C*, its host gene *ASTN2* and the host gene of *ACA59B* (*SLC47A1*) could be detected in HCT-116 cells in which the corresponding CpG island was methylated (**Fig. 3.7**). Normal colon mucosa expressed the three transcripts (**Fig. 3.7**). Most importantly, the expression for *ACA59B* (*SLC47A1*) was restored upon treatment with the DNA demethylating agent 5-aza-2´-deoxycytidine in the HCT-116 cell line (**Fig. 3.7**). These results were confirmed using an alternative model of an isogenic HCT-116 cell line in which the two major DNA methyltransferases, DNMT1 and DNMT3B, had been genetically disrupted (HCT116 DKO)[565]. The CpG island for *ACA59B* (*SLC47A1*) was hypomethylated in DKO cells (**Fig. 3.7**), but hypermethylated in the HCT116 parental cell line. *ACA59B* (*SLC47A1*) expression was restored in DKO cells (**Fig. 3.7**), reinforcing the link between CpG island hypermethylation and snoRNA silencing.

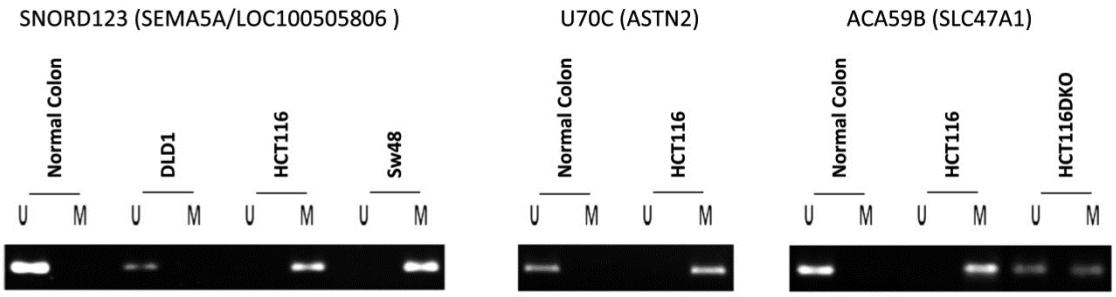


Figure 3.5. Methyl Specific PCR of normal colon and colorectal cancer cell lines. The presence of a band under the "U" lane indicates the presence of unmethylated alleles; the presence of a band under the "M" lane indicates the presence of methylated alleles; and M represent unmethylated and methylated specific amplification, respectively.

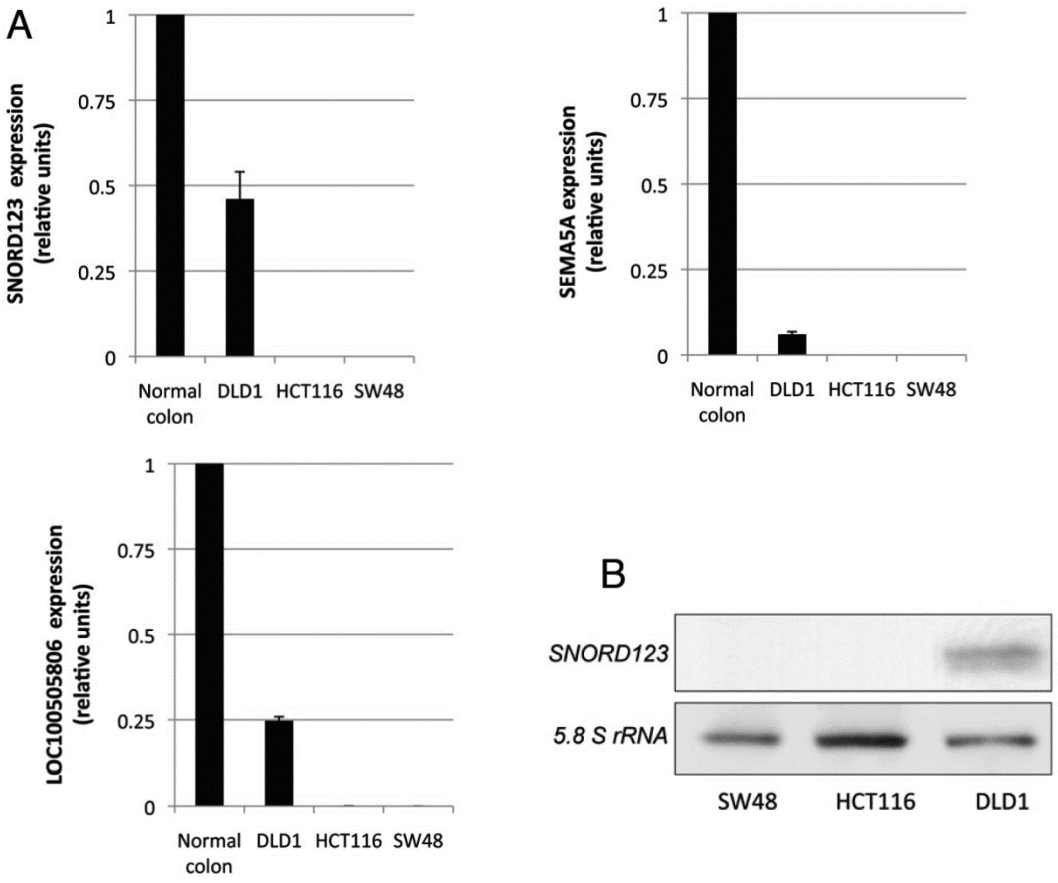


Figure 3.6. Expression analyses of of the transcripts derived from the SNORD123 / *LOC100505806* / *SEMA5A* CpG island. A, Quantitative RT-PCR of SNORD123, *LOC10050580* and *SEMA5A* showed loss of expression in the CpG island hypermethylated HCT-116 and SW48 cells. *SNORD123, ACA59B* and *U70C* are expressed in the unmethylated DLD1 cancer cells and in normal colon mucosa. B, northern-blot analysis shows the absence of the *SNORD123* transcript in the hypermethylated HCT-116 and SW48 cells and its presence in the unmethylated DLD1 cells.

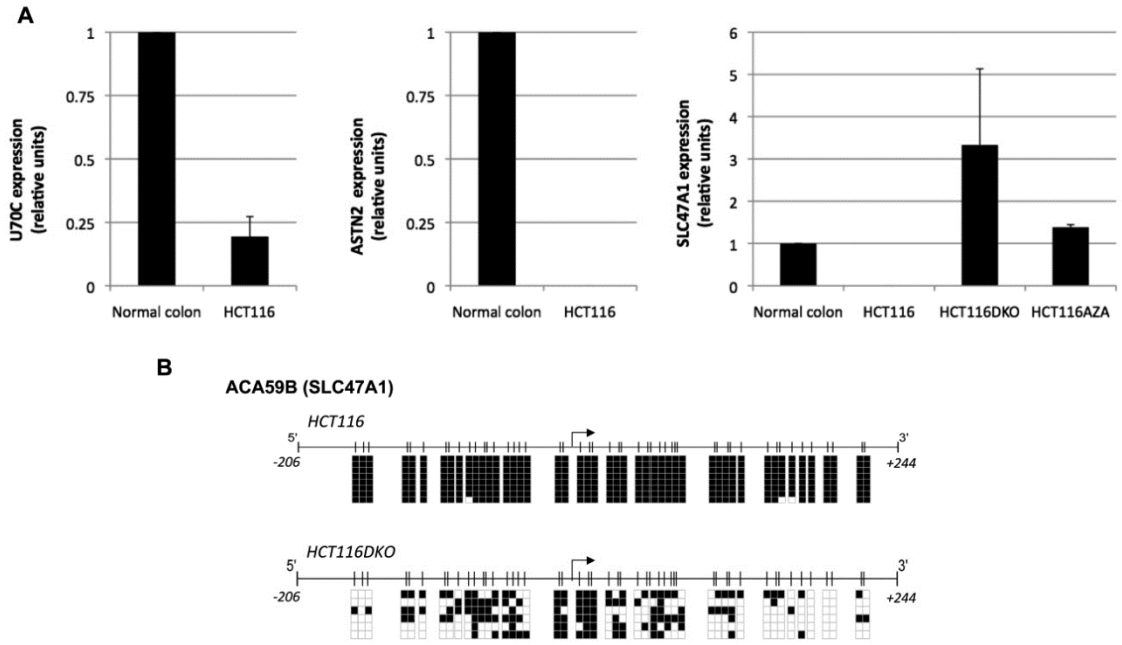


Figure 3.7. A, Quantitative RT-PCR for the snoRNA *U70C*, its host gene *ASTN2* and the host gene of *ACA59B* (*SLC47A1*) show loss of expression for the three transcripts in HCT-116 cells in which the corresponding CpG islands were methylated. Normal colon mucosa expressed the three transcripts. The expression for *ACA59B* (*SLC47A1*) was restored upon treatment with the DNA demethylating agent in the HCT-116 cell line. These results were confirmed using an alternative model of an isogenic HCT-116 cell line in which the two major DNA methyltransferases, DNMT1 and DNMT3b, had been genetically disrupted (HCT116 DKO). *ACA59B* (*SLC47A1*) expression was restored in DKO cells. B, The CpG island for *ACA59B* (*SLC47A1*) was hypomethylated in DKO cells, but hypermethylated in the HCT116 parental cell line.

4.3 Profile of snoRNA hypermethylation in different tumor types

To address the cancer-specific hypermethylation event in snoRNAs, we ruled out the possible presence of *SNORD123*, *U70C* and *ACA59B* CpG island methylation in a panel of normal tissues from the colon (n = 5), breast (n = 7) and lung (n = 22), in addition to four blood samples from healthy volunteers, using a DNA methylation microarray approach (**Fig. 3.8**)[561]. The observed high-throughput DNA methylation platform included five, two and four CpG sites corresponding to the bisulfite genomic sequenced CpG islands for *SNORD123*, *U70C* and *ACA59B*, respectively (**Fig. 3.8**). The presence of *SNORD123*, *U70C* and *ACA59B* cancer-specific CpG island hypermethylation and transcriptional silencing was not a unique feature of the colorectal cancer cell line HCT-116; analyzing the NCI60 panel of human cancer cell lines (n = 60) from nine tumor

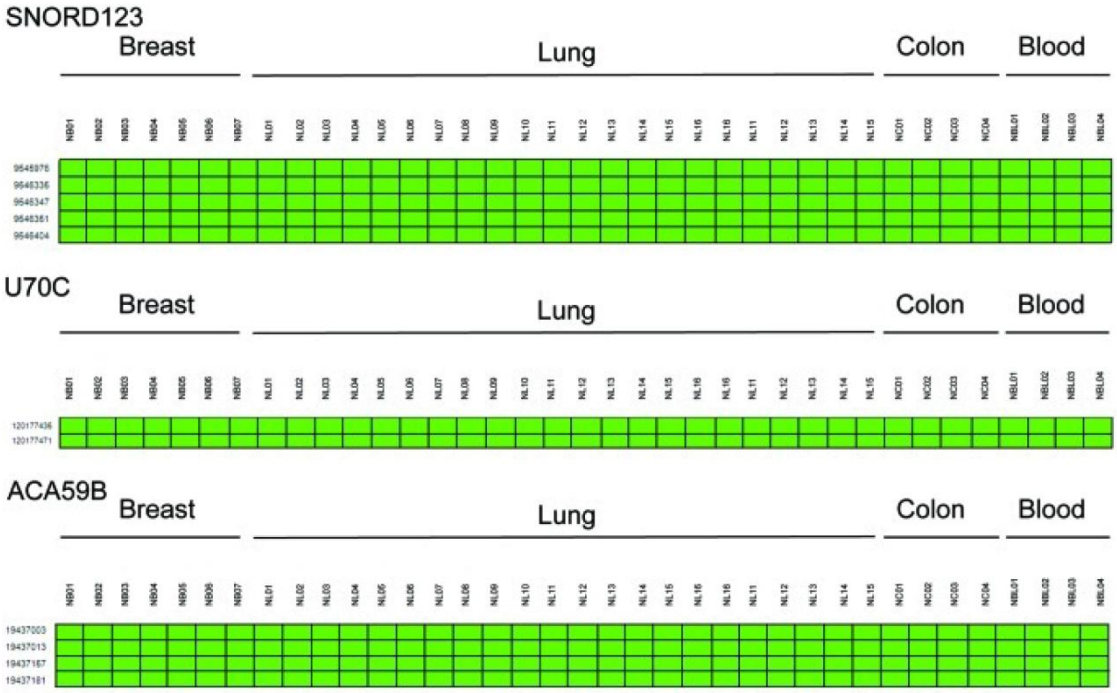


Figure 3.8. *SNORD123*, *U70C* and *ACA59B* CpG island methylation in normal tissues. DNA methylation microarray analyses of five, two and four CpG sites corresponding to the bisulfite genomic sequenced CpG islands for *SNORD123*, *U70C* and *ACA59B* in normal breast, lung and colon tissues and blood samples. Each square represents a single CpG: green square, unmethylated CpG; red square, methylated CpG. The three snoRNA-associated CpG islands were unmethylated in all the normal tissues tested.

types, we also found them in other colon cancer cell lines and lung, breast, prostate, ovarian, renal, melanoma, lymphoma and leukemia cells (**Fig. 3.9**).

Most importantly, the CpG island hypermethylation of *SNORD123*, *U70C* and *ACA59B* was not an in vitro phenomenon. Having noted the high frequency of CpG island hypermethylation for the three snoRNAs in leukemia cell lines (**Fig. 3.9**), we examined the presence of these epigenetic events in 48 primary samples from acute lymphoblastoid leukemia (ALL) patients, 43 had a B-cell phenotype while five were T-ALLs (**Fig. 3.10**). Using the described DNA methylation platform, we observed *SNORD123*, *U70C* and *ACA59B* hypermethylation in 27% (13 of 48), 39% (19 of 48) and 29% (14 of 48) of acute lymphoblastoid leukemia cases, respectively (**Fig. 3.10**). We did not observe any association between CpG Island hypermethylation of the three studied snoRNAs and disease-free survival or overall survival in these patients (Log rank Mantel-Cox test $p > 0.05$). We also extended the analyses of snoRNA DNA

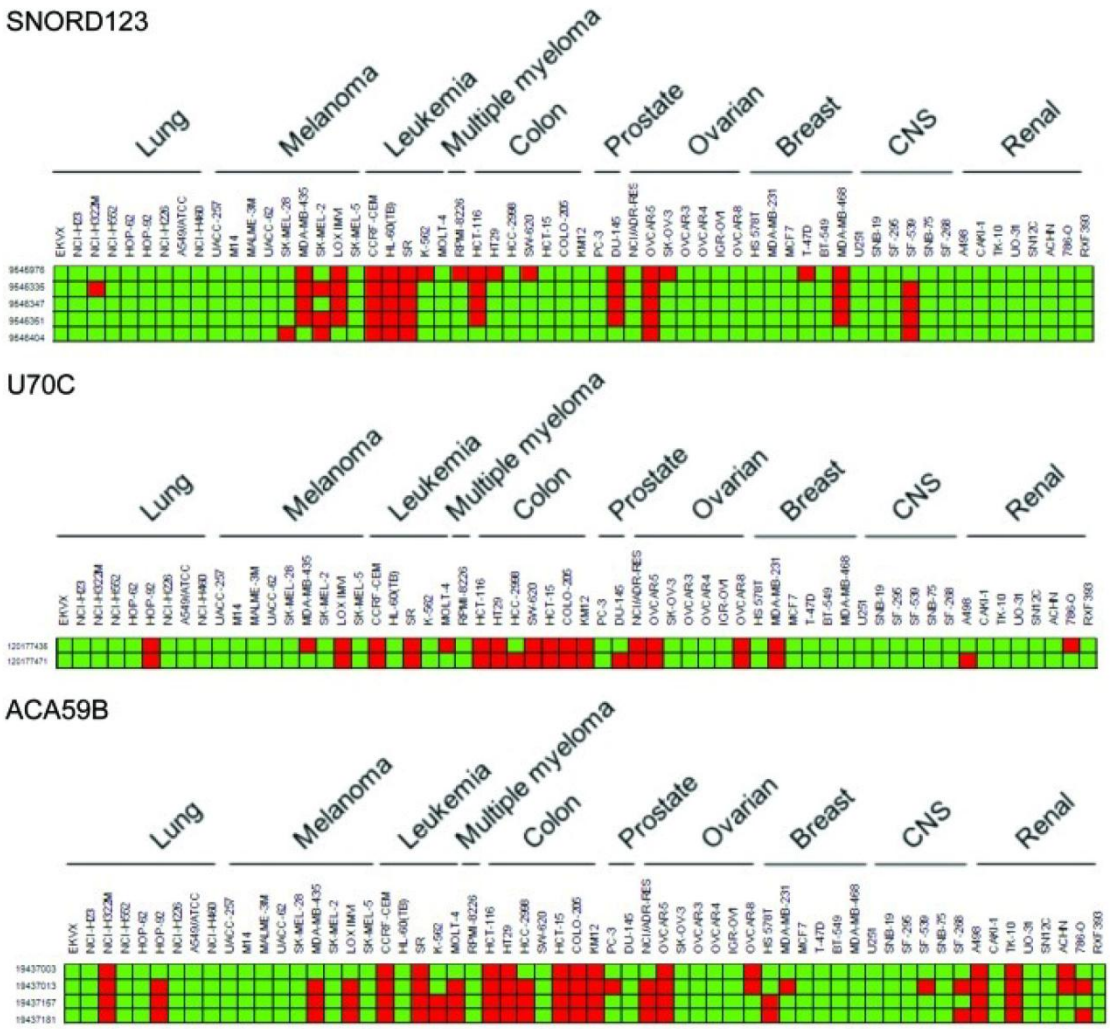


Figure 3.9. *SNORD123*, *U70C* and *ACA59B* CpG island methylation in human cancer cell lines. DNA methylation microarray analyses of five, two and four CpG sites corresponding to the bisulfite genomic sequenced CpG islands for *SNORD123*, *U70C* and *ACA59B* in the NCI60 panel of seven different tumor types. Each square represents a single CpG: green square, unmethylated CpG; red square, methylated CpG. The three snoRNA-associated CpG islands were methylated in the originally studied HCT-116 cells, but hypermethylation events were also observed in other classes of malignancies, such as leukemias.

methylation to primary acute myelogenous leukemia (AML) samples. Among 16 primary acute myelogenous leukemia cases, we observed *SNORD123* CpG island hypermethylation in 25% (4 of 16) of samples, while *ACA59B* and *U70C* were unmethylated in all cases. Finally, among 20 primary multiple myeloma cases, we observed *ACA59B* CpG island hypermethylation in 10% (2 of 20) of samples, while *SNORD123* and *U70C* were unmethylated in all cases.

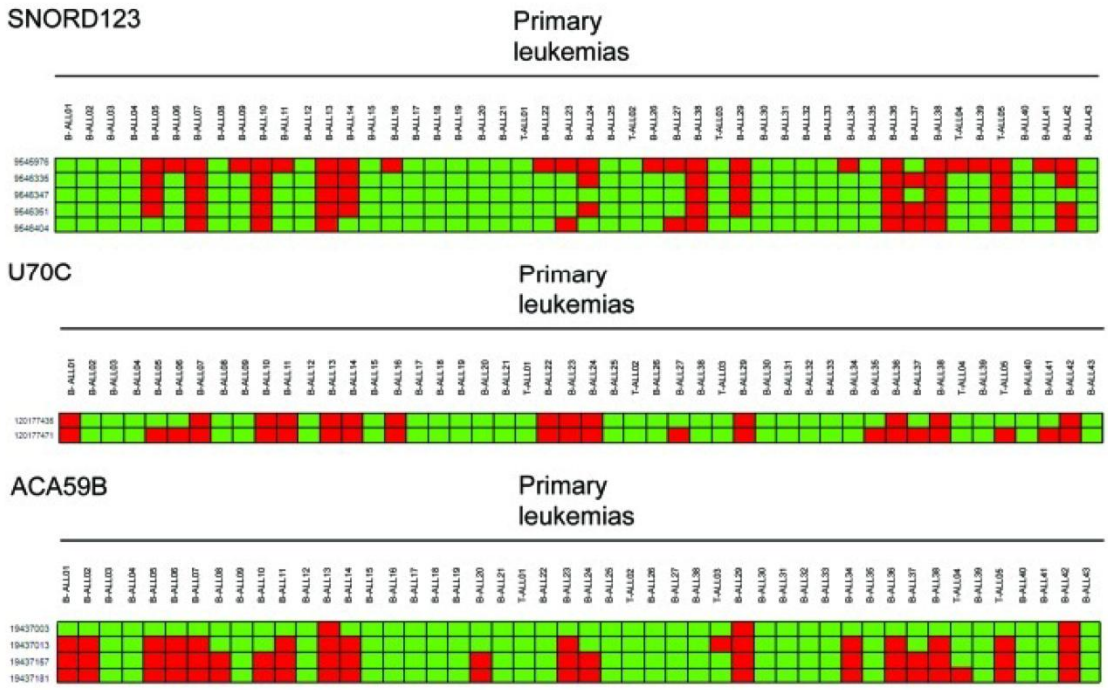


Figure 3.10. *SNORD123*, *U70C* and *ACA59B* CpG island methylation in primary acute lymphoblastoid leukemias. DNA methylation microarray analyses of five, two and four CpG sites corresponding to the bisulfite genomic sequenced CpG islands for *SNORD123*, *U70C* and *ACA59B* in leukemia patients demonstrated hypermethylation. Each square represents a single CpG: green square, unmethylated CpG; red square, methylated CpG. The B-cell or T-cell phenotype of the studied ALLs is indicated.

Overall, our results reveal the existence of cancer-specific hypermethylation events in CpG islands associated with snoRNAs that lead to their transcriptional inactivation in transformed cells. Despite our increasing knowledge about the biological roles of snoRNAs, one of the main challenges in cancer research into ncRNAs is the identification of a particular function that is relevant for cellular transformation. As coding genes[49,194], microRNAs[533] and T-UCRs[537] undergoing cancer-specific CpG island hypermethylation-associated silencing are known to have tumor suppressor roles, it is possible that snoRNAs act in a similar manner. This additional level of complexity is really true for the epigenetic silencing of two of the identified snoRNAs, *SNORD123*[566] and *ACA59B* (also known as *SNORA59B*)[567], because their target RNAs are unknown. *ACA59B* resides in an intron of the *solute carrier family 47 member 1* (*SLC47A1*) gene, while *SNORD123* is a C/D box snoRNA that resides within a long ncRNA transcript (*LOC100505806*) while in opposite direction is transcribed from the same CpG island the coding gene *SEMA5A*, adding another level of complexity in this

case. For *U70C* (also known as *SNORA70* or *ACA70*), the task might be a little easier. U70C was originally cloned from a cervical cancer cell line and belongs to the H/ACA box class of snoRNAs, having the predicted hairpin-hinge-hairpin-tail structure, conserved H/ACA-box motifs, and an association with the GAR1 protein[566,568,569]. The snoRNA *U70C* resides in an intron of the *astrotactin 2* (*ASTN2*) gene in the sense orientation and it serves as a guide for the pseudouridylation of selected bases of rRNA by forming short duplexes with the 18S rRNA U1692, the target for this snoRNA[566,568,569]. A role for 18S rRNA in tumorigenesis is starting to emerge[570-572], and our findings provide additional information about this role.

The enormous task of understanding the mechanisms by which snoRNA epigenetic silencing contributes to the origin and progression of human tumors still lies ahead. In the meantime, our observation that epigenetic inactivation by CpG island hypermethylation of a subset of snoRNAs, such as *SNORD123*, *U70C* and *ACA59B*, occurs across a wide spectrum of human cancer cell lines and primary tumors of diverse cellular and tissue origin provides clear support for the concept that major disruption of ncRNA programming is a common feature of cancer cells.

5. Acknowledgements

This work was supported by the European Research Council (ERC) Advanced Grant EPINORC, the Ministerio de Ciencia e Innovación (MICINN) grant SAF2011–22803, Fondo de Investigaciones Sanitarias Grant PI08–1345 and the Health Department of the Catalan Government (Generalitat de Catalunya). HF is a Portuguese Foundation for Science and Technology PhD Fellow (SFRH/BD/33887/2009. ME is an ICREA Research Professor.

CHAPTER IV

Epigenetic loss of the PIWI/piRNA machinery in human testicular tumorigenesis

Epigenetics. 2014;9(1):113-118.

Humberto J Ferreira

Holger Heyn

Xavier Garcia del Muro

August Vidal

Sara Larriba

Clara Muñoz

Alberto Villanueva

Manel Esteller

1. Abstract

Although most cancer research has focused in mRNA, non-coding RNAs are also an essential player in tumorigenesis. In addition to the well-recognized microRNAs, recent studies have also shown that epigenetic silencing by CpG island hypermethylation of other classes of non-coding RNAs, such as transcribed ultraconserved regions (T-UCRs) or small nucleolar RNAs (snoRNAs), also occur in human neoplasia. Herein we have studied the putative existence of epigenetic aberrations in the activity of PIWI proteins, an Argonaute family protein subclass, and the small regulatory PIWI-interacting RNAs (piRNAs) in testicular cancer, as the PIWI/piRNA pathway plays a critical role in male germline development. We have observed the existence of promoter CpG island hypermethylation-associated silencing of *PIWIL1*, *PIWIL2*, *PIWIL4*, and *TDRD1* in primary seminoma and non-seminoma testicular tumors, in addition to testicular germ cell tumor cell lines. Most importantly, these epigenetic lesions occur in a context of piRNA downregulation and loss of DNA methylation of the LINE-1 repetitive sequences, one of the target genomic loci where the PIWI/piRNA machinery acts as a caretaker in non-transformed cells.

2. Introduction

piRNAs (PIWI-interacting RNAs) are a peculiar class of small non-coding RNAs of 24–30 nt in length that are mainly expressed in the germline[503,573]. They are transcribed from regions in the genome that contain transcribed transposable elements and other repetitive elements[503,573]. piRNAs are Dicer-independent and bind the PIWI subfamily of Argonaute family proteins[503,574]. The PIWI/piRNA pathway seem to be involved in the maintenance of genomic stability and germ cell function by two different mechanisms that have been described in *Drosophila melanogaster*: cleavage of transposable element transcripts by PIWI proteins, a process that is mediated through piRNA base-pairing recognition, and heterochromatin-mediated transcriptional silencing associated with a gain of DNA methylation[503,573,574]. PIWI-class proteins, such as PIWIL2 and PIWIL4, are also involved in a so-denominated "ping-pong" amplification cycle, creating antisense piRNAs that are capable of repressing the transcript of origin[503,573,574]. In fact, knockout mice models for the proteins involved in the piRNA biogenesis (such as, MIWI, MILI, and MIWI2) revealed a recovery of transposon activity, which is thought to be the cause of the observed sterility due to meiotic arrest[575,576]. In this regard, DNA methylation regulated expression of piRNAs occurs in human spermatozoa[577] and non-genetic infertility syndromes in males could be associated with epigenetic disruption of PIWI proteins[578].

Interestingly, piRNAs have recently been detected also in human cancer cells and somatic cells, and it has been suggested that piRNAs control gene expression more broadly than previously[573]. Herein, we have wondered if in their natural functional context (the normal human testis), the PIWI/piRNA pathway undergoes aberrant DNA methylation events that compromise their activity in the corresponding transformed cells (human testicular germ cell tumors). To this aim, we have analyzed the 5′end promoter CpG islands of the main PIWI-class protein genes (*PIWIL1*, *PIWIL2*, and *PIWIL4*) and its associated protein TDRD1 (tudor domain containing protein 1). PIWIL1 is expressed after birth in the pachytene stage and acts in translational control in the latest stages of spermatogenesis[579]. PIWIL2 has essential roles in the initial phases of spermatogenesis: transposon silencing in fetal gonocytes, germline stem cell self-renewal and early prophase of mammalian testis[580]. Furthermore, PIWIL2 has been implicated in translational regulation of many genes during early spermatogenesis since

it binds piRNAs and mRNAs[580]. TDRD1 binds directly to PIWIL2 and PIWIL1[581]. Although it does not affect the ability of PIWI proteins to associate with piRNAs in embryonic testes, it controls the entry of correct transcripts into the normal pool of piRNAs[582]. Our data indicate that epigenetic disruption of the entire PIWI/piRNA pathway is indeed a hallmark for the development of testicular tumors.

3. Materials and Methods

Cell lines and patient samples

833K, SuSa and GCT27 testicular germ cell tumor lines[583] were cultured in RPMI 1640 medium supplemented with 20% (w/v) fetal bovine serum and penicillin/streptomycin (all from Invitrogen). All the cell lines were cultured at 37°C in a humidified atmosphere with 5% (v/v) CO_2. A representative subset of primary testicular germ cell tumors were obtained from the IDIBELL Bank of tumors, supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncología de Catalunya (XBTC). The study was approved by the relevant institutional review boards and ethics committees and informed consent was obtained from each patient. Seventeen primary pure seminomas (SE) and 19 non-seminomas (NSE) including either pure tumors (embryonal carcinomas, choriocarcinomas and yolk-sac tumors) as mix tumors composed by two or more components were included in the study. RNA and DNA from cell lines and primary tumor RNA and DNA from cell lines, testicular cancer samples and healthy controls were extracted using TRIzol (Invitrogen) and Phenol:Chloroform:Isoamylalcohol (Sigma), respectively. Bisulphite modification of genomic DNA was performed with the EZ DNA Methylation Kit (Zymo) following the manufacturer's protocol. Total RNA from adult normal testis (R1234260-50-BC) was purchased from BioCat.

Pyrosequencing

The set of primers for PCR amplification and sequencing were designed using a specific software pack (PyroMark assay design version 2.0.01.15). PIWIL1: FF 5′-GTTTGYGGGG TAGTTAATGA GGG, RV 5′-ACCTCCCAAA ACCTCCTT, SEQ 5′-CCCAAAACCT CCTTC; PIWIL2: FF 5′-ATGGGTTAAT TAGATAGTTT GTTTTTGTGA, RV 5′- AACCCACATA CTCCAAAACC AATTTC, SEQ 5′-

AATTAGATAG TTTGTTTTTG TGAA; PIWIL4: FF 5′-AAGTAATGGG AAGAAAAAGG AAGTT, RV 5′-CCAAAATAAC CACAAAAACT ACAAAATCTC, SEQ 5′-CAACATCCAC AACCA; TDRD1: FF 5′-AAGGAATTTT TTGAGTTTGT AATTAGAGTA, RV 5′-ATACAAACCC TCTCCCTCCC CTATA, SEQ 5′-TTGTAATTAG AGTATAAGTT GTTT. LINE-1 was quantified using the PyroMark Q96 LINE-1 assay (Qiagen) that target four CpG sites in positions 331 to 305 (GenBank accession no X58075) within the human LINE-1 transposon DNA consensus sequence. PCR was performed with primers biotinylated to convert the PCR product to single-stranded DNA templates. We used the Vacuum Prep Tool (Biotage) to prepare single-stranded PCR products according to manufacturer's instructions. Pyrosequencing reactions and methylation quantification were performed in a PyroMark Q24 System version 2.0.6 (Qiagen). Graphic representation of DNA methylation values shows the averaged values over multiple CpG sites.

Real-time qRT-PCR

Total RNA was reverse transcribed using ThermoScript RT-PCR System (Invitrogen, Life Technologies) under standard conditions with hexanucleotide random primers, according to the manufacturer's instructions. cDNA of protein coding genes was amplified by real-time PCR using Taqman assays: *PIWIL1* (Hs01041737), *PIWIL2* (Hs00216263), *PIWIL4* (Hs00381509) and *TDRD1* (Hs00229805) and expression values were normalized to *GAPDH* (AB Assay ID 4333764F). Reverse transcription and amplification of 3 randomly chosen piRNAs was performed using a custom-designed TaqMan assays (Applied Biosystems), providing specificity for the mature RNA target. Expression values of piRNAs *DQ598918* (CSQJAPI), *DQ589977* (CSMSF6U) and *DQ601609* (CSVI3FS) were normalized to *RNU19* (AB Assay ID 001003).

Infinium 450K Methylation array

All DNA samples were assessed for integrity, quantity and purity by electrophoresis in a 1.3% agarose gel, picogreen quantification, and nanodrop measurements. All samples were randomly distributed into 96 well plates. Bisulfite conversion of 500 ng of genomic DNA was performed using the EZ DNA methylation kit (Zymo Research) following the manufacturer's instructions. 200 ng of bisulfite-converted DNA were used

for hybridization on the HumanMethylation450 BeadChip (Illumina). Briefly, samples were whole-genome amplified followed by enzymatic end-point fragmentation, precipitation and resuspension. The resuspended samples were hybridized onto the BeadChip for 16 h at 48 °C, and then washed. A single nucleotide extension with labeled dideoxy-nucleotides was performed and repeated rounds of staining were applied with a combination of labeled antibodies differentiating between biotin and DNP. Chip analysis was performed using Illumina HiScan SQ fluorescent scanner. The intensities of the images were extracted and the data was normalized using GenomeStudio (2010.3) Methylation module (1.8.5) software.

Statistical analysis

Statistical analyses were performed using GraphPad PRISM (v5.04). The nonparametric Mann-Whitney U test was used to analyze differences in absolute expression and methylation level in cancer patient groups compared with controls. A value of $P < 0.05$ was considered significant.

4. Results and Discussion

4.1 Gain of 5′end promoter CpG island methylation for the *PIWIL1, PIWIL2, PIWIL4* and *TDRD1* genes occurs in primary testicular tumors in association with their transcriptional silencing

The *PIWIL1, PIWIL2, PIWIL4,* and *TDRD1* genes have 5′end CpG islands surrounding the corresponding transcription start sites and, thus, they are candidate genes to be epigenetically inactivated by promoter CpG island hypermethylation in human cancer[63,431]. Using bisulfite modification of DNA coupled with pyrosequencing, we observed that the four genes undergo hypermethylation events in primary testicular cancer in comparison to normal testicular tissues that show unmethylated CpG islands. The gain of 5′end CpG island methylation of *PIWIL1, PIWIL2, PIWIL4,* and *TDRD1* occurred both in seminoma and non-seminoma tumors (**Fig. 4.1**). To demonstrate the transcriptional silencing of these PIWI-class protein genes in cancer cells in association with the presence of CpG island hypermethylation, we measured *PIWIL1*, *PIWIL2*, *PIWIL4*, and *TDRD1* levels by quantitative RT-PCR. The expression of the four studied PIWI/piRNA pathway genes was significantly downregulated in testicular tumors in

comparison to normal testis. The diminished expression of *PIWIL1*, *PIWIL2*, *PIWIL4*, and *TDRD1* occurred both in seminoma and non-seminoma tumors (**Fig. 4.1**).

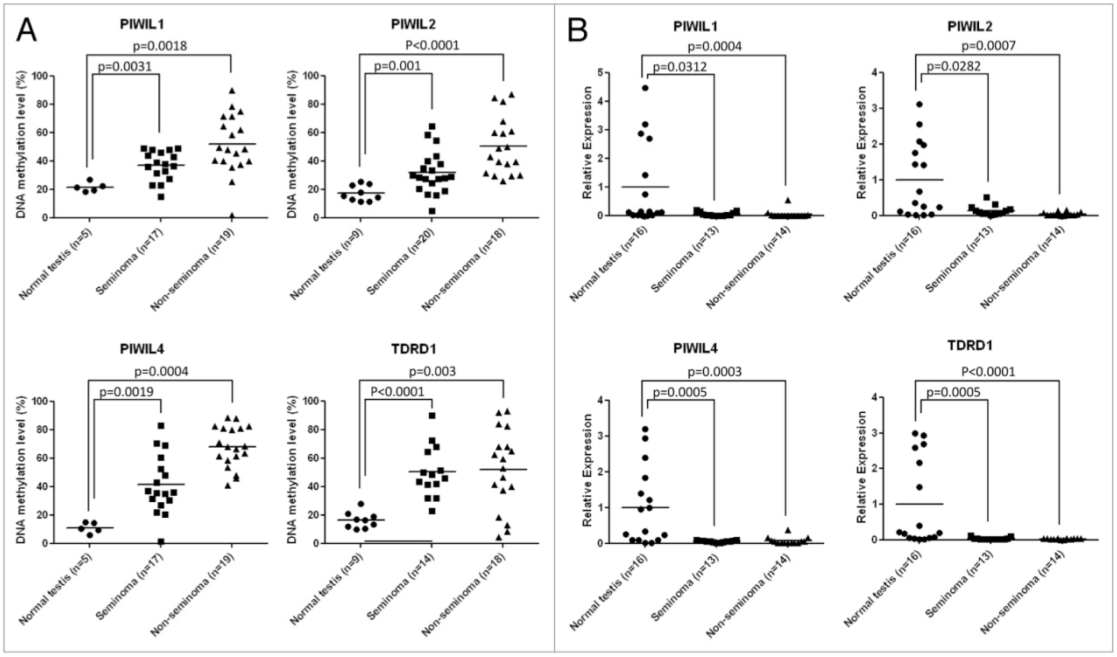


Figure 4.1. Epigenetic inactivation of genes encoding piRNA-related proteins in primary testicular germ cell tumors. (**A**) DNA methylation levels at the 5′end CpG islands of the *PIWIL1*, *PIWIL2*, *PIWIL4*, and *TDRD1* genes determined by sodium bisulfite modification coupled to pyrosequencing. (**B**) mRNA expression levels of the *PIWIL1*, *PIWIL2*, *PIWIL4*, and *TDRD1* genes determined by quantitative reverse transcription PCR.

We further tightened the link between CpG island hypermethylation of the studied PIWI-class protein genes and transcriptional inactivation by the analyses of testicular cancer cell lines. Using a DNA methylation microarray[561] that contains numerous CpG sites located in the *PIWIL1, PIWIL2, PIWIL4,* and *TDRD1* CpG islands (**Fig. 4.2**), we found that the human testicular germ cell tumor lines 833K, GCT27 and SuSa showed dense promoter CpG island hypermethylation for the described genes. Most importantly, we did not observe expression of the four studied PIWI/piRNA pathway genes in any of the three studied cancer cell lines, while normal testis expressed the *PIWIL1, PIWIL2, PIWIL4* and *TDRD1* transcripts (**Fig. 4.2**).

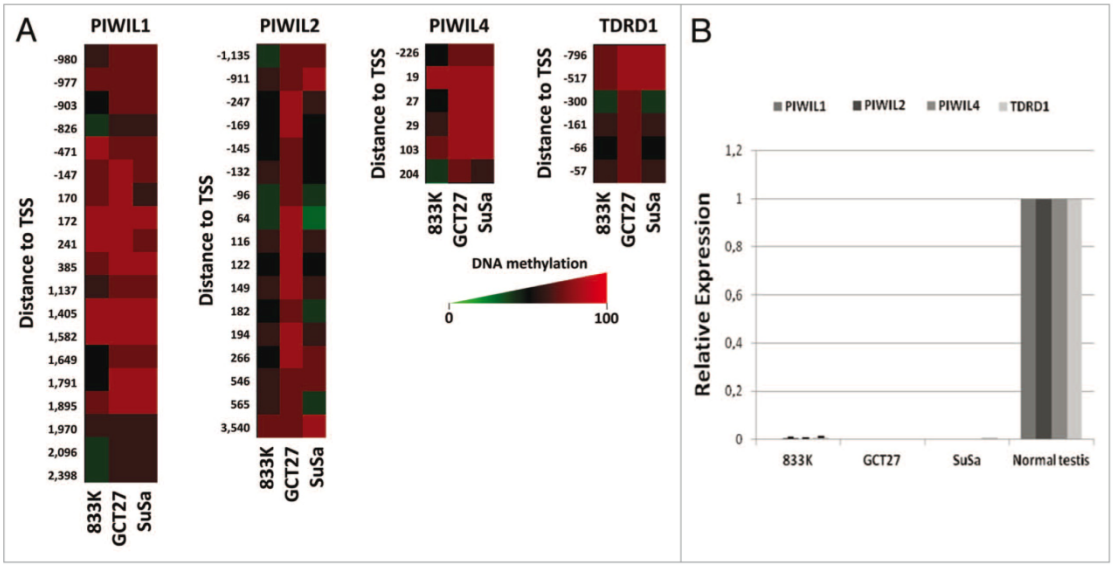


Figure 4.2. Epigenetic inactivation of genes encoding piRNA-related proteins in testicular germ cell tumor lines. (**A**) DNA methylation levels at the 5′end CpG islands of the *PIWIL1*, *PIWIL2*, *PIWIL4* and *TDRD1* genes determined by sodium bisulfite modification coupled to hybridization to a DNA microarray (450K Illumina). DNA methylation levels are color-coded (red: high, green: low). Probe distances to the transcription start site (TSS) are indicated. (**B**) mRNA expression levels of the *PIWIL1*, *PIWIL2*, *PIWIL4* and *TDRD1* genes determined by quantitative reverse transcription PCR.

4.2 Epigenetic inactivation of PIWI-class protein genes is associated with diminished piRNA expression and hypomethylation events at LINE-1 loci

Once we had determined the existence of promoter CpG island hypermethylation events in the described PIWI-class protein genes and the diminished expression of their corresponding transcripts in testicular tumorigenesis, both in primary tumors and cultured transformed cells, we wondered about the downstream impact of the described epigenetic inactivation. We first analyzed the expression levels of piRNAs in the same samples studied for *PIWIL1, PIWIL2, PIWIL4,* and *TDRD1* CpG island methylation and transcription. piRNAs show a high diversity in their genomic sequences and are transcribed from a relatively small number of genomic regions called piRNAs clusters[584]. After a primary RNA is generated, piRNA accumulation requires amplification by the above mentioned ping-pong mechanism involving at least two distinct Piwi proteins, a process that occurs in the cytoplasm[503,573]. Herein, we randomly selected three piRNAs transcribed from different genomic loci to sample the global levels of expression of these small regulatory ncRNAs: *DQ598918* (chr7:99 691 656–

99 691 686), *DQ589977* (chr17:79 479 330–79 479 358) and *DQ601609* (chr1:179 557 005–179 557 036). Using quantitative RT-PCR to study the expression levels of the three piRNAs, we found diminished expression of all of them in primary testicular tumors in comparison to normal testis (**Fig. 4.3**). The diminished expression of the piRNAs *DQ598918*, *DQ589977* and *DQ601609* occurred both in seminoma and non-seminoma tumors (**Fig. 4.3**).

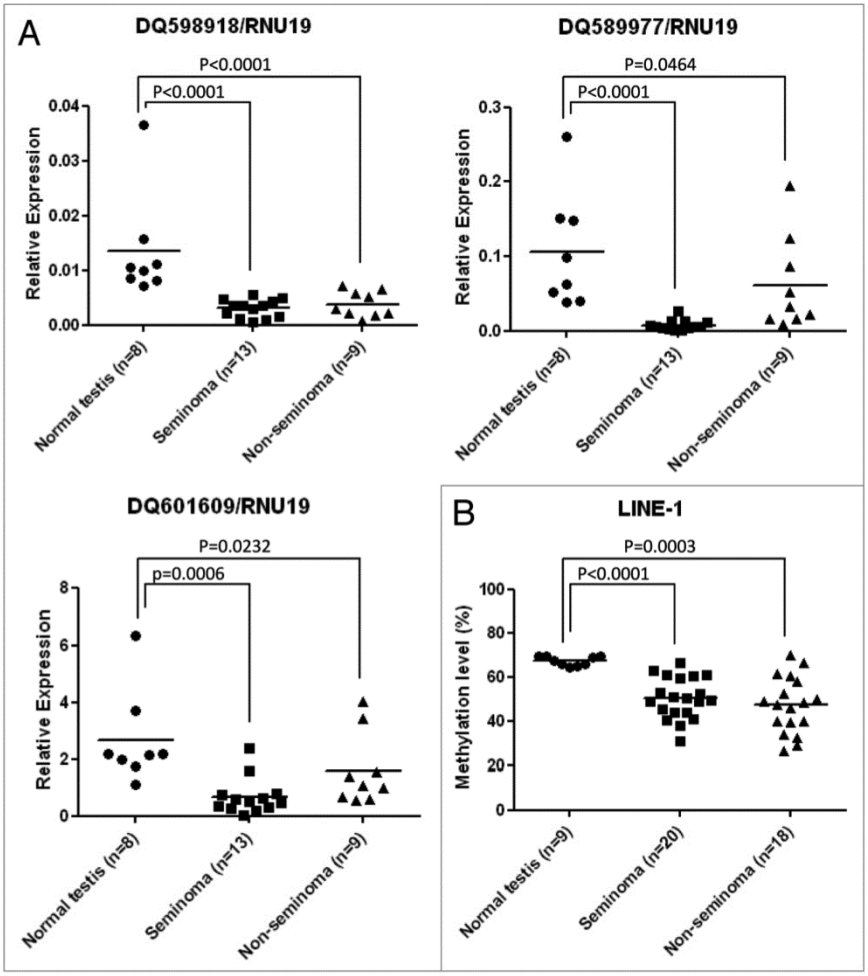


Figure 4.3. Molecular environment of the epigenetic loss of PIWI-protein genes in primary testicular tumors: diminished piRNA expression and LINE1 hypomethylation. (**A**) Expression levels of the piRNAs *DQ598918*, *DQ589977* and *DQ601609* determined by quantitative reverse transcription PCR. (**B**) DNA methylation levels at the LINE1 sequence determined by sodium bisulfite modification coupled to pyrosequencing.

Interestingly, the detected aberrant hypermethylation and diminished expression of PIWI-family genes, together with the downregulated piRNAs, might have the ability to provoke further DNA methylation changes of additional loci. In this regard, genetic and molecular analyses have identified interactions between DNA methyltransferases

(DNMTs) and piRNA pathway members. The PIWI/DNMT3L complex targets genomic loci, sequence-guided by small RNAs[585]. DNMT3L[585] as well as PIWIL2[586] and TDRD1[582] null models revealed a loss of DNA methylation at LINE-1 and intracisternal A-particle (IAP) transposons, leading to reactivation of the repetitive sequences. Thus, we proceeded to determine in our studied set of primary testicular tumors that harbor *PIWIL1, PIWIL2, PIWIL4,* and *TDRD1* promoter CpG island methylation-associated silencing and piRNA diminished levels, had also undergone DNA methylation changes at the LINE1 sequence. Using bisulfite modification of DNA coupled with pyrosequencing, we observed that the LINE1 sequences experimented hypomethylation events in testicular cancer in comparison to normal testis that show a more methylated LINE1 sequence (**Fig. 4.3**). The loss of LINE1 methylation occurred both in primary seminoma and non-seminoma tumors (**Fig. 4.3**).

Overall, our results show the existence of cancer specific hypermethylation events in the CpG islands of genes associated with piRNAs that leads to their transcriptional inactivation in testicular cancer. Most importantly, the epigenetic inactivation of PIWI-class protein genes (*PIWIL1, PIWIL2,* and *PIWIL4*) and its associated protein TDRD1 in human testicular tumorigenesis occurs in a molecular context characterized by a diminished expression of the piRNAs and the DNA hypomethylation of LINE1, a PIWI/piRNA target sequence. Interestingly, epigenetic disruption of PIWI proteins also occurs in non-genetic infertility syndromes in males[578] and there is an epidemiological association between male infertility and testicular cancer[587,588]. Thus, the epigenetic disruption of the PIWI/piRNA pathway could be the missing common link in the genesis of both pathologies. A model for the PIWI/piRNA pathway in normal testis and its disruption in testicular tumors is also shown in **Figure 4.4**. Although much work lies ahead to understand the role of the PIWI/piRNA machinery in cellular transformation, the current work provide intriguing clues for further developments and discoveries in this area.

5. Acknowledgments

The research leading to these results has received funding from the European Research Council under the European Community's Seventh Framework Programme (FP7/2007–2013)/ERC grant agreement no. 268626—EPINORC project, the Spanish Ministry of Economy and Competitiveness (MINECO Project no. SAF2011-22803), the Cellex

Foundation and the Health and Science Departments of the Catalan Government (Generalitat de Catalunya). HJF is a Portuguese Foundation for Science and Technology (FCT) Ph.D. Fellow. ME is an ICREA Research Professor.

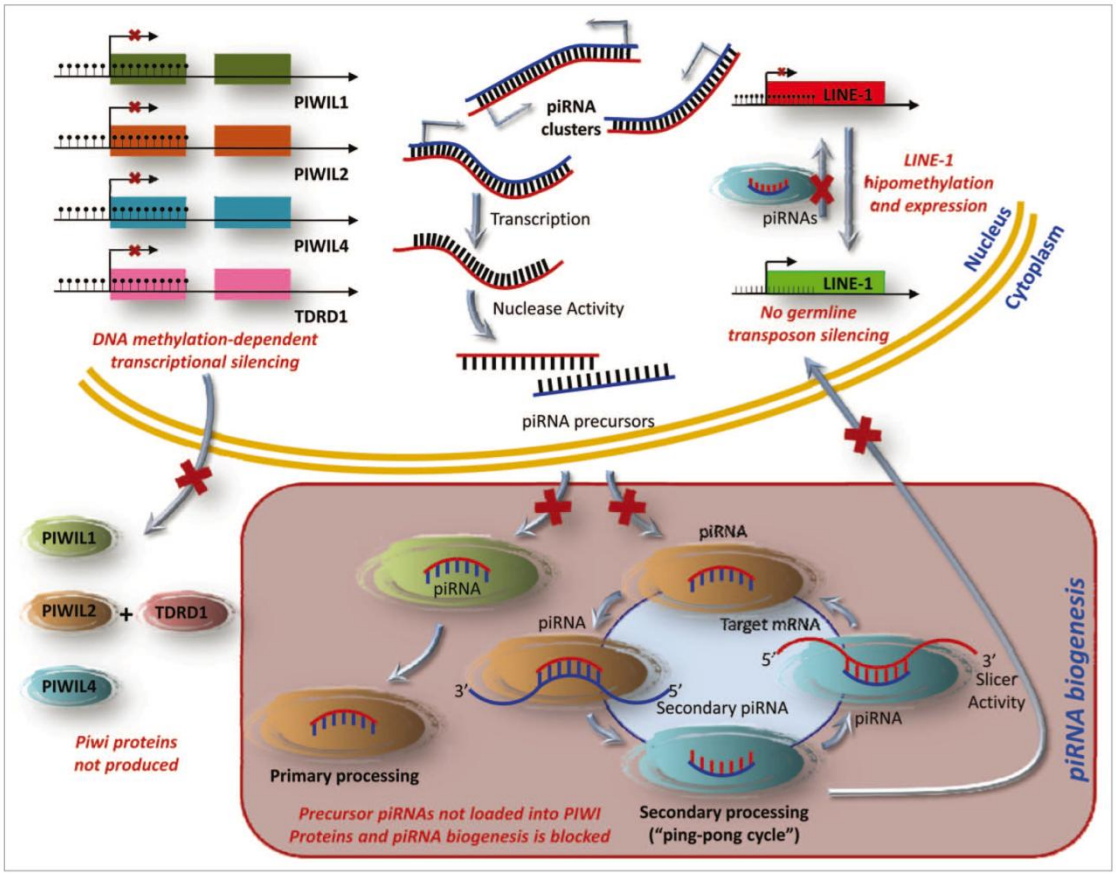


Figure 4.4. A model for the PIWI/piRNA pathway in germline cells and its disruption in testicular tumors. piRNAs are small non-coding RNAs that are mainly transcribed as single-stranded intergenic RNAs from well-conserved mono- and bi-directional clusters of repetitive elements. These piRNA precursors translocate into the cytoplasm, where they mature into functional piRNAs. The PIWI proteins catalyze a self-amplification loop, "ping-pong" cycle. Their incorporation into the PIWI ribonucleoprotein (piRNP) complex targets repetitive elements through target degradation and epigenetic silencing. In testicular cancer types, piRNA biogenesis and function are disrupted by DNA hypermethylation mediated transcriptional silencing of PIWI-proteins, leading to the expression of germline transposons.

CHAPTER V

DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia

Oncogene (2016) 35, 3079–3082.

Humberto J Ferreira

Holger Heyn

Miguel Vizoso

Cátia Moutinho

Enrique Vidal

Antonio Gomez

Anna Martinez-Cardús

Laia Simó-Riudalbas

Sebastian Moran

Edgar Jost

Manel Esteller

1. Abstract

Close to half of *de novo* acute myeloid leukemia (AML) cases do not exhibit any cytogenetic aberrations. In this regard, distortion of the DNA methylation setting and the presence of mutations in epigenetic modifier genes can also be molecular drivers of the disease. In recent years, somatic missense mutations of the *DNA methyltransferase 3A* (*DNMT3A*) have been reported in ~20% of AML patients; however, no obvious critical downstream gene has been identified that could explain the role of DNMT3A in the natural history of AML. Herein, using whole-genome bisulfite sequencing and DNA methylation microarrays, we have identified a key gene undergoing promoter hypomethylation-associated transcriptional reactivation in *DNMT3A* mutant patients, the leukemogenic HOX cofactor MEIS1. Our results indicate that, in the absence of mixed lineage leukemia fusions, an alternative pathway for engaging an oncogenic MEIS1-dependent transcriptional program can be mediated by *DNMT3A* mutations.

2. Introduction

Acute myeloid leukemia (AML) comprises a group of hematopoietic malignancies derived from myeloid precursors that have a highly heterogeneous clinical course and response to therapy. AML is characterized by greater proliferation and lower differentiation of the hematopoietic progenitor cells. Non-random cytogenetic aberrations are the single most important prognostic factor of the disease, but close to half of *de novo* AML cases do not exhibit any[589]. Many molecular drivers with potential prognostic significance have been described particularly for this last group, such as mutations in *nucleophosmin*, *fms-related tyrosine kinase 3* and *CCAAT/enhancer-binding protein-α*. In recent years, distortion of the DNA methylation setting and the presence of mutations in epigenetic modifier genes, such as *Tet methylcytosine dioxygenase 2* and *isocitrate dehydrogenase 1/2*, have been directly implicated in the pathogenesis of AML[169]. In this regard, somatic missense mutations of the *DNA methyltransferase 3A* (*DNMT3A*) have also been reported in ~20% of AML patients, in whom they are usually associated with an unfavorable prognosis[471,472,590,591].

DNMT3A is a *de novo* DNA methyltransferase that catalyzes the transfer of a methyl group onto the 5′-position of cytosine of CpG dinucleotides. Most of the *DNMT3A* mutations present in AMLs are heterozygous, with a great predominance of missense alterations in the R882 residue located in the catalytic domain[471,472,590,591]. R882H DNMT3A has recently been shown to act as a dominant negative that inhibits wild-type DNMT3A[473]. In this context, AML samples carrying *DNMT3A* mutations have been found to be associated with DNA methylation changes[592,593]. However, no clear and common epigenetic signature has so far emerged and, most importantly, no obvious critical downstream gene has been identified that could explain the role of DNMT3A in the natural history of AML.

3. Materials and Methods

Samples description

OCI-AML3 and OCI-AML5 cell lines were obtained from the Leibniz Institute DSMZ-German Collection of Microorganisms and Cell Cultures (Braunschweig, Germany) and cultured according to supplier's specifications at 37ºC in a humidified atmosphere with

5% (v/v) CO_2. Genomic DNA and RNA were extracted by the Phenol:Chloroform: Isoamylalcohol methodology (Sigma) and by automated purification with the Maxwell® 16 LEV simplyRNA Kit (Promega, Madison, WI, USA) following the manufacturer's protocol, respectively. DNA from primary AML samples was provided by the Department of Hematology of the RWTH Aachen University (Aachen, Germany) after written consent according to the "Biobank" rules of the medical faculty and approval by the local ethics committee (Permit Number: EK206/09).

Whole Genome Bisulfite Sequencing

We spiked genomic DNA (1 or 2 µg) with unmethylated λ DNA (5 ng of λ DNA per µg of genomic DNA) (Promega). We sheared DNA by sonication to 50–500 bp with a Covaris E220 and selected 150- to 300-bp fragments using AMPure XP beads (Agencourt Bioscience Corp.). We constructed genomic DNA libraries using the TruSeq Sample Preparation kit (Illumina Inc.) following Illumina's standard protocol. After adaptor ligation, we treated DNA with sodium bisulfite using the EpiTect Bisulfite kit (Qiagen) following the manufacturer's instructions for formalin-fixed and paraffin-embedded (FFPE) tissue samples. We performed two rounds of conversion to achieve >99% conversion. We enriched adaptor-ligated DNA through seven cycles of PCR using the PfuTurboCx Hotstart DNA polymerase (Stratagene). We monitored library quality using the Agilent 2100 BioAnalyzer (Agilent) and determined the concentration of viable sequencing fragments (molecules carrying adapters at both extremities) by quantitative PCR using the Library Quantification Kit from KAPA Biosystems. We performed paired-end DNA sequencing (two reads of 100 bp each) using the Illumina Hi-Seq 2000.

Sequencing quality was assessed using the Illumina Sequencing Analysis Viewer and FastQC software. We ensured the raw reads used in subsequent analyses were within the standard parameters set by the Illumina protocol. Positional quality along the reads was confirmed to be QC>30, and we excluded biases towards specific motifs or GC-enriched regions in the PCR amplification or hybridization. Sequence alignment and DNA methylation calling of WGBS reads were performed using Bismark V.0.7.4 software[594]. SAM/BAM and BED file handling was done using SAMtools, bedtools[595] and Tabix[596]. Statistical analysis and graphic representation was performed with R (http://www.R-project.org) and multicore and ggplot2 libraries. We smoothed the DNA

methylation profiles using a previously described method for processing WGBS data[597]. Briefly, the method assumes that the DNA methylation profile is defined by a varying function of the genomic location that can be estimated with a local likelihood smoother. We used HG19 as the reference genome and retrieved genomic information from Biomart[598] and Gencode V.16. The TSS was considered to be the most upstream base of all the annotated transcript variants of the gene.

Infinium HumanMethylation450 BeadChip

All DNA samples were assessed for integrity, quantity and purity by electrophoresis in a 1.3% agarose gel, picogreen quantification, and nanodrop measurements. All samples were randomly distributed into 96-well plates. Bisulfite conversion of 500 ng of genomic DNA was done using the EZ DNA methylation kit (Zymo Research), following the manufacturer's instructions. 200 ng of bisulfite-converted DNA were used for hybridization on the HumanMethylation450 BeadChip (Illumina). The HumanMethylation450 BeadChip data were processed using the Bioconductor minfi package. We performed the "Illumina" procedure, which corrects for background signal and normalizes it, taking the first array of the plate as a reference. The methylation level (β) for each of the 485,577 CpG sites was calculated as the ratio of methylated signal divided by the sum of methylated and unmethylated signals plus 100. After the normalization step, we removed probes related to X and Y chromosomes. All analyses were performed in human genome version 19 (HG19).

Expression and chromatin immunoprecipitation (ChIP) analysis

Total RNA was reverse transcribed with the oligo-dT ThermoScript RT-PCR system (Invitrogen, Life Technologies, USA), according to the manufacturer's instructions. cDNA was amplified by real-time PCR using SYBR (Applied Biosystems) green detection and *PPIA* and *GAPDH* were used as housekeeping genes for normalization (**Table 5.1**). Protein lysates were obtained using Laemli Buffer 1X after washing the cells with cold PBS and the respective concentrations were determined using the Bio-Rad DC protein assay (Bio-Rad Laboratories, Hercules, CA). Subsequent to standard techniques of western blot, membranes were incubated with Anti-MEIS1 antibody (ab19867, ABCAM) and β-actin-HRP (A3854, SIGMA). ChIP analysis for DNMT3A was performed as previously described[599]. Primers are available upon request.

Table 5.1. Sequence of Quantitative RT-PCR Primers

Oligo Name	**Sequence**
Hs-*MEIS1*-F	GACAATTTCTGCCACCGGTAT
Hs-*MEIS1*-R	TGATCTCTGTTCCAAGAGGGC
Hs-*HOXA11*-F	AGCCTCCCTTCTTTTCTGCC
Hs-*HOXA11*-R	GGCTCAATGGCGTACTCTCT
Hs-*IRF8*-F	CACGCTGGCAAGCAAGATTA
Hs-*IRF8*-R	CGGTCCGTCACTTCCTCAAA
Hs-*HOXB2*-F	CAAGAAACCCAGCCAATCCG
Hs-*HOXB2*-R	CAGCTGCGTGTTGGTGTAAG
Hs-*NRG4*-F	TCAACCCTACTCTCTTGACCA
Hs-*NRG4*-R	AACGACTTGTGACTGGGACC
Hs-*ADAMTS5*-F	GCATCTAAGCCCTGGTCCAA
Hs-*ADAMTS5*-R	TCGTGGTAGGTCCAGCAAAC
Hs-*KLF2*-F	TTCGCATCTGAAGGCGCATC
Hs-*KLF2*-R	GAGAAGGCACGATCGCACAG
Hs-*CADM1*-F	TCTGCTGTTGCTCTTCTCCG
Hs-*CADM1*-R	GGTCTGCCTGTTGGGATTCA
Hs-*PRKCDBP*-F	AGCTCCACGTTCTGCTCTTC
Hs-*PRKCDBP*-R	GCTCTGGTGCCTTCTGGAAA
Hs-*TRIP6*-F	GCAGGAAGAGGAAGAGGAGG
Hs-*TRIP6*-R	ACACTGGCCAAAGTACTCCC
Hs-*RNASE6*-F	CAACAGCTTCTGAGCTTTGGAC
Hs-*RNASE6*-R	GCTTAGGCCAAGCATGAAGT
Hs-*PLD6*-F	CGACTACATGGCCCTCAACG
Hs-*PLD6*-R	GTTGTTCTGGATGGCTTGCG
Hs-*GAPDH*-F	TGCACCACCAACTGCTTAGC
Hs-*GAPDH*-R	GGCATGGACTGTGGTCATGAG
Hs-*PPIA*-F	ATGGTCAACCCCACCGTGT
Hs-*PPIA*-R	TCTGCTGTCTTTGGGACCTTG

4. Results and Discussion

To find downstream hypomethylated targets mediated by the *DNMT3A* mutational event, we have taken an unbiased epigenetic approach to examine the entire DNA methylome at the single-nucleotide level of a well known *DNMT3A* AML mutant cell line (OCI-AML3, which harbors a heterozygous R882C mutation)[600] and a widely used *DNMT3A* wild-type AML cell line (OCI-AML5). Using whole-genome bisulfite sequencing, we generated 476 146 848 and 497 572 515 sequencing reads, of which 74.3% (353 777 108) and 80.4% (400 048 302) mapped uniquely to the human genome, respectively. Genome wide, we achieved a base coverage of 23.1x for OCI-AML3 and 26.1x for OCI-AML5 and 32.8x and 32.5x at CpG dinucleotides, respectively, enabling us to interrogate DNA methylation levels for >25 M CpG sites genome wide (>5 reads per site). The complete whole-genome bisulfite sequencing data from OCI-AML3 and OCI-AML5 are illustrated in **Figure 5.1a**, and are available for download from NCBI GEO (National Center for Biotechnology Information Gene Expression Omnibus):

http://www-ncbi-nlm-nih-gov.sire.ub.edu/geo/query/acc.cgi?token=crcvqguqdhwxrsf&acc=GSE62303

We observed that *DNMT3A* mutant AML cells had a 9% (66.1% vs 75.1%) decrease in average DNA methylation level and fewer methylated CpG dinucleotides than did the *DNMT3A* wild-type cells (**Fig. 5.1b and c**). The diminished methylated CpG dinucleotide content in OCI-AML3 observed with respect to OCI-AML5 cells is consistent with the reduced DNA methyltransferase activity associated with the mutations described in DNMT3A[471,472,590,591]. To find specific target genes affected by the DNA hypomethylation events noted in the AML cells harboring the *DNMT3A* mutation, we searched for particular differentially methylated regions (DMRs) between the two AML cell lines. These were defined as consecutively and consistently differentially methylated loci located beyond the 95% confidence interval (CI) of the smoothed methylation profiles. Using these criteria, we identified 182 800 DMRs between OCI-AML3 and OCI-AML5 cells. The most common DMR change was the presence of a methylated sequence in OCI-AML5 that was unmethylated in OCI-AML3: 156 919 hypomethylated events that represented 86% of the identified DMRs (**Fig. 5.1d**). We focused on those hypomethylated DMRs located in unique candidate 5′-end regulatory promoters, which corresponded to a total of 1416 genes. To identify the hypomethylated promoters that had a transcriptional effect on the respective associated

genes, we complemented the whole-genome bisulfite sequencing data with the results of an expression microarray experiment for the OCI-AML3 and OCI-AML5 cell lines (http://www.cancerrxgene.org/downloads/). This approach yielded 292 genes with transcriptional activation associated with promoter hypomethylation in *DNMT3A* mutant cells relative to wild-type cells (**Fig. 5.1e** and **Table 5.2**).

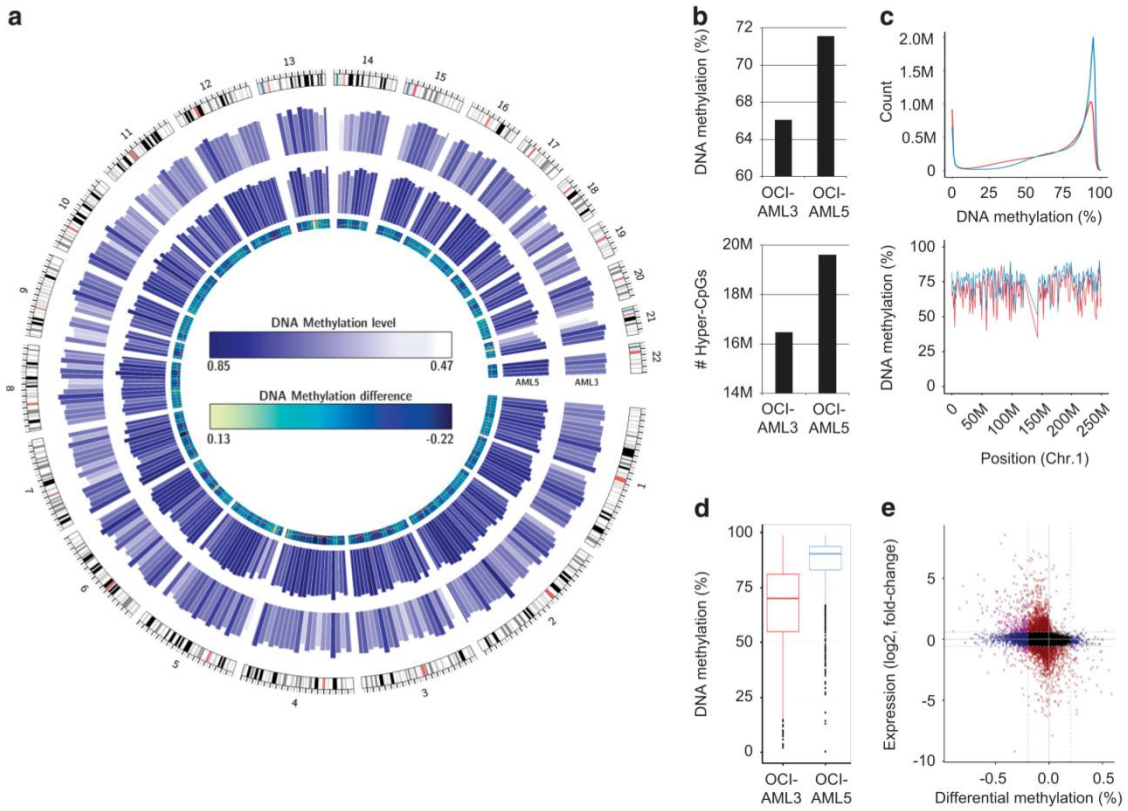


Figure 5.1. Complete DNA methylomes of *DNMT3A* wild-type and mutant AML cell lines. (**a**) Global DNA methylation levels in the *DNMT3A* mutant (OCI-AML3, outer circle) and wild-type (OCI-AML5, inner circle) cell lines analyzed by whole-genome bisulfite sequencing. Mean DNA methylation levels are displayed for 10 Mb genomic segments and all chromosomes. (**b**) Genome-wide analysis of DNA methylation levels at the CpG level (upper panel) and absolute number of hypermethylated (>0.66 methylation level; lower panel) CpG dinucleotides. (**c**) Genome-wide CpG methylation levels (upper panel) and DNA methylation profile exemplified by chromosome 1 (lower panel). (**d**) DNA methylation levels in DMRs hypomethylated in AML3. (**e**) Difference in promoter methylation (x axis; OCI-AML3 vs OCI-AML5) is associated with differential gene expression (y axis; OCI-AML3 vs OCI-AML5). Applied thresholds are indicated by dotted lines (δ DNA methylation >0.2; 1.5-fold change in gene expression). The 292 identified hypomethylated and overexpressed genes are highlighted in purple (upper left quadrant).

Table 5.2 (Part 1). List of hypomethylated differentially expressed genes (OCI-AML3 vs OCI-AML5).

	Gene Expression			DNA methylation
Gene symbol	OCI-AML3	AML5	Fold change (log2)	AML3-AML5
AARS	10.00	7.69	2.32	-0.30
ABCB4	5.82	3.14	2.67	-0.45
ACAP1	4.77	3.40	1.37	-0.48
ACTN1	8.73	7.41	1.32	-0.20
ADAMTS5	5.52	2.94	2.58	-0.36
ADCK4	4.52	3.80	0.72	-0.24
ADORA3	5.23	3.64	1.59	-0.30
ADRBK1	6.16	5.07	1.10	-0.34
AFF3	6.17	4.69	1.48	-0.47
ALDH2	6.86	3.06	3.80	-0.49
ALDH9A1	9.02	7.54	1.47	-0.21
AMN1	6.44	3.91	2.53	-0.84
ANKRD13D	6.85	5.01	1.84	-0.32
ANPEP	5.00	3.22	1.77	-0.23
ANXA2	7.89	5.32	2.57	-0.24
APOBR	6.44	4.97	1.47	-0.38
APOC1	7.64	4.23	3.41	-0.25
ARHGAP1	6.71	5.73	0.99	-0.20
ARHGAP22	3.86	2.76	1.10	-0.38
ARHGEF17	4.13	3.36	0.77	-0.30
ART3	6.19	2.95	3.24	-0.29
ASGR2	8.25	3.30	4.96	-0.21
ATF4	12.84	11.74	1.10	-0.24
ATF5	7.74	4.26	3.48	-0.37
ATM	7.24	5.22	2.02	-0.47
AZI2	8.57	6.95	1.62	-0.24
B3GNT2	8.93	6.96	1.97	-0.26
B3GNT8	4.41	2.93	1.47	-0.35
BAK1	4.88	3.57	1.32	-0.24
BCL2A1	5.71	2.69	3.03	-0.28
BEX1	11.94	3.42	8.52	-0.41
BHLHB9	4.38	3.71	0.67	-0.39
BMP2	8.97	3.15	5.82	-0.26
BRSK1	4.74	3.54	1.20	-0.23
BTF3L4	8.62	7.35	1.27	-0.30
C11orf48	3.80	3.14	0.65	-0.28
C19orf59	9.12	4.70	4.42	-0.23
C1orf162	9.43	4.29	5.14	-0.32
C22orf46	4.85	4.10	0.75	-0.21
C2orf69	8.43	7.82	0.62	-0.34
C9orf64	6.92	4.57	2.34	-0.23
C9orf85	8.48	6.26	2.22	-0.51
CABP4	5.42	3.54	1.88	-0.42
CACNA2D4	8.30	4.25	4.05	-0.41
CADM1	6.58	3.37	3.20	-0.41

Table 5.2 (Part 2). List of hypomethylated differentially expressed genes (OCI-AML3 vs OCI-AML5).

CAMK1	5.55	3.03	2.53	-0.49
CANX	10.54	9.72	0.82	-0.25
CCDC28B	6.01	4.14	1.87	-0.23
CCR1	7.48	3.91	3.57	-0.39
CD1D	10.30	5.94	4.36	-0.21
CD226	4.44	2.79	1.65	-0.23
CD320	5.87	3.52	2.35	-0.58
CD70	8.24	7.02	1.23	-0.29
CDA	5.09	3.11	1.98	-0.37
CDKN1B	7.84	7.17	0.67	-0.27
CDKN2B	3.60	2.97	0.63	-0.90
CECR6	9.56	3.58	5.98	-0.28
CENPH	8.93	7.80	1.13	-0.21
CFD	10.97	10.34	0.63	-0.25
CLSTN2	4.47	3.26	1.21	-0.23
CNBP	10.71	10.02	0.69	-0.31
COL15A1	6.05	3.55	2.50	-0.29
COPZ2	7.46	5.96	1.49	-0.32
CPNE8	7.72	5.06	2.66	-0.22
CRCP	6.70	5.67	1.03	-0.24
CRIM1	3.75	3.00	0.75	-0.20
CRK	6.41	5.79	0.62	-0.21
CTSG	12.92	9.25	3.67	-0.66
CYB5R2	4.72	3.17	1.55	-0.67
DENND2D	7.41	5.05	2.36	-0.26
DHRS9	11.20	9.84	1.36	-0.42
DNAJC4	5.82	5.14	0.69	-0.21
DOK2	7.75	4.84	2.91	-0.65
DUSP12	9.47	7.93	1.54	-0.25
EDEM1	5.39	4.69	0.71	-0.20
EEF1A1	6.66	6.05	0.60	-0.22
EFEMP2	5.26	3.88	1.37	-0.46
EIF2A	11.09	10.46	0.62	-0.35
EIF3M	8.87	8.26	0.61	-0.32
ELANE	8.35	3.06	5.29	-0.67
EMR2	4.54	3.82	0.71	-0.20
F2RL3	4.58	3.82	0.76	-0.30
FASTK	6.20	5.51	0.70	-0.22
FCER1G	11.17	3.34	7.84	-0.45
FCRLB	4.44	3.06	1.37	-0.22
FEZ1	7.65	3.14	4.51	-0.31
FEZ2	9.74	8.08	1.65	-0.37
FSTL4	3.91	3.30	0.61	-0.25
FUT7	5.85	4.09	1.76	-0.40
GALM	6.69	3.30	3.39	-0.46
GGPS1	7.60	6.93	0.67	-0.35
GNG12	3.57	2.98	0.59	-0.21
GPSM3	5.99	4.28	1.71	-0.25
GRAMD4	5.89	5.11	0.78	-0.31
GRSF1	8.47	7.70	0.77	-0.26

Table 5.2 (Part 3). List of hypomethylated differentially expressed genes (OCI-AML3 vs OCI-AML5).

GSTT1	4.94	4.13	0.81	-0.33
GYG2	4.67	3.33	1.34	-0.46
HADHA	9.03	8.36	0.68	-0.44
HADHB	8.66	7.98	0.68	-0.43
HAL	7.58	5.35	2.22	-0.51
HENMT1	7.85	3.50	4.35	-0.80
HERPUD1	10.28	7.14	3.14	-0.26
HIVEP3	5.63	4.81	0.82	-0.27
HNMT	7.75	3.18	4.57	-0.33
HOXA11	8.78	2.99	5.79	-0.25
HOXA13	8.33	3.07	5.26	-0.23
HOXB2	6.70	3.03	3.67	-0.31
HS3ST4	4.33	2.92	1.41	-0.41
HSD17B8	6.67	3.75	2.92	-0.37
HYLS1	7.80	5.82	1.98	-0.35
ICAM3	7.16	5.35	1.81	-0.39
IFI16	9.57	5.12	4.44	-0.21
IGFBP3	4.65	3.10	1.55	-0.25
IL11RA	4.48	3.65	0.83	-0.26
IL31RA	4.53	3.39	1.14	-0.24
INPP4A	6.04	5.31	0.73	-0.30
IRF8	11.32	9.28	2.04	-0.40
IRX5	8.89	6.99	1.90	-0.54
ITGB2	11.35	10.55	0.80	-0.23
ITPKC	4.31	3.62	0.68	-0.25
KCTD11	4.33	3.69	0.64	-0.48
KDM2A	6.64	5.82	0.82	-0.30
KIAA0226L	4.12	3.31	0.81	-0.27
KIAA0930	8.62	7.66	0.95	-0.32
KIF7	4.62	3.77	0.85	-0.35
KLF11	5.28	4.06	1.23	-0.35
KLF15	5.27	4.41	0.86	-0.41
KLF2	7.65	5.37	2.28	-0.25
LAMP5	9.68	6.35	3.33	-0.24
LAT2	9.73	8.69	1.04	-0.24
LDHC	6.58	3.02	3.56	-0.62
LLGL2	5.14	4.43	0.71	-0.42
LPCAT1	10.19	8.05	2.13	-0.29
LPHN2	4.68	3.03	1.65	-0.23
LPPR3	6.36	5.23	1.13	-0.39
LRRC58	9.50	8.83	0.67	-0.24
LTBP3	3.96	3.31	0.64	-0.26
LTBR	7.49	3.60	3.89	-0.48
LXN	5.04	4.11	0.93	-0.36
MACROD1	5.25	4.45	0.79	-0.31
MAD1L1	8.54	5.96	2.58	-0.21
MAP1LC3A	5.46	2.95	2.51	-0.49
MBNL1	8.80	8.16	0.64	-0.23
MCOLN2	6.09	3.57	2.52	-0.29
MDM4	7.48	6.35	1.13	-0.22

Table 5.2 (Part 4). List of hypomethylated differentially expressed genes (OCI-AML3 vs OCI-AML5).

MEIS1	5.63	3.44	2.19	-0.43
METTL8	8.33	7.61	0.71	-0.29
MGA	6.61	5.87	0.73	-0.60
MKX	9.76	2.91	6.85	-0.32
MLST8	8.35	5.34	3.01	-0.20
MLXIPL	5.69	4.42	1.27	-0.32
MMP14	7.32	4.79	2.52	-0.32
MMP2	4.52	3.38	1.14	-0.39
MNDA	11.22	9.38	1.85	-0.21
MS4A3	11.55	9.76	1.79	-0.33
MSN	11.01	9.85	1.17	-0.50
MST1	5.21	4.05	1.16	-0.26
MX2	7.04	6.23	0.81	-0.34
NAA38	8.52	5.89	2.63	-0.25
NAP1L5	5.95	2.67	3.27	-0.26
NAPG	6.95	5.68	1.28	-0.20
NCF4	10.82	9.34	1.48	-0.24
NEFH	5.59	2.91	2.68	-0.23
NEK8	5.56	4.20	1.36	-0.21
NFKBIL1	7.04	4.40	2.65	-0.35
NLRP3	8.31	7.59	0.72	-0.23
NME3	8.66	7.40	1.26	-0.26
NOLC1	7.93	7.32	0.61	-0.28
NPY1R	5.06	3.29	1.77	-0.30
NR1H3	4.64	4.00	0.64	-0.33
NRG4	7.30	3.33	3.97	-0.23
NT5E	5.81	3.16	2.65	-0.58
NT5M	7.31	5.94	1.37	-0.31
NUDT13	4.97	4.21	0.76	-0.25
NUMB	5.69	4.96	0.73	-0.23
NXF3	6.92	3.25	3.67	-0.26
ONECUT2	9.00	4.99	4.01	-0.23
OSBPL11	7.73	5.94	1.79	-0.33
OSBPL5	5.11	3.63	1.48	-0.44
P4HB	10.38	8.78	1.60	-0.31
PCM1	7.41	6.00	1.41	-0.31
PDGFRL	4.18	3.40	0.78	-0.41
PGBD5	4.04	3.36	0.68	-0.20
PID1	6.41	3.63	2.78	-0.21
PIGU	7.56	6.20	1.36	-0.27
PIK3AP1	8.56	7.73	0.83	-0.24
PLD3	4.98	3.05	1.93	-0.23
PLD6	7.02	4.19	2.83	-0.20
PLEC	4.90	4.12	0.78	-0.46
PPAPDC3	4.51	3.36	1.15	-0.34
PPIL3	10.40	8.88	1.52	-0.25
PPP4R2	7.82	7.02	0.80	-0.22
PRKCDBP	6.50	3.46	3.03	-0.77
PROCA1	4.71	3.23	1.49	-0.22
PSD4	4.68	3.73	0.96	-0.23

Table 5.2 (Part 5). List of hypomethylated differentially expressed genes (OCI-AML3 vs OCI-AML5).

PSMD12	8.68	8.04	0.64	-0.22
PTCD2	4.16	3.14	1.02	-0.26
PTGER3	4.11	3.15	0.96	-0.31
PTPN7	7.33	6.00	1.33	-0.22
RAB5A	7.27	6.50	0.77	-0.27
RAC1	9.48	8.82	0.66	-0.26
RBBP9	6.74	5.22	1.52	-0.24
RBM38	7.76	6.47	1.29	-0.33
RBM47	5.65	3.40	2.25	-0.25
RFX7	5.15	4.50	0.65	-0.20
RINL	5.78	3.65	2.14	-0.37
RNASE6	10.46	3.98	6.48	-0.33
RNF114	7.96	7.25	0.71	-0.23
RNF157	4.31	3.59	0.71	-0.22
ROGDI	6.21	3.64	2.57	-0.21
RPGRIP1	3.76	3.14	0.62	-0.26
RPL36	7.04	6.39	0.65	-0.22
RPL39L	7.04	3.41	3.63	-0.20
RPL7L1	8.97	8.26	0.71	-0.25
RPS18	5.37	4.39	0.97	-0.37
RRAS	5.01	3.72	1.29	-0.23
RUSC1	5.24	4.30	0.94	-0.27
RXRA	7.61	6.73	0.88	-0.32
S100A11	10.36	7.21	3.15	-0.49
S100A13	5.45	4.03	1.41	-0.36
SAMHD1	7.77	4.61	3.16	-0.22
SCPEP1	8.62	7.48	1.14	-0.22
SEC23B	9.50	8.48	1.02	-0.29
SERPINH1	6.88	4.05	2.83	-0.23
SETDB1	6.76	6.13	0.63	-0.31
SFMBT2	6.28	3.74	2.54	-0.63
SGTB	6.39	4.47	1.92	-0.38
SIKE1	7.23	5.94	1.29	-0.25
SKA2	7.46	6.01	1.46	-0.29
SLC35D2	5.97	3.77	2.20	-0.62
SLC36A4	8.11	7.14	0.98	-0.23
SLC3A2	9.26	6.46	2.81	-0.24
SLC9A3	3.65	3.02	0.63	-0.20
SLFN12	4.74	2.64	2.10	-0.62
SOWAHC	7.16	3.28	3.88	-0.26
SP110	6.48	5.54	0.94	-0.23
SPAG9	5.63	4.86	0.77	-0.20
SPATC1L	5.41	4.62	0.79	-0.22
SPTLC2	9.01	7.12	1.89	-0.22
SRM	9.30	8.59	0.72	-0.20
STX12	7.20	6.44	0.77	-0.35
SUSD3	5.33	3.02	2.30	-0.46
SYAP1	6.96	6.06	0.91	-0.30
TAF6	4.12	3.52	0.59	-0.25
TARS	11.70	9.76	1.95	-0.21

Table 5.2 (Part 6). List of hypomethylated differentially expressed genes (OCI-AML3 vs OCI-AML5).

