Dissertation

submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Diplom-Biologin Katharina Heilmann

born in Schweinfurt, Germany

Date of oral examination:……

Epigenetic characterization of the C3(1) SV40T mouse model of human breast cancer

Referees: PD Dr. Karin Müller-Decker

PD Dr. Dieter Weichenhan

Contributions

Several sections of this work are part of a manuscript *Heilmann et al.* (submitted to *EMBO Molecular Medicine*) with the title "**Genome-wide Screen for Differentially Methylated Long Noncoding RNAs identifies *Esrp2* and lncRNA *Esrp2-as* as Coordinately Expressed from a Bidirectional Promoter and Regulated by Enhancer DNA Methylation". I**n particular, the introduction chapter 1.1.3.2, the Material and Methods chapters 3.4, 3.5, 3.6.1, 3.7, 3.8, 3.12, 3.13, 3.14, the Results chapter 4.3 and the Discussion chapter 5.3 contain suggestions and corrections from co-authors. The corresponding text passages are marked in blue with underline.

Next generation sequencing and analysis of sequencing read quality was conducted by the DKFZ Genomic and Proteomics Core Facility. The Core Facility also prepared sequencing libraries for MCIp samples as indicated in chapter 3.4.

Dr. Yassen Assenov calculated DMRs for human subtypes from 450k data used for comparison of DNA methylation similarity between human breast cancer (from TCGA data) and C3(1) mouse DMRs and prepared the respective Figure 4-12.

Declaration

Declarations according to § 8 (3) b) and c) of the doctoral degree regulations:

a) I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those expressly indicated,

b) I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

Heidelberg

_____

(Katharina Heilmann)

Every journey starts with one small step:

To all the people that helped me on the way!

Table of Contents

Summary

Breast cancer is a heterogeneous disease, and various subtypes have been defined at the level of gene expression and epigenetic modifications, such as DNA methylation. Epigenetic alterations are attractive candidates for the development of novel biomarkers or as targets for new therapeutic approaches. Mouse models allow monitoring of tumor development from early time points of initiation to final stages of tumorigenesis, but are often poorly characterized with respect to alterations of the epigenetic landscape. Therefore, the aim of this thesis was to generated genome-wide profiles of DNA methylation and histone modifications for the C3(1)SV40TAg (C3(1)) mouse model of basal-like breast cancer and to investigate the epigenetic regulation of long noncoding RNAs.

Using a genome-wide screen with Methyl CpG Immunoprecipitation followed by next generation sequencing, we identified several thousand regions with recurrent methylation alterations at different stages of C3(1) tumorigenesis. Differentially methylated genes pointed towards a luminal progenitor as tumor cell of origin in the C3(1) model, and we confirmed a link between DNA methylation and gene expression for five of these genes. Comparisons at the level of promoter methylation revealed general similarity of the C3(1) methylome to human breast cancer.

Generation of a chromatin map of the C3(1) model from four histone modifications (H3K4me3, H3K4me1, H3K27ac, H3K27me3) allowed an integrative investigation of methylation changes at breast tissue-specific enhancer regions. We linked tissue-specific alterations of chromatin states in combination with DNA methylation to changes in the expression of genes with importance for mammary gland development and breast cancer. These results unveiled a potential involvement of transcription factors Etv4 (ETS variant 4) and Runx1 (runt related transcription factor 1) in C3(1) tumorigenesis that might help to understand tumor development in basal-like breast cancer.

An emerging theme in epigenetic research is the capacity of long noncoding RNAs (lncRNAs) to modulate gene expression by recruitment of gene-silencing or activating complexes. Since regulation of lncRNAs expression is poorly characterized, we investigated DNA methylation changes during carcinogenesis at lncRNA promoters and their influence on neighboring protein-coding genes. Exemplarily, we demonstrated coordinated overexpression of *Esrp2* (*Epithelial splicing regulatory protein 2*) and the lncRNA *Esrp2-as* (*Esrp2-antisense*) in C3(1) tumors that was inversely correlated with DNA methylation levels. Knockdown and overexpression of the transcripts did not provide evidence for reciprocal regulation of transcript expression. In contrast, luciferase reporter assays suggested that co-expression of both transcripts is controlled by differential methylation at a common enhancer region. These results are of clinical relevance as high levels of *ESRP2* expression in human breast cancer are linked to unfavorable prognosis.

Zusammenfassung

Brustkrebs ist eine heterogene Erkrankung und basierend auf Unterschieden in der Genexpression oder von epigenetischen Modifikationen, wie zum Beispiel von DNA Methylierung, wurden verschiedene Subtypen definiert. Epigenetische Veränderungen sind attraktive Kandidaten für die Entwicklung neuer Biomarker oder neuer Ansätze zur gezielten Krebstherapie. Der Einsatz von Mausmodellen ermöglicht es dabei, die gesamte Tumorentwicklung von Beginn an, bis zu den finalen Stadien zu beobachten. Allerdings sind Mausmodelle in Bezug auf epigenetische Veränderungen bisher unzureichend charakterisiert. Deshalb war es das Ziel dieser Arbeit, genomweite Profile für Veränderungen der DNA Methylierung und von Histonmodifikationen im C3(1)SV40TAg (C3(1)) Modell für basalen Brustkrebs zu erstellen, sowie die epigenetische Regulation von langen nicht-kodierenden RNAs zu untersuchen.

Mithilfe von Methyl-CpG-Immunpräzipitation gefolgt von Sequenzierung der angereicherten Fragmente haben wir genomweit mehrere tausend Regionen identifiziert, deren DNA Methylierung sich über verschiedene Stadien der C3(1) Tumorentwicklung konsistent verändert hat. Die von differenzieller Methylierung betroffenen Gene ließen darauf schließen, dass eine Population von luminalen Vorläuferzellen den Ausgangspunkt der Tumorentstehung im C3(1) Modell darstellen könnte. Bei fünf von diesen Genen konnten wir eine Verbindung zwischen DNA Methylierung und Genexpression bestätigen. Der Vergleich der Promoter Methylierung ergab außerdem, dass eine generelle Ähnlichkeit zwischen dem C3(1) Methylom und dem von humanem Brustkrebs besteht.

Um eine Karte der Chromatinlandschaft zu erstellen, wurden vier Histonmodifikationen (H3K4me3, H3K4me1, H3K27ac, H3K27me3) auf genomweiter Ebene analysiert. Dies erlaubte die integrative Untersuchung von Veränderungen in der Methylierung an brustgewebsspezifischen Enhancer-Regionen. Dabei fanden wir einen Zusammenhang zwischen Genexpression und gewebstypischen epigenetischen Veränderungen, insbesondere bei Genen, die eine Rolle bei der Entwicklung der Brust und von Brustkrebs spielen. Ebenso konnte ein potenzieller Einfluss der Transkriptionsfaktoren *Etv4* (engl. *ETS variant 4*) und *Runx1* (engl. *Runt related transcription factor 1*) für die C3(1) Tumorigenese identifiziert werden. Diese Ergebnisse könnten dabei helfen die Tumorentstehung in basalem Brustkrebs besser zu verstehen.

Ein Thema von zunehmender Bedeutung für die epigenetische Forschung ist die Fähigkeit langer nicht-kodierender RNAs (lncRNAs), aktivierende oder hemmende Komplexe an die DNA zu rekrutieren, um dadurch die Genexpression zu regulieren. Da die Regulation der Expression für lncRNAs schlecht charakterisiert ist, haben wir krebsspezifische Veränderungen in der Methylierung von lncRNA Promotoren untersucht und wie diese benachbarte protein-kodierende Gene beeinflussen. In diesem Zusammenhang bestätigten wir exemplarisch die koordinierte

Überexpression von *Esrp2* (engl. *Epithelial splicing regulators protein2*) und der lncRNA *Esrp2-as* (engl. *Esrp2-antisense*) in C3(1) Tumoren. Expressionslevel waren dabei invers korreliert mit DNA Methylierung. Knockout sowie gezielte Überexpression der Transkripte erbrachten keine Hinweise für eine wechselseitige Regulation der Expression. Dagegen konnten Untersuchungen mit Luciferasereportern die Hypothese stützten, dass die Ko-Expression beider Transkripte durch differenzielle Methylierung einer gemeinsamen Enhancer-Region kontrolliert wird. Diese Ergebnisse haben auch klinische Relevanz, da hohe Expression von *ESRP2 in* humanem Brustkrebs mit einer ungünstigen Prognose assoziiert ist.

List of abbreviations

450K	Infinium HumanMethylation 450 bead Chip array
5caC	5-carboxylcytosine
5fC	5-formylcytosine
5hmC	5-hydroxymethylcytosine
5mC	5-methylcytosine
AML	Acute myeloid leukemia
BT DNA	Bisulfite treated DNA
C3(1)	C3(1) SV40TAg mouse model of breast cancer
CAGE	Cap analysis of gene expression
cDNA	Complementary DNA
CGI	CpG island
ChIP	Chromatin immunoprecipitation
CLDN4	Claudin4
CpG	Cytosine-guanine dinucleotide
CT	Cycle threshold
DAC	Decitabine (5-aza-2'-deoxycytosine)
DCIS	Ductal carcinoma in situ
DMR	Differentially methylated region
DMSO	Dimethylsulfoxide
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
EGR1	Early growth response 1
ELF3	E74 like transcription factor
ESRP2	Epithelial splicing regulatory protein 2
ESRP2-AS	Epithelial splicing regulatory protein 2 –antisense
ETOH	Ethanol
ETS	E26 transformation factor
ETV4	ETS variant 4
FANTOM	Functional ANnoTation Of the Mammalian genome
GFP	Green fluorescent protein
GSN	Gelsolin
H3K27ac	Histone 3 acetylated lysine 27
H3K27me3	Histone 3 trimethylated lysine 27
H3K4me1	Histone 3 monomethylated lysine 4
H3K4me3	Histone 3 trimethylated lysine 4
HDAC	Histone deacetylase
HOXA9	Homeobox A9
IDC	Invasive ductal carcinoma
IGFBP6	Insulin like growth factor binding protein 6
LNA	Locked nuceic acid
lncRNA	Long noncoding RNA
MALDI-TOF	Matrix assisted laser desorption ionization time-of-flight
MBD	Methyl CpG binding domain
MCIp	Methl CpG binding protein immunoprecipitation
MDS	Myelodysplastic syndrome

MEC	Mammary epithelial cell
MG	Mammary gland
miRNA	Micro RNA
NGS	Next generation sequencing
PBS	Phosphate buffered saline
PCA	Principal Component Analysis
PCR	Polymerase chain reaction
PKP3	Plakophilin 3
RNA	Ribonucleic acid
RT-qPCR	Reverse transcription quantitative PCR
RUNX1	Runt related transcription factor 1
SD	Standard deviation
SEM	Standard error of the mean
SEQ	Next generation sequencing
TCGA	The Cancer Genome Atlas
TG	Transgenic
TSS	Transcription start sit
UNC5B	Unc-5 netrin receptor b
UTR	Untranslated region
WT	Wildtype

1. Introduction

1.1 Epigenetics

Despite the genome being the blueprint for living cells, the DNA sequence is not the only feature to pass on information to the next generation of cells. An additional level of regulation is encoded by epigenetic traits, with the term ‘epigenetics’ being based on the Greek word ‘epi’, meaning ‘upon’, ‘above’, ‘over’ -genetics. Therefore, epigenetics is defined as heritable changes that influence which particular genes are expressed at a specific time, but are not encoded by the sequence of nucleotides [1]. With respect to the different levels of gene regulation, three main mechanisms are considered, namely DNA methylation, posttranslational histone modifications and the influence of noncoding RNAs.

1.1.1 DNA methylation

DNA methylation was the first epigenetic mark to be discovered and occurs at cytosines (5methylcytosine, 5mC) mainly in the context of CpG dinucleotides. The distribution of CpGs is not evenly spread through the genome, but accumulates in regions of high CpG density, so called CpG islands (CGIs) [2] (Figure 1-1). The regions that flank the CGIs are defined as shores and extend about 0.2-2kb up- and downstream of the CGIs [3]. While CGIs in the promoter of genes are generally unmethylated in normal cells, thus allowing expression of the respective gene, hypermethylation of promoter CGIs in cancer leads to gene silencing [4]. In contrast, CpGs located in genic or intergenic regions that also include repetitive sequences, are generally methylated in normal cells, but are targeted by global hypomethylation during carcinogenesis [4]. This was associated with genomic instability in cancer [4]. Variable methylation at CGI shores was also described to influence gene expression levels, in particular tissue-specific expression [3], while methylation in exonic regions was linked to splicing of RNA rather than overall transcript levels [5, 6]. In addition, a regulatory role of DNA methylation can be exerted by more distal elements, namely enhancers, which are less CpG dense than the CGIs and are highly tissue-specific in their function [7].

The exact mechanism how DNA methylation regulates gene expression is not well understood, but in many cases, DNA methylation is thought to decrease gene expression by blocking the binding of transcription factors such as for n-MYC or YY-1 [8, 9]. Whether the absence of 5mC is then the cause or the consequence of transcription factor binding is still debated. Some cases are described where DNA methylation ameliorates binding of proteins,

such as for CEBPA [10] or even is required for DNA binding, as known for methyl-CpG binding domain (MBD) proteins, such as MECP2 or MBD1-4 [11].

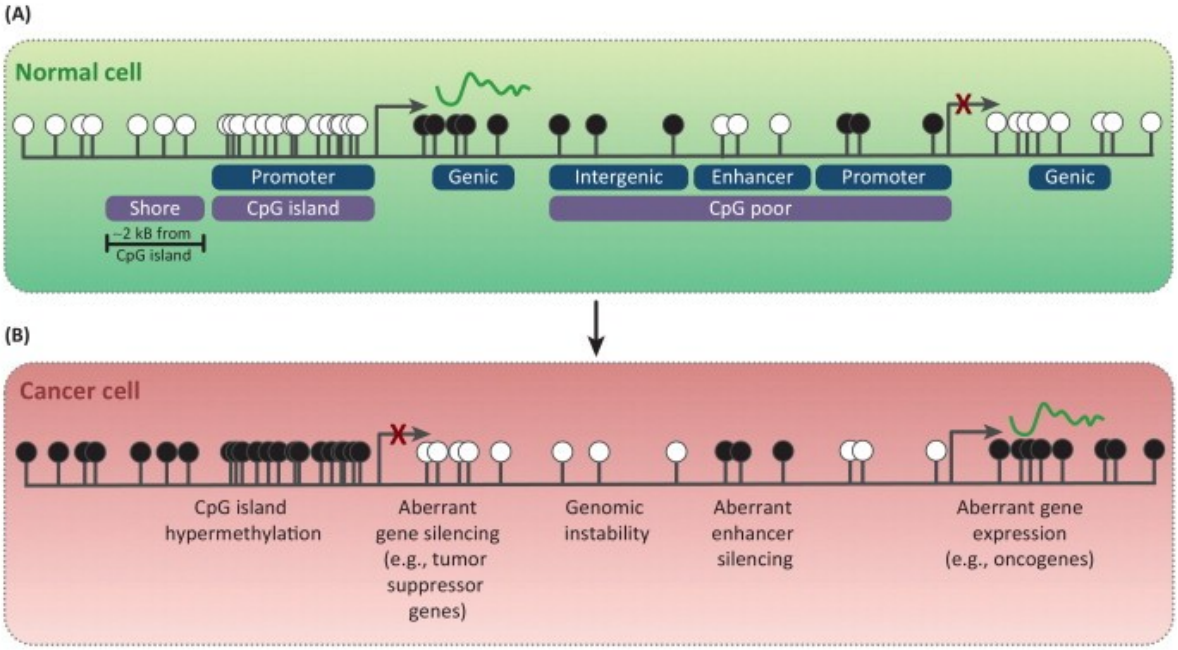


Figure 1-1 Schematic representation of the distribution of methylated and unmethylated CpGs in the genomic context. Influence of DNA methylation on gene expression in normal cells A) and the effects of deregulated DNA methylation in cancer cells B). White circle: unmethylated CpG; Black circle: methylated CpG. Scheme taken from [2].

1.1.1.1 Establishment of DNA methylation patterns

There are three main enzymes responsible for establishing DNA methylation, the maintenance DNA methyltransferase 1 (DNMT1), which binds preferentially to hemimethylated DNA during the replication process [12] and the *de novo* methylating enzymes DNMT3A and DNMT3B [13]. Knockout of DNMT1 and DNMT3A/3B in mice results in embryonic or postnatal lethality, which further underlines the importance of these enzymes and of correct methylation patterns for mammalian development [13, 14]. The mechanism for active demethylation is still more debated, but involves a step of hydroxylation to 5-hydroxymethylcytosine (5hmC) catalyzed by the ten-eleven translocation (TET) dioxygenase enzymes (TET1-3), followed by further oxidation to 5-formyl-(5fC) and 5-carboxyl-cytosine (5caC) [15-17]. By means of the base excision mismatch repair machinery, the thymine DNA glycosylase (TDG) then removes the modified cytidine base leaving an abasic site that is substituted for an unmodified cytosine [17] (Figure 1-2). In addition, the existence of a decarboxylase enzyme is postulated, which could directly remove the carboxyl residue and

restore the unmodified cytosine, though experimental data is still inconclusive in this context [18].

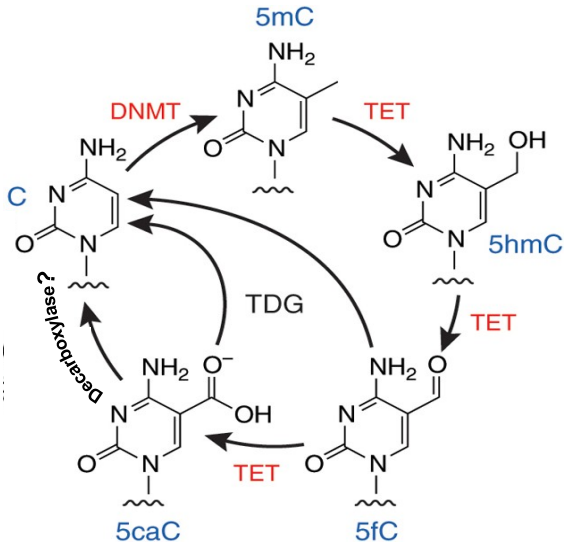


Figure 1-2 Cycle of DNA methylation and demethylation. Enzymes involved in this cycle include DNA methyltransferase (DNMT), TET dioxygenases (TET) and thymine DNA glycosylase (TDG). Demethylation of 5mC involves several oxidation steps via 5hmC, 5fC, and 5caC. Modified from [18]

1.1.1.2 DNA methylation in cancer

Many studies in different types of cancers provided evidence for aberrant methylation in specific genes, which are silenced by hypermethylation in CGI promoters, such as for *CDKN2A (cyclin dependent kinase inhibitor 2a), MLH1 (mutL homolog 1), BRCA1 (breast cancer associated 1)*, and *VHL (von Hippel-Lindau tumor suppressor*) [19]. The identification of such events suggests that several candidates hold promise as potential biomarkers, e.g. for early diagnosis (*GSTP1* (*glutathione S-transferase pi 1*) in prostate cancer), prognosis (*DAPK1 (death associated protein kinase 1)* in brain cancers), or to predict response to therapy (*MGMT (O-6-methylguanine-DNA methyltransferase*) in gliomas) [19]. Since overexpression of DNMTs is a frequent event in many cancers, such as in lung, breast, stomach, or colon cancer [20], and also mutations were detected in hematological cancers and other diseases [21], the inhibition of these enzymes is considered as an attractive therapeutic approach. The DNMT inhibitor Decitabine (Dac, 5-aza-2'-deoxycytidine) is a cytidine analogue, which is incorporated into newly synthesized DNA, where it covalently binds DNMTs and leads to the depletion of the methyltransferase [20]. Loss of the enzyme prevents the methylation of the DNA during cell division, thus preventing the establishment of methylation patterns by passive demethylation. Dac treatment already received FDA-approval for use in the treatment of myelodysplastic syndrome (MDS) and acute myeloid leukemia (AML) [20].

1.1.2 Histone modifications

Another level of epigenetic regulation is encoded by posttranslational modifications of proteins that support the organization of the genome into sets of transcriptional active chromatin (euchromatin) and inactive chromatin (heterochromatin) [22]. The responsible proteins are called histones and form a nucleosome consisting of an octamer of each two H2A, H2B, H3 and H4 histone molecules together with 147bp of DNA wrapped around the octamer [23]. The histone tails that protrude from the nucleosome can be modified at specific amino acids, mainly lysines (K) and arginines (R), and carry many different modifications such as phosphorylation, methylation, acetylation, ubiquitinylation, sumoylation, or ADP-ribosylation (reviewed in [22]). The deposition of these marks is a tightly controlled process, which applies the interaction of enzymes that either place or remove specific residues at the histone tails. In this process, histone acetyl transferases (HATs) catalyze the deposition of acetyl modifications to histone tails from acetyl CoA, and histone deactylases (HDACs) remove them. Since lysines can be methylated to a varying degree from mono- (me1), over di- (me2) to trimethylation (me3), sometimes more than one enzyme is required to catalyze the different steps. For example, while EZH2 (Enhancer of zest homolog 2) methylates H3K27 up to the trimethylated state, SET7/9 (SET domain containing lysine methyltransferase 7/9) only catalyzes the formation of H3K4me1, and further methylation to H3K4me3 requires activity of the MLL (Mixed lineage leukemia)-family of methyltransferases [22, 24, 25]. At some histone tail positions different marks were observed, such as methylation or acetylation at H3K27 or H3K9, and for others only one type of modification was described so far, for example phosphorylation of histone 3 at serine 10 (H3S10) [24] (Figure 1-3).

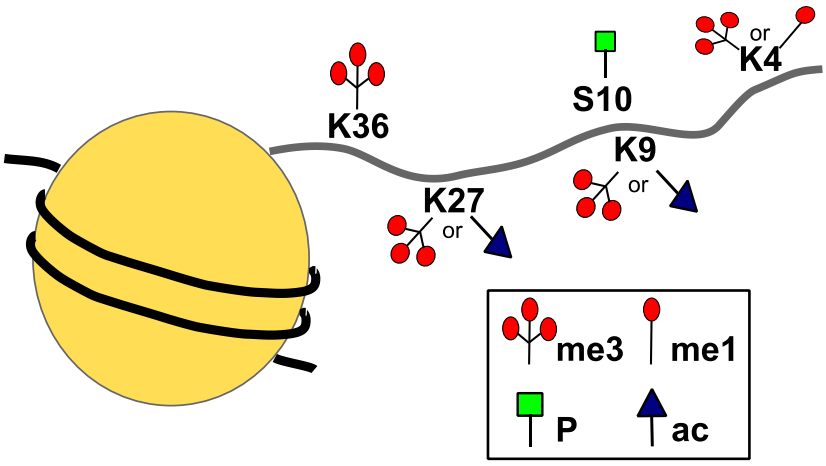


Figure 1-3 Selected positions for histone modifications on the tail of histone H3.

Depending on their particular function specific modifications are found at different genomic locations. The respective genomic sites were determined by chromatin immunoprecipitation followed by next generation sequencing (ChIP-seq) and in this context the modifications of the H3 tail were intensely studied. H3K4me1 is found at enhancers and H3K4me3 is a mark for active transcription found mostly at transcription start sites (TSS) [26, 27]. H3K27ac is another active mark, which is found both at enhancers and promoters [28, 29], and H3K36me3 covers the complete length of transcribed genes [26]. In contrast, H3K27me3 is a repressive mark found genome-wide in regions of heterochromatin. When found in combination with H3K4me3, it marks bivalent promoters that are in a poised state. This is characteristic for embryonic stem cells to maintain a pluripotency status [30]. Since some modifications are capable to influence the deposition of other marks on the same or on neighboring nucleosomes, this crosstalk between modifications reveals an additional level of complexity for regulation [24].

1.1.2.1 Characterization of chromatin states

The diversity of different histone marks suggests that functional implications might be translated not only by one specific, but by several different modifications or a combination thereof. Based on the number of different marks that are known, multiple combinations are possible in theory, but only specific combinations might actually occur in living cells and even fewer might have a distinct functional meaning [31]. For example, H3K4me3 marks active promoters when co-occurring with H3K27ac [28], but represents a poised promoter together with the H3K27me3 mark [30]. The two marks for H3K27 are mutually exclusive [28]. Thus, it is necessary to characterize the combination of different histone marks and to define the

specific biological role of chromatin states to crack the 'histone code' [32, 33]. Several studies defined such chromatin states according to patterns of histone modifications and applied different approaches to develop models ranging from 4 to 51 different states, depending on the nature and number of the studied marks [31, 34-36]. Ernst and Kellis used a multivariate hidden Markov model (HMM) to define 15 states according to the occurrence of nine histone modifications determined by ChIP-seq in nine cell types and defined functional categories such as promoters, enhancers or polycomb-repressed regions [35]. The publicly available ChromHMM algorithm annotates each genomic region with a specific state for different cell types and was used by several studies including the ENCODE (Encyclopedia of DNA elements) consortium [37-40]. Irrespective of the approach used for the definition of different states, the real challenge is their functional characterization. This challenge is faced by increasing number of studies in this area [32].

1.1.2.2 Histone modifications in cancer

In contrast to DNA methylation, alterations in histone modifications during cancer development with respect to their genomic locations and affected genes are not well investigated. Fraga *et al*. discovered that many cancer cell lines and also tumors have reduced levels of acetylation at K16 and trimethylation at K20 for histone H4, in particular at repetitive sequences [41]. Other evidence for changes of histone modifications in cancer is provided rather indirectly, as several of the histone modifying enzymes were found to be differentially expressed in various cancers [24]. The methyltransferase EZH2 is located in the multi protein polycomb repressor complex 2 (PRC2) and is frequently overexpressed, but can also be downregulated in tumors, suggesting a role either as an oncogene or a tumor suppressor [42]. Several other members of the PRC2 also harbor mutations or structural aberrations in cancer that cause disrupted patters of H3K27me3. Furthermore, several of the 18 HDAC proteins known in humans were shown to be overexpressed in cancers, such as in gastric, pancreatic, lung and colon cancer [43]. Thus, HDAC inhibitors are attractive candidates as anticancer drugs. Vorinostat and romidepsin already received FDA approval for rare cutaneous T cell lymphoma and other hematological diseases, while other inhibitors that target histone modifying enzymes are still tested for their applicability in the clinics [19].

1.1.3 Noncoding RNAs

The third major mechanism of epigenetic regulation for gene expression is contributed by noncoding RNAs, meaning RNAs, which are not translated into proteins. There are two main categories of noncoding RNAs, which are defined by their length into small and long noncoding RNAs (lncRNA).

1.1.3.1 MicroRNAs

Among the small RNAs the microRNAs (miRNAs) are best characterized with respect to their function. Mature miRNAs with a length of ≈22 nucleotides bind to the 3' UTR of mRNAs, thereby marking the bound transcripts for degradation or inhibiting translation into proteins. Similar to mRNAs, miRNA expression is deregulated in cancer and deregulation can be controlled by DNA methylation [44].

1.1.3.2 Long noncoding RNAs (LncRNAs)

lncRNAs can range from 200bp to several kb or even longer and have not been assigned into functional subgroups yet. The role of lncRNAs can range from transcriptional regulation to formation of structural scaffolds, as well as involvement in various cellular processes from apoptosis to differentiation. Studies of both developmental processes and diseases revealed that deregulated expression of lncRNAs can influence expression or stability of protein-coding RNAs [45, 46]. With respect to specificity, lncRNAs have been shown to control expression of genes in a localized, gene specific fashion, e.g. the repression of cell cycle regulator p21 by the lncRNA *FAL1* (*focally amplified lncRNA on chromosome 1*) [47], as well as by targeting large chromosomal regions, e.g. X-chromosome inactivation by *XIST* (*X inactive specific transcript*) (reviewed in [48]). LncRNAs further act as host genes encoding micro RNAs [49], or as a decoy for miRNAs, thus preventing miRNA binding to protein-coding RNAs [50]. Several studies have demonstrated that lncRNAs influence DNA methylation or the chromatin landscape by interacting with modifiers of epigenetic marks, thereby recruiting these modifiers to specific DNA loci and resulting in gene silencing or activation [47, 51-56].

The identification and functional evaluation of lncRNAs has become an area of substantial scientific interest. Screening strategies include the analysis of differential expression using lncRNA specific microarrays, or tiling arrays that cover genomic regions of interest, such as the *Hox* cluster, for *de novo* identification of lncRNAs [57, 58]. In addition, chromatin marks for active transcription (H3K4me3, H3K36me3) were combined with tiling microarray

analysis to locate novel lncRNAs [59]. Nowadays, RNA-seq allows the detection of lncRNAs at a genome-wide unbiased scale [57, 58].

1.2 Breast cancer

Breast cancer was the most frequent cancer in women worldwide in 2012 with 1.7 Million newly diagnosed cases and more than 520.000 deaths [60]. Despite the common name, breast cancer is a heterogeneous disease consisting of different subtypes requiring the use of different classification criteria, which include histology, hormone receptor status or molecular gene expression profiles [61]. The heterogeneity of breast cancer subtypes is partly attributable to the development and organization of the breast with respect to its different cell types. The two main structures in the breast are the terminal lobular units and the ducts that connect them. Both structures are lined in the lumen by a single epithelial layer of luminal cells, which are coated by a layer of basal or myoepithelial cells (Figure 1-4). The basement membrane separates these cells from the surrounding tissue, and is not penetrated by cells of premalignant or early malignant stages such as mammary intraepithelial neoplasia (MIN) or ductal carcinoma in situ (DCIS) (Figure 1-4). This feature distinguishes DCIS from invasive carcinomas that cross this barrier into surrounding tissue and further into the blood stream to form distant metastases in other organs like the lung, brain, liver or bone [61].

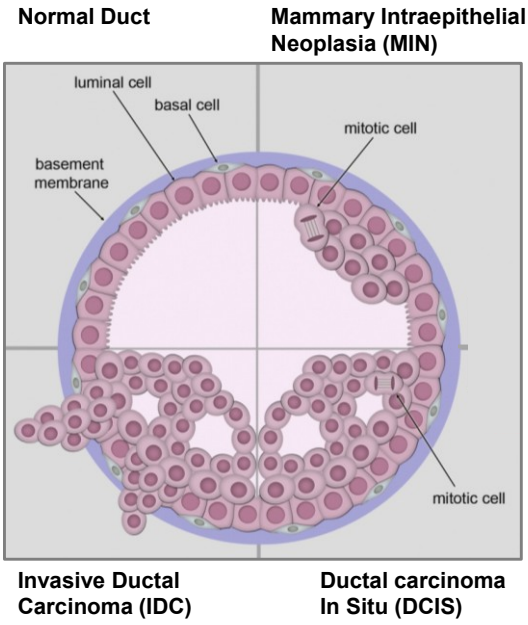


Figure 1-4 Schematic representation of breast duct and lobular composition and deregulated proliferation during breast cancer progression. Figure modified from [62].

The histological classification of the two main categories of breast cancer is defined by the location of occurrence in either the ductal or the breast lobular unit as invasive ductal carcinoma, not otherwise specified (IDC NOS) or invasive lobular carcinoma (ILC). These subtypes represent about 75% and 10% of breast cancer cases followed by a number of special, rather rare types, such as inflammatory, tubular, mucinous or medullary carcinoma [61]. Especially for the IDC NOS and ILC cases the histological classification alone contains only limited value to estimate risk of recurrence or to predict the response to different adjuvant therapies due to the high degree of heterogeneity within the groups. In order to address this issue, immunohistochemical analysis of mainly three receptors, namely estrogen receptor (ER), progesterone receptor (PR), and human epidermal receptor 2 (HER2) is applied. ER positive tumors are dependent on hormone signals mainly from estrogen and thus, patients often benefit from treatment with selective estrogen receptor antagonists like tamoxifen or fulvestrant that block ER signaling [63]. PR staining is mainly concordant with ER and provides only little, if any additional benefit to predict response to anti-estrogen treatment [64]. HER2 detection is a surrogate for aggressive tumors that were originally linked to bad prognosis. However, since the effects of HER2 caused by overexpression or amplification can be targeted by HER2-specific blocking antibodies such as trastuzumab, survival rates for these patients improved [64]. Tumors that are negative for all three markers are defined as triple negative breast cancer (TNBC) and are found in patients with the worst prognosis, mostly due to a lack of targeted therapy options [61].

A further category for classification was introduced by means of gene expression profiling, which extended the set of three marker genes to several hundreds of genes. Perou *et al*. [65] were the first to identify clusters of molecular subtypes, which were confirmed for their robustness by several studies and cohorts over the years [65-67]. Specific clusters of genes that distinguished the tumor samples displayed enrichment in ‘gene signatures’ characteristic for specific cell types like immune cells, adipose/normal cells, breast luminal epithelial or breast basal cells, and ultimately suggested corresponding names for the tumor subgroups. The number of genes required for this gene expression-based assignment was reduced from several hundred to a set of 50 distinguishing genes (Prediction Analysis of Microarray, Pam50), which can be analyzed by RT-qPCR, microarray or more recently by next generation sequencing [64]. The assignment to subtypes resulted in subclassification of breast cancer into luminal A and B, HER2-enriched, and basal-like subgroups, which are sometimes supplemented by a normal-like category. The existence of the last group is debated, as this assignment might reflect contamination with normal nonmalignant cells. Luminal A and luminal B contain mainly ER/PR double positive tumors, although luminal B can stain positive for HER2 and more frequently demonstrates higher proliferation signature

and worse prognosis than luminal A. Classification according to a HER2-enriched molecular profile is largely concordant with the immunohistochemical biomarker, especially for ER negative cases [64]. Although the terms TNBC and basal-like breast cancer are often used interchangeably, between 15 -45 % of basal-like tumors express ER or HER2, thus are not TNBC by definition. Around 15% of TNBC display a molecular subtype other than basal [68]. This degree of heterogeneity also complicates the identification of targeted therapy options for basal-like as well as triple negative breast cancer patients and will probably require further subdivision into better characterized subgroups using additional classifiers such as chromosomal aberrations or genetic mutations. For example, breast cancer patients with *BRCA1* germline mutations often display gene expression patterns similar to the basal-like subtype, maybe representing such a subgroup [69]. Common characteristics of basal-like breast cancer include frequent *TP53* mutations, genomic instability, and inactivation of the RB1 (Retinoblastoma 1) pathway [68, 70].

1.3 Epigenetics and breast cancer

1.3.1 DNA methylation in breast cancer

Similar to other cancer types, aberrations in epigenetic patterns are also frequently identified in breast cancer. This suggests additional possibilities to define breast cancer subgroups for prognosis and therapy options. Early evidence for aberrant DNA methylation was supplied by candidate gene-based studies, which found hypermethylation of *CDKN2A, RASSF1 (ras association domain family member 1*), *ESR1 (estrogen receptor 1*), *RARβ* (*retinoic acid receptor beta*), *CDH1 (cadherin 1), or BRCA1* (reviewed in [71]). Some of these genes actually showed potential as biomarkers (reviewed in [72]) and helped to characterize molecular mechanisms in tumors. For example, breast tumors with hypermethylated *BRCA1* were phenotypically similar to *BRCA1*-mutated tumors, arguing for a functional analogy of hypermethylation and mutation in this gene [73]. Furthermore, treatment of ER negative breast cancer cells with DNMT and HDAC inhibitors resulted in reexpression of epigenetically silenced *ESR1* and restored sensitivity to antiestrogen therapy [74, 75].

These candidate-based approaches with only few target genes were recently extended to a more genome-wide level. Several studies screened breast cancer samples for DNA methylation patterns, mainly using microarrays that cover CpG islands and promoter regions [76]. The identified methylation patterns were capable to distinguish breast cancer subgroups and in particular discriminated ER positive vs ER negative cases [77]. Furthermore, it was possible to reconstruct the molecular subtypes employing DNA

methylation patterns, at least for luminal A and B and basal-like subtypes [78]. Further studies even detected subclusters within the original subtypes [77].

The basal-like breast cancer samples often displayed a distinctly different methylation pattern form the other subtypes. This difference was later confirmed by data generated from 'The Cancer Genome Atlas '(TCGA) Network, which detected basal-like breast cancer to be hypomethylated, while a cluster of mainly luminal B tumors was characterized by hypermethylation [70].

1.3.2 Histone modifications in breast cancer

The role of altered histone modification patterns for the transcription of specific genes is less well examined in breast cancer. Nevertheless, treatment with HDAC inhibitors increased expression of *ESR1* and other silenced genes, which suggests a role of chromatin marks in the regulation of gene expression [76]. In particular, histone modifiers like HDACs or EZH2 displayed differential expression in breast cancer, but the consequences of expression differences for both types of enzymes were described with variable implications for prognosis [76, 79]. In breast cancer, PRC2 targets are frequently hypermethylated, especially at CGI promoters. This process of 'epigenetic switching' occurs in different types of cancers and describes the change from polycomb- to methylation-mediated gene silencing, as this change of epigenetic marks does not influence gene expression [80, 81].

1.3.3 Noncoding RNAs in breast cancer

Expression levels for various noncoding RNAs are deregulated in breast cancer. Recent studies identified miRNAs with both oncogenic (miRNA-21, miRNA-155) and tumor suppressor function (miRNA-200, miRNA-126 or miRNA-335) respective of their target gene [82]. Prediction of lncRNA function and prediction of their target genes is not possible for most examples and requires individual examination for each lncRNA candidate. Nevertheless, lncRNAs with oncongenic and tumor suppressive functions in breast cancer have been discovered. *HOTAIR (HOX transcript antisense RNA)* and *MALAT1 (metastasis associated lung adenocarcinoma transcript 1)* were described as oncogenic lncRNAs, whereas *MEG3 (maternally expressed 3)* and *GAS5 (growth arrest specific 5*) have tumor suppressive functions [83]. Only recently a study confirmed differential expression of lncRNAs to be distinctive for individual breast cancer subgroups similar to mRNA expression profiles [84].

1.4 C3(1) SV40T mouse model of basal-like breast cancer

In the C3(1)SV40T (C3(1)) transgenic mouse model of breast cancer, tumorigenesis is induced by the simian virus 40 early region under the control of the rat prostatic steroid binding protein, which directs transgene expression to the mammary epithelial cells [85, 86]. In particular, the large T antigen (SV40T) of the SV40 early region exhibits cell transforming potential, mainly due to its binding to the cell cycle regulators p53 and Rb1 (depicted in Figure 1-5) (reviewed in [87, 88]).

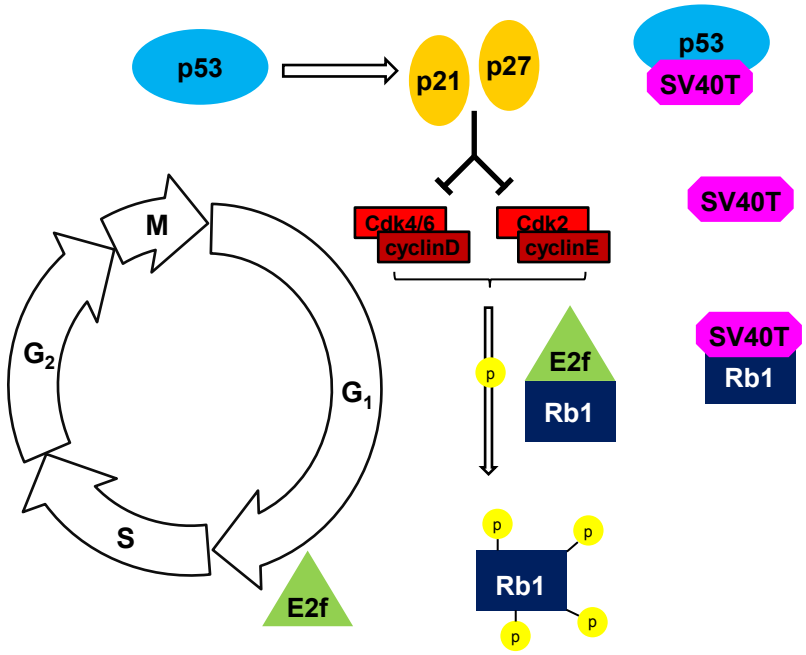


Figure 1-5 Influence on cell cycle regulation by Simian Virus 40 large T antigen (SV40T). p53 induces cyclin dependent kinase inhibitors p21 and p27 to initiate cell cycle arrest. Cyclin dependent kinases (Cdks) in complex with cyclins phosphorylate Rb1, thereby releasing E2f transcription factors to drive G1/S-Phase progression. SV40T binds to p53 and Rb1 to induce cell cycle progression. Figure modified from [85].

Transcription factor p53 induces cell cycle arrest or apoptosis upon stress by inducing the cyclin dependent kinase inhibitors p21 and p27. SV40T disrupts p53 binding to DNA and prevents apoptosis [87, 88]. During normal cell cycle progression Rb1 binds E2f transcription factors, which are released by Rb1 phosphorylation to induce transcription of E2f dependent genes during cell cycle progression. SV40T binding to Rb1 has an equivalent effect to phosphorylation by complexes of cyclins and cyclin dependent kinases, thus releasing E2f.

SV40T further interacts with other proteins like the transcriptional coactivator complex CBP/p300 and the ubiquitin ligase Cul7, which might aid in the transformation process [87, 88]. The deregulation of these cell cycle checkpoints results in abnormal proliferation and produces full invasive carcinomas in 100% of virgin female C3(1) mice [85]. In contrast to other transgenic breast cancer models driven by the mouse mammary tumor virus (MMTV) specific long terminal repeat or whey acidic protein (WAP), the C3(1) promoter region is not sensitive to estrogen exposure [89], thus obliterating the need for pregnancy or hormone treatment to induce tumors. Instead, tumors arise spontaneously and several stages were observed that histologically resemble human mammary intraepithelial neoplasia (MIN, ≈8 weeks), ductal carcinoma in situ (DCIS, ≈12 weeks) and finally progress to the stage of invasive carcinomas at around 16 weeks of age [85]. Gene expression analysis grouped the C3(1) mouse model together with samples of the basal-like subtype of human breast cancer [90]. This was supported by Deeb *et al*., who analyzed C3(1) mice together with other SV40T driven cancer models to define an intrinsic T-antigen signature and described a similar gene expression signature in human breast cancers of the basal-like and luminal B subtype [91]. Profiling for miRNA expression confirmed clustering of C3(1) tumors together with other genetically modified mouse models of the basal-like breast cancer subtype [92]. Various studies have utilized the C3(1) mouse model to demonstrate chemopreventive efficacy of various compounds. These compounds included synthetic drugs like a combination treatment with an inhibitor of ornithine decarboxylase and a precursor steroid for androgens and estrogens (DFMO/DHEA), treatment with the anti-inflammatory agent bindarit, treatment with the retinoid X receptor inhibitor LGD1069 or the cyclooxygenase-2 inhibitor celecoxib [93-96]. In addition, several natural compounds also showed chemopreventive potential, such as green tea and the flavonoids genistein and quercetin [97-99].

2. Aims

Breast cancer is a heterogeneous disease and at least five molecular subtypes have been described, with variable implications for prognosis, metastasis risk and treatment options. Therapy options for breast cancer will be improved by identifying both new markers for early detection as well as novel targets for therapeutic approaches. The characterization of epigenetic modifications, such as DNA methylation, histone modifications or lncRNAs, can serve as an approach to identify potential candidates. Mouse models of cancer are instrumental in investigating epigenetic alterations during tumorigenesis. However, with the exception of microRNA expression, epigenetic patterns in mouse models of breast cancer are still poorly characterized and require further investigations.

