Epigenetic alterations and ectopic activation of tissue-specific gene clusters in lung adenocarcinoma

ANNA DIACOFOTAKI

JUNE 2022

Thesis submitted for the degree of Doctor of Philosophy in Biomedical and Pharmaceutical Sciences

Supervisor: Professor Charles De Smet

Laboratory of Genetic and Epigenetic Alterations of the Genome
de Duve Institute - Université Catholique de Louvain

Advisory Committee

Supervisor

Professor Charles De Smet

de Duve Institute, laboratory of Genetic and Epigenetic Alterations of the Genome
Université Catholique de Louvain, Brussels, Belgium

President of the jury

Professor Guido Bommer

de Duve Institute, laboratory of Physiological Chemistry
Université Catholique de Louvain, Brussels, Belgium

Internal jury members

Professor Pierre Coulie

de Duve Institute, laboratory of Cellular Genetics
Université Catholique de Louvain, Brussels, Belgium

Professor Emmanuel Hermans

Institute of Neurosciences, laboratory of Neuropharmacology
Université Catholique de Louvain, Brussels, Belgium

External jury members

Professor Pierre-Antoine Defossez

Epigenetics and Cell Fate Centre, laboratory of DNA Methylation Dynamics of Eukaryotic Genome
Université de Paris, Paris, France

Professor Bernard Thienpont

Department of Human Genetics, laboratory of Functional Epigenetics
Katholieke Universiteit Leuven, Leuven, Belgium

This thesis was supported by UCLouvain-FSR, FNRS-FRIA, and UCLouvain-Patrimoine grants.

ACKNOWLEDGEMENTS

Coming to the end of this journey, it is with great emotion that I would like to express my deepest gratitude to all the people that helped me come this far, that allowed me to grow both scientifically and personally.

I would like to thank the members of my jury, who followed me through the years and allowed this project to go further, *Profs Guido Bommer, Pierre Coulie and Emmanuel Hermans.* Thank you for your advice and support during the committee meetings and the grant applications. I would also wish to thank the external examiners that agreed to be part of the advisory panel of my thesis and evaluate my work: *Prof. Pierre-Antoine Defossez* and *Prof. Bernard Thienpont*. It is a great honor to have the opportunity to talk about our science with the experts in the field. I am looking forward to our scientific discussion and I am certain that future projects and new perspectives will arise from our work. I thank you in advance for your availability, time, and help.

I am more than grateful to the funding agencies that made this thesis possible. First, to the FSR-UCLouvain grant that secured my first year in the doctoral school, then the FNRS-FRIA grant that allowed me to pursue my research adventure and finally to the Patrimoine grant that allowed to complete this scientific journey.

On a more personal note...

I want to warmly thank my supervisor, *Prof. Charles De Smet,* who gave me the chance to be a member of the GEPI team. Dear Charles, a thank you is not enough for everything that you have done for me throughout these years. Thank you for giving me the opportunity to join your lab, to work with you and learn from you. Since day one, you showed me a world that I didn't know existed and helped me become part of it. Thank you for your trust, for all the time you devoted to train me and guide me, to structure my thoughts, define "the" question and help me connect the pieces of the puzzle. Thank you for giving me the freedom to think and try, and for getting me back on track when I needed. Thank you for your availability, your door was always open for me. I hope you enjoyed our DDICian and Brainy adventures as much as I did (even if it took 3 weeks for an analysis to be done, in

parallel with new “je pense” theories, decorated with your extravagant face expressions). Thank you for everything.

I would also like to thank *Dr Axelle Loriot*. Axelle, thank you for sharing the DDIC project with me. For your support, advice, kind words, I am grateful. Thank you for entrusting me with leading the project, for the brainstorming sessions, the knowledge that you shared with me both experimentally and through your programming scripts. It was of great help and source of knowledge, thank you.

I also want to thank *Jean Fain* (aka HRH Mr Assistant), my partner in crime! Thank you for your craziness, for being there for me and your support throughout the ups, and especially the downs of this journey, throughout all the experimental failures, deadlines, and presentations. You always found the right words to make me go further and lift me up. Thank you, it meant a lot to me.

A big thank you goes to *Dr Hara Episkopou*, I hope you know how much you contributed to my growth. Thank you for taking the time to train me, to listen and give me advice both experimentally and organizationally. I can see so much more clearly now, and it helped me get through both projects and the thesis writing. Thank you for your constant guidance and support through the years, for all the scientific discussions, and most importantly, your friendship.

I would deeply like to thank the Telo-lab of our GEPI unit. It wouldn’t have been the same without all of you! Eating the donkey tail was tough, but more enjoyable in your presence! I want to thank *Dr Anabelle Decottignies* for her advice and her role as a woman in STEM. Thank you for the Telo-adoption, your availability and feedback for grant writing, and presentations. *Eloïse Claude,* “ma chère collègue” and I think I can now say my friend. Thank you for being there for me, particularly in this last rough year. We were in this together since the beginning, and I am very grateful to have shared this journey with you. I wish you the best for the end! We are almost there! Thank you as well for the time you took to proofread this manuscript and your support until the end! I also wish to thank *Dan Geelen* for all the meticulous “Dan moments” regarding the typos of this work. I would also like to thank the rest of the super-duper GEPI team! Thank you for the constant help and laughter with protocols, reagents, experiments, troubleshooting, plant culture, and for all the moments shared inside and outside the lab, especially during the

difficult period of the pandemic. I wish you the best of luck for everything you undertake in the future.

I also wish to thank *Prof. Laurent Gatto*. Thank you, Laurent, for the amazing teacher that you are and your dedication! Thank you for finding the time to listen and give me advice regarding analyses and statistics. Because of your valuable help my work took a very different turn and I thank you for that. Thank you for encouraging me to attend the Bioinformatics Summer School and Bioconductor meetings. Reproducible research is truly a part of me now.

I would also want to thank the LPAD team members for the lovely atmosphere that reigns on our floor, the team building sessions and drinks that we shared. I particularly want to thank, *Angeline Fages and Dr Lila Gannoun*, the dream duo, for their support, and their kind words. I also would like to thank *Dr Sophie Pirenne* for the time she took to explain pathology data for tumor classification. I also thank *Dr Ilianna Zoi* for the great advice given on academic journeys.

I also wish to thank my parents from the bottom of my heart for giving me the opportunity to pursue a university degree despite all the sacrifices they made for my siblings and me. I also want to thank my siblings for always being there for me, for their unconditional love and support, for encouraging me and letting me fall on them when I needed to. Thank you in particular Christina, for the time you devoted to edit this manuscript, it was of tremendous help. Dear family, I hope you know that I wouldn't have made it without any of you! Finally, I also want to express my deepest gratitude to you *Dr. Isabelle Schiltz*. Your constant presence throughout the years keeps me going. You have always been here through everything, and I couldn't ask for more. Clearly, a thank you is not even close to enough. I also want to thank *Dr. Athena Cavounidis*, my driver example, for showing me.. more! Thank you for your support, the editing tips for this manuscript and your friendship. It helped me more than you can imagine in these last months of this journey.

Ευχαριστώ τον καθένα από εσάς,
Με εκτίμηση,

Άννα

TABLE OF CONTENTS

SUMMARY

Tumor development is characterized by both genetic and epigenetic alterations that lead to gene expression changes. One of such transcriptomic changes is the aberrant activation of a group of germline-specific genes in a variety of tumors of somatic origin. The spurious activation of this group of genes is due to the DNA demethylation of their promoter region in tumor cells, a phenomenon associated with the global loss of DNA methylation that often accompanies tumorigenesis. Several of these "cancer-germline" genes were found to exert pro-tumoral functions, or evoke an immune response. Cancer-germline genes are therefore attracting much interest, as they constitute promising targets for anti-cancer therapies with the hope of limited side effects due to their highly restricted pattern of expression.

Whether additional tissue-specific gene clusters become activated in tumors, be it by DNA demethylation or not, remains unanswered. Most studies, indeed, have used transcriptomic approaches that were biased towards testis-specific genes. In the present work, we aimed to fill this gap by combining omics data of lung adenocarcinoma (LUAD) tissues and cell lines, as well as normal lung and alveolar type II cells (AT2), the cells of origin of LUAD, without imposing any criteria of expression specificity to the ectopically activated genes we were looking for.

We divided our work in two complementary axes. In the first part, we asked whether the global loss of DNA demethylation in tumors is also associated with the ectopic activation of somatic-specific gene clusters in LUAD. We found that, indeed, promoter DNA demethylation in these cells leads to the transcriptional induction of genes displaying selective expression in gastro-intestinal tissues (colon and small intestine), and of genes specific to stratified squamous epithelia (esophagus, skin and vagina). Interestingly, our data show that ectopic activation of this latter group of genes in LUAD is associated with poor prognosis of the cancer patients. In the second part of our work, we assessed the extent of tissue-specific gene clusters that become ectopically activated in LUAD, irrespective of the regulatory mechanism involved in their activation, i.e., not especially as a result of DNA demethylation. The results revealed that besides the above identified tissue-specific gene clusters, a large number of brain-related genes were ectopically activated in LUAD tissues

and cell lines. Multiple regulatory mechanisms appeared to contribute to the ectopic activation of these brain-related genes in LUAD, including ASCL1-driven neuroendocrine differentiation, promoter DNA demethylation, and dysregulation of the REST silencing complex.

Our work confirms and extends previous observations of illegitimate transcription of germline and somatic-specific gene clusters in LUAD. We also shed light on the regulatory mechanisms involved in their aberrant activation, that contribute thereby to the “cellular identity crisis” of tumor cells.

ABBREVIATIONS

2-OG	*2-OxoGlutarate*
5-azadC	*5-aza-2'-deoxycytidine*
5hmC	*5-hydroxymethylcytosine*
5mC	*5-methylcytosine*
ADD	*ATRX-DNMT3-DNMT3L domain*
AID	*Activation-Induced cytidine Deaminase*
APOBEC	*APOlipoprotein B mRNA Editing enzyme, Catalytic polypeptide-like*
AT2	*Alveolar Type II cells*
BER	*Base Excision Repair*
BORIS	*Brother Of the Regulator of Imprinted Sites*
CB	*Cancer-Brain*
CBT	*Cancer-Brain/Testis*
CG	*Cancer-Germline*
CGI	*CpG Island*
ChIP	*Chromatin ImmunoPrecipitation*
chrX	*chromosome X*
CpG	*Cytosine-phosphate-Guanine*
CT	*Cancer-Testis*
CTA	*Cancer-Testis Antigens*
CTL	*Cytolytic T Lymphocytes*
CxxC	*Cysteine-residue-residue-Cysteine*
Cys	*Cysteine-rich domain*
DBTSS	*DataBase of Transcriptional Start Sites*
DDIC	*DNA Demethylation-associated Induction in Cancer*
DLL3	*Delta-Like canonical Ligand 3*
DNMT	*DNA methyl-transferase*
DSBH	*Double-stranded-beta-helix domain*
EctopEx	*Ectopically Expressed*
ENCODE	*Encyclopedia of DNA Elements*
EZH1	*Enhancer of Zeste Homolog 1*
EZH2	*Enhancer of Zeste Homolog 2*
Fe(II)	*Fe2+*
GABA	*γ-aminobutyric acid*
GABRA3	*γ-aminobutyric acid type A Receptor subunit Alpha 3*
GI	*Gastro-Intestinal-specific gene cluster*
GPR19	*G Protein-coupled Receptor 19*
GTEx	*Genotype-Tissue Expression*
HAT	*Histone AcetylTransferase*
HDAC	*Histone DeACetylase*
HKDM	*Histone lysine (K) DeMethylase*
HKMT	*Histone lysine (K) MethylTransferase*
HP1	*Heterochromatin Protein 1*
HPA	*Human Protein Atlas*
IDH	*Isocitrate Deshydrogenase*

KO	*Knock-Out*
lncRNA	*long non-coding RNA*
LUAD	*LUng ADenocarcinoma*
LUSC	*LUng Squamous Carcinoma*
MAGEA1/3/6	*Melanoma AntiGEn 1/3/6*
MBD	*Methyl Binding Domain*
MBP	*Methyl Binding Proteins*
mESC	*mouse Embryonic Stem Cells*
MLTC	*Mixed Lymphocyte Tumor cell Culture*
MS-qPCR	*Methylation-Specific qPCR*
MTase	*MethylTransferase domain*
mTOR	*mammalian Target Of Rapamycin*
NE	*NeuroEndocrine*
NKX2-1	*NK2 homeoboX 1*
NLS	*Nuclear Localisation Signal*
NSCLC	*Non-Small Cell Lung Cancer*
PGC	*Primordial Germ Cells*
PRC2	*Polycomb Repressor Complex 2*
PTM	*Post-Translational Modification*
PWWP	*Pro-Trp-Trp-Pro*
RE1	*Repressive Element 1*
REST	*RE1-Silencing Transcription factor*
RFTS	*Replication Foci-Targeting Sequence*
RNA polII	*RNA polymerase II*
SCLC	*Small Cell Lung Cancer*
SE	*Stratified squamous Epithelia-specific gene cluster*
seq	*sequencing*
SEREX	*SERological analysis of cDNA EXpression libraries*
SETD2	*SET Domain containing 2 enzyme*
SRA	*SET- and RING-associated domain*
TCGA	*The Cancer Genome Atlas*
TDG	*Thymine DNA Glycosylase*
TE	*Transposable Elements*
TET	*Ten Eleven Translocation enzyme*
TJP1	*Tight Junction Protein 1*
TNF	*Tumor Necrosis Factor*
TSS	*Transcriptional Start Site*
TTF1	*Thyroid Transcription Factor 1*
UBL	*Ubiquitin-Like domain*
UHRF1	*Ubiquitin like with PHD and Ring Finger domains 1*
WGBS	*Whole Genome Bisulfite Sequencing*
ZIC2	*ZInc finger of the Cerebellum 2*

GENERAL INTRODUCTION

A. A brief introduction of Epigenetics in health and disease

1. Epigenetics in physiology: maintaining cell identity

The complex nature of a cell and subsequently the architecture of the corresponding tissue and organ are achieved through gene regulatory networks highly coordinated in space and time. The fine-tuned mechanisms that orchestrate the set-up and maintenance of these networks are all encoded by a specific succession of the four nucleotides A-T-C-G of the genome. When and where the cell will be able to "read" and "interpret" precise genomic sequences is key to establish and maintain their identity throughout cell divisions. Using the same genetic code inherited from the fertilized oocyte, we observe a portfolio of gene expression landscapes that allow the development of hundreds of different cell types, each harboring specialized functions. There exists therefore an additional layer of inherited information that allows the regulation of gene expression in a highly ordered spatiotemporal dimension. The study of this heritable information that achieves this cellular identity and memory has been coined "Epigenetics" (literally meaning "onto genetics", as carried by the Greek prefix "epi").

Historically, the field of Epigenetics took different turns in their definition, depending on the scientific field that used the term (Deans and Maggert 2015; Deichmann 2016; Greally 2018). In fact, Epigenetics definition was a matter of debate, with multiple researchers using it interchangeably with the notion "transcriptional regulation". Conrad Waddington's initial definition in 1942 complemented with Robin Holliday's notion of inheritance came together to define this field (Holliday 2006). Accordingly, in this work, we define it as the chemical modifications of the DNA and the associated proteins, that do not affect the underlying nucleotide sequence, and that are stably transmitted through cell divisions, even in the absence of the original cue that established them in the first place. These modifications, therefore, maintain a stable cell identity.

Today, we could classify epigenetic modifications in three main categories. The first and most widely studied epigenetic mark is DNA methylation, followed by post-

translational modifications of the associated histone proteins. These epigenetic modifications decorate the underlying genomic sequence and influence the degree of chromatin compaction inside the nucleus. This results in the study of the third epigenetic subcategory, a highly ordered and non-random occupancy of chromatin inside the nuclear envelope. This is a fast-expanding area of research following the rise of chromosome-conformation capture technologies and bio-informatics analysis tools.

Epigenetic modifications are stable in normal cells. However, they are altered in diseased tissue compared to their healthy counterpart, including in cancer. In the present work, we are focusing primarily on DNA methylation as the epigenetic mark of interest, and in particular, on the consequences of their alterations on the cell identity of tumor cells.

2. *Epigenetics in tumor biology: breaking cell identity*

A change of perspective: DNA mutations vs DNA epi-mutations

Previously and traditionally, tumor biology was viewed under a dichotomous search of oncogenes and tumor suppressor genes, whose alterations resulted in driver events in the cell of origin of the tumor. Common examples are activating *KRAS* mutations found in a wide variety of tumor types, including in 93% of pancreatic adenocarcinoma tissues (Cancer Genome Atlas Research Network 2017), or inactivating mutations of *TP53* found in 42% of Pan-cancer tumor tissues of TCGA (Kandoth et al. 2013). However, it is well established today that besides these genetic alterations that undoubtedly contribute to tumor initiation and development, tumor biology is also characterized by profound changes of their epigenome (You and Jones 2012).

The earliest epigenetic alteration described was the global reduction of DNA methylation of tumor cells compared to the tissue of origin of the tumor (Gama-Sosa et al. 1983a; Feinberg and Vogelstein 1983a). This phenomenon occurs in a wide variety of tumor types, although to different extents. With the emergence of new high throughput technologies, histone post-translational modifications were also found to be altered in tumor cells, contributing therefore to perturbed

nucleosomal dynamics. Indeed, studies have shown that about half of human cancers display mutations in proteins that install or interpret chromatin modifications (You and Jones 2012; Rodriguez-Paredes and Esteller 2011). Additionally, there is a synergistic effect between genetic and epigenetic alterations, where these so-called epimutations can provide the second hit for gene silencing in tumor cells (Herman et al. 1994). Finally, it was also observed that some types of cancer such as ependydomas display extensive gains of DNA methylation, and this intriguingly, without harboring any somatic mutations (Mack et al. 2014).

These genetic and epigenetic alterations in tumors both lead to gene expression changes in these cells, and thus, to an altered cell identity compared to their normal counterparts (Roy and Hebrok 2015). It is believed that this permissive chromatin state confers high adaptability and sustainability of tumor cells in their microenvironment (Hanahan 2022; Flavahan, Gaskell, and Bernstein 2017). These data support that genetics and epigenetics should be viewed as tightly linked and not as independent processes in tumor cells.

In the present work, we focus on epigenetic alterations in tumors, and how they affect the gene expression landscape of the cells.

B. Chromatin organization determines cell identity

In eukaryotes, DNA molecules are localized inside the nuclear envelope of the cell. They are associated with proteins, including histones, altogether forming a structure called the chromatin. Each diploid human cell contains two meters of DNA tightly and meticulously packed inside the six µm-diameter of the nucleus. This DNA condensation is achieved via different levels of chromatin compaction, precisely organized in space and time to allow transcriptional activation or stable repression of specific genomic sequences. Indeed, the chromatin can be divided into euchromatin, a conformation promoting gene expression, and heterochromatin, a refractory state to gene transcription.

The basic unit of the chromatin is the nucleosome (**Fig. 1A**). This unit is defined as a structure composed of the negatively charged DNA rolled twice (+/- 146bp) around a positively charged histone core octamer. This histone octamer is composed of homodimers of H2A, H2B, H3 and H4 histone proteins (Luger et al. 1997). The density of the nucleosomes of the chromatin fiber dictates whether a region will be "open" to gene transcription, when containing a low density of nucleosomes, or a region will be "closed" to gene transcription, when the nucleosomes are tightly packed to one another. Electron microscopy experiments have shown that a loose chromatin can be viewed as a pearl fiber of 10nm in their lowest level of compaction (**Fig. 1A**). Successive foldings of these structures lead to higher levels of chromatin compaction (**Fig. 1B and 1C**) resulting in the final chromatin conformation of the nucleus and a precise 3D location of the chromosomes (**Fig. 1D**).

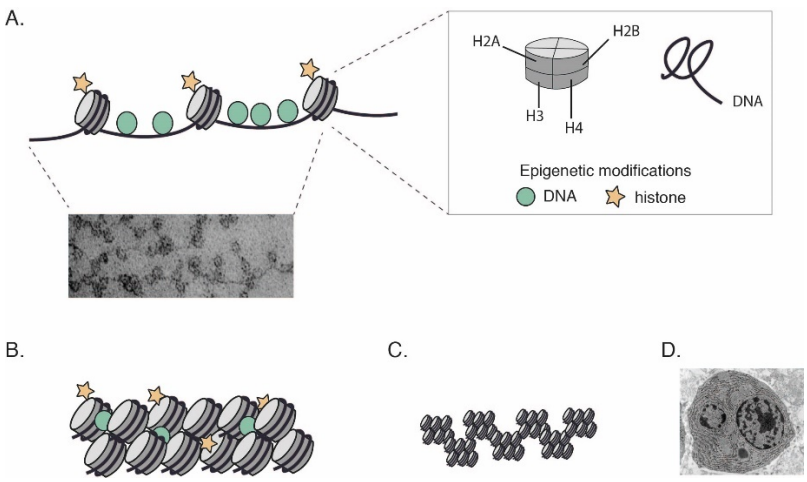


***Figure 1: Chromatin compaction states**. A) Top: the basic unit of the chromatin is the nucleosome composed of DNA (black thread) wrapped around an octamer of histone proteins (grey cylinder). Bottom: observation of the 10nm chromatin fiber conformation under electron microscopy, nicknamed "pearl-necklace" structure (image adapted from (Olins and Olins 2003)). B) Additional chromatin compaction state by the folding of chromatin upon itself. C) A higher degree of compaction is achieved by further folding of the previous structures. D) In the interphase, chromatin takes precise locations inside the nucleus with lighter colored structures representing transcriptional prone regions (euchromatin) vs darker colored regions representing transcriptionally silent regions (heterochromatin) (image adapted from webpathology website of electronic microscopy of a myeloma cell).*

The layout of the chromatin structure is prescribed by the modifications present on the DNA and the associated histones (Bannister and Kouzarides 2011). It is the presence or absence of these marks that ensures correct nucleosome positioning, recruitment of chromatin remodelers, interaction with the replication machinery, and thus, regulation of gene expression and replication in the cell (Li, Carey, and Workman 2007).

1. Histone tail post-translational modifications

Histone proteins constitute the core of the nucleosome structure (**Fig. 1A**). They interact with the negatively charged phosphate groups of the DNA by their basic amino acid side chain. They are organized in a globular domain, with their unique characteristic of the N terminus protruding outside of the nucleosome core (Luger et al. 1997). These so-called histone tails are platforms for different post-translational modifications (PTM) such as acetylation, methylation, phosphorylation or ubiquitination. The presence or absence of these modifications on the histone tails leads to the compaction or loosening of the underlying chromatin structure. Histone PTMs are therefore involved in different cellular processes such as transcription, DNA replication or DNA repair (Kouzarides 2007).

Several histone PTMs have been described, and have been compiled with the use of a defined nomenclature: the modified histone is written first, followed by the amino acid and their position on which the modification is covalently bound, and finally altogether suffixed by the chemical modification itself (Jenuwein and Allis 2001). For instance, “H3K27me3” designates the trimethylation state of the lysine residue at position 27 of histone H3. These modifications are installed and removed by enzymes described as “writers” and “erasers” respectively.

From a gene transcription perspective, histone marks can be classified as activating or repressing based on the consequences they have on gene expression. Among these marks, the acetylation and methylation of lysine residues on histone H3 are the most extensively studied and will be briefly described here (**Table 1**).

- ***Acetylation of histone tails***

The acetylation of lysine residues on histone tails results in the neutralization of their positive charge. Therefore, the interaction between histones and DNA is weakened, which leads to a more accessible state of the DNA (Kouzarides 2007).

Acetylation is catalyzed by histone acetyl-transferase enzymes (HAT). HATs use the acetyl group of acetyl-coenzyme A as a substrate to acetylate the ε-amine of the lysine residue. In contrast, the removal of the acetyl group from the lysine is catalyzed by histone deacetylase (HDAC) enzymes and correlates with transcriptional repression.

These enzymes responsible for histone acetylation turn-over have different family members. For instance, HATs can be divided into two major classes (A and B), with the former further subdivided in three families (GNAT, MYST and CBP/p300) (Sterner and Berger 2000; Bannister and Kouzarides 2011). Even with these multiple actors intervening in the acetylation process, these enzymes do not appear to have specific histones or lysine residues as targets, rather, they have a broad capacity of acetylation. Additionally, these enzymes are usually part of larger multiprotein complexes. HDAC1, for example, was described as a component of the REST repressor complex, responsible for the silencing of neuronal-specific genes in non-neuronal cell types (Huang, Myers, and Dingledine 1999).

It is presumed that after histone acetylation, the more open chromatin state favors the access of transcription factors to their binding sites. Indeed, it was observed that histone tail acetylation coincides with promoter regions (Rando 2007). Recently, it was shown that this modification can also be a consequence of transcriptional activation. Indeed, in *S. cerevisiae,* it was shown that RNA polII has the ability to recruit HATs and to promote their function (Martin et al. 2021). Either way, histone acetylation is associated with transcriptional induction, through the disruption of histone and DNA interactions.

- ***Methylation of histone tails***

Methylation of lysine residues, on the other hand, preserves the positive charge of the histone tail, conserving therefore the interaction of histones and DNA. This mark is set by histone lysine methyltransferase enzymes (HKMT) that use the

methyl group of S-adenosylmethionine as a donor. Lysine can be mono-, di- or tri-methylated. Conversely, these modifications are removed by histone lysine demethylase (HKDM). Compared to HAT/HDAC, these methylation-specific enzymes appear to have precise histone and lysine residues as targets (Kouzarides 2007) (**Table 1**).

The effect of lysine methylation on transcription can be both activating as well as repressing, depending on the position of the lysine on the histone tail that is being modified.

Activating histone marks:

In promoter regions, transcriptional activation coincides with an enrichment of the tri-methylation of the lysine at position 4 of histone 3 (H3K4me3) at the transcriptional start site (TSS) of the gene (Santos-Rosa et al. 2002). Indeed, it was found that the more the gene is transcriptionally active, the more the H3K4me3 peak is high in the promoter region of these genes (Howe et al. 2017). Mounting evidence suggest that this mark is set following transcriptional induction, rather than having a role in the process of initiation itself (Howe et al. 2017). For a subset of genes, however, H3K4me3 seems to have an initiation instructive role by interacting with general transcription factors (Lauberth et al. 2013). Furthermore, H3K4me3 also seems to have a protective role against *de novo* methylation of active TSS by impeding the catalytic activity of the *de novo* DNA methylation enzymes DNMT3/A (see section B.2a1).

Along gene bodies, the deposition of H3K36me3 has been observed. This mark is installed by the SET Domain containing 2 enzyme (SETD2), associated with the transcriptional machinery. It was shown that one of the roles of H3K36me3 is to interact with HDAC and lead therefore to histone deacetylation of the underlying chromatin. This condensed chromatin state would serve as a protective mechanism against illegitimate transcription initiations in these regions (Carrozza et al. 2005).

Repressive histone marks:

Two repressive histone modifications are extensively studied: H3K9me3 and H3K27me3. Methylation of the lysine 9 of H3 (H3K9me1/2/3) is found at

constitutive heterochromatin regions such as centromeres, telomeric repeats or transposable elements. H3K27me3 is found at genomic loci of facultative heterochromatin. H3K27me3 is deposited by the Enhancer of Zeste Homolog 1 and 2 (EZH1 and EZH2). These enzymes are members of the large Polycomb Repressor Complex 2 (PRC2) involved in the silencing of developmental genes, as well as in lineage specific differentiation (Margueron and Reinberg 2011).

Co-occurrence of active and repressive histone marks:

Several promoter and enhancer regions were found to display both activating H3K4me3 and repressing H3K27me3 marks, in particular in embryonic stem cells (Voigt, Tee, and Reinberg 2013). These bivalent marks keep the transcriptional activity in a poised state, i.e., capable of responding quickly to cellular cues that would establish either silencing or activation.

- ***Readers of histone PTM***

Histone tail modifications are instructive to guide the degree of compaction of the chromatin. However, it is not the marks themselves *per se* that drive nucleosome remodeling. Rather, they are viewed as platforms for the recruitment of "reader" proteins, that will interpret the installed histone marks. For instance, H3K9me3 repressive mark can be bound by Heterochromatin Protein 1 (HP1) (**Fig. 2**). Then, HP1 can lead to the recruitment of histone methyltransferases propagating therefore a repressive chromatin context (Schotta et al. 2004). HP1 can also lead to chromatin compaction by oligomerizing with neighboring H3K9me3-bound HP1 (Canzio et al. 2011). Additionally, chromatin remodeling complexes can also be directly recruited by the histone PTMs. They are ATP-dependent complexes that restructure the nucleosomes to serve either chromatin loosening or condensation (de la Serna, Ohkawa, and Imbalzano 2006).

Table 1: Examples of histone tail modifications. *Examples of some writer and eraser enzymes are described (they respectively establish and remove the corresponding histone mark; non-exhaustive list).*

Histone mark	Genomic location	Transcriptional association	Writers	Erasers
H3ac	Promoter/Enhancer	Activation	HAT	HDAC
H3K4me3	Promoter	Activation	SET1, MLL	LSD1
H3K36me3	Gene body	Activation	SETD2	KDM1a
H3K9me3	Promoter/Gene body	Repression	G9A	JMJD2
H3K27me3	Promoter/Enhancer	Repression	EZH2	UTX

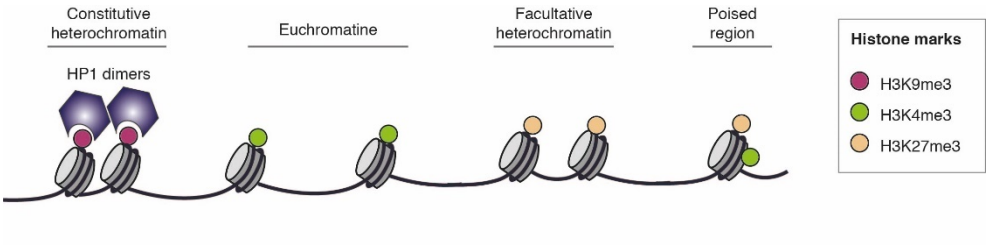


Figure 2: Schematic representation of activating and repressive Histone 3 modifications. *H3K4me3 histone mark is associated with transcriptional activation, while H3K9me3 and H3K27me3 are associated with transcriptional repression.*

To summarize, chromatin compaction is dictated by the underlying code of histone modifications. Epigenetic mechanisms, however, involve another level of control, as these histone tails PTM closely interact with modifications on the DNA, namely DNA methylation.

2. *DNA methylation*

DNA methylation is a chemical modification of the DNA that consists in a methyl group covalently attached to the 5th carbon of the pyrimidine ring of a cytosine (**Fig. 3**). This results in a modified cytosine nucleotide, termed 5-methylcytosine (5mC). This modification occurs predominantly at a cytosine followed by a guanine, also referred to as CpG (cytosine-phosphate-guanine) site. This particular CpG context

renders the mark symmetrical on the DNA strands and allows the faithful maintenance of their methylation state through cell divisions. It is a conserved epigenetic mark across vertebrates, bacteria and plants. However, some common laboratory models are devoid of this mark such as *C. elegans, D. melanogaster,* or *S. pombe.*

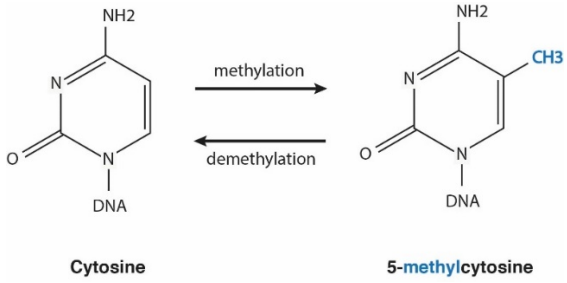


Figure 3: Chemical structures of cytosine and 5-methylcytosine. *The methyl group is highlighted in blue.*

DNA methylation is established by the DNA Methyl-Transferase enzymes (DNMTs) and enzymatically removed by the Ten Eleven Translocation family (TET). DNA methylation is a potent and vital epigenetic mark. Indeed, when deleting DNMT enzymes, embryos die during embryogenesis (*Dnmt1* and *Dnmt3B* knockout (KO) mice models) or after birth (*Dnmt3A* KO mouse model) (Li, Bestor, and Jaenisch 1992; Okano et al. 1999).

Throughout this work, when referring to the 5mC we will use the term DNA methylation out of simplicity.

a) Actors of DNA methylation turn-over

a.1) Writers of DNA methylation

The catalysis of the transfer of a methyl group onto the 5th carbon of the cytosine is performed by the DNMT family of enzymes. These enzymes use S-adenosyl-methionine as the donor of the methyl group and carry out the reaction by their conserved C terminal catalytic domain.

There are mainly two categories of DNMTs that can be distinguished based on the ultimate function of the deposition of the mark: the *de novo* enzymes, DNMT3A and DNMT3B, and the maintenance enzyme, DNMT1 (**Fig. 4**).

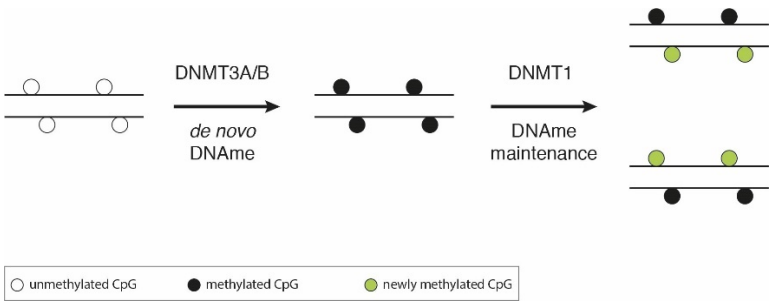


Figure 4: Illustration of de novo DNA methylation and maintenance of DNA methylation *(DNAme). Circles represent CpG dinucleotides. Empty circles correspond to unmethylated CpG and filled colors to methylated CpG.*

DNMT enzymes share multiple conserved protein domains, whose functions will be described below (**Fig. 5**).

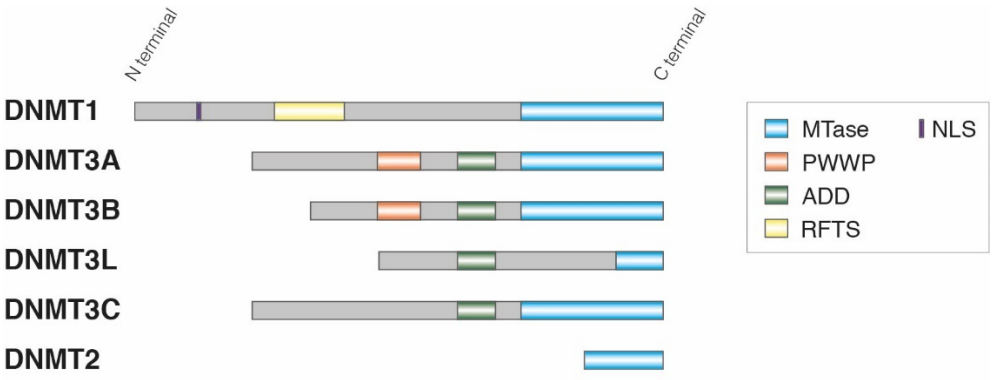


Figure 5: Schematic representation of DNMT protein structures. *Protein domains are represented in color-coded boxes with them being MTase: methyltransferase; PWWP: Pro-Trp-Trp-Pro; ADD: ATRX-DNMT3-DNMT3L; RFTS: Replication Foci Targeting Sequence; NLS: Nuclear Localization Signal.*

De novo DNA methylation:

DNMT3A and DNMT3B are responsible for the establishment of DNA methylation onto the genome during embryogenesis, where the mark was initially absent. They were therefore found to be highly expressed in the developmental embryo, while their expression levels to remain low in adult somatic tissues (Okano et al. 1999).

From a functional point of view, these two enzymes exist in an auto-inhibitory mode (Zhang et al. 2010; Dhayalan et al. 2010). Indeed, the ATRX-DNMT3-DNMT3L (ADD) domain of these proteins binds the neighboring MethylTransferase (MTase) domain, and via this interaction, the catalytic activity of DNMT3A/B cannot occur (**Fig. 6A**). This interaction is impeded when the ADD domain binds an unmethylated H3K4 residue, and therefore, the MTase domain can exert their catalytic activity and methylate the underlying DNA sequence (**Fig. 6B**). However, when the H3K4 residue is methylated, the ADD domain cannot bind, maintaining the inhibitory state of the MTase domain, and leaving therefore, the underlying DNA sequence unmethylated. Transcriptionally active promoters are enriched in H3K4me3, and therefore, appear to exhibit an inverse correlation with their methylation state, i.e., they remain unmethylated. Additionally, DNMT3A/B can also interact with other chromatin modifications (**Fig. 6C**). Their Pro-Trp-Trp-Pro (PWWP) domain was shown to bind H3K36me3, a mark deposited over the transcribed DNA region, implicating these enzymes in the methylation of gene bodies (Dhayalan et al. 2010).

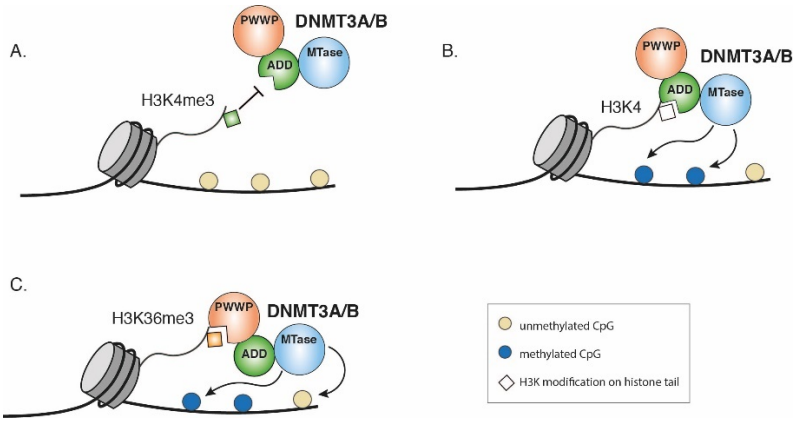


Figure 6: Schematic representation of the mechanisms of de novo DNA methylation performed by DNMT3A and DNMT3B. *A) and B) represent the interaction of DNMT3A/B domains (PWWP, ADD and MTase) in promoter regions, harboring or not a trimethylated H3K4 post-translational modification. C) Same as in A and B, this time in gene bodies.*

An enzymatic cofactor, DNMT3L, is also required for the catalytic activity of DNMT3A/B (**Fig. 5**). DNMT3L lacks the MTase domain and therefore shows no intrinsic catalytic capability. This enzyme has evolved as a cofactor to enhance the reaction speed of DNMT3A and DNMT3B in the germline (Bourc'his et al. 2001).

Recently, another DNMT3 family member has been identified: DNMT3C (**Fig. 5**). DNMT3C is only present in rodents and originates from a duplication of DNMT3B (Barau et al. 2016; Jain et al. 2017). Their function is specific to the male germline of rodents, where it was shown to be required to methylate evolutionary young retrotransposons (constituting about 1% of the mouse genome). DNMT3C is absent from female rodent counterpart cells.

Maintenance of DNA methylation:

Once DNA methylation has been set, the mark needs to be propagated throughout cell divisions to ensure a correct and stable cellular gene expression program. This process is catalyzed by the DNA methylation maintenance enzyme DNMT1. This enzyme is ubiquitously expressed, and their expression reaches their highest level during the S phase of the cell cycle. After DNA replication, the resulting DNA molecule is composed of the methylated template strand and of the newly synthesized daughter strand. At this point, the DNA is qualified as hemi-methylated DNA (**Fig. 7**). DNMT1 has a high affinity for hemi-methylated strands (Hermann, Goyal, and Jeltsch 2004). This implies that only symmetrical DNA methylation can be faithfully maintained during DNA replication. As in the case of DNMT3A/B, DNMT1 also exists in an auto-inhibitory state. Here, their Replication Foci-Targeting Sequence (RFTS) domain binds the catalytic MTase domain and therefore compromises the methylation catalysis capacity of the enzyme (in the same manner as for ADD domains of DNMT3A/B described above) (Song et al. 2011).

Maintenance of DNA methylation during DNA replication also requires the Ubiquitin like with PHD and Ring Finger domains 1 (UHRF1), a partner of DNMT1. It binds the hemi-methylated CpG site via its SET- and RING-associated domain (SRA). UHRF1 also has a Ubiquitin-like domain (UBL) that recruits DNMT1 at the replication foci, breaking the RFTS-MTase interaction and allowing therefore the methylation

of the initially unmethylated daughter strand, re-establishing the methylation symmetry on DNA strands (**Fig. 7**).

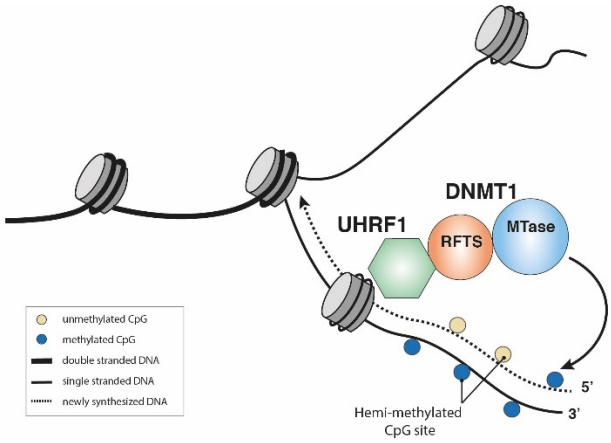


Figure 7: Faithful maintenance of CpG methylation by DNMT1 during DNA replication. *UHRF1 recruits DNMT1 to the replication foci, breaking the inhibitory interaction between the Replication Foci-Targeting Sequence (RFTS) domain and the catalytic MethylTransferase (MTase) domain of DNMT1, leading to the methylation of CpG sites on the newly synthesized DNA strand.*

Finally, there is also an additional enzyme that shows sequence homology with the previously cited DNMTs and was therefore named DNMT2 (**Fig. 5**). However, about a decade ago, it has been shown that this enzyme is a tRNA methyltransferase enzyme, instead of a DNA methyltransferase. Specifically, it catalyzes the methylation of an aspartic residue at the anticodon loop of tRNAs (Goll et al. 2006).

a.2) Erasers of DNA methylation

DNA methylation marks are described as stable during cell divisions. However, in some precise genomic contexts, DNA methylation has to be removed. For instance, during development, this mark is erased in order for the developing embryo to establish their own DNA methylation landscape. Indeed, it is thought that this process would limit the inheritance of altered parental epi-alleles. Additionally, in a more targeted way, the erasure of DNA methylation was proposed to limit the improper deposition of this mark by the DNMTs (Williams, Christensen, and Helin 2011). TETs would therefore act as “protectors” of faulty DNA methylated sequences. Indeed, it was reported that TET1 can bind several CpG rich promoters

in mouse Embryonic Stem Cells (mESC), and that this binding correlates with the presence of 5hmC (5-hydroxymethylcytosine) in these sites, as well as with their unmethylated CpG state. Other studies have reported that it is not promoter regions that are the main targets of TET in mESC, rather, enhancer regions, since *Tet1/2/3* triple KO results in their hypermethylation (Rasmussen and Helin 2016).

Three mechanisms of DNA methylation erasure have been described: one of them is dependent on the process of DNA replication and involves DNMT1, while the other two implicate the TET family of enzymes.

Passive erasure of DNA methylation:

Passive demethylation of DNA consists in blocking the access of DNMT1 to the replication fork. In this erasure process, DNMT1 can no longer exert their DNA methylation maintenance activity on the daughter DNA strand during genome replication. This phenomenon occurs for instance in embryogenesis where DNMT1 and UHRF1 are excluded from the zygote cell nucleus leading to a progressive global DNA methylation loss up to the blastocyst stage (Cardoso and Leonhardt 1999).

Active erasure of DNA methylation:

Active DNA demethylation is independent of DNA replication and implicates specific DNA demethylation enzymes of the TET dioxygenase protein family (Tahiliani et al. 2009). The successful removal of the methyl residue from the methylated cytosine consists in a series of oxidation reactions of the methyl group by TET enzymes followed by base excision repair mechanisms (**Fig. 8**).

Three members of this family have been described: TET1, TET2 and TET3. From a protein structural view, TET family members share a conserved catalytic domain at their C terminus. This domain is composed of a cysteine-rich domain and a double-stranded-β-helix domain (DSBH) containing binding sites for their cofactors, Fe^{2+} and 2-oxoglutarate (2-OG). Additionally, TET1 and TET3 also share an N-terminal CxxC zinc finger domain that allows their binding to the DNA sequence (**Fig. 9**).

TET enzymes convert 5mC to 5hmC using O_2, Fe^{2+} and 2-OG as cofactors. Successive oxidations convert 5hmC to 5-formyl-cytosine and 5-carboxy-cytosine (**Fig. 8**) (Ito et al. 2011).

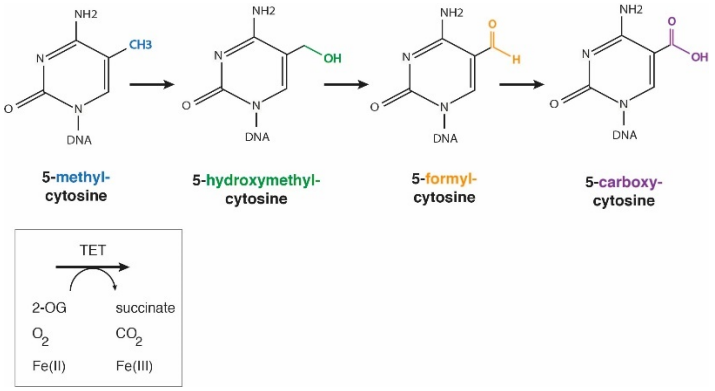


Figure 8: Oxidation of 5-methylcytosine and their derivatives by TET demethylating enzymes. *Top: chemical structures of the oxidation intermediates of 5mC by TET catalysis. Bottom: catalytic mechanism of TET enzymes. TET use 2-oxoglutarate (2-OG) and Fe^{2+} as cofactors for the oxidation reactions.*

After sequential oxidations, the modified cytosine residue is recognized by the Thymine DNA Glycosylase (TDG) enzyme. This recognition leads to the excision of the oxidized cytosine from the DNA strand. The abasic site is repaired by the Base Excision Repair (BER) mechanism that successfully reestablishes a cytosine residue, resulting overall to the demethylation of the initially methylated cytosine (Zhang et al. 2012). Another mechanism that has been proposed for the active demethylation of 5mC by TETs, is the deamination of 5hmC by the Activation-Induced Deaminase/Apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like proteins (AID/APOBEC). This deamination process results in a 5-hydroxyuracil nucleotide, leading consequently to a nucleotide mismatch with the opposite guanine. This is once more recognized by the TDG enzyme, and the BER mechanism subsequently restores an unmethylated cytosine (Guo et al. 2011). However, this process was later contested by studies where no deaminase activity of AID/APOBEC could be observed *in vitro* (Nabel et al. 2012).

Of note, it is assumed that 5mC oxidative derivates could also facilitate passive DNA demethylation, since DNMT1 cannot efficiently methylate a strand where their template was 5hmC, instead of 5mC (Wu and Zhang 2014).

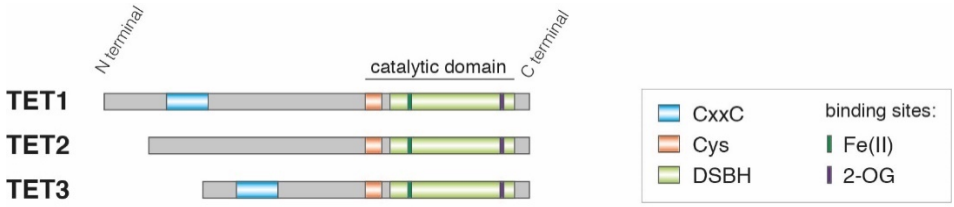


Figure 9: Schematic representation of TET protein structures. *Colour-coded boxes represent protein domains: a N terminal cytosine-x-x-cytosine (CxxC) zinc finger domain binding the DNA, a C terminal conserved catalytic domain comprising a cysteine-rich domain (Cys) and a Double-stranded-β-helix (DSBH) domain containing the binding sites of TET cofactor (iron and 2-oxoglutarate, Fe(II) and 2-OG respectively).*

b) Establishment of DNA methylation

Although a small chemical DNA modification, the presence of the methyl group on CpGs is crucial for proper development. Altering DNA methylation profiles during embryogenesis leads to lethality (**Table 2**), either during gestation or following birth as observed by KO experiments in mouse models (Okano et al. 1999; Li, Bestor, and Jaenisch 1992).

Table 2: Knockout mice phenotype after deletion of writers of DNA methylation and UHRF1 partner protein. *DNAme: DNA methylation.*

Protein	Function	KO phenotype in mice	Reference
DNMT1	DNAme maintenance	Early embryonic lethality	*Li et al. 1992*
UHRF1	DNMT1 cofactor	Early embryonic lethality	*Maenohara et al. 2017*
DNMT3A	*de novo* DNAme	Lethality after birth	*Okano et al. 1999*
DNMT3B	*de novo* DNAme	Lethality at mid-gestation	*Okano et al. 1999*

DNA methylation patterns are established during embryogenesis and are maintained throughout life. During development, there is a dynamic reshaping of the methylome of both pronuclei of sperm and oocyte, and later of the embryo for

the correct installment of global DNA methylation profiles in somatic tissues and germline (**Fig. 10**). Development is essentially marked by two waves of global DNA methylation erasures occurring in these cells. The first wave takes place between the zygote and the blastocyst stage, while the second occurs during the migration of premature germ cells (primordial germ cells) to the future gonads. It is presumed that the reason these methylation rewirings take place is to establish pluripotency of the cells after fertilization, as well as limit the inheritance of altered DNA epi-alleles.

b.1) Methylation dynamics in the embryo

The oocyte and sperm have their own level of DNA methylation, the sperm being highly methylated (+/- 80%) compared to the oocyte (+/- 40%) (Molaro et al. 2011; Smallwood et al. 2011). After fertilization, the male pronucleus loses DNA methylation, prior to the first genome replication (one-cell stage in the mouse). Studies have shown that this is achieved predominantly by an active process of DNA demethylation mediated by TET3 (Iqbal et al. 2011). In parallel, the oocyte also exhibits loss of DNA methylation, this time progressively by passive demethylation resulting from the eviction of DNMT1 from the nucleus (Cardoso and Leonhardt 1999). This passive demethylation process of the zygote continues until the blastocyst stage where the global methylation level reaches their lowest point. Imprinted genes escape this demethylation wave and maintain therefore their inherited parental-specific methylation pattern (Wang et al. 2014). Some retrotransposable elements were found to resist this process as well (Lane et al. 2003). Subsequently, once the blastocyst is implanted on the uterine wall, there is a progressive remethylation of the genome dependent on the *de novo* DNMT3A/B that occurs up to the epiblast stage, prior to somatic tissue differentiation. This allows the establishment of the definitive methylome levels of the embryo in their somatic tissues.

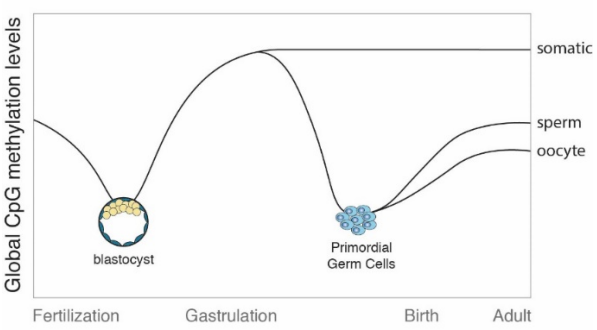


Figure 10: Schematic representation of the dynamic methylome reshaping during development.

b.2) Methylation dynamics in the developing germ cells

At the epiblast stage, a group of cells, the Primordial Germ Cells (PGC), emerge and are directed towards germ cell specification. PGCs extensively proliferate and migrate to the genital ridges, the future gonads. During this process, the second wave of DNA demethylation occurs (**Fig. 10**). This results in the removal of methylation patterns transmitted from the epiblast. Contrary to the first wave, imprinted genes are subjected to this demethylation process in order to erase parental asymmetry of the methylome and establish new methylation patterns based on the embryo's gender. Global demethylation is mainly due to a passive mechanism, with DNMT1 and UHRF1 being excluded from the nucleus. TET1 and TET2 proteins appear to be involved in this process as well, since they are highly expressed in PGCs and the latter display detectable levels of 5hmC (Hackett et al. 2013). Subsequently, genome remethylation takes place for the establishment of methylome patterns specific to male or female gametes. Male gametes are extensively methylated prior to birth, reaching their definitive methylation as in somatic tissues. This is achieved by the *de novo* activity of DNMT3A and their cofactor DNMT3L, as well as DNMT3C. In comparison, in the oocyte, methylation level remains low and is acquired after meiosis and prior to ovulation through the action of DNMT3A and DNMT3L.

c) Distribution of DNA methylation

Genome-wide analyses of methylomes have shown that DNA methylation is a widespread DNA modification of the genome. Indeed, on average, 80% of CpGs are methylated in somatic cells of mammals and this proportion is stable between cell types (Lister et al. 2009).

The distribution of CpG sites onto the genome is neither random nor even. The genome is generally depleted of CpGs (1/80) compared to the expected theoretical CpG frequency (1/16). Indeed, the methylation state of a cytosine renders this nucleotide unstable, and a target of mutation by spontaneous deamination events in the cell (Bird 1980). Nevertheless, there are genomic regions that exhibit the expected density of CpG sites (Bird 1980), and are therefore resistant to this evolutionary erosion. These regions were termed CpG islands (CGI) and are defined as regions of about 500bp, with a high density of CpGs, and a C+G base composition ratio of 0.6. The existence of these conserved frequency of CpGs relies on the fact that they are maintained methylation-free in germline cells, ensuring therefore their stable transmission to the offspring.

There are about 25.000 CGIs in the human genome, and 2/3 of them coincide with promoter regions (Saxonov, Berg, and Brutlag 2006). CGIs are also found in gene bodies and intergenic regions. Several CGIs in these two latter categories are not annotated as transcriptional start sites (TSS), but nevertheless, show evidence of such activity (Illingworth et al. 2010). These sequences were termed orphan CGIs (**Fig. 11**). These data therefore suggest a close relationship between CGIs and transcription.

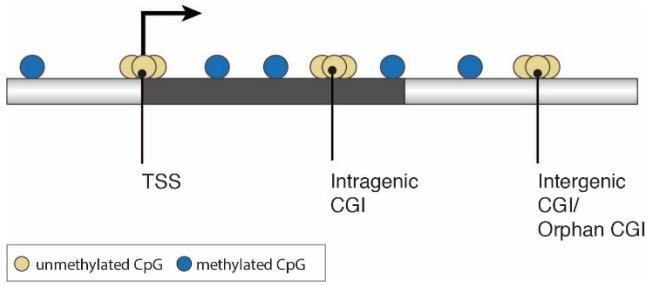


Figure 11: Genomic distribution of CpG islands. *The grey box represents a gene locus, with the broken arrow indicating the transcriptional start site (TSS); CGI: CpG island.*

d) DNA methylation is associated with transcriptional repression

It has long been established that DNA methylation exerts stable and potent repression of transcription (Jones and Takai 2001). However, the mechanisms by which the repression is achieved are not fully understood. It has been shown that DNA methylation can act through two main modes of actions to mediate transcriptional silencing: restraining access of activating transcription factors, and recruiting of chromatin modifying complexes (Klose and Bird 2006). First, the addition of the methyl-group onto the CpG site can physically impair the access of transcription factors to their genomic binding sites (**Fig. 12A and B**). Of note, a recent study has found that the CpG methylation can also increase the affinity of some transcription factors to their recognition site when methylated (Yin et al. 2017). Secondly, DNA methylation can be recognized by Methyl-Binding Proteins (MBP). MBP have been classified into three families: MBD, SRA and Zinc fingers (Buck-Koehntop and Defossez 2013). For instance, UHRF1 partner of DNMT1 in the DNA methylation maintenance process, (**Fig. 7**) is a member of the SRA family. These MBP have the ability to interact with chromatin remodeling complexes as well as with histone modifying enzymes (**Fig. 12C**). Through the recruitment of MBP to methylated genomic loci, there is consequently an overall nucleosome remodeling of the region leading to a closed chromatin state (Hendrich and Bird 1998; Nan et al. 1998). Together, the interaction of methylated cytosines and chromatin components conveys the transcriptional silencing of specific genomic regions.