TCIRG1	7.85	4.79	3.05	-0.30
TGFB1I1	4.72	3.89	0.83	-0.27
TLE3	6.36	4.71	1.65	-0.28
TLE6	4.79	3.25	1.54	-0.23
TMC6	4.66	3.72	0.94	-0.39
TMEM115	6.41	5.82	0.59	-0.32
TMEM128	7.50	6.38	1.12	-0.40
TMEM138	5.51	4.60	0.91	-0.26
TMEM167A	10.82	9.76	1.06	-0.32
TMEM209	7.32	5.79	1.52	-0.22
TMEM41B	6.43	5.51	0.92	-0.47
TMEM47	6.35	3.26	3.09	-0.33
TNF	7.28	3.35	3.93	-0.49
TOMM20	9.29	8.23	1.06	-0.28
TOX2	4.06	3.32	0.74	-0.22
TRIB2	8.37	3.34	5.03	-0.21
TRIP13	8.32	7.64	0.69	-0.22
TRIP6	9.17	4.29	4.87	-0.61
TRIT1	6.49	5.43	1.07	-0.20
TRPM2	5.79	4.47	1.32	-0.45
TSEN54	7.20	6.42	0.78	-0.25
TTC38	4.78	3.98	0.81	-0.26
TUBG2	6.27	4.48	1.79	-0.22
UBE2C	9.98	9.32	0.66	-0.33
UBE2M	9.27	8.44	0.83	-0.21
UBE2Z	7.37	6.48	0.89	-0.23
UGGT1	5.99	5.22	0.77	-0.27
UNC13D	4.92	3.88	1.04	-0.22
USP4	10.15	9.56	0.59	-0.26
VIT	7.23	3.34	3.89	-0.42
VKORC1L1	6.04	5.30	0.74	-0.24
VLDLR	7.71	4.12	3.59	-0.21
VPS4B	5.81	5.19	0.62	-0.21
WDR54	8.53	7.19	1.34	-0.24
WDR55	6.58	5.62	0.96	-0.27
XPOT	11.19	10.24	0.95	-0.21
YDJC	9.52	8.40	1.12	-0.29
YPEL3	7.14	4.08	3.06	-0.44
ZNF133	6.60	5.33	1.27	-0.22
ZNF331	6.98	4.92	2.06	-0.38
ZNF439	3.82	2.61	1.21	-0.41
ZNF451	5.67	4.99	0.68	-0.23
ZNF513	5.36	4.49	0.87	-0.22
ZNF532	5.47	3.65	1.82	-0.26
ZNF562	4.38	3.59	0.79	-0.24
ZNHIT1	9.57	8.85	0.72	-0.20
ZXDC	5.52	4.89	0.63	-0.23

We next examined how the profile of genes with hypomethylation-associated expression derived from the *DNMT3A* AML cell line models translated to primary samples obtained from AML patients. To this end, we screened sixty-eight AML patients (whose clinical information is summarized in **Table 5.3**) for *DNMT3A* mutations in exons 10–23 by direct Sanger sequencing; we also hybridized these samples to a comprehensive DNA methylation microarray that interrogates ~450 000 CpG sites. We detected 14 *DNMT3A* mutations (21%) in our AML group, a similar percentage to that reported previously[471,472,590,591], consisting of 13 R882 mutations (7 R882H, 4 R882S and 2 R882P) and 1 S525C mutation. *DNMT3A* mutations were enriched in the AML cases that showed no cytogenetic abnormalities (Fisher's exact test, P=0.0353). None of our AML cases had *mixed lineage leukemia* (*MLL*) translocations. AML patients with *DNMT3A* mutations had a shorter 5-year overall survival (OS) (log-rank test; P=0.046; hazard ratio, 95% CI: 2.02, 1.00–4.10).

When we combined the *DNMT3A* mutational status data with the DNA methylation analysis of our 292 identified hypomethylated-activated genes, we were able to define a signature of 12 hypomethylated gene promoters that were significantly enriched in the primary AML cases carrying the *DNMT3A* mutations (Wilcoxon's test; P<0.01) (**Figure 5.2a** and **Table 5.4**). Interestingly, this 12-gene hypomethylation signature was also associated with worse OS (log-rank test; P=0.037; HR, 95% CI: 1.92, 1.03–3.60) (**Fig. 5.2a**). The DNA hypomethylation signature of the 12 genes was validated in an independent cohort of primary AML patient samples (n=194)[169], in which the described hypomethylated CpG sites were enriched in *DNMT3A* mutant patient samples (Fisher's exact test, P<0.01; **Fig. 5.2b**). The 12-gene hypomethylated signature was also associated with shorter OS in this validation group (log-rank test; P=0.014; HR, 95% CI: 1.89, 1.12–3.18) (**Fig. 5.2b**). We further confirmed by quantitative reverse transcription–PCR that the hypomethylated status of these candidate genes in the *DNMT3A* mutant OCI-AML3 cells was associated with a high level of expression of the corresponding transcripts, whereas their methylated status in *DNMT3A* wild-type OCI-AML5 cells was linked to transcriptional repression (**Fig. 5.3a**). These target genes include two homeobox genes (*HOXA11* and *HOXB2*), members of a family of transcription factors involved in differentiation that, it has been suggested, are hypomethylated in *DNMT3A* mutant AML[592,593].

Table 5.3. Clinical characteristics of the studied AML patient cohort.

Characteristics	N	%	DNMT3A mutational status			
			Wild-type		Mutated	
			N	%	N	%
Gender						
Male	32	47.1%	26	81.3%	6	18.7%
Female	32	47.1%	24	75.0%	8	25.0%
Unknown	4	5.8%	4	100%	0	0%
Age						
< 50	19	28.0%	12	63.2%	7	36.8%
> 50	45	66.2%	38	84.4%	7	15.6%
Unknown	4	5.8%	4	100%	0	0%
Risk group						
Favorable	10	14.7%	10	100%	0	0%
Intermediate	39	57.3%	26	66.7%	13	33.3%
Adverse	8	11.8%	7	87.5%	1	12.5%
Unknown	11	16.2%	11	100%	0	0%
Cytogenetics						
Normal	31	45.6%	20	64.5%	11	35.5%
Complex	4	5.9%	3	75.0%	1	25.0%
Others	22	32.3%	20	90.9%	2	9.1%
Unknown	11	16.2%	11	100%	0	0%
FAB subtype						
M0 to M2	34	50.0%	27	79.4%	7	20.6%
M3 to M4	24	35.3%	19	79.2%	5	20.8%
M5 to M7	6	8.9%	4	66.7%	2	33.3%
Unknown	4	5.8%	4	100%	0	0%
5-year Survival						
Alive	24	35.3%	21	87.5%	3	12.5%
Exitus	40	58.9%	29	72.5%	11	27.5%
Unknown	4	5.8%	4	100%	0	0%

However, most importantly, the highest-ranked candidate gene, whose five significantly differentially methylated CpG sites spanned the largest region among the 12 genes (**Table 5.4**), was the leukemogenic HOX cofactor *MEIS1*. The reactivation of the *MEIS1* gene at the protein level in the *DNMT3A* mutant cell line context was also confirmed (**Fig. 5.3b**). The targeting of the DNMT3A protein to the *MEIS1* gene was observed using the chromatin immunoprecipitation assay (**Fig. 5.3b**). We also noted the impact of DNMT3A mutant-mediated *MEIS1* hypomethylation in the context of the primary patient[169], whereby AML *DNMT3A* mutant patients were hypomethylated and had a higher level of expression of *MEIS1* (Spearman's correlation; ρ=−0.71, P<0.01)

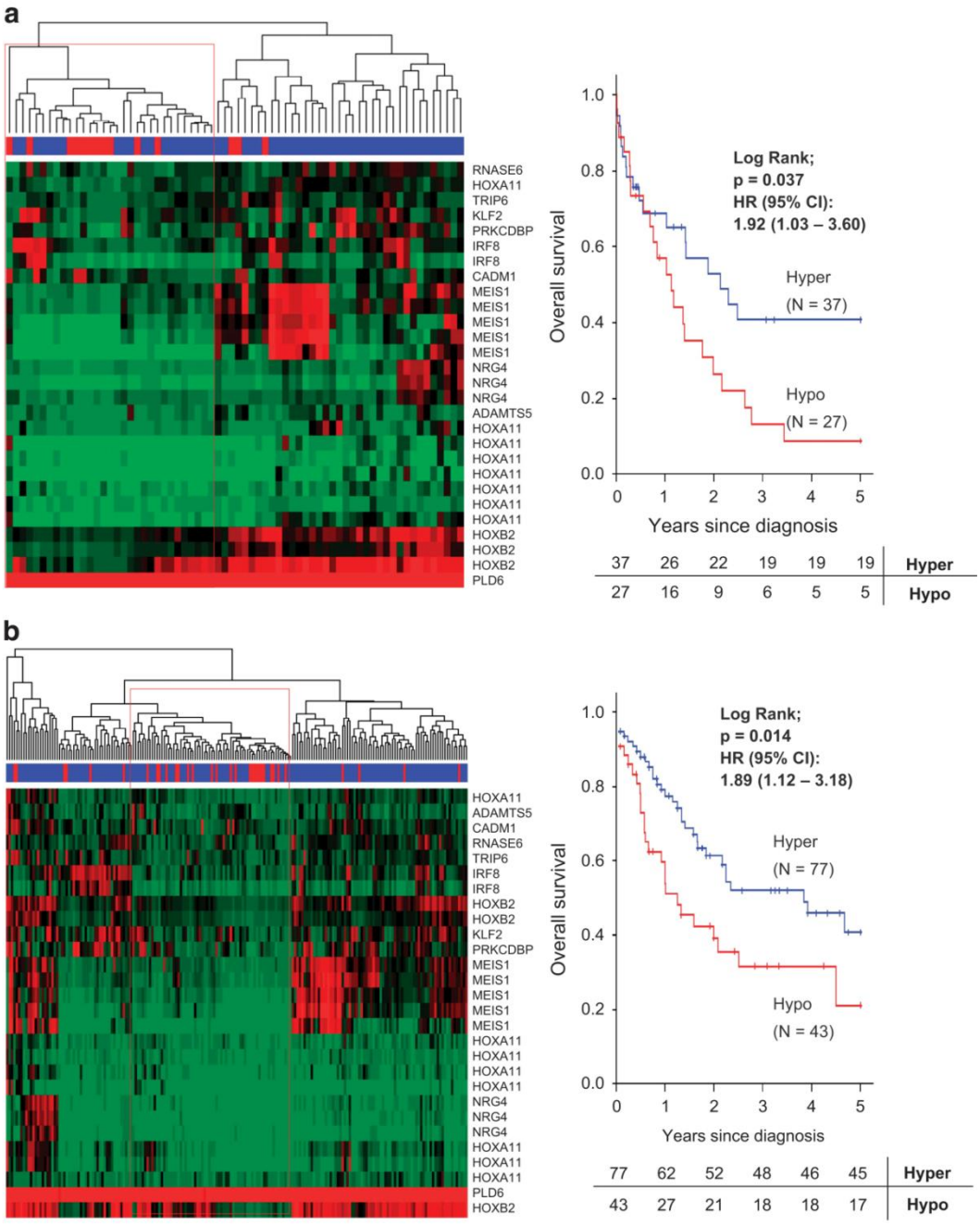


Figure 5.2. *DNMT3A* mutations in AML are associated with a DNA hypomethylation signature characterized by poor patient survival and *MEIS1* induction. (**a**) Hierarchical clustered DNA methylation levels (green: 0%; red: 100%) of 68 AML samples and 28 CpG sites significantly differentially methylated between *DNMT3A* mutant (red) and wild-type patients (blue). The red boxes indicate samples assigned to the DNMT3A mutant-related hypomethylated cluster. Differential survival analysis (5-year OS) of patients within (red line) or outside (blue line) the identified hypomethylated cluster (*n*=64, right panel). (**b**) Hierarchical cluster of the 28 CpG sites related to DNMT3A mutation in 194 primary AML patient samples[169]. The red boxes indicate samples assigned to the *DNMT3A* mutant-related hypomethylated cluster. Differential survival analysis (5-year OS) of patients within (red line) or outside (blue line) the identified hypomethylated cluster in the independent patient cohort (*n*=139, right panel).

Table 5.4. Differentially methylated genes in 68 AML patients (Wilcoxon test, $p<0.01$).

Gene symbol	DMR size (bp)	Expression fold change (log2)	CpG ID	# probes (450K)	Genomic location
MEIS1	6926	2.19	cg05877497, cg22731271, cg26537478, cg01271812, cg08238215	5	chr2:66667059-66673985
HOXA11	3147	5.79	cg00418216, cg20994254, cg24446586, cg05311410, cg12997720, cg26857670, cg13352750, cg17950095	8	chr7:27224700-27227847
IRF8	186	2.04	cg16705546, cg04599946	2	chr16:85936480-85936666
HOXB2	90	3.67	cg01882880, cg21097733, cg17573933	3	chr17:46623729-46623819
NRG4	69	3.97	cg00438616, cg00371829, cg22211154	3	chr15:76304796-76304865
ADAMTS5	1	2.58	cg02462195	1	chr21:28340304
KLF2	1	2.28	cg05906166	1	chr19:16437057
CADM1	1	3.20	cg10193817	1	chr11:115375226
PRKCDBP	1	3.03	cg16776065	1	chr11:6340486
TRIP6	1	4.87	cg21406967	1	chr7:100464553
RNASE6	1	6.48	cg23595621	1	chr14:21250201
PLD6	1	2.83	cg26043529	1	chr17:17106266

(**Fig. 5.3b**). MEIS1 is critical for the development of hematopoietic cells and has highly regulated transcriptional activity with high levels observed in hematopoietic stem cells and early progenitor cells, but downregulated expression in later stages of hematopoietic development[601]. This latter pattern appears to be disrupted in leukemogenesis, as persistent overexpression of *MEIS1* has been consistently observed in association with poor prognosis in acute leukemia patients[602]. In addition, *MEIS* overexpression causes shorter latency and accelerated progression in different leukemogenic models[603,604]. Interestingly, the common translocations in AML that involve MLL drive the activation of *MEIS1* that is essential for the initiation and maintenance of MLL-rearranged AML[603]. In this regard, *MEIS1* overexpression in murine bone marrow progenitor generates an AML with features in common with those driven by the MLL-fusion proteins[604].

Our results suggest that, in the absence of MLL fusions, as in our cases, an alternative pathway for engaging a leukemogenic MEIS1-dependent transcriptional program can be mediated by *DNMT3A* mutations. Under these circumstances, those AML patients carrying the alteration in the DNA methyltransferase would undergo a hypomethylation event at the *MEIS1* promoter that would lead to the overexpression of this key oncogene in leukemia[605].

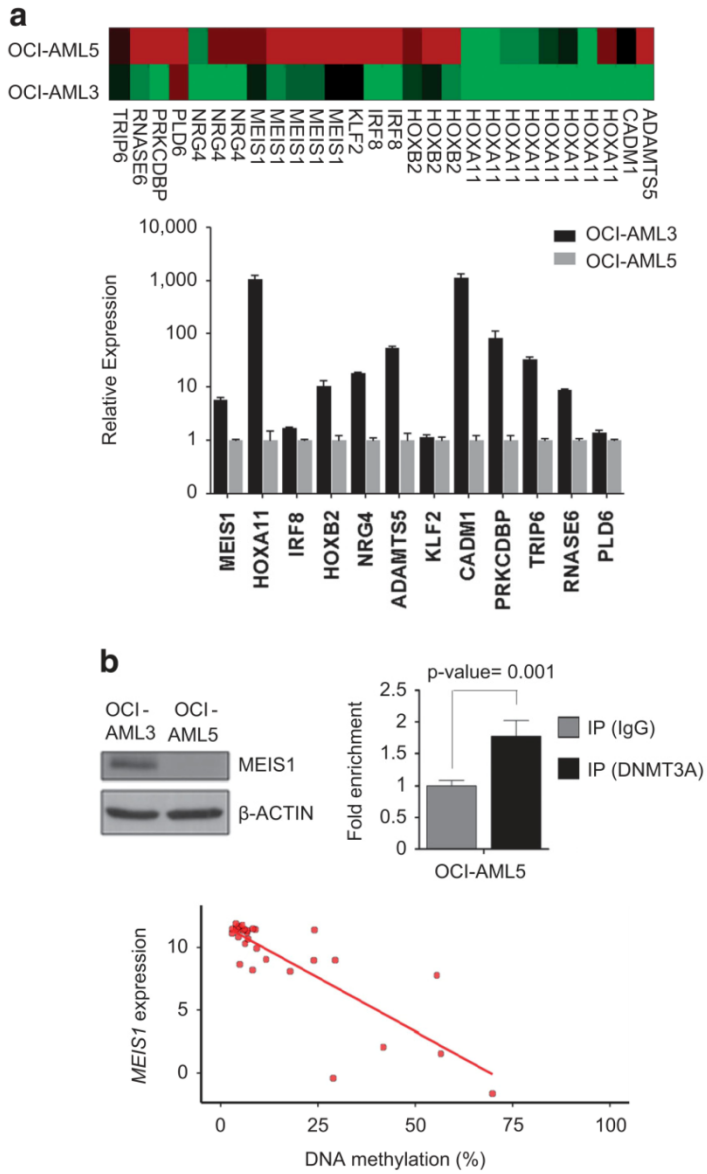


Figure 5.3. *DNMT3A* mutations in AML are associated with a DNA hypomethylation signature characterized by *MEIS1* induction. (**a**) DNA methylation level of the 28 CpG sites among the 12 candidate genes in the OCI-AML3 and AML5 cell lines analyzed by DNA methylation array (upper panel). Relative gene expression levels of the 12 genes related to *DNMT3A* mutation profiled by quantitative PCR in OCI-AML3 (black) and AML5 (gray) cell lines (lower panel). (**b**) Protein levels of MEIS1 in OCI-AML3 and AML5 cells analyzed by immunoblotting (top left panel). Quantitative chromatin immunoprecipitation assay to assess DNMT3A occupancy at the *MEIS1* studied CpG sites in AML5 cells. Data are presented as fold enrichment±s.e.m. Data of four independent experiments are shown. Significance of Student's *t*-tests is shown. IgG, immunoglobulin G (top right panel). Sites in *MEIS1* and transcriptional activity using matched data. Standard deviations are indicated by error bars. Bottom panel, association between the *DNMT3A* mutation-related differentially methylated CpG from 170 primary AML samples (Cancer Genome Atlas Research Network[169]). Mean DNA methylation levels over the five CpG sites of *MEIS1* and gene expression levels in *DNMT3A* mutant patients are shown.

5. Acknowledgements

This work was supported by the European Community's Seventh Framework Program (FP7/2007-2013) under grant agreements HEALTH-F5-011-282510—BLUEPRINT and ERC no. 268626—EPINORC project; the Cellex Foundation; the MINECO Project no. SAF2011-22803 and AGAUR 2014SGR633 grants; and the Health and Science Departments of the Catalan government (Generalitat de Catalunya). HJF was supported by the Program in Experimental Biology and Biomedicine, University of Coimbra, Portugal and ME is an ICREA Research Professor.

6. Unpublished data

There are some considerations to be done about the integration of this study on this PhD thesis. The initial aim of this project was the identification and characterization of known or predicted non-coding RNAs, up-regulated in AML cell lines and in patients harboring *DNMT3A* mutations.

Our hypothesis considered that the inhibited or limited activity of the *de novo* DNA methyltransferase DNMT3A would lead to an absence of DNA methylation in several loci, which in healthy conditions undergo *de novo* DNA methylation to mediate specific gene silencing. We speculated that between those genes, the lack of methylation in non-coding RNAs and their continuous expression would create a selective advantage assisting the malignant process. By analyzing the data obtained by WGBS and through the Infinium HumanMethylation450 BeadChip Array, we first identified several DMRs between the two cell lines. By investigating the genomic location of such DMRs, we verified that *MEIS1* DMR overlapped genomically with the promoter region of other genes encoded in the same locus in the sense and antisense strand. We selected one RefSeq (NCBI Reference Sequence Database) non-coding RNA and two Ensembl predicted non-coding genes, expecting their possible implication in the regulation of the protein-coding gene *MEIS1*. To test our hypothesis two pairs of primers were designed to amplify the cDNA of each one of the predicted candidate genes. However none of the three selected candidates were successfully amplified in our AML cell lines : OCI-AML3 harboring a mutation in *DNMT3A* and OCI-AML5, wild type for *DNMT3A*. Additionally, OCI-AML2 with only one copy of mutated *DNMT3A* was used in this validation but the predicted transcripts were not detected **(Fig. 5.4)**.

The lack of validation of our non-coding candidate genes and further analysis, prompted us to focus our attention on the role of this DMR in the direct epigenetic regulation of *MEIS1*, leading to the above transcribed original publication.

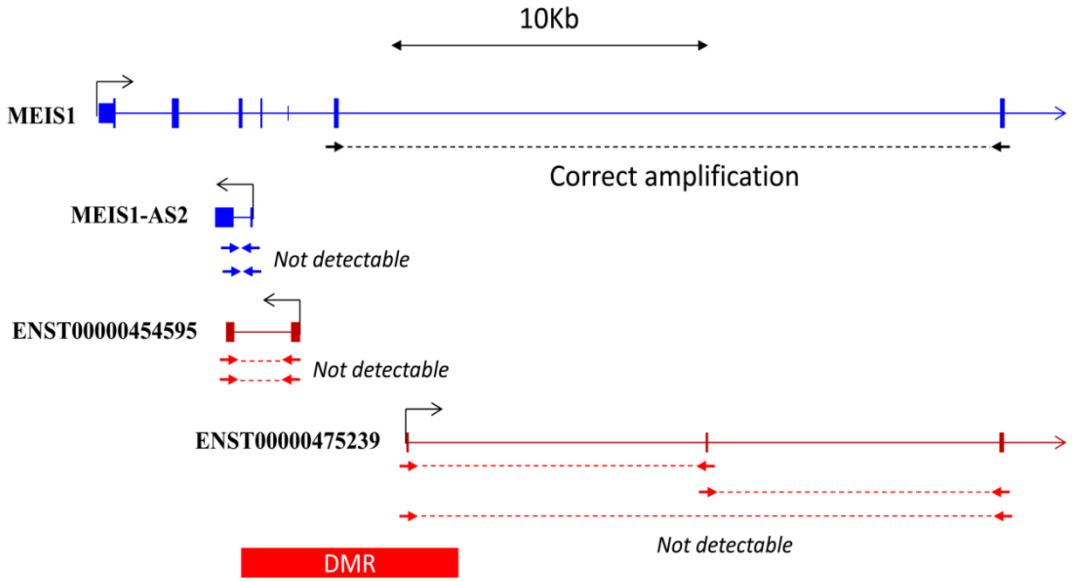
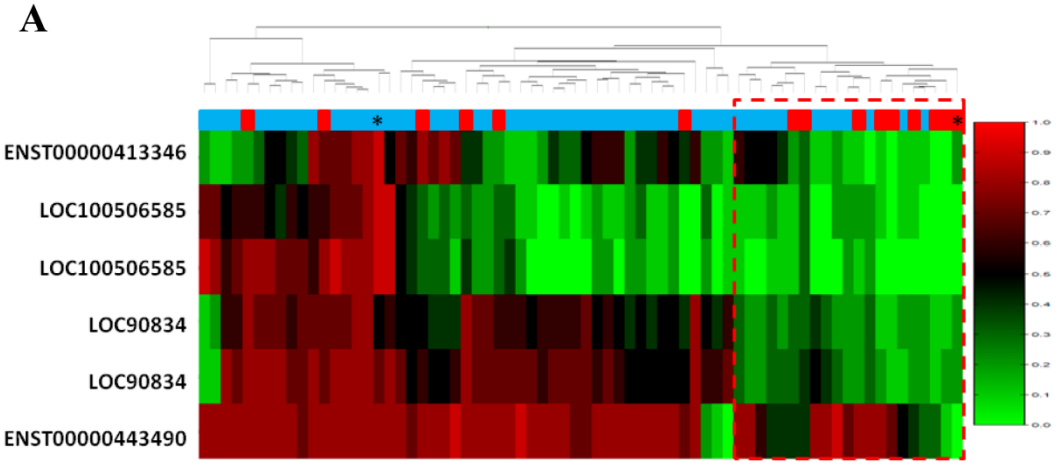


Figure 5.4. Localization of the DMR between OCI-AML3 and OCI-AML5, overlapping the promoter region and the TSS of some intragenic non-coding RNAs in the antisense strand of *MEIS1* gene. RefSeq genes are represented in blue and Ensembl predicted genes in red. Arrows in black, blue and red represent the primers used for validation. Amplification of *MEIS1* cDNA was used as a positive control.

Based on the genome-wide DNA methylation data, we concomitantly interrogated the genomic proximity of the previously identified highest-ranked DMRs to the promoter region of validated or predicted non-coding transcripts. Using a similar approach to the one used in the mentioned publication, we considered the promoter region of ncRNAs up to 2Kb of their TSS in both directions, comparing OCI-AML3 and OCI-AML5 cell lines. We were able to establish a DNA hypomethylation signature at 6 CpG differentially methylated sites (4 different lncRNAs). Expanding our analysis to patient samples harboring the wild type or the mutated form of *DNMT3A*, we also verified a tendency for their grouping in terms of DNA methylation profile (**Fig. 5.5A**).

We hypothesized that the identified lncRNAs undergo a hypomethylation-mediated transcriptional reactivation, expanding our analysis to a fifth lncRNA with partial genomic overlapping with one of the previously identified lncRNAs (**Fig. 5.5B**). By

A



B

Non-coding gene	Database	Strand	Chr	TSS (GRCh37/hg19)	CpG ID
ENST00000413346	Ensembl	+	1	231157357	cg02464665
LOC100506585	Refseq	+	7	157647277	cg06352807, cg12600749
LOC90834	UCSC	-	22	50173958	cg17923377, cg23659250
ENST00000443490	Ensembl	+	22	46476192	cg24361808
Quantified by partial overlapping with ENST00000443490					
MIRLET7BHG	Refseq	+	22	46481877	

C

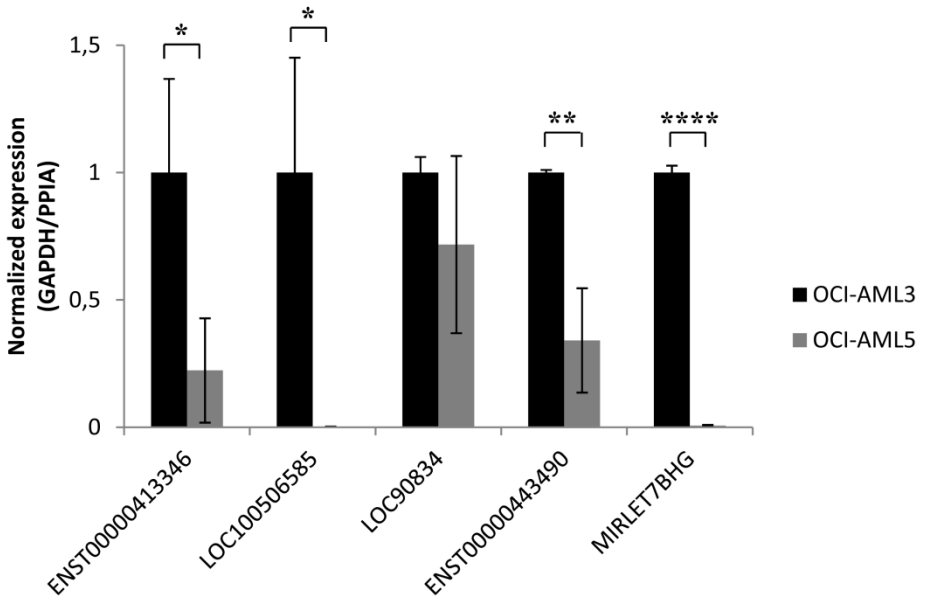


Figure 5.5. *DNMT3A* mutations in AML are associated with a DNA hypomethylation signature of non-coding RNAs. (**A**) Hierarchical clustered DNA methylation levels (green: 0%; red: 100%) of 68 AML samples at 6 CpG differentially methylated sites located up to 2Kb of the TSS of the ncRNAs listed, plus OCI-AML3 and OCI-AML5 cell lines (asterisks). *DNMT3A* mutant samples (red) have a tendency to cluster together (red dashed box). Samples harboring the wild type form of *DNMT3A* are represented in blue. (**B**) Genomic features and (**C**) quantification of the ncRNAs analysed.* p<0,05; **p<0,01; ****p<0,0001 Unpaired t-test, two-tailed.

analyzing the expression of the lncRNAs *ENST00000413346*, *LOC100506585*, *LOC90834*, *ENST00000443490* and *MIRLET7BHG* in OCI-AML3 and OCI-AML5, we verified that four of them were significantly more expressed in the *DNMT3A* mutant cell line OCI-AML3 ($p<0,05$) (**Fig. 5.5C**). Primers are listed in the **Table 5.5**.

In this parallel study, we were able to validate four lncRNAs undergoing a hypomethylation-associated transcriptional reactivation in *DNMT3A* mutant cells, namely *ENST00000413346*, *LOC100506585*, *ENST00000443490* and *MIRLET7BHG*. Our results suggested that these ncRNAs could be reactivated by the deficient activity of DNMT3A harboring the known driver mutation in AML, on Arginine 882. The establishment of correlations between the DNA methylation of the promoter regions of these lncRNAs and their expression in a larger group of cell lines or patient samples, as well as, the assessment of DNMT3A occupancy at their promoter region would help to define whether the possible reactivation of these lncRNAs are driven by *DNMT3A* mutations in AML. Moreover further studies are required to interrogate the possible functions of these lncRNAs in the leukemogenic process. *MIRLET7BHG* is the host

Table 5.5. List of oligos used in this complementary study.

Oligo Name	Sequence
Hs-*ENST00000413346*-F	TTTAGAGCCTGTGGTGGTTGA
Hs-*ENST00000413346*-R	CTCTGTCCCTGGACAACAACT
Hs-*ENST00000443490*-F	CCCAGGGACGTCATTTTCAC
Hs-*ENST00000443490*-R	CACGATGGGCTCCTGATGTC
Hs-*ENST00000454595*-F1	CAGCCATTCGGAACCCTCC
Hs-*ENST00000454595*-R1	AAAATGGCTGCTGGAAACGC
Hs-*ENST00000454595*-F2	TGCAAATGTATCAGAAGGCTGAA
Hs-*ENST00000454595*-R2	GCTTGTCGGGTGGGAAACTT
Hs-*ENST00000475239*-F1	CCAAGTTTCAGCTGGTCGGA
Hs-*ENST00000475239*-R1	CTTGCTCACTGCTGTTGTCC
Hs-*ENST00000475239*-F2	CTCGCTTGTGTGACCCCCT
Hs-*ENST00000475239*-R2	AGACTGTTTGCTTCCCAGTGTTT
Hs-*LOC100506585*-F	TGTCCTCCCTGCCTGATTCT
Hs-*LOC100506585*-R	CTCCTCCCTAGGGGATGCTC
Hs-*LOC90834*-F	CTTGAGTCTGTGGAGGGCAG
Hs-*LOC90835*-R	ACCCGCTTCATACACTTGCT
Hs-*MEIS1-AS2*-F1	TGAGCTCTCCACGTTTGCTT
Hs-*MEIS1-AS3*-R1	ATTTCTCCGGCCAGATACGC
Hs-*MEIS1-AS2*-F2	AGCTACAACCGTGGTGTTCT
Hs-*MEIS1-AS3*-R2	GATACCGTCCATGGCTCAGG
Hs-*MIRLET7BHG*-F	GCCCTTCAAAGTCCGTGTGG
Hs-*MIRLET7BHG*-R	TCCTGATGTCTCGGGTGGT

gene of *let-7a-3* and *let-7b*. Interestingly 6 members of the let-7 family, including *let-7a* and *let-7b* were described to be highly expressed in two AML cell lines, namely THP-1 and HL60[606]. Additionally, overexpression of *let-7a-3* was associated with poor prognosis in AML[607] and it was previously reported that the *let-7a-3* locus is methylated in normal human tissues and hypomethylated in some lung adenocarcinomas. Despite the later study was focused in a different genomic location, it suggested that the loss of methylation was associated with *let-7a-3* transcriptional activation and enhanced tumor phenotype[608].

Based on our results, we suggest that *DNMT3A* mutations could contribute to the hypomethylation-associated reactivation of the host gene that carries both *let-7a-3* and *let-7b*, connecting these ncRNAs to the associated worse prognosis of patients.

Conclusion: Altered expression of ncRNAs in DNMT3a mutant AML patients might have a role in oncogenesis.

Note: This section is not part of the Short Communication "DNMT3A mutations mediate the epigenetic reactivation of the leukemogenic factor MEIS1 in acute myeloid leukemia"

CHAPTER VI

Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer

Genome Biol. 2016; 17:11.

Holger Heyn*

Enrique Vidal*

Humberto J. Ferreira

Miguel Vizoso

Sergi Sayols

Antonio Gomez

Sebastian Moran

Raquel Boqué-Sastre

Sonia Guil

Anna Martinez-Cardús

Charles Y. Lin

Romina Royo

Jose V. Sanchez-Mut

Ramon Martinez

Marta Gut

David Torrents

Modesto Orozco

Ivo Gut

Richard A. Young

Manel Esteller

**co-authors*

1. Abstract

One of the hallmarks of cancer is the disruption of gene expression patterns. Many molecular lesions contribute to this phenotype, and the importance of aberrant DNA methylation profiles is increasingly recognized. Much of the research effort in this area has examined proximal promoter regions and epigenetic alterations at other loci are not well characterized.

Using whole genome bisulfite sequencing to examine uncharted regions of the epigenome, we identify a type of far-reaching DNA methylation alteration in cancer cells of the distal regulatory sequences described as super-enhancers. Human tumors undergo a shift in super-enhancer DNA methylation profiles that is associated with the transcriptional silencing or the overactivation of the corresponding target genes. Intriguingly, we observe locally active fractions of super-enhancers detectable through hypomethylated regions that suggest spatial variability within the large enhancer clusters. Functionally, the DNA methylomes obtained suggest that transcription factors contribute to this local activity of super-enhancers and that *trans*-acting factors modulate DNA methylation profiles with impact on transforming processes during carcinogenesis.

We develop an extensive catalogue of human DNA methylomes at base resolution to better understand the regulatory functions of DNA methylation beyond those of proximal promoter gene regions. CpG methylation status in normal cells points to locally active regulatory sites at super-enhancers, which are targeted by specific aberrant DNA methylation events in cancer, with putative effects on the expression of downstream genes.

2. Introduction

The naked DNA sequence alone cannot explain the different cellular functions or phenotypes of cells and organisms with identical genetic sequences, such as the presence of different tissues within the same individual[609], monozygotic twins[177], and cloned animals[610]. This is even more pertinent when we try to explain the pathophysiology of the most common human diseases with their multifactorial causes. The existence of different chemical marks, such as DNA methylation and post-translational modifications of histones, that regulate gene activity in the epigenetic layers has taken center stage in biology and medicine[611]. However, many studies have taken a biased approach in examining the regulatory sequences nearest to the transcriptional start sites of the studied genes and, with rare exceptions[179,314,612], other potentially important regions have been neglected in attempts to address the role of epigenomics in tissue identity and disease. In this context, the existence of super-enhancers[515] or locus control regions[613,614], large clusters of transcriptional enhancers that drive expression of genes that define cell identity, has been described. Most importantly, disease-associated variation is especially enriched in the super-enhancers of the corresponding cell types[615], and new super-enhancers for oncogenes and other transforming genes have been identified in cancer cells[516-519]. Herein, we present human DNA methylomes at single-nucleotide resolution of normal and cancer cells to identify epigenetic shifts in super-enhancers associated with these diseases.

3. Materials and Methods

Whole genome bisulfite sequencing

Cancer cell lines were obtained from the American Type Culture Collection (ATCC) and cultivated according to the provider's recommendations. All primary samples analyzed in this study were approved for research use by the respective ethics committees and were evaluated by trained personal before entering this study. DNA from cell lines or fresh frozen healthy and tumor samples was extracted using Phenol: Chloroform: Isoamylalcohol (Sigma).

We spiked genomic DNA (1 or 2 μg) with unmethylated λ DNA (5 ng of λ DNA per μg of genomic DNA) (Promega). We sheared DNA by sonication to 50–500 bp with a

Covaris E220 and selected 150- to 300-bp fragments using AMPure XP beads (Agencourt Bioscience Corp.). We constructed genomic DNA libraries using the TruSeq Sample Preparation kit (Illumina Inc.) following Illumina's standard protocol. After adaptor ligation, we treated DNA with sodium bisulfite using the EpiTect Bisulfite kit (Qiagen) following the manufacturer's instructions for formalin-fixed and paraffin-embedded tissue samples. We performed two rounds of conversion to achieve >99 % conversion. We enriched adaptor-ligated DNA through seven cycles of PCR using the PfuTurboCx Hotstart DNA polymerase (Stratagene). We monitored library quality using the Agilent 2100 BioAnalyzer and determined the concentration of viable sequencing fragments (molecules carrying adapters at both extremities) by quantitative PCR using the Library Quantification Kit from KAPA Biosystems. We performed paired-end DNA sequencing (two reads of 100 bp each) using the Illumina HiSeq 2000.

Sequencing quality was assessed using the Illumina Sequencing Analysis Viewer and FastQC software (http://www.bioinformatics.babraham.ac.uk/projects/fastqc). We ensured the raw reads used in subsequent analyses were within the standard parameters set by the Illumina protocol. Positional quality along the reads was confirmed to be QC > 30, and we excluded biases towards specific motifs or GC-enriched regions in the PCR amplification or hybridization. Sequence alignment and DNA methylation calling of WGBS reads were performed using Bismark V.0.7.4 software[594]. SAM/BAM and BED file handling was done using SAMtools[616], BEDtools[595] and Tabix[596]. Statistical analysis and graphical representation were performed with R[617] and multicore and ggplot2 libraries. We smoothed the DNA methylation profiles using a previously described method for processing WGBS data[597]. Briefly, the method assumes that the DNA methylation profile is defined by a varying function of the genomic location that can be estimated with a local likelihood smoother. We used hg19 as the reference genome and retrieved genomic information from Biomart[598] and GENCODE V.16 [618]. The TSS was considered to be the most upstream base of all the annotated transcript variants of the gene. The DNA methylation data sets for the two breast cancer cell lines (MDA-MB-468PT and MDA-MB-468LN) were previously published and are available under accession code GSE56763, Gene Expression Omnibus (GEO).

Hypomethylated regions

HMRs were identified as previously described[619]. Briefly, the raw methylated and unmethylated read counts of each CpG site, modeled with a beta-binomial distribution, provided the input for a hidden Markov segmentation model with two states (high and low methylation). Subsequently, a score was computed for each identified hypomethylated region as the number of CpG sites minus the sum of their methylation values. Further, the resulting regions were filtered on the basis of the 99th percentile of the score obtained by randomly permuting CpG sites. Differential DNA methylation in super-enhancers was calculated as difference (δ) in HMR occupancy (regions overlapping HMRs) between two samples.

In order to identify large HMRs, we followed a similar strategy to that described for identifying histone mark-defined super-enhancers[615], identifying regions that are substantially larger than their normal counterparts. We initially extracted HMRs with an average smoothed DNA methylation level of <0.2 and sorted the regions by genomic size. Secondly, we scaled the size and sorting index to map them to values over a 0–1 range. We then plotted the scaled region size (y axis) against the scaled region index (x axis) and examined a subset of the data (above the 90th percentile of size, high-scaled region index) and fitted a linear model with the log of the scaled size as outcome and the logistically transformed scaled index as predictor. Using the fitted parameter values, we reverted the variable transformation and identified the region index for which the derivative of the curve was 1 (i.e., a line with slope of 1 was the tangent to the curve). HMRs above this point were defined as large HMRs. This procedure was performed for each sample separately.

DNA methylation of super-enhancers

Super-enhancer coordinates were obtained from Hnisz, D. et al., 2013[615]. For the set of genomic regions defined as super-enhancers, we extended to each side by 50 % of the total length to include equally sized flanking regions in downstream analyses. Further, we scaled the position of each region to the center (0), the edges of the original region (−1 and 1), and the edges of the extended region (−2 and 2). We then retrieved the smoothed methylation information for each CpG inside the super-enhancers and flanking regions. Differential DNA methylation levels inside super-enhancers and

flanking regions were analyzed by Fisher's exact test, classifying CpGs as hypomethylated (<0.33 DNA methylation) or hypermethylated (>0.66 DNA methylation). Tissue-specificity of the DNA methylation profiles within super-enhancers was determined by assessing the tissue-matched DNA methylation profile, as described above, and their characteristics in an unmatched tissue context. Differences in DNA methylation (flanking region versus super-enhancer region) between tissues were analyzed by ANOVA.

WGBS-based tissue-specific hypomethylated super-enhancers were defined by identifying super-enhancers with an absolute HMR occupancy >20 % and a difference in HMR occupancy between the corresponding tissue and the remaining normal tissues >10 %. Each of these selected regions was considered as validated if the average beta value (HumanMethylation450 BeadChip) in the corresponding tissue samples was <33 % and the Student's *t*-test FDR comparing the corresponding tissue samples against the remaining samples was <0.05.

ChIP-sequencing data of the histone mark H3K27ac were retrieved from Hnisz, D. et al., 2013[615]. We computed the H3K27ac signal (ChIP versus input) and averaged the smoothed DNA methylation values in 50-bp windows. To define associations between histone signals and DNA methylation, we performed a Wilcoxon rank-sum test for the H3K27ac signal between hypomethylated (average <0.33) and hypermethylated (average >0.66) windows. Subsequently, we fitted a multivariate linear model with H3K27ac signal as response variable, DNA methylation status (hypo/hyper) and CpG density as predictors to assess the impact of CpG density on the association.

Differential DNA methylation analysis in cancer was done by computing the proportion of super-enhancers covered by HMRs. For each cancer sample and super-enhancer, we calculated the difference in HMR occupancy (δ HMR; cancer versus corresponding normal tissue). In order to assess overall differences between normal and cancer samples in super-enhancers, we performed a paired *t*-test for the reduction in DNA methylation (DNA methylation flanking super-enhancers versus DNA methylation inside super-enhancers) between the normal and cancer samples.

Expression analysis

The relationship between DNA methylation and gene expression was assessed using data obtained from RNA sequencing and public data sets. Raw RNA sequencing FASTQ reads from the breast cancer cell lines (MCF10A, MDA-MB-468PT and MDA-MB-468LN) were aligned against the human hg19 reference sequence using the TopHat read-mapping algorithm[620]. Conversion to BAM format was carried out using SAMtools[616]. Counts of alignments for each gene using BAM files were generated using BEDtools multicov[595]. In a subsequent analysis, the non-transformed cell line MCF10A was considered as control. Data from primary tumor samples were obtained from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). The analyzed samples included 110 normal breast samples and 30 matched invasive breast carcinomas (BRCAs), 12 normal colon and 258 adenocarcinomas (COADs), and 57 matched normal lung and adenocarcinomas (LUADs). To study the association of super-enhancer DNA methylation and gene expression, we obtained TCGA RNA-sequencing data (level 3) at the gene level and performed a Spearman's correlation test. Correlation analysis of gene expression and differential DNA methylation (normal versus cancer, $\delta > 0.1$) were performed using a Spearman's correlation test. Alternatively, we assigned the super-enhancers to the closest gene TSS, excluding those super-enhancers without a TSS within 1 Mb. We fit a log-linear model with RNA-Sequencing by Expectation Maximization-normalized gene expression as the response variable and average super-enhancer DNA methylation as predictor. The association between differential super-enhancer DNA methylation and gene expression was determined by fitting a linear model with the log fold-change of gene expression (cancer versus normal) as response and the δ HMR occupancy for all the super-enhancers gaining DNA methylation (δ HMR occupancy >0 %) or by Spearman's correlation test.

For microRNA quantification the Taqman microRNA Reverse Transcription kit and microRNA specific Taqman assays (Applied Biosytems) were used. The expression level was evaluated by real-time quantitative PCR using the 7900HT Fast Real-Time PCR System (Applied Biosystems). Expression values are reported as relative microRNA expression levels normalized to RNU6B expression.

Infinium HumanMethylation450 BeadChip

DNA from fresh frozen healthy and tumor samples was extracted using phenol:chloroform:isoamylalcohol (Sigma). All DNA samples were assessed for integrity, quantity and purity by electrophoresis in a 1.3 % agarose gel, picogreen quantification, and nanodrop measurements. All samples were randomly distributed into 96-well plates. Bisulfite conversion of 500 ng of genomic DNA was done using the EZ DNA Methylation Kit (Zymo Research), following the manufacturer's instructions. Bisulfite-converted DNA (200 ng) were used for hybridization on the HumanMethylation450 BeadChip (Illumina).

The HumanMethylation450 BeadChip data were processed using the Bioconductor minfi package[621]. We performed the "llumina" procedure that mimics the method of GenomeStudio (Illumina); specifically, it performs a background correction and a normalization taking as a reference the first array of the plate. We removed probes with one or more single nucleotide polymorphisms (SNPs) with a minor allele frequency (MAF) >1 % (1000 Genomes) in the first 10 bp of the interrogated CpG, based on Chapuy, B. et al., 2013[622]. In order to minimize batch effect, we used ComBat normalization[623]. The methylation level (β) for each of the 485,577 CpG sites was calculated as the ratio of methylated signal divided by the sum of methylated and unmethylated signals plus 100. After the normalization step, we removed probes related to X and Y chromosomes. All analyses were performed in human genome version 19 (hg19).

We identified HMRs within super-enhancer-overlapping probes (≥3) on the BeadChip and computed the average DNA methylation level for super-enhancers (HMR located probes) per sample (tissue-wise). Differences in DNA methylation levels at hypomethylated super-enhancer regions were determined using Student's *t*-test (FDR < 0.05). Selected super-enhancers were hierarchically clustered using Manhattan distance and median clustering algorithms. Finally, we assessed the BeadChip-based CpG methylation levels of common differentially methylated super-enhancers and performed hierarchical clustering using Canberra distance and Ward clustering algorithms with CpG-level data. The DNA methylation data for lung adenocarcinomas and lung squamous cell carcinomas were previously published and are available under accession code GSE39279, Gene Expression Omnibus (GEO).

The DNA hypomethylation observed at cancer-related super-enhancers was validated using data obtained from TCGA data portal (https://tcga-data.nci.nih.gov/tcga/). The analyzed samples included 41 matched normal and colorectal cancer samples. We obtained TCGA DNA methylation data from the HumanMethylation450 BeadChip (level 3) and averaged DNA methylation levels per super-enhancer containing ≥3 probes in the hypomethylated region. Significant differences between normal and cancer samples were assessed using a Wilcoxon test, with values of $p < 0.01$ considered to be significant.

CNV analysis

To test for biases in DNA methylation analysis due to CNV in cancer samples, we applied two independent approaches based on DNA methylation or SNP array data. For the 714 primary cancer samples analyzed using the HumanMethylation450 BeadChip, we performed a copy number analysis comparing cancer and normal samples using Bioconductor and the CopyNumber450K R package for CNV inference using the Illumina 450 k DNA methylation assay. We defined a region to be aberrant if >50 % of the region presented a significant copy number alteration as reported by the software (FDR < 0.05). Alternatively, for TCGA data set of colorectal adenocarcinomas[624], we used level 3 CNV data and defined a region to be aberrant if >50 % of the super-enhancer region presented copy numbers <1.5 or >2.5. For the WGBS cancer samples, we hybridized genomic DNA on the HumanOmni5 SNP array (Illumina) and performed a copy number analysis based on GenomeStudio software (V.2011.1) routines for the HumanOmni5-4v1_B chips.

Ethics

The Clinical Research Ethics Committee of the Bellvitge University Hospital approved the current study under the reference PR055/10. All patients who supplied the primary tumor samples have given written informed consent. The experimental methods comply with the Helsinki Declaration.

Availability of supporting data

The bisulfite sequencing data sets supporting the results of this article are available in the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under

accession number SRP033252 and to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE52272). All HumanMethylation450 BeadChip data from this study are available in GEO under accession number GSE52272.

Functional enrichment analysis

Gene Ontology (GO) enrichment analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID; v6.7)[625,626]. The results were corrected for multiple hypotheses testing using the Bonferroni p-value adjustment method.

ChIA-PET Data Integration

We retrieved MCF7 Pol2 ChIA-PET data from Li, G. et al., 2012[627] (4 replicates) and kept anchors that were present in >1 replicate and where any of the anchor pairs overlapped a gene promoter (consensus anchor regions). Then, super-enhancers (SEs) not overlapping promoters were assigned to consensus ChIA-PET anchor regions. Finally, we assessed the association between SE DNA methylation and target gene expression with a Spearman's correlation test using TCGA data from normal breast samples[628].

5-Hydroxy DNA Methylation Profiling

Genomic DNA was quantified using Quibit fluorometer (Qubit® dsDNA BR Assay Kit). 4 µg of DNA in 150µL were sheared using g-TUBE (Covaris, Inc.) to 10kbp fragments, by centrifugation at RT for 1 min at 6,000 rpm (Eppendorf 5424). g-TUBEs were then inverted in the centrifuge, and centrifuged again for 1 min at 6,000 rpm (RT). Sheared DNA was recovered from the screw-cap, and processed through GeneJET purification kit to reduce volume from 150µL to 40µL. GeneJET columns were prepared with 50 mM NaOH twice, followed by Binding Buffer. Once GeneJET columns were prepared, 150µL of sheared DNA were processed on GeneJET columns, followed by 3 washes with 80% acetonitrile. DNA was recovered from the columns using 40µL of pre-warmed ultra-pure water (65°C). Concentrated sheared DNA was then quantified using Qubit, and all samples were spiked-in with 0.5% (w/w) of Digestion Control according to the quantification of the DNA.

Each sample was divided in two aliquots of 20µL each, and were processed following TrueMethyl 24 Kit User Guide (Version 3.1 July 2013, CEGX, UK). The aliquot intended for the oxidative bisulfite conversion (OxBs), underwent a pre-cleaning step, using BioRad P6 Micro-Bio spin columns to remove possible contaminants that could interfere the oxidation reaction. Then both aliquots were denatured for 30 min at 37°C with Denaturing Solution provided in the kit, and immediately kept on ice. To each aliquot intended for OxBs, 1µL of Oxidant Solution was added, while the aliquots intended for normal bisulfite conversion (Bs) 1µL of water was added instead. Samples were incubated for 30 min at 40°C, and centrifuge 10 min at 14,000 x g, to remove precipitates. Supernatant was then used for bisulfite conversion process by adding 5µL of Bisulfite Additive and 170µL of Bisulfite Reagent, and incubated 2 cycles in a PCR machine (5 min @ 95°C; 20 min @ 60°C; 5 min @ 95°C; 40 min @ 60°C; 5 min @ 95°C; 165 min @ 60°C).

Bisulfite converted DNA was clean-up by pelleting precipitated salts, then subjected to a desulfonation reaction and finally clean-up again using Amicon Ultra 0.5 30kDa filter columns as indicated on the protocol with the modification on step 4.15 to increase the centrifugation time to 65 min, as indicated on TrueMethyl workflow for 450k Analysis (CEGX Version 1.1 December 2013). Both aliquots were quantified using Qubit ssDNA Assay Kit and 160 ng of bisulfite converted DNA were processed following standard protocol for HumanMethylation450 microarray.

Raw methylation data (idats) were background corrected and normalized using Methylation module (v1.9.0) of Illumina GenomeStudio software (V2011.1) in order to compute β-values. Hydroxymethylation levels were computed by subtracting β-values resulting from OxBs aliquot (5hmC+5mC) to the β-values from Bs aliquot (5mC).

Transcription Factor Occupancy Analysis

We used the ENCODE transcription factor binding site (TFBS) data available at UCSC, comprising 91 cell lines and 188 antibodies[497]. Given a sample and its corresponding super-enhancer (SE), we extended each SE 50 % of its size on each direction and split the resulting regions in 60 windows according to the relative distance to the center of the SE. We then computed the average methylation and the proportion covered by any of the TFBS (occupancy) among the window for each window and SE. As the

distribution of both methylation and TFBS occupancy was bimodal, we categorized these values using a cutoff of 50 % and performed a Fisher's exact test separately for the windows inside the SE.

Transcription Factor Enrichment Analysis

The statistically over-represented transcription factor binding sites (TFBS) in a set of sequences were compared with a background set using the CLOVER algorithm (Cis-eLement OVERrepresentation)[629] employing the following procedure: the JASPAR 2009 CORE collection of TFBS pattern matrices[630] were downloaded and converted to CLOVER format using PERL scripts, then the subset of sequences to search for over-represented TFBS and the background of sequences with which to compare them were specified. Specifically, we defined the hypomethylated colon cancer-related SE regions as the target set and the entire set of colon cancer-related SE loci as the background set. CLOVER compares each motif in turn with the sequence set and calculates a raw score that indicates how well represented the motif is in the subset. CLOVER also determines the statistical significance of the raw scores. Therefore, for the background set, CLOVER repeatedly extracts random fragments matched by length to the target subset of sequences, and calculates raw scores for these fragments. The proportion of times that the raw score of a fragment set exceeds or equals the raw score of the target set, is taken as the probability, P, that the motif's presence in the target set can be explained by chance alone. For each motif, a separate probability is calculated for each background file. Values of $P<0.05$ were considered to be statistically significant.

Chromatin Immunoprecipitation (ChIP) to Assess FOXQ1 Occupancy at Binding Sites and Super-enhancers Regions

Previous to ChIP, FOXQ1 monoexonic cDNA (lacking UTR regions) was amplified from HCT116 genomic DNA using specific primers with end adaptors containing EcoRI and BamH1 sequences, a Kozak sequence, and the N-terminal flag-tag (DYKDDDDK). The PCR products were cloned in bacteria, polymorphisms and mutations verified by Sanger-sequencing, and ligated into pLVX-IRES-ZsGreen1 plasmid from Clontech using EcoRI and BamH1 restriction enzymes (refseq: NM_033260). 10 µg of each plasmid were mixed with 7.5 µg of PS-PAX2 and 2.5 µg of PMD2.G plasmid in 1 ml jetPRIME buffer and 50 µl of jetPRIME reagent were

added (114-15, Polyplus transfection). After 20 min of RT incubation, the mix was diluted over a 10 cm disk containing 10 ml of DMEM and 293T cells at 80% confluence. After 48 h, medium containing viruses was recovered and 45-µm filtered. 3 ml of this medium plus polybrene (8 µM) was added to six-well plates containing the host cells at 80% confluence. After 48-72 h, cells were expanded and green positive cells were isolated by flow cytometry. Ectopic expression of FOXQ1 protein was evaluated by western blot using anti-Flag-HRP antibody (A8592, Sigma).

For ChIP, fresh cultures (1.5-2.0 x 10^7 cells) were cross-linked with 1% formaldehyde for 8 min and the reaction was blocked by adding glycine to a final concentration of 0.125 M. After washing twice with ice-cold PBS, cell pellets were resuspended in 1ml of Farnham lysis buffer (PIPES 5mM pH8.0, KCl 85mM, NP-40 0.5%) supplemented with protease inhibitor cocktail (Complete EDTA-free, Roche) and kept on ice for 10 min. The nuclear pellet was then resuspended in 1ml RIPA buffer (Tris-HCl 50 mM pH8.0, EDTA 20 mM, SDS 1%) supplemented with protease inhibitor cocktail (Complete EDTA-free, Roche) and kept on ice for 10 min. Samples were subsequently sonicated with S220 Covaris ultrasonicator for 18 min (peak incident power: 75W, duty factor: 10%, cycles per burst: 200). The chromatin size of the fragments obtained was 250-500 bp. Samples were diluted with dilution buffer (SDS 0.01%, Triton X-100 1.1%, EDTA 1.2 mM, NaCl 165 mM, Tris-HCl 16.7 mM pH 8.1). Magnetic beads were used for pre-clearing diluted chromatin (overnight at 4ºC) and for incubation with anti-Flag antibody (F1804, Sigma). Non-related mouse IgG antibody (12-371B, Millipore) was used as a negative control. The bead-antibody complexes were then incubated with pre-cleared chromatin for 8 h at 4ºC with rotation. The immune complexes were washed at 4ºC in rotation: twice with low-salt buffer (Tris-HCl 50 mM pH 8.0, NaCl 150 mM, SDS 0.1%, NP-40 1%, EDTA 1 mM, deoxycholate Na 0.5%), twice with high-salt buffer (Tris-HCl 50 mM pH 8.0, NaCl 500 mM, SDS 0.1%, NP-40 1%, EDTA 1 mM, deoxycholate Na 0.5%), twice with LiCl buffer (Tris-HCl 50 mM pH 8.0, LiCl 250 mM, SDS 0.1%, NP-40 1%, EDTA 1 mM, deoxycholate Na 0.5%) and once with TE Buffer (Tris-HCl 10 mM pH 8.0, EDTA 0.25 mM), 2 min each. Cross-linked chromatin was then eluted from the magnetic beads by adding elution buffer ($NaHCO_3$ 100 mM, SDS 1%). Samples were de-crosslinked overnight at 65ºC and incubated with proteinase K at 50 µg/ml final concentration for 1 h. Finally, DNA was purified with a PCR

purification kit (28106, Qiagen). The following SybrGreen gene-specific primer pairs were used:

MYC forward, 5'- GGATTTTTCCAATGGACACG-3',

MYC reverse: 5'- AAACAAAGCCAGACCTCAGC-3';

GPR forward: 5'- TCTTCTCATTCTGGGTCCACT-3',

GPR reverse: 5'- GGAAGTCAAAGATTCCTCAAGCA-3';

RNF forward: 5'- GCTTCCGTTTCAGAAAGCCA-3',

RNF reverse: 5'- TCCTCTTCTCTGCCCAATCA-3'.

Primer amplification efficiency and primer dimer formation was tested previously to ChIP. Western blot from ChIP pull downs were run to evaluate the presence of FOXQ1. ChIP data are presented as percentage of input ± SEM, $n \geq 3$. Significance of Student's *t*-tests (equal variance, one tail) is shown.

Experimental Validation of Transcription Factor Effects

For qRT-PCR experiments, total RNA from the colorectal cancer cell line SW1116 transfected with shRNAs against the respective transcription factors or a scrambled control was extracted using MAXwell (Promega) and retro-transcribed using the ThermoScript™ RT–PCR System (Invitrogen). Gene expression was determined by quantitative real-time PCR using SYBR Green (Applied Biosystems) according to the manufacturer's recommendations. Target gene expression levels were normalized to two housekeeping genes (*PPIA* and *B2M*).

Cell proliferation was determined for the colorectal cancer cell line SW1116 transfected with shRNAs against the respective transcription factors or a scrambled control by a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay. The cell viability was quantified over 6 days, staining the cells with MTT for 3 hours and blocking the reaction adding lysis buffer (HCl 20 mM, acetic acid 2.5%, SDS 20%, dimethylformamide 50%, pH 4.7). Measurements were performed at 560nm after overnight incubation at 37ºC.

Super-enhancer disruption with JQ1

After overnight incubation, the colorectal cancer cell lines HCT116 and SW1116 were treated with a 2-fold dilution series of JQ1 (A1910, Apexbio) or vehicle alone (DMSO at final concentration 0.2%). After 48h, cell viability was determined by a 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide (MTT) assay. To determine sublethal concentrations, we calculate the EC50 for JQ1 for both cell lines using the mean of four replicates in respect to vehicle treated cells. Viability curves were generated using a sigmoidal dose-response and a variable slope model (GraphPad Prism 5 software) from which the EC_{50} values were extracted. To investigate the effect of JQ1 treatment at the DNA methylation level, HCT116 and SW1116 cells were treated with a sub-lethal JQ1 concentration (10 μM) for 48h. After 24h, culture medium and drug were renewed and cells were harvested after 48h. RNA and DNA were extracted to evaluate mRNA expression level changes of *MYC*, *RNF43* and *GPRC5A* and for DNA methylation analysis on the Infinium HumanMethylation450 BeadChip, respectively.

4. Results and Discussion

We performed whole genome bisulfite sequencing (WGBS) to obtain unique DNA methylation data sets for five normal tissues and eight associated cancer samples (**Table 6.1**). Normal samples (n = 5) included brain, blood (CD19+), breast, lung and colon specimens. In order to enable the analysis of DNA methylation variance from different perspectives, we produced references data sets for cancer samples that involved both primary tumors (n = 2) and cancer cell lines (n = 6). These included a donor-matched primary colon triplet (normal tissue, primary cancer, liver metastasis) and matched primary and metastasis breast cancer cell lines, enabling us to analyze changes during tumor progression. The epigenetic peculiarities that could be present in cancer cell lines were addressed through replication experiments in an additional set of 78 normal tissue samples and 714 primary tumors using the HumanMethylation450 BeadChip (**Table 6.2**). The obtained data were also validated using the DNA methylation microarray profiles available for 208 normal samples and 675 primary tumor samples in The Cancer Genome Atlas (TCGA) projects (**Table 6.2**)[43,628,631].

Table 6.1. Whole genome bisulfite sequencing of 13 human samples.

Sample ID	Status	Tissue	Origin	Total reads	Coverage genome	Coverage CpG	Average methylation	SE[a]	SE covered[b]
CD19	Normal	B cells	Primary	318714023	6.0	14.1	76.0	688	99.0 %
Brain	Normal	Brain (white matter)	Primary	557237398	11.1	7.0	77.1	1067	99.6 %
Breast	Normal	Breast	Primary	606872747	15.1	32.1	73.0	1099	99.5 %
Colon	Normal	Colon	Primary	609043678	13.7	24.3	69.6	1023	99.4 %
Lung	Normal	Lung	Primary	333333332	7.2	8.7	74.4	1286	99.1 %
Colon_P	Cancer	Colorectal cancer	Primary	670281443	16.7	24.6	66.5	1023	99.4 %
Colon_M	Cancer	Colorectal cancer metastasis	Primary	652566967	16.3	24.7	62.4	1023	99.4 %
MDA-MB-468PT	Cancer	Breast cancer	Cell line	626288553	15.4	37.6	57.1	1099	99.4 %
MDA-MB-468LN	Cancer	Breast cancer metastasis	Cell line	600134926	14.3	37.1	42.8	1099	99.5 %
U87MG	Cancer	Glioblastoma	Cell line	281524883	6.3	8.5	55.7	1067	99.6 %
H1437	Cancer	Lung adenocarcinoma	Cell line	333333332	7.9	10.3	48.1	1286	99.1 %
H1672	Cancer	Small cell lung cancer	Cell line	329691560	7.4	10.5	65.6	1286	99.1 %
H157	Cancer	Lung squamous cell cancer	Cell line	333333332	7.8	10.7	41.8	1286	99.2 %

[a]*SE* is the number of super-enhancer regions determined in the respective normal tissue samples[615]
[b]SE covered is the percentage of super-enhancers covered by WGBS (>50 % of CpG sites)

Table 6.2. Genome-scale DNA methylation analysis of 78 normal tissue samples, 714 primary tumors and 24 metastasis samples (HumanMethylation450 BeadChip) and combined expression/DNA methylation analysis of 208 normal and 675 primary tumor samples (TCGA).

Cancer type	Status	Origin	Number of samples	Number of samples TCGA
Lung	Normal	Primary sample	26	57
Colon	Normal	Primary sample	18	41
Breast	Normal	Primary sample	19	110
Brain (white matter)	Normal	Primary sample	10	-
Blood (CD19+)	Normal	Primary sample	5	-
Lung adenocarcinoma	Cancer	Primary sample	321	216
Lung squamous cell carcinoma	Cancer	Primary sample	120	-
Colorectal cancer	Cancer	Primary sample	103	258
Colorectal cancer metastasis	Metastasis	Primary sample	24	-
Breast cancer	Cancer	Primary sample	66	201
Small cell lung cancer	Cancer	Primary sample	56	-
Glioblastoma	Cancer	Primary sample	48	-

Aligning uniquely mapping bisulfite sequencing reads (mean ~480 million reads per sample) of the original 13 samples undergoing whole genome single-nucleotide resolution analysis resulted in a median genomic coverage of 11.1× (14.1× CpG coverage) per sample. Consistent with previous reported results, apart from bimodal DNA methylation levels at promoter sites, the genomes presented high methylation levels, which were globally reduced in cancer samples (**Table 6.3** and **Fig. 6.1**)[314,612]. To estimate the relationship between super-enhancers and DNA methylation levels, we determined DNA methylation profiles for enhancer regions within their respective tissue types. From the super-enhancers previously described in our normal tissue types through the histone modification H3K27ac (identified as a superior and sufficient mark for the identification of super-enhancers[615], we could examine 99.3 % (5128 of 5163; >50 % CpGs covered; **Table 6.3**) using our WGBS data. We found significant enrichment of unmethylated DNA sequences within the super-enhancers compared with the flanking genomic regions (Fisher's exact test, odds ratio (OR) 5.6, $p < 0.001$), supporting the relevance of the features in the here interrogated context. In particular, the edges of the enhancers were CpG-unmethylated, clearly marking the boundaries of the regulatory regions (**Fig. 6.2a,b**), a phenomenon that was consistent throughout the analyzed tissue types (**Fig. 6.3**) and that could not be observed in traditional enhancers (**Fig. 6.4a,b**)[615]. Moreover, super-enhancers were significantly more hypomethylated than traditional enhancers (Fisher's exact test, OR 1.8, $p < 0.001$), further supporting DNA methylation to specifically indicate functionality in this enhancer subtype.

The fact that super-enhancer edges show lower DNA methylation levels compared with their center could be related to an enrichment of transcription factor binding sites at the extreme parts of the regions (Fisher's exact test, OR 5.33, $p = 1.0 \times 10^{-11}$; **Fig. 6.4c**)[632]. Indeed, DNA hypomethylation and transcription factor occupancy revealed a significant relationship (Fisher's exact test, OR 11.3, $p = 2.2 \times 10^{-16}$; **Fig. 6.4d**), consistent with previous reports describing a co-dependency of both regulatory mechanisms[633,634].

The extent of tissue-specific DNA methylation differences in the super-enhancer regions was low, with only 12.6 % (644 out of 5111) of them showing CpG methylation differences from different normal tissues (δ hypomethylated regions (HMRs) occupancy >10 %; **Fig. 6.5a** and **Supplementary Table 6.1**). We assessed variance in super-enhancer DNA methylation profiles by differential analysis of HMRs, focal sites of low

Table 6.3. Whole genome bisulfite sequencing of 13 human samples.

	Average DNA methylation (SD)				Super-enhancers covered by WGBS			
	Promoters	Exons	Introns	Intergenic	>5% CpGs	>20% CpGs	>50% CpGs	Total
Colon	51.4 (30.62)	81.7 (11.69)	81.51 (10.5)	74.78 (10.35)	99.8%	99.5%	99.4%	1023
Colon_P	49.82 (29.4)	78.88 (14.38)	78.67 (13.3)	69.32 (12.37)	99.8%	99.5%	99.4%	1023
Colon_M	47.47 (29.07)	75.94 (16.75)	75.61 (15.69)	64.13 (14)	99.8%	99.5%	99.4%	1023
Breast	50.92 (31.57)	82.47 (11.87)	82.92 (10.27)	76.95 (10.75)	100.0%	99.8%	99.5%	1099
468LN	42.3 (29.93)	62.81 (31.32)	62.77 (30.88)	41.86 (26.92)	100.0%	99.8%	99.5%	1099
468PT	49.14 (31.9)	75.33 (26.65)	74.67 (26.78)	59.45 (28.91)	99.8%	99.6%	99.4%	1099
Lung	52.98 (31.36)	84.14 (11.08)	84.13 (9.71)	78.3 (10.19)	99.8%	99.6%	99.1%	1286
H1437	42.08 (30.82)	68.65 (28.49)	67.85 (28.34)	47.52 (26.64)	99.8%	99.5%	99.1%	1286
H157	40.94 (31.33)	65.85 (33.87)	65.69 (33.54)	37.96 (26.85)	99.8%	99.5%	99.2%	1286
H1672	50.36 (32.54)	79.44 (22.5)	79.97 (21.54)	63.67 (27.99)	99.8%	99.6%	99.1%	1286
Brain	54.51 (32.2)	85.58 (12.2)	85.82 (10.44)	82.14 (10.89)	100.0%	99.9%	99.6%	1067
U87MG	47.28 (29.61)	71.43 (25.9)	71.73 (24.88)	52.85 (26.19)	100.0%	99.9%	99.6%	1067
CD19	55.14 (31.8)	86.13 (11.6)	85.86 (10.18)	80.14 (10.38)	100.0%	99.6%	99.0%	688

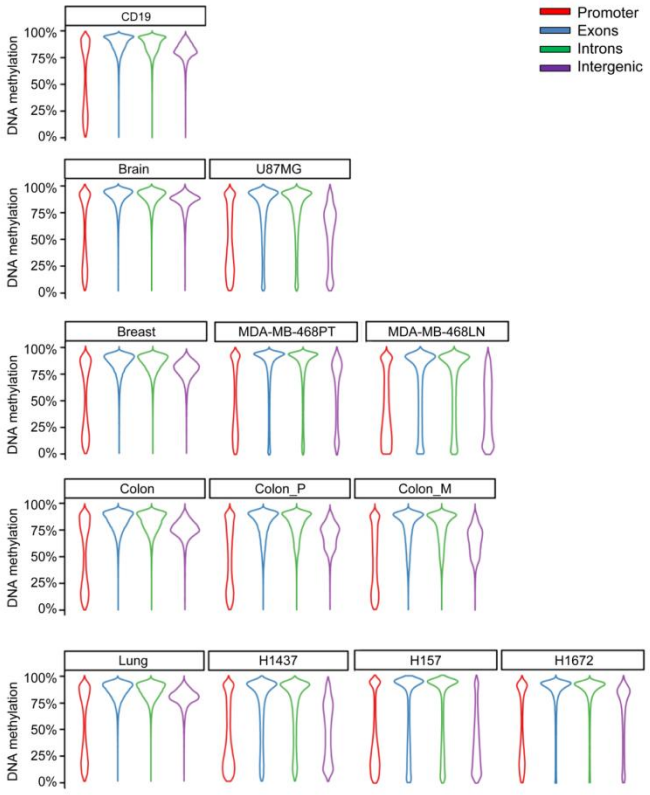


Figure 6.1. Genome-wide CpG methylation levels of 13 samples determined by whole genome bisulfite shotgun sequencing (WGBS). Displayed are CpG methylation levels at promoters (red), exons (blue), introns (green) and intergenic regions (purple). Samples are organized according to their tissue types.

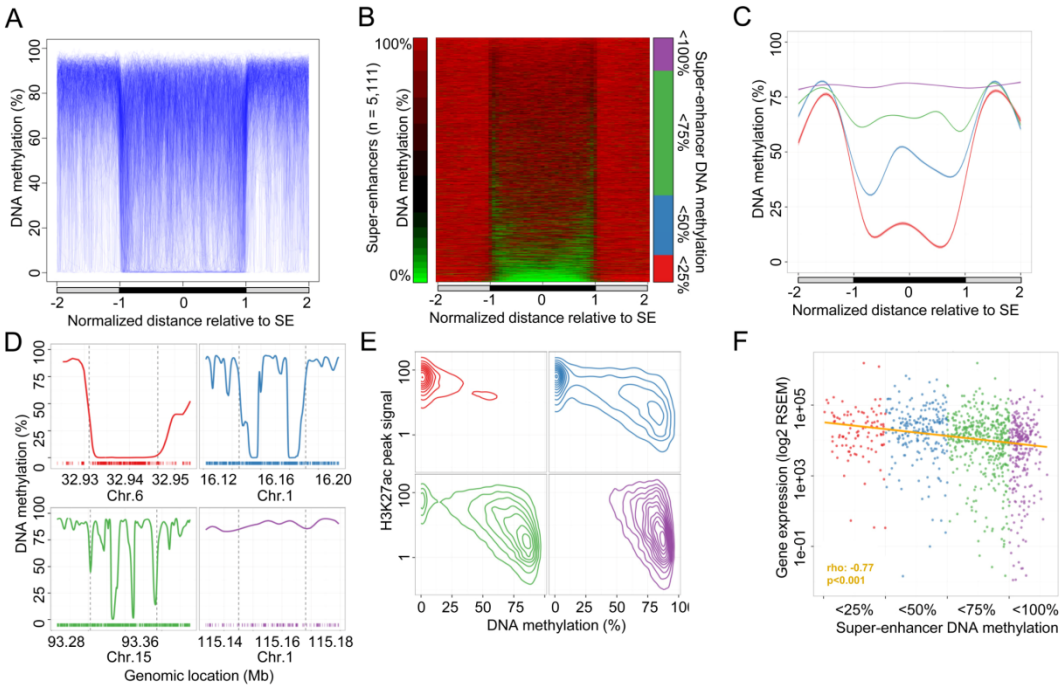


Figure 6.2. DNA methylation profile of super-enhancer regions derived from normal tissues determined by whole genome bisulfite sequencing (WGBS). **a** Scaled DNA methylation profile of 5111 super-enhancers (*SE*) in their respective normal tissues (n = 5). Each super-enhancer is represented by a single line (*blue*) and smoothed DNA methylation levels inside the super-enhancer (*black bar*) and equally sized flanking sequences (*gray bar*) are displayed. **b** DNA methylation levels of super-enhancers in their respective normal tissues (n = 5) in equally sized windows (*green*, 0 %; *red*, 100 %). Each horizontal line represents a single super-enhancer, ordered by average DNA methylation levels. Super-enhancers are grouped according to their average DNA methylation levels (*red*, <25 %; *blue*, <50 %; *green*, <75 %; *purple*, <100 %). **c** Smoothed average DNA methylation profile of all super-enhancers categorized into four groups on the basis of DNA methylation levels. **d** Examples of the DNA methylation profiles of breast super-enhancers representing the defined subgroups. Genomic locations of the super-enhancers (*dashed vertical lines*) and equally sized flanking regions are displayed and CpG dinucleotides locations are indicated (*bottom*, *colored bars*). **e** Association between DNA methylation levels and H3K27ac peak signals[615] in normal breast tissues and breast super-enhancers (n = 1091) displayed as averaged values (50-bp windows). Super-enhancers were classified into previously defined subgroups. **f** Gene expression levels of target transcripts in normal breast tissues. Scaled averaged expression levels of genes associated with breast super-enhancers (n = 1091) in normal breast tissue samples (n = 110; TCGA[628]). Super-enhancers were grouped according to their average DNA methylation levels. Significance of a Spearman's correlation test is indicated. *RSEM* RNA-Sequencing by Expectation Maximization.

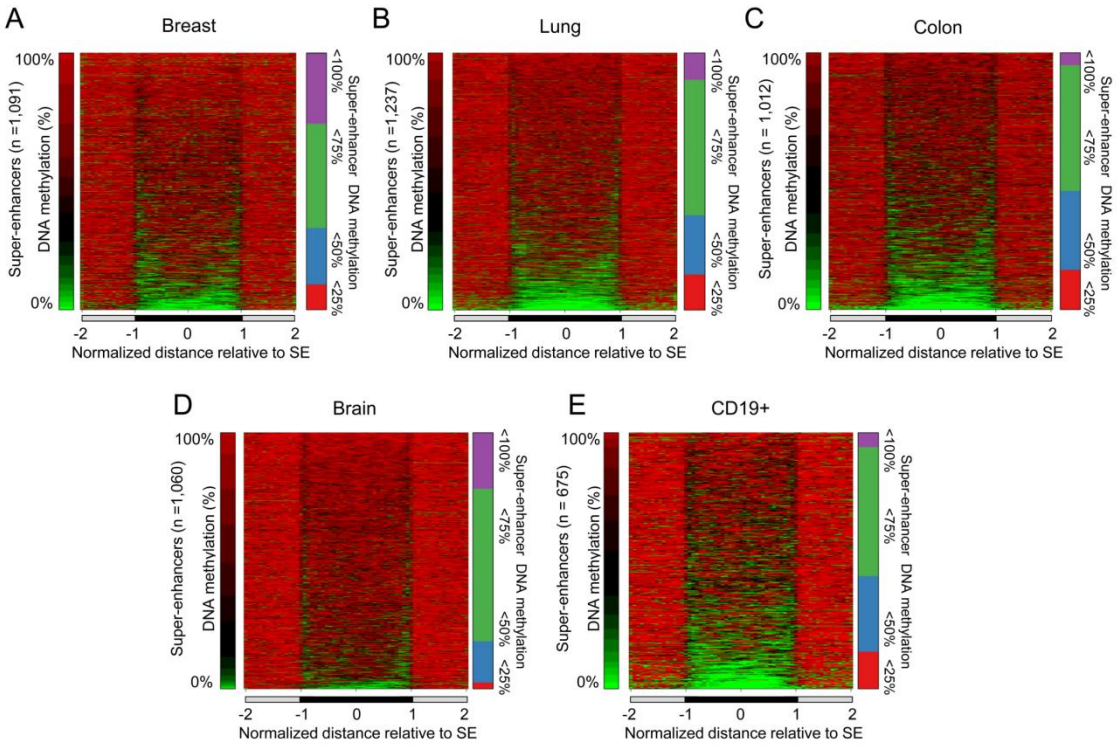


Figure 6.3. DNA methylation profiles of super-enhancer regions derived from the displayed normal tissues[615] determined by whole genome bisulfite shotgun sequencing (WGBS). Scaled DNA methylation profile of 1,091 super-enhancers detected in normal breast (**a**) 1,237 in normal lung (**b**), 1,012 in normal colon (**c**), 1,060 in normal brain (**d**) tissue and 675 in normal CD19+ cells (**e**). DNA methylation levels in equally sized windows inside super-enhancers (black bar) and flanking sequences (grey bar) are shown. Each horizontal line represents a single super-enhancer ordered by average DNA methylation levels (0%, green; 100%, red). Super-enhancers are grouped according to their average DNA methylation levels (red, <25%; blue, <50%; green, <75%; purple, <100%).

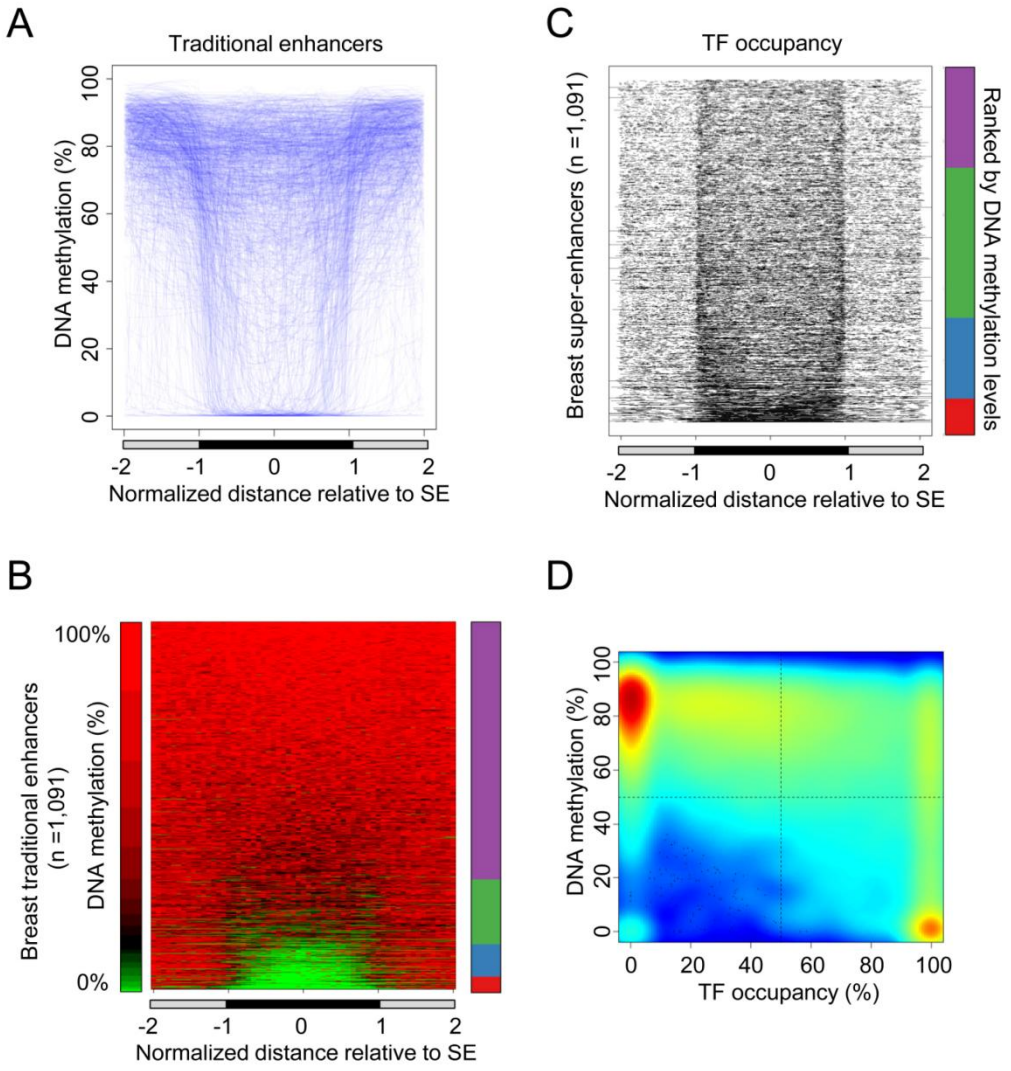


Figure 6.4. DNA methylation profiles of super-enhancer regions reveal a specific epigenetic profile. (**a**) Scaled DNA methylation profiles of a random set of 1,091 traditional enhancers (determined by H3K27ac in Human Mammary Epithelial Cells, HMECs[615]) in normal breast tissue. Each enhancer is represented by a single line (blue) and smoothed DNA methylation levels inside the enhancer (black bar) and equally sized flanking sequences (grey bar) are displayed. (**b**) DNA methylation levels of a random set of 1,091 traditional enhancers enhancers (determined by H3K27ac in Human Mammary Epithelial Cells, HMECs[615]) in normal breast tissue in equally sized windows (0%, green; 100%, red). Each horizontal line represents a single enhancer, ordered by average DNA methylation levels. Enhancers are grouped according to their average DNA methylation levels (red, <25%; blue, <50%; green, <75%; purple, <100%). (**c**) Transcription factor (TF) occupancy at scaled 1,091 breast super-enhancers (TF determined in 91 cell lines[632]) in equally sized windows. Each horizontal line represents a single super-enhancer, ordered by previous defined average DNA methylation levels (Figure 6.1). The proportion of the windows covered by TFs is color-coded (100% TF occupancy, black; no TF binding detected, white). (**d**) Correlation analysis of DNA methylation levels and TF occupancy in scaled 1,091 breast super-enhancer regions and equally sized windows. The density of windows is displayed gradiently (high, red; low, blue).