Consequently, the first aim of this thesis was to characterize the C3(1) mouse model of breast cancer at the level of DNA methylation. We planned to perform a genome-wide screen with MCIp-seq on age matched WT and C3(1) mammary glands and tumors from different age groups starting at the age of 8 weeks up to 24 weeks. We intended to validate differentially methylated candidate regions with the quantitative EpiTYPER MassARRAY and also to evaluate a link between DNA methylation and gene expression. We expected to clarify, whether DNA methylation differences between mammary glands and tumors in the mouse model were consistent with methylation changes in human breast cancer.

The second aim was to extend the epigenetic characterization of the C3(1) model to the level of histone modifications. Generating ChIP-seq data for three active and one repressive histone mark, we wanted to establish a map of chromatin states in the C3(1) tumors. Overlapping these states with the DNA methylation data should help to characterize genomic regions beside known promoters, such as tissue-specific enhancer regions, and their role in regulating gene expression during tumorigenesis.

The third aim was to investigate epigenetically regulated lncRNAs and evaluate their influence on neighboring protein-coding genes using one candidate pair of coding and noncoding RNA as an example. We planned to validate differential methylation and expression and to examine a potential interaction between the protein-coding and noncoding RNA as well as their regulation using functional assays.

3. Material & Methods

3.1 Materials

3.1.1 Instrumentation, chemicals, reagents

Table 3-1 Instruments used in addition to standard laboratory instrumentation

Instruments	Manufacturer
Agilent 2100 Bioanalyzer	Agilent Technologies, Santa Clara, CA, USA
LightCycler® 480	Roche Diagnostics, Mannheim, Germany
MassARRAY Compact System	Sequenom, San Diego, CA, USA
MassARRAY Nanodispenser	Sequenom, San Diego, CA, USA
Mastercycler® ep gradient	Eppendorf AG, Hamburg, Germany
Mastercycler® pro 384	Eppendorf AG, Hamburg, Germany
Nanodrop spectrophotometer (ND-1000)	PeqLab, Erlangen, Germany
Spectramax 250 microplate reader	GMI Inc., Cold Spring, USA
Qubit® 2.0 Fluorometer	Invitrogen, Life Technologies, Darmstadt, Germany
Covaris S220 sonicator	Covaris, Woburn, MA, USA
SX-8G IP-Star® Automated System	Diagenode, Liège, Belgium
Tissue lyser	Qiagen, Hilden, Germany
Dismembrator S	Sartorius, Göttingen, Germany

Table 3-2 General chemicals and reagents

Material	Manufacturer
100bp DNA ladder	Fermentas, Life Technologies, Darmstadt, Germany
Agencourt AMPure XP beads	Beckman Coulter GmbH, Krefeld, Germany
Ampicilin	Sigma-Aldrich, Taufkirchen, Germany
Bacto Agar	BD Bioscience, San Jose, CA, USA
Bacto Yeast extract	BD Bioscience, San Jose, CA, USA
Bactor Tryptone	BD Bioscience, San Jose, CA, USA
Blasticin S	Sigma-Aldrich, Taufkirchen, Germany
Calciumchloride ($CaCl_2$)	Merck, Darmstadt, Germany
Coelenterazin (COE)	Sigma-Aldrich, Taufkirchen, Germany
Complete Protease inhibitor cocktail, EDTA free	Roche Diagnostics, Mannheim, Germany
Decitabine (5-aza-2'-deoxycytidine)	Sigma-Aldrich, Taufkirchen, Germany
Dharmafect transfection reagent 1	Invitrogen, Life Technologies, Darmstadt, Germany
Diamag protein A coated paramagnetic beads	Diagenode, Liège, Belgium
Dimethylsulfoxide (DMSO)	Merck, Darmstadt, Germany
D-Luciferin	Biosynth AG, Staad, Switzerland
DMEM/F12	Gibco, LifeTechnologies, Darmstadt, Germany
Dulbecco's modified Eagle Medium (DMEM)	Gibco, LifeTechnologies, Darmstadt, Germany
Ethidiumbromide	Sigma-Aldrich, Taufkirchen, Germany
Ethylene diamine tetraacetic acid (EDTA)	Sigma-Aldrich, Taufkirchen, Germany
Fetal bovine serum (FCS)	Biochrome, Berlin, Germany
Formaldehyde, methanol free (16%)	ThermoScientific, Darmstadt, Germany

Material (Table 3-2 continued)	Manufacturer
Gene Ruler DNA ladder mix	Fermentas, Life Technologies, Darmstadt, Germany
GenJet (Ver.II-LnCAP)	SignaGen Laboratories, Rockville, MD, USA
Glutamine	Gibco, LifeTechnologies, Darmstadt, Germany
Glycine (1.25M)	Diagenode, Liège, Belgium
Hank's balanced salt solution (HBSS)	Gibco, LifeTechnologies, Darmstadt, Germany
Linear polyethylenimine(PEI) (MW25000)	Alfa Aesar, Karlsruhe, Germany
Loading Dye (6x)	Fermentas, Life Technologies, Darmstadt, Germany
Locked nucleic acid (LNA) Gapmers	Exiqon, Vedbaek, Denmark
Magnesium sulfate (MgSO4)	Sigma-Aldrich, Taufkirchen, Germany
Magnesiumchloride (MgCl2)	Sigma-Aldrich, Taufkirchen, Germany
Nonidet P-40 (NP-40)	Sigma-Aldrich, Taufkirchen, Germany
Oxalic acid	Sigma-Aldrich, Taufkirchen, Germany
Phenylacetic acid	Carl Roth, Karlsruhe, Germany
Phenylmethylsulfonylfluoride (PMSF)	Sigma-Aldrich, Taufkirchen, Germany
Phosphate buffered saline (PBS)	Gibco, LifeTechnologies, Darmstadt, Germany
Potassium chloride (KCl)	Sigma-Aldrich, Taufkirchen, Germany
Sodium chloride (NaCl)	Sigma-Aldrich, Taufkirchen, Germany
Sodium deoxycholate	Sigma-Aldrich, Taufkirchen, Germany
Sodium dodecyl sulfate (SDS)	Carl Roth, Karlsruhe, Germany
TransIT-LT1	Mirus Bio LLC, Madison, WI, USA
Tris-HCl	Sigma-Aldrich, Taufkirchen, Germany
Triton X-100	Sigma-Aldrich, Taufkirchen, Germany
Trypsin EDTA (0.02%)	Gibco, LifeTechnologies, Darmstadt, Germany
β-mercaptoethanol	Applichem, Darmstadt, Germany

Table 3-3 Enzymes and reaction buffers

Material	Manufacturer
10x PCR Buffer	Qiagen, Hilden, Germany
5x first strand synthesis buffer	Invitrogen, Life Technologies, Darmstadt, Germany
Collagenase A	Roche Diagnostics, Mannheim, Germany
dNTPs (dATP, dTTP, dCTP, dGTP)	Fermentas, Life Technologies, Darmstadt, Germany
HotstarTaq DNA polymerase Kit	Qiagen, Hilden, Germany
Hyaluronidase type I-S	Sigma-Aldrich, Taufkirchen, Germany
Immolase Buffer	Bioline, Luckenwalde, Germany
Immolase DNA Polymerase	Bioline, Luckenwalde, Germany
M.SssI methyltransferase	Fermentas, Life Technologies, Darmstadt, Germany
MNase (300U/µl)	Fermentas, Life Technologies, Darmstadt, Germany
Phusion buffer GC	Fermentas, Life Technologies, Darmstadt, Germany
Phusion polymerase	Fermentas, Life Technologies, Darmstadt, Germany
Proteinase K	Qiagen, Hilden , Germany
Recombinant restriction enzymes: HindIII-HF, KpnI-HF, NheI-HF, SalI-HF, XhoI	New England Biolabs, Ipswich, USA
SuperScript II reverse transcriptase	Invitrogen, Life Technologies, Darmstadt, Germany
T4 DNA Ligase	Fermentas, Life Technologies, Darmstadt, Germany
T4 DNA Ligase Buffer	Fermentas, Life Technologies, Darmstadt, Germany
Universal Probe library and master	Roche Diagnostics, Mannheim, Germany
Yellow Sub™ DNA loading dye	Geneo Bioproducts, Hamburg Germany

Table 3-4 Kits

Kit name	Manufacturer
DNeasy Blood & Tissue kit	Qiagen, Hilden, Germany
AllPrep DNA & RNA miniprep kit	Qiagen, Hilden, Germany
RNeasy Miniprep Kit	Qiagen, Hilden, Germany
Qiaquick Gel extraction kit	Qiagen, Hilden, Germany
MinElute PCR purification kit	Qiagen, Hilden, Germany
REPLI-g mini kit	Qiagen, Hilden, Germany
Quantitect SYBR Green	Qiagen, Hilden, Germany
Qiaprep Spin Miniprep kit	Qiagen, Hilden, Germany
EZ DNA methylation Kit	Zymo Research, Irvine, CA, USA
Agilent RNA 6000 Nano Kit	Agilent Technologies, Santa Clara, CA, USA
Agilent High sensitivity DNA Kit	Agilent Technologies, Santa Clara, CA, USA
NEBNext Ultra DNA Library Prep Kit for Illumina	New England Biolabs, Ipswich, USA
Auto ChIP Kit- A 100	Diagenode, Liège, Belgium
RNase-free DNase Set	Qiagen, Hilden, Germany
MassCLEAVE T7 kit (T Cleavage)	Sequenom, San Diego, CA, USA
Qubit dsDNA HS assay kit	Invitrogen, Life Technologies, Darmstadt, Germany
TOPO TA cloning Kit	Qiagen, Hilden, Germany

Table 3-5 Consumables

Consumables	Manufacturer
15ml, 50ml falcon tubes	Greiner Bio-One, Frickenhausen, Germany
Filter tips and normal tips for pipettes (10µl, 20µl, 200µl, 1000µl)	Biozym, Hessisch Oldendorf, Germany
8-well single cap PCR strips	Biozym, Hessisch Oldendorf, Germany
384 well PCR plates, transparent or white tubes for LC480	Steinbrenner, Wiesenbach, Germany
384 well plates, white flat bottom	Greiner Bio-One, Frickenhausen, Germany
Sterile serological pipettes (5ml, 10ml, 25ml)	BD Bioscience, San Jose, CA, USA
Cell culture flasks, 25cm2 and 75cm2	Greiner Bio-One, Frickenhausen, Germany
6-, 12-, 96- well tissue culture plates	Greiner Bio-One, Frickenhausen, Germany
10 cm petri dish	Greiner Bio-One, Frickenhausen, Germany
Reaction tubes (1.5ml, 2ml, 5ml)	Eppendorf AG, Hamburg, Germany
Scalpels	Feather Safety Razor, Osaka, Japan
Cell scraper	Sigma-Aldrich, Taufkirchen, Germany

3.1.2 Software and databases

Table 3-6 Software and databases

Software	URL
BEDTools Suite of tools v2.17.0	http://bedtools.readthedocs.io/en/latest/index.html
ChromHMM	http://compbio.mit.edu/ChromHMM/
EpiDesigner	http://epidesigner.com/
EpiTYPER	http://www.MassARRAY-EpiTYPER.software.informer.com/
Gene Set Enrichment Analysis	http://www.broadinstitute.org/gsea/index.jsp
GraphPad Prism®	http://www.graphpad.com/
HOMER Suite of tools	http://homer.salk.edu/homer/
Inkscape	https://inkscape.org/de/
Molecular Signatures Database v5.1	http://software.broadinstitute.org/gsea/msigdb
Multi Experiment Viewer Version 4.9.0	http://www.tm4.org/mev.html
Qlucore Omics Explorer v3.1	http://www.qlucore.com/
R	http://www.r-project.org/
TCGA data download	https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp
UCSC Cancer Genomics browser	https://genome-cancer.ucsc.edu/
UCSC Genome Browser	http://genome.ucsc.edu/
UPL assay design center	https://lifescience.roche.com/webapp/wcs/stores/servlet/CategoryDisplay?tab=Assay+Design+Center&identifier=Universal+Probe+Library&langId=-1
Zenbu Genome browser	http://fantom.gsc.riken.jp/zenbu/

3.1.3 Primers

Table 3-7 RT-qPCR primers

Gene name	Forward primer	Reverse Primer	Probe
Actb	AAGGCCAACCGTGAAAAGAT	GTGGTACGACCAGAGGCATAC	56
Cldn4	CATCCAAATTGCTGGTGGAT	CCAGCTTGCGCCTCTACT	29
Elf3	CCAGAAAGCTGAGCAAGGAA	CTCGGATAAACTCCCACAGG	95
Esrp2	TCAGTGTCTTTCAAGCCTACCA	CAACAGGCATCAGAGTGGTG	92
Esrp2-as v1+2	AACACTCATTACAACTCAGTCATGG	CGAGACCGACTTAATCCTCCT	22
Esrp2-as v1-4	CCTAACTCATGCCCAAGGAA	TGTGTGGACAAACCCAGAAG	106
Gsn v1	CCCAAAGTCGGGTGTCTG	CTTCCCTGCCTTCAGGAAT	22
Gsn v2+4	CCTTGTGCAGCCTGTAAGC	CTTCCCTGCCTTCAGGAAT	22
Hprt1	TCCTCCTCAGACCGCTTTT	CCTGGTTCATCATCGCTAATC	95
Igfbp6	AACCCCGAGAGAACGAAGAG	GGGGTTTGCTCTCCTTTGTAG	81
Pkp3	CTAGTCGCCCCTGTCACTATG	AGGGAGCAAACACCAGTCTC	18
Tbp	CGGTCGCGTCATTTTCTC	GGGTTATCTTCACACACCATGA	107
Unc5b	TGGAGGACACACCTGTAGCA	CAAGTAGCCACCCAGAGTCC	1

Table 3-8 EpiTYPER MassARRAY primer

Amplicon	Forward primer [a)]	Reverse primer T7 [b)]
Esrp2 A1	GGTTTAATATTGTTAGGTGGGTGTG	AACCCATAAATCTCAACCCTTACTC
Esrp2 A2	TTATTGTTAGGATTGGGTGAGTGTT	ACACCCACCTAACAATATTAAACCC
Esrp2 A3	TTGTGTTATGGTGTTTATTGTGAGG	AAACCAACTTTACAACCTAAATCCC
Esrp2 A4	GTGTGGGTTAAAGTTTAGGTTGTTG	ATTCAACCCTACCCTACCTAAAAAA
Esrp2 A5	AGGTTTGTTTGTTTTAGTTGATTGG	AACAACCTAAACTTTAACCCACACC
Esrp2 A6	GGTTAGTTAGTTGGGATAGGTAGGT	TACTAAACCATTCCCAAAACCAC
Esrp2 A7	GGTTTTATTATTTGTTAAATTAGGAGGA	AATCTCTACAACTCCCACTACACCC
Esrp2 A8	TGTTTTTTGTTGAATTTTATGGGTT	AAAAAACCAAAACCCTACCCAA
Esrp2 A9	TGGAATTTAATAGGAGGTAATTTAGGG	AAAAACAAAAACAAACAAACAAACC
Esrp2 A10	GGAGGGGTTTTATTTTAAGGTTTT	CCCATTCATCCCATACAATTTAATA
Esrp2 A11	AGTATTTGGGAGGTAGAGGTAGGTG	TCCTTCCACAAATAAATCCTAAACA
Esrp2 A12	TTGGGGTGGTAGAAATAGGTAGATT	CCTACCTTTAATCCCAACACTAAAA
Esrp2 A13	TATTGTTATTTATGGGTAGGTTGGG	AATAACATCAAATCACTCCCTTTCT
Esrp2 A14	TTAGGGTAGGGTTTTTAGGAGAGA	CCTTCAATAACCAATAACCAAAAAC
Cbx8_intergenic	TGTTTATGGATTTAATTTTTTGGGA	ACTACAATCAACCTACACAAACCAA
Cdh4	TGGGTTTTGATAGATTATTGGATTG	AAAACTATACCTCTAACTCTCCCTTC
Ch16_intergenic	GGTTATTTGGTTTTTTAGAAATTTTG	TCAACTATCAACACATAATCCCTC
Cldn4	GGTGGGTAATTGTAGTAGAGGAGGT	CAAAACACAAATCAAATACAATACAAAA
Cldn6	TTAAATGTTGGAGTTTTGATTTTGG	TCCACCTTACCCCTAAAAATAACAT
Elf3	GTTTTTGAGGTTATAGGAAGGGAAG	AAACAACCTAAAACACAAATACCCA
Espin	TGTTGTATTAGGTAGGTAGGTTGG	ACACCACCCAAAACTAACAAAAAAT
Galnt13	TATTAAGAAGAGAAGAGAGAGAAGAAAGAT	TAATCCTAACCCAATAACAAACCTC
Gsn	TGTAGGAAGGAAAGAGTTTTGGTTT	AAACCCAAATATCTCAAAAATCCC
Hoxa5	TTTTGTTTGATGATTTTTAGAGGTAAAT	CCTCACAATTAATACATCCTAATAAAACTA
Iffo1	TAGAGGATGTTAAGGGTTTAGTTGG	TCTTCCTCCTACAACAAAAACAACA
Igfbp6	GGTTGGTATTGTAGTTTTGGGG	AACCATAACCTAAAATAAACTACCCA
Klf15	AAGTATTTTGTAGTTTTTGGGGAGG	AATTCTAAAATCCACTAAAAACCC
Mab21l1	GAGAAATGTTAGGTTAGGAAAGTTGTTAT	CCTAAACCACCAATATCTAAAACCT
Msi1	AGGGTTATGTAGTTGAAGAAATTGG	ACCCCACCTCACCTACTAAAACTAT
Omp	GGGGTAGGGTAGTTAGGGTTGTTAT	AACCAAAACCTAACCCAACAAATAC
Pkp3	GGGAAGGAGAGATGTTTTATTTAGG	AACTTCCTACTATCTACCCTACAACC
Tmem39	GAGGGGATAGGATTTTATAGTG	CAAAACAACTCACACCAAATAAACA
Tor2a	ATTTTTGTTTTTTGGTTTGGTTTTT	AACCAAACTCCCTTCTCTCTACCTA
Unc5b	ATATGGGATGTGAGTGTTGGTATTT	CACCCAAACCCAATAAAATAAATC

[a)] 10bp balancing tag is attached 5' to forward primer (aggaagagag)

[b)] T7 RNA polymerase recognition site is attached 5' of the reverse primer (cagtaatacgactcactatagggagaaggct)

Table 3-9 Cloning Primers

Primer name	Sequence	Purpose
Esrp2-as v4 fwd	TATGGTACCTGCTGAGGCGTGCGCTCGCCGGC TGGAGTCCCACCTCTTGCT	Overexpression *Esrp2-as*
Esrp2-as v1 fwd	TATGGTACCCAAGGACAGAGCTTCATACTTGGC CACTCGCAG	Overexpression *Esrp2-as*
Esrp2-as v1+4 rev	TATGTCGACTTTGAAGTTCTGGAATTGTTTTAAT ATGGGGAATGCCTGGTTGAG	Overexpression *Esrp2-as*
Esrp2-as P4/E4 fwd [a)]	GAGCTCGCTAGCTTACGAATTCCTCCCCTGTG	Luciferase Promoter + Enhancer
Esrp2-as P3/E3 fwd [a)]	GAGCTCGCTAGCCCACTTCTGGCTTCGAGAAC	Luciferase Promoter + Enhancer
Esrp2-as P2/E2 fwd [a)]	GAGCTCGCTAGCCATCATGCACATGCTTGGAC	Luciferase Promoter + Enhancer
Esrp2-as P1/E1 fwd [a)]	GAGCTCGCTAGCCCCCTAGCCTAGTTCCCTGT	Luciferase Promoter + Enhancer
Esrp2-as P1-4 rev [b)]	ATTGCCAAGCTTGCTCTGTCGGCCTTAGATTG	Luciferase Promoter
Esrp2 P1 fwd [b)]	ATTGCCAAGCTTCTACTCCGGCGAGTTACCTG	Luciferase Promoter
Esrp2 P1 rev [a)]	GAGCTCGCTAGCCCTTGCAGAATGGAAAGAGG	Luciferase Promoter
Esrp2 P2 rev [a)]	GAGCTCGCTAGCACTCTAGCCGCTCTCTGCTG	Luciferase Promoter
Esrp2 P3 rev [a)]	GAGCTCGCTAGCCATTCCAGGCTTATCGTGGT	Luciferase Promoter
Esrp2 P1rev fwd [a)]	GAGCTCGCTAGCCTACTCCGGCGAGTTACCTG	Luciferase Promoter
Esrp2 P1rev rev [b)]	ATTGCCAAGCTTCCTTGCAGAATGGAAAGAGG	Luciferase Promoter
Esrp2 E4 rev [b)]	ATTGCCAAGCTTGGCAATCATCAGACCAGGAT	Luciferase Enhancer
Esrp2 E3 rev [b)]	TAAGCCAAGCTTGCCGTATGGGCTGTATGAAT	Luciferase Enhancer
Esrp2 E2 rev [b)]	ATTGCCAAGCTTCTAGGGGTGGGGATTAGGAG	Luciferase Enhancer
Esrp2 E1 rev [b)]	ATTGCCAAGCTTCTGTCCTTGGTGGCTCTGTT	Luciferase Enhancer
Esrp2 E4 rev [a)]	GAGCTCGCTAGCGGCAATCATCAGACCAGGAT	Luciferase Enhancer

[a)] NheI restriction site attached [b)] HindIII restriction site attached

Table 3-10 Additional Primers

Primer name	Sequence	Purpose
IMR_0069	GGACAAACCACAACTAGAATGCAG	Genotyping C3(1)
IMR_0068	CAGAGCAGAATTGTGGAGTGG	Genotyping C3(1)
IMR_0016	GTCAGTCGAGTGCACAGTTT	Genotyping internal control
IMR_0015	CAAATGTTGCTTGTCTGGTG	Genotyping internal control
mouse*Sall3* Fwd	GTTATTTTAGATTTTATTTAGTAGTG	Bisulfite conversion
mouse*Sall3* Rev	TAAAAATAAACCTTCAAATTACCCTT	Bisulfite conversion
mouse*Mest* fwd	CAGACGCCACCTCCGATCC	Control for MCIp enrichment
mouse*Mest* rev	GGCCGCATTATCCCATGCC	Control for MCIp enrichment
mChIP-K4me3_fwd	GGGCCCACGAGTGTCTAC	control of ChIP pulldown with H3K4me3 (UPL #5)
mChIP-K4me3_rev	CAGCCAACTTTACGCCTAGC	control of ChIP pulldown with H3K4me3 (UPL #5)
mChIP-K4me1_fwd	TGCCAGTCCTGCAAGTCA	control of ChIP pulldown with H3K4me1 (UPL #96)
mChIP-K4me1_rev	GGAATACTCTGGGCTCTCCTTAT	control of ChIP pulldown with H3K4me1 (UPL #96)

3.2 Mouse work

3.2.1 Breeding and genotyping

Mouse work was supported by Jana Petersen and Monika Helf, Division of Epigenomics and Cancer Risk Factors, DKFZ, Heidelberg. The FVB-Tg(C3-1-TAg)cJeg/J (C3(1)) mouse model for breast cancer was described previously [85, 86] and 4 founder animals were purchased from Jackson laboratories. The animals were bred and housed in the animal facility of the German Cancer Research Center (DKFZ) under a controlled 14h light/10h dark cycle, at 22°C±1°C, relative humidity 50-55% with access to chow (phytoestrogen-low R/M-H, V1554-703, Ssniff, Soest, Germany) and water *ad libitum*. Mice were bred on either complete transgenic background (TG) or mixed with wildtype (WT) in a ratio of 2 females : 1 male.

For genotyping, tail biopsies were taken at time of weaning to isolate DNA and perform genotyping PCR (Table 3-11).

Table 3-11 PCR set-up and cycling protocol for C3(1) genotyping

Component	Amount
Mouse tail DNA	2µl
10x Immolase Buffer	1.2µl
dNTPs (10mM)	0.96µl
$MgCl_2$ (50mM)	0.6µl
IMR068 (20µM)	0.6µl
IMR069 (20µM)	0.6µl
IMR015 (20µM)	0.3µl
IMR016 (20µM)	0.3µl
Yellow Sub ™	1.6µl
Immolase DNA Polymerase (5U/µl)	0.28µl
H_2O	Add to 12µl

Temperature	Time	
94°C	2min	
94°C	20sec	12x
64°C: -0.5°C/ cycle	30sec	12x
72°C	1min	12x
94°C	20sec	25x
58°C	30sec	25x
72°C	40sec	25x
72°C	5min	
10°C	hold	

The band at 200bp represents the internal PCR control and is present in both TG and WT samples. The band of 500bp represents the SV40T product in transgenic animals (Figure 3-1).

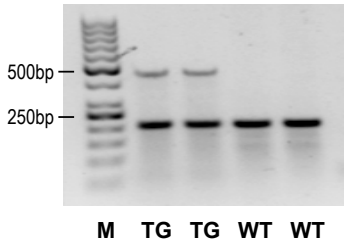


Figure 3-1 Representative gel of C3(1) genotyping PCR products.

M: DNA ladder; TG: C3(1) animal, WT: wild type animal

3.2.2 Sample collection

For the analysis of DNA methylation kinetics, female mice of WT or TG animals were sacrificed in age intervals of 4 weeks starting at 8 weeks up to 24 weeks of age. After sacrifice of animals by CO_2, individual tumors or mammary glands were resected before snap freezing in liquid nitrogen. The animal work was approved by the state Animal Care and Use Committee (Regierungspräsidium Karlsruhe) as regulated by German federal law for animal welfare under the registration number 35-9185.82/A-15/08.

3.3 Nucleic acid isolation and quantification

DNA and RNA were isolated from tissues or cell lines with the DNeasy Blood & Tissue kit, RNeasy Mini Kit or AllPrep DNA&RNA kit from Qiagen according to manufacturer's instructions. For removal of residual DNA during RNA isolation, an additional on-column DNase treatment with RNase-Free DNase Set was performed according to the manufacturer's recommendation. For disruption and homogenization of tissue pieces during AllPrep isolation, 3-4 cycles of tissue lyser shaking (30s, frequency 30 1/s) in buffer RLT plus were conducted. Quantity and purity of nucleic acids were measured with Nanodrop spectrophotometer (ND-1000). For RNA quality, the RNA integrity (RIN) was assessed by Agilent RNA 600 Nano chip and only samples with a RIN> 6 were analyzed by RT-qPCR. In case of library preparation for next generation sequencing, double stranded DNA is required, which can be measured by means of a DNA intercalating fluorophore on a Qubit 2.0 Fluorometer using Qubit dsDNA HS assay kit according to the manufacturer's instructions.

3.4 Methyl CpG binding domain immunoprecipitation (MCIp)

For the analysis of genome-wide DNA methylation differences, Methyl CpG binding domain immunoprecipitation followed by next generation sequencing (MCIp-seq) was applied [100, 101]. This exploits the capability of the MBD2-Fc fusion protein to bind methylated DNA fragments by the MBD2 domain and to protein A-coated paramagnetic beads by the Fc part. Binding of DNA fragments to the MBD2 domain is dependent on the methylation level and CpG density of fragments. Washing with five solutions of increasing salt concentration elutes fragments bound to the MBD2 domain, thus enriching for highly methylated fragments. The highest salt concentration contains the fragments with highest methylation levels and was used for enrichment of both TG as well as WT samples of different ages to compare methylation changes occurring in the course of tumorigenesis.

MCIp was conducted following the protocol described by Sonnet *et al*. [100] with minor modifications using a SX8G-V52 robot for automated processing of a maximum of 8 samples

in parallel with the program MCIP30_ETipRedCorr3.HLD. In brief, 5µg DNA in 120µl EB buffer was sheared to ≈150bp using a Covaris S2 sonicator for 6 cycles (duty cycle 20% intensity 5, burst per cycle 200, time: 60sec), checking in between cycles for formation of air bubbles in the sonication tube which leads to uneven fragmentation. Size distribution was confirmed on an Agilent DNA High sensitivity Chip before proceeding with the MCIp reaction in the robot. 60µg of MBD2-Fc protein and 40µl of protein A-coated paramagnetic beads were used for the reaction.

Several washing steps with wash buffer (0.1% NP-40, 20mM Tris-HCl pH 8.0, 2mM $MgCl_2$, 0.5mM EDTA), containing increasing NaCl concentrations (300mM - 1M), were performed to remove unbound fragments and enrich for highly methylated DNA, which exhibited the strongest affinity for MBD2. Thus, the highest salt concentration (1M) was at the same time the final eluate and was used in the following steps. DNA of the final eluate was purified with the MinElute Kit in 20µl dH_2O. Enrichment of methylated DNA was confirmed by SYBR Green based RT-qPCR (Table 3-12) for the imprinted *Mest* gene with the program of 95°C for 15min followed by 45 cycles of 94°C for 15sec and 60°C for 30sec on a Light Cycler480. Enrichment of *Mest* fragments is found mostly in the low salt fraction (300mM, unmethylated allele) and in the highest salt fraction (1M, methylated allele). Relative enrichment was calculated with help of a standard curve generated from serial dilutions of fragmented, unprecipitated input DNA (1:10, 1:100, 1:1000). Fragment quantity and size distribution were measured by Qubit and DNA High sensitivity chip.

Table 3-12 Enrichment RT-qPCR

Component	Amount
2x Quantitect SYBR Green	5µl
Mest Fwd (10µM)	0.3µl
Mest Rev (10µM)	0.3µl
DNA (diluted 1:10)	2µl
dH_2O	2.4µl

The quality controlled DNA was processed by the DKFZ Genomics and Proteomics Core Facility for library preparation and next generation sequencing. Library preparation was performed using the Illumina ChIP-Seq DNA Sample Prep Kit (Catalog IDs: IP-102-1001) according to the manufacturer's protocol (#11257047 Rev. A) with minor adaptations for ligation of 6mer barcoded adapters, size selection via E-gel, clean up via AmpureXP beads (Agencourt) and PCR with primers PE 1.0 (AATGATACGGCGACCACCGAGATCTACACT CTTTCCCTACACGACGCTCTTCCGATCT) and PE 2.0: (CAAGCAGAAGACGGCATACGA GATCGGTCTCGGCATTCCTGCTGAACCGCTCTTCCGATCT) according to Lefrancois *et al.* [102]. Barcoded adapter and primer sequences were kindly provided by Megumi Onishi

Seebacher. Samples were 18-plexed and sequenced on 2 lanes of an Illumina HiSeq 2000 sequencer (single read 50bp).

3.5 Bioinformatic data mining

3.5.1 Alignment and quality control

Alignment of reads to the mouse reference genome mm10 was performed with Burrows Wheeler aligner (BWA) [103] and removal of duplicates as well as bad quality reads was conducted with Picard (https://broadinstitue.github.io/picard) and Samtools [104], respectively. Saturation efficiency and CG-coverage was calculated as an additional control step by the R-MEDIPs package [105].

3.5.2 Calling of DMRs (HOMER)

For peak calling of DMRs, reads for animals of the same genotype and the same age group were combined and analyzed with HOMER (findPeaks: FDR< 0.001, p-value< 0.0001, size 150bp, minDist 300bp) [106] using either normal (hypermethylation) or tumor reads (hypomethylation) as background. DMRs in the 24w age group were selected for co-occurrence with DMRs in either the 16 or 20w group or both (mergePeaks –cobound). Promoter DMRs were required to overlap with a region of 2kb upstream and 0.5kb downstream of the TSS of RefSeq annotated genes (GRCm38/mm10, http://genome.ucsc.edu/ [107]) by the bedtools intersect option [108]. Application of the HOMER mergePeaks command (-d 300bp, -venn) on DMRs obtained from the individual age groups could further divide DMRs into age group unique DMRs and recurrent DMRs, the later being further classified into i) progressive (occurring from their first appearance to the latest age group at 24 weeks), ii) continuous (occurring in neighboring age groups, but not in the 24w group) and iii) discontinuous DMRs (occurring in non-neighboring age groups). Regions were further characterized for their localization with respect to the closest genomic feature (e.g. exon, CpG-island, repeat element) with the annotatePeaks command. Coverage at certain regions of interest (DMRs, tissue specific chromatin states) by MCIp-seq was obtained using the annotatePeaks –hist option for the respective regions.

3.5.3 Principal component analysis of recurrent DMRs in C3(1) tumors and mammary glands

Recurrent DMRs were annotated for the MCIp-seq tag densities in individual animals with HOMER (see section 3.5.2) and similarity of samples was examined by principal component analysis (PCA) with Qlucore Omics Explorer software v 3.1.

3.6 Quantitative methylation analysis by EpiTYPER MassARRAY

3.6.1 Bisulfite conversion and EpiTYPER PCR

For the assessment of DNA methylation levels on a quantitative basis, EpiTYPER MassARRAY was performed as previously described [81, 109]. In brief, 500ng-1µg DNA was bisulfite treated using the ZYMO EZ DNA methylation kit according to manufacturer's instructions except changing the denaturing conditions to 15 cycles of 1h at 50°C and 10sec at 98°C followed by an additional 1h at 50°C and increasing the elution volume to 2x 30µl M-elution buffer. Successful conversion was confirmed by PCR for *Sall3* amplicon which only gives a product for converted cytosines (Primer in Table 3-10). In a PCR, BT DNA was amplified with target specific primers (Table 3-8) that attach a T7 polymerase recognition site to the PCR product (Table 3-13).

Table 3-13 PCR set-up and cycling protocol for amplification of BT DNA for EpiTYPER MassARRAY analysis

Component	Amount
BT DNA	1µl
10x PCR Buffer	0.5µl
Primer mix (10µM)	0.1µl
dNTPs (10mM)	0.1µ
Hotstar Taq (5U/µl)	0.04µl
dH_2O	add to 5µl

Temperature	Time	
95°C	15min	
94°C	30sec	48x
56 - 60°C	30sec	
72°C	30-60sec	
72°C	5min	
4°C	hold	

3.6.2 Dephosphorylation, *in vitro* transcription, and desalting with Resin

After dephosphorylation of unincorporated dNTPs by shrimp alkaline phosphatase (SAP) treatment for 20min at 37°C followed by heat inactivation for 5min at 80°C, the product was *in vitro* transcribed by T7 polymerase (MassCLEAVE T7 kit (T cleavage)) for 3h at 37°C in combination with RNaseA mediated base-specific fragmentation of RNA (Table 3-14).

Table 3-14 In-vitro transcription mix (MassCLEAVE T7 kit)

Component	Amount
SAP-treated PCR product	2µl
5x T7 Polymerase buffer	0.89µl
T-Cleavage mix	0.22µl
DTT (100mM)	0.22µl
T7-polymerase	0.4µl
RNaseA (10mg/ml)	0.06µl
dH_2O	Add to 7µl

The product was diluted with 20µl and desalted with 6mg resin rotating for 30min at room temperature before fragments were analyzed by matrix assisted light desorption/ionization time-of-flight mass spectrometry (MALDI-TOF). The EpiTYPER software compares the obtained mass spectra with the expected spectra and thus, identifies individual peaks. The difference between methylated and unmethylated cytosine is translated from a C/T transition after bisulfite conversion and PCR to an A/G difference after *in vitro* transcription, which results in a mass shift of 16Da per CpG site in the fragment. The software quantifies the ratio and area abundance for peaks of methylated and unmethylated fragments resulting in a relative methylation value. In addition, spectra were visually inspected for correct peak recognition in comparison with a 6-point *in vitro* methylated standard (0-100%) (Preparation see chapter 3.6.3) before inclusion of CpG units into the final dataset.

3.6.3 Preparation of 6-point *in vitro* methylated standard

For the generation of a 6-point *in vitro* methylated standard, genomic DNA was whole-genome amplified to generate unmethylated DNA using the REPLI-G mini kit according to the manufacturer's instructions. For the generation of methylated DNA, 15µg genomic DNA was incubated with M.SssI for 1h at 37°C (Table 3-15). *In vitro* methylated genomic DNA and whole-genome amplified DNA was cleaned using the DNeasy Blood & Tissue kit according to the manufacturer's specifications. To obtain the 6-point standard, methylated and unmethylated DNA were mixed in different proportions to obtain 0%, 20%, 40%, 60%, 80% and 100% methylation levels. Standard was bisulfite converted before use in EpiTYPER PCR.

Table 3-15 Reaction mix for *in vitro* methylation of genomic DNA

Component	Amount
Genomic DNA	15µg
M.SssI	10µl
50 x SAM	13µl
10x M.SssI Buffer	65µl
dH_2O	Add to 650µl

3.7 Relative quantification of nucleic acids by real time PCR (RT-qPCR)

3.7.1 cDNA synthesis

cDNA was generated from isolated RNA by means of reverse transcription (RT) using SuperScriptII Reverse Transcriptase. 0.5-1µg RNA were incubated with 200ng of random hexamers for 5min at 65°C and cooled for 1min on ice. Then, reverse transcription mix (Table 3-16) was added and incubated as follows: 22°C for 10min, 42°C 50min and 72°C for 15min. cDNA was stored undiluted at -20°C or depending on application diluted 1:5 or 1:10 with dH_2O.

Table 3-16 Reverse transcription mix and incubation conditions

Components	Amount
5x First-Strand buffer	4µl
DTT (0.1M)	2µl
dNTPs (10mM)	1µl
SuperScript II RT (200U/µl)	0.2µl

3.7.2 Quantitative real time PCR

Quantitative real time PCR (qPCR) analysis was performed using the Roche Lightcycler 480 and the Universal Probe Library system, which employs a fluorescently labeled probe that specifically binds the target of interest. In the reaction setup, 2.5µl sample (cDNA, ChIP DNA) were mixed with 3.5µl Probes Master, 1µl primer mix (0.5µM each primer) and 0.05µl universal probe, for running the following Lightcycler program of 15min at 95°C followed by 45 cycles of 10s at 95°C, 20s at 55°C and 10s at 72°C. Primer design was performed with the web-based Roche Universal Probe library assay design center to identify compatible probe-primer combinations. C_T values were calculated by the Lightcycler 480 software with the Abs quant/2nd derivative max option. Relative expression levels of target genes were normalized to three housekeeping genes (*Hprt1, Tbp, Actb*) according to the Livak $^{\Delta\Delta}$CT method [110]. For calculation of enrichment of ChIP DNA fragments, DNA values were normalized to the 5% input control. Primers and respective probe numbers are listed in Table 3-7.

3.8 Cell culture experiments

3.8.1 Cell lines and cell culture

The $M28_{N2}$, $M27_{H4}$, M6 and M6C cell lines were derived from different stages of mammary gland tumors of C3(1) origin [111] and were kindly provided by Cheryl Jorcyk (Boise State University). 3T3-L1 mouse preadipocytes [112], platinum-E (Plat-E) retroviral packaging cells

[113] and NMuMG mouse mammary gland epithelial cells [114] were a kind gift from Daniel Mathow and Hepa1.6 murine hepatoma cells [115] were generously provided by Ursula Klingmüller (both DKFZ Heidelberg). Cell lines were cultivated in DMEM +10% FCS in a humidified atmosphere at 5% CO_2 and 37°C. For NMuMG cells, medium contained 10mg/ml insulin. DNA and RNA from MC38 murine colon carcinoma cells and CMT93 murine rectum carcinoma cells were provided by Christoph Weigel (DKFZ Heidelberg). For propagation, cells were dislodged from the tissue culture surface by trypsin-EDTA (0.02%) treatment. Cryoconservation of cells was conducted in medium containing 10% DMSO followed by freezing at a rate of -1°C/min in freezing containers before ultimate storage in liquid nitrogen.

3.8.2 Decitabine treatment

In order to evaluate the effect of methylation changes on gene expression, cell lines were treated with the DNMT inhibitor Dac. Cells were seeded at densities of 2-2.5x 10^5 cells per six well 24h before initiation of treatment and Dac dissolved in dimethylsulfoxide (DMSO) was added to a final concentration of 0.1µM (M6) and 1µM ($M27_{H4}$) (max. DMSO concentration of 1µM respectively). Medium containing Dac or DMSO was replaced every 24h for a total treatment duration of 96h. Changes in expression were measured by RT-qPCR and demethylation was confirmed by EpiTYPER MassARRAY. The experiment was repeated two or three times.

3.9 Comparison of DMRs in the C3(1) mouse model and human breast cancer samples

3.9.1 Analysis of 450k DNA methylation data for TCGA breast cancer samples

DNA methylation obtained for human breast samples by Infinium HumanMethylation 450k bead Chip array (450k) were downloaded from the TCGA data portal (http://cancergenome.nih.gov). For genes with a mouse DMR in the promoter region, all probes overlapping the human promoter (2kb upstream and 0.5kb downstream of the TSS) of the respective gene were selected for analysis. Pam50 classification of samples was used according to RNA microarray data [70]. Data were visualized by Qlucore Omics Explorer software v3.1 to generate principal PCA plots and heatmaps. Heatmaps for clustering applied weighted average linkage principle and a two-group t-test comparison between basal-like breast cancer samples and the other Pam50 defined subtypes (without normal control samples) identified the top 5% of Infinium probes that best distinguish the basal-like subgroup.

For validation of DMRs between C3(1) and the human subtypes, the promoter methylation was calculated by averaging the methylation values for all probes of the respective gene. The methylation was independently compared for each individual pair (gene, subtype) to normal controls by student's t-test, correcting for multiple testing by Benjamini-Hochberg correction. Hyper- and hypomethylation was defined as a significant ($p< 0.01$) increase or decrease in methylation. The direction of methylation change was compared with DMRs from the C3(1) model.

3.9.2 Hierarchical clustering of mouse promoters and progressive DMRs

For comparison with DMRs in human breast cancer samples, tag density was calculated over the complete promoter regions in contrast to specifically for the DMRs, and animals of the age groups 20-24w were analyzed with hierarchical clustering (average linkage, Spearman rank correlation) using Multiexperiment Viewer (MeV) Version 4.9.0. In order to examine the dynamics of DMR formation during tumorigenesis, read densities for progressive DMRs were normalized according to the mean and variance across all MCIp-seq samples with Qlucore and were clustered with average linkage to generate heatmaps.

3.10 Gene set enrichment analysis (GSEA)

In order to investigate whether differentially methylated genes were enriched in previously defined gene sets, the Molecular Signatures Database v5.1 (MSigDB) was queried for overlaps [116] (http://software.broadinstitute.org/gsea/msigdb/annotate.jsp). Furthermore, gene sets defined by differences in DNA methylation and chromatin states were analyzed for enrichment in the published C3(1) gene expression data set [90] with GSEA Java desktop application [117]. The analysis was conducted with standard parameters using a cutoff for p-value < 0.5 and false discovery rate (FDR) < 0.25.

3.11 Chromatin Immunoprecipitation (ChIP)

3.11.1 Mammary epithelial cell (MEC) enrichment

For chromatin immunoprecipitation in WT mammary glands we isolated and enriched MECs, following an adapted protocol for isolation and cultivation of mammospheres by Dr. Ansam Sinjab [118]. In particular, axillary and inguinal mammary glands were dissected and collected in ice-cold DMEM/F12. Glands were washed by dipping into 70% ethanol and

DMEM/F12 before mincing on a petri dish to a size of ≈1mm^3 with scissors. Minced glands were digested while rotating at 37°C for 3-4h in digestion cocktail (10ml per 5 glands; 200mM glutamine, 5% FCS, 500µl collagenase A (stock 25mg/ml in HBSS), 1.25ml hyaluronidase (1mg/ml freshly prepared in HBSS), add DMEM/F12 to 10ml). The mix became more turbid with organoids sinking to the bottom and fat tissue floating on top. Organoids were sedimented by 4 centrifugation steps at 4°C for 1min increasing speed from 80g, 160g, 320g, to 450g. Supernatant was removed without disturbing the pellet. The pellet was washed once with 10ml cold DMEM/F12 and once with 10ml cold HBSS with centrifugation at 450g for 5min at 4°C. Cells in the pellet were enriched for mammary epithelial cells and were freshly cross-linked for chromatin preparation.