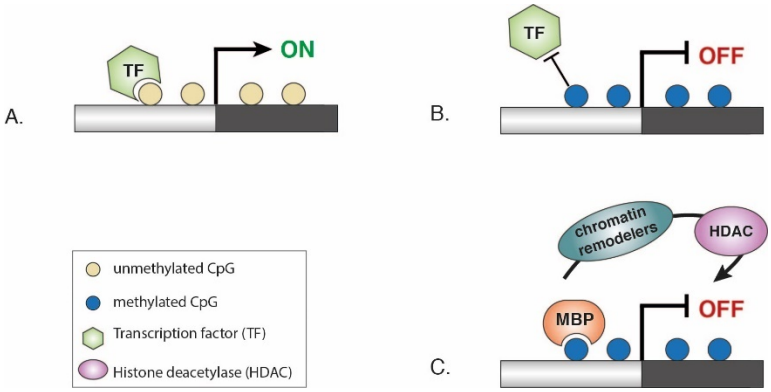


Figure 12: DNA methylation is associated with transcriptional repression. *A) CpG sites in the promoter region are unmethylated, allowing therefore the transcription factors to recognize their binding motif which results in transcriptional activation of the underlying gene. When the promoter*

region is methylated, transcriptional repression occurs through two mechanisms: methylation of the binding motif of a transcription factor impedes their binding to the promoter region (B), or methylated CpGs are recognized by methyl-binding domain proteins (MBP) that can recruit chromatin remodelers and histone deacetylase which results in a refractory state of transcription.

e) The roles of DNA methylation silencing in somatic tissues

As stated above, DNA methylation is found in different genomic contexts, including promoter regions and gene bodies, with the deposition of this mark being associated with transcriptional repression. Initially, it was thought that DNA methylation constituted the essential mechanism of tissue-specific gene silencing in cells. However, evidence for this regulation remained scarce (Bestor, Edwards, and Boulard 2015). Indeed, it was observed that many tissue-specific gene promoters are CpG rich regions (Saxonov, Berg, and Brutlag 2006; Weber et al. 2007), and remain unmethylated even when silent, excluding therefore a primordial role of DNA methylation in their stable silencing (**Fig. 13**).

Nevertheless, DNA methylation appears to be the essential component of transcriptional repression for a subset of genomic sequences, as will be described below.

e.1) Gene promoters

Imprinted genes:

There are particular genomic loci that exhibit differential methylation of their alleles based on their parental origin. This results in the monoallelic expression of the two gene copies (Li, Beard, and Jaenisch 1993). These genomic regions are known as imprinted genes and, to date, about a hundred have been reported in humans. Methylation marks on these genes are acquired in a gender-specific manner during either early oogenesis or spermatogenesis. After fertilization, DNA methylation on these alleles is stably maintained through cell divisions in somatic tissues during embryogenesis, as well as after birth. Indeed, imprinted genes escape the first wave of DNA demethylation in the early embryo. However, their methylation marks are erased during the second wave of DNA methylation

occurring in primordial germ cells during their migration to the genital ridges. This event allows the establishment of new DNA methylation marks on the allele based on the gender of the embryo. Mounting evidence in mouse models suggest that these genes are involved in a wide variety of functions such as the regulation of metabolism, body temperature and feeding (Glaser et al. 2022; Peters 2014). Indeed, alterations in the imprinted alleles result in imprinted disorders such as the Prader-Willi syndrome. This syndrome consists of a range of symptoms such as cognitive impairment, developmental delay, and obesity among others. It is caused by the loss of expression of the paternal loci located on 15q11-13.

Transposable elements:

DNA methylation is also involved in the stable repression of mobile genomic elements. Indeed, approximately half of the mammalian genome has been invaded by transposable elements (TE), the vast majority of them being retrotransposons. In mouse models, it was shown that young retrotransposons conserved the ability of their genomic mobility, and these insertions can cause up to 10% of genomic mutations, threatening therefore genetic integrity (Maksakova et al. 2006). DNA methylation appears as a potent repressive mechanism of these TEs. Studies have shown that in *Dnmt1 -/-* mouse models, the loss of genome-wide DNA methylation leads to a dramatic induction of these elements resulting in lethality at embryonic day 9.5 (Walsh, Chaillet, and Bestor 1998). It is hypothesized that DNA methylation has evolved to specifically control TE activation, and therefore serves primarily as a host defense mechanism.

Germline-specific genes:

Promoter DNA methylation also appears as an essential regulatory mechanism to repress a subset of germline genes in somatic tissues. The promoter of these genes is heavily methylated in somatic cells, and demethylated in germ cells. The group of germline genes using this DNA methylation-dependent mode of regulation is of great importance to our work and will therefore be extensively described in a later subchapter (see section C.3).

e.2) Gene bodies

Despite the mutagenic property of DNA methylation, this epigenetic mark is also found in gene bodies, which appear heavily methylated. Indeed, they were described majorly as CpG-poor regions, harboring repetitive elements and transposons (Jones 2012). Methylation at these sites would therefore operate as a repressor of the transcription of these mobile genomic elements as described above (Yoder, Walsh, and Bestor 1997). Some gene bodies, however, also display CpG-rich regions. These regions are not associated with any annotated genomic feature and are qualified "orphan CGIs" as described above. They display promoter features, and therefore were proposed to act as cryptic promoters that would be silenced by DNA methylation (Illingworth et al. 2010). Recently, by studying severely hypomethylated mouse embryos after a *Dnmt1* KO or *Dnmt3a/b* double KO, the contribution of DNA methylation in the repression of this illegitimate transcription was confirmed (Dahlet et al. 2020). Of note, the methylation of gene bodies does not appear to impede transcriptional elongation of upstream initiated transcripts. Rather, it appears to block transcriptional initiation (Jones 1999). Therefore, in gene bodies there is an inverse correlation between CpG methylation and repression, compared to what is observed in promoter regions. Finally, another role attributed to DNA methylation in gene bodies is the regulation of alternative splicing. In particular, an enrichment of DNA methylation is observed in certain gene exons compared to their neighboring introns, which are associated with exon inclusion during transcription (Gelfman and Ast 2013).

e.3) Chromosome X inactivation

Female mammals display two X chromosomes (chrX) compared to males, that harbor only one. To compensate for this imbalance of genes, female somatic cells randomly inactivate one of the two chrX (Lyon 1961). This results in a mosaic phenotype in females, with cells inactivating either the paternal or the maternal chrX. The process of chrX inactivation is initiated by the expression of a long non-coding RNA (lncRNA) *XIST*, transcribed from the chrX that will be inactivated (Brockdorff et al. 1991). *XIST* transcripts gradually coat the whole chrX, and this subsequently results in a large chromatin remodeling of the coated chromosome. Activating histone marks such as H3K4me3 and histone acetylation are removed,

while repressive histone marks such as H3K27me3 and H2AK119Ub are deposited (Zylicz et al. 2019). Finally, DNA methylation, including at CGI promoters, comes into play in the final stages of the inactivation, as an additional and indispensable layer of stable gene repression (Riggs 1975; Disteche and Berletch 2015). Overall, these successive chromatin modifications lead to a highly condensed state of the chrX (Barr and Bertram 1949).

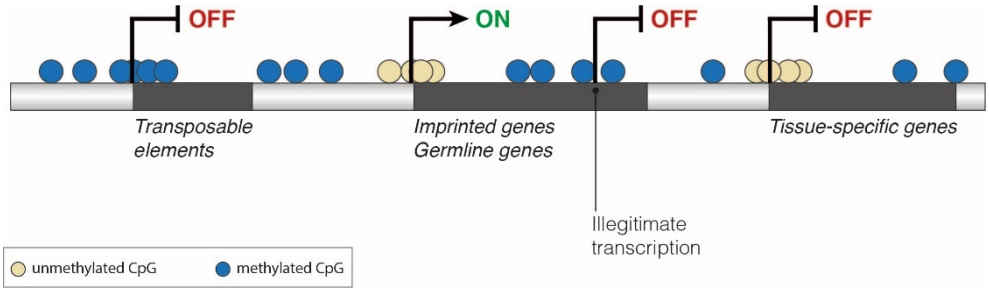


Figure 13: Functions of DNA methylation in transcriptional silencing. *Broken arrows indicate transcriptional start sites. Black boxes delineate gene units. Colored circles indicate the methylation state of a CpG site. Transcriptional activation or repression is indicated by "ON" and "OFF" respectively.*

f) Analyzing DNA methylation

To study DNA methylation, the challenge is to precisely distinguish the methylated from the unmethylated state of the cytosine residue in a particular CpG site. The gold standard technique used in the field is the sodium bisulfite treatment of genomic DNA. The sodium bisulfite will convert an unmethylated cytosine to uracil by a series of successive reactions. Then, by PCR amplification, the uracil will be replaced by a thymine. This conversion is blocked by the methylation state of the cytosine, leaving therefore 5mC unmodified (**Fig. 14**).

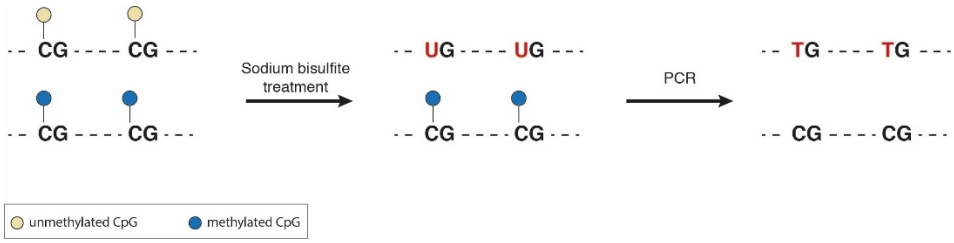


Figure 14: Schematic representation of cytosine conversion to thymine after sodium bisulfite treatment.

Subsequently, targeting methods or genome-wide methods can be applied to investigate the cytosine methylation state of a region of interest. For example, methylation-specific qPCR (MS-qPCR) can be used to study a small subset of genes. It is based on the design of two primer couples, one recognizing the unmethylated region of interest (i.e., all cytosines would be converted to thymines), while the other would recognize the methylated region of interest (cytosines remain cytosines). Amplification reactions are carried out by both primer couples. Finally, the ratio of methylation vs unmethylation percentage of the region of interest is computed.

To investigate the state of cytosine methylation in a genome-wide way, multiple methods are available. Here, we review the two methods that were used in the present work.

Infinium methylation 450K:

This assay is based on an array technology used to study 450K genomic sites such as promoter regions, gene bodies, or intergenic regions. It has become a cost-effective technique nowadays, commonly used in methylomic analyses. The array consists of two probes for each interrogated CpG site: one probe is designed to recognize the methylated state of the cytosine of interest, while the other probe recognizes the unmethylated cytosine. After the treatment of the DNA with sodium bisulfite, the DNA is denatured and hybridized onto the array. Then a reaction of single-base extension will occur efficiently if the hybridization of the cytosine state with the corresponding probe is achieved. Nucleotides used for the extension are coupled to fluorochromes therefore resulting in two colors for each CpG site,

corresponding to the hybridization of the methylated and unmethylated state of each region. Finally, as in the case of MS-PCR, a ratio of methylated vs unmethylated signals obtained from the analysis is computed. A limited amount of CpG sites can therefore be studied using this technique.

Whole Genome Bisulfite Sequencing (WGBS):

After bisulfite treatment, DNA can also be subjected to sequencing. This results in a single CpG resolution of the methylation ratio of reads that were found to be methylated or unmethylated genome-widely for a sample.

Data generated by laboratories are deposited in publicly available platforms such as GEO from NCBI, and can therefore be exploited by researchers. In the present work, we combined multiple methylomic datasets of such repositories.

The main drawback of the use of sodium bisulfite for the analysis of methylation states is that it leads to DNA degradation. Large DNA quantities of starting material are therefore needed, which can be cumbersome, for example in single cell methylation analyses. New techniques are currently being developed to resolve this, such as the TET-assisted pyridine sequencing (Liu et al. 2019), that uses TET DNA demethylation and the conversion of oxidized cytosines to thymine nucleotides, which does not alter DNA.

C. Alterations of DNA methylation in tumors

After the erasure of the parental methylome during development and the re-establishment of DNA methylation patterns in somatic and germ cells of the embryo, the methylation landscape of normal cells remains stable throughout cell divisions. However, this stability is profoundly altered in disease, in particular, in cancer. As a matter of fact, it is clearly established today that cancer is as a genetic, as well as an epigenetic disease. As mentioned in part A, mutations are found in

genes implicated in different epigenetic regulation mechanisms including factors involved in DNA methylation, histone modifications, and chromatin organization. In the present work, we focus on epigenetic alterations affecting DNA methylation (**Fig. 15**).

The first epigenetic alteration observed in tumors, was the aberrant DNA methylation levels in these cells. Indeed, it was found that tumors exhibited a global reduction of their overall DNA methylation content compared to their normal counterpart (Feinberg and Vogelstein 1983a; Gama-Sosa et al. 1983a). This phenomenon was termed DNA **hypo-**methylation. At the same time, tumor cells were also found to exhibit local gains of DNA methylation that affected CGIs genomic regions (Baylin et al. 1986). These events were termed DNA **hyper-**methylation events. Both hypo- and hyper-methylation coexist in the same tumor cell, and are present in virtually all tumor types, although to different extents. Therefore, tumor methylomes, because of these epigenetic phenomena are extensively altered. However, it is not clear today which events are "driver-methylation" events or "passenger-methylation" events in tumors, i.e., impact or not tumor progression. Furthermore, the mechanisms leading to these two epigenetic alterations remain largely unresolved today.

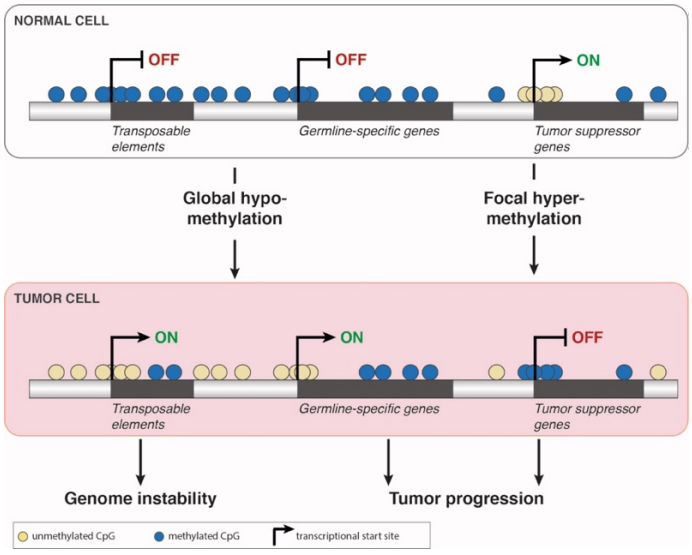


Figure 15: Consequences of genome-wide DNA hypomethylation and focal DNA hypermethylation in tumor cells compared to their normal counterparts. *Broken arrows indicate transcriptional start sites of genes. Transcriptional activation or repression is indicated by "ON" and "OFF" respectively.*

1. *Focal DNA hypermethylation*

Gains of DNA methylation in tumors are found in genomic regions that were initially unmethylated in the normal tissue, i.e., CGIs. Since these CGIs are discrete genomic regions, these hypermethylation gains were described as local. Some tumor types appear to display a higher degree of these events, a phenotype termed CpG island methylator phenotype (Weisenberger et al. 2006). As stated in part B, DNA methylation is associated with gene repression, and CGIs are frequently found in promoter regions. It was therefore hypothesized that the hypermethylation events of CGIs would be involved in tumor suppressor gene silencing (**Fig. 15**) (Esteller 2002). Indeed, studies have shown that multiple genes involved in cell proliferation (*CDKN2A*), DNA repair (*BRCA1, MLH1*), or in hypoxia-regulated pathway (*VHL*) are targets of *de novo* DNA methylation in tumor cells (Catteau et al. 1999; Herman et al. 1994; Kane et al. 1997; Merlo et al. 1995). These observations indicate that in addition to genetic alterations occurring at these loci, DNA hypermethylation is an alternative mechanism of their silencing. It was indeed observed, that CGI hypermethylation can provide the second 'hit' of tumor suppressor inactivation (Esteller 2007).

Mechanisms leading to CGIs DNA hypermethylation:

The exact causes that result in focal DNA hypermethylation of CGIs in tumor cells remain largely unknown. Emerging evidence suggest that they involve a reduction of the TET DNA demethylase activity in these cells (Wu and Zhang 2017). For instance, in hematological malignancies, TET2 inactivating mutations were found in up to 30% of acute myeloid leukemia patients (Scourzic, Mouly, and Bernard 2015). As mentioned previously, TET enzymes are dependent on 2-OG and oxygen for the successive oxidation reactions of the methylated cytosine. Indeed, it was shown that the availability of these TET cofactors can be altered in tumor cells, resulting in the decrease of TET enzymatic activity. For instance, in hematological malignancies, missense mutations of the isocitrate dehydrogenases, the enzymes that generate 2-OG, lead to the conversion of 2-OG to 2-hydroxyglutarate. It was shown that this metabolite impedes TET activity subsequently leading to DNA hypermethylation. Contrastingly, it appears that solid tumors, which harbor very few inactivating

mutations of TET, display a reduction in TET activity due to the hypoxic tumor microenvironment. Indeed, for their proper enzymatic activity, TET enzymes need a precise oxygen concentration level. As tumor cells develop, and the blood supply cannot follow at the same pace, oxygen levels drops, resulting in a diminished TET activity and hypermethylation of gene promoters (Thienpont et al. 2016).

Intriguingly, it was also observed that CGI hypermethylation targets gene promoters that were initially silent in the cell from which the tumor originated. It was found that the promoter regions susceptible to these hypermethylation events, were "pre-marked" with H3K27me3, or were in a poised state, exhibiting both H3K27me3 and H3K4me3 histone modifications (Schlesinger et al. 2007). It is presumed that this event would serve as a stable silencing mechanism that would impair the induction of the initially silent genes under certain environmental cues (Ehrlich 2019). However, the mechanisms connecting these two epigenetic modifications, DNA methylation and H3K27me3, in these promoter regions remain undetermined.

2. *Genome-wide DNA hypomethylation*

Tumor cells also exhibit DNA hypomethylation patterns compared to their normal tissue of origin (Gama-Sosa et al. 1983a; Feinberg and Vogelstein 1983a). Since DNA methylation regulates repeated elements that constitute up to 50% of the human genome, extensive losses of DNA methylation are found in these regions (**Fig. 15**) (Rauch et al. 2008). These losses were therefore qualified as genome-wide. Indeed, the methylation level of these genomic elements are often used as a mirror of the overall DNA methylation changes in tumor cells.

DNA hypomethylation occurs virtually in all tumors, although to different extents. Tumor tissues can show a reduction up to 60% of their DNA methylation content compared to their normal counterparts (Ehrlich 2002). For instance, melanoma and lung cancer were found to exhibit profound losses of DNA methylation, whereas in hematological malignancies, these losses were modest (Ehrlich 2009a).

Compared to CGI DNA hypermethylation which was extensively studied in tumor cells, the consequences of DNA hypomethylation remain largely undetermined. A widely recognized effect of DNA hypomethylation in tumors is the promotion of genomic instability (Gaudet et al. 2003; Schulz et al. 2002), where losses of methylation were found to be associated with extensive chromosomal rearrangements. Another presumed consequence of DNA methylation losses is the activation of mobile genomic elements. However, such evidence remains scarce due to the difficulties of studying repetitive sequences, and only beginning to emerge with the rise of new technologies, invoking a role of retrotransposons as active promoter regions, or reporting their insertion in tumor suppressor genes, such as *APC* (Cajuso et al. 2019; Jiang and Upton 2019). Finally, it was long presumed that DNA hypomethylation in tumor cells would lead to the massive aberrant activation of genes, and in particular oncogenes. However, evidence for such a scenario remains scarce, since most tissue-specific genes rely on other mechanisms than DNA methylation for their regulation. DNA hypomethylation does however, associate with the ectopic activation of cancer-germline (CG) genes in tumors. Several of these genes appear to exert pro-tumoral functions, implying a promoting role of DNA hypomethylation in tumorigenesis via the aberrant activation of CG genes. This will be further discussed in section C3 (**Fig. 15**).

Mechanisms leading to global DNA hypomethylation:

DNA hypomethylation was discovered about 40 years ago (Feinberg and Vogelstein 1983a; Gama-Sosa et al. 1983a). However, still today, the exact mechanisms resulting in this extensive DNA methylation loss remain undetermined. This is partly due to the fact that DNA demethylation mainly targets repetitive genomic sequences and seems to implicate a transient DNA demethylation initiator event (Loriot et al. 2006). Multiple hypotheses have been proposed as triggering events of DNA hypomethylation (De Smet and Loriot 2010). One possibility frequently debated in the field is the combination of a high proliferative rate of tumor cells, and the maintenance fidelity of DNMT1. Studies from Tanay and Berman groups have found that DNA methylation is gradually lost in late-replicating domains (Zhou et al. 2018; Meir et al. 2020). Another hypothesis suggests that low levels of DNA methylation in tumor cells would reflect a pre-malignant senescent state of these

cells. Indeed, it was found that cells reaching replicative senescence show a decrease in global DNA methylation levels, in parallel with a decreased activity of DNMT1. Cells that would have mutated p53 to bypass this checkpoint, would continue to proliferate maintaining a lower genomic methylation content. Another possibility would be the depletion of S-adenosylmethionine, the donor of methyl groups, in tumor cells, leading to a decrease in DNMT activity. Finally, it was recently shown that in a model of T cell lymphoma, TET depletion results in the paradoxal hypomethylation of the heterochromatic compartment of the nucleus (Lopez-Moyado et al. 2019). This was associated with a displacement of DNMT3A from the heterochromatin to the euchromatin compartment.

Of note, DNA hyper- and hypo-methylation have long been considered unrelated to one another from a mechanistic point of view. However, a recent study of our group revealed that they can be linked through the activation of a CG lncRNA (Fain et al. 2021) (see section C.3.d).

Overall, for both hyper- and hypo-methylation events in tumor cells, the mechanisms that initiate these two common phenomena, remain far from being completely elucidated. It is possible that there is no unique mechanism explaining all these alterations but rather, a combination of events happening in different cell states of the tumor cells prior to them reaching their complete malignancy potential.

Our group aims to understand the consequences of DNA hypomethylation in tumor cells. To do so, our workhorse is the group of tissue-specific genes of the germline lineage.

3. The strange case of cancer-germline genes

CG genes are a group of germline-specific genes found ectopically activated in tumor cells of a variety of histological origins (for a detailed definition, refer to section C.3.b below). About a hundred of genes presenting this restricted expression profile in germ cells and tumor cells have been described to date as belonging to the CG family. CG genes are compiled in a public database established by the Ludwig Institute for Cancer Research with the aim of providing a reference

list and characterization for further study of CG genes (Almeida et al. 2009). Intriguingly, the majority of CG genes are localized on chrX. It has been shown that these genes are often expressed in the spermatogonial stage of spermatogenesis, while the non-X-CG genes were found mostly to be expressed during meiosis (Simpson et al. 2005).

In tumor cells, CG genes exhibit a variable level and activation frequency depending on the tumor type. For instance, they are frequently found activated in lung cancer or melanoma, whereas weakly activated in liquid cancers (Scanlan et al. 2002; Sahin et al. 1998). Additionally, CG genes often appear to be co-activated in tumor cells. The frequency of this co-activation suggests a common regulatory mechanism of transcriptional regulation. Indeed, for a subset of CG genes, it was shown that promoter DNA methylation was found to be their primary regulatory mechanism (De Smet, Lurquin, Lethe, et al. 1999) (see section C.3.e).

a) A brief historical perspective on the discovery of CG genes

The first human tumor antigen was isolated in 1991 in our institute by Boon and colleagues (van der Bruggen et al. 1991). The genetic approach used for the identification of the tumor antigen was based on the fact that cytolytic T lymphocytes (CTL) are capable of lysing tumor cells coming from the same patient (autologous) *in vitro.*

MZ2-MEL cell line was established from MZ2 melanoma patient, from whom CTLs were also obtained (**Fig. 16**). The mixed culture of tumor cells and CTL clones resulted in the lysis of MZ2-MEL. For some CTL clones, however, MZ2-MEL cells escaped their cytolytic action. These MZ2-MEL cells had lost the antigen presented in their surface (termed antigen E) and therefore could not be recognized by the CTL. This MZ2-MEL subpopulation was qualified as antigen-loss tumor cell variant. To thus identify the lost antigen E, a genomic cosmid library of MZ2-MEL was constructed and was stably transfected into the MZ2-MEL antigen-loss variant. Once the peptide antigen was again at the surface of MZ2-MEL, tumor cells were sensitive to the lysis by the CTL. The cytolytic activity was assessed by the detection of Tumor Necrosis Factor (TNF) release by the CTL.

By following this procedure, the first tumor antigen had been identified. The corresponding gene was named *Melanoma Antigen 1* (*MAGE1*). Intriguingly, when assessing the expression profile of *MAGE1* in normal tissues, the gene was silent in all somatic tissues, and expressed in the testis. Soon, it was discovered that the genes encoding antigens recognized by CTLs shared the same expression pattern and therefore, based on this feature, these genes were termed "cancer-germline genes" (CG genes). It was quickly shown that *MAGE1* belonged to a family of genes sharing a high percentage of sequence identity and therefore was renamed *MAGEA1. MAGEA1* identification paved the way for the discovery of additional genes exhibiting a cancer-testis expression profile, and activated aberrantly in tumor cells.

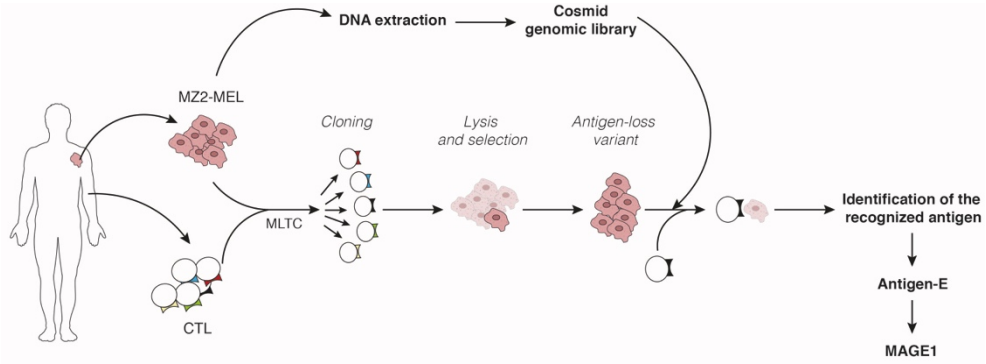


Figure 16: Schematic representation of the genetic approach used by Boon and colleagues for the identification of MAGEA1. *Melanoma cell line depicted in dashed lines and lighter color represent lysed cells. CTL: cytotoxic T lymphocytes. MLTC: Mixed Lymphocyte Tumor cell Culture; MZ2-MEL: melanoma cell line.*

Subsequently, to restrain the time of procedure for the identification of tumor antigens recognized by CTLs, the above approach was modified by using a cDNA library instead of a genomic library (**Fig. 17**) (De Plaen et al. 1997). RNA was extracted from established melanoma cell lines, converted to cDNA and cloned into plasmids for the generation of a cDNA library in *E. Coli*. Then, library pools were co-transfected into the simian COS-7 cells, along with the HLA molecule of interest. The capacity of the CTL to recognize and lyse the co-transfected COS-7 was again

measured by the release of TNF. This technique allowed the identification of *BAGE* and *GAGE* CG genes starting from MZ2-MEL cell line (Boel et al. 1995; Van den Eynde et al. 1995).

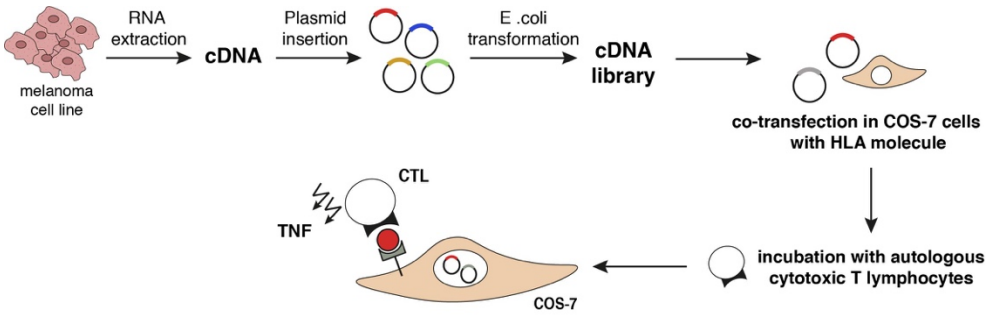


Figure 17: Schematic representation of an alternative genetic approach for the identification of tumor antigens based on cDNA libraries. *CTL: cytotoxic T lymphocytes; COS-7: simian cell line; TNF: Tumor Necrosis Factor.*

Finally, another genetic approach for the identification of tumor antigens was developed by the group of Pfreundschuh (Sahin et al. 1995). This method was based on the presence of antibodies in the serum of patient directed against these tumor antigens. The method was called Serological analysis of cDNA expression libraries (SEREX). SEREX technique consisted in creating a phage cDNA library of tumor tissue biopsies and subsequently infecting *E. Coli.* The objective of the assay was to assess whether there are antibodies in the autologous serum recognizing the expressed proteins in bacteria.

With the emergence of parallel sequencing, the identification of CG genes became easier. Today, with the rise of high throughput techniques and their availability to the public by worldwide consortia or research groups, CG gene identification has been greatly facilitated in terms of cost and time (Chen et al. 2005; da Silva et al. 2017; Hofmann et al. 2008; Scanlan et al. 2002; Wang et al. 2016; Yokoe et al. 2008). There was therefore a tremendous shift in the time of analysis of CG genes: it went from several weeks to identify potential CG gene candidates with the SEREX technique to couple of minutes to filter for a testis-specific expression of genes in

a panel of normal tissues of GTEx (for a description of the database, refer to section D.3).

b) Insights into the definition of cancer-germline genes

Multiple research groups identified CG genes and multiple terms to describe this group of genes such as “cancer-testis genes”, “cancer-testis antigens”, “cancer-germline antigens” were coined. The definition seems therefore, lab-culture oriented, with some authors even using these terms interchangeably. In a vast number of studies, the term “cancer-testis antigens”, or CTAs, prevails. However, by integrating all the studies over the years regarding CTAs, multiple issues arise with the usage of this term.

First, studies have shown that the protein expression of CTAs is restricted to germ cells and not somatic cells composing the tissue of the testis (Djureinovic et al. 2016; Jungbluth et al. 2000). More specifically, it was shown that multiple genes of this family are expressed in spermatogonia, the pre-meiotic cells of the testis (Loriot, Boon, and De Smet 2003; Takahashi et al. 1995). With the emergence of single-cell RNA-seq technologies these observations can be extended to several genes of this family (**Fig. 18**). These data therefore show that the majority of their expression is specific to germ cells rather than their supporting microenvironment of Leydig, Sertoli cells or additional stromal cells, which are included in the term “testis”.

Secondly, several of these “testis”-specific genes were also found expressed in fetal ovary, their expression therefore including both genders (Jungbluth et al. 2007; Koslowski et al. 2004). Additionally, some of these genes were also expressed in placenta.

Thirdly, CTAs were initially identified as evoking an immune response in cancer patients. However, today, with the description of hundreds of genes belonging to the CG family, the immunogenicity of the vast majority of them still remains to be determined. Furthermore, by using the term “antigen” to characterize this group of genes, it implies that these genes encode antigens presented in the surface of germ cells. However, this is not the case since germ cells lack HLA class I molecules.

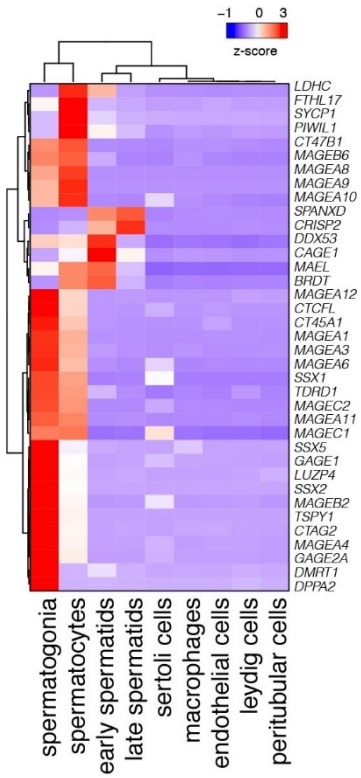


Figure 18: Cancer-Germline gene expression in cell types of the human testis. *Single cell expression data of the HPA were interrogated for the germline-specific CG genes list obtained from Van Tongelen, Loriot and De Smet 2017. Gene expression is depicted by a z-score of TPM values. Gene expression in testis cell types was clustered using Ward's method and Euclidean distance.*

Another issue that has been raised with the emergence of publicly available RNA-seq data of normal tissue samples, is that the tissue specificity of several genes catalogued as CG genes, is not restrained to the testis, ovary, nor placenta. Indeed, an initial study of Hide and colleagues interrogating cDNA libraries of massive parallel sequencing expression or cap analysis gene expression, combined with RT-PCR experiments, classified CTAs as testis-restricted, testis-brain-restricted, and testis-selective genes (Hofmann et al. 2008). In line with this observation, our group recently re-assessed the CT genes compiled in CTdatabase using RNA-seq data of GTEx portal. It was shown that 77% of the genes listed as CG genes presented a testis-restricted expression profile (RPKM > 0.5 in testis and/or ovary, and RPKM < 0.5 in the rest of the somatic tissues), whereas the remaining 23% of genes either exhibited their highest expression in the testis and additional expression in other

tissues (termed testis-preferential), or their expression level in the testis was similar to the one found in several somatic tissues (Van Tongelen, Loriot, and De Smet 2017).

Finally, some studies also attribute the term CG to genes that present a restricted expression to germ cells and that are induced by a demethylating treatment in normal and tumor cell lines. Several CG genes were indeed found to be regulated by promoter DNA-methylation. Nevertheless, it remains to be studied the exact fraction of CG genes solely dependent on DNA methylation. Indeed, additional mechanisms were found to regulate this group of genes (Akers, Odunsi, and Karpf 2010) (refer to subsection C.3.e).

Definition of "Cancer-Germline genes":

Based on the above observations, we believe that the term "cancer-germline genes" is better suited to describe the group of genes that are physiologically restricted to germ cells of the testis and fetal ovary, occasionally to placenta tissue, and ectopically activated in tumors, without any inference on the immunogenicity of the gene, or the epigenetic mechanism involved in their regulation. Therefore, throughout this work, the term cancer-germline genes (CG genes) will be used when referring to this intriguing group of genes.

c) Physiological function of CG genes

CG genes display a physiological expression site in germline cells of both testis and ovaries (occasionally, in placenta tissue). Specifically, they are found primarily in the first steps of spermatogenesis, in pre-meiotic germ cells (**Fig. 18**) (Jungbluth et al. 2000).

Insights into their physiological function have come from mouse models established for the depletion of several CG genes. Mice are viable, however, for most invalidated CG genes, they are infertile. This indicates that CG genes have a crucial role in gamete formation (Fon Tacer et al. 2019). Studies have found that CG genes can contribute to different features of the gametes including their

metabolism, RNA regulation, movement, meiosis and more recently, response to genotoxic stress (Whitehurst 2014; Gantchev et al. 2020; Fon Tacer et al. 2019).

d) Oncogenic functions of CG genes

The fact that several of these genes have been discovered by their ability to unveil tumor cells to the immune system seems quite counterintuitive. It would therefore be reasonable to expect that their activation exerts pro-tumoral functions in order to find them activated in these cells, in primary tumor samples, as well as in metastatic lesions.

Emerging data support a pro-oncogenic function of this group of genes (Gibbs and Whitehurst 2018; Van Tongelen, Loriot, and De Smet 2017). Several CG genes encode proteins, others are lncRNAs, or were found to host CG miRNAs.

CG Proteins:

Mounting evidence suggest that CG proteins can participate in tumor progression through a variety of cellular functions (Gibbs and Whitehurst 2018; Van Tongelen, Loriot, and De Smet 2017). Indeed, when compiling data of different studies performed on different CG genes, they appear to contribute to multiple cancer hallmarks. For instance, BORIS, a CG transcription factor was found to compete with CTCF transcription factor at the promoter region of *hTERT*, one of the two subunits of telomerase. This, in turn, results in the transcriptional activation of hTERT (Renaud et al. 2011). BORIS consequently appears to participate in tumor immortalization cancer hallmark. CG proteins were also found to participate in the deregulation of cellular energetics via their involvement in the AMPK pathway. Studies reported that MAGEA3/6 interact with TRIM28, a ubiquitin ligase. This interaction leads to the ubiquitination of AMPK and thus its degradation in tumor cells. This results in an increase in mammalian Target Of Rapamycin (mTOR) activity, and subsequently protein synthesis and cell viability (Pineda et al. 2015). In the same way as for MAGEA3/6, it was shown that MAGEC2 can also interact with TRIM28, leading to the ubiquitination of p53 transcription factor (Doyle et al. 2010), and consequently to apoptosis escape.

CG miRNAs:

CG transcripts can also harbor miRNAs. Our group previously showed that *CT-GABRA3* is a lncRNA arising from an alternative promoter of *GABRA3* gene (Loriot et al. 2014) (**Fig. 19**). *GABRA3* encodes a subunit of the γ-aminobutyric acid receptor (GABA). *CT-GABRA3* is ectopically activated in lung cancer and melanoma due to their promoter demethylation in these cells. This transcript contains the sequence of two CG miRNAs (miR-105 and miR-767) which are aberrantly expressed mirroring the expression pattern of their host. Studies support an oncogenic promoting role of these two miRNAs. First, regarding miR-105, it was shown that this miRNA targets the 3'UTR of *Tight Junction Protein 1* (*TJP1*), a protein implicated in the maintenance of tight junctions between cells (Zhou et al. 2014). In a mouse model, it was shown that tumor cells secrete exosomes containing miR-105 that target *TJP1* of endothelial cells. This leads to the weakening of the endothelial barrier, and therefore, to the dissemination of tumor cells in the bloodstream. miR-767, on the other hand, was found to target TET1/3 demethylating enzymes in melanoma cell lines (Loriot et al. 2014). This results in a decrease of the global levels of 5hmC, contributing to tumor epigenetic alterations in these cells.

CG lncRNAs:

CT-GABRA3 is a lncRNA that, in addition to harboring two miRNAs exerting pro-tumorigenic functions, was used as a model to establish an unexpected connection between DNA hypomethylation and DNA hypermethylation in tumor cells (Fain et al. 2021) (**Fig. 19**). Indeed, it was found that promoter DNA demethylation of *CT-GABRA3* results in their induction in lung and melanoma tumors. This transcriptional activation triggers the methylation of the *GABRA3* promoter, initially unmethylated in these cells, and the deposition of H3K36me3, a mark associated with gene bodies. Therefore, the ectopic activation of *CT-GABRA3* leads to aberrant hypermethylation of genomic regions encompassing the transcriptional unit of *CT-GABRA3*, including imbedded promoter regions such as *GABRA3*, by, presumably, the passage of the transcriptional machinery.

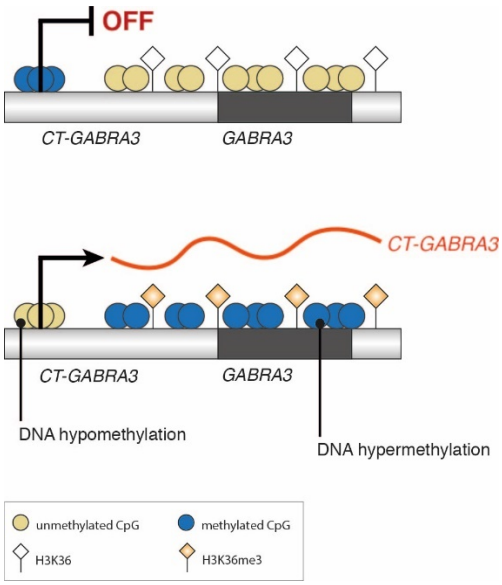


Figure 19: DNA hypomethylation-induced activation of CT-GABRA3 leads to the DNA hypermethylation of downstream GABRA3 promoter region. *Broken arrows indicate the TSS of CT-GABRA3. Transcriptional activation or repression is indicated by "ON" and "OFF" respectively.*

To summarize, CG genes seem to be implicated in different cellular processes in both normal and tumor cells. In the case of tumor cells, multiple CG genes were found to participate to cancer hallmarks (**Fig. 20**). Of note, as stated above, these genes are often co-activated in tumor cells. Hence, a combined effect of CG genes on tumor cells would be expected.

Despite the continuous efforts to characterize CG genes, most CG genes have unknown functions in both gametes and tumor cells.

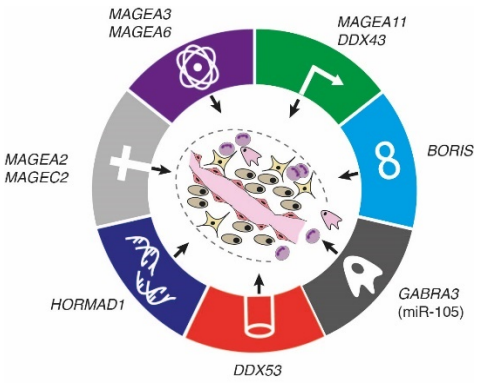


Figure 20: Examples of CG genes contributing to cancer hallmarks. *For a detailed function of each of these genes, refer to the text above and Van Tongelen, Loriot and De Smet 2017.*

e) Regulation of CG genes

DNA methylation:

Insights into the regulation mechanisms of CG genes come from experiments performed on tumor cells using *MAGEA1,* the founding member of this CG family, as a model. Weber *et al*. provided the first evidence that *MAGEA1* could be induced by increasing concentrations of 5-azadC in treated melanoma cell lines (Weber et al. 1994). Subsequently, it was shown that transcriptional activation was strictly correlated with promoter DNA demethylation of *MAGEA1* gene. Indeed, it was observed that *MAGEA1* promoter was unmethylated in germ cells, melanoma and sarcoma cell lines, where transcriptional activation was on. In contrast, *MAGEA1* promoter was methylated when the gene was transcriptionally silent in somatic tissues and tumor cell lines (De Smet, Lurquin, Lethe, et al. 1999). Interestingly, demethylation of *MAGEA1* promoter region was also observed when tumor cells, initially silent for the expression of *MAGEA1,* were treated with 5-azadC and thus induced the expression of the gene. This inverse correlation was furthermore extended for several of these genes to melanoma and lung cancer tissue samples. A mechanistic link came from transfection experiments where *MAGEA1* promoter was cloned in front of a transgene. Transfection in tumor cell lines that endogenously expressed *MAGEA1* revealed that the transgene could not be

expressed when the promoter was methylated. Additionally, treatment with siRNAs targeting DNMT1, the DNA methylation maintenance enzyme, lead to the transcriptional induction of CG genes (Loriot et al. 2006). More recently, in a high throughput manner, experiments in hTERT-immortalized fibroblasts transfected with shDNMT1s also showed an induction of CG genes (O'Neill et al. 2018). Of note, as stated above, the DNA demethylation of the promoter region of these genes was associated with the global phenomenon of DNA hypomethylation occurring in these cells (De Smet et al. 1996b).

Additional mechanisms:

The fact that transcriptional induction of CG genes occurs following their promoter DNA demethylation in cells, or the fact that the expression of an unmethylated transgene in an endogenously repressed *MAGEA1* melanoma cell line, suggest that ubiquitous transcription factors are involved in the regulation of these genes. Consistently, studies have identified the binding sites of E26-transformation specific (ETS) transcription factor family in the promoter of *MAGEA1* gene (De Smet, Lurquin, Lethe, et al. 1999). Indeed, mutagenesis of the two ETS sites resulted in a decreased *MAGEA1* promoter activity. Nevertheless, even after the DNA demethylation of their promoter region, some CG genes do not display any transcriptional activation. This suggests the involvement of additional mechanisms for their activation. For example, BORIS, was found to be involved in the transcriptional activation of other CG genes such as *MAGEA2* and *MAGEA3* (Bhan et al. 2011). The regulation of the latter genes appears therefore conditioned by the presence, and thus the initial ectopic activation of *BORIS* in these cells. Finally, the involvement of histone modifications was also investigated in the induction of CG genes. It appears that activating and repressing modifications are not a cause but rather a consequence of transcriptional activity in these cells (Cannuyer et al. 2013b).

Together, the above data point to a primary mechanism of promoter DNA methylation in the activation of CG genes.

f) Interest for the study of CG genes

The discovery of CG genes exhibited high enthusiasm for cancer immunotherapy. Indeed, the fact that these genes are recognized as non-self from the immune system, in combination with their highly tissue-restricted expression profile provide new candidates for targeted therapeutic approaches.

A testis-biased approach for the identification of genes ectopically activated in tumors?

Studies that aimed to identify aberrantly activated genes in tumors, for instance, for the search of tumor biomarkers or prognosis factors, or the search of driver events in these cells, usually hold a bias towards the tissue-specificity of the candidate genes. Indeed, based on the existence and the characteristics of the group of CG genes, studies often impose a filter of a physiological expression site to the testis (Wang et al. 2016; Rousseaux et al. 2013). This therefore excludes the possibility that genes belonging to other tissue-specific clusters and ectopically activated in tumors would be identified. Furthermore, other studies set as criteria regulatory restrictions, such as the presence/absence of particular transcription factors, or the methylation state of the promoter region of potential candidates in these cells. These filters therefore do not allow a fair study of the aberrant transcriptome of tumor cells compared to their normal counterparts.

D. Lung adenocarcinoma as a model system

In the present work, we raised a number of research questions (see Research Objectives chapter) that we attempted to address in chapters 1 and 2. We used lung adenocarcinoma (LUAD) as a study model to get insights into ectopic activation of tissue-specific gene clusters and their corresponding regulatory mechanisms in LUAD. Therefore, here below, some characteristics of this disease are outlined.

1. *Lung cancer features*

Lung cancer is the second most prevalent form of cancer and the leading cause of death of cancer worldwide (Sung et al. 2021). The 5-year survival rate, all lung cancer subtypes considered, is estimated at about 17%. One of the causes of this high death rate is the lack of early lung cancer detection. The late diagnosis is attributed to the absence of precise symptoms pointing to this particular disease. Indeed, there is a broad unspecific symptom spectrum in the early stages of lung cancer such as cough, or shortness of breath (Latimer and Mott 2015). Additionally, lungs do not display nociceptors, and thus, no pain can be detected early-on. Lung cancer diagnosis is therefore often established after tumor cells have already metastasized in distant organs and have started altering their function. Common symptoms, in addition to those specific to the metastatic sites, include hemoptysis, compression of the vena cava due to the growth of tumor cells themselves or to related adenopathies, or paraneoplastic syndromes. Another reason for the high mortality rate is the fact that treatment options are still limited today. In fact, surgery for tumor excision is only available for early tumor stage, while once the tumor has disseminated, only systemic treatment strategies are recommended. Finally, lung cancer, as will be described below, can be classified into different subtypes, for which different therapies can be an option or not.

Lung cancer diagnosis is established by pathological evidence resulting from lung biopsy (Latimer and Mott 2015). The most recognized and major risk factor of lung cancer is tobacco smoking. Other risk factors include passive smoking or environmental exposure to heavy metals or asbestos (Malhotra et al. 2016).

Treatment options are dependent on the extent of the tumor dissemination, as well as the histological subtype of the tumor. They include surgical excision of the lesion when the mass is localized, whereas, when the site is inoperable or the disease has advanced, a combination of platinum-based chemotherapeutic agents, radiotherapy, targeted molecular therapies or antibody-based immunotherapy is chosen depending on the patient features (Heist and Engelman 2012).

Histological classification of lung cancer:

From a pathological view, lung tumors are classified into mainly two groups (Nicholson et al. 2022)(**Fig. 21**):

- **Non-small cell lung cancer** (NSCLC) that represents about 85% of lung cancer cases.
- **Small cell lung cancer** (SCLC) that represents about 15% of lung cancer cases. They constitute a type of lethal neuroendocrine tumors that can be divided into further subtypes based on the transcription factor overexpressed in the tumor cells (*ASCL1*, *NEUROD1*, or none of those) (Sabari et al. 2017). They appear to arise from a rare subpopulation of neuroendocrine cells of the lung (Ferone et al. 2020).

NSCLC is the prevailing type of lung tumors. This lung cancer subtype can be subclassified into three main categories: lung squamous cell carcinoma (LUSC), lung adenocarcinoma (LUAD) and large cell carcinoma. LUAD and LUSC constitute the predominant categories with 40% and 25% of cases respectively (Travis et al. 2015). In the present study, we will focus on LUAD.

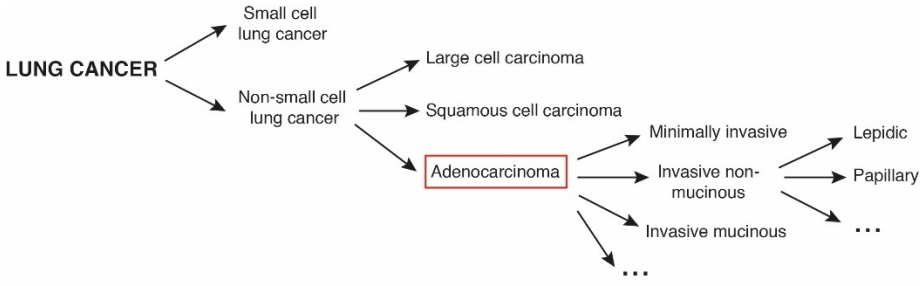


Figure 21: Schematic classification of lung cancer tumors. *Adenocarcinoma, our study model, is highlighted in red. Three dots represent further subcategories of lung tumor classification that are not depicted on the scheme. The lung cancer classification is based on the 2021 WHO classification of lung tumors (Nicholson et al. 2022).*

2. *LUAD features*

a) Pathological classification

Pathological diagnosis of LUAD is based on the search of protein markers on tissue biopsies, and more specifically, Thyroid Transcription Factor 1 (TTF1, also known as NKX2-1), a transcription factor involved in the differentiation of lung alveolar cells. Positive staining for TTF1 confirms the adenocarcinoma nature of the sample (Travis et al. 2015). However, this marker was shown to detect LUAD in only 80% of the cases. TTF1 is indeed lost in a fraction of LUAD, and this has been associated with a worse prognosis (Zhang et al. 2015).

To complicate matters, it was also shown that LUAD can undergo transdifferentiation events. Indeed, several independent studies have found that 10-20% of LUAD tissue samples display neuroendocrine (NE) features such as positive staining of classical NE markers: synaptophysin, chromogranin A among others (Bhattacharjee et al. 2001; Kosari et al. 2014). It was shown that this transdifferentiation phenotype is governed by the aberrant expression of *ASCL1,* a transcription factor responsible for the development of NE pulmonary cells (Borges et al. 1997). Besides this NE transdifferentiation event, it was also observed that some LUAD samples can be directed towards a squamous phenotype. Indeed, LUAD tissue samples displayed simultaneously features of adenocarcinoma (TTF1 positive stain) and squamous tumors (p63 positive stain) (Li and Lu 2018). This subtype is rare, reported as 0.4-4% of total LUAD tissue samples. In mouse models, this transition appears to be governed by KrasG12V activating mutations and a successive combinatory loss of the pioneer transcription factors Ttf1 and FoxA1/2 (Camolotto et al. 2018).

Based on histological features of LUAD tissues, LUAD can be further divided into subtypes such as minimally invasive adenocarcinoma, or invasive non-mucinous adenocarcinoma among others. The latter constitutes the most prevalent pattern of LUAD and presents additional histological subtypes such as lepidic, acinar, or papillary adenocarcinoma. This classification is based on the predominant pattern of cell morphology composing the LUAD tissue sample (Travis et al. 2015) (**Fig. 20**). For instance, when the lesion presents a glandular differentiation with papillary infoldings of thin connective tissue protruding into the air space, they are classified as papillary LUAD.

Overall, multiple LUAD subtypes have been described from a histological point of view, and it was shown that some of them correlate with a poorer patient survival (Makinen et al. 2017). Nevertheless, to date in the clinic, these LUAD subtypes are not considered for therapeutic approaches of LUAD patients. As stated above, approaches depend on the lung cancer major subtype (such as LUAD or LUSC) and the extent of tumor dissemination.

With the emergence of omics technologies, studies have focused on building comprehensive maps of tumors, including LUAD, in order to better characterize their genomic, as well as epigenomic alterations, with the ultimate aim to stratify tumor samples and customize patient care.

b) Genetic classification

In 2014, TCGA described a detailed profile of genetic alterations occurring in LUAD (n=230) (Cancer Genome Atlas Research 2014). They extended previous observations on the nature of the genetic mutations in LUAD and their prevalence in these samples. For instance, they identified that *KRAS* mutations occur in 33% of LUAD tissue samples and observed that mutations in this oncogene are mutually exclusive with those in *EGFR*. Other mutations in classical driver oncogenes included *BRAF* (10%), while mutations in tumor suppressor genes included *STK11* (17%), *RB1* (4%) or *CDKN2A* (4%). Importantly, among the studied cohort, 24% of samples did not harbor any known genetic driver alterations, suggesting the existence of yet undescribed tumorigenic mechanisms in LUAD. Further studies in a different LUAD tissue sample cohort than the one of TCGA, also confirm the genetic alterations of these driver genes (Chen et al. 2020), indicating that there are common and recurrent events in LUAD development. However, the frequency of a particular mutation differs between ethnicities (such as Caucasian and Asian populations).

Characterization of genetic alterations occurring in tumors paved the way for targeted therapies in LUAD, specifically the use of tyrosine kinase inhibitors against EGFR, ALK, ROS1, BRAF (Paez et al. 2004; Kwak et al. 2010; Bergethon et al. 2012). This has revolutionized LUAD therapies, that, for decades, were based on

conventional platinum-based chemotherapies resulting in heavy side effects of LUAD patients.

c) Transcriptomic classification

Early on, studies have attempted to cluster LUAD tissue samples based on their transcriptomic landscape. Initial work has showed that LUAD samples could be grouped into subclasses enriched for NE genes, and pneumocyte II genes (Bhattacharjee et al. 2001). Follow-up studies using other clustering approaches identified three main transcriptomic subtypes of LUAD named: bronchioid, squamoid and magnoid (Hayes et al. 2006). More recently, the LUAD cohort of TCGA, using a larger number of tumor samples refined this classification and gave rise to the following three subtypes: terminal respiratory unit, proximal-inflammatory and proximal-proliferative (Cancer Genome Atlas Research 2014). This classification is widely found in the literature, although with the rising number of transcriptomic studies in LUAD, the definition of these subtypes does not always converge.

d) Methylomic classification

DNA methylation levels of CGI promoter regions were also examined in the LUAD cohort of TCGA. This allowed to group LUAD in two main categories: one exhibiting a high CGI methylator phenotype, and the other reflecting CGI DNA methylation levels closer to the ones in normal lung tissue (low CGI methylator phenotype). In accordance, in the high CGI methylator phenotype subgroup, hypermethylation events of tumor suppressor genes such as at the promoter region of *CDKN2A* were detected. In the case of *CDKN2A*, this leads to gene silencing by an epigenetic mechanism, in addition to the genetic mutations described above.

To date, no consistent association could be observed between LUAD classification subtypes and particular gene expressions, patient survival or resistance to therapies, that could be translated in the clinic. Therefore, as stated previously, none of these classifications are taken into account in routine diagnosis.