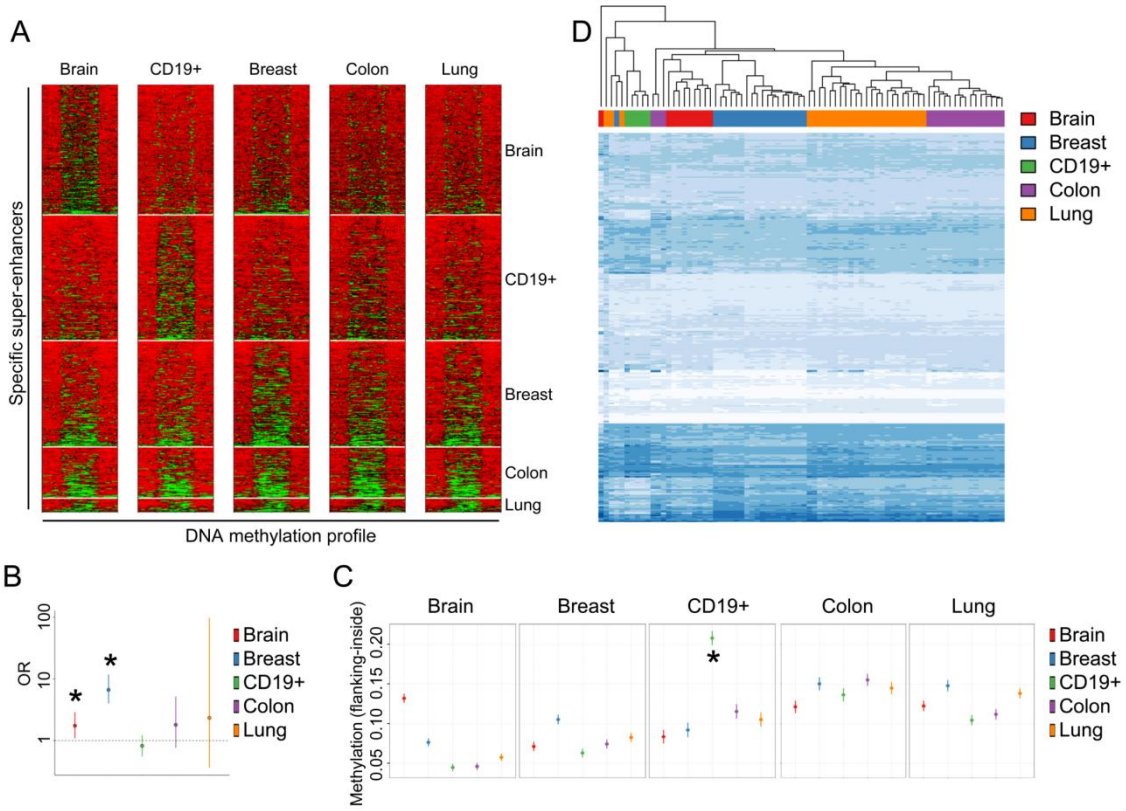


Figure 6.5. DNA methylation profiles of super-enhancer regions derived from normal tissues that display tissue-specific DNA methylation patterns. (**a**) DNA methylation levels of super-enhancers determined by WGBS are displayed for related and foreign tissue contexts. Scaled DNA methylation profiles of normal brain, CD19+, breast, colon and lung tissues in equally sized windows inside super-enhancers and flanking sequences are shown. Each horizontal line represents a single super-enhancer ordered by average DNA methylation levels of the specific tissues (0%, green; 100%, red). (**b**) Enrichment analysis of tissue-specific super-enhancers located in promoter regions compared to non-specific super-enhancer locations. Significant enrichments are indicated (*,$p<0.001$; Fisher's exact test). (**c**) Average DNA methylation differences between outside and inside super-enhancer regions (points) and their corresponding 95% CIs (segments). Each panel corresponds to a set of normal tissue-derived super-enhancers and normal tissue samples are color-coded. (**d**) Hierarchical clustering of DNA methylation levels in tissue-specific super-enhancers (rows) of normal breast (n=20), normal lung (n=26), normal colon (n=18), normal brain (n=9) tissues and sorted CD19+ cells (n=5), analyzed using the Human DNA methylation BeadChip. Average DNA methylation levels of probes (≥5) located in hypomethylated regions within tissue-specific super-enhancers are displayed (0%, white; 100%, blue).

DNA methylation levels that mark active regulatory loci[619,635,636], to account for the high heterogeneity at the large genomic regions represented by super-enhancers. Remarkably, tissue-specific HMRs at breast and blood super-enhancers were significantly enriched in specific transcription factor binding within the respective tissues, as measured by the occupancy of ten commonly profiled factors determined in CD19+ (GM12878; Fisher's exact test, OR = 2.81, $p < 0.001$) and breast cells (MCF7; Fisher's exact test, OR = 1.64, $p = 0.007$)[632]. Moreover, super-enhancers with tissue-specific DNA methylation levels in breast and brain samples were enriched at promoter regions compared with non-specific super-enhancers, in contrast to previous results that suggest tissue-specific DNA methylation to be enriched in *cis*-elements (Fisher's exact test, OR 6.64, $p < 0.001$ and OR 1.74, $p = 0.018$, respectively; **Fig. 6.5b**)[609]. The sample with the greatest DNA methylation difference compared with normal tissues was that of the CD19+ cell-related super-enhancers (ANOVA, $p < 0.001$; **Fig. 6.5c**), which was the only representative of a non-solid tissue type. It is of note that the presence of tissue-specific DNA methylation in this minor fraction of super-enhancers could be validated by genome-scale analysis using DNA methylation microarrays (HumanMethylation450 BeadChip). Of the normal tissue-derived super-enhancers, 75.5 % (486 of 644) were represented by at least three probes, in a unique set of 78 normal samples (**Table 6.2**), representing the analyzed tissue types, of which 71.4 % (347 of 486) showed significant difference between the respective tissue types (Student's *t*-test, false discovery rate (FDR) < 0.05; **Fig. 6.5d** and **Supplementary Table 6.2**). As examples of super-enhancer tissue-specific DNA methylation we can cite the genes encoding the RNA-binding protein QKI (involved in myelinization and oligodendrocyte differentiation), which is unmethylated in white brain matter but heavily methylated in all other normal tissues (**Fig. 6.6a**), and *lymphoblastic leukemia-associated hematopoiesis regulator 1* (*LYL1*; plays a role in blood vessel maturation and hematopoiesis), which is unmethylated in CD19+ cells but hypermethylated in all other normal tissues (**Fig. 6.6b**).

From the 5111 super-enhancers studied we established four categories based on their average DNA methylation levels (**Fig. 6.2b,c**). Remarkably, we determined striking differences between DNA methylation profiles at super-enhancers, ranging from fully hypermethylated to completely unmethylated (**Fig. 6.2d**). Moreover, focal hypomethylated regions pointed to spatial differences in DNA methylation within

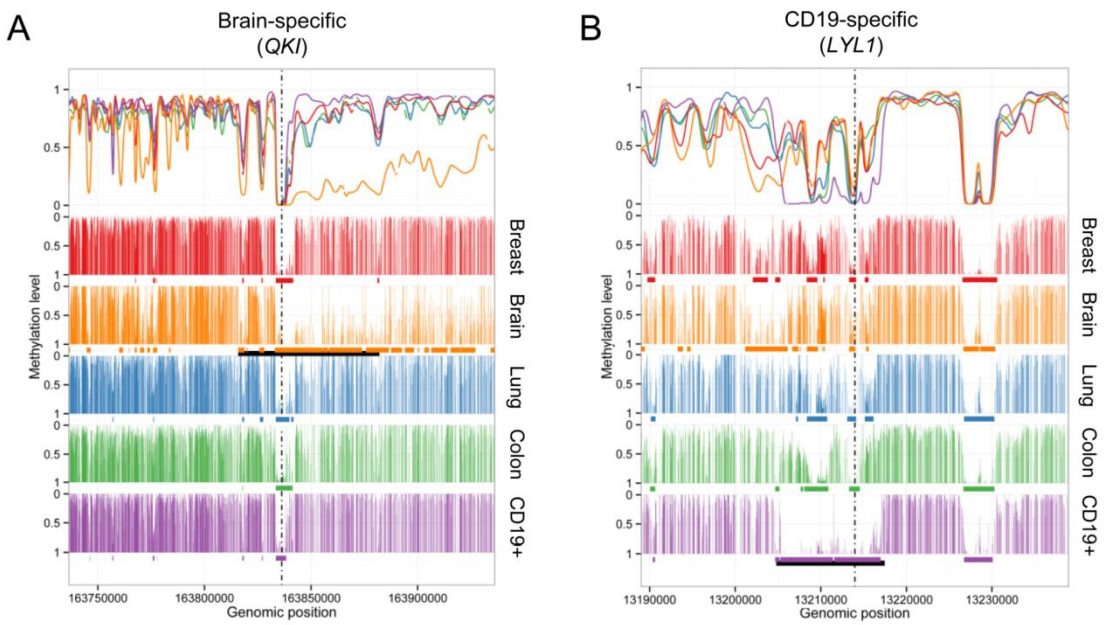


Figure 6.6. DNA methylation profiles of the brain-specific super-enhancer region associated with *QKI* (**a**) and the CD19+ specific region related to *LYL1* (**b**). Smoothed (colored line) and raw (colored bars) CpG methylation levels are indicated for normal breast (red), normal brain (orange), normal lung (blue), normal colon (green) and sorted CD19+ (purple) cells. Hypomethylated regions (HMRs, colored bars) and super-enhancers (black bar) are indicated. The respective transcription start sites are highlighted (broken lines).

super-enhancers, suggesting local variability in their activity. Accordingly and in contrast to previous assumptions, the focal variability of the here studied epigenetic mark supports the action of independent regulatory units and challenges the conjoint activity of enhancer clusters for this subset of super-enhancer regions.

From an epigenetic perspective, the CpG unmethylated status was significantly correlated with H3K27ac occupancy (Spearman's correlation test, rho 0.535, $p < 0.001$; **Fig. 6.2e**) and, to a lesser extent, with H3K4me1 (Spearman's correlation test, rho 0.278, $p < 0.001$), further supporting the former mark as sufficiently bookmarking super-enhancer functionality. This association was independent of the local CpG density, suggesting a sequence-independent connection between the two epigenetic marks (multivariate linear model, $p < 0.001$; **Fig. 6.7**). Most importantly, unmethylated status was significantly associated with increased transcriptional activity of the regulated target genes, indicating that DNA methylation levels at these sequences may be of value as surrogate marks of super-enhancer functionality (Spearman's correlation

test, rho −0.77, $p < 0.001$; **Fig. 6.2f**). Although, functional DNA methylation variance at enhancer sites has been reported previously[637-640], we observed a stronger effect of differential DNA methylation on gene expression levels of super-enhancer-related targets (**Fig. 6.8a**). It is of note that the increased correlation between DNA methylation and gene expression at super-enhancers compared with traditional enhancers was observed for enhancer sites overlapping promoter regions and those distal to the target gene transcription start site (TSS), suggesting an elevated effect of differential super-enhancer DNA methylation independent of the distance to its target (**Fig. 6.8a**). Moreover, DNA methylation levels at super-enhancers overlapping promoters showed significantly higher correlation at regions flanking the proximal (±2 kb of the TSS) promoter (Spearman's correlation test, rho 0.26 versus 0.18), further suggesting that enhancer-specific dynamics drive gene regulation. It is noteworthy that we did not observe a correlation between super-enhancers and target promoter-related CpG island DNA methylation levels (Spearman's correlation test, rho 0.0001, $p = 0.99$), although both genomic features independently correlated significantly with gene expression (Spearman's correlation test, rho 0.31, $p < 0.001$ and rho 0.16, $p < 0.001$, respectively), suggesting an independent function of both regulatory elements. Furthermore, the effect of enhancers on gene expression was closely related to the enhancer size, with DNA methylation levels at super-enhancers presenting the highest correlation with target gene expression compared with smaller sized counterparts (**Fig. 6.8b**).

For *cis*-acting super-enhancers, we observed that the assignment of the closest gene as target resulted in better correlations between super-enhancer DNA methylation and gene expression than a chromatin conformation-based method (ChIA-PET Pol2 in MCF-7 cells, Spearman's correlation test, rho −0.048, $p = 0.4$; **Fig. 6.8c**)[627]. However, both strategies clearly include falsely assigned enhancer–target pairs and more suitable methodologies have yet to be defined.

4.1 Aberrant DNA methylation profiles of super-enhancers in human cancer

Considering the association between DNA methylation status and super-enhancer activity in normal tissues, we wondered whether the observed epigenetic pattern was significantly altered in human cancer. We observed that 14 % (727 out of 5111) of the super-enhancers studied underwent CpG methylation changes in their respective human tumor types, e.g., normal breast versus breast cancer cell lines (**Fig. 6.9a**). The most

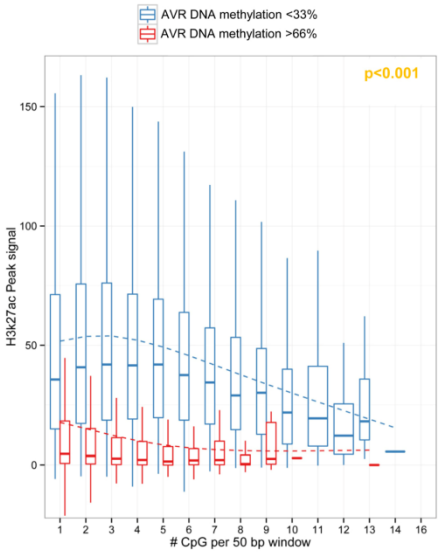


Figure 6.7. Hypomethylation level at the H3K27ac peak was independent of CpG density. CpG density of super-enhancers (x-axis) and the associated H3K27ac peak signal (y-axis) in 50-bp windows. Super-enhancer windows are split into high (>66%, red) and low (<33%, blue) DNA methylation levels. The significant differences of peak signals between high and low DNA methylation levels in super-enhancer windows were assessed with a multivariate linear model, adjusting for CpG density (p<0.001).

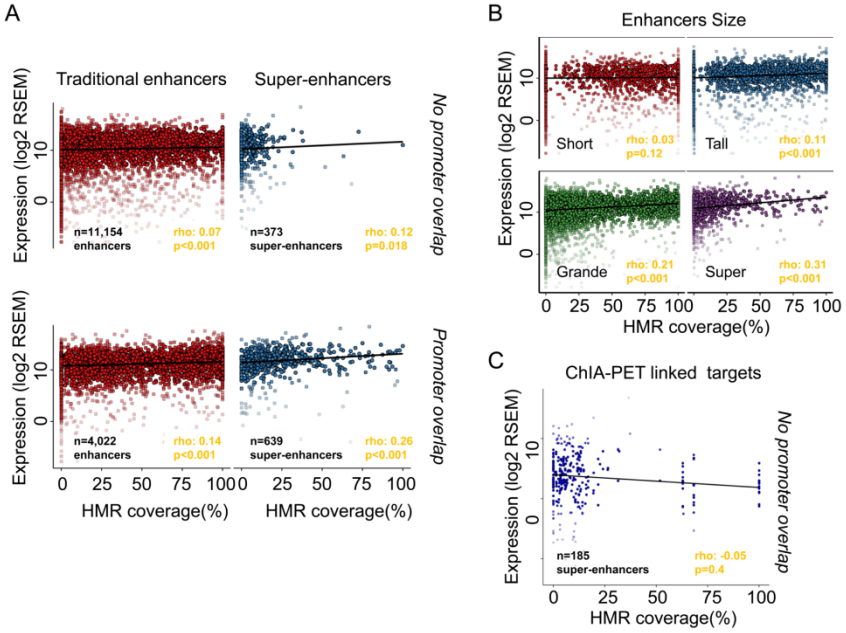


Figure 6.8. Super-enhancer DNA methylation level correlate with gene expression of the nearest gene. (**a**) Correlation analysis of enhancer HMR coverage levels and gene expression of the closest target gene. Enhancers are split in traditional enhancers (H3K27ac, red) and super-enhancer (blue) and *cis*-acting (no promoter overlap, upper panel) and promoter overlapping (lower panel). Expression data was log transformed and significances of a Spearman's correlation test are indicated. (**b**) Correlation analysis of enhancer HMR coverage levels and gene expression considering the enhancer size. Enhancers are split in short (red, 355-1,512 bp), tall (blue, 1,513-2,931 bp), grande (green, 2,932-58,999 bp) and super-enhancers (purple, 2,932-119,310 bp). Expression data was log transformed and significances of a Spearman's correlation test are indicated. (**c**) Correlation analysis of HMR coverage in *cis*-acting (no promoter overlap) super-enhancer regions and the expression of putative target genes defined by ChIA-PET analysis of polymerase II binding in MCF7 cells[627]. Expression data was log transformed and significances of a Spearman's correlation test are indicated.

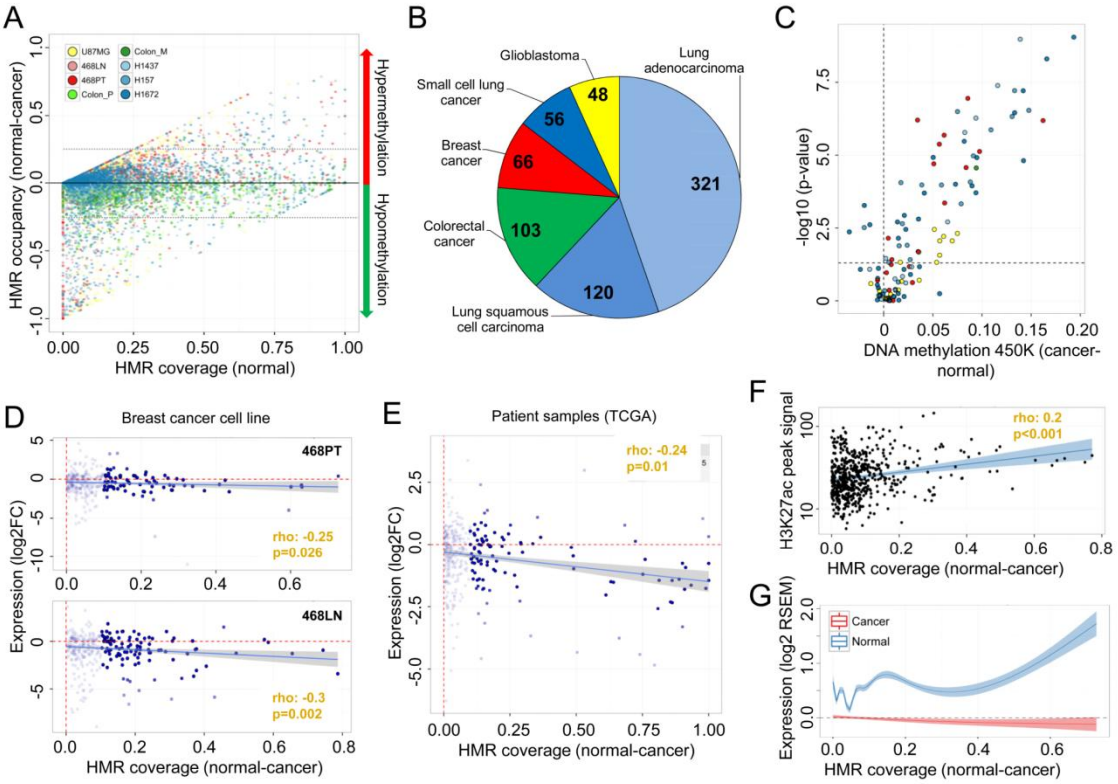


Figure 6.9. Cancer-specific alterations in DNA methylation within super-enhancer regions determined using WGBS. **a** Difference in DNA methylation levels (occupancy of hypomethylated regions (*HMRs*)) between cancer (n = 8) and normal (n = 5) samples paired within their respective tissue contexts (y-axis). HMR occupancy of normal tissues is indicated (x-axis) and cancer sample types are color-coded and the threshold indicated (*dotted line*; δ HMR occupancy 25 %). **b** Sample distribution of 714 cancer samples analyzed on the HumanMethylation450 BeadChip. **c** Validation of DNA hypermethylation at super-enhancers in 714 cancer samples using the HumanMethylation450 BeadChip (450 K). Significance was assessed by differential DNA methylation levels and the Student's *t*-test (*p* value), comparing normal and cancer samples and averaging over the analyzed CpG (≥3) within a super-enhancer region (FDR < 0.05). The cancer samples are color-coded as defined in (**b**). **d** The association between HMR occupancy (WGBS) and target gene expression (RNA-seq) is assessed comparing normal breast (MCF10A) and the primary (468PT, *upper panel*) and metastatic (468LN, *lower panel*) breast cancer cell lines. Expression data are displayed as log transformed fold-change (*log2FC*) and significances of a Spearman's correlation test are indicated. **e** Differences in HMR occupancy (WGBS) and target gene expression (RNA-seq, scaled log expression) are displayed comparing matched normal breast and primary carcinoma samples (TCGA[628], n = 25). **f** Association of H3K27ac signal (ChIP-seq) and differential HMR occupancy (WGBS) at hypermethylated super-enhancers. H3K27ac signals were retrieved from normal breast tissue[615]. **g** Smoothed (GAM) scaled log expression values of super-enhancer-related genes in matched normal and cancer samples (TCGA[628]], n = 25) plotted against the difference in HMR occupancy (WGBS) for all super-enhancers gaining methylation in cancer. *GAM* Generalized Additive Model, *RSEM* RNA-Sequencing by Expectation Maximization.

common DNA methylation shift was the loss of CpG methylation in the cancer sample, which was noted in 75.4 % (548 of 727) of cases, whilst 24.6 % (179 of 727) of super-enhancers gained DNA methylation across the eight tissue-matched cancer samples (δ HMR occupancy >25 %; **Fig. 6.9a**; **Fig. 6.10a** and **Supplementary Tables 6.3 and 6.4**). Interestingly, the hypomethylation events were rather unspecific, as they were associated with the global loss of DNA methylation usually observed in cancer samples (paired *t*-test, $p > 0.05$)[314,612,641], the only notable exception being colorectal tumors, in which they were significantly super-enhancer locus-specific (average flanking regions versus super-enhancer reduction 29.8 % [tumor] and 33.9 % [metastasis], paired *t*-test, $p < 0.001$; **Fig. 6.10b** and **Supplementary Table 6.4**). Thus, to determine functional epigenetic alterations, we decided to initially focus on the hypermethylated events, which were enriched in genes associated with transcriptional and metabolic processes and angiogenesis (FDR < 0.01; **Supplementary Table 6.5**). Importantly, hypermethylation events were also replicated using DNA methylation microarray analyses in a unique cohort of 714 primary cancer samples (**Table 6.2** and **Fig. 6.9b**), where 58.1 % (68 of 117) of the interrogated DNA hypermethylation events at super-enhancers were confirmed (Student's *t*-test, FDR < 0.05; **Fig. 6.9c**; **Supplementary Table 6.6**). These results further suggest that the hypermethylation events observed in the cancer cell line models are mirroring altered DNA methylation profiles at super-enhancer regions in primary tumors. Hypermethylated super-enhancers in cancer included genes previously related to cellular transformation (e.g., *CIC*, *FOXA2*, *FOXP1*, *RUNX1* and *TBX3*)[642]. Importantly, we excluded that copy number variations (CNVs) have confounded our analysis of the primary cancer samples by detecting significant differences in DNA methylation levels between normal and CNV samples in only a very minor fraction of the super-enhancers (4.3 %, 5/117; Student's *t*-test, FDR < 0.05; **Supplementary Table 6.6**).

It is of note that, using oxidative bisulfite (ox-BS) treatment coupled with DNA methylation microarray analyses, we could exclude the gain of DNA methylation observed in cancer to be due to an increase of 5-hydroxy methylation (5-hmC), a specific cytosine modification that confounds with 5-methylation (5-mC) in bisulfite (BS)-based analyses and found to be enriched in traditional enhancer regions (**Fig. 6.11**)[643]. In order to test a significant contribution of the 5-hmC to the methylation gain in super-enhancers, we compared the methylation values obtained from BS-treated

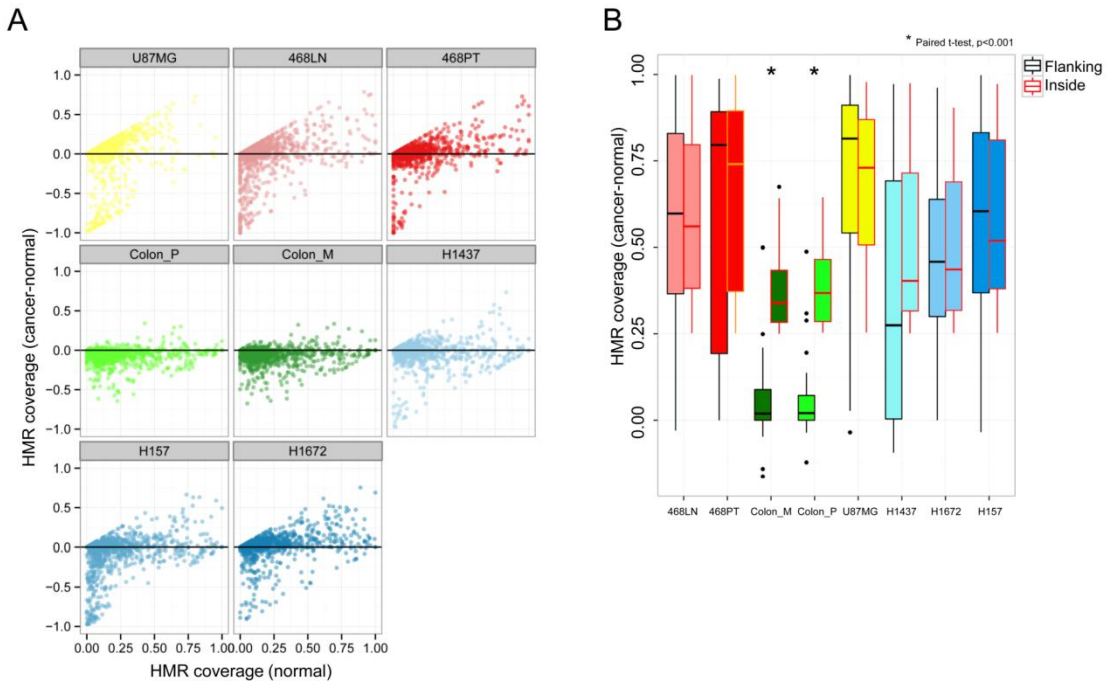


Figure 6.10. Cancer-specific alterations in DNA methylation within super-enhancer regions determined using WGBS. (**a**) Super-enhancers of normal tissue gaining DNA methylation in a cancer context. Difference in DNA methylation levels (coverage of hypomethylated regions, HMRs) between cancer (n=8) and normal (n=5) samples paired within their respective tissue context (y axis). HMR coverage of normal tissues is indicated (x axis) and one plot per normal–tumor pair is displayed. (**b**) DNA methylation levels (coverage of hypomethylated regions, HMRs) of hypomethylated super-enhancers and the indicated cancer samples (δ HMR occupancy >25%) inside (red) and flanking (black) the super-enhancer regions.

against ox-BS-treated cancer samples, enabling us to estimate the 5-hmC levels[644]. With the alternative hypothesis being that the ox-BS values were greater than 0, we did not observe a significant presence of 5-hmC in any cancer sample (paired one-tailed Wilcoxon test).

To further elucidate the functional consequences associated with the identified cancer-specific super-enhancer DNA methylation shifts, we investigated the impact of the tumor-associated gains of super-enhancer DNA methylation on gene expression. We first used a breast cancer model that included the paired breast cancer cell lines MDA-MB-468PT (derived from the primary tumor) and MDA-MB-468LN (derived from a lymph node metastasis) and the untransformed immortalized breast epithelial cell line MCF10A, associating differential gene expression (RNA sequencing, RNA-seq) with super-enhancer DNA methylation levels. As has been observed for the proximal

regulatory gene regions, where a general repressive effect of DNA methylation is widely recognized[645], we found an association between DNA methylation gain in breast super-enhancer regions and gene repression of the associated genes for both MDA-MB-468PT (Spearman's correlation test, rho −0.25, p = 0.026) and MDA-MB-468LN (Spearman's correlation test, rho −0.3, p = 0.002; **Fig. 6.9d**) cell lines.

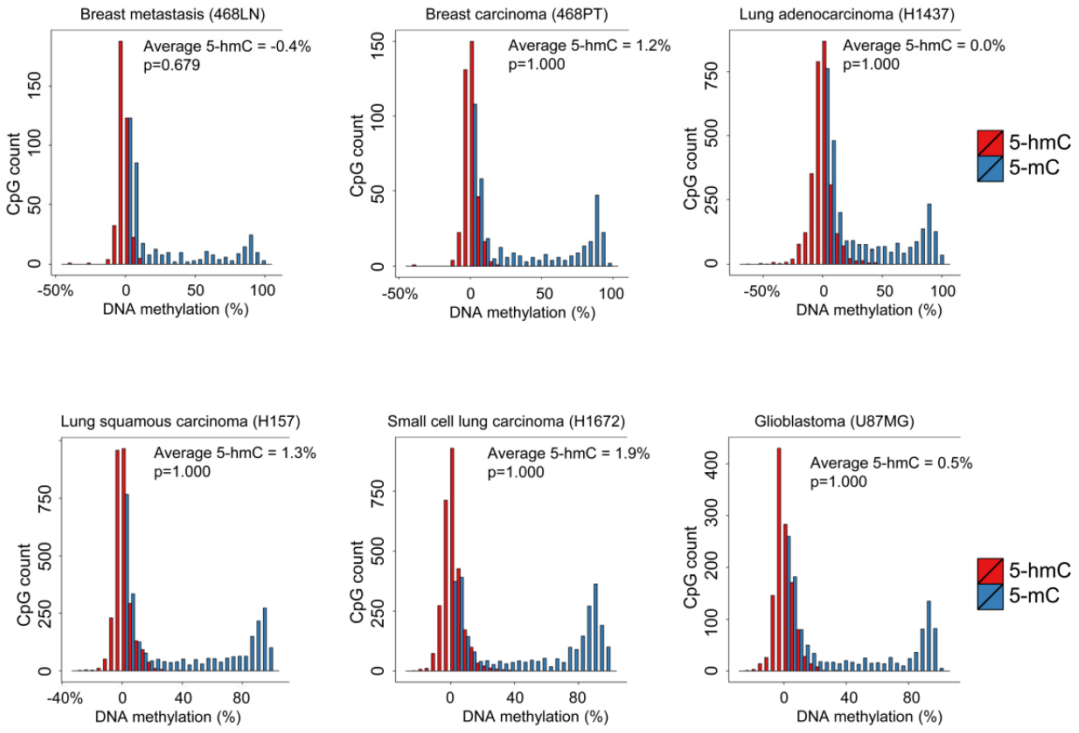


Figure 6.11. Absence of 5-hydroxy CpG methylation (5-hmC) at hypermethylated super-enhancer loci in cancer samples. We compared the methylation values obtained from bisulfite treated cancer samples against oxidative bisulfite treated samples to assess 5-hmC levels and to distinguish between 5-hmC (red) and 5-mC (blue) levels. Significant contribution of the 5-hmC was excluded for CpG sites (450K probes) in hypermethylated super-enhancer regions within their respective cancer samples (Supplementary Table 6.4) comparing 5-hmC and 5-mC levels (paired one tailed Wilcoxon test).

We extended these observations to primary breast tumors from the TCGA[628], whose expression patterns have also been determined by RNA-seq. We confirmed the significant association between the DNA methylation gains of super-enhancers identified in our breast cancer cell line data set and gene repression observed in the matched TCGA breast cancer samples (Spearman's correlation test, rho −0.24, $p = 0.01$; **Fig. 6.9e**). Interestingly, the super-enhancers that became hypermethylated in breast

cancer were those that, in normal breast epithelial cells, were the most enriched in the H3K27ac histone mark (Spearman's correlation test, rho 0.2, $p < 0.001$; **Fig. 6.9f**), which defines these particular distal regulatory regions[515,517,615], and the H3K4me1 enhancer mark (Spearman's correlation test, rho 0.2, $p < 0.001$). Remarkably, the most hypermethylated super-enhancers had also the highest level of expression for the respective associated genes in normal breast epithelial cells (linear slope 1.23, $p < 0.001$; **Fig. 6.9g**).

We were able to validate the link between cancer-specific super-enhancer hypermethylation and the transcriptional inactivation of the corresponding genes beyond the breast tumor type. In the lung tumorigenesis samples from the H1437 (lung adenocarcinoma) and H157 (lung squamous cell carcinoma) cancer cell lines, we found evidence that lung super-enhancer gain of DNA methylation was associated with the downregulation of the target genes (linear slope −3.06, $p < 0.001$ and −2.09, $p = 0.004$, respectively; **Fig. 6.12a,b**) determined by publically available expression microarrays[646]. We also extended these findings to primary lung adenocarcinoma and lung squamous cell carcinoma tumors from the TCGA[631], in which expression of the candidate genes originates from RNA-seq experiments. In this setting, we observed a significant association between lung super-enhancer hypermethylation identified in our lung cancer cell lines and gene downregulation found in the matched primary lung cancer samples (Spearman's correlation test, rho −0.19, $p = 0.012$ and rho −0.25, $p < 0.001$, respectively; **Fig. 6.12c,d**). The significant association between cancer-specific DNA methylation of super-enhancers and gene repression was also noted in the glioblastoma cell line U87MG (Spearman correlation test, rho −0.26, $p < 0.001$; **Fig. 6.12e**), in which we performed an expression microarray experiment. Thus, the results overall suggest that a tumor-related gain of DNA methylation in super-enhancers has a transcriptionally repressive effect on the corresponding related genes.

We next considered the commonality among different tumor types within super-enhancer DNA methylation changes, and the type of genes and pathways affected by these aberrant epigenetic shifts. We first observed that within regions of commonly hypomethylated super-enhancers in normal contexts, the cancer samples (**Table 6.2**) clustered by tumor type (**Fig. 6.13a**), a phenomenon we previously identified for DNA methylation events in proximal promoters among distinct human tumors[647].

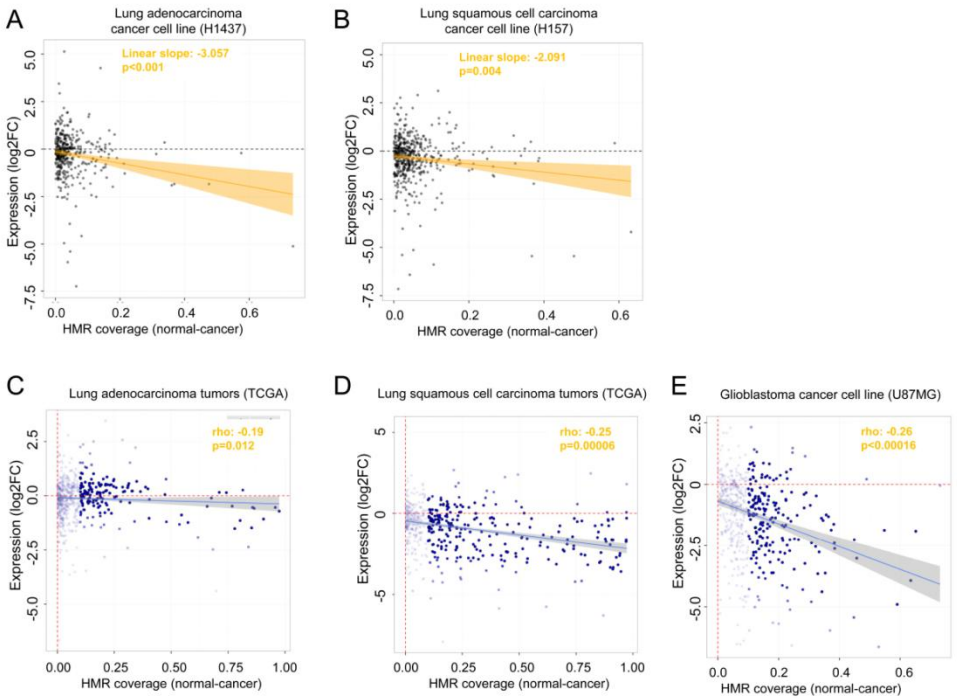


Figure 6.12. Association between DNA methylation levels of hypermethylated super-enhancers and gene expression of target genes. (**a,b**) Differential HMR coverage (x-axis, WGBS) and target gene expression (y-axis) are displayed, comparing normal lung samples and the primary lung adenocarcinoma (H1437, **a**) and lung squamous cell carcinoma (H157, **b**) cancer cell lines. Significance of the linear regression model is indicated. (**c,d**) Associations of hypermethylated super-enhancers and target gene expression using primary lung adenocarcinoma (**c**) and lung squamous cell carcinoma. (**d**) samples (TCGA). Differential HMR coverage (x-axis, WGBS) and target gene expression (y-axis, RNAseq, scaled log expression) are shown, comparing matched normal lung and primary carcinoma samples. Significances of Spearman's correlation test are indicated. (**e**) Differential HMR coverage (x-axis, WGBS) and target gene expression (y-axis) are displayed, comparing normal brain (white matter) samples and the glioblastoma cell line. Significance of a Spearman's correlation test is indicated.

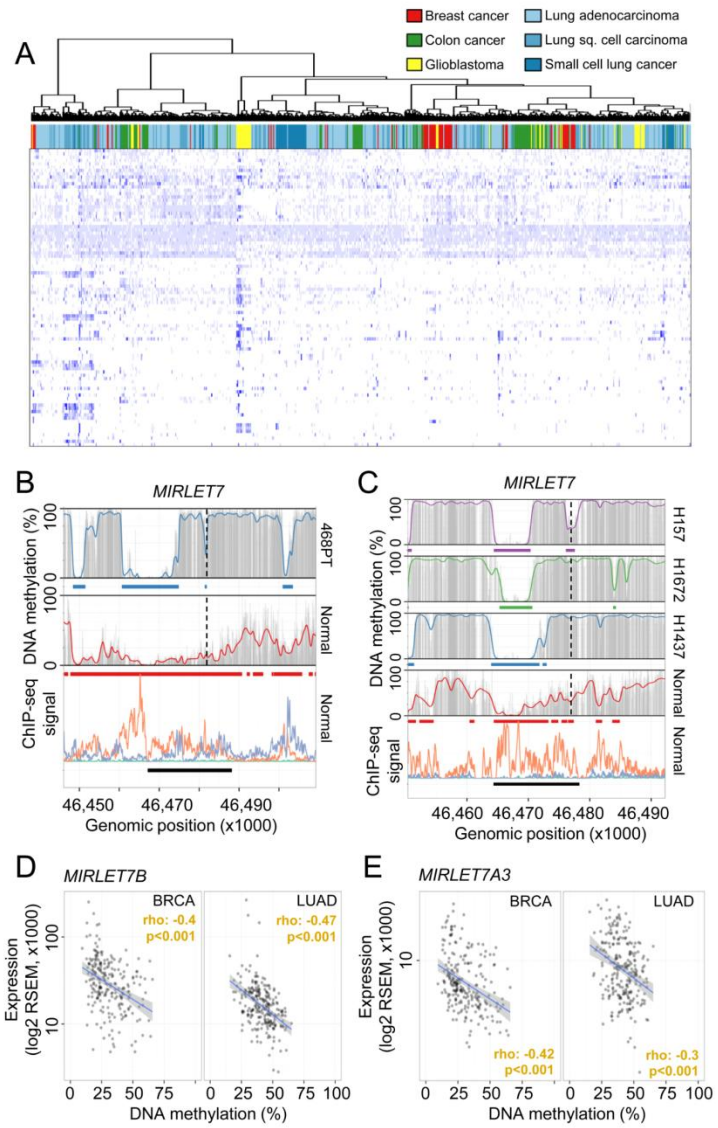


Figure 6.13. Cancer type-specific alterations of DNA methylation signatures at super-enhancer loci. **a** Hierarchical clustering of common hypomethylated super-enhancer regions in normal tissues (rows, <25 % average DNA methylation) in 714 cancer samples (columns). Average CpG methylation levels in common regions were clustered using Canberra distances and the Ward cluster method. DNA methylation levels are color-coded from 0 % (*light blue*) to 100 % (*dark blue*) and the different cancer types are color-coded. **b**, **c** DNA methylation profiles of the super-enhancer regions associated with *MIRLET7* in normal tissues and cell lines derived from breast (**b**) and lung cancer (**c**). Smoothed (*colored line*), raw (*gray bars*) CpG methylation levels, hypomethylated regions (*colored bars*) and super-enhancers (*black bars*) are indicated. The enhancer-related histone marks (*bottom panel*) H3K27ac (*orange*) and H3K4me1 (*purple*) are displayed as ChIP-seq signal intensities[615]. Transcription start sites are indicated (*broken line*). **d**, **e** Association of DNA methylation levels (TCGA, HumanMethylation450 BeadChip, averaged probe levels within the super-enhancer) and gene expression (TCGA, RNA-seq, absolute expression values) related to the *MIRLET7* super-enhancer and targeted microRNAs *MIRLET7B* (**d**) and *MIRLET7A3* (**e**) in breast (n = 201) and lung (n = 216) cancer samples. Significances of a Spearman's correlation test are indicated. *RSEM* RNA-Sequencing by Expectation Maximization.

Interestingly, despite the clear presence of super-enhancer DNA methylation that is associated with the cancer type, there are hypermethylated super-enhancers shared by common epithelial tumors such as the breast and lung samples (**Fig. 6.14a**). This is the case for the super-enhancer of the tumor suppressor microRNA *MIRLET7*, where hypomethylation of the super-enhancer was diminished by a gain of CpG methylation in a fraction of the regulatory region (**Fig. 6.13b,c**; **Fig. 6.14b,c**). It is of note that the large highly hypomethylated super-enhancer regions displayed focal gains in DNA methylation in cancer, suggesting that distinct segments might exhibit specific functions in healthy and cancer contexts. Consistent with the suspected regulatory function, hypermethylation of the *MIRLET7*-associated super-enhancer region was associated with transcriptional silencing of *MIRLET7B* and *MIRLET7A3*, two family members coded within the affected pri-microRNA (**Fig. 6.14d**). Moreover, microRNAs *MIRLET7B* and *MIRLET7A3* were repressed in primary breast carcinomas (TCGA[628]; Wilcoxon test, $p = 0.001$ and $p = 0.033$, respectively) and lung adenocarcinomas (TCGA[631]; Wilcoxon test, $p < 0.001$ and $p < 0.001$, respectively) (**Fig. 6.14e,f**) and hypermethylation at super-enhancers was significantly correlated with microRNA repression in breast carcinomas (Spearman correlation test, rho −0.4 and −0.42, $p < 0.001$ and $p < 0.001$, respectively) and lung adenocarcinomas (Spearman correlation test, rho −0.47 and − 0.3, $p < 0.001$ and $p < 0.001$, respectively) (**Fig. 6.13d,e**).

4.2 Cancer-specific super-enhancers coincide with regional hypomethylation

Until now we have focused our attention on those sequences described as being super-enhancers that ensure cell and tissue identity in normal tissues[515,615]. However, a new class of super-enhancer sequences has recently been described that only play this *de novo* regulatory role in transformed cells to drive the cancer phenotype and its associated hallmarks[43,517,615]. We examined the DNA methylation changes occurring in the super-enhancers of colorectal cancer (HCT-116, $n = 387$), in which we obtained 99 % coverage using our WGBS approach. We observed that these newly developed tumor-related super-enhancers were associated with DNA hypomethylation events ($n = 23$, δ HMR occupancy >25 %) at these sequences in the transformed cells compared with normal colorectal mucosa (**Fig. 6.15a**; **Table 6.4**). Most notably, the super-enhancer hypomethylation shift was independent of the global loss of DNA methylation generally found in cancer cells (paired *t*-test, $p < 0.001$)[314,612,641] and rather

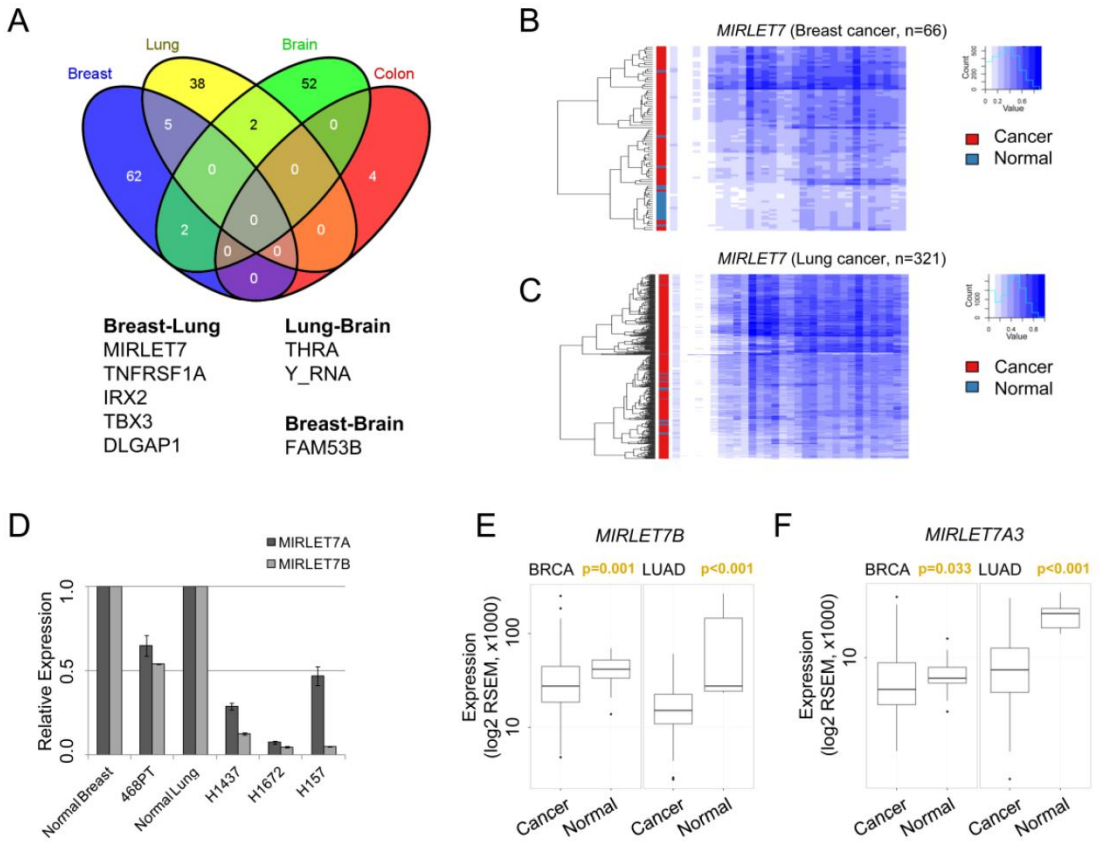


Figure 6.14. (**a**) Recurrently affected genes associated with hypermethylated super-enhancers in cancer tissue. Target genes of different samples within tissue types were merged with respect to the origin of the tumor. (**b,c**) Hierarchical clustering of CpG methylation levels (HumanMethylation450 BeadChip) within the super-enhancer regions associated with breast (**b**) and lung (**c**) tissue in 66 primary breast and 321 lung adenocarcinoma samples. CpG methylation levels were clustered using Euclidian distances and the Complete cluster method. DNA methylation levels are color-coded from 0% (white) to 100% (dark blue) and normal (blue) and cancer (red) samples are color-coded. (**d**) Repression of *MIRLET7B* and *MIRLET7A3* in breast and lung cancer cell lines (indicated). Expression levels were determined by quantitative real-time PCR and are displayed relative to the respective normal controls. (**e,f**) Repression of *MIRLET7B* and *MIRLET7A3* in primary in (**e**) breast (TCGA[628]) and (**f**) lung (TCGA[631]) cancer samples. Significances of a Wilcoxon test are indicated.

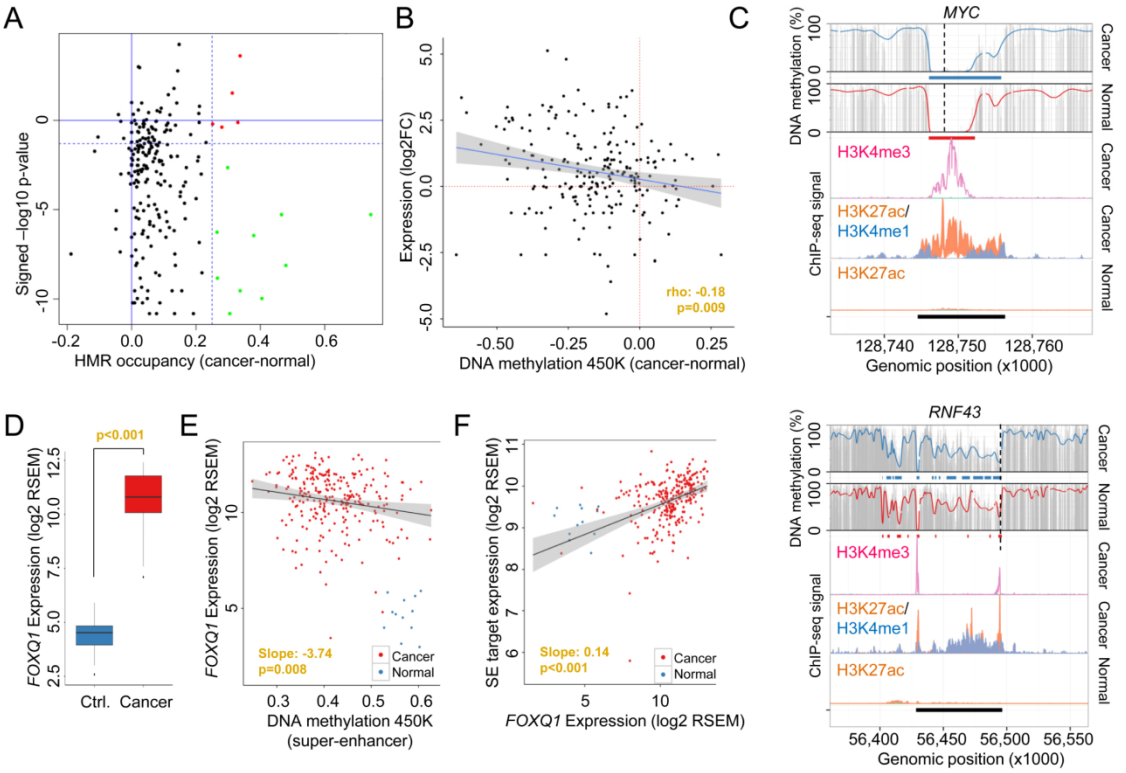


Figure 6.15. Hypomethylation at cancer-related super-enhancers in colorectal tumors. **a** Differential DNA methylation (occupancy of hypomethylated regions (*HMRs*)) at colorectal cancer-related super-enhancers between normal mucosa and primary colorectal cancer samples (WGBS, x-axis). Differentially methylated super-enhancers are indicated (*colored dots*, δ HMR occupancy >25 %). Results were validated in a cohort of matched normal and primary colorectal tumor samples (TCGA, n = 41, HumanMethylation450 BeadChip) and significant differences assessed by the Wilcoxon test (*green dots*, $p < 0.05$, y-axis). **b** Hypomethylation at super-enhancers was associated with increased target gene expression analyzed by HumanMethylation450 BeadChip (450 K, x-axis) and RNA-seq (y-axis) in matched primary colorectal cancer samples (n = 12, TCGA). Expression data are displayed as log transformed fold-change (*log2FC*). **c** DNA methylation profiles of the super-enhancer regions associated with *MYC* and *RNF43* in normal and colorectal cancer samples (WGBS). Smoothed (*colored line*), raw (*gray bars*) CpG methylation levels, hypomethylated regions (*colored bars*) and super-enhancers (*black bars*) are indicated. The enhancer-related histone marks H3K27ac (*orange*) and H3K4me1 (*blue*) and the promoter-related mark H3K4me3 (*pink*) are displayed as ChIP-seq signal intensities (*bottom panels*)[615]. The transcription start sites are indicated (*broken line*). **d** Gene expression levels of the transcription factor *FOXQ1* in normal (*blue*) and colorectal cancer (*red*) samples (TCGA). **e, f** Association of *FOXQ1* expression and DNA methylation levels (HumanMethylation450 BeadChip, 450 K) at hypomethylated super-enhancer regions (**e**) or expression levels of associated target genes (**f**) in colorectal cancer in normal (*blue*) and colorectal cancer (*red*) samples (TCGA). Significance was assessed from a linear regression model applied solely to the cancer samples. *RSEM* RNA-Sequencing by Expectation Maximization.

Table 6.4. Hypomethylated cancer-related super-enhancers in colorectal cancer cells.

Enhancer					% DNA hypomethylation			450K (AVR)			CNV	CNV	#TFBS
Chr.	Start	End	Enhancer ID (Hnisz et al., 2013)	Gene symbol	Normal	Tumor	Delta	Normal	Tumor	-log10	FDR	WGBS	FOXQ1
1	1079431	1101911	7_MACS_peak_34_lociStitched	MIR200b	0.21	0.51	-0.31	0.56	0.36	-10.82	NA	loss	18
1	1365609	1378027	3_MACS_peak_73_lociStitched	VWA1	0.28	0.75	-0.47	NA	NA	NA	NA	loss	0
1	45270646	45276647	MACS_peak_1200	PLK3	0.18	0.52	-0.34	0.22	0.34	3.63	NA	NA	7
2	70310047	70317059	2_MACS_peak_15169_lociStitched	PCBP1	0.54	0.95	-0.40	0.54	0.40	-9.97	NA	NA	0
3	49934498	49944642	2_MACS_peak_18795_lociStitched	MST1R	0.09	0.57	-0.48	0.40	0.22	-8.11	0.79	NA	7
7	130563226	130576863	MACS_peak_24364	MIR29a/b	0.00	0.27	-0.27	0.74	0.36	-8.83	NA	NA	7
7	130577071	130604705	MACS_peak_24365	MIR29a/b	0.21	0.95	-0.74	0.55	0.31	-5.29	0.71	NA	11
8	128744519	128756323	2_MACS_peak_25390_lociStitched	MYC	0.53	0.83	-0.30	0.46	0.31	-2.65	0.71	NA	5
10	101664881	101695667	4_MACS_peak_3541_lociStitched	DNMBP	0.02	0.30	-0.28	NA	NA	NA	0.71	NA	22
11	309864	321533	2_MACS_peak_3964_lociStitched	IFITM3	0.69	0.94	-0.25	0.40	0.39	-0.20	NA	NA	2
12	7069196	7075991	MACS_peak_5450	MIR200c	0.39	0.73	-0.34	0.68	0.31	-9.54	NA	NA	4
12	13022807	13063891	9_MACS_peak_5515_lociStitched	GPRC5A	0.05	0.38	-0.33	0.44	0.44	-0.11	NA	NA	35
14	69245317	69265057	3_MACS_peak_7661_lociStitched	ZFP36L1	0.47	0.80	-0.33	NA	NA	NA	0.64	NA	9
14	105552749	105561461	1_MACS_peak_8032_lociStitched	GPR123	0.53	0.81	-0.28	NA	NA	NA	NA	loss	4
16	1137094	1142491	MACS_peak_8969	C1QTNF8	0.41	0.71	-0.30	NA	NA	NA	NA	loss	6
16	89622859	89633240	MACS_peak_10212	SNORD68	0.39	0.67	-0.28	0.51	0.51	-0.38	NA	NA	6
17	56428488	56496439	10_MACS_peak_11452_lociStitched	RNF43	0.11	0.58	-0.47	0.52	0.44	-5.29	NA	NA	51
17	75275055	75285678	1_MACS_peak_11905_lociStitched	SEPT9	0.27	0.73	-0.46	NA	NA	NA	NA	NA	5
17	79301530	79306049	MACS_peak_12086	TMEM105	0.24	0.56	-0.31	0.52	0.60	1.54	NA	loss	0
20	62305260	62334757	6_MACS_peak_17235_lociStitched	ARFRP1	0.11	0.37	-0.27	0.59	0.19	-6.28	0.71	loss	15
22	38138277	38150317	MACS_peak_18004	TRIOBP	0.20	0.45	-0.25	NA	NA	NA	0.71	NA	8
22	38176901	38182015	MACS_peak_18012	PDCL3	0.22	0.60	-0.38	0.64	0.31	-6.46	NA	NA	3
22	46464429	46478530	MACS_peak_18258	MIRLET7	0.53	0.99	-0.46	NA	NA	NA	NA	NA	3

represented a focal DNA demethylation event within the super-enhancer regions (**Fig. 6.16**). As we did with the aforementioned normal tissue super-enhancers, we validated the DNA hypomethylation changes in these *de novo* cancer super-enhancers using a cohort of matched normal colon and primary colorectal tumors (TCGA[624], $n = 41$) analyzed by DNA methylation microarrays (**Fig. 6.15a**; **Table 6.4**). Noteworthy, we again excluded potential biases included by CNV in these regions (**Table 6.4**). In this setting, we further confirmed that the loss of DNA methylation in these emerging cancer super-enhancers was significantly associated with an increase in expression of the corresponding regulated genes in the primary colon tumors in comparison with the matched normal colon mucosa (TCGA[624]; Spearman's correlation test, rho -0.18, $p = 0.009$; **Fig. 6.15b**). Examples within the most hypomethylated cancer super-enhancers include those sequences regulating the *MYC* and *RNF43*[648] oncogenes (**Fig. 6.15c**; **Fig. 6.17a,b**), regions not affected by CNV in the primary colorectal cancer sample analyzed by WGBS (**Table 6.4**). Importantly, DNA methylation changes affected solely regions specifically marked by H3K27ac in colon cancer and widely excluded H3K4me3, further indicating that alterations in super-enhancers occur predominantly distal to the core promoter regions (**Fig. 6.15c**).

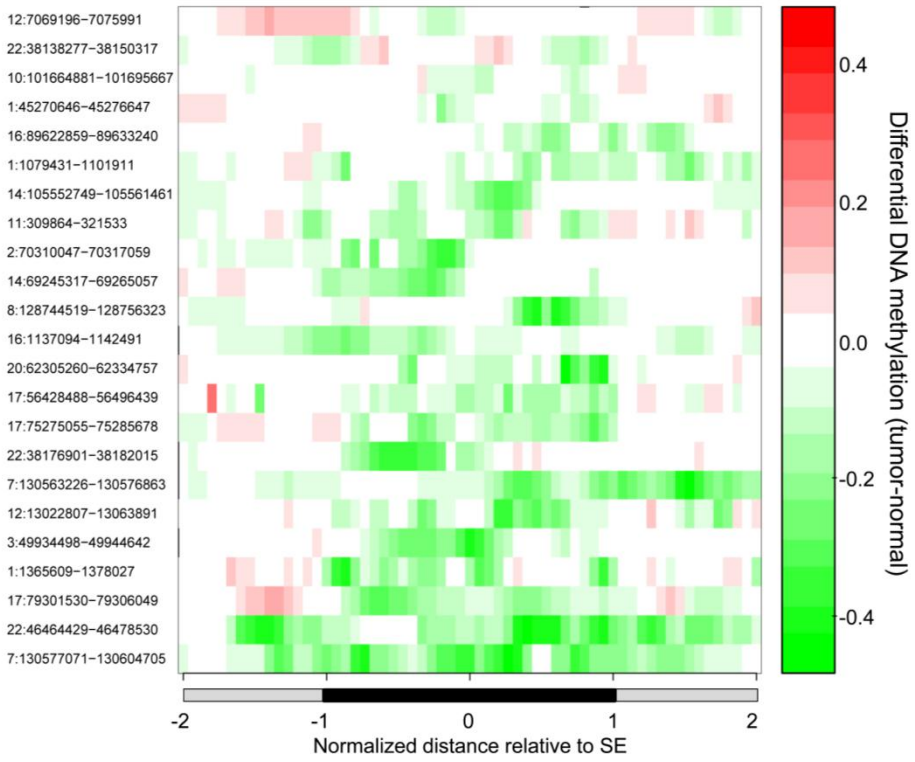


Figure 6.16. Focal DNA demethylation events within colorectal super-enhancer regions. Scaled 23 hypomethylated super-enhancers and flanking regions in equally spaced windows. Differential average CpG methylation levels (primary tumor - normal sample) are displayed for the respective windows with colors indicating hypo- (green) and hypermethylation (red).

An interesting matter arising from these results is their value for identifying putative mechanisms that create such specific patterns of oncogenic super-enhancer hypomethylation. It has been proposed that the availability and binding of transcription factors (TFs) to regulatory regions might be able to impact on the DNA methylome and that it is not the transcriptional activity per se that alters the DNA methylation profile of regulatory elements[633,634]. Herein, we have studied the putative enrichment of TF binding sites in these colorectal cancer-specific hypomethylated enhancers and we observed a significant enrichment for specific TF binding motifs (**Fig. 6.18a**). From these factors, specifically FOXQ1 (forkhead box Q1; $p = 0.013$), a member of the FOX gene family that is involved in tumorigenesis[649], was the most overexpressed TF in primary colorectal cancer samples and showed multiple binding sites (**Table 6.4**) and a significant enrichment at hypomethylated super-enhancer loci (**Fig. 6.18b**). In relation to this point, *FOXQ1* had a 73-fold greater expression in primary colorectal cancer

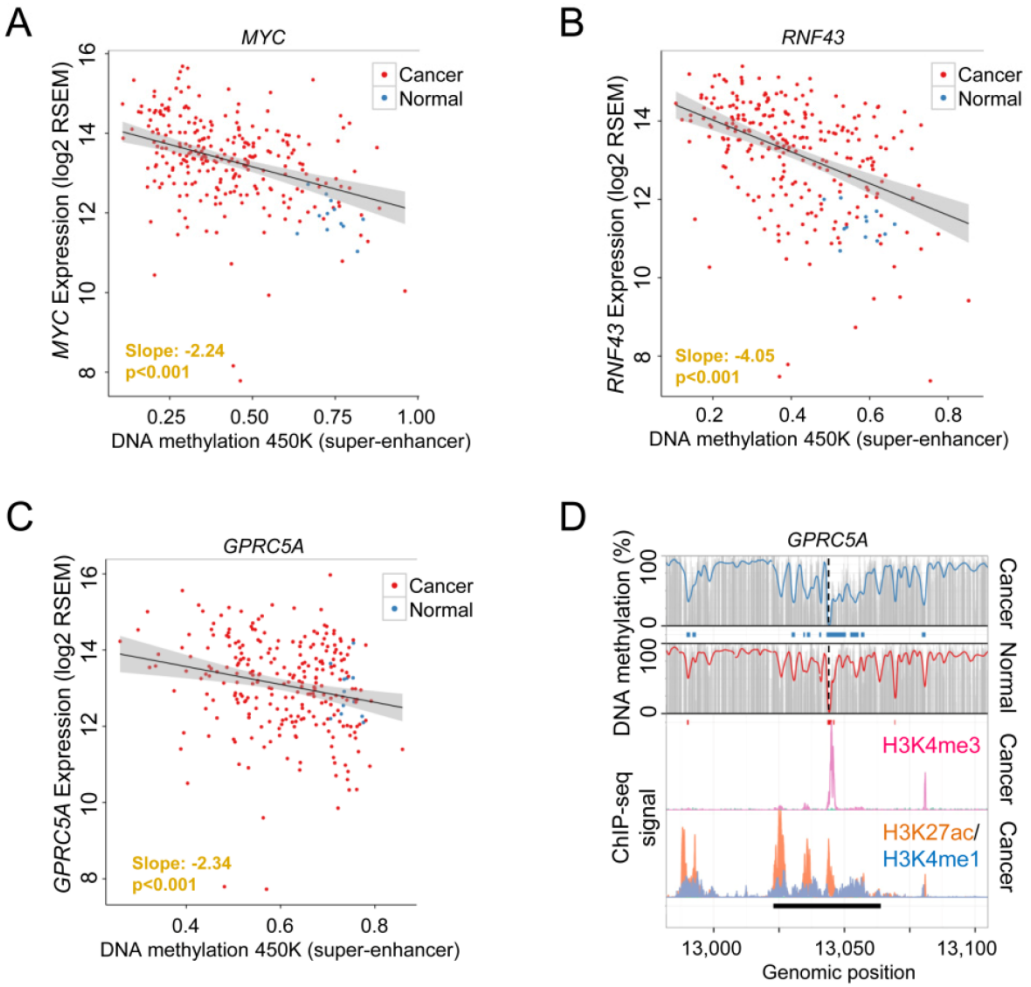


Figure 6.17. (**a-c**) Association between transcriptional activity (y-axis) of *MYC* (**a**), *RNF43* (**b**) and *GPRC5A* (**c**) and hypomethylation (x-axis, HumanMethylation450 BeadChip, 450K) of their related super-enhancer in normal (blue, n=12) and colorectal cancer (red, n=258) samples (TCGA[624]). Significance was assessed using a linear regression model applied solely to the cancer samples. (**d**) DNA methylation profiles of the super-enhancer region associated with *GPRC5A* in normal mucosa (red) and the colorectal cancer sample (blue). Smoothed (colored line), raw (grey bars) CpG methylation levels, hypomethylated regions (colored bars) and the super-enhancer (black bars) are indicated. The enhancer-related histone marks H3K27ac (orange) and H3K4me1 (blue) and the promoter-related mark H3K4me3 (pink) are displayed as ChIP-seq signal intensities (bottom panels)[615]. The transcription start site is indicated (broken line).

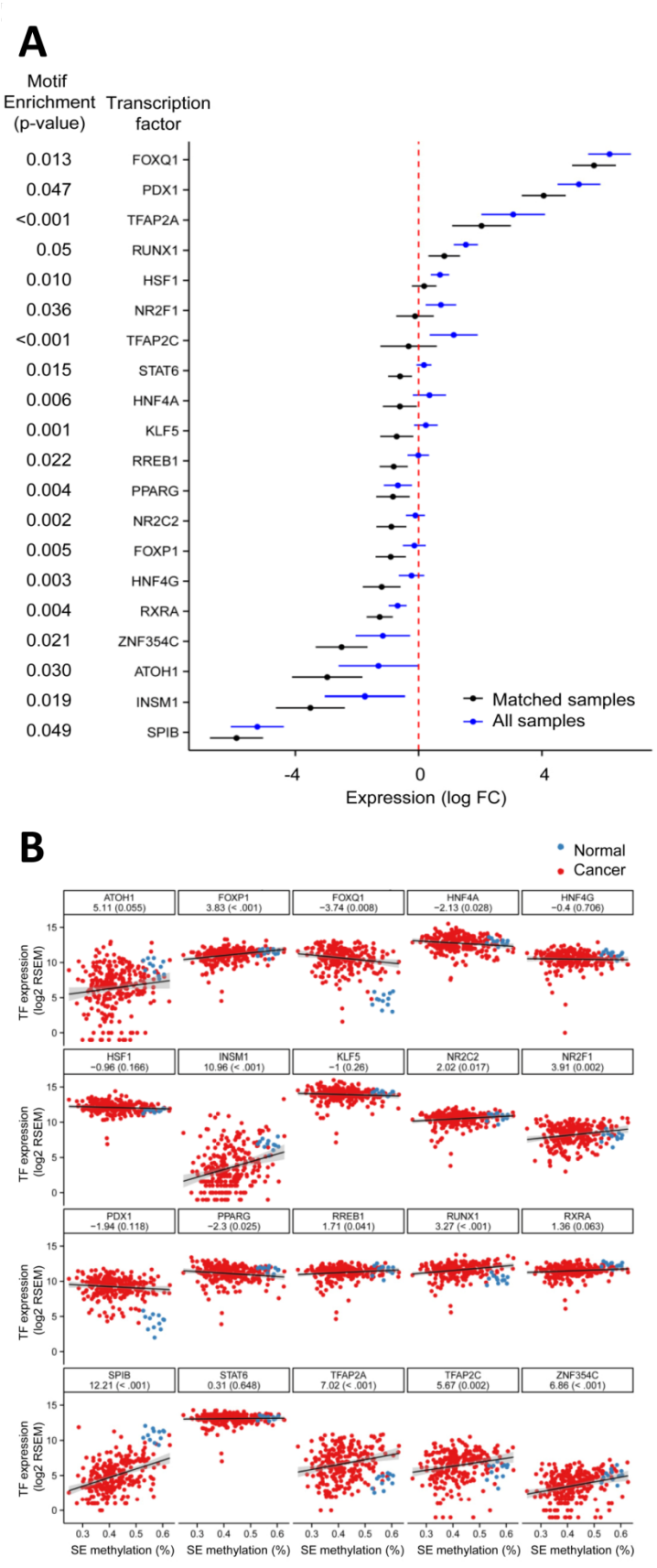


Figure 6.18. *FOXQ1* is overexpressed in primary colorectal tumor samples and correlates with the loss of DNA methylation in super-enhancers. (**a**) Gene expression levels of transcription factors with binding motifs significantly enriched ($p \leq 0.05$) at hypomethylated super-enhancers in the primary colorectal cancer sample (Supplementary Table 6.5). Displayed are log fold-change (FC) expression levels comparing matched (black) unmatched (blue) primary colorectal cancer samples (COAD, TCGA). (**b**) Correlation analysis between DNA methylation levels in hypomethylated super-enhancer regions (Supplementary Table 6.4) and gene expression of enriched transcription factors. Significant associations were assessed using a linear regression model on the cancer samples (red). Slopes and p-values are displayed blow the annotation of the tested transcription factors. Matched DNA methylation was determined by averaging over the DNA methylation array probes (TCGA) falling into the hypomethylated regions of the super-enhancers.

samples than in matched control samples (TCGA624; Wilcoxon test, $p < 0.001$; **Fig. 6.15d**). Furthermore, the stronger *FOXQ1* expression was significantly associated with hypomethylation of the previously defined super-enhancers (linear slope −3.74, $p = 0.008$; **Fig. 6.15e**) and the activation of associated target genes (linear slope 0.14, $p < 0.001$; **Fig. 6.15f**), such as the well-known oncogenes *MYC* and *RNF43* (**Fig. 6.19a,b** and **Fig. 6.20a,b**). Interestingly, the presence of cancer-specific super-enhancer hypomethylation and the tumorigenic effect mediated by the presence of FOXQ1 binding sites could be useful for identifying new candidate oncogenes, such as *GPRC5A* (*G protein-coupled receptor, class C, group 5, member A*; **Fig. 6.17c,d**, **Fig. 6.19c** and **Fig. 6.20c**), which, by mediating between retinoid acid and G protein signaling pathways, has a role in epithelial cell differentiation[650].

Importantly, we experimentally validated the association between *FOXQ1* expression and target gene regulation in a colorectal cancer cell line model system (HCT116 and SW1116 cancer cell lines). Initially, we confirmed the occupancy of FOXQ1 at binding sites within the super-enhancer regions of the previous described target genes *MYC*, *RNF43* and *GPRC5A* (**Fig. 6.21a**). Furthermore, following small hairpin RNA (shRNA)-mediated knockdown of the TF, we observed significant downregulation of *MYC*, *RNF43* and *GPRC5A*, suggesting a direct regulatory role of FOXQ1 (**Fig. 6.21b**). In line with the oncogenic role of FOXQ1 targets in colorectal cancer settings, knockdown of the TF reduced cell proliferation of the colorectal cancer cell line (**Fig. 6.21c**). Remarkably, in addition to FOXQ1, we could also experimentally confirm the regulatory effect of other enriched TFs, whose expression correlated significantly with super-enhancer hypomethylation level ($p < 0.05$; **Fig. 6.18b**). Specifically, we experimentally confirmed the regulatory effect of the TFs *HNF4A* and *PPARG* on *RNF43* and *GPRC5A* expression (**Fig. 6.22a,b**). Herein, knockdown of the TFs repressed *RNF43* and *GPRC5A* expression (**Fig. 6.22c**) and resulted in reduced cell viability (**Fig. 6.22d**), further supporting the accuracy of the functional prediction based on super-enhancer DNA methylation levels (**Fig. 6.18b**).

Further, we were interested if disruption of the super-enhancer structure would interfere with the DNA methylation levels in the respective regions. Therefore, we treated the colorectal cancer cell lines HCT116 and SW1116 at sub-lethal concentrations with the BET-bromodomain inhibitor JQ1, a small molecule targeting BRD4, a key component

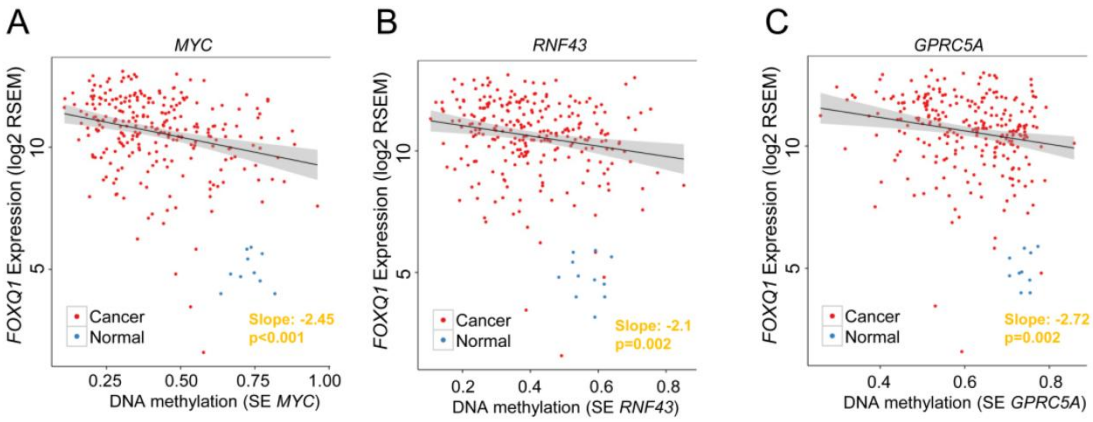


Figure 6.19. Association of *FOXQ1* expression levels and DNA methylation levels at previously defined hypomethylated super-enhancer regions of *MYC* (**a**), *RNF43* (**b**) and *GPRC5A* (**c**) in colorectal cancer in normal (blue) and colorectal cancer (red) samples (TCGA). Significance was assessed using a linear regression model applied solely to the cancer samples. Matched DNA methylation was determined by averaging over the DNA methylation array probes (TCGA) falling into the hypomethylated regions of the respective super-enhancers.

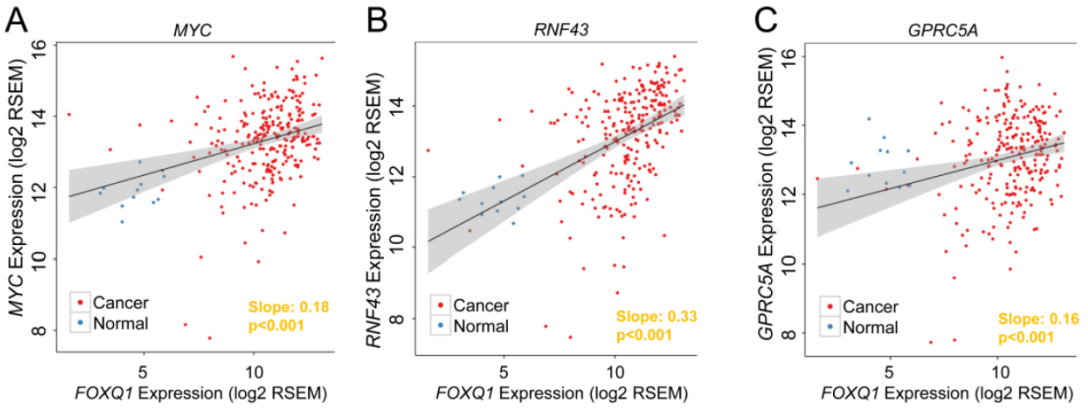


Figure 6.20. Association of *FOXQ1* expression levels and expression levels of genes associated with hypomethylated cancer-related super-enhancers in colorectal cancer, specifically, *MYC* (**a**), *RNF43* (**b**) and *GPRC5A* (**c**), in colorectal cancer in normal (blue) and colorectal cancer (red) samples (TCGA). Significance was assessed using a linear regression model applied solely to the cancer samples. Matched DNA methylation was determined by averaging over the DNA methylation array probes (TCGA) falling into the hypomethylated regions of the respective super-enhancers.

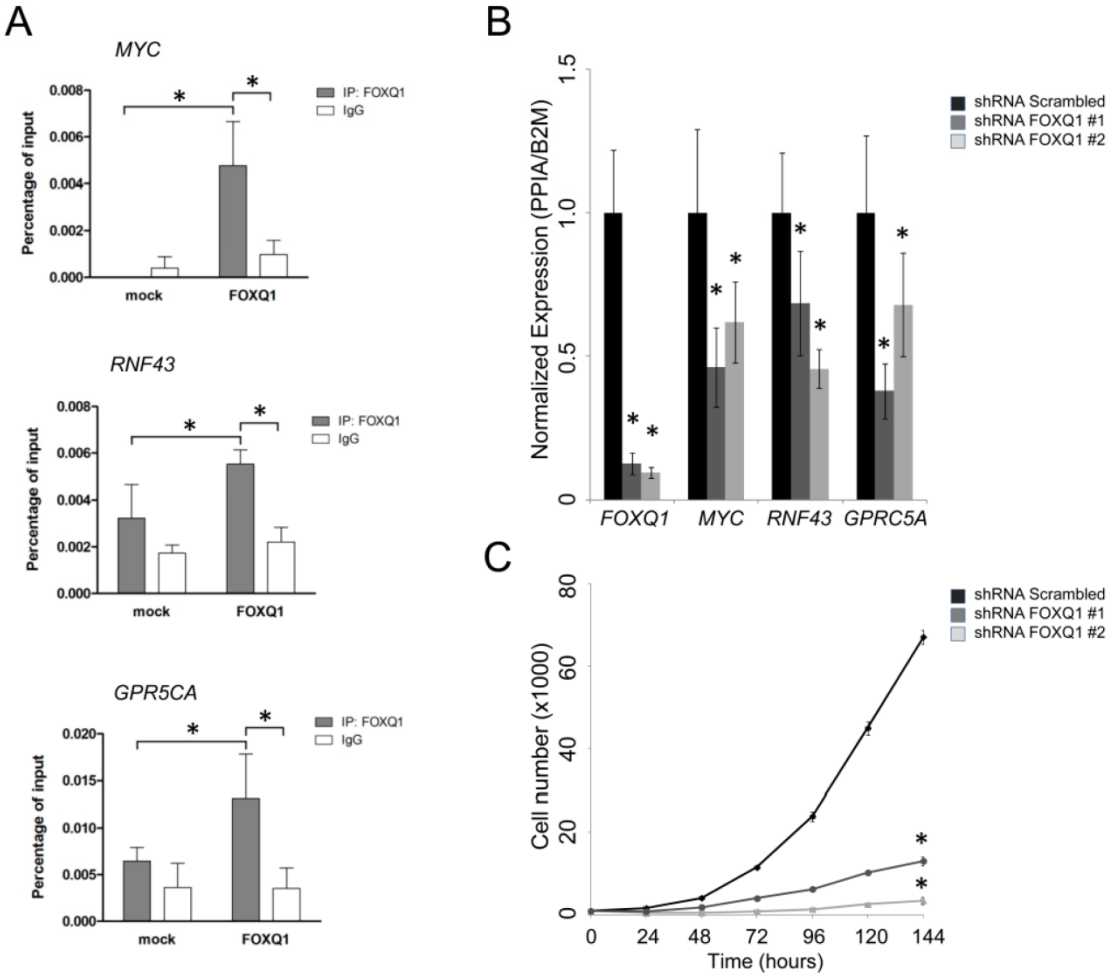


Figure 6.21. Functional validation of the effect of FOXQ1 on their predicted target genes *MYC*, *RNF43* and *GPRC5A*. (**a**) Chromatin immunoprecipitation (IP) experiments for FOXQ1 (flag-tagged) on predicted binding site in the super-enhancer regions related to *MYC*, *RNF43* and *GPRC5A* in HCT116 colorectal cancer cells. Significant enrichments were assigned using a Student's *t*-test (*). Enrichments were assessed comparing IP results of FOXQ1-flag expressing to untransfected cells or to unspecific antibody binding (IgG). (**b**) Small hairpin RNA (shRNA) mediated knockdown of *FOXQ1* led to a reduced expression of the predicted target genes *MYC*, *RNF43* and *GPRC5A* in SW1116 colorectal cancer cells. Significant differences in expression levels were assigned using a Student's *t*-test (*). (**c**) Small hairpin RNA (shRNA) mediated knockdown of *FOXQ1* in SW1116 resulted in a reduced viability of the colorectal cancer cells assessed using an MTT viability assays and analyzed over six consecutive days. Significant differences were assigned using a Student's *t*-test (*).

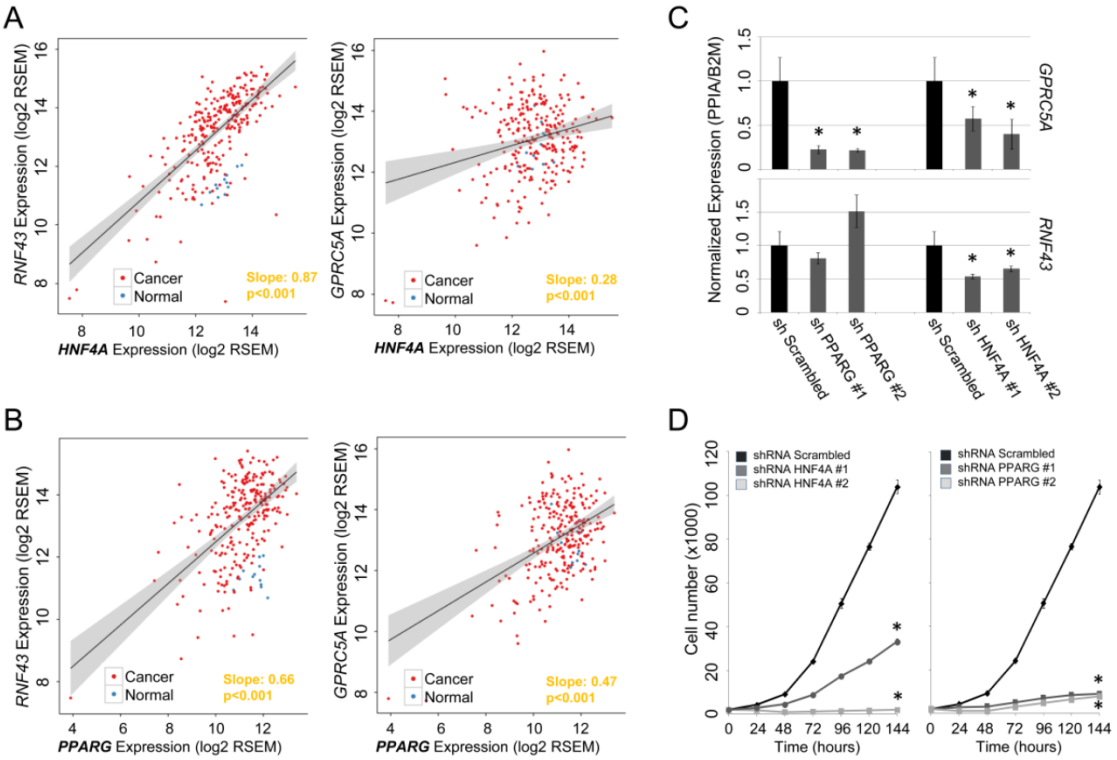


Figure 6.22. Validation of the effect of *HNF4A* and *PPARG* on their predicted target genes *RNF43* and *GPRC5A*. Association of TF expression levels and expression levels of genes associated with hypomethylated cancer-related super-enhancers in colorectal cancer, specifically, *RNF43* (**a**) and *GPRC5A* (**b**), in colorectal cancer in normal (blue) and colorectal cancer (red) samples (TCGA). Significance was assessed using a linear regression model applied solely to the cancer samples. Matched DNA methylation was determined by averaging over the DNA methylation array probes (TCGA) falling into the hypomethylated regions of the respective super-enhancers. (**c**) Small hairpin RNA (shRNA) mediated knockdown of *HNF4A* and *PPARG* led to a reduced expression of the predicted target genes *RNF43* and *GPRC5A* in SW1116 colorectal cancer cells. Significant differences were assigned using a Student's *t*-test (*). (**d**) Small hairpin RNA (shRNA) mediated knockdown of *HNF4A* and *PPARG* resulted in a reduced viability of SW1116 colorectal cancer cells assessed using an MTT viability assays and analyzed over six consecutive days. Significant differences were assigned using a Student's *t*-test (*).

of the secondary super-enhancer structure (**Fig. 6.23a,b**)[517]. Interestingly, although the treatment with JQ1 decreased the expression of super-enhancer gene targets, such as *MYC*, *RNF43* or *GPRC5A*, we could not detect an effect on DNA methylation levels at super-enhancer-related CpG sites (**Fig. 6.23c,d**). The lack of DNA methylation variance following JQ1 treatment suggests that the secondary super-enhancer structure per se is not a determinant of DNA methylation profiles, but that it is the binding of TFs to the DNA that locally establishes CpG methylation levels.