3.11.2 Chromatin preparation

For ChIP of C3(1) tumors, tissue was disrupted and homogenized in liquid nitrogen using a tissue dismembrator S to generate a homogenous tissue powder. For crosslinking of chromatin and chromatin-bound proteins, approximately 60-80mg tissue powder was thawed on ice for 10min before adding 1ml freshly prepared 1% formaldehyde (methanol-free) diluted in PBS and incubation for 10min at room temperature. During incubation, homogenization of solution was enhanced by two cycles of tissue lyser shaking (30sec, 1/30sec frequency) before the reaction was quenched by adding glycine to a final concentration of 125mM and incubation for 5min at room temperature. Samples were pelleted by centrifugation at 300g for 5min at 4°C and washed three times in PBS/0.5mM PMSF. The pellet was resuspended in 900µl MNase Buffer (Table 3-17) with addition of 10U MNase/15mg tissue before incubation for 15min at 37°C. The reaction mix was then complemented by 100µl 10x sonication buffer (Table 3-17). Suspension was cooled for 15 min on ice before sonication using a Covaris S2 sonicator for 40 cycles (cycles per burst 200, duty cycle 20%, intensity 8, time 60s). An aliquot was analyzed on an Agilent High Sensitivity DNA Chip for quality assessment, with good quality indicated by a large peak at 150bp indicating mononucleosomes and only a minor tail towards larger fragments of polynucleosomes (Figure 3-2).

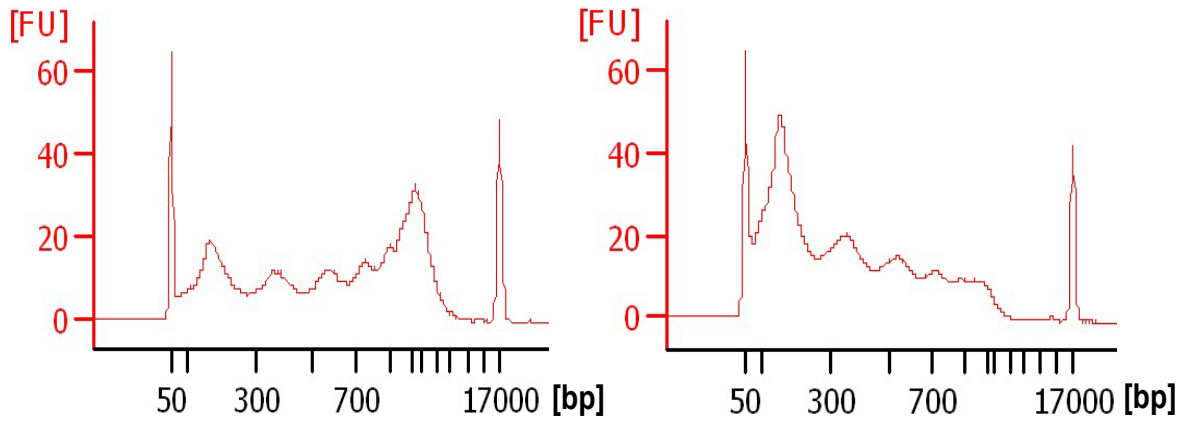


Figure 3-2. Agilent chip results for chromatin samples after MNase digestion and ultrasonication. If fragmentation was incomplete (left), the highest peak was seen at fragment sizes of polysomes (four nucleosomes or more). Completely fragmented chromatin (right) displayed mainly mononucleosomes sized (≈150bp) fragments and only fewer large fragments.

For MEC chromatin, cells were freshly crosslinked with 10ml of 1% formaldehyde in PBS shaking at room temperature for 10min followed by quenching and washing as with tumor samples. Cells were counted before the last washing step and aliquoted before snap-freezing the pellet in liquid nitrogen. For fragmentation, MECs were resuspendend in 900µl MNase buffer and homogenized by 2x2 cycles of tissue lyser shaking before incubation at 37°C with 5U MNase/ $1x10^6$ cells for 40min at 600rpm. Reaction was complemented with 10x sonication buffer as for tumors and sonicated at the same conditions for 20 cycles.

Chromatin was centrifuged at full speed for 15min followed by preclearing the supernatant with 20µl pre-equilibrated protein A paramagnetic beads and 1µl of rabbit negative control IgG antibody by rotating for 2h at 4°C.

Table 3-17 Buffer composition for Chromatin imunoprecipitiation

MNase Buffer: 100mM Tris pH 8.0, 2M NaCl, 10mM EDTA 0.1% SDS 0.2% Na deoxycholate, 1x protease inhibitor complete EDTA-free
10x sonication buffer: 100mM Tris pH 8.0, 2M NaCl, 10mM EDTA 01% SDS 0.2% Na deoxycholate, 1x protease inhibitor complete EDTA-free
ChIP elution buffer: 10mM Tris pH 8.0, 5mM EDTA 300mM NaCl, 0.5% SDS

3.11.3 Automated Chromatin Immunoprecipitation

Immunoprecipitation reaction was performed with the Auto ChIP Kit and the SX8G-V52 robot according to the manufacturer's specifications using the IP and beads incubation 100µl program [119]. 10µg Chromatin DNA were incubated with antibodies against histone modifications (Table 3-18) by constant pipetting of the robot at 4°C overnight for 11-13h before precipitation of the formed complex by pre-washed protein A coupled paramagnetic

beads. Further washing steps to reduce unspecific DNA binding was followed by elution in ChIP elution buffer (Table 3-17) and de-crosslinking for 4h at 65°C with 5µl proteinase K and 3µl RNase A (10mg/ml) for 30min at 37°C. Finally, DNA was cleaned up with AMPure XP SPRI beads 1.8x, 3 wash steps with 80% EtOH and eluted in 22µl EB buffer before initial analysis of enrichment with RT-qPCR (chapter 3.7.2). For next generation sequencing, 20ng DNA of precipitated samples as well as from 5% input control were used for library preparation for NGS according to the NEBnext Ultra DNA Library Prep Kit for Illumina and sequenced on an Illumina HiSeq2000 (single read 50bp).

Table 3-18 Antibodies used for ChIP

Antibody	Amount used per precipitation (10µg chromatin DNA)	Company
H3K4me3 (ab8580)	2µg	Abcam, Cambridge, UK
H3K4me1 (ab8895)	1µg	Abcam, Cambridge, UK
H3K27ac (ab6002)	5µg	Abcam, Cambridge, UK
H3K27me3 (07-449)	3µg	Merck Milipore, Darmstadt, Germany
negative control IgG (AIB-103-110)	1µg/ chromatin preparation	Diagenode, Liège, Belgium

3.11.4 ChIP-seq library preparation

Library preparation was performed according to the NEBnext Ultra DNA Library Prep Kit for Illumina protocol diluting the NEBNext Adaptor 1:10 and including size selection. PCR amplification, which included the addition of index primers, was monitored by RT-qPCR on a LightCyler 480 by addition of SybrGreen to the reaction. PCR amplification was stopped in the exponential phase shortly before entering the linear phase of the reaction and if the fluorescence intensity was higher than 7, which occurred mostly between 9-13 cycles. After clean-up with AMPure XP beads, quantity and size distribution of libraries was controlled by Qubit and DNA High sensitivity Chip, before they were multiplexed and sequenced on an Illumina HiSeq2000 (single read 50bp) by the DKFZ Genomics and Proteomics Core Facility. Alignment and quality control of ChIP-seq reads were performed as for MCIp-seq reads (see chapter 3.5.2) only omitting the saturation analysis.

3.11.5 Classification of chromatin states

The ChIP-seq data obtained for MECs and the three tumor samples for the four histone marks, H3K4me3, H3K4me1, H3K27ac, and H3K27me3 were examined with the ChromHMM algorithm [37] to train a Hidden Markov model, which allows the assignment of chromatin states to the different positions of the genome. The input signal combined for the

four samples was used for corrections and ChromHMM was run with standard parameters to train a single joint model over all samples. The ChromHMM CompareModels option was used to compare different models that were trained for different numbers of states with the 16-state model (16 is the maximal possible number of states when analyzing data obtained with four different marks). When reducing the number of states from the 16-states model to 11-states, none or only one additional state was lost at each step due to low correlation (correlation < 0.9 or < 0.8). Since four additional states were lost between the 10- and the 11-state model, the later model was chosen for analysis based on the available data. The basic features to calculate for enrichment that were provided by ChromHMM, like CpG islands or RefSeq Genes, were extended by FANTOM5 CAGE-seq data and CAGE-seq based enhancers [120] (downloaded from: http://fantom.gsc.riken.jp/5/data/), chromatin defined enhancers from 19 mouse tissues and cell types by Shen and colleagues [121] and H3K36me3 ChIP-seq data for MECs [122]. Respective regions were converted from NCBI37/mm9 assembly to GRCm38/mm10 with LiftOver (http://genome.ucsc.edu/). Another enrichment analysis was performed to look for overlap with DMRs, in particular progressive DMRs (details for selection are provided in chapter 3.5.2). With the Rmerge Peaks script (by Manuela Zuckncik) regions with the same chromatin state in at least two of the three tumor samples were identified as common for tumors. For the definition of a region as being a tissue-specific chromatin state (either tumor- or MEC-specific), the regions with common states in tumor were reduced by those that overlapped with regions of the same state in MECs by means of the bedtools subtract (option –A) command or vice versa. With the bedtools intersect command, the tissue-specific chromatin state regions that overlap with DMRs could then be identified and used for a HOMER Motif analysis (command: findMotifsGenome.pl –size 200 –S10) [106].

3.12 Antisense LNA Gapmer mediated knockdown of *Esrp2-as*

For Locked nucleic acid (LNA) Gapmer mediated knockdown, custom designed antisense oligos against *Esrp2-as* and the negative control oligo A were reverse transfected into cell lines M6, $M28_{N2}$ and NMuMG with Dharmafect 1 solution. LNAs activate RNaseH mediated RNA degradation and allow for strand-specific targeting of RNA molecules, which in total reduces off-target effects. LNAs at a final concentration of 20nM and Dharmafect 1 solution (2µl/well for 12 well plates) were separately diluted in 100µl serum-free DMEM medium and incubated for 5min before mixing the components. After additional 20min of incubation, the transfection mix was added to the cell suspension (≈ $5x10^4$ cells per well in 1ml medium) in the wells and incubated for 96h (M6, $M28_{N2}$) or 72h (NMuMG). In M6 cells, a long term

knockdown series was performed by repeatedly transfecting cells after 96h for four times. Knockdown experiments were performed once.

3.13 Overexpression of lncRNA *Esrp2-as* and *Esrp2*

3.13.1 Cloning of *Esrp2-as* for overexpression

For cloning of *Esrp2-as*, 3µg murine liver RNA was reverse transcribed using 200ng random hexamers and SuperScript II reverse transcriptase as described in section 3.4.2. The transcript was PCR amplified with primers to attach KpnI and SalI restriction sites to the product (Table 3-19).

Table 3-19 PCR setup and cycling protocol for amplification of *Esrp2-as*

Component	Amount
cDNA	1µl
5x Phusion buffer GC	4µl
dNTPs (10mM, 2.5mM each)	0.4µl
Primer fwd (10µM)	0.5µl
Primer rev (10µM)	0.5µl
Phusion Polymerase (2U/µl)	0.2µl
Q-solution (Qiagen) 5x	4µl
Template cDNA	1µl
dH2O	Add to 20µl

Temperature	Time	
98°C	3min	
98°C	10sec	32x
66°C	30sec	
72°C	1min	
72°C	5min	
4°C	hold	

The PCR reaction for transcript variants v1 and v4 were confirmed for correct size on an agarose gel and consequently gel purified using Qiaquick gel extraction Kit. After A-tailing, the PCR product was inserted into the pCR2.1 vector following the TOPO-TA (Invitrogen) protocol. Chemical competent TOP10 E.coli were transformed for propagation following the manufacturer's instruction. The insert was cut at the introduced KpnI and SalI restriction sites, and after gel purification ligated into the respectively digested pCRII-cGFP-bGH vector (kindly provided by Sven Diederichs, DKFZ). Correct sequence of clones was confirmed with Sanger sequencing by GATC Biotech (Konstanz, Germany).

3.13.2 Transfection of cell lines for overexpression of *Esrp2* and *Esrp2-as*

For overexpression of *Esrp2-as* in $M27_{H4}$, Hepa 1.6, and 3T3-L1 cells, constructs (v1 or v4) were diluted in serum-free DMEM (2.5µg) and mixed with respectively diluted TransIT-Lt1 (7.5µl) transfection reagent solution at a ratio of 3:1 transfection reagent to plasmid DNA. 20-

30min after mixing, the transfection mix was evenly distributed to cells, which were seeded 18-24h before (≈1x 10^5 cells per well in 6 well plates). For transient transfection, cells were harvested after 72h ($M27_{H4}$) or 48h (Hepa 1.6 and 3T3-L1) and analyzed by RT-qPCR.

The retroviral vector pMXs-IRES-Blast-*Esrp2*-FF [123] and the corresponding empty vector with GFP were a gift from Russ P Carstens (University of Pennsylvania). Retroviral infection of $M27_{H4}$ and 3T3-L1 cells was performed as in [124], followed by selection with Blasticidin S at 5µg/ml for stable incorporation.

3.14 Luciferase Reporter assays

3.14.1 Cloning for dual luciferase reporter assays

For dual luciferase reporter assays, the sequences covering regions upstream of the *Esrp2* and the *Esrp2-as* TSS were PCR amplified from genomic liver DNA with primers attaching NheI and HindIII restriction sites (Table 3-20).

Table 3-20 PCR set-up and cycling protocol for amplification of reporter fragments

Component	Amount
10x PCR buffer	3µl
dNTPs (10mM, 2.5mM each)	0.6µl
Primer fwd (10µM)	0.75µl
Primer rev (10µM)	0.75µl
Hotstar Taq polymerase (5U/µl)	0.2µl
Q-solution (Qiagen) 5x	6µl
Template DNA (55ng/µl)	2µl
dH_2O	Add to 30µl

Temperature	Time	
95°C	15min	
94°C	30sec	35x
60°C	30sec	
72°C	1min/kb	
72°C	10 min	
4°C	hold	

After gel purification with the Qiaquick Gel extraction kit, the PCR products were digested with NheI-HF and HindIII-HF restriction enzymes for 3h at 37°C and subsequently purified with the MinElute PCR purification kit and ligated in the respectively cut and gel-purified target vectors pGL4.10 or pGL4.23. Transformation and sequence confirmation by Sanger sequencing was conducted as before (chapter 3.13.1). Primers are listed in Table 3-9.

3.14.2 Promoter and enhancer evaluation with dual luciferase reporter assays

For the assessment of promoter and enhancer potential, Hepa1.6 cells were reverse transfected with 40ng of pGL4.10 or pGL4.23 firefly luciferase reporter constructs using 0.2µl GenJet Ver.II-LnCAP and 10ng CMV-*Renilla* luciferase as a transfection normalization

control. The plasmids were separately adjusted to 5µl with plain DMEM and incubated for 5min before mixing. The transfection reagent was diluted in 5µl DMEM and after short mixing immediately added to the plasmids. After incubation for 15min, the transfection mixture was added to the wells of a white 384 well flat bottom plate with ≈6000 cells per well. Luciferase activity was measured after 48h on a Spectramax microplate reader as previously described [125]. Measurements were taken for eight technical transfection replicates of four independent experiments and normalized to the respective pGL4.10 or pGL4.23 empty vector (EV).

3.15 Statistical analyses

Statistical analyses were conducted with GraphPad Prism v5.04 or Excel 2007. Experiments were evaluated by statistical means for experiments with n ≥3 as indicated in Figure legends.

4. Results

4.1 Genome-wide analysis of DNA methylation changes in the C3(1) mouse model

A major focus of this work was the evaluation of the DNA methylation kinetics in the C3(1) mouse model of breast cancer. We chose this model because of its continuous progression of tumor formation with stages that resemble human breast cancer at the histological level. Besides, there is a set of cell lines available for functional analysis that was established from various stages of tumorigenesis in this model [111]. In order to study the kinetics of tumor formation, we collected tumors and age-matched WT mammary glands at 4-week (w) intervals starting at 8 weeks up to 24 weeks of age (Figure 4-1).

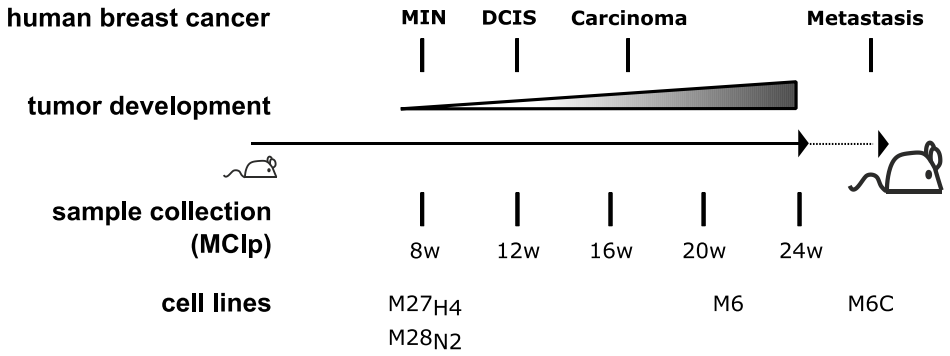


Figure 4-1 Schematic representation of tumor progression in the C3(1) mouse model. The model resembles stages of human breast cancer (MIN, DCIS, invasive carcinoma, metastasis) and the respective time points for sample collection for MCIp are indicated. The cell lines are derived from C3(1) lesions and represent the stages at the specific time points. w: weeks MIN: mammary intraepithelial neoplasia; DCIS: ductal carcinoma in situ.

Three samples per age group of mammary glands and tumors were used in a genome-wide approach to enrich for methylated regions using MCIp-seq. The quality of the data obtained by sequencing was tested by a number of steps that involved bioinformatics followed by technical and biological validation (Figure 4-2).

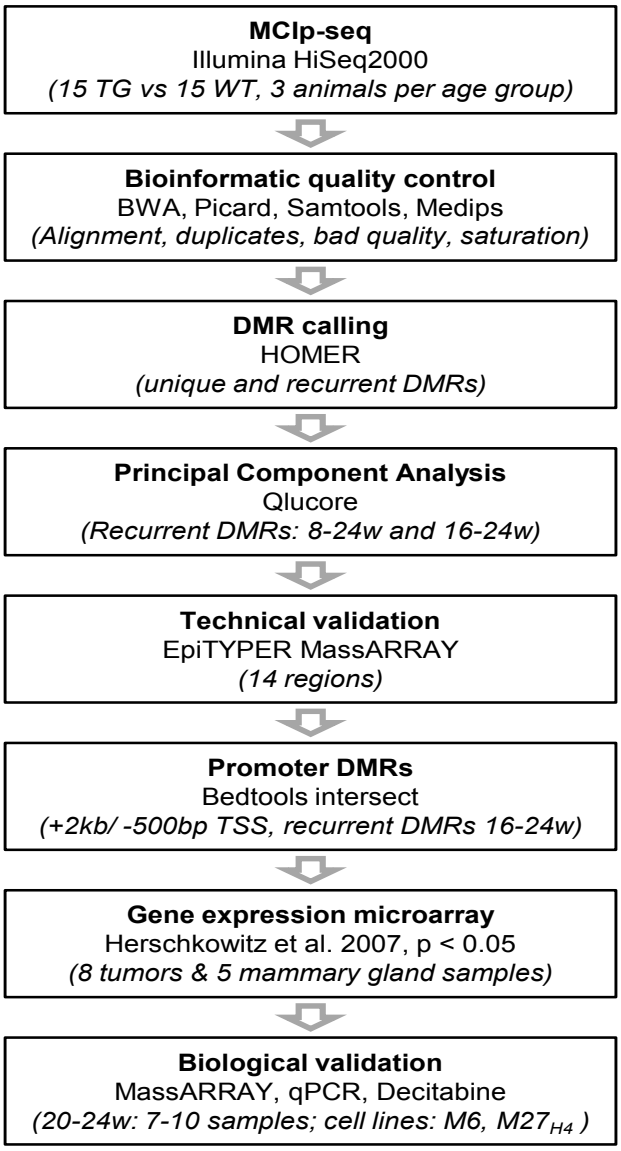


Figure 4-2 Workflow for generation and analysis of MCIp-seq data and validation of DMRs. Detailed description in chapter 4.1.1 - 4.1.3

4.1.1 DNA methylome mining using MCIp-seq data

4.1.1.1 Bioinformatic quality control

The process of bioinformatic quality control with respect to unique alignment and removal of duplicate and bad quality reads is described in detail in the material and methods section (chapter 3.5.1).

Individual samples contained 2-4.5 Mio unique reads with saturation correlation values of 0.75-0.92 (Supplementary Table 1). Saturation values estimate the reproducibility and complexity of the analyzed MCIp-seq library. Values below 0.7 are considered as low quality,

although no general rule exists. Since we measured three animals per age group and at least two animals in each group had saturation values higher than 0.85, we decided to include even samples with lower saturation in our analysis.

4.1.1.2 DMR calling and principal component analysis

After bioinformatic quality control of the sequencing data, we wanted to identify differentially methylated regions, which are defined for MCIp as differences in the enrichment between the samples. Calling of DMRs was achieved with HOMER findPeaks looking for a 2-fold change in enrichment between TG and WT samples in an age-matched fashion. We distinguished between DMRs unique for a single age group or those recurrent in more than one age group. The overall number of DMRs increased with age for both the recurrent and individual DMRs (Figure 4-3) starting with approximately 10.000 DMRs in the young animals (8-12w) and increasing up to 30.000 DMRs for 24w animals. The recurrent DMRs contributed about one third to the DMRs in each age group. The largest increase in DMRs occurred between 12 and 16 weeks of age, indicating this stage as an important period for DNA methylation changes during tumorigenesis.

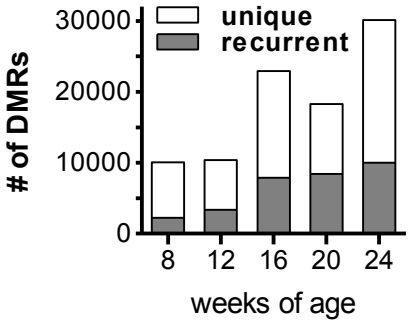


Figure 4-3 Increase in DMRs during tumor progression in age-matched samples of the C3(1) mouse model. Unique DMRs only occur in a single age group and recurrent DMRs occur in more than one age group.

Since we considered the recurrent DMRs as biologically most relevant, we annotated read counts to each individual animal and performed a principal component analysis (PCA). Young TG animals grouped closely together with most of the WT animals (Figure 4-4, left panel, indicated by a green circle). The older animals showed a tendency to group further away from this central group both for the TG animals (16-24w, red circle) and for some of the 20-24w WT animals (blue circle). In order to investigate the age groups with respect to distinction between tumor and normal tissue, we limited our analysis to the older animals (16-24w) and indeed observed a clear separation of TG and WT animals (Figure 4-4, right

panel). This suggested that the largest and most consistent methylation differences occur in these age groups. Therefore, in our further analyses we focused on these animals.

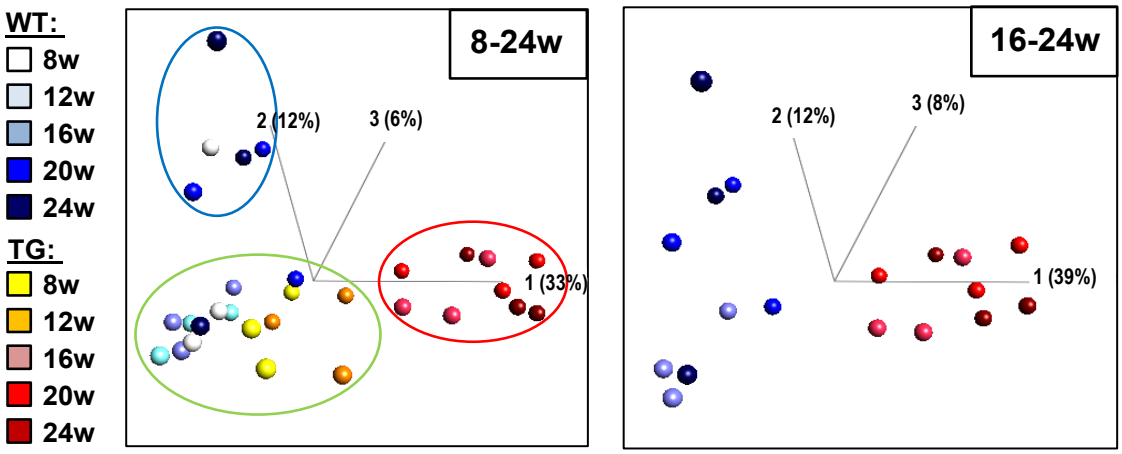


Figure 4-4 Principal component analysis of recurrent DMRs. DMRs (in reads per animal) for age groups from 8-24 weeks (w) distinguish tumors (TG) from normal mammary glands (WT). TG (in shades of red) and WT samples (in shades of blue) separate into two distinct groups from an age of 16 weeks onwards.

4.1.2 Validation of differential methylation with quantitative EpiTYPER MassARRAY technology

4.1.2.1 Technical validation of randomly selected candidate regions

Since MCIp is an enrichment-based method and detects methylation changes in a relative manner, we next tested the reproducibility of the identified DMRs at a quantitative level. We applied the quantitative EpiTYPER MassARRAY technology to the same samples that had been analyzed by MCIp-seq. We selected 14 DMRs occurring in the 24w age group, which overlapped with at least one other age group (16 or 20w). When comparing MCIp-seq reads (as genome browser tracks) with heatmaps of EpiTYPER results, we saw a high concordance on a sample-wise basis. A CGI overlapping with *Cldn6 (Claudin6)* is exemplarily shown as a hypermethylated region, and a CGI overlapping with exon 1 of *Omp (olfactory marker protein)* depicts an example hypomethylated region (Figure 4-5). For specific genomic regions small differences were reproducible between the two techniques, e.g. for the *Cldn6* amplicon, WT animals had lower DNA methylation levels and lower enrichment in the left half of the region and higher enrichment corresponding to higher DNA methylation levels in the right half (Figure 4-5). In order to further confirm concordance between both techniques, we correlated average amplicon methylation levels and

normalized read counts of the DMRs and found significant, positive correlations (> 0.45) for all but one (*Cdh4*) of the examined regions (Figure 4-5+ Supplementary Table 2 Genomic Regions for technical and biological validation, genomic location of all selected DMRs in Supplementary Figure 1). Thus, an overall validation rate of > 90% confirmed a high reproducibility of results between the relative and absolute techniques and confirmed the validity of our DMR selection criteria.

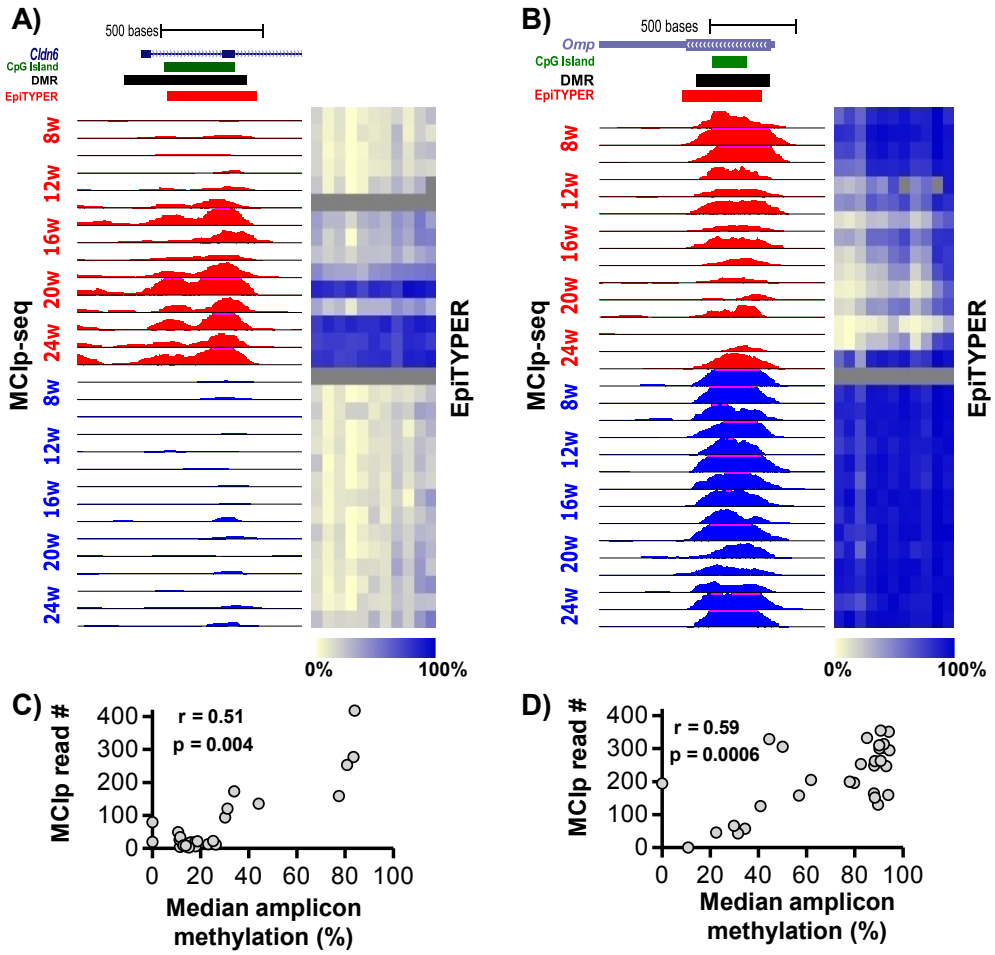


Figure 4-5 Sample-wise technical validation of DNA methylation assessed by MCIp-seq and EpiTYPER MassARRAY. Depicted are recurrent DMRs associated with *Cldn6* A) and *Omp* B) as examples of genomic regions gaining or losing methylation during tumorigenesis. The top part displays the localization of the DMR (black) identified by MCIp in relation to the gene locus (blue), CpG islands (green), and MassARRAY amplicons used for technical comparison (red). Genome browser tracks (left) represent the MCIp reads for three animals per age group (TG: red; WT: blue) and are compared on a sample-wise basis with quantitative MassARRAY data (heatmap, right). Methylation levels in the heatmap are depicted by a color gradient from 0% methylation (light yellow) to 100% methylation (blue). Each square represents one CpG unit consisting of at maximum 4 individual CpGs; grey squares depict failed measurements. Spearman rank correlations were calculated for MCIp reads and median amplicon methylation of the *Cldn6* (C) and *Omp* (D) DMRs.

4.1.2.2 Selection and technical validation of promoter DMRs

Hypermethylation of promoter regions during carcinogenesis is commonly known to result in gene silencing [1]. To investigate this effect in a genome-wide manner, we overlapped DMRs recurrent in the 16-24w age groups with promoter regions (2kb upstream to 500bp downstream of TSSs) of RefSeq-annotated genes (GRC38/mm10) and detected 291 hypo- and 159 hypermethylated promoters (Supplementary Table 3 and Supplementary Table 4). Using published gene expression microarray data for the C3(1) model [2], we retrieved relative expression levels for 322 of those genes. We found significant expression differences (Student's two sided t-test, $p < 0.05$) between tumor and normal tissue for 114 genes. Of these, more than 70% (80 genes) had hypomethylated promoters, which were associated with upregulation of 50 genes (62.5%) and downregulation of 30 genes (37.5%). Among the genes with hypermethylated promoters, the majority (24 genes, 70%) was downregulated, and 10 genes (30%) were upregulated in tumors versus normal mammary glands.

Among the hypermethylated genes with significant expression differences, we selected the top three most downregulated ones, namely *Gsn* (*Gelsolin*), *Igfbp6* (*Insulin-like growth factor binding protein 6*), and *Rbpms2* (*RNA binding protein with multiple splicing 2)* for further investigation (Supplementary Table 3). The selection of these regions was supported by gene expression changes in human basal-like breast cancer [2], for which we required the same trend as in the mouse model. This criterion was true for the top hypermethylated candidate genes, but led to the exclusion of some top hypomethylated candidate genes, which were upregulated in the mouse model, but downregulated in human samples.

Due to its dependence on enrichment for detection of methylation differences, MCIp-seq requires a consistent read coverage for reliable detection of DMRs. Especially hypomethylation can be over-interpreted if enrichment for background methylation (in our case in normal mammary glands) is highly variable. Consequentially, some of the top upregulated candidate genes (*Marveld2, Rtkn, Vasn*) were disregarded due to high enrichment variability for the respective regions in WT samples (Supplementary Table 4). Cell-type specific marker proteins such as *Krt7* or *Krtcap3* were also excluded. Finally, we selected *Pkp3* (*Plakophilin 3*), *Elf3* (*E74 like factor 3*), *Unc5b* (*Unc-5 netrin receptor b*) and *Cldn4* (*Claudin 4*) as hypomethylated candidate genes.

All EpiTYPER assays for the selected regions performed well, except for *Rbpms2,* which had a strong PCR bias for methylated DNA. Methylation patterns of the *Rbpms2* amplicon were also highly heterogeneous; therefore the gene was excluded from analysis. For the other regions EpiTYPER methylation levels and MCIp sequencing reads were significantly

positively correlated with Spearman rho coefficients from 0.57 to 0.82 (Supplementary Table 2). While for all amplicons methylation levels in the WT samples were relatively stable over time, we confirmed a more or less gradual loss or gain in methylation for TG animals with increasing progression of tumor development (Figure 4-6 and Table 4-1).

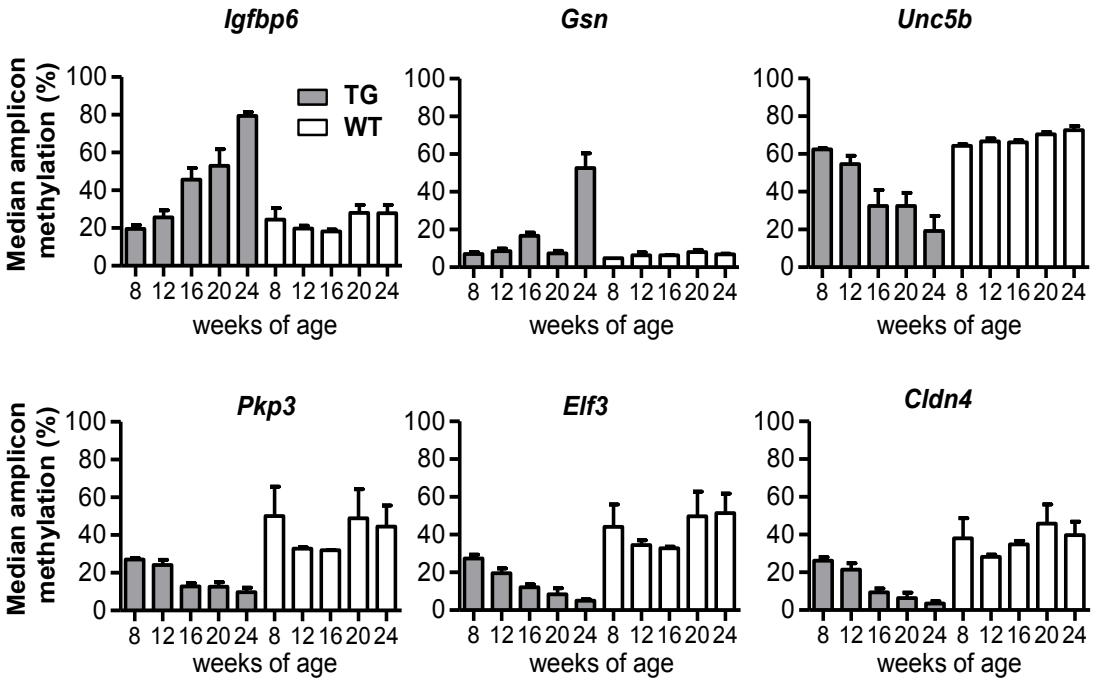


Figure 4-6 Carcinogenesis-induced DNA methylation changes in *Igfbp6, Gsn, Unc5b, Pkp3, Elf3,* and *Cldn4* promoters. Validation of gradual changes in DNA methylation by EpiTYPER MassARRAY for candidate promoter DMRs detected by MCIp-seq. Depicted is the average + SEM of median amplicon methylation for 2-3 individual animals per age group. Statistical analysis in Table 4-1.

Kruskal Wallis analyses confirmed significant changes in DNA methylation in the TG animals over time, whereas methylations levels in the WT groups did not significantly change (Table 4-1).

Table 4-1 Kruskal Wallis group-wise statistical analysis of methylation changes in selected promoter regions.

Gene	*Igfbp6*	*Gsn*	*Unc5b*	*Pkp3*	*Elf3*	*Cldn4*
TG #	0.01 *	0.03 *	0.03 *	0.04 *	0.02 *	0.02 *
WT #	0.18 ns	0.21 ns	0.07 ns	0.51 ns	0.86 ns	0.56 ns

reported are p-values of the Kruskal-Wallis analysis over the age groups in either TG or WT animals.

* p< 0.05, ns: not significant.

4.1.3 Promoter methylation and functional impact on gene expression

4.1.3.1 Biological validation of promoter methylation and gene expression differences

We selected an additional set of ten TG and seven WT samples from the 20 and 24w age groups to estimate the biological reproducibility of promoter methylation levels and to confirm expression differences indicated from the microarray data. Variability of median methylation levels in WT samples was low for *Igfbp6, Gsn* and *Unc5b,* but deviating up to 30% around the group average for *Pkp3, Elf3,* and *Cldn4* (Figure 4-7A). Nevertheless, we confirmed the DNA methylation levels measured at 20-24w in the discovery samples (marked in red) and detected significant methylation differences between TG and WT samples of approximately 40% for all candidates regions, with the exception of *Gsn* for which the difference was only about 20% (Figure 4-7A). This lower difference in methylation levels for *Gsn* can be explained by the fact that *Gsn* methylation increased only in the 24w group during carcinogenesis, whereas methylation levels at 20 weeks were still in the range of the WT animals (Figure 4-6).

When evaluating mRNA expression levels of the candidate genes, we detected a significant increase in tumor tissue in the range of 4-6 fold for *Unc5b, Pkp3, Elf3,* and *Cldn4,* and 20-10,000-fold downregulation for *Igfbp6* and *Gsn*. For *Gsn,* we investigated two transcript variants (v1 and v2/4) and only the variant v1, which displayed promoter hypermethylation, was downregulated in tumors (Figure 4-7B).

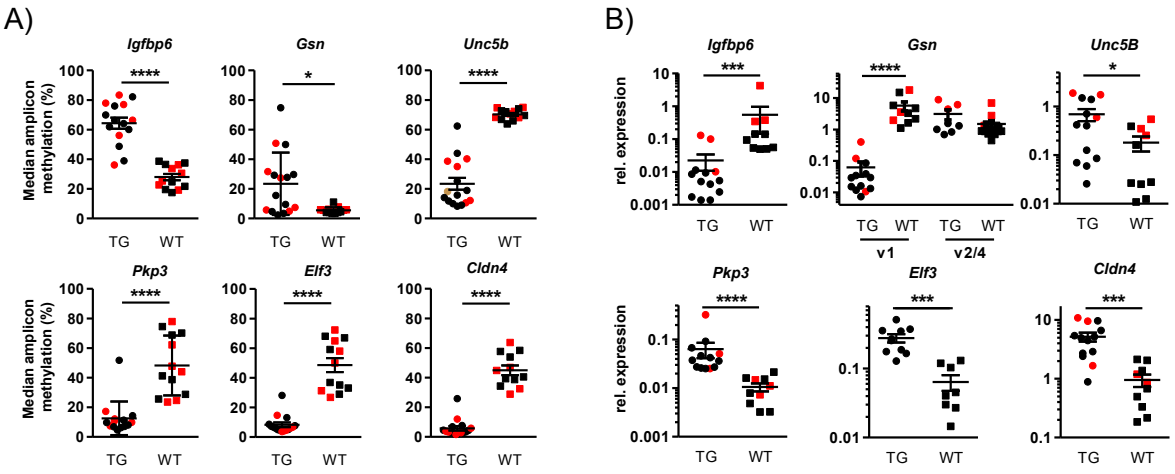


Figure 4-7 Promoter methylation and expression levels of candidate regions in animals aged 20-24w. A) Median amplicon methylation levels determined by EpiTYPER MassARRAY as in Figure 4-5. B) Gene expression levels relative to three reference genes (*Hprt1, Tbp, Actb*). Circles: tumors (TG); squares: mammary glands (WT). Red symbols indicate samples used for discovery of DMRs by MCIp-seq. For *Elf3,* RT-qPCR measurements were only conducted for the validation sample set due to insufficient RNA availability for the discovery set. Depicted are the individual samples and the mean ± SEM. Student's t-test (two-sided), * p< 0.05, *** p< 0.001, **** p< 0.0001.

4.1.3.2 Correlation between DNA methylation and gene expression

For four out of six of the investigated genes, we measured a significant inverse correlation between expression levels and promoter DNA methylation (Figure 4-8) with Spearman correlation coefficients in the range of -0.62 to -0.82. Correlation analysis of *Unc5b* did not indicate an influence of promoter methylation on expression levels. For *Gsn* the situation appeared to be different: all the TG animals (circles) had low expression levels irrespective of their methylation level (≈ 5%- 75%), whereas the region was unmethylated in WT animals, and the generally high expression levels of *Gsn* v1 varied up to 10-fold. This might indicate that expression of *Gsn* is regulated in WT animals by a transcription factor, which binds the unmethylated region and modulates expression levels. Absence of the respective factor in tumors would explain the low levels of *Gsn* expression and allow a gradual gain in promoter methylation during carcinogenesis.