Together, these data indicate that besides the classical alterations of known drivers in LUAD, there is a fair share of tissue samples that exhibit none of these genetic aberrations. Additionally, there appears to be a high heterogeneity of LUAD tissue samples due to extensive transcriptional and methylation changes in these cells.

3. LUAD as a model of study

As stated above, in the present work, LUAD was chosen as a model system. This was justified for the following reasons: first, the study of ectopic gene activation implies to locate the precise cell of origin of the tumor to be able to confirm the silent expression of genes of interest in these cells, compared to the tumor tissue. For LUAD, this has been extensively studied using mouse models and targeting single or a combination of mutations to distinct cell types of the lung (Blanpain 2013; Ferone et al. 2020). Indeed, it was found that the predominant cell type from which LUAD arises is alveolar type II (AT2) cells, located in the alveolar space of the lung. Consistently with this location, LUAD lesions are localized in the periphery of the lung. Secondly, multiple genome-wide transcriptomic, methylomic and proteomic experiments have been performed in a wide variety of materials related to LUAD such as: AT2 cells, normal lung, LUAD tissues and LUAD cell lines (Cancer Genome Atlas Research 2014; Consortium 2013; Suzuki et al. 2014; Uhlen et al. 2015) (**Table 3**). Therefore, we can perform multiple comparisons on different omic levels between normal and tumor lung tissues. Finally, LUAD tissues were found to exhibit extensive hypomethylation events, and therefore are a consistent model to study the consequence of the phenomenon of the global loss of DNA methylation occurring in tumors (Ikeda et al. 2013).

Throughout this work, we will refer to multiple databases that compile expression data (RNA-seq), methylation data (Infinium, WGBS) and immunohistochemical stainings of normal tissue samples, tumor cell lines and tumor tissue samples. These datasets were either obtained from data shared publicly by other research laboratories or are part of worldwide consortia (**Table 3**).

Normal tissue databases:

- The **Genotype-Tissue Expression project** (GTEx) aims to extensively characterize normal tissues. GTEx collects data of normal samples these samples in terms of RNA-seq (bulk and more recently single-cell), as well as H3K27ac chromatin immunoprecipitation (ChIP)-seq and DNA methylation (WGBS). The current release (v8) harbors data from 54 post-mortem normal tissue samples gathered from about a thousand individuals, with a minimum of 70 samples per tissue. Individuals come from different age groups, gender (with 67% being male) and ethnicities (with a high prevalence of Caucasian ethnicity, 85%) (Consortium 2013).

- The **Human Protein Atlas** (HPA) (Uhlen et al. 2015) compiles bulk and single-cell RNA-seq data, as well as proteomics data coming from antibody-based tissue-imaging of normal tissues and cell lines. The goal of the project is to meticulously characterize the human proteome at the cellular, tissue and organ level in healthy and tumor tissues.

- The **Encyclopedia of DNA elements** (ENCODE) (Consortium 2012) project gathers data of a wide variety of omics technology, such as expression, methylation, ChIP experiments, chromatin accessibility experiments in normal tissues, primary cells, and tumor cell lines. They provide extensive resources with the aim of identifying functional genomic elements.

Tumor databases:

- **The Cancer Genome Atlas** (TCGA) (Cancer Genome Atlas Research 2014) collects data from tumor and normal matching tissues of a wide variety of cancer types (n=33), both solid and liquid. The consortium aims to build a comprehensive profiling of tumor cells by characterizing these tissues to different extents: on the genomic level (exome-seq, ATAC-seq, copy-number variation, methylation data), transcriptomic level (bulk RNA-seq data, miR-seq data) and proteomics level (reverse phase protein array). There are also clinical data available of the patients from whom the samples were collected (such as vital status, age, treatment strategy followed, smoking history, etc.) to build associations between clinical and specific tumor features.

- The **DataBase of Transcriptional Start Sites** (DBTSS) (Suzuki et al. 2014): originally created to assemble TSS data, DBTSS provides data of RNA-seq

and matching methylation of 26 LUAD cell lines. It also allows the study of other chromatin features by providing ChIP data of activating and repressing histone modifications as well as RNA polII.

All the above studies are publicly available either via their portal websites, or via public repositories such as GEO-NCBI. These consortia are constantly being updated with new data being released.

Table 3: Tools for transcriptomic, methylomic and proteomic analyses in LUAD. *Databases of normal and LUAD tissues, and the corresponding data used in this study. WGBS: Whole Genome Bisulfite Sequencing; NA: Not Available.*

	Projects				
	GTEx	**ENCODE**	**HPA**	**TCGA**	**DBTSS**
Pathology	Normal	Normal	Normal/ Tumor	Normal/ Tumor	Tumor
Type	Tissues	Tissues/Cells	Tissues	Tissues	Cells
Histology	Variety	Variety	Variety	Variety	Variety
Expression					
Technology	RNA-seq	RNA-seq	RNA-seq	RNA-seq	RNA-seq
Files	Normalized	Normalized	Normalized	Normalized	Raw
Methylation					
Technology	WGBS	WGBS	NA	Infinium	Methyl-seq
Files	Normalized	Normalized	NA	Normalized	Raw

Introduction references

Akers, S. N., K. Odunsi, and A. R. Karpf. 2010. 'Regulation of cancer germline antigen gene expression: implications for cancer immunotherapy', *Future Oncol*, 6: 717-32.

Almeida, L. G., N. J. Sakabe, A. R. deOliveira, M. C. Silva, A. S. Mundstein, T. Cohen, Y. T. Chen, R. Chua, S. Gurung, S. Gnjatic, A. A. Jungbluth, O. L. Caballero, A. Bairoch, E. Kiesler, S. L. White, A. J. Simpson, L. J. Old, A. A. Camargo, and A. T. Vasconcelos. 2009. 'CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens', *Nucleic Acids Res*, 37: D816-9.

Bannister, A. J., and T. Kouzarides. 2011. 'Regulation of chromatin by histone modifications', *Cell Res*, 21: 381-95.

Barau, J., A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, F. Guillou, and D. Bourc'his. 2016. 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354: 909-12.

Barr, M. L., and E. G. Bertram. 1949. 'A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis', *Nature*, 163: 676.

Baylin, S. B., J. W. Hoppener, A. de Bustros, P. H. Steenbergh, C. J. Lips, and B. D. Nelkin. 1986. 'DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas', *Cancer Res*, 46: 2917-22.

Bergethon, K., A. T. Shaw, S. H. Ou, R. Katayama, C. M. Lovly, N. T. McDonald, P. P. Massion, C. Siwak-Tapp, A. Gonzalez, R. Fang, E. J. Mark, J. M. Batten, H. Chen, K. D. Wilner, E. L. Kwak, J. W. Clark, D. P. Carbone, H. Ji, J. A. Engelman, M. Mino-Kenudson, W. Pao, and A. J. Iafrate. 2012. 'ROS1 rearrangements define a unique molecular class of lung cancers', *J Clin Oncol*, 30: 863-70.

Bestor, T. H., J. R. Edwards, and M. Boulard. 2015. 'Notes on the role of dynamic DNA methylation in mammalian development', *Proc Natl Acad Sci U S A*, 112: 6796-9.

Bhan, S., S. S. Negi, C. Shao, C. A. Glazer, A. Chuang, D. A. Gaykalova, W. Sun, D. Sidransky, P. K. Ha, and J. A. Califano. 2011. 'BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer', *Clin Cancer Res*, 17: 4267-76.

Bhattacharjee, A., W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson. 2001. 'Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses', *Proc Natl Acad Sci U S A*, 98: 13790-5.

Bird, A. P. 1980. 'DNA methylation and the frequency of CpG in animal DNA', *Nucleic Acids Res*, 8: 1499-504.

Blanpain, C. 2013. 'Tracing the cellular origin of cancer', *Nat Cell Biol*, 15: 126-34.

Borges, M., R. I. Linnoila, H. J. van de Velde, H. Chen, B. D. Nelkin, M. Mabry, S. B. Baylin, and D. W. Ball. 1997. 'An achaete-scute homologue essential for neuroendocrine differentiation in the lung', *Nature*, 386: 852-5.

Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor. 2001. 'Dnmt3L and the establishment of maternal genomic imprints', *Science*, 294: 2536-9.

Brockdorff, N., A. Ashworth, G. F. Kay, P. Cooper, S. Smith, V. M. McCabe, D. P. Norris, G. D. Penny, D. Patel, and S. Rastan. 1991. 'Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome', *Nature*, 351: 329-31.

Buck-Koehntop, B. A., and P. A. Defossez. 2013. 'On how mammalian transcription factors recognize methylated DNA', *Epigenetics*, 8: 131-7.

Cajuso, T., P. Sulo, T. Tanskanen, R. Katainen, A. Taira, U. A. Hanninen, J. Kondelin, L. Forsstrom, N. Valimaki, M. Aavikko, E. Kaasinen, A. Ristimaki, S. Koskensalo, A. Lepisto, L. Renkonen-Sinisalo, T. Seppala, T. Kuopio, J. Bohm, J. P. Mecklin, O. Kilpivaara, E. Pitkanen, K. Palin, and L. A. Aaltonen. 2019.

'Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival', *Nat Commun*, 10: 4022.

Camolotto, S. A., S. Pattabiraman, T. L. Mosbruger, A. Jones, V. K. Belova, G. Orstad, M. Streiff, L. Salmond, C. Stubben, K. H. Kaestner, and E. L. Snyder. 2018. 'FoxA1 and FoxA2 drive gastric differentiation and suppress squamous identity in NKX2-1-negative lung cancer', *Elife*, 7.

Cancer Genome Atlas Research, Network. 2014. 'Comprehensive molecular profiling of lung adenocarcinoma', *Nature*, 511: 543-50.

Cancer Genome Atlas Research Network. Electronic address, andrew aguirre dfci harvard edu, and Network Cancer Genome Atlas Research. 2017. 'Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma', *Cancer Cell*, 32: 185-203 e13.

Cannuyer, J., A. Loriot, G. K. Parvizi, and C. De Smet. 2013. 'Epigenetic hierarchy within the MAGEA1 cancer-germline gene: promoter DNA methylation dictates local histone modifications', *PLoS One*, 8: e58743.

Canzio, D., E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar, and B. Al-Sady. 2011. 'Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly', *Mol Cell*, 41: 67-81.

Cardoso, M. C., and H. Leonhardt. 1999. 'DNA methyltransferase is actively retained in the cytoplasm during early development', *J Cell Biol*, 147: 25-32.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. 'Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription', *Cell*, 123: 581-92.

Catteau, A., W. H. Harris, C. F. Xu, and E. Solomon. 1999. 'Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics', *Oncogene*, 18: 1957-65.

Chen, J., H. Yang, A. S. M. Teo, L. B. Amer, F. G. Sherbaf, C. Q. Tan, J. J. S. Alvarez, B. Lu, J. Q. Lim, A. Takano, R. Nahar, Y. Y. Lee, C. Z. J. Phua, K. P. Chua, L. Suteja, P. J. Chen, M. M. Chang, T. P. T. Koh, B. H. Ong, D. Anantham, A. A. L. Hsu, A. Gogna, C. W. Too, Z. W. Aung, Y. F. Lee, L. Wang, T. K. H. Lim, A. Wilm, P. S. Choi, P. Y. Ng, C. K. Toh, W. T. Lim, S. Ma, B. Lim, J. Liu, W. L. Tam, A. J. Skanderup, J. P. S. Yeong, E. H. Tan, C. L. Creasy, D. S. W. Tan, A. M. Hillmer, and W. Zhai. 2020. 'Genomic landscape of lung adenocarcinoma in East Asians', *Nat Genet*, 52: 177-86.

Chen, Y. T., A. O. Gure, and M. J. Scanlan. 2005. 'Serological analysis of expression cDNA libraries (SEREX): an immunoscreening technique for identifying immunogenic tumor antigens', *Methods Mol Med*, 103: 207-16.

Chen, Y. T., M. J. Scanlan, C. A. Venditti, R. Chua, G. Theiler, B. J. Stevenson, C. Iseli, A. O. Gure, T. Vasicek, R. L. Strausberg, C. V. Jongeneel, L. J. Old, and A. J. Simpson. 2005. 'Identification of cancer/testis-antigen genes by massively parallel signature sequencing', *Proc Natl Acad Sci U S A*, 102: 7940-5.

Consortium, Encode Project. 2012. 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489: 57-74.

Consortium, G. TEx. 2013. 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet*, 45: 580-5.

da Silva, V. L., A. F. Fonseca, M. Fonseca, T. E. da Silva, A. C. Coelho, J. E. Kroll, J. E. S. de Souza, B. Stransky, G. A. de Souza, and S. J. de Souza. 2017. 'Genome-wide identification of cancer/testis genes and their association with prognosis in a pan-cancer analysis', *Oncotarget*, 8: 92966-77.

Dahlet, T., A. Argueso Lleida, H. Al Adhami, M. Dumas, A. Bender, R. P. Ngondo, M. Tanguy, J. Vallet, G. Auclair, A. F. Bardet, and M. Weber. 2020. 'Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity', *Nat Commun*, 11: 3153.

de la Serna, I. L., Y. Ohkawa, and A. N. Imbalzano. 2006. 'Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers', *Nat Rev Genet*, 7: 461-73.

De Smet, C., O. De Backer, I. Faraoni, C. Lurquin, F. Brasseur, and T. Boon. 1996. 'The activation of human gene MAGE-1 in

tumor cells is correlated with genome-wide demethylation', *Proc Natl Acad Sci U S A*, 93: 7149-53.

De Smet, C., and A. Loriot. 2010. 'DNA hypomethylation in cancer: epigenetic scars of a neoplastic journey', *Epigenetics*, 5: 206-13.

De Smet, C., C. Lurquin, B. Lethe, V. Martelange, and T. Boon. 1999. 'DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter', *Mol Cell Biol*, 19: 7327-35.

Deans, C., and K. A. Maggert. 2015. 'What do you mean, "epigenetic"?', *Genetics*, 199: 887-96.

Deichmann, U. 2016. 'Epigenetics: The origins and evolution of a fashionable topic', *Dev Biol*, 416: 249-54.

Dhayalan, A., A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin, and A. Jeltsch. 2010. 'The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation', *J Biol Chem*, 285: 26114-20.

Disteche, C. M., and J. B. Berletch. 2015. 'X-chromosome inactivation and escape', *J Genet*, 94: 591-9.

Djureinovic, D., B. M. Hallstrom, M. Horie, J. S. M. Mattsson, L. La Fleur, L. Fagerberg, H. Brunnstrom, C. Lindskog, K. Madjar, J. Rahnenfuhrer, S. Ekman, E. Stahle, H. Koyi, E. Branden, K. Edlund, J. G. Hengstler, M. Lambe, A. Saito, J. Botling, F. Ponten, M. Uhlen, and P. Micke. 2016. 'Profiling cancer testis antigens in non-small-cell lung cancer', *JCI Insight*, 1: e86837.

Doyle, J. M., J. Gao, J. Wang, M. Yang, and P. R. Potts. 2010. 'MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases', *Mol Cell*, 39: 963-74.

Ehrlich, M. 2002. 'DNA methylation in cancer: too much, but also too little', *Oncogene*, 21: 5400-13.

———. 2009. 'DNA hypomethylation in cancer cells', *Epigenomics*, 1: 239-59.

———. 2019. 'DNA hypermethylation in disease: mechanisms and clinical relevance', *Epigenetics*, 14: 1141-63.

Esteller, M. 2002. 'CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future', *Oncogene*, 21: 5427-40.

———. 2007. 'Epigenetic gene silencing in cancer: the DNA hypermethylome', *Hum Mol Genet*, 16 Spec No 1: R50-9.

Fain, J. S., A. Loriot, A. Diacofotaki, A. Van Tongelen, and C. De Smet. 2021. 'Transcriptional overlap links DNA hypomethylation with DNA hypermethylation at adjacent promoters in cancer', *Sci Rep*, 11: 17346.

Feinberg, A. P., and B. Vogelstein. 1983. 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, 301: 89-92.

Ferone, G., M. C. Lee, J. Sage, and A. Berns. 2020. 'Cells of origin of lung cancers: lessons from mouse studies', *Genes Dev*, 34: 1017-32.

Flavahan, W. A., E. Gaskell, and B. E. Bernstein. 2017. 'Epigenetic plasticity and the hallmarks of cancer', *Science*, 357.

Fon Tacer, K., M. C. Montoya, M. J. Oatley, T. Lord, J. M. Oatley, J. Klein, R. Ravichandran, H. Tillman, M. Kim, J. P. Connelly, S. M. Pruett-Miller, A. L. Bookout, E. Binshtock, M. M. Kaminski, and P. R. Potts. 2019. 'MAGE cancer-testis antigens protect the mammalian germline under environmental stress', *Sci Adv*, 5: eaav4832.

Gama-Sosa, M. A., V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke, and M. Ehrlich. 1983. 'The 5-methylcytosine content of DNA from human tumors', *Nucleic Acids Res*, 11: 6883-94.

Gantchev, J., A. Martinez Villarreal, S. Gunn, M. Zetka, N. Odum, and I. V. Litvinov. 2020. 'The ectopic expression of meiCT genes promotes meiomitosis and may facilitate carcinogenesis', *Cell Cycle*, 19: 837-54.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, and R. Jaenisch. 2003. 'Induction of tumors in mice by genomic hypomethylation', *Science*, 300: 489-92.

Gelfman, S., and G. Ast. 2013. 'When epigenetics meets alternative splicing: the roles of DNA methylation and GC architecture', *Epigenomics*, 5: 351-3.

Gibbs, Z. A., and A. W. Whitehurst. 2018. 'Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors', *Trends Cancer*, 4: 701-12.

Glaser, J., J. Iranzo, M. Borensztein, M. Marinucci, A. Gualtieri, C. Jouhanneau, A. Teissandier, C. Gaston-Massuet, and D. Bourc'his.

2022. 'The imprinted Zdbf2 gene finely tunes control of feeding and growth in neonates', *Elife*, 11.

Goll, M. G., F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, and T. H. Bestor. 2006. 'Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2', *Science*, 311: 395-8.

Greally, J. M. 2018. 'A user's guide to the ambiguous word 'epigenetics'', *Nat Rev Mol Cell Biol*, 19: 207-08.

Guo, J. U., Y. Su, C. Zhong, G. L. Ming, and H. Song. 2011. 'Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain', *Cell*, 145: 423-34.

Hackett, J. A., R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, and M. A. Surani. 2013. 'Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine', *Science*, 339: 448-52.

Hanahan, D. 2022. 'Hallmarks of Cancer: New Dimensions', *Cancer Discov*, 12: 31-46.

Heist, R. S., and J. A. Engelman. 2012. 'SnapShot: non-small cell lung cancer', *Cancer Cell*, 21: 448 e2.

Hendrich, B., and A. Bird. 1998. 'Identification and characterization of a family of mammalian methyl-CpG binding proteins', *Mol Cell Biol*, 18: 6538-47.

Herman, J. G., F. Latif, Y. Weng, M. I. Lerman, B. Zbar, S. Liu, D. Samid, D. S. Duan, J. R. Gnarra, W. M. Linehan, and et al. 1994. 'Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma', *Proc Natl Acad Sci U S A*, 91: 9700-4.

Hermann, A., R. Goyal, and A. Jeltsch. 2004. 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites', *J Biol Chem*, 279: 48350-9.

Hofmann, O., O. L. Caballero, B. J. Stevenson, Y. T. Chen, T. Cohen, R. Chua, C. A. Maher, S. Panji, U. Schaefer, A. Kruger, M. Lehvaslaiho, P. Carninci, Y. Hayashizaki, C. V. Jongeneel, A. J. Simpson, L. J. Old, and W. Hide. 2008. 'Genome-wide analysis of cancer/testis gene expression', *Proc Natl Acad Sci U S A*, 105: 20422-7.

Howe, F. S., H. Fischl, S. C. Murray, and J. Mellor. 2017. 'Is H3K4me3 instructive for transcription activation?', *Bioessays*, 39: 1-12.

Huang, Y., S. J. Myers, and R. Dingledine. 1999. 'Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes', *Nat Neurosci*, 2: 867-72.

Ikeda, K., K. Shiraishi, A. Eguchi, H. Shibata, K. Yoshimoto, T. Mori, Y. Baba, H. Baba, and M. Suzuki. 2013. 'Long interspersed nucleotide element 1 hypomethylation is associated with poor prognosis of lung adenocarcinoma', *Ann Thorac Surg*, 96: 1790-4.

Illingworth, R. S., U. Gruenewald-Schneider, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews, and A. P. Bird. 2010. 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS Genet*, 6: e1001134.

Iqbal, K., S. G. Jin, G. P. Pfeifer, and P. E. Szabo. 2011. 'Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine', *Proc Natl Acad Sci U S A*, 108: 3642-7.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang. 2011. 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*, 333: 1300-3.

Jain, D., C. Meydan, J. Lange, C. Claeys Bouuaert, N. Lailler, C. E. Mason, K. V. Anderson, and S. Keeney. 2017. 'rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline', *PLoS Genet*, 13: e1006964.

Jenuwein, T., and C. D. Allis. 2001. 'Translating the histone code', *Science*, 293: 1074-80.

Jiang, J. C., and K. R. Upton. 2019. 'Human transposons are an abundant supply of transcription factor binding sites and promoter activities in breast cancer cell lines', *Mob DNA*, 10: 16.

Jones, P. A. 1999. 'The DNA methylation paradox', *Trends Genet*, 15: 34-7.

———. 2012. 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nat Rev Genet*, 13: 484-92.

Jones, P. A., and D. Takai. 2001. 'The role of DNA methylation in mammalian epigenetics', *Science*, 293: 1068-70.

Jungbluth, A. A., W. A. Silva, Jr., K. Iversen, D. Frosina, B. Zaidi, K. Coplan, S. K. Eastlake-Wade, S. B. Castelli, G. C. Spagnoli, L. J. Old, and M. Vogel. 2007. 'Expression of cancer-testis (CT) antigens in placenta', *Cancer Immun*, 7: 15.

Jungbluth, A. A., E. Stockert, Y. T. Chen, D. Kolb, K. Iversen, K. Coplan, B. Williamson, N. Altorki, K. J. Busam, and L. J. Old. 2000. 'Monoclonal antibody MA454 reveals a heterogeneous expression pattern of MAGE-1 antigen in formalin-fixed paraffin embedded lung tumours', *Br J Cancer*, 83: 493-7.

Kandoth, C., M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. M. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, and L. Ding. 2013. 'Mutational landscape and significance across 12 major cancer types', *Nature*, 502: 333-39.

Kane, M. F., M. Loda, G. M. Gaida, J. Lipman, R. Mishra, H. Goldman, J. M. Jessup, and R. Kolodner. 1997. 'Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines', *Cancer Res*, 57: 808-11.

Klose, R. J., and A. P. Bird. 2006. 'Genomic DNA methylation: the mark and its mediators', *Trends Biochem Sci*, 31: 89-97.

Kosari, F., C. M. Ida, M. C. Aubry, L. Yang, I. V. Kovtun, J. L. Klein, Y. Li, S. Erdogan, S. C. Tomaszek, S. J. Murphy, L. C. Bolette, C. P. Kolbert, P. Yang, D. A. Wigle, and G. Vasmatzis. 2014. 'ASCL1 and RET expression defines a clinically relevant subgroup of lung adenocarcinoma characterized by neuroendocrine differentiation', *Oncogene*, 33: 3776-83.

Koslowski, M., C. Bell, G. Seitz, H. A. Lehr, K. Roemer, H. Muntefering, C. Huber, U. Sahin, and O. Tureci. 2004. 'Frequent nonrandom activation of germ-line genes in human cancer', *Cancer Res*, 64: 5988-93.

Kouzarides, T. 2007. 'Chromatin modifications and their function', *Cell*, 128: 693-705.

Kwak, E. L., Y. J. Bang, D. R. Camidge, A. T. Shaw, B. Solomon, R. G. Maki, S. H. Ou, B. J. Dezube, P. A. Janne, D. B. Costa, M. Varella-Garcia, W. H. Kim, T. J. Lynch, P. Fidias, H. Stubbs, J. A. Engelman, L. V. Sequist, W. Tan, L. Gandhi, M. Mino-Kenudson, G. C. Wei, S. M. Shreeve, M. J. Ratain, J. Settleman, J. G. Christensen, D. A. Haber, K. Wilner, R. Salgia, G. I. Shapiro, J. W. Clark, and A. J. Iafrate. 2010. 'Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer', *N Engl J Med*, 363: 1693-703.

Lane, N., W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, and W. Reik. 2003. 'Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse', *Genesis*, 35: 88-93.

Latimer, K. M., and T. F. Mott. 2015. 'Lung cancer: diagnosis, treatment principles, and screening', *Am Fam Physician*, 91: 250-6.

Lauberth, S. M., T. Nakayama, X. Wu, A. L. Ferris, Z. Tang, S. H. Hughes, and R. G. Roeder. 2013. 'H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation', *Cell*, 152: 1021-36.

Li, B., M. Carey, and J. L. Workman. 2007. 'The role of chromatin during transcription', *Cell*, 128: 707-19.

Li, C., and H. Lu. 2018. 'Adenosquamous carcinoma of the lung', *Onco Targets Ther*, 11: 4829-35.

Li, E., C. Beard, and R. Jaenisch. 1993. 'Role for DNA methylation in genomic imprinting', *Nature*, 366: 362-5.

Li, E., T. H. Bestor, and R. Jaenisch. 1992. 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell*, 69: 915-26.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker. 2009. 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature*, 462: 315-22.

Liu, Y., P. Siejka-Zielinska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Bockler, and C. X. Song. 2019. 'Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution', *Nat Biotechnol*, 37: 424-29.

Lopez-Moyado, I. F., A. Tsagaratou, H. Yuita, H. Seo, B. Delatte, S. Heinz, C. Benner, and A.

Rao. 2019. 'Paradoxical association of TET loss of function with genome-wide DNA hypomethylation', *Proc Natl Acad Sci U S A*, 116: 16933-42.

Loriot, A., T. Boon, and C. De Smet. 2003. 'Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia', *Int J Cancer*, 105: 371-6.

Loriot, A., E. De Plaen, T. Boon, and C. De Smet. 2006. 'Transient down-regulation of DNMT1 methyltransferase leads to activation and stable hypomethylation of MAGE-A1 in melanoma cells', *J Biol Chem*, 281: 10118-26.

Loriot, A., A. Van Tongelen, J. Blanco, S. Klaessens, J. Cannuyer, N. van Baren, A. Decottignies, and C. De Smet. 2014. 'A novel cancer-germline transcript carrying pro-metastatic miR-105 and TET-targeting miR-767 induced by DNA hypomethylation in tumors', *Epigenetics*, 9: 1163-71.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. 'Crystal structure of the nucleosome core particle at 2.8 A resolution', *Nature*, 389: 251-60.

Lyon, M. F. 1961. 'Gene action in the X-chromosome of the mouse (Mus musculus L.)', *Nature*, 190: 372-3.

Mack, S. C., H. Witt, R. M. Piro, L. Gu, S. Zuyderduyn, A. M. Stutz, X. Wang, M. Gallo, L. Garzia, K. Zayne, X. Zhang, V. Ramaswamy, N. Jager, D. T. Jones, M. Sill, T. J. Pugh, M. Ryzhova, K. M. Wani, D. J. Shih, R. Head, M. Remke, S. D. Bailey, T. Zichner, C. C. Faria, M. Barszczyk, S. Stark, H. Seker-Cin, S. Hutter, P. Johann, S. Bender, V. Hovestadt, T. Tzaridis, A. M. Dubuc, P. A. Northcott, J. Peacock, K. C. Bertrand, S. Agnihotri, F. M. Cavalli, I. Clarke, K. Nethery-Brokx, C. L. Creasy, S. K. Verma, J. Koster, X. Wu, Y. Yao, T. Milde, P. Sin-Chan, J. Zuccaro, L. Lau, S. Pereira, P. Castelo-Branco, M. Hirst, M. A. Marra, S. S. Roberts, D. Fults, L. Massimi, Y. J. Cho, T. Van Meter, W. Grajkowska, B. Lach, A. E. Kulozik, A. von Deimling, O. Witt, S. W. Scherer, X. Fan, K. M. Muraszko, M. Kool, S. L. Pomeroy, N. Gupta, J. Phillips, A. Huang, U. Tabori, C. Hawkins, D. Malkin, P. N. Kongkham, W. A. Weiss, N. Jabado, J. T. Rutka, E. Bouffet, J. O. Korbel, M. Lupien, K. D. Aldape, G. D. Bader, R. Eils, P. Lichter, P. B. Dirks, S. M. Pfister, A. Korshunov, and M. D. Taylor. 2014. 'Epigenomic alterations define lethal CIMP-positive ependymomas of infancy', *Nature*, 506: 445-50.

Makinen, J. M., K. Laitakari, S. Johnson, R. Makitaro, R. Bloigu, P. Paakko, E. Lappi-Blanco, and R. Kaarteenaho. 2017. 'Histological features of malignancy correlate with growth patterns and patient outcome in lung adenocarcinoma', *Histopathology*, 71: 425-36.

Maksakova, I. A., M. T. Romanish, L. Gagnier, C. A. Dunn, L. N. van de Lagemaat, and D. L. Mager. 2006. 'Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line', *PLoS Genet*, 2: e2.

Malhotra, J., M. Malvezzi, E. Negri, C. La Vecchia, and P. Boffetta. 2016. 'Risk factors for lung cancer worldwide', *Eur Respir J*, 48: 889-902.

Margueron, R., and D. Reinberg. 2011. 'The Polycomb complex PRC2 and its mark in life', *Nature*, 469: 343-9.

Martin, B. J. E., J. Brind'Amour, A. Kuzmin, K. N. Jensen, Z. C. Liu, M. Lorincz, and L. J. Howe. 2021. 'Transcription shapes genome-wide histone acetylation patterns', *Nat Commun*, 12: 210.

Meir, Z., Z. Mukamel, E. Chomsky, A. Lifshitz, and A. Tanay. 2020. 'Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells', *Nat Genet*, 52: 709-18.

Merlo, A., J. G. Herman, L. Mao, D. J. Lee, E. Gabrielson, P. C. Burger, S. B. Baylin, and D. Sidransky. 1995. '5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers', *Nat Med*, 1: 686-92.

Molaro, A., E. Hodges, F. Fang, Q. Song, W. R. McCombie, G. J. Hannon, and A. D. Smith. 2011. 'Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates', *Cell*, 146: 1029-41.

Nabel, C. S., H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, and R. M. Kohli. 2012. 'AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation', *Nat Chem Biol*, 8: 751-8.

Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. 1998. 'Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex', *Nature*, 393: 386-9.

Nicholson, A. G., M. S. Tsao, M. B. Beasley, A. C. Borczuk, E. Brambilla, W. A. Cooper, S. Dacic, D. Jain, K. M. Kerr, S. Lantuejoul, M. Noguchi, M. Papotti, N. Rekhtman, G. Scagliotti, P. van Schil, L. Sholl, Y. Yatabe, A. Yoshida, and W. D. Travis. 2022. 'The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015', *J Thorac Oncol*, 17: 362-87.

O'Neill, K. M., R. E. Irwin, S. J. Mackin, S. J. Thursby, A. Thakur, C. Bertens, L. Masala, J. E. P. Loughery, D. G. McArt, and C. P. Walsh. 2018. 'Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression', *Epigenetics Chromatin*, 11: 12.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell*, 99: 247-57.

Olins, D. E., and A. L. Olins. 2003. 'Chromatin history: our view from the bridge', *Nat Rev Mol Cell Biol*, 4: 809-14.

Paez, J. G., P. A. Janne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, and M. Meyerson. 2004. 'EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy', *Science*, 304: 1497-500.

Peters, J. 2014. 'The role of genomic imprinting in biology and disease: an expanding view', *Nat Rev Genet*, 15: 517-30.

Pineda, C. T., S. Ramanathan, K. Fon Tacer, J. L. Weon, M. B. Potts, Y. H. Ou, M. A. White, and P. R. Potts. 2015. 'Degradation of AMPK by a cancer-specific ubiquitin ligase', *Cell*, 160: 715-28.

Rando, O. J. 2007. 'Global patterns of histone modifications', *Curr Opin Genet Dev*, 17: 94-9.

Rasmussen, K. D., and K. Helin. 2016. 'Role of TET enzymes in DNA methylation, development, and cancer', *Genes Dev*, 30: 733-50.

Rauch, T. A., X. Zhong, X. Wu, M. Wang, K. H. Kernstine, Z. Wang, A. D. Riggs, and G. P. Pfeifer. 2008. 'High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer', *Proc Natl Acad Sci U S A*, 105: 252-7.

Renaud, S., D. Loukinov, L. Alberti, A. Vostrov, Y. W. Kwon, F. T. Bosman, V. Lobanenkov, and J. Benhattar. 2011. 'BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells', *Nucleic Acids Res*, 39: 862-73.

Riggs, A. D. 1975. 'X inactivation, differentiation, and DNA methylation', *Cytogenet Cell Genet*, 14: 9-25.

Rodriguez-Paredes, M., and M. Esteller. 2011. 'Cancer epigenetics reaches mainstream oncology', *Nat Med*, 17: 330-9.

Rousseaux, S., A. Debernardi, B. Jacquiau, A. L. Vitte, A. Vesin, H. Nagy-Mignotte, D. Moro-Sibilot, P. Y. Brichon, S. Lantuejoul, P. Hainaut, J. Laffaire, A. de Reynies, D. G. Beer, J. F. Timsit, C. Brambilla, E. Brambilla, and S. Khochbin. 2013. 'Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers', *Sci Transl Med*, 5: 186ra66.

Roy, N., and M. Hebrok. 2015. 'Regulation of Cellular Identity in Cancer', *Dev Cell*, 35: 674-84.

Sabari, J. K., B. H. Lok, J. H. Laird, J. T. Poirier, and C. M. Rudin. 2017. 'Unravelling the biology of SCLC: implications for therapy', *Nat Rev Clin Oncol*, 14: 549-61.

Sahin, U., O. Tureci, Y. T. Chen, G. Seitz, C. Villena-Heinsen, L. J. Old, and M. Pfreundschuh. 1998. 'Expression of multiple cancer/testis (CT) antigens in breast cancer and melanoma: basis for polyvalent CT vaccine strategies', *Int J Cancer*, 78: 387-9.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. 2002. 'Active genes are tri-methylated at K4 of histone H3', *Nature*, 419: 407-11.

Saxonov, S., P. Berg, and D. L. Brutlag. 2006. '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.

Scanlan, M. J., C. M. Gordon, B. Williamson, S. Y. Lee, Y. T. Chen, E. Stockert, A. Jungbluth, G. Ritter, D. Jager, E. Jager, A. Knuth, and L. J. Old. 2002. 'Identification of cancer/testis genes by database mining and mRNA expression analysis', *Int J Cancer*, 98: 485-92.

Schlesinger, Y., R. Straussman, I. Keshet, S. Farkash, M. Hecht, J. Zimmerman, E. Eden, Z. Yakhini, E. Ben-Shushan, B. E. Reubinoff, Y. Bergman, I. Simon, and H. Cedar. 2007. 'Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer', *Nat Genet*, 39: 232-6.

Schotta, G., M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg, and T. Jenuwein. 2004. 'A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin', *Genes Dev*, 18: 1251-62.

Schulz, W. A., J. P. Elo, A. R. Florl, S. Pennanen, S. Santourlidis, R. Engers, M. Buchardt, H. H. Seifert, and T. Visakorpi. 2002. 'Genomewide DNA hypomethylation is associated with alterations on chromosome 8 in prostate carcinoma', *Genes Chromosomes Cancer*, 35: 58-65.

Scourzic, L., E. Mouly, and O. A. Bernard. 2015. 'TET proteins and the control of cytosine demethylation in cancer', *Genome Med*, 7: 9.

Simpson, A. J., O. L. Caballero, A. Jungbluth, Y. T. Chen, and L. J. Old. 2005. 'Cancer/testis antigens, gametogenesis and cancer', *Nat Rev Cancer*, 5: 615-25.

Smallwood, S. A., S. Tomizawa, F. Krueger, N. Ruf, N. Carli, A. Segonds-Pichon, S. Sato, K. Hata, S. R. Andrews, and G. Kelsey. 2011. 'Dynamic CpG island methylation landscape in oocytes and preimplantation embryos', *Nat Genet*, 43: 811-4.

Song, J., O. Rechkoblit, T. H. Bestor, and D. J. Patel. 2011. 'Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation', *Science*, 331: 1036-40.

Sterner, D. E., and S. L. Berger. 2000. 'Acetylation of histones and transcription-related factors', *Microbiol Mol Biol Rev*, 64: 435-59.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. 2021. 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA Cancer J Clin*, 71: 209-49.

Suzuki, A., H. Makinoshima, H. Wakaguri, H. Esumi, S. Sugano, T. Kohno, K. Tsuchihara, and Y. Suzuki. 2014. 'Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines', *Nucleic Acids Res*, 42: 13557-72.

Tahiliani, M., K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. 2009. 'Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1', *Science*, 324: 930-5.

Takahashi, K., S. Shichijo, M. Noguchi, M. Hirohata, and K. Itoh. 1995. 'Identification of MAGE-1 and MAGE-4 proteins in spermatogonia and primary spermatocytes of testis', *Cancer Res*, 55: 3478-82.

Thienpont, B., J. Steinbacher, H. Zhao, F. D'Anna, A. Kuchnio, A. Ploumakis, B. Ghesquiere, L. Van Dyck, B. Boeckx, L. Schoonjans, E. Hermans, F. Amant, V. N. Kristensen, K. Peng Koh, M. Mazzone, M. Coleman, T. Carell, P. Carmeliet, and D. Lambrechts. 2016. 'Tumour hypoxia causes DNA hypermethylation by reducing TET activity', *Nature*, 537: 63-68.

Travis, W. D., E. Brambilla, A. G. Nicholson, Y. Yatabe, J. H. M. Austin, M. B. Beasley, L. R. Chirieac, S. Dacic, E. Duhig, D. B. Flieder, K. Geisinger, F. R. Hirsch, Y. Ishikawa, K. M. Kerr, M. Noguchi, G. Pelosi, C. A. Powell, M. S. Tsao, I. Wistuba, and W. H. O. Panel. 2015. 'The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification', *J Thorac Oncol*, 10: 1243-60.

Uhlen, M., L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten. 2015. 'Proteomics. Tissue-based map of the human proteome', *Science*, 347: 1260419.

van der Bruggen, P., C. Traversari, P. Chomez, C. Lurquin, E. De Plaen, B. Van den Eynde, A. Knuth, and T. Boon. 1991. 'A gene encoding an antigen recognized by

cytolytic T lymphocytes on a human melanoma', *Science*, 254: 1643-7.

Van Tongelen, A., A. Loriot, and C. De Smet. 2017. 'Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes', *Cancer Lett*, 396: 130-37.

Voigt, P., W. W. Tee, and D. Reinberg. 2013. 'A double take on bivalent promoters', *Genes Dev*, 27: 1318-38.

Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nat Genet*, 20: 116-7.

Wang, C., Y. Gu, K. Zhang, K. Xie, M. Zhu, N. Dai, Y. Jiang, X. Guo, M. Liu, J. Dai, L. Wu, G. Jin, H. Ma, T. Jiang, R. Yin, Y. Xia, L. Liu, S. Wang, B. Shen, R. Huo, Q. Wang, L. Xu, L. Yang, X. Huang, H. Shen, J. Sha, and Z. Hu. 2016. 'Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types', *Nat Commun*, 7: 10499.

Wang, L., J. Zhang, J. Duan, X. Gao, W. Zhu, X. Lu, L. Yang, J. Zhang, G. Li, W. Ci, W. Li, Q. Zhou, N. Aluru, F. Tang, C. He, X. Huang, and J. Liu. 2014. 'Programming and inheritance of parental DNA methylomes in mammals', *Cell*, 157: 979-91.

Weber, J., M. Salgaller, D. Samid, B. Johnson, M. Herlyn, N. Lassam, J. Treisman, and S. A. Rosenberg. 1994. 'Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine', *Cancer Res*, 54: 1766-71.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan, and D. Schubeler. 2007. 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nat Genet*, 39: 457-66.

Weisenberger, D. J., K. D. Siegmund, M. Campan, J. Young, T. I. Long, M. A. Faasse, G. H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A. J. French, S. N. Thibodeau, J. Jass, R. Haile, and P. W. Laird. 2006. 'CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer', *Nat Genet*, 38: 787-93.

Whitehurst, A. W. 2014. 'Cause and consequence of cancer/testis antigen activation in cancer', *Annu Rev Pharmacol Toxicol*, 54: 251-72.

Williams, K., J. Christensen, and K. Helin. 2011. 'DNA methylation: TET proteins-guardians of CpG islands?', *EMBO Rep*, 13: 28-35.

Wu, H., and Y. Zhang. 2014. 'Reversing DNA methylation: mechanisms, genomics, and biological functions', *Cell*, 156: 45-68.

Wu, X., and Y. Zhang. 2017. 'TET-mediated active DNA demethylation: mechanism, function and beyond', *Nat Rev Genet*, 18: 517-34.

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schubeler, C. Vinson, and J. Taipale. 2017. 'Impact of cytosine methylation on DNA binding specificities of human transcription factors', *Science*, 356.

Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. 'Cytosine methylation and the ecology of intragenomic parasites', *Trends Genet*, 13: 335-40.

Yokoe, T., F. Tanaka, K. Mimori, H. Inoue, T. Ohmachi, M. Kusunoki, and M. Mori. 2008. 'Efficient identification of a novel cancer/testis antigen for immunotherapy using three-step microarray analysis', *Cancer Res*, 68: 1074-82.

You, J. S., and P. A. Jones. 2012. 'Cancer genetics and epigenetics: two sides of the same coin?', *Cancer Cell*, 22: 9-20.

Zhang, L., X. Lu, J. Lu, H. Liang, Q. Dai, G. L. Xu, C. Luo, H. Jiang, and C. He. 2012. 'Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA', *Nat Chem Biol*, 8: 328-30.

Zhang, Y., R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle, and A. Jeltsch. 2010. 'Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail', *Nucleic Acids Res*, 38: 4246-53.

Zhang, Y., R. Wang, Y. Li, Y. Pan, H. Hu, Y. Zhang, H. Li, L. Shen, Y. Yu, Y. Sun, and H. Chen. 2015. 'Negative Thyroid Transcription Factor 1 Expression Defines an Unfavorable Subgroup of Lung Adenocarcinomas', *J Thorac Oncol*, 10: 1444-50.

Zhou, W., H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird, and B. P. Berman. 2018. 'DNA methylation loss in late-replicating domains is linked to mitotic cell division', *Nat Genet*, 50: 591-602.

Zhou, W., M. Y. Fong, Y. Min, G. Somlo, L. Liu, M. R. Palomares, Y. Yu, A. Chow, S. T. O'Connor, A. R. Chin, Y. Yen, Y. Wang, E. G. Marcusson, P. Chu, J. Wu, X. Wu, A. X. Li, Z. Li, H. Gao, X. Ren, M. P. Boldin, P. C. Lin, and S. E. Wang. 2014. 'Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis', *Cancer Cell*, 25: 501-15.

Zylicz, J. J., A. Bousard, K. Zumer, F. Dossin, E. Mohammad, S. T. da Rocha, B. Schwalb, L. Syx, F. Dingli, D. Loew, P. Cramer, and E. Heard. 2019. 'The Implication of Early Chromatin Changes in X Chromosome Inactivation', *Cell*, 176: 182-97 e23.

RESEARCH OBJECTIVES

Tumor cells display genetic, as well as epigenetic alterations compared to their normal counterparts. These two alterations lead to the deregulation of gene expression profiles in these cells. The present study is part of the overall aim of our group to better characterize the consequences of epigenetic deregulations occurring in tumors, and specifically, on their transcriptome.

One of the most extensively studied gene expression alterations in tumors, is the ectopic expression of germline genes. The aberrant activation of this germline-specific gene cluster in tumors is due in part to the loss of DNA methylation in their promoter region. This epigenetic deregulation is associated with the global loss of DNA methylation occurring in virtually all tumor cells. Because of their restricted expression site, the fact that several of these genes were found to elicit an immune response and display oncogenic functions, these genes represent ideal targets for anti-tumor therapies. Studies have consequently focused on better characterizing this group of genes, leading to a constant germline-tissue bias of the studies.

A detailed and tissue-unbiased inventory of the ectopic activation of tissue-specific gene clusters in tumors is still lacking today. The aim of the present work was therefore to **determine whether tumor cells spuriously activate tissue-specific gene clusters,** besides the known germline group of genes.

To address this question, we followed two complementary approaches, and this, without applying any initial restraining criteria on physiological tissue expression sites. First, we took a **mechanistic-centered approach**: we wondered whether the phenomenon of global DNA hypomethylation in tumors could be associated with the promoter DNA demethylation of additional tissue-specific genes, leading therefore, to their transcriptional induction in these cells. Our results revealed that, besides the germline group of genes, two somatic tissue-specific gene clusters are aberrantly activated in tumors: a first gene cluster exhibiting a restricted profile to gastro-intestinal tract tissues, and a second gene cluster showing a specific expression to stratified squamous epithelia. Our findings are described in Chapter 1 of this study.

In light of these results, we then wondered whether additional epigenetic deregulations, other than promoter DNA methylation, could also contribute to the activation of tissue-specific genes in tumor cells. In Chapter 2, we therefore

followed a **transcriptomic-centered approach**, i.e., without restricting our analyses to any regulatory mechanism in tumor cells as a criterion. This analysis led to the identification of a large number of genes, ectopically activated in tumors, with a physiological expression site in brain tissues.

CHAPTER 1

Identification of tissue-specific gene clusters induced by DNA demethylation in lung adenocarcinoma: more than germline genes

Anna Diacofotaki [1], Axelle Loriot [1,2] and Charles De Smet [1]

[1] Group of Genetics and Epigenetics, de Duve Institute, Université Catholique de Louvain; Brussels, Belgium

[2] Group of Computational Biology and Bioinformatics, de Duve Institute, Université Catholique de Louvain ; Brussels Belgium

>> Cancers 2022, 14, 1007. https://doi.org/10.3390/cancers14041007

Simple Summary: Loss of DNA methylation is often observed in human tumors, but how this epigenetic alteration impacts the transcriptome of cancer cells remains largely undefined. So far, DNA hypomethylation in tumors has been associated with aberrant activation of a germline-specific gene expression program. Here, we exploited transcriptomic and methylomic datasets of lung adenocarcinoma to investigate the possibility that other gene expression programs also become ectopically activated in hypomethylated tumors. Remarkably, we found that DNA hypomethylation in lung adenocarcinoma is associated with ectopic activation of not only germline-specific genes, but also gene clusters displaying specific expression in the gastrointestinal tract, or in stratified epithelia. Interestingly, expression of genes in this latter group was of prognostic value. Together, our study brings novel insight into the transcriptomic changes associated with DNA hypomethylation in tumors, and is an incentive to explore the value of hypomethylated DNA sequences as cancer biomarkers.

Abstract: Genome-wide loss of DNA methylation is commonly observed in human cancers, but its impact on the tumor transcriptome remains ill-defined. Previous studies demonstrated that this epigenetic alteration causes aberrant activation of a germline-specific gene expression program. Here, we examined if DNA hypomethylation in tumors also leads to de-repression of gene clusters with other tissue specificities. To this end, we explored transcriptomic and methylomic datasets from human lung adenocarcinoma (LUAD) cell lines, normal lung, and lung alveolar type II cells, considered as the origin of LUAD. Interestingly, DNA demethylation in LUAD cell lines was associated with activation of not only germline-specific (CG) genes, but also gene clusters displaying specific expression in the gastrointestinal tract (GI), or in stratified epithelia (SE). Consistently, genes from all three clusters showed highly specific patterns of promoter methylation among normal tissues and cell types, and were generally sensitive to induction by a DNA demethylating agent. Analysis of TCGA datasets confirmed that demethylation and activation of CG, GI and SE genes also occurs in vivo in LUAD tumor tissues, in association with global genome hypomethylation. For genes of the GI cluster, we demonstrated that HNF4A is a necessary factor for transcriptional activation following promoter demethylation. Interestingly, expression of several SE genes, in particular *FAM83A*, correlated with both tumor grade and reduced patient survival. Together, our study uncovers novel cell-type specific gene clusters that become aberrantly activated in LUAD tumors in association with genome-wide hypomethylation.

Keywords: *epigenetics; DNA methylation; DNA hypomethylation; cancer-germline genes; lung adenocarcinoma*

1. Introduction

Cell type-specific patterns of gene expression are in part determined by epigenetic mechanisms, involving chemical modifications of DNA and associated histones, which modulate the accessibility of transcription factors to the chromatin (Bird 2002; Jenuwein and Allis 2001). Once installed, epigenetic modifications are maintained through cell divisions, and exert thereby durable effects. Methylation of CpG dinucleotides in the DNA, for instance, has been associated with a closed chromatin structure, and hence long-term repression of gene transcription (Bird 2002). It was therefore proposed that DNA methylation contributes to the establishment of specific gene expression programs during differentiation of the various cell lineages (Riggs 1989).

It is now clear that cancer development is driven in part by epigenetic dysregulations, which cause changes in gene expression patterns (Baylin and Jones 2011). Loosening of gene regulatory processes is thought to confer increased plasticity to cancer cells, thereby accelerating their progression towards a state of enhanced malignancy (Feinberg 2014).

Global loss of DNA methylation marks was the first epigenetic alteration to be described in human tumors (Feinberg and Vogelstein 1983b; Gama-Sosa et al. 1983b). Intriguingly, DNA hypomethylation in tumors has been merely associated with ectopic activation of a cluster of germline-specific genes, including more than 100 testis-specific genes being aberrantly expressed in a wide variety of tumor types (De Smet and Loriot 2013). Studies confirmed that such “cancer-testis” (CT) or “cancer-germline” (CG) genes rely primarily on DNA methylation for repression in non-expressing cells, and that genome demethylation is a sufficient trigger for their concerted activation in most tumors (Cannuyer et al. 2013a; De Smet, Lurquin, Lethé, et al. 1999; Grunwald et al. 2006; Sigalotti et al. 2002). Due to their highly restricted pattern of expression, CG genes were exploited as biomarkers in cancer detection, and as antigenic determinants in anti-cancer vaccinations (Akers, Odunsi, and Karpf 2010). Moreover, evidence has accumulated indicating that some CG genes contribute to tumor development, notably by regulating processes of cell proliferation, death resistance, metabolic adaptation, and DNA repair (Van Tongelen, Loriot, and De Smet 2017; Whitehurst 2014).

While DNA hypomethylation in tumors is evidently associated with aberrant activation of a large group of germline-specific genes, it is less clear if it can also lead to de-repression of gene clusters that normally show restricted expression in somatic tissues. Several examples of genes with specific expression in somatic

tissues were shown to become activated in tumors in association with promoter demethylation (Ehrlich 2009b; Szyf, Pakneshan, and Rabbani 2004). However, the few genes identified could not be grouped into tissue-specific clusters, and their activation in tumors appeared less stringently associated with DNA demethylation, as compared with CG genes (Dokun et al. 2008). In fact, the very notion that DNA methylation can serve as a primary mechanism of regulation for genes with specific expression in somatic tissues remains a matter of debate (Bestor, Edwards, and Boulard 2015).

In the present study, we decided to reinvestigate the possibility that DNA hypomethylation in tumors might be associated with ectopic activation of not only a germline-specific expression program, but also of somatic expression programs. To this end, we screened publicly available transcriptomic and methylomic datasets derived from human lung adenocarcinoma (LUAD) cell lines, normal lung tissues, and alveolar type II (AT2) cells, the cells at the origin of adenocarcinoma of the lung (Blanpain 2013). Computational analyses were performed on these datasets to identify transcripts showing coincident transcriptional de-repression and promoter demethylation in cancer cells. Of note, compared with other studies, our screening procedure did not take into account up-regulated transcripts, but focused on transcripts that are initially absent in normal cells and become induced de novo in tumor cells. The tissue specificity of the isolated transcripts was then determined by analyzing transcriptomic and methylomic databases of normal human tissues and cells. Interestingly, besides germline-specific genes, we identified other clusters of tissue-specific genes, and confirmed their activation in vivo in hypomethylated LUAD tumor tissues. The role of DNA methylation and transcription factors in regulating such genes was evaluated experimentally. Finally, we investigated if expression of the genes we identified was associated with tumor grade and patient survival in LUAD.

2. Materials and Methods

2.1. LUAD Cell Lines Datasets

For expression and methylation analyses in the 26 LUAD cell lines, FastQ files of bulk RNA-seq and methyl-seq were obtained from Suzuki and colleagues (Suzuki et al. 2014). RNA-seq FastQ files were processed as previously described (Fain et al. 2019), to obtain expression levels of individual referenced and unreferenced transcripts in each cell line. For methylation analysis, read quality control and trimming of low-quality reads were performed using Trim Galore! software v0.5.0

(https://github.com/FelixKrueger/TrimGalore) (accessed on 10 February 2022). Read alignment and methylation calling were done using Bismark v0.20.0 (Krueger and Andrews 2011). LUAD cell lines were considered positive for the expression of a gene when the TPM was ≥2, and negative when it was <1. All accession numbers are listed in the Supplementary Table S1.

2.2. Procedure for the Identification of DDIC and Non-DDIC Transcripts in LUAD Cell Lines

1) Merging of transcripts originating from the same promoter region in LUAD cell lines. Transcripts arising from a transcriptional start site (TSS) located less than 50bp apart were considered to originate from the same promoter region. Their expression levels were therefore summed for transcript quantification. For each pooled transcript group, the TSS located at the closest 5' end was chosen as reference.

2) Selection of repressed transcripts in normal lung. Transcripts that were initially silent in AT2 cells (RNA-seq data, (Marconett et al. 2017)) and normal lung (RNA-seq, Roadmap Epigenomics (Roadmap Epigenomics et al. 2015)), and exhibited a methylation level of their promoter region ≥60% in normal lung (WGBS, Roadmap Epigenomics) and AT2 cells (WGBS, (Zuber et al. 2016)) were retained. Transcript repression was defined as TPM <1 and/or lack of RNA-seq reads mapping to the genomic region of each transcript. The level of methylation of the promoter region was computed by averaging the methylation values of each CpG located at TSS -/+ 400bp for each transcript. Only genomic regions harboring at least 3 CpGs were considered for this analysis, and when coverage information was available, CpGs covered by at least 2 reads were analyzed.

3) Selection of activated transcripts in LUAD cell lines. Among the previous transcript selection, we selected those that fulfilled the following criteria: first, transcripts were repressed in at least 2/26 cell lines (expression quantile $10 \leq 0.1$ TPM) and activated in at least 1/26 cell line with their maximum expression $\geq$2 TPM. Secondly, the minimum value of the level of methylation of their promoter region was below 40% and the maximum above 60%. These transcripts were considered as being repressed in normal lung tissue and ectopically expressed in LUAD cell lines.

4) Correlation between transcriptional activation and methylation of promoter region. We computed a Pearson correlation of the expression and promoter methylation status of each transcript in LUAD cell lines. At least 7/26 LUAD cell lines had to have both transcript expression and promoter methylation information available for a correlation to be computed. We selected transcripts that showed a

significant inverse correlation between their transcriptional activation and promoter methylation ($r \leq -0.4$, p-value < 0.05).

5) Manual curation. We visualized BAM files of the LUAD cell lines using IGV (Integrative Genomics Viewer, (Robinson et al. 2011)) to determine the accuracy of the chosen TSS, confirm the transcriptional activation status, and determine transcript backbone for each of the selected transcripts. Finally, we also verified the correlation between expression and promoter methylation of each transcript by generating heatmaps (ComplexHeatmap R package v2.8.0) depicting the methylation state of each CpG located at TSS -/+ 400bp. Transcripts that passed all the above filters were referred to as DNA-Demethylation-associated Induction in Cancer transcripts (DDIC transcripts). Non-DDIC transcripts were selected using the same procedure with different criteria for the following parameters: transcripts were originating from a gene giving rise only to that transcript (i.e., not alternative gene promoters), the corresponding gene is repressed in normal lung (lung tissue samples of GTEx show a median TPM <0.5 and the one from Roadmap Epigenomics <1 TPM) and repressed in AT2 cells (TPM <1). They showed no correlation between transcriptional activation and promoter demethylation ($-0.1 \leq r \leq 0.1$). No manual curation step was performed for non-DDIC transcripts.