4.3 Large-scale hypomethylation marks potential cancer drivers

Finally, we wondered whether DNA methylation data obtained from WGBS could be used to identify new candidate cancer regulatory regions beyond the histone-based super-enhancer loci[515,615]. In line, extended hypomethylated regions were previously established as important regulatory elements in hematopoietic cells with a function in leukemogenesis[651]. To test this hypothesis, we ranked all the *de novo* formed hypomethylated DNA regions (<20 % average DNA methylation) in our colorectal cancer samples by size, having shown above that HMRs in colorectal tumorigenesis presented locus-specific properties (**Fig. 6.10b** and **Supplementary Table 6.4**). In this setting, we did observe an unequal distribution of HMR sizes, as previously reported for the super-enhancer-defining mark H3K27ac (**Fig. 6.24a**). Importantly, these large HMRs were mutually exclusive to the presence of super-enhancers in the respective regions, suggesting they represent an independent epigenetic feature to histone defined regulatory elements. Intriguingly, large HMRs mainly spanned gene promoter regions (22/26; **Supplementary Table 6.7**), a phenomenon previously described for genes activated in medulloblastoma patients, where an extensive expanded hypomethylation beyond the proximal promoter was observed, which might be a general feature of cancer-related gene activation[652]. Further, most of the HMRs that were present only in the metastatic cancer samples presented features suggesting a role in tumorigenesis. For example, the largest observed HMR (34.1 kb) in the metastatic colorectal cancer sample corresponded to beta-catenin (*CTNNB1*), a key component of the WNT pathway and driver of epithelial–mesenchymal transition (**Fig. 6.24b**)[653]. *AXIN2*, another key member of the WNT signaling pathway[654], was also among the top identified HMRs and is, together with an additional illustrative example, displayed in **Figure 6.24c,d**. Importantly, these findings were validated in an independent cohort of colorectal

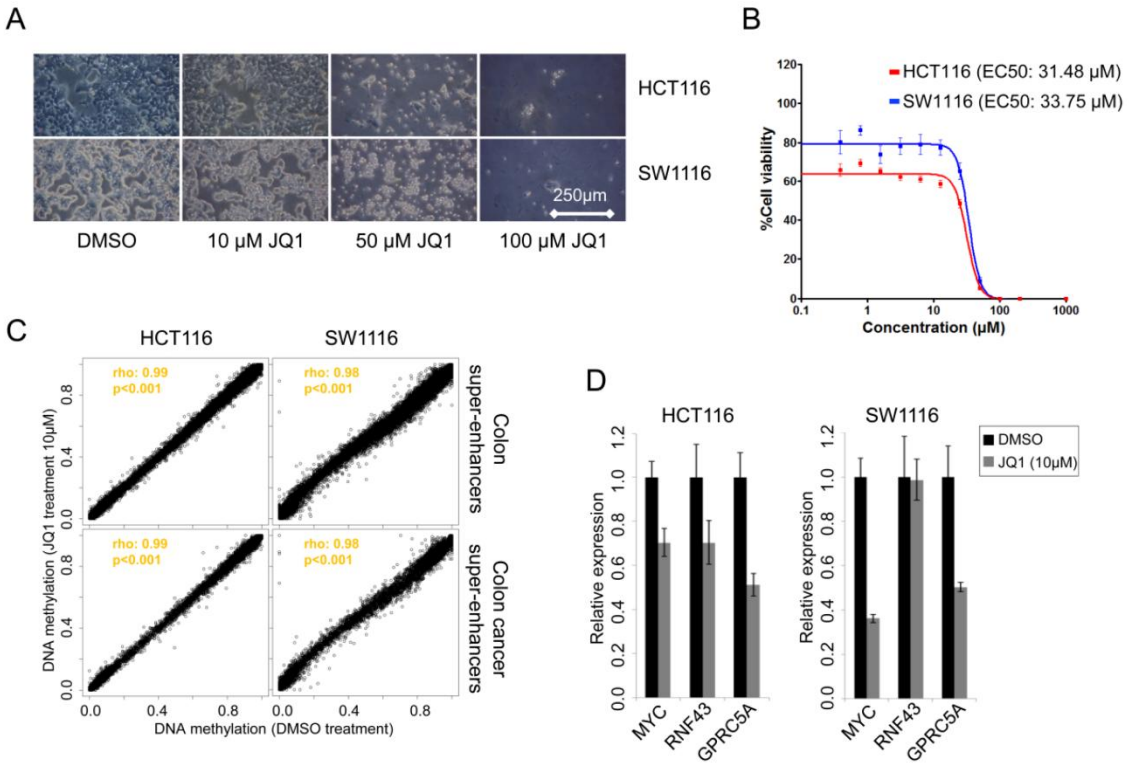


Figure 6.23. Super-enhancer disruption with the BRD4 inhibitor JQ1 does not affect DNA methylation profiles. (**a**) Phenotypic alterations of colorectal cancer cells after JQ1 treatment. HCT116 and SW1116 cells after 48h of treatment with DMSO (vehicle control) or 10µM, 50µM and 100µM of JQ1. Cell viability decreased at 10µM and showed high toxicity at concentration ≥50µM. (**b**) Cell viability assay after treatment with crescent concentrations of JQ1 for 48h to determine drug EC50 values (indicated). Each data point corresponds to the mean of four replicates and represents the percentage of JQ1-treated cells in respect to vehicle treated cells. Error bars ±SD are displayed. Viability curves were generated using a sigmoidal dose-response and a variable slope model. (**c**) DNA methylation levels of CpG sites within normal colon (up) or colorectal (down) super-enhancers comparing JQ1 (y-axis) and vehicle (x-axis) treated cells. Experiments were performed in HCT116 (left) and SW1116 (right) cancer cell lines and DNA methylation was determined using the Infinium HumanMethylation450 BeadChip. Similarity was assessed using a Spearman's correlation test (indicated). (**d**) Expression levels of *MYC*, *RNF43* and *GPRC5A* in HCT116 and SW1116 cells determined by qRT-PCR following 48h treatment with JQ1 or vehicle control (DMSO).

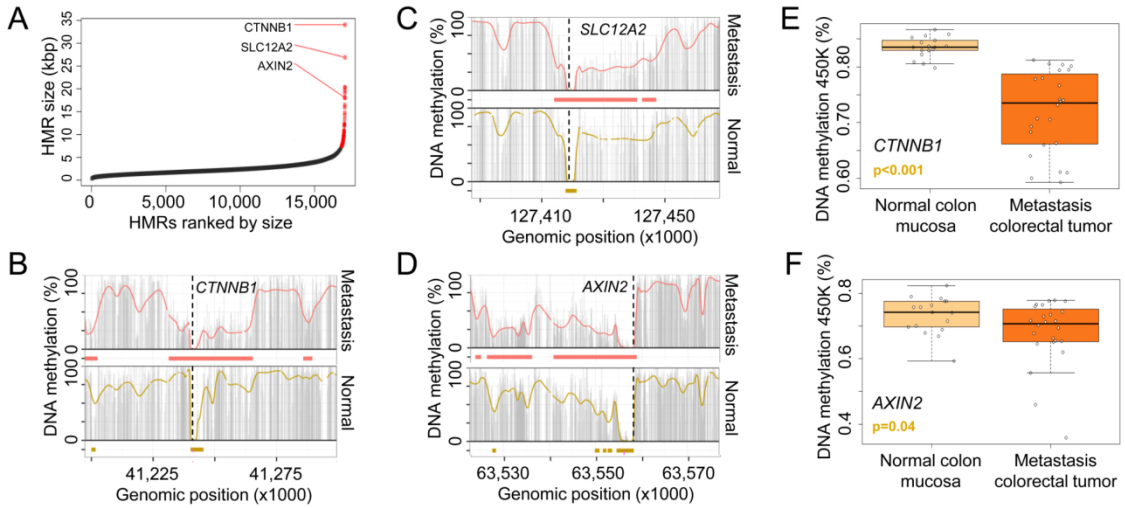


Figure 6.24. Large hypomethylated regions in colorectal metastasis. **a** HMRs derived from the metastatic colorectal cancer sample ranked by genomic size. Large HMRs are indicated (*red dots*). **b–d** DNA methylation profile of the HMRs spanning *CTNNB1* (**b**), *SLC12A2* (**c**) and *AXIN2* (**d**) in the normal (*yellow*) and metastatic (*red*) samples. Smoothed (*colored line*), raw (*gray bars*) CpG methylation levels and hypomethylated regions (*colored bars*) are indicated. **e, f** Validation of the large HMRs associated with *CTNNB1* (**e**) and *AXIN2* (**f**) using the HumanMethylation450 BeadChip. Displayed are average CpG methylation levels of *CTNNB1* and *AXIN2* for 18 normal colon mucosa and 24 colorectal metastasis samples. Significant differences were assessed using Student's *t*-test.

metastasis samples (n = 24) using DNA methylation microarray analysis (Student's *t*-test, $p < 0.05$; **Fig. 6.24e,f**). Thus, these findings suggest that large cancer-specific HMRs are likely candidate markers for identifying sequences that could act as *de novo* activators in a super-enhancer-like manner.

4.4 Conclusions

Overall, our findings indicate that super-enhancers, regulatory regions critical for cell identity and function, are partially regulated by their CpG methylation status in normal cells, and that they are targeted by specific aberrant DNA methylation events in cancer, with putative effects for the expression of the downstream-controlled genes. Further, we determined spatial differences of healthy and transformed DNA methylation profiles within these large enhancer clusters, suggesting local differences in activity in super-enhancer regions.

We hypothesize that local changes in TF binding act on super-enhancer DNA methylation profiles with subsequent effects on target gene expression. Accordingly, super-enhancer DNA methylation levels indicate regulatory activity and, moreover, point to implicated TFs. In cancer, the perturbed expression of key TFs establishes novel super-enhancers that drive oncogene expression, a scenario that we partially delineated through the identification of FOXQ1 as a putative factor driving the differential DNA methylation at colorectal cancer-specific super-enhancers and the overexpression of key oncogenes, such as *MYC* and *RNF43*.

Our results also emphasize that developing more extensive catalogues of human DNA methylomes at base resolution would help us gain a better understanding of the regulatory functions of DNA methylation beyond those of the most widely studied proximal promoter gene regions.

5. Acknowledgments

The research leading to these results received funding from: the European Research Council (ERC), grant EPINORC, under agreement number 268626; MICINN Projects – SAF2011-22803 and BFU2011-28549; Ministerio de Economía y Competitividad (MINECO), co-financed by the European Development Regional Fund, 'A way to achieve Europe' ERDF, under grant number SAF2014-55000-R; the Cellex Foundation; AGAUR Catalan Government Project #2009SGR1315; the Institute of Health Carlos III (ISCIII), under the Spanish Cancer Research Network (RTICC) number RD12/0036/0039, the Integrated Project of Excellence number PIE13/00022 (ONCOPROFILE) and the research grant PI11/00321; the Sandra Ibarra Foundation, under IV ghd Grants for breast cancer research; the Olga Torres Foundation; the European Community's Seventh Framework Programme (FP7/2007-2013), grant HEALTH-F5-2011-282510 – BLUEPRINT, and the Health and Science Departments of the Generalitat de Catalunya. H.H. is a Miguel Servet (CP14/00229) researcher funded by the Spanish Institute of Health Carlos III (ISCIII). D.T. and M.E. are ICREA Research Professors.

6. Authors' contributions

H.H. and M.E. conceived and directed the study. H.H. and E.V. analyzed the data with support of S.S. and A.G. **H.J.F**. and M.V. **performed experiments**. A.M.C., J.V.S.M. and R.M. provided clinical samples. C.Y.L., R.R., D.T., M.O. and R.A.Y. contributed to data analysis. R.S.B. and S.G. performed RNA sequencing, I.G. and M.G. DNA sequencing and S.M. the Infinium 450 k microarray experiments and primary data analysis. H.H. and M.E. wrote the manuscript. All authors read and approved the final manuscript.

7. Supplementary information

Note: For a more user friendly lecture of this chapter, the following tables were not inserted next to the place where they were mentioned. Moreover, to not considerably affect the structure of this thesis they were resized. This information is also accessible in the online version of the article (http://dx.doi.org/10.1186/s13059-016-0879-2).

Supplementary Table 6.1: Tissue-specific DNA methylation at super-enhancers in normal tissue types.

	Super-Enhancer					HMR coverage (%)		
Tissue	Chr.	Start	End	Transcript ID	Gene	Tissue	Others	Difference
Brain	6	163815822	163882069	ENST00000537883.1	QKI	0.79	0.13	0.66
Brain	7	29516606	29530680	ENST00000410098.1	CHN2	0.63	0.04	0.59
Brain	1	161160323	161173936	ENST00000478866.1	NDUFS2	0.80	0.21	0.59
Brain	13	78291692	78329377	ENST00000358679.2	SLAIN1	0.59	0.00	0.59
Brain	12	110152218	110181906	ENST00000358906.3	FAM222A	0.66	0.08	0.58
Brain	17	79097124	79111085	ENST00000575363.1	AATK	0.59	0.03	0.56
Brain	22	46467856	46491390	ENST00000360737.3	MIRLET7	0.95	0.42	0.54
Brain	18	74684916	74780601	ENST00000583266.1	MBP	0.53	0.04	0.49
Brain	9	131142171	131157912	ENST00000483206.1	URM1	0.49	0.01	0.48
Brain	11	61520065	61534843	ENST00000265460.5	MYRF	0.53	0.06	0.47
Brain	7	124359038	124407316	ENST00000303921.2	GPR37	0.51	0.05	0.46
Brain	18	13610336	13640017	ENST00000435606.1	LDLRAD4	0.53	0.08	0.45
Brain	22	46457577	46466999	ENST00000396008.2	MIRLET7	0.94	0.50	0.44
Brain	22	46457577	46466999	ENST00000333761.1	MIRLET7	0.94	0.50	0.44
Brain	9	130697475	130743184	ENST00000494606.1	FAM102A	0.61	0.17	0.44
Brain	10	14577374	14624047	ENST00000495292.1	FAM107B	0.46	0.02	0.44
Brain	13	67782754	67806323	ENST00000377861.3	PCDH9	0.59	0.20	0.39
Brain	3	33685210	33704614	ENST00000486796.1	CLASP2	0.45	0.06	0.39
Brain	5	36599886	36615091	ENST00000509272.1	SLC1A3	0.50	0.11	0.38
Brain	21	17441409	17461338	ENST00000478932.1	USP25	0.43	0.04	0.38
Brain	8	120648911	120687423	ENST00000427067.2	ENPP2	0.41	0.04	0.37
Brain	7	22212818	22261220	ENST00000420196.1	RAPGEF5	0.37	0.00	0.37
Brain	16	1013186	1042763	ENST00000566034.1	SOX8	0.55	0.19	0.36
Brain	22	38361166	38471689	ENST00000494434.1	PICK1	0.39	0.03	0.36
Brain	9	140081775	140094696	ENST00000477345.1	TPRN	0.56	0.20	0.35
Brain	18	60417206	60470491	ENST00000497351.1	PHLPP1	0.35	0.00	0.35
Brain	8	26427962	26519962	ENST00000474808.1	DPYSL2	0.45	0.10	0.35
Brain	1	198871424	198907549	ENST00000491302.1	PTPRC	0.48	0.13	0.35
Brain	17	36570707	36622668	ENST00000443378.1	ARHGAP23	0.54	0.20	0.34
Brain	3	72424289	72464834	ENST00000477973.1	RYBP	0.34	0.00	0.34
Brain	6	134488737	134505092	ENST00000525700.1	SGK1	0.60	0.27	0.34
Brain	10	81137824	81212605	ENST00000372333.3	ZCCHC24	0.43	0.09	0.33
Brain	4	96452170	96474433	ENST00000504962.1	UNC5C	0.48	0.15	0.33
Brain	16	75266933	75302422	ENST00000393422.2	BCAR1	0.51	0.18	0.33
Brain	3	39487852	39573984	ENST00000479860.1	MOBP	0.34	0.02	0.32
Brain	4	154142334	154219000	ENST00000433687.1	TRIM2	0.38	0.05	0.32
Brain	6	163659272	163687676	ENST00000419182.1	AL078585.1	0.36	0.04	0.32
Brain	10	88421359	88444547	ENST00000542786.1	LDB3	0.33	0.02	0.31
Brain	12	48167031	48182319	ENST00000599515.1	AC004466.1	0.40	0.10	0.31
Brain	2	145129528	145283099	ENST00000475115.1	ZEB2	0.39	0.08	0.30
Brain	4	134065738	134085565	ENST00000511112.1	PCDH10	0.73	0.42	0.30
Brain	1	160068777	160077945	ENST00000448417.1	IGSF8	0.44	0.14	0.30
Brain	2	127806748	127912517	ENST00000409400.1	BIN1	0.36	0.06	0.30
Brain	3	181403893	181455295	ENST00000325404.1	SOX2	0.84	0.56	0.28
Brain	16	19841311	19898686	ENST00000564449.1	GPRC5B	0.35	0.06	0.28
Brain	6	30844750	30864481	ENST00000514434.1	DDR1	0.45	0.18	0.27
Brain	4	115547254	115584118	ENST00000394511.3	UGT8	0.27	0.00	0.27
Brain	14	77765872	77789479	ENST00000555093.1	GSTZ1	0.39	0.12	0.27
Brain	10	650150	686281	ENST00000441152.2	PRR26	0.31	0.03	0.27
Brain	8	53318133	53342376	ENST00000520716.1	ST18	0.29	0.02	0.27
Brain	2	220142124	220157631	ENST00000460801.1	PTPRN	0.43	0.16	0.27
Brain	3	170135527	170167574	ENST00000488989.1	CLDN11	0.34	0.07	0.27
Brain	17	68152783	68183190	ENST00000243457.3	KCNJ2	0.38	0.11	0.27
Brain	21	48018786	48058971	ENST00000440086.1	PRMT2	0.35	0.09	0.26
Brain	19	13093539	13211813	ENST00000588680.1	NFIX	0.47	0.21	0.26
Brain	14	50976072	51001976	ENST00000441560.2	ATL1	0.36	0.11	0.26
Brain	5	173313778	173354104	ENST00000519467.1	CPEB4	0.34	0.08	0.25
Brain	3	134072004	134096230	ENST00000502491.1	AMOTL2	0.44	0.19	0.25
Brain	22	37940770	37968154	ENST00000434728.1	CDC42EP1	0.50	0.25	0.25
Brain	2	164566271	164594418	ENST00000482917.1	FIGN	0.35	0.11	0.25
Brain	15	64974641	64996542	ENST00000559665.2	OAZ2	0.36	0.11	0.24
Brain	1	65712720	65745163	ENST00000483402.1	DNAJC6	0.34	0.10	0.24
Brain	13	101281681	101326566	ENST00000480433.1	TMTC4	0.26	0.02	0.24
Brain	11	67173218	67188540	ENST00000531388.1	CARNS1	0.45	0.21	0.24
Brain	2	158160059	158191916	ENST00000419116.2	ERMN	0.24	0.00	0.24
Brain	11	111845328	111865086	ENST00000530645.1	DIXDC1	0.33	0.09	0.24
Brain	5	126625320	126648566	ENST00000274473.6	MEGF10	0.30	0.06	0.24
Brain	1	221879013	221917508	ENST00000477026.1	DUSP10	0.32	0.09	0.23
Brain	10	22858862	22973344	ENST00000432610.1	PIP4K2A	0.25	0.02	0.23
Brain	7	28724048	28800494	ENST00000468813.1	CREB5	0.27	0.04	0.23
Brain	2	105022517	105068208	ENST00000361360.2	POU3F3	0.24	0.01	0.23
Brain	22	39853661	39873571	ENST00000418314.1	MGAT3	0.34	0.11	0.23
Brain	9	97831315	97861405	ENST00000468164.1	C9orf3	0.25	0.02	0.23
Brain	1	202090299	202115217	ENST00000272217.2	ARL8A	0.39	0.17	0.23
Brain	16	56277138	56298622	ENST00000563440.1	GNAO1	0.24	0.02	0.22
Brain	5	132105879	132116081	ENST00000492490.1	8-Sep	0.43	0.21	0.22
Brain	13	107142225	107162158	ENST00000245323.4	EFNB2	0.29	0.07	0.22
Brain	2	102312024	102465976	ENST00000421882.1	MAP4K4	0.26	0.04	0.22
Brain	10	73722545	73806874	ENST00000373115.4	CHST3	0.28	0.06	0.22
Brain	4	48622532	48657738	ENST00000302806.5	FRYL	0.22	0.00	0.21
Brain	1	160047565	160068437	ENST00000460351.1	IGSF8	0.35	0.14	0.21
Brain	10	126400866	126432679	ENST00000392754.3	FAM53B	0.40	0.19	0.21
Brain	9	133698900	133750533	ENST00000318560.5	ABL1	0.29	0.08	0.21
Brain	17	40114067	40162277	ENST00000591153.1	DNAJC7	0.27	0.07	0.20
Brain	16	57277838	57336233	ENST00000569059.1	PLLP	0.30	0.10	0.20
Brain	10	64127809	64167185	ENST00000421210.1	ZNF365	0.23	0.03	0.20
Brain	3	133460192	133484110	ENST00000498622.1	TF	0.24	0.04	0.20
Brain	1	109889550	109945440	ENST00000471996.1	SORT1	0.26	0.05	0.20
Brain	14	65146034	65196928	ENST00000556801.1	PLEKHG3	0.31	0.11	0.20
Brain	10	128242317	128261039	ENST00000368674.1	C10orf90	0.24	0.04	0.20
Brain	17	72226877	72248627	ENST00000526858.1	TTYH2	0.23	0.03	0.20
Brain	22	45603349	45638976	ENST00000417906.1	KIAA0930	0.27	0.08	0.20
Brain	1	205461954	205494873	ENST00000505932.1	CDK18	0.26	0.06	0.20
Brain	2	74591442	74620177	ENST00000440727.1	DCTN1	0.27	0.07	0.20
Brain	10	124216441	124276238	ENST00000420892.1	HTRA1	0.27	0.07	0.20
Brain	19	54921528	54950915	ENST00000467939.1	TTYH1	0.38	0.18	0.20
Brain	20	35445944	35496093	ENST00000357779.3	SOGA1	0.30	0.11	0.20
Brain	6	41649397	41702285	ENST00000433032.1	TFEB	0.30	0.10	0.20
Brain	2	95678307	95745847	ENST00000349807.3	MAL	0.27	0.08	0.19
Brain	20	31025912	31074699	ENST00000326071.4	C20orf112	0.35	0.16	0.19
Brain	15	45655307	45688573	ENST00000558118.1	GATM	0.24	0.05	0.19
Brain	4	113802567	113832361	ENST00000503271.1	ANK2	0.21	0.02	0.19
Brain	4	113802567	113832361	ENST00000503423.1	ANK2	0.21	0.02	0.19
Brain	4	113802567	113832361	ENST00000506722.1	ANK2	0.21	0.02	0.19
Brain	1	214156400	214205391	ENST00000261454.4	PROX1	0.32	0.13	0.19
Brain	10	126679464	126714232	ENST00000486955.2	CTBP2	0.23	0.04	0.19
Brain	1	109816196	109827024	ENST00000471740.1	PSRC1	0.36	0.17	0.19
Brain	1	66414430	66466650	ENST00000526666.1	PDE4B	0.20	0.02	0.18
Brain	1	90285126	90311060	ENST00000527156.1	LRRC8D	0.33	0.14	0.18
Brain	11	124785353	124817909	ENST00000528971.1	HEPACAM	0.20	0.02	0.18
Brain	1	230932874	231004412	ENST00000366663.5	C1orf198	0.21	0.03	0.18
Brain	1	204796255	204866707	ENST00000514644.1	NFASC	0.21	0.03	0.18
Brain	7	30174536	30225450	ENST00000455738.1	C7orf41	0.23	0.06	0.18
Brain	17	79345155	79376776	ENST00000307745.7	BAHCC1	0.54	0.37	0.17
Brain	17	73682578	73702695	ENST00000579005.1	SAP30BP	0.22	0.05	0.17
Brain	5	124064380	124097267	ENST00000512940.1	ZNF608	0.41	0.24	0.17
Brain	19	50985631	51019282	ENST00000598657.1	ASPDH	0.28	0.11	0.17
Brain	7	32078425	32118455	ENST00000464881.1	PDE1C	0.23	0.06	0.17
Brain	12	123317747	123395518	ENST00000371248.3	VPS37B	0.30	0.13	0.17
Brain	9	135976307	136024639	ENST00000372062.3	RALGDS	0.33	0.16	0.17
Brain	17	56589712	56620235	ENST00000582976.1	4-Sep	0.36	0.19	0.17
Brain	11	63801514	63883871	ENST00000246841.3	FLRT1	0.21	0.05	0.17
Brain	6	69337682	69367340	ENST00000370598.1	BAI3	0.33	0.16	0.17
Brain	12	26265040	26280746	ENST00000534829.1	SSPN	0.64	0.47	0.17
Brain	20	45187487	45213480	ENST00000464518.1	SLC13A3	0.22	0.05	0.17
Brain	3	11024732	11080079	ENST00000495636.1	SLC6A1	0.23	0.06	0.17
Brain	1	160161970	160191690	ENST00000481831.1	DCAF8	0.25	0.09	0.17
Brain	1	160019557	160042341	ENST00000368089.3	KCNJ10	0.23	0.07	0.16
Brain	18	13548488	13588929	ENST00000592657.1	LDLRAD4	0.21	0.05	0.16
Brain	9	8840393	8880342	ENST00000481079.1	PTPRD	0.24	0.08	0.16
Brain	7	141373362	141403139	ENST00000482493.1	KIAA1147	0.22	0.06	0.16
Brain	1	149967151	149985599	ENST00000417191.1	OTUD7B	0.27	0.11	0.16
Brain	4	77099124	77138142	ENST00000512895.1	FAM47E	0.30	0.14	0.16
Brain	22	46445235	46456789	ENST00000381051.2	MIRLET7	0.63	0.47	0.16
Brain	10	699249	735394	ENST00000280886.6	DIP2C	0.21	0.05	0.16
Brain	10	73574974	73621362	ENST00000394934.1	PSAP	0.21	0.05	0.15
Brain	17	27071068	27095625	ENST00000584059.1	FAM222B	0.29	0.14	0.15
Brain	11	134254603	134298147	ENST00000531510.1	B3GAT1	0.26	0.11	0.15
Brain	13	97862565	97936838	ENST00000445661.2	MBNL2	0.24	0.09	0.15
Brain	14	90845063	90876551	ENST00000556721.1	CALM1	0.35	0.20	0.15
Brain	2	193010308	193063844	ENST00000409056.3	TMEFF2	0.21	0.06	0.15
Brain	3	156845617	156856684	ENST00000467849.1	CCNL1	0.30	0.15	0.15
Brain	1	155828272	155864174	ENST00000539162.1	SYT11	0.21	0.06	0.15
Brain	7	45111554	45132881	ENST00000490531.2	NACAD	0.22	0.07	0.15
Brain	6	166689003	166721330	ENST00000322583.3	PRR18	0.20	0.05	0.15
Brain	1	205626881	205650899	ENST00000367145.3	SLC45A3	0.29	0.14	0.15
Brain	9	14299740	14327190	ENST00000493697.1	NFIB	0.57	0.42	0.15
Brain	17	38216903	38234316	ENST00000577486.1	THRA	0.63	0.48	0.15
Brain	9	124022263	124092639	ENST00000477553.1	GSN	0.26	0.12	0.15
Brain	13	114786201	114844487	ENST00000389544.4	RASA3	0.21	0.06	0.15
Brain	11	6400761	6441884	ENST00000533407.1	APBB1	0.21	0.07	0.15
Brain	1	205195003	205257756	ENST00000481950.1	TMCC2	0.23	0.09	0.14
Brain	1	33789227	33859751	ENST00000468406.1	PHC2	0.25	0.11	0.14
Brain	6	52225169	52265846	ENST00000360726.3	PAQR8	0.21	0.07	0.14
Brain	5	115856316	115912087	ENST00000512156.1	SEMA6A	0.23	0.08	0.14
Brain	17	17638844	17657613	ENST00000395774.1	RAI1	0.30	0.16	0.14
Brain	5	139012918	139146609	ENST00000505812.1	CXXC5	0.36	0.22	0.14
Brain	16	15234031	15244187	ENST00000448014.2	NPIPP1	0.27	0.13	0.14
Brain	16	15234031	15244187	ENST00000358815.3	NPIPP1	0.27	0.13	0.14
Brain	17	76857381	76931557	ENST00000262768.7	TIMP2	0.23	0.09	0.14
Brain	17	57405868	57485233	ENST00000582813.1	YPEL2	0.20	0.06	0.14
Brain	19	51033029	51070726	ENST00000599957.1	LRRC4B	0.30	0.17	0.14
Brain	17	42161429	42202224	ENST00000587135.1	HDAC5	0.23	0.10	0.14
Brain	1	110071934	110092553	ENST00000369851.4	GNAI3	0.31	0.18	0.13
Brain	12	49579740	49630368	ENST00000552125.1	TUBA1C	0.23	0.09	0.13
Brain	1	61541625	61624014	ENST00000496712.1	NFIA	0.24	0.11	0.13
Brain	14	51524492	51564122	ENST00000360392.4	TRIM9	0.20	0.07	0.13
Brain	3	123128250	123170404	ENST00000462833.1	ADCY5	0.20	0.08	0.13
Brain	11	118463235	118534958	ENST00000528823.1	PHLDB1	0.23	0.10	0.13
Brain	10	102752997	102775915	ENST00000454422.1	LZTS2	0.44	0.32	0.13
Brain	11	73357627	73375282	ENST00000544282.1	PLEKHB1	0.20	0.08	0.13
Brain	7	86272604	86301105	ENST00000421579.1	GRM3	0.22	0.09	0.13
Brain	1	32792421	32824727	ENST00000574315.1	TSSK3	0.27	0.14	0.12
Brain	15	37371561	37406086	ENST00000559129.1	MEIS2	0.46	0.33	0.12
Brain	15	43800300	43824728	ENST00000300231.5	MAP1A	0.28	0.16	0.12
Brain	5	171829492	171881486	ENST00000519643.1	SH3PXD2B	0.21	0.09	0.12
Brain	11	119211319	119250680	ENST00000527843.1	USP2	0.22	0.10	0.12
Brain	5	646895	697805	ENST00000360578.5	TPPP	0.23	0.12	0.12
Brain	2	232526039	232554844	ENST00000466801.1	PTMA	0.34	0.23	0.12
Brain	10	72971329	73044360	ENST00000373192.4	UNC5B	0.21	0.09	0.12
Brain	10	24989499	25014855	ENST00000463892.1	ARHGAP21	0.33	0.22	0.12
Brain	2	9768796	9795405	ENST00000460093.1	YWHAQ	0.21	0.09	0.11
Brain	16	1064221	1098971	ENST00000397547.2	SSTR5	0.24	0.13	0.11
Brain	19	17858364	17882158	ENST00000600393.1	FCHO1	0.22	0.11	0.11
Brain	12	6639174	6665533	ENST00000396830.2	IFFO1	0.40	0.29	0.11
Brain	10	111967550	111992938	ENST00000369612.1	MXI1	0.40	0.29	0.11
Brain	17	48914258	48946511	ENST00000509385.1	TOB1	0.28	0.17	0.11
Brain	17	38258962	38270871	ENST00000246672.3	NR1D1	0.42	0.31	0.11
Brain	9	139405020	139447486	ENST00000277541.6	NOTCH1	0.32	0.21	0.11
Brain	6	15245470	15269892	ENST00000397311.3	JARID2	0.36	0.25	0.11
Brain	12	1752842	1776949	ENST00000545747.1	WNT5B	0.24	0.13	0.11
Brain	6	128798636	128843173	ENST00000368202.4	PTPRK	0.20	0.09	0.11
Brain	10	88108134	88126246	ENST00000327946.7	GRID1	0.29	0.19	0.11
Brain	1	156353833	156405678	ENST00000495000.1	C1orf61	0.29	0.18	0.11
Brain	21	34399873	34453980	ENST00000498799.1	OLIG1	0.28	0.18	0.11
Brain	20	45922895	45992027	ENST00000441977.1	ZMYND8	0.24	0.13	0.11
Brain	21	43639119	43683392	ENST00000340588.4	ABCG1	0.22	0.11	0.11
Brain	17	76703832	76732967	ENST00000590300.1	CYTH1	0.24	0.14	0.10
Brain	3	69130585	69168856	ENST00000485444.1	ARL6IP5	0.22	0.12	0.10
Brain	19	30713867	30786928	ENST00000591488.1	ZNF536	0.21	0.11	0.10
Brain	12	48125991	48140168	ENST00000395360.2	RAPGEF3	0.20	0.10	0.10
Breast	22	46467114	46488203	ENST00000360737.3	MIRLET7	1.00	0.47	0.53
Breast	5	131593100	131603174	ENST00000418373.1	PDLIM4	0.60	0.08	0.51
Breast	11	8828385	8837469	ENST00000531237.1	ST5	0.63	0.12	0.51
Breast	22	46448058	46467049	ENST00000381051.2	MIRLET7	1.00	0.50	0.50
Breast	9	33157786	33167403	ENST00000379731.4	B4GALT1	0.92	0.42	0.50
Breast	12	10868156	10877012	ENST00000540747.1	CSDA	0.79	0.29	0.50
Breast	1	157964217	157991843	ENST00000416935.2	KIRREL	0.50	0.04	0.47
Breast	12	6479020	6487672	ENST00000539925.1	LTBR	0.57	0.12	0.46
Breast	3	29321550	29334286	ENST00000456853.1	RBMS3	0.65	0.21	0.45
Breast	5	142773722	142785691	ENST00000502500.1	NR3C1	0.95	0.51	0.44
Breast	1	153579895	153590703	ENST00000469571.1	S100A14	0.49	0.05	0.44
Breast	15	42562799	42569003	ENST00000567421.1	GANC	0.91	0.51	0.40
Breast	4	146653244	146658422	ENST00000510096.1	C4orf51	0.68	0.29	0.39
Breast	1	45270631	45276266	ENST00000482715.1	BTBD19	0.65	0.27	0.38
Breast	1	156065471	156100930	ENST00000392353.3	LMNA	0.56	0.18	0.38
Breast	2	239193425	239200323	ENST00000431832.1	PER2	0.73	0.36	0.37
Breast	2	36579464	36605312	ENST00000473403.1	CRIM1	0.55	0.19	0.35
Breast	18	3622231	3626476	ENST00000486430.1	DLGAP1	0.52	0.17	0.35
Breast	21	36252652	36264673	ENST00000399237.2	RUNX1	0.93	0.58	0.35
Breast	1	33792562	33815439	ENST00000473158.1	PHC2	0.49	0.16	0.33
Breast	10	24737918	24757794	ENST00000376451.2	KIAA1217	0.37	0.04	0.33
Breast	8	26465475	26470176	ENST00000523690.1	DPYSL2	0.73	0.40	0.32
Breast	3	141083866	141089271	ENST00000507657.1	ZBTB38	0.39	0.08	0.32
Breast	1	154941228	154949531	ENST00000473344.1	CKS1B	0.82	0.52	0.31
Breast	7	115993193	115998892	ENST00000465451.1	CAV2	0.37	0.07	0.30
Breast	14	77489392	77513614	ENST00000238647.3	IRF2BPL	0.67	0.37	0.29
Breast	14	69402525	69445588	ENST00000553659.1	ACTN1	0.35	0.06	0.29
Breast	20	51581498	51597746	ENST00000371497.4	TSHZ2	0.45	0.16	0.29
Breast	7	55085115	55095372	ENST00000463948.1	EGFR	0.50	0.21	0.29
Breast	9	14298144	14323961	ENST00000493697.1	NFIB	0.71	0.42	0.28
Breast	1	27830256	27900464	ENST00000487743.2	AHDC1	0.37	0.09	0.28
Breast	8	145008274	145031982	ENST00000527816.1	PLEC	0.62	0.34	0.27
Breast	11	62306073	62328249	ENST00000531324.1	AHNAK	0.61	0.34	0.27
Breast	7	27134577	27144793	ENST00000222718.5	HOXA2	0.95	0.68	0.27
Breast	20	10634437	10656669	ENST00000254958.4	JAG1	0.49	0.22	0.27
Breast	15	93425946	93432559	ENST00000555520.1	CHD2	0.87	0.61	0.26
Breast	11	6339234	6344196	ENST00000530979.1	PRKCDBP	0.61	0.34	0.26
Breast	8	116659885	116682409	ENST00000451156.1	TRPS1	0.42	0.17	0.26
Breast	1	59245181	59252471	ENST00000371222.2	JUN	0.86	0.61	0.25
Breast	12	120661774	120703620	ENST00000547772.1	PXN	0.36	0.10	0.25
Breast	11	114162169	114180527	ENST00000545255.1	NNMT	0.26	0.03	0.24
Breast	10	122913683	122920721	ENST00000478567.1	WDR11	0.38	0.16	0.22
Breast	4	77506454	77513253	ENST00000484236.1	SHROOM3	0.31	0.09	0.22
Breast	12	59300121	59315174	ENST00000548968.1	LRIG3	0.39	0.18	0.21
Breast	1	100109973	100129273	ENST00000263174.4	PALMD	0.24	0.03	0.21
Breast	1	60135366	60141479	ENST00000471169.1	FGGY	0.24	0.03	0.21
Breast	5	121515651	121520476	ENST00000514925.1	ZNF474	0.26	0.06	0.21
Breast	6	74224616	74234296	ENST00000455918.1	EEF1A1	0.79	0.59	0.21
Breast	11	9586185	9600407	ENST00000524612.1	WEE1	0.46	0.26	0.21
Breast	20	43963799	43990114	ENST00000537976.1	SDC4	0.33	0.13	0.21
Breast	19	13123147	13172690	ENST00000358552.3	NFIX	0.50	0.29	0.20
Breast	14	105432978	105441663	ENST00000555122.1	AHNAK2	0.43	0.23	0.20
Breast	17	2294949	2311496	ENST00000571836.2	MNT	0.62	0.42	0.20
Breast	11	57528367	57568939	ENST00000534647.1	CTNND1	0.31	0.11	0.20
Breast	1	144706472	144710218	ENST00000338347.4	NBPF9	0.32	0.12	0.20
Breast	1	144706472	144710218	ENST00000440491.2	NBPF9	0.32	0.12	0.20
Breast	11	65238595	65276583	ENST00000309775.7	AP000769.1	0.63	0.43	0.20
Breast	14	55568260	55575863	ENST00000553493.1	LGALS3	0.32	0.13	0.19
Breast	6	121756401	121769000	ENST00000282561.3	GJA1	0.38	0.18	0.19
Breast	22	30792614	30822843	ENST00000407550.3	MTFP1	0.36	0.17	0.19
Breast	12	6432600	6452223	ENST00000538363.1	TNFRSF1A	0.50	0.30	0.19
Breast	2	112249648	112262948	ENST00000371162.4	AC108463.3	0.27	0.08	0.19

Breast	1	181055984	181089589	ENST00000367577.4	IER5	0.32	0.13	0.19
Breast	16	56638741	56649586	ENST00000567300.1	MT2A	0.46	0.27	0.19
Breast	5	1881509	1890183	ENST00000513692.1	IRX4	0.97	0.78	0.19
Breast	3	187453456	187468930	ENST00000496823.1	BCL6	0.81	0.63	0.19
Breast	1	207190041	207209398	ENST00000461135.2	C1orf116	0.22	0.03	0.19
Breast	6	111906089	111928358	ENST00000528599.1	TRAF3IP2	0.33	0.15	0.18
Breast	22	27868674	27873807	ENST00000497225.1	MN1	0.23	0.04	0.18
Breast	22	36718824	36814131	ENST00000216181.5	MYH9	0.34	0.16	0.18
Breast	19	49375182	49379790	ENST00000600406.1	PPP1R15A	0.78	0.60	0.18
Breast	3	193849125	193860857	ENST00000476918.1	HES1	0.79	0.61	0.18
Breast	11	118779500	118810391	ENST00000534788.1	UPK2	0.63	0.45	0.18
Breast	19	1154329	1184542	ENST00000587655.1	SBNO2	0.28	0.10	0.18
Breast	18	3245538	3270102	ENST00000584539.1	MYL12B	0.43	0.25	0.18
Breast	14	74207241	74227902	ENST00000421708.1	ELMSAN1	0.58	0.41	0.18
Breast	14	23292133	23323062	ENST00000547596.1	MMP14	0.33	0.15	0.17
Breast	10	95184875	95242959	ENST00000488645.1	MYOF	0.21	0.04	0.17
Breast	14	55537362	55546494	ENST00000553976.1	MAPK1IP1L	0.24	0.07	0.17
Breast	20	52443340	52447648	ENST00000422805.1	BCAS1	0.63	0.46	0.17
Breast	17	46122636	46152031	ENST00000581319.1	NFE2L1	0.28	0.11	0.17
Breast	11	66821179	66830989	ENST00000532559.1	RHOD	0.33	0.16	0.17
Breast	20	34326330	34331841	ENST00000493853.1	RBM39	0.91	0.74	0.17
Breast	13	48971911	48988013	ENST00000482024.1	LPAR6	0.24	0.07	0.17
Breast	8	62622640	62634995	ENST00000379449.6	ASPH	0.38	0.21	0.17
Breast	2	12841318	12863362	ENST00000405331.3	TRIB2	0.35	0.19	0.17
Breast	5	67510759	67574308	ENST00000520675.1	PIK3R1	0.27	0.10	0.16
Breast	19	16177568	16192210	ENST00000588483.1	TPM4	0.43	0.26	0.16
Breast	10	112252889	112273435	ENST00000468749.1	DUSP5	0.37	0.21	0.16
Breast	19	47598365	47618174	ENST00000594526.1	SAE1	0.46	0.30	0.16
Breast	20	17537085	17560729	ENST00000449141.2	DSTN	0.27	0.12	0.16
Breast	17	38627225	38717027	ENST00000254051.6	TNS4	0.21	0.05	0.16
Breast	11	65678179	65686589	ENST00000530188.1	C11orf68	0.30	0.14	0.16
Breast	17	80831491	80848486	ENST00000572984.1	TBCD	0.21	0.05	0.16
Breast	1	27320120	27340181	ENST00000289166.5	FAM46B	0.33	0.18	0.16
Breast	1	19247868	19283180	ENST00000416166.1	IFFO2	0.29	0.14	0.15
Breast	10	121410843	121447500	ENST00000450186.1	BAG3	0.21	0.06	0.15
Breast	7	42257859	42278946	ENST00000428534.1	GLI3	0.33	0.17	0.15
Breast	7	869464	876882	ENST00000469755.1	SUN1	0.22	0.07	0.15
Breast	15	50644604	50649357	ENST00000560896.1	GABPB1	0.96	0.80	0.15
Breast	17	7736868	7749068	ENST00000570632.1	KDM6B	0.92	0.77	0.15
Breast	18	46449301	46483832	ENST00000586093.1	SMAD7	0.47	0.32	0.15
Breast	17	75274996	75288850	ENST00000587237.1	9-Sep	0.30	0.15	0.15
Breast	12	109225745	109251401	ENST00000548522.1	SSH1	0.24	0.09	0.15
Breast	12	52608969	52643707	ENST00000544024.1	KRT86	0.21	0.06	0.15
Breast	2	114081148	114089147	ENST00000429538.3	PAX8	0.30	0.16	0.15
Breast	9	124030410	124052489	ENST00000545652.1	GSN	0.40	0.26	0.15
Breast	17	75387047	75426957	ENST00000586521.1	9-Sep	0.22	0.08	0.15
Breast	11	61732568	61749363	ENST00000601917.1	AP003733.1	0.38	0.24	0.14
Breast	7	41734099	41745805	ENST00000442711.1	INHBA	0.40	0.26	0.14
Breast	6	138172161	138196002	ENST00000433680.1	TNFAIP3	0.29	0.14	0.14
Breast	1	16274460	16281222	ENST00000494020.1	ZBTB17	0.21	0.07	0.14
Breast	19	41220087	41229010	ENST00000263370.2	ITPKC	0.69	0.55	0.14
Breast	15	68548590	68582156	ENST00000566739.1	FEM1B	0.25	0.12	0.13
Breast	2	106009478	106028025	ENST00000447958.1	FHL2	0.21	0.08	0.13
Breast	10	74051120	74097380	ENST00000473051.1	DNAJB12	0.27	0.14	0.13
Breast	19	42771661	42788152	ENST00000575839.1	CIC	0.57	0.44	0.13
Breast	6	10398835	10421638	ENST00000486038.1	TFAP2A	0.81	0.68	0.13
Breast	11	57045039	57094324	ENST00000527207.1	TNKS1BP1	0.23	0.09	0.13
Breast	3	71104592	71120542	ENST00000497553.1	FOXP1	0.41	0.28	0.13
Breast	20	19954385	20001261	ENST00000481837.1	NAA20	0.22	0.09	0.13
Breast	19	13948994	13959464	ENST00000591727.1	NANOS3	0.68	0.56	0.13
Breast	16	23152633	23161835	ENST00000219689.7	USP31	0.37	0.24	0.12
Breast	1	16156416	16179578	ENST00000375759.3	SPEN	0.49	0.37	0.12
Breast	4	169752417	169771288	ENST00000507735.1	PALLD	0.28	0.16	0.12
Breast	3	9436165	9466747	ENST00000468208.1	SETD5	0.30	0.18	0.12
Breast	1	207221596	207227576	ENST00000367079.2	PFKFB2	0.61	0.48	0.12
Breast	2	43444985	43456496	ENST00000282388.3	ZFP36L2	1.00	0.88	0.12
Breast	1	33207559	33241755	ENST00000373480.1	KIAA1522	0.22	0.10	0.12
Breast	4	95366945	95386196	ENST00000511767.1	PDLIM5	0.24	0.11	0.12
Breast	7	36426266	36435508	ENST00000418118.1	ANLN	0.36	0.23	0.12
Breast	5	124075983	124085579	ENST00000512940.1	ZNF608	0.90	0.78	0.12
Breast	1	155097310	155110100	ENST00000484027.1	SLC50A1	0.63	0.51	0.12
Breast	1	201404474	201474025	ENST00000532460.1	CSRP1	0.22	0.10	0.12
Breast	18	3446318	3460082	ENST00000472042.1	TGIF1	0.68	0.56	0.12
Breast	6	136605683	136612552	ENST00000528229.1	BCLAF1	0.54	0.42	0.12
Breast	1	6637843	6670728	ENST00000496707.1	KLHL21	0.28	0.17	0.12
Breast	11	65185034	65198304	ENST00000531151.1	FRMD8	1.00	0.88	0.12
Breast	5	140887653	140909039	ENST00000468119.1	DIAPH1	0.23	0.11	0.12
Breast	10	106080462	106112294	ENST00000458723.1	ITPRIP	0.31	0.19	0.12
Breast	4	103736345	103751180	ENST00000503282.1	UBE2D3	0.48	0.36	0.11
Breast	5	2739262	2759651	ENST00000502957.1	IRX2	0.92	0.80	0.11
Breast	15	75930589	75963120	ENST00000569152.1	SNX33	0.27	0.16	0.11
Breast	2	87744763	87759323	ENST00000359481.4	PLGLB2	0.21	0.09	0.11
Breast	10	82213990	82238676	ENST00000372156.1	TSPAN14	0.25	0.14	0.11
Breast	1	144981791	144985538	ENST00000596396.1	AL590452.1	0.28	0.17	0.11
Breast	12	109082915	109097018	ENST00000547170.1	CORO1C	0.31	0.21	0.11
Breast	10	73723394	73770620	ENST00000373115.4	CHST3	0.25	0.15	0.11
Breast	12	125388867	125425378	ENST00000542416.1	UBC	0.39	0.28	0.11
Breast	1	27183345	27192842	ENST00000339276.4	SFN	0.33	0.22	0.11
Breast	17	17723596	17746027	ENST00000435530.2	SREBF1	0.36	0.25	0.11
Breast	14	69241271	69268143	ENST00000408913.2	ZFP36L1	0.42	0.32	0.10
Breast	10	33226394	33252825	ENST00000439974.3	ITGB1	0.30	0.19	0.10
Breast	16	15735275	15740014	ENST00000396353.2	NDE1	0.64	0.54	0.10
Breast	17	17578482	17616584	ENST00000471135.2	RAI1	0.26	0.15	0.10
Breast	11	12694725	12726400	ENST00000334310.6	TEAD1	0.27	0.17	0.10
Breast	1	8062975	8088860	ENST00000487559.1	ERRFI1	0.25	0.14	0.10
Breast	9	130280649	130347137	ENST00000373314.3	FAM129B	0.20	0.10	0.10
Breast	10	74002370	74038384	ENST00000471240.1	DDIT4	0.27	0.17	0.10
Breast	17	27134005	27140821	ENST00000582059.1	FAM222B	0.45	0.35	0.10
Breast	7	115849775	115868787	ENST00000485009.1	TES	0.21	0.11	0.10
Breast	2	46523740	46563334	ENST00000475822.1	EPAS1	0.23	0.13	0.10
CD19	19	13204800	13217457	ENST00000590120.1	LYL1	0.94	0.34	0.60
CD19	14	106313303	106330974	ENST00000482999.1	KIAA0125	0.51	0.00	0.51
CD19	17	73289760	73295202	ENST00000580273.1	SLC25A19	0.51	0.00	0.51
CD19	8	134509931	134514822	ENST00000520020.1	ST3GAL1	0.60	0.12	0.48
CD19	12	57869667	57876031	ENST00000552249.1	ARHGAP9	0.56	0.14	0.42
CD19	2	232255961	232260646	ENST00000287590.5	B3GNT7	0.54	0.13	0.41
CD19	22	50627294	50632183	ENST00000395829.1	TRABD	0.69	0.28	0.41
CD19	2	182168501	182179942	ENST00000476089.1	ITGA4	0.40	0.00	0.40
CD19	6	31547188	31556914	ENST00000464044.1	LST1	0.57	0.18	0.40
CD19	17	43299369	43312992	ENST00000591434.1	FMNL1	0.55	0.16	0.39
CD19	1	26609716	26620076	ENST00000319041.6	SH3BGRL3	0.40	0.03	0.37
CD19	17	47814002	47827686	ENST00000515240.1	FAM117A	0.35	0.00	0.35
CD19	19	16693911	16706114	ENST00000596492.1	MED26	0.52	0.17	0.35
CD19	7	50344148	50370755	ENST00000426121.1	IKZF1	0.41	0.07	0.35
CD19	4	26883290	26890121	ENST00000478049.1	STIM2	0.34	0.00	0.34
CD19	2	197013308	197048452	ENST00000449152.1	STK17B	0.42	0.07	0.34
CD19	21	45660565	45668530	ENST00000400377.3	ICOSLG	0.72	0.38	0.34
CD19	14	64700800	64703816	ENST00000554928.1	SYNE2	0.42	0.08	0.34
CD19	9	71251305	71262079	ENST00000541509.1	PIP5K1B	0.33	0.00	0.33
CD19	9	134593692	134613095	ENST00000372189.3	RAPGEF1	0.40	0.08	0.32
CD19	11	9708104	9716296	ENST00000524817.1	SWAP70	0.39	0.07	0.32
CD19	20	42307887	42312191	ENST00000217026.4	MYBL2	0.32	0.00	0.32
CD19	15	81576278	81610588	ENST00000559953.1	IL16	0.33	0.01	0.32
CD19	11	58340596	58348101	ENST00000389919.4	ZFP91-CNTF	0.83	0.52	0.31
CD19	17	56394652	56423683	ENST00000583624.1	BZRAP1	0.46	0.16	0.31
CD19	8	56751994	56758932	ENST00000519728.1	LYN	0.54	0.24	0.30
CD19	1	111401311	111424334	ENST00000476408.1	CD53	0.29	0.00	0.29
CD19	14	61937696	61947119	ENST00000536400.1	PRKCH	0.33	0.04	0.29
CD19	1	167584400	167603373	ENST00000537350.1	RCSD1	0.37	0.08	0.29
CD19	2	31436778	31445756	ENST00000403897.3	CAPN14	0.28	0.00	0.28
CD19	5	133262873	133277639	ENST00000509913.1	C5orf15	0.28	0.00	0.28
CD19	7	86840957	86850205	ENST00000455575.1	TMEM243	0.59	0.31	0.28
CD19	19	35694706	35707859	ENST00000324675.3	FAM187B	0.34	0.06	0.28
CD19	6	13258837	13278907	ENST00000481706.1	PHACTR1	0.29	0.01	0.28
CD19	18	77219226	77227908	ENST00000591089.1	NFATC1	0.29	0.01	0.28
CD19	19	7400756	7417433	ENST00000576789.1	CTB-133G6.1	0.32	0.05	0.27
CD19	11	117872360	117887673	ENST00000525467.1	IL10RA	0.28	0.01	0.27
CD19	9	112727706	112739040	ENST00000434623.2	AKAP2	0.27	0.00	0.27
CD19	14	105524940	105537562	ENST00000546679.1	GPR132	0.32	0.05	0.27
CD19	8	101505352	101515105	ENST00000519316.1	ANKRD46	0.31	0.04	0.27
CD19	19	49827519	49845306	ENST00000600121.1	CD37	0.38	0.12	0.27
CD19	1	53096945	53108122	ENST00000424164.1	FAM159A	0.35	0.08	0.27
CD19	14	106282350	106287664	ENST00000482999.1	KIAA0125	0.27	0.00	0.27
CD19	15	75066062	75093388	ENST00000569321.1	CSK	0.51	0.25	0.27
CD19	1	31201550	31236569	ENST00000476492.1	LAPTM5	0.30	0.04	0.26
CD19	10	11184898	11223017	ENST00000537122.1	CELF2	0.32	0.06	0.26
CD19	6	32402294	32411438	ENST00000374982.5	HLA-DRA	0.36	0.10	0.26
CD19	15	86229664	86252170	ENST00000560482.1	AKAP13	0.31	0.05	0.26
CD19	8	11340662	11367244	ENST00000529894.1	BLK	0.26	0.00	0.26
CD19	11	35097683	35106244	ENST00000598940.1	AL356215.1	0.31	0.05	0.26
CD19	4	40174339	40216705	ENST00000503978.1	RHOH	0.30	0.04	0.26
CD19	19	42374270	42392833	ENST00000337665.4	ARHGEF1	0.44	0.19	0.25
CD19	8	56790857	56809092	ENST00000520220.1	LYN	0.35	0.10	0.25
CD19	11	61120450	61130574	ENST00000423772.2	TMEM138	0.45	0.20	0.25
CD19	10	49874683	49897623	ENST00000360890.2	WDFY4	0.25	0.00	0.25
CD19	22	24821855	24835079	ENST00000439591.1	ADORA2A	0.26	0.02	0.25
CD19	16	29346657	29372711	ENST00000356328.3	SNX29P2	0.24	0.00	0.24
CD19	17	3807403	3821299	ENST00000571637.1	P2RX1	0.34	0.10	0.24
CD19	19	35807029	35840479	ENST00000593704.1	CD22	0.26	0.02	0.24
CD19	9	134127597	134154499	ENST00000464831.1	FAM78A	0.35	0.11	0.24
CD19	6	149785200	149817360	ENST00000462655.1	ZC3H12D	0.28	0.04	0.24
CD19	11	128582503	128610944	ENST00000534087.1	FLI1	0.28	0.04	0.24
CD19	1	21618391	21627146	ENST00000527991.1	ECE1	0.26	0.02	0.24
CD19	11	72850513	72872020	ENST00000422375.1	FCHSD2	0.41	0.17	0.24
CD19	16	81803418	81877366	ENST00000569523.1	PLCG2	0.26	0.02	0.24
CD19	2	233923908	233972771	ENST00000467393.1	INPP5D	0.29	0.05	0.24
CD19	16	85921635	85951991	ENST00000569607.1	IRF8	0.29	0.05	0.24
CD19	17	74476430	74497452	ENST00000590288.1	RHBDF2	0.36	0.13	0.23
CD19	20	4789416	4805261	ENST00000379400.3	RASSF2	0.41	0.18	0.23
CD19	22	47159882	47174367	ENST00000406733.1	TBC1D22A	0.27	0.03	0.23
CD19	5	88114708	88142027	ENST00000509373.1	MEF2C	0.24	0.01	0.23
CD19	20	46038978	46050263	ENST00000262975.4	ZMYND8	0.24	0.01	0.23
CD19	19	2610475	2632474	ENST00000587867.1	GNG7	0.31	0.09	0.22
CD19	22	42304034	42337858	ENST00000472374.2	CENPM	0.44	0.22	0.22
CD19	6	32807846	32814107	ENST00000414474.1	PSMB9	0.52	0.30	0.22
CD19	5	1479672	1506754	ENST00000507282.1	LPCAT1	0.23	0.01	0.22
CD19	6	150937029	150965213	ENST00000367326.1	PLEKHG1	0.25	0.03	0.22
CD19	18	2956157	2983444	ENST00000581568.1	LPIN2	0.23	0.02	0.22
CD19	6	159459555	159485569	ENST00000338313.5	TAGAP	0.27	0.05	0.22
CD19	7	35764330	35771796	ENST00000396081.1	HERPUD2	0.22	0.00	0.22
CD19	7	35764330	35771796	ENST00000311350.3	HERPUD2	0.22	0.00	0.22
CD19	11	118739056	118768630	ENST00000292174.4	CXCR5	0.25	0.04	0.22
CD19	7	150100445	150107156	ENST00000474836.1	ZNF775	0.56	0.35	0.21
CD19	19	10712789	10716244	ENST00000407327.4	SLC44A2	0.80	0.58	0.21
CD19	16	21367569	21393569	ENST00000542817.1	NPIPL3	0.21	0.00	0.21
CD19	4	56812204	56816635	ENST00000257287.4	CEP135	0.73	0.52	0.21
CD19	17	38004143	38027814	ENST00000377940.3	ZPBP2	0.26	0.05	0.21
CD19	20	43204843	43247481	ENST00000349959.3	PKIG	0.24	0.03	0.21
CD19	14	89768030	89775675	ENST00000555658.1	FOXN3	0.30	0.09	0.21
CD19	14	69147805	69153516	ENST00000555997.1	ZFP36L1	0.23	0.03	0.21
CD19	3	5047573	5059145	ENST00000460806.1	BHLHE40	0.23	0.02	0.20
CD19	11	58967706	58992697	ENST00000361050.3	MPEG1	0.22	0.02	0.20
CD19	16	10959821	11016093	ENST00000572665.1	CIITA	0.28	0.07	0.20
CD19	1	150122093	150138999	ENST00000485470.1	PLEKHO1	0.34	0.13	0.20
CD19	20	55964442	55977216	ENST00000371219.2	RBM38	0.53	0.33	0.20
CD19	3	196350844	196371448	ENST00000426755.1	PIGX	0.34	0.14	0.20
CD19	17	45276453	45285339	ENST00000571981.1	MYL4	0.23	0.04	0.20
CD19	15	57571264	57599710	ENST00000560948.1	TCF12	0.24	0.04	0.19
CD19	22	18476769	18494405	ENST00000424046.1	MICAL3	0.32	0.12	0.19
CD19	1	206730430	206756416	ENST00000304534.8	RASSF5	0.25	0.05	0.19
CD19	15	74666045	74697205	ENST00000268053.6	CYP11A1	0.31	0.12	0.19
CD19	11	122930787	122942297	ENST00000531063.1	HSPA8	0.53	0.34	0.19
CD19	16	85965145	85983944	ENST00000569607.1	IRF8	0.21	0.02	0.19
CD19	9	126959034	126982108	ENST00000373603.1	NEK6	0.20	0.02	0.19
CD19	9	126959034	126982108	ENST00000540326.1	NEK6	0.20	0.02	0.19
CD19	18	60804706	60831957	ENST00000590515.1	BCL2	0.25	0.07	0.19
CD19	17	61992715	62032882	ENST00000584310.1	SCN4A	0.22	0.04	0.19
CD19	2	136871650	136898787	ENST00000466288.1	CXCR4	0.33	0.14	0.18
CD19	1	27984903	27990153	ENST00000339145.4	IFI6	0.51	0.32	0.18
CD19	22	37613897	37642449	ENST00000401529.3	RAC2	0.28	0.09	0.18
CD19	1	207078366	207106998	ENST00000487208.1	PIGR	0.23	0.05	0.18
CD19	10	11703130	11728646	ENST00000379237.1	USP6NL	0.23	0.05	0.18
CD19	6	32904722	32916498	ENST00000416244.2	HLA-DMB	0.27	0.09	0.18
CD19	14	64966362	64978672	ENST00000555321.1	ZBTB1	0.55	0.37	0.18
CD19	2	71290385	71301563	ENST00000524537.1	NAGK	0.39	0.21	0.18
CD19	3	58318904	58343543	ENST00000491164.1	PXK	0.25	0.07	0.18
CD19	5	163334703	163344988	ENST00000521838.1	MAT2B	0.20	0.03	0.18
CD19	6	41671523	41703442	ENST00000433032.1	TFEB	0.33	0.15	0.18
CD19	3	10232905	10242463	ENST00000256458.4	IRAK2	0.33	0.16	0.18
CD19	1	172348783	172369092	ENST00000465374.1	PIGC	0.21	0.03	0.17
CD19	2	128144112	128148979	ENST00000409179.2	MAP3K2	0.62	0.45	0.17
CD19	21	44817042	44836022	ENST00000478426.1	SIK1	0.22	0.05	0.17
CD19	12	48196588	48213374	ENST00000445237.2	HDAC7	0.40	0.23	0.17
CD19	11	67034303	67056724	ENST00000447274.2	ANKRD13D	0.40	0.23	0.17
CD19	22	22510876	22526241	ENST00000403807.3	VPREB1	0.23	0.07	0.17
CD19	9	139115225	139137596	ENST00000455222.1	QSOX2	0.22	0.05	0.17
CD19	7	138778458	138805019	ENST00000471652.1	ZC3HAV1	0.29	0.12	0.17
CD19	14	106023918	106072429	ENST00000431372.1	TMEM121	0.21	0.04	0.17
CD19	21	45556978	45601335	ENST00000598764.1	AP001055.1	0.28	0.12	0.17
CD19	3	122291690	122308644	ENST00000483793.1	PARP15	0.20	0.04	0.16
CD19	18	74764261	74781577	ENST00000580473.1	MBP	0.29	0.13	0.16
CD19	16	84618823	84644362	ENST00000561707.1	COTL1	0.20	0.04	0.16
CD19	9	35616119	35651731	ENST00000474403.1	SIT1	0.23	0.07	0.16
CD19	1	38937882	38948718	ENST00000474456.1	RRAGC	0.22	0.06	0.16
CD19	6	143160621	143189869	ENST00000367604.1	HIVEP2	0.22	0.06	0.16
CD19	6	143160621	143189869	ENST00000012134.2	HIVEP2	0.22	0.06	0.16
CD19	6	143160621	143189869	ENST00000474532.1	HIVEP2	0.22	0.06	0.16
CD19	3	107841403	107856868	ENST00000468021.1	IFT57	0.22	0.06	0.16
CD19	6	167362856	167374276	ENST00000496851.2	RNASET2	0.36	0.20	0.16
CD19	9	130665468	130681316	ENST00000479747.1	ST6GALNAC4	0.33	0.18	0.16
CD19	19	50917413	50938549	ENST00000357701.5	MYBPC2	0.42	0.26	0.16
CD19	13	46942847	46975976	ENST00000480935.1	KIAA0226L	0.24	0.08	0.16
CD19	2	238580170	238610065	ENST00000473815.1	LRRFIP1	0.35	0.19	0.16
CD19	12	92527263	92581890	ENST00000552315.1	BTG1	0.30	0.14	0.15
CD19	3	13034134	13065411	ENST00000473088.1	IQSEC1	0.30	0.15	0.15
CD19	11	73076425	73101853	ENST00000393580.2	RELT	0.21	0.06	0.15
CD19	11	82763568	82785588	ENST00000533276.1	RAB30	0.31	0.16	0.15
CD19	2	46761453	46803094	ENST00000473428.1	RHOQ	0.24	0.09	0.15
CD19	14	23018284	23039986	ENST00000557595.1	AE000662.92	0.25	0.10	0.15
CD19	12	7044970	7072648	ENST00000537533.1	PTPN6	0.36	0.21	0.15
CD19	4	82389749	82394691	ENST00000507538.1	RASGEF1B	0.74	0.59	0.14
CD19	12	113486672	113531373	ENST00000553140.1	DTX1	0.23	0.09	0.14
CD19	4	185452534	185460044	ENST00000393593.3	IRF2	0.39	0.25	0.14
CD19	12	9788063	9838708	ENST00000544322.1	CLEC2D	0.20	0.06	0.14
CD19	4	40289700	40319643	ENST00000310169.2	CHRNA9	0.21	0.06	0.14
CD19	17	38473833	38487686	ENST00000582914.1	RARA	0.58	0.44	0.14
CD19	7	142490667	142511468	ENST00000463701.1	PRSS1	0.26	0.12	0.14
CD19	1	111734737	111768995	ENST00000445067.2	CHI3L2	0.20	0.06	0.14
CD19	6	149077573	149093955	ENST00000367463.4	UST	0.21	0.07	0.14
CD19	2	231729691	231744201	ENST00000492029.1	ITM2C	0.33	0.19	0.14
CD19	6	144654443	144673365	ENST00000421035.1	UTRN	0.20	0.07	0.14
CD19	16	15728212	15738278	ENST00000396353.2	NDE1	0.42	0.28	0.14
CD19	10	90144149	90152479	ENST00000437752.1	RNLS	0.29	0.15	0.14
CD19	1	40838755	40871047	ENST00000539317.1	SMAP2	0.22	0.08	0.14
CD19	17	27064086	27073896	ENST00000584944.1	TRAF4	0.48	0.35	0.13
CD19	19	2040263	2096653	ENST00000395307.2	IZUMO4	0.32	0.19	0.13
CD19	9	3524176	3529109	ENST00000449190.1	RFX3	0.90	0.77	0.13
CD19	17	29814146	29838313	ENST00000578694.1	RAB11FIP4	0.24	0.11	0.13

CD19	13	41534432	41594336	ENST00000239882.3	ELF1	0.21	0.08	0.13
CD19	19	18278723	18292546	ENST00000600463.1	IFI30	0.23	0.10	0.13
CD19	6	106957303	106996547	ENST00000487681.1	AIM1	0.26	0.13	0.13
CD19	2	196511447	196525260	ENST00000418005.1	SLC39A10	0.39	0.26	0.13
CD19	17	75396031	75469717	ENST00000593189.1	9-Sep	0.28	0.15	0.12
CD19	14	50327610	50337580	ENST00000298310.5	NEMF	0.50	0.38	0.12
CD19	12	46657445	46665336	ENST00000546519.1	SLC38A1	0.57	0.45	0.12
CD19	16	11758829	11785379	ENST00000575349.1	TXNDC11	0.22	0.10	0.12
CD19	1	154914194	154929302	ENST00000490230.1	PBXIP1	0.29	0.16	0.12
CD19	1	2477758	2482788	ENST00000426449.1	TNFRSF14	0.72	0.59	0.12
CD19	2	54783925	54824402	ENST00000333896.5	SPTBN1	0.27	0.15	0.12
CD19	11	9621558	9642085	ENST00000527848.1	WEE1	0.21	0.09	0.12
CD19	15	45001549	45023195	ENST00000561237.1	TRIM69	0.36	0.24	0.12
CD19	20	44736001	44751095	ENST00000461171.1	CD40	0.21	0.09	0.12
CD19	12	110431794	110453837	ENST00000261739.4	ANKRD13A	0.28	0.16	0.12
CD19	2	202106787	202128005	ENST00000429881.1	CASP8	0.21	0.10	0.11
CD19	8	29937591	29951059	ENST00000521083.1	TMEM66	0.29	0.18	0.11
CD19	22	31680149	31688920	ENST00000443175.1	PIK3IP1	0.44	0.33	0.11
CD19	9	123686905	123707487	ENST00000540010.1	TRAF1	0.30	0.19	0.11
CD19	10	112618797	112633729	ENST00000444997.1	PDCD4	0.35	0.25	0.11
CD19	16	28921620	28948628	ENST00000567368.1	CD19	0.20	0.10	0.11
CD19	14	105116134	105145145	ENST00000330634.7	INF2	0.23	0.12	0.11
CD19	5	122109910	122117182	ENST00000514949.1	SNX2	0.41	0.31	0.11
CD19	18	2634081	2658907	ENST00000261598.8	SMCHD1	0.25	0.15	0.10
CD19	19	39887963	39904206	ENST00000438123.1	PLEKHG2	0.70	0.60	0.10
CD19	1	66795175	66818506	ENST00000526197.1	PDE4B	0.28	0.18	0.10
Colon	20	43968274	43977802	ENST00000537976.1	SDC4	0.77	0.29	0.48
Colon	1	2505891	2510796	ENST00000493183.1	FAM213B	0.61	0.15	0.46
Colon	17	57904024	57931778	ENST00000587470.1	VMP1	0.49	0.14	0.35
Colon	6	74224863	74234116	ENST00000455918.1	EEF1A1	0.90	0.58	0.32
Colon	7	27210777	27221249	ENST00000396344.4	HOXA10	0.76	0.45	0.30
Colon	17	37780840	37791262	ENST00000580029.1	PPP1R1B	0.49	0.19	0.30
Colon	10	3846234	3854755	ENST00000542957.1	KLF6	0.42	0.13	0.29
Colon	1	19966426	19978402	ENST00000427894.1	NBL1	0.54	0.25	0.28
Colon	15	75072089	75083506	ENST00000567571.1	CSK	0.80	0.52	0.28
Colon	5	79542438	79554102	ENST00000513907.1	SERINC5	0.51	0.26	0.25
Colon	17	76164003	76173458	ENST00000592456.1	SYNGR2	0.53	0.29	0.24
Colon	21	42931018	42955389	ENST00000454499.1	TMPRSS2	0.24	0.01	0.23
Colon	1	207111967	207121081	ENST00000491503.1	PIGR	0.23	0.00	0.23
Colon	13	28526778	28556061	ENST00000548877.1	CDX2	0.94	0.72	0.23
Colon	15	86121044	86133400	ENST00000560340.1	AKAP13	0.40	0.18	0.22
Colon	18	3446352	3459783	ENST00000472042.1	TGIF1	0.77	0.55	0.22
Colon	14	50465553	50471375	ENST00000529902.1	C14orf182	0.66	0.45	0.22
Colon	16	29830344	29838029	ENST00000570234.1	MVP	0.48	0.28	0.19
Colon	3	193848397	193860794	ENST00000476918.1	HES1	0.80	0.61	0.19
Colon	12	7051408	7057319	ENST00000538318.1	PTPN6	0.59	0.40	0.18
Colon	1	159888888	159896553	ENST00000397334.2	TAGLN2	0.63	0.45	0.18
Colon	17	7161661	7168259	ENST00000573745.1	CLDN7	0.63	0.45	0.18
Colon	6	31700974	31708664	ENST00000493662.2	MSH5-SAPCD1	0.87	0.70	0.18
Colon	20	52195018	52241000	ENST00000540425.1	ZNF217	0.41	0.24	0.17
Colon	3	49056900	49060957	ENST00000480392.1	NDUFAF3	0.84	0.67	0.17
Colon	12	56472347	56482109	ENST00000546748.1	ERBB3	0.44	0.27	0.17
Colon	11	130013608	130061419	ENST00000530376.1	ST14	0.22	0.05	0.16
Colon	2	200306225	200342782	ENST00000463386.1	SATB2	0.56	0.40	0.16
Colon	16	2028149	2036628	ENST00000567719.1	GFER	0.59	0.43	0.16
Colon	4	38662929	38689928	ENST00000436901.1	AC021860.1	0.45	0.29	0.16
Colon	8	126438172	126449618	ENST00000519576.1	TRIB1	0.52	0.36	0.16
Colon	1	1092807	1104591	ENST00000506177.1	TTLL10	0.35	0.19	0.16
Colon	19	49463688	49472130	ENST00000331825.6	FTL	0.54	0.39	0.15
Colon	1	37936771	37953471	ENST00000471012.1	ZC3H12A	0.49	0.34	0.15
Colon	7	150052314	150083105	ENST00000486297.1	ZNF775	0.36	0.21	0.15
Colon	2	173291273	173330716	ENST00000409080.1	ITGA6	0.25	0.10	0.15
Colon	19	13948788	13956320	ENST00000591727.1	NANOS3	0.69	0.55	0.15
Colon	11	307666	312937	ENST00000399815.2	IFITM2	0.70	0.55	0.14
Colon	11	118659296	118664159	ENST00000533239.1	DDX6	1.00	0.86	0.14
Colon	10	134259441	134282106	ENST00000392630.3	C10orf91	0.26	0.12	0.14
Colon	10	134259441	134282106	ENST00000321248.2	C10orf91	0.26	0.12	0.14
Colon	4	102266379	102270939	ENST00000529296.1	AP001816.1	0.94	0.80	0.14
Colon	1	150531037	150543263	ENST00000369035.2	C1orf138	0.51	0.37	0.14
Colon	17	27275334	27280852	ENST00000577182.1	PIPOX	1.00	0.86	0.14
Colon	1	149819766	149826428	ENST00000403683.1	HIST2H3A	0.28	0.14	0.14
Colon	11	71748625	71754435	ENST00000535947.1	NUMA1	0.81	0.67	0.13
Colon	5	96268784	96273156	ENST00000231368.5	LNPEP	0.95	0.81	0.13
Colon	19	1256417	1263739	ENST00000589161.1	CIRBP	0.65	0.52	0.13
Colon	11	64654606	64662582	ENST00000457202.1	EHD1	0.23	0.09	0.13
Colon	2	201980044	201998492	ENST00000460961.1	CFLAR	0.57	0.44	0.13
Colon	3	9436663	9444883	ENST00000406341.1	SETD5	0.78	0.65	0.13
Colon	17	4845115	4854686	ENST00000519300.1	ENO3	0.72	0.58	0.13
Colon	1	153503987	153511480	ENST00000462951.2	S100A6	0.39	0.26	0.13
Colon	10	90638872	90664402	ENST00000371924.1	STAMBPL1	0.20	0.07	0.13
Colon	9	140187219	140215940	ENST00000356628.2	NRARP	0.53	0.40	0.13
Colon	5	180668526	180674519	ENST00000514318.1	GNB2L1	0.82	0.70	0.13
Colon	2	174828021	174831622	ENST00000490182.1	SP3	1.00	0.88	0.12
Colon	1	1675407	1679081	ENST00000246421.4	SLC35E2	0.74	0.62	0.12
Colon	2	220109452	220119946	ENST00000392088.2	TUBA4A	0.53	0.40	0.12
Colon	12	53762866	53777584	ENST00000548560.1	SP1	0.41	0.29	0.12
Colon	7	1486269	1507327	ENST00000297508.7	MICALL2	0.35	0.23	0.12
Colon	19	1852750	1865843	ENST00000592313.1	KLF16	0.63	0.51	0.12
Colon	6	33538390	33559780	ENST00000374458.1	GGNBP1	0.27	0.16	0.12
Colon	4	40185249	40208107	ENST00000503978.1	RHOH	0.25	0.13	0.12
Colon	7	104643451	104656449	ENST00000474203.1	MLL5	0.62	0.51	0.12
Colon	8	103816742	103826153	ENST00000518697.1	AZIN1	0.54	0.43	0.12
Colon	2	151323963	151344439	ENST00000454202.1	RND3	0.28	0.17	0.11
Colon	1	169069278	169085972	ENST00000367813.3	ATP1B1	0.45	0.34	0.11
Colon	17	38213719	38232994	ENST00000577486.1	THRA	0.55	0.44	0.11
Colon	12	6559071	6563067	ENST00000543567.1	TAPBPL	0.75	0.64	0.11
Colon	17	4033794	4049258	ENST00000573984.1	CYB5D2	0.27	0.16	0.11
Colon	5	179244886	179250297	ENST00000360718.5	SQSTM1	0.68	0.57	0.11
Colon	7	100608082	100627662	ENST00000536621.1	MUC12	0.30	0.20	0.11
Colon	17	7459695	7466311	ENST00000429205.2	SENP3	0.74	0.63	0.11
Colon	7	27202748	27210249	ENST00000489695.1	HOXA9	0.95	0.85	0.10
Colon	6	32934747	32951807	ENST00000482838.1	BRD2	0.61	0.51	0.10
Colon	3	50373024	50380264	ENST00000490675.1	ZMYND10	0.49	0.39	0.10
Lung	20	23060957	23069729	ENST00000246006.4	CD93	0.23	0.01	0.21
Lung	17	7378645	7384073	ENST00000380599.4	ZBTB4	0.65	0.45	0.20
Lung	12	6441703	6453465	ENST00000538363.1	TNFRSF1A	0.61	0.45	0.16
Lung	2	66659175	66674430	ENST00000475239.1	MEIS1	0.78	0.63	0.16
Lung	22	46464333	46478321	ENST00000443490.1	MIRLET7	0.85	0.71	0.15
Lung	20	22536539	22565230	ENST00000319993.4	FOXA2	0.69	0.54	0.15
Lung	1	145434374	145443479	ENST00000486597.1	TXNIP	0.71	0.57	0.15
Lung	12	92526891	92540701	ENST00000552315.1	BTG1	0.71	0.57	0.14
Lung	14	105938037	105944737	ENST00000548309.1	CRIP2	0.80	0.66	0.14
Lung	5	43033834	43045550	ENST00000314890.3	ANXA2R	0.57	0.43	0.14
Lung	4	174427206	174460857	ENST00000505300.1	HAND2	0.74	0.60	0.14
Lung	9	35726029	35733751	ENST00000486056.1	CREB3	0.65	0.53	0.12
Lung	1	154939647	154948901	ENST00000473344.1	CKS1B	0.61	0.48	0.12
Lung	19	45244285	45262369	ENST00000473473.1	BCL3	0.46	0.34	0.12
Lung	12	119613625	119628282	ENST00000542496.1	HSPB8	0.22	0.10	0.12
Lung	1	2160113	2167961	ENST00000508416.1	SKI	0.66	0.55	0.12
Lung	16	56638742	56646907	ENST00000567300.1	MT2A	0.50	0.39	0.12
Lung	11	67803279	67809641	ENST00000533947.1	TCIRG1	0.45	0.34	0.11
Lung	19	13948720	13956318	ENST00000591727.1	NANOS3	0.66	0.55	0.11
Lung	11	305282	312843	ENST00000399815.2	IFITM2	0.50	0.40	0.10

Supplementary Table 6.2: Validation of tissue-specific DNA methylation at super-enhancers in normal tissue types.