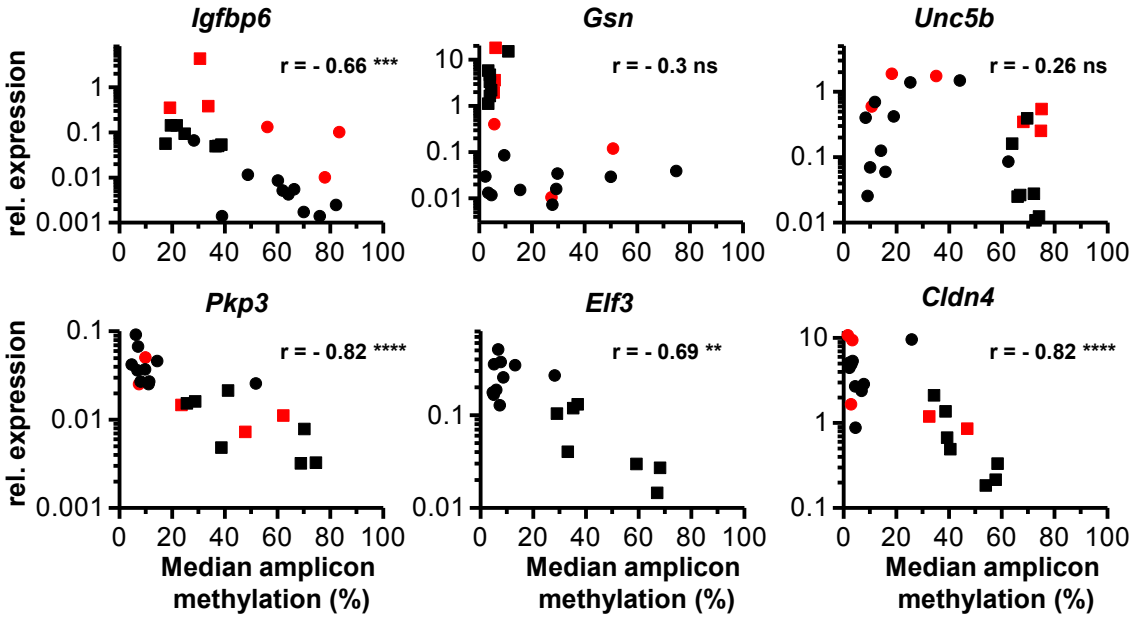


Figure 4-8 DNA methylation and gene expression levels are inversely correlated for candidate genes. For *Gsn*, correlation was calculated for expression of transcript variant 1. Samples and symbols as in Figure 4-7. Spearman's rank correlation, ** p< 0.01, *** p< 0.001, **** p< 0.0001

4.1.3.3 Decitabine treatment of C3(1) cell lines

Spearman correlation results supported a role of DNA methylation for regulation of expression of most of our candidate genes. In order to evaluate this relationship, we treated cell lines derived from C3(1) tumors (M6, $M27_{H4}$) with Decitabine (Dac) and measured expression changes. The M6 cell line was chosen as it represented the late stages of tumor development in the C3(1) model and was highly methylated for *Igfbp6* and *Gsn* (> 90%) and

weaker methylated for *Unc5b* (40%). Many of the hypomethylated candidates (*Pkp3, Elf3, Cldn4*) were completely unmethylated in the M6 cells and therefore were studied in the $M27_{H4}$ cell line with basal methylation levels between 20% (*Pkp3*) and 50-60% (*Elf3, Cldn4*) (see Figure 4-9A and compare Figure 4-1). Dac treatment resulted in a decrease in methylation of the target regions in $M27_{H4}$ cells by 5-15% and in M6 cells by 20-60%.

Upon loss of methylation after Dac treatment we observed upregulation of all transcripts at variable levels except for *Unc5b*. This confirmed our previous observations (Figure 4-8) that expression levels of *Unc5b* were not dependent on promoter methylation.

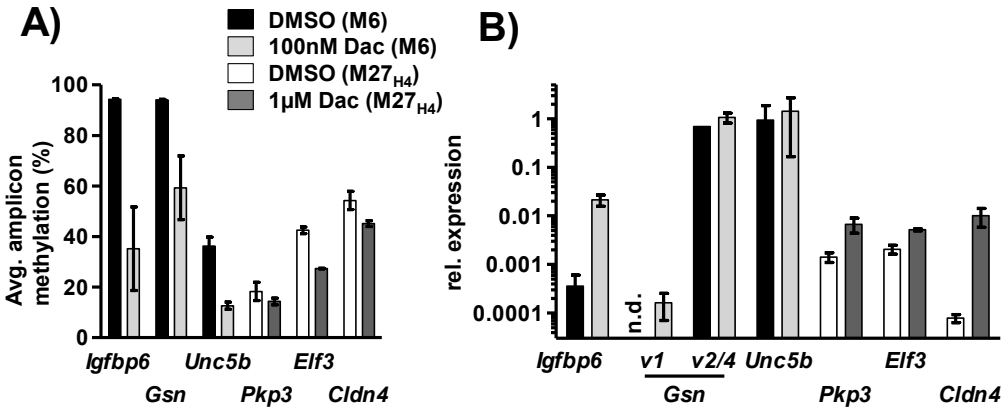


Figure 4-9 Promoter demethylation induces reexpression of candidate genes. A) Treatment of cells with 100nM (M6) and 1µM ($M27_{H4}$) Dac decreased DNA methylation levels for candidate genes. Depicted are the average amplicon methylation levels with the mean ± SD of 2 independent experiments. B) Dac treatment induces reexpression of candidate genes in M6 and $M27_{H4}$ cells. Depicted are expression levels relative to three reference genes (*Hprt1, Tbp, Actb*) as the mean ± SD of 2 independent experiments and thus data were not evaluated by statistical means.

The low level of methylation for *Pkp3* was probably the reason for the minor methylation changes upon Dac treament, but the change was sufficient to induce a 4-fold upregulation of the transcript (Figure 4-9B). Basal expression of *Gsn* v1 in M6 cells was below the detection limit of RT-qPCR, but upon Dac exposure *Gsn* v1 levels passed the detection threshold, while the other *Gsn* transcripts were unaffected and remained highly expressed. This suggested a role of DNA methylation in the regulation of this gene in the cell line, despite the fact that we could not observe significant correlation between expression levels and DNA methylation in tissue samples (Figure 4-8). *Igfbp6, Elf3,* and *Cldn4* transcripts were upregulated ranging between 2-fold to more than 120-fold upon treatment with Dac. Taken together, these results suggested that promoter methylation in these genes contributes to the regulation of their expression.

4.1.4 Comparison of methylation profiles for C3(1) and human breast cancer

So far, the C3(1) model was compared to human breast cancer only at the level of histology, gene- and miRNA expression [90, 92], but not for epigenetic modifications. We selected promoter regions as a set of candidate regions for interspecies comparison of DNA methylation, since they were found to have higher functional conservation in form of transcription factor binding between mouse and human tissues than other more distal regulatory elements [126]. We evaluated 450k data for human breast cancer samples published by 'The Cancer Genome Atlas' consortium (TCGA) [70]. From the genes with a DMR in the C3(1) model, we selected those genes with a known human homologue and chose the 4149 probes in the respective promoters of 355 genes, among which 133 were hypermethylated and 222 were hypomethylated in the mouse (for more details see Material & Methods chapter 3.9). From the sample set of 664 primary tumors and 96 normal control samples, we only investigated 215 tumors with known Pam50 classification to analyze the similarity of DNA methylation patterns between mouse and human breast cancer subtypes. A principal component analysis of tumor samples with all 4149 probes distinguished the 41 basal-like breast cancer samples (red) from all other breast cancer subgroups (Figure 4-10A).

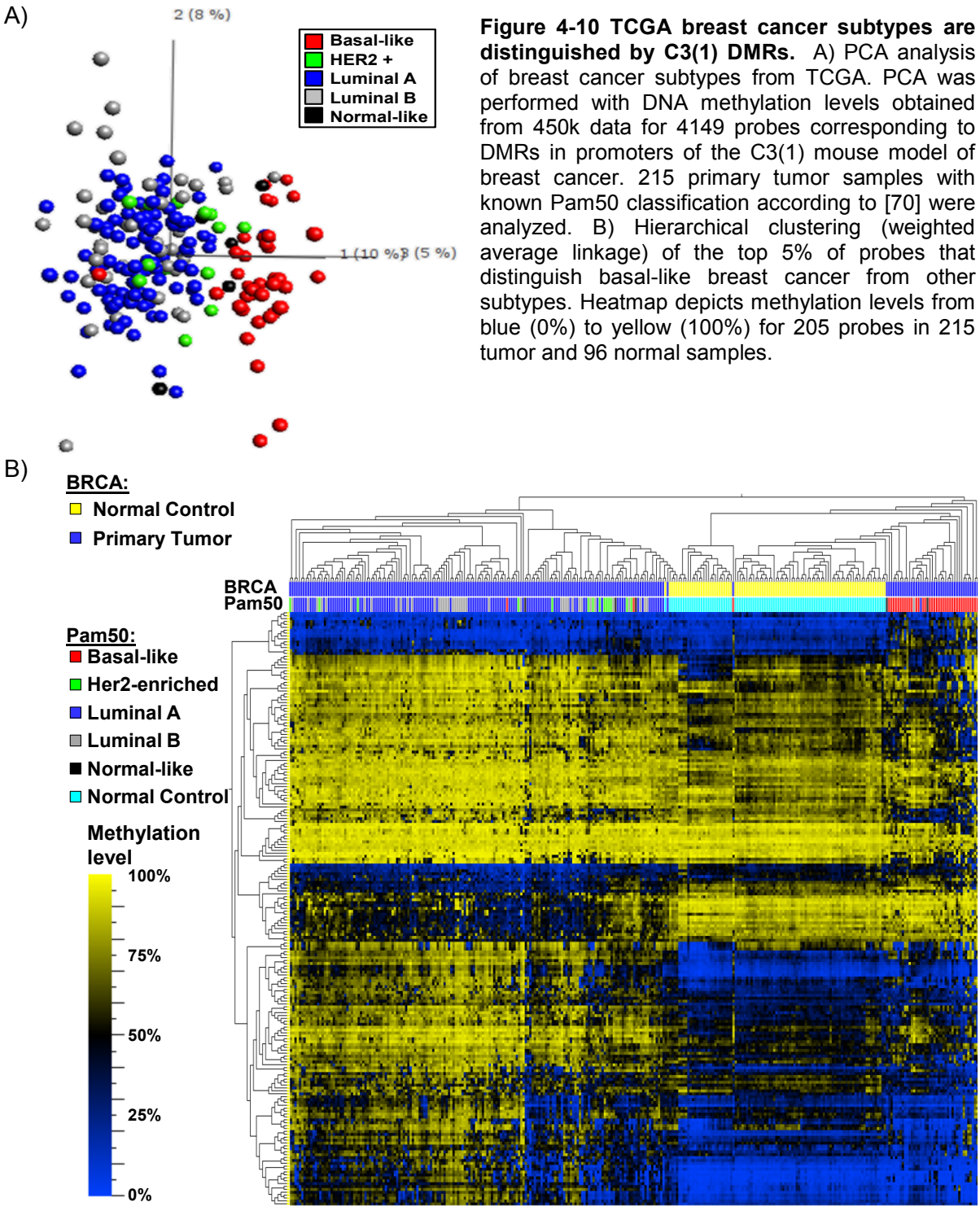


Figure 4-10 TCGA breast cancer subtypes are distinguished by C3(1) DMRs. A) PCA analysis of breast cancer subtypes from TCGA. PCA was performed with DNA methylation levels obtained from 450k data for 4149 probes corresponding to DMRs in promoters of the C3(1) mouse model of breast cancer. 215 primary tumor samples with known Pam50 classification according to [70] were analyzed. B) Hierarchical clustering (weighted average linkage) of the top 5% of probes that distinguish basal-like breast cancer from other subtypes. Heatmap depicts methylation levels from blue (0%) to yellow (100%) for 205 probes in 215 tumor and 96 normal samples.

Since the 450k dataset covered the complete promoter regions (2kb upstream to 500bp downstream of the TSS with an average of 12 probes per gene), many CpG sites depicted similar methylation levels across all samples. Thus, for higher discriminatory power between the subtypes, we selected the top 5% probes that discriminated basal-like samples from the other subtypes. When including the normal control samples in a hierarchical clustering analysis, the basal-like group was distinct from other breast cancer subtypes and clustered close to the control samples (Figure 4-10B). The top 5% of probes were located in a set of 85 genes. We compared this gene list for overlap with published gene sets available in the Molecular Signatures Database collection (MSigDB v5.1). We found overlaps with several gene sets from the C2 collection of curated gene sets (GCP: chemical and genetic perturbations, 3396 gene sets). The top 15 most significantly enriched sets are listed in Table 4-2. These sets included high or intermediate CpG density promoters including those that were marked by H3K4me3 or H3K27me3 in ES cells [127], but also gene sets linked to breast development [128] and breast cancer, in particular the basal-like subtype [129].

Table 4-2 Gene set enrichment analysis for genes with promoter DMRs among the top 5% probes that distinguish basal-like breast cancer.

	Geneset name	Description	Ratio *	p-value	Genes
1	MIKKELSEN_ES_ICP_WITH_H3K4ME3	Genes with intermediate-CpG-density (ICP) promoters bearing histone H3 K4 trimethylation mark (H3K4me3) in embryonic stem cells (ES).	0.0167	4.41E-09	KRT17, TRIM29, TRIM2, PTGES, C1QTNF1, ALDOC, NME5, ZP3, ASPRV1, GDF5, HNF1A, CPSF4L
2	SMID_BREAST_CANCER_BASAL_UP	Genes upregulated in basal subtype of breast cancer samples.	0.017	1.75E-08	KRT17, TRIM29, TRIM2, CSPG4, SCHIP1, KRT4, DZIP1, MFAP2, HSPB2, SPIB, COL11A2
3	ISSAEVA_MLL2_TARGETS	Genes downregulated in HeLa cells upon knockdown of MLL2 [GeneID=8085] by RNAi.	0.0806	8.43E-08	KRT17, TRIM29, CSPG4, SCHIP1, GSN
4	KOINUMA_TARGETS_OF_SMAD2_OR_SMAD3	Genes with promoters occupied by SMAD2 or SMAD3 [GeneID=4087, 4088] in HaCaT cells (keratinocyte) according to a ChIP-chip analysis.	0.0133	1.96E-07	KRT17, PTGES, KRT4, BCAT1, PLEC, SAT1, DKK3, CLU, KRT8, TNFRSF1A, HOXD11
5	HATADA_METHYLATED_IN_LUNG_CANCER_UP	Genes with hypermethylated DNA in lung cancer samples.	0.0205	4.26E-07	C1QTNF1, ALDOC, BCAT1, HPN, BCOR, GFPT2, JPH2, ST6GAL2
6	PURBEY_TARGETS_OF_CTBP1_NOT_SATB1_DN	Genes downregulated in HEK-293 cells (fibroblast) upon knockdown of CTBP1 but not of SATB1 [GeneID=1487, 6304] by RNAi.	0.0179	1.20E-06	NME5, KRT4, DZIP1, PLEC, SAT1, DKK3, TBX1, SLC5A1
7	LEI_MYB_TARGETS	Myb-regulated genes in MCF7 (breast cancer) and lung epithelial cell lines overexpressing MYBL2, MYBL1 or MYB [GeneID=4605;4603;4602].	0.022	1.47E-06	KRT17, PLEC, SAT1, CLU, KRT8, SLC16A3, NUPR1
8	NIKOLSKY_BREAST_CANCER_17Q21_Q25_AMPLICON	Genes within amplicon 17q21-q25 identified in a copy number alterations study of 191 breast tumor samples.	0.0209	2.07E-06	C1QTNF1, SLC16A3, HOXB3, HOXB8, HOXB1, HOXB9, UTS2R
9	MCBRYAN_PUBERTAL_BREAST_3_4WK_UP	Genes upregulated during pubertal mammary gland development between weeks 3 and 4.	0.028	2.14E-06	ALDOC, CSPG4, GSN, CLU, KRT8, TNFRSF12A
10	MIKKELSEN_NPC_HCP_WITH_H3K27ME3	Genes with high-CpG-density promoters (HCP) bearing histone H3 trimethylation mark at K27 (H3K27me3) in neural progenitor cells (NPC).	0.0205	2.33E-06	HOXD11, HOXB3, HOXB8, HOXB1, HOXB9, PRKCZ, GALNT13
11	BENPORATH_ES_WITH_H3K27ME3	Set 'H3K27 bound': genes possessing the trimethylated H3K27 (H3K27me3) mark in their promoters in human embryonic stem cells, as identified by ChIP on chip.	0.0098	3.81E-06	GSN, PLEC, TBX1, HOXB3, HOXB8, HOXB1, RARA, INA, MAB21L1, GBX2, HIC1
12	SMID_BREAST_CANCER_LUMINAL_B_DN	Genes downregulated in the luminal B subtype of breast cancer.	0.0142	6.54E-06	KRT17, TRIM29, TRIM2, DZIP1, MFAP2, HSPB2, SPIB, SAT1
13	SMIRNOV_CIRCULATING_ENDOTHELIOCYTES_IN_CANCER_UP	Genes upregulated in circulating endothelial cells (CEC) from cancer patients compared to those from healthy donors.	0.0316	8.80E-06	PTGES, BCAT1, DKK3, CLU, RARA
14	OXFORD_RALA_OR_RALB_TARGETS_DN	Genes downregulated after knockdown of RALA or RALB [GeneiD=5898;5899], which were also differentially expressed in bladder cancer compared to normal bladder urothelium tissue.	0.125	1.00E-05	KRT17, CLU, SLC16A3
15	LIM_MAMMARY_STEM_CELL_DN	Genes consistently downregulated in mammary stem cells both in mouse and human species.	0.0164	1.02E-05	KRT8, HPN, SLC5A1, PRKCZ, PSME1, RORC, FAM110A

*(# of genes in overlap/ #genes in gene set)

Clustering according to the mouse DMRs revealed that the basal-like samples are different from other subtypes and similar to normal breast tissue. Since the DMRs in the mouse covered only part of the promoter region we extended the analysis to the full 2.5kb regions and annotated the average MCIp read density in the 20 and 24w animals (Figure 4-11). Hierarchical clustering of the promoter reads still separated the TG and WT animals into two clusters. For some promoter regions the MCIp-seq coverage in the animals was low, which most likely indicated that the DMR was weak or only restricted to a narrow part of the promoter. However, the analysis still showed a higher prevalence of hypomethylated regions in the tumors, which resembled the pattern of hypomethylation in basal-like breast cancer [70]. According to the read coverage, the complete promoter regions were similar to the pattern of the DMRs.

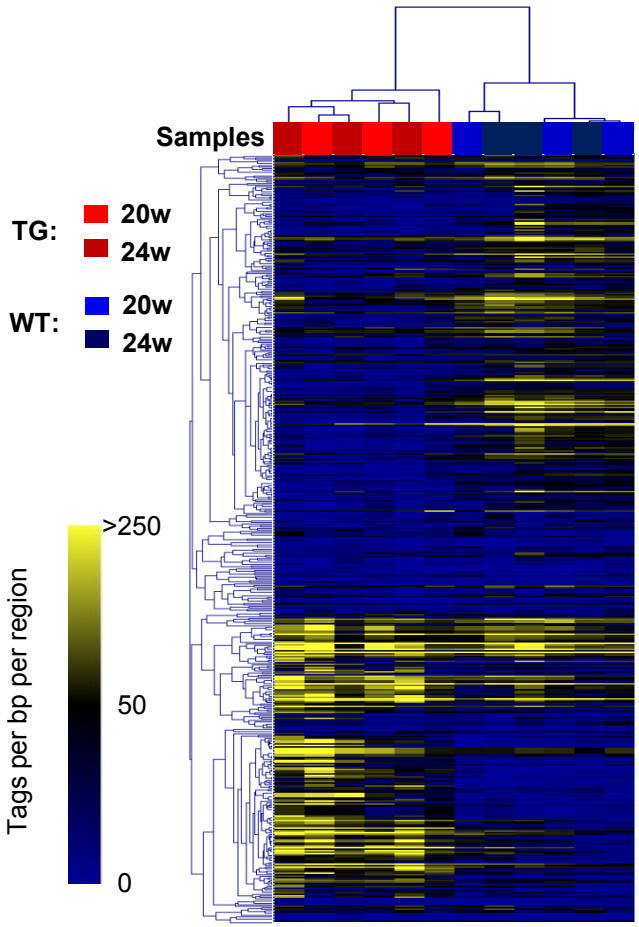


Figure 4-11 Hierarchical clustering of differentially methylated promoter regions according to MCIp-seq coverage. Each row depicts the MCIp-seq coverage in tags per bp per region for an individual promoter according to a colorcode from blue to yellow. The columns depict the TG (red) and WT (blue) samples. Hierarchical clustering was performed using average linkage according to Spearman's rank correlation.

For a further validation of mouse DMRs in human breast cancer, we first investigated the degree of differential methylation at the 355 genes for the four molecular subtypes compared to normal controls (chapter 3.9.1). The normal-like subtype only contained five samples and thus was excluded from the comparison. The highest number of significant DMRs ($p< 0.01$) among the 355 genes was detected for the luminal subtypes with 261 and 240 DMRs for

type and type B respectively (Figure 4-12). In contrast, for the basal-like samples only 50% of the genes were differentially methylated and even less in the HER2 subtype. Since methylation levels for basal-like samples were described to be similar to normal breast tissue, which we also observed in the hierarchical cluster analysis, this explained the relatively low number of DMRs for this subtype. The HER2 group only consisted of 13 patients, which might account for the low number of DMRs in this subgroup. Overall, about 25% of the evaluated genes displayed concordant loss or gain of methylation between C3(1) tumors and all human breast cancer subtypes, with an equal proportion of hyper- and hypomethylated regions. This suggested low, but evident similarity between the mouse model and human breast cancer at the level of DNA methylation. About the same number of genes also had a similar methylation status among all subtypes, but did not reflect the methylation change detected in the C3(1) samples. For the remainder of the genes the methylation signature was variable between the subtypes. However, an assignment of the C3(1) model to a specific molecular subtype was not possible. The apparent high degree of similarity in terms of methylation levels of the basal-like subtype and normal controls introduced a bias when analyzing human DMRs, as regions might be variable between the subtypes, but not between basal-like tumors and normal controls. As the frequency of DMRs is already much higher in the luminal subtypes, the likelihood to observe a consistent methylation change between a human subtype and the C3(1) model would be higher for the luminal than for the basal-like subgroups. Thus, subtype assignment would be more reliable when based on quantitative methylation levels for both the mouse model and the human samples to allow a direct comparison.

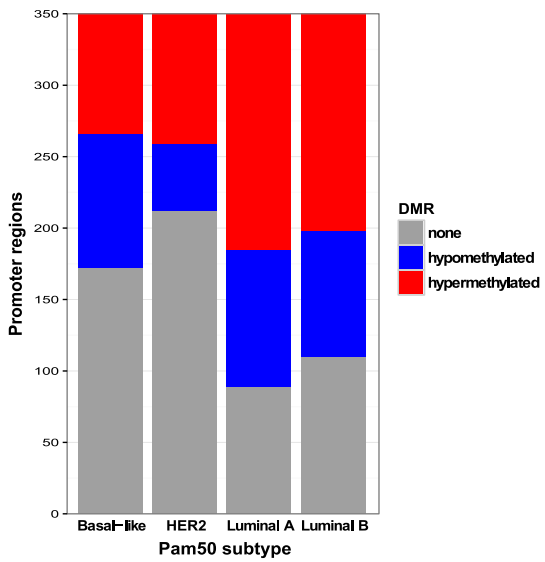


Figure 4-12 Promoter DMRs in human breast cancer subtypes. Depicted is the respective number of DMRs for the 355 genes with mouse promoter DMRs in the different molecular breast cancer subgroups. Hypermethylated regions are in red, hypomethylated regions in blue and in grey regions with no significant methylation change. n= number of samples in subtype.

This figure was prepared by Dr. Yassen Assenov.

4.1.5 Characterization of recurrent DMRs

In our analysis so far, we characterized the DMRs without considerations for the kinetics or direction of methylation changes. Thus, for a more detailed evaluation we separated the DMRs into hyper- and hypomethylated regions. Similar to our previous observation, the number of DMRs increased with tumor progression over time. With increasing age, we observed more hypomethylated than hypermethylated regions (Figure 4-13A). While the number of unique DMRs with gain or loss of DNA methylation was relatively similar per age group, recurrently hypomethylated regions were twice as frequent as hypermethylated ones, especially in age groups 16-24w and when combining all age groups (Figure 4-13B). Currently we cannot exclude that the unique DMRs represent technical noise due to variable enrichment by MCIp.

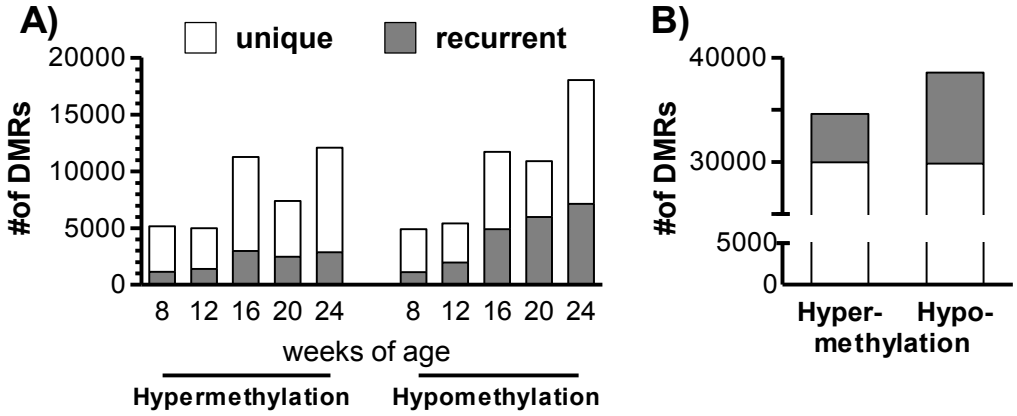


Figure 4-13 Frequency of hyper- and hypomethylated regions. Frequency of DMRs is depicted per age groups A) and in total B).

4.1.5.1Recurrent DMRs form specific subgroups

In order to subgroup the recurrent DMRs further, we defined a group of i) **progressive DMRs** that occur in all age groups up to 24 weeks from the time point of first occurrence, ii) **continuous DMRs** with occurrence in two or more neighboring age groups except for the 24w group, and iii) **discontinuous DMRs**, for which the sequential occurrence of DMRs was interrupted (Figure 4-14A). Again, the frequency of DMRs per subgroup was different for hyper- and hypomethylated regions. Especially the progressive DMRs were more frequently hypomethylated, contributing by almost 50% to the recurrent DMRs (Figure 4-14B).

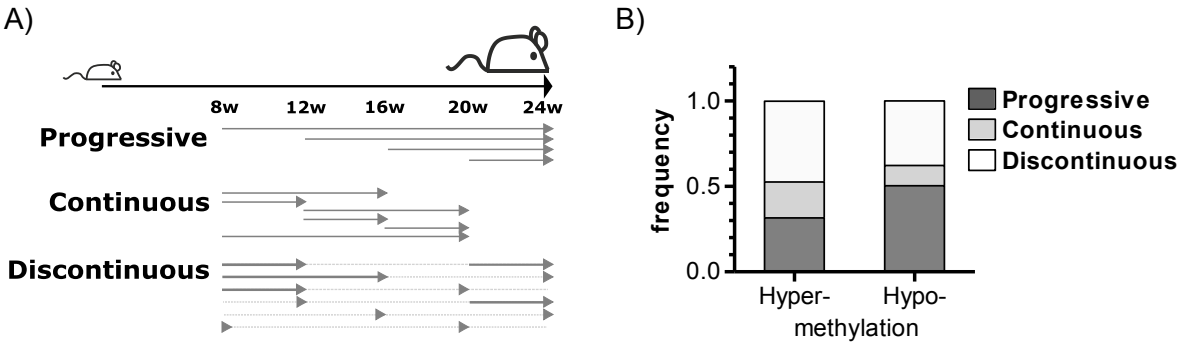


Figure 4-14 Characterization of recurrent DMRs. A) Scheme for categorizing recurrent DMRs according to their occurrence in neighboring age groups into progressive, continuous and discontinuous DMRs. B) Distribution frequency of recurrent DMRs among hyper- and hypomethylated regions.

In addition, when comparing the MCIp-seq coverage at the three recurrent DMR types between the different age groups, the largest changes occurred in the progressive DMRs of TG animals (Figure 4-15A). In this DMR subtype, the loss and gain of methylation progressed gradually in even steps from one age group to the next and the differences in enrichment were most prominent. In contrast, in the WT animals, we observed a gradual age-related drop in MCIp-seq coverage mainly for the continuous and discontinuous DMRs, whereas the profiles for the progressive DMRs were similar in all age groups (Figure 4-15B). This development of the DMRs during tumorigenesis was also observed for the individual animals of the different age groups in the progressive DMRs (Figure 4-15C). Thus, we focused further analyses on the progressive subgroup of the recurrent DMRs.

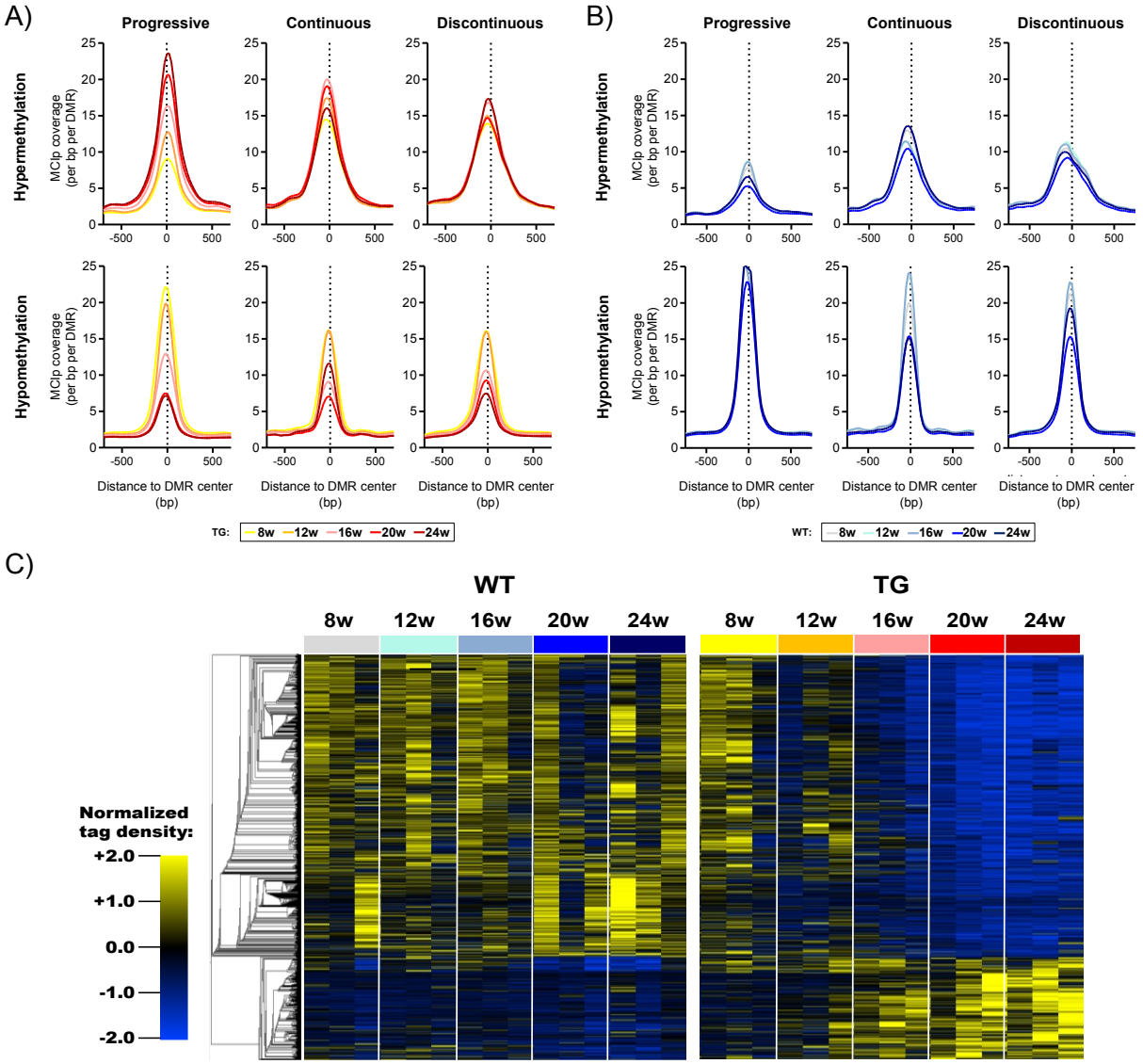


Figure 4-15 Sequencing coverage of recurrent DMRs. A,B) Average MCIp tag coverage (per bp per DMR) of the different age groups around the center of different recurrent DMR subgroups for TG animals A) and WT animals B). Depicted is the average coverage of three animals in the age group. C) MCIp coverage of progressive DMRs for the individual animals of the different age groups. Tag density was normalized according to the mean and the variance of each individual DMR and is depicted in a color coded gradient from blue to yellow.

4.1.5.2 Relative distribution of DMRs in the context of genomic features

In order to evaluate potential functions for the DMRs, we annotated the regions according to their closest genomic feature using HOMER. The distribution of hyper- and hypomethylated **unique DMRs** was similar (Figure 4-16A). Only 3% of the roughly 30.000 regions were located in TSS-associated features (CpG island, promoter, 5' untranslated region (5'UTR)), while 41-48% were located in exons and introns. The remainder of regions was distributed

between repeat elements (1/4 to 1/3) and intergenic regions (1/5). This highly similar distribution and lack of enrichment for any feature further hints that the unique DMRs might represent background noise.

For the **recurrent DMRs,** we focused on the **progressive subgroup** and observed almost three times more hypomethylated regions than hypermethylated (Figure 4-16B). Again, about half of the differential hyper- and hypomethylated regions were located in exons and introns. The largest difference was seen for regions in TSS-associated features: 20% of the hypermethylated regions, but only 4% of the hypomethylated regions were annotated to this feature, consistent with a gradual gain in methylation in gene promoters during carcinogenesis. In contrast to the unique DMRs, only few of the progressive DMRs (7-12%) mapped to repetitive sequences. Almost one third of the hypomethylated regions were located in intergenic areas. The biological function of these intergenic regions is less well defined, but distal regulatory elements such as enhancers can be situated in this genomic context. Therefore, limiting the analysis of DMRs to promoter regions is likely to miss important regions.

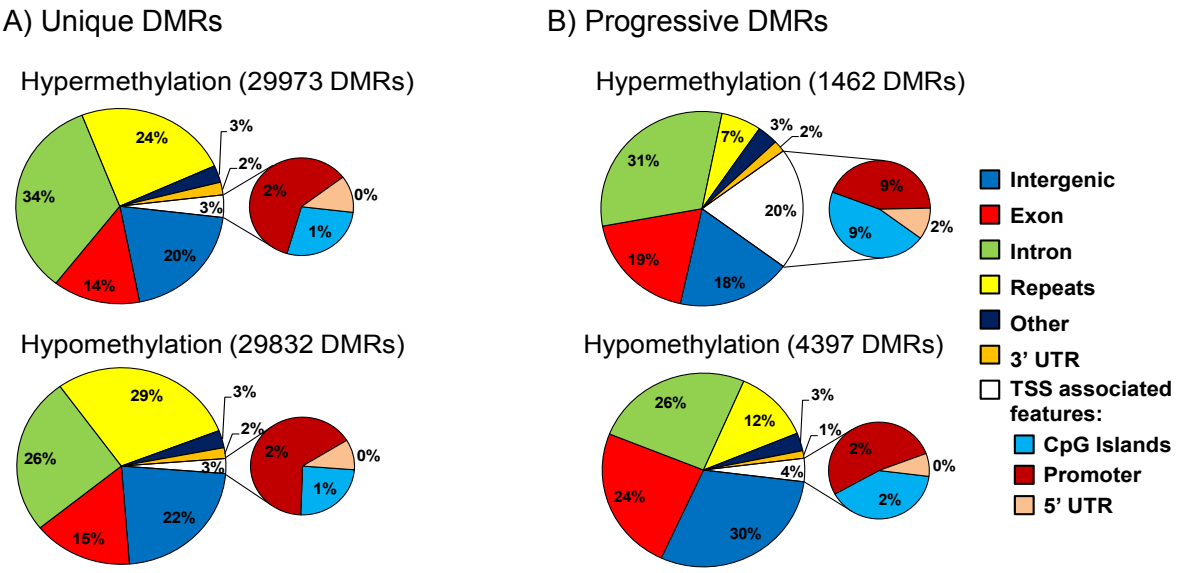


Figure 4-16 Genomic distribution of hyper- and hypomethylated regions. A) DMRs with unique occurrence in any age group. B) Recurrent DMRs with progressive occurrence in all age groups up to 24w of age.

Despite the fact that promoter DMRs only cover part of the changes that occur during tumorigenesis, we wanted to investigate some potential functions of the respective genes and performed a gene set enrichment analysis using the HOMER annotation for the closest gene of the promoter and 5' UTR features. While the hypermethylated genes were enriched for targets of the PRC2 complex and/or marked by H3K27me3 in embryonic stem (ES) cells

(*LRRC3B, ESAM, TCEA3, HOXD1, GNAS, CDH11, NRG1, EPHA5, GABRA2, POU3F4, CDH7, DOK6, HES7, ARL5C, ELAVL2, C1QL1, GABRG3, PRRT1, GSN, PRKCB*), the hypomethylated genes displayed the active H3K4me3 mark at promoters in ES cells (*KRT17, ZMYND8, HNF1A, OPLAH, PKP3, C1QTNF1, GDF5, PADI4, ZNF750, ADAM18, CPSF4L, MYBPC2, MCTS1, TNK1*). Furthermore, we observed opposite patterns for differentially expressed genes in mammary stem cells (MaSCs). MaSC-specific upregulated genes were found to be hypermethylated in tumors of the C3(1) model (*BCOR, GSN, NRG1, FAM184A, LRRN1, PCDH7, SCHIP1, CSPG4, SORBS1, SLC6A8, TMEM47, TSPYL2, DKK3*) and genes downregulated in MaSCs were enriched in hypomethylated promoters (*OPLAH, ZNF750, CLDN7, C1orf210, LYPD3, NIPSNAP3A*). This suggested that the differentiation of MaSCs might be prone to epigenetic deregulation during tumorigenesis in the C3(1) model.

4.1.6 Summary and outlook

In conclusion, we could validate the DNA methylation changes that we identified with our genome-wide approach using MCIp-seq. We further confirmed a link between DNA methylation levels at the promoter and gene expression for five selected candidate genes. Comparison of promoter DMRs from the C3(1) model with TCGA data of human breast cancer samples revealed similarity between mouse and human DNA methylation across different molecular breast cancer subtypes. Moreover, recurrent DMRs were separated into three distinct subgroups with the progressive DMRs demonstrating the largest differences in methylation levels and were almost three times more frequent in hypomethylated regions than in hypermethylated. Gene set enrichment analyses pointed to antagonistic functions of genes with progressively hyper- and hypomethylated promoter regions.

Beside exonic and intronic regions, a large proportion of the progressive DMRs were localized in intergenic regions. In order to better characterize these intergenic DMRs, in particular to identify possible enhancer regions, data for histone marks at these regions are required. It is known that enhancers are highly tissue-specific [35, 121]. Therefore, patterns of enhancer-specific histone marks have to be acquired for the tissue of interest. Two previous studies described changes in the chromatin landscape between mammary epithelial cells of virgin and pregnant mice using ChIP-seq for histone marks H3K4me2, H3K36me3, and H3K27me3 [122, 130], later supplemented by data for H3K4me3 [131]. Since chromatin changes during tumorigenesis are likely to be distinct from those occurring during pregnancy and the histone marks provided in these studies were only of limited suitability to study enhancers, we generated independent ChIP-seq data for the C3(1) mouse model.

4.2 Characterization of the chromatin landscape in the C3(1) model

Our analysis of genome-wide changes in DNA methylation in the C3(1) mouse model revealed that the majority of DMRs occurred in exons, introns or intergenic regions. In order to better understand these epigenetic changes and to characterize potential enhancer regions, we generated a map of chromatin states from ChIP-seq data for histone modifications.

4.2.1 ChIP-seq of histone modifications in the C3(1) mouse model

For the creation of a chromatin map of histone marks, we evaluated three tumor samples from the age group of 20-24w. For comparison, we enriched mammary epithelial cells (MECs) by enzymatic tissue dissociation from a pool of mammary glands of seven age-matched WT animals. We selected mammary epithelial cells, since the SV40T transgene is expressed in this cell type, and tumor formation is correlated with transgene expression [85, 86]. We analyzed by ChIP-seq H3K4me3, H3K4me1, and H3K27ac as histone marks of active chromatin and H3K27me3 as a repressive mark. Alignment and bioinformatic quality control of sequencing reads was performed as for MCIp-seq, omitting the saturation analysis. Depending on the histone marks, we obtained around 17-30 Mio reads per sample (Supplementary Table 5).

Initial examination of candidate regions revealed comparable enrichment of the histone marks for the three tumor samples. We also noticed that gene expression differences for selected genes were reflected in chromatin mark differences between tumors and WT epithelial cells, as depicted exemplary for *Foxc1* or *Mybl2* (Figure 4-17). These two genes are included in the Pam50 classification of breast cancers subtypes and show high expression in human basal-like samples [132]. Expression levels for these genes were also elevated in tumors of the C3(1) model. Consistently, we observed higher enrichment of the active marks H3K4me3 and H3K27ac and loss of repressive H3K27me3 in tumors compared to MECs. For *Foxc1*, the active histone marks were not only limited to the direct proximity of the TSS, but even spanned several kb around the complete transcript. This initial plausibility check of our data supported the validity of our ChIP-seq data.

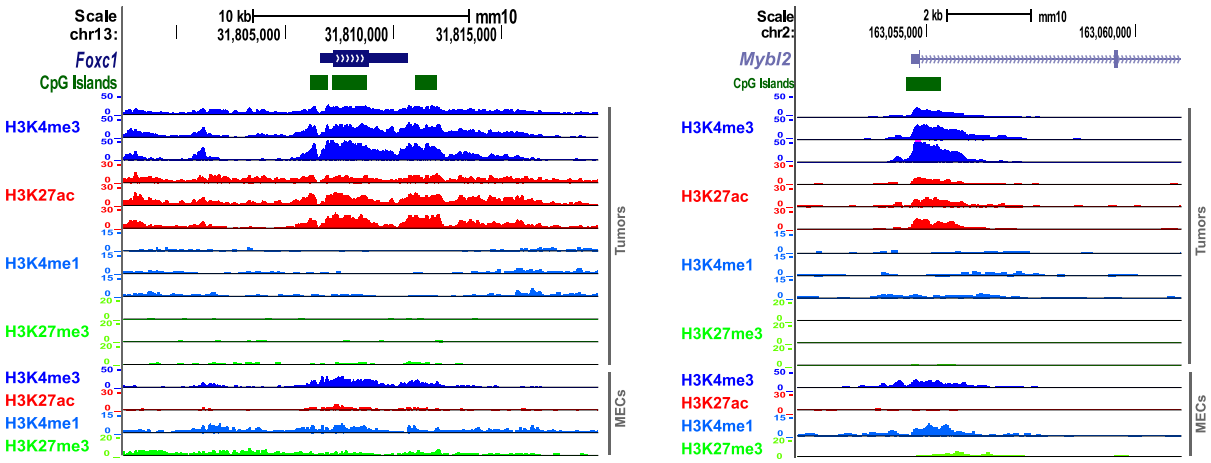


Figure 4-17 Genome browser tracks for ChIP-seq data of histone modifications in three tumors and one MEC pool for regions of *Foxc1* and *Mybl2*. Color code for histone modifications is as follows: H3K4me3, blue; H3K27ac, red; H3K4me1, light blue; H3K27me3, green.

4.2.2 A hidden Markov model defines chromatin states in the C3(1) model

The functional units in the genome are generally marked by different chromatin states that consist of combinations of various marks. We applied the ChromHMM algorithm to train a hidden Markov model of such states on our ChIP-seq data [37] (chapter 3.11.5). With the four analyzed histone modifications the maximum number of possible combinations in a model was 16, but not all theoretically possible combinations actually occur in a biological setting or are relevant. Consequently, the most appropriate model would include the minimal number necessary to define all functional states, but at the same time ignoring redundant states. Therefore, we evaluated models with different numbers of states and compared the correlation of the individual states to the 16-state model with the aim to define the optimal number of states in the model with a low number of states with correlations < 0.9. We selected the 11-state model, since we observed a decrease in the correlations of individual states with the 16-state model for the 10-state model and all other models with lower numbers of states (Figure 4-18).