2.3. RNA-Seq Public Datasets

1) Bulk RNA-seq. Normalized expression data of samples coming from 50 normal tissues were obtained from GTEx portal (v8, (Consortium 2013)). Both gene expression of individual samples per tissue and median expression of all samples per tissue were downloaded. FastQ files of the following cells and tissues were downloaded from Sequence Read Archive (SRA): AT2, lung, testis, sigmoid colon, small intestine, stomach, pancreas, liver, esophagus, skin, adipose, cerebral cortex, heart and thyroid. Files were processed as previously described (Fain et al. 2021). All accession numbers are listed in Supplementary Table S3.

2) Single-cell RNA-seq. Normalized expression data of skin single-cell study (scRNA-seq, (Sole-Boldo et al. 2020)) were obtained from the Human Protein Atlas (HPA) v20.1 (Uhlen et al. 2005). All accession numbers are listed in Supplementary Table S3.

2.4. RNA-Seq of Cell Lines after DNA Demethylation Treatment

FastQ files of MCF7 breast cancer cell line, HMLER immortalized mammary epithelial cells, and TS603 glioma cell line, treated or not with a demethylating agent, were downloaded from SRA (D'Alessio, Weaver, and Szyf 2007; Grandin et al. 2016; Park et al. 2021). Read quality control was performed using FastQC

software (v0.11.8) and low-quality reads were discarded using Trimmomatic software v0.38 (Bolger, Lohse, and Usadel 2014). Reads were aligned onto hg38 genome using HISAT2 v2.1.0 (Kim et al. 2019), and gene quantification was carried out with featureCounts software from the subread package v2.0.0 (Liao, Smyth, and Shi 2014). Expression data are depicted in TPM. For differential expression analysis, gene expression counts generated by featureCounts of each cell line were used as input for DESeq2 (v1.32.0) R package (Love, Huber, and Anders 2014). Genes with an adjusted *p*-value < 0.05 were considered differentially expressed between conditions. All accession numbers are listed in Supplementary Table S4.

2.5. DNA Methylation Public Datasets

1) Normal tissues. For cerebral cortex samples, FastQ files of whole genome bisulfite-seq (WGBS) were downloaded from SRA. Trim Galore! software v0.5.0 was used to read quality control and trimming of low-quality reads. Read alignment and methylation calling were performed using Bismark v0.20.0. For liver sample, BAM file of WGBS was downloaded from ENCODE consortium database (Consortium 2012), and methylation calling was performed using Bismark v0.20.0. For the rest of the normal tissues, normalized hg38 WGBS data were obtained from the ENCODE consortium database.

2) Primary cells. For keratinocytes and sperm cells, WGBS hg19 normalized data were obtained from SRA (Roadmap Epigenomics et al. 2015). Data were converted to hg38 coordinates using liftOver v1.10.0 R package. Normalized hg38 WGBS data of HUES64 were obtained from the ENCODE consortium database. For AT2 cells, FastQ files of WGBS were downloaded from SRA (Zuber et al. 2016). Trim Galore! software v0.5.0 was used to read quality control and trimming of low-quality reads. Read alignment onto hg38 genome and methylation calling were performed using Bismark v0.20.0. For all the above studies, when available, methylation information for the same CpG sequenced in forward and reverse strand were averaged. For duodenal crypt cell samples, normalized Infinium HumanMethylation 450 assays and EPICarrays data were obtained (Lewis et al. 2020). All accession numbers are listed in Supplementary Table S3.

2.6. Immunohistochemical Data of Normal Tissues

Immunohistochemical images of the esophagus, skin and vagina were obtained from the HPA. Antibodies for the detection of A2ML1 and SERPINB5 proteins were selected based on the fact that immunohistochemical staining patterns mirrored the gene expression profile as established by RNA-seq studies of GTEx and HPA consortia. Antibodies and tissue codes are listed in Supplementary Table S5.

2.7. Cell Culture

LXF289 lung adenocarcinoma cell line was purchased from CLS Standard and cultured in RPMI medium (CLS Standard). SKMEL23 melanoma cell line, and LB996RCC renal carcinoma cell line were obtained from the Brussels branch of the Ludwig Institute Cancer Research, and were cultured as previously described (Cannuyer et al. 2013a; De Smet, Lurquin, Lethé, et al. 1999). HEK293 human embryonic kidney cells were purchased from Thermo Fischer and cultured in DMEM (Life Technologies, Carlsbad, CA, USA). HFF2 foreskin fibroblasts were kindly provided by Dr Decottignies (de Duve Institute, UCLouvain) and cultured in DMEM. All media were supplemented with 10% Fetal Bovine Serum (Sigma, Burlington, MA, USA) and 1% penicillin/streptomycin (Life Technologies).

2.8. Cell Treatment with 5-azadC and siRNAs

For 5-azadC treatments, tumor and normal cell lines were grown to 60–70% confluency and then treated with a single dose of 2 µM of 5-aza-2′-deoxycytidine diluted in 1:1 acetic acid/water. Cells were maintained in culture for 6 days and then harvested for RNA extraction. For combined treatment with 5-azadC and siRNAs, LXF289 cells were seeded at 200,000 cells/well of a 6w plate and were reverse transfected with siRNAs directed against HNF4A (siHNF4A), or control siRNAs directed against luciferase (siLuc, (Tilman et al. 2012)) at a final concentration of 100 nM, using Lipofectamine 2000 (Invitrogen). The following day, cells were treated with either 2 µM 5-azadC or 1:1 acetic acid/water as a control. Cells were harvested at day 3 and day 5 post-transfection for RNA and protein analysis. siRNA sequences are listed in Supplementary Table S6.

2.9. RT-PCR and qPCR Analyses

RNA of normal tissue samples (lung, testis, esophagus and colon) was purchased from Ambion (Life Technologies). RNA of normal and tumor cell lines was extracted using TriPure isolation reagent (Roche, Basel, Switzerland). Reverse transcription was performed using MML-V reverse transcriptase kit (Invitrogen), random hexamers (Invitrogen), Ribolock RNAse inhibitor (Invitrogen, 20U) and 2 µg of total RNA per reaction in a final volume of 20 µL. PCR analyses were carried out using DreamTaq kit (ThermoFischer Scientific, Waltham, MA, USA) with 1/40 of the reverse transcription solution engaged per reaction in a total volume of 20 µL. qPCR analyses were performed using KAPA SYBR FAST kit (Sigma-Aldrich) with 1/40 of the reverse transcription solution engaged per reaction in a final volume of 10 µL. All reactions were carried out according to the manufacturer's instructions. Primer sequences and experimental conditions are listed in Supplementary Table S7. Different positive control cDNAs were used depending on the gene: testis (*NAP1L1*,

NAA11, *MAGEA1*, *ACTB*), colon (*EPS8L3*, *VIL1*), esophagus (*GJB5*, *SERPINB5*, *PGLYRP3*).

2.10. Western Blot

At day 3 post-transfection LXF289 cells were harvested for protein extraction. Cells were lysed and proteins recovered using RIPA buffer (150 mM NaCl, 1% TritonX-100, 0.5% sodium deoxycholate, 0.1% SDS and 50 mM Tris pH8) supplemented with 1× cOmplete Mini protease inhibitor cocktail (Roche) and 1× PhosSTOP phosphatase inhibitor cocktail (Roche). 30 µg of proteins were loaded onto an 8% SDS-PAGE gel and subsequently transferred for 2 h, 240 mA onto a PVDF membrane (Millipore, Burlington, MA, USA). The membrane was blocked with 5% milk-TBS-Tween 0.1%, 1 h at RT and then incubated overnight with anti-HNF4a antibody (1:1000 CST3113; diluted in 5% milk-TBS-Tween 0.1%). The membrane was then washed 3× in TBS-Tween 0.1% and incubated at RT 1 h with a goat anti-rabbit antibody conjugated to HRP (1:10000 Enzo Life Sciences, Farmingdale, NY, USA, ADI-SAB-300J, diluted in TBS-Tween 0.1%). The membrane was washed 3× in TBS-Tween 0.1% and revealed using SuperSignal West Pico PLUS Chemiluminescent substrate (ThermoFisher) and CL-Xposure films (Life Technologies). For Vinculin reveal, membrane was stripped with 4× NaOH 0.4M solution, blocked 1 h at RT with 5% milk-TBS-Tween 0.1%, and incubated for 1 h with anti-VCL antibody (1:100000, Millipore 05386). A goat anti-mouse secondary antibody conjugated to HRP (1:10000 ab205719 diluted in TBS-Tween 0.1%) was used to reveal the membrane as described above.

2.11. The Cancer Genome Atlas Consortium Datasets and Analyses

Bulk RNA-seq, hg19 Infinium HumanMethylation450 assay of LUAD tumor samples, and LUAD patients' vital status were obtained through TCGAbiolinks R package v2.14.1 (Colaprico et al. 2016). A list of pathology grades of TCGA-LUAD tumor samples data was compiled and kindly provided by Drs Yu and Snyder (Stanford University, Stanford, CA, USA, (Yu et al. 2016)).

1) Expression analysis. FPKM expression data were converted to TPM by dividing each FPKM gene expression in a sample by the sum of all gene expressions for that sample ($\times 10^6$). LUAD tumor samples were considered positive for the expression of a gene if they exhibited a TPM $\geq$2. They were considered negative when they showed a TPM <1. To define activating and non-activating LUAD tumor groups for each DDIC, expression quantile 20 and 80 of each gene were used as activation cut-offs (i.e., tumor samples that show a DDIC expression TPM <q20 were considered as repressing, and conversely, when TPM >q80 were considered as activating).

When q80 value was <1 TPM, then all tumor samples showing a gene expression ≥1 TPM were considered positive.

2) Tumor grade analysis. Tumor grades were compared in the activating and non-activating tumor groups using a Chi-squared test. Comparison was computed if there were at least 15 tumors in each group. LUAD samples qualified as grades 1.5 and 2.5 were categorized as grade 1 and grade 2, respectively.

3) Survival analysis. Patient median survival time were compared between the activating and non-activating LUAD tumor groups using a Log rank test from the survival R package v3.2-11. At least 15 tumor samples had to constitute each group, to compare survival time between patients. When median survival time was not reached for a tumor group, the length of the TCGA-LUAD survival study was taken as the median survival time.

4) Methylation analysis. Infinium CG probes in hg19 coordinates were converted to hg38 coordinates using liftOver v1.10.0 R package. For correlation analyses between gene expression and promoter DNA methylation, LUAD tumor samples that exhibited information for both expression and promoter methylation were analyzed. CG probes located at -/+ 400bp of TSS of each gene were averaged to define the level of methylation of the promoter region. Pearson correlation was computed for each gene.

5) Global DNA methylation analysis. Autosomal probes that showed an average of methylation higher than 0.7 in all normal lung samples (n = 32) were selected as a proxy for the assessment of global methylation levels in each tissue sample (n = 137,954 probes). To assess promoter methylation status in regard to global methylation levels, we defined two tumor subgroups based on the promoter methylation values of each DDIC in all LUAD tumor samples. We considered a first group of tumor samples that showed a methylated promoter region of each gene (i.e., the promoter methylation level was ≥quantile 80 of methylation values for that gene) and a second group that showed a demethylated promoter of each gene (promoter methylation ≤ quantile 20). Then, the global methylation level was compared between these two tumor groups using a Student's *t*-test.

2.12. Statistical Analysis and Graphical Representations

Statistical analysis was computed in R v4.1.0 (http://www.R-project.org) (accessed on 10 February 2022) or GraphPad Prism (v5.0). All *p*-values were adjusted using the Benjamini-Hochberg method. For clustering of DDIC expression in normal tissues of GTEx and LUAD samples, Ward's method with Euclidean distance were used. DDIC genes were clustered inside their tissue specific categories using the

same parameters. Clustering results are represented using the ComplexHeatmap R package (v2.8.0).

3. Results

3.1. Search for Transcripts Showing DNA Demethylation-Associated Induction in LUAD Cell Lines

In order to identify transcripts that are induced in lung cancer cells in association with promoter DNA demethylation, we conducted an integrative analysis of transcriptomic and methylomic datasets from LUAD cell lines (n = 26, DBTSS, Supplementary Table S1), normal lung (n = 1, Roadmap Epigenomics) and lung AT2 cells (n = 1, (Zuber et al. 2016)). A selection procedure was applied, using the following criteria (**Figure 1A**): (i) transcripts are not expressed and their promoter region is methylated in normal lung and AT2 cells; (ii) their expression is observed in a fraction of the LUAD cell lines (at least one cell line) and is inversely correlated with their promoter DNA methylation level; (iii) accuracy of expression specificity, TSS position, and the structure of the transcripts was validated after visualization of RNA-seq data with the Integrative Genome Viewer (IGV). This led to a final list of 171 transcripts that show transcriptional repression and promoter DNA methylation in normal lung and AT2 cells, but are transcriptionally induced in association with promoter demethylation in LUAD cell lines (**Figure 1A, B; Supplementary Table S2**). Transcripts with this profile were qualified as "DDIC" (*DNA Demethylation-associated Induction in Cancer*). Of note, transcripts that showed ectopic activation in LUAD cell lines, but without coincident loss of promoter DNA methylation (n = 169) were retained as a control group for subsequent analyses, and termed non-DDIC (*non-DNA Demethylation-associated Induction in Cancer*, **Figure 1A, B; Supplementary Table S2**).

Among the 171 DDIC transcripts, 131 (77%) corresponded to previously referenced transcript variants of known genes, 19 (11%) to unreferenced transcript variants of known genes and 22 (12%) to unreferenced transcripts originating from previously undescribed genes (**Supplementary Figure S1**). Together, these analyses led to the identification of a set of transcripts that are induced in LUAD cells in association with DNA demethylation of their promoter region.

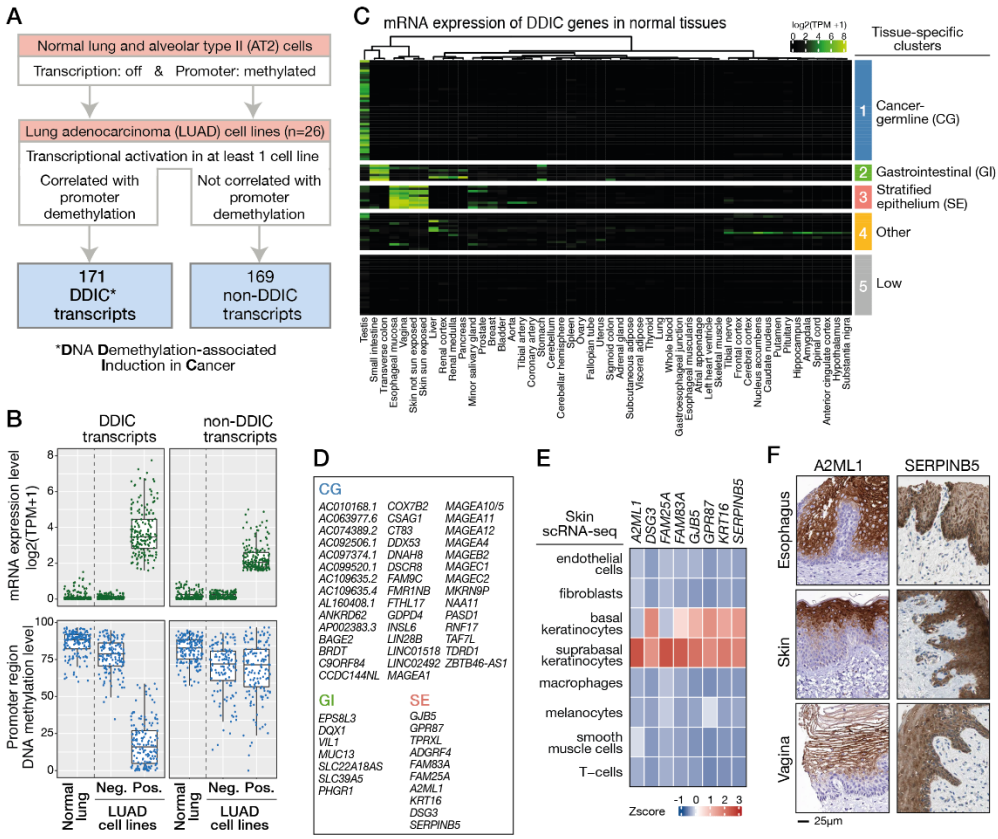


Figure 1. DNA demethylation in LUAD cell lines is associated with transcriptional induction of multiple tissue-specific gene expression programs. *(A) Workflow for the identification of genes exhibiting DNA demethylation-associated induction (DDIC genes) in LUAD cell lines. (B) Levels of mRNA expression (RNA-seq) and promoter region methylation (Methyl-seq) of isolated DDIC and control non-DDIC transcripts were evaluated in normal lung and LUAD cell lines. For each transcript, LUAD cell lines were divided in two subgroups according to the status of expression of the transcript (Positive or Negative). Dots represent mean mRNA level (green) and mean promoter methylation level (blue) of a single transcript in the corresponding subgroup of LUAD cell lines. (C) Median expression of DDIC genes in normal tissue samples (GTEx). Unsupervised hierarchical clustering was used to categorize DDIC genes according to their specificity of expression in normal tissue samples. (D) List of DDIC genes belonging to CG, GI and SE categories. (E) single-cell RNA-seq data (HPA) were used to determine relative expression levels of SE-DDIC genes in cells composing the skin tissue. (F) Immunohistochemical images from the HPA, which were available for two SE-DDIC proteins (A2ML1 and SERPINB5), show expression of these proteins in the stratified epithelium of the esophagus, skin and vagina.*

3.2. DDIC Genes Belonging to Tissue-Specific Expression Programs

We next determined the pattern of expression of DDIC transcripts among normal tissues. To this end, we explored RNA-seq data from a series of normal human tissues (GTEx). To avoid confounding interpretations which may occur from genes harboring multiple transcript variants, only DDIC genes producing a single transcript were retained for the analysis (n = 103). Unsupervised hierarchical clustering identified three clusters of DDIC genes exhibiting restricted expression among normal tissues. The largest cluster of DDIC genes (Cluster 1, n = 44) displayed specific or preferential expression in testis (**Figure 1C, D**), and more specifically in testicular germ cells according to single cell RNA-seq data (**Supplementary Figure S2**). Many of these genes corresponded to previously characterized CG genes. Another cluster of DDIC genes (Cluster 2, n = 7) displayed maximal expression in the small intestine and colon (**Figure 1C, D**), suggesting that DNA demethylation might contribute to induce a gastrointestinal (GI) gene expression program in LUAD cells. A third cluster of DDIC genes was identified (Cluster 3, n = 10), which showed preferential expression in esophagus, skin and vagina (**Figure 1C, D**). A common characteristic of these seemingly unrelated tissues is that they are composed of a multilayered stratified epithelium (Billingham and Silvers 1963). Consistently, single-cell RNA-seq (skin) and immunohistochemical data (esophagus, skin and vagina), which were available from the Human Protein Atlas, revealed specific expression of genes and proteins belonging to cluster 3 in the epithelial cell layers (**Figure 1E, F**). This suggested therefore that DDIC genes in Cluster 3 belong to an expression program associated with the formation of stratified epithelia (SE). Notably, expression of the lung alveolar cell marker *NKX2-1* was detected in tumors expressing GI- or SE-DDIC genes, indicating that these genes do not define subgroups of tumors originating from a different type of normal precursor cells (**Supplementary Figure S2**). Finally, DDIC genes that did not belong to cluster 1, 2 or 3, showed either disparate patterns of expression among analyzed tissues (Other, n = 16), or expression levels that were below the defined threshold (<2 TPM) in all analyzed tissues (Low, n = 26).

Together, these observations suggested that DNA demethylation in tumor cells activates not only a germline expression program (CG), but might also be associated with induction of genes belonging to a gastrointestinal (GI) and a stratified epithelium (SE) transcription program.

3.3. DDIC Promoters Show Tissue- and Cell Type-Specific DNA Demethylation

Genes regulated by DNA methylation are expected to exhibit matching patterns of expression and promoter demethylation in normal tissues. We searched to determine if this is the case for DDIC genes of the CG, GI, and SE clusters. First, the mean number of CpG sites comprised in the promoter region (TSS -/+ 400bp) of DDIC genes was calculated (**Figure 2A**). DDIC genes in the GI and SE categories showed a lower density of CpGs in their promoter region, as compared with CG-DDIC genes, but nevertheless contained a mean amount of 14 CpGs within the 800 bp segment. Analysis of transcriptomic and methylomic data of normal human tissues (ENCODE database) revealed that DDIC genes categorized in the CG, GI, or SE clusters exhibited reduced DNA methylation levels of their promoter region in the tissues where they are expressed, thereby supporting a link between DNA demethylation and expression for these genes (**Figure 2B**).

We further explored the relationship between DNA demethylation and activation of CG, GI and SE transcription programs at the cellular level, by analyzing methylomic data derived from specific cell types, including human lung AT2 cells, embryonic stem cells, spermatozoa, skin keratinocytes and duodenum crypt cells (**Figure 2C**). In AT2 and embryonic stem (ES) cells, the promoter region of all three categories of DDIC genes exhibited comparably high levels of DNA methylation. In spermatozoa, as expected, the mean promoter DNA methylation levels of CG genes appeared much lower than that of the two other gene categories. In skin keratinocytes, which compose the stratified epithelium of the epidermis, the lowest level of DNA methylation was instead observed for DDIC genes belonging to the SE cluster. Genes of the GI cluster, on the other hand, showed lowest DNA methylation levels in intestinal (duodenum) crypt cells. Together these results revealed that DDIC genes of the CG, GI and SE categories exhibit highly distinct DNA demethylation patterns among normal cell types, and that these match with their tissue-specific expression profiles.

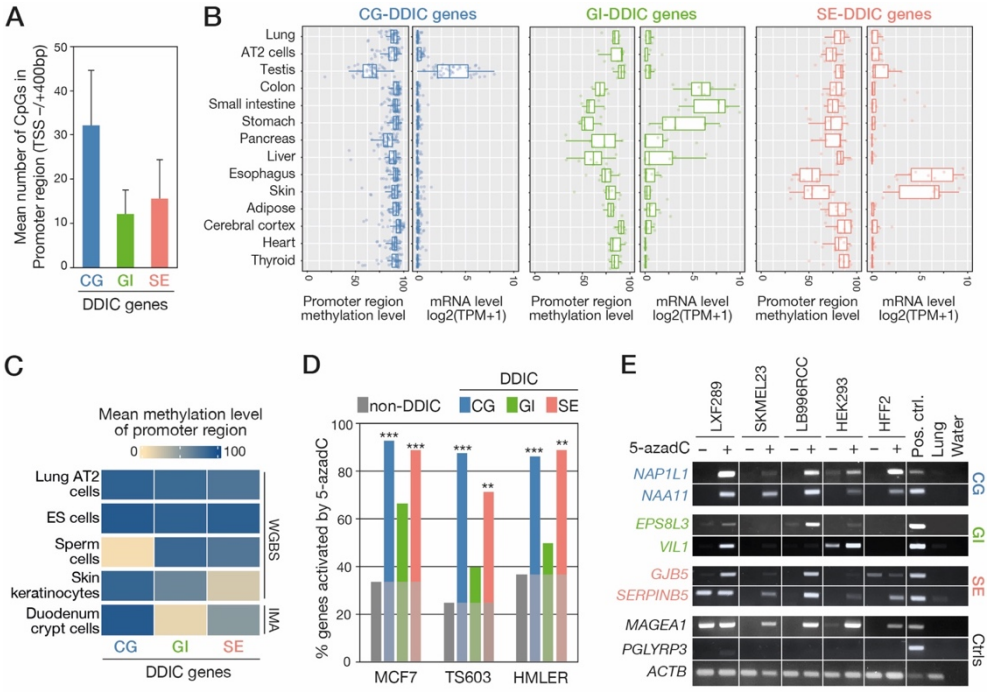


Figure 2. Promoter DNA methylation is involved in the regulation of DDIC genes of the CG, GI and SE clusters. *(A) Mean number of CpGs in the promoter region of DDIC genes of the indicated categories (mean ± SD). (B) Transcriptomic and methylomic data from the ENCODE database were used to compare mean levels of mRNA expression and promoter methylation of DDIC genes in a panel of normal tissues. Each dot represents the value for a single DDIC gene of the indicated category. (C) Publicly available methylomic data (whole genome bisulfite sequencing, WGBS; Infinium Methylation Assay, IMA) were examined to compare mean promoter methylation levels of CG-, GI- and SE-DDIC gene clusters in the indicated cell types. (D) RNA-seq data obtained from tumor cell lines (MCF7, TS603) and a mammary epithelial cell culture (HMLER) exposed to 5-azadC were explored, and percentages of genes in each DDIC category that were significantly upregulated by the demethylating agent were measured, and compared to non-DDIC genes. Fischer's exact test. ***: p < 0.001; **: p < 0.01. (E) RT-PCR experiments were performed to evaluate induction of indicated DDIC genes in 5-azadC-treated tumor cell lines or normal fibroblast culture (HFF2). Control genes included a previously described CG gene (MAGEA1), a non-DDIC gene (PGLYRP3), and a ubiquitously expressed gene (ACTB).*

3.4. Experimental Evaluation of the Role of DNA Methylation in Regulating DDIC Transcripts

In order to establish a direct role of DNA demethylation in the transcriptional induction of DDIC transcripts, we decided to explore RNA-seq data obtained from human cell lines that had been exposed to the DNA demethylating agent 5-aza-

deoxycytidine (5-azadC). The analyzed datasets derived from three different cell lines: the TS603 glioblastoma cell line, the MCF7 breast cancer cell line, and the HMLER transformed mammary epithelial cell line [34,45,46]. We evaluated the frequency of transcriptional induction of DDIC transcripts of the CG, GI, and SE categories upon 5-azadC treatment in these cells (**Figure 2D, Supplementary Table S2**). For comparison, we also examined the impact of 5-azadC on non-DDIC transcripts, the expression of which was not correlated with promoter demethylation in LUAD cells (**Figure 1A, B**). Contrasting with the low proportion (25–37%) of non-DDIC transcripts exhibiting upregulation following 5-azadC, a large proportion (86–93%) of DDIC transcripts in the CG category showed significant induction in treated cells (**Figure 2D**). This was consistent with the well demonstrated role of DNA methylation in the regulation of CG genes. Interestingly, DDIC transcripts of the SE group also showed a high frequency of induction upon 5-azadC treatment (71–89%), thereby supporting an important role of DNA methylation in their regulation. For DDIC transcripts of the GI category, we observed an intermediate frequency of 5-azadC induction (40–67%), suggesting a less direct impact of DNA methylation on their regulation.

To verify the validity of these in silico observations, we conducted 5-azadC induction experiments in four tumor cell lines, as well as in normal cultured fibroblasts, and assessed the expression of several DDIC genes by RT-PCR. The results confirmed that 5-azadC treatment was more often associated with induction of CG- and SE-DDIC genes, than with GI-DDIC genes (**Figure 2D**). Contrastingly, 5-azadC had very little effect on the expression of a control non-DDIC gene. Together, the results indicate that DNA methylation acts as an essential component in the transcriptional regulation of DDIC genes belonging to the CG and SE categories, whereas it appeared to play a more auxiliary role in the control of expression of GI-DDIC genes.

3.5. Validating DNA Demethylation-Associated Induction of DDIC Genes In Vivo in LUAD Tumors

We next searched to verify if the DDIC genes we identified in LUAD cell lines also become induced in vivo in LUAD tissue samples, and whether this is also correlated with promoter demethylation. We first explored RNA-seq datasets of The Cancer Genome Atlas (TCGA), which were obtained from large series of LUAD (n = 510) and normal lung (n = 58) tissue samples (Cancer Genome Atlas Research 2014). Analysis of the data confirmed absent or very low mRNA expression of all DDIC genes in normal lung tissues, and significant induction in multiple tumor samples (**Figure**

3A). Immunohistochemical data which were available for several of the corresponding proteins (Human Protein Atlas) indicated that induction of DDIC genes in LUAD was associated with expression of the protein (**Supplementary Figure S4**). DDIC genes of the GI and SE clusters showed a tendency to be induced in a higher proportion of LUAD samples, as compared with CG genes (**Figure 3A, B; Supplementary Table S2**). However, in the tumors where they were induced, all three categories of DDIC genes displayed similar ranges of mRNA expression levels (**Figure 3B**). Unsupervised hierarchical clustering did not reveal groups of tumors expressing exclusively one or another category of DDIC genes.

To investigate the role of DNA demethylation in the induction of CG-, GI- and SE-DDIC genes in LUAD tissues, we analyzed methylomic data (Infinium methylation assay, IMA) from the TCGA. IMA probes interrogating the methylation status of the promoter region were available for 48 out of the 61 CG-, GI, or SE-DDIC genes. For each of these genes, the mRNA expression level and promoter region DNA methylation level were compared in the different LUAD samples, and a correlation coefficient (Pearson) was calculated (**Figure 3C, D; Supplementary Table S2**). Consistent with a role for DNA demethylation in transcriptional induction, DDIC genes generally displayed a negative correlation coefficient (mean = −0.51; **Figure 3D**). Contrastingly, control non-DDIC genes exhibited a mean correlation coefficient close to 0.

Together, these results confirm that induction of DDIC genes also occurs in vivo in LUAD tissues, and is associated with DNA demethylation of the promoter region. For several genes of the GI cluster, however, we observed that in a fraction of the tumors, mRNA expression remained absent or low despite DNA demethylation of the promoter region (**Figure 3C**), thereby suggesting that other factors might be necessary for transcriptional induction of these genes.

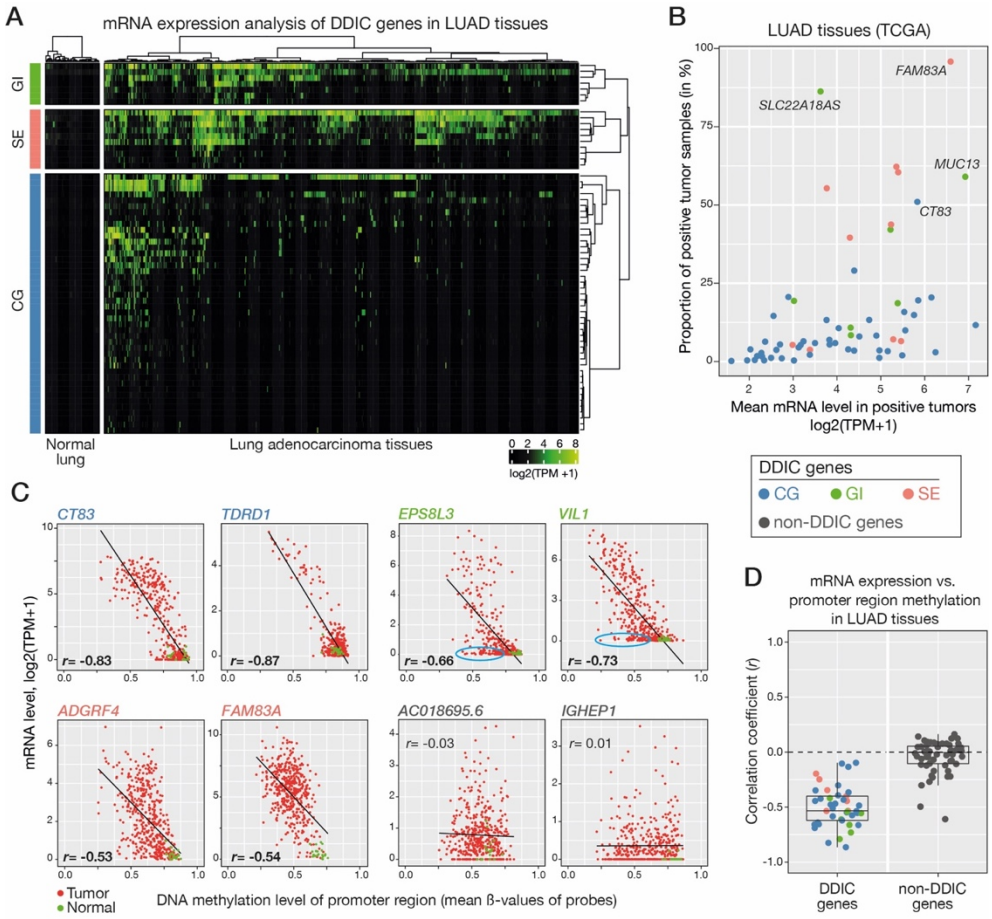


Figure 3. DDIC gene clusters are induced in vivo in LUAD tissues in association with promoter demethylation. *(A) RNA-seq from TCGA were explored to evaluate expression levels of CG, GI and SE gene clusters in LUAD (n = 510) and normal lung tissues (n = 58). A heatmap was generated, where samples are clustered according to DDIC gene expression profiles. (B) Frequency of transcriptional induction in LUAD samples, and mRNA level in positive LUAD samples (each dot represents a single gene with the indicated color code for categories). (C) Correlations between mRNA expression and promoter demethylation of DDIC genes in LUAD tissues: representative plots are shown, including two non-DDIC genes (red dots: tumor samples, green dots: normal lung tissues). Pearson's correlation coefficients (r) and linear regression lines are shown. Tumor samples showing demethylation of the gene promoter without transcriptional induction are encircled in blue. (D) Correlation coefficients, as determined in panel C, were obtained for all CG-, GI- and SE-DDIC genes and compared with that of non-DDIC genes.*

3.6. Global Genome Hypomethylation is Associated with Local Demethylation of DDIC Genes

It has been demonstrated previously for CG genes that local DNA demethylation of their promoter region in tumors coincides with a process of global genome demethylation (Cadieux et al. 2006; De Smet et al. 1996a; Woloszynska-Read et al. 2008). Here we searched to determine if demethylation of DDIC gene promoters similarly correlates with genome-wide DNA hypomethylation in LUAD tumors. To this end, we established a procedure based on methylomic data from the TCGA to evaluate global genome methylation levels in LUAD tissue samples. We first searched to identify CpGs that are initially methylated in normal lung tissues. CpGs located on the X chromosome were ignored in order to avoid gender-related differences resulting from the X inactivation process in female cells. Among the >450k CpG sites that are interrogated in TCGA datasets, we identified 137,954 CpGs that displayed high methylation levels (mean methylation ß value = 0.84) in all normal lung tissue samples (n = 32). We then calculated the mean level of methylation of all 137,954 CpGs in each LUAD sample, and used it as an estimator of overall genome methylation level. The results revealed that a majority (82%) of LUAD tissue samples exhibited discernable genome hypomethylation as compared with normal lung tissues (ß value < quantile 25 of normal tissues: 0.827; **Figure 4A**).

We next searched to determine if local DNA demethylation of DDIC gene promoters correlates with global genome hypomethylation. For each DDIC gene, LUAD samples were separated in two subgroups depending on the methylation status of the promoter region (Methylated or Unmethylated), and global genome methylation levels were compared between the two subgroups (b). The results revealed a significant association between global genome hypomethylation and local promoter demethylation not only for CG genes, as expected, but also for all DDIC genes of the GI and SE categories (**Figure 4C; Supplementary Table S2**). Demethylation of GI and SE gene promoters, however, was associated with a decrease in global DNA methylation that was generally less pronounced than for CG genes (**Figure 4C**). To further address this issue, we divided LUAD samples into four groups depending on the range of global DNA hypomethylation, and determined in every group the mean level of methylation of each category of DDIC gene promoters (**Figure 4D**). Consistent with previous data, the results showed that CG-DDIC gene promoters only become substantially demethylated (mean promoter ß-value < 0.6) in the LUAD samples that show a marked decrease in global DNA methylation (mean global ß value < 0.70). In comparison, GI- and SE-DDIC gene promoters already exhibited a similar level of local demethylation in tumors that only show an intermediate level of genome hypomethylation (mean global ß value < 0.77; **Figure 4D**). Moreover, we noted that in normal lung tissues, the mean

methylation level of GI- and SE-DDIC gene promoters (0.74 and 0.72 mean ß values, respectively) was generally lower than that of CG-DDIC gene promoters (0.84; **Figure 4D**).

Together, these results indicate that local DNA demethylation of the promoters of all three categories of DDIC genes coincides with a process of global genome demethylation. However, compared with CG gene promoters, promoters of the GI and SE gene categories appear to be less extensively methylated in the normal lung tissue, and to already exhibit local DNA demethylation in tumors displaying a lower extent of genome hypomethylation.

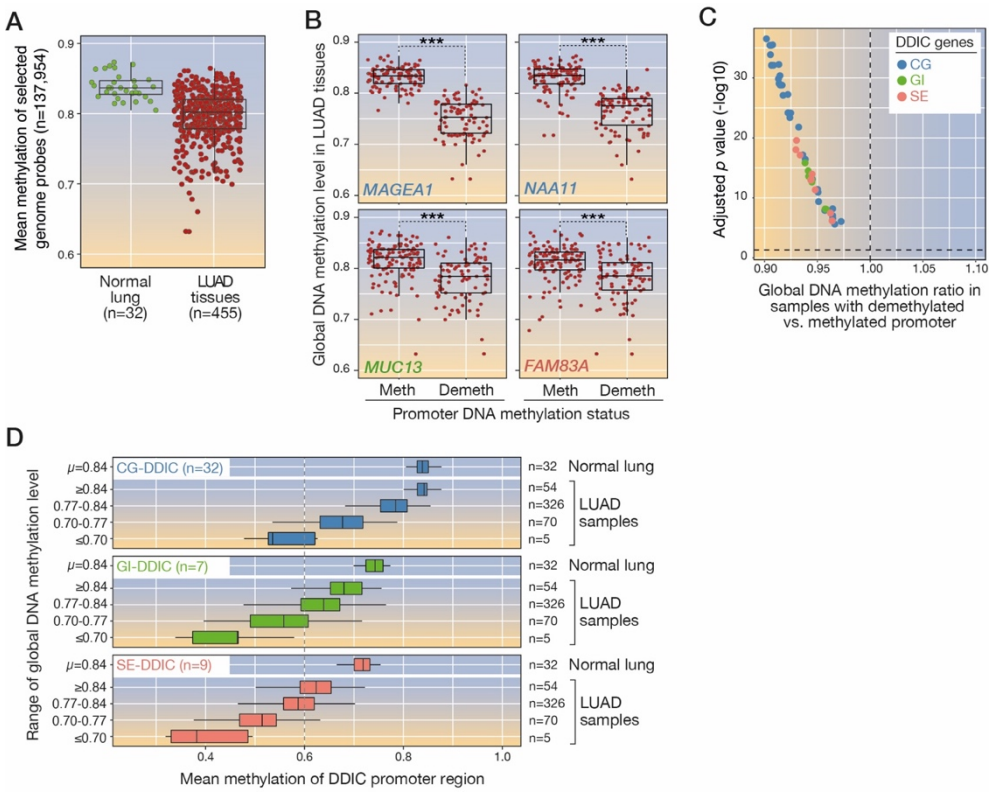


Figure 4. Focal DDIC promoter demethylation is associated with genome-wide loss of DNA methylation in LUAD tissue samples. *(A) Global DNA methylation estimates of LUAD and normal matching lung tissue samples of TCGA, based on the analysis of a selected pool of CpGs dispersed over the entire genome. (B) Tumor samples exhibiting either methylation (Meth) or demethylation (Demeth) of the indicated gene promoter were separated, and their level of global DNA methylation was compared (Student's t-test, ***: p < 0.001). The results for four genes are illustrated. (C) Statistical values determined as shown in panel C, were obtained for all CG-, GI-, and SE-DDIC genes and plotted accordingly. (D) LUAD tissue samples were divided into subgroups according to their range of global*

genome methylation level. In each subgroup of samples, the mean level of methylation of gene promoters belonging to the indicated DDIC category was calculated, and is represented by a box plot. The same analysis was conducted in normal lung tissue samples.

3.7. Assessing the Role of Transcription Factors in the Activation of GI-DDIC Genes

As mentioned before, the correlation between promoter demethylation and transcriptional activation that we observed for DDIC genes was less strict in the case of genes belonging to the GI cluster. In particular, we observed that in several LUAD samples, the expression of GI-DDIC genes remained absent despite demethylation of their promoter region (**Figure 3C**). This suggested that transcriptional induction of these genes might be conditioned by the presence of specific transcription factors. We therefore searched to identify transcription factors involved in the regulation of GI-DDIC genes. To this end we used i-*cis*Target, a prediction tool integrating sequence conservation, transcription factor binding motifs, and ChIP-seq data (Imrichova et al. 2015). Analysis with i-*cis*Target revealed highly significant enrichment of consensus binding sites for Hepatocyte Nuclear Factor 4 Alpha (HNF4A), which were detected in conserved regions of 6 out of the 7 GI-DDIC genes (**Figure 5A**). Moreover, ChIP-seq data confirmed binding of HNF4A in the promoter region of 3 of these genes (*EPS8L3*, *MUC13*, *VIL1*; **Figure 5A**). HNF4A is a highly tissue-specific transcriptional activator, which plays a role in development of the liver and the gastrointestinal tract.

Analysis of GTEx RNA-seq datasets confirmed that *HNF4A* is highly expressed in colon and small intestine, and instead absent in other normal tissues, including lung (**Figure 5B**). Importantly, analysis of RNA-seq data from TCGA revealed ectopic activation of *HNF4A* in 23% of LUAD tissue samples (≥2 TPM, **Figure 5B**). To evaluate the role of HNF4A in the induction of *EPS8L3*, *MUC13*, and *VIL1* genes in LUAD, we first analyzed RNA-seq data to find out if their expression is correlated with that of *HNF4A*. The results showed positive correlation between *HNF4A* and all three GI-DDIC genes in LUAD samples, with highest correlation scores for *EPS8L3* and *MUC13* (**Figure 5C**). To further assess the potential role of HNF4A in regulating these genes, we tested the effect of its depletion on the ability to induce expression of *EPS8L3*, *MUC13* and *VIL1* genes with 5-azadC. Thus, LXF289 LUAD cells, which express HNF4A, were transfected with either control siRNAs (siLuc) or siRNAs directed against HNF4A (siHNF4A, **Figure 5D, Figure S5**), and were thereafter exposed to 5-azadC. The results revealed that induction of *MUC13* and *EPS8L3* (but not of *VIL1*) by 5-azadC was significantly impaired when HNF4A was downregulated (**Figure 5E**). Of note, HNF4A downregulation had instead no impact on 5-azadC induction of a CG gene (*CT-GABRA3*, **Figure 5E**). Together, these results illustrate the case where

induction of GI-DDIC genes requires the presence of specific transcription factors, and identify HNF4A as one such factor.

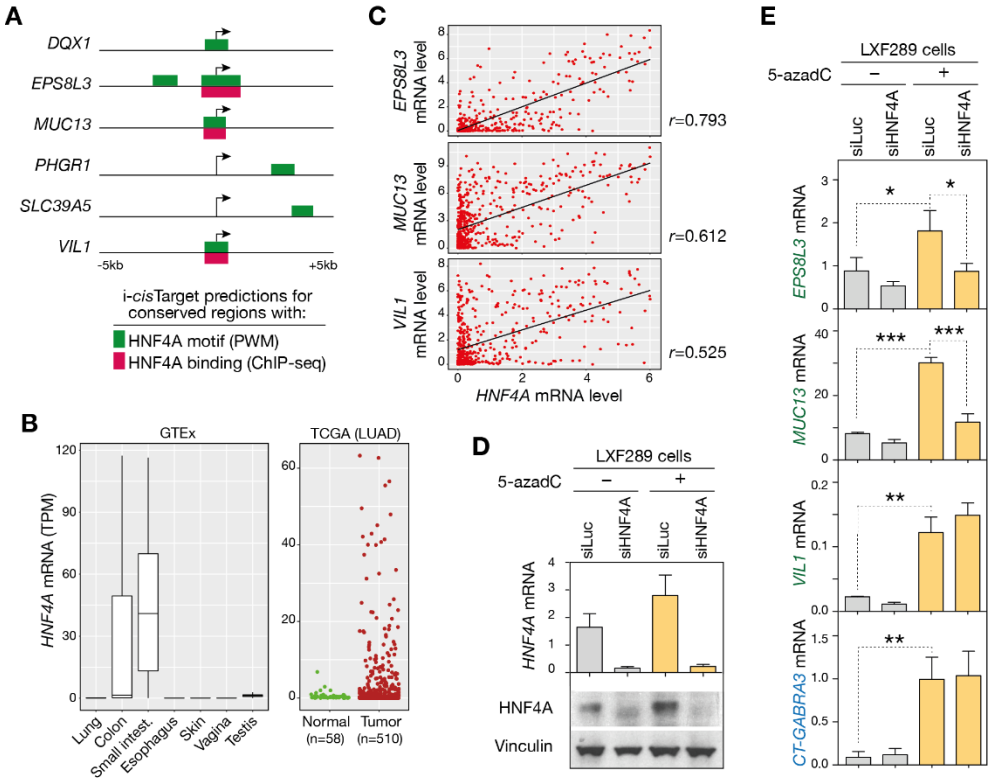


Figure 5. Transcription factor HNF4A contributes to transcriptional induction of DDIC genes of the GI cluster. *(A) Schematic representation of i-cisTarget predictions for GI-DDIC genes: conserved regions with HNF4A consensus binding motifs (based on position weight matrices, PWM) and ChIP-seq binding peaks (obtained from LoVo colon tumor cells) are depicted, with the broken arrow corresponding to the TSS. (B) RNA-seq data were analyzed to determine HNF4A expression in a panel of normal tissues (GTEx), as well as in LUAD and normal matching lung tissue samples (TCGA). (C) Expression (RNA-seq, log2(TPM+1)) of HNF4a was compared to that of EPS8L3, MUC13, and VIL1 genes in LUAD samples (Pearson's correlation coefficients (r) are indicated). (D) LXF289 LUAD cells were transfected with siRNAs against HNF4a (siHNF4A) or control siRNAs (siLuc), and then exposed to 5-azadC (+) or vehicle (–). HNF4A expression in the different cell groups was evaluated at the mRNA level by RT-qPCR (n = 3, mean ± SEM), and at the protein level by Western blot (Vinculin was used as loading control). (E) Cells exposed to these experimental conditions were also submitted to RT-qPCR analyses to evaluate levels of expression EPS8L3, MUC13, and VIL1 genes (relative to ACTB; CT-GABRA3 was used as a HNF4A-independent control gene). Values represent mean of 3 independent experiments ±SEM (One-way ANOVA, Tukey's multiple comparison; ***: p <0.001; **: p <0.01; *: p < 0.05).*

3.8. Associating Genome Hypomethylation and DDIC Gene Activation with Tumor Grade and Patient Survival

It has been proposed that epigenetic alterations, including genome-wide DNA hypomethylation, might be associated with the increased plasticity of tumor cells, which often deviate from their original differentiation status. Tumor grading systems provide an estimate of the degree of cellular deviation, with the highest grade (4) corresponding to histological images showing cells with the most divergent appearance. Tumor grades for LUAD samples are available in TCGA datasets, and we first examined if global genome hypomethylation is indeed associated with loss of cellular differentiation, i.e., with a higher grade. The results revealed that global DNA hypomethylation in LUAD was significantly correlated with higher tumor grades (**Figure 6A**).

We then examined each DDIC gene of the CG, GI and SE clusters, to find out if their expression in LUAD also correlates with tumor grade. Interestingly, expression of DDIC genes of the CG and SE categories, but not of the GI category, was significantly associated with higher tumor grades (**Figure 6A, B; Supplementary Figure S2**). Highest correlation scores were observed for *ADGRF4* and *FAM83A* genes of the SE category, and *LIN28B* of the CG category.

We also exploited clinical data from the TCGA to examine if genome hypomethylation and expression of DDIC genes are associated with reduced survival of LUAD patients. The results showed lack of correlation between genome hypomethylation and patient survival (**Figure 6C**). Instead, expression of several DDIC genes of the SE category, and most significantly of *FAM83A*, was significantly associated with reduced survival of LUAD patients (**Figure 6D; Supplementary Table S2**).

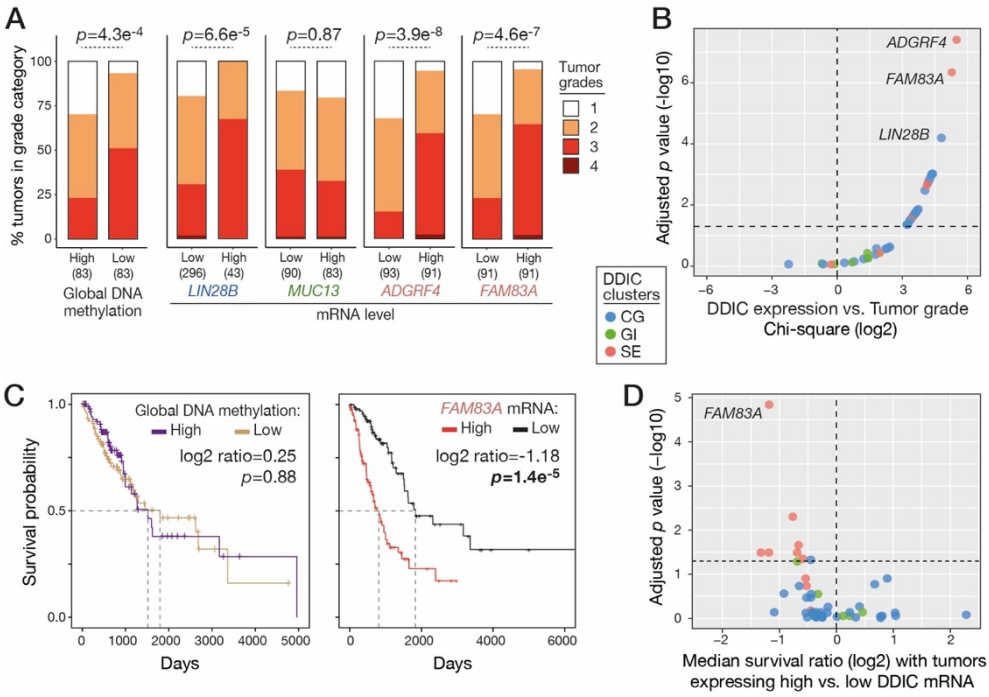


Figure 6. Associations of tumor grade and patient survival with genome hypomethylation and DDIC gene expression in LUAD. *(A) Comparison of tumor grades in LUAD samples grouped according to global DNA methylation levels or mRNA levels of DDIC genes (low: q20; high: q80; number of samples are indicated below columns). Comparisons were performed using a Chi-squared test (adjusted p-value). (B) Tumor grade comparisons as described in A, were conducted for all CG-, GI- and SE-DDIC genes. Dots on the graph represent Chi-squared statistic vs. p-adjusted values. (C) Survival curves (Kaplan–Meier) of LUAD patients (TCGA) according to high or low global DNA methylation levels, and to high or low FAM83A mRNA levels (log-rank test). Dashed lines mark median survival time of the two subgroups, and the ratio between the two median times is indicated (log2). (D) Patient survival analyses, as described in C, were conducted for all CG-, GI- and SE-DDIC genes. Dots on the graph represent ratios of median survival in high/low subgroups (log2) and p-adjusted values.*

4. Discussion

Most tumor cells show genome-wide losses of DNA methylation, and previous studies, including from our group, have shown that this epigenetic alteration is associated with the aberrant activation of a group of germline-specific (CG) genes (De Smet and Loriot 2013). Whether gene groups sharing other tissue specificities also become activated in association with genome demethylation in tumors

remained an open question. Here, we investigated this issue by conducting an unbiased exploration of publicly available transcriptomic and methylomic datasets. Lung adenocarcinoma was chosen for this analysis, as datasets from both cell lines and tissues were available. Moreover, the normal precursor cells of this type of tumor has been identified (AT2 cells) (Blanpain 2013), and transcriptomic and methylomic datasets from these cells were also available.

Our analyses confirm that CG genes constitute the dominant group of genes that become activated in tumor cells as a result of genome hypomethylation, since 43% of the genes that we identified as being induced in association with DNA demethylation (DDIC) showed specific expression in testicular germ cells. Importantly, however, we also identified two other tissue-specific gene clusters showing DNA demethylation-associated induction in lung tumor cell lines: one with specific expression in the gastrointestinal tract (GI), and the other displaying restricted expression in tissues comprising a stratified squamous epithelium (SE). Genes belonging to these two groups were however less abundant than CG genes, as they represented only 7% (GI) and 10% (SE) of all isolated DDIC genes.

Analysis of in vivo tumor samples confirmed aberrant induction of all three group of genes in LUAD tissues. CG genes, however, were activated in a lower proportion of tumor samples, as compared with GI- or SE-DDIC genes. Analysis of global DNA methylation levels indicates that CG genes only become activated in tumors that show extensive genome hypomethylation. In comparison, GI- and SE-DDIC genes already showed local promoter demethylation and transcriptional induction in tumors with less pronounced global genome hypomethylation. This is likely due to the fact that, compared with CG genes, GI- and SE-DDIC genes contain promoters displaying a lower amount of CpGs, and an initially lower level of methylation in normal tissues.

Transcriptional activation of DNA methylation-regulated genes relies not only on DNA demethylation, but also on the availability of transcription factors capable of inducing efficient promoter activity. It has been demonstrated for CG genes that such factors are present in most cells (De Smet et al. 1995). DNA demethylation is therefore a sufficient trigger for induction of these genes in most tumors. Our data suggest that this is also the case for SE-DDIC genes. A representative gene in this cluster is *SERPINB5*, the expression of which was previously shown to be regulated primarily by DNA methylation (Futscher et al. 2002). For GI-DDIC genes instead, we observed a more conditional link between DNA demethylation and transcriptional induction, which could be explained by a more restricted pattern of expression of activating transcription factors. Our data indicate that HNF4A constitutes one such factor, and show that aberrant induction of several GI-DDIC genes in LUAD tumors

is dependent on ectopic upregulation of this liver- and gut-specific factor in addition to DNA demethylation. The precise mechanism by which HNF4A mediates induction of GI-DDIC genes in association with promoter DNA demethylation in LUAD tumors cells remains to be determined. It has been demonstrated previously that HNF4A is upregulated in a fraction of LUAD tumors, and contributes to a gene expression signature that characterizes a subtype of LUAD, which was defined as invasive mucinous adenocarcinoma (Sugano et al. 2013). It is likely that the genes we characterized here represent a restricted subgroup of this signature, for which transcriptional induction in LUAD relies not only on the presence of HNF4A, but also on promoter DNA demethylation.

Our results indicate that genome hypomethylation in LUAD is associated with loss of cellular differentiation, which is in line with previous observations in different tumor types (Soares et al. 1999; Zhang et al. 2020). Consistently, transcriptional induction of several DDIC genes was observed predominantly in poorly differentiated tumors. We found instead no association between the level of genome hypomethylation and the probability of survival of LUAD patients. Reduced patient survival was instead significantly associated with high expression of several genes belonging to the SE-DDIC cluster. Among these, *FAM83A* was most significantly associated with poor survival of LUAD patients. Ectopic expression of *FAM83A* has been previously described in different tumor types, including lung adenocarcinoma (Cai et al. 2019). It has been demonstrated in a model of breast cancer that FAM83A acts downstream of the EGFR signaling and exerts oncogenic properties by promoting tumor growth, and conferring resistance to EGFR-tyrosine kinase inhibitors (Lee et al. 2012). Mounting evidence suggests that *FAM83A* plays a predominant role in the development of a variety of tumors (Snijders et al. 2017). Our study now identifies promoter DNA demethylation as a critical contributor of aberrant upregulation of this important oncogene in cancer cells. Future experiments will be required to determine if other DDIC genes can act as cancer drivers in LUAD.