	Super-Enhancer					HMR coverage (%)			DNA Methylation 450K		Significance	
Tissue	Chr.	Start	End	Transcript ID	Gene	Tissue	Others	Diff.	AVR Tissue	AVR Others	-log10(p-value)	Validated
Brain	10	126400866	126432679	ENST00000392754.3	FAM53B	0.40	0.19	0.21	0.27	0.32	8.42	TRUE
Brain	11	67173218	67188540	ENST00000531388.1	CARNS1	0.45	0.21	0.24	0.19	0.25	5.85	TRUE
Brain	1	205195003	205257756	ENST00000481950.1	TMCC2	0.23	0.09	0.14	0.23	0.27	4.99	TRUE
Brain	17	72226877	72248627	ENST00000526858.1	TTYH2	0.23	0.03	0.20	0.28	0.45	4.55	TRUE
Brain	1	205626881	205650899	ENST00000367145.3	SLC45A3	0.29	0.14	0.15	0.06	0.07	4.38	TRUE
Brain	22	37940770	37968154	ENST00000434728.1	CDC42EP1	0.50	0.25	0.25	0.20	0.26	4.33	TRUE
Brain	7	124359038	124407316	ENST00000303921.2	GPR37	0.51	0.05	0.46	0.23	0.30	4.25	TRUE
Brain	1	221879013	221917508	ENST00000477026.1	DUSP10	0.32	0.09	0.23	0.16	0.19	3.98	TRUE
Brain	17	27071068	27095625	ENST00000584059.1	FAM222B	0.29	0.14	0.15	0.12	0.16	3.98	TRUE
Brain	7	28724048	28800494	ENST00000468813.1	CREB5	0.27	0.04	0.23	0.26	0.35	3.87	TRUE
Brain	14	65146034	65196928	ENST00000556801.1	PLEKHG3	0.31	0.11	0.20	0.20	0.26	3.83	TRUE
Brain	19	51033029	51070726	ENST00000599957.1	LRRC4B	0.30	0.17	0.14	0.32	0.43	3.80	TRUE
Brain	6	41649397	41702285	ENST00000433032.1	TFEB	0.30	0.10	0.20	0.30	0.38	3.78	TRUE
Brain	21	48018786	48058971	ENST00000440086.1	PRMT2	0.35	0.09	0.26	0.27	0.32	3.64	TRUE
Brain	1	149967151	149985599	ENST00000417191.1	PTUD7B	0.27	0.11	0.16	0.15	0.17	3.55	TRUE
Brain	11	73357627	73375282	ENST00000544282.1	PLEKHB1	0.20	0.08	0.13	0.18	0.21	3.31	TRUE
Brain	10	81137824	81212605	ENST00000372333.3	ZCCHC24	0.43	0.09	0.33	0.23	0.27	3.26	TRUE
Brain	12	49579740	49630368	ENST00000552125.1	TUBA1C	0.23	0.09	0.13	0.18	0.24	3.22	TRUE
Brain	1	109889550	109945440	ENST00000471996.1	SORT1	0.26	0.05	0.20	0.23	0.30	3.11	TRUE
Brain	6	30844750	30864481	ENST00000514434.1	DDR1	0.45	0.18	0.27	0.28	0.37	3.10	TRUE
Brain	11	118463235	118534958	ENST00000528823.1	PHLDB1	0.23	0.10	0.13	0.26	0.31	3.09	TRUE
Brain	17	57405868	57485233	ENST00000582813.1	YPEL2	0.20	0.06	0.14	0.15	0.19	3.03	TRUE
Brain	1	214156400	214205391	ENST00000261454.4	PROX1	0.32	0.13	0.19	0.22	0.25	3.01	TRUE
Brain	4	77099124	77138142	ENST00000512895.1	FAM47E	0.30	0.14	0.16	0.16	0.19	3.01	TRUE
Brain	1	202090299	202115217	ENST00000272217.2	ARL8A	0.39	0.17	0.23	0.25	0.35	2.95	TRUE
Brain	7	141373362	141403139	ENST00000482493.1	KIAA1147	0.22	0.06	0.16	0.25	0.29	2.93	TRUE
Brain	13	67782754	67806323	ENST00000377861.3	PCDH9	0.59	0.20	0.39	0.18	0.23	2.91	TRUE
Brain	2	232526039	232554844	ENST00000466801.1	PTMA	0.34	0.23	0.12	0.18	0.20	2.87	TRUE
Brain	17	76857381	76931557	ENST00000262768.7	TIMP2	0.23	0.09	0.14	0.33	0.39	2.79	TRUE
Brain	1	160019557	160042341	ENST00000368089.3	KCNJ10	0.23	0.07	0.16	0.19	0.22	2.72	TRUE
Brain	5	173313778	173354104	ENST00000519467.1	CPEB4	0.34	0.08	0.25	0.17	0.23	2.70	TRUE
Brain	3	156845617	156856684	ENST00000467849.1	CCNL1	0.30	0.15	0.15	0.23	0.30	2.70	TRUE
Brain	16	57277838	57336233	ENST00000569059.1	PLLP	0.30	0.10	0.20	0.13	0.16	2.70	TRUE
Brain	10	111967550	111992938	ENST00000369612.1	MXI1	0.40	0.29	0.11	0.15	0.16	2.68	TRUE
Brain	2	164566271	164594418	ENST00000482917.1	FIGN	0.35	0.11	0.25	0.20	0.25	2.63	TRUE

Brain	15	45655307	45688573	ENST00000558118.1	GATM	0.24	0.05	0.19	0.20	0.28	2.62	TRUE
Brain	1	204796255	204866707	ENST00000514644.1	NFASC	0.21	0.03	0.18	0.30	0.37	2.60	TRUE
Brain	2	127806748	127912517	ENST00000409400.1	BIN1	0.36	0.06	0.30	0.21	0.26	2.53	TRUE
Brain	6	134488737	134505092	ENST00000525700.1	SGK1	0.60	0.27	0.34	0.24	0.29	2.51	TRUE
Brain	5	139012918	139146609	ENST00000505812.1	CXXC5	0.36	0.22	0.14	0.23	0.26	2.43	TRUE
Brain	9	140081775	140094696	ENST00000477345.1	TPRN	0.56	0.20	0.35	0.25	0.29	2.39	TRUE
Brain	14	77765872	77789479	ENST00000555093.1	GSTZ1	0.39	0.12	0.27	0.19	0.22	2.37	TRUE
Brain	5	646895	697805	ENST00000360578.5	TPPP	0.23	0.12	0.12	0.27	0.32	2.34	TRUE
Brain	3	134072004	134096230	ENST00000502491.1	AMOTL2	0.44	0.19	0.25	0.19	0.22	2.33	TRUE
Brain	17	40114067	40162277	ENST00000591153.1	DNAJC7	0.27	0.07	0.20	0.33	0.38	2.32	TRUE
Brain	9	130697475	130743184	ENST00000494606.1	FAM102A	0.61	0.17	0.44	0.27	0.32	2.32	TRUE
Brain	1	230932874	231004412	ENST00000366663.5	C1orf198	0.21	0.03	0.18	0.29	0.34	2.31	TRUE
Brain	10	72971329	73044360	ENST00000373192.4	UNC5B	0.21	0.09	0.12	0.20	0.23	2.30	TRUE
Brain	1	160047565	160068437	ENST00000460351.1	IGSF8	0.35	0.14	0.21	0.29	0.31	2.25	TRUE
Brain	12	26265040	26280746	ENST00000534829.1	SSPN	0.64	0.47	0.17	0.14	0.16	2.12	TRUE
Brain	1	109816196	109827024	ENST00000471740.1	PSRC1	0.36	0.17	0.19	0.16	0.18	2.11	TRUE
Brain	19	13093539	13211813	ENST00000588680.1	NFIX	0.47	0.21	0.26	0.29	0.34	2.10	TRUE
Brain	14	90845063	90876551	ENST00000556721.1	CALM1	0.35	0.20	0.15	0.09	0.10	2.08	TRUE
Brain	3	11024732	11080079	ENST00000495636.1	SLC6A1	0.23	0.06	0.17	0.20	0.24	2.07	TRUE
Brain	1	32792421	32824727	ENST00000574315.1	TSSK3	0.27	0.14	0.12	0.18	0.19	2.07	TRUE
Brain	17	17638844	17657613	ENST00000395774.1	RAI1	0.30	0.16	0.14	0.25	0.29	2.06	TRUE
Brain	19	17858364	17882158	ENST00000600393.1	FCHO1	0.22	0.11	0.11	0.21	0.24	2.01	TRUE
Brain	7	45111554	45132881	ENST00000490531.2	NACAD	0.22	0.07	0.15	0.27	0.31	1.99	TRUE
Brain	2	220142124	220157631	ENST00000460801.1	PTPRN	0.43	0.16	0.27	0.31	0.35	1.96	TRUE
Brain	11	134254603	134298147	ENST00000531510.1	B3GAT1	0.26	0.11	0.15	0.22	0.26	1.94	TRUE
Brain	6	163815822	163882069	ENST00000537883.1	QKI	0.79	0.13	0.66	0.17	0.18	1.89	TRUE
Brain	19	50985631	51019282	ENST00000598657.1	ASPDH	0.28	0.11	0.17	0.28	0.34	1.86	TRUE
Brain	1	155828272	155864174	ENST00000539162.1	SYT11	0.21	0.06	0.15	0.14	0.17	1.85	TRUE
Brain	1	205461954	205494873	ENST00000505932.1	CDK18	0.26	0.06	0.20	0.23	0.26	1.84	TRUE
Brain	4	154142334	154219000	ENST00000433687.1	TRIM2	0.38	0.05	0.32	0.31	0.33	1.80	TRUE
Brain	7	30174536	30225450	ENST00000455738.1	C7orf41	0.23	0.06	0.18	0.30	0.34	1.80	TRUE
Brain	9	124022263	124092639	ENST00000477553.1	GSN	0.26	0.12	0.15	0.30	0.34	1.77	TRUE
Brain	5	36599886	36615091	ENST00000509272.1	SLC1A3	0.50	0.11	0.38	0.17	0.20	1.76	TRUE
Brain	1	61541625	61624014	ENST00000496712.1	NFIA	0.24	0.11	0.13	0.12	0.14	1.76	TRUE
Brain	17	56589712	56620235	ENST00000582976.1	SEPT4	0.36	0.19	0.17	0.21	0.23	1.72	TRUE
Brain	10	124216441	124276238	ENST00000420892.1	HTRA1	0.27	0.07	0.20	0.30	0.34	1.64	TRUE
Brain	8	26427962	26519962	ENST00000474808.1	DPYSL2	0.45	0.10	0.35	0.33	0.37	1.59	TRUE
Brain	9	14299740	14327190	ENST00000493697.1	NFIB	0.57	0.42	0.15	0.23	0.27	1.56	TRUE
Brain	6	52225169	52265846	ENST00000360726.3	PAQR8	0.21	0.07	0.14	0.12	0.15	1.50	TRUE
Brain	2	145129528	145283099	ENST00000475115.1	ZEB2	0.39	0.08	0.30	0.21	0.24	1.50	TRUE
Brain	1	90285126	90311060	ENST00000527156.1	LRRC8D	0.33	0.14	0.18	0.05	0.06	1.48	TRUE
Brain	12	123317747	123395518	ENST00000371248.3	VPS37B	0.30	0.13	0.17	0.31	0.33	1.47	TRUE
Brain	1	160161970	160191690	ENST00000481831.1	DCAF8	0.25	0.09	0.17	0.16	0.19	1.44	TRUE
Brain	14	50976072	51001976	ENST00000441560.2	ATL1	0.36	0.11	0.26	0.14	0.15	1.34	TRUE
Brain	5	132105879	132116081	ENST00000492490.1	SEPT8	0.43	0.21	0.22	0.21	0.25	1.31	TRUE
Brain	22	46445235	46456789	ENST00000381051.2	MIRLET7	0.63	0.47	0.16	0.19	0.23	1.30	TRUE
Brain	20	31025912	31074699	ENST00000326071.4	C20orf112	0.35	0.16	0.19	0.11	0.13	1.27	FALSE
Brain	1	160068777	160077945	ENST00000448417.1	IGSF8	0.44	0.14	0.30	0.26	0.28	1.25	FALSE
Brain	17	42161429	42202224	ENST00000587135.1	HDAC5	0.23	0.10	0.14	0.17	0.18	1.24	FALSE
Brain	6	15245470	15269892	ENST00000397311.3	JARID2	0.36	0.25	0.11	0.08	0.08	1.24	FALSE
Brain	2	95678307	95745847	ENST00000349807.3	MAL	0.27	0.08	0.19	0.21	0.24	1.24	FALSE
Brain	5	126625320	126648566	ENST00000274473.6	MEGF10	0.30	0.06	0.24	0.27	0.32	1.16	FALSE
Brain	12	6639174	6665533	ENST00000396830.2	IFFO1	0.40	0.29	0.11	0.18	0.20	1.15	FALSE
Brain	20	45922895	45992027	ENST00000441977.1	ZMYND8	0.24	0.13	0.11	0.16	0.19	1.14	FALSE
Brain	10	102752997	102775915	ENST00000454422.1	LZTS2	0.44	0.32	0.13	0.20	0.22	1.09	FALSE
Brain	7	32078425	32118455	ENST00000464881.1	PDE1C	0.23	0.06	0.17	0.22	0.26	1.07	FALSE
Brain	1	110071934	110092553	ENST00000369851.4	GNAI3	0.31	0.18	0.13	0.02	0.03	1.05	FALSE
Brain	10	64127809	64167185	ENST00000421210.1	ZNF365	0.23	0.03	0.20	0.24	0.26	1.04	FALSE
Brain	17	48914258	48946511	ENST00000509385.1	TOB1	0.28	0.17	0.11	0.17	0.18	1.01	FALSE
Brain	15	43800300	43824728	ENST00000300231.5	MAP1A	0.28	0.16	0.12	0.17	0.18	1.01	FALSE
Brain	1	33789227	33859751	ENST00000468406.1	PHC2	0.25	0.11	0.14	0.30	0.33	0.95	FALSE
Brain	6	69337682	69367340	ENST00000370598.1	BAI3	0.33	0.16	0.17	0.09	0.09	0.83	FALSE
Brain	20	35445944	35496093	ENST00000357779.3	SOGA1	0.30	0.11	0.20	0.30	0.31	0.82	FALSE
Brain	9	133698900	133750533	ENST00000318560.5	ABL1	0.29	0.08	0.21	0.33	0.36	0.81	FALSE
Brain	15	64974641	64996542	ENST00000559665.2	OAZ2	0.36	0.11	0.24	0.16	0.17	0.80	FALSE
Brain	2	9768796	9795405	ENST00000460093.1	YWHAQ	0.21	0.09	0.11	0.14	0.15	0.75	FALSE
Brain	3	181403893	181455295	ENST00000325404.1	SOX2	0.84	0.56	0.28	0.22	0.22	0.72	FALSE
Brain	11	111845328	111865086	ENST00000530645.1	DIXDC1	0.33	0.09	0.24	0.32	0.34	0.70	FALSE
Brain	6	128798636	128843173	ENST00000368202.4	PTPRK	0.20	0.09	0.11	0.08	0.08	0.67	FALSE
Brain	13	97862565	97936838	ENST00000445661.2	MBNL2	0.24	0.09	0.15	0.21	0.23	0.67	FALSE
Brain	17	38258962	38270871	ENST00000246672.3	NR1D1	0.42	0.31	0.11	0.15	0.16	0.64	FALSE
Brain	22	46457577	46466999	ENST00000396008.2	MIRLET7	0.94	0.50	0.44	0.31	0.35	0.64	FALSE
Brain	5	124064380	124097267	ENST00000512940.1	ZNF608	0.41	0.24	0.17	0.31	0.37	0.62	FALSE
Brain	12	1752842	1776949	ENST00000545747.1	WNT5B	0.24	0.13	0.11	0.29	0.31	0.37	FALSE
Brain	2	193010308	193063844	ENST00000409056.3	TMEFF2	0.21	0.06	0.15	0.22	0.23	0.37	FALSE
Brain	3	69130585	69168856	ENST00000485444.1	ARL6IP5	0.22	0.12	0.10	0.18	0.19	0.35	FALSE
Brain	5	115856316	115912087	ENST00000512156.1	SEMA6A	0.23	0.08	0.14	0.16	0.17	0.33	FALSE
Brain	12	48167031	48182319	ENST00000599515.1	AC004466.1	0.40	0.10	0.31	0.04	0.04	0.32	FALSE
Brain	10	73574974	73621362	ENST00000394934.1	PSAP	0.21	0.05	0.15	0.19	0.20	0.29	FALSE
Brain	21	43639119	43683392	ENST00000340588.4	ABCG1	0.22	0.11	0.11	0.19	0.20	0.28	FALSE
Brain	9	135976307	136024639	ENST00000372062.3	RALGDS	0.33	0.16	0.17	0.23	0.24	0.27	FALSE
Brain	19	54921528	54950915	ENST00000467939.1	TTYH1	0.38	0.18	0.20	0.20	0.22	0.24	FALSE
Brain	9	8840393	8880342	ENST00000481079.1	PTPRD	0.24	0.08	0.16	0.22	0.23	0.22	FALSE
Brain	4	134065738	134085565	ENST00000511112.1	PCDH10	0.73	0.42	0.30	0.16	0.15	0.16	FALSE
Brain	15	37371561	37406086	ENST00000559129.1	MEIS2	0.46	0.33	0.12	0.24	0.22	0.15	FALSE
Brain	3	170135527	170167574	ENST00000488989.1	CLDN11	0.34	0.07	0.27	0.32	0.30	0.13	FALSE
Brain	7	86272604	86301105	ENST00000421579.1	GRM3	0.22	0.09	0.13	0.12	0.11	0.12	FALSE
Brain	17	38216903	38234316	ENST00000577486.1	THRA	0.63	0.48	0.15	0.21	0.21	0.09	FALSE
Brain	4	96452170	96474433	ENST00000504962.1	UNC5C	0.48	0.15	0.33	0.12	0.11	0.03	FALSE
Brain	19	30713867	30786928	ENST00000591488.1	ZNF536	0.21	0.11	0.10	0.15	0.14	0.02	FALSE
Brain	10	24989499	25014855	ENST00000463892.1	ARHGAP21	0.33	0.22	0.12	0.15	0.16	-0.02	FALSE
Brain	14	51524492	51564122	ENST00000360392.4	TRIM9	0.20	0.07	0.13	0.22	0.21	-0.09	FALSE
Brain	17	68152783	68183190	ENST00000243457.3	KCNJ2	0.38	0.11	0.27	0.09	0.08	-0.10	FALSE
Brain	1	65712720	65745163	ENST00000483402.1	DNAJC6	0.34	0.10	0.24	0.24	0.21	-0.16	FALSE
Brain	11	6400761	6441884	ENST00000533407.1	APBB1	0.21	0.07	0.15	0.09	0.09	-0.24	FALSE
Brain	1	156353833	156405678	ENST00000495000.1	C1orf61	0.29	0.18	0.11	0.24	0.21	-0.27	FALSE
Brain	17	79345155	79376776	ENST00000307745.7	BAHCC1	0.54	0.37	0.17	0.18	0.17	-0.61	FALSE
Brain	21	34399873	34453980	ENST00000498799.1	OLIG1	0.28	0.18	0.11	0.19	0.17	-0.63	FALSE
Brain	11	119211319	119250680	ENST00000527843.1	USP2	0.22	0.10	0.12	0.19	0.17	-1.23	FALSE
Brain	10	88108134	88126246	ENST00000327946.7	GRID1	0.29	0.19	0.11	0.20	0.16	-1.70	FALSE
Breast	10	121410843	121447500	ENST00000450186.1	BAG3	0.21	0.06	0.15	0.26	0.39	18.11	TRUE
Breast	10	95184875	95242959	ENST00000488645.1	MYOF	0.21	0.04	0.17	0.27	0.42	15.89	TRUE
Breast	11	65678179	65686589	ENST00000530188.1	C11orf68	0.30	0.14	0.16	0.32	0.39	15.28	TRUE
Breast	1	207221596	207227576	ENST00000367079.2	PFKFB2	0.61	0.48	0.12	0.08	0.12	15.17	TRUE
Breast	2	106009478	106028025	ENST00000447958.1	FHL2	0.21	0.08	0.13	0.14	0.25	14.92	TRUE
Breast	11	6339234	6344196	ENST00000530979.1	PRKCDBP	0.61	0.34	0.26	0.14	0.25	14.91	TRUE
Breast	1	156065471	156100930	ENST00000392353.3	LMNA	0.56	0.18	0.38	0.16	0.23	14.72	TRUE
Breast	11	66821179	66830989	ENST00000532559.1	RHOD	0.33	0.16	0.17	0.17	0.26	14.21	TRUE
Breast	19	1154329	1184542	ENST00000587655.1	SBNO2	0.28	0.10	0.18	0.30	0.46	14.12	TRUE
Breast	11	114162169	114180527	ENST00000545255.1	NNMT	0.26	0.03	0.24	0.30	0.58	13.61	TRUE
Breast	2	112249648	112262948	ENST00000371162.4	AC108463.3	0.27	0.08	0.19	0.03	0.21	13.38	TRUE
Breast	19	13948994	13959464	ENST00000591727.1	NANOS3	0.68	0.56	0.13	0.19	0.30	13.12	TRUE
Breast	11	62306073	62328249	ENST00000531324.1	AHNAK	0.61	0.34	0.27	0.17	0.28	13.12	TRUE
Breast	2	87744763	87759323	ENST00000359481.4	PLGLB2	0.21	0.09	0.11	0.03	0.18	13.06	TRUE
Breast	1	6637843	6670728	ENST00000496707.1	KLHL21	0.28	0.17	0.12	0.16	0.25	12.95	TRUE
Breast	14	69241271	69268143	ENST00000408913.2	ZFP36L1	0.42	0.32	0.10	0.09	0.13	12.42	TRUE
Breast	20	51581498	51597746	ENST00000371497.4	TSHZ2	0.45	0.16	0.29	0.17	0.26	12.06	TRUE

Breast	8	62622640	62634995	ENST00000379449.6	ASPH	0.38	0.21	0.17	0.18	0.23	11.50	TRUE
Breast	4	169752417	169771288	ENST00000507735.1	PALLD	0.28	0.16	0.12	0.10	0.14	11.46	TRUE
Breast	1	16156416	16179578	ENST00000375759.3	SPEN	0.49	0.37	0.12	0.08	0.11	11.33	TRUE
Breast	1	181055984	181089589	ENST00000367577.4	IER5	0.32	0.13	0.19	0.16	0.19	11.21	TRUE
Breast	8	145008274	145031982	ENST00000527816.1	PLEC	0.62	0.34	0.27	0.21	0.28	11.19	TRUE
Breast	14	105432978	105441663	ENST00000555122.1	AHNAK2	0.43	0.23	0.20	0.14	0.23	10.98	TRUE
Breast	15	75930589	75963120	ENST00000569152.1	SNX33	0.27	0.16	0.11	0.11	0.16	10.81	TRUE
Breast	15	68548590	68582156	ENST00000566739.1	FEM1B	0.25	0.12	0.13	0.17	0.25	10.74	TRUE
Breast	20	43963799	43990114	ENST00000537976.1	SDC4	0.33	0.13	0.21	0.11	0.16	10.63	TRUE
Breast	2	36579464	36605312	ENST00000473403.1	CRIM1	0.55	0.19	0.35	0.21	0.32	10.63	TRUE
Breast	12	125388867	125425378	ENST00000542416.1	UBC	0.39	0.28	0.11	0.12	0.15	10.54	TRUE
Breast	12	10868156	10877012	ENST00000540747.1	CSDA	0.79	0.29	0.50	0.08	0.15	10.38	TRUE
Breast	6	111906089	111928358	ENST00000528599.1	TRAF3IP2	0.33	0.15	0.18	0.18	0.25	10.38	TRUE
Breast	17	2294949	2311496	ENST00000571836.2	MNT	0.62	0.42	0.20	0.12	0.14	10.21	TRUE
Breast	10	33226394	33252825	ENST00000439974.3	ITGB1	0.30	0.19	0.10	0.09	0.12	10.06	TRUE
Breast	14	55568260	55575863	ENST00000553493.1	LGALS3	0.32	0.13	0.19	0.17	0.41	10.04	TRUE
Breast	11	57045039	57094324	ENST00000527207.1	TNKS1BP1	0.23	0.09	0.13	0.11	0.20	9.99	TRUE
Breast	1	27830256	27900464	ENST00000487743.2	AHDC1	0.37	0.09	0.28	0.33	0.43	9.98	TRUE
Breast	16	56638741	56649586	ENST00000567300.1	MT2A	0.46	0.27	0.19	0.20	0.31	9.81	TRUE
Breast	22	36718824	36814131	ENST00000216181.5	MYH9	0.34	0.16	0.18	0.18	0.23	9.73	TRUE
Breast	1	27183345	27192842	ENST00000339276.4	SFN	0.33	0.22	0.11	0.29	0.37	9.65	TRUE
Breast	1	201404474	201474025	ENST00000532460.1	CSRP1	0.22	0.10	0.12	0.19	0.24	9.49	TRUE
Breast	3	187453456	187468930	ENST00000496823.1	BCL6	0.81	0.63	0.19	0.09	0.11	9.49	TRUE
Breast	11	8828385	8837469	ENST00000531237.1	ST5	0.63	0.12	0.51	0.32	0.48	9.25	TRUE
Breast	2	239193425	239200323	ENST00000431832.1	PER2	0.73	0.36	0.37	0.20	0.28	9.19	TRUE
Breast	19	47598365	47618174	ENST00000594526.1	SAE1	0.46	0.30	0.16	0.13	0.16	9.18	TRUE
Breast	11	9586185	9600407	ENST00000524612.1	WEE1	0.46	0.26	0.21	0.09	0.13	8.91	TRUE
Breast	14	77489392	77513614	ENST00000238647.3	IRF2BPL	0.67	0.37	0.29	0.17	0.24	8.75	TRUE
Breast	5	131593100	131603174	ENST00000418373.1	PDLIM4	0.60	0.08	0.51	0.29	0.40	8.73	TRUE
Breast	11	61732568	61749363	ENST00000601917.1	AP003733.1	0.38	0.24	0.14	0.09	0.11	8.53	TRUE
Breast	1	45270631	45276266	ENST00000482715.1	BTBD19	0.65	0.27	0.38	0.33	0.40	8.47	TRUE
Breast	19	42771661	42788152	ENST00000575839.1	CIC	0.57	0.44	0.13	0.15	0.17	8.26	TRUE
Breast	12	52608969	52643707	ENST00000544024.1	KRT86	0.21	0.06	0.15	0.27	0.36	8.01	TRUE
Breast	7	869464	876882	ENST00000469755.1	SUN1	0.22	0.07	0.15	0.29	0.44	7.97	TRUE
Breast	2	46523740	46563334	ENST00000475822.1	EPAS1	0.23	0.13	0.10	0.13	0.17	7.61	TRUE
Breast	20	17537085	17560729	ENST00000449141.2	DSTN	0.27	0.12	0.16	0.15	0.20	7.60	TRUE
Breast	9	130280649	130347137	ENST00000373314.3	FAM129B	0.20	0.10	0.10	0.29	0.37	7.59	TRUE
Breast	22	30792614	30822843	ENST00000407550.3	MTFP1	0.36	0.17	0.19	0.11	0.13	7.58	TRUE
Breast	12	109225745	109251401	ENST00000548522.1	SSH1	0.24	0.09	0.15	0.16	0.24	7.09	TRUE
Breast	14	23292133	23323062	ENST00000547596.1	MMP14	0.33	0.15	0.17	0.13	0.19	7.09	TRUE
Breast	1	27320120	27340181	ENST00000289166.5	FAM46B	0.33	0.18	0.16	0.14	0.17	7.04	TRUE
Breast	1	8062975	8088860	ENST00000487559.1	ERRFI1	0.25	0.14	0.10	0.16	0.21	6.92	TRUE
Breast	1	154941228	154949531	ENST00000473344.1	CKS1B	0.82	0.52	0.31	0.14	0.19	6.71	TRUE
Breast	19	41220087	41229010	ENST00000263370.2	ITPKC	0.69	0.55	0.14	0.09	0.11	6.65	TRUE
Breast	7	41734099	41745805	ENST00000442711.1	INHBA	0.40	0.26	0.14	0.13	0.19	6.64	TRUE
Breast	15	42562799	42569003	ENST00000567421.1	GANC	0.91	0.51	0.40	0.06	0.08	6.58	TRUE
Breast	5	142773722	142785691	ENST00000502500.1	NR3C1	0.95	0.51	0.44	0.10	0.15	6.55	TRUE
Breast	1	16274460	16281222	ENST00000494020.1	ZBTB17	0.21	0.07	0.14	0.32	0.48	6.16	TRUE
Breast	1	155097310	155110100	ENST00000484027.1	SLC50A1	0.63	0.51	0.12	0.09	0.11	6.11	TRUE
Breast	11	118779500	118810391	ENST00000534788.1	UPK2	0.63	0.45	0.18	0.14	0.20	5.75	TRUE
Breast	17	27134005	27140821	ENST00000582059.1	FAM222B	0.45	0.35	0.10	0.05	0.08	5.73	TRUE
Breast	22	46467114	46488203	ENST00000360737.3	MIRLET7	1.00	0.47	0.53	0.25	0.45	5.64	TRUE
Breast	1	33207559	33241755	ENST00000373480.1	KIAA1522	0.22	0.10	0.12	0.22	0.27	5.58	TRUE
Breast	22	46448058	46467049	ENST00000381051.2	MIRLET7	1.00	0.50	0.50	0.20	0.30	5.57	TRUE
Breast	7	55085115	55095372	ENST00000463948.1	EGFR	0.50	0.21	0.29	0.19	0.24	5.29	TRUE
Breast	11	65238595	65276583	ENST00000309775.7	AP000769.1	0.63	0.43	0.20	0.14	0.18	5.25	TRUE
Breast	10	82213990	82238676	ENST00000372156.1	TSPAN14	0.25	0.14	0.11	0.13	0.17	5.20	TRUE
Breast	17	46122636	46152031	ENST00000581319.1	NFE2L1	0.28	0.11	0.17	0.13	0.15	5.18	TRUE
Breast	11	57528367	57568939	ENST00000534647.1	CTNND1	0.31	0.11	0.20	0.16	0.30	5.02	TRUE
Breast	17	17723596	17746027	ENST00000435530.2	SREBF1	0.36	0.25	0.11	0.14	0.16	4.99	TRUE
Breast	9	124030410	124052489	ENST00000545652.1	GSN	0.40	0.26	0.15	0.21	0.23	4.78	TRUE
Breast	3	141083866	141089271	ENST00000507657.1	ZBTB38	0.39	0.08	0.32	0.17	0.34	4.74	TRUE
Breast	17	17578482	17616584	ENST00000471135.2	RAI1	0.26	0.15	0.10	0.18	0.21	4.67	TRUE
Breast	11	12694725	12726400	ENST00000334310.6	TEAD1	0.27	0.17	0.10	0.16	0.19	4.63	TRUE
Breast	12	6432600	6452223	ENST00000538363.1	TNFRSF1A	0.50	0.30	0.19	0.05	0.08	4.48	TRUE
Breast	17	7736868	7749068	ENST00000570832.1	KDM6B	0.92	0.77	0.15	0.07	0.08	4.31	TRUE
Breast	18	3446318	3460082	ENST00000472042.1	TGIF1	0.68	0.56	0.12	0.06	0.07	4.19	TRUE
Breast	20	19954385	20001261	ENST00000481837.1	NAA20	0.22	0.09	0.13	0.15	0.18	4.18	TRUE
Breast	1	19247868	19263180	ENST00000410166.1	IFFO2	0.29	0.14	0.15	0.21	0.23	3.72	TRUE
Breast	3	29321550	29334286	ENST00000456853.1	RBMS3	0.65	0.21	0.45	0.05	0.07	3.66	TRUE
Breast	6	138172161	138196002	ENST00000433680.1	TNFAIP3	0.29	0.14	0.14	0.07	0.10	3.62	TRUE
Breast	12	59300121	59315174	ENST00000548968.1	LRIG3	0.39	0.18	0.21	0.07	0.09	3.48	TRUE
Breast	18	3245538	3270102	ENST00000584539.1	MYL12B	0.43	0.25	0.18	0.06	0.08	3.41	TRUE
Breast	3	9436165	9466747	ENST00000468208.1	SETD5	0.30	0.18	0.12	0.11	0.13	3.16	TRUE
Breast	11	65185034	65198304	ENST00000531151.1	FRMD8	1.00	0.88	0.12	0.10	0.14	2.99	TRUE
Breast	18	46449301	46483832	ENST00000586093.1	SMAD7	0.47	0.32	0.15	0.15	0.18	2.93	TRUE
Breast	9	14298144	14323961	ENST00000493697.1	NFIB	0.71	0.42	0.28	0.22	0.27	2.74	TRUE
Breast	3	193849125	193860857	ENST00000476918.1	HES1	0.79	0.61	0.18	0.07	0.08	2.64	TRUE
Breast	7	42257859	42278946	ENST00000428534.1	GLI3	0.33	0.17	0.15	0.20	0.20	2.56	TRUE
Breast	21	36252652	36264673	ENST00000399237.2	RUNX1	0.93	0.58	0.35	0.13	0.17	2.54	TRUE
Breast	19	13123147	13172690	ENST00000358552.3	NFIX	0.50	0.29	0.20	0.32	0.35	2.35	TRUE
Breast	10	74002370	74038384	ENST00000471240.1	DDIT4	0.27	0.17	0.10	0.11	0.12	2.32	TRUE
Breast	4	95366945	95386196	ENST00000511767.1	PDLIM5	0.24	0.11	0.12	0.08	0.10	2.31	TRUE
Breast	19	16177568	16192210	ENST00000588483.1	TPM4	0.43	0.26	0.16	0.16	0.18	2.23	TRUE
Breast	6	10398835	10421638	ENST00000486038.1	TFAP2A	0.81	0.68	0.13	0.19	0.22	2.02	TRUE
Breast	2	43444985	43456496	ENST00000282388.3	ZFP36L2	1.00	0.88	0.12	0.05	0.06	1.99	TRUE
Breast	15	93425946	93432559	ENST00000555520.1	CHD2	0.87	0.61	0.26	0.01	0.02	1.91	TRUE
Breast	7	36426266	36435508	ENST00000418118.1	ANLN	0.36	0.23	0.12	0.06	0.07	1.86	TRUE
Breast	1	100109973	100129273	ENST00000263174.4	PALMD	0.24	0.03	0.21	0.16	0.23	1.86	TRUE
Breast	14	74207241	74227902	ENST00000421708.1	ELMSAN1	0.58	0.41	0.18	0.12	0.13	1.80	TRUE
Breast	19	49375182	49379790	ENST00000600406.1	PPP1R15A	0.78	0.60	0.18	0.07	0.09	1.63	TRUE
Breast	20	10634437	10656669	ENST00000254958.4	JAG1	0.49	0.22	0.27	0.06	0.06	1.58	TRUE
Breast	1	59245181	59252471	ENST00000371222.2	JUN	0.86	0.61	0.25	0.05	0.05	1.42	TRUE
Breast	4	103736345	103751180	ENST00000503282.1	UBE2D3	0.48	0.36	0.11	0.05	0.06	1.39	TRUE
Breast	20	34326330	34331841	ENST00000493853.1	RBM39	0.91	0.74	0.17	0.05	0.05	1.38	TRUE
Breast	10	106080462	106112294	ENST00000458723.1	ITPRIP	0.31	0.19	0.12	0.19	0.19	1.27	FALSE
Breast	16	23152633	23161835	ENST00000219689.7	USP31	0.37	0.24	0.12	0.02	0.03	1.01	FALSE
Breast	4	146653244	146658422	ENST00000510096.1	C4orf51	0.68	0.29	0.39	0.04	0.05	0.98	FALSE
Breast	6	74224616	74234296	ENST00000455918.1	EEF1A1	0.79	0.59	0.21	0.07	0.08	0.87	FALSE
Breast	6	136605683	136612552	ENST00000528229.1	BCLAF1	0.54	0.42	0.12	0.03	0.04	0.87	FALSE
Breast	6	121756401	121769000	ENST00000282561.3	GJA1	0.38	0.18	0.19	0.15	0.20	0.80	FALSE
Breast	15	50644604	50649357	ENST00000560896.1	GABPB1	0.96	0.80	0.15	0.04	0.04	0.79	FALSE
Breast	16	15735275	15740014	ENST00000396353.2	NDE1	0.64	0.54	0.10	0.04	0.04	0.76	FALSE
Breast	10	74051120	74097380	ENST00000473051.1	DNAJB12	0.27	0.14	0.13	0.13	0.13	0.68	FALSE
Breast	5	1881509	1890183	ENST00000513692.1	IRX4	0.97	0.78	0.19	0.19	0.22	0.63	FALSE
Breast	7	115849775	115868787	ENST00000485009.1	TES	0.21	0.11	0.10	0.05	0.06	0.60	FALSE
Breast	2	12841318	12863362	ENST00000405331.3	TRIB2	0.35	0.19	0.17	0.09	0.11	0.26	FALSE
Breast	9	33157786	33167403	ENST00000379731.4	B4GALT1	0.92	0.42	0.50	0.14	0.14	-0.02	FALSE
Breast	10	112252889	112273435	ENST00000468749.1	DUSP5	0.37	0.21	0.16	0.24	0.26	-0.15	FALSE
Breast	5	67510759	67574308	ENST00000520675.1	PIK3R1	0.27	0.10	0.16	0.08	0.07	-0.34	FALSE
Breast	3	71104592	71120542	ENST00000497553.1	FOXP1	0.41	0.28	0.13	0.13	0.12	-0.62	FALSE
Breast	5	2739262	2759651	ENST00000502957.1	IRX2	0.92	0.80	0.11	0.23	0.20	-1.21	FALSE
Breast	5	140887653	140909039	ENST00000468119.1	DIAPH1	0.23	0.11	0.12	0.32	0.26	-2.92	FALSE
Breast	8	116659885	116682409	ENST00000451156.1	TRPS1	0.42	0.17	0.26	0.20	0.15	-3.13	FALSE
Breast	12	109082915	109097018	ENST00000547170.1	CORO1C	0.31	0.21	0.11	0.24	0.14	-14.35	FALSE
CD19	17	56394652	56423683	ENST00000583624.1	BZRAP1	0.46	0.16	0.31	0.20	0.41	39.73	TRUE
CD19	1	111401311	111424334	ENST00000476408.1	CD53	0.29	0.00	0.29	0.12	0.41	37.85	TRUE

CD19	16	85965145	85983944	ENST00000569607.1	IRF8	0.21	0.02	0.19	0.30	0.57	37.13	TRUE
CD19	6	13258837	13278907	ENST00000481706.1	PHACTR1	0.29	0.01	0.28	0.19	0.58	36.71	TRUE
CD19	2	233923908	233972771	ENST00000467393.1	INPP5D	0.29	0.05	0.24	0.19	0.42	35.52	TRUE
CD19	1	31201550	31236569	ENST00000476492.1	LAPTM5	0.30	0.04	0.26	0.31	0.56	31.68	TRUE
CD19	22	50627294	50632183	ENST00000395829.1	TRABD	0.69	0.28	0.41	0.15	0.42	30.04	TRUE
CD19	18	74764261	74781577	ENST00000580473.1	MBP	0.29	0.13	0.16	0.28	0.54	29.54	TRUE
CD19	10	11184898	11223017	ENST00000537122.1	CELF2	0.32	0.06	0.26	0.13	0.29	28.19	TRUE
CD19	6	32807846	32814107	ENST00000414474.1	PSMB9	0.52	0.30	0.22	0.16	0.23	25.92	TRUE
CD19	1	167584400	167603373	ENST00000537350.1	RCSD1	0.37	0.08	0.29	0.20	0.36	25.14	TRUE
CD19	1	150122093	150138999	ENST00000485470.1	PLEKHO1	0.34	0.13	0.20	0.10	0.21	24.45	TRUE
CD19	9	134127597	134154499	ENST00000464831.1	FAM78A	0.35	0.11	0.24	0.16	0.30	24.20	TRUE
CD19	1	206730430	206756416	ENST00000304534.8	RASSF5	0.25	0.05	0.19	0.16	0.27	22.59	TRUE
CD19	13	41534432	41594336	ENST00000239882.3	ELF1	0.21	0.08	0.13	0.09	0.31	21.93	TRUE
CD19	12	57869667	57876031	ENST00000552249.1	ARHGAP9	0.56	0.14	0.42	0.26	0.40	21.40	TRUE
CD19	14	61937696	61947119	ENST00000536400.1	PRKCH	0.33	0.04	0.29	0.04	0.29	19.40	TRUE
CD19	7	50344148	50370755	ENST00000426121.1	IKZF1	0.41	0.07	0.35	0.20	0.42	19.07	TRUE
CD19	21	44817042	44836022	ENST00000478426.1	SIK1	0.22	0.05	0.17	0.30	0.58	18.82	TRUE
CD19	11	128582503	128610944	ENST00000534087.1	FLI1	0.28	0.04	0.24	0.14	0.36	17.88	TRUE
CD19	3	122291690	122308644	ENST00000483793.1	PARP15	0.20	0.04	0.16	0.13	0.29	17.63	TRUE
CD19	17	43299369	43312992	ENST00000591434.1	FMNL1	0.55	0.16	0.39	0.17	0.30	17.34	TRUE
CD19	22	37613897	37642449	ENST00000401529.3	RAC2	0.28	0.09	0.18	0.07	0.12	17.11	TRUE
CD19	2	238580170	238610065	ENST00000473815.1	LRRFIP1	0.35	0.19	0.16	0.13	0.24	16.12	TRUE
CD19	20	55964442	55977216	ENST00000371219.2	RBM38	0.53	0.33	0.20	0.15	0.24	15.94	TRUE
CD19	21	45660565	45668530	ENST00000400377.3	ICOSLG	0.72	0.38	0.34	0.08	0.14	15.33	TRUE
CD19	14	105116134	105145145	ENST00000330634.7	INF2	0.23	0.12	0.11	0.14	0.28	15.28	TRUE
CD19	1	26609716	26620076	ENST00000319041.6	SH3BGRL3	0.40	0.03	0.37	0.24	0.41	14.66	TRUE
CD19	1	53096945	53108122	ENST00000424164.1	FAM159A	0.35	0.08	0.27	0.19	0.26	13.94	TRUE
CD19	22	24821855	24835079	ENST00000439591.1	ADORA2A	0.26	0.02	0.25	0.32	0.60	13.11	TRUE
CD19	19	2610475	2632474	ENST00000587867.1	GNG7	0.31	0.09	0.22	0.10	0.17	12.71	TRUE
CD19	13	46942847	46975976	ENST00000480935.1	KIAA0226L	0.24	0.08	0.16	0.07	0.10	12.28	TRUE
CD19	20	4789416	4805261	ENST00000379400.3	RASSF7	0.41	0.18	0.23	0.22	0.36	12.12	TRUE
CD19	4	56812204	56816635	ENST00000257287.4	CEP135	0.73	0.52	0.21	0.11	0.14	12.05	TRUE
CD19	8	56790857	56809092	ENST00000520220.1	LYN	0.35	0.10	0.25	0.13	0.21	11.96	TRUE
CD19	6	159459555	159485569	ENST00000338313.5	TAGAP	0.27	0.05	0.22	0.13	0.25	11.68	TRUE
CD19	17	29814146	29838313	ENST00000578694.1	RAB11FIP4	0.24	0.11	0.13	0.14	0.26	11.44	TRUE
CD19	19	35807029	35840479	ENST00000593704.1	CD22	0.26	0.02	0.24	0.22	0.56	11.25	TRUE
CD19	19	13204800	13217457	ENST00000590120.1	LYL1	0.94	0.34	0.60	0.21	0.44	11.15	TRUE
CD19	11	67034303	67056724	ENST00000447274.2	ANKRD13D	0.40	0.23	0.17	0.12	0.23	10.45	TRUE
CD19	11	73076425	73101853	ENST00000393580.2	RELT	0.21	0.06	0.15	0.13	0.18	10.21	TRUE
CD19	11	9621558	9642085	ENST00000527848.1	WEE1	0.21	0.09	0.12	0.06	0.11	9.86	TRUE
CD19	9	35616119	35651731	ENST00000474403.1	SIT1	0.23	0.07	0.16	0.15	0.23	9.35	TRUE
CD19	15	74666045	74697205	ENST00000268053.6	CYP11A1	0.31	0.12	0.19	0.26	0.34	8.30	TRUE
CD19	1	111734737	111768995	ENST00000445067.2	CHI3L2	0.20	0.06	0.14	0.09	0.12	8.24	TRUE
CD19	6	41671523	41703442	ENST00000433032.1	TFEB	0.33	0.15	0.18	0.18	0.23	8.08	TRUE
CD19	6	32904722	32916498	ENST00000416244.2	HLA-DMB	0.27	0.09	0.18	0.16	0.31	7.98	TRUE
CD19	1	66795175	66818506	ENST00000526197.1	PDE4B	0.28	0.18	0.10	0.13	0.25	7.81	TRUE
CD19	16	85921635	85951991	ENST00000569607.1	IRF8	0.29	0.05	0.24	0.22	0.37	7.79	TRUE
CD19	3	196350844	196371448	ENST00000426755.1	PIGX	0.34	0.14	0.20	0.12	0.18	7.65	TRUE
CD19	9	130665468	130681316	ENST00000479747.1	ST6GALNAC4	0.33	0.18	0.16	0.09	0.13	7.17	TRUE
CD19	6	31547188	31556914	ENST00000464044.1	LST1	0.57	0.18	0.40	0.19	0.25	6.59	TRUE
CD19	6	149785200	149817360	ENST00000462655.1	ZC3H12D	0.28	0.04	0.24	0.17	0.35	6.58	TRUE
CD19	12	46657445	46665336	ENST00000546519.1	SLC38A1	0.57	0.45	0.12	0.10	0.17	6.45	TRUE
CD19	11	58967706	58992697	ENST00000361050.3	MPEG1	0.22	0.02	0.20	0.29	0.50	6.35	TRUE
CD19	1	207078366	207106998	ENST00000487208.1	PIGR	0.23	0.05	0.18	0.20	0.38	6.25	TRUE
CD19	12	48196588	48213374	ENST00000445237.2	HDAC7	0.40	0.23	0.17	0.11	0.15	6.14	TRUE
CD19	2	54783925	54824402	ENST00000333896.5	SPTBN1	0.27	0.15	0.12	0.08	0.13	6.13	TRUE
CD19	2	197013308	197048452	ENST00000449152.1	STK17B	0.42	0.07	0.34	0.07	0.12	5.81	TRUE
CD19	2	71290385	71301563	ENST00000524537.1	NAGK	0.39	0.21	0.18	0.07	0.10	5.74	TRUE
CD19	6	32402294	32411438	ENST00000374982.5	HLA-DRA	0.36	0.10	0.26	0.24	0.36	5.34	TRUE
CD19	19	42374270	42392833	ENST00000337665.4	ARHGEF1	0.44	0.19	0.25	0.15	0.23	5.12	TRUE
CD19	10	49874683	49897623	ENST00000360890.2	WDFY4	0.25	0.00	0.25	0.30	0.59	5.11	TRUE
CD19	17	74476430	74497452	ENST00000590288.1	RHBDF2	0.36	0.13	0.23	0.27	0.38	4.87	TRUE
CD19	1	154914194	154929302	ENST00000490230.1	PBXIP1	0.29	0.16	0.12	0.16	0.21	4.81	TRUE
CD19	20	44736001	44751095	ENST00000461171.1	CD40	0.21	0.09	0.12	0.06	0.15	4.79	TRUE
CD19	3	13034134	13065411	ENST00000473088.1	IQSEC1	0.30	0.15	0.15	0.16	0.23	4.63	TRUE
CD19	11	72850513	72872020	ENST00000422375.1	FCHSD2	0.41	0.17	0.24	0.10	0.13	4.54	TRUE
CD19	19	18278723	18292546	ENST00000600463.1	IFI30	0.23	0.10	0.13	0.15	0.25	4.49	TRUE
CD19	2	136871650	136898787	ENST00000466288.1	CXCR4	0.33	0.14	0.18	0.06	0.08	4.31	TRUE
CD19	11	61120450	61130574	ENST00000423772.2	TMEM136	0.45	0.20	0.25	0.21	0.23	4.29	TRUE
CD19	15	75066062	75093388	ENST00000569321.1	CSK	0.51	0.25	0.27	0.13	0.18	4.20	TRUE
CD19	16	11758829	11785379	ENST00000575349.1	TXNDC11	0.22	0.10	0.12	0.10	0.11	4.16	TRUE
CD19	16	10959821	11016093	ENST00000572665.1	CIITA	0.28	0.07	0.20	0.11	0.23	4.05	TRUE
CD19	15	57571264	57599710	ENST00000560948.1	TCF12	0.24	0.04	0.19	0.16	0.22	4.02	TRUE
CD19	14	106313303	106330974	ENST00000482999.1	KIAA0125	0.51	0.00	0.51	0.31	0.59	4.00	TRUE
CD19	19	2040263	2096653	ENST00000395307.2	IZUMO4	0.32	0.19	0.13	0.18	0.24	3.98	TRUE
CD19	16	81803418	81877366	ENST00000569523.1	PLCG2	0.26	0.02	0.24	0.24	0.26	3.96	TRUE
CD19	8	56751994	56758932	ENST00000519728.1	LYN	0.54	0.24	0.30	0.04	0.08	3.70	TRUE
CD19	2	231729691	231744201	ENST00000492029.1	ITM2C	0.33	0.19	0.14	0.24	0.34	3.70	TRUE
CD19	10	112618797	112633729	ENST00000444997.1	PDCD4	0.35	0.25	0.11	0.08	0.11	3.69	TRUE
CD19	14	105524940	105537562	ENST00000546679.1	GPR132	0.32	0.05	0.27	0.29	0.48	3.56	TRUE
CD19	6	106957303	106996547	ENST00000487681.1	AIM1	0.26	0.13	0.13	0.11	0.17	3.39	TRUE
CD19	1	38937882	38948718	ENST00000474456.1	RRAGC	0.22	0.06	0.16	0.19	0.32	3.28	TRUE
CD19	22	42304034	42337858	ENST00000472374.2	CENPM	0.44	0.22	0.22	0.22	0.28	3.26	TRUE
CD19	2	202106787	202128005	ENST00000429881.1	CASP8	0.21	0.10	0.11	0.15	0.20	3.18	TRUE
CD19	17	27064086	27073896	ENST00000584944.1	TRAF4	0.48	0.35	0.13	0.12	0.13	3.14	TRUE
CD19	12	7044970	7072648	ENST00000537533.1	PTPN6	0.36	0.21	0.15	0.20	0.27	3.12	TRUE
CD19	4	82389749	82394691	ENST00000507538.1	RASGEF1B	0.74	0.59	0.14	0.08	0.09	3.09	TRUE
CD19	12	92527263	92581890	ENST00000552315.1	BTG1	0.30	0.14	0.15	0.07	0.07	3.06	TRUE
CD19	16	28921620	28948628	ENST00000567368.1	CD19	0.20	0.10	0.11	0.20	0.30	2.96	TRUE
CD19	4	185452534	185460044	ENST00000393593.3	IRF2	0.39	0.25	0.14	0.15	0.19	2.90	TRUE
CD19	22	18476769	18494405	ENST00000424046.1	MICAL3	0.32	0.12	0.19	0.18	0.31	2.86	TRUE
CD19	15	86229664	86252170	ENST00000560482.1	AKAP13	0.31	0.05	0.26	0.24	0.38	2.74	TRUE
CD19	17	3807403	3821299	ENST00000571637.1	P2RX1	0.34	0.10	0.24	0.30	0.47	2.74	TRUE
CD19	15	45001549	45023195	ENST00000561237.1	TRIM69	0.36	0.24	0.12	0.09	0.13	2.68	TRUE
CD19	2	46761453	46803094	ENST00000473428.1	RHOQ	0.24	0.09	0.15	0.05	0.06	2.66	TRUE
CD19	12	9788063	9838708	ENST00000544322.1	CLEC2D	0.20	0.06	0.14	0.15	0.17	2.63	TRUE
CD19	6	167362856	167374276	ENST00000496851.2	RNASET2	0.36	0.20	0.16	0.11	0.12	2.16	TRUE
CD19	7	142490667	142511468	ENST00000463701.1	PRSS1	0.26	0.12	0.14	0.12	0.14	2.14	TRUE
CD19	7	86840957	86850205	ENST00000455575.1	TMEM243	0.59	0.31	0.28	0.07	0.08	2.08	TRUE
CD19	1	2477758	2482788	ENST00000426449.1	TNFRSF14	0.72	0.59	0.12	0.11	0.14	2.03	TRUE
CD19	7	150100445	150107156	ENST00000474836.1	ZNF775	0.56	0.35	0.21	0.15	0.21	1.88	TRUE
CD19	4	40289700	40319643	ENST00000310169.2	CHRNA9	0.21	0.06	0.14	0.08	0.15	1.71	TRUE
CD19	8	134509931	134514822	ENST00000520020.1	ST3GAL1	0.60	0.12	0.48	0.25	0.29	1.71	TRUE
CD19	17	75396031	75469717	ENST00000593189.1	SEPT9	0.28	0.15	0.12	0.27	0.31	1.48	TRUE
CD19	11	58340596	58348101	ENST00000389919.4	ZFP91-CNTF	0.83	0.52	0.31	0.07	0.07	1.29	FALSE
CD19	9	3524176	3529109	ENST00000449190.1	RFX3	0.90	0.77	0.13	0.06	0.07	1.25	FALSE
CD19	7	138778458	138805019	ENST00000471652.1	ZC3HAV1	0.29	0.12	0.17	0.10	0.12	1.22	FALSE
CD19	2	232255961	232260646	ENST00000287590.5	B3GNT7	0.54	0.13	0.41	0.32	0.37	1.15	FALSE
CD19	19	50917413	50938549	ENST00000357701.5	MYBPC2	0.42	0.26	0.16	0.22	0.24	1.12	FALSE
CD19	14	23018284	23039986	ENST00000557595.1	AE000662.92	0.25	0.10	0.15	0.05	0.05	1.04	FALSE
CD19	12	110431794	110453837	ENST00000261739.4	ANKRD13A	0.28	0.16	0.12	0.07	0.07	0.71	FALSE
CD19	14	50327610	50337580	ENST00000298310.5	NEMF	0.50	0.38	0.12	0.04	0.04	0.70	FALSE
CD19	3	58318904	58343543	ENST00000491164.1	PXK	0.25	0.07	0.18	0.02	0.03	0.61	FALSE
CD19	17	38004143	38027814	ENST00000377940.3	ZPBP2	0.26	0.05	0.21	0.32	0.34	0.52	FALSE
CD19	8	29937591	29951059	ENST00000521083.1	TMEM66	0.29	0.18	0.11	0.04	0.04	0.15	FALSE
CD19	19	10712789	10716244	ENST00000407327.4	SLC44A2	0.80	0.58	0.21	0.16	0.17	0.14	FALSE
CD19	2	196511447	196525260	ENST00000418005.1	SLC39A10	0.39	0.26	0.13	0.03	0.03	0.00	FALSE

CD19	5	122109910	122117182	ENST00000514949.1	SNX2	0.41	0.31	0.11	0.08	0.07	-0.04	FALSE
CD19	16	15728212	15738278	ENST00000396353.2	NDE1	0.42	0.28	0.14	0.04	0.04	-0.07	FALSE
CD19	19	39887963	39904206	ENST00000438123.1	PLEKHG2	0.70	0.60	0.10	0.08	0.07	-0.13	FALSE
CD19	18	2634081	2658907	ENST00000261598.8	SMCHD1	0.25	0.15	0.10	0.10	0.09	-0.24	FALSE
CD19	22	31680149	31688920	ENST00000443175.1	PIK3IP1	0.44	0.33	0.11	0.08	0.08	-0.27	FALSE
CD19	17	38473833	38487686	ENST00000582914.1	RARA	0.58	0.44	0.14	0.16	0.15	-0.31	FALSE
CD19	11	122930787	122942297	ENST00000531063.1	HSPA8	0.53	0.34	0.19	0.15	0.15	-0.44	FALSE
CD19	11	82763568	82785588	ENST00000533276.1	RAB30	0.31	0.16	0.15	0.05	0.05	-0.74	FALSE
CD19	2	128144112	128148979	ENST00000409179.2	MAP3K2	0.62	0.45	0.17	0.14	0.12	-0.84	FALSE
CD19	14	64966362	64978672	ENST00000555321.1	ZBTB1	0.55	0.37	0.18	0.11	0.09	-1.91	FALSE
CD19	1	40838755	40871047	ENST00000539317.1	SMAP2	0.22	0.08	0.14	0.10	0.07	-2.81	FALSE
CD19	1	27984903	27990153	ENST00000339145.4	IFI6	0.51	0.32	0.18	0.14	0.10	-3.04	FALSE
Colon	5	79542438	79554102	ENST00000513907.1	SERINC5	0.51	0.26	0.25	0.13	0.18	14.54	TRUE
Colon	10	90638872	90664402	ENST00000371924.1	STAMBPL1	0.20	0.07	0.13	0.09	0.13	13.17	TRUE
Colon	15	75072089	75083506	ENST00000567571.1	CSK	0.80	0.52	0.28	0.08	0.11	10.57	TRUE
Colon	10	134259441	134282106	ENST00000392630.3	C10orf91	0.26	0.12	0.14	0.26	0.32	9.61	TRUE
Colon	1	37936771	37953471	ENST00000471012.1	ZC3H12A	0.49	0.34	0.15	0.16	0.19	9.15	TRUE
Colon	19	1256417	1263739	ENST00000589161.1	CIRBP	0.65	0.52	0.13	0.15	0.18	8.78	TRUE
Colon	19	1852750	1865843	ENST00000592313.1	KLF16	0.63	0.51	0.12	0.07	0.09	8.29	TRUE
Colon	3	49056900	49060957	ENST00000480392.1	NDUFAF3	0.84	0.67	0.17	0.10	0.11	8.01	TRUE
Colon	17	27275334	27280852	ENST00000577182.1	PIPOX	1.00	0.86	0.14	0.08	0.09	7.92	TRUE
Colon	16	2028149	2036628	ENST00000567719.1	GFER	0.59	0.43	0.16	0.16	0.19	7.14	TRUE
Colon	11	130013608	130061419	ENST00000530376.1	ST14	0.22	0.05	0.16	0.21	0.26	7.07	TRUE
Colon	6	33538390	33559780	ENST00000374458.1	GGNBP1	0.27	0.16	0.12	0.13	0.15	6.84	TRUE
Colon	2	220109452	220119946	ENST00000392088.2	TUBA4A	0.53	0.40	0.12	0.07	0.09	6.77	TRUE
Colon	9	140187219	140215940	ENST00000356628.2	NRARP	0.53	0.40	0.13	0.24	0.29	5.89	TRUE
Colon	7	150052314	150083105	ENST00000486297.1	ZNF775	0.36	0.21	0.15	0.14	0.16	5.67	TRUE
Colon	14	50465553	50471375	ENST00000529902.1	C14orf182	0.66	0.45	0.22	0.11	0.12	5.40	TRUE
Colon	6	74224863	74234116	ENST00000455918.1	EEF1A1	0.90	0.58	0.32	0.10	0.12	5.13	TRUE
Colon	6	31700974	31708664	ENST00000493662.2	MSH5-SAPCD1	0.87	0.70	0.18	0.11	0.12	4.92	TRUE
Colon	1	1092807	1104591	ENST00000506177.1	TTLL10	0.35	0.19	0.16	0.29	0.34	4.42	TRUE
Colon	7	1486269	1507327	ENST00000297508.7	MICALL2	0.35	0.23	0.12	0.09	0.11	4.22	TRUE
Colon	17	4845115	4854686	ENST00000519300.1	ENO3	0.72	0.58	0.13	0.11	0.12	3.71	TRUE
Colon	16	29830344	29838029	ENST00000570234.1	MVP	0.48	0.28	0.19	0.28	0.35	3.49	TRUE
Colon	1	153503987	153511480	ENST00000462951.2	S100A6	0.39	0.26	0.13	0.07	0.09	3.38	TRUE
Colon	4	102266379	102270939	ENST00000529296.1	AP001816.1	0.94	0.80	0.14	0.05	0.06	3.03	TRUE
Colon	7	100608082	100627662	ENST00000536621.1	MUC12	0.30	0.20	0.11	0.27	0.29	2.91	TRUE
Colon	1	159888888	159896553	ENST00000397334.2	TAGLN2	0.63	0.45	0.18	0.09	0.10	2.80	TRUE
Colon	1	149819766	149826428	ENST00000403683.1	HIST2H3A	0.28	0.14	0.14	0.07	0.08	2.76	TRUE
Colon	7	27202748	27210249	ENST00000489695.1	HOXA9	0.95	0.85	0.10	0.19	0.23	2.73	TRUE
Colon	7	27210777	27221249	ENST00000396344.4	HOXA10	0.76	0.45	0.30	0.26	0.29	2.60	TRUE
Colon	19	13948788	13956320	ENST00000591727.1	NANOS3	0.69	0.55	0.15	0.24	0.29	2.55	TRUE
Colon	7	104643451	104656449	ENST00000474203.1	MLL5	0.62	0.51	0.12	0.06	0.07	2.40	TRUE
Colon	1	150531037	150543263	ENST00000369035.2	C1orf138	0.51	0.37	0.14	0.15	0.16	2.35	TRUE
Colon	1	19966426	19978402	ENST00000427894.1	NBL1	0.54	0.25	0.28	0.15	0.17	2.18	TRUE
Colon	5	180668526	180674519	ENST00000514318.1	GNB2L1	0.82	0.70	0.13	0.07	0.08	2.01	TRUE
Colon	17	57904024	57931778	ENST00000587470.1	VMP1	0.49	0.14	0.35	0.25	0.33	1.96	TRUE
Colon	11	118659296	118664159	ENST00000533239.1	DDX6	1.00	0.86	0.14	0.05	0.06	1.96	TRUE
Colon	12	7051408	7057319	ENST00000538318.1	PTPN6	0.59	0.40	0.18	0.14	0.14	1.93	TRUE
Colon	20	52195018	52241000	ENST00000540425.1	ZNF217	0.41	0.24	0.17	0.22	0.23	1.74	TRUE
Colon	6	32934747	32951807	ENST00000482838.1	BRD2	0.61	0.51	0.10	0.05	0.06	1.58	TRUE
Colon	17	7161661	7168259	ENST00000573745.1	CLDN7	0.63	0.45	0.18	0.12	0.13	1.40	TRUE
Colon	15	86121044	86133400	ENST00000560340.1	AKAP13	0.40	0.18	0.22	0.20	0.26	1.38	TRUE
Colon	2	201980044	201998492	ENST00000460961.1	CFLAR	0.57	0.44	0.13	0.12	0.13	1.25	FALSE
Colon	2	174828021	174831622	ENST00000490182.1	SP3	1.00	0.88	0.12	0.07	0.07	1.19	FALSE
Colon	17	4033794	4049258	ENST00000573984.1	CYB5D2	0.27	0.16	0.11	0.07	0.07	1.17	FALSE
Colon	20	43968274	43977802	ENST00000537976.1	SDC4	0.77	0.29	0.48	0.18	0.19	1.08	FALSE
Colon	2	173291273	173330716	ENST00000409080.1	ITGA6	0.25	0.10	0.15	0.17	0.18	0.90	FALSE
Colon	3	9436663	9444883	ENST00000406341.1	SETD5	0.78	0.65	0.13	0.09	0.10	0.86	FALSE
Colon	5	96268784	96273156	ENST00000231368.5	LNPEP	0.95	0.81	0.13	0.03	0.03	0.76	FALSE
Colon	8	103816742	103826153	ENST00000518697.1	AZIN1	0.54	0.43	0.12	0.11	0.11	0.75	FALSE
Colon	8	126438172	126449618	ENST00000519576.1	TRIB1	0.52	0.36	0.16	0.06	0.05	0.64	FALSE
Colon	17	38213719	38232994	ENST00000577486.1	THRA	0.55	0.44	0.11	0.15	0.16	0.64	FALSE
Colon	18	3446352	3459783	ENST00000472042.1	TGIF1	0.77	0.55	0.22	0.09	0.09	0.62	FALSE
Colon	1	1675407	1679081	ENST00000246421.4	SLC35E2	0.74	0.62	0.12	0.03	0.04	0.54	FALSE
Colon	3	193848397	193860794	ENST00000476918.1	HES1	0.80	0.61	0.19	0.08	0.08	0.47	FALSE
Colon	2	200306225	200342782	ENST00000463386.1	SATB2	0.56	0.40	0.16	0.16	0.17	0.46	FALSE
Colon	17	7459695	7466311	ENST00000429205.2	SENP3	0.74	0.63	0.11	0.20	0.19	0.26	FALSE
Colon	1	169069278	169085972	ENST00000367813.3	ATP1B1	0.45	0.34	0.11	0.13	0.12	0.19	FALSE
Colon	12	6559071	6563067	ENST00000543567.1	TAPBPL	0.75	0.64	0.11	0.06	0.05	-0.04	FALSE
Colon	12	56472347	56482109	ENST00000546748.1	ERBB3	0.44	0.27	0.17	0.15	0.14	-0.05	FALSE
Colon	11	307666	312937	ENST00000399815.2	IFITM2	0.70	0.55	0.14	0.16	0.16	-0.08	FALSE
Colon	3	50373024	50380264	ENST00000490675.1	ZMYND10	0.49	0.39	0.10	0.16	0.14	-0.13	FALSE
Colon	13	28526778	28556061	ENST00000548877.1	CDX2	0.94	0.72	0.23	0.23	0.23	-0.27	FALSE
Colon	2	151323963	151344439	ENST00000454202.1	RND3	0.28	0.17	0.11	0.12	0.11	-0.59	FALSE
Colon	4	38662929	38689928	ENST00000436901.1	AC021860.1	0.45	0.29	0.16	0.14	0.13	-0.59	FALSE
Colon	19	49463688	49472130	ENST00000331825.6	FTL	0.54	0.39	0.15	0.19	0.18	-0.82	FALSE
Colon	5	179244886	179250297	ENST00000360718.5	SQSTM1	0.68	0.57	0.11	0.07	0.06	-1.19	FALSE
Colon	11	71748625	71754435	ENST00000535947.1	NUMA1	0.81	0.67	0.13	0.14	0.12	-1.54	FALSE
Colon	12	53762866	53777584	ENST00000548560.1	SP1	0.41	0.29	0.12	0.11	0.10	-1.97	FALSE
Colon	11	64654606	64662582	ENST00000457202.1	EHD1	0.23	0.09	0.13	0.14	0.12	-5.94	FALSE
Lung	4	174427206	174460857	ENST00000505300.1	HAND2	0.74	0.60	0.14	0.24	0.31	9.72	TRUE
Lung	2	66659175	66674430	ENST00000475239.1	MEIS1	0.78	0.63	0.16	0.20	0.28	9.54	TRUE
Lung	1	2160113	2167961	ENST00000508416.1	SKI	0.66	0.55	0.12	0.17	0.26	5.43	TRUE
Lung	20	22536539	22565230	ENST00000319993.4	FOXA2	0.69	0.54	0.15	0.17	0.22	2.36	TRUE
Lung	16	56638742	56646907	ENST00000567300.1	MT2A	0.50	0.39	0.12	0.22	0.27	2.26	TRUE
Lung	12	6441703	6453465	ENST00000538363.1	TNFRSF1A	0.61	0.45	0.16	0.05	0.08	1.37	TRUE
Lung	12	119613625	119628282	ENST00000542496.1	HSPB8	0.22	0.10	0.12	0.14	0.19	1.36	TRUE
Lung	11	305282	312843	ENST00000399815.2	IFITM2	0.50	0.40	0.10	0.16	0.17	1.10	FALSE
Lung	1	154939647	154948901	ENST00000473344.1	CKS1B	0.61	0.48	0.12	0.15	0.17	0.69	FALSE
Lung	22	46464333	46478321	ENST00000443490.1	MIRLET7	0.85	0.71	0.15	0.20	0.23	0.32	FALSE
Lung	11	67803279	67809641	ENST00000533947.1	TCIRG1	0.45	0.34	0.11	0.10	0.12	0.29	FALSE
Lung	17	7378645	7384073	ENST00000380599.4	ZBTB4	0.65	0.45	0.20	0.22	0.24	0.06	FALSE
Lung	5	43033834	43045550	ENST00000314890.3	ANXA2R	0.57	0.43	0.14	0.20	0.21	-0.09	FALSE
Lung	19	45244285	45262369	ENST00000473473.1	BCL3	0.46	0.34	0.12	0.08	0.09	-0.21	FALSE
Lung	12	92526891	92540701	ENST00000552315.1	BTG1	0.71	0.57	0.14	0.12	0.11	-0.21	FALSE
Lung	14	105938037	105944737	ENST00000548309.1	CRIP2	0.80	0.66	0.14	0.16	0.17	-0.32	FALSE
Lung	1	145434374	145443479	ENST00000486597.1	TXNIP	0.71	0.57	0.15	0.10	0.10	-0.61	FALSE
Lung	19	13948720	13956318	ENST00000591727.1	NANOS3	0.66	0.55	0.11	0.27	0.28	-1.91	FALSE
Lung	9	35726029	35733751	ENST00000486056.1	CREB3	0.65	0.53	0.12	0.13	0.12	-2.30	FALSE

Supplementary Table 6.3: Hypermethylated super-enhancers in cancer samples (δ HMR > 25%).

	Super-Enhancer			HMR coverage (%)			
Cancer type	**Chr.**	**Start**	**End**	**Normal**	**Cancer**	**δ**	**Gene symbol**
Primary breast tumor (468PT)	7	27134577	27144793	0.95	0.22	0.73	HOXA2
Primary breast tumor (468PT)	5	1881509	1890183	0.97	0.29	0.67	IRX4
Primary breast tumor (468PT)	3	29321550	29334286	0.65	0.00	0.65	RBMS3
Primary breast tumor (468PT)	5	142773722	142785691	0.95	0.32	0.63	NR3C1
Primary breast tumor (468PT)	11	8828385	8837469	0.63	0.00	0.63	ST5
Primary breast tumor (468PT)	22	46467114	46488203	1.00	0.39	0.61	MIRLET7
Primary breast tumor (468PT)	5	2739262	2759651	0.92	0.31	0.60	IRX2
Primary breast tumor (468PT)	5	131593100	131603174	0.60	0.00	0.60	PDLIM4
Primary breast tumor (468PT)	22	46448058	46467049	1.00	0.50	0.50	MIRLET7
Primary breast tumor (468PT)	7	27175056	27185979	0.61	0.17	0.44	HOXA5
Primary breast tumor (468PT)	21	36252652	36264673	0.93	0.50	0.42	RUNX1
Primary breast tumor (468PT)	1	45270631	45276266	0.65	0.25	0.41	BTBD19
Primary breast tumor (468PT)	1	33792562	33815439	0.49	0.10	0.39	PHC2
Primary breast tumor (468PT)	10	122913683	122920721	0.38	0.00	0.38	WDR11
Primary breast tumor (468PT)	20	51581498	51597746	0.45	0.08	0.37	TSHZ2
Primary breast tumor (468PT)	14	74207241	74227902	0.58	0.22	0.36	MIR4505
Primary breast tumor (468PT)	12	92527863	92539838	0.63	0.28	0.35	BTG1
Primary breast tumor (468PT)	1	154941228	154949531	0.82	0.49	0.34	MIR4258
Primary breast tumor (468PT)	7	41734099	41745805	0.40	0.07	0.34	INHBA
Primary breast tumor (468PT)	11	6339234	6344196	0.61	0.28	0.32	PRKCDBP
Primary breast tumor (468PT)	1	144706472	144710218	0.32	0.00	0.32	PDE4DIP
Primary breast tumor (468PT)	1	27830256	27900464	0.37	0.05	0.32	WASF2
Primary breast tumor (468PT)	15	93425946	93432559	0.87	0.56	0.31	CHD2
Primary breast tumor (468PT)	20	34326330	34331841	0.91	0.61	0.30	RBM39
Primary breast tumor (468PT)	15	50644604	50649357	0.96	0.66	0.29	GABPB1
Primary breast tumor (468PT)	12	6432600	6452223	0.50	0.21	0.29	TNFRSF1A
Primary breast tumor (468PT)	3	71104592	71120542	0.41	0.12	0.28	FOXP1
Primary breast tumor (468PT)	1	157964217	157991843	0.50	0.22	0.28	KIRREL
Primary breast tumor (468PT)	1	144981791	144985538	0.28	0.00	0.28	PDE4DIP
Primary breast tumor (468PT)	14	69402525	69445588	0.35	0.08	0.28	ACTN1
Primary breast tumor (468PT)	19	42771661	42788152	0.57	0.30	0.27	ERF
Primary breast tumor (468PT)	7	74264047	74268456	0.26	0.00	0.26	GTF2IRD2
Primary breast tumor (468PT)	10	126406768	126435377	0.34	0.08	0.26	FAM53B
Breast tumor metastasis (468LN)	5	2739262	2759651	0.92	0.13	0.79	IRX2
Breast tumor metastasis (468LN)	5	1881509	1890183	0.97	0.21	0.76	IRX4
Breast tumor metastasis (468LN)	7	27134577	27144793	0.95	0.21	0.74	HOXA2
Breast tumor metastasis (468LN)	1	45270631	45276266	0.65	0.00	0.65	BTBD19
Breast tumor metastasis (468LN)	5	131593100	131603174	0.60	0.00	0.60	PDLIM4
Breast tumor metastasis (468LN)	2	239193425	239200323	0.73	0.15	0.59	PER2
Breast tumor metastasis (468LN)	12	6479020	6487672	0.57	0.00	0.57	LTBR
Breast tumor metastasis (468LN)	9	33157786	33167403	0.92	0.36	0.56	B4GALT1
Breast tumor metastasis (468LN)	18	3622231	3626476	0.52	0.00	0.52	DLGAP1
Breast tumor metastasis (468LN)	16	54959639	54972666	0.93	0.43	0.51	IRX5
Breast tumor metastasis (468LN)	7	27175056	27185979	0.61	0.11	0.49	HOXA5
Breast tumor metastasis (468LN)	7	5457613	5470399	0.98	0.49	0.49	TNRC18
Breast tumor metastasis (468LN)	4	146653244	146658422	0.68	0.21	0.47	MMAA
Breast tumor metastasis (468LN)	3	193849125	193860857	0.79	0.34	0.46	HES1
Breast tumor metastasis (468LN)	12	10868156	10877012	0.79	0.35	0.44	CSDA
Breast tumor metastasis (468LN)	1	153579895	153590703	0.49	0.06	0.43	S100A14
Breast tumor metastasis (468LN)	1	59245181	59252471	0.86	0.44	0.43	JUN
Breast tumor metastasis (468LN)	11	62306073	62328249	0.61	0.20	0.42	AHNAK
Breast tumor metastasis (468LN)	3	71104592	71120542	0.41	0.00	0.41	FOXP1
Breast tumor metastasis (468LN)	10	122913683	122920721	0.38	0.00	0.38	WDR11
Breast tumor metastasis (468LN)	6	121756401	121769000	0.38	0.00	0.38	GJA1
Breast tumor metastasis (468LN)	4	128702346	128706700	0.55	0.17	0.38	HSPA4L
Breast tumor metastasis (468LN)	20	10634437	10656669	0.49	0.11	0.37	JAG1
Breast tumor metastasis (468LN)	6	35692791	35701120	0.55	0.18	0.37	FKBP5
Breast tumor metastasis (468LN)	22	46448058	46467049	1.00	0.63	0.37	MIRLET7
Breast tumor metastasis (468LN)	11	8828385	8837469	0.63	0.26	0.37	ST5
Breast tumor metastasis (468LN)	3	29321550	29334286	0.65	0.29	0.36	RBMS3
Breast tumor metastasis (468LN)	19	13123147	13172690	0.50	0.14	0.36	TRMT1
Breast tumor metastasis (468LN)	14	77489392	77513614	0.67	0.31	0.35	C14orf4
Breast tumor metastasis (468LN)	1	27830256	27900464	0.37	0.02	0.35	WASF2
Breast tumor metastasis (468LN)	12	115105999	115123428	0.60	0.24	0.35	TBX3
Breast tumor metastasis (468LN)	11	66821179	66830989	0.33	0.00	0.33	RHOD
Breast tumor metastasis (468LN)	11	65238595	65276583	0.63	0.31	0.32	MALAT1
Breast tumor metastasis (468LN)	1	144706472	144710218	0.32	0.00	0.32	PDE4DIP
Breast tumor metastasis (468LN)	15	93425946	93432559	0.87	0.56	0.31	CHD2
Breast tumor metastasis (468LN)	7	55085115	55095372	0.50	0.19	0.31	EGFR
Breast tumor metastasis (468LN)	1	37936765	37953189	0.42	0.11	0.31	ZC3H12A
Breast tumor metastasis (468LN)	4	77506454	77513253	0.31	0.00	0.31	STBD1
Breast tumor metastasis (468LN)	7	41734099	41745805	0.40	0.10	0.31	INHBA
Breast tumor metastasis (468LN)	8	145008274	145031982	0.62	0.32	0.30	PLEC1
Breast tumor metastasis (468LN)	17	75274996	75288850	0.30	0.01	0.29	SEPT9
Breast tumor metastasis (468LN)	6	10398835	10421638	0.81	0.52	0.29	TFAP2A
Breast tumor metastasis (468LN)	12	6432600	6452223	0.50	0.21	0.29	TNFRSF1A
Breast tumor metastasis (468LN)	1	144981791	144985538	0.28	0.00	0.28	PDE4DIP
Breast tumor metastasis (468LN)	10	74051120	74097380	0.27	0.00	0.27	DDIT4
Breast tumor metastasis (468LN)	22	36718824	36814131	0.34	0.07	0.27	MYH9
Breast tumor metastasis (468LN)	8	116659885	116682409	0.42	0.15	0.27	TRPS1
Breast tumor metastasis (468LN)	6	30843042	30858880	0.32	0.05	0.27	DDR1
Breast tumor metastasis (468LN)	5	95291493	95300367	0.41	0.14	0.27	ELL2
Breast tumor metastasis (468LN)	5	121515651	121520476	0.26	0.00	0.26	ZNF474
Breast tumor metastasis (468LN)	3	187453456	187468930	0.81	0.56	0.26	BCL6
Breast tumor metastasis (468LN)	2	36579464	36605312	0.55	0.29	0.26	CRIM1
Breast tumor metastasis (468LN)	12	46653450	46664706	0.32	0.07	0.25	SLC38A1
Breast tumor metastasis (468LN)	19	12888234	12905680	0.68	0.43	0.25	JUNB
Lung adenocarcinoma (H1437)	5	2738590	2759482	0.83	0.10	0.74	IRX2
Lung adenocarcinoma (H1437)	5	172655376	172676296	0.74	0.17	0.58	NKX2-5
Lung adenocarcinoma (H1437)	3	128199015	128216666	0.79	0.25	0.53	GATA2
Lung adenocarcinoma (H1437)	4	174427206	174460857	0.74	0.25	0.48	HAND2
Lung adenocarcinoma (H1437)	1	145434374	145443479	0.71	0.24	0.48	TXNIP
Lung adenocarcinoma (H1437)	2	66659175	66674430	0.78	0.33	0.45	MEIS1
Lung adenocarcinoma (H1437)	19	13122379	13134630	0.43	0.04	0.39	NFIX
Lung adenocarcinoma (H1437)	1	145448015	145457936	0.37	0.00	0.37	TXNIP
Lung adenocarcinoma (H1437)	17	7378645	7384073	0.65	0.29	0.36	ZBTB4
Lung adenocarcinoma (H1437)	14	105938037	105944737	0.80	0.46	0.34	CRIP2
Lung adenocarcinoma (H1437)	9	35726029	35733751	0.65	0.34	0.31	CREB3
Lung adenocarcinoma (H1437)	22	46445515	46454339	0.63	0.34	0.29	MIRLET7
Lung adenocarcinoma (H1437)	22	46464333	46478321	0.85	0.58	0.27	MIRLET7
Lung adenocarcinoma (H1437)	10	77154286	77170840	0.72	0.45	0.26	ZNF503
Lung squamous cell carcinoma (H157)	4	174427206	174460857	0.74	0.07	0.67	HAND2
Lung squamous cell carcinoma (H157)	5	2738590	2759482	0.83	0.20	0.63	IRX2
Lung squamous cell carcinoma (H157)	3	128199015	128216666	0.79	0.16	0.63	GATA2
Lung squamous cell carcinoma (H157)	2	8815972	8826211	0.96	0.37	0.59	ID2
Lung squamous cell carcinoma (H157)	2	66659175	66674430	0.78	0.26	0.52	MEIS1
Lung squamous cell carcinoma (H157)	14	36971679	36995027	1.00	0.51	0.49	SFTA3
Lung squamous cell carcinoma (H157)	1	145434374	145443479	0.71	0.23	0.48	TXNIP
Lung squamous cell carcinoma (H157)	12	92526891	92540701	0.71	0.27	0.43	BTG1
Lung squamous cell carcinoma (H157)	19	13122379	13134630	0.43	0.00	0.43	NFIX
Lung squamous cell carcinoma (H157)	12	6441703	6453465	0.61	0.23	0.39	TNFRSF1A
Lung squamous cell carcinoma (H157)	11	118659939	118664216	0.92	0.54	0.38	DDX6

Lung squamous cell carcinoma (H157)	10	77154286	77170840	0.72	0.33	0.38	ZNF503
Lung squamous cell carcinoma (H157)	17	7378645	7384073	0.65	0.28	0.37	ZBTB4
Lung squamous cell carcinoma (H157)	1	145448015	145457936	0.37	0.00	0.37	TXNIP
Lung squamous cell carcinoma (H157)	12	115105978	115131359	0.52	0.15	0.36	TBX3
Lung squamous cell carcinoma (H157)	2	201724772	201732443	0.74	0.39	0.35	Y_RNA
Lung squamous cell carcinoma (H157)	17	38222943	38234141	0.48	0.13	0.35	THRA
Lung squamous cell carcinoma (H157)	12	7032586	7039696	0.64	0.30	0.34	ATN1
Lung squamous cell carcinoma (H157)	9	35726029	35733751	0.65	0.33	0.32	CREB3
Lung squamous cell carcinoma (H157)	22	46464333	46478321	0.85	0.53	0.32	MIRLET7
Lung squamous cell carcinoma (H157)	16	54956715	54973307	0.73	0.41	0.32	IRX5
Lung squamous cell carcinoma (H157)	14	105938037	105944737	0.80	0.48	0.32	CRIP2
Lung squamous cell carcinoma (H157)	17	7459853	7466333	0.71	0.44	0.26	SENP3
Lung squamous cell carcinoma (H157)	22	46445515	46454339	0.63	0.37	0.26	MIRLET7
Lung squamous cell carcinoma (H157)	5	42990572	43020949	0.39	0.13	0.26	ANXA2R
Lung squamous cell carcinoma (H157)	3	48504418	48515502	0.26	0.00	0.26	SHISA5
Small cell lung cancer (H1672)	11	65182856	65196355	0.89	0.13	0.76	NEAT1
Small cell lung cancer (H1672)	14	36971679	36995027	1.00	0.31	0.69	SFTA3
Small cell lung cancer (H1672)	12	6441703	6453465	0.61	0.00	0.61	TNFRSF1A
Small cell lung cancer (H1672)	19	13948720	13956318	0.66	0.13	0.52	MIR23
Small cell lung cancer (H1672)	20	22536539	22565230	0.69	0.17	0.51	FOXA2
Small cell lung cancer (H1672)	4	174427206	174460857	0.74	0.23	0.50	HAND2
Small cell lung cancer (H1672)	3	128199015	128216666	0.79	0.29	0.50	GATA2
Small cell lung cancer (H1672)	17	38222943	38234141	0.48	0.00	0.48	THRA
Small cell lung cancer (H1672)	22	46464333	46478321	0.85	0.38	0.47	MIRLET7
Small cell lung cancer (H1672)	14	105938037	105944737	0.80	0.33	0.47	CRIP2
Small cell lung cancer (H1672)	11	118777304	118801687	0.71	0.25	0.46	UPK2
Small cell lung cancer (H1672)	2	66659175	66674430	0.78	0.33	0.45	MEIS1
Small cell lung cancer (H1672)	22	46445515	46454339	0.63	0.19	0.45	MIRLET7
Small cell lung cancer (H1672)	12	114806250	114852976	0.45	0.02	0.43	TBX5
Small cell lung cancer (H1672)	16	54956715	54973307	0.73	0.33	0.40	IRX5
Small cell lung cancer (H1672)	6	30646634	30659820	0.63	0.23	0.39	NRM
Small cell lung cancer (H1672)	12	92526891	92540701	0.71	0.33	0.38	BTG1
Small cell lung cancer (H1672)	11	67803279	67809641	0.45	0.09	0.37	TCIRG1
Small cell lung cancer (H1672)	1	145448015	145457936	0.37	0.00	0.37	TXNIP
Small cell lung cancer (H1672)	9	35726029	35733751	0.65	0.29	0.36	CREB3
Small cell lung cancer (H1672)	17	7378645	7384073	0.65	0.29	0.36	ZBTB4
Small cell lung cancer (H1672)	17	79475627	79487805	0.63	0.28	0.36	ACTG1
Small cell lung cancer (H1672)	5	2738590	2759482	0.83	0.48	0.35	IRX2
Small cell lung cancer (H1672)	11	118659939	118664216	0.92	0.57	0.35	DDX6
Small cell lung cancer (H1672)	2	201724772	201732443	0.74	0.40	0.35	Y_RNA
Small cell lung cancer (H1672)	19	1256391	1263727	0.50	0.16	0.34	CIRBP
Small cell lung cancer (H1672)	19	13122379	13134630	0.43	0.10	0.34	NFIX
Small cell lung cancer (H1672)	1	153537691	153541885	0.41	0.10	0.32	S100A2
Small cell lung cancer (H1672)	1	1364851	1372426	0.46	0.14	0.32	VWA1
Small cell lung cancer (H1672)	18	3591992	3606653	0.46	0.15	0.31	DLGAP1
Small cell lung cancer (H1672)	22	38708087	38715695	0.57	0.27	0.31	CSNK1E
Small cell lung cancer (H1672)	6	31694582	31706840	0.45	0.14	0.30	CLIC1
Small cell lung cancer (H1672)	19	1247145	1256325	0.68	0.38	0.30	MIDN
Small cell lung cancer (H1672)	19	41219787	41228823	0.64	0.35	0.29	ITPKC
Small cell lung cancer (H1672)	10	77154286	77170840	0.72	0.43	0.29	ZNF503
Small cell lung cancer (H1672)	1	145434374	145443479	0.71	0.42	0.29	TXNIP
Small cell lung cancer (H1672)	7	1569896	1580027	0.40	0.11	0.28	MAFK
Small cell lung cancer (H1672)	9	14302438	14323420	0.55	0.27	0.28	NFIB
Small cell lung cancer (H1672)	12	115105978	115131359	0.52	0.24	0.28	TBX3
Small cell lung cancer (H1672)	10	88726022	88733482	0.37	0.10	0.27	AGAP11
Small cell lung cancer (H1672)	5	172655376	172676296	0.74	0.48	0.26	NKX2-5
Small cell lung cancer (H1672)	16	56638742	56646907	0.50	0.25	0.25	MT2A
Small cell lung cancer (H1672)	9	14340900	14353877	0.32	0.06	0.25	NFIB
Primary colon tumor (Colon_P))	17	46631233	46657163	0.43	0.09	0.34	HOXB3
Primary colon tumor (Colon_P)	7	27178057	27189568	0.69	0.39	0.30	HOXA6
Primary colon tumor (Colon_P)	5	134361067	134376085	0.64	0.36	0.29	PITX1
Primary colon metastasis (Colon_M)	7	27178057	27189568	0.69	0.36	0.33	HOXA6
Primary colon metastasis (Colon_M)	17	46631233	46657163	0.43	0.11	0.32	HOXB3
Primary colon metastasis (Colon_M)	5	134361067	134376085	0.64	0.36	0.28	PITX1
Primary colon metastasis (Colon_M)	15	75072089	75083506	0.80	0.53	0.27	CSK
Glioblastoma (U87MG)	1	161160323	161173936	0.80	0.07	0.73	NDUFS2
Glioblastoma (U87MG)	6	163815822	163882069	0.79	0.14	0.65	QKI
Glioblastoma (U87MG)	7	29516606	29530680	0.63	0.00	0.63	CHN2
Glioblastoma (U87MG)	13	78291692	78329377	0.59	0.00	0.59	SLAIN1
Glioblastoma (U87MG)	9	130697475	130743184	0.61	0.07	0.55	FAM102A
Glioblastoma (U87MG)	18	74684916	74780601	0.53	0.00	0.53	MBP
Glioblastoma (U87MG)	17	79097124	79111085	0.59	0.09	0.50	AATK
Glioblastoma (U87MG)	9	131142171	131157912	0.49	0.00	0.49	MIR219-2
Glioblastoma (U87MG)	11	61520065	61534843	0.53	0.04	0.49	MYRF
Glioblastoma (U87MG)	10	14577374	14624047	0.46	0.00	0.46	FAM107B
Glioblastoma (U87MG)	3	33685210	33704614	0.45	0.00	0.45	CLASP2
Glioblastoma (U87MG)	13	67782754	67806323	0.59	0.14	0.45	PCDH9
Glioblastoma (U87MG)	22	38361166	38471689	0.39	0.01	0.39	SOX10
Glioblastoma (U87MG)	17	36570707	36622668	0.54	0.16	0.39	AC124789.1
Glioblastoma (U87MG)	6	134488737	134505092	0.60	0.22	0.38	SGK1
Glioblastoma (U87MG)	11	67173218	67188540	0.45	0.06	0.38	CARNS1
Glioblastoma (U87MG)	10	81137824	81212605	0.43	0.05	0.38	ZCCHC24
Glioblastoma (U87MG)	8	26427962	26519962	0.45	0.08	0.37	DPYSL2
Glioblastoma (U87MG)	7	22212818	22261220	0.37	0.00	0.37	RAPGEF5
Glioblastoma (U87MG)	3	181403893	181455295	0.84	0.47	0.36	SOX2
Glioblastoma (U87MG)	17	38216903	38234316	0.63	0.27	0.36	THRA
Glioblastoma (U87MG)	18	60417206	60470491	0.35	0.00	0.35	PHLPP1
Glioblastoma (U87MG)	12	48167031	48182319	0.40	0.06	0.35	AC004466.1
Glioblastoma (U87MG)	9	140081775	140094696	0.56	0.21	0.35	TPRN
Glioblastoma (U87MG)	3	170135527	170167574	0.34	0.00	0.34	CLDN11
Glioblastoma (U87MG)	3	72424289	72464834	0.34	0.00	0.34	RYBP
Glioblastoma (U87MG)	2	145129528	145283099	0.39	0.05	0.34	ZEB2-AS1
Glioblastoma (U87MG)	10	88421359	88444547	0.33	0.00	0.33	LDB3
Glioblastoma (U87MG)	5	36599886	36615091	0.50	0.17	0.33	SLC1A3
Glioblastoma (U87MG)	22	37940770	37968154	0.50	0.18	0.32	CDC42EP1
Glioblastoma (U87MG)	2	127806748	127912517	0.36	0.04	0.32	BIN1
Glioblastoma (U87MG)	4	154142334	154219000	0.38	0.06	0.32	Y_RNA
Glioblastoma (U87MG)	6	163659272	163687676	0.36	0.04	0.32	snoU13
Glioblastoma (U87MG)	19	13093539	13211813	0.47	0.16	0.31	NFIX
Glioblastoma (U87MG)	1	202090299	202115217	0.39	0.08	0.31	ARL8A
Glioblastoma (U87MG)	16	19841311	19898686	0.35	0.04	0.31	GPRC5B
Glioblastoma (U87MG)	14	29226576	29255674	0.61	0.30	0.30	RP11-966I7.2
Glioblastoma (U87MG)	10	126400866	126432679	0.40	0.10	0.30	FAM53B
Glioblastoma (U87MG)	17	17638844	17657613	0.30	0.00	0.30	RAI1-AS1
Glioblastoma (U87MG)	15	64974641	64996542	0.36	0.06	0.30	RP11-330L19.4
Glioblastoma (U87MG)	16	75266933	75302422	0.51	0.22	0.29	BCAR1
Glioblastoma (U87MG)	2	232526039	232554844	0.34	0.05	0.29	Metazoa_SRP
Glioblastoma (U87MG)	13	107142225	107162158	0.29	0.00	0.29	EFNB2
Glioblastoma (U87MG)	17	68152783	68183190	0.38	0.09	0.29	KCNJ2
Glioblastoma (U87MG)	10	102752997	102775915	0.44	0.16	0.28	LZTS2
Glioblastoma (U87MG)	14	77765872	77789479	0.39	0.10	0.28	GSTZ1
Glioblastoma (U87MG)	8	120648911	120687423	0.41	0.13	0.28	ENPP2
Glioblastoma (U87MG)	18	13610336	13640017	0.53	0.25	0.28	LDLRAD4
Glioblastoma (U87MG)	17	79345155	79376776	0.54	0.27	0.27	MIR4740
Glioblastoma (U87MG)	4	115547254	115584118	0.27	0.00	0.27	MIR577
Glioblastoma (U87MG)	16	15234031	15244187	0.27	0.00	0.27	Y_RNA
Glioblastoma (U87MG)	5	132105879	132116081	0.43	0.16	0.27	8-Sep
Glioblastoma (U87MG)	10	650150	686281	0.31	0.04	0.26	RP11-809C18.3
Glioblastoma (U87MG)	9	135976307	136024639	0.33	0.07	0.26	RALGDS
Glioblastoma (U87MG)	11	111845328	111865086	0.33	0.07	0.25	DIXDC1
Glioblastoma (U87MG)	10	22858862	22973344	0.25	0.00	0.25	PIP4K2A
Glioblastoma (U87MG)	1	198871424	198907549	0.48	0.23	0.25	RP11-31E23.1

Supplementary Table 6.4: Hypomethylated super-enhancers in cancer samples (δ HMR > 25%).