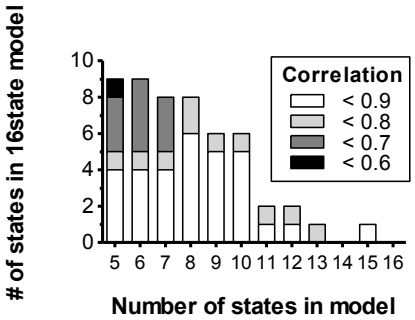


Figure 4-18 Correlation of individual chromatin states in models with different numbers of states to the 16-state model. For each model, the number of states from the 16-state model is depicted that have a correlation < 0.9 or less.

The 11 different states of the model were classified according to known biological functions of histone mark combinations as well as the association of particular states with genomic features (Figure 4-19). For example, H3K27me3 occupancy is known to mark a repressive chromatin state. By annotating the respective regions with HOMER for genomic context we discriminated a genic and an intergenic repressed state (blue states). Following a similar approach, we classified poised/inactive (dark pink) and active promoter states (red). Moreover, we examined enrichment of the states at particular genomic features such as TSSs, CGIs or published enhancer regions in MECs and tumor samples (see Figure 4-20). This enabled the allocation of enhancer states, in particular the enhancer state marked in yellow, for which the enrichment patterns varied between tissue types. While for MECs this state associated with promoters and enhancer regions, in tumors the strongest enrichment was found at known enhancers. Since this state was also defined via enrichment of H3K4me1, which is more often associated with enhancers than with promoters, the yellow state was classified as an enhancer rather than promoter state. The transcription/low signal state (green) covered only few regions weakly associating with H3K27ac and H3K4me1 (Figure 4-19). It enriched strongest for known H3K36me3 peaks (defined by ChIP-seq in MECs [122]), therefore association with transcription is likely. However, since enrichment for the green state was overall quite weak (Figure 4-20), and only few regions of this state had coverage for any of our chipped histone marks, it was also assigned as low signal.

	H3 K27me3	H3 K4me3	H3 K27ac	H3 K4me1
Active Promoter	1	99	95	9
Weak Promoter	2	92	35	87
Weak Promoter (poised)	1	63	9	10
Inactive/poised Promoter	89	76	23	58
Enhancer (CAGE/strong)	1	90	98	88
Enhancer (weak-1)	1	1	13	64
Enhancer (weak-2)	0	5	74	53
Repressed (intergenic)	10	0	0	0
Repressed (genic)	73	0	0	2
Transcription/low signal	0	0	4	3
Low signal	0	0	0	0

Figure 4-19 A combined hidden Markov model for 11 chromatin states. The model was determined by ChromHMM and functional interpretation of chromatin states is depicted by a color code. Each row represents an individual chromatin state of the model with its assigned functional category and the frequency (%) of regions in that state, which are occupied by the respective histone modification.

When comparing the chromatin landscape of MECs and tumors, we observed in tumors a transition from active to inactive/poised promoter states and loss of the strong enhancer state at CGIs, TSS, and CAGE-seq peaks (Figure 4-20). States that overlapped with genes and the H3K36me3 peaks were consistent between MECs and tumors.

A) MECs:

	CpG islands	TSS (Refseq)	TSS ± 2kb	CAGE-seq (Fantom5)	Enhancer (Fantom5)	Enhancer (Shen 2012)	Genes (Refseq)	H3K36me3 peaks
Active Promoter	111	81	26	70	5.8	0.4	1.7	0.3
Weak Promoter	8.9	7.8	14	8.3	7.5	3.2	2.1	0.8
Weak Promoter (poised)	4.7	7.9	14	6.1	4.0	1.3	1.5	0.9
Inactive/poised Promoter	48	23	16	19	6.1	2.3	1.6	0.2
Enhancer (CAGE/strong)	25	15	15	18	16	4.1	2.0	0.4
Enhancer (weak-1)	0.1	0.6	0.9	1.4	4.2	7.7	2.0	0.9
Enhancer (weak-2)	0.2	0.8	1.2	2.7	11	9.1	1.7	1.0
Repressed (intergenic)	0.1	0.6	0.9	0.5	0.7	0.9	0.9	0.2
Repressed (genic)	1.0	1.7	2.5	1.4	1.5	2.2	1.1	0.2
Transcription/low signal	0.2	0.4	0.8	1.0	1.9	2.8	1.7	2.3
Low signal	0.0	0.3	0.4	0.2	0.4	0.3	0.8	1.0

B) Tumors:

	CpG islands	TSS (Refseq)	TSS ± 2kb	CAGE-seq (Fantom5)	Enhancer (Fantom5)	Enhancer (Shen 2012)	Genes (Refseq)	H3K36me3 peaks
Active Promoter	70	48	23	43	8.2	1.2	1.8	0.4
Weak Promoter	4.2	5.2	12	6.2	7.3	3.9	2.1	0.9
Weak Promoter (poised)	4.7	7.0	17	5.6	4.7	1.5	1.8	0.7
Inactive/poised Promoter	76	31	17	26	6.5	2.2	1.6	0.1
Enhancer (CAGE/strong)	4.5	5.2	6.4	7.6	14	7.0	1.8	0.5
Enhancer (weak-1)	0.2	1.0	1.4	1.6	4.5	6.7	1.7	0.7
Enhancer (weak-2)	0.1	1.0	1.5	2.0	6.2	6.2	1.5	0.9
Repressed (intergenic)	0.1	0.6	1.0	0.6	0.8	0.9	1.0	0.1
Repressed (genic)	1.7	1.8	2.8	1.7	1.9	2.1	1.1	0.1
Transcription/low signal	0.2	0.5	1.2	1.1	2.1	3.2	1.5	2.0
Low signal	0.0	0.3	0.4	0.3	0.5	0.5	0.9	1.0

Figure 4-20 Enrichment of chromatin states at genomic features. Enrichment was calculated for MECs A) and the average of three tumor samples B). Numbers represent the individual enrichment score of the chromatin state for the respective genomic feature or regions defined by other studies such as CAGE-seq peaks (FANTOM 5 [120]), enhancers defined by CAGE-seq (FANTOM 5 [120]), enhancers defined by histone modifications (Shen *et al*. [121]) or H3K36me3 peaks in MECs (Lemay *et al*. [122]). A color gradient from white (low) to blue (high) represents the intensity of chromatin state enrichment for each individual feature.

4.2.3 Chromatin state enrichment at progressive DMRs For a more comprehensive description of the epigenetic differences between tumors and MECs, we analyzed the enrichment of chromatin states at progressive DMRs (Figure 4-21). Regions with progressive hypermethylation were strongest enriched for inactive promoter marks in both MECs and tumors. Enrichment for enhancer states was reduced in tumors compared to MECs at regions that gained methylation, suggesting methylation-mediated silencing of these active regions during tumorigenesis. Although this analysis only calculates enrichment and does not investigate which exact genomic positions are affected, this might explain the loss of enhancer state enrichment at promoters and TSS for tumors (Figure 4-20). In contrast, the association of hypomethylated regions with strong enhancer states increased 2-3 fold in tumors compared to MECs. This was also true for the association with active and

weak promoter states. Overall, recurrent loss of DNA methylation was accompanied by a gain in active chromatin states.

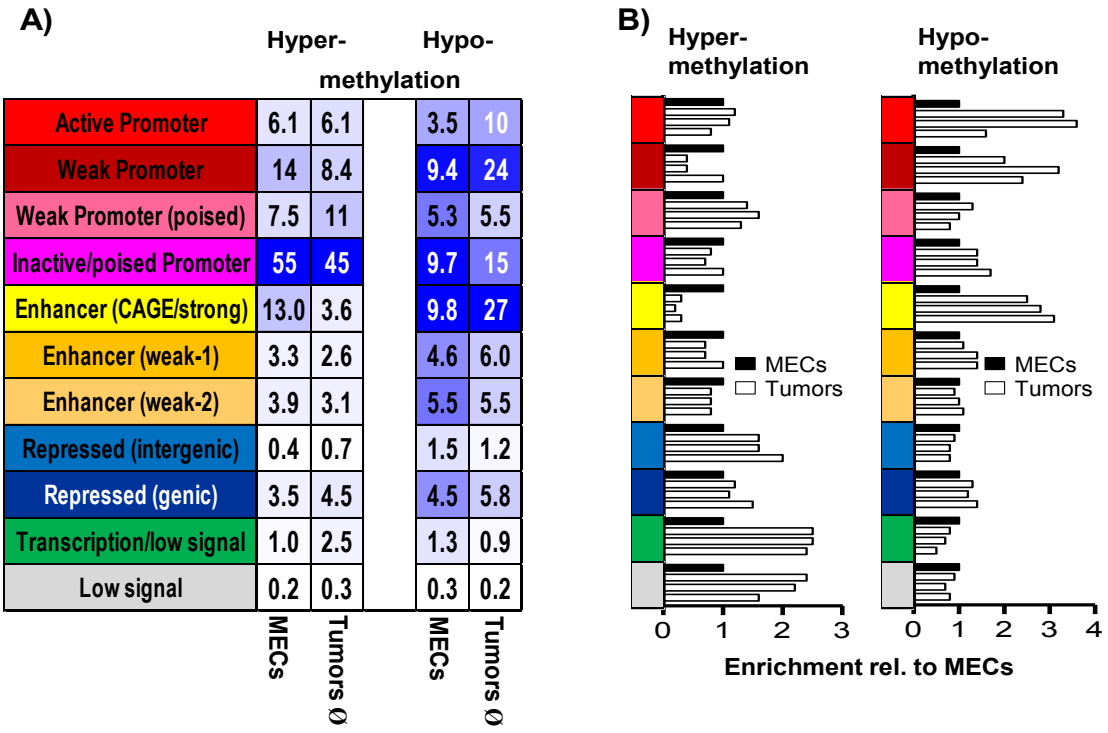


Figure 4-21 Enrichment of chromatin states at progressive DMRs. Enrichment of states was calculated for hyper- and hypomethylated regions in MECs and three tumors and is displayed as absolute enrichment scores A) or relative to the enrichment score of MECs B). Color gradient from white (low) to blue (high) represents the intensity of chromatin state enrichment separately for MECs and tumors.

In order to compare changes in DNA methylation and chromatin states during tumorigenesis in more detail, we focused on regions with MEC- or tumor-specific chromatin states. MCIp-seq coverage was annotated to active, weak and inactive promoters as well as to strong enhancer states with highest overall enrichment for progressive DMRs to investigate DNA methylation changes over time in the different age groups and genotypes (Figure 4-22). MCIp enrichment patterns mostly displayed a pronounced peak at the center of the chromatin feature. In WT samples, MCIp-seq coverage remained constant with increasing age, confirming stable DNA methylation in these selected regions. In TG samples, MCIp coverage increased at MEC-specific inactive promoters and to a weaker degree also at the strong enhancer sites over the course of tumor formation, indicating cancer-associated hypermethylation. At tumor-specific states (active/weak promoters, strong enhancers), we

observed hypomethylation with tumor progression, which was even more pronounced than gain in methylation at MEC-specific enhancers and inactive promoters.

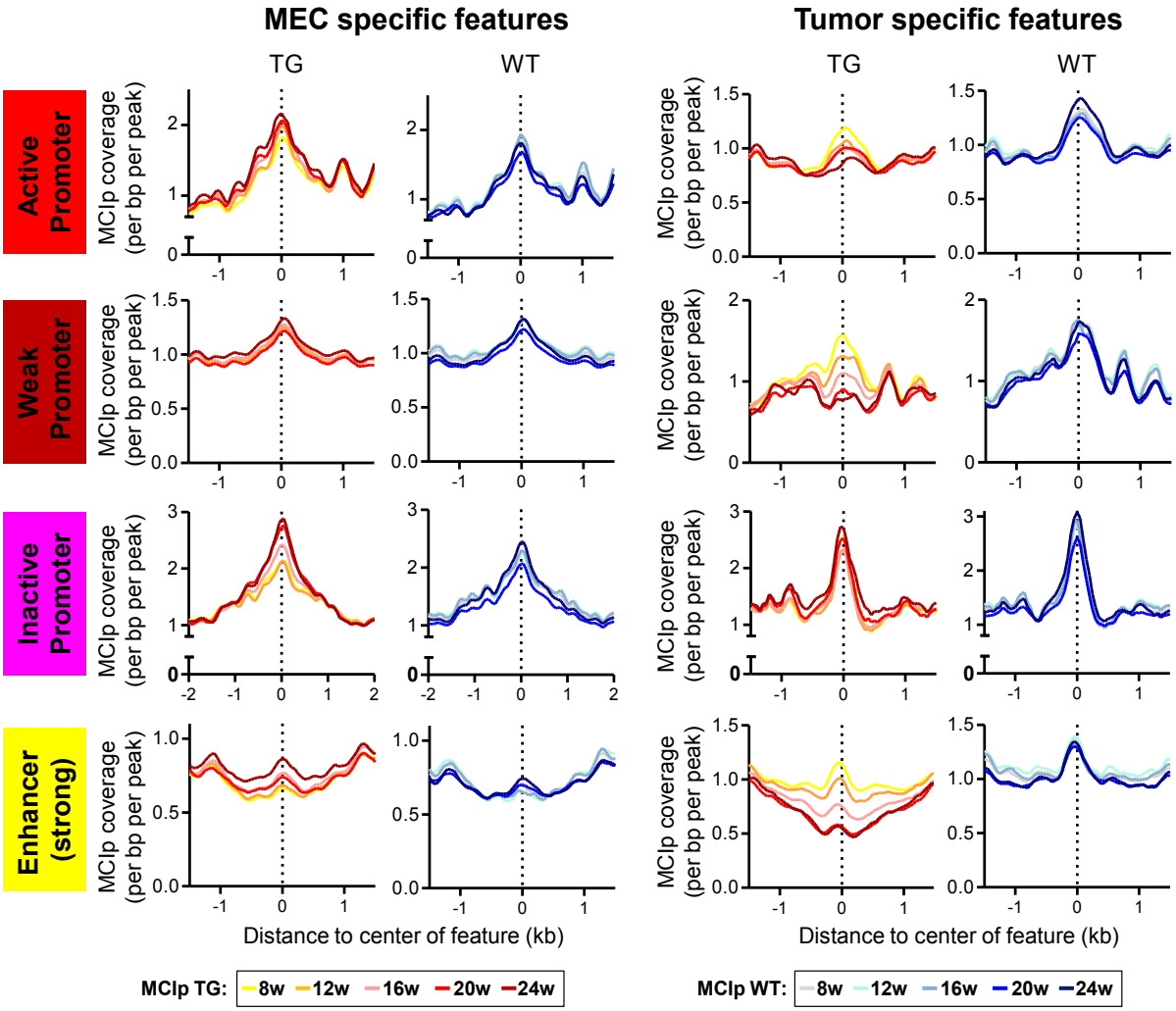


Figure 4-22 MCIp-sequencing coverage at MEC- or tumor-specific chromatin states. Depicted is the average of the MCIp-seq coverage (per bp per peak) for samples at active, weak and inactive promoters, as well as for strong enhancer states. Chromatin states are separated in MEC- (left) and tumor-specific states (right). States common in both sample types were omitted. Methylation coverage at specific age groups is depicted in separate graphs for WT (yellow to red lines) and TG animals (grey to blue lines).

4.2.4 Functional characterization of tissue-specific chromatin states

4.2.4.1 Association of chromatin states with gene expression

In order to further characterize the interaction of chromatin and DNA methylation changes in the regulation of gene expression during carcinogenesis in the C3(1) model, we focused on regions where alterations in DNA methylation and chromatin stages directly overlapped. A

summary of the degree of overlap between tissue-specific chromatin states and progressive DMRs is listed in Table 4-3.

Table 4-3 Tissue-specific chromatin states overlapping with DMRs

Chromatin state	# of MEC-specific states overlapping with hypermethylated regions	# of tumor-specific states overlapping with hypomethylated regions
Active promoter	32	365
Weak promoter	197	45
Inactive/poised promoter	197	59
Enhancer strong	101	441

We identified 365 hypomethylated regions that overlapped with tumor-specific active promoter states. Surprisingly, only a minor fraction of these promoter states was actually annotated as promoter or a TSS-associated feature by HOMER. Instead, almost 75% of these regions were annotated as exons, introns or intergenic regions, similar to the pattern we observed when examining all progressively hypomethylated regions (Figure 4-23A). This might be due to the fact that the DMRs identified with MCIp are biased by their requirements of CpG density and methylation degree. Therefore, we might have missed DMRs in the promoter context and instead detected neighboring DMRs located closer to an exon or intron. Since the chromatin states often covered larger regions than the relatively narrow MCIp peaks, promoter regions might be better covered by the chromatin state than by the MCIp signals. This is exemplary depicted for DMRs associated with *Krt8* and *Col7a1*, which are upregulated in prepubertal breast (Figure 4-23B, C). *Krt8* is hypomethylated at a promoter- and an intron-associated DMR ≈2.3kb downstream of the TSS, both covered by the active promoter chromatin state. *Col7a1* only displayed hypomethylation in an exonic region, although the DMR is only about 500bp downstream of the TSS and also is covered by an active promoter state. In addition, promoter states in intergenic regions might hint towards lncRNA promoters, which are not annotated so far.

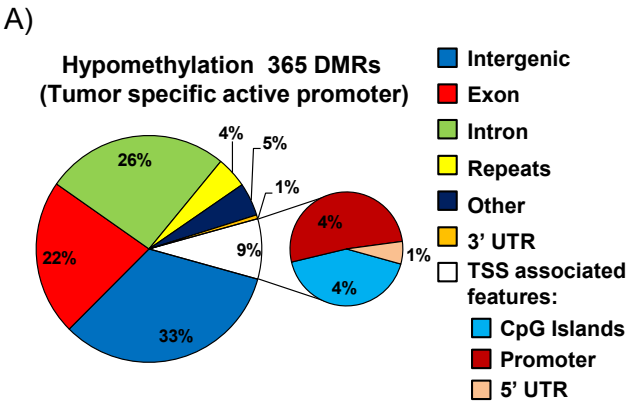
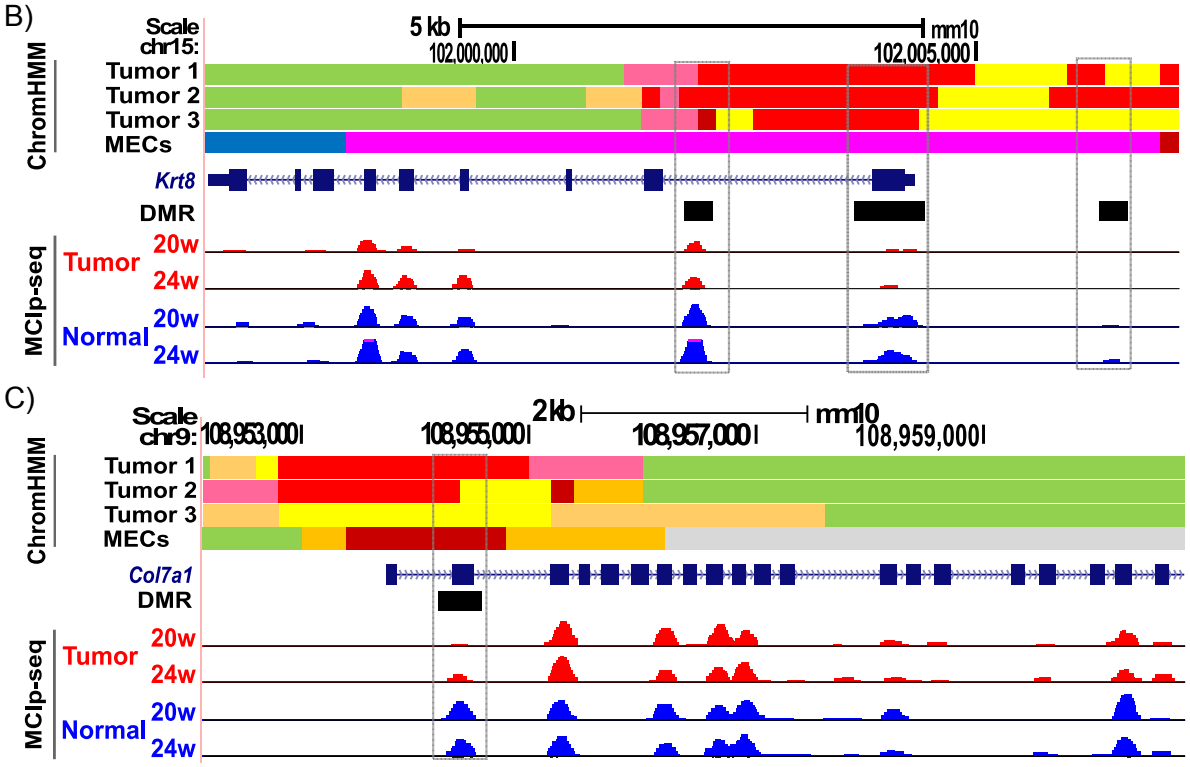


Figure 4-23 Tumor-specific active promoter states. A) Genomic distribution of progressively hypomethylated regions overlapping with a tumor-specific active promoter state. Annotation of regions according to HOMER. B, C) Tumor-specific active promoter states overlap with hypomethylated regions for *Krt8* (B) and *Col7a1* (C). Color code for chromatin states as in Figure 4-19. MCIp-seq reads are depicted as average of three animals per age group in tumors (red) and normal mammary glands (blue). Black boxes mark DMRs while RefSeq genes are depicted in blue. Overlap between tissue-specific chromatin states and DMRs is emphasized by dotted lines.



Annotating the respective regions according to the closest gene with HOMER allowed us to find common functions for these genes by performing GSEA (Table 4-4). We found genes enriched in prepubertal mammary gland development or upregulated in ERBB2 tumors (e.g. *Cln3, Cldn7, Krt8*), as well as genes that could be either upregulated in MaSCs, such as *Bmp7, Wif1,* and *Krt14* or downregulated in MaSCs such as *Krt8, Krt18* or *Cldn8*. In addition genes were enriched that were upregulated after overexpression of *TRP53* or *TRP63* in human mammary epithelial cells (HMEC), such as *Egr2*, *Smad6,* and *Wnt7b*. Since the SV40T transgene is known to inactivate p53 and Rb1, we would have rather expected to find enrichment of upregulated genes after knockdown of *Tp53*. However, when SV40T binds to

p53 it stabilizes the protein and prevents degradation [133]. This could mimic overexpression of p53. Furthermore, several genes were upregulated in preneoplastic stages and showed continuously high expression in breast tumors, like *Cldn3, Cldn7, Spint 1* or *Vdr*.

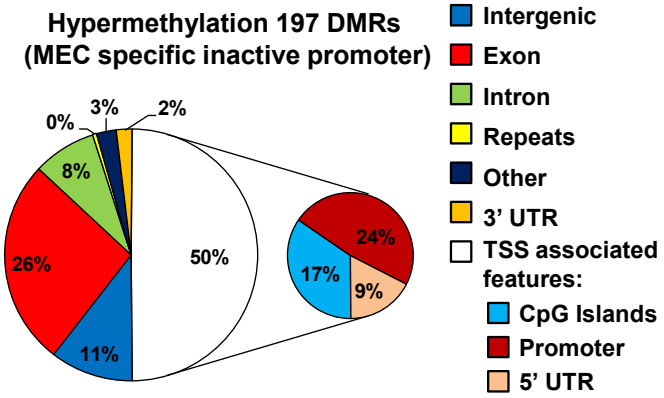
Table 4-4 Gene set enrichment analysis of hypomethylated tumor-specific target genes with active promoter state

	Gene set name (# of overlaps) *	Description	ratio	P-value
1	MCBRYAN_PUBERTAL_BREAST_ 4_5WK_UP (24)	Genes upregulated during pubertal mammary gland development between week 4 and 5.	0.089	5.68E-22
2	MEISSNER_BRAIN_HCP_WITH_ H3K4ME3_AND_H3K27ME3 (38)	Genes with high-CpG-density promoters (HCP) bearing histone H3 dimethylation at K4 (H3K4me2) and trimethylation at K27 (H3K27me3) in brain.	0.036	4.63E-20
3	LANDIS_ERBB2_BREAST_ TUMORS_324_UP (17)	Upregulated genes from the 324 genes identified by two analytical methods as changed in the mammary tumors induced by transgenic expression of ERBB2 [GeneID=2064].	0.113	9.80E-18
4	GRAESSMANN_APOPTOSIS_BY_ DOXORUBICIN_DN (44)	Genes downregulated in ME-A cells (breast cancer) undergoing apoptosis in response to doxorubicin [PubChem=31703].	0.025	3.63E-17
5	WONG_ADULT_TISSUE_STEM_ MODULE (28)	The 'adult tissue stem' module: genes coordinately upregulated in a compendium of adult tissue stem cells.	0.039	5.82E-16
6	KOINUMA_TARGETS_OF_SMAD2_ OR_SMAD3 (29)	Genes with promoters occupied by SMAD2 or SMAD3 [GeneID=4087, 4088] in HaCaT cells (keratinocyte) according to a ChIP-chip analysis.	0.035	2.16E-15
7	CUI_TCF21_TARGETS_2_DN (27)	All significantly downregulated genes in kidney glomeruli isolated from TCF21 [Gene ID=6943] knockout mice.	0.033	1.37E-13
8	GROSS_HYPOXIA_VIA_ELK3_AND _HIF1A_UP (13)	Genes upregulated in SEND cells (skin endothelium) at hypoxia after knockdown of ELK3 [GeneID=2004] and HIF1A [GeneID=3091] by RNAi.	0.092	1.19E-12
9	PEREZ_TP63_TARGETS (18)	Genes upregulated in the HMEC cells (primary mammary epithelium) upon expression of the transcriptionally active isoform of TP63 [GeneID=8626] off adenoviral vector.	0.051	1.40E-12
10	DACOSTA_UV_RESPONSE_VIA_ ERCC3_DN (26)	Genes downregulated in fibroblasts expressing mutant forms of ERCC3 [GeneID=2071] after UV irradiation.	0.030	1.81E-12
11	LANDIS_BREAST_CANCER_ PROGRESSION_UP (9)	Genes upregulated in preneoplastic mammary tissues and whose expression is maintained in tumors.	0.205	2.09E-12
12	PEREZ_TP53_TARGETS (30)	Genes upregulated in the HMEC cells (primary mammary epithelium) upon expression of TP53 [GeneID=7157] off adenoviral vector.	0.026	2.60E-12
13	LIM_MAMMARY_STEM_CELL_DN (19)	Genes consistently downregulated in mammary stem cells both in mouse and human species.	0.044	3.36E-12
14	LIM_MAMMARY_STEM_CELL_UP (20)	Genes consistently upregulated in mammary stem cells both in mouse and human species.	0.041	3.95E-12

* full list of overlapping genes in Supplementary Table 6.

We identified 197 hypermethylated regions that overlapped with MEC-specific inactive promoters. As both DNA hypermethylation and enrichment of the H3K27me3 mark represent a gene-repressive state, gain in methylation at these regions might reflect an epigenetic switch from chromatin-mediated repression in normal tissue to methylation-mediated gene silencing in tumors. A role in gene silencing was further supported by the fact that half of the regions were annotated as TSS-associated features (Figure 4-24).

A)



B)

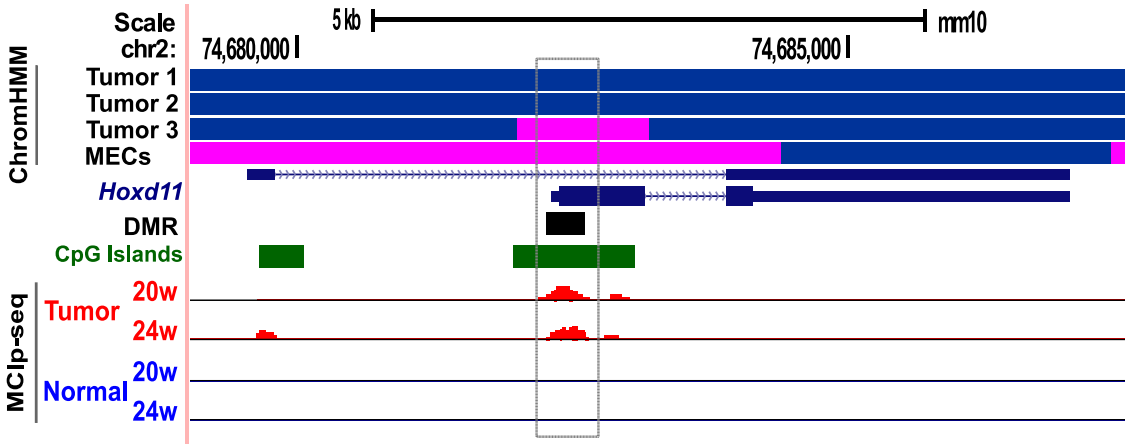


Figure 4-24 MEC-specific inactive poised promoter states. A) Genomic distribution of progressively hyper-methylated regions overlapping with MEC-specific inactive poised promoters. Annotation of regions according to HOMER. B) *Hoxd11* is hypermethylated at a MEC-specific inactive/poised promoter state. Color code and samples as in Figure 4-23.

Consistently, in GSEA analyses genes associated with these MEC-specific inactive promoter regions were highly enriched for targets of PRC2 members and genes marked by bivalent chromatin in brain and neural progenitor cells (Table 4-5) (e.g. *Hoxc4, Hoxd11, Foxc1, Lhx2, Pcdhx11*). These genes are generally expressed only during development and silenced in differentiated cells.

Table 4-5 Gene set enrichment analysis of hypermethylated MEC-specific target genes with inactive/poised promoter state

	Gene set name (# of overlaps)*	Description	ratio	P-value
1	BENPORATH_SUZ12_TARGETS (26)	Set 'Suz12 targets': genes identified by ChIP on chip as targets of the Polycomb protein SUZ12 [GeneID=23512] in human embryonic stem cells.	0.025	2.90E-17
2	BENPORATH_EED_TARGETS (24)	Set 'Eed targets': genes identified by ChIP on chip as targets of the Polycomb protein EED [GeneID=8726] in human embryonic stem cells.	0.0226	5.32E-15
3	MEISSNER_BRAIN_HCP_WITH_ H3K4ME3_AND_H3K27ME3 (24)	Genes with high-CpG-density promoters (HCP) bearing histone H3 dimethylation at K4 (H3K4me2) and trimethylation at K27 (H3K27me3) in brain.	0.0225	6.14E-15
4	BENPORATH_PRC2_TARGETS (19)	Set 'PRC2 targets': identified by ChIP on chip on human embryonic stem cells as genes that: posess the trimethylated H3K27 mark in their promoters and are bound by SUZ12 [GeneID=23512] and EED [GeneID=8726] Polycomb proteins.	0.0291	5.52E-14
5	MEISSNER_NPC_HCP_WITH_H3 K4ME2_AND_H3K27ME3 (15)	Genes with high-CpG-density promoters (HCP) bearing histone H3 dimethylation mark at K4 (H3K4me2) and trimethylation mark at K27 (H3K27me3) in neural precursor cells (NPC).	0.043	1.15E-13

* full list of overlapping genes in Supplementary Table 7.

Of the regions classified as strong enhancers, 101 MEC-specific ones were hypermethylated, whereas 441 tumor-specific ones were hypomethylated during tumorigenesis (Figure 4-25). Both hypo- and hypermethylated enhancer regions were mostly found in introns, exons and intergenic regions. In total, this comprised 75-85% of all regions. In addition, 20% of the MEC-specific enhancers that gained methylation were located close to TSS-associated features.

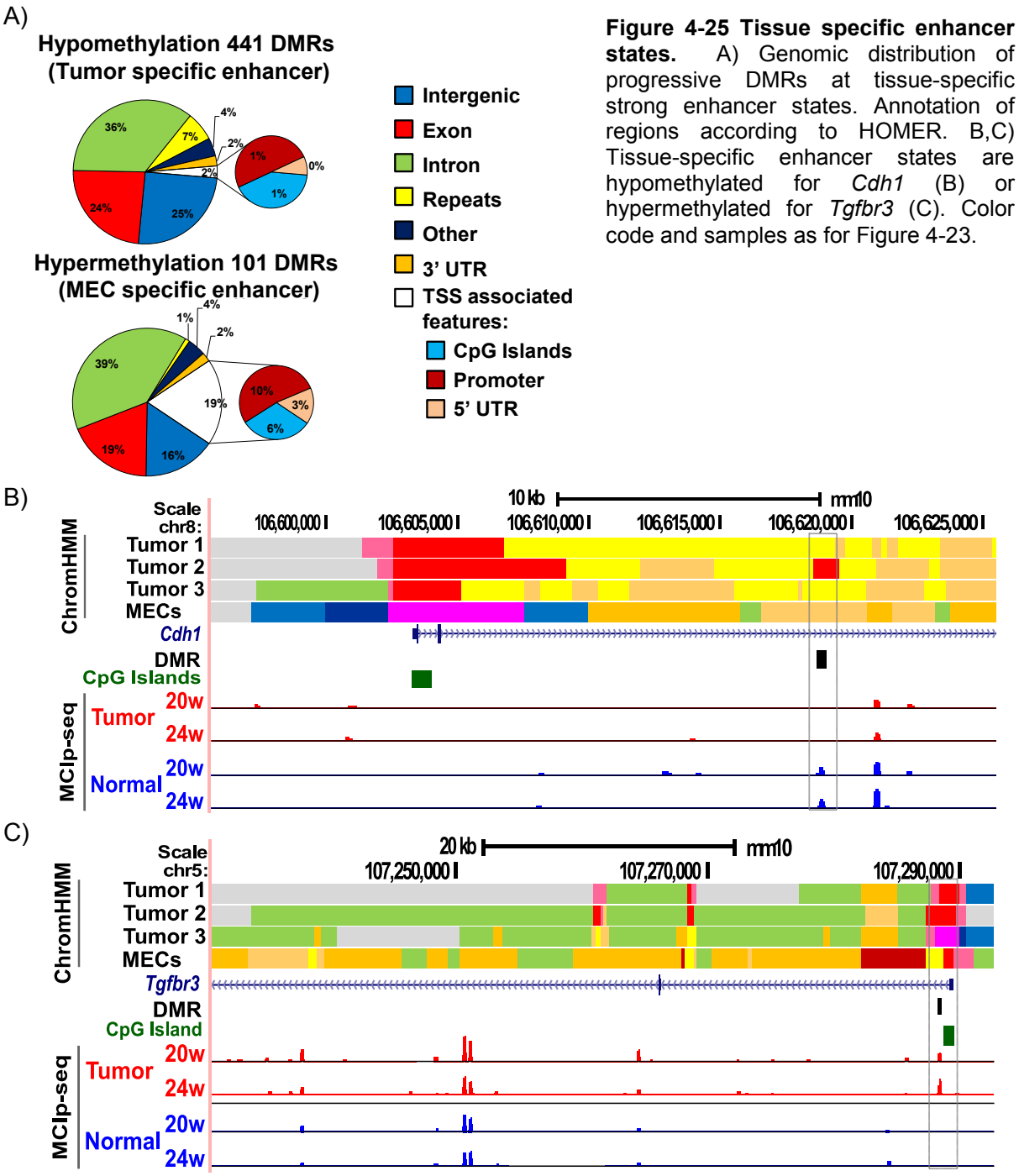


Figure 4-25 Tissue specific enhancer states. A) Genomic distribution of progressive DMRs at tissue-specific strong enhancer states. Annotation of regions according to HOMER. B,C) Tissue-specific enhancer states are hypomethylated for *Cdh1* (B) or hypermethylated for *Tgfbr3* (C). Color code and samples as for Figure 4-23.

A high proportion of differentially methylated enhancer regions were annotated to genic features (exons, introns), suggesting that these enhancers could influence expression of the closest gene. Therefore, we used the closest gene annotation according to HOMER to select genes for enrichment analysis with known gene sets of the MSigDB.

The top five most enriched gene sets for the hypomethylated tumor-specific enhancers included three sets of known RB1 and p53 targets (Table 4-6) that were downregulated after knockdown in skin cells. In the C3(1) model, which is driven by disturbing Rb1 and p53

signaling via expression of SV40T, these genes were rather upregulated, but as mentioned before, SV40T-binding to p53 leads to higher protein stability, which could explain the increased expression. Additional gene sets indicated upregulation in basal-like breast cancer and involvement in pubertal breast development.

Table 4-6 Gene set enrichment analysis of hypomethylated tumor-specific target genes with strong enhancer state

	Gene set name (# of overlaps) *	Description	ratio	p-value
1	MARTINEZ_RB1_AND_TP53_ TARGETS_DN (38)	Genes downregulated in mice with skin specific double knockout of both RB1 and TP53 [GeneID=5925;7157] by Cre-lox.	0.0643	1.22E-26
2	MARTINEZ_TP53_TARGETS_ DN (36)	Genes downregulated in mice with skin specific knockout of TP53 [GeneID=7157].	0.0607	2.00E-24
3	MEISSNER_BRAIN_HCP_WITH _H3K4ME3_AND_H3K27ME3 (41)	Genes with high-CpG-density promoters (HCP) bearing histone H3 dimethylation at K4 (H3K4me2) and trimethylation at K27 (H3K27me3) in brain.	0.0384	2.85E-20
4	BLALOCK_ALZHEIMERS_ DISEASE_UP (41)	Genes upregulated in brain from patients with Alzheimer's disease.	0.029	4.99E-19
5	MARTINEZ_RB1_TARGETS_ DN (28)	Genes downregulated in mice with skin specific knockout of RB1 [GeneID=5925] by Cre-lox.	0.0516	2.31E-17
6	JAEGER_METASTASIS_DN (20)	Genes downregulated in metastases from malignant melanoma compared to the primary tumors.	0.0775	4.46E-16
7	WONG_ADULT_TISSUE_ STEM_MODULE (30)	The 'adult tissue stem' module: genes coordinately upregulated in a compendium of adult tissue stem cells.	0.0416	5.28E-16
8	DODD_NASOPHARYNGEAL_ CARCINOMA_UP (46)	Genes upregulated in nasopharyngeal carcinoma (NPC) compared to the normal tissue.	0.0253	1.18E-15
9	PEREZ_TP63_TARGETS (22)	Genes upregulated in the HMEC cells (primary mammary epithelium) upon expression of the transcriptionally active isoform of TP63 [GeneID=8626] off adenoviral vector.	0.062	1.60E-15
10	SMID_BREAST_CANCER_ BASAL_UP (28)	Genes upregulated in basal subtype of breast cancer samples.	0.0432	2.04E-15
11	MCBRYAN_PUBERTAL_ BREAST_4_5WK_UP (19)	Genes upregulated during pubertal mammary gland development between week 4 and 5.	0.0701	1.54E-14

* full list of overlapping genes in Supplementary Table 8.

Genes associated with the hypermethylated MEC-specific enhancers were enriched in genes deregulated in various types of cancer, such as thyroid cancer or acute myeloid leukemia (Table 4-7). This included the transforming growth factor β receptor III (TGFBR3), which could have both tumor promoting [134] as well as tumor suppressive functions [135] in breast cancer depending on the subtype. Surprisingly the MEC-specific enhancer, which overlaps with the progressively hypermethylated region, displayed an active promoter state in two tumors and an inactive promoter state in the third tumor. Thus further testing will be required to evaluate if this represents an example of intratumor heterogeneity and which epigenetic mark will be the dominant factor to influence gene expression.

Table 4-7 Gene set enrichment analysis of genes associated with hypermethylated MEC-specific strong enhancers

	Gene set name (#of overlaps)*	Description	ratio	P-value
1	CHYLA_CBFA2T3_TARGETS_UP (11)	Genes upregulated in immature bone marrow progenitor cells upon knock out of CBFA2T3 [GeneID=863].	5.33E-11	1.81E-07
2	RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_DN (13)	Genes downregulated in poorly differentiated thyroid carcinoma (PDTC) compared to normal thyroid tissue.	8.71E-10	1.48E-06
3	LIM_MAMMARY_STEM_CELL_UP (10)	Genes consistently upregulated in mammary stem cells both in mouse and human species.	9.85E-09	1.12E-05
4	RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC_DN (10)	Genes downregulated in anaplastic thyroid carcinoma (ATC) compared to normal thyroid tissue.	2.38E-08	1.89E-05
5	WEST_ADRENOCORTICAL_TUMOR_ DN (10)	Downregulated genes in pediatric adrenocortical tumors (ACT) compared to the normal tissue.	2.78E-08	1.89E-05
6	MULLIGHAN_MLL_SIGNATURE_2_UP (9)	The 'MLL signature 2': genes upregulated in pediatric AML (acute myeloid leukemia) with rearranged MLL [GeneID=4297] compared to the AML cases with intact MLL and NPM1 [GeneID=4869].	3.65E-08	2.07E-05
7	DACOSTA_UV_RESPONSE_VIA_ ERCC3_DN (11)	Genes downregulated in fibroblasts expressing mutant forms of ERCC3 [GeneID=2071] after UV irradiation.	1.90E-07	9.20E-05
8	MULLIGHAN_MLL_SIGNATURE_1_UP (8)	The 'MLL signature 1': genes upregulated in pediatric AML (acute myeloid leukemia) with rearranged MLL [GeneID=4297] compared to all AML cases with the intact gene.	2.60E-07	1.10E-04
9	HAN_SATB1_TARGETS_UP (8)	Genes upregulated in MDA-MB-231 cells (breast cancer) after knockdown of SATB1 [GeneID=6304] by RNAi.	3.48E-07	1.31E-04
10	BOYLAN_MULTIPLE_MYELOMA_D_ CLUSTER_DN (4)	Genes from cluster 4: downregulated in group D of tumors arising from overexpression of BCL2L1 and MYC [GeneID=598;4609] in plasma cells.	6.36E-07	2.04E-04
11	JOHNSTONE_PARVB_TARGETS_ 3_UP (9)	Genes upregulated upon overexpression of PARVB [GeneID=29780] in MDA-MB-231 cells (breast cancer) cultured in 3D Matrigel only.	6.59E-07	2.04E-04

* full list of overlapping genes in Supplementary Table 9.

With the lists of differentially methylated, tissue-specific chromatin state associated genes, we performed additional GSEA on the published gene expression microarray data set of the C3(1) model [90]. We could confirm that differential methylation for genes of these lists was associated with gene expression changes that link to the phenotype. This was true for the MEC-specific hypermethylated inactive promoters, the hypomethylated tumor-specific active promoters and both of the enhancer gene sets (Figure 4-26). For the enhancers, hypomethylation was linked with upregulation of gene expression in the C3(1) model, while the hypermethylated regions were rather downregulated.

In contrast, enrichment was not significant for genes of hypomethylated tumor-specific weak promoters, probably due to the low number of regions (45 regions, Table 4-3) in this gene list.

These results emphasize the importance of DNA methylation changes at the strong enhancer regions for gene expression and tumorigenesis in the C3(1) model.