5. Conclusions

To conclude, our study reveals that genome hypomethylation in tumors is associated not only with aberrant activation of germline-specific genes, but also with ectopic induction of gene expression programs that are specific of defined somatic tissues. Several genes in these newly identified gene clusters were

associated with poor survival of lung cancer patients. Moreover, our results reveal that these genes contain CpGs displaying highly contrasted methylation states in normal and tumoral lung tissues, and are therefore amenable to the development of DNA methylation biomarkers of prognostic significance. Several DNA methylation biomarkers have been successfully translated into clinical practice, but so far, these markers correspond to DNA sequences that show hypermethylation in tumor cells (Locke et al. 2019). Our results are an incentive to explore the value of hypomethylated DNA sequences as cancer biomarkers.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/cancers14041007/s1, Figure S1: Identified DDIC transcripts derive from referenced and unreferenced genes and transcript variants, Figure S2: CG-DDIC genes are predominantly expressed in testicular germ cells, Figure S3: GI- and SE-DDIC genes are coexpressed with NKX2-1, an AT2-specific marker, Figure S4: GI- and SE-DDIC proteins are ectopically expressed in LUAD tissue samples, Figure S5: Raw images of western-blot films corresponding to Figure 5D, Table S1: Methylomic and transcriptomic datasets of LUAD cell lines used in this study, Table S2: CG-, GI- and SE-DDIC gene list, Table S3: Methylomic and transcriptomic datasets of normal cells and tissues, Table S4: Transcriptomic datasets of cell lines treated with 5-azadC, Table S5: Antibodies and tissue section codes from HPA, Table S6: siRNAs used for transfection experiments, Table S7: Primer sequences, PCR and qPCR reagents and conditions.

Author Contributions: Conceptualization, A.D., A.L. and C.D.S.; funding acquisition, A.D. and C.D.S.; investigation, A.D., A.L. methodology, A.D. and A.L.; supervision, C.D.S.; writing—original draft, A.D. and C.D.S.; writing—review and editing, A.D., A.L. and C.D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by grants from the Fonds de la recherche scientifique (FRS-FNRS), Belgium (ref. J.0032.19), the D.G. Higher Education and Scientific Research of the French Community of Belgium (Action de Recherches Concertées), and the Fonds special de recherche (FSR) of the UCLouvain, Belgium. A.D. was supported by a fellowship of the FRS-FNRS-FRIA, Belgium (ref.1.E008.19). A.L. was supported by the de Duve Institute, Brussels, Belgium.

Data Availability Statement: Accessions to archived datasets analyzed in this study are listed in Tables S1, S3 and S4.

Acknowledgments: The authors would like to acknowledge Kun-Hsing Yu and Michael Snyder (Stanford University, CA) for providing lists of tumor grades of LUAD tissue samples from the TCGA.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

References

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **2002**, *16*, 6–21.
2. Jenuwein, T.; Allis, C.D. Translating the histone code. *Science* **2001**, *293*, 1074–1080.
3. Riggs, A.D. DNA methylation and cell memory. *Cell Biophys.* **1989**, *15*, 1–13.
4. Baylin, S.B.; Jones, P.A. A decade of exploring the cancer epigenome—biological and translational implications. *Nat. Rev. Cancer* **2011**, *11*, 726–734. https://doi.org/10.1038/nrc3130.
5. Feinberg, A.P. Epigenetic stochasticity, nuclear structure and cancer: The implications for medicine. *J. Intern. Med.* **2014**, *276*, 5–11. https://doi.org/10.1111/joim.12224.
6. Feinberg, A.P.; Vogelstein, B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* **1983**, *301*, 89–92.
7. Gama-Sosa, M.A.; Slagel, V.A.; Trewyn, R.W.; Oxenhandler, R.; Kuo, K.C.; Gehrke, C.W.; Ehrlich, M. The 5-methylcytosine content of DNA from human tumors. *Nucl. Acids Res.* **1983**, *11*, 6883–6894.
8. De Smet, C.; Loriot, A. DNA hypomethylation and activation of germline-specific genes in cancer. *Adv. Exp. Med. Biol* **2013**, *754*, 149–166.
9. Cannuyer, J.; Loriot, A.; Parvizi, G.K.; De Smet, C. Epigenetic hierarchy within the MAGEA1 cancer-germline gene: Promoter DNA methylation dictates local histone modification. *PLoS ONE* **2013**, *8*, e58743.
10. De Smet, C.; Lurquin, C.; Lethé, B.; Martelange, V.; Boon, T. DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter. *Mol. Cell. Biol.* **1999**, *19*, 7327–7335.
11. Grunwald, C.; Koslowski, M.; Arsiray, T.; Dhaene, K.; Praet, M.; Victor, A.; Morresi-Hauf, A.; Lindner, M.; Passlick, B.; Lehr, H.A.; et al. Expression of multiple epigenetically regulated cancer/germline genes in nonsmall cell lung cancer. *Int. J. Cancer* **2006**, *118*, 2522–2528.
12. Sigalotti, L.; Coral, S.; Nardi, G.; Spessotto, A.; Cortini, E.; Cattarossi, I.; Colizzi, F.; Altomonte, M.; Maio, M. Promoter methylation controls the expression of MAGE2, 3 and 4 genes in human cutaneous melanoma. *J. Immunother.* **2002**, *25*, 16–26.
13. Akers, S.N.; Odunsi, K.; Karpf, A.R. Regulation of cancer germline antigen gene expression: Implications for cancer immunotherapy. *Future Oncol.* **2010**, *6*, 717–732.
14. Van Tongelen, A.; Loriot, A.; De Smet, C. Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes. *Cancer Lett.* **2017**, *396*, 130–137. https://doi.org/10.1016/j.canlet.2017.03.029.
15. Whitehurst, A.W. Cause and consequence of cancer/testis antigen activation in cancer. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 251–272. https://doi.org/10.1146/annurev-pharmtox-011112-140326.
16. Ehrlich, M. DNA hypomethylation in cancer cells. *Epigenomics* **2009**, *1*, 239–259.
17. Szyf, M.; Pakneshan, P.; Rabbani, S.A. DNA methylation and breast cancer. *Biochem. Pharm.* **2004**, *68*, 1187–1197.
18. Dokun, O.Y.; Florl, A.R.; Seifert, H.H.; Wolff, I.; Schulz, W.A. Relationship of SNCG, S100A4, S100A9 and LCN2 gene expression and DNA methylation in bladder cancer. *Int. J. Cancer* **2008**, *123*, 2798–2807.
19. Bestor, T.H.; Edwards, J.R.; Boulard, M. Notes on the role of dynamic DNA methylation in mammalian development. *Proc. Natl. Acad.*

Sci. USA **2015**, *112*, 6796–6799. https://doi.org/10.1073/pnas.1415301111.
20. Blanpain, C. Tracing the cellular origin of cancer. *Nat. Cell Biol.* **2013**, *15*, 126–134. https://doi.org/10.1038/ncb2657.
21. Suzuki, A.; Makinoshima, H.; Wakaguri, H.; Esumi, H.; Sugano, S.; Kohno, T.; Tsuchihara, K.; Suzuki, Y. Aberrant transcriptional regulations in cancers: Genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines. *Nucl. Acids Res.* **2014**, *42*, 13557–13572. https://doi.org/10.1093/nar/gku885.
22. Fain, J.S.; Van Tongelen, A.; Loriot, A.; De Smet, C. Epigenetic Coactivation of MAGEA6 and CT-GABRA3 Defines Orientation of a Segmental Duplication in the Human X Chromosome. *Cytogenet. Genome Res.* **2019**, *159*, 12–18. https://doi.org/10.1159/000502933.
23. Krueger, F.; Andrews, S.R. Bismark: A flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **2011**, *27*, 1571–1572. https://doi.org/10.1093/bioinformatics/btr167.
24. Marconett, C.N.; Zhou, B.; Sunohara, M.; Pouldar, T.M.; Wang, H.; Liu, Y.; Rieger, M.E.; Tran, E.; Flodby, P.; Siegmund, K.D.; et al. Cross-Species Transcriptome Profiling Identifies New Alveolar Epithelial Type I Cell-Specific Genes. *Am. J. Respir. Cell Mol. Biol.* **2017**, *56*, 310–321. https://doi.org/10.1165/rcmb.2016-0071OC.
25. Roadmap Epigenomics, C.; Kundaje, A.; Meuleman, W.; Ernst, J.; Bilenky, M.; Yen, A.; Heravi-Moussavi, A.; Kheradpour, P.; Zhang, Z.; Wang, J.; et al. Integrative analysis of 111 reference human epigenomes. *Nature* **2015**, *518*, 317–330. https://doi.org/10.1038/nature14248.
26. Zuber, V.; Marconett, C.N.; Shi, J.; Hua, X.; Wheeler, W.; Yang, C.; Song, L.; Dale, A.M.; Laplana, M.; Risch, A.; et al. Pleiotropic Analysis of Lung Cancer and Blood Triglycerides. *J. Natl. Cancer Ins.* **2016**, *108*, djw167. https://doi.org/10.1093/jnci/djw167.
27. Robinson, J.T.; Thorvaldsdottir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. https://doi.org/10.1038/nbt.1754.
28. Consortium, G.T. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. https://doi.org/10.1038/ng.2653.
29. Fain, J.S.; Loriot, A.; Diacofotaki, A.; Van Tongelen, A.; De Smet, C. Transcriptional overlap links DNA hypomethylation with DNA hypermethylation at adjacent promoters in cancer. *Sci. Rep.* **2021**, *11*, 17346. https://doi.org/10.1038/s41598-021-96844-0.
30. Sole-Boldo, L.; Raddatz, G.; Schutz, S.; Mallm, J.P.; Rippe, K.; Lonsdorf, A.S.; Rodriguez-Paredes, M.; Lyko, F. Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming. *Commun. Biol.* **2020**, *3*, 188. https://doi.org/10.1038/s42003-020-0922-4.
31. Uhlen, M.; Bjorling, E.; Agaton, C.; Szigyarto, C.A.; Amini, B.; Andersen, E.; Andersson, A.C.; Angelidou, P.; Asplund, A.; Asplund, C.; et al. A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol. Cell Proteom.* **2005**, *4*, 1920–1932. https://doi.org/10.1074/mcp.M500279-MCP200.
32. D'Alessio, A.C.; Weaver, I.C.; Szyf, M. Acetylation-induced transcription is required for active DNA demethylation in methylation-silenced genes. *Mol. Cell Biol.* **2007**, *27*, 7462–7474.
33. Grandin, M.; Mathot, P.; Devailly, G.; Bidet, Y.; Ghantous, A.; Favrot, C.; Gibert, B.; Gadot, N.; Puisieux, I.; Herceg, Z.; et al. Inhibition of DNA methylation promotes breast tumor sensitivity to netrin-1 interference. *EMBO Mol. Med.* **2016**, *8*, 863–877. https://doi.org/10.15252/emmm.201505945.
34. Park, J.W.; Sahm, F.; Steffl, B.; Arrillaga-Romany, I.; Cahill, D.; Monje, M.; Herold-Mende, C.; Wick, W.; Turcan, S. TERT and DNMT1 expression predict sensitivity to decitabine in gliomas. *Neuro. Oncol.* **2021**, *23*, 76–87. https://doi.org/10.1093/neuonc/noaa207.
35. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.
36. Kim, D.; Paggi, J.M.; Park, C.; Bennett, C.; Salzberg, S.L. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat. Biotechnol.* **2019**, *37*,

907–915. https://doi.org/10.1038/s41587-019-0201-4.

37. Liao, Y.; Smyth, G.K.; Shi, W. featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **2014**, *30*, 923–930. https://doi.org/10.1093/bioinformatics/btt656.
38. Love, M.I.; Huber, W.; Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **2014**, *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.
39. Consortium, E.P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **2012**, *489*, 57–74. https://doi.org/10.1038/nature11247.
40. Lewis, S.K.; Nachun, D.; Martin, M.G.; Horvath, S.; Coppola, G.; Jones, D.L. DNA Methylation Analysis Validates Organoids as a Viable Model for Studying Human Intestinal Aging. *Cell Mol. Gastroenterol. Hepatol.* **2020**, *9*, 527–541. https://doi.org/10.1016/j.jcmgh.2019.11.013.
41. Tilman, G.; Arnoult, N.; Lenglez, S.; Van Beneden, A.; Loriot, A.; De Smet, C.; Decottignies, A. Cancer-linked satellite 2 DNA hypomethylation does not regulate Sat2 non-coding RNA expression and is initiated by heat shock pathway activation. *Epigenetics* **2012**, *7*, 903–913. https://doi.org/10.4161/epi.21107.
42. Colaprico, A.; Silva, T.C.; Olsen, C.; Garofano, L.; Cava, C.; Garolini, D.; Sabedot, T.S.; Malta, T.M.; Pagnotta, S.M.; Castiglioni, I.; et al. TCGAbiolinks: An R/Bioconductor package for integrative analysis of TCGA data. *Nucl. Acids Res.* **2016**, *44*, e71. https://doi.org/10.1093/nar/gkv1507.
43. Yu, K.H.; Zhang, C.; Berry, G.J.; Altman, R.B.; Re, C.; Rubin, D.L.; Snyder, M. Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features. *Nat. Commun.* **2016**, *7*, 12474. https://doi.org/10.1038/ncomms12474.
44. Billingham, R.E.; Silvers, W.K. The origin and conservation of epidermal specificities. *N. Engl. J. Med.* **1963**, *268*, 539–545. https://doi.org/10.1056/NEJM196303072681008.
45. D'Anna, F.; Van Dyck, L.; Xiong, J.; Zhao, H.; Berrens, R.V.; Qian, J.; Bieniasz-Krzywiec, P.; Chandra, V.; Schoonjans, L.; Matthews, J.; et al. DNA methylation repels binding of hypoxia-inducible transcription factors to maintain tumor immunotolerance. *Genome Biol.* **2020**, *21*, 182. https://doi.org/10.1186/s13059-020-02087-z.
46. Devailly, G.; Grandin, M.; Perriaud, L.; Mathot, P.; Delcros, J.G.; Bidet, Y.; Morel, A.P.; Bignon, J.Y.; Puisieux, A.; Mehlen, P.; et al. Dynamics of MBD2 deposition across methylated DNA regions during malignant transformation of human mammary epithelial cells. *Nucl. Acids Res.* **2015**, *43*, 5838–5854. https://doi.org/10.1093/nar/gkv508.
47. Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550. https://doi.org/10.1038/nature13385.
48. Cadieux, B.; Ching, T.T.; VandenBerg, S.R.; Costello, J.F. Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation. *Cancer Res.* **2006**, *66*, 8469–8476.
49. De Smet, C.; De Backer, O.; Faraoni, I.; Lurquin, C.; Brasseur, F.; Boon, T. The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation. *Proc. Natl. Acad. Sci. USA* **1996**, *93*, 7149–7153.
50. Woloszynska-Read, A.; Mhawech-Fauceglia, P.; Yu, J.; Odunsi, K.; Karpf, A.R. Intertumor and intratumor NY-ESO-1 expression heterogeneity is associated with promoter-specific and global DNA methylation status in ovarian cancer. *Clin. Cancer Res.* **2008**, *14*, 3283–3290.
51. Imrichova, H.; Hulselmans, G.; Atak, Z.K.; Potier, D.; Aerts, S. i-cisTarget 2015 update: Generalized cis-regulatory enrichment analysis in human, mouse and fly. *Nucl. Acids Res.* **2015**, *43*, W57–64. https://doi.org/10.1093/nar/gkv395.
52. De Smet, C.; Courtois, S.J.; Faraoni, I.; Lurquin, C.; Szikora, J.P.; De Backer, O.; Boon, T. Involvement of two Ets binding sites in the transcriptional activation of the MAGE1 gene. *Immunogenetics* **1995**, *42*, 282–290.
53. Futscher, B.W.; Oshiro, M.M.; Wozniak, R.J.; Holtan, N.; Hanigan, C.L.; Duan, H.; Domann, F.E. Role for DNA methylation in the control of cell type specific maspin expression. *Nat. Genet.* **2002**, *31*, 175–179. https://doi.org/10.1038/ng886.

54. Sugano, M.; Nagasaka, T.; Sasaki, E.; Murakami, Y.; Hosoda, W.; Hida, T.; Mitsudomi, T.; Yatabe, Y. HNF4alpha as a marker for invasive mucinous adenocarcinoma of the lung. *Am. J. Surg. Pathol.* **2013**, *37*, 211–218. https://doi.org/10.1097/PAS.0b013e31826be303.
55. Soares, J.; Pinto, A.E.; Cunha, C.V.; Andre, S.; Barao, I.; Sousa, J.M.; Cravo, M. Global DNA hypomethylation in breast carcinoma: Correlation with prognostic factors and tumor progression. *Cancer* **1999**, *85*, 112–118.
56. Zhang, W.; Klinkebiel, D.; Barger, C.J.; Pandey, S.; Guda, C.; Miller, A.; Akers, S.N.; Odunsi, K.; Karpf, A.R. Global DNA Hypomethylation in Epithelial Ovarian Cancer: Passive Demethylation and Association with Genomic Instability. *Cancers* **2020**, *12*, 764. https://doi.org/10.3390/cancers12030764.
57. Cai, L.; Luo, D.; Yao, B.; Yang, D.M.; Lin, S.; Girard, L.; DeBerardinis, R.J.; Minna, J.D.; Xie, Y.; Xiao, G. Systematic Analysis of Gene Expression in Lung Adenocarcinoma and Squamous Cell Carcinoma with a Case Study of FAM83A and FAM83B. *Cancers* **2019**, *11*, 886. https://doi.org/10.3390/cancers11060886.
58. Lee, S.Y.; Meier, R.; Furuta, S.; Lenburg, M.E.; Kenny, P.A.; Xu, R.; Bissell, M.J. FAM83A confers EGFR-TKI resistance in breast cancer cells and in mice. *J. Clin. Invest.* **2012**, *122*, 3211–3220. https://doi.org/10.1172/JCI60498.
59. Snijders, A.M.; Lee, S.Y.; Hang, B.; Hao, W.; Bissell, M.J.; Mao, J.H. FAM83 family oncogenes are broadly involved in human cancers: An integrative multi-omics approach. *Mol. Oncol.* **2017**, *11*, 167–179. https://doi.org/10.1002/1878-0261.12016.
60. Locke, W.J.; Guanzon, D.; Ma, C.; Liew, Y.J.; Duesing, K.R.; Fung, K.Y.C.; Ross, J.P. DNA Methylation Cancer Biomarkers: Translation to the Clinic. *Front. Genet.* **2019**, *10*, 1150. https://doi.org/10.3389/fgene.2019.01150.

Supplementary materials, Diacofotaki et al. 2022

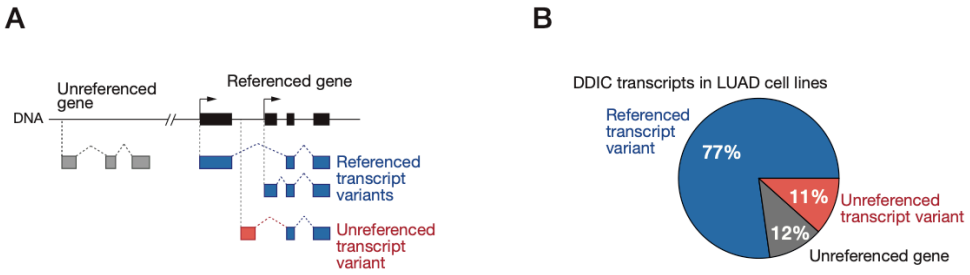


Figure S1: Identified DDIC transcripts derive from referenced and unreferenced genes and transcript variants. *A) DDIC transcripts which were identified in our screening of RNA-seq data were compared to the annotated reference genome. The meaning of referenced vs unreferenced genes and transcript variants is shown below. B) Proportions of DDIC in each of these categories is presented.*

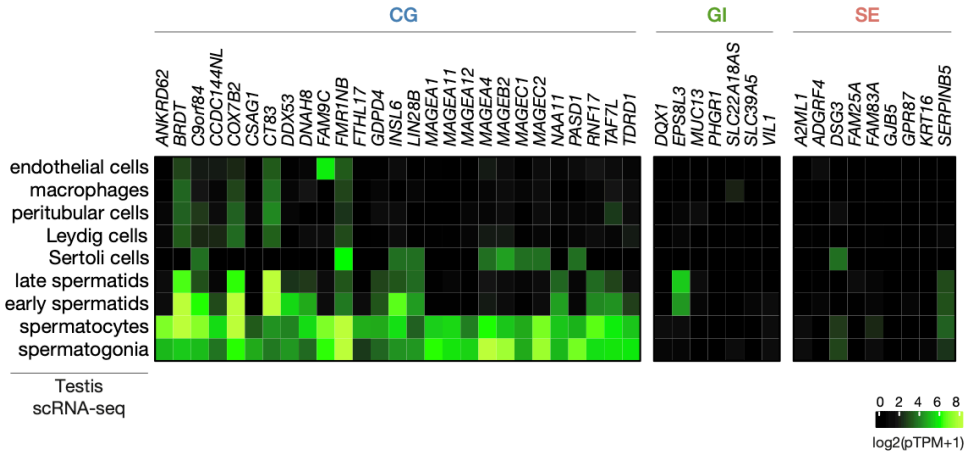


Figure S2: CG-DDIC genes are predominantly expressed in testicular germ cells. *Heatmap representing the expression level of CG-, GI- and SE-DDIC genes in cells constituting the tissue of the testis (scRNA-seq data from the HPA; pTPM: protein Transcripts Per Million).*

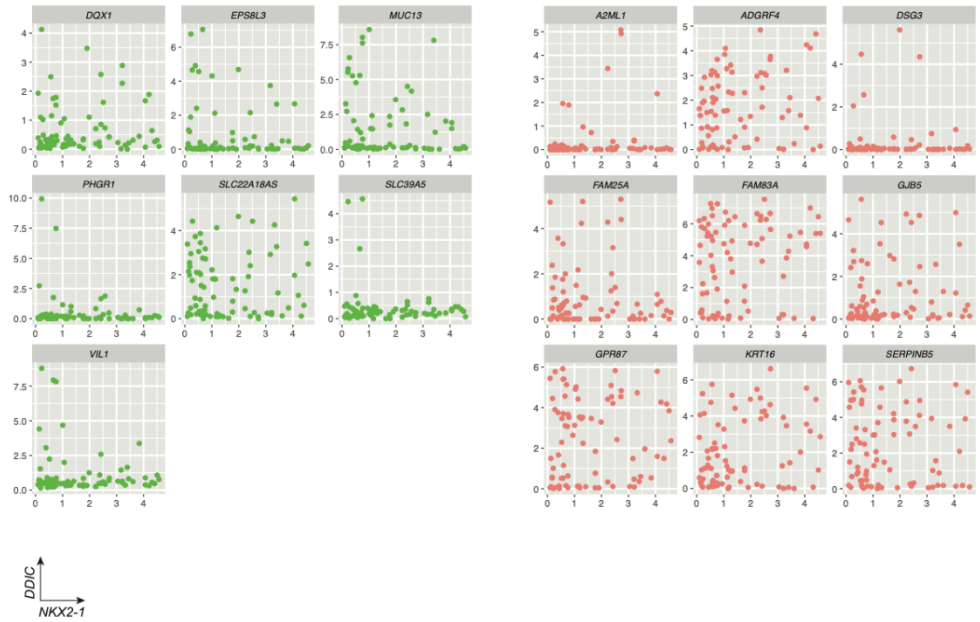


***Figure S3**: **GI- and SE-DDIC genes are co-expressed with NKX2-1, an AT2-specific marker**. Each dot represents a LUAD cell line for which DDIC gene and NKX2-1 expressions are reported on the y and x axis respectively. The DDIC gene name is reported on top of each graph. Expression data (in log2(TPM+1)) were obtained via DepMap portal (https://depmap.org/portal/) and correspond to 19Q3 release.*

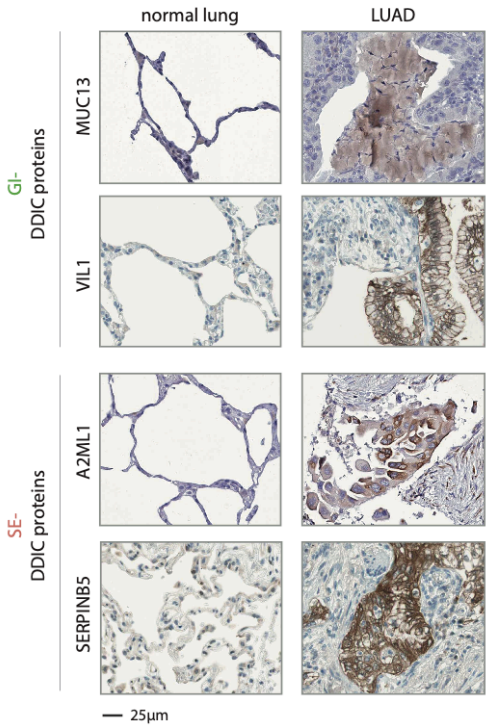
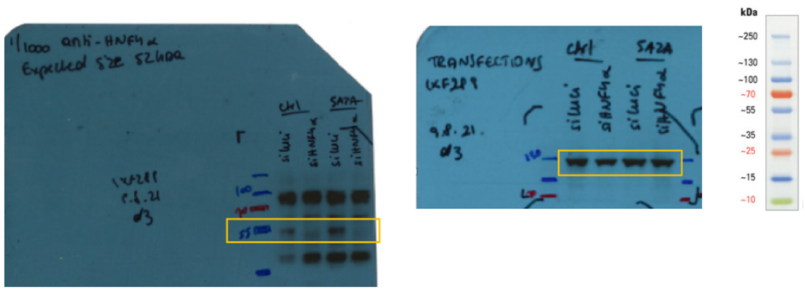


***Figure S4**: **GI- and SE-DDIC proteins are ectopically expressed in LUAD tissue samples**. Immunohistochemical images from the HPA of normal lung and LUAD tissue samples stained with antibodies targeting the above-mentioned GI- and SE-DDIC proteins. Antibodies and tissue codes are listed in supplementary table S5 and were selected using the same criteria as described in the material and methods section.*

A. B. C.



D.

Condition	Treatment	Lanes	VCL area	HNF4A area	HNF4A/VCL area ratio	% siLuc condition
siLuc	ctrl	1	42345,936	16245,108	0,384	100
siHNF4a	ctrl	2	35563,016	11172,915	0,314	81,9
siLuc	5-azadC	3	42110,714	35563,057	0,845	100
siHNF4a	5-azadC	4	38624,844	5159,388	0,134	15,8

Figure S5: Raw images of western-blot films corresponding to Fig. 5D. *Original uncropped images of HNF4A (A) and VCL (B) detections. Yellow rectangles delineate the image parts shown in the main figure. C) Protein size bands of the marker used (in kDa; PageRuler ThermoFischer). Protein sizes are indicated in red and blue on the film images. D) Signal quantifications of the lanes of interest (A) using ImageJ software (v1.43).*

Supplementary tables

Table S1: Methylomic and transcriptomic datasets of LUAD cell lines used in this study.

Sample	Type	Pathology	Accession	File	Assay	BioProject	Method	Layout	Platform	Reference
A427	cell line	lung adenocarcinoma	DRR016652	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
A549	cell line	lung adenocarcinoma	DRR016653	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
ABC1	cell line	lung adenocarcinoma	DRR016654	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1299	cell line	lung adenocarcinoma	DRR016655	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1437	cell line	lung adenocarcinoma	DRR016656	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1648	cell line	lung adenocarcinoma	DRR016657	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1650	cell line	lung adenocarcinoma	DRR016658	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1703	cell line	lung adenocarcinoma	DRR016659	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1819	cell line	lung adenocarcinoma	DRR016660	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1975	cell line	lung adenocarcinoma	DRR016661	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2126	cell line	lung adenocarcinoma	DRR016662	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2228	cell line	lung adenocarcinoma	DRR016663	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2347	cell line	lung adenocarcinoma	DRR016664	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H322	cell line	lung adenocarcinoma	DRR016665	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
II18	cell line	lung adenocarcinoma	DRR016666	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
LC2ad	cell line	lung adenocarcinoma	DRR016667	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC14	cell line	lung adenocarcinoma	DRR016668	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC3	cell line	lung adenocarcinoma	DRR016669	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC7	cell line	lung adenocarcinoma	DRR016670	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC9	cell line	lung adenocarcinoma	DRR016671	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad1	cell line	lung adenocarcinoma	DRR016672	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad2	cell line	lung adenocarcinoma	DRR016673	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_KJ	cell line	lung adenocarcinoma	DRR016674	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_MS	cell line	lung adenocarcinoma	DRR016675	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_OK	cell line	lung adenocarcinoma	DRR016676	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
VMRC_LCD	cell line	lung adenocarcinoma	DRR016677	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
A427	cell line	lung adenocarcinoma	DRR016694	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
A549	cell line	lung adenocarcinoma	DRR016695	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
ABC1	cell line	lung adenocarcinoma	DRR016696	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1299	cell line	lung adenocarcinoma	DRR016697	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1437	cell line	lung adenocarcinoma	DRR016698	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1648	cell line	lung adenocarcinoma	DRR016699	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1650	cell line	lung adenocarcinoma	DRR016700	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1703	cell line	lung adenocarcinoma	DRR016701	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1819	cell line	lung adenocarcinoma	DRR016702	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1975	cell line	lung adenocarcinoma	DRR016703	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2126	cell line	lung adenocarcinoma	DRR016704	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2228	cell line	lung adenocarcinoma	DRR016705	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2347	cell line	lung adenocarcinoma	DRR016706	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H322	cell line	lung adenocarcinoma	DRR016707	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
II18	cell line	lung adenocarcinoma	DRR016708	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
LC2ad	cell line	lung adenocarcinoma	DRR016709	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC14	cell line	lung adenocarcinoma	DRR016710	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC3	cell line	lung adenocarcinoma	DRR016711	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC7	cell line	lung adenocarcinoma	DRR016712	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC9	cell line	lung adenocarcinoma	DRR016713	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad1	cell line	lung adenocarcinoma	DRR016714	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad2	cell line	lung adenocarcinoma	DRR016715	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_KJ	cell line	lung adenocarcinoma	DRR016716	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_MS	cell line	lung adenocarcinoma	DRR016717	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_OK	cell line	lung adenocarcinoma	DRR016718	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
VMRC_LCD	cell line	lung adenocarcinoma	DRR016719	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014

Table S2: CG-, GI- and SE-DDIC gene list.

							5-azadC induction			LUAD tissue samples			Association with LUAD global DNA methylation		Association with LUAD tumor grade		Association with LUAD patient survival	
Gene name	Cluster	Strand	Chr	CpG number	ENSG identifier	TSS	MCF7	TS603	HMLER	Positive tumors (%)	Mean TPM in positive tumors	Correlation coefficient	Ratio	Adjusted p-value	Chi-squared statistic	adjusted p-value	Ratio	Adjusted p-value
AC010168.1	CG	+	12	18	ENSG00000214772	14665655	yes	yes	NA	14,51	4,87	NA	NA	NA	12,04	0,0192	0,82	0,8783
AC063977.6	CG	-	19	53	ENSG00000269072	51181966	yes	NA	NA	1,57	3,90	-0,35	0,92	7,00E-28	NA	NA	0,79	0,8783
AC074389.2	CG	+	7	38	ENSG00000231476	1619444	yes	NA	no	3,92	18,10	-0,62	0,92	1,47E-29	3,38	0,4350	0,78	0,9338
AC092506.1	CG	+	1	24	ENSG00000215869	104072983	yes	no	NA	3,92	3,07	-0,55	0,92	4,21E-24	1,72	0,7381	1,32	0,5351
AC097374.1	CG	-	2	24	ENSG00000259848	94947341	yes	no	no	0,39	2,87	NA	NA	NA	NA	NA	NA	NA
AC099520.1	CG	-	5	23	ENSG00000251574	105392970	no	no	NA	0,39	4,06	NA	NA	NA	NA	NA	NA	NA
AC109635.2	CG	+	11	59	ENSG00000254518	50298579	yes	NA	NA	1,18	4,65	-0,40	0,97	2,33E-06	NA	NA	NA	NA
AC109635.4	CG	-	11	68	ENSG00000255042	50298452	yes	no	no	3,33	34,70	-0,45	0,96	8,47E-09	2,32	0,6080	0,78	0,9455
AL160408.1	CG	+	1	9	ENSG00000228044	234656125	yes	NA	yes	5,88	14,57	NA	NA	NA	3,50	0,2749	1,01	0,9184
ANKRD62	CG	+	18	55	ENSG00000181626	12093843	yes	NA	yes	0,39	3,37	-0,47	0,95	3,89E-12	NA	NA	NA	NA
AP002383.3	CG	+	11	31	ENSG00000256779	95038372	yes	NA	yes	2,16	9,42	-0,42	0,95	7,83E-14	NA	NA	1,71	0,8973
BAGE2	CG	+	21	30	ENSG00000187172	10413532	yes	yes	no	5,49	5,51	NA	NA	NA	9,44	0,0452	2,04	0,7440
BRDT	CG	+	1	39	ENSG00000137948	91949371	yes	no	yes	29,02	20,01	-0,66	0,96	2,59E-07	12,74	0,0160	1,59	0,1679
C9ORF84	CG	-	9	32	ENSG00000165181	111795008	yes	yes	yes	20,59	6,43	NA	0,95	4,19E-10	13,28	0,0143	0,90	0,5449
CCDC144NL	CG	-	17	20	ENSG00000205212	20868225	yes	NA	yes	1,76	3,53	NA	NA	NA	NA	NA	0,75	0,7236
COX7B2	CG	-	4	28	ENSG00000170516	46909207	yes	yes	yes	13,33	25,62	-0,45	0,91	5,74E-33	18,33	0,0018	1,75	0,8783
CSAG1	CG	-	X	26	ENSG00000198930	152733691	yes	yes	yes	20,39	70,18	-0,66	0,91	5,74E-33	20,22	0,0011	0,83	0,7236
CT83	CG	-	X	23	ENSG00000204019	116462976	no	yes	NA	50,98	55,98	-0,83	0,96	8,93E-08	0,21	0,8996	0,90	0,7440
DDX53	CG	+	X	27	ENSG00000184735	22999960	yes	yes	yes	3,73	5,14	-0,49	0,93	1,65E-22	10,18	0,0334	1,71	0,9711
DNAH8	CG	+	6	28	ENSG00000124721	38723122	yes	yes	yes	0,20	2,01	-0,10	0,91	2,25E-36	NA	NA	NA	NA
DSCR8	CG	+	21	31	ENSG00000198054	38121451	yes	yes	yes	8,04	21,78	NA	NA	NA	9,36	0,0452	0,80	0,8098
FAM9C	CG	-	X	34	ENSG00000187268	13044602	yes	yes	yes	1,18	30,26	-0,30	0,97	8,33E-07	NA	NA	NA	NA
FMR1NB	CG	+	X	37	ENSG00000176988	147981337	yes	yes	yes	1,96	43,87	-0,43	0,91	1,47E-29	NA	NA	0,47	0,7236
FTHL17	CG	-	X	59	ENSG00000132446	31072041	yes	yes	yes	2,94	74,92	-0,69	0,94	3,94E-17	NA	NA	1,27	0,9455
GDPD4	CG	-	11	30	ENSG00000178795	77301649	yes	yes	yes	0,39	7,11	-0,13	0,96	1,22E-08	NA	NA	NA	NA
INSL6	CG	-	9	50	ENSG00000120210	5185639	yes	NA	yes	3,53	19,98	-0,55	0,94	7,48E-18	4,90	0,2750	0,70	0,9455
LIN28B	CG	+	6	18	ENSG00000187772	104936616	yes	yes	yes	6,47	8,42	NA	NA	NA	27,77	6,61E-05	0,73	0,3389
LINC01518	CG	-	10	26	ENSG00000233515	42691721	yes	no	yes	13,33	12,57	NA	NA	NA	16,43	0,0035	0,78	0,7236
LINC02492	CG	-	4	22	ENSG00000250590	187672755	no	NA	NA	6,27	4,14	NA	NA	NA	9,27	0,0454	0,74	0,2798
MAGEA1	CG	+	X	23	ENSG00000198681	153179284	yes	yes	yes	14,90	53,06	-0,58	0,90	2,23E-37	12,77	0,0160	0,73	0,0470
MAGEA10-MAGEA5	CG	-	X	35	ENSG00000266560	152138493	NA	NA	NA	2,75	3,85	-0,48	0,91	2,66E-36	5,32	0,2462	0,64	0,1838
MAGEA11	CG	+	X	32	ENSG00000185247	149712060	yes	yes	yes	5,49	13,24	NA	NA	NA	11,57	0,0220	0,70	0,3389
MAGEA12	CG	+	X	28	ENSG00000213401	152733757	yes	yes	yes	19,61	56,72	-0,66	0,91	5,74E-33	20,69	0,0010	0,84	0,9455
MAGEA4	CG	+	X	23	ENSG00000147381	151912495	yes	yes	yes	11,57	142,20	-0,44	0,91	9,70E-35	12,17	0,0192	0,69	0,7519
MAGEB2	CG	+	X	34	ENSG00000099399	30215563	yes	yes	yes	10,00	46,08	-0,57	0,92	7,04E-25	19,38	0,0014	0,84	0,8783
MAGEC1	CG	+	X	27	ENSG00000155495	141903894	yes	yes	yes	8,24	28,82	-0,51	0,91	3,09E-31	11,98	0,0192	0,84	0,9455
MAGEC2	CG	-	X	22	ENSG00000046774	142205290	yes	yes	yes	15,88	45,32	-0,54	0,91	2,65E-36	20,99	0,0010	0,75	0,7236
MKRN9P	CG	-	12	23	ENSG00000258128	87784568	yes	yes	no	4,51	7,74	-0,64	0,95	1,18E-11	5,31	0,2462	2,06	0,8783
NAA11	CG	-	4	27	ENSG00000156269	79326061	yes	yes	yes	5,88	10,31	-0,53	0,92	5,02E-28	1,26	0,8049	4,86	0,8288
PASD1	CG	+	X	25	ENSG00000166049	151563729	yes	yes	yes	3,53	30,31	-0,37	0,91	7,87E-30	0,64	0,8996	0,53	0,2723
RNF17	CG	+	13	37	ENSG00000132972	24764169	yes	yes	yes	0,98	5,69	-0,39	0,91	1,29E-30	NA	NA	NA	NA
TAF7L	CG	-	X	29	ENSG00000102387	101293057	yes	yes	yes	6,86	13,18	-0,10	0,91	4,18E-33	11,10	0,0251	1,18	0,7236
TDRD1	CG	+	10	36	ENSG00000095627	114179270	yes	yes	yes	10,59	15,44	-0,86	0,96	6,63E-09	10,47	0,0306	1,86	0,1243
ZBTB46-AS1	CG	+	20	52	ENSG00000231208	63808076	yes	NA	yes	5,29	8,02	-0,73	0,92	6,87E-25	4,71	0,2796	0,84	0,8973
DQX1	GI	-	2	10	ENSG00000144045	74526191	yes	no	yes	19,41	7,11	-0,79	0,94	2,41E-13	1,60	0,7502	0,76	0,8783
EPS8L3	GI	-	1	12	ENSG00000198758	109763923	yes	NA	NA	18,63	40,76	-0,66	0,94	7,55E-14	2,62	0,5562	0,62	0,0512
MUC13	GI	-	3	10	ENSG00000173702	124934747	no	yes	yes	59,02	120,55	-0,61	0,96	7,91E-09	0,92	0,8734	0,80	0,2798
PHGR1	GI	+	15	7	ENSG00000233041	40351033	NA	NA	no	8,43	18,94	-0,42	0,94	1,56E-16	2,69	0,5544	0,78	0,7519
SLC22A18AS	GI	-	11	23	ENSG00000254827	2903490	no	yes	no	86,27	11,34	-0,54	0,96	7,16E-09	2,62	0,3693	1,38	0,7236
SLC39A5	GI	+	12	7	ENSG00000139540	56230063	no	no	no	10,78	18,86	-0,56	0,94	2,71E-14	0,61	0,8049	1,09	0,8783
VIL1	GI	+	2	11	ENSG00000127831	218419123	yes	no	yes	42,16	36,29	-0,73	0,94	3,27E-15	3,79	0,3775	1,19	0,8783
A2ML1	SE	+	12	10	ENSG00000166535	8822621	yes	no	NA	6,47	43,07	-0,45	0,94	9,07E-14	3,91	0,3693	0,70	0,1838
ADGRF4	SE	+	6	8	ENSG00000153294	47698580	yes	no	no	55,29	12,62	-0,53	0,95	1,17E-14	45,29	3,94E-08	0,62	0,0322
DSG3	SE	+	18	12	ENSG00000134757	31447741	NA	yes	yes	7,06	37,95	-0,33	0,95	4,74E-12	11,03	0,0251	0,44	0,0322
FAM25A	SE	+	10	12	ENSG00000188100	87020294	no	NA	yes	3,73	9,43	-0,20	0,93	2,82E-20	18,10	0,0019	0,69	0,1243
FAM83A	SE	+	8	32	ENSG00000147689	123183158	yes	yes	yes	95,88	95,37	-0,53	0,96	6,47E-07	38,85	4,58E-07	0,44	0,0000
GJB5	SE	+	1	13	ENSG00000189280	34755047	yes	NA	yes	39,61	18,62	-0,35	0,93	9,98E-19	4,75	0,2796	0,66	0,0446
GPR87	SE	-	3	8	ENSG00000138271	151316820	yes	NA	yes	62,16	39,92	-0,40	0,96	3,43E-08	18,81	0,0016	0,40	0,0322
KRT16	SE	-	17	27	ENSG00000186832	41612899	yes	yes	yes	60,39	41,10	-0,25	0,94	1,13E-13	10,99	0,0251	0,59	0,0050
SERPINB5	SE	+	18	24	ENSG00000206075	63476958	yes	yes	yes	43,73	36,59	-0,25	0,93	7,97E-18	17,53	0,0023	0,63	0,0217
TPRXL	SE	+	3	11	ENSG00000180438	14017401	yes	yes	no	5,29	6,93	NA	NA	NA	0,81	0,8823	0,73	0,6652

Table S3: Methylomic and transcriptomic datasets of normal cells and tissues.

Sample	Type	Pathology	Accession	File	Assay	Genome built	Method	Layout	Platform	Reference
alveolar type II	primary cells	normal	SRR1773103	FastQ	methylation	not applicable	Whole genome bisulfite-seq	paired	Illumina HiSeq 2000	Zuber et al. 2016
sperm	primary cells	normal	GSM1127119	wig	methylation	hg19	Whole genome bisulfite-seq	paired	Illumina HiSeq 2000	Roadmap Epigenomics 2015
HUES64	primary cells	normal	ENCFF770UYJ	bed	methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq 2000	Roadmap Epigenomics 2015
keratinocytes	primary cells	normal	GSM1127056	wig	methylation	hg19	Whole genome bisulfite-seq	paired	Illumina HiSeq 2000	Roadmap Epigenomics 2015
duodenum crypt	primary cells	normal	GSE141254	txt	methylation	not applicable	Infinium methylation assay	not applicable	HumanMethylation450 BeadChips Infinium MethylationEPIC arrays	Lewis et al. 2020
adipose	tissue	normal	ENCFF318AMC	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
esophagus	tissue	normal	ENCFF625GVK	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
heart (left ventricule)	tissue	normal	ENCFF536RSX	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
liver	tissue	normal	ENCFF356KGQ	bam	methylation	not applicable	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
lung	tissue	normal	ENCFF039JFT	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
pancreas	tissue	normal	ENCFF763RUE	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
pre-frontal cortex	tissue	normal	SRR3278486_SRR3278482	FastQ	methylation	not applicable	Whole genome bisulfite-seq	paired	Illumina HiSeq 2500	Jenkinson et al. 2017 (3)
sigmoid colon	tissue	normal	ENCFF157POM	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
skin (lower leg)	tissue	normal	ENCFF219GCQ	bed	methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq X Ten	ENCODE Project Consortium, 2012
small intestine	tissue	normal	ENCFF241AQC	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
stomach	tissue	normal	ENCFF497YOO	bed	methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
testis	tissue	normal	ENCFF715DMX	bed	methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq X Ten	ENCODE Project Consortium, 2012
thyroid	tissue	normal	ENCFF223JJW	bed	methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq X Ten	ENCODE Project Consortium, 2012
alveolar type II	primary cells	normal	SRR1852851	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Marconett et al. 2017
adipose	tissue	normal	ERR315332_ERR315431 ERR315343_ERR315342 ERR315378	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014 (4)
colon	tissue	normal	ERR315357_ERR315484 ERR315462_ERR315400 ERR315348_ERR315403	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
cerebral cortex	tissue	normal	ERR315455_ERR315432 ERR315477	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
esophagus	tissue	normal	ERR315411_ERR315398 ERR315489_ERR315434 ERR315472_ERR315362	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
heart	tissue	normal	ERR315384_ERR315328 ERR315356_ERR315367 ERR315413_ERR315435 ERR315389_ERR315331 ERR315430	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
small intestine	tissue	normal	ERR315344_ERR315381 ERR315408_ERR315409 ERR315423_ERR315388 ERR315419_ERR315364	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
liver	tissue	normal	ERR315327_ERR315414 ERR315463_ERR315394 ERR315451	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
lung	tissue	normal	SRR577579_SRR577582	FastQ	expression	not applicable	bulk RNA-seq		Illumina HiSeq 2000	Roadmap Epigenomics, 2015
pancreas	tissue	normal	ERR315466_ERR315436 ERR315479_ERR315429	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
skin	tissue	normal	ERR315339_ERR315401 ERR315460_ERR315376 ERR315372_ERR315464	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
stomach	tissue	normal	ERR315379_ERR315369 ERR315467_ERR315485	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
testis	tissue	normal	ERR315352_ERR315415 ERR315492	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
thyroid	tissue	normal	ERR315363_ERR315397 ERR315358_ERR315412 ERR315428_ERR315491 ERR315483_ERR315422 ERR315337	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014

Table S4: Transcriptomic datasets of cell lines treated with 5-azadC.

Sample	Type	Pathology	Name	Accession	File	Assay	Method	Layout	Platform	Reference
MCF7	cell line	tumor	MCF7_RNAseq_control_1day-Normoxia_r1a	SRR7822252	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_control_1day-Normoxia_r1b	SRR7822253	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_control_1day-Normoxia_r2a	SRR7822254	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_control_1day-Normoxia_r2b	SRR7822255	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_control_1day-Normoxia_r3a	SRR7822256	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_control_1day-Normoxia_r3b	SRR7822257	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_aza_1day-Normoxia_r1a	SRR7822264	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_aza_1day-Normoxia_r1b	SRR7822265	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_aza_1day-Normoxia_r2a	SRR7822266	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_aza_1day-Normoxia_r2b	SRR7822267	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_aza_1day-Normoxia_r3a	SRR7822268	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
MCF7	cell line	tumor	MCF7_RNAseq_aza_1day-Normoxia_r3b	SRR7822269	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2500	D'Anna et al. 2020
TS603	cell line	tumor	TS603_DMSO_rep1	SRR12105780	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Park et al. 2021
TS603	cell line	tumor	TS603_DMSO_rep2	SRR12105781	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Park et al. 2021
TS603	cell line	tumor	TS603_DAC_rep1	SRR12105782	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Park et al. 2021
TS603	cell line	tumor	TS603_DAC_rep2	SRR12105783	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Park et al. 2021
HMLER	cell line	normal immortalized	HMLER_DAC_1	SRR3362409	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Grandin et al. 2016
HMLER	cell line	normal immortalized	HMLER_DAC_2	SRR3362410	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Grandin et al. 2016
HMLER	cell line	normal immortalized	HMLER_Ctrl1	SRR3362411	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Grandin et al. 2016
HMLER	cell line	normal immortalized	HMLER_Ctrl2	SRR3362412	FastQ	expression	bulk RNA-seq	single	Illumina HiSeq 2000	Grandin et al. 2016

Table S5: Antibodies and tissue section codes from HPA.

Protein	DDIC	Antibody	Tissue	Pathology	Gender	Age	Patient_ID
MUC13	GI	HPA045163	Lung	Normal	Male	65	1470
VIL1	GI	HPA006884	Lung	Normal	Female	49	2268
A2ML1	SE	HPA038847	Esophagus	Normal	Female	66	3399
A2ML1	SE	HPA038847	Skin	Normal	Male	52	3338
A2ML1	SE	HPA038847	Vagina	Normal	Female	40	2276
A2ML1	SE	HPA038847	Lung	Normal	Male	65	1470
SERPINB5	SE	CAB009570	Esophagus	Normal	Male	54	3197
SERPINB5	SE	CAB009570	Skin	Normal	Male	16	2549
SERPINB5	SE	CAB009570	Vagina	Normal	Female	44	2480
SERPINB5	SE	CAB009570	Lung	Normal	Female	49	2268
MUC13	GI	HPA045163	LUAD	Tumor	Female	69	2777
VIL1	GI	HPA006884	LUAD	Tumor	Female	70	3391
A2ML1	SE	HPA038847	LUAD	Tumor	Female	76	448
SERPINB5	SE	CAB009570	LUAD	Tumor	Female	70	3391

Table S6: siRNAs used for transfection experiments.

siRNA	Sequence (5'-3')	Reference	Manufacturer
siHNF4A	GAC-CGG-AUC-AGC-ACU-CGA-A CGG-AAG-AAC-CAC-AUG-UAC-U GGG-CUG-GCA-UGA-AGA-AGG-A CCA-AGU-ACA-UCC-CAG-CUU-U	L-003406-00-0020 ON-TARGETplus SMARTpool	Dharmacon
siLuciferase	CUUACGCUGAGUACUUCGA	Tilman et al. 2012	Eurogentec

Table S7: Primer sequences, PCR and qPCR reagents and conditions.

Gene	Forward primer (5'-3')	Reverse primer (5'-3')	Assay	Kit	Manufacturer	Temperature/Time	Cycles
NAP1L1	CCT-GGA-TCT-GAG-AGC-TTC-TCT-T	ACA-CCG-CTC-GCG-ATC-CAA-T	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
NAA11	GCA-GTG-ACA-GCA-AAG-AAC-CTA	GAT-CCC-AGC-AGG-ATA-TGT-GAA	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
EPS8L3	TGA-GCT-CGT-ACA-CAT-CCT-CTT	ACA-GGG-TCC-TGG-TAT-CCT-A	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
VIL1	CGT-GTT-CAA-TGC-TAA-CAG-CAA-C	ATG-AGA-CCC-TAC-AAT-CAG-GGT-A	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
GJB5	GCA-GGC-TCT-GTC-CTG-GAA-ACA	CGA-GTA-TTG-CAG-TCG-AAG-TCC-T	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
SERPINB5	TTC-CAG-GAT-AAC-TGT-GAC-T	TCC-AAA-GGG-TAC-ATC-TTT-GAC-A	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
MAGEA1	GCC-GAA-GGA-ACC-TGA-CC	ACT-GGG-TTG-CCT-CTG-TCG	PCR	DreamTaq	ThermoFischer	Annealing: 62°C, 30s Extension: 72°C, 1min	35
PGLYRP3	CGT-CTA-CAC-CAT-AGG-CTG-GT	CCT-TCT-GGA-TGG-CAT-AGG-AGA-TCA	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	36
ACTB	CCC-TGG-ACT-TCG-AGC-AAG-AGA-T	AAG-GTA-GTT-TCG-TGG-ATG-CCA-CA	PCR	DreamTaq	ThermoFischer	Annealing: 60°C, 30s Extension: 72°C, 30s	20
HNF4A	CAT-ACG-CAT-CCT-TGA-CGA-GCT	GAT-GAA-CTG-GAT-CTG-CTC-GAT	qPCR	KAPA SYBR FAST	Sigma Aldrich	Annealing: 60°C, 30s Extension: 60°C, 30s	40
MUC13	GCC-ATC-ATT-CAT-CTT-ACT-CTT-CT	TCA-CTG-TCT-GCA-GCA-GTA-GGT	qPCR	KAPA SYBR FAST	Sigma Aldrich	Annealing: 60°C, 30s Extension: 60°C, 30s	40
CT-GABRA3	GAA-AGA-AAG-AAA-GGT-CAC-AGG-TCT-CT	GCC-AAT-GTC-CTG-CTT-CAC-AAA-GT	qPCR	KAPA SYBR FAST	Sigma Aldrich	Annealing: 60°C, 30s Extension: 60°C, 30s	40
ACTB	CCC-TGG-ACT-TCG-AGC-AAG-AGA-T	AAG-GTA-GTT-TCG-TGG-ATG-CCA-CA	qPCR	KAPA SYBR FAST	Sigma Aldrich	Annealing: 60°C, 30s Extension: 60°C, 30s	40
EPS8L3	TGA-GCT-CGT-ACA-CAT-CCT-CTT	ACA-GGG-TCC-TGG-TAT-CCT-A	qPCR	KAPA SYBR FAST	Sigma Aldrich	Annealing: 60°C, 30s Extension: 60°C, 30s	40
VIL1	CGT-GTT-CAA-TGC-TAA-CAG-CAA-C	ATG-AGA-CCC-TAC-AAT-CAG-GGT-A	qPCR	KAPA SYBR FAST	Sigma Aldrich	Annealing: 60°C, 30s Extension: 60°C, 30s	40

CHAPTER 2

An abundance of brain-specific genes showing ectopic activation in lung adenocarcinoma

Anna Diacofotaki [1], Axelle Loriot [1,2] and Charles De Smet [1]

[1] Group of Genetics and Epigenetics, de Duve Institute, Université Catholique de Louvain; Brussels, Belgium

[2] Group of Computational Biology and Bioinformatics, de Duve Institute, Université Catholique de Louvain ; Brussels Belgium

Abstract

Tumor cells exhibit genetic and epigenetic alterations which perturb the original gene expression landscape. It has been proposed that this may lead to a so-called "cancer identity crisis", resulting from the activation of unscheduled tissue-specific gene expression programs. To address this issue, we performed a comprehensive analysis integrating transcriptomic data from lung adenocarcinoma (LUAD) tissues and cell lines, normal lung, and alveolar type 2 cells, which are considered as the origin of LUAD. Our search led to the identification of four tissue-specific gene clusters showing ectopic activation in LUAD. Three of these had been previously described, as they corresponded to the cancer-germline, gastro-intestinal, and stratified epithelium gene clusters, which become activated in LUAD in association with the process of genome-wide DNA demethylation. The fourth cluster we identified included a large group of genes with specific expression in brain (CB genes, n=63) or in brain and testis (CBT genes, n=28). In the brain, CB and CBT genes are predominantly expressed in neuronal cells, and many of them exert functions related to signaling transmission. Multiple mechanisms appear to contribute to induction of CB and CBT genes in LUAD. For 7% of these genes, activation was correlated with upregulation of *ASCL1*, a master regulator governing neuroendocrine transdifferentiation in 10-20% of LUAD. Analysis of methylomic data suggested that about 10% of CB/CBT genes are induced in LUAD following demethylation of their promoter. Finally, the use of a transcription factor prediction tool (i-*cis*Target) identified the REST repressor as a likely regulator of at least 25% of the CB/CBT genes. Together, our data suggest that alterations in various regulatory mechanisms contribute to the activation of a set of brain-related genes in LUAD, raising the possibility that acquisition of neuronal functions might support tumor cell development.

1. Introduction

Tumoral transformation is commonly associated with extensive changes in the gene expression program of the cell. These alterations are caused on the one hand by genetic alterations, such as mutations, deletions, and translocations, which impact signaling pathways, and on the other hand, by epigenetic alterations that remodel the chromatin landscape (You and Jones 2012; Chan and Hughes 2015). It is now clear that the epigenetic versatility of tumor cells contributes to their capacity to adapt to the evolving environment, and accelerates acquisition of malignant hallmarks (Flavahan, Gaskell, and Bernstein 2017; Hanahan 2022).