	Super-Enhancer			HMR coverage (%)			
Tissue	Chr.	Start	End	Normal	Cancer	δ	Gene symbol
Primary breast tumor (468PT)	2	91785127	91788546	0.00	1.00	1.00	AC233263.1
Primary breast tumor (468PT)	7	107947705	107953802	0.00	1.00	1.00	NRCAM
Primary breast tumor (468PT)	16	72320261	72327455	0.00	1.00	1.00	PMFBP1
Primary breast tumor (468PT)	3	74578493	74588238	0.00	1.00	1.00	CNTN3
Primary breast tumor (468PT)	10	118176539	118201896	0.00	0.99	0.99	PNLIPRP3
Primary breast tumor (468PT)	11	131923942	131950946	0.00	0.98	0.98	NTM
Primary breast tumor (468PT)	20	22354832	22372608	0.00	0.96	0.96	FOXA2
Primary breast tumor (468PT)	20	22354832	22372608	0.00	0.96	0.96	FOXA2
Primary breast tumor (468PT)	8	32082261	32138267	0.02	0.97	0.94	NRG1
Primary breast tumor (468PT)	8	32082261	32138267	0.02	0.97	0.94	NRG1
Primary breast tumor (468PT)	8	32082261	32138267	0.02	0.97	0.94	NRG1
Primary breast tumor (468PT)	8	32082261	32138267	0.02	0.97	0.94	NRG1
Primary breast tumor (468PT)	8	32082261	32138267	0.02	0.97	0.94	NRG1
Primary breast tumor (468PT)	8	32082261	32138267	0.02	0.97	0.94	NRG1
Primary breast tumor (468PT)	5	167229136	167249797	0.06	1.00	0.94	TENM2
Primary breast tumor (468PT)	4	166884826	166896752	0.00	0.93	0.93	TLL1
Primary breast tumor (468PT)	11	41071460	41077058	0.00	0.91	0.91	LRRC4C
Primary breast tumor (468PT)	2	17760510	17781417	0.06	0.97	0.91	VSNL1
Primary breast tumor (468PT)	21	28941125	28955796	0.00	0.90	0.90	ADAMTS5
Primary breast tumor (468PT)	15	84511818	84524836	0.06	0.97	0.90	ADAMTSL3
Primary breast tumor (468PT)	16	82659814	82694376	0.05	0.94	0.89	CDH13
Primary breast tumor (468PT)	8	32160365	32202004	0.05	0.93	0.87	NRG1
Primary breast tumor (468PT)	8	32160365	32202004	0.05	0.93	0.87	NRG1
Primary breast tumor (468PT)	8	32160365	32202004	0.05	0.93	0.87	NRG1
Primary breast tumor (468PT)	8	32160365	32202004	0.05	0.93	0.87	NRG1
Primary breast tumor (468PT)	8	32160365	32202004	0.05	0.93	0.87	NRG1
Primary breast tumor (468PT)	8	32160365	32202004	0.05	0.93	0.87	NRG1
Primary breast tumor (468PT)	12	29884879	29936970	0.04	0.90	0.86	TMTC1
Primary breast tumor (468PT)	3	74647878	74690393	0.05	0.91	0.86	CNTN3
Primary breast tumor (468PT)	11	122026659	122068967	0.07	0.93	0.86	BLID
Primary breast tumor (468PT)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Primary breast tumor (468PT)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Primary breast tumor (468PT)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Primary breast tumor (468PT)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Primary breast tumor (468PT)	5	61211861	61217024	0.00	0.83	0.83	C5orf64
Primary breast tumor (468PT)	11	109930034	109968654	0.06	0.87	0.81	ZC3H12C
Primary breast tumor (468PT)	10	108182232	108202133	0.00	0.81	0.81	SORCS1
Primary breast tumor (468PT)	5	167124831	167172069	0.01	0.81	0.80	TENM2
Primary breast tumor (468PT)	12	18411546	18429687	0.00	0.80	0.80	PIK3C2G
Primary breast tumor (468PT)	4	177902138	177945974	0.00	0.76	0.76	VEGFC
Primary breast tumor (468PT)	3	115502074	115525187	0.08	0.84	0.75	LSAMP
Primary breast tumor (468PT)	8	49607646	49623337	0.07	0.82	0.74	EFCAB1
Primary breast tumor (468PT)	8	123139023	123153617	0.09	0.81	0.72	HAS2
Primary breast tumor (468PT)	7	55760378	55777116	0.00	0.68	0.68	VOPP1
Primary breast tumor (468PT)	8	32403683	32442699	0.19	0.86	0.66	NRG1
Primary breast tumor (468PT)	20	1789115	1811959	0.13	0.77	0.64	SIRPA
Primary breast tumor (468PT)	21	39831004	39862029	0.02	0.65	0.63	ERG
Primary breast tumor (468PT)	17	5729361	5740922	0.00	0.61	0.61	WSCD1
Primary breast tumor (468PT)	3	62571130	62594277	0.00	0.53	0.53	CADPS
Primary breast tumor (468PT)	5	1930374	1976494	0.12	0.63	0.50	IRX4
Primary breast tumor (468PT)	10	92593455	92620169	0.06	0.55	0.49	HTR7
Primary breast tumor (468PT)	11	114162169	114180527	0.26	0.75	0.49	NNMT
Primary breast tumor (468PT)	6	121756401	121769000	0.38	0.86	0.48	GJA1
Primary breast tumor (468PT)	1	162875345	162886688	0.00	0.45	0.45	C1orf110
Primary breast tumor (468PT)	1	162875345	162886688	0.00	0.45	0.45	C1orf110
Primary breast tumor (468PT)	5	149314305	149321676	0.00	0.38	0.38	PDE6A
Primary breast tumor (468PT)	5	149314305	149321676	0.00	0.38	0.38	PDE6A
Primary breast tumor (468PT)	3	193545523	193550492	0.00	0.37	0.37	OPA1
Primary breast tumor (468PT)	17	39674641	39706091	0.17	0.52	0.35	KRT19
Primary breast tumor (468PT)	2	143614960	143638844	0.02	0.35	0.34	KYNU
Primary breast tumor (468PT)	10	65310940	65315929	0.00	0.33	0.33	REEP3
Primary breast tumor (468PT)	10	65310940	65315929	0.00	0.33	0.33	REEP3
Primary breast tumor (468PT)	15	83945999	83992250	0.08	0.41	0.33	BNC1
Primary breast tumor (468PT)	17	57902558	57931624	0.25	0.56	0.31	VMP1
Primary breast tumor (468PT)	7	130643572	130647502	0.19	0.49	0.30	MKLN1
Primary breast tumor (468PT)	1	218648751	218673670	0.00	0.29	0.29	C1orf143
Primary breast tumor (468PT)	7	18547133	18551508	0.00	0.29	0.29	HDAC9
Primary breast tumor (468PT)	17	18621223	18626374	0.00	0.29	0.29	TRIM16L
Primary breast tumor (468PT)	11	15986587	15993130	0.00	0.29	0.29	SOX6
Primary breast tumor (468PT)	13	73629841	73641096	0.29	0.57	0.28	KLF5
Primary breast tumor (468PT)	1	8062975	8088860	0.25	0.51	0.27	ERRFI1
Primary breast tumor (468PT)	1	160982545	160992918	0.21	0.46	0.25	F11R
Breast tumor metastasis (468LN)	2	91785127	91788546	0.00	1.00	1.00	AC233263.1
Breast tumor metastasis (468LN)	2	213706261	213711784	0.00	1.00	1.00	IKZF2
Breast tumor metastasis (468LN)	3	74578493	74588238	0.00	1.00	1.00	CNTN3
Breast tumor metastasis (468LN)	20	22354832	22372608	0.00	1.00	1.00	FOXA2
Breast tumor metastasis (468LN)	20	22354832	22372608	0.00	1.00	1.00	FOXA2
Breast tumor metastasis (468LN)	4	166884826	166896752	0.00	1.00	1.00	TLL1
Breast tumor metastasis (468LN)	17	5729361	5740922	0.00	0.99	0.99	WSCD1
Breast tumor metastasis (468LN)	13	106796195	106810591	0.00	0.97	0.97	EFNB2
Breast tumor metastasis (468LN)	4	74869808	74873810	0.00	0.96	0.96	CXCL5
Breast tumor metastasis (468LN)	3	62571130	62594277	0.00	0.95	0.95	CADPS
Breast tumor metastasis (468LN)	1	115150842	115162363	0.00	0.95	0.95	DENND2C
Breast tumor metastasis (468LN)	5	66176502	66195546	0.00	0.94	0.94	MAST4
Breast tumor metastasis (468LN)	5	167229136	167249797	0.06	1.00	0.94	TENM2
Breast tumor metastasis (468LN)	5	167124831	167172069	0.01	0.93	0.93	TENM2
Breast tumor metastasis (468LN)	11	41071460	41077058	0.00	0.91	0.91	LRRC4C
Breast tumor metastasis (468LN)	16	82089108	82117271	0.00	0.91	0.91	HSD17B2
Breast tumor metastasis (468LN)	11	93740640	93763165	0.05	0.96	0.90	HEPHL1
Breast tumor metastasis (468LN)	3	74647878	74690393	0.05	0.93	0.88	CNTN3
Breast tumor metastasis (468LN)	12	18411546	18429687	0.00	0.88	0.88	PIK3C2G
Breast tumor metastasis (468LN)	15	84511818	84524836	0.06	0.94	0.88	ADAMTSL3
Breast tumor metastasis (468LN)	12	52886804	52920914	0.07	0.95	0.88	KRT5
Breast tumor metastasis (468LN)	19	46683563	46726215	0.07	0.94	0.86	IGFL1
Breast tumor metastasis (468LN)	5	73526769	73548658	0.00	0.86	0.86	ARHGEF28
Breast tumor metastasis (468LN)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Breast tumor metastasis (468LN)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Breast tumor metastasis (468LN)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Breast tumor metastasis (468LN)	12	19149492	19185438	0.00	0.85	0.85	PLEKHA5
Breast tumor metastasis (468LN)	11	119990825	120009715	0.09	0.94	0.85	TRIM29
Breast tumor metastasis (468LN)	18	10411107	10423068	0.04	0.89	0.85	APCDD1
Breast tumor metastasis (468LN)	18	38445567	38458816	0.00	0.84	0.84	PIK3C3
Breast tumor metastasis (468LN)	5	169370816	169407860	0.09	0.92	0.83	FAM196B
Breast tumor metastasis (468LN)	6	12313555	12330214	0.04	0.87	0.83	EDN1
Breast tumor metastasis (468LN)	5	61211861	61217024	0.00	0.83	0.83	C5orf64
Breast tumor metastasis (468LN)	8	32082261	32138267	0.02	0.84	0.82	NRG1
Breast tumor metastasis (468LN)	8	32082261	32138267	0.02	0.84	0.82	NRG1
Breast tumor metastasis (468LN)	8	32082261	32138267	0.02	0.84	0.82	NRG1
Breast tumor metastasis (468LN)	8	32082261	32138267	0.02	0.84	0.82	NRG1
Breast tumor metastasis (468LN)	8	32082261	32138267	0.02	0.84	0.82	NRG1
Breast tumor metastasis (468LN)	8	32082261	32138267	0.02	0.84	0.82	NRG1
Breast tumor metastasis (468LN)	7	18547133	18551508	0.00	0.82	0.82	HDAC9
Breast tumor metastasis (468LN)	4	13908076	13924205	0.00	0.81	0.81	BOD1L1
Breast tumor metastasis (468LN)	6	143645048	143661635	0.00	0.81	0.81	AL031320.1
Breast tumor metastasis (468LN)	2	70570838	70616677	0.00	0.80	0.80	FAM136A
Breast tumor metastasis (468LN)	9	91379222	91406810	0.00	0.80	0.80	C9orf47
Breast tumor metastasis (468LN)	7	55000692	55038342	0.02	0.81	0.79	EGFR
Breast tumor metastasis (468LN)	4	74454199	74487257	0.06	0.85	0.79	RASSF6
Breast tumor metastasis (468LN)	2	121929305	121969683	0.03	0.81	0.78	TFCP2L1
Breast tumor metastasis (468LN)	15	71567004	71589795	0.10	0.88	0.78	THSD4
Breast tumor metastasis (468LN)	17	69398086	69438343	0.00	0.77	0.77	AC007461.1
Breast tumor metastasis (468LN)	12	70358689	70362371	0.00	0.77	0.77	MYRFL
Breast tumor metastasis (468LN)	8	32160365	32202004	0.05	0.81	0.76	NRG1
Breast tumor metastasis (468LN)	8	32160365	32202004	0.05	0.81	0.76	NRG1
Breast tumor metastasis (468LN)	8	32160365	32202004	0.05	0.81	0.76	NRG1
Breast tumor metastasis (468LN)	8	32160365	32202004	0.05	0.81	0.76	NRG1
Breast tumor metastasis (468LN)	8	32160365	32202004	0.05	0.81	0.76	NRG1
Breast tumor metastasis (468LN)	8	32160365	32202004	0.05	0.81	0.76	NRG1
Breast tumor metastasis (468LN)	5	1930374	1976494	0.12	0.88	0.76	IRX4
Breast tumor metastasis (468LN)	20	1445625	1487649	0.07	0.83	0.76	SIRPB2
Breast tumor metastasis (468LN)	10	108182232	108202133	0.00	0.76	0.76	SORCS1
Breast tumor metastasis (468LN)	5	66283790	66335622	0.08	0.83	0.75	MAST4
Breast tumor metastasis (468LN)	7	22893734	22901595	0.14	0.88	0.74	TOMM7
Breast tumor metastasis (468LN)	6	80234368	80256281	0.23	0.95	0.73	LCA5
Breast tumor metastasis (468LN)	22	27868674	27873807	0.23	0.95	0.72	MN1
Breast tumor metastasis (468LN)	21	39831004	39862029	0.02	0.72	0.71	ERG
Breast tumor metastasis (468LN)	6	54710175	54759167	0.04	0.74	0.70	FAM83B
Breast tumor metastasis (468LN)	3	123517889	123561019	0.03	0.72	0.69	MYLK
Breast tumor metastasis (468LN)	8	128672774	128685618	0.05	0.71	0.66	MYC
Breast tumor metastasis (468LN)	3	188984945	189053159	0.01	0.66	0.65	TPRG1
Breast tumor metastasis (468LN)	2	17760510	17781417	0.06	0.69	0.64	VSNL1
Breast tumor metastasis (468LN)	13	25237454	25247430	0.18	0.81	0.64	ATP12A
Breast tumor metastasis (468LN)	7	36690919	36727628	0.02	0.65	0.63	AOAH
Breast tumor metastasis (468LN)	3	189540773	189562008	0.08	0.69	0.62	TP63
Breast tumor metastasis (468LN)	3	189540773	189562008	0.08	0.69	0.62	TP63
Breast tumor metastasis (468LN)	3	189540773	189562008	0.08	0.69	0.62	TP63
Breast tumor metastasis (468LN)	3	189540773	189562008	0.08	0.69	0.62	TP63
Breast tumor metastasis (468LN)	3	189540773	189562008	0.08	0.69	0.62	TP63
Breast tumor metastasis (468LN)	17	39734929	39745765	0.08	0.69	0.61	KRT14
Breast tumor metastasis (468LN)	3	9208190	9218379	0.08	0.68	0.61	SRGAP3
Breast tumor metastasis (468LN)	8	66840219	66887130	0.06	0.66	0.61	DNAJC5B

Breast tumor metastasis (468LN)	8	66840219	66887130	0.06	0.66	0.61	DNAJC5B
Breast tumor metastasis (468LN)	5	148155468	148172554	0.00	0.60	0.60	ADRB2
Breast tumor metastasis (468LN)	3	111296776	111315059	0.00	0.59	0.59	ZBED2
Breast tumor metastasis (468LN)	5	64484499	64507100	0.00	0.59	0.59	ADAMTS6
Breast tumor metastasis (468LN)	5	52540140	52559537	0.03	0.62	0.59	MOCS2
Breast tumor metastasis (468LN)	11	12130227	12162030	0.10	0.67	0.57	MICAL2
Breast tumor metastasis (468LN)	1	94425779	94461050	0.03	0.60	0.56	ABCA4
Breast tumor metastasis (468LN)	15	83945999	83992250	0.08	0.65	0.56	BNC1
Breast tumor metastasis (468LN)	6	107163126	107200634	0.01	0.57	0.56	QRSL1
Breast tumor metastasis (468LN)	1	209527622	209609482	0.05	0.61	0.56	CAMK1G
Breast tumor metastasis (468LN)	9	124030410	124052489	0.40	0.95	0.55	GSN
Breast tumor metastasis (468LN)	3	44923928	44941628	0.03	0.58	0.54	TGM4
Breast tumor metastasis (468LN)	8	128402145	128416279	0.12	0.66	0.54	POU5F1B
Breast tumor metastasis (468LN)	7	130643572	130647502	0.19	0.73	0.53	MKLN1
Breast tumor metastasis (468LN)	10	82213990	82238676	0.25	0.78	0.53	TSPAN14
Breast tumor metastasis (468LN)	11	58340393	58348640	0.48	1.00	0.52	ZFP91-CNTF
Breast tumor metastasis (468LN)	2	1708289	1747387	0.12	0.62	0.50	PXDN
Breast tumor metastasis (468LN)	10	6762937	6792166	0.11	0.61	0.50	PRKCQ
Breast tumor metastasis (468LN)	4	100367626	100385121	0.00	0.50	0.50	ADH7
Breast tumor metastasis (468LN)	12	109114581	109129684	0.20	0.69	0.49	CORO1C
Breast tumor metastasis (468LN)	17	34104001	34120836	0.05	0.54	0.49	MMP28
Breast tumor metastasis (468LN)	7	134297576	134309091	0.00	0.48	0.48	BPGM
Breast tumor metastasis (468LN)	2	70745929	70782514	0.07	0.54	0.47	TGFA
Breast tumor metastasis (468LN)	2	208102152	208124172	0.08	0.55	0.47	KLF7
Breast tumor metastasis (468LN)	18	20393690	20449701	0.01	0.48	0.47	RBBP8
Breast tumor metastasis (468LN)	1	162875345	162886688	0.00	0.47	0.47	C1orf110
Breast tumor metastasis (468LN)	1	162875345	162886688	0.00	0.47	0.47	C1orf110
Breast tumor metastasis (468LN)	8	29196390	29211042	0.48	0.94	0.46	DUSP4
Breast tumor metastasis (468LN)	12	52820050	52849388	0.03	0.49	0.46	KRT6B
Breast tumor metastasis (468LN)	9	118011886	118026827	0.00	0.45	0.45	1-Dec
Breast tumor metastasis (468LN)	5	33299342	33322096	0.04	0.49	0.44	TARS
Breast tumor metastasis (468LN)	4	15893942	15940788	0.01	0.45	0.44	FGFBP1
Breast tumor metastasis (468LN)	16	68669444	68711125	0.14	0.58	0.44	CDH3
Breast tumor metastasis (468LN)	19	33840783	33869285	0.07	0.49	0.42	CEBPG
Breast tumor metastasis (468LN)	11	12085413	12114013	0.03	0.45	0.42	MICAL2
Breast tumor metastasis (468LN)	2	99524648	99538593	0.14	0.56	0.42	KIAA1211L
Breast tumor metastasis (468LN)	3	131851786	131856052	0.02	0.44	0.42	CPNE4
Breast tumor metastasis (468LN)	7	48120763	48135456	0.12	0.54	0.42	UPP1
Breast tumor metastasis (468LN)	22	25338358	25391181	0.11	0.52	0.41	TMEM211
Breast tumor metastasis (468LN)	6	143716698	143732923	0.04	0.44	0.40	AL031320.1
Breast tumor metastasis (468LN)	1	221611253	221636061	0.04	0.43	0.40	DUSP10
Breast tumor metastasis (468LN)	10	118176539	118201896	0.00	0.39	0.39	PNLIPRP3
Breast tumor metastasis (468LN)	18	28809492	28837065	0.06	0.44	0.38	DSG1
Breast tumor metastasis (468LN)	10	24737918	24757794	0.37	0.76	0.38	KIAA1217
Breast tumor metastasis (468LN)	20	51581498	51597746	0.45	0.82	0.37	TSHZ2
Breast tumor metastasis (468LN)	17	39674641	39706091	0.17	0.54	0.37	KRT19
Breast tumor metastasis (468LN)	4	11615367	11638574	0.00	0.37	0.37	HS3ST1
Breast tumor metastasis (468LN)	7	41136094	41161972	0.04	0.41	0.37	C7orf10
Breast tumor metastasis (468LN)	6	17863327	17868787	0.00	0.36	0.36	KIF13A
Breast tumor metastasis (468LN)	8	32403683	32442699	0.19	0.55	0.36	NRG1
Breast tumor metastasis (468LN)	11	128362001	128396377	0.21	0.57	0.36	ETS1
Breast tumor metastasis (468LN)	11	18328431	18352028	0.10	0.45	0.36	GTF2H1
Breast tumor metastasis (468LN)	X	105509168	105546610	0.00	0.35	0.35	MUM1L1
Breast tumor metastasis (468LN)	12	59433027	59479164	0.05	0.40	0.34	LRIG3
Breast tumor metastasis (468LN)	20	56261686	56292224	0.14	0.47	0.33	PMEPA1
Breast tumor metastasis (468LN)	5	168586808	168610831	0.05	0.38	0.33	SLIT3
Breast tumor metastasis (468LN)	11	12183685	12222874	0.04	0.36	0.33	MICAL2
Breast tumor metastasis (468LN)	21	30003125	30011412	0.00	0.33	0.33	N6AMT1
Breast tumor metastasis (468LN)	21	30003125	30011412	0.00	0.33	0.33	N6AMT1
Breast tumor metastasis (468LN)	21	30003125	30011412	0.00	0.33	0.33	N6AMT1
Breast tumor metastasis (468LN)	6	53475369	53500722	0.04	0.36	0.33	GCLC
Breast tumor metastasis (468LN)	5	95554865	95565313	0.00	0.32	0.32	PCSK1
Breast tumor metastasis (468LN)	11	114162169	114180527	0.26	0.58	0.32	NNMT
Breast tumor metastasis (468LN)	4	186225739	186230197	0.00	0.31	0.31	SNX25
Breast tumor metastasis (468LN)	2	206514186	206569665	0.11	0.41	0.30	NRP2
Breast tumor metastasis (468LN)	18	59613226	59621595	0.06	0.36	0.30	RNF152
Breast tumor metastasis (468LN)	11	15986587	15993130	0.00	0.30	0.30	SOX6
Breast tumor metastasis (468LN)	5	86394511	86425151	0.00	0.30	0.30	AC008394.1
Breast tumor metastasis (468LN)	17	805322	839104	0.10	0.40	0.29	NXN
Breast tumor metastasis (468LN)	21	40386166	40424480	0.08	0.37	0.29	PSMG1
Breast tumor metastasis (468LN)	6	116684394	116720580	0.05	0.34	0.29	DSE
Breast tumor metastasis (468LN)	6	111906089	111928358	0.33	0.61	0.28	TRAF3IP2
Breast tumor metastasis (468LN)	21	44720362	44794103	0.13	0.41	0.28	SIK1
Breast tumor metastasis (468LN)	3	111518531	111530936	0.08	0.35	0.27	PHLDB2
Breast tumor metastasis (468LN)	3	111518531	111530936	0.08	0.35	0.27	PHLDB2
Breast tumor metastasis (468LN)	21	44907375	44943674	0.06	0.33	0.27	SIK1
Breast tumor metastasis (468LN)	13	73984215	74005603	0.00	0.27	0.27	KLF5
Breast tumor metastasis (468LN)	9	111295882	111325768	0.00	0.27	0.27	ACTL7B
Breast tumor metastasis (468LN)	11	93860386	93869783	0.22	0.48	0.26	PANX1
Breast tumor metastasis (468LN)	10	99443070	99451242	0.24	0.49	0.25	AVPI1
Lung adenocarcinoma (H1437)	12	108664152	108729705	0.02	0.99	0.97	CMKLR1
Lung adenocarcinoma (H1437)	12	2321941	2506285	0.01	0.96	0.96	CACNA1C
Lung adenocarcinoma (H1437)	6	164505820	164531835	0.01	0.92	0.91	QKI
Lung adenocarcinoma (H1437)	3	16923413	16959538	0.07	0.94	0.86	PLCL2
Lung adenocarcinoma (H1437)	12	2160784	2308190	0.03	0.89	0.86	CACNA1C
Lung adenocarcinoma (H1437)	11	128560125	128616737	0.09	0.94	0.85	FLI1
Lung adenocarcinoma (H1437)	8	12937916	12990887	0.08	0.92	0.84	DLC1
Lung adenocarcinoma (H1437)	1	153579237	153590776	0.04	0.86	0.81	S100A14
Lung adenocarcinoma (H1437)	14	23898802	23913794	0.00	0.80	0.80	MYH7
Lung adenocarcinoma (H1437)	10	11188477	11246900	0.03	0.83	0.80	CELF2
Lung adenocarcinoma (H1437)	16	49857837	49893960	0.17	0.95	0.78	ZNF423
Lung adenocarcinoma (H1437)	17	32547489	32583972	0.05	0.82	0.77	CCL2
Lung adenocarcinoma (H1437)	11	67775074	67783359	0.10	0.84	0.74	UNC93B1
Lung adenocarcinoma (H1437)	14	23870473	23890571	0.00	0.72	0.72	MYH6
Lung adenocarcinoma (H1437)	5	175084363	175123938	0.05	0.76	0.71	HRH2
Lung adenocarcinoma (H1437)	6	45914072	45984711	0.02	0.72	0.70	CLIC5
Lung adenocarcinoma (H1437)	11	124745703	124772028	0.07	0.76	0.68	ROBO4
Lung adenocarcinoma (H1437)	15	68564041	68574856	0.24	0.91	0.68	FEM1B
Lung adenocarcinoma (H1437)	20	43963996	43977803	0.09	0.75	0.66	SDC4
Lung adenocarcinoma (H1437)	3	14834008	14933715	0.04	0.66	0.62	FGD5
Lung adenocarcinoma (H1437)	12	4377579	4414053	0.22	0.79	0.57	CCND2
Lung adenocarcinoma (H1437)	10	17256938	17276860	0.22	0.77	0.55	VIM
Lung adenocarcinoma (H1437)	11	74858188	74908024	0.00	0.54	0.54	SLCO2B1
Lung adenocarcinoma (H1437)	12	114806250	114852976	0.45	0.93	0.48	TBX5
Lung adenocarcinoma (H1437)	1	8063878	8088662	0.20	0.68	0.48	ERRFI1
Lung adenocarcinoma (H1437)	17	79337961	79404678	0.22	0.70	0.47	BAHCC1
Lung adenocarcinoma (H1437)	1	151960740	151974470	0.24	0.65	0.40	S100A10
Lung adenocarcinoma (H1437)	6	139690764	139698884	0.53	0.94	0.40	CITED2
Lung adenocarcinoma (H1437)	7	150642159	150687879	0.17	0.57	0.40	KCNH2
Lung adenocarcinoma (H1437)	16	637852	644850	0.41	0.79	0.39	RAB40C
Lung adenocarcinoma (H1437)	16	29165405	29212441	0.01	0.39	0.38	SNX29P2
Lung adenocarcinoma (H1437)	22	50323334	50364887	0.25	0.61	0.37	PIM3
Lung adenocarcinoma (H1437)	3	53269245	53309568	0.07	0.42	0.35	TKT
Lung adenocarcinoma (H1437)	1	153537691	153541885	0.41	0.76	0.34	S100A2
Lung adenocarcinoma (H1437)	2	145236716	145282659	0.29	0.63	0.34	ZEB2
Lung adenocarcinoma (H1437)	9	132241137	132264724	0.28	0.61	0.33	NTMT1
Lung adenocarcinoma (H1437)	6	82458588	82464329	0.56	0.89	0.33	FAM46A
Lung adenocarcinoma (H1437)	17	48131999	48143308	0.25	0.58	0.33	ITGA3
Lung adenocarcinoma (H1437)	17	79300771	79323073	0.29	0.61	0.32	TMEM105
Lung adenocarcinoma (H1437)	3	64669880	64677423	0.57	0.88	0.32	ADAMTS9
Lung adenocarcinoma (H1437)	3	71625820	71634865	0.45	0.77	0.32	FOXP1
Lung adenocarcinoma (H1437)	21	47373962	47413579	0.16	0.48	0.31	COL6A1
Lung adenocarcinoma (H1437)	5	180668274	180675859	0.53	0.84	0.31	GNB2L1
Lung adenocarcinoma (H1437)	17	80223721	80257686	0.16	0.47	0.31	CSNK1D
Lung adenocarcinoma (H1437)	16	86504285	86554076	0.52	0.83	0.31	FOXF1
Lung adenocarcinoma (H1437)	22	19830490	19844157	0.25	0.55	0.30	C22orf29
Lung adenocarcinoma (H1437)	16	19117268	19151258	0.11	0.40	0.29	CTD-2349B8.1
Lung adenocarcinoma (H1437)	1	153504009	153510224	0.41	0.70	0.29	S100A6
Lung adenocarcinoma (H1437)	3	9437015	9444925	0.71	1.00	0.29	SETD5
Lung adenocarcinoma (H1437)	12	52536940	52547179	0.09	0.37	0.29	KRT80
Lung adenocarcinoma (H1437)	12	7032586	7039696	0.64	0.92	0.28	ATN1
Lung adenocarcinoma (H1437)	2	127412308	127428737	0.14	0.42	0.27	GYPC
Lung adenocarcinoma (H1437)	6	74224953	74234140	0.53	0.79	0.26	EEF1A1
Lung adenocarcinoma (H1437)	10	33226742	33251655	0.24	0.50	0.26	ITGB1
Lung adenocarcinoma (H1437)	19	13956569	13963420	0.41	0.66	0.25	NANOS3
Lung squamous cell carcinoma (H157)	14	23898802	23913794	0.00	0.97	0.97	MYH7
Lung squamous cell carcinoma (H157)	1	164717131	164764470	0.01	0.98	0.97	PBX1
Lung squamous cell carcinoma (H157)	15	92456353	92524948	0.01	0.98	0.97	SLCO3A1
Lung squamous cell carcinoma (H157)	12	2321941	2506285	0.01	0.98	0.97	CACNA1C
Lung squamous cell carcinoma (H157)	12	108664152	108729705	0.02	0.99	0.97	CMKLR1
Lung squamous cell carcinoma (H157)	7	51322792	51385835	0.03	0.98	0.95	COBL
Lung squamous cell carcinoma (H157)	22	33196018	33254838	0.06	1.00	0.94	TIMP3
Lung squamous cell carcinoma (H157)	6	11345367	11394452	0.01	0.95	0.94	NEDD9
Lung squamous cell carcinoma (H157)	2	218653998	218760935	0.01	0.95	0.93	TNS1
Lung squamous cell carcinoma (H157)	3	14834008	14933715	0.04	0.97	0.93	FGD5
Lung squamous cell carcinoma (H157)	22	32924539	32976379	0.01	0.93	0.91	SYN3
Lung squamous cell carcinoma (H157)	12	2160784	2308190	0.03	0.94	0.91	CACNA1C
Lung squamous cell carcinoma (H157)	9	71319153	71371192	0.06	0.97	0.91	PIP5K1B
Lung squamous cell carcinoma (H157)	22	34206955	34272052	0.03	0.94	0.91	LARGE
Lung squamous cell carcinoma (H157)	3	123047106	123169548	0.02	0.93	0.91	ADCY5
Lung squamous cell carcinoma (H157)	22	33011575	33041335	0.07	0.97	0.90	SYN3
Lung squamous cell carcinoma (H157)	11	117730279	117757325	0.09	0.99	0.90	FXYD6
Lung squamous cell carcinoma (H157)	10	116233610	116285340	0.00	0.89	0.89	ABLIM1
Lung squamous cell carcinoma (H157)	11	19719061	19814464	0.05	0.94	0.89	NAV2
Lung squamous cell carcinoma (H157)	5	175084363	175123938	0.05	0.94	0.89	HRH2
Lung squamous cell carcinoma (H157)	15	92394913	92441672	0.06	0.95	0.89	SLCO3A1

Lung squamous cell carcinoma (H157)	1	175138842	175191990	0.07	0.95	0.88	KIAA0040
Lung squamous cell carcinoma (H157)	3	16923413	16959538	0.07	0.95	0.88	PLCL2
Lung squamous cell carcinoma (H157)	6	45914072	45984711	0.02	0.89	0.87	CLIC5
Lung squamous cell carcinoma (H157)	9	677510	756481	0.04	0.90	0.86	KANK1
Lung squamous cell carcinoma (H157)	4	23855744	23906109	0.05	0.91	0.86	PPARGC1A
Lung squamous cell carcinoma (H157)	10	11188477	11246900	0.03	0.89	0.86	CELF2
Lung squamous cell carcinoma (H157)	10	97137580	97321838	0.02	0.87	0.85	SORBS1
Lung squamous cell carcinoma (H157)	1	208373068	208418917	0.11	0.96	0.85	PLXNA2
Lung squamous cell carcinoma (H157)	6	112519021	112620635	0.03	0.88	0.84	LAMA4
Lung squamous cell carcinoma (H157)	10	116297903	116338177	0.00	0.83	0.83	ABLIM1
Lung squamous cell carcinoma (H157)	11	11556367	11614464	0.00	0.82	0.82	GALNT18
Lung squamous cell carcinoma (H157)	1	201326267	201371297	0.03	0.84	0.81	LAD1
Lung squamous cell carcinoma (H157)	14	23870473	23890571	0.00	0.81	0.81	MYH6
Lung squamous cell carcinoma (H157)	2	218814905	218887533	0.05	0.85	0.80	TNS1
Lung squamous cell carcinoma (H157)	1	208057100	208099004	0.05	0.83	0.79	CD34
Lung squamous cell carcinoma (H157)	13	113612236	113642719	0.07	0.84	0.78	MCF2L
Lung squamous cell carcinoma (H157)	6	46859103	46921360	0.04	0.81	0.77	GPR116
Lung squamous cell carcinoma (H157)	11	128560125	128616737	0.09	0.86	0.77	FLI1
Lung squamous cell carcinoma (H157)	2	1701232	1756057	0.07	0.83	0.76	PXDN
Lung squamous cell carcinoma (H157)	2	206546300	206580305	0.14	0.90	0.76	NRP2
Lung squamous cell carcinoma (H157)	3	46885844	46916759	0.08	0.83	0.75	MYL3
Lung squamous cell carcinoma (H157)	10	45484961	45497498	0.13	0.87	0.74	ZNF22
Lung squamous cell carcinoma (H157)	16	49857837	49893960	0.17	0.91	0.74	ZNF423
Lung squamous cell carcinoma (H157)	9	73070671	73110216	0.03	0.77	0.74	KLF9
Lung squamous cell carcinoma (H157)	3	13031469	13061226	0.11	0.85	0.73	IQSEC1
Lung squamous cell carcinoma (H157)	1	203236433	203297636	0.16	0.87	0.71	BTG2
Lung squamous cell carcinoma (H157)	22	18913160	18924266	0.07	0.78	0.71	PRODH
Lung squamous cell carcinoma (H157)	2	42273101	42300251	0.08	0.78	0.70	PKDCC
Lung squamous cell carcinoma (H157)	1	167585748	167668005	0.01	0.71	0.69	RCSD1
Lung squamous cell carcinoma (H157)	10	116356071	116457594	0.02	0.70	0.68	ABLIM1
Lung squamous cell carcinoma (H157)	6	164505820	164531835	0.01	0.69	0.67	QKI
Lung squamous cell carcinoma (H157)	15	64259521	64312980	0.00	0.66	0.66	DAPK2
Lung squamous cell carcinoma (H157)	3	123371190	123441485	0.03	0.67	0.64	MYLK
Lung squamous cell carcinoma (H157)	6	167010067	167054751	0.08	0.71	0.63	RPS6KA2
Lung squamous cell carcinoma (H157)	1	16338365	16348730	0.11	0.73	0.63	CLCNKA
Lung squamous cell carcinoma (H157)	11	113906599	114092628	0.05	0.67	0.62	ZBTB16
Lung squamous cell carcinoma (H157)	14	100196476	100226407	0.15	0.75	0.61	EML1
Lung squamous cell carcinoma (H157)	17	32547489	32583972	0.05	0.64	0.59	CCL2
Lung squamous cell carcinoma (H157)	11	58938916	58992734	0.05	0.64	0.59	MPEG1
Lung squamous cell carcinoma (H157)	2	159810283	159870494	0.06	0.61	0.55	TANC1
Lung squamous cell carcinoma (H157)	5	148793738	148806368	0.07	0.62	0.55	IL17B
Lung squamous cell carcinoma (H157)	17	53342089	53373406	0.13	0.67	0.54	HLF
Lung squamous cell carcinoma (H157)	3	138066340	138119923	0.04	0.57	0.53	MRAS
Lung squamous cell carcinoma (H157)	10	88420822	88454560	0.01	0.55	0.53	LDB3
Lung squamous cell carcinoma (H157)	7	150642159	150687879	0.17	0.69	0.52	KCNH2
Lung squamous cell carcinoma (H157)	3	128718996	128746653	0.14	0.66	0.52	EFCC1
Lung squamous cell carcinoma (H157)	9	73168217	73230803	0.02	0.54	0.52	KLF9
Lung squamous cell carcinoma (H157)	5	141183253	141261839	0.14	0.65	0.51	PCDH1
Lung squamous cell carcinoma (H157)	2	233923287	233954214	0.11	0.62	0.51	INPP5D
Lung squamous cell carcinoma (H157)	1	26386415	26449049	0.04	0.55	0.51	PDIK1L
Lung squamous cell carcinoma (H157)	15	66905412	66955707	0.04	0.55	0.51	RP11-321F6.1
Lung squamous cell carcinoma (H157)	19	33751162	33802757	0.14	0.64	0.50	CEBPA
Lung squamous cell carcinoma (H157)	22	37562253	37596293	0.18	0.67	0.49	C1QTNF6
Lung squamous cell carcinoma (H157)	1	41939307	41981913	0.03	0.52	0.49	EDN2
Lung squamous cell carcinoma (H157)	16	46764273	46798191	0.03	0.51	0.49	MYLK3
Lung squamous cell carcinoma (H157)	1	47676896	47710031	0.21	0.70	0.49	TAL1
Lung squamous cell carcinoma (H157)	15	65346572	65383634	0.08	0.56	0.48	KBTBD13
Lung squamous cell carcinoma (H157)	3	53174020	53206578	0.08	0.55	0.47	PRKCD
Lung squamous cell carcinoma (H157)	12	114806250	114852976	0.45	0.90	0.45	TBX5
Lung squamous cell carcinoma (H157)	10	45452698	45465905	0.11	0.55	0.45	RASSF4
Lung squamous cell carcinoma (H157)	15	85359756	85408023	0.01	0.46	0.45	ALPK3
Lung squamous cell carcinoma (H157)	12	109547560	109572829	0.04	0.49	0.44	ACACB
Lung squamous cell carcinoma (H157)	5	148650125	148713310	0.02	0.46	0.44	AFAP1L1
Lung squamous cell carcinoma (H157)	2	220473619	220511327	0.05	0.49	0.44	SLC4A3
Lung squamous cell carcinoma (H157)	3	126068503	126077592	0.29	0.73	0.44	KLF15
Lung squamous cell carcinoma (H157)	16	29165405	29212441	0.01	0.45	0.43	SNX29P2
Lung squamous cell carcinoma (H157)	12	52288433	52321075	0.05	0.47	0.43	ACVRL1
Lung squamous cell carcinoma (H157)	5	148513364	148558772	0.02	0.44	0.42	ABLIM3
Lung squamous cell carcinoma (H157)	6	11194797	11293348	0.06	0.48	0.42	NEDD9
Lung squamous cell carcinoma (H157)	11	62646517	62690197	0.11	0.52	0.41	CHRM1
Lung squamous cell carcinoma (H157)	15	57840518	57983809	0.01	0.42	0.41	GCOM1
Lung squamous cell carcinoma (H157)	2	30440179	30497935	0.08	0.48	0.40	LBH
Lung squamous cell carcinoma (H157)	1	94485792	94536633	0.03	0.43	0.40	ABCA4
Lung squamous cell carcinoma (H157)	5	142177195	142264107	0.00	0.40	0.39	ARHGAP26
Lung squamous cell carcinoma (H157)	16	57653421	57683065	0.02	0.41	0.39	GPR56
Lung squamous cell carcinoma (H157)	2	159936457	160087427	0.01	0.40	0.39	TANC1
Lung squamous cell carcinoma (H157)	8	12610121	12624953	0.34	0.73	0.38	LONRF1
Lung squamous cell carcinoma (H157)	12	117479472	117535649	0.01	0.39	0.38	TESC

Lung squamous cell carcinoma (H157)	16	58484914	58536190	0.08	0.46	0.38	NDRG4
Lung squamous cell carcinoma (H157)	20	19737283	19767891	0.13	0.50	0.38	AL121761.2
Lung squamous cell carcinoma (H157)	10	73469788	73536258	0.05	0.42	0.37	C10orf54
Lung squamous cell carcinoma (H157)	3	156802993	156855459	0.06	0.43	0.37	CCNL1
Lung squamous cell carcinoma (H157)	6	13383397	13489697	0.04	0.41	0.37	AL583828.1
Lung squamous cell carcinoma (H157)	9	116236515	116397394	0.04	0.41	0.37	RGS3
Lung squamous cell carcinoma (H157)	1	55007351	55053202	0.05	0.41	0.36	ACOT11
Lung squamous cell carcinoma (H157)	16	66392035	66422229	0.05	0.41	0.36	CDH5
Lung squamous cell carcinoma (H157)	17	4379415	4447094	0.09	0.44	0.35	MYBBP1A
Lung squamous cell carcinoma (H157)	10	105494540	105555321	0.01	0.36	0.34	SH3PXD2A
Lung squamous cell carcinoma (H157)	11	74858188	74908024	0.00	0.34	0.34	SLCO2B1
Lung squamous cell carcinoma (H157)	10	71158153	71242818	0.05	0.38	0.33	TSPAN15
Lung squamous cell carcinoma (H157)	1	151960740	151974470	0.24	0.57	0.33	S100A10
Lung squamous cell carcinoma (H157)	22	36541286	36558381	0.18	0.51	0.33	APOL3
Lung squamous cell carcinoma (H157)	1	2056832	2083663	0.06	0.38	0.32	PRKCZ
Lung squamous cell carcinoma (H157)	6	33659980	33715998	0.04	0.37	0.32	IP6K3
Lung squamous cell carcinoma (H157)	20	24972518	25040392	0.09	0.40	0.31	ACSS1
Lung squamous cell carcinoma (H157)	17	56583934	56616821	0.16	0.47	0.31	4-Sep
Lung squamous cell carcinoma (H157)	14	89852016	89897880	0.11	0.42	0.31	FOXN3-AS1
Lung squamous cell carcinoma (H157)	9	129783892	129898008	0.01	0.32	0.31	ANGPTL2
Lung squamous cell carcinoma (H157)	1	31839527	31868307	0.05	0.36	0.31	FABP3
Lung squamous cell carcinoma (H157)	12	6221157	6351408	0.03	0.34	0.30	CD9
Lung squamous cell carcinoma (H157)	17	48981354	49028680	0.06	0.36	0.30	SPAG9
Lung squamous cell carcinoma (H157)	3	46731656	46739274	0.17	0.47	0.30	ALS2CL
Lung squamous cell carcinoma (H157)	14	75724170	75782540	0.16	0.46	0.30	FOS
Lung squamous cell carcinoma (H157)	3	125978012	126033776	0.04	0.33	0.29	KLF15
Lung squamous cell carcinoma (H157)	17	72422406	72460049	0.19	0.47	0.28	GPRC5C
Lung squamous cell carcinoma (H157)	15	68564041	68574856	0.24	0.52	0.28	FEM1B
Lung squamous cell carcinoma (H157)	19	2567152	2637706	0.08	0.35	0.28	GNG7
Lung squamous cell carcinoma (H157)	17	21178575	21194211	0.42	0.70	0.28	MAP2K3
Lung squamous cell carcinoma (H157)	16	53534624	53554437	0.15	0.41	0.26	AKTIP
Lung squamous cell carcinoma (H157)	11	72851228	72894684	0.09	0.36	0.26	FCHSD2
Lung squamous cell carcinoma (H157)	4	89203835	89219912	0.17	0.43	0.26	PPM1K
Lung squamous cell carcinoma (H157)	8	145021828	145029514	0.53	0.78	0.25	PLEC
Small cell lung cancer (H1672)	17	32547489	32583972	0.05	0.95	0.90	CCL2
Small cell lung cancer (H1672)	11	58938916	58992734	0.05	0.95	0.90	MPEG1
Small cell lung cancer (H1672)	14	23898802	23913794	0.00	0.89	0.89	MYH7
Small cell lung cancer (H1672)	16	88262997	88280525	0.07	0.94	0.88	BANP
Small cell lung cancer (H1672)	14	23870473	23890571	0.00	0.82	0.82	MYH6
Small cell lung cancer (H1672)	12	10862096	10876772	0.16	0.97	0.81	CSDA
Small cell lung cancer (H1672)	1	145208293	145266278	0.07	0.87	0.80	NOTCH2NL
Small cell lung cancer (H1672)	19	48218393	48248060	0.12	0.87	0.75	EHD2
Small cell lung cancer (H1672)	7	100727881	100783471	0.09	0.84	0.74	SERPINE1
Small cell lung cancer (H1672)	16	29259027	29284865	0.01	0.74	0.73	SNX29P2
Small cell lung cancer (H1672)	11	44746277	44803353	0.04	0.77	0.72	TSPAN18
Small cell lung cancer (H1672)	12	4377579	4414053	0.22	0.94	0.72	CCND2
Small cell lung cancer (H1672)	19	2567152	2637706	0.08	0.78	0.71	GNG7
Small cell lung cancer (H1672)	11	44584904	44654491	0.04	0.74	0.69	CD82
Small cell lung cancer (H1672)	X	64886373	64932289	0.26	0.96	0.69	MSN
Small cell lung cancer (H1672)	18	19743086	19793872	0.19	0.88	0.69	GATA6
Small cell lung cancer (H1672)	21	47373962	47413579	0.16	0.85	0.68	COL6A1
Small cell lung cancer (H1672)	7	116138842	116200757	0.10	0.76	0.66	CAV1
Small cell lung cancer (H1672)	19	6055959	6111185	0.10	0.70	0.60	RFX2
Small cell lung cancer (H1672)	8	11532909	11620420	0.32	0.90	0.59	C8orf49
Small cell lung cancer (H1672)	1	151960740	151974470	0.24	0.76	0.52	S100A10
Small cell lung cancer (H1672)	10	46950950	46984234	0.12	0.63	0.52	SYT15
Small cell lung cancer (H1672)	1	51433106	51444971	0.49	1.00	0.51	CDKN2C
Small cell lung cancer (H1672)	21	46711726	46750687	0.07	0.57	0.50	POFUT2
Small cell lung cancer (H1672)	21	46711726	46750687	0.07	0.57	0.50	POFUT2
Small cell lung cancer (H1672)	21	46711726	46750687	0.07	0.57	0.50	POFUT2
Small cell lung cancer (H1672)	2	241925890	241950654	0.12	0.61	0.49	SNED1
Small cell lung cancer (H1672)	12	10364085	10373949	0.26	0.73	0.47	GABARAPL1
Small cell lung cancer (H1672)	1	144923313	145010556	0.02	0.48	0.46	AL590452.1
Small cell lung cancer (H1672)	16	3102086	3157272	0.10	0.56	0.46	ZSCAN10
Small cell lung cancer (H1672)	19	18401834	18406232	0.52	0.97	0.45	JUND
Small cell lung cancer (H1672)	16	86504285	86554076	0.52	0.96	0.45	FOXF1
Small cell lung cancer (H1672)	11	62646517	62690197	0.11	0.55	0.44	CHRM1
Small cell lung cancer (H1672)	19	45244285	45262369	0.46	0.89	0.43	BCL3
Small cell lung cancer (H1672)	11	1770660	1805042	0.08	0.51	0.43	CTSD
Small cell lung cancer (H1672)	19	2462980	2496458	0.29	0.70	0.42	GADD45B
Small cell lung cancer (H1672)	19	33751162	33802757	0.14	0.55	0.41	CEBPA
Small cell lung cancer (H1672)	3	46885844	46916759	0.08	0.48	0.40	MYL3
Small cell lung cancer (H1672)	13	114821455	114835116	0.12	0.52	0.40	RASA3
Small cell lung cancer (H1672)	7	143071647	143093995	0.11	0.51	0.40	EPHA1
Small cell lung cancer (H1672)	13	110937717	111086117	0.06	0.46	0.40	COL4A2
Small cell lung cancer (H1672)	22	37562253	37596293	0.18	0.56	0.38	C1QTNF6
Small cell lung cancer (H1672)	17	81037486	81068828	0.21	0.59	0.38	METRNL
Small cell lung cancer (H1672)	13	114863283	114898583	0.15	0.52	0.37	RASA3

Small cell lung cancer (H1672)	12	53602263	53626034	0.27	0.62	0.36	RARG
Small cell lung cancer (H1672)	11	6338518	6348300	0.22	0.56	0.34	PRKCDBP
Small cell lung cancer (H1672)	16	88809516	88852538	0.11	0.43	0.32	PIEZO1
Small cell lung cancer (H1672)	19	45347579	45383849	0.17	0.49	0.32	PVRL2
Small cell lung cancer (H1672)	9	132161805	132179421	0.11	0.42	0.32	NTMT1
Small cell lung cancer (H1672)	16	29165405	29212441	0.01	0.33	0.31	SNX29P2
Small cell lung cancer (H1672)	1	2475405	2484449	0.45	0.76	0.31	TNFRSF14
Small cell lung cancer (H1672)	10	134321138	134365757	0.10	0.41	0.31	INPP5A
Small cell lung cancer (H1672)	6	139690764	139698884	0.53	0.84	0.31	CITED2
Small cell lung cancer (H1672)	8	11663395	11764350	0.05	0.35	0.30	CTSB
Small cell lung cancer (H1672)	10	17256938	17276860	0.22	0.52	0.30	VIM
Small cell lung cancer (H1672)	20	24972518	25040392	0.09	0.39	0.30	ACSS1
Small cell lung cancer (H1672)	6	30880966	30905424	0.07	0.37	0.30	SFTA2
Small cell lung cancer (H1672)	22	39616903	39707444	0.13	0.42	0.29	AL031590.1
Small cell lung cancer (H1672)	3	128718996	128746653	0.14	0.43	0.29	EFCC1
Small cell lung cancer (H1672)	1	12192038	12247305	0.08	0.36	0.28	TNFRSF1B
Small cell lung cancer (H1672)	13	110868117	110924162	0.03	0.30	0.27	COL4A2
Small cell lung cancer (H1672)	13	111155320	111212932	0.02	0.29	0.27	COL4A2
Small cell lung cancer (H1672)	3	9437015	9444925	0.71	0.97	0.26	SETD5
Small cell lung cancer (H1672)	6	3053230	3070286	0.27	0.52	0.25	RIPK1
Primary colon tumor (Colon_P)	11	64654606	64662582	0.23	0.87	0.64	EHD1
Primary colon tumor (Colon_P)	11	34644271	34676099	0.09	0.70	0.61	EHF
Primary colon tumor (Colon_P)	17	75277484	75285351	0.20	0.78	0.57	9-Sep
Primary colon tumor (Colon_P)	1	1092807	1104591	0.35	0.90	0.56	TTLL10
Primary colon tumor (Colon_P)	12	6478956	6487347	0.24	0.78	0.54	LTBR
Primary colon tumor (Colon_P)	19	35605339	35613006	0.23	0.76	0.53	FXYD3
Primary colon tumor (Colon_P)	1	207111967	207121081	0.23	0.74	0.51	PIGR
Primary colon tumor (Colon_P)	17	46021666	46045798	0.16	0.64	0.48	CDK5RAP3
Primary colon tumor (Colon_P)	9	140082154	140094655	0.20	0.67	0.47	TPRN
Primary colon tumor (Colon_P)	1	153579813	153590695	0.05	0.52	0.47	S100A14
Primary colon tumor (Colon_P)	1	234949066	234963582	0.06	0.52	0.46	IRF2BP2
Primary colon tumor (Colon_P)	1	234949066	234963582	0.06	0.52	0.46	IRF2BP2
Primary colon tumor (Colon_P)	11	72485443	72499721	0.10	0.56	0.46	STARD10
Primary colon tumor (Colon_P)	1	155900806	155917642	0.20	0.63	0.44	RXFP4
Primary colon tumor (Colon_P)	16	70091792	70101175	0.34	0.75	0.41	RP11-419C5.2
Primary colon tumor (Colon_P)	15	40634401	40644052	0.05	0.44	0.39	PHGR1
Primary colon tumor (Colon_P)	22	37406895	37441393	0.08	0.47	0.39	MPST
Primary colon tumor (Colon_P)	17	56404613	56418558	0.34	0.73	0.38	BZRAP1
Primary colon tumor (Colon_P)	8	145725241	145730866	0.32	0.70	0.38	GPT
Primary colon tumor (Colon_P)	1	167049094	167060421	0.09	0.46	0.38	GPA33
Primary colon tumor (Colon_P)	10	3816041	3829259	0.31	0.68	0.38	KLF6
Primary colon tumor (Colon_P)	21	43876507	43895539	0.03	0.39	0.37	SLC37A1
Primary colon tumor (Colon_P)	1	3369826	3378616	0.30	0.66	0.37	ARHGEF16
Primary colon tumor (Colon_P)	22	46479313	46489352	0.06	0.43	0.36	FLJ27365
Primary colon tumor (Colon_P)	7	130570945	130647038	0.16	0.51	0.36	MKLN1
Primary colon tumor (Colon_P)	1	202072763	202088434	0.10	0.43	0.33	GPR37L1
Primary colon tumor (Colon_P)	17	7161661	7168259	0.63	0.95	0.32	CLDN7
Primary colon tumor (Colon_P)	16	31137728	31147935	0.17	0.47	0.31	PRSS8
Primary colon tumor (Colon_P)	1	46632072	46655214	0.08	0.38	0.30	TSPAN1
Primary colon tumor (Colon_P)	20	62312602	62332044	0.16	0.45	0.29	TNFRSF6B
Primary colon tumor (Colon_P)	19	49115902	49124026	0.39	0.67	0.29	SPHK2
Primary colon tumor (Colon_P)	15	45726853	45749919	0.09	0.38	0.29	C15orf48
Primary colon tumor (Colon_P)	2	173291273	173330716	0.25	0.53	0.28	ITGA6
Primary colon tumor (Colon_P)	20	62153361	62177846	0.18	0.46	0.28	PTK6
Primary colon tumor (Colon_P)	2	200306225	200342782	0.56	0.84	0.28	SATB2
Primary colon tumor (Colon_P)	20	25200574	25236386	0.08	0.36	0.28	PYGB
Primary colon tumor (Colon_P)	21	40159118	40212366	0.18	0.44	0.27	ETS2
Primary colon tumor (Colon_P)	1	226052374	226087213	0.19	0.46	0.27	LEFTY1
Primary colon tumor (Colon_P)	3	58642829	58682800	0.06	0.32	0.26	FAM3D
Primary colon tumor (Colon_P)	16	574288	586492	0.27	0.52	0.26	SOLH
Primary colon tumor (Colon_P)	4	6747543	6755479	0.18	0.44	0.26	KIAA0232
Primary colon tumor (Colon_P)	14	50465553	50471375	0.66	0.92	0.25	C14orf182
Primary colon metastasis (Colon_M)	15	40634401	40644052	0.05	0.73	0.67	PHGR1
Primary colon metastasis (Colon_M)	11	64654606	64662582	0.23	0.87	0.64	EHD1
Primary colon metastasis (Colon_M)	16	70091792	70101175	0.34	0.96	0.61	RP11-419C5.2
Primary colon metastasis (Colon_M)	17	75277484	75285351	0.20	0.80	0.60	9-Sep
Primary colon metastasis (Colon_M)	11	34644271	34676099	0.09	0.64	0.55	EHF
Primary colon metastasis (Colon_M)	17	56404613	56418558	0.34	0.87	0.53	BZRAP1
Primary colon metastasis (Colon_M)	7	130570945	130647038	0.16	0.66	0.51	MKLN1
Primary colon metastasis (Colon_M)	7	50410776	50436416	0.00	0.49	0.49	IKZF1
Primary colon metastasis (Colon_M)	9	140082154	140094655	0.20	0.68	0.49	TPRN
Primary colon metastasis (Colon_M)	1	153579813	153590695	0.05	0.53	0.49	S100A14
Primary colon metastasis (Colon_M)	11	72485443	72499721	0.10	0.57	0.47	STARD10
Primary colon metastasis (Colon_M)	1	234949066	234963582	0.06	0.50	0.45	IRF2BP2
Primary colon metastasis (Colon_M)	1	234949066	234963582	0.06	0.50	0.45	IRF2BP2
Primary colon metastasis (Colon_M)	10	3816041	3829259	0.31	0.75	0.44	KLF6
Primary colon metastasis (Colon_M)	16	637814	644973	0.45	0.89	0.44	RAB40C
Primary colon metastasis (Colon_M)	5	169688532	169740245	0.01	0.45	0.43	LCP2
Primary colon metastasis (Colon_M)	22	46479313	46489352	0.06	0.49	0.42	FLJ27365
Primary colon metastasis (Colon_M)	1	155900806	155917642	0.20	0.62	0.42	RXFP4
Primary colon metastasis (Colon_M)	4	6747543	6755479	0.18	0.60	0.41	KIAA0232
Primary colon metastasis (Colon_M)	1	1092807	1104591	0.35	0.75	0.41	TTLL10
Primary colon metastasis (Colon_M)	7	142493655	142510845	0.11	0.51	0.40	PRSS1
Primary colon metastasis (Colon_M)	1	3369826	3378616	0.30	0.69	0.39	ARHGEF16
Primary colon metastasis (Colon_M)	7	50345823	50370722	0.06	0.45	0.39	IKZF1
Primary colon metastasis (Colon_M)	2	219149842	219160653	0.44	0.82	0.38	TMBIM1
Primary colon metastasis (Colon_M)	1	202072763	202088434	0.10	0.47	0.37	GPR37L1
Primary colon metastasis (Colon_M)	3	156836974	156855523	0.13	0.49	0.36	CCNL1
Primary colon metastasis (Colon_M)	8	145725241	145730866	0.32	0.68	0.35	GPT
Primary colon metastasis (Colon_M)	16	31137728	31147935	0.17	0.51	0.35	PRSS8
Primary colon metastasis (Colon_M)	13	27523605	27558934	0.13	0.47	0.34	USP12
Primary colon metastasis (Colon_M)	17	46021666	46045798	0.16	0.50	0.34	CDK5RAP3
Primary colon metastasis (Colon_M)	6	41031558	41042616	0.34	0.67	0.33	NFYA
Primary colon metastasis (Colon_M)	1	113240401	113250840	0.29	0.62	0.32	RHOC
Primary colon metastasis (Colon_M)	18	72918561	72932349	0.40	0.72	0.32	TSHZ1
Primary colon metastasis (Colon_M)	17	7161661	7168259	0.63	0.94	0.31	CLDN7
Primary colon metastasis (Colon_M)	3	9938065	9959891	0.17	0.47	0.31	IL17RC
Primary colon metastasis (Colon_M)	17	81030950	81060926	0.29	0.59	0.30	METRNL
Primary colon metastasis (Colon_M)	2	200306225	200342782	0.56	0.86	0.30	SATB2
Primary colon metastasis (Colon_M)	19	49115902	49124026	0.39	0.68	0.30	SPHK2
Primary colon metastasis (Colon_M)	1	211499374	211527884	0.09	0.38	0.30	TRAF5
Primary colon metastasis (Colon_M)	12	6478956	6487347	0.24	0.53	0.29	LTBR
Primary colon metastasis (Colon_M)	12	89739271	89785574	0.15	0.44	0.29	DUSP6
Primary colon metastasis (Colon_M)	20	62312602	62332044	0.16	0.45	0.29	TNFRSF6B
Primary colon metastasis (Colon_M)	11	307666	312937	0.70	0.98	0.28	IFITM2
Primary colon metastasis (Colon_M)	5	175959707	175981606	0.36	0.64	0.28	CDHR2
Primary colon metastasis (Colon_M)	21	43876507	43895539	0.03	0.31	0.28	SLC37A1
Primary colon metastasis (Colon_M)	19	54970768	54983186	0.35	0.63	0.28	LENG9
Primary colon metastasis (Colon_M)	1	226052374	226087213	0.19	0.47	0.27	LEFTY1
Primary colon metastasis (Colon_M)	15	31546396	31565926	0.03	0.31	0.27	KLF13
Primary colon metastasis (Colon_M)	14	55558308	55599128	0.10	0.37	0.27	LGALS3
Primary colon metastasis (Colon_M)	2	145252863	145282644	0.47	0.74	0.27	ZEB2
Primary colon metastasis (Colon_M)	1	46632072	46655214	0.08	0.34	0.26	TSPAN1
Primary colon metastasis (Colon_M)	15	63333068	63347067	0.46	0.71	0.26	TPM1
Primary colon metastasis (Colon_M)	15	45726853	45749919	0.09	0.35	0.26	C15orf48
Primary colon metastasis (Colon_M)	11	1865282	1903941	0.12	0.38	0.26	LSP1
Primary colon metastasis (Colon_M)	14	50465553	50471375	0.66	0.92	0.25	C14orf182
Primary colon metastasis (Colon_M)	22	37406895	37441393	0.08	0.33	0.25	MPST
Primary colon metastasis (Colon_M)	2	47563284	47609335	0.08	0.33	0.25	EPCAM
Primary colon metastasis (Colon_M)	17	76402998	76414113	0.00	0.25	0.25	PGS1
Glioblastoma (U87MG)	15	77985239	78036242	0.02	1.00	0.98	LINGO1
Glioblastoma (U87MG)	1	112417478	112468665	0.03	1.00	0.97	KCND3
Glioblastoma (U87MG)	17	9811786	9891020	0.02	0.99	0.97	GAS7
Glioblastoma (U87MG)	11	89364627	89408904	0.00	0.97	0.97	NOX4
Glioblastoma (U87MG)	4	16653926	16697144	0.00	0.97	0.96	LDB2
Glioblastoma (U87MG)	11	7592529	7626823	0.06	1.00	0.94	PPFIBP2
Glioblastoma (U87MG)	9	19549453	19728789	0.03	0.97	0.94	SLC24A2
Glioblastoma (U87MG)	9	19549453	19728789	0.03	0.97	0.94	SLC24A2
Glioblastoma (U87MG)	6	97413825	97477513	0.03	0.96	0.93	KLHL32
Glioblastoma (U87MG)	17	9903813	9965252	0.05	0.98	0.93	GAS7
Glioblastoma (U87MG)	2	21473055	21505937	0.06	0.99	0.93	TDRD15
Glioblastoma (U87MG)	3	35676346	35724308	0.05	0.97	0.92	ARPP21
Glioblastoma (U87MG)	15	32065846	32164147	0.06	0.98	0.91	OTUD7A
Glioblastoma (U87MG)	11	119942989	119964740	0.06	0.97	0.91	TRIM29
Glioblastoma (U87MG)	9	137967864	138040365	0.09	0.99	0.90	OLFM1
Glioblastoma (U87MG)	4	7342499	7398832	0.09	0.99	0.90	SORCS2
Glioblastoma (U87MG)	10	1224882	1259633	0.08	0.97	0.89	ADARB2
Glioblastoma (U87MG)	11	83518536	83549646	0.00	0.89	0.89	DLG2
Glioblastoma (U87MG)	20	58177200	58286056	0.10	0.99	0.89	PHACTR3
Glioblastoma (U87MG)	7	154558028	154588504	0.08	0.97	0.89	DPP6
Glioblastoma (U87MG)	8	143575390	143626966	0.10	0.97	0.88	BAI1
Glioblastoma (U87MG)	11	49190901	49236799	0.04	0.92	0.88	FOLH1
Glioblastoma (U87MG)	5	175104513	175139823	0.03	0.90	0.88	HRH2
Glioblastoma (U87MG)	19	57200811	57239668	0.06	0.93	0.87	ZNF835
Glioblastoma (U87MG)	11	7650333	7699694	0.06	0.93	0.87	CYB5R2
Glioblastoma (U87MG)	1	110693399	110756870	0.07	0.94	0.87	KCNC4
Glioblastoma (U87MG)	11	117399960	117681020	0.12	0.99	0.87	DSCAML1
Glioblastoma (U87MG)	1	111123465	111163650	0.13	1.00	0.87	KCNA2
Glioblastoma (U87MG)	1	4712562	4775202	0.12	0.99	0.87	AJAP1
Glioblastoma (U87MG)	11	83939015	84004173	0.03	0.89	0.86	DLG2
Glioblastoma (U87MG)	4	42605460	42661658	0.12	0.98	0.86	ATP8A1
Glioblastoma (U87MG)	4	6440647	6473745	0.14	1.00	0.86	PPP2R2C
Glioblastoma (U87MG)	11	119843908	119906650	0.11	0.97	0.86	TRIM29
Glioblastoma (U87MG)	4	7296865	7329947	0.06	0.91	0.85	SORCS2
Glioblastoma (U87MG)	4	7296865	7329947	0.06	0.91	0.85	SORCS2
Glioblastoma (U87MG)	11	117721606	117754344	0.17	1.00	0.83	FXYD6

Glioblastoma (U87MG)	20	37432328	37531273	0.11	0.94	0.82	PPP1R16B
Glioblastoma (U87MG)	5	743055	802933	0.08	0.90	0.82	ZDHHC11B
Glioblastoma (U87MG)	11	64379909	64428733	0.07	0.89	0.82	NRXN2
Glioblastoma (U87MG)	11	133874209	133916196	0.12	0.93	0.81	JAM3
Glioblastoma (U87MG)	11	133874209	133916196	0.12	0.93	0.81	JAM3
Glioblastoma (U87MG)	6	97344914	97390757	0.16	0.96	0.80	KLHL32
Glioblastoma (U87MG)	11	114090121	114132225	0.07	0.86	0.80	NNMT
Glioblastoma (U87MG)	7	51292392	51394212	0.18	0.97	0.80	CDBL
Glioblastoma (U87MG)	2	60731415	60797670	0.15	0.95	0.80	BCL11A
Glioblastoma (U87MG)	19	29973373	30022034	0.17	0.96	0.79	VSTM2B
Glioblastoma (U87MG)	3	118677505	118754932	0.10	0.89	0.79	IGSF11
Glioblastoma (U87MG)	9	19747340	19790384	0.15	0.94	0.78	SLC24A2
Glioblastoma (U87MG)	11	131430276	131505034	0.15	0.92	0.78	NTM
Glioblastoma (U87MG)	13	113650879	113757822	0.08	0.85	0.77	AL137002.1
Glioblastoma (U87MG)	15	77895466	77967694	0.09	0.84	0.75	LINGO1
Glioblastoma (U87MG)	17	72226877	72248627	0.23	0.98	0.75	TTYH2
Glioblastoma (U87MG)	1	160019557	160042341	0.23	0.98	0.75	KCNJ10
Glioblastoma (U87MG)	9	8840393	8880342	0.24	0.98	0.74	PTPRD
Glioblastoma (U87MG)	9	19457142	19518307	0.09	0.83	0.74	ACER2
Glioblastoma (U87MG)	20	45187487	45213480	0.22	0.96	0.74	SLC13A3
Glioblastoma (U87MG)	8	9753618	9790735	0.15	0.89	0.74	MSRA
Glioblastoma (U87MG)	11	113926954	114070258	0.08	0.81	0.73	ZBTB16
Glioblastoma (U87MG)	11	134254603	134298147	0.26	0.99	0.73	B3GAT1
Glioblastoma (U87MG)	13	25710619	25792315	0.15	0.86	0.71	AMER2
Glioblastoma (U87MG)	4	36164927	36295153	0.11	0.82	0.71	DTHD1
Glioblastoma (U87MG)	3	11024732	11080079	0.23	0.94	0.71	SLC8A1
Glioblastoma (U87MG)	11	64441886	64486253	0.13	0.83	0.70	NRXN2
Glioblastoma (U87MG)	8	27089725	27139876	0.09	0.77	0.67	STMN4
Glioblastoma (U87MG)	5	141125726	141169237	0.03	0.70	0.67	ARAP3
Glioblastoma (U87MG)	6	166689003	166721330	0.20	0.87	0.67	PRR18
Glioblastoma (U87MG)	10	88016137	88053122	0.07	0.73	0.66	GRID1
Glioblastoma (U87MG)	8	143531641	143555608	0.30	0.96	0.65	BAI1
Glioblastoma (U87MG)	12	110034532	110070399	0.16	0.81	0.65	MVK
Glioblastoma (U87MG)	16	934803	971974	0.12	0.76	0.64	LMF1
Glioblastoma (U87MG)	10	88108134	88126248	0.29	0.94	0.64	GRID1
Glioblastoma (U87MG)	16	5005930	5093187	0.14	0.77	0.63	ALG1
Glioblastoma (U87MG)	10	98094383	98136176	0.18	0.80	0.62	OPALIN
Glioblastoma (U87MG)	16	1064221	1098971	0.24	0.85	0.61	SSTR5
Glioblastoma (U87MG)	17	3402742	3440702	0.12	0.70	0.58	TRPV3
Glioblastoma (U87MG)	1	204796255	204866707	0.21	0.79	0.58	NFASC
Glioblastoma (U87MG)	2	105022517	105068208	0.24	0.82	0.58	POU3F3
Glioblastoma (U87MG)	19	31043362	31082430	0.06	0.62	0.56	ZNF536
Glioblastoma (U87MG)	19	51033029	51070726	0.30	0.86	0.56	LRRC4B
Glioblastoma (U87MG)	3	39487852	39573984	0.34	0.88	0.54	MOBP
Glioblastoma (U87MG)	17	26265479	26324131	0.04	0.57	0.54	LYRM9
Glioblastoma (U87MG)	7	86272604	86301105	0.22	0.75	0.53	GRM3
Glioblastoma (U87MG)	3	10505723	10550300	0.05	0.56	0.51	ATP2B2
Glioblastoma (U87MG)	20	24441648	24488844	0.15	0.66	0.51	SYNDIG1
Glioblastoma (U87MG)	6	110640550	110689914	0.02	0.53	0.51	METTL24
Glioblastoma (U87MG)	21	43639119	43683392	0.22	0.71	0.49	ABCG1
Glioblastoma (U87MG)	15	89891804	89952636	0.35	0.82	0.47	POLG
Glioblastoma (U87MG)	6	69337682	69367340	0.33	0.79	0.46	BAI3
Glioblastoma (U87MG)	1	175825389	175858645	0.12	0.57	0.45	RFWD2
Glioblastoma (U87MG)	11	72284488	72365197	0.06	0.49	0.43	PDE2A
Glioblastoma (U87MG)	11	1405512	1442659	0.20	0.63	0.43	BRSK2
Glioblastoma (U87MG)	3	133460192	133484110	0.24	0.67	0.42	TF
Glioblastoma (U87MG)	1	156353833	156405678	0.29	0.70	0.41	C1orf61
Glioblastoma (U87MG)	15	45655307	45688573	0.24	0.65	0.41	GATM
Glioblastoma (U87MG)	14	78431807	78458522	0.14	0.54	0.41	ADCK1
Glioblastoma (U87MG)	1	204888687	205049666	0.13	0.51	0.38	CNTN2
Glioblastoma (U87MG)	2	164566271	164594418	0.35	0.72	0.36	FIGN
Glioblastoma (U87MG)	X	103025990	103049724	0.64	1.00	0.36	PLP1
Glioblastoma (U87MG)	1	200978848	201012524	0.14	0.49	0.35	KIF21B
Glioblastoma (U87MG)	12	113534226	113581503	0.11	0.46	0.35	RASAL1
Glioblastoma (U87MG)	11	131522182	131565052	0.19	0.53	0.34	NTM
Glioblastoma (U87MG)	8	53318133	53342376	0.29	0.62	0.34	ST18
Glioblastoma (U87MG)	13	113365298	113413184	0.12	0.45	0.33	ATP11A
Glioblastoma (U87MG)	7	37232585	37330045	0.11	0.44	0.32	ELMO1
Glioblastoma (U87MG)	4	17510272	17521459	0.21	0.53	0.32	CLRN2
Glioblastoma (U87MG)	8	65275409	65295224	0.24	0.55	0.31	BHLHE22
Glioblastoma (U87MG)	15	84214391	84255934	0.08	0.38	0.30	SH3GL3
Glioblastoma (U87MG)	3	42724031	42791782	0.06	0.36	0.30	CCDC13
Glioblastoma (U87MG)	1	20755054	20821566	0.19	0.48	0.30	CAMK2N1
Glioblastoma (U87MG)	21	34399873	34453980	0.28	0.57	0.29	OLIG1
Glioblastoma (U87MG)	11	133987173	134023589	0.17	0.46	0.28	JAM3
Glioblastoma (U87MG)	19	50985631	51019282	0.28	0.53	0.25	ASPDH

Supplementary Table 6.5: GO enrichment analysis of hypermethylated super-enhancers (GOTERM_BP_5, p<0.01).