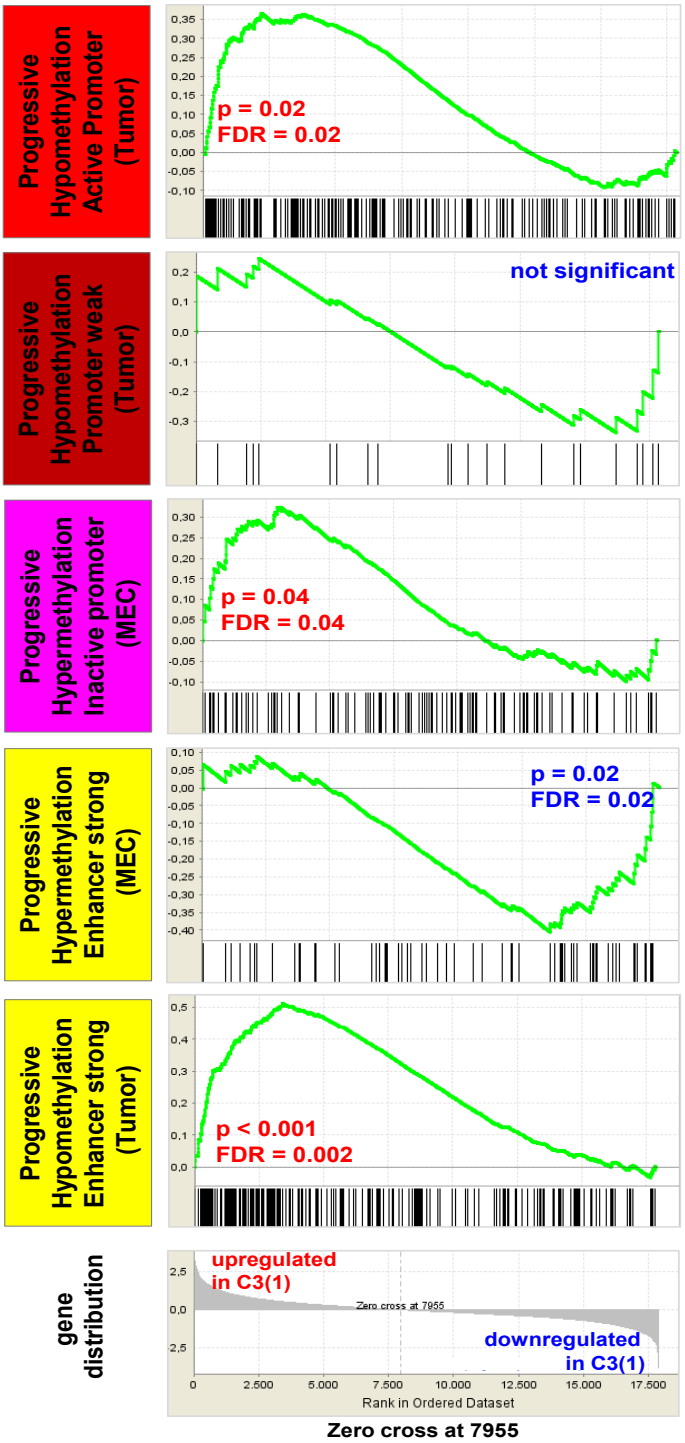
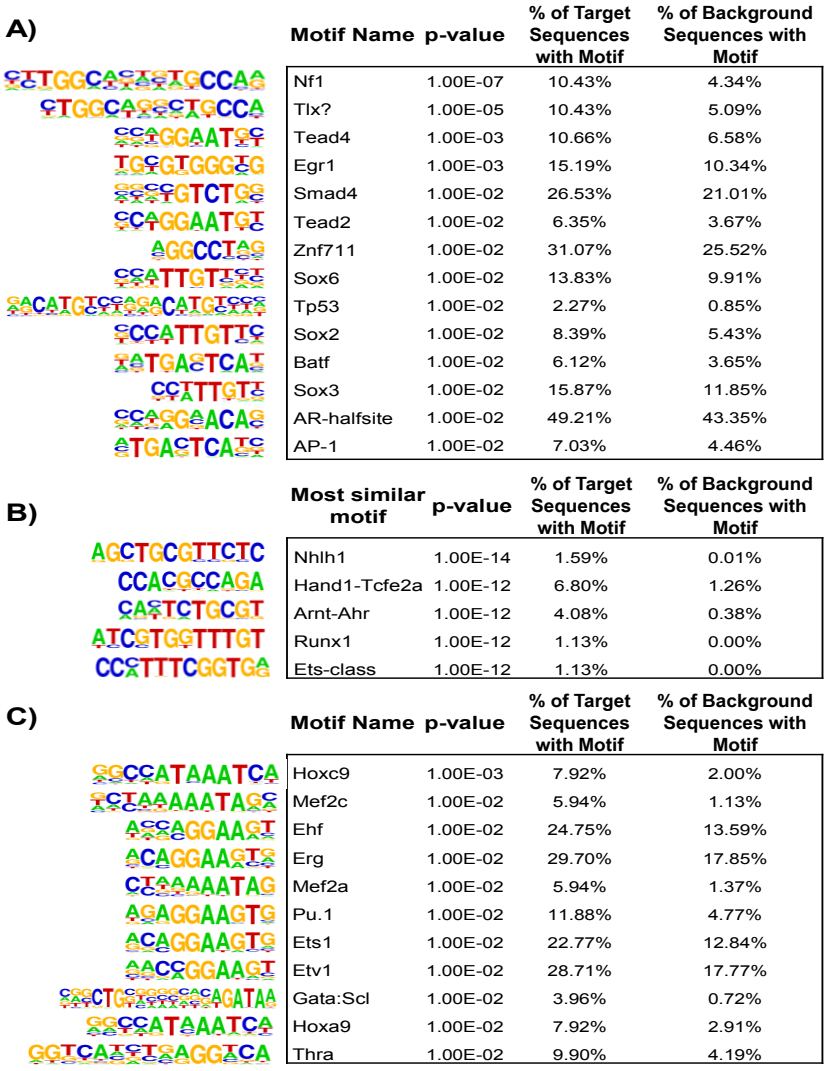


Figure 4-26 Gene set enrichment analysis (GSEA) of gene sets comprising progressive DMRs overlapping with tissue specific chromatin states. Enrichment was calculated on gene expression microarray data [90] for eight C3(1) tumors and five WT mammary gland samples. Depicted are enrichment score (ES) graphs (green, top), gene set distribution (black line graphs, below the ES graph), p-values and false discovery rate q-values (FDR). Significance cutoffs were p< 0.05 and FDR < 0.25.

4.2.4.2 Transcription factor binding sites at tissue-specific chromatin states

Since gene expression can be influenced by binding of transcription factors, we investigated whether the enhancer regions were enriched for transcription factor binding sites. For the 441 hypomethylated enhancer regions, HOMER motif search predicted enrichment for 14 known transcription factor motifs and five motifs with similarity to previously defined motifs (Figure 4-27A, B). Another 11 known motifs were enriched in the hypermethylated MEC-specific enhancer regions (Figure 4-27C).



A)

Motif Name	p-value	% of Target Sequences with Motif	% of Background Sequences with Motif
Nf1	1.00E-07	10.43%	4.34%
Tlx?	1.00E-05	10.43%	5.09%
Tead4	1.00E-03	10.66%	6.58%
Egr1	1.00E-03	15.19%	10.34%
Smad4	1.00E-02	26.53%	21.01%
Tead2	1.00E-02	6.35%	3.67%
Znf711	1.00E-02	31.07%	25.52%
Sox6	1.00E-02	13.83%	9.91%
Tp53	1.00E-02	2.27%	0.85%
Sox2	1.00E-02	8.39%	5.43%
Batf	1.00E-02	6.12%	3.65%
Sox3	1.00E-02	15.87%	11.85%
AR-halfsite	1.00E-02	49.21%	43.35%
AP-1	1.00E-02	7.03%	4.46%

B)

Most similar motif	p-value	% of Target Sequences with Motif	% of Background Sequences with Motif
Nhlh1	1.00E-14	1.59%	0.01%
Hand1-Tcfe2a	1.00E-12	6.80%	1.26%
Arnt-Ahr	1.00E-12	4.08%	0.38%
Runx1	1.00E-12	1.13%	0.00%
Ets-class	1.00E-12	1.13%	0.00%

C)

Motif Name	p-value	% of Target Sequences with Motif	% of Background Sequences with Motif
Hoxc9	1.00E-03	7.92%	2.00%
Mef2c	1.00E-02	5.94%	1.13%
Ehf	1.00E-02	24.75%	13.59%
Erg	1.00E-02	29.70%	17.85%
Mef2a	1.00E-02	5.94%	1.37%
Pu.1	1.00E-02	11.88%	4.77%
Ets1	1.00E-02	22.77%	12.84%
Etv1	1.00E-02	28.71%	17.77%
Gata:Scl	1.00E-02	3.96%	0.72%
Hoxa9	1.00E-02	7.92%	2.91%
Thra	1.00E-02	9.90%	4.19%

Figure 4-27 HOMER list of enriched motifs for transcription factors in progressive DMRs at tissue-specific strong enhancers. A, B) Hypomethylated tumor-specific strong enhancers. A) Known transcription factor binding motifs with hypergeometric motif logo and the respective name of the transcription factor. Additional values depict the p-value for motif enrichment, as well as the percentage of sequences that contain the motif in either the target regions or the background regions. B) *De novo* motif search. Depicted are motif logos that were calculated by a *de novo* motif search in the target regions. Motifs are named according to the transcription factors whose motif most closely resembles the new one. Additional information as in A). C) Hypermethylated MEC-specific strong enhancers. Information for motifs depicted as in A).

The p-values for the enrichment of these motifs were quite low and thus, these results need to be handled with caution. Therefore, we assessed the expression levels of the respective transcription factors in the C3(1) model to ascertain whether motif enrichment was accompanied by changes in expression of the respective transcription factors between WT and TG animals. We limited our selection to inverse correlations, meaning that loss of enhancer methylation was linked to increased transcription factor expression and vice versa. We found significant upregulation of *Tead4, Egr1, Arnt, Runx1, and Etv4 (Ets-class)* in tumors, as well as significant downregulation of *Mef2c* and *Hoxa9* (Figure 4-28).

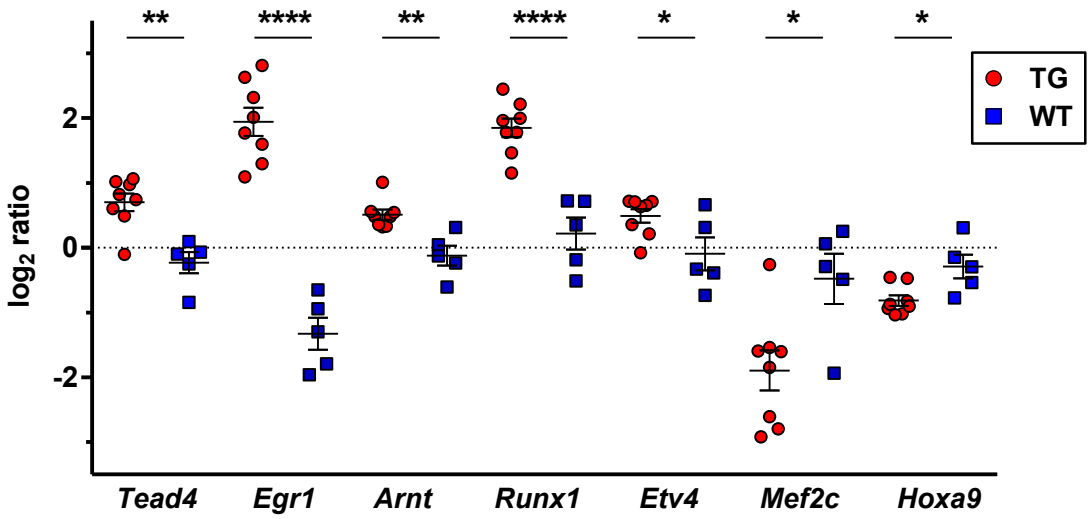


Figure 4-28 Expression levels of selected transcription factors associated with motif enrichment. Expression levels are depicted as microarray intensities for eight tumors (TG, red circle) and five normal mammary glands (WT, blue square). Samples were measured relative to whole mouse RNA of day one pups [90].Student's t-test (two-sided), * p< 0.05, ** p< 0.01, **** p< 0.0001.

Of the analyzed transcription factors, *Egr1* and *Runx1* were the two factors with the highest increase in tumor issue. Examining the chromatin states and DMRs for these transcription factors, we found hypomethylated regions overlapping with active promoter states in tumors that were defined as weak promoter (*Egr1*) and inactive promoter states (*Runx1*) in MECs (Figure 4-29). In addition, we detected three hypomethylated regions overlapping with tumor-specific intragenic enhancer states in *Runx1.* Together with other genes of the hypomethylated enhancers, *Runx1* was also enriched in gene sets for bivalent brain promoters and nasopharyngeal carcinoma upregulated genes (Supplementary Table 8). Our data suggested that the expression of these factors is coordinately regulated by both the chromatin state and the methylation levels at their promoters.

For *Etv4* we observed a change from inactive to active promoter state around the TSS without accompanying loss of methylation. However, we found a similar change in chromatin

states about 25kb upstream coinciding with hypomethylation, which might confer an alternative promoter for *Etv4*.

For *Hoxa9*, we detected a hypermethylated region located downstream of the gene with an inactive/poised promoter state changing into a completely repressed state in tumors, consistent with its downregulation in tumors.

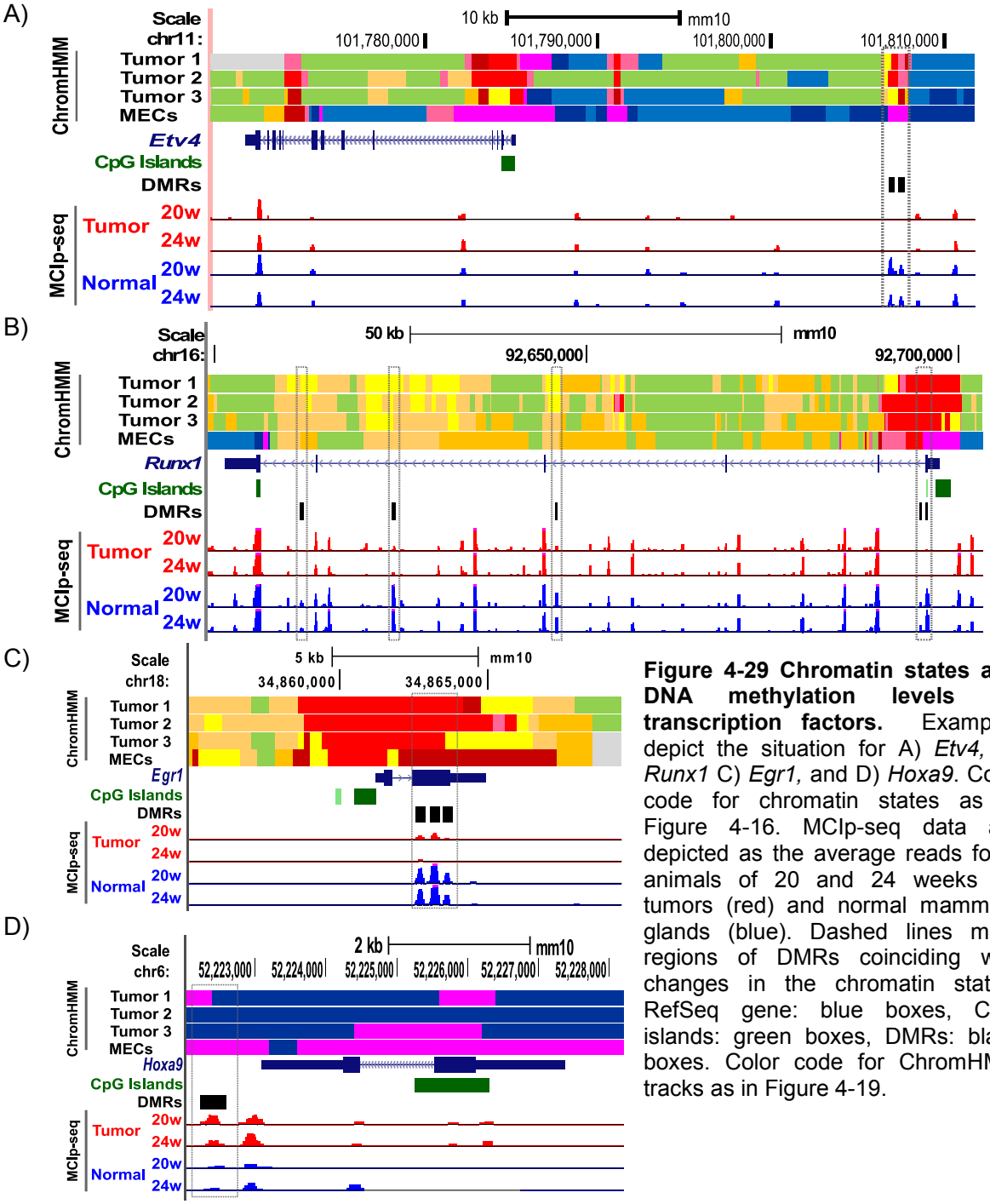


Figure 4-29 Chromatin states and DNA methylation levels for transcription factors. Examples depict the situation for A) *Etv4,* B) *Runx1* C) *Egr1,* and D) *Hoxa9*. Color code for chromatin states as in Figure 4-16. MCIp-seq data are depicted as the average reads for 3 animals of 20 and 24 weeks for tumors (red) and normal mammary glands (blue). Dashed lines mark regions of DMRs coinciding with changes in the chromatin states. RefSeq gene: blue boxes, CpG islands: green boxes, DMRs: black boxes. Color code for ChromHMM tracks as in Figure 4-19.

The enrichment of binding motifs for these transcription factors in hypomethylated enhancers suggests that the epigenetic changes might be either caused by overexpression of the factors, or the binding of the factors is facilitated due to the changes in epigenetics. Our data indicates that up- and downregulation of these transcription factors might be regulated by epigenetic changes themselves and that these factors might be involved in processes that are important for tumorigenesis in the C3(1) mouse model.

4.2.5 Summary

In conclusion, we created a map of chromatin states for the C3(1) tumors and MECs and observed tissue specific changes in these states. When we overlapped these tissue specific chromatin states with progressive DMRs we observed several concordant changes between the chromatin composition and DNA methylation. Moreover, we found that a combination of these marks can be linked to gene expression changes of the C3(1) model especially for enhancer and active promoter states. Finally, we observed enriched transcription factor motifs in differentially methylated enhancers and expression of these factors in turn might be regulated by epigenetic mechanisms. These transcription factors might play an important role during the tumorigenesis in the C3(1) model.

4.3 Evaluation of differentially methylated lncRNA *Esrp2-as* (*1810019D21RIK*) and protein-coding gene *Esrp2*

4.3.1 Genome-wide screen to identify differentially methylated lncRNAs

In order to evaluate the third level of epigenetic regulation of gene expression, we investigated the regulation of lncRNAs. For a screen of differentially methylated lncRNAs, we utilized the MCIp-seq dataset for samples of mice at 20 and 24 weeks of age to define DMRs common in both age groups (screening strategy in Figure 4-30). This sample set contained 1614 hyper- and 4956 hypomethylated regions. By overlapping these DMRs with promoters (2kb upstream to 500bp downstream of the TSS) of 3639 mouse RefSeq annotated lncRNAs (GRCm38/mm10), we identified 37 hyper- and 32 hypomethylated lncRNA promoters (Table 4-8).

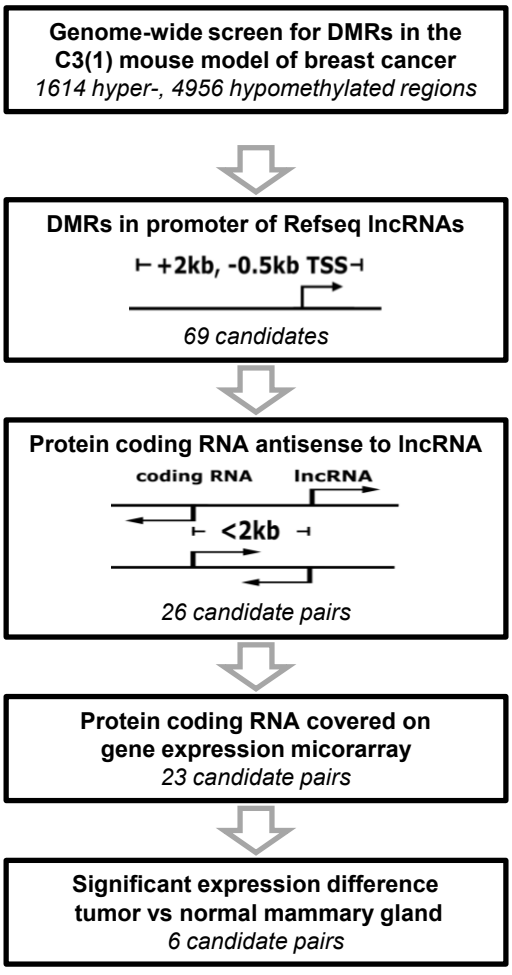


Figure 4-30 Schematic representation of the screening strategy for identification of differentially methylated lncRNAs.

lncRNAs can function on protein-coding genes in direct neighborhood or on more distal genes, sometimes located even on different chromosomes. Following the 'guilt by association' principle, we focused on lncRNAs in the vicinity of protein-coding genes and selected candidates with neighboring mRNAs in antisense orientation and the TSSs of both RNAs not more than 2kb apart. We thus identified 26 pairs of protein-coding and noncoding RNAs (Table 4-8).

Studies that describe the lncRNA function of these candidates are scarce, except for *Haglr* (previously designated as *Hoxd-as1*), *Fendrr*, and *Hoxa11as*, which were all shown to regulate expression of protein-coding genes, either in the direct neighborhood or located more distantly [53, 55, 136-138]. Therefore, we investigated the neighboring protein-coding genes for overlap with published gene sets of the Molecular Signature Database v5.1 (MSigDB). The protein-coding genes located close to hypermethylated lncRNAs were known targets of the PRC2 members EED and SUZ12 (*Foxf1, Foxd2, Hoxd1, Evx1, Dlx4, Sox21, Magi2, Itpkb*) or were marked by the repressive histone modification H3K27me3 in progenitor or even differentiated cells (*Aldh1a2, Gabrg3, Cdh8, Vstmb2, Nrg3, Otud7a*) [127, 139, 140]. The hypomethylated genes were either amplified (*Hoxa11, Hoxa2*) [141] or methylated (*Irs2, Thbs1*) [142] in breast cancer or were downregulated in metastasis from malignant melanoma compared to primary tumors (*Esrp2*, *Palmd, Lsr*) [143].

Using published gene expression microarray data [90], we identified 6 pairs with significant expression differences of the protein-coding RNA between tumor and normal mammary gland samples. We could not analyze expression of the lncRNAs as they were mostly not covered on the array. Among the differentially expressed genes we selected *Esrp2* and *Esrp2-as* for further analysis. Loss of *Esrp2* has been associated with epithelial-to-mesenchymal transition [144-146], a process that plays an important role in metastasis formation.

Table 4-8 Candidate lncRNAs with differentially methylated promoters and neighboring protein-coding genes

DMR	lncRNA		protein-coding gene		expression difference		reference
Genomic position (mm10)	**name**	**ID**	**name**	**ID**	**Tumor - Normal[(a)]**	**p-value [(b)]**	
Hypermethylation							
chr1:180332417-180332819	*Gm5069*	NR_003623	*Itpkb*	NM_001081175			
chr2:74762966-74763423	*Haglr*	NR_110445	*Hoxd1*	NM_010467	-0.10	0.55	
chr3:37896886-37897206	*Gm20755*	NR_040559					
chr3:82876021-82876531	*Rbm46os*	NR_040382	*Rbm46*	NM_001277170			
chr4:114907033-114907726	*Foxd2os*	NR_030721	*Foxd2*	NM_008593	-0.34	0.09	
chr4:145463820-145464350	*Smarca5-ps*	NR_002888					
chr4:152697098-152697448	*Gm833*	NR_033138					
chr5:19226691-19226991	*4921504A21Rik*	NR_102341	*Magi2*	NM_001170746	0.64	0.23	
chr5:57718135-57718436	*4932441J04Rik*	NR_015588	*Pcdh7*	NM_018764	-1.26	0.0007 ***	[147]
chr6:47943214-47943743	*Zfp783*	NR_027963					
chr6:52315701-52316027	*5730457N03Rik*	NR_038163	*Evx1*	NM_007966	-0.34	0.10	
chr6:6864808-6865158	*Dlx6os2*	NR_002839					
chr7:14623181-14623487	*Nlrp5-ps*	NR_045119					
chr7:40899025-40899517	*A230077H06Rik*	NR_040329	*Vstm2b*	NM_021387	-0.49	0.13	
chr7:51511083-51511470	*Ano5*	NR_073508					
chr7:57386794-57387157	*Gm9962*	NR_033504	*Gabrg3*	NM_008074	-0.45	0.04 *	[148]
chr7:63444419-63444734	*4930554H23Rik*	NR_131089	*Otud7a*	NM_130880	2.36	0.00009 ****	[149]
chr7:79515301-79515977	*AI854517*	NR_040312					
chr8:120230039-120230397	*A330074K22Rik*	NR_110496					
chr8:121084994-121085403	*Fendrr*	NR_045471	*Foxf1*	NM_010426	0.30	0.11	
chr8:84912426-84912992	*Gm38426*	NR_103491					
chr8:99416005-99416420	*A330008L17Rik*	NR_132435	*Cdh8*	NM_007667	0.01	0.97	
chr9:71216962-71217270	*Gm3458*	NR_110518	*Aldh1a2*	NM_009022	-0.39	0.13	
chr10:93336092-93336432	*Gm17745*	NR_038014					

chr11:95143377-95143691	*Dlx4os*	NR_040279	*Dlx4*	NM_007867	0.05	0.90	
chr13:111868178-111868548	*Gm15326*	NR_130345					
chr13:72816340-72816825	*D730050B12Rik*	NR_046196					
chr14:118234395-118234797	*LOC105245869*	NR_131969	*Sox21*	NM_177753	-0.27	0.37	
chr14:39472112-39472839	*LOC432842*	NR_131973	*Nrg3*	NM_008734	-0.15	0.59	
chr15:83366702-83367300	*1700001L05Rik*	NR_027980					
chr16:98081806-98082497	*A630089N07Rik*	NR_015491					
chr17:34094655-34095214	*BC051537*	NR_046183					
chr17:80373020-80373349	*Gm10190*	NR_028385					
chr19:30539362-30539867	*Ppp1r2-ps3*	NR_003650					
chrX:12159988-12160665	*2900008C10Rik*	NR_045434					
chrX:12761797-12762195	*Gm14634*	NR_045852	*Med14*	NM_012005	-0.31	0.13	
Hypomethylation							
chr1:133269974-133270300	*Gm19461*	NR_037984					
chr1:136696173-136696502	*Platr22*	NR_037986					
chr1:71888158-71888468	*Gm8883*	NR_027658					
chr2:118112952-118113258	*Gm13986*	NR_126479	*Thbs1*	NM_011580			
chr3:116968252-116968565	*4930455H04Rik*	NR_040596	*Palmd*	NM_023245	-0.09	0.84	
chr5:112206149-112206560	*1700028D13Rik*	NR_045377					
chr5:112206149-112206560	*1700028D13Rik*	NR_045378					
chr5:135000770-135001082	*Wbscr25*	NR_026907					
chr5:143758048-143758457	*D130017N08Rik*	NR_015486					
chr6:52165042-52165362	*Hoxaas2*	NR_131182	*Hoxa2*	NM_010451	-0.39	0.28	
chr6:52243386-52243706	*Hoxa11os*	NR_015348	*Hoxa11*	NM_010450	-0.31	0.05 *	[150, 151]
chr7:28496984-28497589	*1700028B04Rik*	NR_033605					
chr7:30971903-30972229	*Fam187b*	NR_038860	*Lsr*	NM_017405	3.57	0.00004 ****	[152]
chr8:106133504-106133810	*1810019D21Rik (Esrp2-as)*	NR_040344	*Esrp2*	NM_176838	1.55	0.0005 ***	[144, 145]

chr8:11005553-11005971	*9530052E02Rik*	NR_046017	*Irs2*	NM_001081212			
chr8:70774591-70774966	*2010320M18Rik*	NR_029440	*Pik3r2*	NM_008841	0.47	0.11	
chr9:40333291-40333624	*1700110K17Rik*	NR_040728					
chr11:102377699-102378035	*Bloodlinc*	NR_131196					
chr11:112711182-112711489	*BC006965*	NR_024085					
chr12:11204247-11204597	*9530020I12Rik*	NR_131083					
chr12:33149419-33149899	*F730043M19Rik*	NR_015602	*Atxn7l1*	NM_028139	-0.18	0.53	
chr13:44216558-44216865	*A330076C08Rik*	NR_045088					
chr13:63296450-63296765	*Gm16907*	NR_045794					
chr14:105589344-105589691	*9330188P03Rik*	NR_102319					
chr14:118363552-118363938	*1700044C05Rik*	NR_045624					
chr14:118922434-118923043	*Dzip1*	NR_130725					
chr14:55071799-55072110	*Zfhx2os*	NR_004444					
chr15:103147888-103148190	*D930007P13Rik*	NR_045743					
chr16:95928970-95929274	*1600002D24Rik*	NR_040484					
chr17:56554141-56554483	*Gm20219*	NR_130128					
chr19:53076038-53076348	*1700054A03Rik*	NR_045320					

[(a)] Expression data were taken from [90] and include eight TG tumor samples and five WT mammary glands. Values are reported as the differences between tumor and normal samples in log2 array intensities relative to a pool of whole mouse embryo RNA.

[(b)] p-values were calculated using two sided Student's t-test, * $p < 0.05$, *** $p < 0.001$, **** $p < 0.0001$

4.3.2 *Esrp2* and *Esrp2-as* are coordinately overexpressed in C3(1) tumors

Esrp2-as is expressed as four annotated transcripts with a length between 1.2 and 1.6kb (Figure 4-31). The three longer transcripts (v1-3) share a common TSS approximately 1.6kb downstream of the *Esrp2* TSS. In contrast, the TSS of the short variant (v4) is located about 100bp upstream of the *Esrp2* TSS. Thus, the three longer lncRNA transcripts overlap partly with *Esrp2*, but still share two exons and an intronic region with the short variant. The DMR identified by MCIp-seq was located about 1.7kb upstream of the v1-3 TSS overlapping with exon 8 of *Esrp2*. The TSS of the short variant v4 was located at the border of a CpG island, which also covers exon 1 and 2 of *Esrp2*.

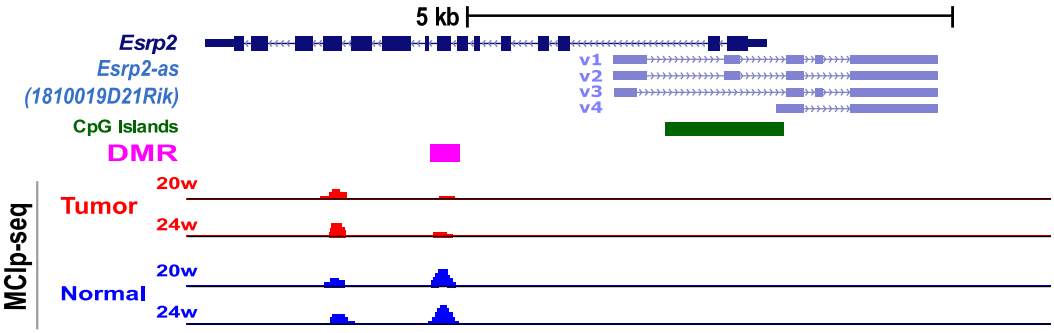


Figure 4-31 DMR identified by MCIp-seq at the *Esrp2* region. Depicted is the genomic organization of *Esrp2* (dark blue) and *Esrp2-as* variants 1-4 (v1-4, light blue). Enrichment of methylated fragments by MCIp is shown for tumors (red) and normal WT mammary glands (blue), each lane representing the average of sequencing reads obtained for three samples at age 20 and 24 weeks. The DMR (pink) is located ≈ 1.7kb upstream from the TSS of *Esrp2-as* v1-3. The CpG island (green) covers exon 1 and 2 of *Esrp2*.

In order to validate the reported difference in *Esrp2* expression [90] in our own data set and to examine expression levels of *Esrp2-as*, we performed RT-qPCR analysis and confirmed a significant 3-fold higher expression of *Esrp2* in tumors vs. normal tissue (Figure 4-32A). Since the different splice variants of *Esrp2-as* could not be detected individually by RT-qPCR, we designed assays to detect either the long variants 1+2 (v1+2) by an assay spanning the first intron, or all 4 variants together (v1-4) using primers located in the common last exon. When detecting all four variants together, *Esrp2-as* was about 20-fold higher expressed than when analyzing only the long variants v1+2. Publicly accessible FANTOM5 CAGE-seq (Cap Analysis of Gene Expression) data [153, 154] confirmed transcription initiation for both the short and the long variants of *Esrp2-as* in mammary glands of lactating and pregnant mice (Figure 4-32B), supporting expression of both the long and the short transcript variants. *Esrp2-as* v1-4 levels were significantly increased by about 2-fold in tumor tissue, whereas the long variants (v1+2) did not show differential expression.

This indicated that in tumorigenesis the short variant v4 might be stronger induced than the long variants, which were kept at a constant level of expression.

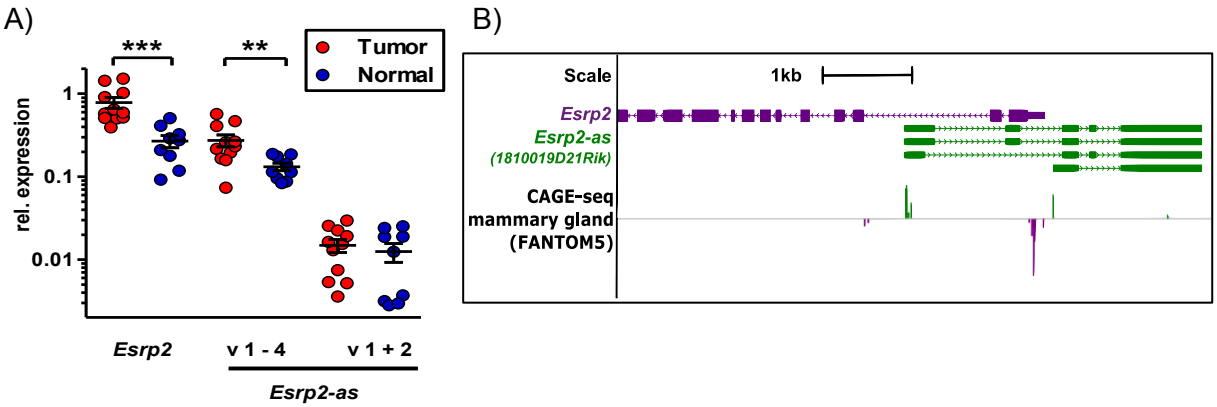


Figure 4-32 Gene expression in C3(1) tumors and normal mammary glands. A) Relative expression levels as determined by RT-qPCR are significantly different between tumor (n=11) and normal samples (n=9) for *Esrp2* and *Esrp2-as* (v1-4), but not for *Esrp2-as* (v1+2). Samples were derived from animals aged 20-24 weeks, and expression levels were normalized to three reference genes (*Hprt1, Tbp, β-Actin*). Mann-Whitney-U test, ** p< 0.01, *** p< 0.001. B) CAGE-seq data from FANTOM 5 were obtained for mammary gland samples (CNhs 10480 & CNhs 10476) via the Zenbu genome browser [154, 155] in the region of *Esrp2*. Color coding indicates strand specificity of CAGE signals (green: + strand, purple: - strand).

Correlation analysis indicated co-expression of *Esrp2* and the antisense transcripts in tumor tissue and normal mammary glands (*Esrp2-as* v1-4, Spearman r= 0.85, p< 0.0001). When extending the expression analysis to spleen and liver samples derived from C3(1) mice as well as various mouse cell lines, we confirmed strong, positive correlation for *Esrp2* and the lncRNA transcripts (Spearman r= 0.88, p< 0.0001 (*Esrp2-as* v1-4) and r= 0.60, p< 0.0001 (*Esrp2-as* v1+2)), but also correlation between the different antisense transcripts (Spearman r= 0.66, p< 0.0001 (Figure 4-33 and Figure 4-34)). These results suggested either mutual regulation of expression between the coding and noncoding RNA, or a common control mechanism.

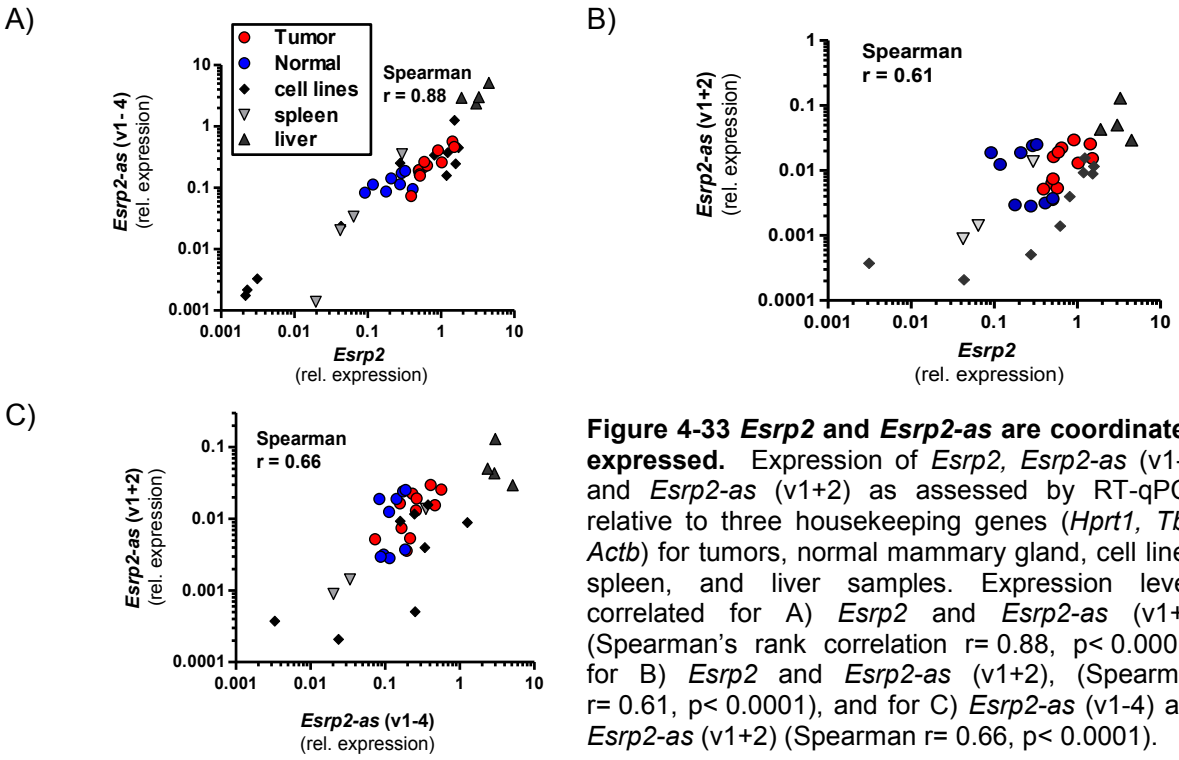


Figure 4-33 *Esrp2* and *Esrp2-as* are coordinately expressed. Expression of *Esrp2, Esrp2-as* (v1-4) and *Esrp2-as* (v1+2) as assessed by RT-qPCR relative to three housekeeping genes (*Hprt1, Tbp, Actb*) for tumors, normal mammary gland, cell lines, spleen, and liver samples. Expression levels correlated for A) *Esrp2* and *Esrp2-as* (v1+4) (Spearman's rank correlation r= 0.88, p< 0.0001), for B) *Esrp2* and *Esrp2-as* (v1+2), (Spearman r= 0.61, p< 0.0001), and for C) *Esrp2-as* (v1-4) and *Esrp2-as* (v1+2) (Spearman r= 0.66, p< 0.0001).

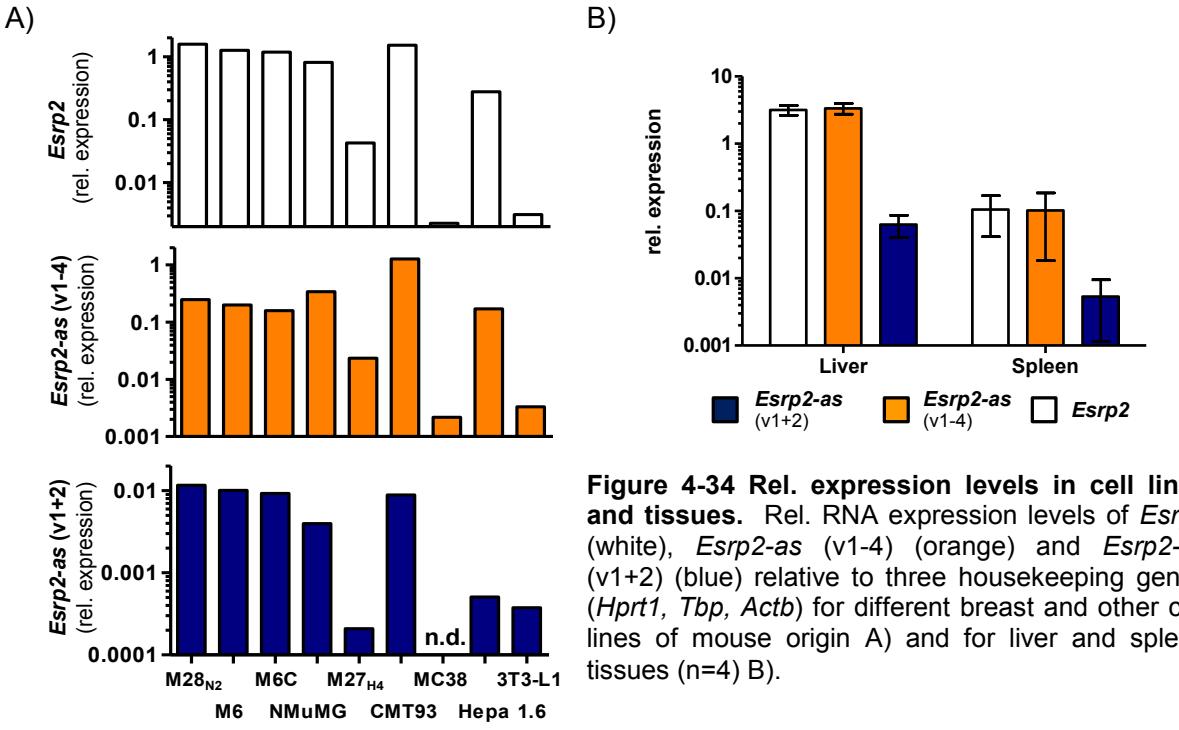


Figure 4-34 Rel. expression levels in cell lines and tissues. Rel. RNA expression levels of *Esrp2* (white), *Esrp2-as* (v1-4) (orange) and *Esrp2-as* (v1+2) (blue) relative to three housekeeping genes (*Hprt1, Tbp, Actb*) for different breast and other cell lines of mouse origin A) and for liver and spleen tissues (n=4) B).

4.3.3 CGI shores of *Esrp2* are differentially methylated in tumors and cell lines

Using published whole genome bisulfite sequencing (WGBS) and RNA-seq datasets of different mouse tissues [156-159] (Figure 4-35), we detected that the CGI located in the

Esrp2 promoter region was generally unmethylated in all tissues, whereas the CGI shores displayed variable methylation between tissues that express *Esrp2* and *Esrp2-as* (liver, kidney, stomach, lung) and those that do not (spleen, heart). This indicated that differential methylation was not only limited to the region that we identified by MCIp-seq.