The earliest epigenetic alteration that was observed in tumor cells, was the genome-wide loss of DNA methylation, a mark present on cytosines in CpG sequences that mediates long-term transcriptional repression (Feinberg and Vogelstein 1983a; Gama-Sosa et al. 1983a). DNA hypomethylation was observed in a vast majority of tumors, although to different extents, and has been associated with genome instability, probably due to the demethylation of repeated sequences, and with changes in gene expression (Ehrlich 2009a). A remarkable consequence of DNA hypomethylation in tumors is the ectopic activation of gene clusters that rely on DNA methylation for their tissue-restricted expression. It was observed indeed that DNA hypomethylation is associated with the aberrant activation of a group of germline-specific genes in a variety of tumors of somatic origin (De Smet and Loriot 2013). More recently, two other tissue-specific gene clusters were found to be activated in association with DNA hypomethylation in lung tumors: one displaying restricted expression in the gastro-intestinal tract, and the other, showing specific expression in stratified squamous epithelia (Diacofotaki, Loriot, and De Smet 2022). These observations illustrate the concept of “cellular identity crisis” in cancer, where tumor cells activate unscheduled gene expression programs with potential impacts on their phenotypic characteristics (Roy and Hebrok 2015).

To our knowledge, there have been only few studies aiming at characterizing the tissue-specific gene clusters that become ectopically activated in tumors of unrelated histological origins (Axelsen et al. 2007; Lotem et al. 2005). Such gene clusters are nevertheless of particular interest, as their ectopic activation in tumors offers diagnostic and prognostic opportunities (Shukla et al. 2017; Somerville et al. 2018; Rousseaux et al. 2013). Some of these genes may also exert cellular

functionalities that contribute to tumor development (Van Tongelen, Loriot, and De Smet 2017).

Taking advantage of the vast amount of publicly available transcriptomic datasets that were generated from normal and tumoral tissues and cells, we decided to re-investigate the existence of tissue-specific gene clusters showing ectopic activation in tumor cells. We chose lung adenocarcinoma (LUAD) for this exploration, because multiple omics datasets are available for LUAD tissues and cell lines. Moreover, the cell at the origin of this tumor (lung alveolar type II cells, AT2) has been characterized (Blanpain 2013; Ferone et al. 2020), and has been analyzed at the transcriptomic and methylomic levels (Weeden et al. 2017; Zuber et al. 2016). qWe thus performed an integrative analysis of RNA-seq data from LUAD tissues and cell lines, normal lung, and AT2 cells, as well as of single-cell RNA-seq of normal tissues. Remarkably, we observed activation of a neuronal-specific transcriptional network in LUAD cells. The mechanisms underlying ectopic activation of this particular gene cluster were investigated.

2. Materials and methods

Public datasets and files

- *Normal cells and tissue samples*

Alveolar type 2 cells: FastQ files of bulk RNA-seq experiments of alveolar type 2 cells originating from small (n=3) and large (n=3) airways of normal lung were obtained from (Weeden et al. 2017). Quality control of reads was assessed using FastQC v0.11.8 (Brabraham institute) and reads of low quality were discarded using Trimmomatics v0.3 (Bolger, Lohse, and Usadel 2014). Reads were aligned onto hg38 genome using HISAT2 v2.1.0, (Kim et al. 2019). FeatureCounts from the subread package v2.0.0 (Liao, Smyth, and Shi 2014) was used for gene quantification. Expression data for each gene were averaged for small and large airway tissue samples (n=6 in total) and measured in Transcripts Per Million (TPM). For CpG methylation analyses, FastQ files of whole genome bisulfite-seq of AT2 cells were obtained from the study of (Zuber et al. 2016). To assess read quality control and discard low-quality reads, Trim Galore! v0.5.0 was used

(https://github.com/FelixKrueger/TrimGalore). Read alignment onto the hg38 genome and methylation calling were performed using Bismark v0.20.0 (Krueger and Andrews 2011). When available, methylation percentages of the same CpG site sequenced in both forward and reverse strands were averaged. Accession numbers of all files are listed in Table S1.

Normal tissue samples: For methylation analyses of normal lung, cerebral cortex, testis, adipose tissue, colon, esophagus, heart, small intestine, liver, skeletal muscle, pancreas, skin, stomach and thyroid, files were obtained and processed as previously described (Diacofotaki, Loriot, and De Smet 2022). For the visualization of mapped reads onto the genome of normal cerebral cortex and testis tissue samples, corresponding FastQ files were downloaded from SRA and processed as previously described (Diacofotaki, Loriot, and De Smet 2022). Accession numbers of all files are listed in Table S1.

GTEx consortium: Normalized bulk RNA-seq data of normal tissues were downloaded from the GTEx portal v7 (Consortium 2013). Expression data for each tissue are measured in median TPM of all samples per tissue analyzed.

Human Protein Atlas consortium: Normalized single cell RNA-seq expression data were downloaded from the Human Protein Atlas (HPA) v20.1 (Uhlen et al. 2005). Expression data from different clusters of the same cell type were averaged to obtain a mean expression of each gene per cell type. Data are expressed in pTPM (transcripts per million of protein coding genes).

- *LUAD cell lines and tissue samples*

LUAD cell lines: FastQ files of bulk RNA-seq and methyl-seq experiments of 26 LUAD cell lines were obtained from the DBTSS portal (Suzuki et al. 2014) and processed as previously described (Fain et al. 2021). All accession numbers are listed in Table S2.

The Cancer Genome Atlas consortium: Normalized bulk RNA-seq FPKM (fragments per kilobase per million) data of the LUAD project (Cancer Genome Atlas Research 2014) were downloaded using the TCGAbiolinks R package v2.14.1 (Colaprico et al. 2016). Data were converted to TPM by dividing each FPKM gene expression in a sample by the sum of all gene expressions for that sample ($\times 10^6$). LUAD tissues were

defined as positive for the expression of a specific gene when TPM ≥ 2, and negative when TPM < 1. All accession numbers are listed in Table S1.

Identification of genes ectopically activated in LUAD cells

a) *Selection of genes initially repressed in AT2 cells and normal lung*. Genes that displayed an average expression TPM < 0.5 in AT2 cells and normal lung tissue samples of TCGA (n=58) and GTEx (n=578), were considered transcriptionally inactive in these cells and tissues.
b) *Selection of genes activated in LUAD tissue samples.* Among the above selected genes, we retained those that showed an expression TPM ≥ 2 in at least 5% of the total number of LUAD tissue samples available from TCGA (n=510).
c) *Selection of genes activated in LUAD cells.* To restrict gene expression to tumor cells and not their surrounding micro-environment, we retained the genes that were activated in at least one LUAD cell line and repressed in at least two LUAD cell lines of the 26 LUAD cell lines available from DBTSS. Activation in these cells was defined as TPM ≥ 2, and repression as TPM ≤ 0.1.

Definition of tissue-specific gene clusters in LUAD

Cancer-Brain (CB) genes were defined as follows: they showed either an expression TPM ≥ 2 in at least one brain tissue from those available in GTEx and TPM < 1 in all other tissue samples, or they showed a median expression ≥ 1 TPM in at least one brain tissue and in maximum one other tissue (other than the testis). Genes that displayed a median expression TPM ≥ 1 in at least one of the brain tissues available from GTEx and in testis, and exhibited a median expression TPM < 1 in all other tissues, were defined as Cancer-Brain/Testis (CBT) genes. Genes that showed a median expression TPM ≥ 2 only in testis of GTEx, or a median expression TPM ≥ 1 in testis and in maximum one other tissue, were defined as Cancer-Testis (CT) genes.

For the enrichment analysis of observed vs expected genes, the expected CB and CBT genes were defined in the same way as the isolated CB and CBT genes, this time among all genes that are initially silent in normal lung and AT2 cells.

Overrepresentation analyses using Gene Ontology annotation

To evaluate a potential enrichment of molecular functions and biological process subcategories of Gene Ontology annotations for a gene list, we used clusterProfiler R package v4.0.2 (Yu et al. 2012).

Analysis of the methylation status of promoter regions

To accurately identify the TSS of each CB and CBT gene and confirm their activation in LUAD cells, we visualized BAM files of cerebral cortex, testis tissue samples and of 26 LUAD cell lines in the IGV software (Robinson et al. 2011). To measure promoter DNA methylation of each gene of interest, we averaged the methylation of each CpG site located at TSS -/+ 400 bp in each LUAD cell line.

i-cis-Target analysis for transcription factor binding

To identify transcription factors involved in the regulation of CB and CBT genes, we used i-*cis*Target webtool with default parameters and ENSEMBL gene names as input. We selected the following ChIP studies against REST: HCT116 (ENCFF001UEC), MCF7 (ENCFF001UNI) and PANC1 (ENCFF001UOD). When multiple conserved regions of the same gene displayed an enrichment for REST, the rank selected for the enrichment ranking percentage calculation was the one coming from the region closest to the TSS of that gene.

RNA-seq of C2 adult dermal fibroblasts transfected with sh-REST

FastQ files of C2 cells transfected with a sh-RNA directed against REST or control were obtained from the study of (Drouin-Ouellet et al. 2017). FastQ files were processed as described above for AT2 cells. For differential expression analysis, the DESeq2 R package v1.32.0 was used, with counts generated by Featurecounts as input for the analysis.

Statistical analyses and graphical representations

Statistical analyses were performed in R 4.1.0 (http://www.R-project.org). p-values were adjusted using the Benjamini-Hochberg method and a p-adjusted value < 0.05 was considered statistically significant. Graphs were generated with the following R packages: ggplot2 v.3.3.5, ComplexHeatmap v2.8.0 (Gu, Eils, and Schlesner 2016) and clusterProfile v4.0.2. For unsupervised hierarchical clustering analyses of gene

expression in GTEx and TCGA tissue samples or HPA single cells, the Ward.D method and Euclidean distance were used as arguments in the Heatmap function.

3. Results

1. LUAD cells activate germline and somatic tissue-specific expression clusters

In order to uncover tissue-specific gene clusters showing ectopic activation in LUAD cells, we integrated RNA-seq datasets of AT2 cells, normal tissues (GTEx), LUAD cell lines (DBTSS) and LUAD tissue samples (TCGA), and applied the following selection workflow (**Fig. 1A**). We first selected genes that are silent in AT2 cells and in normal lung. Among these genes, we retained those that were activated in at least 5% of the available LUAD tissue samples. Finally, in order to isolate genes with ectopic activation in tumor cells, rather than in nonmalignant cells of the tumor microenvironment, we selected the genes that were also activated in at least one out of the 26 analyzed LUAD cell lines. Our procedure resulted in the selection of 1051 genes that were silent in AT2 cells and normal lung, and that showed ectopic activation in LUAD cells (EctopEx genes).

In the next step, we searched to determine the expression profile of EctopEx genes among healthy tissues by examining RNA-seq datasets from a panel of normal tissue samples (GTEx; **Fig. 1B**). Unsupervised hierarchical clustering of gene expression in these tissues was applied, and revealed several groups of EctopEx genes showing congruent expression in defined tissues. First, there was a group of genes that showed their highest expression in the testis. Closer examination of the genes in this category revealed the presence of multiple genes of the cancer-germline (CG) group of genes. Another group of EctopEx genes showed maximum expression in the esophagus, skin and vagina, and was therefore reminiscent of the stratified epithelium (SE) gene cluster that we described previously (Diacofotaki, Loriot, and De Smet 2022). A third group of genes with predominant expression in colon and small intestine, and occasionally in liver, stomach and/or pancreas, corresponded to the previously identified gastrointestinal (GI) gene cluster (Diacofotaki, Loriot, and De Smet 2022; Shukla et al. 2017). Intriguingly, a large

cluster of EctopEx genes displayed their highest expression in brain tissues, including the hypothalamus, pituitary gland, cerebellum, and/or basal ganglia (putamen, nucleus accumbens or caudate nucleus) (**Fig. 1B**). Several of these genes appeared to be simultaneously expressed in testis. The remaining EctopEx genes displayed disparate patterns of expression among the analyzed tissues, and could therefore not be classified into a defined cluster, or showed expression levels that were below the fixed threshold (TPM < 2) in all tissues. Together, these results revealed that, besides the previously identified CG, SE and GI genes, a large cluster of genes with high expression in the brain become ectopically activated in LUAD cells.

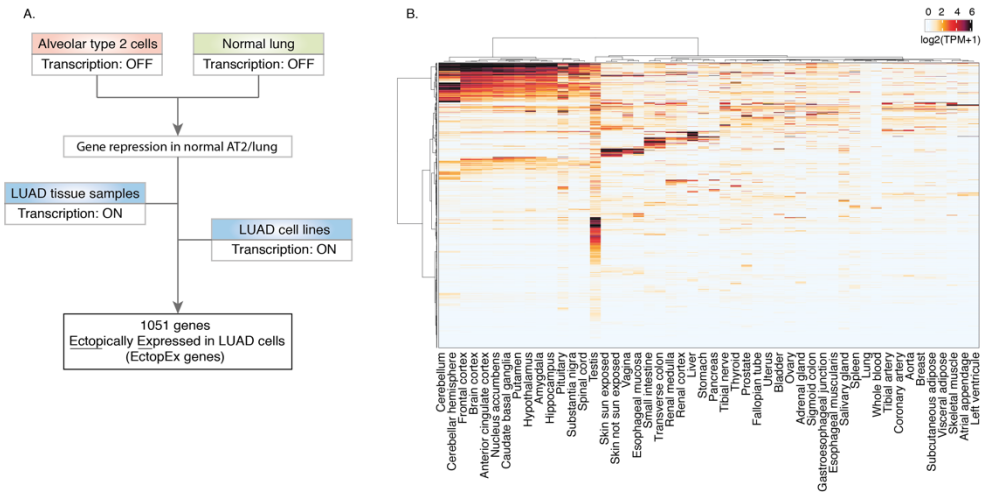


Figure 1: LUAD cells show aberrant activation of genes belonging to tissue-specific expression clusters. *A) Selection criteria used to identify genes ectopically activated in LUAD cells (EctopEx genes). B) EctopEx gene expression in a panel of normal tissues available from GTEx was clustered using Ward's method and Euclidean distance. Expression data are depicted in log2 transformed median TPM + pseudo-count.*

2. *Brain and Brain/Testis genes with ectopic activation in LUAD*

To get further insight into the group of brain-expressed genes showing ectopic activation in LUAD, we first specified a defined list of EctopEx genes belonging to this category by applying the following criteria: genes had to be expressed in at least one brain tissue sample (GTEx) and could also be expressed in maximum one

other tissue. This resulted in the selection of 91 EctopEx genes with preferential expression in the brain. Strikingly, a significant proportion of these genes (28/91) showed expression in testis in addition to the brain. We therefore defined two gene subgroups: the "Cancer-Brain" (CB) genes showing expression in the brain but not in testis (n=63), and the "Cancer-Brain/Testis" (CBT) genes displaying expression in both brain and testis (n=28; **Fig. 2**).

Cancer-Brain genes (CB, n=63)						
ABHD17AP6	*CSMD2*	*IFNE*	*NETO1*	*RP1-161N10.1*	*RP11-98D18.1*	*SLC6A15*
AC018641.7	*DLX6-AS1*	*KCNK12*	*NKAIN2*	*RP1-232L24.3*	*RP13-977J11.8*	*SOX11*
ANKRD34B	*FEZF1-AS1*	*KCNK9*	*OPRD1*	*RP1-90G24.11*	*RP3-470B24.5*	*TMEM74*
C1QL2	*FGF20*	*KISS1R*	*OPRK1*	*RP11-100E13.1*	*RP4-555D20.4*	
CALB1	*FOXB1*	*KRT83*	*PAX7*	*RP11-350E12.4*	*RP5-827C21.2*	
CALCR	*FSTL5*	*LCN15*	*PCDHA6*	*RP11-38M8.1*	*RPSAP69*	
CGA	*FUT9*	*LHX1*	*PCDHA7*	*RP11-416I2.1*	*SIX3OS1-1*	
CNGB1	*GABRG2*	*LHX5*	*POU3F2*	*RP11-525A16.4*	*SIX3OS1-2*	
CNTNAP2	*HCN1*	*LINC00460*	*RGS20*	*RP11-713C5.1*	*SLC18A3*	
CREG2	*HTR2C*	*LINC01833*	*RNU1-122P*	*RP11-863P13.3*	*SLC26A4-AS1*	

Cancer-Brain/Testis genes (CBT, n=28)		
C1orf61	*GPR158*	*RTBDN*
CACNA1B	*GPR19*	*SIX3*
CDH18	*GRIN2A*	*SYNGR4*
CSMD1	*HAR1B*	*SYT16*
CTD-2162K18.4	*KCNQ2*	*SYT5*
DISP3	*KHDC1L*	*ZFR2*
DLL3	*LHFPL4*	*ZIC2*
DPF1	*PCDHA1*	*ZIC5*
ELAVL2	*RP11-317N12.1*	
GABRA3	*RP11-417L19.2*	

Figure 2: Cancer-Brain and Cancer Brain/Testis genes list selected as being ectopically activated in LUAD cells.

One possible explanation for the bimodal expression of CBT genes could be the usage of alternative promoters, one that is active in brain and the other in testis. This was previously described in the *GABRA3* locus, where it was found that transcriptional activation of the gene in tumor cells originated specifically from the testis-specific promoter (Loriot et al. 2014). We therefore decided to precisely determine the transcription start sites (TSS) used by CBT genes in normal brain cortex, testis and LUAD cell lines by visualizing RNA-seq reads with the Integrative Genomics Viewer (IGV). For the majority of CBT genes that could be analyzed (20/25), we observed that transcription in the brain, testis and LUAD cells originated from one same TSS. For only five CBT genes, we observed alternative

usage of different TSS in brain and testis: activation in LUAD cells occurring at the testis-specific TSS in four cases, and at the brain-specific TSS in one case (**Fig. 3**).

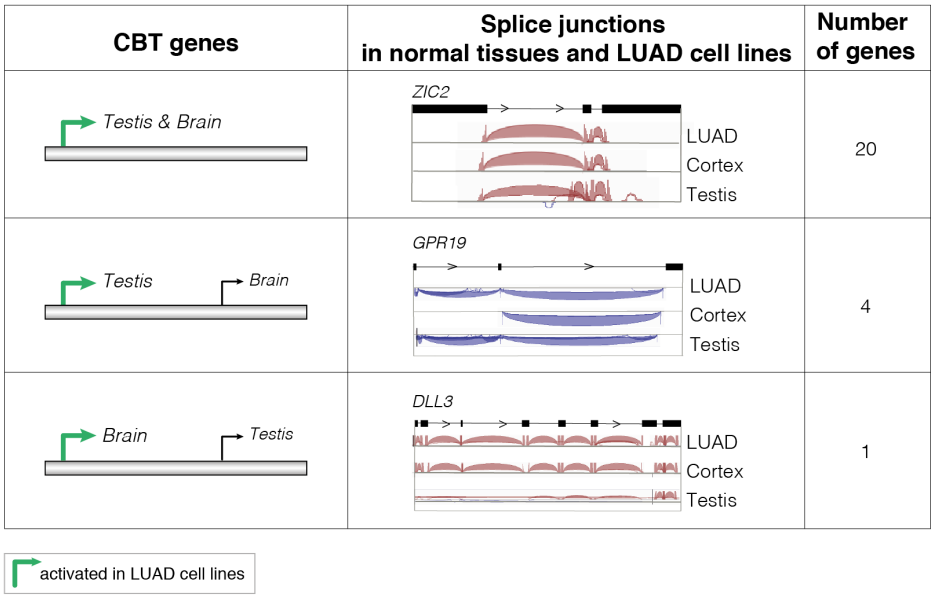


Figure 3: Cancer Brain/Testis genes majorly display a shared transcriptional start site in cerebral cortex and testis tissues activated in LUAD. *The table depicts the different cases of transcriptional start sites (TSS) usage of cancer Brain/Testis (CBT) genes in cerebral cortex, testis and LUAD cell lines. Left: schematic representation of the promoter region of CBT genes with broken arrows depicting the TSS. Middle: screenshot of visualized BAM files of normal cerebral cortex, testis and LUAD cell lines for the precise determination of the tissue-specific TSS activated in LUAD. Black boxes represent gene exons and the connecting lines between them, gene introns. Splice junctions of RNA-seq reads are viewed underneath (red: gene located on the forward strand; blue: gene located on the reverse strand). Right: number of CBT genes that activate a specific TSS in LUAD cells. For 24/28 CBT genes the TSS could be clearly determined.*

3. A majority of CB and CBT genes display neuronal expression and function

As the brain tissue is composed of a heterogeneous cell population, it was important to determine if CB and CBT genes are expressed in a particular cell type. To this end, the expression of CB and CBT genes were examined across single cell RNA-seq data derived from a brain tissue sample (HPA). The results showed that most CB and CBT genes (70%) are expressed in neuronal cells, and only few in

supporting glial cells or in microglial cells (**Fig. 4A**). Consistently, examination of cellular functions associated with CB and CBT genes revealed enrichment for biological processes such as regulation of membrane potential and neural precursor cell proliferation, and for molecular functions related to neuronal signaling transmission (**Fig. 4B**).

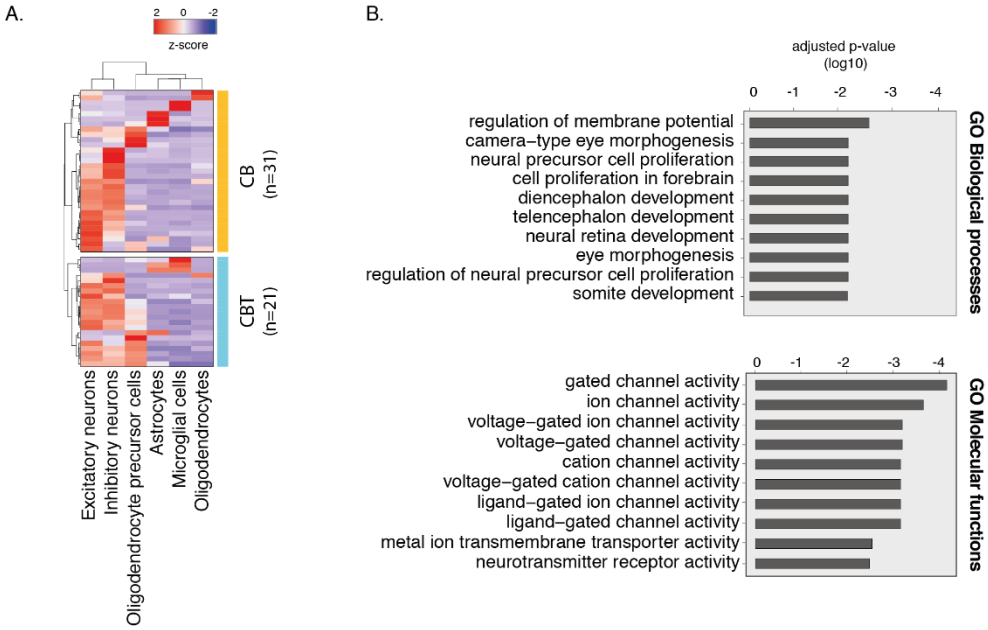


Figure 4: CB and CBT genes are predominantly neuronal-specific genes. *A) Single-cell RNA-seq data of brain tissue available from HPA were explored to determine cell type-specificity of CB and CBT genes. Gene expression was clustered using Ward's method and Euclidean distance in the available cell types of the brain. Data are depicted using a z-score of protein coding transcripts per million. B) Results of overrepresentation analysis of CB and CBT genes using "biological process" (top) and "molecular function" (bottom) subcategories of Gene Ontology annotation. The top10 statistically significant results are depicted (p-adjusted <0.05).*

4. Frequent activation of CB and CBT genes in LUAD samples

The frequency of activation and level of expression of CB and CBT genes in LUAD tissues of TCGA were investigated. Individual CB and CBT gene activation ranged from 5% to 70% in LUAD tissues, with an average activation frequency of 13.8% and a mean expression level of 11.4 TPM (**Fig. 5A**). This was similar to what was

observed for testis-specific genes in these tissues, with an average activation frequency and a mean expression level of 13.6% and 13.2 TPM respectively. The aberrant activation of CB and CBT genes appears therefore as a frequent event in LUAD tissue samples.

The substantial amount of CB and CBT genes that we found to become activated in LUAD suggested that genes with such expression profiles might be selectively activated in tumor cells. Another plausible explanation, however, was that the high number of CB and CBT genes only reflected the preponderance of human genes with specific or preferential expression in brain and testis, as was previously demonstrated (Uhlen et al. 2015). To test these hypotheses, we compared the observed numbers of CB and CBT genes with the numbers of brain- and brain/testis-specific genes that would be expected to become activated by chance in LUAD. The results showed that there was a significant enrichment of CB and CBT genes among the genes ectopically activated in LUAD, that is, CB and CBT genes exceeded the expected numbers if activation of brain and brain/testis-specific genes in LUAD was a random process (**Fig. 5B**). Of note, such an enrichment was not observed for EctopEx genes showing specific expression in testis (**Fig. 5B**).

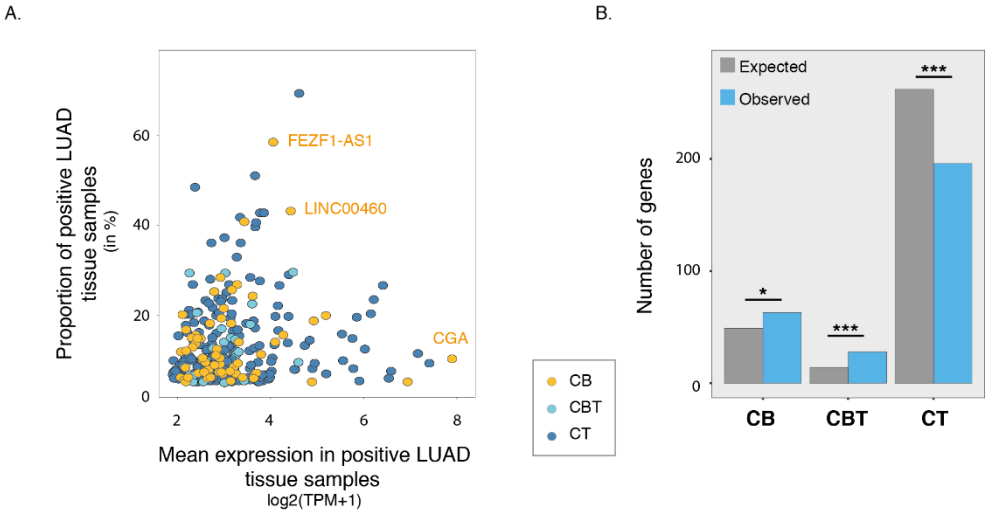


Figure 5: LUAD show a significant activation of CB and CBT genes. *A) Activation frequency and mean expression level of CB, CBT and testis-specific genes were computed in LUAD tissue samples of TCGA. B) Comparison of the observed number of CB, CBT and CT genes selected as being ectopically activated in LUAD, and the corresponding number of expected brain-, brain/testis and testis-specific genes that would be obtained by a random activation in LUAD cells (Chi-squared test, ***: p-value < 0.001, *: p-value < 0.05).*

5. Activation of a small group of CB and CBT genes coincides with neuroendocrine transdifferentiation of LUAD

Previously, studies have reported that about 10 to 20% of LUAD tissue samples exhibit neuroendocrine features (Bhattacharjee et al. 2001; Kosari et al. 2014). This neuroendocrine phenotype is governed by the ectopic expression of *ASCL1* in LUAD cells, a transcription factor responsible for the development of sensory organs and neuroendocrine pulmonary cells (Borges et al. 1997). We therefore wondered whether the observed activation of CB and CBT genes would delineate a neuroendocrine transdifferentiation of LUAD.

We first sought evidence of *ASCL1* activation in LUAD tissues of TCGA, in order to use the expression of this marker as a surrogate for the neuroendocrine phenotype. *ASCL1* activation was detected in 21% of LUAD tissues (107/510 tissue samples, TPM $\geq$ 2), a proportion in line with literature reports (**Fig. 6A**). To determine whether expression of CB or CBT genes overlaps that of *ASCL1* in LUAD tissues, we computed a Pearson correlation coefficient between the two expressions (**Fig. 6B**). Data show that 5/63 CB and 1/28 CBT genes are positively correlated with *ASCL1* expression in LUAD tissues ($\rho > 0.4$, p-adjusted < 0.05). The gene that showed the highest correlation coefficient was *DLL3,* a gene involved in the Notch signaling pathway that has already been described as a target gene of *ASCL1* in small cell lung cancer (Borromeo et al. 2016; Casarosa, Fode, and Guillemot 1999). Overall, our analysis suggests that for a minority of CB and CBT genes (~ 7%) ectopic activation in LUAD coincides with the known process of neuroendocrine trans-differentiation.

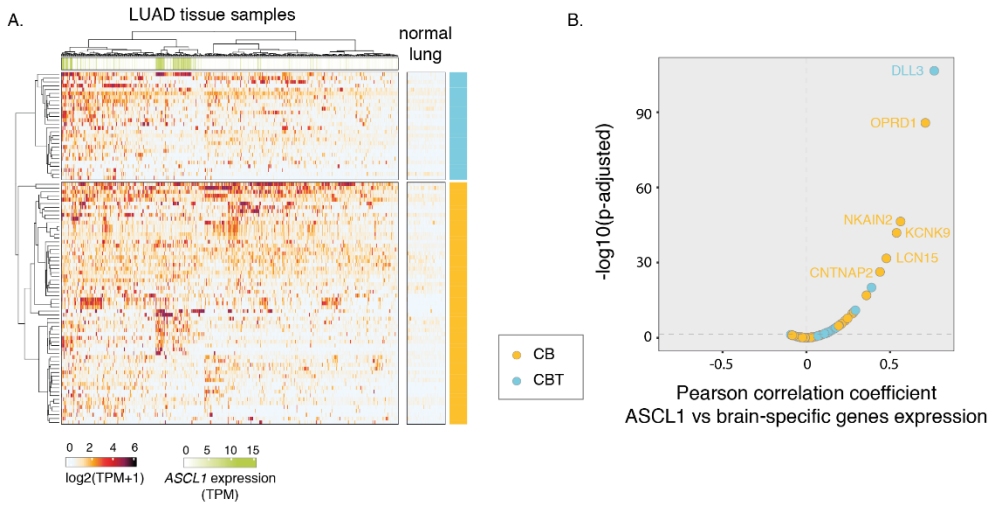


Figure 6: A subgroup of LUAD tissues activating CB and CBT genes is directed towards a neuroendocrine phenotype. *A) Expression data of CB and CBT genes in LUAD tissues of the TCGA were clustered using Ward's method and Euclidean distance inside their tissue-specific categories. ASCL1 expression, a marker of neuroendocrine transdifferentiation of LUAD, is depicted below the heatmap in a light green color gradient. B) Correlation coefficient between ASCL1 and CB and CBT gene expressions in LUAD tissue samples.*

6. *Promoter DNA methylation is associated with the ectopic activation of a subgroup of CB and CBT genes in LUAD cells*

Considering the well described role of DNA hypomethylation in the ectopic activation of tissue-specific gene clusters in tumors (De Smet and Loriot 2013; Diacofotaki, Loriot, and De Smet 2022), we sought to determine whether the activation of some CB and CBT genes in LUAD is also caused by promoter DNA demethylation. To do so, we first examined the initial level of promoter DNA methylation of these genes in AT2 cells and a panel of normal tissues (ENCODE), including cerebral cortex and testis (**Fig. 7A**).

Whereas most EctopEx genes with testis-specific expression showed a methylated promoter region in the assessed cells and tissues (median methylation of promoter region > 50%), CB and CBT genes showed a predominantly unmethylated promoter. Consistently, we found that these two latter groups of genes are enriched in CpG sites (**Fig. 7B**), in accordance with the notion that CpG rich promoters remain methylation-free in somatic cells even when inactive (Bird 1986; Weber et al. 2007).

Nevertheless, we noticed that a subset of CB and CBT genes (11/91) displayed a high level of DNA methylation (> 30%) within their promoter region in most tissues, which was compatible with a potential role of DNA methylation in the regulation of these genes (**Fig. 7C**).

To determine if promoter DNA methylation is implicated in the aberrant activation of this subgroup of CB and CBT genes, we compared the mean methylation level of their promoter region in LUAD cells where they are expressed, and in LUAD cells where they are not (**Fig. 7D**). The results showed that the promoter regions of the subgroup of CB and CBT genes were generally less methylated in LUAD cell lines that express these genes. Together, our data suggest that the transcriptional activation of a subset of CB and CBT genes in LUAD is associated with a process of promoter DNA demethylation.

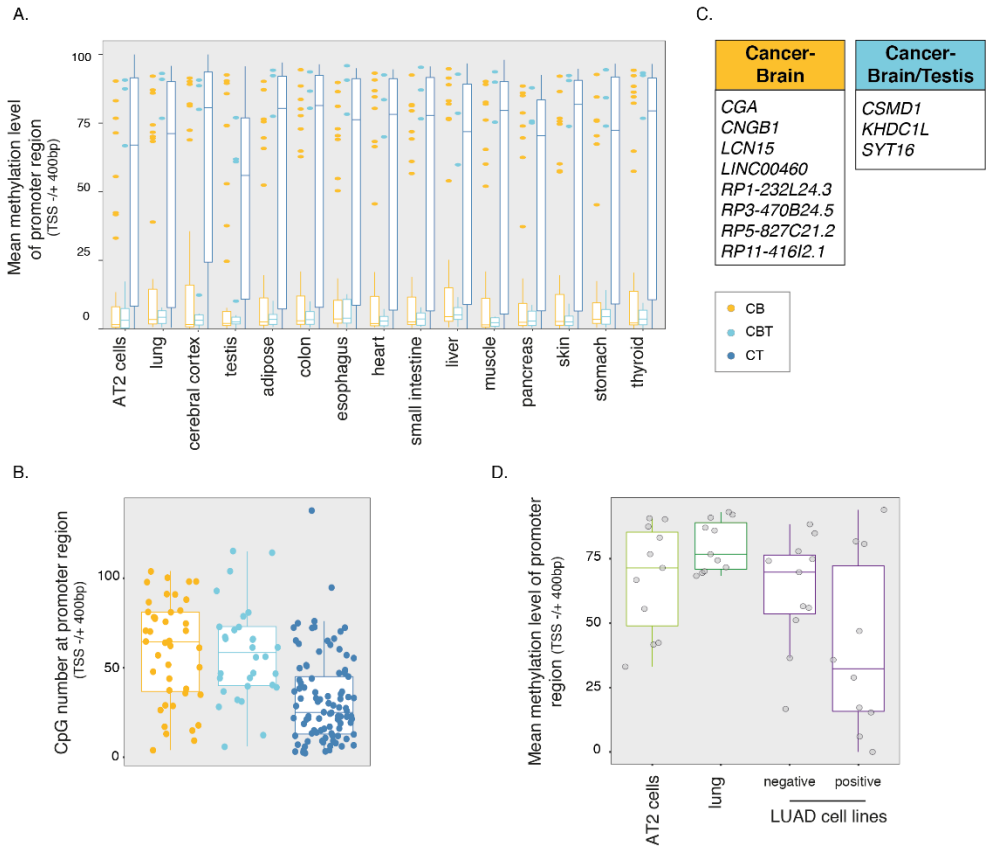


Figure 7: Ectopic activation of a subset of CB and CBT genes is correlated with promoter DNA demethylation in LUAD cells. *A) WGBS methylation data of a panel of normal tissues and AT2 cells*

were explored to determine the initial level of promoter DNA methylation of BT and CBT genes. Each dot represents the mean methylation of CpG sites located at TSS -/+ 400bp, color-coded by their tissue-specificity. B) Each dot represents the mean number of CpG sites for a gene located at TSS -/+ 400bp, color-coded by their tissue-specificity. C) List of CB and CBT genes that showed a methylated promoter in AT2 cells (mean methylation of CpG sites located at TSS -/+ 400bp > 30%). *D) Each dot represents the mean methylation of CpG sites at TSS -/+ 400bp for each of CB and CBT gene that displayed an initially methylated promoter in AT2 cells. LUAD cells were grouped into expressing (positive) and non-expressing (negative) subgroups for each of CB and CBT gene.*

7. The REST repressor is involved in the regulation of a subset of CB and CBT genes

To get further insight into the mechanisms leading to the ectopic activation of CB and CBT genes in LUAD cells, we decided to look for transcription factors possibly involved in their regulation. We used i-*cis*Target, an online portal compiling ChIP-seq data and transcription factor binding site motifs for conserved genomic regions. i-*cis*Target analyses showed enriched peaks of ChIP against the REST transcriptional repressor for both CB (n=12) and CBT (n=8) genes in a variety of tumor cell lines (**Fig. 8**). The prediction tool also revealed enrichment of REST binding motifs in CB and CBT genes. The REST repressor is involved in the silencing of neuronal genes in non-neuronal tissues, as well as in undifferentiated neuronal progenitor cells (Schoenherr and Anderson 1995; Qureshi, Gokhan, and Mehler 2010). REST therefore restrains expression of neuronal genes in mature neurons of the nervous system. To evaluate the involvement of REST in controlling the expression of CB and CBT genes, we analyzed RNA-seq data deriving from adult dermal fibroblasts that were exposed to an shRNA directed against REST. For a subset of CB and CBT genes we observed an increase of their expression following *REST* depletion compared to the control condition (**Fig. 8B**). This transcriptional induction was not observed for testis-specific EctopEx genes, which is consistent with a targeted effect of REST on neuronal genes. Overall, our data suggest that the REST repressor is involved in the regulation of a significant fraction (>25%) of CB and CBT genes.

A.

REST enrichment ranking (%)

		HCT116	MCF7	PANC1
CB	*C1QL2*	*0,046*	*0,026*	*NE*
	CALB1	*0,226*	*0,194*	*0,111*
	CNTNAP2	*0,418*	*0,184*	*0,257*
	FSTL5	*0,352*	*0,007*	*0,299*
	GABRG2	*0,101*	*0,261*	*NE*
	HCN1	*0,198*	*0,090*	*0,146*
	KCNK12	*0,233*	*0,158*	*0,121*
	LHX1	*NE*	*0,742*	*0,462*
	LHX5	*NE*	*0,032*	*0,011*
	POU3F2	*1,138*	*NE*	*1,276*
	SLC18A3	*0,023*	*0,089*	*0,039*
	SOX11	*2,373*	*NE*	*NE*
CBT	*CACNA1B*	*0,064*	*0,054*	*0,023*
	ELAVL2	*0,337*	*0,210*	*0,275*
	GPR158	*0,065*	*0,025*	*0,009*
	GRIN2A	*NE*	*NE*	*0,378*
	KCNQ2	*0,259*	*0,001*	*0,316*
	LHFPL4	*0,130*	*0,056*	*0,236*
	RTBDN	*0,018*	*0,173*	*0,073*
	SYT5	*0,147*	*0,197*	*0,124*

B.

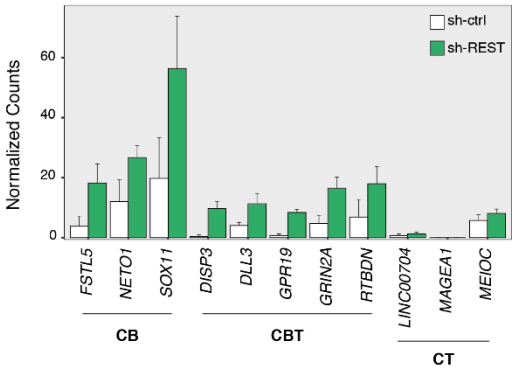


Figure 8: REST chromatin repressor complex is involved in the regulation of a subset of CB and CBT genes. *A) Enrichment ranking of CB and CBT genes in HCT116 (colon adenocarcinoma), MCF7 (breast carcinoma) and PANC1 (pancreatic adenocarcinoma) cell lines after ChIP against REST (NE: no enrichment found). B) Normalized counts of RNA-seq of adult dermal fibroblasts transfected with an shRNA directed against REST or sh-ctrl. CB and CBT genes depicted were found significantly induced in sh-REST compared to sh-ctrl condition (p-value < 0.05). Three CT are shown as control genes (i.e. loss of REST is not expected to induce them).*

4. Discussion and perspectives

In order to thoroughly investigate the extent of ectopic activation of tissue-specific genes in LUAD, we integrated RNA-seq data of LUAD tissue samples and cell lines, as well as their normal lung counterpart and cell of origin, AT2 cells. By applying stringent selection criteria, with no bias towards any physiological expression site or regulatory mechanism, we found that 1051 genes, which are originally silent in normal lung and AT2 cells, show ectopic activation in a significant fraction (≥ 5%) of LUAD samples. This number of so-called EctopEx genes corresponds to 4.5% of the total number of transcripts expressed in LUAD tissues.

Several tissue-specific gene clusters were identified among the EctopEx genes, including the known group of germline-specific genes, as well as gene clusters specific to the gastro-intestinal tract and stratified squamous epithelia, which have been recently described (Diacofotaki, Loriot, and De Smet 2022). Unexpectedly, our data also revealed that, besides these known tissue-specific expression clusters, LUAD cells are enriched for genes that are normally expressed in brain tissues. Our data show that these "brain genes" are frequently activated in LUAD in a similar proportion as testis-specific genes.

Interestingly, a subset of these brain genes also displays expression in the testis. We therefore defined two gene clusters: the "cancer-brain" (CB) genes and the "cancer-brain/testis" (CBT) genes. The existence of genes with a cancer-brain/testis expression profile had already been described by Hoffman et al. in their effort to provide a genome-wide expression classification of cancer-germline genes (Hofmann et al. 2008). We hereby confirm the existence of this group using RNA-seq data of normal and LUAD tissues and extend the gene members of this category. In fact, we noticed no overlap between the CBT genes identified by our selection and the genes listed by Hoffman and colleagues. Indeed, a subset of these genes (*PASD1, MAGEC2*) was identified as being expressed only in the testis by our selection criteria, and not displaying expression in the brain. The rest of the genes identified by Hoffman et al. belongs to gene families that share high sequence identity, such as the *GAGE* family. In RNA-seq data processing, reads originating from such genes are undistinguishable and can therefore not be precisely mapped to a particular genomic location. They are therefore eliminated from the analysis, which is a potential reason of their absence among our list of genes.

We found no evidence of concurrent activation of CB and CBT genes in a subset of LUAD samples suggesting that ectopic de-repression of these genes may result from multiple mechanisms of transcriptional deregulation. We found that activation of a small subset of CB and CBT genes coincides with that of *ASCL1,* a gene involved in the process of neuroendocrine transdifferentiation, which is known to occur in several LUAD lesions. Another subgroup of CB and CBT genes appeared to be regulated by DNA methylation as their activation in LUAD correlated with promoter DNA demethylation. Further experimentation is required to determine if DNA methylation has a primary role in the regulation of these genes, for instance by evaluating the effect of DNA methylation inhibitors on their

expression. DNA demethylation did not appear to underlie the activation of the other CB and CBT genes, since their promoter region was already unmethylated in normal AT2 and lung cells. A search for possible transcription factors involved in the regulation of CB and CBT genes identified REST as a likely candidate, both on the basis of a prediction tool and because depletion of REST correlated with their transcriptional induction. REST is a well-known transcription factor that targets a chromatin repressor complex onto neuronal genes in non-neuronal tissues. Importantly, dysfunction of REST-associated regulatory mechanisms has been reported in several types of tumors and associated with increased malignancy (Coulson 2005). The process leading to impaired REST repression in tumors is unclear, as it may involve alterations of not only the REST protein itself, but also of REST-associated co-repressors, including histone modifying enzymes (Coulson 2005; Dietrich et al. 2012; Huang and Bao 2012). CB and CBT genes may therefore represent valuable markers of dysfunctional REST regulation in tumors. It remains, however, to be elucidated how precisely REST regulates CB and CBT genes. Together, our analyses concerning the regulatory mechanisms involved in the activation of CB and CBT genes in LUAD tumors, point to multiple modes of transcriptional deregulation (**Fig. 9**). The mechanism of de-repression remains completely unexplained for a number of CB and CBT genes. Future investigation is required to address this issue.

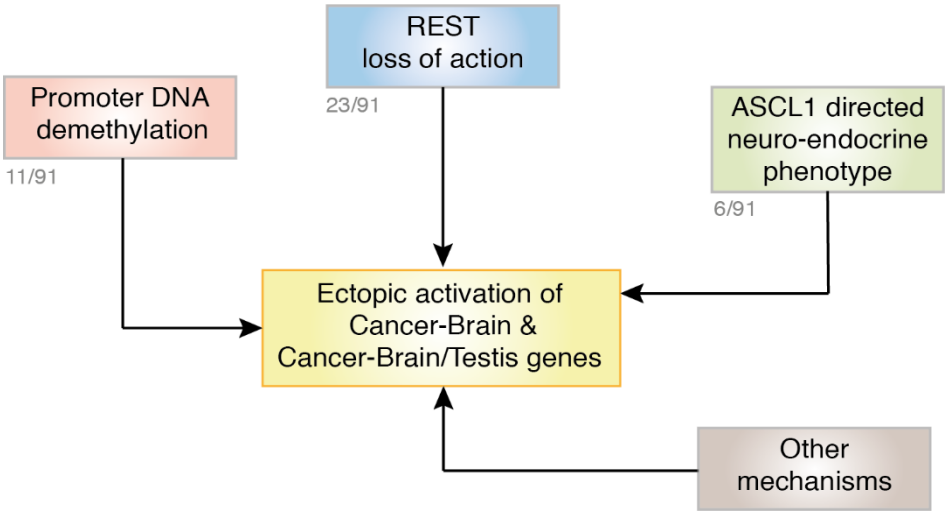


Figure 9: Summary of mechanisms involved in the spurious activation of CB and CBT genes in LUAD.

The abundance of brain genes that become ectopically activated in LUAD and the multiplicity of mechanisms contributing to their de-repression suggest that the activation of these genes might be positively selected in tumors, possibly because they exert tumor-promoting functions. A quick survey of the CancerMine literature-mined database identified several CB and CBT genes with potential oncogenic functions. *POU3F2,* for instance, encodes a transcription factor that acts as an initiation event for the transdifferentiation of prostate tumor cells into a lethal neuroendocrine phenotype in a mouse model (Bishop et al. 2017). *KCNK9* was reported to be amplified and overexpressed in breast and lung tumors. The gene encodes a potassium channel that regulates neuronal cell excitability, and its overexpression in cancer cells was shown to promote resistance to hypoxia and serum deprivation (Mu et al. 2003). *GPR19* encodes a G-protein-coupled receptor with a distinct role in circadian rhythm regulation orchestrated by the suprachiasmatic nucleus in the brain (Yamaguchi et al. 2021). In lung cancer cells, GPR19 was shown to promote cell proliferation by accelerating the G2-M transition of the cell cycle (Kastner et al. 2012).

Emerging evidence supports that tumor cells can interact with the peripheral nervous system, and that this promotes their growth and progression. Indeed, using a mouse model of prostate tumor, Magnon and colleagues have shown that the infiltration of adrenergic fibers into tumor tissues increases their growth rate, while cholinergic innervation favors tumor dissemination (Magnon et al. 2013). In other mouse tumor models, Zhao and colleagues demonstrated that surgical or pharmacological denervation of the stomach markedly reduced gastric cancer incidence and progression, possibly through inhibition of Wnt signaling and suppression of stem cell expansion (Zhao et al. 2014). We hypothesize that the expression of brain-related genes by tumor cells may help them to respond to neuronal-related signals.

Brain metastases originate from primary tumors of various histological types, lung cancer being the most frequent source followed by breast cancer (Khan et al. 2020). Evidence indicates that the propensity of a tumor to form a metastasis in a distant site depends on its ability to interact with cells and signaling molecules of the host organ (Gao et al. 2019). For instance, metastatic cells in the brain sometimes upregulate receptors and transporters of the γ-aminobutyric acid neurotransmitter, which they can use as an energy source for protein synthesis

(Neman et al. 2014). *GABRA3,* which encodes a GABA receptor subunit, is among the CBT genes we identified. Another study demonstrated that, by activating the expression of brain-specific NMDAR receptors, metastatic cells interact with neurons through the glutamate-stimulated signaling, increasing thereby their proliferation capacity (Zeng et al. 2019). Among the CBT genes, *GRIN2A* encodes an NMDAR receptor. It is therefore tempting to propose that the activation of CB and CBT genes in LUAD contributes to the capacity to form brain metastases. Evidence for this, could be provided by investigating if LUAD tumors that metastasized to the brain, express higher frequencies of activation of CB and CBT genes, when compared to LUAD tumors that formed metastases in other organs.

References

Axelsen, J. B., J. Lotem, L. Sachs, and E. Domany. 2007. 'Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles', *Proc Natl Acad Sci U S A*, 104: 13122-7.

Bhattacharjee, A., W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson. 2001. 'Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses', *Proc Natl Acad Sci U S A*, 98: 13790-5.

Bird, A. P. 1986. 'CpG-rich islands and the function of DNA methylation', *Nature*, 321: 209-13.

Bishop, J. L., D. Thaper, S. Vahid, A. Davies, K. Ketola, H. Kuruma, R. Jama, K. M. Nip, A. Angeles, F. Johnson, A. W. Wyatt, L. Fazli, M. E. Gleave, D. Lin, M. A. Rubin, C. C. Collins, Y. Wang, H. Beltran, and A. Zoubeidi. 2017. 'The Master Neural Transcription Factor BRN2 Is an Androgen Receptor-Suppressed Driver of Neuroendocrine Differentiation in Prostate Cancer', *Cancer Discov*, 7: 54-71.

Blanpain, C. 2013. 'Tracing the cellular origin of cancer', *Nat Cell Biol*, 15: 126-34.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30: 2114-20.

Borges, M., R. I. Linnoila, H. J. van de Velde, H. Chen, B. D. Nelkin, M. Mabry, S. B. Baylin, and D. W. Ball. 1997. 'An achaete-scute homologue essential for neuroendocrine differentiation in the lung', *Nature*, 386: 852-5.

Borromeo, M. D., T. K. Savage, R. K. Kollipara, M. He, A. Augustyn, J. K. Osborne, L. Girard, J. D. Minna, A. F. Gazdar, M. H. Cobb, and J. E. Johnson. 2016. 'ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs', *Cell Rep*, 16: 1259-72.

Cancer Genome Atlas Research, Network. 2014. 'Comprehensive molecular profiling of lung adenocarcinoma', *Nature*, 511: 543-50.

Casarosa, S., C. Fode, and F. Guillemot. 1999. 'Mash1 regulates neurogenesis in the ventral telencephalon', *Development*, 126: 525-34.

Chan, B. A., and B. G. Hughes. 2015. 'Targeted therapy for non-small cell lung cancer: current standards and the promise of the future', *Transl Lung Cancer Res*, 4: 36-54.

Colaprico, A., T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, and H. Noushmehr. 2016. 'TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data', *Nucleic Acids Res*, 44: e71.

Consortium, G. TEx. 2013. 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet*, 45: 580-5.

Coulson, J. M. 2005. 'Transcriptional regulation: cancer, neurons and the REST', *Curr Biol*, 15: R665-8.

De Smet, C., and A. Loriot. 2013. 'DNA hypomethylation and activation of germline-specific genes in cancer', *Adv Exp Med Biol*, 754: 149-66.

Diacofotaki, A., A. Loriot, and C. De Smet. 2022. 'Identification of Tissue-Specific Gene Clusters Induced by DNA Demethylation in Lung Adenocarcinoma: More Than Germline Genes', *Cancers (Basel)*, 14.

Dietrich, N., M. Lerdrup, E. Landt, S. Agrawal-Singh, M. Bak, N. Tommerup, J. Rappsilber, E. Sodersten, and K. Hansen. 2012. 'REST-mediated recruitment of polycomb repressor complexes in mammalian cells', *PLoS Genet*, 8: e1002494.

Drouin-Ouellet, J., S. Lau, P. L. Brattas, D. Rylander Ottosson, K. Pircs, D. A. Grassi, L. M. Collins, R. Vuono, A. Andersson Sjoland, G. Westergren-Thorsson, C. Graff, L. Minthon, H. Toresson, R. A. Barker, J. Jakobsson, and M. Parmar. 2017. 'REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways', *EMBO Mol Med*, 9: 1117-31.

Ehrlich, M. 2009. 'DNA hypomethylation in cancer cells', *Epigenomics*, 1: 239-59.

Fain, J. S., A. Loriot, A. Diacofotaki, A. Van Tongelen, and C. De Smet. 2021. 'Transcriptional overlap links DNA hypomethylation with DNA hypermethylation at adjacent promoters in cancer', *Sci Rep*, 11: 17346.

Feinberg, A. P., and B. Vogelstein. 1983. 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, 301: 89-92.

Ferone, G., M. C. Lee, J. Sage, and A. Berns. 2020. 'Cells of origin of lung cancers: lessons from mouse studies', *Genes Dev*, 34: 1017-32.

Flavahan, W. A., E. Gaskell, and B. E. Bernstein. 2017. 'Epigenetic plasticity and the hallmarks of cancer', *Science*, 357.

Gama-Sosa, M. A., V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke, and M. Ehrlich. 1983. 'The 5-methylcytosine content of DNA from human tumors', *Nucleic Acids Res*, 11: 6883-94.

Gao, Y., I. Bado, H. Wang, W. Zhang, J. M. Rosen, and X. H. Zhang. 2019. 'Metastasis Organotropism: Redefining the Congenial Soil', *Dev Cell*, 49: 375-91.

Gu, Z., R. Eils, and M. Schlesner. 2016. 'Complex heatmaps reveal patterns and correlations in multidimensional genomic data', *Bioinformatics*, 32: 2847-9.

Hanahan, D. 2022. 'Hallmarks of Cancer: New Dimensions', *Cancer Discov*, 12: 31-46.

Hofmann, O., O. L. Caballero, B. J. Stevenson, Y. T. Chen, T. Cohen, R. Chua, C. A. Maher, S. Panji, U. Schaefer, A. Kruger, M. Lehvaslaiho, P. Carninci, Y. Hayashizaki, C. V. Jongeneel, A. J. Simpson, L. J. Old, and W. Hide. 2008. 'Genome-wide analysis of cancer/testis gene expression', *Proc Natl Acad Sci U S A*, 105: 20422-7.

Huang, Z., and S. Bao. 2012. 'Ubiquitination and deubiquitination of REST and its roles in cancers', *FEBS Lett*, 586: 1602-5.

Kastner, S., T. Voss, S. Keuerleber, C. Glockel, M. Freissmuth, and W. Sommergruber. 2012. 'Expression of G protein-coupled receptor 19 in human lung cancer cells is triggered by entry into S-phase and supports G(2)-M cell-cycle progression', *Mol Cancer Res*, 10: 1343-58.

Khan, M., S. Arooj, R. Li, Y. Tian, J. Zhang, J. Lin, Y. Liang, A. Xu, R. Zheng, M. Liu, and Y. Yuan. 2020. 'Tumor Primary Site and Histology Subtypes Role in Radiotherapeutic Management of Brain Metastases', *Front Oncol*, 10: 781.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. 'Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype', *Nat Biotechnol*, 37: 907-15.

Kosari, F., C. M. Ida, M. C. Aubry, L. Yang, I. V. Kovtun, J. L. Klein, Y. Li, S. Erdogan, S. C. Tomaszek, S. J. Murphy, L. C. Bolette, C. P. Kolbert, P. Yang, D. A. Wigle, and G. Vasmatzis. 2014. 'ASCL1 and RET expression defines a clinically relevant subgroup of lung adenocarcinoma characterized by neuroendocrine differentiation', *Oncogene*, 33: 3776-83.

Krueger, F., and S. R. Andrews. 2011. 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*, 27: 1571-2.

Liao, Y., G. K. Smyth, and W. Shi. 2014. 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30: 923-30.

Loriot, A., A. Van Tongelen, J. Blanco, S. Klaessens, J. Cannuyer, N. van Baren, A. Decottignies, and C. De Smet. 2014. 'A novel cancer-germline transcript carrying pro-metastatic miR-105 and TET-targeting miR-767 induced by DNA hypomethylation in tumors', *Epigenetics*, 9: 1163-71.

Lotem, J., D. Netanely, E. Domany, and L. Sachs. 2005. 'Human cancers overexpress genes that are specific to a variety of normal human tissues', *Proc Natl Acad Sci U S A*, 102: 18556-61.

Magnon, C., S. J. Hall, J. Lin, X. Xue, L. Gerber, S. J. Freedland, and P. S. Frenette. 2013. 'Autonomic nerve development contributes to prostate cancer progression', *Science*, 341: 1236361.