GO term	# genes	% genes	Enrichment	p-value*	Sample**
Regulation of gene expression	16	57,1	3,6	8,7E-6	468PT
Regulation of transcription	16	57,1	4,0	9,1E-6	468PT
Regulation of macromolecule biosynthetic process	16	57,1	3,7	1,0E-5	468PT
Regulation of cellular biosynthetic process	16	57,1	3,5	1,1E-5	468PT
Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	16	57,1	3,7	1,4E-5	468PT
Regulation of transcription, DNA-dependent	13	46,4	4,7	3,0E-5	468PT
Regulation of RNA metabolic process	13	46,4	4,6	3,3E-5	468PT
Regulation of RNA metabolic process	18	35,3	3,4	6,1E-4	468LN
Vasculature development	8	15,7	10,9	7,9E-4	468LN
Regulation of transcription, DNA-dependent	18	35,3	3,5	8,9E-4	468LN
Blood vessel development	8	15,7	11,1	9,0E-4	468LN
Blood vessel morphogenesis	7	13,7	11,3	2,9E-3	468LN
Gland development	6	11,8	15,2	3,0E-3	468LN
Chordate embryonic development	8	15,7	8,3	3,2E-3	468LN
Angiogenesis	6	11,8	13,8	4,0E-3	468LN
Regulation of transcription	19	37,3	2,5	4,6E-3	468LN
Regulation of cellular biosynthetic process	20	39,2	2,3	6,0E-3	468LN
Regulation of gene expression	16	76,2	3,6	6,5E-6	H157
Regulation of transcription	16	76,2	4,0	6,8E-6	H157
Regulation of macromolecule biosynthetic process	16	76,2	3,7	7,6E-6	H157
Regulation of cellular biosynthetic process	16	76,2	3,5	8,1E-6	H157
Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	16	76,2	3,7	1,0E-5	H157
Regulation of transcription, DNA-dependent	12	57,1	4,4	2,4E-4	H157
Regulation of RNA metabolic process	12	57,1	4,3	2,5E-4	H157
Regulation of transcription	19	51,4	3,1	3,1E-4	H1672
Regulation of gene expression	19	51,4	2,8	3,2E-4	H1672
Regulation of transcription, DNA-dependent	15	40,5	3,6	3,4E-4	H1672
Positive Regulation of RNA metabolic process	9	24,3	7,9	3,6E-4	H1672
Regulation of macromolecule biosynthetic process	19	51,4	2,8	3,8E-4	H1672
Positive Regulation of transcription, DNA-dependent	9	24,3	8,0	3,9E-4	H1672
Regulation of RNA metabolic process	15	40,5	3,5	3,9E-4	H1672
Regulation of cellular biosynthetic process	19	51,4	2,7	4,2E-4	H1672
Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	19	51,4	2,9	5,2E-4	H1672
Positive Regulation of transcription	9	24,3	6,8	8,0E-4	H1672
Positive Regulation of gene expression	9	24,3	6,6	9,0E-4	H1672
Positive Regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	9	24,3	6,1	1,4E-3	H1672
Positive Regulation of nitrogen compound metabolic process	9	24,3	5,9	1,6E-3	H1672
Positive Regulation of macromolecule biosynthetic process	9	24,3	5,8	1,6E-3	H1672
Positive Regulation of cellular biosynthetic process	9	24,3	5,6	2,1E-3	H1672
Positive Regulation of biosynthetic process	9	24,3	5,5	2,2E-3	H1672
Regulation of striated muscle tissue development	4	10,8	33,9	3,5E-3	H1672
Positive Regulation of macromolecule metabolic process	9	24,3	4,5	8,2E-3	H1672
Tissue development	8	21,6	5,1	8,8E-3	H1672
Positive Regulation of cellular metabolic process	9	24,3	4,3	9,3E-3	H1672

*** Bonferroni corrected**

**** Only samples with significantly enriched GO terms**

Supplementary Table 6.6: Validation of hypermethylated super-enhancers in cancer samples (HumanMethylation450 BeadChip) .

Cancer type	Super-enhancer ID	Gene	Chr.	Start	End	Validation set δ AVR	FDR	WGBS δ HMR	CNV FDR
Breast cancer	1_MACS_peak_2257_lociStitched	CKS1B	1	154941228	154949531	0.05	9.26E-08	0.34	9.97E-01
Breast cancer	2_MACS_peak_23111_lociStitched	RBMS3	3	29321550	29334286	0.06	9.30E-06	0.65	9.97E-01
Breast cancer	2_MACS_peak_23764_lociStitched	FOXP1	3	71104592	71120542	0.01	4.33E-01	0.28	9.97E-01
Breast cancer	1_MACS_peak_27137_lociStitched	IRX4	5	1881509	1890183	0.08	1.20E-04	0.67	9.97E-01
Breast cancer	3_MACS_peak_27146_lociStitched	IRX2	5	2739262	2759651	0.03	4.58E-02	0.60	9.97E-01
Breast cancer	1_MACS_peak_28648_lociStitched	NR3C1	5	142773722	142785691	0.09	8.80E-12	0.63	9.97E-01
Breast cancer	1_MACS_peak_31912_lociStitched	INHBA	7	41734099	41745805	0.06	1.94E-06	0.34	9.97E-01
Breast cancer	6_MACS_peak_5342_lociStitched	FAM53B	10	126406768	126435377	0.01	1.84E-01	0.26	9.97E-01
Breast cancer	1_MACS_peak_5527_lociStitched	PRKCDBP	11	6339234	6344196	0.10	1.04E-08	0.32	9.97E-01
Breast cancer	3_MACS_peak_7389_lociStitched	TNFRSF1A	12	6432600	6452223	0.01	5.72E-03	0.29	9.97E-01
Breast cancer	2_MACS_peak_8659_lociStitched	BTG1	12	92527863	92539838	0.03	1.27E-05	0.35	9.97E-01
Breast cancer	4_MACS_peak_10930_lociStitched	ELMSAN1	14	74207241	74227902	-0.01	7.93E-02	0.36	9.97E-01
Breast cancer	1_MACS_peak_11720_lociStitched	GABPB1	15	50644604	50649357	0.00	4.10E-02	0.29	2.33E-01
Breast cancer	MACS_peak_12476	CHD2	15	93425946	93432559	0.01	2.01E-03	0.31	9.97E-01
Breast cancer	3_MACS_peak_17254_lociStitched	CIC	19	42771661	42788152	0.03	5.75E-12	0.27	9.97E-01
Breast cancer	1_MACS_peak_21219_lociStitched	RBM39	20	34326330	34331841	0.00	5.85E-04	0.30	9.97E-01
Breast cancer	2_MACS_peak_21515_lociStitched	TSHZ2	20	51561498	51597746	0.06	7.91E-08	0.37	9.97E-01
Breast cancer	1_MACS_peak_21859_lociStitched	RUNX1	21	36252652	36264673	0.04	2.53E-03	0.42	9.97E-01
Breast cancer	MACS_peak_22783	MIRLET7	22	46467114	46488203	0.16	2.61E-07	0.61	2.33E-01
Colorectal cancer	3_MACS_peak_30801_lociStitched	PITX1	5	134361067	134376085	0.09	2.24E-06	0.29	8.18E-01
Glioblastoma	13_MACS_peak_22858_lociStitched	BIN1	2	127806748	127912517	0.06	1.66E-02	0.32	5.54E-01
Glioblastoma	15_MACS_peak_22991_lociStitched	ZEB2	2	145129528	145283099	0.00	8.77E-01	0.34	9.23E-01
Glioblastoma	6_MACS_peak_23868_lociStitched	PTMA	2	232526039	232554844	0.07	6.33E-04	0.29	9.23E-01
Glioblastoma	6_MACS_peak_29107_lociStitched	SOX2	3	181403893	181455295	0.01	8.20E-01	0.36	4.22E-01
Glioblastoma	1_MACS_peak_31287_lociStitched	SLC1A3	5	36599886	36615091	0.00	8.95E-01	0.33	3.74E-02
Glioblastoma	MACS_peak_32260	SEPT8	5	132105879	132116081	0.05	1.29E-02	0.27	9.71E-01
Glioblastoma	2_MACS_peak_34748_lociStitched	SGK1	6	134488737	134505092	0.06	2.19E-02	0.38	9.73E-01
Glioblastoma	12_MACS_peak_35205_lociStitched	QKI	6	163815822	163882069	0.01	8.17E-01	0.65	9.71E-01
Glioblastoma	5_MACS_peak_40731_lociStitched	RALGDS	9	135976307	136024639	0.02	1.30E-01	0.26	9.23E-01
Glioblastoma	1_MACS_peak_40902_lociStitched	TPRN	9	140081775	140094896	0.05	3.59E-03	0.35	9.23E-01
Glioblastoma	11_MACS_peak_5386_lociStitched	ZCCHC24	10	81137824	81212605	0.04	6.60E-03	0.38	9.23E-01
Glioblastoma	3_MACS_peak_5703_lociStitched	LZTS2	10	102752997	102775915	0.04	3.35E-02	0.28	4.22E-01
Glioblastoma	3_MACS_peak_7435_lociStitched	CARNS1	11	67173218	67188540	0.07	5.40E-05	0.38	9.23E-01
Glioblastoma	MACS_peak_9242	AC004466.1	12	48167031	48182319	0.00	7.19E-01	0.35	9.23E-01
Glioblastoma	3_MACS_peak_11368_lociStitched	PCDH9	13	67782754	67806323	0.01	8.20E-01	0.45	5.54E-01
Glioblastoma	1_MACS_peak_12563_lociStitched	GSTZ1	14	77765872	77789479	0.02	2.56E-01	0.28	9.58E-01
Glioblastoma	4_MACS_peak_13804_lociStitched	OAZ2	15	64974641	64996542	0.02	2.43E-02	0.30	9.23E-01
Glioblastoma	MACS_peak_17165	THRA	17	38216903	38234316	0.01	7.56E-01	0.36	9.23E-01
Glioblastoma	4_MACS_peak_17901_lociStitched	KCNJ2	17	68152783	68183190	0.00	7.48E-01	0.29	9.23E-01
Glioblastoma	5_MACS_peak_18355_lociStitched	BAHCC1	17	79345155	79376776	-0.01	2.32E-01	0.27	9.23E-01
Glioblastoma	3_MACS_peak_26411_lociStitched	CDC42EP1	22	37940770	37968154	0.06	5.93E-04	0.32	9.23E-01
Lung adenocarcinoma	MACS_peak_3232	TXNIP	1	145434374	145443479	0.00	8.52E-01	0.48	5.29E-01
Lung adenocarcinoma	3_MACS_peak_30019_lociStitched	MEIS1	2	66659175	66674430	0.14	4.25E-18	0.45	no CNV
Lung adenocarcinoma	2_MACS_peak_39322_lociStitched	GATA2	3	128199015	128216666	0.08	2.96E-08	0.53	no CNV
Lung adenocarcinoma	4_MACS_peak_42669_lociStitched	HAND2	4	174427206	174460857	0.12	3.22E-12	0.48	no CNV
Lung adenocarcinoma	5_MACS_peak_42987_lociStitched	IRX2	5	2738590	2759482	0.07	1.60E-05	0.74	5.29E-01
Lung adenocarcinoma	5_MACS_peak_45552_lociStitched	NKX2-5	5	172655376	172676296	0.09	4.62E-08	0.58	no CNV
Lung adenocarcinoma	MACS_peak_54700	CREB3	9	35726029	35733751	-0.01	2.73E-02	0.31	NA
Lung adenocarcinoma	3_MACS_peak_7162_lociStitched	ZNF503-AS2	10	77154286	77170840	0.02	2.08E-02	0.26	no CNV
Lung adenocarcinoma	MACS_peak_17492	CRIP2	14	105938037	105944737	0.00	6.84E-01	0.34	no CNV
Lung adenocarcinoma	MACS_peak_22262	ZBTB4	17	7378645	7384073	0.01	1.38E-01	0.36	no CNV
Lung adenocarcinoma	MACS_peak_36631	MIRLET7	22	46445515	46454339	0.02	1.82E-01	0.29	NA
Lung adenocarcinoma	MACS_peak_36635	MIRLET7	22	46464333	46478321	0.03	1.82E-02	0.27	NA
Lung squamous cell carcinoma	MACS_peak_3232	TXNIP	1	145434374	145443479	0.02	1.32E-01	0.48	no CNV
Lung squamous cell carcinoma	2_MACS_peak_28599_lociStitched	ID2	2	8815972	8826211	0.00	5.29E-01	0.59	9.86E-07
Lung squamous cell carcinoma	3_MACS_peak_30019_lociStitched	MEIS1	2	66659175	66674430	0.11	4.14E-13	0.52	1.91E-01
Lung squamous cell carcinoma	3_MACS_peak_32032_lociStitched	CLK1	2	201724772	201732443	0.02	5.41E-06	0.35	4.78E-02
Lung squamous cell carcinoma	2_MACS_peak_37947_lociStitched	SHISA5	3	48504418	48515502	0.01	7.22E-01	0.26	2.92E-01
Lung squamous cell carcinoma	2_MACS_peak_39322_lociStitched	GATA2	3	128199015	128216666	0.08	5.59E-07	0.63	1.91E-01
Lung squamous cell carcinoma	4_MACS_peak_42669_lociStitched	HAND2	4	174427206	174460857	0.13	6.38E-14	0.67	NA
Lung squamous cell carcinoma	5_MACS_peak_42987_lociStitched	IRX2	5	2738590	2759482	0.15	6.38E-14	0.63	NA
Lung squamous cell carcinoma	4_MACS_peak_43416_lociStitched	ANXA2R	5	42990572	43020949	0.09	2.87E-07	0.26	2.79E-02
Lung squamous cell carcinoma	MACS_peak_54700	CREB3	9	35726029	35733751	-0.02	4.14E-03	0.32	1.15E-01
Lung squamous cell carcinoma	3_MACS_peak_7162_lociStitched	ZNF503-AS2	10	77154286	77170840	0.04	1.91E-05	0.38	1.91E-01
Lung squamous cell carcinoma	1_MACS_peak_11040_lociStitched	DDX6	11	118659939	118664216	0.01	5.77E-01	0.38	3.75E-01
Lung squamous cell carcinoma	MACS_peak_11710	TNFRSF1A	12	6441703	6453465	0.02	1.82E-01	0.39	NA
Lung squamous cell carcinoma	MACS_peak_11762	ATN1	12	7032586	7039696	0.00	5.35E-01	0.34	NA
Lung squamous cell carcinoma	5_MACS_peak_13097_lociStitched	BTG1	12	92526891	92540701	0.00	9.18E-01	0.43	NA
Lung squamous cell carcinoma	6_MACS_peak_13707_lociStitched	TBX3	12	115105978	115131359	0.08	3.09E-10	0.36	NA
Lung squamous cell carcinoma	7_MACS_peak_15952_lociStitched	NKX2-1	14	36971679	36995027	0.13	7.34E-12	0.49	2.92E-01
Lung squamous cell carcinoma	MACS_peak_17492	CRIP2	14	105938037	105944737	0.00	4.53E-01	0.32	no CNV
Lung squamous cell carcinoma	4_MACS_peak_20804_lociStitched	IRX5	16	54956715	54973307	0.09	1.03E-11	0.32	NA
Lung squamous cell carcinoma	MACS_peak_22262	ZBTB4	17	7378645	7384073	0.02	2.86E-03	0.37	NA
Lung squamous cell carcinoma	MACS_peak_22267	SENP3	17	7459853	7466333	0.01	1.78E-01	0.26	NA
Lung squamous cell carcinoma	MACS_peak_23136	THRA	17	38222943	38234141	0.06	3.97E-05	0.35	no CNV
Lung squamous cell carcinoma	MACS_peak_36631	MIRLET7	22	46445515	46454339	0.01	4.99E-01	0.26	NA
Lung squamous cell carcinoma	MACS_peak_36635	MIRLET7	22	46464333	46478321	0.03	3.82E-02	0.32	NA
Small cell lung cancer	MACS_peak_42	VWA1	1	1364851	1372426	0.02	4.71E-01	0.32	9.12E-01
Small cell lung cancer	MACS_peak_3232	TXNIP	1	145434374	145443479	0.01	5.20E-01	0.29	2.81E-01
Small cell lung cancer	MACS_peak_3436	S100A2	1	153537691	153541885	-0.03	3.10E-02	0.32	4.02E-01
Small cell lung cancer	3_MACS_peak_30019_lociStitched	MEIS1	2	66659175	66674430	0.19	1.70E-24	0.45	9.12E-01
Small cell lung cancer	3_MACS_peak_32032_lociStitched	CLK1	2	201724772	201732443	0.01	2.01E-03	0.35	4.52E-01
Small cell lung cancer	2_MACS_peak_39322_lociStitched	GATA2	3	128199015	128216666	0.09	8.82E-07	0.50	9.12E-01
Small cell lung cancer	4_MACS_peak_42669_lociStitched	HAND2	4	174427206	174460857	0.17	1.74E-15	0.50	9.12E-01
Small cell lung cancer	5_MACS_peak_42987_lociStitched	IRX2	5	2738590	2759482	0.13	3.90E-10	0.35	9.12E-01
Small cell lung cancer	5_MACS_peak_45552_lociStitched	NKX2-5	5	172655376	172676296	0.14	2.02E-11	0.26	7.99E-01
Small cell lung cancer	2_MACS_peak_46644_lociStitched	NRM	6	30646634	30659820	0.04	8.43E-04	0.39	1.42E-02
Small cell lung cancer	2_MACS_peak_46715_lociStitched	CLIC1	6	31694582	31706840	0.02	7.73E-02	0.30	8.38E-01
Small cell lung cancer	1_MACS_peak_49158_lociStitched	MAFK	7	1569896	1580027	0.01	6.46E-01	0.28	9.12E-01
Small cell lung cancer	4_MACS_peak_54476_lociStitched	NFIB	9	14302438	14323420	0.01	3.41E-01	0.28	1.57E-01
Small cell lung cancer	4_MACS_peak_54483_lociStitched	NFIB	9	14340900	14353877	0.06	7.73E-02	0.25	9.12E-01
Small cell lung cancer	MACS_peak_54700	CREB3	9	35726029	35733751	-0.01	4.66E-01	0.36	9.12E-01
Small cell lung cancer	3_MACS_peak_7162_lociStitched	ZNF503-AS2	10	77154286	77170840	0.04	7.31E-05	0.29	7.99E-01
Small cell lung cancer	MACS_peak_7364	C10orf116	10	88726022	88733482	0.01	6.46E-01	0.27	2.39E-01
Small cell lung cancer	MACS_peak_9994	FRMD8	11	65182856	65196355	0.02	1.82E-01	0.76	4.52E-01
Small cell lung cancer	MACS_peak_10139	TCIRG1	11	67803279	67809641	0.01	6.09E-01	0.37	1.57E-01
Small cell lung cancer	1_MACS_peak_11040_lociStitched	DDX6	11	118659939	118664216	0.01	1.50E-01	0.35	9.12E-01
Small cell lung cancer	2_MACS_peak_11044_lociStitched	UPK2	11	118777304	118801687	0.00	8.56E-01	0.46	6.46E-01
Small cell lung cancer	MACS_peak_11710	TNFRSF1A	12	6441703	6453465	0.02	2.30E-02	0.61	5.80E-01
Small cell lung cancer	5_MACS_peak_13097_lociStitched	BTG1	12	92526891	92540701	0.00	9.62E-01	0.38	9.81E-01
Small cell lung cancer	9_MACS_peak_13696_lociStitched	TBX5	12	114806250	114852976	0.14	6.63E-10	0.43	7.99E-01
Small cell lung cancer	6_MACS_peak_13707_lociStitched	TBX3	12	115105978	115131359	0.07	9.04E-07	0.28	9.12E-01
Small cell lung cancer	7_MACS_peak_15952_lociStitched	NKX2-1	14	36971679	36995027	0.09	2.49E-05	0.69	9.12E-01
Small cell lung cancer	MACS_peak_17492	CRIP2	14	105938037	105944737	-0.01	5.23E-01	0.47	9.81E-01
Small cell lung cancer	4_MACS_peak_20804_lociStitched	IRX5	16	54956715	54973307	0.08	2.07E-09	0.40	9.12E-01
Small cell lung cancer	1_MACS_peak_20826_lociStitched	MT2A	16	56638742	56646907	0.04	3.34E-02	0.25	9.12E-01
Small cell lung cancer	MACS_peak_22262	ZBTB4	17	7378645	7384073	0.05	6.28E-07	0.36	7.24E-01
Small cell lung cancer	MACS_peak_23136	THRA	17	38222943	38234141	0.03	1.15E-01	0.48	2.39E-01
Small cell lung cancer	1_MACS_peak_24752_lociStitched	ACTG1	17	79475627	79487805	-0.01	9.09E-02	0.36	9.12E-01
Small cell lung cancer	3_MACS_peak_24980_lociStitched	DLGAP1	18	3591992	3606653	-0.01	6.37E-01	0.31	3.82E-01
Small cell lung cancer	MACS_peak_26165	MIDN	19	1247145	1256325	0.02	9.48E-03	0.30	no CNV
Small cell lung cancer	MACS_peak_26166	CIRBP	19	1256391	1263727	-0.01	5.54E-01	0.34	NA
Small cell lung cancer	MACS_peak_27682	ITPKC	19	41219787	41228823	-0.02	1.21E-02	0.29	no CNV
Small cell lung cancer	6_MACS_peak_33435_lociStitched	FOXA2	20	22536539	22565230	0.09	5.88E-06	0.51	8.38E-01
Small cell lung cancer	MACS_peak_36245	CSNK1E	22	38708087	38715695	0.00	8.64E-01	0.31	9.81E-01
Small cell lung cancer	MACS_peak_36631	MIRLET7	22	46445515	46454339	-0.02	2.34E-01	0.45	9.12E-01
Small cell lung cancer	MACS_peak_36635	MIRLET7	22	46464333	46478321	-0.01	7.51E-01	0.47	7.88E-01

Supplementary Table 6.7: Large hypomethylated regions in primary and metastatic colorectal cancer samples.

Sample type	Chr.	Start	End	Size (bp)	Meth. level	Gene symbol	Distance TSS	% HMR normal colon
Primary tumor	1	218336382	218349940	13558	0.09	TGFB2	168736	0.28
Primary tumor	13	73628949	73641647	12698	0.15	KLF5	0	0.30
Primary tumor	9	74373398	74385279	11881	0.10	TMEM2	0	0.28
Primary tumor	2	91663698	91675203	11505	0.10	IGKV1OR2-118	3693	0.46
Primary tumor	8	117884649	117895539	10890	0.13	RAD21	0	0.25
Primary tumor	2	90448882	90459317	10435	0.09	AC233263.3	53454	0.19
Primary tumor	21	34437234	34447500	10266	0.19	OLIG1	0	0.36
Primary tumor	5	127414083	127424041	9958	0.06	SLC12A2	0	0.36
Primary tumor	1	207921099	207930539	9440	0.13	CD46	0	0.20
Primary tumor	1	203289774	203299097	9323	0.16	U6	1547	0.40
Primary tumor	20	25227101	25236365	9264	0.14	PYGB	0	0.31
Primary tumor	2	173291608	173300419	8811	0.12	ITGA6	0	0.42
Primary tumor	22	37410887	37419122	8235	0.19	MPST	0	0.23
Primary tumor	11	46933497	46941707	8210	0.11	LRP4	0	0.39
Primary tumor	15	63795181	63803381	8200	0.04	USP3	0	0.47
Primary tumor	21	33244343	33252482	8139	0.09	HUNK	0	0.43
Primary tumor	18	55463636	55471500	7864	0.13	ATP8B1	0	0.32
Primary tumor	2	230577345	230585201	7856	0.11	DNER	0	0.36
Primary tumor	10	32338334	32346117	7783	0.12	KIF5B	0	0.45
Primary tumor	6	124122411	124130163	7752	0.18	NKAIN2	0	0.42
Primary tumor	12	33044638	33052291	7653	0.13	PKP2	0	0.30
Primary tumor	5	94615687	94623331	7644	0.11	MCTP1	0	0.39
Primary tumor	1	207222956	207230560	7604	0.13	PFKFB2	0	0.45
Primary tumor	6	128835158	128842574	7416	0.06	PTPRK	0	0.39
Primary tumor	7	17337977	17345308	7331	0.07	AHR	0	0.32
Primary tumor	14	55591597	55598891	7294	0.14	LGALS3	0	0.26
Metastasis	3	41231323	41265376	34053	0.14	CTNNB1	0	0.15
Metastasis	5	127414053	127440949	26896	0.10	SLC12A2	0	0.13
Metastasis	22	46459859	46480180	20321	0.14	MIRLET7	0	0.43
Metastasis	9	29192909	29213215	20306	0.18	LINGO2	41208	0.20
Metastasis	5	94614719	94634334	19615	0.17	MCTP1	0	0.15
Metastasis	21	40174290	40193387	19097	0.18	ETS2	0	0.46
Metastasis	17	63540803	63558867	18064	0.12	AXIN2	0	0.35
Metastasis	2	90442156	90460214	18058	0.08	AC233263.3	52557	0.27
Metastasis	3	87025252	87041710	16458	0.19	VGLL3	0	0.10
Metastasis	4	75229835	75245790	15955	0.20	EREG	0	0.20
Metastasis	2	91667887	91682638	14751	0.08	IGKV1OR2-118	0	0.30
Metastasis	1	218336537	218350719	14182	0.09	TGFB2	167958	0.26
Metastasis	14	94639297	94652604	13307	0.16	PPP4R4	0	0.31
Metastasis	9	74372950	74385279	12329	0.07	TMEM2	0	0.27
Metastasis	12	89739454	89751618	12164	0.05	DUSP6	0	0.43
Metastasis	13	73628491	73640649	12158	0.09	KLF5	0	0.31
Metastasis	17	56404994	56417144	12150	0.10	MIR4736	0	0.40
Metastasis	21	47062453	47074317	11864	0.14	PCBP3	0	0.18
Metastasis	1	207919635	207930507	10872	0.09	CD46	0	0.17
Metastasis	22	46433823	46444602	10779	0.18	MIRLET7	0	0.16
Metastasis	2	132767266	132777949	10683	0.17	RP1-156L9.1	15532	0.07
Metastasis	5	88172728	88183358	10630	0.18	CTC-454M9.1	0	0.49
Metastasis	11	46931142	46941707	10565	0.10	LRP4	0	0.30
Metastasis	11	25440840	25450991	10151	0.18	AC015820.1	107833	0.00
Metastasis	8	106324990	106335123	10133	0.12	ZFPM2	0	0.26
Metastasis	14	66972981	66982851	9870	0.08	GPHN	0	0.29
Metastasis	1	142928205	142937914	9709	0.20	RP11-423O2.7	18802	0.00
Metastasis	21	33243902	33253577	9675	0.11	HUNK	0	0.36
Metastasis	21	10105053	10114480	9427	0.20	CR381653.1	66688	0.16
Metastasis	3	79060095	79069497	9402	0.09	ROBO1	0	0.43
Metastasis	1	203289774	203299097	9323	0.13	U6	1547	0.40
Metastasis	1	11530768	11540086	9318	0.12	PTCHD2	0	0.18
Metastasis	2	41704857	41713858	9001	0.12	AC010739.1	109668	0.00
Metastasis	1	37500826	37509794	8968	0.19	GRIK3	1096	0.46
Metastasis	14	88781784	88790680	8896	0.16	KCNK10	0	0.27
Metastasis	1	151960764	151969653	8889	0.10	AL591893.1	0	0.50
Metastasis	6	69935822	69944658	8836	0.18	BAI3	0	0.13
Metastasis	7	102787637	102796362	8725	0.15	RP11-401L13.5	0	0.20
Metastasis	2	106007479	106016181	8702	0.18	FHL2	0	0.15
Metastasis	7	26191459	26200010	8551	0.08	NFE2L3	0	0.37
Metastasis	2	230577164	230585636	8472	0.11	DNER	0	0.34
Metastasis	13	67803752	67812201	8449	0.13	PCDH9	0	0.28
Metastasis	3	89155675	89164093	8418	0.13	EPHA3	0	0.39
Metastasis	13	78265037	78273411	8374	0.04	MIR3665	0	0.45
Metastasis	18	53249062	53257434	8372	0.10	TCF4	0	0.38
Metastasis	9	177326	185696	8370	0.16	CBWD1	0	0.24
Metastasis	8	92610199	92618527	8328	0.19	7SK	3006	0.29
Metastasis	15	79596231	79604549	8318	0.15	TMED3	0	0.19
Metastasis	12	33044638	33052875	8237	0.10	PKP2	0	0.28
Metastasis	12	89912441	89920652	8211	0.08	GALNT4	0	0.31
Metastasis	20	5583918	5592121	8203	0.08	GPCPD1	0	0.42
Metastasis	7	30322505	30330689	8184	0.06	MIR550A1	0	0.32
Metastasis	1	235092612	235100792	8180	0.15	RP11-443B7.1	0	0.31
Metastasis	14	65167106	65175127	8021	0.12	PLEKHG3	0	0.39
Metastasis	10	32338291	32346311	8020	0.07	Y_RNA	0	0.44
Metastasis	15	63795181	63803166	7985	0.04	USP3	0	0.49
Metastasis	12	12937531	12945460	7929	0.13	APOLD1	0	0.15
Metastasis	11	94276085	94283992	7907	0.06	FUT4	0	0.23
Metastasis	8	117884817	117892656	7839	0.07	MIR3610	0	0.32
Metastasis	3	12328878	12336686	7808	0.09	PPARG	0	0.44
Metastasis	5	129239340	129247107	7767	0.10	CHSY3	0	0.48
Metastasis	1	207222956	207230681	7725	0.10	PFKFB2	0	0.45
Metastasis	7	77320936	77328644	7708	0.12	RSBN1L	0	0.33
Metastasis	10	11055794	11063495	7701	0.07	CELF2	0	0.32
Metastasis	20	31170742	31178437	7695	0.09	RP11-410N8.4	0	0.41
Metastasis	1	46595249	46602845	7596	0.08	RP4-533D7.5	0	0.42
Metastasis	2	69607934	69615506	7572	0.11	GFPT1	0	0.29
Metastasis	12	71832196	71839719	7523	0.05	LGR5	0	0.41
Metastasis	2	91995135	92002654	7519	0.05	AC127391.1	3143	0.15

CHAPTER VII

Discussion

1. Epigenetic Transcriptional Silencing of ncRNAs

The classical view of cancer interprets the malignant phenomena based on *genetics* and on the tiny part of our genome that encodes *protein-coding genes*. Its standard pharmacological treatment is generally indiscriminate, targeting both cancer and healthy cells, causing severe side effects[655]. However specific treatment biomarkers (e.g. DNA methylation status of the promoter of *MGMT* in gliomas[656]) and the development of new specific drugs that target genetic alterations (e.g. *BCR-ABL* fusion gene in CML[657]; or *BRAF* activating mutations in melanoma[658]) may improve the clinical outcome. However, due to the fact that these alterations are not exclusive, cancer cells frequently escape the inhibited pathways, activating alternative ones, to survive and to proliferate.

The accumulation of irreversible genetic mutations during tumor progression nourishes a tremendous challenge in cancer: the reversible modulation of gene expression in order to control tumor progression and/or metastasis. During this PhD thesis we aimed to uncover new layers of gene regulation, namely those related to ncRNAs that are continuously being identified as key regulators in tumorigenesis and are themselves able to modulate the expression of other RNAs. For instance, several studies have demonstrated that miRNAs and lncRNAs with tumor suppressor features become commonly silenced by CGI hypermethylation[236,659-661]. Additionally, genome-wide studies realized that the originally assigned "junk DNA" encodes for non-coding transcripts[497]. The identification and better understanding of epigenetic pathways altered during the carcinogenesis process would help to expose part of the complexity of a cancer cell, uncovering new biomarkers and therapeutic targets.

snoRNAs are localized in the nucleolus, guiding post-transcriptional modifications of spliceosomal and ribosomal RNAs[540,662-664]. Consequently, at ribosome level, snoRNAs cooperate for their correct assembly and function[665]. Nevertheless, snoRNAs have shown to have unpredicted functions in oncogenesis, being disrupted by copy number variation, mutations, altered expression and chromosomal translocations in several malignancies[551,552,666,667]. Exemplarily, mutations in *dyskerin (DKC1)*, the gene that codes for the enzyme that associates with H/ACA box snoRNAs to catalyze the pseudouridylation of rRNAs, increases cancer susceptibility[668]. On the other hand, modifications in ribosome biogenesis is implicated in tumor progression[669,670],

suggesting that snoRNAs could be involved in cancer by an altered guiding of post-transcriptional modifications of rRNA.

In our first study we identified three snoRNAs that are recurrently repressed by hypermethylation of the CGI overlapping the promoter region of their host gene, in a panel of cancer cell lines and clinical samples. We first noticed that the snoRNAs *SNORD123*, *U70C* and *ACA59B* became heavily hypermethylated in HCT-116 colorectal cancer cell line in comparison with normal colon mucosa. At transcriptional level, we verified that the hypermethylation of snoRNA-related CGIs was associated with the gene silencing of both host genes and associated snoRNAs. Interestingly, we observed that one of these CGIs was able to modulate the expression of three different RNA entities at the same time, the snoRNA *SNORD123*, its host lncRNA (*LOC100505806*) and *SEMA5A*, the last one transcribed in the opposite direction relative to the same differentially methylated CGI. Taking this into account, we amplified our study to other tissues, verifying a specific CGI hypermethylation in cell lines derived from different types of cancer. Interestingly, besides confirming the hypermethylation of these CGIs in other colorectal cancer cell lines we noticed a similar profile in a substantial proportion of leukemia cell lines. Moving to patient samples, we observed concordant DNA hypermethylation of these snoRNAs in a considerable fraction of ALL samples. In AML and primary multiple myeloma samples, CGI hypermethylation was observed for *SNORD123* or *ACA59B*, respectively.

Similarly to some described coding and non-coding genes holding tumor suppressor functions[49,194,533,537], for the first time we reported three different snoRNAs that similarly undergo cancer-specific CGI hypermethylation-associated silencing. Accordingly, these events also lead to the transcriptional inactivation of their host genes and some studies revealed that some snoRNA host genes are deregulated in cancer, playing an important role in tumorigenesis. For instance, the snoRNA host lncRNA *ZNFX1 antisense RNA 1* (*ZFAS1*) was recognized as a tumor suppressor gene in breast cancer[505,671], being an oncogene in hepatocellular[672] and colorectal[673] carcinomas. Similarly, the snoRNA host lncRNA *growth arrest specific 5* (*GAS5*) was described as a tumor suppressor in a variety of solid tumors[550,674-682]. Hence, we observed a down-regulation of the *SNORD123*-host lncRNA in colorectal cancer, suggesting its potential involvement in cancer.

As it happens to *HBII-52*, reported to regulate alternative splicing[554], we hypothesized that the epigenetic repressed *SNORD123* and *ACA59B* could contribute to tumorigenesis by an unknown mechanism not related with ribosomal and spliceosomal RNA guided-modifications. They are conserved across vertebrates and expressed at least in normal colon mucosa but they do not have a known target (*orphan* snoRNAs). By contrast, the epigenetically repressed snoRNA *U70C* was also shown to be down-regulated in CLL patients[683] and deregulated in X-linked dyskeratosis congenita cells[684], associated with cancer susceptibility. Curiously, this snoRNA guide a modification of 18S rRNA, that in turn is suggested to be involved in cancer[570-572].

Taking into account that the epigenetic transcriptional repression of tumor suppressor genes is a frequent trait of cancer, our results suggest that some snoRNAs and host genes down-regulated by DNA methylation may contribute to tumorigenesis, being concomitantly potential biomarkers for cancer diagnosis. Accordingly, some snoRNAs were shown to be downregulated in NSCLC[553], AML and ALL[685], compared to normal cells. Other studies support a possible role of snoRNAs in gene silencing, by promoting pre-mRNA degradation or inhibiting splicing and/or transport of the RNA, acting by an antisense-like mechanism in the nucleus[686,687]. Thus, our findings support the hypothesis of an important role for the snoRNAs in oncogenesis, similarly to what was described for miRNAs[533] and T-UCRs[537,688], through their classical functions in ribosome biogenesis or through so far unknown functions of the *orphan* snoRNAs.

Other ncRNAs that we found to be epigenetically regulated in malignant cells were the piRNAs. This class of ncRNA is mainly expressed in germline cells[503,573], assisting the maintenance of genomic stability and germ cell function. Accordingly, they have an important role in transposon silencing by DNA methylation, guiding also the cleavage of transposable element transcripts by PIWI proteins, protecting the genome against adverse transposon-induced insertional mutations[579,585]. Recently it was suggested that piRNAs could also induce gene-specific DNA methylation at non-transposable element genetic loci[689]. For instance, a recent study, in neurons, associated the piwi/piRNA complex with the methylation of a CGI in the promoter region of *cyclic AMP-responsive element-binding protein 2* (*CREB2*)[690].

Owing to our interest in the deregulation of the PIWI/piRNA pathway in cancer, we decided to interrogate aberrant DNA methylation events in primary testicular tumors,

due to the fact that both PIWI proteins and piRNAs are known to co-exist in testis. We hypothesized that both entities could be epigenetically silenced. However, we centered our attention in the epigenetic regulation of the PIWI-proteins involved in the biogenesis of piRNAs, since the expression levels of the last ones would be affected if their machinery of biogenesis was disrupted, independently on their methylation profile. We observed an epigenetic transcriptional repression of *piwi like RNA-mediated gene silencing 1* (*PIWIL1), PIWIL2*, *PIWIL4* and *TDRD1* by hypermethylation in primary testicular tumors (SE and NSE) and in three germ cell tumor cell lines. Curiously, the epigenetic disruption of PIWI proteins also occurs in non-genetic infertility syndromes in males[578], that have been epidemiologically associated with testicular cancer[587,588], being a common hallmark of both pathologies. We also demonstrated that the epigenetic silencing of the PIWI-protein genes involved in the biogenesis machinery of piRNAs were associated with a consistent decrease in piRNA levels and hypomethylation events at LINE-1 loci. In accordance with our study, recent data from small RNA sequencing on 22 human testicular germ cell tumor samples have confirmed a global loss of piRNAs on this type of tumors[691].

Other groups have corroborated piRNA-related proteins transcriptional repression in some cancers and several mutations were also described across different cancer types[579]. In sarcoma patients, the expression decrease of *PIWIL2* and *PIWIL4* was associated with worse prognosis[692]. Other study demonstrated the increase in cell proliferation and a decrease in the expression of tumor suppressor genes upon the knockdown of *PiwiL2* in murine bone marrow mesenchymal stem cells[693]. In NSCLC tumors, *PIWIL2* and *PIWIL4* were found to be downregulated, while *PIWIL1* was over-expressed in some tumors in comparison with normal tissue, being demonstrated that the latter could be regulated in part by DNA methylation[694]. Controversially, despite very few studies have reported the presence of piRNAs in normal or cancer somatic tissues, several other studies have reported the up-regulation of PIWI-proteins in somatic tumors[579].

The transcriptional silencing of tumor suppressor genes by DNA hypermethylation is a common and well-studied event in cancer. Although additional studies need to be performed, we suggest that both snoRNAs and piRNAs might be involved in carcinogenesis since they are inactivated by the same mechanism. Thereby, they are

potential biomarkers that should be exploited to the improvement of cancer diagnosis and personalized treatment selection.

2. DNA Methylation Mechanisms in Cancer

The CGI hypermethylation-dependent silencing of tumor suppressor genes in cancer is already very well described. Accordingly, we detected two different classes of ncRNAs repressed in cancer, prompting us to think about the mechanisms that overall govern the DNA methylation and expression profiles of a cancer cell. Moreover, in an independent study we have described the epigenetic activation of a cryptic *TBC1D16* transcript that enhance melanoma progression[453]. In a new layer of epigenetic research in which we were interested, we further hypothesized the existence of a hypomethylation-associated transcriptional activation of both coding and non-coding oncogenic genes.

One of the best clinical models illustrating the deregulation of DNA methylation is AML, where several mutations in epigenetic modifier genes such as *TET2*, *IDH1/2*, *ASXL1*, *MLL*, *EZH2* and *DNMT3A* have been described[168]. We wondered to what extent they could change the epigenetic landscape of a cancer cell. Consequently, in our first approach, we hypothesized that DNA hypomethylation events might have clinical implications by gene expression regulation of both coding and non-coding transcripts implicated in leukemogenesis. Since the *de novo* DNA methyltransferase *DNMT3A* gene harbor a missense mutation in approximately 20% of AML patients, acting as a dominant negative that inhibits wild-type DNMT3A protein and being associated with DNA methylation changes[471-473,590-593], we focused our attention on this gene. To establish cause-consequence events, between the mutational status of *DNMT3A* and downstream hypomethylated regions able to regulate leukemogenic-genes, we first depict the entire DNA methylome at single-base resolution in two AML cell lines, OCI-AML5 and OCI-AML3. These cell lines harbor the wild type and the mutated form (heterozygous R882C[600]) of DNMT3A, correspondingly, being the Arg882 (R882) site one of the known driver mutations in AML[471]. Unsurprisingly, we found 182 800 DMRs between the two cell lines, 86% of which correspondent to hypomethylated events in *DNMT3A* mutants cells. Accordingly, we observed a global decrease in the DNA methylation of this cell line (OCI-AML3). Of particular interest due to their relative location, one of these DMRs appears in the 5' intergenic locus of *MEIS1*, overlapping with the TSS of some predicted sense and antisense lncRNAs. We first

focused our attention in the possible role of this DMR governing the expression of these lncRNAs, nonetheless we were unable to validate their existence in our model. Curiously, by combining WGBS with the expression microarray data for both cell lines, coupled to the analysis of the mutational status of *DNMT3A* and the methylation profile of primary AML patients, we identified a set of twelve candidate target loci for DNMT3A in AML. The hypomethylation-associated transcriptional expression of these target genes were then confirmed in OCI-AML3 cell line, harboring the heterozygous mutation for *DNMT3A*.

Our analysis resulted in the identification of the key leukemogenic gene *MEIS1*. We suggested that this highest-ranked candidate gene is actively expressed due to the absence of a functional tetramer of DNMT3A[473] in patients harboring its mutated form. Curiously, it was reported that the expression of the transcription factor MEIS1 was strongly associated with the expression of the lncRNA *NR_033375*[695], suggesting a role of MEIS1 in the transcriptional regulation of ncRNAs that, in turn, could have important functions in the leukemogenic process. Moreover, in our preliminary studies we also described a signature of four hypomethylation-associated transcriptional reactivated lncRNAs in DNMT3A mutant cells (OCI-AML3), namely *ENST00000413346, LOC100506585, ENST00000443490* and *MIRLET7BHG* (harbors *let-7a-3* and *let-7b*). Despite complementary studies are required, the highest-ranked DMRs that we associated to the regulation of these lncRNAs, in cell lines, have a tendency to be established in AML patients harboring the *DNMT3A* wild type compared to the mutated forms. In line, recent reports have shown that *let-7a-3* and *let-7b* were highly expressed in AML cell lines[606] and that increased expression of let-7a-3 was associated with poor prognosis in AML[607].

In our AML WGBS analysis we have observed that the largest majority of DMRs do not overlap 5'-end regulatory promoters. Several other genomic regions with gene expression regulatory potential were differentially methylated between the *DNMT3A* wild type and mutant cells. According to several scientific reports, these loci could have important roles in cancer, by regulating the expression of crucial players of the disease[516-519,613-615]. Our interest in these regions led us to expand our studies to other tissues, to look for genome wide DNA methylation events in cancer. Accordingly, we performed WGBS in five normal tissues and eight associated cancer samples, supported

by DNA methylation microarray analyses of a large group of patient samples, to establish associations between super-enhancers DNA methylation status and their cancer-related activity.

In our previous discussed studies, we observed a transcriptional repressive effect of hypermethylation events in the proximal regulatory gene regions. Similarly, we established a correlation between tumor-related hypermethylation of super-enhancers and transcriptional silencing of the corresponding related genes. Curiously, in breast cancer these events occur in the super-enhancers that in normal breast epithelial cells are the most enriched in the super-enhancer-defining histone mark H3K27ac. Considering that super-enhancers are regulatory elements able to drive the expression of genes, ensuring cell and tissue identity in normal tissues[515,615], the methylation profile of super-enhancers was also associated with cancer type. However, some hypermethylated super-enhancers are shared by epithelial tumors of different origin. For instance, the super-enhancer regulating *MIRLET7BHG* is associated with the transcriptional silencing of the tumor suppressors *let-7a-3* and *let-7b*, in both lung and breast cancers.

In cancer, apart from the repression of super-enhancers that define cell identity in normal cells, super-enhancers with a *de novo* regulatory role in malignant cells were recently described[43,517,615]. In colorectal cancer, we have demonstrated that these tumor-related super-enhancers undergo DNA hypomethylation events with concomitant transcriptional activation of the corresponding regulated genes, such as *v-Myc avian myelocytomatosis viral oncogene homolog* (*MYC*) and *ring finger protein 43* (*RNF43*)[*648*] oncogenes.

In order to integrate our findings about the transcriptional gene regulation of super-enhancers in cancer, we focused our attention in the regulators of their methylation profiles, hypothesizing that transcription factors could modulate this scenario. Our study suggested that the disturbed expression and binding of transcription factors promote the establishment of novel super-enhancers, driving the expression of key players in the tumorigenic process. Moreover we have identified FOXQ1 as a putative transcription factor affecting the DNA hypomethylation at colorectal cancer-specific super-enhancers with an associated transcriptional overexpression of *MYC* and *RNF43* oncogenes.

The description of a model supporting the expression of oncogenes through these regulatory regions, prompted us to think that an elegant therapeutic approach against cancer would be the disruption of these novel cancer-specific super-enhancers. We used JQ1 to target BRD4, a key component of the secondary super-enhancer structure[517], being able to decreased the expression of some of the super-enhancer gene targets, namely *MYC*, *RNF43* and *GPRC5A*. Nevertheless we were unable to change the methylation profile of such super-enhancers, supporting the idea that transcription factors binding should drive those changes and that the disturbance of their secondary structure by itself is not decisive.

Our data in AML suggested the existence of a transcriptional reactivation of *MIRLET7BHG* by an associated DNA hypomethylation of its upstream region, derived from the presence of the mutant form of DNMT3A. Despite the TSS of *MIRLET7BHG* is located at more than 2kb (~ 4kb) downstream of the identified DMR in *DNMT3A* mutant cells, we have included this miRNA host gene in our study because of its partial genomic overlapping with the predicted lncRNA *ENST00000443490*, being possible an association between their regulation. We noticed that both transcripts were transcriptionally reactivated by hypomethylation in OCI-AML3, without knowing we were studying the super-enhancer analyzed in the following study. Through the validation of the differentially expression of both transcripts, we hypothesize that *ENST00000443490* is acting as an eRNA[696], assisting *MIRLET7BHG* expression. On the other hand, our last study indicated that *MIRLET7BHG* undergo a super-enhancer hypermethylation-associated transcriptional silencing in both breast and lung cancers. It is important to refer that there are controversial studies about the expression and function of *let-7a-3* and *let-7b* in different malignancies. For instance, one of these studies interrogated the proximal upstream and overlapping region of the *let-7a-3* pri-miRNA, associating its hypomethylation with the transcriptional activation and enhanced tumor phenotype in lung adenocarcinomas. The oncogenic function classification was derived from the overexpression of *let-7a-3* in A549 lung adenocarcinoma cell line. Nevertheless, the correlation between the DNA methylation of this locus and the transcriptional activation of the related miRNA was established using a colorectal cancer cell line. Importantly, the expression levels of let-7a-3 were not analyzed in patient samples (normal *versus* tumor) or in lung cancer cell lines with differences in terms of DNA methylation at this locus. It was also shown that the

methylation of this region, different from the genomic location that was analyzed in our study, was associated to the function of both DNMT1 and DNMT3B, but the role of DNMT3A was not analysed[608]. Controversially, another study had already described the downregulation of let-7 family of miRNAs in human lung cancer, comparing both normal and paired primary tumor samples by Northern blot analysis and quantitative real-time PCR. They described that patients with lower levels of let-7 had worse prognosis and that let-7a overexpression in A549 resulted in a reduced tumor phenotype[697].

Our results support the hypothesis of a DNA methylation-dependent transcriptional regulation of the host gene encoding both *let-7a-3* and *let-7b* that is tissue dependent; and a dual role of these players in cancer, dependent on the cellular context, promoting a worst prognosis in AMLs by their overexpression and a more aggressive tumorigenesis in lung and breast cancer through their downregulation.

CHAPTER VIII

Concluding Remarks and Future Perspectives

1. Concluding Remarks

Cancer is an *iceberg* in the form of a *guillotine* and the little that is already known is continuously changing in the cellular context, challenging the therapeutic approaches. In this PhD thesis, we have unveiled a small portion of this iceberg, giving new insights on the characterization of the epigenetic landscape that overall govern the malignancy of a cancer cell. Our findings show by the first time that two classes of ncRNAs, snoRNAs and piRNAs, can be transcriptionally repressed in cancer by DNA hypermethylation of the promoter region of their host gene or of the proteins responsible for their biogenesis, respectively.

Moreover by high-throughput methods we show that, in AML, changes in the epigenetic landscape caused by driver mutations in *DNMT3A* gene are associated with the transcriptional silencing of the key leukemogenic gene *MEIS1*. Moreover, we speculate through preliminary studies that some lncRNAs can also be disrupted by this mechanism. In a larger study comprising solid malignancies, in one hand we describe the DNA hypermethylation of tissue-specific super-enhancers and in the other, *de novo* formed super-enhancers associated to DNA hypomethylation, in malignant cells. We hypothesize that the last ones derive from a disturbed expression and binding of transcription factors, such as FOXQ1 in colorectal cancer. In our last model we also support the idea that miRNAs can be silenced through the DNA hypermethylation of super-enhancers, such as the one controlling the expression of the *let-7a-3* and *let-7b* lncRNA-host transcript.

In summary, in addition to protein-coding genes, we provide new insights into the epigenetic regulation of ncRNAs that might have a role in cancer, such as snoRNAs, piRNAs, lncRNAs and miRNAs. Importantly, ncRNAs are able to modulate the cellular epigenetic landscape or regulate the activity of other RNAs in a cellular context-dependent manner. Thus, epigenetics and ncRNAs are two challenging targets in cancer.

2. Future Perspectives

One of the biggest concerns for public health is cancer. The very old foe remains *insatiable* and continues to threaten our lives, increasingly. Despite new advances in medicine, public health campaigns, early diagnosis and more effective treatments, there

is no substantial breaking in cancer related mortality. The higher incidence is presently explained by the higher average life expectancy, exposure to uncategorized carcinogens and our lifestyle, mainly in industrialized countries. The oncologic field is demanding the discovery of new *biomarkers* to detect cancer at very early stages; the characterization of cancer-related pathways, aiming the development of related new pharmacological approaches throughout specific drug design; and the molecular characterization of tumor subtypes, seeking a more personalized treatment.

The translational application of our actual knowledge on the modulation of epigenetic mechanisms is still very limited due to the genome wide effects, using for instance, hypomethylating agents. However, ncRNAs themselves could be exploited not only as potential *biomarkers* for cancer diagnosis or treatment selection, but also as specific targets in cancer. Due to the wide range of molecular functions of ncRNAs, the development of challenging approaches to modulate their expression or function, in cancer, is starting to emerge, establishing a hope to cure or make chronic this malignant disease. Theoretically, in a sequence-based approach it would be possible the specific targeting of almost any deregulated RNA, avoiding genome wide adverse side effects[698-700]. For instance, the silencing of miR-122, required for Hepatitis C virus infection, was already approached by locked nucleic acid (LNA)-modified oligonucleotides (Miravirsen)[701], being the first miRNA-targeting drug reaching a phase II clinical trial[702]. On the other hand, miRNA replacement was approached by a double-stranded RNA mimic of the tumor suppressor miR-34, encapsulated in a liposomal nanoparticle formulation (MRX34), being currently in a phase I clinical trial[703].

Promising RNA-based therapeutic approaches are starting to emerge and would be even more promising taking into account the development of therapeutic strategies with oral bioavailability[704-706], similarly to some chronic treatments to control diabetes or hypertension.

CHAPTER IX

References

1. WHO. Cancer, Fact sheet N°297, World Health Organization. (2015).
2. Foulds, L. The experimental study of tumor progression: a review. *Cancer Res* **14**, 327-39 (1954).
3. Nowell, P.C. The clonal evolution of tumor cell populations. *Science* **194**, 23-8 (1976).
4. Mukherjee, S. *The Emperor of All Maladies: A Biography of Cancer*, (Scribner, New York, 2010).
5. Wagener, D.J.T. *The History of Oncology*, (Bohn Stafleu van Loghum, 2009).
6. Hajdu, S.I. A note from history: landmarks in history of cancer, part 1. *Cancer* **117**, 1097-102 (2011).
7. Haddow, A. Historical Notes on Cancer from the MSS. of Louis Westenra Sambon: (Section of the History of Medicine). *Proc R Soc Med* **29**, 1015-28 (1936).
8. Garrison, F.H. The history of cancer. *Bull N Y Acad Med* **2**, 179-85 (1926).
9. Hajdu, S.I. Medieval pathfinders in surgical oncology. *Cancer* **101**, 879-82 (2004).
10. Kumar, D.R., Hanlin, E., Glurich, I., Mazza, J.J. & Yale, S.H. Virchow's contribution to the understanding of thrombosis and cellular biology. *Clin Med Res* **8**, 168-72 (2010).
11. Javier, R.T. & Butel, J.S. The history of tumor virology. *Cancer Res* **68**, 7693-706 (2008).
12. Sudhakar, A. History of Cancer, Ancient and Modern Treatment Methods. *J Cancer Sci Ther* **1**, 1-4 (2009).
13. Hajdu, S.I. A note from history: landmarks in history of cancer, part 2. *Cancer* **117**, 2811-20 (2011).
14. Hajdu, S.I. A note from history: landmarks in history of cancer, part 4. *Cancer* **118**, 4914-28 (2012).
15. Hajdu, S.I. A note from history: landmarks in history of cancer, part 3. *Cancer* **118**, 1155-68 (2012).
16. Hajdu, S.I. & Darvishian, F. A note from history: landmarks in history of cancer, part 5. *Cancer* **119**, 1450-66 (2013).
17. Boveri, T. The Origin of Malignant Tumors. *Translated by Marcella Boveri, Baltimore: Williams and Wilkins Company, 1929* (1914).
18. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature* **171**, 737-8 (1953).
19. Huebner, R.J. & Todaro, G.J. Oncogenes of RNA tumor viruses as determinants of cancer. *Proc Natl Acad Sci U S A* **64**, 1087-94 (1969).
20. Murphree, A.L. & Benedict, W.F. Retinoblastoma: clues to human oncogenesis. *Science* **223**, 1028-33 (1984).
21. Soto, A.M. & Sonnenschein, C. The somatic mutation theory of cancer: growing problems with the paradigm? *Bioessays* **26**, 1097-107 (2004).
22. Hanahan, D. & Weinberg, R.A. The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
23. Ferlay J, S.I., Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 [Internet]. Lyon, France: International Agency for Research on Cancer. (2013).
24. Anand, P. et al. Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* **25**, 2097-116 (2008).
25. Felton, J.S. The heritage of Bernardino Ramazzini. *Occup Med (Lond)* **47**, 167-79 (1997).
26. Loeb, L.A., Ernster, V.L., Warner, K.E., Abbotts, J. & Laszlo, J. Smoking and lung cancer: an overview. *Cancer Res* **44**, 5940-58 (1984).
27. Loeb, L.A. Endogenous carcinogenesis: molecular oncology into the twenty-first century--presidential address. *Cancer Res* **49**, 5489-96 (1989).
28. Boffetta, P., Hashibe, M., La Vecchia, C., Zatonski, W. & Rehm, J. The burden of cancer attributable to alcohol drinking. *Int J Cancer* **119**, 884-7 (2006).
29. IARC. Alcohol consumption and ethyl carbamate. *IARC Monogr Eval Carcinog Risks Hum* **96**, 3-1383 (2010).

30. Bagnardi, V. et al. Light alcohol drinking and cancer: a meta-analysis. *Ann Oncol* **24**, 301-8 (2013).
31. Bagnardi, V., Blangiardo, M., La Vecchia, C. & Corrao, G. A meta-analysis of alcohol drinking and cancer risk. *Br J Cancer* **85**, 1700-5 (2001).
32. Bouvard, V. et al. Carcinogenicity of consumption of red and processed meat. *Lancet Oncol* **16**, 1599-600 (2015).
33. Colditz, G.A., Sellers, T.A. & Trapido, E. Epidemiology - identifying the causes and preventability of cancer? *Nat Rev Cancer* **6**, 75-83 (2006).
34. Marshall, B.J. & Warren, J.R. Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration. *Lancet* **1**, 1311-5 (1984).
35. Parsonnet, J. et al. Helicobacter pylori infection and the risk of gastric carcinoma. *N Engl J Med* **325**, 1127-31 (1991).
36. WHO, I.A.f.R.o.C. IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. (1994).
37. Ch'In, K.Y., Lei, A.T. & Wang, T.Y. Primary mucinous carcinoma of liver associated with Clonorchis sinensis infection. *Chin Med J* **73**, 26-35 (1955).
38. WHO. Cancer prevention, World Health Organization. (2016).
39. Angelos, P. et al. Contralateral Prophylactic Mastectomy: Challenging Considerations for the Surgeon. *Ann Surg Oncol* **22**, 3208-12 (2015).
40. Brandberg, Y. et al. Psychological reactions, quality of life, and body image after bilateral prophylactic mastectomy in women at high risk for breast cancer: a prospective 1-year follow-up study. *J Clin Oncol* **26**, 3943-9 (2008).
41. MedlinePlus. Radical prostatectomy, U.S. National Library of Medicine. (2015).
42. Cavallo, F., De Giovanni, C., Nanni, P., Forni, G. & Lollini, P.L. 2011: the immune hallmarks of cancer. *Cancer Immunol Immunother* **60**, 319-26 (2011).
43. Hanahan, D. & Weinberg, R.A. Hallmarks of cancer: the next generation. *Cell* **144**, 646-74 (2011).
44. Grivennikov, S.I., Greten, F.R. & Karin, M. Immunity, inflammation, and cancer. *Cell* **140**, 883-99 (2010).
45. Diaz-Cano, S.J. Tumor heterogeneity: mechanisms and bases for a reliable application of molecular marker design. *Int J Mol Sci* **13**, 1951-2011 (2012).
46. Kondo, Y. et al. Gene silencing in cancer by histone H3 lysine 27 trimethylation independent of promoter DNA methylation. *Nat Genet* **40**, 741-50 (2008).
47. Kondo, Y. Epigenetic cross-talk between DNA methylation and histone modifications in human cancers. *Yonsei Med J* **50**, 455-63 (2009).
48. Berdasco, M. & Esteller, M. Aberrant epigenetic landscape in cancer: how cellular identity goes awry. *Dev Cell* **19**, 698-711 (2010).
49. Jones, P.A. & Baylin, S.B. The epigenomics of cancer. *Cell* **128**, 683-92 (2007).
50. Esteller, M. Cancer epigenomics: DNA methylomes and histone-modification maps. *Nat Rev Genet* **8**, 286-98 (2007).
51. Croce, C.M. Oncogenes and cancer. *N Engl J Med* **358**, 502-11 (2008).
52. Myllykangas, S. et al. DNA copy number amplification profiling of human neoplasms. *Oncogene* **25**, 7324-32 (2006).
53. ACS, A.C.S. Colon/Rectum Cancer-Detailed Guide-Colorectal cancer risk factors. (2016).
54. Kastrinos, F. & Syngal, S. Inherited colorectal cancer syndromes. *Cancer J* **17**, 405-15 (2011).
55. Ellis, C.N. Inherited colorectal cancer syndromes. *Clin Colon Rectal Surg* **18**, 150-62 (2005).
56. Bolocan, A., Ion, D., Stoian, R.V. & Serban, M.B. Map syndrome (MYH Associated Polyposis) colorectal cancer, etiopathological connections. *J Med Life* **4**, 109-11 (2011).
57. Fearon, E.R. & Vogelstein, B. A genetic model for colorectal tumorigenesis. *Cell* **61**, 759-67 (1990).

58. Manne, U., Shanmugam, C., Katkoori, V.R., Bumpers, H.L. & Grizzle, W.E. Development and progression of colorectal neoplasia. *Cancer Biomark* **9**, 235-65 (2010).
59. Lin, O.S. Acquired risk factors for colorectal cancer. *Methods Mol Biol* **472**, 361-72 (2009).
60. Krause, W.F. & DuBois, R.N. The molecular basis for prevention of colorectal cancer. *Clin Colorectal Cancer* **1**, 47-54 (2001).
61. Davies, R.J., Miller, R. & Coleman, N. Colorectal cancer screening: prospects for molecular stool analysis. *Nat Rev Cancer* **5**, 199-209 (2005).
62. Sharma, S., Kelly, T.K. & Jones, P.A. Epigenetics in cancer. *Carcinogenesis* **31**, 27-36 (2010).
63. Esteller, M. Epigenetics in cancer. *N Engl J Med* **358**, 1148-59 (2008).
64. Bonasio, R., Tu, S. & Reinberg, D. Molecular signals of epigenetic states. *Science* **330**, 612-6 (2010).
65. Markowitz, S.D. & Bertagnolli, M.M. Molecular origins of cancer: Molecular basis of colorectal cancer. *N Engl J Med* **361**, 2449-60 (2009).
66. Pancione, M., Remo, A. & Colantuoni, V. Genetic and epigenetic events generate multiple pathways in colorectal cancer progression. *Patholog Res Int* **2012**, 509348 (2012).
67. Kinzler, K.W. & Vogelstein, B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* **386**, 761, 763 (1997).
68. Vogelstein, B. et al. Genetic alterations during colorectal-tumor development. *N Engl J Med* **319**, 525-32 (1988).
69. Miyaki, M. et al. Characteristics of somatic mutation of the adenomatous polyposis coli gene in colorectal tumors. *Cancer Res* **54**, 3011-20 (1994).
70. Fearon, E.R. et al. Identification of a chromosome 18q gene that is altered in colorectal cancers. *Science* **247**, 49-56 (1990).
71. Pino, M.S. & Chung, D.C. The chromosomal instability pathway in colon cancer. *Gastroenterology* **138**, 2059-72 (2010).
72. Howlader N, N.A., Krapcho M, Miller D, Bishop K, Altekruse SF, Kosary CL, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA (eds). SEER Cancer Statistics Review, 1975-2013, National Cancer Institute. Bethesda, MD, http://seer.cancer.gov/csr/1975_2013/. *based on November 2015 SEER data submission, posted to the SEER web site* (April 2016).
73. Stein, A. & Schmoll, H.J. Systemic treatment of liver metastases from colorectal cancer. *Ther Adv Med Oncol* **5**, 193-203 (2013).
74. Sheth, K.R. & Clary, B.M. Management of hepatic metastases from colorectal cancer. *Clin Colon Rectal Surg* **18**, 215-23 (2005).
75. Kanas, G.P. et al. Survival after liver resection in metastatic colorectal cancer: review and meta-analysis of prognostic factors. *Clin Epidemiol* **4**, 283-301 (2012).
76. Borasio, P. et al. Role of surgical resection in colorectal lung metastases: analysis of 137 patients. *Int J Colorectal Dis* **26**, 183-90 (2011).
77. Hammoud, M.A., McCutcheon, I.E., Elsouki, R., Schoppa, D. & Patt, Y.Z. Colorectal carcinoma and brain metastasis: distribution, treatment, and survival. *Ann Surg Oncol* **3**, 453-63 (1996).
78. Katoh, M., Unakami, M., Hara, M. & Fukuchi, S. Bone metastasis from colorectal cancer in autopsy cases. *J Gastroenterol* **30**, 615-8 (1995).
79. Aoyagi, T., Terracina, K.P., Raza, A. & Takabe, K. Current treatment options for colon cancer peritoneal carcinomatosis. *World J Gastroenterol* **20**, 12493-500 (2014).
80. Erroi, F. et al. Ovarian metastasis from colorectal cancer: prognostic value of radical oophorectomy. *J Surg Oncol* **96**, 113-7 (2007).
81. Danesi, R., Bocci, G. & Di Paolo, A. Biologic basis of ovarian metastasis of colorectal cancer. *Clin Colorectal Cancer* **3**, 223-4 (2004).

82. Welter, S., Jacobs, J., Krbek, T., Poettgen, C. & Stamatis, G. Prognostic impact of lymph node involvement in pulmonary metastases from colorectal cancer. *Eur J Cardiothorac Surg* **31**, 167-72 (2007).
83. Fujii, T. et al. Process of distant lymph node metastasis in colorectal carcinoma: implication of extracapsular invasion of lymph node metastasis. *BMC Cancer* **11**, 216 (2011).
84. Viatori, M. Testicular cancer. *Semin Oncol Nurs* **28**, 180-9 (2012).
85. Gilbert, D., Rapley, E. & Shipley, J. Testicular germ cell tumours: predisposition genes and the male germ cell niche. *Nat Rev Cancer* **11**, 278-88 (2011).
86. Winter, C. & Albers, P. Testicular germ cell tumors: pathogenesis, diagnosis and treatment. *Nat Rev Endocrinol* **7**, 43-53 (2011).
87. Hayes-Lattin, B. & Nichols, C.R. Testicular cancer: a prototypic tumor of young adults. *Semin Oncol* **36**, 432-8 (2009).
88. Kollmannsberger C, D.S., Hansen E, Corless CL, Roth BJ, Nichols CR. Testis Cancer. In: Cancer Medicine, 8th Edition, Kufe DW, Bast RC, Hait WN, Hong WK, Pollock RE, Weichselbaum RR,
Holland JF, Frei E (eds.). *B. C. Decker, Inc; Hamilton, pp. 1263-1289,* (2010).
89. Huyghe, E., Matsuda, T. & Thonneau, P. Increasing incidence of testicular cancer worldwide: a review. *J Urol* **170**, 5-11 (2003).
90. Dieckmann, K.P. & Pichlmeier, U. Clinical epidemiology of testicular germ cell tumors. *World J Urol* **22**, 2-14 (2004).
91. Mikuz, G. [Testicular cancer - a matter of geography? Epidemiology and etiopathogenesis of germ cell tumors]. *Pathologe* **35**, 211-7 (2014).
92. Depue, R.H., Pike, M.C. & Henderson, B.E. Estrogen exposure during gestation and risk of testicular cancer. *J Natl Cancer Inst* **71**, 1151-5 (1983).
93. Weir, H.K., Marrett, L.D., Kreiger, N., Darlington, G.A. & Sugar, L. Pre-natal and peri-natal exposures and risk of testicular germ-cell cancer. *Int J Cancer* **87**, 438-43 (2000).
94. Henderson, B.E., Bernstein, L., Ross, R.K., Depue, R.H. & Judd, H.L. The early in utero oestrogen and testosterone environment of blacks and whites: potential effects on male offspring. *Br J Cancer* **57**, 216-8 (1988).
95. Strader, C.H., Weiss, N.S., Daling, J.R., Karagas, M.R. & McKnight, B. Cryptorchism, orchiopexy, and the risk of testicular cancer. *Am J Epidemiol* **127**, 1013-8 (1988).
96. Dieckmann, K.P. & Loy, V. Prevalence of contralateral testicular intraepithelial neoplasia in patients with testicular germ cell neoplasms. *J Clin Oncol* **14**, 3126-32 (1996).
97. Hemminki, K. & Li, X. Familial risk in testicular cancer as a clue to a heritable and environmental aetiology. *Br J Cancer* **90**, 1765-70 (2004).
98. Skakkebaek, N.E. Testicular dysgenesis syndrome. *Horm Res* **60 Suppl 3**, 49 (2003).
99. Jacobsen, R. et al. Risk of testicular cancer in men with abnormal semen characteristics: cohort study. *Bmj* **321**, 789-92 (2000).
100. Aguirre, D. et al. Extragonadal germ cell tumors are often associated with Klinefelter syndrome. *Hum Pathol* **37**, 477-80 (2006).
101. Queipo, G. et al. Intracranial germ cell tumors: association with Klinefelter syndrome and sex chromosome aneuploidies. *Cytogenet Genome Res* **121**, 211-4 (2008).
102. Holl, K. et al. Maternal Epstein-Barr virus and cytomegalovirus infections and risk of testicular cancer in the offspring: a nested case-control study. *Apmis* **116**, 816-22 (2008).
103. Trabert, B. et al. Baldness, acne and testicular germ cell tumours. *Int J Androl* **34**, e59-67 (2011).
104. Meeks, J.J., Sheinfeld, J. & Eggener, S.E. Environmental toxicology of testicular cancer. *Urol Oncol* **30**, 212-5 (2012).
105. Oosterhuis, J.W. & Looijenga, L.H. Testicular germ-cell tumours in a broader perspective. *Nat Rev Cancer* **5**, 210-22 (2005).
106. Rajpert-De Meyts, E., McGlynn, K.A., Okamoto, K., Jewett, M.A. & Bokemeyer, C. Testicular germ cell tumours. *Lancet* **387**, 1762-1774 (2016).

107. Oosterhuis, J.W. & Looijenga, L.H. Current views on the pathogenesis of testicular germ cell tumours and perspectives for future research: highlights of the 5th Copenhagen Workshop on Carcinoma in situ and Cancer of the Testis. *Apmis* **111**, 280-9 (2003).
108. Bredael, J.J., Vugrin, D. & Whitmore, W.F., Jr. Autopsy findings in 154 patients with germ cell tumors of the testis. *Cancer* **50**, 548-51 (1982).
109. Katsube, Y., Berg, J.W. & Silverberg, S.G. Epidemiologic pathology of ovarian tumors: a histopathologic review of primary ovarian neoplasms diagnosed in the Denver Standard Metropolitan Statistical Area, 1 July-31 December 1969 and 1 July-31 December 1979. *Int J Gynecol Pathol* **1**, 3-16 (1982).
110. Krag Jacobsen, G. et al. Testicular germ cell tumours in Denmark 1976-1980. Pathology of 1058 consecutive cases. *Acta Radiol Oncol* **23**, 239-47 (1984).
111. Ulbright, T.M. Germ cell tumors of the gonads: a selective review emphasizing problems in differential diagnosis, newly appreciated, and controversial issues. *Mod Pathol* **18 Suppl 2**, S61-79 (2005).
112. O'Callaghan, A. & Mead, G.M. Testicular carcinoma. *Postgrad Med J* **73**, 481-6 (1997).
113. Zafarana, G. et al. 12p-amplicon structure analysis in testicular germ cell tumors of adolescents and adults by array CGH. *Oncogene* **22**, 7695-701 (2003).
114. Skakkebaek, N.E. et al. Germ cell cancer and disorders of spermatogenesis: an environmental connection? *Apmis* **106**, 3-11; discussion 12 (1998).
115. Gobel, U. et al. Germ-cell tumors in childhood and adolescence. GPOH MAKEI and the MAHO study groups. *Ann Oncol* **11**, 263-71 (2000).
116. Gonzalez-Crussi, F., Winkler, R.F. & Mirkin, D.L. Sacrococcygeal teratomas in infants and children: relationship of histology and prognosis in 40 cases. *Arch Pathol Lab Med* **102**, 420-5 (1978).
117. Bahrami, A., Ro, J.Y. & Ayala, A.G. An overview of testicular germ cell tumors. *Arch Pathol Lab Med* **131**, 1267-80 (2007).
118. Trabert, B., Stang, A., Cook, M.B., Rusner, C. & McGlynn, K.A. Impact of classification of mixed germ-cell tumours on incidence trends of non-seminoma. *Int J Androl* **34**, e274-7 (2011).
119. Aschim, E.L., Haugen, T.B., Tretli, S., Daltveit, A.K. & Grotmol, T. Risk factors for testicular cancer--differences between pure non-seminoma and mixed seminoma/non-seminoma? *Int J Androl* **29**, 458-67 (2006).
120. Mannuel, H.D., Mitikiri, N., Khan, M. & Hussain, A. Testicular germ cell tumors: biology and clinical update. *Curr Opin Oncol* **24**, 266-71 (2012).
121. Sandberg, A.A., Meloni, A.M. & Suijkerbuijk, R.F. Reviews of chromosome studies in urological tumors. III. Cytogenetics and genes in testicular tumors. *J Urol* **155**, 1531-56 (1996).
122. Skotheim, R.I. et al. Novel genomic aberrations in testicular germ cell tumors by array-CGH, and associated gene expression changes. *Cell Oncol* **28**, 315-26 (2006).
123. Rodriguez, S. et al. Expression profile of genes from 12p in testicular germ cell tumors of adolescents and adults associated with i(12p) and amplification at 12p11.2-p12.1. *Oncogene* **22**, 1880-91 (2003).
124. Dmitrovsky, E. et al. Isochromosome 12p in non-seminoma cell lines: karyologic amplification of c-ki-ras2 without point-mutational activation. *Oncogene* **5**, 543-8 (1990).
125. Nathanson, K.L. et al. The Y deletion gr/gr and susceptibility to testicular germ cell tumor. *Am J Hum Genet* **77**, 1034-43 (2005).
126. Repping, S. et al. Polymorphism for a 1.6-Mb deletion of the human Y chromosome persists through balance between recurrent mutation and haploid selection. *Nat Genet* **35**, 247-51 (2003).
127. Turnbull, C. & Rahman, N. Genome-wide association studies provide new insights into the genetic basis of testicular germ-cell tumour. *Int J Androl* **34**, e86-96; discussion e96-7 (2011).

128. Biermann, K. et al. c-KIT is frequently mutated in bilateral germ cell tumours and down-regulated during progression from intratubular germ cell neoplasia to seminoma. *J Pathol* **213**, 311-8 (2007).
129. Looijenga, L.H. et al. Stem cell factor receptor (c-KIT) codon 816 mutations predict development of bilateral testicular germ-cell tumors. *Cancer Res* **63**, 7674-8 (2003).
130. McIntyre, A. et al. Activating mutations and/or expression levels of tyrosine kinase receptors GRB7, RAS, and BRAF in testicular germ cell tumors. *Neoplasia* **7**, 1047-52 (2005).
131. Rapley, E.A. et al. Localization to Xq27 of a susceptibility gene for testicular germ-cell tumours. *Nat Genet* **24**, 197-200 (2000).
132. Ling, H., Krassnig, L., Bullock, M.D. & Pichler, M. MicroRNAs in Testicular Cancer Diagnosis and Prognosis. *Urol Clin North Am* **43**, 127-34 (2016).
133. Zhou, A.D. et al. beta-Catenin/LEF1 transactivates the microRNA-371-373 cluster that modulates the Wnt/beta-catenin-signaling pathway. *Oncogene* **31**, 2968-78 (2012).
134. Gillis, A.J. et al. Targeted serum miRNA (TSmiR) test for diagnosis and follow-up of (testicular) germ cell cancer patients: a proof of principle. *Mol Oncol* **7**, 1083-92 (2013).
135. Dieckmann, K.P. et al. MicroRNAs miR-371-3 in serum as diagnostic tools in the management of testicular germ cell tumours. *Br J Cancer* **107**, 1754-60 (2012).
136. Tan, B.T., Park, C.Y., Ailles, L.E. & Weissman, I.L. The cancer stem cell hypothesis: a work in progress. *Lab Invest* **86**, 1203-7 (2006).
137. Collins, D.H. & Rose, W.M. The nature of anaemia in leukaemia. *J Pathol Bacteriol* **60**, 63-74 (1948).
138. Ross, J.F., Crockett, C.L., Jr., Finch, S.C. & Emerson, C.P. The mechanism of anemia in leukemia and malignant lymphoma. *Bmq* **5**, 8-10 (1954).
139. Hotchkiss, D.J., Jr. The Anemia Of Lymphomas And Leukemias. *Med Clin North Am* **48**, 1479-92 (1964).
140. Aderka, D., Praff, G., Santo, M., Weinberger, A. & Pinkhas, J. Bleeding due to thrombocytopenia in acute leukemias and reevaluation of the prophylactic platelet transfusion policy. *Am J Med Sci* **291**, 147-51 (1986).
141. Psaila, B. et al. Differences in platelet function in patients with acute myeloid leukemia and myelodysplasia compared to equally thrombocytopenic patients with immune thrombocytopenia. *J Thromb Haemost* **9**, 2302-10 (2011).
142. Naresh, K.N., Sivasankaran, P. & Veliath, A.J. Platelet function in acute leukemias. *J Assoc Physicians India* **41**, 377-8 (1993).
143. Carrillo, J.M., Jimenez, E. & Jimenez, R. [Infections in the child with acute leukemia]. *Bol Med Hosp Infant Mex* **38**, 313-22 (1981).
144. DeGregorio, M.W., Lee, W.M., Linker, C.A., Jacobs, R.A. & Ries, C.A. Fungal infections in patients with acute leukemia. *Am J Med* **73**, 543-8 (1982).
145. Flynn, P.M., Marina, N.M., Rivera, G.K. & Hughes, W.T. Candida tropicalis infections in children with leukemia. *Leuk Lymphoma* **10**, 369-76 (1993).
146. Sung, L., Buxton, A., Gamis, A., Woods, W.G. & Alonzo, T.A. Life-threatening and fatal infections in children with acute myeloid leukemia: a report from the Children's Oncology Group. *J Pediatr Hematol Oncol* **34**, e30-5 (2012).
147. Hersh, E.M., Bodey, G.P., Nies, B.A. & Freireich, E.J. Causes Of Death In Acute Leukemia: A Ten-Year Study Of 414 Patients From 1954-1963. *Jama* **193**, 105-9 (1965).
148. Chang, H.Y. et al. Causes of death in adults with acute leukemia. *Medicine (Baltimore)* **55**, 259-68 (1976).
149. Davis, A.S., Viera, A.J. & Mead, M.D. Leukemia: an overview for primary care. *Am Fam Physician* **89**, 731-8 (2014).
150. Smith, M. et al. Adult acute myeloid leukaemia. *Crit Rev Oncol Hematol* **50**, 197-222 (2004).
151. Lichtman, M.A. Obesity and the risk for a hematological malignancy: leukemia, lymphoma, or myeloma. *Oncologist* **15**, 1083-101 (2010).

152. Finn, L. et al. Epidemiology of adult acute myeloid leukemia: Impact of exposures on clinical phenotypes and outcomes after therapy. *Cancer Epidemiol* **39**, 1084-92 (2015).
153. ACS, A.C.S. Leukemia-Acute Myeloid (Myelogenous). (2016).
154. Lane, S.W., Scadden, D.T. & Gilliland, D.G. The leukemic stem cell niche: current concepts and therapeutic opportunities. *Blood* **114**, 1150-7 (2009).
155. Porwit, A. & Vardiman, J.W. Acute myeloid leukemia with expanded erythropoiesis. *Haematologica* **96**, 1241-3 (2011).
156. WHO. Acute Myelogenous Leukemia And Acute Promyelocytic Leukemia, Union for International Cancer Control, World Health Organization. *Review of Cancer Medicines on the WHO List of Essential Medicines* (2014).
157. Bennett, J.M. et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol* **33**, 451-8 (1976).
158. Vardiman, J.W. et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* **114**, 937-51 (2009).
159. Mrozek, K., Heinonen, K. & Bloomfield, C.D. Prognostic value of cytogenetic findings in adults with acute myeloid leukemia. *Int J Hematol* **72**, 261-71 (2000).
160. Redner, R.L. Variations on a theme: the alternate translocations in APL. *Leukemia* **16**, 1927-32 (2002).
161. Zhou, G.B., Zhao, W.L., Wang, Z.Y., Chen, S.J. & Chen, Z. Retinoic acid and arsenic for treating acute promyelocytic leukemia. *PLoS Med* **2**, e12 (2005).
162. Martens, J.H. & Stunnenberg, H.G. The molecular signature of oncofusion proteins in acute myeloid leukemia. *FEBS Lett* **584**, 2662-9 (2010).
163. Slovak, M.L. et al. Karyotypic analysis predicts outcome of preremission and postremission therapy in adult acute myeloid leukemia: a Southwest Oncology Group/Eastern Cooperative Oncology Group Study. *Blood* **96**, 4075-83 (2000).
164. Grimwade, D. et al. The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. *Blood* **92**, 2322-33 (1998).
165. Byrd, J.C. et al. Pretreatment cytogenetic abnormalities are predictive of induction success, cumulative incidence of relapse, and overall survival in adult patients with de novo acute myeloid leukemia: results from Cancer and Leukemia Group B (CALGB 8461). *Blood* **100**, 4325-36 (2002).
166. Mrozek, K., Heinonen, K. & Bloomfield, C.D. Clinical importance of cytogenetics in acute myeloid leukaemia. *Best Pract Res Clin Haematol* **14**, 19-47 (2001).
167. Dohner, H. et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood* **115**, 453-74 (2010).
168. Hou, H.A. & Tien, H.F. Mutations in epigenetic modifiers in acute myeloid leukemia and their clinical utility. *Expert Rev Hematol* **9**, 447-69 (2016).
169. TCGA. The Cancer Genome Atlas Research Network, Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* **368**, 2059-74 (2013).
170. Waddington, C.H. The epigenotype. 1942. *Int J Epidemiol* **41**, 10-3 (2012).
171. Waddington, C.H. The strategy of the genes: A discussion of some aspects of theoretical Biology. (1957).
172. Barros, S.P. & Offenbacher, S. Epigenetics: connecting environment and genotype to phenotype and disease. *J Dent Res* **88**, 400-8 (2009).
173. Choi, S.W. & Friso, S. Epigenetics: A New Bridge between Nutrition and Health. *Adv Nutr* **1**, 8-16 (2010).
174. Cooney, C.A., Dave, A.A. & Wolff, G.L. Maternal methyl supplements in mice affect epigenetic variation and DNA methylation of offspring. *J Nutr* **132**, 2393S-2400S (2002).
175. Miller, G. Epigenetics. The seductive allure of behavioral epigenetics. *Science* **329**, 24-7 (2010).