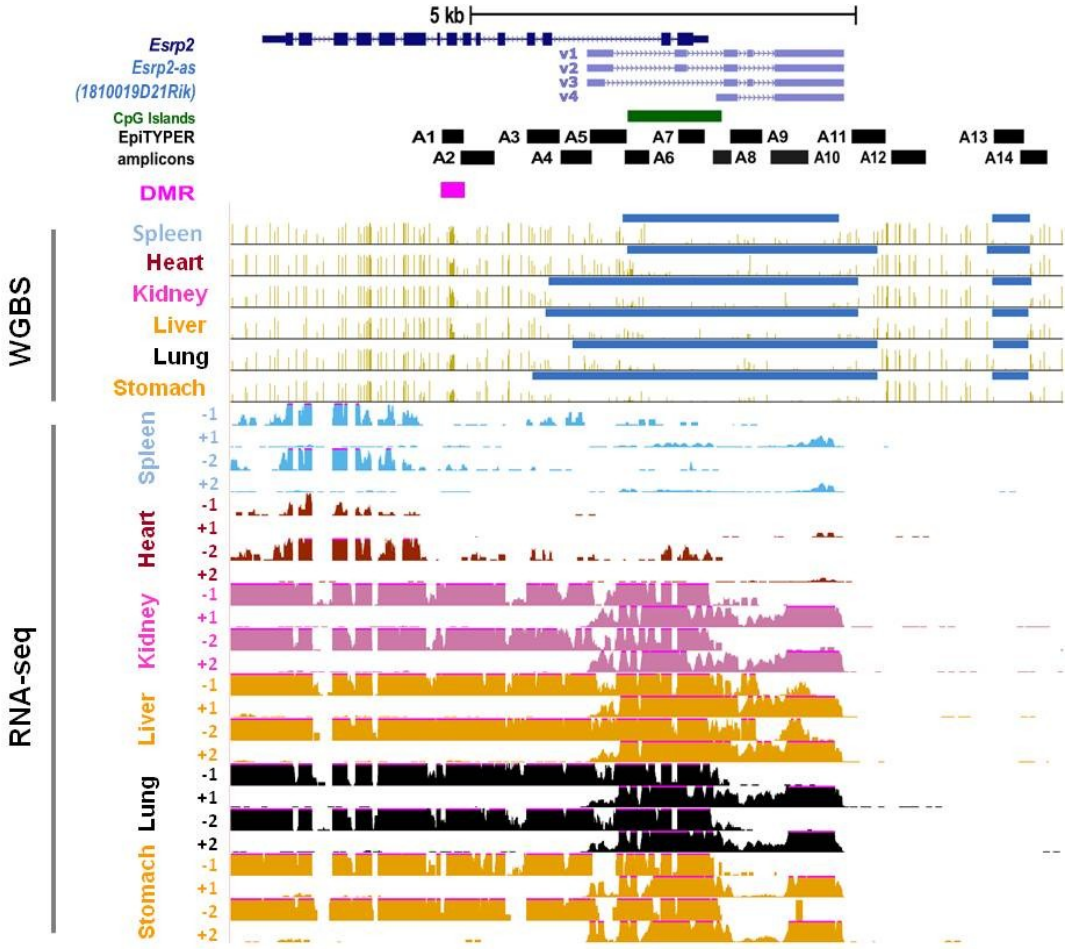


Figure 4-35 DNA methylation and RNA-seq levels for different mouse tissues in the *Esrp2* region. Data tracks were downloaded from the UCSC genome browser. Whole genome bisulfite sequencing tracks (WGBS) were taken from the Hon_2013 track [156, 157] and RNA-seq data from CSHL_Long_RNA-seq track [158, 159]. Blue horizontal bars represent regions of hypomethylation, while DNA methylation levels at individual CpG sites are depicted by orange vertical lines. RNA-seq reads are reported strand specifically (+, -) for replicates 1 and 2 by color codes for the different tissues. The genomic region is depicted as in Figure 4-31, with the addition of amplicons A1-A14 (black boxes) for EpiTYPER MassARRAY analysis.

In order to analyze methylation levels in a quantitative manner, we designed EpiTYPER MassARRAY amplicons covering the DMR (Figure 4-35, Amplicon A1), the CGI (A6-A8), as well as the shore regions on both sides of the CGI (A2-5, A9-14). MassARRAY analyses confirmed that the DMR (A1) was hypomethylated in tumors vs. normal mammary glands with an average methylation difference of 35% (Figure 4-36A).

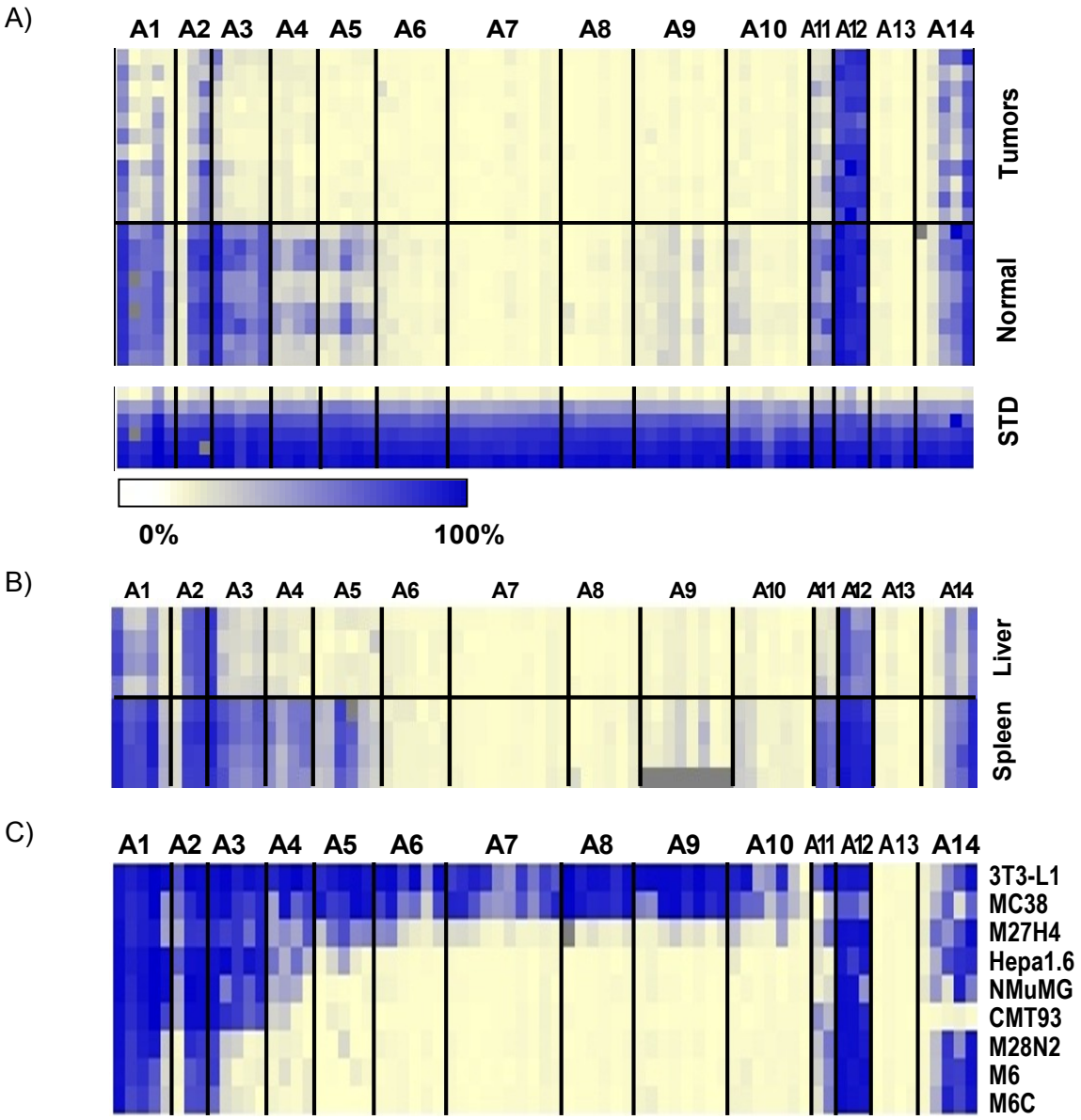


Figure 4-36 DNA methylation at the *Esrp2* region. DNA methylation levels for tumor samples (n=11) and normal mammary gland tissue (n=9) A), for liver and spleen samples (n=4) B) or various cell lines C) are determined by quantitative EpiTYPER MassARRAY technology. Heatmaps display DNA methylation with each row representing one individual sample and each column one CpG unit comprising of 1 to 4 individual CpG sites. Locations of the different amplicons relative to the genes are depicted in Figure 4-35. DNA methylation is displayed by a color coded gradient from 0% (light yellow) to 100% methylation (blue). Gray squares represent failed measurements.

The central region spanning the CGI (A6-A8) was unmethylated in both tumor and normal tissue. In contrast, individual CpG units in the shore regions A3-A5 and A9-A11 were 10 to 50% less methylated in tumors than in normal mammary glands. As WGBS tracks are not yet available for murine mammary gland tissue, we selected four spleen and four liver samples of the C3(1) mouse model to determine concordance of MassARRAY data with published WGBS data (Figure 4-36B). Indeed, methylation values for spleen and liver

correlated well between techniques (Spearman r= 0.67, p< 0.0001 (liver) and r= 0.77, p< 0.0001 (spleen)). For comparison, we also examined DNA methylation levels in various cell lines (Figure 4-36C). Whereas most of the cell lines had a methylation profile either similar to normal mammary glands ($M27_{H4}$ (mammary cells derived from the C3(1) mouse model), Hepa1.6 (murine hepatoma), NMuMG (murine mammary gland), CMT93 (murine rectal carcinoma)) or to tumor tissue ($M28_{N2}$, M6, M6C, all three derived from the C3(1) mouse model), two cell lines (3T3-L1 (murine preadipocytes), MC38 (murine colon carcinoma)) were completely methylated (70-90% methylation) in the analyzed region, except for amplicon A13 which was unmethylated in all samples. Consistently, 3T3-L1 and MC38 cell lines displayed also the lowest expression levels of *Esrp2* and *Esrp2-as* (Figure 4-34).

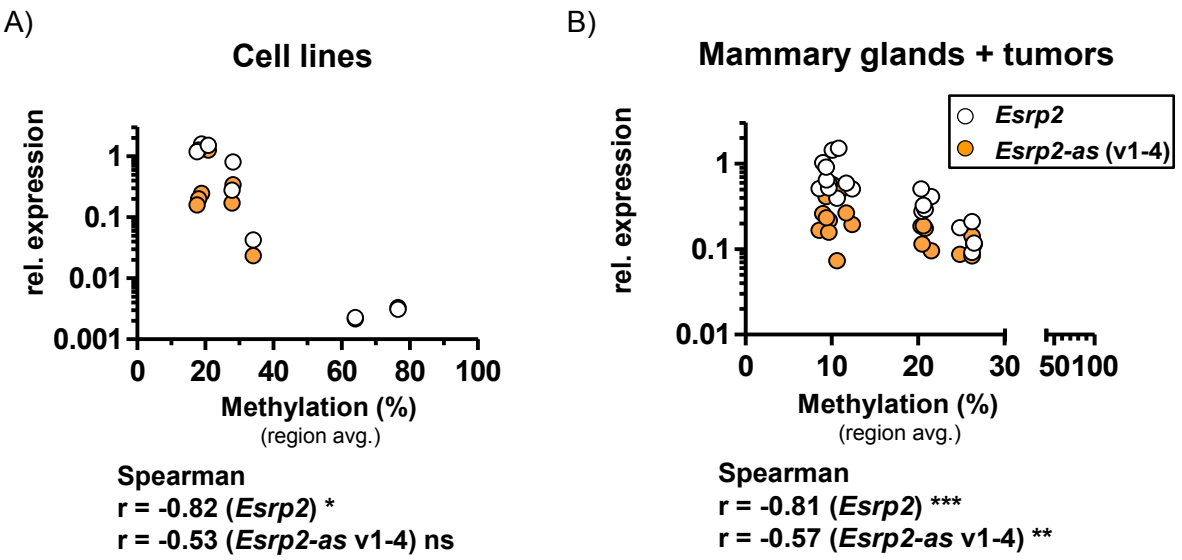


Figure 4-37 DNA methylation inversely correlates with gene expression of *Ersp2* and *Esrp2-as*. Correlation between DNA methylation levels (average over complete region A1-A14 as shown in Figure 4-36) and expression levels normalized to three housekeeping genes (*Hprt1, Tbp, Actb*) was calculated by Spearman's rank correlation for cell lines A) and tumor/normal tissues B). * p< 0.05, ** p< 0.01, *** p< 0.001.

We observed strong, negative correlation between expression and DNA methylation in cell lines (Spearman r= -0.82, p= 0.02 (*Esrp2*) and r= -0.53, p= 0.1 (*Esrp2-as* v1-4)) (Figure 4-37A), as well as levels in tumor samples and normal tissue (Spearman r= -0.81, p< 0.0001 (*Esrp2*) and r= -0.57, p = 0.008 (*Esrp2-as* v1-4)) (Figure 4-37B). These results confirmed that the region around *Esrp2* was differentially methylated, and methylation inversely correlated with expression.

4.3.4 *ESRP2* in human breast cancer

In order to test whether our findings regarding *Esrp2* are also relevant for human breast cancer, we extracted DNA methylation levels in the region around the TSS of *ESRP2* from TCGA 450k. Breast tumors were indeed significantly hypomethylated in the region downstream of the *ESRP2* TSS compared to normal breast tissue (Figure 4-38A). Besides, *ESRP2* was significantly overexpressed in TCGA breast cancer samples (Figure 4-38B), and expression inversely correlated with average DNA methylation levels in that region (Spearman's rank correlation r= -0.29, p< 0.0001). Furthermore, high expression of *ESRP2* was significantly associated with lower disease-free survival (hazard ratio 1.24, 95% confidence interval 1.098 – 1.409) (Figure 4-38C)) as indicated by Breastmark data [160]. So far, a human homologue of *Esrp2-as* has not been annotated yet, but CAGE-seq data for MCF7 human breast cancer cells (FANTOM5 [153]) suggest transcription initiation at genomic locations corresponding to the TSSs of mouse *Esrp2-as*.

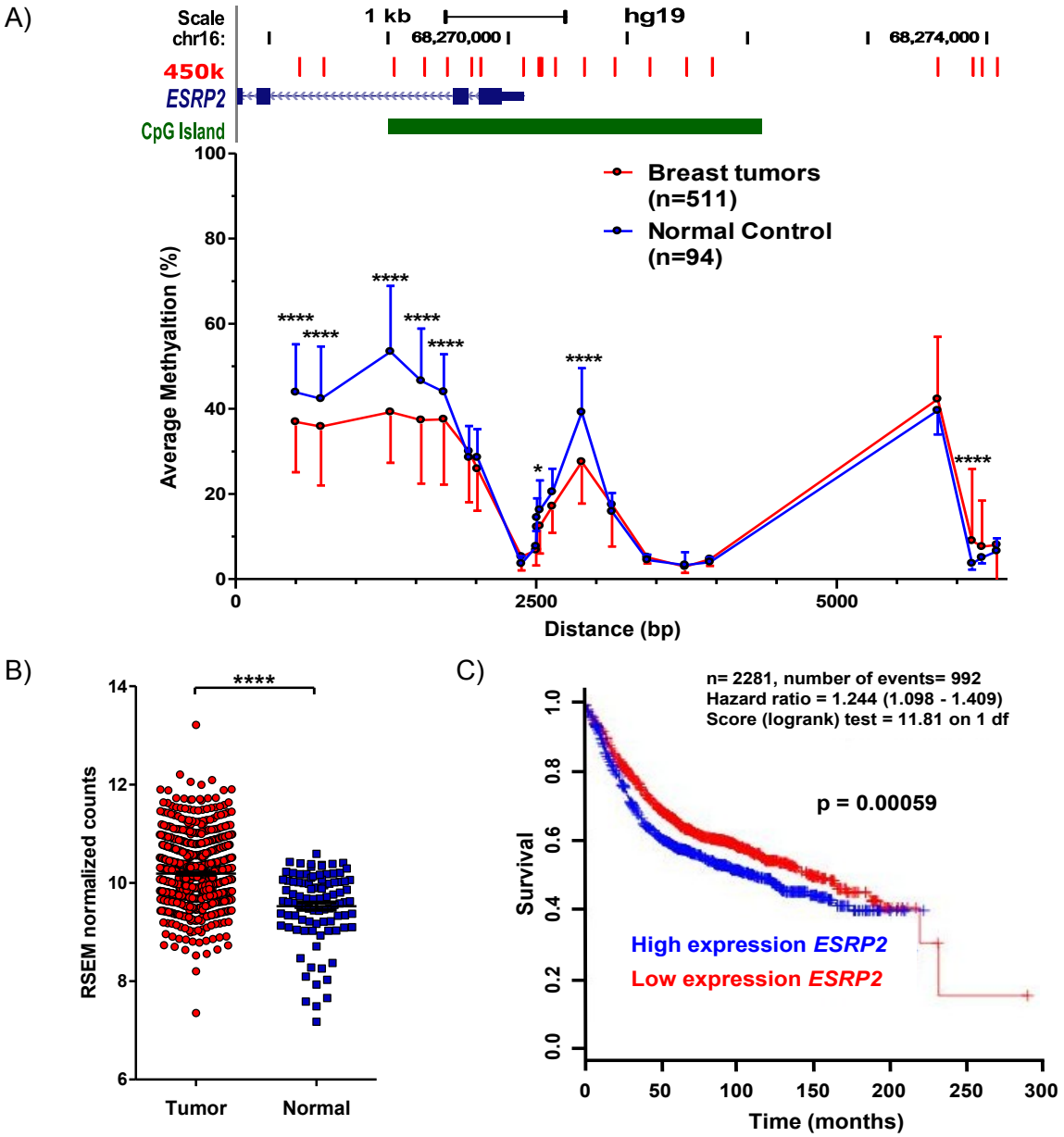


Figure 4-38 *ESRP2* hypomethylation and overexpression lowers chances of disease-free survival in human breast cancer. A) 450k array data were downloaded for 511 breast tumors (red circles) with available Pam50 classification from TCGA data download and compared with 94 normal control samples (blue squares). Values are depicted as mean ± SD. The scheme (top) depicts the genomic location of human *ESRP2* (blue) and respective positions of 450k methylation probes (red). The location of the CGI is indicated in green. B) *ESRP2* is overexpressed in human breast cancer samples from the TCGA data set. RNA-seq gene expression data were downloaded from UCSC cancer browser (TCGA BRCA gene expression (IlluminaHiSeq)) for 825 tumor samples with available Pam50 classification and were compared to 96 normal control samples by two-sided Student's t-test **** p< 0.0001. RSEM: RNA-Seq by Expectation-Maximization. C) Kaplan Meier curves of disease-free survival are plotted for *ESRP2* expression. A total of 2281 breast cancer patients with an overall number of 992 events are separated by median *ESRP2* expression into a high (blue) and low (red) expressing group. p = 0.00059 (logrank test). Data obtained from Breastmark [160].

4.3.5 Demethylation induces reexpression of *Esrp2* and *Esrp2-as*

In order to analyze whether changes in DNA methylation would result in gene expression changes, we performed demethylation experiments in the $M27_{H4}$ murine breast cancer cell line treated with 1µM Dac. DNA methylation levels were reduced by 10-30% in amplicons with methylation levels above 20% in the DMSO control, with A1 being among the strongest demethylated amplicons (Figure 4-39A). Both *Esrp2* and *Esrp2-as* v1-4 transcript levels increased by 3- and 4-fold respectively, whereas the increase in *Esrp2-as* v1+2 expression was only marginal (Figure 4-39B), which further supports the importance of the short variant v4.

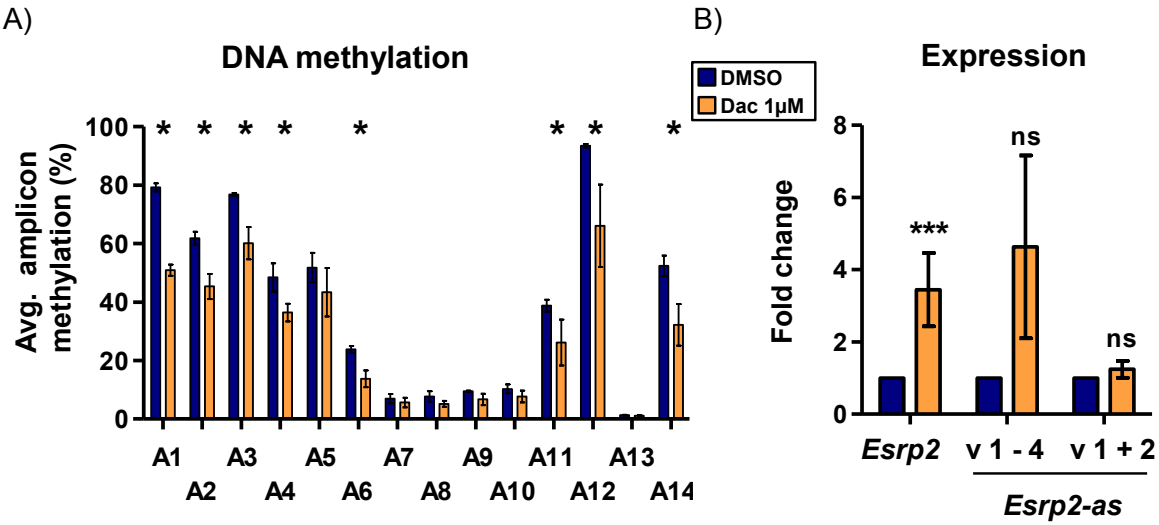


Figure 4-39 Decitabine (Dac) treatment induces demethylation and reexpression of *Esrp2* and *Esrp2-as*. A) Treatment of $M27_{H4}$ cells with 1µM Dac decreased DNA methylation levels. Depicted are the average amplicon methylation levels for A1-A14 with the mean ± SD of 3 independent experiments. Mann Whitney test (one-sided), * $p< 0.05$. B) Dac treatment induces reexpression of *Esrp2* and *Esrp2-as* in $M27_{H4}$ cells. Expression is measured relative to three housekeeping genes (*Hprt1, Tbp, Actb*) and normalized to DMSO control. Depicted is the mean ± SD of 3 independent experiments. Unpaired Student's t-test (one-sided). *** $p< 0.001$, ns: not significant.

4.3.6 Knockdown and overexpression of *Esrp2-as* does not affect *Esrp2* expression

lncRNAs have been associated with gene silencing functions by interacting with modifiers of epigenetic marks, such as the polycomb repressor complexes or DNA methyltransferases. More recently, lncRNAs were also reported to target gene activating functions to specific DNA loci, such as the lncRNA TARID involved in demethylation and expression of tumor suppressor gene TCF21 (transcription factor 21) [54].

Dac treatment experiments did not allow to discriminate whether reexpression of the lncRNA transcripts upregulated expression of the coding RNA or whether both RNAs were regulated by a common mechanism. To test a mutual regulation of both RNAs, we knocked down expression of *Esrp2-as* with locked nucleic acid antisense Gapmers (LNAs). Transfection of M6 murine breast cancer cells with two different LNAs targeting the last exon of the *Esrp2-as* variants reduced antisense transcript levels by 60-80% after 96h, whereas *Esrp2* levels were not influenced (Figure 4-40A).

In case that the lncRNA influences *Esrp2* expression, this could either happen at the level of *de novo* transcription or at the level of RNA stability, which might require more time to manifest. We therefore performed a series of four repeated knockdowns by transfecting the cells with the LNAs every 96h. Neither the knockdown efficiency of the lncRNA nor its effect on *Esrp2* expression was enhanced after this knockdown series (Figure 4-40A). Knockdown of *Esrp2-as* in the $M28_{N2}$ and NMUMG cell line also did not lead to a reexpression of *Esrp2* (Figure 4-40B).

Conversely, transient overexpression of *Esrp2-as* v1 or v4 in $M27_{H4}$ cells resulted on average in a 30-fold and 2500-fold higher expression of *Esrp2-as* v1-4 and *Esrp2-as* v1+2 respectively, but *Esrp2* transcript levels were unaffected (Figure 4-40C). Additional attempts of overexpression in the Hepa1.6 and 3T3-L1 cell line produced similar results (Figure 4-40D).

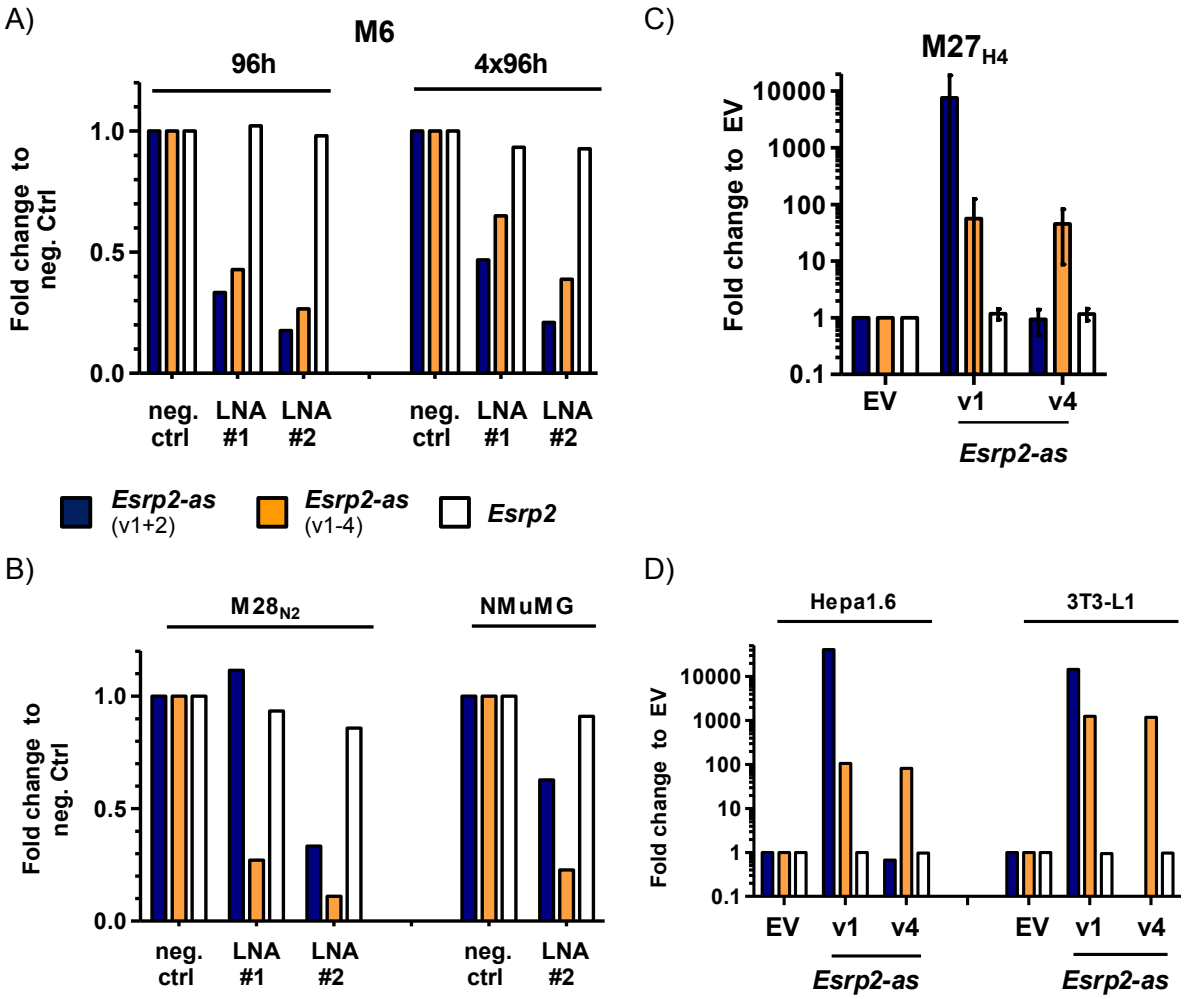


Figure 4-40 Knockdown and overexpression of *Esrp2-as*. A) Knockdown of *Esrp2-as* by Locked Nucleic Acid antisense Gapmers (LNAs) in the M6 cell line for 96h or 4x 96h. Expression levels are measured relative to three housekeeping genes (*Hprt1, Tbp, Actb*) and are normalized to cells transfected with negative control LNA. Experiments were performed once. B) $M28_{N2}$ and NMuMG cells were transiently transfected with LNAs against *Esrp2-as* and examined with RT-qPCR after 96h ($M28_{N2}$) and 72h (NMuMG) for changes in expression levels. Experiments were conducted once. C) $M27_{H4}$ cells were transiently transfected with constructs containing *Esrp2-as* v1 or v4 for overexpression. Expression levels were assessed by RT-qPCR 72h post-transfection and normalized to cells transfected with the empty vector (EV). Mean ± SD of 7 independent experiments, student's t-test did not show significant changes in *Esrp2* expression. D) Overexpression of *Esrp2-as* in additional cell lines. Hepa 1.6 (left) and 3T3-L1 cells (right) were transiently transfected with *Esrp2-as* v1 or v4 constructs and expression was determined after 48h. Depicted is the mean of 2 independent experiments and thus data were not evaluated by statistical means.

In order to exclude the possibility that expression of the coding transcript regulates the noncoding RNA, we also generated a $M27_{H4}$ mammary cell and 3T3-L1 preadipocyte cell population that stably overexpressed *Esrp2* (Figure 4-41). Under these conditions, *Esrp2* was about 100-fold higher expressed than in the EV control in $M27_{H4}$ and almost 2000-fold higher in the 3T3-L1 cell line, but *Esrp2-as* levels were not changed.

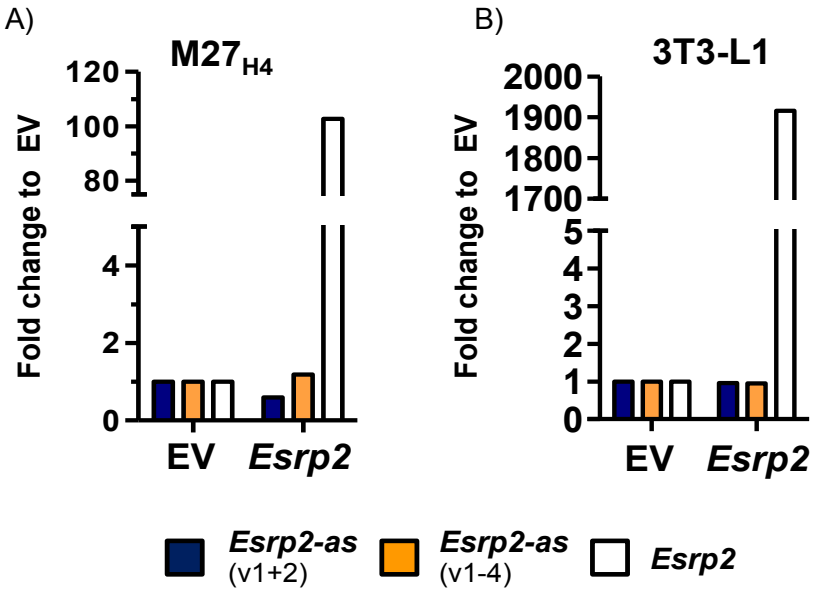


Figure 4-41 Overexpression of *Esrp2*. $M27_{H4}$ A) and 3T3-L1 B) cells were retrovirally infected with either EV or pMXs-*Esrp2* and were selected for incorporation of the gene by Blasiticidin S to generate stable cell lines. RT-qPCR measurements were normalized to EV. Experiments were performed once.

These data taken together suggested that neither *Esrp2* nor *Esrp2-as* are controlled by the respective transcript on the opposite strand and instead indicated a common regulation of expression, which is controlled by DNA methylation.

4.3.7 Luciferase reporter assays confirm a bidirectional promoter and an enhancer region

In order to identify the regions involved in the regulation of *Esrp2* and *Esrp2-as* expression, we performed dual luciferase reporter assays using promoter constructs (P) of *Esrp2* and the long variants of *Esrp2-as*, v1-3 (Figure 4-42A). Reporter assay results confirmed promoter activity for the *Esrp2* P-fragments. The P1 fragment located closest to the *Esrp2* TSS was associated with strong 4-fold induction of luciferase activity compared to EV (Figure 4-42B). For *Esrp2-as* constructs P1-P4, luciferase activity was comparable to that of EV. These findings were concordant with the weaker expression levels of the long variants *Esrp2-as* v1+2 compared to the combined transcripts v1-4 and further indicated that the short variant v4 might be the major *Esrp2-as* transcript. Indeed, luciferase assays with *Esrp2* P1 in reverse orientation relative to the luciferase gene (sense to the *Esrp2-as* v4 transcript) revealed almost equally high luciferase activity as for the *Esrp2* P1 fragment (Figure 4-42C).

Therefore, we concluded that the *Esrp2* P1 region has bidirectional promoter activity and is relevant for expression of both *Esrp2* and *Esrp2-as* v4.

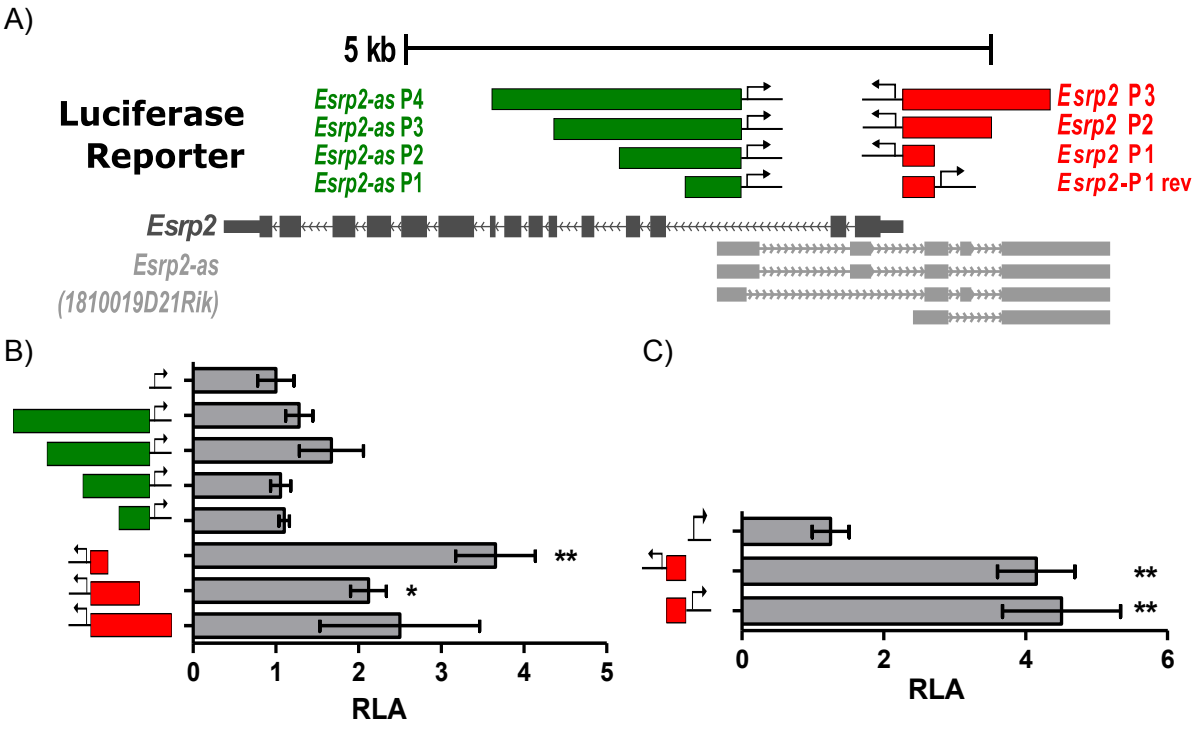


Figure 4-42 Luciferase reporter assays confirm bidirectional promoter activity at *Esrp2*. A) Schematic representation of luciferase reporter constructs to determine promoter activity of fragments covering the *Esrp2* region. Location of *Esrp2* and *Esrp2-as* are depicted in grey and promoter constructs are marked in red and green. B,C) Reporter plasmids with firefly luciferase for promoter B) and bidirectional promoter C) activity were transiently transfected into Hepa1.6 cells and luciferase activity was measured 48h post transfection and normalized to a co-transfected CMV-*Renilla* luciferase construct. Transfections were conducted parallel in eight technical replicates and reported is the mean ± SEM of four independent experiments normalized to the pGL4.10 EV. Unpaired Student's t-test (two-sided), asterisks represent comparisons against EV, * $p < 0.05$, ** $p < 0.01$. RLA: relative luciferase activity.

DNA methylation levels in the *Esrp2* P1 region (Figure 4-35 and Figure 4-36, amplicon A7-A8) are very low in both normal mammary glands and tumors as well as in all other tissues with publicly available methylation data and thus failed to explain the observed differences in expression between tumor and normal tissue and among different tissues. Examining our data for chromatin states we observed tumor-specific enhancer states next to the TSS of *Esrp2-as* (Figure 4-43A). This suggested that in tumors these regions might have enhancer rather than promoter function. Therefore, we analyzed *Esrp2-as* E1-E4 regions in luciferase vectors with a minimal promoter (pGL4.23) to test for enhancer activity. The E4 region demonstrated the strongest activity of the enhancer constructs and increased the luciferase expression from the minimal promoter significantly by about 5-fold (Figure 4-43B). For confirmation of enhancer activity on the specific *Esrp2* P1 promoter, we also positioned the

E4 fragment in front of the bidirectional *Esrp2* P1 fragments. In this setting, we could not confirm added enhancer activity of E4 (Figure 4-43C). This might be due to the fact that the activity of the bidirectional promoter fragment was already higher than that of the minimal promoter of the pGL4.23 empty vector and thus might not respond to a similar extent to the additional enhancer effect.

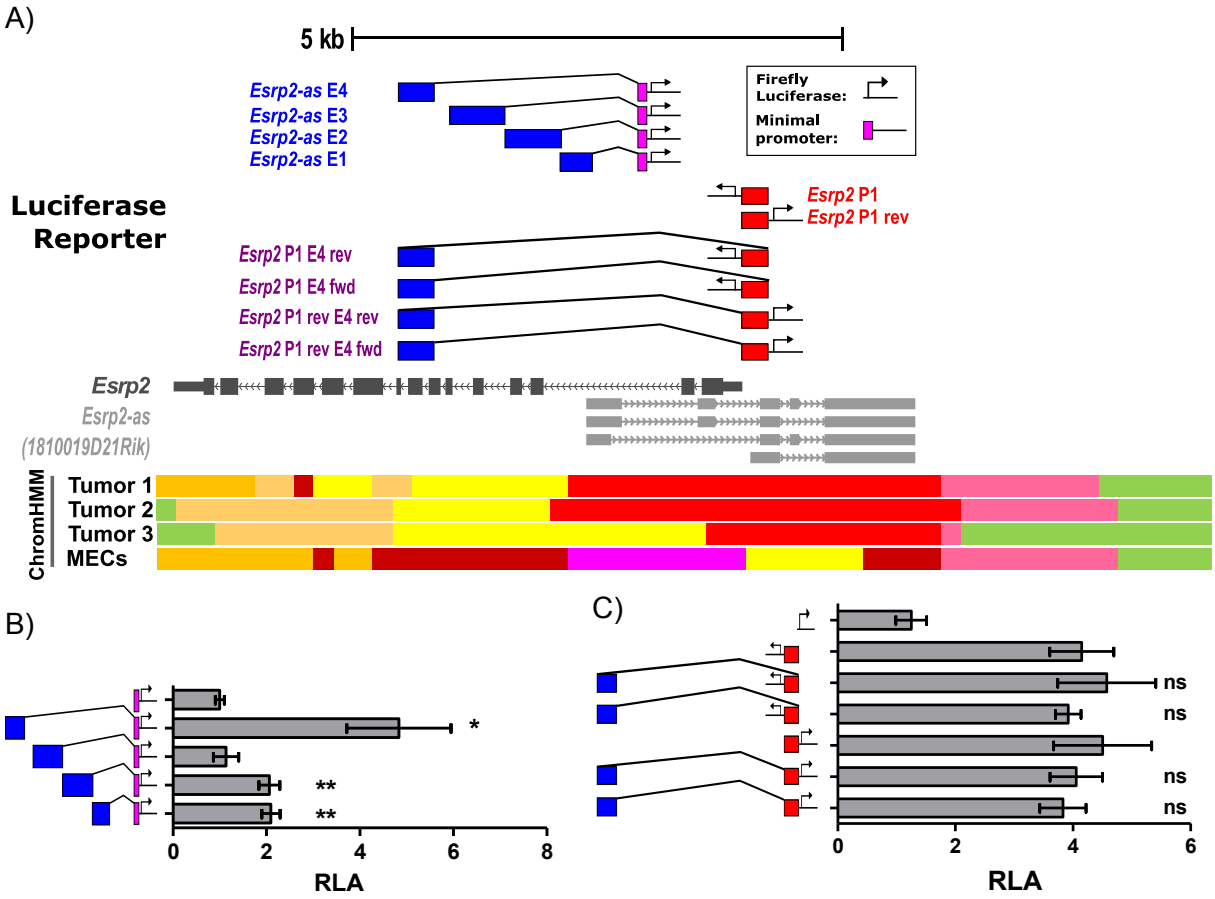


Figure 4-43 *Esrp2-as* upstream regions exhibit enhancer activity. A) Schematic representation of enhancer reporter fragments with respect to genomic location and association with chromatin states. Enhancer reporters are displayed as blue boxes combined with a minimal promoter (pink). *Esrp2* promoter fragment P1 is depicted in red, enhancer fragment E4 is indicated in blue for P1 specific enhancer constructs. Fwd (forward) and rev (reverse) describe the orientation of the fragments relative to the natural orientation. Location of *Esrp2* and *Esrp2-as* is depicted in grey. Color code for chromatin states obtained by ChromHMM as for Figure 4-19. B, C) Transfections and measurements were conducted and reported as in Figure 4-43 and were normalized to the pGL4.23 EV B) or the pGL4.10 EV C). Unpaired Student's t-test (two-sided), asterisks represent comparisons against pGL4.23 EV * $p < 0.05$, ** $p < 0.01$, ns: not significant. RLA: relative luciferase activity; EV: empty vector.

5. Discussion

5.1 DNA methylation changes in the C3(1) model

In order to gain information on genome-wide alterations in DNA methylation in the C3(1) mouse model, we conducted MCIp-seq using mammary gland and tumor tissue of different age groups. We confirmed a link between DNA methylation and gene expression at recurrent DMRs and demonstrated the general similarity of the methylation profile to human breast cancer samples.

5.1.1 Genome-wide DNA methylation studies in mouse models of cancer

The importance of DNA methylation for cancer formation in genetically engineered mouse models (GEMMs) has been frequently investigated by interfering with epigenetic modifying enzymes either by demethylating drugs such as decitabine and zebularine or by knockout of the epigenetic protein of interest such as DNMTs or MBD protein (reviewed in [161]). In breast cancer models, the relevance of DNA methylation was demonstrated by knockout of DNMT1 in the C3(1) and in the MMTV-Neu-Tg model, and by Zebularine treatment in the MMTV polyoma middle T antigen (PyMT) model. In all three models, depletion of DNMT1 reduced tumor formation and tumor burden, therefore distinguishing DNA methylation as an essential factor for tumorigenesis [162, 163].