Mu, D., L. Chen, X. Zhang, L. H. See, C. M. Koch, C. Yen, J. J. Tong, L. Spiegel, K. C. Nguyen, A. Servoss, Y. Peng, L. Pei, J. R. Marks, S. Lowe, T. Hoey, L. Y. Jan, W. R. McCombie, M. H. Wigler, and S. Powers. 2003. 'Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene', *Cancer Cell*, 3: 297-302.

Neman, J., J. Termini, S. Wilczynski, N. Vaidehi, C. Choy, C. M. Kowolik, H. Li, A. C. Hambrecht, E. Roberts, and R. Jandial. 2014. 'Human breast cancer metastases

to the brain display GABAergic properties in the neural niche', *Proc Natl Acad Sci U S A*, 111: 984-9.

Qureshi, I. A., S. Gokhan, and M. F. Mehler. 2010. 'REST and CoREST are transcriptional and epigenetic regulators of seminal neural fate decisions', *Cell Cycle*, 9: 4477-86.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. 'Integrative genomics viewer', *Nat Biotechnol*, 29: 24-6.

Rousseaux, S., A. Debernardi, B. Jacquiau, A. L. Vitte, A. Vesin, H. Nagy-Mignotte, D. Moro-Sibilot, P. Y. Brichon, S. Lantuejoul, P. Hainaut, J. Laffaire, A. de Reynies, D. G. Beer, J. F. Timsit, C. Brambilla, E. Brambilla, and S. Khochbin. 2013. 'Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers', *Sci Transl Med*, 5: 186ra66.

Roy, N., and M. Hebrok. 2015. 'Regulation of Cellular Identity in Cancer', *Dev Cell*, 35: 674-84.

Schoenherr, C. J., and D. J. Anderson. 1995. 'The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes', *Science*, 267: 1360-3.

Shukla, S., J. Cyrta, D. A. Murphy, E. G. Walczak, L. Ran, P. Agrawal, Y. Xie, Y. Chen, S. Wang, Y. Zhan, D. Li, E. W. P. Wong, A. Sboner, H. Beltran, J. M. Mosquera, J. Sher, Z. Cao, J. Wongvipat, R. P. Koche, A. Gopalan, D. Zheng, M. A. Rubin, H. I. Scher, P. Chi, and Y. Chen. 2017. 'Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance', *Cancer Cell*, 32: 792-806 e7.

Somerville, T. D. D., Y. Xu, K. Miyabayashi, H. Tiriac, C. R. Cleary, D. Maia-Silva, J. P. Milazzo, D. A. Tuveson, and C. R. Vakoc. 2018. 'TP63-Mediated Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma', *Cell Rep*, 25: 1741-55 e7.

Suzuki, A., H. Makinoshima, H. Wakaguri, H. Esumi, S. Sugano, T. Kohno, K. Tsuchihara, and Y. Suzuki. 2014. 'Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines', *Nucleic Acids Res*, 42: 13557-72.

Uhlen, M., E. Bjorling, C. Agaton, C. A. Szigyarto, B. Amini, E. Andersen, A. C. Andersson, P. Angelidou, A. Asplund, C. Asplund, L. Berglund, K. Bergstrom, H. Brumer, D. Cerjan, M. Ekstrom, A. Elobeid, C. Eriksson, L. Fagerberg, R. Falk, J. Fall, M. Forsberg, M. G. Bjorklund, K. Gumbel, A. Halimi, I. Hallin, C. Hamsten, M. Hansson, M. Hedhammar, G. Hercules, C. Kampf, K. Larsson, M. Lindskog, W. Lodewyckx, J. Lund, J. Lundeberg, K. Magnusson, E. Malm, P. Nilsson, J. Odling, P. Oksvold, I. Olsson, E. Oster, J. Ottosson, L. Paavilainen, A. Persson, R. Rimini, J. Rockberg, M. Runeson, A. Sivertsson, A. Skollermo, J. Steen, M. Stenvall, F. Sterky, S. Stromberg, M. Sundberg, H. Tegel, S. Tourle, E. Wahlund, A. Walden, J. Wan, H. Wernerus, J. Westberg, K. Wester, U. Wrethagen, L. L. Xu, S. Hober, and F. Ponten. 2005. 'A human protein atlas for normal and cancer tissues based on antibody proteomics', *Mol Cell Proteomics*, 4: 1920-32.

Uhlen, M., L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten. 2015. 'Proteomics. Tissue-based map of the human proteome', *Science*, 347: 1260419.

Van Tongelen, A., A. Loriot, and C. De Smet. 2017. 'Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes', *Cancer Lett*, 396: 130-37.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan, and D. Schubeler. 2007. 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nat Genet*, 39: 457-66.

Weeden, C. E., Y. Chen, S. B. Ma, Y. Hu, G. Ramm, K. D. Sutherland, G. K. Smyth, and M. L. Asselin-Labat. 2017. 'Lung Basal Stem Cells Rapidly Repair DNA Damage Using the Error-Prone Nonhomologous End-

Joining Pathway', *PLoS Biol*, 15: e2000731.

Yamaguchi, Y., I. Murai, K. Goto, S. Doi, H. Zhou, G. Setsu, H. Shimatani, H. Okamura, T. Miyake, and M. Doi. 2021. 'Gpr19 is a circadian clock-controlled orphan GPCR with a role in modulating free-running period and light resetting capacity of the circadian clock', *Sci Rep*, 11: 22406.

You, J. S., and P. A. Jones. 2012. 'Cancer genetics and epigenetics: two sides of the same coin?', *Cancer Cell*, 22: 9-20.

Yu, G., L. G. Wang, Y. Han, and Q. Y. He. 2012. 'clusterProfiler: an R package for comparing biological themes among gene clusters', *OMICS*, 16: 284-7.

Zeng, Q., I. P. Michael, P. Zhang, S. Saghafinia, G. Knott, W. Jiao, B. D. McCabe, J. A. Galvan, H. P. C. Robinson, I. Zlobec, G. Ciriello, and D. Hanahan. 2019. 'Synaptic proximity enables NMDAR signalling to promote brain metastasis', *Nature*, 573: 526-31.

Zhao, C. M., Y. Hayakawa, Y. Kodama, S. Muthupalani, C. B. Westphalen, G. T. Andersen, A. Flatberg, H. Johannessen, R. A. Friedman, B. W. Renz, A. K. Sandvik, V. Beisvag, H. Tomita, A. Hara, M. Quante, Z. Li, M. D. Gershon, K. Kaneko, J. G. Fox, T. C. Wang, and D. Chen. 2014. 'Denervation suppresses gastric tumorigenesis', *Sci Transl Med*, 6: 250ra115.

Zuber, V., C. N. Marconett, J. Shi, X. Hua, W. Wheeler, C. Yang, L. Song, A. M. Dale, M. Laplana, A. Risch, A. Witoelar, W. K. Thompson, A. J. Schork, F. Bettella, Y. Wang, S. Djurovic, B. Zhou, Z. Borok, H. F. van der Heijden, J. de Graaf, D. Swinkels, K. K. Aben, J. McKay, R. J. Hung, H. Bikeboller, V. L. Stevens, D. Albanes, N. E. Caporaso, Y. Han, Y. Wei, M. A. Panadero, J. I. Mayordomo, D. C. Christiani, L. Kiemeney, O. A. Andreassen, R. Houlston, C. I. Amos, N. Chatterjee, I. A. Laird-Offringa, I. G. Mills, and M. T. Landi. 2016. 'Pleiotropic Analysis of Lung Cancer and Blood Triglycerides', *J Natl Cancer Inst*, 108.

Supplementary tables

Table S1: Transcriptomic and methylomic studies of normal tissues, AT2 cells and normal adult fibroblasts C2 cells

Sample	Type	Pathology	Accession	File	Assay	Genome built	Method	Layout	Platform	Reference
large airway AT2	primary cells	normal	SRR3680940	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Weeden et al. 2017
large airway AT2	primary cells	normal	SRR3680941	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Weeden et al. 2017
large airway AT2	primary cells	normal	SRR3680942	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Weeden et al. 2017
small airaway AT2	primary cells	normal	SRR3680946	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Weeden et al. 2017
small airaway AT2	primary cells	normal	SRR3680947	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Weeden et al. 2017
small airaway AT2	primary cells	normal	SRR3680948	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Weeden et al. 2017
cerebral cortex	tissue	normal	ERR315455, ERR315432 ERR315477 ERR315352, ERR315415	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
testis	tissue	normal	ERR315492	FastQ	expression	not applicable	bulk RNA-seq	paired	Illumina HiSeq 2000	Fagerberg et al. 2014
AT2	primary cells	normal	SRR1773103	FastQ	CpG methylation	not applicable	whole genome bisulfite seq	paired	Illumina HiSeq 2000	Zuber et al. 2016
adipose	tissue	normal	ENCFF318AMC	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
esophagus	tissue	normal	ENCFF625GVK	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
heart (left ventricule)	tissue	normal	ENCFF536RSX	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
liver	tissue	normal	ENCFF356KGQ	bam	CpG methylation	not applicable	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
lung	tissue	normal	ENCFF039IFT	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
pancreas	tissue	normal	ENCFF763RUE	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
pre-frontal cortex	tissue	normal	SRR3278486_SRR3278482	FastQ	CpG methylation	not applicable	Whole genome bisulfite-seq	paired	Illumina HiSeq 2500	Jenkinson et al. 2017
sigmoid colon	tissue	normal	ENCFF157POM	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
skin (lower leg)	tissue	normal	ENCFF219CCQ	bed	CpG methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq X Ten	ENCODE Project Consortium, 2012
small intestine	tissue	normal	ENCFF241AQC	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
stomach	tissue	normal	ENCFF497YOO	bed	CpG methylation	hg38	Whole genome bisulfite-seq	single	Illumina HiSeq 2500	Roadmap Epigenomics, 2015
testis	tissue	normal	ENCFF715DMX	bed	CpG methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq X Ten	ENCODE Project Consortium, 2012
thyroid	tissue	normal	ENCFF223LJW	bed	CpG methylation	hg38	Whole genome bisulfite-seq	paired	Illumina HiSeq X Ten	ENCODE Project Consortium, 2012
C2-shctrl	adult dermal fibroblasts	normal	SRR5036502	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Drouin-Ouellet et al. 2017
C2-shctrl	adult dermal fibroblasts	normal	SRR5036503	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Drouin-Ouellet et al. 2017
C2-shctrl	adult dermal fibroblasts	normal	SRR5036504	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Drouin-Ouellet et al. 2017
C2-shREST	adult dermal fibroblasts	normal	SRR5036511	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Drouin-Ouellet et al. 2017
C2-shREST	adult dermal fibroblasts	normal	SRR5036512	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Drouin-Ouellet et al. 2017
C2-shREST	adult dermal fibroblasts	normal	SRR5036513	FastQ	expression	not applicable	bulk RNA-seq	single	Illumina HiSeq 2000	Drouin-Ouellet et al. 2017

Table S2: Methylomic and transcriptomic datasets of LUAD cell lines

Sample	Type	Pathology	Accession	File	Assay	BioProject	Method	Layout	Platform	Reference
A427	cell line	lung adenocarcinoma	DRR016652	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
A549	cell line	lung adenocarcinoma	DRR016653	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
ABC1	cell line	lung adenocarcinoma	DRR016654	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1299	cell line	lung adenocarcinoma	DRR016655	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1437	cell line	lung adenocarcinoma	DRR016656	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1648	cell line	lung adenocarcinoma	DRR016657	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1650	cell line	lung adenocarcinoma	DRR016658	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1703	cell line	lung adenocarcinoma	DRR016659	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1819	cell line	lung adenocarcinoma	DRR016660	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1975	cell line	lung adenocarcinoma	DRR016661	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2126	cell line	lung adenocarcinoma	DRR016662	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2228	cell line	lung adenocarcinoma	DRR016663	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2347	cell line	lung adenocarcinoma	DRR016664	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H322	cell line	lung adenocarcinoma	DRR016665	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
II18	cell line	lung adenocarcinoma	DRR016666	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
LC2ad	cell line	lung adenocarcinoma	DRR016667	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC14	cell line	lung adenocarcinoma	DRR016668	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC3	cell line	lung adenocarcinoma	DRR016669	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC7	cell line	lung adenocarcinoma	DRR016670	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC9	cell line	lung adenocarcinoma	DRR016671	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad1	cell line	lung adenocarcinoma	DRR016672	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad2	cell line	lung adenocarcinoma	DRR016673	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_KJ	cell line	lung adenocarcinoma	DRR016674	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_MS	cell line	lung adenocarcinoma	DRR016675	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_OK	cell line	lung adenocarcinoma	DRR016676	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
VMRC_LCD	cell line	lung adenocarcinoma	DRR016677	FastQ	methylation	PRJDB1928	Bisulfite-sequencing	paired	Illumina HiSeq 2500	Suzuki et al. 2014
A427	cell line	lung adenocarcinoma	DRR016694	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
A549	cell line	lung adenocarcinoma	DRR016695	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
ABC1	cell line	lung adenocarcinoma	DRR016696	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1299	cell line	lung adenocarcinoma	DRR016697	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1437	cell line	lung adenocarcinoma	DRR016698	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1648	cell line	lung adenocarcinoma	DRR016699	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1650	cell line	lung adenocarcinoma	DRR016700	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1703	cell line	lung adenocarcinoma	DRR016701	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1819	cell line	lung adenocarcinoma	DRR016702	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H1975	cell line	lung adenocarcinoma	DRR016703	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2126	cell line	lung adenocarcinoma	DRR016704	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2228	cell line	lung adenocarcinoma	DRR016705	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H2347	cell line	lung adenocarcinoma	DRR016706	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
H322	cell line	lung adenocarcinoma	DRR016707	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
II18	cell line	lung adenocarcinoma	DRR016708	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
LC2ad	cell line	lung adenocarcinoma	DRR016709	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC14	cell line	lung adenocarcinoma	DRR016710	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC3	cell line	lung adenocarcinoma	DRR016711	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC7	cell line	lung adenocarcinoma	DRR016712	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
PC9	cell line	lung adenocarcinoma	DRR016713	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad1	cell line	lung adenocarcinoma	DRR016714	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_Ad2	cell line	lung adenocarcinoma	DRR016715	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_KJ	cell line	lung adenocarcinoma	DRR016716	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_MS	cell line	lung adenocarcinoma	DRR016717	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
RERF_LC_OK	cell line	lung adenocarcinoma	DRR016718	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014
VMRC_LCD	cell line	lung adenocarcinoma	DRR016719	FastQ	expression	PRJDB2256	bulk RNA-seq	paired	Illumina HiSeq 2500	Suzuki et al. 2014

GENERAL DISCUSSION

Gene expression profiles of tumor cells are heavily altered compared to their normal tissue counterpart. Today, it is clearly established that these alterations are due to genetic, as well as epigenetic aberrations occurring in these cells. One of the consequences of these aberrations in tumors is the spurious activation of tissue-specific genes, initially silent in the cell of origin of the tumor. The ectopic activation of genes in tumor cells presents new opportunities to establish markers of the disease. Additionally, their potential contribution to tumor progression could lead to new therapeutic strategies. One such tissue-specific gene cluster activated in tumors is the group of cancer-germline (CG) genes found ectopically expressed in a wide variety of non-germline tumors. The group of CG genes is transcriptionally induced following their promoter DNA demethylation in tumors. This promoter DNA demethylation event is associated with the phenomenon of global DNA hypomethylation occurring in these cells.

The group of CG genes is the only tissue-specific expression program activated in tumors. This is partly due to the fact that studies aiming to identify misexpressed genes in tumors frequently hold a bias towards the physiological expression site of the candidate genes, i.e., their restrained expression in testis. Another reason could be that studies often focus on genes that are extensively up- or down-regulated in tumors compared to the tissue from which the tumor originated. This implies that the genes are initially expressed, at least at some level, in the cell of origin of the tumor, excluding therefore, a spurious activation in tumors. Overall, a clear answer to whether somatic-specific gene clusters are found ectopically activated in tumors has remained unknown so far.

In the present work, **we aimed to reassess the extent of ectopic tissue-specific gene activation in tumors,** and in particular in lung adenocarcinoma (LUAD), **without imposing any initial tissue restriction on their physiological expression sites, or on their transcriptional regulatory mechanisms**.

In this general discussion chapter, we start by summarizing the key findings obtained in chapters 1 and 2. We continue by describing how these findings are integrated in the current state of the art in regard to the role of DNA methylation in regulating tissue-specific genes, followed by the possibility that the ectopic activation event of tissue-specific gene clusters is not restricted to LUAD. Finally, we discuss about the way these genes could contribute to tumor progression.

1. *LUAD cells show ectopic activation of somatic-specific gene clusters*

By comparing the transcriptome of LUAD tissues with the transcriptome of their normal counterpart (lung tissue and AT2, cells of origin of LUAD), we found that, besides the known gametogenic gene cluster, LUAD show aberrant activation of three somatic-specific group of genes (**Fig. 22**).

In chapter 1, we found that the global loss of DNA methylation occurring in LUAD was associated with promoter DNA demethylation and, therefore, transcriptional induction of the following two gene clusters: one displaying preferential expression in normal colon and small intestine (we termed this gene cluster, gastro-intestinal cluster or GI cluster), and the other exhibiting a preferential expression in stratified squamous epithelia (esophagus, skin and vagina; termed stratified squamous epithelia cluster or SE cluster). Nevertheless, this regulation dependence on promoter DNA methylation was less pronounced for the GI cluster, which appeared to also require ectopic activation of gastro-intestinal specific transcription factors, such as HNF4A, to ensure enhanced transcriptional activation.

In chapter 2, we observed that LUAD are enriched for a large group of genes that show physiological expression sites in a wide variety of brain tissues. Our results indicate that a significant subset of these genes is activated following the deregulation of REST neuronal repressor complex in tumors. REST deregulation is shown to be a common event in tumor cells, associated with a tumor suppressive function (Coulson 2005; Majumder 2006). Other mechanisms for the transcriptional induction of small subsets of brain genes appeared to involve promoter DNA demethylation, and an association with a neuroendocrine transdifferentiation phenotype of LUAD.

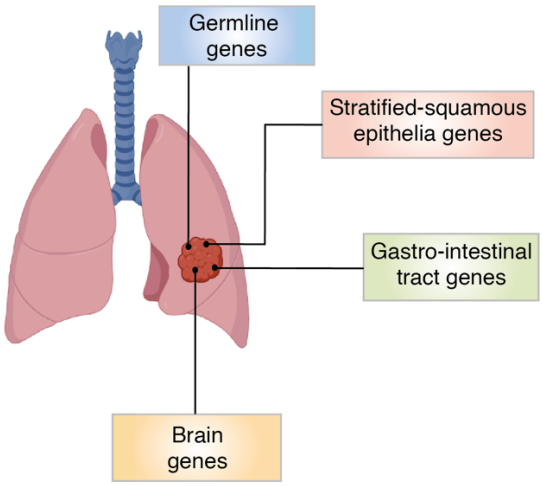


Figure 22: Summary of the ectopic activation of tissue-specific gene clusters occurring in LUAD.
Lung image retrieved from Biorender.com (2022).

2. *A role of promoter DNA methylation in the silencing of somatic specific genes*

It is widely accepted today that methylation has an essential role in chrX inactivation, monoallelic expression of imprinted genes, and stable silencing of repetitive genomic elements. Nevertheless, it is still highly debatable whether DNA methylation is also involved in the regulation of tissue-specific genes in somatic cells (Wilkinson 2015; Martienssen 1998; Bestor, Edwards, and Boulard 2015), as it was initially proposed by Riggs, Holliday and Pugh (Riggs 1975; Holliday and Pugh 1975). As stated in the general introduction chapter, evidence supports such regulatory mechanism for a subset of germline-specific genes (De Smet, Lurquin, Lethe, et al. 1999). However, data that extend this observation to somatic genes have, so far, remained scarce.

Arguments in favor of the existence of such regulation consisted first, in an inverse correlation observed between promoter DNA methylation and the activation state of the corresponding genes in expressing cells. Secondly, the fact that genes were

transcriptionally induced after 5-azadC demethylation treatment, provided an additional hint linking gene methylation and gene expression. However, these observations were strongly counterargued by the following facts: firstly, transcription factor binding appears to prevent *de novo* DNA methylation of the nearby genomic sequences (Macleod et al. 1994). Additionnally, some transcription factors were found to be capable of inducing active local DNA demethylation in the sequences encompassing their binding sites (Suzuki et al. 2017). These data therefore suggest that the observed initial anti-correlation would be a consequence, rather than a cause, of transcriptional activation in these cells. Furthermore, it was shown that several genes induced following 5-azadC treatment initially exhibited a demethylated promoter region in non-expressing cells, excluding therefore a role of DNA methylation in their regulation. Such induction could be explained by an indirect regulatory mechanism, i.e., 5-azadC induces a factor resulting in the activation of the gene of interest. Another possibility for the observed correlation between promoter DNA demethylation and gene expression would be that the gene of interest, initially not methylated, acquired *de novo* DNA methylation in non-expressing cells, a phenomenon known to occur to repressed genes in cultured cells (Antequera, Boyes, and Bird 1990), and therefore, would now become activable by 5-azadC. This phenomenon is known as an epigenetic switch, where a gene switches their initial regulation mechanism to another. Finally, in *Dnmt1 -/-* mouse embryos that exhibited severe genome-wide DNA hypomethylation, no tissue-specific gene activation could be observed (Walsh and Bestor 1999).

Overall, these data clearly indicate, that no definite answer has been obtained as to the implication of promoter DNA methylation in the regulation of tissue-specific genes in normal somatic cells, or in tumors. Our study provides evidence that such regulation does exist. We found that GI and SE gene clusters display initially a methylated promoter in non-expressing cells, and a non-methylated promoter in expressing cells, consistent with a possible role of DNA methylation in their regulation. Furthermore, this inverse correlation was extended to expressing and non-expressing LUAD cell lines, as well as LUAD tissues. Experimental DNA demethylation treatment using 5-azadC, also lead to the transcriptional induction of these genes in normal and tumor cell lines. Finally, promoter DNA demethylation of GI and SE gene clusters was associated with a global loss of genomic methylation, with the LUAD tissue samples showing a demethylated promoter region of these

genes, displaying an overall lower content of DNA methylation compared to the LUAD tissues that showed a methylated promoter region of the same genes.

Our data therefore support that DNA methylation is involved in the regulation of a small fraction of genes of the gastrointestinal tract and stratified squamous epithelia. Evidently, the SE and GI gene clusters that we identified in our work as induced by promoter DNA demethylation in LUAD are only a part of a greater transcriptional network of these tissues, regulated by other mechanisms than promoter DNA methylation, and implicating multiple regulatory players. This is in line with previous observations showing that tissue-specific gene activation by a single epigenetic alteration is not enough to achieve transcriptional induction in normal cells (Naciri et al. 2019).

3. *Evidence of ectopic activation of GI, SE and brain-specific genes in a wide variety of tumors*

The ectopic activation of GI, SE and brain-specific gene clusters was identified in LUAD, which we selected as a study model. As mentioned in the general introduction chapter, we chose this lung tumor subtype for the following reasons: firstly, the cell of origin from which LUAD arise has been identified as AT2 cells (Ferone et al. 2020). This is of great importance since studies focusing on ectopic gene activation imply that the genes are initially repressed in the cell of origin of the tumor. Secondly, a wide variety of transcriptomic and methylomic data of AT2 cells, normal lung tissue, LUAD tissues and cell lines are available either through worldwide consortia or public repository platforms. This allows a potent comparison of gene expression between normal and tumor cells and tissues. Finally, this tumor subtype has been described as subjected to extensive DNA methylation alterations, and transcriptional changes compared to their normal counterpart (Cancer Genome Atlas Research 2014). Whether GI, SE and brain-specific gene clusters are also found ectopically expressed in additional tumor types is currently unknown.

The method that we applied for the identification of tissue-specific gene clusters in LUAD could be used for other tumor models, provided that they display the above-

described data availability as in the case of LUAD. This analysis would be of great interest to see whether tumor types from different origins display an ectopic activation of tissue-specific gene clusters, and whether they are of the same tissue-specificity as the ones described above.

Evidence in the literature suggest that, indeed, tumor tissues, other than LUAD, display spurious activation of genes that, in normal cells, are expressed in the gastrointestinal tract, or in stratified squamous epithelia or are involved in neuronal signaling. For instance, it was found that prostate tumor cells exhibited an aberrant activation of a gastrointestinal transcriptional program that was associated with castration resistance (Shukla et al. 2017). Other studies have found that a subtype of pancreatic adenocarcinoma is characterized by a squamous transdifferentiation phenotype. It was shown that this is partially governed by TP63 (Somerville et al. 2018), a transcription factor implicated in the installment of epithelia stratification. Moreover, it was also shown that breast cancer cells that have metastasized to the brain showed ectopic activation of genes related to GABA signaling (Neman et al. 2014). Finally, preliminary data that we have recently generated show that in TCGA tumor cohorts, GI and SE gene clusters are also activated in different tumor types, such as pancreatic and prostate adenocarcinomas, and, to a lesser extent, in brain tumors or hematological malignancies (data not shown).

Taken together, these data provide evidence that the ectopic activation of genes belonging to the three somatic transcriptional programs that we identified is not a phenomenon restricted to LUAD. It goes without saying that additional regulatory mechanisms than the ones we incriminated in our work, are also involved in their ectopic activation in tumor cells. These observations do not address the issue of whether additional tissue-specific genes, other than GI, SE, and brain-specific genes, are also aberrantly expressed in tumors. It would be of interest to investigate whether, depending on the tumor cell of origin, different tissue-specific transcriptional networks could be activated in tumors based on their initial epigenomic background.

4. *Potential pro-tumoral role of the ectopic activation of tissue-specific gene clusters in tumors*

As stated in the general introduction chapter, epigenetic mechanisms stabilize chromatin conformation. This leads, therefore, to a stable gene expression profile throughout cell division and to the maintenance of a stable cellular identity. However, in tumor cells, chromatin states are "unlocked" allowing stochastic activation and repression of genes. This event is described as a cellular "plastic state" which contributes to the acquisition of unscheduled gene expression patterns. These altered expression profiles can participate to tumor progression by allowing an increased adaptability of the cell to any environmental cue (Flavahan, Gaskell, and Bernstein 2017).

Nevertheless, the extensive alterations of the epigenome observed in tumor cells makes it difficult to distinguish driver from passenger epigenetic alterations, i.e., alterations participating in tumor development or bearing no consequences in this process. The observed ectopic expression of tissue-specific genes could therefore be only a consequence of the removal of the restraining barriers of gene regulation without impacting tumor progression in any way. Indeed, in both Chapter 1 and Chapter 2, we could not group LUAD tissues into discrete subgroups based on the expression of one or the other tissue gene cluster. Rather, these genes displayed a disparate activation in LUAD tissues.

However, the fact that ectopic genes are found highly, frequently, and recurrently activated in tumors points to a potential pro-tumoral function. This was the case for *FAM83A,* a gene belonging to the SE gene cluster. We found that this gene was activated in nearly all LUAD tissue samples, and that their expression was associated with a higher histological grade of these tumors, and a poorer prognosis of LUAD patients. As briefly discussed in Chapter 1, *FAM83A* was found activated in a wide variety of tumors, including pancreatic adenocarcinoma, breast or ovarian cancer (Snijders et al. 2017). Multiple studies have shown that this protein is associated with an increased activation of PI3K/AKT/ and MAPK signaling pathways (Lee et al. 2012; Bartel and Jackson 2017). Indeed, *FAM83A* overexpression in breast cancer cells was found to promote proliferation and invasion, as well as a resistance to EGFR-Tyrosine Kinase inhibitors. Similar effects were extended to

pancreatic cancer cell lines (Chen et al. 2017), implying an oncogenic function in tumors of different histological origins. Overall, these data support that the ectopic activation of tissue-specific genes in tumors could have an impactful effect on tumor biology, and not just be described as another tumor epiphenomenon.

Our study also revealed that LUAD cells show an enrichment of the activation of brain-specific genes to what would be expected by chance. The abundance of this group of genes in LUAD could therefore lead to the hypothesis that brain-specific gene activation could promote advantageous features to tumor cells.

Interestingly, as briefly stated above, a study found that metastatic breast tissues in the brain overexpress genes related to GABA signaling. In a cellular model of breast cancer, the authors found that this ectopic expression allowed breast tumor cells to use GABA as an alternative metabolite and therefore, ensure their proliferative capability (Neman et al. 2014). A recent study also supports similar findings with a different neurotransmitter, glutamate (Zeng et al. 2019). Breast cancer cells that had metastasized to the brain exhibited upregulation of the corresponding glutamate receptor (NMDAR). Intriguingly, breast to brain metastases is a common feature of breast cancer (Gao et al. 2019). Of note, LUAD also show a “brain-preferential” metastatic site. It is therefore tempting to propose that one of the causes of this metastatic organotropism is the capability of tumor cells to adapt to the specific microenvironment of the brain after their seeding in this tissue.

Furthermore, it is now clearly recognized that neurons are an important cell type of the tumor microenvironment (Monje et al. 2020). As a matter of fact, it was shown that for multiple types of tumors, the higher the density of nerve fibers present in tumor lesions, the more aggressive the phenotype of the tumor. Indeed, tumor innervation has been described as an alternative path of metastatic spread, where tumor cells have the ability to migrate through the nerves of their microenvironment. This phenomenon is known as perineural invasion. This concept was well illustrated in prostate adenocarcinoma using mouse models. Indeed, Magnon and colleagues observed that the autonomic adrenergic fibers of the mass regulated tumor development by the release of catecholamine in the tumor microenvironment (Magnon et al. 2013). In contrast, via the activation of the cholinergic fibers, tumor cells were able to promote their metastatic spread. In

accordance with this, in mouse models of gastric adenocarcinoma, it was shown that denervation of the stomach resulted in a decrease of tumor progression (Zhao et al. 2014). Additionally, it was reported that interactions between tumor cells and the nervous system can be achieved through different levels. Firstly, via a direct communication of axons and tumor cells by forming synapses, and secondly, by a paracrine signaling of the microenvironment (Monje et al. 2020). These data indicate that there exists a crosstalk of tumor cells with the nerve fibers of their microenvironment that promotes their development.

Together, these studies suggest that tumor cells of different cancer types, through the activation of genes related to neuronal functions and signaling, could adapt their metabolism to be compatible with the metastatic brain niche, or could exploit nerves as a signaling source to enable their development, and as an alternative route of metastasis. It is therefore enticing to hypothesize that an enrichment in the activation of brain-specific genes in tumors cells would promote and serve the above-described mechanisms.

Overall, these data suggest that tumor cells are able to highjack multiple normal tissue-specific expression programs in order to sustain their development in different environmental conditions.

5. *Ectopic activation of brain-specific genes in tumors: a surrogate of the loss of REST activity*

Our project revealed that for a fraction of brain-specific genes, REST repressor complex is implicated in their transcriptional regulation. REST is involved in the repression of neuronal genes in non-neuronal tissues via their binding to DNA at Repressive element 1 (RE1) sites (Schoenherr and Anderson 1995). REST is part of a larger multiprotein complex composed of several corepressor proteins such as CoREST, HDAC enzymes, H3K4 demethylase (LSD1), or H3K9 methylase (G9a) (Ooi and Wood 2007). Therefore, the recruitment of REST on the genome results in the erasure of activating histone marks and the writing of repressing histone marks, that ultimately lead to a closed chromatin compaction state (Ooi and Wood 2007).

REST activity was reported to be altered in multiple tumor types. Indeed, it was found that REST exerts a tumor suppressive function in epithelial cells (Westbrook et al. 2005). From a mechanistic point of view, the loss of REST function was found to correlate with an increase of the PI3K-AKT signaling pathway in breast carcinoma (Negrini et al. 2013). Another study found that REST would exert their tumor suppressive function via the recruitment of G9a H3K9 methylase onto genomic regions. It was shown that depletion of REST or CDYL, a protein that physically bridges REST to G9a, leads to the transformation of human mammary epithelial cells (Mulligan et al. 2008). Essentially, it is presumed that it is not REST *per se* that drives the tumor suppressive function in cells, but rather the proteins with which they interact. Indeed, REST is viewed as a hub for the recruitment of a myriad of chromatin players.

The alterations of REST activity in tumors were found to be caused by different mechanisms. First, the expression of *REST* was reported to be lost in tumor cells. It was described for instance, that the chromosomal region 4q12, which includes the *REST* locus, is frequently deleted in colon cancer (Westbrook et al. 2005). It was also reported that *REST* is subjected to multiple alternative splicing events, with some of them leading to premature termination and thus, protein truncation. Additionally, other REST protein isoforms appeared to act as dominant negative of REST, impeding therefore REST canonical function (Coulson 2005). Furthermore, it was also shown that REST mRNA and protein levels can be intact in tumor cells. Their activity appears to be altered because of the loss of function of their chromatin remodeler partners such as the SWI/SNF complex (Huang and Bao 2012). Finally, there also appears to be a fine-tuning between the ubiquitination and deubiquitination of REST, important in maintaining the differentiation of neuronal stem cells (Huang and Bao 2012). However, such balancing mechanism has not been shown to play a role in REST protein levels in tumors yet.

Together, these data suggest that there is a plethora of mechanisms that lead to REST deregulation in tumor cells. Since REST was found to exert tumor suppressive functions in non-neuronal tumors, it is important to be able to detect their loss of activity in these cells. The brain-specific genes identified as ectopically activated in tumors by REST deregulation could be used as a surrogate of REST activity in these cells. In accordance, in prostate tumor cells, it was shown that the emergence of a neuroendocrine phenotype coincides with the loss of REST activity (Coulson 2005).

Therefore, determining the brain-specific genes induced by REST loss of function across multiple tumor types, could provide new markers to monitor REST activity in these cells.

Conclusion and future steps

Our study shows that epigenetic alterations in tumor cells, consisting of promoter DNA demethylation alone or combined with the ectopic activation of tissue-specific transcription factors (such as *HNF4A*), or the loss of chromatin complex activity (such as REST), lead to ectopic activation of tissue-specific expression programs in tumors (**Fig. 23**). This results in an altered cell identity compared to their normal counterpart. Literature-based evidence suggest that several genes belonging to these tissue-specific gene clusters could exert pro-tumoral functions, and this, in different types of tumors. Their ectopic expression could therefore be exploited as a marker of the deregulation events that lead to their activation in these cells, and offer new therapeutic opportunities.

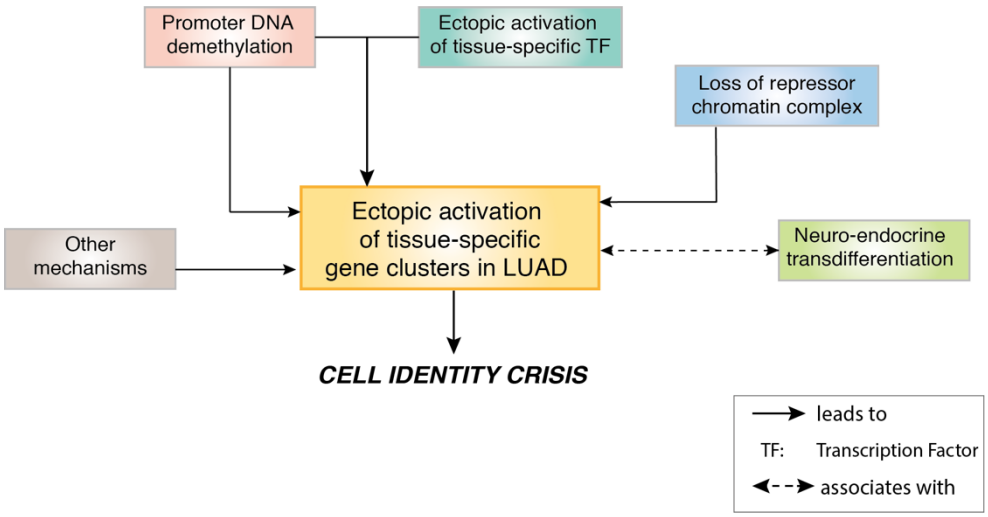


Figure 23: Summary of the epigenetic alterations identified in our study leading to ectopic activation of tissue-specific gene clusters in LUAD.

Future analyses will be focusing on the ectopic activation of brain-specific genes in tumors described in Chapter 2 and will be studying the causes of REST deregulation in LUAD. In particular, we will focus on expression analyses of *REST* and their partners in LUAD, to assess the extent of their transcriptional repression in these cells. Additionally, we will also use exon and exon-junction quantification data, available from the LUAD-TCGA project, in order to investigate potential REST alternative splicing events leading to their deregulation in LUAD.

Finally, regarding the possibility that the activation of brain-specific genes would be a determinant factor in the metastatic process to the brain, we will compare the transcriptome of LUAD primary tissues that have metastasized or not in the brain. We will seek to evaluate whether LUAD tissues that presented brain metastases displayed a more frequent activation of brain-specific genes compared to the LUAD tissues that did not (data obtained from GEO-NCBI and generated by the Taipei Veterans General Hospital, Taiwan).

Discussion references

Akers, S. N., K. Odunsi, and A. R. Karpf. 2010. 'Regulation of cancer germline antigen gene expression: implications for cancer immunotherapy', *Future Oncol*, 6: 717-32.

Almeida, L. G., N. J. Sakabe, A. R. deOliveira, M. C. Silva, A. S. Mundstein, T. Cohen, Y. T. Chen, R. Chua, S. Gurung, S. Gnjatic, A. A. Jungbluth, O. L. Caballero, A. Bairoch, E. Kiesler, S. L. White, A. J. Simpson, L. J. Old, A. A. Camargo, and A. T. Vasconcelos. 2009. 'CTdatabase: a knowledge-base of high-throughput and curated data on cancer-testis antigens', *Nucleic Acids Res*, 37: D816-9.

Antequera, F., J. Boyes, and A. Bird. 1990. 'High levels of de novo methylation and altered chromatin structure at CpG islands in cell lines', *Cell*, 62: 503-14.

Axelsen, J. B., J. Lotem, L. Sachs, and E. Domany. 2007. 'Genes overexpressed in different human solid cancers exhibit different tissue-specific expression profiles', *Proc Natl Acad Sci U S A*, 104: 13122-7.

Bannister, A. J., and T. Kouzarides. 2011. 'Regulation of chromatin by histone modifications', *Cell Res*, 21: 381-95.

Barau, J., A. Teissandier, N. Zamudio, S. Roy, V. Nalesso, Y. Herault, F. Guillou, and D. Bourc'his. 2016. 'The DNA methyltransferase DNMT3C protects male germ cells from transposon activity', *Science*, 354: 909-12.

Barr, M. L., and E. G. Bertram. 1949. 'A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis', *Nature*, 163: 676.

Bartel, C. A., and M. W. Jackson. 2017. 'HER2-positive breast cancer cells expressing elevated FAM83A are sensitive to FAM83A loss', *PLoS One*, 12: e0176778.

Baylin, S. B., J. W. Hoppener, A. de Bustros, P. H. Steenbergh, C. J. Lips, and B. D. Nelkin. 1986. 'DNA methylation patterns of the calcitonin gene in human lung cancers and lymphomas', *Cancer Res*, 46: 2917-22.

Baylin, S. B., and P. A. Jones. 2011. 'A decade of exploring the cancer epigenome - biological and translational implications', *Nat Rev Cancer*, 11: 726-34.

Bergethon, K., A. T. Shaw, S. H. Ou, R. Katayama, C. M. Lovly, N. T. McDonald, P. P. Massion, C. Siwak-Tapp, A. Gonzalez, R. Fang, E. J. Mark, J. M. Batten, H. Chen, K. D. Wilner, E. L. Kwak, J. W. Clark, D. P. Carbone, H. Ji, J. A. Engelman, M. Mino-Kenudson, W. Pao, and A. J. Iafrate. 2012. 'ROS1 rearrangements define a unique molecular class of lung cancers', *J Clin Oncol*, 30: 863-70.

Bestor, T. H., J. R. Edwards, and M. Boulard. 2015. 'Notes on the role of dynamic DNA methylation in mammalian development', *Proc Natl Acad Sci U S A*, 112: 6796-9.

Bhan, S., S. S. Negi, C. Shao, C. A. Glazer, A. Chuang, D. A. Gaykalova, W. Sun, D. Sidransky, P. K. Ha, and J. A. Califano. 2011. 'BORIS binding to the promoters of cancer testis antigens, MAGEA2, MAGEA3, and MAGEA4, is associated with their transcriptional activation in lung cancer', *Clin Cancer Res*, 17: 4267-76.

Bhattacharjee, A., W. G. Richards, J. Staunton, C. Li, S. Monti, P. Vasa, C. Ladd, J. Beheshti, R. Bueno, M. Gillette, M. Loda, G. Weber, E. J. Mark, E. S. Lander, W. Wong, B. E. Johnson, T. R. Golub, D. J. Sugarbaker, and M. Meyerson. 2001. 'Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses', *Proc Natl Acad Sci U S A*, 98: 13790-5.

Billingham, R. E., and W. K. Silvers. 1963. 'The origin and conservation of epidermal specificities', *N Engl J Med*, 268: 539-45 concl.

Bird, A. 2002. 'DNA methylation patterns and epigenetic memory', *Genes Dev*, 16: 6-21.

Bird, A. P. 1980. 'DNA methylation and the frequency of CpG in animal DNA', *Nucleic Acids Res*, 8: 1499-504.

———. 1986. 'CpG-rich islands and the function of DNA methylation', *Nature*, 321: 209-13.

Bishop, J. L., D. Thaper, S. Vahid, A. Davies, K. Ketola, H. Kuruma, R. Jama, K. M. Nip, A. Angeles, F. Johnson, A. W. Wyatt, L. Fazli, M. E. Gleave, D. Lin, M. A. Rubin, C. C. Collins, Y. Wang, H. Beltran, and A.

Zoubeidi. 2017. 'The Master Neural Transcription Factor BRN2 Is an Androgen Receptor-Suppressed Driver of Neuroendocrine Differentiation in Prostate Cancer', *Cancer Discov*, 7: 54-71.

Blanpain, C. 2013. 'Tracing the cellular origin of cancer', *Nat Cell Biol*, 15: 126-34.

Boel, P., C. Wildmann, M. L. Sensi, R. Brasseur, J. C. Renauld, P. Coulie, T. Boon, and P. van der Bruggen. 1995. 'BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes', *Immunity*, 2: 167-75.

Bolger, A. M., M. Lohse, and B. Usadel. 2014. 'Trimmomatic: a flexible trimmer for Illumina sequence data', *Bioinformatics*, 30: 2114-20.

Borges, M., R. I. Linnoila, H. J. van de Velde, H. Chen, B. D. Nelkin, M. Mabry, S. B. Baylin, and D. W. Ball. 1997. 'An achaete-scute homologue essential for neuroendocrine differentiation in the lung', *Nature*, 386: 852-5.

Borromeo, M. D., T. K. Savage, R. K. Kollipara, M. He, A. Augustyn, J. K. Osborne, L. Girard, J. D. Minna, A. F. Gazdar, M. H. Cobb, and J. E. Johnson. 2016. 'ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs', *Cell Rep*, 16: 1259-72.

Bourc'his, D., G. L. Xu, C. S. Lin, B. Bollman, and T. H. Bestor. 2001. 'Dnmt3L and the establishment of maternal genomic imprints', *Science*, 294: 2536-9.

Brockdorff, N., A. Ashworth, G. F. Kay, P. Cooper, S. Smith, V. M. McCabe, D. P. Norris, G. D. Penny, D. Patel, and S. Rastan. 1991. 'Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome', *Nature*, 351: 329-31.

Buck-Koehntop, B. A., and P. A. Defossez. 2013. 'On how mammalian transcription factors recognize methylated DNA', *Epigenetics*, 8: 131-7.

Cadieux, B., T. T. Ching, S. R. VandenBerg, and J. F. Costello. 2006. 'Genome-wide hypomethylation in human glioblastomas associated with specific copy number alteration, methylenetetrahydrofolate reductase allele status, and increased proliferation', *Cancer Res*, 66: 8469-76.

Cai, L., D. Luo, B. Yao, D. M. Yang, S. Lin, L. Girard, R. J. DeBerardinis, J. D. Minna, Y. Xie, and G. Xiao. 2019. 'Systematic Analysis of Gene Expression in Lung Adenocarcinoma and Squamous Cell Carcinoma with a Case Study of FAM83A and FAM83B', *Cancers (Basel)*, 11.

Cajuso, T., P. Sulo, T. Tanskanen, R. Katainen, A. Taira, U. A. Hanninen, J. Kondelin, L. Forsstrom, N. Valimaki, M. Aavikko, E. Kaasinen, A. Ristimaki, S. Koskensalo, A. Lepisto, L. Renkonen-Sinisalo, T. Seppala, T. Kuopio, J. Bohm, J. P. Mecklin, O. Kilpivaara, E. Pitkanen, K. Palin, and L. A. Aaltonen. 2019. 'Retrotransposon insertions can initiate colorectal cancer and are associated with poor survival', *Nat Commun*, 10: 4022.

Camolotto, S. A., S. Pattabiraman, T. L. Mosbruger, A. Jones, V. K. Belova, G. Orstad, M. Streiff, L. Salmond, C. Stubben, K. H. Kaestner, and E. L. Snyder. 2018. 'FoxA1 and FoxA2 drive gastric differentiation and suppress squamous identity in NKX2-1-negative lung cancer', *Elife*, 7.

Cancer Genome Atlas Research, Network. 2014. 'Comprehensive molecular profiling of lung adenocarcinoma', *Nature*, 511: 543-50.

Cancer Genome Atlas Research Network. Electronic address, andrew aguirre dfci harvard edu, and Network Cancer Genome Atlas Research. 2017. 'Integrated Genomic Characterization of Pancreatic Ductal Adenocarcinoma', *Cancer Cell*, 32: 185-203 e13.

Cannuyer, J., A. Loriot, G. K. Parvizi, and C. De Smet. 2013a. 'Epigenetic hierarchy within the MAGEA1 cancer-germline gene: promoter DNA methylation dictates local histone modification', *PLoS One*, 8: e58743.

———. 2013b. 'Epigenetic hierarchy within the MAGEA1 cancer-germline gene: promoter DNA methylation dictates local histone modifications', *PLoS One*, 8: e58743.

Canzio, D., E. Y. Chang, S. Shankar, K. M. Kuchenbecker, M. D. Simon, H. D. Madhani, G. J. Narlikar, and B. Al-Sady. 2011. 'Chromodomain-mediated oligomerization of HP1 suggests a nucleosome-bridging mechanism for heterochromatin assembly', *Mol Cell*, 41: 67-81.

Cardoso, M. C., and H. Leonhardt. 1999. 'DNA methyltransferase is actively retained in the cytoplasm during early development', *J Cell Biol*, 147: 25-32.

Carrozza, M. J., B. Li, L. Florens, T. Suganuma, S. K. Swanson, K. K. Lee, W. J. Shia, S. Anderson, J. Yates, M. P. Washburn, and J. L. Workman. 2005. 'Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription', *Cell*, 123: 581-92.

Casarosa, S., C. Fode, and F. Guillemot. 1999. 'Mash1 regulates neurogenesis in the ventral telencephalon', *Development*, 126: 525-34.

Catteau, A., W. H. Harris, C. F. Xu, and E. Solomon. 1999. 'Methylation of the BRCA1 promoter region in sporadic breast and ovarian cancer: correlation with disease characteristics', *Oncogene*, 18: 1957-65.

Chan, B. A., and B. G. Hughes. 2015. 'Targeted therapy for non-small cell lung cancer: current standards and the promise of the future', *Transl Lung Cancer Res*, 4: 36-54.

Chen, J., H. Yang, A. S. M. Teo, L. B. Amer, F. G. Sherbaf, C. Q. Tan, J. J. S. Alvarez, B. Lu, J. Q. Lim, A. Takano, R. Nahar, Y. Y. Lee, C. Z. J. Phua, K. P. Chua, L. Suteja, P. J. Chen, M. M. Chang, T. P. T. Koh, B. H. Ong, D. Anantham, A. A. L. Hsu, A. Gogna, C. W. Too, Z. W. Aung, Y. F. Lee, L. Wang, T. K. H. Lim, A. Wilm, P. S. Choi, P. Y. Ng, C. K. Toh, W. T. Lim, S. Ma, B. Lim, J. Liu, W. L. Tam, A. J. Skanderup, J. P. S. Yeong, E. H. Tan, C. L. Creasy, D. S. W. Tan, A. M. Hillmer, and W. Zhai. 2020. 'Genomic landscape of lung adenocarcinoma in East Asians', *Nat Genet*, 52: 177-86.

Chen, S., J. Huang, Z. Liu, Q. Liang, N. Zhang, and Y. Jin. 2017. 'FAM83A is amplified and promotes cancer stem cell-like traits and chemoresistance in pancreatic cancer', *Oncogenesis*, 6: e300.

Chen, Y. T., M. J. Scanlan, C. A. Venditti, R. Chua, G. Theiler, B. J. Stevenson, C. Iseli, A. O. Gure, T. Vasicek, R. L. Strausberg, C. V. Jongeneel, L. J. Old, and A. J. Simpson. 2005. 'Identification of cancer/testis-antigen genes by massively parallel signature sequencing', *Proc Natl Acad Sci U S A*, 102: 7940-5.

Colaprico, A., T. C. Silva, C. Olsen, L. Garofano, C. Cava, D. Garolini, T. S. Sabedot, T. M. Malta, S. M. Pagnotta, I. Castiglioni, M. Ceccarelli, G. Bontempi, and H. Noushmehr. 2016. 'TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data', *Nucleic Acids Res*, 44: e71.

Consortium, Encode Project. 2012. 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489: 57-74.

Consortium, G. TEx. 2013. 'The Genotype-Tissue Expression (GTEx) project', *Nat Genet*, 45: 580-5.

Coulson, J. M. 2005. 'Transcriptional regulation: cancer, neurons and the REST', *Curr Biol*, 15: R665-8.

D'Alessio, A. C., I. C. Weaver, and M. Szyf. 2007. 'Acetylation-induced transcription is required for active DNA demethylation in methylation-silenced genes', *Mol Cell Biol*, 27: 7462-74.

da Silva, V. L., A. F. Fonseca, M. Fonseca, T. E. da Silva, A. C. Coelho, J. E. Kroll, J. E. S. de Souza, B. Stransky, G. A. de Souza, and S. J. de Souza. 2017. 'Genome-wide identification of cancer/testis genes and their association with prognosis in a pan-cancer analysis', *Oncotarget*, 8: 92966-77.

Dahlet, T., A. Argueso Lleida, H. Al Adhami, M. Dumas, A. Bender, R. P. Ngondo, M. Tanguy, J. Vallet, G. Auclair, A. F. Bardet, and M. Weber. 2020. 'Genome-wide analysis in the mouse embryo reveals the importance of DNA methylation for transcription integrity', *Nat Commun*, 11: 3153.

de la Serna, I. L., Y. Ohkawa, and A. N. Imbalzano. 2006. 'Chromatin remodelling in mammalian differentiation: lessons from ATP-dependent remodellers', *Nat Rev Genet*, 7: 461-73.

De Plaen, E., C. Lurquin, B. Lethe, P. van der Bruggen, V. Brichard, J. C. Renauld, P. Coulie, A. Van Pel, and T. Boon. 1997. 'Identification of genes coding for tumor antigens recognized by cytolytic T lymphocytes', *Methods*, 12: 125-42.

De Smet, C., S. J. Courtois, I. Faraoni, C. Lurquin, J. P. Szikora, O. De Backer, and T. Boon. 1995. 'Involvement of two Ets binding sites in the transcriptional activation of the MAGE1 gene', *Immunogenetics*, 42: 282-90.

De Smet, C., O. De Backer, I. Faraoni, C. Lurquin, F. Brasseur, and T. Boon. 1996a. 'The activation of human gene MAGE-1 in

tumor cells is correlated with genome-wide demethylation', *Proc Natl Acad Sci U S A*, 93: 7149-53.

———. 1996b. 'The activation of human gene MAGE-1 in tumor cells is correlated with genome-wide demethylation', *Proc Natl Acad Sci U S A*, 93: 7149-53.

De Smet, C., and A. Loriot. 2010. 'DNA hypomethylation in cancer: epigenetic scars of a neoplastic journey', *Epigenetics*, 5: 206-13.

———. 2013. 'DNA hypomethylation and activation of germline-specific genes in cancer', *Adv Exp Med Biol*, 754: 149-66.

De Smet, C., C. Lurquin, B. Lethe, V. Martelange, and T. Boon. 1999. 'DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter', *Mol Cell Biol*, 19: 7327-35.

De Smet, C., C. Lurquin, B. Lethé, V. Martelange, and T. Boon. 1999. 'DNA methylation is the primary silencing mechanism for a set of germ line- and tumor-specific genes with a CpG-rich promoter.', *Mol. Cell. Biol.*, 19: 7327-35.

Deans, C., and K. A. Maggert. 2015. 'What do you mean, "epigenetic"?', *Genetics*, 199: 887-96.

Deichmann, U. 2016. 'Epigenetics: The origins and evolution of a fashionable topic', *Dev Biol*, 416: 249-54.

Dhayalan, A., A. Rajavelu, P. Rathert, R. Tamas, R. Z. Jurkowska, S. Ragozin, and A. Jeltsch. 2010. 'The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation', *J Biol Chem*, 285: 26114-20.

Diacofotaki, A., A. Loriot, and C. De Smet. 2022. 'Identification of Tissue-Specific Gene Clusters Induced by DNA Demethylation in Lung Adenocarcinoma: More Than Germline Genes', *Cancers (Basel)*, 14.

Dietrich, N., M. Lerdrup, E. Landt, S. Agrawal-Singh, M. Bak, N. Tommerup, J. Rappsilber, E. Sodersten, and K. Hansen. 2012. 'REST-mediated recruitment of polycomb repressor complexes in mammalian cells', *PLoS Genet*, 8: e1002494.

Disteche, C. M., and J. B. Berletch. 2015. 'X-chromosome inactivation and escape', *J Genet*, 94: 591-9.

Djureinovic, D., B. M. Hallstrom, M. Horie, J. S. M. Mattsson, L. La Fleur, L. Fagerberg, H. Brunnstrom, C. Lindskog, K. Madjar, J. Rahnenfuhrer, S. Ekman, E. Stahle, H. Koyi, E. Branden, K. Edlund, J. G. Hengstler, M. Lambe, A. Saito, J. Botling, F. Ponten, M. Uhlen, and P. Micke. 2016. 'Profiling cancer testis antigens in non-small-cell lung cancer', *JCI Insight*, 1: e86837.

Dokun, O. Y., A. R. Florl, H. H. Seifert, I. Wolff, and W. A. Schulz. 2008. 'Relationship of SNCG, S100A4, S100A9 and LCN2 gene expression and DNA methylation in bladder cancer', *Int J Cancer*, 123: 2798-807.

Doyle, J. M., J. Gao, J. Wang, M. Yang, and P. R. Potts. 2010. 'MAGE-RING protein complexes comprise a family of E3 ubiquitin ligases', *Mol Cell*, 39: 963-74.

Drouin-Ouellet, J., S. Lau, P. L. Brattas, D. Rylander Ottosson, K. Pircs, D. A. Grassi, L. M. Collins, R. Vuono, A. Andersson Sjoland, G. Westergren-Thorsson, C. Graff, L. Minthon, H. Toresson, R. A. Barker, J. Jakobsson, and M. Parmar. 2017. 'REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways', *EMBO Mol Med*, 9: 1117-31.

Ehrlich, M. 2002. 'DNA methylation in cancer: too much, but also too little', *Oncogene*, 21: 5400-13.

———. 2009a. 'DNA hypomethylation in cancer cells', *Epigenomics*, 1: 239-59.

———. 2009b. 'DNA hypomethylation in cancer cells', *Epigenomics*, 1: 239-59.

———. 2019. 'DNA hypermethylation in disease: mechanisms and clinical relevance', *Epigenetics*, 14: 1141-63.

Esteller, M. 2002. 'CpG island hypermethylation and tumor suppressor genes: a booming present, a brighter future', *Oncogene*, 21: 5427-40.