176. Zeilinger, S. et al. Tobacco smoking leads to extensive genome-wide changes in DNA methylation. *PLoS One* **8**, e63812 (2013).
177. Fraga, M.F. et al. Epigenetic differences arise during the lifetime of monozygotic twins. *Proc Natl Acad Sci U S A* **102**, 10604-9 (2005).
178. Heyn, H. et al. DNA methylation profiling in breast cancer discordant identical twins identifies DOK7 as novel epigenetic biomarker. *Carcinogenesis* **34**, 102-8 (2013).
179. Heyn, H. et al. Distinct DNA methylomes of newborns and centenarians. *Proc Natl Acad Sci U S A* **109**, 10522-7 (2012).
180. Palini, S., De Stefani, S., Scala, V., Dusi, L. & Bulletti, C. Epigenetic regulatory mechanisms during preimplantation embryo development. *Ann N Y Acad Sci* **1221**, 54-60 (2011).
181. Arney, K.L. & Fisher, A.G. Epigenetic aspects of differentiation. *J Cell Sci* **117**, 4355-63 (2004).
182. Inbar-Feigenberg, M., Choufani, S., Butcher, D.T., Roifman, M. & Weksberg, R. Basic concepts of epigenetics. *Fertil Steril* **99**, 607-15 (2013).
183. Elango, N. & Yi, S.V. DNA methylation and structural and functional bimodality of vertebrate promoters. *Mol Biol Evol* **25**, 1602-8 (2008).
184. Schwartz, S., Meshorer, E. & Ast, G. Chromatin organization marks exon-intron structure. *Nat Struct Mol Biol* **16**, 990-5 (2009).
185. Blasco, M.A. The epigenetic regulation of mammalian telomeres. *Nat Rev Genet* **8**, 299-309 (2007).
186. Collins, K. & Mitchell, J.R. Telomerase in the human organism. *Oncogene* **21**, 564-79 (2002).
187. Flores, I., Cayuela, M.L. & Blasco, M.A. Effects of telomerase and telomere length on epidermal stem cell behavior. *Science* **309**, 1253-6 (2005).
188. Choi, J.D. & Lee, J.S. Interplay between Epigenetics and Genetics in Cancer. *Genomics Inform* **11**, 164-73 (2013).
189. Holohan, B. et al. Decreasing initial telomere length in humans intergenerationally understates age-associated telomere shortening. *Aging Cell* **14**, 669-77 (2015).
190. Nagel, I. et al. Deregulation of the telomerase reverse transcriptase (TERT) gene by chromosomal translocations in B-cell malignancies. *Blood* **116**, 1317-20 (2010).
191. Peifer, M. et al. Telomerase activation by genomic rearrangements in high-risk neuroblastoma. *Nature* **526**, 700-4 (2015).
192. Hollister, J.D. & Gaut, B.S. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**, 1419-28 (2009).
193. Le, T.N., Miyazaki, Y., Takuno, S. & Saze, H. Epigenetic regulation of intragenic transposable elements impacts gene transcription in Arabidopsis thaliana. *Nucleic Acids Res* **43**, 3911-21 (2015).
194. Rodriguez-Paredes, M. & Esteller, M. Cancer epigenetics reaches mainstream oncology. *Nat Med* **17**, 330-9 (2011).
195. Shalginskikh, N., Poleshko, A., Skalka, A.M. & Katz, R.A. Retroviral DNA methylation and epigenetic repression are mediated by the antiviral host protein Daxx. *J Virol* **87**, 2137-50 (2013).
196. Lee, Y.C. The Role of piRNA-Mediated Epigenetic Silencing in the Population Dynamics of Transposable Elements in Drosophila melanogaster. *PLoS Genet* **11**, e1005269 (2015).
197. Collick, A., Reik, W., Barton, S.C. & Surani, A.H. CpG methylation of an X-linked transgene is determined by somatic events postfertilization and not germline imprinting. *Development* **104**, 235-44 (1988).
198. Reik, W., Collick, A., Norris, M.L., Barton, S.C. & Surani, M.A. Genomic imprinting determines methylation of parental alleles in transgenic mice. *Nature* **328**, 248-51 (1987).
199. Lyon, M.F. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* **14**, 135-48 (1962).

200. Wutz, A. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. *Nat Rev Genet* **12**, 542-53 (2011).
201. Qiu, J. Epigenetics: unfinished symphony. *Nature* **441**, 143-5 (2006).
202. Ahmed, F. Epigenetics: Tales of adversity. *Nature* **468**, S20 (2010).
203. Painter, R.C. et al. Transgenerational effects of prenatal exposure to the Dutch famine on neonatal adiposity and health in later life. *Bjog* **115**, 1243-9 (2008).
204. Roseboom, T., de Rooij, S. & Painter, R. The Dutch famine and its long-term consequences for adult health. *Early Hum Dev* **82**, 485-91 (2006).
205. Heijmans, B.T. et al. Persistent epigenetic differences associated with prenatal exposure to famine in humans. *Proc Natl Acad Sci U S A* **105**, 17046-9 (2008).
206. Burggren, W.W. Dynamics of epigenetic phenomena: intergenerational and intragenerational phenotype 'washout'. *J Exp Biol* **218**, 80-7 (2015).
207. Portela, A. & Esteller, M. Epigenetic modifications and human disease. *Nat Biotechnol* **28**, 1057-68 (2010).
208. Schones, D.E. et al. Dynamic regulation of nucleosome positioning in the human genome. *Cell* **132**, 887-98 (2008).
209. Cairns, B.R. The logic of chromatin architecture and remodelling at promoters. *Nature* **461**, 193-8 (2009).
210. Han, M. & Grunstein, M. Nucleosome loss activates yeast downstream promoters in vivo. *Cell* **55**, 1137-45 (1988).
211. Chodavarapu, R.K. et al. Relationship between nucleosome positioning and DNA methylation. *Nature* **466**, 388-92 (2010).
212. Kim, D.H., Saetrom, P., Snove, O., Jr. & Rossi, J.J. MicroRNA-directed transcriptional gene silencing in mammalian cells. *Proc Natl Acad Sci U S A* **105**, 16230-5 (2008).
213. Szenthe, K. et al. The 5' regulatory sequences of active miR-146a promoters are hypomethylated and associated with euchromatic histone modification marks in B lymphoid cells. *Biochem Biophys Res Commun* **433**, 489-95 (2013).
214. Matouk, C.C. & Marsden, P.A. Epigenetic regulation of vascular endothelial gene expression. *Circ Res* **102**, 873-87 (2008).
215. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nat Genet* **33 Suppl**, 245-54 (2003).
216. Bohmdorfer, G. & Wierzbicki, A.T. Control of Chromatin Structure by Long Noncoding RNA. *Trends Cell Biol* **25**, 623-32 (2015).
217. Magistri, M., Faghihi, M.A., St Laurent, G., 3rd & Wahlestedt, C. Regulation of chromatin structure by long noncoding RNAs: focus on natural antisense transcripts. *Trends Genet* **28**, 389-96 (2012).
218. Mattill, H.A. The Protamines and Histones (Kossel, Albrecht). *Journal of Chemical Education 1928 5 (12), 1709* (1928).
219. Kornberg, R.D. Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-71 (1974).
220. Kornberg, R.D. & Thomas, J.O. Chromatin structure; oligomers of the histones. *Science* **184**, 865-8 (1974).
221. Azad, G.K. & Tomar, R.S. Proteolytic clipping of histone tails: the emerging role of histone proteases in regulation of various biological processes. *Mol Biol Rep* **41**, 2717-30 (2014).
222. Dawson, M.A. & Kouzarides, T. Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).
223. Kouzarides, T. SnapShot: Histone-modifying enzymes. *Cell* **128**, 802 (2007).
224. Daujat, S., Zeissler, U., Waldmann, T., Happel, N. & Schneider, R. HP1 binds specifically to Lys26-methylated histone H1.4, whereas simultaneous Ser27 phosphorylation blocks HP1 binding. *J Biol Chem* **280**, 38090-5 (2005).
225. Rando, O.J. & Chang, H.Y. Genome-wide views of chromatin structure. *Annu Rev Biochem* **78**, 245-71 (2009).
226. Gaffney, D.J. et al. Controls of nucleosome positioning in the human genome. *PLoS Genet* **8**, e1003036 (2012).

227. Li, B., Carey, M. & Workman, J.L. The role of chromatin during transcription. *Cell* **128**, 707-19 (2007).
228. Bai, L. & Morozov, A.V. Gene regulation by nucleosome positioning. *Trends Genet* **26**, 476-83 (2010).
229. Jiang, C. & Pugh, B.F. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* **10**, 161-72 (2009).
230. Lee, J.S., Smith, E. & Shilatifard, A. The language of histone crosstalk. *Cell* **142**, 682-5 (2010).
231. Simo-Riudalbas, L. & Esteller, M. Targeting the histone orthography of cancer: drugs for writers, erasers and readers. *Br J Pharmacol* **172**, 2716-32 (2015).
232. Kouzarides, T. Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
233. Kulis, M. & Esteller, M. DNA methylation and cancer. *Adv Genet* **70**, 27-56 (2010).
234. Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**, 9-25 (1975).
235. Holliday, R. & Pugh, J.E. DNA modification mechanisms and gene activity during development. *Science* **187**, 226-32 (1975).
236. Suzuki, H., Maruyama, R., Yamamoto, E. & Kai, M. DNA methylation and microRNA dysregulation in cancer. *Mol Oncol* **6**, 567-78 (2012).
237. Lee, H.S. & Chen, Z.J. Protein-coding genes are epigenetically regulated in Arabidopsis polyploids. *Proc Natl Acad Sci U S A* **98**, 6753-8 (2001).
238. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev* **16**, 6-21 (2002).
239. Illingworth, R.S. & Bird, A.P. CpG islands--'a rough guide'. *FEBS Lett* **583**, 1713-20 (2009).
240. Weber, M. & Schubeler, D. Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* **19**, 273-80 (2007).
241. Wu, H., Caffo, B., Jaffee, H.A., Irizarry, R.A. & Feinberg, A.P. Redefining CpG islands using hidden Markov models. *Biostatistics* **11**, 499-514 (2010).
242. Saxonov, S., Berg, P. & Brutlag, D.L. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci U S A* **103**, 1412-7 (2006).
243. Ioshikhes, I.P. & Zhang, M.Q. Large-scale human promoter mapping using CpG islands. *Nat Genet* **26**, 61-3 (2000).
244. Bestor, T.H., Edwards, J.R. & Boulard, M. Notes on the role of dynamic DNA methylation in mammalian development. *Proc Natl Acad Sci U S A* **112**, 6796-9.
245. Kitamura, E. et al. Analysis of tissue-specific differentially methylated regions (TDMs) in humans. *Genomics* **89**, 326-37 (2007).
246. Venter, J.C. et al. The sequence of the human genome. *Science* **291**, 1304-51 (2001).
247. Vinson, C. & Chatterjee, R. CG methylation. *Epigenomics* **4**, 655-63 (2012).
248. Ball, M.P. et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. *Nat Biotechnol* **27**, 361-8 (2009).
249. Vavouri, T. & Lehner, B. Human genes with CpG island promoters have a distinct transcription-associated chromatin organization. *Genome Biol* **13**, R110 (2012).
250. Ji, H. et al. Comprehensive methylome map of lineage commitment from haematopoietic progenitors. *Nature* **467**, 338-42 (2010).
251. Doi, A. et al. Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**, 1350-3 (2009).
252. Irizarry, R.A. et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet* **41**, 178-86 (2009).
253. Straussman, R. et al. Developmental programming of CpG island methylation profiles in the human genome. *Nat Struct Mol Biol* **16**, 564-71 (2009).
254. Paulsen, M. & Ferguson-Smith, A.C. DNA methylation in genomic imprinting, development, and disease. *J Pathol* **195**, 97-110 (2001).

255. Walsh, C.P., Chaillet, J.R. & Bestor, T.H. Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat Genet* **20**, 116-7 (1998).
256. Gilson, E. & Horard, B. Comprehensive DNA methylation profiling of human repetitive DNA elements using an MeDIP-on-RepArray assay. *Methods Mol Biol* **859**, 267-91 (2012).
257. Liang, G. et al. Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol Cell Biol* **22**, 480-91 (2002).
258. Slotkin, R.K. & Martienssen, R. Transposable elements and the epigenetic regulation of the genome. *Nat Rev Genet* **8**, 272-85 (2007).
259. Xie, M. et al. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* **45**, 836-41 (2013).
260. Barton, S.C., Surani, M.A. & Norris, M.L. Role of paternal and maternal genomes in mouse development. *Nature* **311**, 374-6 (1984).
261. Kacem, S. & Feil, R. Chromatin mechanisms in genomic imprinting. *Mamm Genome* **20**, 544-56 (2009).
262. Reik, W. & Lewis, A. Co-evolution of X-chromosome inactivation and imprinting in mammals. *Nat Rev Genet* **6**, 403-10 (2005).
263. Gama-Sosa, M.A. et al. Tissue-specific differences in DNA methylation in various mammals. *Biochim Biophys Acta* **740**, 212-9 (1983).
264. Song, F. et al. Association of tissue-specific differentially methylated regions (TDMs) with differential gene expression. *Proc Natl Acad Sci U S A* **102**, 3336-41 (2005).
265. Bird, A.P. CpG-rich islands and the function of DNA methylation. *Nature* **321**, 209-13 (1986).
266. Strichman-Almashanu, L.Z. et al. A genome-wide screen for normally methylated human CpG islands that can identify novel imprinted genes. *Genome Res* **12**, 543-54 (2002).
267. Lovkvist, C., Dodd, I.B., Sneppen, K. & Haerter, J.O. DNA methylation in human epigenomes depends on local topology of CpG sites. *Nucleic Acids Res* (2016).
268. Jin, B. et al. Linking DNA methyltransferases to epigenetic marks and nucleosome structure genome-wide in human tumor cells. *Cell Rep* **2**, 1411-24 (2012).
269. Pathania, R. et al. DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. *Nat Commun* **6**, 6910 (2015).
270. Song, J., Teplova, M., Ishibe-Murakami, S. & Patel, D.J. Structure-based mechanistic insights into DNMT1-mediated maintenance DNA methylation. *Science* **335**, 709-12 (2012).
271. Day, J.J. & Sweatt, J.D. DNA methylation and memory formation. *Nat Neurosci* **13**, 1319-23 (2010).
272. Okano, M., Bell, D.W., Haber, D.A. & Li, E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-57 (1999).
273. Li, E., Bestor, T.H. & Jaenisch, R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. *Cell* **69**, 915-26 (1992).
274. Baylin, S.B. DNA methylation and gene silencing in cancer. *Nat Clin Pract Oncol* **2 Suppl 1**, S4-11 (2005).
275. Dong, S.M., Kim, H.S., Rha, S.H. & Sidransky, D. Promoter hypermethylation of multiple genes in carcinoma of the uterine cervix. *Clin Cancer Res* **7**, 1982-6 (2001).
276. Kang, S.H. et al. Transcriptional repression of the transforming growth factor-beta type I receptor gene by DNA methylation results in the development of TGF-beta resistance in human gastric cancer. *Oncogene* **18**, 7280-6 (1999).
277. Rodriguez, C. et al. CTCF is a DNA methylation-sensitive positive regulator of the INK/ARF locus. *Biochem Biophys Res Commun* **392**, 129-34 (2010).
278. Perez, A. et al. Impact of methylation on the physical properties of DNA. *Biophys J* **102**, 2140-8 (2012).
279. Wade, P.A. Methyl CpG-binding proteins and transcriptional repression. *Bioessays* **23**, 1131-7 (2001).

280. Spruijt, C.G. & Vermeulen, M. DNA methylation: old dog, new tricks? *Nat Struct Mol Biol* **21**, 949-54 (2014).
281. Bogdanovic, O. & Veenstra, G.J. DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma* **118**, 549-65 (2009).
282. Defossez, P.A. & Stancheva, I. Biological functions of methyl-CpG-binding proteins. *Prog Mol Biol Transl Sci* **101**, 377-98 (2011).
283. Filion, G.J. et al. A family of human zinc finger proteins that bind methylated DNA and repress transcription. *Mol Cell Biol* **26**, 169-81 (2006).
284. Laget, S. et al. The human proteins MBD5 and MBD6 associate with heterochromatin but they do not bind methylated DNA. *PLoS One* **5**, e11982 (2010).
285. Lorsbach, R.B. et al. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia* **17**, 637-41 (2003).
286. Tefferi, A. et al. Frequent TET2 mutations in systemic mastocytosis: clinical, KITD816V and FIP1L1-PDGFRA correlates. *Leukemia* **23**, 900-4 (2009).
287. Abdel-Wahab, O. et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. *Blood* **114**, 144-7 (2009).
288. Tahiliani, M. et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* **324**, 930-5 (2009).
289. Guibert, S. & Weber, M. Functions of DNA methylation and hydroxymethylation in mammalian development. *Curr Top Dev Biol* **104**, 47-83 (2013).
290. Pastor, W.A., Aravind, L. & Rao, A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat Rev Mol Cell Biol* **14**, 341-56 (2013).
291. Williams, K., Christensen, J. & Helin, K. DNA methylation: TET proteins-guardians of CpG islands? *EMBO Rep* **13**, 28-35 (2011).
292. Lister, R. & Ecker, J.R. Finding the fifth base: genome-wide sequencing of cytosine methylation. *Genome Res* **19**, 959-66 (2009).
293. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315-22 (2009).
294. Guo, J.U. et al. Distribution, recognition and regulation of non-CpG methylation in the adult mammalian brain. *Nat Neurosci* **17**, 215-22 (2014).
295. Nady, N. et al. Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J Biol Chem* **286**, 24300-11 (2011).
296. Bashtrykov, P., Jankevicius, G., Jurkowska, R.Z., Ragozin, S. & Jeltsch, A. The UHRF1 protein stimulates the activity and specificity of the maintenance DNA methyltransferase DNMT1 by an allosteric mechanism. *J Biol Chem* **289**, 4106-15 (2014).
297. Avvakumov, G.V. et al. Structural basis for recognition of hemi-methylated DNA by the SRA domain of human UHRF1. *Nature* **455**, 822-5 (2008).
298. Liu, X. et al. UHRF1 targets DNMT1 for DNA methylation through cooperative binding of hemi-methylated DNA and methylated H3K9. *Nat Commun* **4**, 1563 (2013).
299. Jenkins, Y. et al. Critical role of the ubiquitin ligase activity of UHRF1, a nuclear RING finger protein, in tumor cell growth. *Mol Biol Cell* **16**, 5621-9 (2005).
300. Qin, W., Leonhardt, H. & Spada, F. Usp7 and Uhrf1 control ubiquitination and stability of the maintenance DNA methyltransferase Dnmt1. *J Cell Biochem* **112**, 439-44 (2011).
301. Rothbart, S.B. et al. Association of UHRF1 with methylated H3K9 directs the maintenance of DNA methylation. *Nat Struct Mol Biol* **19**, 1155-60 (2012).
302. Hu, L., Li, Z., Wang, P., Lin, Y. & Xu, Y. Crystal structure of PHD domain of UHRF1 and insights into recognition of unmodified histone H3 arginine residue 2. *Cell Res* **21**, 1374-8 (2011).
303. Wang, C. et al. Structural basis for site-specific reading of unmodified R2 of histone H3 tail by UHRF1 PHD finger. *Cell Res* **21**, 1379-82 (2011).
304. Rajakumara, E. et al. PHD finger recognition of unmodified histone H3R2 links UHRF1 to regulation of euchromatic gene expression. *Mol Cell* **43**, 275-84 (2011).
305. Du, J. & Patel, D.J. Structural biology-based insights into combinatorial readout and crosstalk among epigenetic marks. *Biochim Biophys Acta* **1839**, 719-27 (2014).

306. Laherty, C.D. et al. Histone deacetylases associated with the mSin3 corepressor mediate mad transcriptional repression. *Cell* **89**, 349-56 (1997).
307. Nagy, L. et al. Nuclear receptor repression mediated by a complex containing SMRT, mSin3A, and histone deacetylase. *Cell* **89**, 373-80 (1997).
308. Suzuki, M., Yamada, T., Kihara-Negishi, F., Sakurai, T. & Oikawa, T. Direct association between PU.1 and MeCP2 that recruits mSin3A-HDAC complex for PU.1-mediated transcriptional repression. *Oncogene* **22**, 8688-98 (2003).
309. Lyst, M.J. et al. Rett syndrome mutations abolish the interaction of MeCP2 with the NCoR/SMRT co-repressor. *Nat Neurosci* **16**, 898-902 (2013).
310. Jones, P.L. et al. Methylated DNA and MeCP2 recruit histone deacetylase to repress transcription. *Nat Genet* **19**, 187-91 (1998).
311. Le Guezennec, X. et al. MBD2/NuRD and MBD3/NuRD, two distinct complexes with different biochemical and functional properties. *Mol Cell Biol* **26**, 843-51 (2006).
312. Xue, Y. et al. NURD, a novel complex with both ATP-dependent chromatin-remodeling and histone deacetylase activities. *Mol Cell* **2**, 851-61 (1998).
313. Klose, R.J. & Bird, A.P. Genomic DNA methylation: the mark and its mediators. *Trends Biochem Sci* **31**, 89-97 (2006).
314. Hansen, K.D. et al. Increased methylation variation in epigenetic domains across cancer types. *Nat Genet* **43**, 768-75 (2011).
315. Issa, J.P. Epigenetic variation and cellular Darwinism. *Nat Genet* **43**, 724-6 (2011).
316. Esteller, M. Epigenetic gene silencing in cancer: the DNA hypermethylome. *Hum Mol Genet* **16 Spec No 1**, R50-9 (2007).
317. Sawan, C. & Herceg, Z. Histone modifications and cancer. *Adv Genet* **70**, 57-85 (2010).
318. Fraga, M.F. et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. *Nat Genet* **37**, 391-400 (2005).
319. Cohen, I., Poreba, E., Kamieniarz, K. & Schneider, R. Histone modifiers in cancer: friends or foes? *Genes Cancer* **2**, 631-47 (2011).
320. Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H. & Nakatani, Y. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell* **87**, 953-9 (1996).
321. Muraoka, M. et al. p300 gene alterations in colorectal and gastric carcinomas. *Oncogene* **12**, 1565-9 (1996).
322. Gayther, S.A. et al. Mutations truncating the EP300 acetylase in human cancers. *Nat Genet* **24**, 300-3 (2000).
323. Peifer, M. et al. Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nat Genet* **44**, 1104-10 (2012).
324. Bryan, E.J. et al. Mutation analysis of EP300 in colon, breast and ovarian carcinomas. *Int J Cancer* **102**, 137-41 (2002).
325. Ferrari, R. et al. Epigenetic reprogramming by adenovirus e1a. *Science* **321**, 1086-8 (2008).
326. Horwitz, G.A. et al. Adenovirus small e1a alters global patterns of histone modification. *Science* **321**, 1084-5 (2008).
327. Simo-Riudalbas, L. et al. KAT6B Is a Tumor Suppressor Histone H3 Lysine 23 Acetyltransferase Undergoing Genomic Loss in Small Cell Lung Cancer. *Cancer Res* **75**, 3936-45 (2015).
328. Morin, R.D. et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. *Nature* **476**, 298-303 (2011).
329. Gorrini, C. et al. Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. *Nature* **448**, 1063-7 (2007).
330. Gocho, Y. et al. A novel recurrent EP300-ZNF384 gene fusion in B-cell precursor acute lymphoblastic leukemia. *Leukemia* **29**, 2445-8 (2015).
331. Ida, K. et al. Adenoviral E1A-associated protein p300 is involved in acute myeloid leukemia with t(11;22)(q23;q13). *Blood* **90**, 4699-704 (1997).
332. Crowley, J.A., Wang, Y., Rapoport, A.P. & Ning, Y. Detection of MOZ-CBP fusion in acute myeloid leukemia with 8;16 translocation. *Leukemia* **19**, 2344-5 (2005).

333. Yang, X.J. & Ullah, M. MOZ and MORF, two large MYSTic HATs in normal and cancer stem cells. *Oncogene* **26**, 5408-19 (2007).
334. Wang, J. et al. Conditional MLL-CBP targets GMP and models therapy-related myeloproliferative disease. *Embo J* **24**, 368-81 (2005).
335. Ropero, S. & Esteller, M. The role of histone deacetylases (HDACs) in human cancer. *Mol Oncol* **1**, 19-25 (2007).
336. West, A.C. & Johnstone, R.W. New and emerging HDAC inhibitors for cancer treatment. *J Clin Invest* **124**, 30-9 (2014).
337. Yang, H. et al. Overexpression of histone deacetylases in cancer cells is controlled by interplay of transcription factors and epigenetic modulators. *Faseb J* **28**, 4265-79 (2014).
338. Muller, B.M. et al. Differential expression of histone deacetylases HDAC1, 2 and 3 in human breast cancer--overexpression of HDAC2 and HDAC3 is associated with clinicopathological indicators of disease progression. *BMC Cancer* **13**, 215 (2013).
339. Kawai, H., Li, H., Avraham, S., Jiang, S. & Avraham, H.K. Overexpression of histone deacetylase HDAC1 modulates breast cancer progression by negative regulation of estrogen receptor alpha. *Int J Cancer* **107**, 353-8 (2003).
340. Mutze, K. et al. Histone deacetylase (HDAC) 1 and 2 expression and chemotherapy in gastric cancer. *Ann Surg Oncol* **17**, 3336-43 (2010).
341. Burdelski, C. et al. HDAC1 overexpression independently predicts biochemical recurrence and is associated with rapid tumor cell proliferation and genomic instability in prostate cancer. *Exp Mol Pathol* **98**, 419-26 (2015).
342. Gao, D.J. et al. Upregulated histone deacetylase 1 expression in pancreatic ductal adenocarcinoma and specific siRNA inhibits the growth of cancer cells. *Pancreas* **39**, 994-1001 (2010).
343. Hayashi, A. et al. Type-specific roles of histone deacetylase (HDAC) overexpression in ovarian carcinoma: HDAC1 enhances cell proliferation and HDAC3 stimulates cell migration with downregulation of E-cadherin. *Int J Cancer* **127**, 1332-46 (2010).
344. Lin, Z. et al. Combination of proteasome and HDAC inhibitors for uterine cervical cancer treatment. *Clin Cancer Res* **15**, 570-7 (2009).
345. Ler, S.Y. et al. HDAC1 and HDAC2 independently predict mortality in hepatocellular carcinoma by a competing risk regression model in a Southeast Asian population. *Oncol Rep* **34**, 2238-50 (2015).
346. Stypula-Cyrus, Y. et al. HDAC up-regulation in early colon field carcinogenesis is involved in cell tumorigenicity through regulation of chromatin structure. *PLoS One* **8**, e64600 (2013).
347. Spurling, C.C. et al. HDAC3 overexpression and colon cancer cell proliferation and differentiation. *Mol Carcinog* **47**, 137-47 (2008).
348. Wilson, A.J. et al. HDAC4 promotes growth of colon cancer cells via repression of p21. *Mol Biol Cell* **19**, 4062-75 (2008).
349. Kang, Z.H. et al. Histone deacetylase HDAC4 promotes gastric cancer SGC-7901 cells progression via p21 repression. *PLoS One* **9**, e98894 (2014).
350. Wang, L. et al. Increased expression of histone deacetylaces (HDACs) and inhibition of prostate cancer growth and invasion by HDAC inhibitor SAHA. *Am J Transl Res* **1**, 62-71 (2009).
351. Chen, J. et al. HDAC5 promotes osteosarcoma progression by upregulation of Twist 1 expression. *Tumour Biol* **35**, 1383-7 (2014).
352. Cao, C. et al. Functional interaction of histone deacetylase 5 (HDAC5) and lysine-specific demethylase 1 (LSD1) promotes breast cancer progression. *Oncogene* (2016).
353. Liu, J. et al. Both HDAC5 and HDAC6 are required for the proliferation and metastasis of melanoma cells. *J Transl Med* **14**, 7 (2016).
354. Ding, G. et al. HDAC6 promotes hepatocellular carcinoma progression by inhibiting P53 transcriptional activity. *FEBS Lett* **587**, 880-6 (2013).

355. Li, S., Liu, X., Chen, X., Zhang, L. & Wang, X. Histone deacetylase 6 promotes growth of glioblastoma through inhibition of SMAD2 signaling. *Tumour Biol* **36**, 9661-5 (2015).
356. Ouaissi, M. et al. High histone deacetylase 7 (HDAC7) expression is significantly associated with adenocarcinomas of the pancreas. *Ann Surg Oncol* **15**, 2318-28 (2008).
357. Niegisch, G. et al. Changes in histone deacetylase (HDAC) expression patterns and activity of HDAC inhibitors in urothelial cancers. *Urol Oncol* **31**, 1770-9 (2013).
358. Oehme, I. et al. Histone deacetylase 8 in neuroblastoma tumorigenesis. *Clin Cancer Res* **15**, 91-9 (2009).
359. Ropero, S. et al. A truncating mutation of HDAC2 in human cancers confers resistance to histone deacetylase inhibition. *Nat Genet* **38**, 566-9 (2006).
360. Hanigan, C.L. et al. An inactivating mutation in HDAC2 leads to dysregulation of apoptosis mediated by APAF1. *Gastroenterology* **135**, 1654-1664 e2 (2008).
361. Ropero, S. et al. Transforming pathways unleashed by a HDAC2 mutation in human cancer. *Oncogene* **27**, 4008-12 (2008).
362. Kriegl, L., Vieth, M., Kirchner, T. & Menssen, A. Up-regulation of c-MYC and SIRT1 expression correlates with malignant transformation in the serrated route to colorectal cancer. *Oncotarget* **3**, 1182-93 (2012).
363. Hao, C. et al. Overexpression of SIRT1 promotes metastasis through epithelial-mesenchymal transition in hepatocellular carcinoma. *BMC Cancer* **14**, 978 (2014).
364. Mvunta, D.H. et al. Overexpression of SIRT1 is Associated With Poor Outcomes in Patients With Ovarian Carcinoma. *Appl Immunohistochem Mol Morphol* (2016).
365. Derr, R.S. et al. High nuclear expression levels of histone-modifying enzymes LSD1, HDAC2 and SIRT1 in tumor cells correlate with decreased survival and increased relapse in breast cancer patients. *BMC Cancer* **14**, 604 (2014).
366. Kim, Y.R., Kim, S.S., Yoo, N.J. & Lee, S.H. Frameshift mutation of SIRT1 gene in gastric and colorectal carcinomas with microsatellite instability. *Apmis* **118**, 81-2 (2010).
367. Halfed, D.G. et al. SIRT2 Expression Is Higher in Uveal Melanoma than In Ocular Melanocytes. *Ocul Oncol Pathol* **2**, 100-4 (2015).
368. Chen, J. et al. SIRT2 overexpression in hepatocellular carcinoma mediates epithelial to mesenchymal transition by protein kinase B/glycogen synthase kinase-3beta/beta-catenin signaling. *Hepatology* **57**, 2287-98 (2013).
369. Shackelford, R. et al. Nicotinamide phosphoribosyltransferase and SIRT3 expression are increased in well-differentiated thyroid carcinomas. *Anticancer Res* **33**, 3047-52 (2013).
370. Miyo, M. et al. Tumour-suppressive function of SIRT4 in human colorectal cancer. *Br J Cancer* **113**, 492-9 (2015).
371. Wang, L., Zhou, H., Wang, Y., Cui, G. & Di, L.J. CtBP maintains cancer cell growth and metabolic homeostasis via regulating SIRT4. *Cell Death Dis* **6**, e1620 (2015).
372. Yu, H. et al. Overexpression of sirt7 exhibits oncogenic property and serves as a prognostic factor in colorectal cancer. *Clin Cancer Res* **20**, 3434-45 (2014).
373. Ryan, R.J. & Bernstein, B.E. Molecular biology. Genetic events that shape the cancer epigenome. *Science* **336**, 1513-4 (2012).
374. Keppler, B.R. & Archer, T.K. Chromatin-modifying enzymes as therapeutic targets--Part 1. *Expert Opin Ther Targets* **12**, 1301-12 (2008).
375. Yokoyama, Y. et al. Cancer-associated upregulation of histone H3 lysine 9 trimethylation promotes cell motility in vitro and drives tumor formation in vivo. *Cancer Sci* **104**, 889-95 (2013).
376. Keung, E.Z. et al. Increased H3K9me3 drives dedifferentiated phenotype via KLF6 repression in liposarcoma. *J Clin Invest* **125**, 2965-78 (2015).
377. Holm, K. et al. Global H3K27 trimethylation and EZH2 abundance in breast tumor subtypes. *Mol Oncol* **6**, 494-506 (2012).

378. Ngollo, M. et al. The association between histone 3 lysine 27 trimethylation (H3K27me3) and prostate cancer: relationship with clinicopathological parameters. *BMC Cancer* **14**, 994 (2014).
379. Yokoyama, Y. et al. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast Cancer Res* **16**, R66 (2014).
380. Meyer, C. & Marschalek, R. LDI-PCR: identification of known and unknown gene fusions of the human MLL gene. *Methods Mol Biol* **538**, 71-83 (2009).
381. Krivtsov, A.V. & Armstrong, S.A. MLL translocations, histone modifications and leukaemia stem-cell development. *Nat Rev Cancer* **7**, 823-33 (2007).
382. Grasso, C.S. et al. The mutational landscape of lethal castration-resistant prostate cancer. *Nature* **487**, 239-43 (2012).
383. Zang, Z.J. et al. Exome sequencing of gastric adenocarcinoma identifies recurrent somatic mutations in cell adhesion and chromatin remodeling genes. *Nat Genet* **44**, 570-4 (2012).
384. Stransky, N. et al. The mutational landscape of head and neck squamous cell carcinoma. *Science* **333**, 1157-60 (2011).
385. Gui, Y. et al. Frequent mutations of chromatin remodeling genes in transitional cell carcinoma of the bladder. *Nat Genet* **43**, 875-8 (2011).
386. Fujimoto, A. et al. Whole-genome sequencing of liver cancers identifies etiological influences on mutation patterns and recurrent mutations in chromatin regulators. *Nat Genet* **44**, 760-4 (2012).
387. Hammerman PS, e.a. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519-25 (2012).
388. Bracken, A.P. et al. EZH2 is downstream of the pRB-E2F pathway, essential for proliferation and amplified in cancer. *Embo J* **22**, 5323-35 (2003).
389. Lu, C. et al. Regulation of tumor angiogenesis by EZH2. *Cancer Cell* **18**, 185-97 (2010).
390. He, M. et al. Cancer angiogenesis induced by Kaposi sarcoma-associated herpesvirus is mediated by EZH2. *Cancer Res* **72**, 3582-92 (2012).
391. Nikoloski, G. et al. Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes. *Nat Genet* **42**, 665-7 (2010).
392. McCabe, M.T. et al. Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc Natl Acad Sci U S A* **109**, 2989-94 (2012).
393. Cock-Rada, A.M. et al. SMYD3 promotes cancer invasion by epigenetic upregulation of the metalloproteinase MMP-9. *Cancer Res* **72**, 810-20 (2012).
394. Hamamoto, R. et al. SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nat Cell Biol* **6**, 731-40 (2004).
395. Hamamoto, R. et al. Enhanced SMYD3 expression is essential for the growth of breast cancer cells. *Cancer Sci* **97**, 113-8 (2006).
396. Mazur, P.K. et al. SMYD3 links lysine methylation of MAP3K2 to Ras-driven cancer. *Nature* **510**, 283-7 (2014).
397. Sarris, M.E., Moulos, P., Haroniti, A., Giakountis, A. & Talianidis, I. Smyd3 Is a Transcriptional Potentiator of Multiple Cancer-Promoting Genes and Required for Liver and Colon Cancer Development. *Cancer Cell* **29**, 354-66 (2016).
398. Fei, Q. et al. Histone methyltransferase SETDB1 regulates liver cancer cell growth through methylation of p53. *Nat Commun* **6**, 8651 (2015).
399. Ceol, C.J. et al. The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513-7 (2011).
400. Rodriguez-Paredes, M. et al. Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. *Oncogene* **33**, 2807-13 (2014).
401. Lu, Z. et al. Histone-lysine methyltransferase EHMT2 is involved in proliferation, apoptosis, cell invasion, and DNA methylation of human neuroblastoma cells. *Anticancer Drugs* **24**, 484-93 (2013).

402. Cerveira, N. et al. Frequency of NUP98-NSD1 fusion transcript in childhood acute myeloid leukaemia. *Leukemia* **17**, 2244-7 (2003).
403. Wang, G.G., Cai, L., Pasillas, M.P. & Kamps, M.P. NUP98-NSD1 links H3K36 methylation to Hox-A gene activation and leukaemogenesis. *Nat Cell Biol* **9**, 804-12 (2007).
404. Berdasco, M. et al. Epigenetic inactivation of the Sotos overgrowth syndrome gene histone methyltransferase NSD1 in human neuroblastoma and glioma. *Proc Natl Acad Sci U S A* **106**, 21830-5 (2009).
405. Lee, M.G., Wynder, C., Cooch, N. & Shiekhattar, R. An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation. *Nature* **437**, 432-5 (2005).
406. Shi, Y. et al. Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **119**, 941-53 (2004).
407. Shi, Y.J. et al. Regulation of LSD1 histone demethylase activity by its associated factors. *Mol Cell* **19**, 857-64 (2005).
408. Ding, J. et al. LSD1-mediated epigenetic modification contributes to proliferation and metastasis of colon cancer. *Br J Cancer* **109**, 994-1003 (2013).
409. Hayami, S. et al. Overexpression of LSD1 contributes to human carcinogenesis through chromatin regulation in various cancers. *Int J Cancer* **128**, 574-86 (2011).
410. Nagasawa, S. et al. LSD1 overexpression is associated with poor prognosis in basal-like breast cancer, and sensitivity to PARP inhibition. *PLoS One* **10**, e0118002 (2015).
411. Kashyap, V. et al. The lysine specific demethylase-1 (LSD1/KDM1A) regulates VEGF-A expression in prostate cancer. *Mol Oncol* **7**, 555-66 (2013).
412. Lv, T. et al. Over-expression of LSD1 promotes proliferation, migration and invasion in non-small cell lung cancer. *PLoS One* **7**, e35065 (2012).
413. Amente, S., Lania, L. & Majello, B. The histone LSD1 demethylase in stemness and cancer transcription programs. *Biochim Biophys Acta* **1829**, 981-6 (2013).
414. Wissmann, M. et al. Cooperative demethylation by JMJD2C and LSD1 promotes androgen receptor-dependent gene expression. *Nat Cell Biol* **9**, 347-53 (2007).
415. Wu, J. et al. Identification and functional analysis of 9p24 amplified genes in human breast cancer. *Oncogene* **31**, 333-41 (2012).
416. Yang, Z.Q. et al. A novel amplicon at 9p23 - 24 in squamous cell carcinoma of the esophagus that lies proximal to GASC1 and harbors NFIB. *Jpn J Cancer Res* **92**, 423-8 (2001).
417. Northcott, P.A. et al. Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nat Genet* **41**, 465-72 (2009).
418. Ehrbrecht, A. et al. Comprehensive genomic analysis of desmoplastic medulloblastomas: identification of novel amplified genes and separate evaluation of the different histological components. *J Pathol* **208**, 554-63 (2006).
419. Rui, L. et al. Cooperative epigenetic modulation by cancer amplicon genes. *Cancer Cell* **18**, 590-605 (2010).
420. Kooistra, S.M. & Helin, K. Molecular mechanisms and potential functions of histone demethylases. *Nat Rev Mol Cell Biol* **13**, 297-311 (2012).
421. Yamamoto, S. et al. JARID1B is a luminal lineage-driving oncogene in breast cancer. *Cancer Cell* **25**, 762-77 (2014).
422. Hayami, S. et al. Overexpression of the JmjC histone demethylase KDM5B in human carcinogenesis: involvement in the proliferation of cancer cells through the E2F/RB pathway. *Mol Cancer* **9**, 59 (2010).
423. Andersson, A. et al. Microarray-based classification of a consecutive series of 121 childhood acute leukemias: prediction of leukemic and genetic subtype as well as of minimal residual disease status. *Leukemia* **21**, 1198-203 (2007).
424. He, J., Nguyen, A.T. & Zhang, Y. KDM2b/JHDM1b, an H3K36me2-specific demethylase, is required for initiation and maintenance of acute myeloid leukemia. *Blood* **117**, 3869-80 (2011).
425. Sperger, J.M. et al. Gene expression patterns in human embryonic stem cells and human pluripotent germ cell tumors. *Proc Natl Acad Sci U S A* **100**, 13350-5 (2003).

426. Tzatsos, A. et al. KDM2B promotes pancreatic cancer via Polycomb-dependent and -independent transcriptional programs. *J Clin Invest* **123**, 727-39 (2013).
427. van Haaften, G. et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet* **41**, 521-3 (2009).
428. Mar, B.G. et al. Sequencing histone-modifying enzymes identifies UTX mutations in acute lymphoblastic leukemia. *Leukemia* **26**, 1881-3 (2012).
429. Robinson, G. et al. Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488**, 43-8 (2012).
430. Van der Meulen, J. et al. The H3K27me3 demethylase UTX is a gender-specific tumor suppressor in T-cell acute lymphoblastic leukemia. *Blood* **125**, 13-21 (2015).
431. Heyn, H. & Esteller, M. DNA methylation profiling in the clinic: applications and challenges. *Nat Rev Genet* **13**, 679-92 (2012).
432. Feinberg, A.P. & Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **301**, 89-92 (1983).
433. Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455 (2003).
434. Gaudet, F. et al. Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489-92 (2003).
435. Kaneda, A. et al. Frequent hypomethylation in multiple promoter CpG islands is associated with global hypomethylation, but not with frequent promoter hypermethylation. *Cancer Sci* **95**, 58-64 (2004).
436. Badal, V. et al. CpG methylation of human papillomavirus type 16 DNA in cervical cancer cell lines and in clinical specimens: genomic hypomethylation correlates with carcinogenic progression. *J Virol* **77**, 6227-34 (2003).
437. Hur, K. et al. Hypomethylation of long interspersed nuclear element-1 (LINE-1) leads to activation of proto-oncogenes in human colorectal cancer metastasis. *Gut* **63**, 635-46 (2014).
438. Tao, Q., Robertson, K.D., Manns, A., Hildesheim, A. & Ambinder, R.F. The Epstein-Barr virus major latent promoter Qp is constitutively active, hypomethylated, and methylation sensitive. *J Virol* **72**, 7075-83 (1998).
439. Wolff, E.M. et al. Hypomethylation of a LINE-1 promoter activates an alternate transcript of the MET oncogene in bladders with cancer. *PLoS Genet* **6**, e1000917 (2010).
440. Chenais, B. Transposable elements in cancer and other human diseases. *Curr Cancer Drug Targets* **15**, 227-42 (2015).
441. Bourc'his, D. & Bestor, T.H. Meiotic catastrophe and retrotransposon reactivation in male germ cells lacking Dnmt3L. *Nature* **431**, 96-9 (2004).
442. Schulz, W.A. L1 retrotransposons in human cancers. *J Biomed Biotechnol* **2006**, 83672 (2006).
443. Carreira, P.E., Richardson, S.R. & Faulkner, G.J. L1 retrotransposons, cancer stem cells and oncogenesis. *Febs J* **281**, 63-73 (2014).
444. Helman, E. et al. Somatic retrotransposition in human cancer revealed by whole-genome and exome sequencing. *Genome Res* **24**, 1053-63 (2014).
445. Lee, E. et al. Landscape of somatic retrotransposition in human cancers. *Science* **337**, 967-71 (2012).
446. Nishigaki, M. et al. Discovery of aberrant expression of R-RAS by cancer-linked DNA hypomethylation in gastric cancer using microarrays. *Cancer Res* **65**, 2115-24 (2005).
447. Wang, Q. et al. Hypomethylation of WNT5A, CRIP1 and S100P in prostate cancer. *Oncogene* **26**, 6560-5 (2007).
448. Sato, N. et al. Frequent hypomethylation of multiple genes overexpressed in pancreatic ductal adenocarcinoma. *Cancer Res* **63**, 4158-66 (2003).
449. Nakamura, N. & Takenaga, K. Hypomethylation of the metastasis-associated S100A4 gene correlates with gene activation in human colon adenocarcinoma cell lines. *Clin Exp Metastasis* **16**, 471-9 (1998).

450. He, Y. et al. Hypomethylation of the hsa-miR-191 locus causes high expression of hsa-mir-191 and promotes the epithelial-to-mesenchymal transition in hepatocellular carcinoma. *Neoplasia* **13**, 841-53 (2011).
451. Yamada, N. et al. Epigenetic regulation of microRNA-128a expression contributes to the apoptosis-resistance of human T-cell leukaemia jurkat cells by modulating expression of fas-associated protein with death domain (FADD). *Biochim Biophys Acta* **1843**, 590-602 (2014).
452. Mets, E. et al. MicroRNA-128-3p is a novel oncomiR targeting PHF6 in T-cell acute lymphoblastic leukemia. *Haematologica* **99**, 1326-33 (2014).
453. Vizoso, M. et al. Epigenetic activation of a cryptic TBC1D16 transcript enhances melanoma progression by targeting EGFR. *Nat Med* **21**, 741-50 (2015).
454. Bhusari, S., Yang, B., Kueck, J., Huang, W. & Jarrard, D.F. Insulin-like growth factor-2 (IGF2) loss of imprinting marks a field defect within human prostates containing cancer. *Prostate* **71**, 1621-30 (2011).
455. Shetty, P.J. et al. Regulation of IGF2 transcript and protein expression by altered methylation in breast cancer. *J Cancer Res Clin Oncol* **137**, 339-45 (2011).
456. Baba, Y. et al. Hypomethylation of the IGF2 DMR in colorectal tumors, detected by bisulfite pyrosequencing, is associated with poor prognosis. *Gastroenterology* **139**, 1855-64 (2010).
457. Ogawa, O. et al. Relaxation of insulin-like growth factor II gene imprinting implicated in Wilms' tumour. *Nature* **362**, 749-51 (1993).
458. Rao, X. et al. CpG island shore methylation regulates caveolin-1 expression in breast cancer. *Oncogene* **32**, 4519-28 (2013).
459. Lee, H.G. et al. Regulation of HK2 expression through alterations in CpG methylation of the HK2 promoter during progression of hepatocellular carcinoma. *Oncotarget* (2016).
460. Wajed, S.A., Laird, P.W. & DeMeester, T.R. DNA methylation: an alternative pathway to cancer. *Ann Surg* **234**, 10-20 (2001).
461. Baylin, S.B. & Herman, J.G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet* **16**, 168-74 (2000).
462. Sakai, T. et al. Allele-specific hypermethylation of the retinoblastoma tumor-suppressor gene. *Am J Hum Genet* **48**, 880-8 (1991).
463. Esteller, M. et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J Natl Cancer Inst* **92**, 564-9 (2000).
464. Esteller, M., Hamilton, S.R., Burger, P.C., Baylin, S.B. & Herman, J.G. Inactivation of the DNA repair gene O6-methylguanine-DNA methyltransferase by promoter hypermethylation is a common event in primary human neoplasia. *Cancer Res* **59**, 793-7 (1999).
465. Herman, J.G. et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A* **95**, 6870-5 (1998).
466. Esteller, M., Levine, R., Baylin, S.B., Ellenson, L.H. & Herman, J.G. MLH1 promoter hypermethylation is associated with the microsatellite instability phenotype in sporadic endometrial carcinomas. *Oncogene* **17**, 2413-7 (1998).
467. Leung, S.Y. et al. hMLH1 promoter methylation and lack of hMLH1 expression in sporadic gastric carcinomas with high-frequency microsatellite instability. *Cancer Res* **59**, 159-64 (1999).
468. Zhao, Y. et al. High-frequency aberrantly methylated targets in pancreatic adenocarcinoma identified via global DNA methylation analysis using methylCap-seq. *Clin Epigenetics* **6**, 18 (2014).
469. Dudziec, E. et al. Hypermethylation of CpG islands and shores around specific microRNAs and mirtrons is associated with the phenotype and presence of bladder cancer. *Clin Cancer Res* **17**, 1287-96 (2011).
470. Patel, J.P. et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* **366**, 1079-89 (2012).

471. Ley, T.J. et al. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* **363**, 2424-33 (2010).
472. Yan, X.J. et al. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* **43**, 309-15 (2011).
473. Russler-Germain, D.A. et al. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell* **25**, 442-54 (2014).
474. Miremadi, A., Oestergaard, M.Z., Pharoah, P.D. & Caldas, C. Cancer genetics of epigenetic genes. *Hum Mol Genet* **16 Spec No 1**, R28-49 (2007).
475. Jones, J.S. et al. DNMT3b polymorphism and hereditary nonpolyposis colorectal cancer age of onset. *Cancer Epidemiol Biomarkers Prev* **15**, 886-91 (2006).
476. Shen, H. et al. A novel polymorphism in human cytosine DNA-methyltransferase-3B promoter is associated with an increased risk of lung cancer. *Cancer Res* **62**, 4992-5 (2002).
477. Montgomery, K.G., Liu, M.C., Eccles, D.M. & Campbell, I.G. The DNMT3B C-->T promoter polymorphism and risk of breast cancer in a British population: a case-control study. *Breast Cancer Res* **6**, R390-4 (2004).
478. Singal, R., Das, P.M., Manoharan, M., Reis, I.M. & Schlesselman, J.J. Polymorphisms in the DNA methyltransferase 3b gene and prostate cancer risk. *Oncol Rep* **14**, 569-73 (2005).
479. Di Croce, L. et al. Methyltransferase recruitment and DNA hypermethylation of target promoters by an oncogenic transcription factor. *Science* **295**, 1079-82 (2002).
480. Zhu, Y., Brown, H.N., Zhang, Y., Holford, T.R. & Zheng, T. Genotypes and haplotypes of the methyl-CpG-binding domain 2 modify breast cancer risk dependent upon menopausal status. *Breast Cancer Res* **7**, R745-52 (2005).
481. Coutinho, D.F. et al. TET2 expression level and 5-hydroxymethylcytosine are decreased in refractory cytopenia of childhood. *Leuk Res* **39**, 1103-8 (2015).
482. Ko, M. et al. TET proteins and 5-methylcytosine oxidation in hematological cancers. *Immunol Rev* **263**, 6-21 (2015).
483. Delhommeau, F. et al. Mutation in TET2 in myeloid cancers. *N Engl J Med* **360**, 2289-301 (2009).
484. Langemeijer, S.M. et al. Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet* **41**, 838-42 (2009).
485. Haffner, M.C. et al. Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. *Oncotarget* **2**, 627-37 (2011).
486. Murata, A. et al. TET family proteins and 5-hydroxymethylcytosine in esophageal squamous cell carcinoma. *Oncotarget* **6**, 23372-82 (2015).
487. Udali, S. et al. Global DNA methylation and hydroxymethylation differ in hepatocellular carcinoma and cholangiocarcinoma and relate to survival rate. *Hepatology* **62**, 496-504 (2015).
488. Rideout, W.M., 3rd, Coetzee, G.A., Olumi, A.F. & Jones, P.A. 5-Methylcytosine as an endogenous mutagen in the human LDL receptor and p53 genes. *Science* **249**, 1288-90 (1990).
489. Bottema, C.D. et al. The pattern of spontaneous germ-line mutation: relative rates of mutation at or near CpG dinucleotides in the factor IX gene. *Hum Genet* **91**, 496-503 (1993).
490. Goelz, S.E., Vogelstein, B., Hamilton, S.R. & Feinberg, A.P. Hypomethylation of DNA from benign and malignant human colon neoplasms. *Science* **228**, 187-90 (1985).
491. Esteller, M. CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future. *Oncogene* **21**, 5427-40 (2002).
492. Harvey Lodish, A.B., S Lawrence Zipursky, Paul Matsudaira, David Baltimore, and James Darnell. *Chapter 1. The Dynamic Cell, Molecular Cell Biology, 4th edition*, (W. H. Freeman and Company, 2000).

493. Hoagland, M.B., Stephenson, M.L., Scott, J.F., Hecht, L.I. & Zamecnik, P.C. A soluble ribonucleic acid intermediate in protein synthesis. *J Biol Chem* **231**, 241-57 (1958).
494. Holley, R.W. et al. Structure Of A Ribonucleic Acid. *Science* **147**, 1462-5 (1965).
495. Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M.B. Annotating non-coding regions of the genome. *Nat Rev Genet* **11**, 559-71 (2010).
496. Taft, R.J., Pheasant, M. & Mattick, J.S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288-99 (2007).
497. ENCODE. Project Consortium, An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).
498. Djebali, S. et al. Landscape of transcription in human cells. *Nature* **489**, 101-8 (2012).
499. Peschansky, V.J. & Wahlestedt, C. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* **9**, 3-12 (2014).
500. Morris, K.V. & Mattick, J.S. The rise of regulatory RNA. *Nat Rev Genet* **15**, 423-37 (2014).
501. Morris, K.V., Chan, S.W., Jacobsen, S.E. & Looney, D.J. Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305**, 1289-92 (2004).
502. Mendell, J.T. MicroRNAs: critical regulators of development, cellular physiology and malignancy. *Cell Cycle* **4**, 1179-84 (2005).
503. Esteller, M. Non-coding RNAs in human disease. *Nat Rev Genet* **12**, 861-74 (2011).
504. Liz, J. & Esteller, M. lncRNAs and microRNAs with a role in cancer development. *Biochim Biophys Acta* **1859**, 169-76 (2016).
505. Askarian-Amiri, M.E. et al. SNORD-host RNA Zfas1 is a regulator of mammary development and a potential marker for breast cancer. *Rna* **17**, 878-91 (2011).
506. Ronnau, C.G., Verhaegh, G.W., Luna-Velez, M.V. & Schalken, J.A. Noncoding RNAs as novel biomarkers in prostate cancer. *Biomed Res Int* **2014**, 591703 (2014).
507. Fatima, R., Akhade, V.S., Pal, D. & Rao, S.M. Long noncoding RNAs in development and cancer: potential biomarkers and therapeutic targets. *Mol Cell Ther* **3**, 5 (2015).
508. Hayes, J., Peruzzi, P.P. & Lawler, S. MicroRNAs in cancer: biomarkers, functions and therapy. *Trends Mol Med* **20**, 460-9 (2014).
509. Mishra, P.J. MicroRNAs as promising biomarkers in cancer diagnostics. *Biomark Res* **2**, 19 (2014).
510. Dhanasekaran, K., Kumari, S. & Kanduri, C. Noncoding RNAs in chromatin organization and transcription regulation: an epigenetic view. *Subcell Biochem* **61**, 343-72 (2013).
511. Derrien, T. et al. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* **22**, 1775-89 (2012).
512. Friedman, R.C., Farh, K.K., Burge, C.B. & Bartel, D.P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Res* **19**, 92-105 (2009).
513. Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat Rev Genet* **15**, 7-21 (2014).
514. Hu, W., Alvarez-Dominguez, J.R. & Lodish, H.F. Regulation of mammalian cell differentiation by long non-coding RNAs. *EMBO Rep* **13**, 971-83 (2012).
515. Whyte, W.A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307-19 (2013).
516. Chapuy, B. et al. Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777-90 (2013).
517. Loven, J. et al. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320-34 (2013).
518. Mansour, M.R. et al. Oncogene regulation. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. *Science* **346**, 1373-7 (2014).
519. Chipumuro, E. et al. CDK7 inhibition suppresses super-enhancer-linked oncogenic transcription in MYCN-driven cancer. *Cell* **159**, 1126-39 (2014).
520. International Human Genome Sequencing Consortium, Finishing the euchromatic sequence of the human genome. *Nature* **431**, 931-45 (2004).

521. Mattick, J.S. Challenging the dogma: the hidden layer of non-protein-coding RNAs in complex organisms. *Bioessays* **25**, 930-9 (2003).
522. Esteller, M. Non-coding RNAs in human disease. *Nat Rev Genet* **12**, 861-74.
523. He, L. & Hannon, G.J. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**, 522-31 (2004).
524. Calin, G.A. & Croce, C.M. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**, 857-66 (2006).
525. Esquela-Kerscher, A. & Slack, F.J. Oncomirs - microRNAs with a role in cancer. *Nat Rev Cancer* **6**, 259-69 (2006).
526. Hammond, S.M. MicroRNAs as tumor suppressors. *Nat Genet* **39**, 582-3 (2007).
527. Melo, S.A. & Esteller, M. Dysregulation of microRNAs in cancer: playing with fire. *FEBS Lett* **585**, 2087-99 (2011).
528. Thomson, J.M. et al. Extensive post-transcriptional regulation of microRNAs and its implications for cancer. *Genes Dev* **20**, 2202-7 (2006).
529. Chang, T.C. et al. Widespread microRNA repression by Myc contributes to tumorigenesis. *Nat Genet* **40**, 43-50 (2008).
530. Melo, S.A. et al. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet* **41**, 365-70 (2009).
531. Hill, D.A. et al. DICER1 mutations in familial pleuropulmonary blastoma. *Science* **325**, 965 (2009).
532. Melo, S.A. et al. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. *Cancer Cell* **18**, 303-15 (2010).
533. Lopez-Serra, P. & Esteller, M. DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene* **31**, 1609-22 (2012).
534. Bejerano, G. et al. Ultraconserved elements in the human genome. *Science* **304**, 1321-5 (2004).
535. Katzman, S. et al. Human genome ultraconserved elements are ultraselected. *Science* **317**, 915 (2007).
536. Calin, G.A. et al. Ultraconserved regions encoding ncRNAs are altered in human leukemias and carcinomas. *Cancer Cell* **12**, 215-29 (2007).
537. Lujambio, A. et al. CpG island hypermethylation-associated silencing of non-coding RNAs transcribed from ultraconserved regions in human cancer. *Oncogene* **29**, 6390-401 (2010).
538. Gerbi, S.A. Small nucleolar RNA. *Biochem Cell Biol* **73**, 845-58 (1995).
539. Cavaille, J., Nicoloso, M. & Bachellerie, J.P. Targeted ribose methylation of RNA in vivo directed by tailored antisense RNA guides. *Nature* **383**, 732-5 (1996).
540. Kiss-Laszlo, Z., Henry, Y., Bachellerie, J.P., Caizergues-Ferrer, M. & Kiss, T. Site-specific ribose methylation of preribosomal RNA: a novel function for small nucleolar RNAs. *Cell* **85**, 1077-88 (1996).
541. Lestrade, L. & Weber, M.J. snoRNA-LBME-db, a comprehensive database of human H/ACA and C/D box snoRNAs. *Nucleic Acids Res* **34**, D158-62 (2006).
542. Balakin, A.G., Smith, L. & Fournier, M.J. The RNA world of the nucleolus: two major families of small RNAs defined by different box elements with related functions. *Cell* **86**, 823-34 (1996).
543. Bachellerie, J.P., Cavaille, J. & Huttenhofer, A. The expanding snoRNA world. *Biochimie* **84**, 775-90 (2002).
544. Reichow, S.L., Hamma, T., Ferre-D'Amare, A.R. & Varani, G. The structure and function of small nucleolar ribonucleoproteins. *Nucleic Acids Res* **35**, 1452-64 (2007).
545. Heiss, N.S. et al. X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. *Nat Genet* **19**, 32-8 (1998).
546. Walne, A.J. et al. Genetic heterogeneity in autosomal recessive dyskeratosis congenita with one subtype due to mutations in the telomerase-associated protein NOP10. *Hum Mol Genet* **16**, 1619-29 (2007).
547. Vulliamy, T. et al. Mutations in the telomerase component NHP2 cause the premature ageing syndrome dyskeratosis congenita. *Proc Natl Acad Sci U S A* **105**, 8073-8 (2008).

548. Chang, L.S., Lin, S.Y., Lieu, A.S. & Wu, T.L. Differential expression of human 5S snoRNA genes. *Biochem Biophys Res Commun* **299**, 196-200 (2002).
549. Mourtada-Maarabouni, M., Hedge, V.L., Kirkham, L., Farzaneh, F. & Williams, G.T. Growth arrest in human T-cells is controlled by the non-coding RNA growth-arrest-specific transcript 5 (GAS5). *J Cell Sci* **121**, 939-46 (2008).
550. Mourtada-Maarabouni, M., Pickard, M.R., Hedge, V.L., Farzaneh, F. & Williams, G.T. GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. *Oncogene* **28**, 195-208 (2009).
551. Dong, X.Y. et al. Implication of snoRNA U50 in human breast cancer. *J Genet Genomics* **36**, 447-54 (2009).
552. Dong, X.Y. et al. SnoRNA U50 is a candidate tumor-suppressor gene at 6q14.3 with a mutation associated with clinically significant prostate cancer. *Hum Mol Genet* **17**, 1031-42 (2008).
553. Horsthemke, B. & Wagstaff, J. Mechanisms of imprinting of the Prader-Willi/Angelman region. *Am J Med Genet A* **146A**, 2041-52 (2008).
554. Kishore, S. & Stamm, S. The snoRNA HBII-52 regulates alternative splicing of the serotonin receptor 2C. *Science* **311**, 230-2 (2006).
555. Bratkovic, T. & Rogelj, B. Biology and applications of small nucleolar RNAs. *Cell Mol Life Sci* **68**, 3843-51 (2011).
556. Kiss, T. & Filipowicz, W. Exonucleolytic processing of small nucleolar RNAs from pre-mRNA introns. *Genes Dev* **9**, 1411-24 (1995).
557. Tycowski, K.T., Shu, M.D. & Steitz, J.A. A mammalian gene with introns instead of exons generating stable RNA products. *Nature* **379**, 464-6 (1996).
558. Hoeppner, M.P., White, S., Jeffares, D.C. & Poole, A.M. Evolutionarily stable association of intronic snoRNAs and microRNAs with their host genes. *Genome Biol Evol* **1**, 420-8 (2009).
559. Li, T., Zhou, X., Wang, X., Zhu, D. & Zhang, Y. Identification and characterization of human snoRNA core promoters. *Genomics* **96**, 50-6 (2010).
560. Takai, D. & Jones, P.A. The CpG island searcher: a new WWW resource. *In Silico Biol* **3**, 235-40 (2003).
561. Sandoval, J. et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* **6**, 692-702 (2011).
562. Kent, W.J. et al. The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).
563. Saini, H.K., Griffiths-Jones, S. & Enright, A.J. Genomic analysis of human microRNA transcripts. *Proc Natl Acad Sci U S A* **104**, 17719-24 (2007).
564. Zhou, X., Ruan, J., Wang, G. & Zhang, W. Characterization and identification of microRNA core promoters in four model species. *PLoS Comput Biol* **3**, e37 (2007).
565. Rhee, I. et al. DNMT1 and DNMT3b cooperate to silence genes in human cancer cells. *Nature* **416**, 552-6 (2002).
566. Yang, J.H. et al. snoSeeker: an advanced computational package for screening of guide and orphan snoRNA genes in the human genome. *Nucleic Acids Res* **34**, 5112-23 (2006).
567. Kiss, A.M., Jady, B.E., Bertrand, E. & Kiss, T. Human box H/ACA pseudouridylation guide RNA machinery. *Mol Cell Biol* **24**, 5797-807 (2004).
568. Luo, Y. & Li, S. Genome-wide analyses of retrogenes derived from the human box H/ACA snoRNAs. *Nucleic Acids Res* **35**, 559-71 (2007).
569. Weber, M.J. Mammalian small nucleolar RNAs are mobile genetic elements. *PLoS Genet* **2**, e205 (2006).
570. McMahon, M., Ayllon, V., Panov, K.I. & O'Connor, R. Ribosomal 18 S RNA processing by the IGF-I-responsive WDR3 protein is integrated with p53 function in cancer cell proliferation. *J Biol Chem* **285**, 18309-18 (2010).
571. Peng, Q. et al. 1A6/DRIM, a novel t-UTP, activates RNA polymerase I transcription and promotes cell proliferation. *PLoS One* **5**, e14244 (2010).

572. Uemura, M. et al. Overexpression of ribosomal RNA in prostate cancer is common but not linked to rDNA promoter hypomethylation. *Oncogene* **31**, 1254-63 (2012).
573. Ishizu, H., Siomi, H. & Siomi, M.C. Biology of PIWI-interacting RNAs: new insights into biogenesis and function inside and outside of germlines. *Genes Dev* **26**, 2361-73 (2012).
574. Meister, G. Argonaute proteins: functional insights and emerging roles. *Nat Rev Genet* **14**, 447-59 (2013).
575. Deng, W. & Lin, H. miwi, a murine homolog of piwi, encodes a cytoplasmic protein essential for spermatogenesis. *Dev Cell* **2**, 819-30 (2002).
576. Kuramochi-Miyagawa, S. et al. Mili, a mammalian member of piwi family gene, is essential for spermatogenesis. *Development* **131**, 839-49 (2004).
577. Krausz, C. et al. Novel insights into DNA methylation features in spermatozoa: stability and peculiarities. *PLoS One* **7**, e44479 (2012).
578. Heyn, H. et al. Epigenetic disruption of the PIWI pathway in human spermatogenic disorders. *PLoS One* **7**, e47892 (2012).
579. Bamezai, S., Rawat, V.P. & Buske, C. Concise review: The Piwi-piRNA axis: pivotal beyond transposon silencing. *Stem Cells* **30**, 2603-11 (2012).
580. Unhavaithaya, Y. et al. MILI, a PIWI-interacting RNA-binding protein, is required for germ line stem cell self-renewal and appears to positively regulate translation. *J Biol Chem* **284**, 6507-19 (2009).
581. Siomi, M.C., Mannen, T. & Siomi, H. How does the royal family of Tudor rule the PIWI-interacting RNA pathway? *Genes Dev* **24**, 636-46 (2010).
582. Reuter, M. et al. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol* **16**, 639-46 (2009).
583. Fung, M.K. et al. Role of MEK/ERK pathway in the MAD2-mediated cisplatin sensitivity in testicular germ cell tumour cells. *Br J Cancer* **95**, 475-84 (2006).
584. Brennecke, J. et al. Discrete small RNA-generating loci as master regulators of transposon activity in Drosophila. *Cell* **128**, 1089-103 (2007).
585. Aravin, A.A. et al. A piRNA pathway primed by individual transposons is linked to de novo DNA methylation in mice. *Mol Cell* **31**, 785-99 (2008).
586. Aravin, A.A., Sachidanandam, R., Girard, A., Fejes-Toth, K. & Hannon, G.J. Developmentally regulated piRNA clusters implicate MILI in transposon control. *Science* **316**, 744-7 (2007).
587. Hotaling, J.M. & Walsh, T.J. Male infertility: a risk factor for testicular cancer. *Nat Rev Urol* **6**, 550-6 (2009).
588. Peng, X., Zeng, X., Peng, S., Deng, D. & Zhang, J. The association risk of male subfertility and testicular cancer: a systematic review. *PLoS One* **4**, e5591 (2009).
589. Grimwade, D. et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood* **116**, 354-65 (2010).
590. Roller, A. et al. Landmark analysis of DNMT3A mutations in hematological malignancies. *Leukemia* **27**, 1573-8 (2013).
591. Im, A.P. et al. DNMT3A and IDH mutations in acute myeloid leukemia and other myeloid malignancies: associations with prognosis and potential treatment strategies. *Leukemia* **28**, 1774-83 (2014).
592. Hajkova, H. et al. Decreased DNA methylation in acute myeloid leukemia patients with DNMT3A mutations and prognostic implications of DNA methylation. *Leuk Res* **36**, 1128-33 (2012).
593. Qu, Y. et al. Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics* **9**, 1108-19 (2014).
594. Krueger, F. & Andrews, S.R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-2 (2011).

595. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).
596. Li, H. Tabix: fast retrieval of sequence features from generic TAB-delimited files. *Bioinformatics* **27**, 718-9 (2011).
597. Hansen, K.D., Langmead, B. & Irizarry, R.A. BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biol* **13**, R83 (2012).
598. Haider, S. et al. BioMart Central Portal--unified access to biological data. *Nucleic Acids Res* **37**, W23-7 (2009).
599. Jacinto, F.V., Ballestar, E. & Esteller, M. Impaired recruitment of the histone methyltransferase DOT1L contributes to the incomplete reactivation of tumor suppressor genes upon DNA demethylation. *Oncogene* **28**, 4212-24 (2009).
600. Hollink, I.H. et al. Low frequency of DNMT3A mutations in pediatric AML, and the identification of the OCI-AML3 cell line as an in vitro model. *Leukemia* **26**, 371-3 (2012).
601. Pineault, N., Helgason, C.D., Lawrence, H.J. & Humphries, R.K. Differential expression of Hox, Meis1, and Pbx1 genes in primitive cells throughout murine hematopoietic ontogeny. *Exp Hematol* **30**, 49-57 (2002).
602. Kohlmann, A. et al. New insights into MLL gene rearranged acute leukemias using gene expression profiling: shared pathways, lineage commitment, and partner genes. *Leukemia* **19**, 953-64 (2005).
603. Wong, P., Iwasaki, M., Somervaille, T.C., So, C.W. & Cleary, M.L. Meis1 is an essential and rate-limiting regulator of MLL leukemia stem cell potential. *Genes Dev* **21**, 2762-74 (2007).
604. Heuser, M. et al. Cell of origin in AML: susceptibility to MN1-induced transformation is regulated by the MEIS1/AbdB-like HOX protein complex. *Cancer Cell* **20**, 39-52 (2011).
605. Zeisig, B.B. et al. Hoxa9 and Meis1 are key targets for MLL-ENL-mediated cellular immortalization. *Mol Cell Biol* **24**, 617-28 (2004).
606. Xiong, Q. et al. Characterization of miRNomes in acute and chronic myeloid leukemia cell lines. *Genomics Proteomics Bioinformatics* **12**, 79-91 (2014).
607. Li, Y. et al. Overexpressed let-7a-3 is associated with poor outcome in acute myeloid leukemia. *Leuk Res* **37**, 1642-7 (2013).
608. Brueckner, B. et al. The human let-7a-3 locus contains an epigenetically regulated microRNA gene with oncogenic function. *Cancer Res* **67**, 1419-23 (2007).
609. Hon, G.C. et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. *Nat Genet* **45**, 1198-206 (2013).
610. Rideout, W.M., 3rd, Eggan, K. & Jaenisch, R. Nuclear cloning and epigenetic reprogramming of the genome. *Science* **293**, 1093-8 (2001).
611. Baylin, S.B. & Jones, P.A. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-34 (2011).
612. Berman, B.P. et al. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nat Genet* **44**, 40-6 (2011).
613. Stamatoyannopoulos, J.A., Clegg, C.H. & Li, Q. Sheltering of gamma-globin expression from position effects requires both an upstream locus control region and a regulatory element 3' to the A gamma-globin gene. *Mol Cell Biol* **17**, 240-7 (1997).
614. Talbot, D. et al. A dominant control region from the human beta-globin locus conferring integration site-independent gene expression. *Nature* **338**, 352-5 (1989).
615. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934-47 (2013).
616. Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
617. R Development Core Team. R: a language and environment for statistical computing. Vienna. *R Foundation for Statistical Computing* (2009).

618. Harrow, J. et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**, 1760-74 (2012).
619. Molaro, A. et al. Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029-41 (2011).
620. Trapnell, C., Pachter, L. & Salzberg, S.L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105-11 (2009).
621. Aryee, M.J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363-9 (2014).
622. Morris, T.J. et al. ChAMP: 450k Chip Analysis Methylation Pipeline. *Bioinformatics* **30**, 428-30 (2014).
623. Leek, J.T. & Storey, J.D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet* **3**, 1724-35 (2007).
624. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-7 (2012).
625. Huang da, W., Sherman, B.T. & Lempicki, R.A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4**, 44-57 (2009).
626. Huang da, W., Sherman, B.T. & Lempicki, R.A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* **37**, 1-13 (2009).
627. Li, G. et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. *Cell* **148**, 84-98 (2012).
628. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70 (2012).
629. Frith, M.C. et al. Detection of functional DNA motifs via statistical over-representation. *Nucleic Acids Res* **32**, 1372-81 (2004).
630. Sandelin, A., Alkema, W., Engstrom, P., Wasserman, W.W. & Lenhard, B. JASPAR: an open-access database for eukaryotic transcription factor binding profiles. *Nucleic Acids Res* **32**, D91-4 (2004).
631. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543-50 (2014).
632. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74.
633. Stadler, M.B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490-5 (2011).
634. Feldmann, A. et al. Transcription factor occupancy can mediate active turnover of DNA methylation at regulatory regions. *PLoS Genet* **9**, e1003994 (2013).
635. Hodges, E. et al. Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol Cell* **44**, 17-28 (2011).
636. Schlesinger, F., Smith, A.D., Gingeras, T.R., Hannon, G.J. & Hodges, E. De novo DNA demethylation and noncoding transcription define active intergenic regulatory elements. *Genome Res* **23**, 1601-14 (2013).
637. Wiench, M. et al. DNA methylation status predicts cell type-specific enhancer activity. *Embo J* **30**, 3028-39 (2011).
638. Taberlay, P.C., Statham, A.L., Kelly, T.K., Clark, S.J. & Jones, P.A. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. *Genome Res* **24**, 1421-32 (2014).
639. Blattler, A. et al. Global loss of DNA methylation uncovers intronic enhancers in genes showing expression changes. *Genome Biol* **15**, 469 (2014).
640. Aran, D., Sabato, S. & Hellman, A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. *Genome Biol* **14**, R21 (2013).
641. Hon, G.C. et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res* **22**, 246-58 (2012).
642. Rubio-Perez, C. et al. In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* **27**, 382-96 (2015).
643. Hon, G.C. et al. 5mC oxidation by Tet2 modulates enhancer activity and timing of transcriptome reprogramming during differentiation. *Mol Cell* **56**, 286-97 (2014).

644. Stewart, S.K. et al. oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips. *Methods* **72**, 9-15 (2015).
645. Jones, P.A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet* **13**, 484-92 (2012).
646. Byers, L.A. et al. An epithelial-mesenchymal transition gene signature predicts resistance to EGFR and PI3K inhibitors and identifies Axl as a therapeutic target for overcoming EGFR inhibitor resistance. *Clin Cancer Res* **19**, 279-90 (2013).
647. Fernandez, A.F. et al. A DNA methylation fingerprint of 1628 human samples. *Genome Res* **22**, 407-19 (2012).
648. Wang, K. et al. Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat Genet* **46**, 573-82 (2014).
649. Bieller, A. et al. Isolation and characterization of the human forkhead gene FOXQ1. *DNA Cell Biol* **20**, 555-61 (2001).
650. Cheng, Y. & Lotan, R. Molecular cloning and characterization of a novel retinoic acid-inducible gene that encodes a putative G protein-coupled receptor. *J Biol Chem* **273**, 35008-15 (1998).
651. Jeong, M. et al. Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat Genet* **46**, 17-23 (2014).
652. Hovestadt, V. et al. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537-41 (2014).
653. Anastas, J.N. & Moon, R.T. WNT signalling pathways as therapeutic targets in cancer. *Nat Rev Cancer* **13**, 11-26 (2013).
654. Wu, Z.Q. et al. Canonical Wnt suppressor, Axin2, promotes colon carcinoma oncogenic activity. *Proc Natl Acad Sci U S A* **109**, 11312-7 (2012).
655. Kayl, A.E. & Meyers, C.A. Side-effects of chemotherapy and quality of life in ovarian and breast cancer patients. *Curr Opin Obstet Gynecol* **18**, 24-8 (2006).
656. Everhard, S. et al. MGMT methylation: a marker of response to temozolomide in low-grade gliomas. *Ann Neurol* **60**, 740-3 (2006).
657. Druker, B.J. et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. *Nat Med* **2**, 561-6 (1996).
658. Wagle, N. et al. Dissecting therapeutic resistance to RAF inhibition in melanoma by tumor genomic profiling. *J Clin Oncol* **29**, 3085-96 (2011).
659. Lopez-Serra, P. & Esteller, M. DNA methylation-associated silencing of tumor-suppressor microRNAs in cancer. *Oncogene* **31**, 1609-22.
660. Zhang, X. et al. Maternally expressed gene 3, an imprinted noncoding RNA gene, is associated with meningioma pathogenesis and progression. *Cancer Res* **70**, 2350-8 (2010).
661. Cao, J. The functional role of long non-coding RNAs and epigenetics. *Biol Proced Online* **16**, 11 (2014).
662. Tollervey, D. & Kiss, T. Function and synthesis of small nucleolar RNAs. *Curr Opin Cell Biol* **9**, 337-42 (1997).
663. Weinstein, L.B. & Steitz, J.A. Guided tours: from precursor snoRNA to functional snoRNP. *Curr Opin Cell Biol* **11**, 378-84 (1999).
664. Mattick, J.S. & Makunin, I.V. Non-coding RNA. *Hum Mol Genet* **15 Spec No 1**, R17-29 (2006).
665. Decatur, W.A. & Fournier, M.J. rRNA modifications and ribosome function. *Trends Biochem Sci* **27**, 344-51 (2002).
666. Mei, Y.P. et al. Small nucleolar RNA 42 acts as an oncogene in lung tumorigenesis. *Oncogene* **31**, 2794-804 (2012).
667. Tanaka, R. et al. Intronic U50 small-nucleolar-RNA (snoRNA) host gene of no protein-coding potential is mapped at the chromosome breakpoint t(3;6)(q27;q15) of human B-cell lymphoma. *Genes Cells* **5**, 277-87 (2000).
668. Alter, B.P., Giri, N., Savage, S.A. & Rosenberg, P.S. Cancer in dyskeratosis congenita. *Blood* **113**, 6549-57 (2009).

669. Montanaro, L., Trere, D. & Derenzini, M. Nucleolus, ribosomes, and cancer. *Am J Pathol* **173**, 301-10 (2008).
670. Shiue, C.N., Berkson, R.G. & Wright, A.P. c-Myc induces changes in higher order rDNA structure on stimulation of quiescent cells. *Oncogene* **28**, 1833-42 (2009).
671. Zhang, Z. et al. Long non-coding RNA chromogenic in situ hybridisation signal pattern correlation with breast tumour pathology. *J Clin Pathol* **69**, 76-81 (2016).
672. Li, T. et al. Amplification of Long Noncoding RNA ZFAS1 Promotes Metastasis in Hepatocellular Carcinoma. *Cancer Res* **75**, 3181-91 (2015).
673. Thorenoor, N. et al. Long non-coding RNA ZFAS1 interacts with CDK1 and is involved in p53-dependent cell cycle control and apoptosis in colorectal cancer. *Oncotarget* **7**, 622-37 (2016).
674. Pickard, M.R., Mourtada-Maarabouni, M. & Williams, G.T. Long non-coding RNA GAS5 regulates apoptosis in prostate cancer cell lines. *Biochim Biophys Acta* **1832**, 1613-23 (2013).
675. Qiao, H.P., Gao, W.S., Huo, J.X. & Yang, Z.S. Long non-coding RNA GAS5 functions as a tumor suppressor in renal cell carcinoma. *Asian Pac J Cancer Prev* **14**, 1077-82 (2013).
676. Lu, X. et al. Downregulation of gas5 increases pancreatic cancer cell proliferation by regulating CDK6. *Cell Tissue Res* **354**, 891-6 (2013).
677. Liu, Z. et al. Downregulation of GAS5 promotes bladder cancer cell proliferation, partly by regulating CDK6. *PLoS One* **8**, e73991 (2013).
678. Shi, X. et al. A critical role for the long non-coding RNA GAS5 in proliferation and apoptosis in non-small-cell lung cancer. *Mol Carcinog* **54 Suppl 1**, E1-E12 (2015).
679. Sun, M. et al. Decreased expression of long noncoding RNA GAS5 indicates a poor prognosis and promotes cell proliferation in gastric cancer. *BMC Cancer* **14**, 319 (2014).
680. Yin, D. et al. Long noncoding RNA GAS5 affects cell proliferation and predicts a poor prognosis in patients with colorectal cancer. *Med Oncol* **31**, 253 (2014).
681. Gee, H.E. et al. The small-nucleolar RNAs commonly used for microRNA normalisation correlate with tumour pathology and prognosis. *Br J Cancer* **104**, 1168-77 (2011).
682. Cao, S., Liu, W., Li, F., Zhao, W. & Qin, C. Decreased expression of lncRNA GAS5 predicts a poor prognosis in cervical cancer. *Int J Clin Exp Pathol* **7**, 6776-83 (2014).
683. Ronchetti, D. et al. Small nucleolar RNAs as new biomarkers in chronic lymphocytic leukemia. *BMC Med Genomics* **6**, 27 (2013).
684. Bellodi, C. et al. H/ACA small RNA dysfunctions in disease reveal key roles for noncoding RNA modifications in hematopoietic stem cell differentiation. *Cell Rep* **3**, 1493-502 (2013).
685. Valleron, W. et al. Specific small nucleolar RNA expression profiles in acute leukemia. *Leukemia* **26**, 2052-60 (2012).
686. Ono, M. et al. Analysis of human small nucleolar RNAs (snoRNA) and the development of snoRNA modulator of gene expression vectors. *Mol Biol Cell* **21**, 1569-84 (2010).
687. Ono, M. et al. Identification of human miRNA precursors that resemble box C/D snoRNAs. *Nucleic Acids Res* **39**, 3879-91 (2011).
688. Liz, J. et al. Regulation of pri-miRNA processing by a long noncoding RNA transcribed from an ultraconserved region. *Mol Cell* **55**, 138-47 (2014).
689. Fu, A., Jacobs, D.I. & Zhu, Y. Epigenome-wide analysis of piRNAs in gene-specific DNA methylation. *RNA Biol* **11**, 1301-12 (2014).
690. Rajasethupathy, P. et al. A role for neuronal piRNAs in the epigenetic control of memory-related synaptic plasticity. *Cell* **149**, 693-707 (2012).
691. Rounge, T.B. et al. Profiling of the small RNA populations in human testicular germ cell tumors shows global loss of piRNAs. *Mol Cancer* **14**, 153 (2015).

692. Iliev, R. et al. Decreased expression levels of PIWIL1, PIWIL2, and PIWIL4 are associated with worse survival in renal cell carcinoma patients. *Onco Targets Ther* **9**, 217-22 (2016).
693. Wu, Q. et al. Expression of the Argonaute protein PiwiL2 and piRNAs in adult mouse mesenchymal stem cells. *Biochem Biophys Res Commun* **396**, 915-20 (2010).
694. Navarro, A. et al. The significance of PIWI family expression in human lung embryogenesis and non-small cell lung cancer. *Oncotarget* **6**, 31544-56 (2015).
695. Fang, K. et al. A distinct set of long non-coding RNAs in childhood MLL-rearranged acute lymphoblastic leukemia: biology and epigenetic target. *Hum Mol Genet* **23**, 3278-88 (2014).
696. Kim, T.K., Hemberg, M. & Gray, J.M. Enhancer RNAs: a class of long noncoding RNAs synthesized at enhancers. *Cold Spring Harb Perspect Biol* **7**, a018622 (2015).
697. Takamizawa, J. et al. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* **64**, 3753-6 (2004).
698. Liu, X.Q., Song, W.J., Sun, T.M., Zhang, P.Z. & Wang, J. Targeted delivery of antisense inhibitor of miRNA for antiangiogenesis therapy using cRGD-functionalized nanoparticles. *Mol Pharm* **8**, 250-9 (2011).
699. Love, T.M., Moffett, H.F. & Novina, C.D. Not miR-ly small RNAs: big potential for microRNAs in therapy. *J Allergy Clin Immunol* **121**, 309-19 (2008).
700. Novina, C.D. & Chabner, B.A. RNA-Directed Therapy: The Next Step in the miRNA Revolution. *Oncologist* **13**, 1-3 (2008).
701. Lanford, R.E. et al. Therapeutic silencing of microRNA-122 in primates with chronic hepatitis C virus infection. *Science* **327**, 198-201 (2010).
702. Burnett, J.C. & Rossi, J.J. RNA-based therapeutics: current progress and future prospects. *Chem Biol* **19**, 60-71 (2012).
703. Adams, B.D., Parsons, C. & Slack, F.J. The tumor-suppressive and potential therapeutic functions of miR-34a in epithelial carcinomas. *Expert Opin Ther Targets* **20**, 737-53 (2016).
704. Uddin, M.N. et al. Enhanced bioavailability of orally administered antisense oligonucleotide to nuclear factor kappa B mRNA after microencapsulation with albumin. *J Drug Target* **21**, 450-7 (2013).
705. van Putten, M. et al. Preclinical studies on intestinal administration of antisense oligonucleotides as a model for oral delivery for treatment of duchenne muscular dystrophy. *Mol Ther Nucleic Acids* **3**, e211 (2014).
706. Akhtar, S. Oral delivery of siRNA and antisense oligonucleotides. *J Drug Target* **17**, 491-5 (2009).