———. 2007. 'Epigenetic gene silencing in cancer: the DNA hypermethylome', *Hum Mol Genet*, 16 Spec No 1: R50-9.

Fain, J. S., A. Loriot, A. Diacofotaki, A. Van Tongelen, and C. De Smet. 2021. 'Transcriptional overlap links DNA hypomethylation with DNA hypermethylation at adjacent promoters in cancer', *Sci Rep*, 11: 17346.

Fain, J. S., A. Van Tongelen, A. Loriot, and C. De Smet. 2019. 'Epigenetic Coactivation of MAGEA6 and CT-GABRA3 Defines Orientation of a Segmental Duplication in the Human X Chromosome', *Cytogenet Genome Res*, 159: 12-18.

Feinberg, A. P. 2014. 'Epigenetic stochasticity, nuclear structure and cancer: the

implications for medicine', *J Intern Med*, 276: 5-11.

Feinberg, A. P., and B. Vogelstein. 1983a. 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, 301: 89-92.

———. 1983b. 'Hypomethylation distinguishes genes of some human cancers from their normal counterparts', *Nature*, 301: 89-92.

Ferone, G., M. C. Lee, J. Sage, and A. Berns. 2020. 'Cells of origin of lung cancers: lessons from mouse studies', *Genes Dev*, 34: 1017-32.

Flavahan, W. A., E. Gaskell, and B. E. Bernstein. 2017. 'Epigenetic plasticity and the hallmarks of cancer', *Science*, 357.

Fon Tacer, K., M. C. Montoya, M. J. Oatley, T. Lord, J. M. Oatley, J. Klein, R. Ravichandran, H. Tillman, M. Kim, J. P. Connelly, S. M. Pruett-Miller, A. L. Bookout, E. Binshtock, M. M. Kaminski, and P. R. Potts. 2019. 'MAGE cancer-testis antigens protect the mammalian germline under environmental stress', *Sci Adv*, 5: eaav4832.

Futscher, B. W., M. M. Oshiro, R. J. Wozniak, N. Holtan, C. L. Hanigan, H. Duan, and F. E. Domann. 2002. 'Role for DNA methylation in the control of cell type specific maspin expression', *Nat Genet*, 31: 175-9.

Gama-Sosa, M. A., V. A. Slagel, R. W. Trewyn, R. Oxenhandler, K. C. Kuo, C. W. Gehrke, and M. Ehrlich. 1983a. 'The 5-methylcytosine content of DNA from human tumors', *Nucleic Acids Res*, 11: 6883-94.

———. 1983b. 'The 5-methylcytosine content of DNA from human tumors', *Nucleic Acids Res*, 11: 6883-94.

Gantchev, J., A. Martinez Villarreal, S. Gunn, M. Zetka, N. Odum, and I. V. Litvinov. 2020. 'The ectopic expression of meiCT genes promotes meiomitosis and may facilitate carcinogenesis', *Cell Cycle*, 19: 837-54.

Gao, Y., I. Bado, H. Wang, W. Zhang, J. M. Rosen, and X. H. Zhang. 2019. 'Metastasis Organotropism: Redefining the Congenial Soil', *Dev Cell*, 49: 375-91.

Gaudet, F., J. G. Hodgson, A. Eden, L. Jackson-Grusby, J. Dausman, J. W. Gray, H. Leonhardt, and R. Jaenisch. 2003. 'Induction of tumors in mice by genomic hypomethylation', *Science*, 300: 489-92.

Gelfman, S., and G. Ast. 2013. 'When epigenetics meets alternative splicing: the roles of DNA methylation and GC architecture', *Epigenomics*, 5: 351-3.

Gibbs, Z. A., and A. W. Whitehurst. 2018. 'Emerging Contributions of Cancer/Testis Antigens to Neoplastic Behaviors', *Trends Cancer*, 4: 701-12.

Glaser, J., J. Iranzo, M. Borensztein, M. Marinucci, A. Gualtieri, C. Jouhanneau, A. Teissandier, C. Gaston-Massuet, and D. Bourc'his. 2022. 'The imprinted Zdbf2 gene finely tunes control of feeding and growth in neonates', *Elife*, 11.

Goll, M. G., F. Kirpekar, K. A. Maggert, J. A. Yoder, C. L. Hsieh, X. Zhang, K. G. Golic, S. E. Jacobsen, and T. H. Bestor. 2006. 'Methylation of tRNAAsp by the DNA methyltransferase homolog Dnmt2', *Science*, 311: 395-8.

Grandin, M., P. Mathot, G. Devailly, Y. Bidet, A. Ghantous, C. Favrot, B. Gibert, N. Gadot, I. Puisieux, Z. Herceg, J. G. Delcros, A. Bernet, P. Mehlen, and R. Dante. 2016. 'Inhibition of DNA methylation promotes breast tumor sensitivity to netrin-1 interference', *EMBO Mol Med*, 8: 863-77.

Greally, J. M. 2018. 'A user's guide to the ambiguous word 'epigenetics'', *Nat Rev Mol Cell Biol*, 19: 207-08.

Grunwald, C., M. Koslowski, T. Arsiray, K. Dhaene, M. Praet, A. Victor, A. Morresi-Hauf, M. Lindner, B. Passlick, H. A. Lehr, S. C. Schafer, G. Seitz, C. Huber, U. Sahin, and O. Tureci. 2006. 'Expression of multiple epigenetically regulated cancer/germline genes in nonsmall cell lung cancer', *Int J Cancer*, 118: 2522-8.

Gu, Z., R. Eils, and M. Schlesner. 2016. 'Complex heatmaps reveal patterns and correlations in multidimensional genomic data', *Bioinformatics*, 32: 2847-9.

Guo, J. U., Y. Su, C. Zhong, G. L. Ming, and H. Song. 2011. 'Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain', *Cell*, 145: 423-34.

Hackett, J. A., R. Sengupta, J. J. Zylicz, K. Murakami, C. Lee, T. A. Down, and M. A. Surani. 2013. 'Germline DNA demethylation dynamics and imprint erasure through 5-hydroxymethylcytosine', *Science*, 339: 448-52.

Hanahan, D. 2022. 'Hallmarks of Cancer: New Dimensions', *Cancer Discov*, 12: 31-46.

Hayes, D. N., S. Monti, G. Parmigiani, C. B. Gilks, K. Naoki, A. Bhattacharjee, M. A. Socinski, C. Perou, and M. Meyerson. 2006. 'Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts', *J Clin Oncol*, 24: 5079-90.

Heist, R. S., and J. A. Engelman. 2012. 'SnapShot: non-small cell lung cancer', *Cancer Cell*, 21: 448 e2.

Hendrich, B., and A. Bird. 1998. 'Identification and characterization of a family of mammalian methyl-CpG binding proteins', *Mol Cell Biol*, 18: 6538-47.

Herman, J. G., F. Latif, Y. Weng, M. I. Lerman, B. Zbar, S. Liu, D. Samid, D. S. Duan, J. R. Gnarra, W. M. Linehan, and et al. 1994. 'Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma', *Proc Natl Acad Sci U S A*, 91: 9700-4.

Hermann, A., R. Goyal, and A. Jeltsch. 2004. 'The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites', *J Biol Chem*, 279: 48350-9.

Hofmann, O., O. L. Caballero, B. J. Stevenson, Y. T. Chen, T. Cohen, R. Chua, C. A. Maher, S. Panji, U. Schaefer, A. Kruger, M. Lehvaslaiho, P. Carninci, Y. Hayashizaki, C. V. Jongeneel, A. J. Simpson, L. J. Old, and W. Hide. 2008. 'Genome-wide analysis of cancer/testis gene expression', *Proc Natl Acad Sci U S A*, 105: 20422-7.

Holliday, R. 2006. 'Epigenetics: a historical overview', *Epigenetics*, 1: 76-80.

Holliday, R., and J. E. Pugh. 1975. 'DNA modification mechanisms and gene activity during development', *Science*, 187: 226-32.

Howe, F. S., H. Fischl, S. C. Murray, and J. Mellor. 2017. 'Is H3K4me3 instructive for transcription activation?', *Bioessays*, 39: 1-12.

Huang, Y., S. J. Myers, and R. Dingledine. 1999. 'Transcriptional repression by REST: recruitment of Sin3A and histone deacetylase to neuronal genes', *Nat Neurosci*, 2: 867-72.

Huang, Z., and S. Bao. 2012. 'Ubiquitination and deubiquitination of REST and its roles in cancers', *FEBS Lett*, 586: 1602-5.

Ikeda, K., K. Shiraishi, A. Eguchi, H. Shibata, K. Yoshimoto, T. Mori, Y. Baba, H. Baba, and M. Suzuki. 2013. 'Long interspersed nucleotide element 1 hypomethylation is associated with poor prognosis of lung adenocarcinoma', *Ann Thorac Surg*, 96: 1790-4.

Illingworth, R. S., U. Gruenewald-Schneider, S. Webb, A. R. Kerr, K. D. James, D. J. Turner, C. Smith, D. J. Harrison, R. Andrews, and A. P. Bird. 2010. 'Orphan CpG islands identify numerous conserved promoters in the mammalian genome', *PLoS Genet*, 6: e1001134.

Imrichova, H., G. Hulselmans, Z. K. Atak, D. Potier, and S. Aerts. 2015. 'i-cisTarget 2015 update: generalized cis-regulatory enrichment analysis in human, mouse and fly', *Nucleic Acids Res*, 43: W57-64.

Iqbal, K., S. G. Jin, G. P. Pfeifer, and P. E. Szabo. 2011. 'Reprogramming of the paternal genome upon fertilization involves genome-wide oxidation of 5-methylcytosine', *Proc Natl Acad Sci U S A*, 108: 3642-7.

Ito, S., L. Shen, Q. Dai, S. C. Wu, L. B. Collins, J. A. Swenberg, C. He, and Y. Zhang. 2011. 'Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine', *Science*, 333: 1300-3.

Jain, D., C. Meydan, J. Lange, C. Claeys Bouuaert, N. Lailler, C. E. Mason, K. V. Anderson, and S. Keeney. 2017. 'rahu is a mutant allele of Dnmt3c, encoding a DNA methyltransferase homolog required for meiosis and transposon repression in the mouse male germline', *PLoS Genet*, 13: e1006964.

Jenuwein, T., and C. D. Allis. 2001. 'Translating the histone code', *Science*, 293: 1074-80.

Jiang, J. C., and K. R. Upton. 2019. 'Human transposons are an abundant supply of transcription factor binding sites and promoter activities in breast cancer cell lines', *Mob DNA*, 10: 16.

Jones, P. A. 1999. 'The DNA methylation paradox', *Trends Genet*, 15: 34-7.

———. 2012. 'Functions of DNA methylation: islands, start sites, gene bodies and beyond', *Nat Rev Genet*, 13: 484-92.

Jones, P. A., and D. Takai. 2001. 'The role of DNA methylation in mammalian epigenetics', *Science*, 293: 1068-70.

Jungbluth, A. A., W. A. Silva, Jr., K. Iversen, D. Frosina, B. Zaidi, K. Coplan, S. K.

Eastlake-Wade, S. B. Castelli, G. C. Spagnoli, L. J. Old, and M. Vogel. 2007. 'Expression of cancer-testis (CT) antigens in placenta', *Cancer Immun*, 7: 15.

Jungbluth, A. A., E. Stockert, Y. T. Chen, D. Kolb, K. Iversen, K. Coplan, B. Williamson, N. Altorki, K. J. Busam, and L. J. Old. 2000. 'Monoclonal antibody MA454 reveals a heterogeneous expression pattern of MAGE-1 antigen in formalin-fixed paraffin embedded lung tumours', *Br J Cancer*, 83: 493-7.

Kandoth, C., M. D. McLellan, F. Vandin, K. Ye, B. Niu, C. Lu, M. Xie, Q. Zhang, J. F. McMichael, M. A. Wyczalkowski, M. D. M. Leiserson, C. A. Miller, J. S. Welch, M. J. Walter, M. C. Wendl, T. J. Ley, R. K. Wilson, B. J. Raphael, and L. Ding. 2013. 'Mutational landscape and significance across 12 major cancer types', *Nature*, 502: 333-39.

Kane, M. F., M. Loda, G. M. Gaida, J. Lipman, R. Mishra, H. Goldman, J. M. Jessup, and R. Kolodner. 1997. 'Methylation of the hMLH1 promoter correlates with lack of expression of hMLH1 in sporadic colon tumors and mismatch repair-defective human tumor cell lines', *Cancer Res*, 57: 808-11.

Kastner, S., T. Voss, S. Keuerleber, C. Glockel, M. Freissmuth, and W. Sommergruber. 2012. 'Expression of G protein-coupled receptor 19 in human lung cancer cells is triggered by entry into S-phase and supports G(2)-M cell-cycle progression', *Mol Cancer Res*, 10: 1343-58.

Khan, M., S. Arooj, R. Li, Y. Tian, J. Zhang, J. Lin, Y. Liang, A. Xu, R. Zheng, M. Liu, and Y. Yuan. 2020. 'Tumor Primary Site and Histology Subtypes Role in Radiotherapeutic Management of Brain Metastases', *Front Oncol*, 10: 781.

Kim, D., J. M. Paggi, C. Park, C. Bennett, and S. L. Salzberg. 2019. 'Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype', *Nat Biotechnol*, 37: 907-15.

Klose, R. J., and A. P. Bird. 2006. 'Genomic DNA methylation: the mark and its mediators', *Trends Biochem Sci*, 31: 89-97.

Kosari, F., C. M. Ida, M. C. Aubry, L. Yang, I. V. Kovtun, J. L. Klein, Y. Li, S. Erdogan, S. C. Tomaszek, S. J. Murphy, L. C. Bolette, C. P. Kolbert, P. Yang, D. A. Wigle, and G. Vasmatzis. 2014. 'ASCL1 and RET expression defines a clinically relevant subgroup of lung adenocarcinoma characterized by neuroendocrine differentiation', *Oncogene*, 33: 3776-83.

Koslowski, M., C. Bell, G. Seitz, H. A. Lehr, K. Roemer, H. Muntefering, C. Huber, U. Sahin, and O. Tureci. 2004. 'Frequent nonrandom activation of germ-line genes in human cancer', *Cancer Res*, 64: 5988-93.

Kouzarides, T. 2007. 'Chromatin modifications and their function', *Cell*, 128: 693-705.

Krueger, F., and S. R. Andrews. 2011. 'Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications', *Bioinformatics*, 27: 1571-2.

Kwak, E. L., Y. J. Bang, D. R. Camidge, A. T. Shaw, B. Solomon, R. G. Maki, S. H. Ou, B. J. Dezube, P. A. Janne, D. B. Costa, M. Varella-Garcia, W. H. Kim, T. J. Lynch, P. Fidias, H. Stubbs, J. A. Engelman, L. V. Sequist, W. Tan, L. Gandhi, M. Mino-Kenudson, G. C. Wei, S. M. Shreeve, M. J. Ratain, J. Settleman, J. G. Christensen, D. A. Haber, K. Wilner, R. Salgia, G. I. Shapiro, J. W. Clark, and A. J. Iafrate. 2010. 'Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer', *N Engl J Med*, 363: 1693-703.

Lane, N., W. Dean, S. Erhardt, P. Hajkova, A. Surani, J. Walter, and W. Reik. 2003. 'Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse', *Genesis*, 35: 88-93.

Latimer, K. M., and T. F. Mott. 2015. 'Lung cancer: diagnosis, treatment principles, and screening', *Am Fam Physician*, 91: 250-6.

Lauberth, S. M., T. Nakayama, X. Wu, A. L. Ferris, Z. Tang, S. H. Hughes, and R. G. Roeder. 2013. 'H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation', *Cell*, 152: 1021-36.

Lee, S. Y., R. Meier, S. Furuta, M. E. Lenburg, P. A. Kenny, R. Xu, and M. J. Bissell. 2012. 'FAM83A confers EGFR-TKI resistance in breast cancer cells and in mice', *J Clin Invest*, 122: 3211-20.

Lewis, S. K., D. Nachun, M. G. Martin, S. Horvath, G. Coppola, and D. L. Jones. 2020. 'DNA Methylation Analysis Validates Organoids as a Viable Model for

Studying Human Intestinal Aging', *Cell Mol Gastroenterol Hepatol*, 9: 527-41.

Li, B., M. Carey, and J. L. Workman. 2007. 'The role of chromatin during transcription', *Cell*, 128: 707-19.

Li, C., and H. Lu. 2018. 'Adenosquamous carcinoma of the lung', *Onco Targets Ther*, 11: 4829-35.

Li, E., C. Beard, and R. Jaenisch. 1993. 'Role for DNA methylation in genomic imprinting', *Nature*, 366: 362-5.

Li, E., T. H. Bestor, and R. Jaenisch. 1992. 'Targeted mutation of the DNA methyltransferase gene results in embryonic lethality', *Cell*, 69: 915-26.

Liao, Y., G. K. Smyth, and W. Shi. 2014. 'featureCounts: an efficient general purpose program for assigning sequence reads to genomic features', *Bioinformatics*, 30: 923-30.

Lister, R., M. Pelizzola, R. H. Dowen, R. D. Hawkins, G. Hon, J. Tonti-Filippini, J. R. Nery, L. Lee, Z. Ye, Q. M. Ngo, L. Edsall, J. Antosiewicz-Bourget, R. Stewart, V. Ruotti, A. H. Millar, J. A. Thomson, B. Ren, and J. R. Ecker. 2009. 'Human DNA methylomes at base resolution show widespread epigenomic differences', *Nature*, 462: 315-22.

Liu, Y., P. Siejka-Zielinska, G. Velikova, Y. Bi, F. Yuan, M. Tomkova, C. Bai, L. Chen, B. Schuster-Bockler, and C. X. Song. 2019. 'Bisulfite-free direct detection of 5-methylcytosine and 5-hydroxymethylcytosine at base resolution', *Nat Biotechnol*, 37: 424-29.

Locke, W. J., D. Guanzon, C. Ma, Y. J. Liew, K. R. Duesing, K. Y. C. Fung, and J. P. Ross. 2019. 'DNA Methylation Cancer Biomarkers: Translation to the Clinic', *Front Genet*, 10: 1150.

Lopez-Moyado, I. F., A. Tsagaratou, H. Yuita, H. Seo, B. Delatte, S. Heinz, C. Benner, and A. Rao. 2019. 'Paradoxical association of TET loss of function with genome-wide DNA hypomethylation', *Proc Natl Acad Sci U S A*, 116: 16933-42.

Loriot, A., T. Boon, and C. De Smet. 2003. 'Five new human cancer-germline genes identified among 12 genes expressed in spermatogonia', *Int J Cancer*, 105: 371-6.

Loriot, A., E. De Plaen, T. Boon, and C. De Smet. 2006. 'Transient down-regulation of DNMT1 methyltransferase leads to activation and stable hypomethylation of MAGE-A1 in melanoma cells', *J Biol Chem*, 281: 10118-26.

Loriot, A., A. Van Tongelen, J. Blanco, S. Klaessens, J. Cannuyer, N. van Baren, A. Decottignies, and C. De Smet. 2014. 'A novel cancer-germline transcript carrying pro-metastatic miR-105 and TET-targeting miR-767 induced by DNA hypomethylation in tumors', *Epigenetics*, 9: 1163-71.

Lotem, J., D. Netanely, E. Domany, and L. Sachs. 2005. 'Human cancers overexpress genes that are specific to a variety of normal human tissues', *Proc Natl Acad Sci U S A*, 102: 18556-61.

Love, M. I., W. Huber, and S. Anders. 2014. 'Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2', *Genome Biol*, 15: 550.

Luger, K., A. W. Mader, R. K. Richmond, D. F. Sargent, and T. J. Richmond. 1997. 'Crystal structure of the nucleosome core particle at 2.8 A resolution', *Nature*, 389: 251-60.

Lyon, M. F. 1961. 'Gene action in the X-chromosome of the mouse (Mus musculus L.)', *Nature*, 190: 372-3.

Mack, S. C., H. Witt, R. M. Piro, L. Gu, S. Zuyderduyn, A. M. Stutz, X. Wang, M. Gallo, L. Garzia, K. Zayne, X. Zhang, V. Ramaswamy, N. Jager, D. T. Jones, M. Sill, T. J. Pugh, M. Ryzhova, K. M. Wani, D. J. Shih, R. Head, M. Remke, S. D. Bailey, T. Zichner, C. C. Faria, M. Barszczyk, S. Stark, H. Seker-Cin, S. Hutter, P. Johann, S. Bender, V. Hovestadt, T. Tzaridis, A. M. Dubuc, P. A. Northcott, J. Peacock, K. C. Bertrand, S. Agnihotri, F. M. Cavalli, I. Clarke, K. Nethery-Brokx, C. L. Creasy, S. K. Verma, J. Koster, X. Wu, Y. Yao, T. Milde, P. Sin-Chan, J. Zuccaro, L. Lau, S. Pereira, P. Castelo-Branco, M. Hirst, M. A. Marra, S. S. Roberts, D. Fults, L. Massimi, Y. J. Cho, T. Van Meter, W. Grajkowska, B. Lach, A. E. Kulozik, A. von Deimling, O. Witt, S. W. Scherer, X. Fan, K. M. Muraszko, M. Kool, S. L. Pomeroy, N. Gupta, J. Phillips, A. Huang, U. Tabori, C. Hawkins, D. Malkin, P. N. Kongkham, W. A. Weiss, N. Jabado, J. T. Rutka, E. Bouffet, J. O. Korbel, M. Lupien, K. D. Aldape, G. D. Bader, R. Eils, P. Lichter, P. B. Dirks, S. M. Pfister, A. Korshunov, and M. D. Taylor. 2014. 'Epigenomic alterations define lethal CIMP-positive ependymomas of infancy', *Nature*, 506: 445-50.

Macleod, D., J. Charlton, J. Mullins, and A. P. Bird. 1994. 'Sp1 sites in the mouse aprt gene promoter are required to prevent methylation of the CpG island', *Genes Dev*, 8: 2282-92.

Magnon, C., S. J. Hall, J. Lin, X. Xue, L. Gerber, S. J. Freedland, and P. S. Frenette. 2013. 'Autonomic nerve development contributes to prostate cancer progression', *Science*, 341: 1236361.

Majumder, S. 2006. 'REST in good times and bad: roles in tumor suppressor and oncogenic activities', *Cell Cycle*, 5: 1929-35.

Makinen, J. M., K. Laitakari, S. Johnson, R. Makitaro, R. Bloigu, P. Paakko, E. Lappi-Blanco, and R. Kaarteenaho. 2017. 'Histological features of malignancy correlate with growth patterns and patient outcome in lung adenocarcinoma', *Histopathology*, 71: 425-36.

Maksakova, I. A., M. T. Romanish, L. Gagnier, C. A. Dunn, L. N. van de Lagemaat, and D. L. Mager. 2006. 'Retroviral elements and their hosts: insertional mutagenesis in the mouse germ line', *PLoS Genet*, 2: e2.

Malhotra, J., M. Malvezzi, E. Negri, C. La Vecchia, and P. Boffetta. 2016. 'Risk factors for lung cancer worldwide', *Eur Respir J*, 48: 889-902.

Marconett, C. N., B. Zhou, M. Sunohara, T. M. Pouldar, H. Wang, Y. Liu, M. E. Rieger, E. Tran, P. Flodby, K. D. Siegmund, E. D. Crandall, I. A. Laird-Offringa, and Z. Borok. 2017. 'Cross-Species Transcriptome Profiling Identifies New Alveolar Epithelial Type I Cell-Specific Genes', *Am J Respir Cell Mol Biol*, 56: 310-21.

Margueron, R., and D. Reinberg. 2011. 'The Polycomb complex PRC2 and its mark in life', *Nature*, 469: 343-9.

Martienssen, R. 1998. 'Transposons, DNA methylation and gene control', *Trends Genet*, 14: 263-4.

Martin, B. J. E., J. Brind'Amour, A. Kuzmin, K. N. Jensen, Z. C. Liu, M. Lorincz, and L. J. Howe. 2021. 'Transcription shapes genome-wide histone acetylation patterns', *Nat Commun*, 12: 210.

Meir, Z., Z. Mukamel, E. Chomsky, A. Lifshitz, and A. Tanay. 2020. 'Single-cell analysis of clonal maintenance of transcriptional and epigenetic states in cancer cells', *Nat Genet*, 52: 709-18.

Merlo, A., J. G. Herman, L. Mao, D. J. Lee, E. Gabrielson, P. C. Burger, S. B. Baylin, and D. Sidransky. 1995. '5' CpG island methylation is associated with transcriptional silencing of the tumour suppressor p16/CDKN2/MTS1 in human cancers', *Nat Med*, 1: 686-92.

Molaro, A., E. Hodges, F. Fang, Q. Song, W. R. McCombie, G. J. Hannon, and A. D. Smith. 2011. 'Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates', *Cell*, 146: 1029-41.

Monje, M., J. C. Borniger, N. J. D'Silva, B. Deneen, P. B. Dirks, F. Fattahi, P. S. Frenette, L. Garzia, D. H. Gutmann, D. Hanahan, S. L. Hervey-Jumper, H. Hondermarck, J. B. Hurov, A. Kepecs, S. M. Knox, A. C. Lloyd, C. Magnon, J. L. Saloman, R. A. Segal, E. K. Sloan, X. Sun, M. D. Taylor, K. J. Tracey, L. C. Trotman, D. A. Tuveson, T. C. Wang, R. A. White, and F. Winkler. 2020. 'Roadmap for the Emerging Field of Cancer Neuroscience', *Cell*, 181: 219-22.

Mu, D., L. Chen, X. Zhang, L. H. See, C. M. Koch, C. Yen, J. J. Tong, L. Spiegel, K. C. Nguyen, A. Servoss, Y. Peng, L. Pei, J. R. Marks, S. Lowe, T. Hoey, L. Y. Jan, W. R. McCombie, M. H. Wigler, and S. Powers. 2003. 'Genomic amplification and oncogenic properties of the KCNK9 potassium channel gene', *Cancer Cell*, 3: 297-302.

Mulligan, P., T. F. Westbrook, M. Ottinger, N. Pavlova, B. Chang, E. Macia, Y. J. Shi, J. Barretina, J. Liu, P. M. Howley, S. J. Elledge, and Y. Shi. 2008. 'CDYL bridges REST and histone methyltransferases for gene repression and suppression of cellular transformation', *Mol Cell*, 32: 718-26.

Nabel, C. S., H. Jia, Y. Ye, L. Shen, H. L. Goldschmidt, J. T. Stivers, Y. Zhang, and R. M. Kohli. 2012. 'AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation', *Nat Chem Biol*, 8: 751-8.

Naciri, I., M. Laisne, L. Ferry, M. Bourmaud, N. Gupta, S. Di Carlo, A. Huna, N. Martin, L. Peduto, D. Bernard, O. Kirsh, and P. A. Defossez. 2019. 'Genetic screens reveal mechanisms for the transcriptional regulation of tissue-specific genes in normal cells and tumors', *Nucleic Acids Res*, 47: 3407-21.

Nan, X., H. H. Ng, C. A. Johnson, C. D. Laherty, B. M. Turner, R. N. Eisenman, and A. Bird. 1998. 'Transcriptional repression by the methyl-CpG-binding protein MeCP2 involves a histone deacetylase complex', *Nature*, 393: 386-9.

Negrini, S., I. Prada, R. D'Alessandro, and J. Meldolesi. 2013. 'REST: an oncogene or a tumor suppressor?', *Trends Cell Biol*, 23: 289-95.

Neman, J., J. Termini, S. Wilczynski, N. Vaidehi, C. Choy, C. M. Kowolik, H. Li, A. C. Hambrecht, E. Roberts, and R. Jandial. 2014. 'Human breast cancer metastases to the brain display GABAergic properties in the neural niche', *Proc Natl Acad Sci U S A*, 111: 984-9.

Nicholson, A. G., M. S. Tsao, M. B. Beasley, A. C. Borczuk, E. Brambilla, W. A. Cooper, S. Dacic, D. Jain, K. M. Kerr, S. Lantuejoul, M. Noguchi, M. Papotti, N. Rekhtman, G. Scagliotti, P. van Schil, L. Sholl, Y. Yatabe, A. Yoshida, and W. D. Travis. 2022. 'The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015', *J Thorac Oncol*, 17: 362-87.

O'Neill, K. M., R. E. Irwin, S. J. Mackin, S. J. Thursby, A. Thakur, C. Bertens, L. Masala, J. E. P. Loughery, D. G. McArt, and C. P. Walsh. 2018. 'Depletion of DNMT1 in differentiated human cells highlights key classes of sensitive genes and an interplay with polycomb repression', *Epigenetics Chromatin*, 11: 12.

Okano, M., D. W. Bell, D. A. Haber, and E. Li. 1999. 'DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development', *Cell*, 99: 247-57.

Olins, D. E., and A. L. Olins. 2003. 'Chromatin history: our view from the bridge', *Nat Rev Mol Cell Biol*, 4: 809-14.

Ooi, L., and I. C. Wood. 2007. 'Chromatin crosstalk in development and disease: lessons from REST', *Nat Rev Genet*, 8: 544-54.

Paez, J. G., P. A. Janne, J. C. Lee, S. Tracy, H. Greulich, S. Gabriel, P. Herman, F. J. Kaye, N. Lindeman, T. J. Boggon, K. Naoki, H. Sasaki, Y. Fujii, M. J. Eck, W. R. Sellers, B. E. Johnson, and M. Meyerson. 2004. 'EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy', *Science*, 304: 1497-500.

Park, J. W., F. Sahm, B. Steffl, I. Arrillaga-Romany, D. Cahill, M. Monje, C. Herold-Mende, W. Wick, and S. Turcan. 2021. 'TERT and DNMT1 expression predict sensitivity to decitabine in gliomas', *Neuro Oncol*, 23: 76-87.

Peters, J. 2014. 'The role of genomic imprinting in biology and disease: an expanding view', *Nat Rev Genet*, 15: 517-30.

Pineda, C. T., S. Ramanathan, K. Fon Tacer, J. L. Weon, M. B. Potts, Y. H. Ou, M. A. White, and P. R. Potts. 2015. 'Degradation of AMPK by a cancer-specific ubiquitin ligase', *Cell*, 160: 715-28.

Qureshi, I. A., S. Gokhan, and M. F. Mehler. 2010. 'REST and CoREST are transcriptional and epigenetic regulators of seminal neural fate decisions', *Cell Cycle*, 9: 4477-86.

Rando, O. J. 2007. 'Global patterns of histone modifications', *Curr Opin Genet Dev*, 17: 94-9.

Rasmussen, K. D., and K. Helin. 2016. 'Role of TET enzymes in DNA methylation, development, and cancer', *Genes Dev*, 30: 733-50.

Rauch, T. A., X. Zhong, X. Wu, M. Wang, K. H. Kernstine, Z. Wang, A. D. Riggs, and G. P. Pfeifer. 2008. 'High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer', *Proc Natl Acad Sci U S A*, 105: 252-7.

Renaud, S., D. Loukinov, L. Alberti, A. Vostrov, Y. W. Kwon, F. T. Bosman, V. Lobanenkov, and J. Benhattar. 2011. 'BORIS/CTCFL-mediated transcriptional regulation of the hTERT telomerase gene in testicular and ovarian tumor cells', *Nucleic Acids Res*, 39: 862-73.

Riggs, A. D. 1975. 'X inactivation, differentiation, and DNA methylation', *Cytogenet Cell Genet*, 14: 9-25.

———. 1989. 'DNA methylation and cell memory', *Cell Biophys*, 15: 1-13.

Roadmap Epigenomics, Consortium, A. Kundaje, W. Meuleman, J. Ernst, M. Bilenky, A. Yen, A. Heravi-Moussavi, P. Kheradpour, Z. Zhang, J. Wang, M. J. Ziller, V. Amin, J. W. Whitaker, M. D. Schultz, L. D. Ward, A. Sarkar, G. Quon, R. S. Sandstrom, M. L. Eaton, Y. C. Wu, A. R. Pfenning, X. Wang, M. Claussnitzer, Y. Liu, C. Coarfa, R. A. Harris, N. Shoresh, C. B. Epstein, E. Gjoneska, D. Leung, W. Xie, R. D. Hawkins, R. Lister, C. Hong, P. Gascard, A. J. Mungall, R. Moore, E. Chuah, A. Tam, T. K. Canfield, R. S. Hansen, R. Kaul, P. J. Sabo, M. S. Bansal, A. Carles, J. R.

Dixon, K. H. Farh, S. Feizi, R. Karlic, A. R. Kim, A. Kulkarni, D. Li, R. Lowdon, G. Elliott, T. R. Mercer, S. J. Neph, V. Onuchic, P. Polak, N. Rajagopal, P. Ray, R. C. Sallari, K. T. Siebenthall, N. A. Sinnott-Armstrong, M. Stevens, R. E. Thurman, J. Wu, B. Zhang, X. Zhou, A. E. Beaudet, L. A. Boyer, P. L. De Jager, P. J. Farnham, S. J. Fisher, D. Haussler, S. J. Jones, W. Li, M. A. Marra, M. T. McManus, S. Sunyaev, J. A. Thomson, T. D. Tlsty, L. H. Tsai, W. Wang, R. A. Waterland, M. Q. Zhang, L. H. Chadwick, B. E. Bernstein, J. F. Costello, J. R. Ecker, M. Hirst, A. Meissner, A. Milosavljevic, B. Ren, J. A. Stamatoyannopoulos, T. Wang, and M. Kellis. 2015. 'Integrative analysis of 111 reference human epigenomes', *Nature*, 518: 317-30.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, and J. P. Mesirov. 2011. 'Integrative genomics viewer', *Nat Biotechnol*, 29: 24-6.

Rodriguez-Paredes, M., and M. Esteller. 2011. 'Cancer epigenetics reaches mainstream oncology', *Nat Med*, 17: 330-9.

Rousseaux, S., A. Debernardi, B. Jacquiau, A. L. Vitte, A. Vesin, H. Nagy-Mignotte, D. Moro-Sibilot, P. Y. Brichon, S. Lantuejoul, P. Hainaut, J. Laffaire, A. de Reynies, D. G. Beer, J. F. Timsit, C. Brambilla, E. Brambilla, and S. Khochbin. 2013. 'Ectopic activation of germline and placental genes identifies aggressive metastasis-prone lung cancers', *Sci Transl Med*, 5: 186ra66.

Roy, N., and M. Hebrok. 2015. 'Regulation of Cellular Identity in Cancer', *Dev Cell*, 35: 674-84.

Sabari, J. K., B. H. Lok, J. H. Laird, J. T. Poirier, and C. M. Rudin. 2017. 'Unravelling the biology of SCLC: implications for therapy', *Nat Rev Clin Oncol*, 14: 549-61.

Sahin, U., O. Tureci, Y. T. Chen, G. Seitz, C. Villena-Heinsen, L. J. Old, and M. Pfreundschuh. 1998. 'Expression of multiple cancer/testis (CT) antigens in breast cancer and melanoma: basis for polyvalent CT vaccine strategies', *Int J Cancer*, 78: 387-9.

Sahin, U., O. Tureci, H. Schmitt, B. Cochlovius, T. Johannes, R. Schmits, F. Stenner, G. Luo, I. Schobert, and M. Pfreundschuh. 1995. 'Human neoplasms elicit multiple specific immune responses in the autologous host', *Proc Natl Acad Sci U S A*, 92: 11810-3.

Santos-Rosa, H., R. Schneider, A. J. Bannister, J. Sherriff, B. E. Bernstein, N. C. Emre, S. L. Schreiber, J. Mellor, and T. Kouzarides. 2002. 'Active genes are tri-methylated at K4 of histone H3', *Nature*, 419: 407-11.

Saxonov, S., P. Berg, and D. L. Brutlag. 2006. '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.

Scanlan, M. J., C. M. Gordon, B. Williamson, S. Y. Lee, Y. T. Chen, E. Stockert, A. Jungbluth, G. Ritter, D. Jager, E. Jager, A. Knuth, and L. J. Old. 2002. 'Identification of cancer/testis genes by database mining and mRNA expression analysis', *Int J Cancer*, 98: 485-92.

Schlesinger, Y., R. Straussman, I. Keshet, S. Farkash, M. Hecht, J. Zimmerman, E. Eden, Z. Yakhini, E. Ben-Shushan, B. E. Reubinoff, Y. Bergman, I. Simon, and H. Cedar. 2007. 'Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer', *Nat Genet*, 39: 232-6.

Schoenherr, C. J., and D. J. Anderson. 1995. 'The neuron-restrictive silencer factor (NRSF): a coordinate repressor of multiple neuron-specific genes', *Science*, 267: 1360-3.

Schotta, G., M. Lachner, K. Sarma, A. Ebert, R. Sengupta, G. Reuter, D. Reinberg, and T. Jenuwein. 2004. 'A silencing pathway to induce H3-K9 and H4-K20 trimethylation at constitutive heterochromatin', *Genes Dev*, 18: 1251-62.

Schulz, W. A., J. P. Elo, A. R. Florl, S. Pennanen, S. Santourlidis, R. Engers, M. Buchardt, H. H. Seifert, and T. Visakorpi. 2002. 'Genomewide DNA hypomethylation is associated with alterations on chromosome 8 in prostate carcinoma', *Genes Chromosomes Cancer*, 35: 58-65.

Scourzic, L., E. Mouly, and O. A. Bernard. 2015. 'TET proteins and the control of cytosine demethylation in cancer', *Genome Med*, 7: 9.

Shukla, S., J. Cyrta, D. A. Murphy, E. G. Walczak, L. Ran, P. Agrawal, Y. Xie, Y. Chen, S. Wang, Y. Zhan, D. Li, E. W. P. Wong, A. Sboner, H. Beltran, J. M. Mosquera, J. Sher, Z. Cao, J. Wongvipat, R. P. Koche, A.

Gopalan, D. Zheng, M. A. Rubin, H. I. Scher, P. Chi, and Y. Chen. 2017. 'Aberrant Activation of a Gastrointestinal Transcriptional Circuit in Prostate Cancer Mediates Castration Resistance', *Cancer Cell*, 32: 792-806 e7.

Sigalotti, L., S. Coral, G. Nardi, A. Spessotto, E. Cortini, I. Cattarossi, F. Colizzi, M. Altomonte, and M. Maio. 2002. 'Promoter methylation controls the expression of MAGE2, 3 and 4 genes in human cutaneous melanoma', *J Immunother*, 25: 16-26.

Simpson, A. J., O. L. Caballero, A. Jungbluth, Y. T. Chen, and L. J. Old. 2005. 'Cancer/testis antigens, gametogenesis and cancer', *Nat Rev Cancer*, 5: 615-25.

Smallwood, S. A., S. Tomizawa, F. Krueger, N. Ruf, N. Carli, A. Segonds-Pichon, S. Sato, K. Hata, S. R. Andrews, and G. Kelsey. 2011. 'Dynamic CpG island methylation landscape in oocytes and preimplantation embryos', *Nat Genet*, 43: 811-4.

Snijders, A. M., S. Y. Lee, B. Hang, W. Hao, M. J. Bissell, and J. H. Mao. 2017. 'FAM83 family oncogenes are broadly involved in human cancers: an integrative multi-omics approach', *Mol Oncol*, 11: 167-79.

Soares, J., A. E. Pinto, C. V. Cunha, S. Andre, I. Barao, J. M. Sousa, and M. Cravo. 1999. 'Global DNA hypomethylation in breast carcinoma: correlation with prognostic factors and tumor progression', *Cancer*, 85: 112-8.

Sole-Boldo, L., G. Raddatz, S. Schutz, J. P. Mallm, K. Rippe, A. S. Lonsdorf, M. Rodriguez-Paredes, and F. Lyko. 2020. 'Single-cell transcriptomes of the human skin reveal age-related loss of fibroblast priming', *Commun Biol*, 3: 188.

Somerville, T. D. D., Y. Xu, K. Miyabayashi, H. Tiriac, C. R. Cleary, D. Maia-Silva, J. P. Milazzo, D. A. Tuveson, and C. R. Vakoc. 2018. 'TP63-Mediated Enhancer Reprogramming Drives the Squamous Subtype of Pancreatic Ductal Adenocarcinoma', *Cell Rep*, 25: 1741-55 e7.

Song, J., O. Rechkoblit, T. H. Bestor, and D. J. Patel. 2011. 'Structure of DNMT1-DNA complex reveals a role for autoinhibition in maintenance DNA methylation', *Science*, 331: 1036-40.

Sterner, D. E., and S. L. Berger. 2000. 'Acetylation of histones and transcription-related factors', *Microbiol Mol Biol Rev*, 64: 435-59.

Sugano, M., T. Nagasaka, E. Sasaki, Y. Murakami, W. Hosoda, T. Hida, T. Mitsudomi, and Y. Yatabe. 2013. 'HNF4alpha as a marker for invasive mucinous adenocarcinoma of the lung', *Am J Surg Pathol*, 37: 211-8.

Sung, H., J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal, and F. Bray. 2021. 'Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries', *CA Cancer J Clin*, 71: 209-49.

Suzuki, A., H. Makinoshima, H. Wakaguri, H. Esumi, S. Sugano, T. Kohno, K. Tsuchihara, and Y. Suzuki. 2014. 'Aberrant transcriptional regulations in cancers: genome, transcriptome and epigenome analysis of lung adenocarcinoma cell lines', *Nucleic Acids Res*, 42: 13557-72.

Suzuki, T., S. Maeda, E. Furuhata, Y. Shimizu, H. Nishimura, M. Kishima, and H. Suzuki. 2017. 'A screening system to identify transcription factors that induce binding site-directed DNA demethylation', *Epigenetics Chromatin*, 10: 60.

Szyf, M., P. Pakneshan, and S. A. Rabbani. 2004. 'DNA methylation and breast cancer', *Biochem Pharmacol*, 68: 1187-97.

Tahiliani, M., K. P. Koh, Y. Shen, W. A. Pastor, H. Bandukwala, Y. Brudno, S. Agarwal, L. M. Iyer, D. R. Liu, L. Aravind, and A. Rao. 2009. 'Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1', *Science*, 324: 930-5.

Takahashi, K., S. Shichijo, M. Noguchi, M. Hirohata, and K. Itoh. 1995. 'Identification of MAGE-1 and MAGE-4 proteins in spermatogonia and primary spermatocytes of testis', *Cancer Res*, 55: 3478-82.

Thienpont, B., J. Steinbacher, H. Zhao, F. D'Anna, A. Kuchnio, A. Ploumakis, B. Ghesquiere, L. Van Dyck, B. Boeckx, L. Schoonjans, E. Hermans, F. Amant, V. N. Kristensen, K. Peng Koh, M. Mazzone, M. Coleman, T. Carell, P. Carmeliet, and D. Lambrechts. 2016. 'Tumour hypoxia causes DNA hypermethylation by reducing TET activity', *Nature*, 537: 63-68.

Tilman, G., N. Arnoult, S. Lenglez, A. Van Beneden, A. Loriot, C. De Smet, and A. Decottignies. 2012. 'Cancer-linked satellite 2 DNA hypomethylation does

not regulate Sat2 non-coding RNA expression and is initiated by heat shock pathway activation', *Epigenetics*, 7: 903-13.

Travis, W. D., E. Brambilla, A. G. Nicholson, Y. Yatabe, J. H. M. Austin, M. B. Beasley, L. R. Chirieac, S. Dacic, E. Duhig, D. B. Flieder, K. Geisinger, F. R. Hirsch, Y. Ishikawa, K. M. Kerr, M. Noguchi, G. Pelosi, C. A. Powell, M. S. Tsao, I. Wistuba, and W. H. O. Panel. 2015. 'The 2015 World Health Organization Classification of Lung Tumors: Impact of Genetic, Clinical and Radiologic Advances Since the 2004 Classification', *J Thorac Oncol*, 10: 1243-60.

Uhlen, M., E. Bjorling, C. Agaton, C. A. Szigyarto, B. Amini, E. Andersen, A. C. Andersson, P. Angelidou, A. Asplund, C. Asplund, L. Berglund, K. Bergstrom, H. Brumer, D. Cerjan, M. Ekstrom, A. Elobeid, C. Eriksson, L. Fagerberg, R. Falk, J. Fall, M. Forsberg, M. G. Bjorklund, K. Gumbel, A. Halimi, I. Hallin, C. Hamsten, M. Hansson, M. Hedhammar, G. Hercules, C. Kampf, K. Larsson, M. Lindskog, W. Lodewyckx, J. Lund, J. Lundeberg, K. Magnusson, E. Malm, P. Nilsson, J. Odling, P. Oksvold, I. Olsson, E. Oster, J. Ottosson, L. Paavilainen, A. Persson, R. Rimini, J. Rockberg, M. Runeson, A. Sivertsson, A. Skollermo, J. Steen, M. Stenvall, F. Sterky, S. Stromberg, M. Sundberg, H. Tegel, S. Tourle, E. Wahlund, A. Walden, J. Wan, H. Wernerus, J. Westberg, K. Wester, U. Wrethagen, L. L. Xu, S. Hober, and F. Ponten. 2005. 'A human protein atlas for normal and cancer tissues based on antibody proteomics', *Mol Cell Proteomics*, 4: 1920-32.

Uhlen, M., L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen, and F. Ponten. 2015. 'Proteomics. Tissue-based map of the human proteome', *Science*, 347: 1260419.

Van den Eynde, B., O. Peeters, O. De Backer, B. Gaugler, S. Lucas, and T. Boon. 1995. 'A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma', *J Exp Med*, 182: 689-98.

van der Bruggen, P., C. Traversari, P. Chomez, C. Lurquin, E. De Plaen, B. Van den Eynde, A. Knuth, and T. Boon. 1991. 'A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma', *Science*, 254: 1643-7.

Van Tongelen, A., A. Loriot, and C. De Smet. 2017. 'Oncogenic roles of DNA hypomethylation through the activation of cancer-germline genes', *Cancer Lett*, 396: 130-37.

Voigt, P., W. W. Tee, and D. Reinberg. 2013. 'A double take on bivalent promoters', *Genes Dev*, 27: 1318-38.

Walsh, C. P., and T. H. Bestor. 1999. 'Cytosine methylation and mammalian development', *Genes Dev*, 13: 26-34.

Walsh, C. P., J. R. Chaillet, and T. H. Bestor. 1998. 'Transcription of IAP endogenous retroviruses is constrained by cytosine methylation', *Nat Genet*, 20: 116-7.

Wang, C., Y. Gu, K. Zhang, K. Xie, M. Zhu, N. Dai, Y. Jiang, X. Guo, M. Liu, J. Dai, L. Wu, G. Jin, H. Ma, T. Jiang, R. Yin, Y. Xia, L. Liu, S. Wang, B. Shen, R. Huo, Q. Wang, L. Xu, L. Yang, X. Huang, H. Shen, J. Sha, and Z. Hu. 2016. 'Systematic identification of genes with a cancer-testis expression pattern in 19 cancer types', *Nat Commun*, 7: 10499.

Wang, L., J. Zhang, J. Duan, X. Gao, W. Zhu, X. Lu, L. Yang, J. Zhang, G. Li, W. Ci, W. Li, Q. Zhou, N. Aluru, F. Tang, C. He, X. Huang, and J. Liu. 2014. 'Programming and inheritance of parental DNA methylomes in mammals', *Cell*, 157: 979-91.

Weber, J., M. Salgaller, D. Samid, B. Johnson, M. Herlyn, N. Lassam, J. Treisman, and S. A. Rosenberg. 1994. 'Expression of the MAGE-1 tumor antigen is up-regulated by the demethylating agent 5-aza-2'-deoxycytidine', *Cancer Res*, 54: 1766-71.

Weber, M., I. Hellmann, M. B. Stadler, L. Ramos, S. Paabo, M. Rebhan, and D. Schubeler. 2007. 'Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome', *Nat Genet*, 39: 457-66.

Weeden, C. E., Y. Chen, S. B. Ma, Y. Hu, G. Ramm, K. D. Sutherland, G. K. Smyth, and M. L. Asselin-Labat. 2017. 'Lung Basal Stem Cells Rapidly Repair DNA Damage Using the Error-Prone Nonhomologous End-Joining Pathway', *PLoS Biol*, 15: e2000731.

Weisenberger, D. J., K. D. Siegmund, M. Campan, J. Young, T. I. Long, M. A. Faasse, G. H. Kang, M. Widschwendter, D. Weener, D. Buchanan, H. Koh, L. Simms, M. Barker, B. Leggett, J. Levine, M. Kim, A. J. French, S. N. Thibodeau, J. Jass, R. Haile, and P. W. Laird. 2006. 'CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer', *Nat Genet*, 38: 787-93.

Westbrook, T. F., E. S. Martin, M. R. Schlabach, Y. Leng, A. C. Liang, B. Feng, J. J. Zhao, T. M. Roberts, G. Mandel, G. J. Hannon, R. A. Depinho, L. Chin, and S. J. Elledge. 2005. 'A genetic screen for candidate tumor suppressors identifies REST', *Cell*, 121: 837-48.

Whitehurst, A. W. 2014. 'Cause and consequence of cancer/testis antigen activation in cancer', *Annu Rev Pharmacol Toxicol*, 54: 251-72.

Wilkinson, M. F. 2015. 'Evidence that DNA methylation engenders dynamic gene regulation', *Proc Natl Acad Sci U S A*, 112: E2116.

Williams, K., J. Christensen, and K. Helin. 2011. 'DNA methylation: TET proteins-guardians of CpG islands?', *EMBO Rep*, 13: 28-35.

Woloszynska-Read, A., P. Mhawech-Fauceglia, J. Yu, K. Odunsi, and A. R. Karpf. 2008. 'Intertumor and intratumor NY-ESO-1 expression heterogeneity is associated with promoter-specific and global DNA methylation status in ovarian cancer', *Clin Cancer Res*, 14: 3283-90.

Wu, H., and Y. Zhang. 2014. 'Reversing DNA methylation: mechanisms, genomics, and biological functions', *Cell*, 156: 45-68.

Wu, X., and Y. Zhang. 2017. 'TET-mediated active DNA demethylation: mechanism, function and beyond', *Nat Rev Genet*, 18: 517-34.

Yamaguchi, Y., I. Murai, K. Goto, S. Doi, H. Zhou, G. Setsu, H. Shimatani, H. Okamura, T. Miyake, and M. Doi. 2021. 'Gpr19 is a circadian clock-controlled orphan GPCR with a role in modulating free-running period and light resetting capacity of the circadian clock', *Sci Rep*, 11: 22406.

Yin, Y., E. Morgunova, A. Jolma, E. Kaasinen, B. Sahu, S. Khund-Sayeed, P. K. Das, T. Kivioja, K. Dave, F. Zhong, K. R. Nitta, M. Taipale, A. Popov, P. A. Ginno, S. Domcke, J. Yan, D. Schubeler, C. Vinson, and J. Taipale. 2017. 'Impact of cytosine methylation on DNA binding specificities of human transcription factors', *Science*, 356.

Yoder, J. A., C. P. Walsh, and T. H. Bestor. 1997. 'Cytosine methylation and the ecology of intragenomic parasites', *Trends Genet*, 13: 335-40.

Yokoe, T., F. Tanaka, K. Mimori, H. Inoue, T. Ohmachi, M. Kusunoki, and M. Mori. 2008. 'Efficient identification of a novel cancer/testis antigen for immunotherapy using three-step microarray analysis', *Cancer Res*, 68: 1074-82.

You, J. S., and P. A. Jones. 2012. 'Cancer genetics and epigenetics: two sides of the same coin?', *Cancer Cell*, 22: 9-20.

Yu, G., L. G. Wang, Y. Han, and Q. Y. He. 2012. 'clusterProfiler: an R package for comparing biological themes among gene clusters', *OMICS*, 16: 284-7.

Yu, K. H., C. Zhang, G. J. Berry, R. B. Altman, C. Re, D. L. Rubin, and M. Snyder. 2016. 'Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features', *Nat Commun*, 7: 12474.

Zeng, Q., I. P. Michael, P. Zhang, S. Saghafinia, G. Knott, W. Jiao, B. D. McCabe, J. A. Galvan, H. P. C. Robinson, I. Zlobec, G. Ciriello, and D. Hanahan. 2019. 'Synaptic proximity enables NMDAR signalling to promote brain metastasis', *Nature*, 573: 526-31.

Zhang, L., X. Lu, J. Lu, H. Liang, Q. Dai, G. L. Xu, C. Luo, H. Jiang, and C. He. 2012. 'Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA', *Nat Chem Biol*, 8: 328-30.

Zhang, W., D. Klinkebiel, C. J. Barger, S. Pandey, C. Guda, A. Miller, S. N. Akers, K. Odunsi, and A. R. Karpf. 2020. 'Global DNA Hypomethylation in Epithelial Ovarian Cancer: Passive Demethylation and Association with Genomic Instability', *Cancers (Basel)*, 12.

Zhang, Y., R. Jurkowska, S. Soeroes, A. Rajavelu, A. Dhayalan, I. Bock, P. Rathert, O. Brandt, R. Reinhardt, W. Fischle, and A. Jeltsch. 2010. 'Chromatin methylation activity of Dnmt3a and Dnmt3a/3L is guided by interaction of the ADD domain with the histone H3 tail', *Nucleic Acids Res*, 38: 4246-53.

Zhang, Y., R. Wang, Y. Li, Y. Pan, H. Hu, Y. Zhang, H. Li, L. Shen, Y. Yu, Y. Sun, and H. Chen. 2015. 'Negative Thyroid Transcription Factor 1 Expression Defines an Unfavorable Subgroup of Lung Adenocarcinomas', *J Thorac Oncol*, 10: 1444-50.

Zhao, C. M., Y. Hayakawa, Y. Kodama, S. Muthupalani, C. B. Westphalen, G. T. Andersen, A. Flatberg, H. Johannessen, R. A. Friedman, B. W. Renz, A. K. Sandvik, V. Beisvag, H. Tomita, A. Hara, M. Quante, Z. Li, M. D. Gershon, K. Kaneko, J. G. Fox, T. C. Wang, and D. Chen. 2014. 'Denervation suppresses gastric tumorigenesis', *Sci Transl Med*, 6: 250ra115.

Zhou, W., H. Q. Dinh, Z. Ramjan, D. J. Weisenberger, C. M. Nicolet, H. Shen, P. W. Laird, and B. P. Berman. 2018. 'DNA methylation loss in late-replicating domains is linked to mitotic cell division', *Nat Genet*, 50: 591-602.

Zhou, W., M. Y. Fong, Y. Min, G. Somlo, L. Liu, M. R. Palomares, Y. Yu, A. Chow, S. T. O'Connor, A. R. Chin, Y. Yen, Y. Wang, E. G. Marcusson, P. Chu, J. Wu, X. Wu, A. X. Li, Z. Li, H. Gao, X. Ren, M. P. Boldin, P. C. Lin, and S. E. Wang. 2014. 'Cancer-secreted miR-105 destroys vascular endothelial barriers to promote metastasis', *Cancer Cell*, 25: 501-15.

Zuber, V., C. N. Marconett, J. Shi, X. Hua, W. Wheeler, C. Yang, L. Song, A. M. Dale, M. Laplana, A. Risch, A. Witoelar, W. K. Thompson, A. J. Schork, F. Bettella, Y. Wang, S. Djurovic, B. Zhou, Z. Borok, H. F. van der Heijden, J. de Graaf, D. Swinkels, K. K. Aben, J. McKay, R. J. Hung, H. Bikeboller, V. L. Stevens, D. Albanes, N. E. Caporaso, Y. Han, Y. Wei, M. A. Panadero, J. I. Mayordomo, D. C. Christiani, L. Kiemeney, O. A. Andreassen, R. Houlston, C. I. Amos, N. Chatterjee, I. A. Laird-Offringa, I. G. Mills, and M. T. Landi. 2016. 'Pleiotropic Analysis of Lung Cancer and Blood Triglycerides', *J Natl Cancer Inst*, 108.

Zylicz, J. J., A. Bousard, K. Zumer, F. Dossin, E. Mohammad, S. T. da Rocha, B. Schwalb, L. Syx, F. Dingli, D. Loew, P. Cramer, and E. Heard. 2019. 'The Implication of Early Chromatin Changes in X Chromosome Inactivation', *Cell*, 176: 182-97 e23.