Until recently, reports of DNA methylation analysis at a genome-wide scale in mouse models of cancer were limited, in contrast to human cancer studies. The efforts to investigate the genome-wide methylome in human cancer were greatly facilitated by the introduction of the Illumina 27k and 450k DNA methylation arrays. This allowed the evaluation of tens to hundreds of thousands of preselected CpG sites in a quantitative manner in an easily reproducible and relatively cheap manner [164]. An equivalent platform for other species is not on the market so far. Although Wong and colleagues [165] confirmed the applicability of the human arrays also for mouse tissues, this was at the cost of greatly reducing the number of analyzable sites to 1308 and 13715 sites on the 27k and 450k array, respectively. Therefore, genome-wide analysis in mouse models were mainly based on enrichment of methylated fragments by methylation-specific antibodies (MeDIP) or by proteins containing a methyl-CpG binding domain (MethylCap, MCIp) followed by microarray analysis [166, 167] or next generation sequencing [168-170]. Using the MCIp-seq approach in the C3(1) model of breast cancer, we identified about 5000 hyper- and 10.000 hypomethylated regions that were recurrently occurring during tumorigenesis and localized in different genomic regions, such as promoters, exons, or intergenic regions. We observed that a majority of DMRs was located outside of promoter regions and CGIs. Therefore, the sequencing approach provided

a greater coverage of regions than approaches using microarrays that are mostly limited to promoter regions or CGIs. The functional relevance of these DMRs outside of CGIs and promoters is less well understood, although examples were discovered that DNA methylation influences splicing [171] or binding of transcription factors [172]. Nevertheless, MCIp is based on enrichment and requires both a high CG density and intermediate to high levels of methylation for binding of DNA fragments. Despite the increased coverage of regions by MCIp-seq compared to analysis by microarray, interesting candidates might still have escaped our analysis. Furthermore, the identification of candidate DMRs with MCIp is based on relative quantification of differences and requires a quantitative confirmation by other techniques such as pyrosequencing or EpiTYPER. The validation rate with the EpiTYPER technique was higher than 90%. We managed not only to confirm significant methylation differences to age-matched WT animals in late age groups, when tumors had developed, but also demonstrated high correlation between sequencing coverage and DNA methylation levels over different age groups during tumor formation. To overcome the limitation that DMRs determined by MCIp rely on relative changes in enrichment, nowadays the analysis of methylation dynamics is facilitated by quantitative approaches like reduced representation bisulfite sequencing (RRBS) or even more detailed with whole genome bisulfite sequencing.

5.1.2 The role of C3(1) promoter DMRs in the mammary gland during pregnancy and tumorigenesis

Among the list of candidates with promoter DMRs, we confirmed a link between DNA methylation and gene expression of five genes, *Gsn, Igfbp6, Elf3*, *Pkp3*, and *Cldn4*. *ELF3* is an important factor for the differentiation of epithelial cells [173]. *CLDN4* and *PKP3* assist in the formation of cell-cell adhesion structures such as tight junctions or desmosomes [174, 175]. *GSN* regulates actin polymerization thereby controlling the mobility of cells [176] and *IGFBP6* regulates insulin-like growth factor (IGF) functions in particular by inhibiting anti-apoptotic functions of IGF2 [177].

Epigenetic regulation of expression of these genes was further supported by data for the MMTV-PyMT breast cancer mouse model [163]. After treatment of the MMTV-PyMT with the DNMT inhibitor zebularine, two of the genes that were hypermethylated in the C3(1) model, *Gsn* and *Igfbp6*, were upregulated upon inhibitor treatment. Differential methylation also led to differential gene expression in different mammary cell populations or during pregnancy [178, 179]. *Igfbp6* was higher methylated in luminal progenitor and differentiated cells than in

a cell fraction containing MaSCs, myoepithelial and basal cells, while *Cldn4* was hypomethylated in the luminal cells compartment. DMRs were inversely associated with expression for both genes, and *Cldn4* also showed striking upregulation after demethylation treatment in both cell types. Hypomethylation of *Pkp3* was observed in the mammary stem cell population following pregnancy [178] and for *Elf3*, pregnancy-associated hypomethylation was observed for both basal and luminal cell populations [179]. These data and our results indicate that the candidates play an important role in the mammary gland either during tumorigenesis or pregnancy. The genes were also shown to be differentially expressed in human cancers including breast cancer [180-184].

We found additional DMRs co-occuring between the C3(1) model and different mammary gland cell populations or during pregnancy [178]. By evaluating the DNA methylation levels of cell type specific markers, we could deduce the cell type composition of tumors. For example, mammary gland luminal cells are not methylated at the luminal cell marker *Krt8* or *Elf5*, and basal cells are hypomethylated at the basal cell marker *Krt5* [179, 185]. The progressive hypomethylation of *Krt8* and *Elf5* in the C3(1) tumors in comparison to the constant levels of *Krt5* methylation at the different tumor stages suggests an expansion of luminal cells compared to the WT mammary gland. Besides, the changes of *Igfbp6* and *Cldn4* methylation resemble a luminal cell-like pattern and further support a high proportion of luminal cells in tumors. Evaluating potential common characteristics of those regions might help to uncover how physiological processes are highjacked during carcinogenesis.

In addition, comparing the methylation data from the C3(1) model with RRBS data for tumorspheres generated from the MMTV-NeuTg breast cancer model, we found a number of about 400 concurrent DMRs in both models [162]. Since the two models represent two different molecular subtypes, namely basal-like and HER2-enriched, these overlapping regions might reflect a common breast cancer signature. The majority of DMRs in either study was unique to the models, thus supporting the differences between the molecular subtypes of breast cancer in the models also at the level of methylation.

5.1.3 Common DMRs between the C3(1) model and human breast cancer patients

While the similarity of GEMMs to human cancer types is generally established by histological and gene expression analyses, the comparability of epigenetic modifications such as DNA methylation cannot be taken for granted. Diede *et al*. evaluated mouse models of medulloblastoma, Burkitts-lymphoma and also breast cancer and did not find consistent

hypermethylation at CGIs between mouse and human [186]. In contrast, Sonnet *et al*. [166] confirmed promoter hypermethylation in a mouse model of AML for ≈40-45% of those genes with known promoter hypermethylation in human AML or MDS. Furthermore, they discovered nine novel hypermethylated targets in the mouse, which they validated in human sample sets of AML and MDS. In the $APC^{+/Min}$ model of intestinal adenoma, Grimm *et al*. [168] identified a conserved set of promoter DMRs (48 concordantly hyper- and 52 hypomethylated promoters in both species) that they validated in human colon carcinoma. Therefore, depending on the specific model, epigenetic variation might be consistent with the human situation or not, and individual examination of the models is required.

Using data of the C3(1) mouse model we could confirm that a set of promoter DMRs distinguishes the human basal-like subtype from the other Pam50 subtypes. Although we could confirm the resemblance between C3(1) and human breast cancer DMRs at the level of individual DMRs in about 25% of promoters, the reliable assignment to one specific subtype was not possible, partly due to technical limitations of our data. MCIp in mouse was performed on bulk tissue and not on sorted cells, and the cell type composition in tumors is different to that of normal tissue. Due to the nature of MCIp, we might enrich for such changes that reflect differences between cell types rather than between methylomes in tumor and normal epithelial cells. Especially adipose tissue in the breast displayed high differences to breast epithelial cells. For the analysis of 450k methylome data, algorithms were established that can compensate for differences in the cell content by comparison with reference methylomes of sorted cell populations [187]. Applying this algorithm to the TCGA normal control samples, we revealed also considerable heterogeneity in cell type composition of normal control samples (data not shown). While some samples mainly consisted of fat tissue and apparently contained hardly any epithelial cells, others were mainly composed of epithelial cells and stromal mammary fibroblasts. Consequently, some hypomethylated mouse DMRs might actually reflect tissue composition differences and thus might not be detectable as DMR in basal-like samples, since this subtype is already lowly methylated and similar to normal controls. This would hinder a correct assignment of the model to a human breast cancer subtype. Overall luminal subgroups were already more variable in methylation and would thus be more likely to overlap with the mouse DMRs. Using quantitative DNA methylation data for the C3(1) tumors would be a way to overcome this limitation and allow direct comparison of the mouse methylome with that of human breast cancer subtypes.

Several recent studies compared methylation heterogeneity between common breast cancer and detected similar methylation subtypes as well as novel subgroups. In 2010, Holm *et al*. [78] confirmed three of the molecular breast cancer subtypes (Luminal A, Luminal B, Basal-

like) with variations in DNA methylation. Another study discovered specific methylation subtypes between luminal B and basal-like patients [188], while Stirzaker and colleagues even described additional methylation-based subgroups within the group of triple negative breast cancers that are linked to different prognosis [189, 190]. Actually, the definition of methylation subgroups might be influenced by the composition of the cohorts, as even more specific subgroups were recently defined, including a new cluster with HER2-enriched specific DNA methylation [191].

Besides, the genomic location also influenced whether a region would gain or loose methylation [191]. While hypermethylated sites were mostly found in CGIs and promoters, hypomethylation occurred at the CGI shore or at the open sea, which describes regions without enriched CpG content and often in intergenic areas [192]. Since this genomic distribution of methylation changes was also true for methylation subtypes assigned by Holm *et al*. [191], this result has also implications for a correct assignment of our C3(1) mouse model to a subtype. So far, we compared only promoter regions, but the basal-like samples as well as the C3(1) were mostly hypomethylated, which according to Holm *et al*. happens at non-promoter regions. Consequently, a suitable approach for subtype comparisons would require the inclusion of intergenic and distal regulatory regions, which are less conserved among species. In summary, similar to the stratification of the subtypes on the basis of histology or gene expression, further studies and additional cohorts will be required to define the robustness of subtype markers based on DNA methylation.

5.1.4 Defining the cell of origin for the C3(1) model

Over the last years increasing effort has been invested in the identification of the cell of origin of various cancer types. This revealed the concept that different cancer subtypes might arise from different cell populations [193]. Oakes *et al.* described that DNA methylation signatures among chronic lymphocytic leukemia patients closely resembled the different methylation patterns observed during stages of B-cell differentiation [194]. Thus, we evaluated our data to delineate the potential cell of tumor origin in the C3(1) model. We detected enrichment for genes with differential expression in MaSCs. Knockout of DNMT1 in mammary glands ascertained the importance of DNA methylation for the establishment of the MaSC population [162]. DNMT1 knockout also reduced tumor numbers and size in the C3(1) and the MMTV-NeuTg breast cancer models, which underlines the importance of MaSCs and DNA methylation for tumor formation. Intact mammary gland architecture seemed to be a prerequisite for full tumor development in these models, but was already

disturbed by the knockout in DNMT1$^{-/-}$ animals. Therefore, mammary glands from the knockout might not be equally affected by the transgene as WT glands, arguing against the MaSCs as the cell of origin. Transgene expression in the C3(1) model started around puberty at the age of 3-4 weeks [85, 86]. Lineage tracing experiments unraveled that different types of progenitor cells are required to control mammary gland development depending on the respective developmental stage [195]. The enrichment of hypomethylated genes with importance for pubertal mammary gland formation further argues that tumorigenesis in the C3(1) model involves deregulation during the pubertal differentiation process. Since at this time period differentiation of the terminal end buds is driven by luminal progenitors, this supports the hypothesis that this cell type has a role in C3(1) tumor formation.

We detected several candidate genes, for which methylation changes in the C3(1) model resembled the pattern observed in luminal cells or luminal progenitors and suggested these cells as candidates for tumor origin. This is supported by the fact that the expression of SV40T is detected specifically in luminal epithelial cells [85, 86]. A categorization of this model as basal-like breast cancer, which is called ‘basal’ due to is expression profile resembling that of basal cells, seems to be inappropriate. Nevertheless, evidence supporting the assignment of the C3(1) model to the basal-like subtype was presented by studies that compared gene expression profiles of different mammary cell populations with those of the molecular breast cancer subtypes [196-199]. These results suggested ER-negative luminal progenitors as the cell of origin of basal-like breast cancer because the molecular profile of these progenitors resembled this subtype most closely [196, 197]. Furthermore, transformation of different mammary gland cell populations with PyMT specifically produced basal-like tumors when expressed in the luminal progenitors, but not equally exclusive from other cell populations [198]. While the cell of origin for other molecular subtypes is still unclear, luminal progenitors are the most likely cell of origin for basal-like breast cancer and therefore, this subtype should be better referred to as ‘luminal progenitor type’ [196]. According to the microarray gene expression data [90], two prominent markers of luminal progenitors, *Itga6* and *Epcam*, were also upregulated in C3(1) tumors, which suggested expansion of this cell population. Since we also frequently saw DMRs reflecting the pattern of luminal cells or pregnancy-induced changes [178], these observations are in line with a basal-like phenotype of the C3(1) model at the level of DNA methylation.

5.2 A map of chromatin states in the C3(1) mouse model

The influence of epigenetic marks on the regulation of gene expression is not restricted to changes in DNA methylation, but is accompanied by alterations in the chromatin landscape.

When characterizing dynamic methylation changes during carcinogenesis in the C3(1) mouse model, we observed the majority of recurrent DMRs, especially hypomethylated regions, to be located outside of known promoter regions, CGIs or 5'UTRs that would confer a direct regulatory influence on gene expression. Unlike these TSS-associated features, distal regulatory elements like enhancers do not have a fixed and tissue universal position in the genome and required tissue-specific definition and characterization. To that end, we generated a chromatin map with histone modifications for C3(1) tumor samples.

5.2.1 Defining a hidden Markov model of chromatin states in the C3(1) mouse model

Using ChIP-seq data for four histone marks, H3K4me1, H3K4me3, H3K27ac, and H3K27me3, we defined a hidden Markov model of chromatin states using the ChromHMM algorithm [37]. The capability of a model to annotate functions to chromatin states largely depends on the number and the diversity of marks. Our analyses resulted in the definition of an 11-state ChromHMM model, which could define several variations of promoter and enhancer regions, but also of repressed states. Thus, it resembled other chromatin state models that were based on the same histone marks [35, 38, 200-202]. Among all changes between chromatin states observed for the C3(1) model, states for strong enhancers diverged most often between tumors and MECs, followed by active and inactive promoters. Since enhancer states previously were shown to occupy important tissue-specific distal regulatory sites [35, 121, 203], we expected such a process to occur during tumor formation. Ernst *et al*. also described some tissue specificity of promoter states, but hypothesized that the major promoter activity is required to perform general functions across many cell types instead of distinct cell type-specific roles [35].

Since silencing of genes, in particular of tumor suppressors, is a common process in cancer, we expected the loss of active and gain in silent promoter states that we observed at the TSS of RefSeq promoters or CGIs. Thus, the ChromHMM model for the chromatin data followed the expected patterns regarding tissue-specific occurrence and changes between tumor and normal mammary epithelial cells. Further diversity of the ChromHMM model could have been achieved by measuring additional marks, like i) H3K36me3, which marks

transcribed genes over the complete length of the transcript [204], ii) CTCF binding, which is an organizer of chromatin structure and performs insulator functions [205], or iii) DNase hypersensitivity sites to define regions of open chromatin at different regulatory sites including promoters, enhancers and insulators [206, 207]. Since we could assign the main functional chromatin states and observed transitions between stages in tumors and MECs, the 11-state model deduced from the four histone modifications was integrated with changes in DNA methylation during carcinogenesis.

5.2.2 Tissue-specific chromatin states and DNA methylation at promoters

A focus on chromatin states that we found enriched specifically in MECs or in tumors facilitated the bioinformatic comparison with alterations in DNA methylation. Combining both mechanisms could help to better estimate the effect on gene expression than by using information on DNA methylation alone.

We hardly observed a gain in methylation at MEC-specific active promoters during tumor formation that could explain silencing of the active state. However, we detected hypermethylation at MEC-specific inactive promoter states at known targets of the PRC2 complex. This switch from polycomb-mediated repression to DNA methylation at promoters is often described in cancer and is important to silence genes in differentiated cells, which are functionally important only during development [80, 168, 191, 208]. It was also shown that altered DNA methylation levels modified the deposition of repressive histone modifications in particular of H3K27me3 [208-210]. Since DNA methylation and H3K27me3 are often mutually exclusive, especially at CGIs [211], hypermethylation at developmental polycomb targets might release the PRC2 machinery for silencing of other regions and contribute to silencing of tumor suppressor genes.

We also observed a high proportion of tumor-specific active promoter sites in intergenic regions and exons or introns. A possible explanation might be the aberrant activation of alternative or even cryptic promoters during carcinogenesis as it was seen in gastric carcinoma [212]. Such cryptic promoters, which have not been defined in the RefSeq gene annotation will result in an altered transcriptome or in altered patterns of isoform expression. This is not limited to coding genes, but lncRNAs have also been frequently identified in intergenic regions, often in combination with tissue-specific and relatively low levels of expression. Bogu *et al*. [201] combined chromatin maps with RNA-seq data of eight murine tissues and ES cells and defined 2803 novel lncRNAs, several of which were associated with promoter or enhancer chromatin states. 178 of these 2803 novel lncRNAs overlapped in

their promoter region (±1kb of the TSS) with tumor-specific active promoter states defined in our model, suggesting expression of these noncoding transcripts in C3(1) tumors. Bogu and colleagues validated many (up to 85%) of the previously identified lncRNAs. We also identified several candidates with differential methylation in the C3(1) model (see Chapter 4.3). Considering the high tissue specificity of lncRNAs, RNA-seq of C3(1) mammary gland tumors should uncover expression of additional lncRNAs with differential chromatin states and DNA methylation.

5.2.3 Tissue-specific methylation changes at enhancer regions

The highest number of epigenetic alterations both at the level of chromatin states and DNA methylation occurred at enhancer states for which we observed both hyper- and hypomethylation, depending on whether we examined MEC- or tumor-specific states. Gene set enrichment analysis using the microarray expression dataset also confirmed a link of these differentially methylated enhancers with a significant effect on expression of the closest genes. Enhancer hypermethylation was associated with gene silencing and hypomethylation was linked to increased gene expression levels. Enhancer methylation was also frequently described to follow tissue- and cell-type specific patterns similar to changes in histone marks [156, 213, 214]. Methylation at these regions sometimes correlated even better with gene expression than promoter methylation, especially in case of unmethylated promoters with constantly open chromatin [214]. Such enhancer DMRs are not limited to determine physiological differences between tissues, but also appear as deregulated between tumors and their respective normal tissue [214-216]. Recently, Heyn *et al*. [217] described tumor-specific DMRs at so called super enhancers that are especially large enhancer clusters with an important role in regulating the transcriptional cell identity and that are marked by high levels of H3K27ac and H3K4me1. These findings are concurrent with our observations that normal tissue-specific enhancers (MEC-specific) gained methylation and that tumor-specific enhancers coincided with hypomethylation in cancer. Since super enhancers also mark so called ‘master transcription factors’ that control cell identity, it might be possible to discover similar factors for tumorigenesis in the C3(1) model. So far we did not distinguish in our analysis between ‘normal’ enhancers and super-enhancers, which might be required to separate master regulators from supporting transcription factors.

5.2.4 Transcription factors with potential involvement in C3(1) tumorigenesis

Since enhancer regions generally are enriched for transcription factor binding sites [35] and transcription factor binding was even shown to modulate DNA methylation levels at these distal regulatory regions [218], we performed a transcription factor motif enrichment analysis at the differentially methylated enhancer regions. These data suggested accumulation of several candidate transcription factors among which *Runx1, Egr1, Etv4,* and *Hoxa9* were differentially expressed in tumors compared to WT mammary glands concordantly with changes in the epigenetic patterns at enhancers. Since these factors themselves were associated with DMRs and altered chromatin states during carcinogenesis, they might also be epigenetically regulated. Detailed studies of the three-dimensional chromatin architecture will be needed to discover the exact mechanisms of regulation of these factors, especially to prove interaction between the potential enhancers and their promoters.

Enrichment of transcription factor motifs was generally quite low, probably due to a relatively low number of regions used for the enrichment analysis. Moreover, the list was missing important regulators for mammary gland formation during puberty or pregnancy, like *Stat5a* or *Elf5* [131, 179, 219]. Neither of these factors was differentially expressed, which is in line with data for the motif search. However, we found enrichment for a motif similar to that of ETS-class transcription factors. One of these, *Etv4* (also known as *Pea3*), has been shown to be an important regulator of mammary gland development by controlling the differentiation of multipotent progenitor cells [220, 221]. *ETV4* is also overexpressed in triple negative breast cancer patients associated with higher risk of metastasis formation [222]. This is also true for transgenic mouse models of breast cancer. In the MMTV-*neu* model, expression of a dominant-negative Etv4 protein reduced tumor number and size [223, 224].

RUNX1 was originally discovered as a tumor suppressor in acute myeloid leukemia and is known for its role in hematopoiesis [225]. More recently it was discovered to be important in breast carcinogenesis with yet undetermined consequences. Mutation analysis supported tumor suppressor function, while a different study linked high expression of *RUNX1* to poor survival, especially of triple negative breast cancer patients, and suggested an oncogenic function [226]. In addition, *RUNX1* was shown to influence the differentiation of bipotent mammary progenitor cells into mature ductal and alveolar structures [227].

With these two factors, we identified two epigenetically regulated transcription factors that are important for the differentiation of progenitor cells. Overexpression of these factors might be an important driving mechanism for tumorigenesis in the C3(1) mouse model.

The two other factors, *Egr1* and *Hoxa9,* were also differentially expressed in human breast cancer and could be linked to silencing of *BRCA1*. Binding to the *BRCA1* promoter by HOXA9 induces *BRCA1* expression; HOXA9 is downregulated in human breast cancer in a correlated fashion with *BRCA1* [228]. *BRCA1* regulation by EGR1 is more indirect at the posttranscriptional level, since EGR1 induces the oncogenic *mir20b*, which was shown to target *BRCA1* [229]. Consistently, the murine homologue of *mir20b* is upregulated in tumors derived from the C3(1) model [92]. The *BRCA1* promoter also has a binding site for ETV4, although this region could also be bound by other ETS-class transcription factors, which have similar recognition motifs. It was shown that ETS2 can form a repressor complex with the SWI/SNF chromatin remodeling complex to silence the *BRCA1* promoter [230]. Since triple negative and basal-like breast cancers often display mutations or DNA methylation-mediated silencing of BRCA1, these transcription factor-based regulation mechanisms are in line with the classification of the C3(1) mice as a model for basal-like breast cancer. Surprisingly though, *Brca1* was reported as significantly upregulated in C3(1) tumor samples [90], which would disagree with a regulatory influence of *Egr1* and *Hoxa9* in the C3(1) model. Nevertheless, additional evidence suggests that *BRCA1* upregulation can also be linked to aggressiveness in breast cancer samples from the triple-negative or HER2-positive type [231]. This upregulated BRCA1 form is actually a C-terminally truncated splice alternative, named BRCA1-IRIS [232], whose mRNA expression levels are stabilized by the loss of full length BRCA1 [233]. The mouse microarray probe for *Brca1* detects the 3'UTR, which would no longer be present in the BRCA1-IRIS homologue; however, other splice variants could exist in the mouse that contain the 3'UTR and act similar to BRCA1-IRIS. Additional investigations are required to answer this open question. Also, it needs to be further clarified whether these transcription factors are transmitters of the tumorigenic process in the C3(1) model.

5.2.5 The role of SV40T and p53 in the C3(1) model

When examining hypomethylated regions that co-occurred with tumor-specific active promoters and enhancers, we observed strong overlap of the associated genes with targets of p53 and RB1 that were downregulated after knockdown or upregulated after overexpression of p53 and RB1. We even detected the p53 binding motif enriched in the hypomethylated tumor-specific enhancer regions (Figure 4-27).

In the C3(1) model, p53 and Rb1 are functionally inactivated by binding of SV40T, but the potential targets that we identified as hypomethylated were mostly upregulated in tumors. At

a first glance this inverse correlation seems contradictive to the GSEA results. However, the functional inactivation of p53 and Rb is not necessarily equivalent to knockdown or overexpression of a single gene, as SV40T has been shown to interact with many proteins to induce transformation of cells. Hermannstädter and colleagues [234] demonstrated that SV40-mediated transformation of cells was even more effective in p53 positive, compared to p53 negative cells, since this enabled the recruitment of additional proteins including the acetyltransferase p300. Binding of a complex of SV40T and p53 to the promoter of *IGF1* was also described as necessary to induce growth-promoting activity [235]. Wildtype p53 is generally relatively unstable in the absence of cellular stress due to an auto-regulatory loop that controls its degradation by ubiquitinylation, but it is stabilized by binding of SV40T [133, 236]. Increased stability of p53 is often observed in the context of mutant p53, which is similar to the effect of posttranslational modifications that influence protein stability in the physiological context (reviewed in [237]). Mutants of p53 have even been shown to interact with other proteins such as transcription factors like ETS2 to direct DNA-binding activity to new areas [238]. Actually, the binding motifs that we discovered for the transcription factors Runx1 and Etv4 were identified by the *de novo* motif algorithm of HOMER [106] and displayed slight differences to their canonical binding motifs. Since p53 was shown to be redirected to novel proximal and distal regulatory regions upon mutation and can bind other factors, also in combination with SV40T, this could explain the altered target sequence of Runx1 and Etv4 in the C3(1) model. A study of the Wap-Tag mouse model [239] described that mammary tumors expressing SV40T even selected cells for occurrence of such mutations. Therefore, it is likely that the transformation by SV40T is more complex than loss of p53 activity, and additional factors have to be considered. Additional experiments have to clarify whether C3(1) tumors possess mutated p53 or whether SV40T-bound p53 complexes are recruited to sites of hypomethylation and activated transcription.

5.3 Differentially methylated lncRNAs in the C3(1) mouse model

The third part of this work aimed to identify lncRNAs epigenetically regulated during breast carcinogenesis in the C3(1) SV40 Tag mouse model [85, 90]. We found 65 known RefSeq lncRNAs with differentially methylated promoters. Since we were also interested how epigenetic regulation of lncRNAs might influence expression of protein-coding genes in their vicinity, we selected candidates with protein-coding neighbors and detected 21 such pairs.

5.3.1 LncRNAs and neighboring coding RNAs with differential methylation

A recent study by Li *et al*. followed a similar approach in human breast cancer [240]. The authors identified several hundred differentially expressed ncRNAs (including miRNAs) with aberrantly methylated promoters. These ncRNAs had high diagnostic potential and were involved in several pathways dysregulated in human breast cancer. Five of the differentially methylated lncRNAs with protein-coding genes in close vicinity that we identified were also reported as differentially methylated in the human study, namely *EVX1-AS*, *MAGI2-AS*, *FOXD2-AS*, *FENDRR*, and *HOXA11-AS* [240], emphasizing the relevance of the mouse model for human breast cancer. The characterization of the lncRNAs thereby follows the previous data of histologic, transcriptomic and miRNA profiling, which classified the C3(1) as model of the more aggressive luminal B and basal-like subtypes of breast cancer [90, 92, 241].

Our initial genome-wide screen identified over 6500 DMRs in the age groups of 20-24w that were overlapped with about 3600 murine RefSeq lncRNA promoters. In total, less than 2% of the lncRNA promoters were described as differentially methylated. This relatively low fraction might be attributed to the fact that MCIp is an enrichment-based technique and only detects differential methylation in regions with high CpG density and homogeneous methylation patterns [242]. Because of these technical restrictions, additional epigenetically regulated lncRNAs might have escaped detection as we mentioned before.

Gene set enrichment analyses [117, 243] associated the genes near hypermethylated lncRNA with epigenetic silencing mediated by the Polycomb repressive complexes. These genes included several homeobox genes, which represent an essential family of developmental regulators with important influence on cell growth and differentiation [244]. Hypermethylation of homeobox genes was identified as an early epigenetic event in human breast cancer [81, 245]. Gal-yam *et al*. postulated that Polycomb repression and DNA hypermethylation act in parallel to reduce the regulatory plasticity of these key regulatory genes [80]. Our study now adds epigenetic regulation of lncRNAs located close to these genes as an additional layer of complexity.

Although we detected about 3-fold more hypo- than hypermethylated regions, these overlapped with fewer lncRNA promoters than the hypermethylated regions. Since our hypomethylated regions are enriched for enhancers that are often located in intergenic regions distal of gene promoters, the supposed lack of hypomethylated lncRNA promoters can be explained. Nine of the hypomethylated lncRNAs were associated with protein-coding genes in their vicinity, and the functions of these genes were more diverse than those

identified in the proximity of hypermethylated lncRNAs. Three of the genes were reported as downregulated in metastases from malignant melanoma compared to primary tumors (*Esrp2, Palmd, Lsr*) [143]. In C3(1) tumors vs. normal mammary tissue, *Lsr* and *Esrp2* were significantly upregulated. Expression of *LSR* (Lipolysis Stimulated Lipoprotein Receptor) was described to correlate with ERα expression in human breast cancer, and levels were reduced in patient samples with lymph node invasion and distant metastasis [246]. *ESRP2* and its related isoform *ESRP1* were found to be specifically expressed in epithelial cells where they play a crucial role in enforcing an epithelial-specific alternative splicing program [123, 146]. Loss of these splicing factors led to epithelial-to-mesenchymal transition [144, 145, 247], which is an important process involved in development, tumor progression, malignant transformation, and metastasis formation [248].

5.3.2 Regulation of *Esrp2/Esrp2-as* expression

We verified coordinate differential expression of *Esrp2* and *Esrp2-as* between tumor and normal mammary gland tissue of the C3(1) model and in various cell lines inversely correlating with DNA methylation. Demethylation experiments as well as luciferase reporter assays in cell lines demonstrate that coordinate expression of *Esrp2* and *Esrp2-as* is regulated by methylation of a putative enhancer region proximal to a bidirectional promoter. The promoter itself is lowly methylated in both tumor tissue and normal mammary glands, as well as in other murine tissues for which WGBS data are publicly available. This is consistent with the observation that enhancer methylation often correlates even better with gene expression than promoter methylation [214, 249]. Fragment E4 (corresponding to amplicon A1) had enhancer activity in luciferase reporter assays with a minimal promoter, but not in combination with the *Esrp2*-P1 promoter region. This might be due to the fact that luciferase vectors lack three-dimensional chromatin organization. Moreover, we cannot exclude the possibility that under endogenous conditions alternative enhancer regions interact with the bidirectional *Esrp2* promoter, for example fragments E1-E3 that overlap the CGI shore region and are also differentially methylated between tumor and normal breast tissue as well as in murine tissues with high vs. low *Esrp2* expression. Chromosome conformation capture techniques, such as 3C or 4C-seq, might be better suited to delineate these enhancer-promoter interactions [250].

5.3.3 Potential function of *Esrp2-as*

The biological function of *Esrp2-as* still remains unclear. Knockdown or overexpression experiments of *Esrp2-as* so far do not support a role in regulating the mRNA expression of *Esrp2*, as has been described for other lncRNAs [54, 55], but protein levels of *Esrp2* still might be affected. Alternatively, the function of *Esrp2-as* might lie in the act of transcription rather than the transcript per se. lncRNA transcription can stabilize binding of transcription factors as demonstrated by transcription-mediated YY1 binding to the *ARID1A* promoter [251]. Also, lncRNA transcription in antisense might render an open chromatin conformation as seen in the activation of the HOXA cluster [252]. Such a mechanism might explain why expression of *Esrp2-as* from an exogenous plasmid did not induce changes in *Esrp2* mRNA expression. Besides, lncRNAs not always influence expression of the closest protein-coding gene, as described for *RB1* and *ncRNA-RB1* [253]. Although *ncRNA-RB1* acted via a target protein different from *RB1*, both fed into the same tumor-inhibitory pathway, thus representing two arms of one common regulatory network. For *Esrp2-as*, we would expect a similar function as for *Esrp2*. RNA-seq of cells with knockdown or overexpression of *Esrp2-as* will identify possible alternative targets.

Consistent with our observations in the mouse model, *ESRP1* and *ESRP2* were found to be upregulated during human oral squamous cell carcinogenesis [254]. Interestingly, the authors observed plasticity in expression, with low expression levels at the invasive tumor front. Since we analyzed homogenates of whole mammary glands and bulk tumor tissue, similar spatial differences in expression of *Esrp2* and *Esrp2-as* might have escaped our observation. Using public data from TCGA we confirmed that *ESRP2* is also upregulated in human breast cancer and that DNA methylation inversely correlated with expression. *ESRP2* upregulation was associated with an up to 2-fold enhanced hazard ratio for tumor relapse within the follow-up period. Future detailed analyses including distant metastases have to clarify whether *ESRP2* shows a plastic expression profile in human breast cancer similar to oral cancer, or whether alternative functions in addition to regulation of epithelial splicing are related to poor prognosis.

6. Conclusion and outlook

In this work, we generated comprehensive genome-wide profiles of epigenetic modifications in the C3(1) mouse model of breast cancer, which included DNA methylation, chromatin state maps defined by histone modifications, and epigenetic regulation of lncRNAs. The changes in the epigenetic landscape were associated with genes and genomic locations that suggest luminal progenitor cells as the cells of origin of tumorigenesis in the C3(1) model. Comparison with TCGA human breast cancer samples revealed general similarity of the C3(1) tumors at the level of DNA methylation for regions that might be differentially methylated in all breast cancer subtypes. Due to the resemblance of the basal-like methylome to that of normal breast tissue, which produced only low numbers of DMRs, the approach to use DMRs for the assignment of the C3(1) model to a specific breast cancer subtype was not feasible. This limitation could be overcome by using quantitative methylation data for the mouse model. However, the fact that we observed higher frequency of hypomethylation in the C3(1) model, especially at intergenic and distal regulatory regions, still indicates resemblance to human basal-like breast cancer. So far, 450k arrays cover mainly promoters and CpG islands, which are often hypermethylated in cancer. The 450k platform was recently upgraded to contain about 850.000 CpG probes, which will cover also more enhancer regions, and this upgrade might facilitate the comparison to the mouse dataset at distal genomic regions that tend to display higher degrees of differential methylation than promoters.

Combination of chromatin state maps with DNA methylation levels revealed tumor-specific enhancers and several transcription factors with potential binding sites in these enhancers. Since some of these factors were differentially expressed between tumors and normal mammary glands, they might be involved in important regulatory processes for tumorigenesis in the C3(1) model. Future experiments will clarify their specific role and their importance for human breast cancer. In addition, the mechanisms of how enhancer methylation regulates gene expression are still poorly understood and have to integrate the impact of the three dimensional architecture of the genome.

This comprehensive analysis of epigenetic modifications in the C3(1) mouse model can now be used as a resource to characterize novel targets for biomarkers and therapeutic approaches in breast cancer. In our screen, we confirmed several lncRNAs that have a role in regulation of gene expression or are differentially methylated in human cancer studies. This underlines the value of our data for further studies. For our analyses we selected the lncRNA *Esrp2-as,* which was hypomethylated and coordinately overexpressed with its protein-coding neighbor *Esrp2*. The biological function of the lncRNA *Esrp2-as* was not

ascertained so far, but the study of lncRNA activities is still in its early stages and continuously uncovers novel modes of actions for these transcripts. Nevertheless, we confirmed the epigenetic regulation of *Esrp2* and *Esrp2-as* expression. Although no homologous lncRNA to the murine *Esrp2-as* is annotated in the human genome to date, high levels of *Esrp2* expression in breast cancer resulted in increased risk of recurrence and metastasis. Thus, further candidates either of the protein-coding or noncoding type might be identified that improve prognosis and treatment options for breast cancer.

7. References

1. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. Genes & development. 2009;23(7):781-3. Epub 2009/04/03. doi: 10.1101/gad.1787609. PubMed PMID: 19339683; PubMed Central PMCID: PMCPmc3959995.
2. Stirzaker C, Taberlay PC, Statham AL, Clark SJ. Mining cancer methylomes: prospects and challenges. Trends in genetics : TIG. 2014;30(2):75-84. Epub 2013/12/26. doi: 10.1016/j.tig.2013.11.004. PubMed PMID: 24368016.
3. Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C, Onyango P, et al. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. Nat Genet. 2009;41(2):178-86. Epub 2009/01/20. doi: 10.1038/ng.298. PubMed PMID: 19151715; PubMed Central PMCID: PMCPMC2729128.
4. Baylin SB, Jones PA. A decade of exploring the cancer epigenome — biological and translational implications. Nat Rev Cancer. 2011;11(10):726-34.
5. Gelfman S, Cohen N, Yearim A, Ast G. DNA-methylation effect on cotranscriptional splicing is dependent on GC architecture of the exon-intron structure. Genome Res. 2013;23(5):789-99. Epub 2013/03/19. doi: 10.1101/gr.143503.112. PubMed PMID: 23502848; PubMed Central PMCID: PMCPmc3638135.
6. Laurent L, Wong E, Li G, Huynh T, Tsirigos A, Ong CT, et al. Dynamic changes in the human methylome during differentiation. Genome Res. 2010;20(3):320-31. Epub 2010/02/06. doi: 10.1101/gr.101907.109. PubMed PMID: 20133333; PubMed Central PMCID: PMCPmc2840979.
7. Jones PA. Functions of DNA methylation: islands, start sites, gene bodies and beyond. Nat Rev Genet. 2012;13(7):484-92. Epub 2012/05/30. doi: 10.1038/nrg3230. PubMed PMID: 22641018.
8. Perini G, Diolaiti D, Porro A, Della Valle G. In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. Proc Natl Acad Sci U S A. 2005;102(34):12117-22. Epub 2005/08/12. doi: 10.1073/pnas.0409097102. PubMed PMID: 16093321; PubMed Central PMCID: PMCPmc1184034.
9. Kim J, Kollhoff A, Bergmann A, Stubbs L. Methylation-sensitive binding of transcription factor YY1 to an insulator sequence within the paternally expressed imprinted gene, Peg3. Human molecular genetics. 2003;12(3):233-45. Epub 2003/01/30. PubMed PMID: 12554678.
10. Rishi V, Bhattacharya P, Chatterjee R, Rozenberg J, Zhao J, Glass K, et al. CpG methylation of half-CRE sequences creates C/EBPalpha binding sites that activate some tissue-specific genes. Proc Natl Acad Sci U S A. 2010;107(47):20311-6. Epub 2010/11/10. doi: 10.1073/pnas.1008688107. PubMed PMID: 21059933; PubMed Central PMCID: PMCPmc2996703.
11. Dhasarathy A, Wade PA. The MBD protein family-reading an epigenetic mark? Mutation research. 2008;647(1-2):39-43. Epub 2008/08/12. doi: 10.1016/j.mrfmmm.2008.07.007. PubMed PMID: 18692077; PubMed Central PMCID: PMCPmc2670759.
12. Hermann A, Goyal R, Jeltsch A. The Dnmt1 DNA-(cytosine-C5)-methyltransferase methylates DNA processively with high preference for hemimethylated target sites. The Journal of biological chemistry. 2004;279(46):48350-9. Epub 2004/09/02. doi: 10.1074/jbc.M403427200. PubMed PMID: 15339928.
13. Okano M, Bell DW, Haber DA, Li E. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell. 1999;99(3):247-57. Epub 1999/11/11. PubMed PMID: 10555141.
14. Li E, Bestor TH, Jaenisch R. Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell. 1992;69(6):915-26. Epub 1992/06/12. PubMed PMID: 1606615.

15. Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009;324(5929):930-5. Epub 2009/04/18. doi: 10.1126/science.1170116. PubMed PMID: 19372391; PubMed Central PMCID: PMCPmc2715015.
16. Ito S, Shen L, Dai Q, Wu SC, Collins LB, Swenberg JA, et al. Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science. 2011;333(6047):1300-3. Epub 2011/07/23. doi: 10.1126/science.1210597. PubMed PMID: 21778364; PubMed Central PMCID: PMCPmc3495246.
17. He YF, Li BZ, Li Z, Liu P, Wang Y, Tang Q, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011;333(6047):1303-7. Epub 2011/08/06. doi: 10.1126/science.1210944. PubMed PMID: 21817016; PubMed Central PMCID: PMCPmc3462231.
18. Huang Y, Rao A. New functions for DNA modifications by TET-JBP. Nature structural & molecular biology. 2012;19(11):1061-4. Epub 2012/11/08. doi: 10.1038/nsmb.2437. PubMed PMID: 23132381; PubMed Central PMCID: PMCPmc3858004.
19. Rodriguez-Paredes M, Esteller M. Cancer epigenetics reaches mainstream oncology. Nat Med. 2011:330-9.
20. Gnyszka A, Jastrzebski Z, Flis S. DNA methyltransferase inhibitors and their emerging role in epigenetic therapy of cancer. Anticancer research. 2013;33(8):2989-96. Epub 2013/07/31. PubMed PMID: 23898051.
21. Hamidi T, Singh AK, Chen T. Genetic alterations of DNA methylation machinery in human diseases. Epigenomics. 2015;7(2):247-65. Epub 2015/05/06. doi: 10.2217/epi.14.80. PubMed PMID: 25942534.
22. Bannister AJ, Kouzarides T. Regulation of chromatin by histone modifications. Cell research. 2011;21(3):381-95. Epub 2011/02/16. doi: 10.1038/cr.2011.22. PubMed PMID: 21321607; PubMed Central PMCID: PMCPmc3193420.
23. Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2.8 A resolution. Nature. 1997;389(6648):251-60. Epub 1997/09/26. doi: 10.1038/38444. PubMed PMID: 9305837.
24. Cohen I, Poreba E, Kamieniarz K, Schneider R. Histone modifiers in cancer: friends or foes? Genes & cancer. 2011;2(6):631-47. Epub 2011/09/24. doi: 10.1177/1947601911417176. PubMed PMID: 21941619; PubMed Central PMCID: PMCPmc3174261.
25. Ruthenburg AJ, Allis CD, Wysocka J. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. Molecular cell. 2007;25(1):15-30. Epub 2007/01/16. doi: 10.1016/j.molcel.2006.12.014. PubMed PMID: 17218268.
26. Hon GC, Hawkins RD, Ren B. Predictive chromatin signatures in the mammalian genome. Human molecular genetics. 2009;18(R2):R195-201. Epub 2009/10/08. doi: 10.1093/hmg/ddp409. PubMed PMID: 19808796; PubMed Central PMCID: PMCPmc2912651.
27. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat Genet. 2007;39(3):311-8. Epub 2007/02/06. doi: 10.1038/ng1966. PubMed PMID: 17277777.
28. Tie F, Banerjee R, Stratton CA, Prasad-Sinha J, Stepanik V, Zlobin A, et al. CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development (Cambridge, England). 2009;136(18):3131-41. Epub 2009/08/25. doi: 10.1242/dev.037127. PubMed PMID: 19700617; PubMed Central PMCID: PMCPmc2730368.
29. Creyghton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A. 2010;107(50):21931-6. Epub 2010/11/26. doi: 10.1073/pnas.1016071107. PubMed PMID: 21106759; PubMed Central PMCID: PMCPmc3003124.

30. Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, et al. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. Cell. 2006;125(2):315-26. Epub 2006/04/25. doi: 10.1016/j.cell.2006.02.041. PubMed PMID: 16630819.
31. Roudier F, Ahmed I, Berard C, Sarazin A, Mary-Huard T, Cortijo S, et al. Integrative epigenomic mapping defines four main chromatin states in Arabidopsis. The EMBO journal. 2011;30(10):1928-38. Epub 2011/04/14. doi: 10.1038/emboj.2011.103. PubMed PMID: 21487388; PubMed Central PMCID: PMCPmc3098477.
32. Baker M. Making sense of chromatin states. Nature methods. 2011;8(9):717-22. Epub 2011/09/01. doi: 10.1038/nmeth.1673. PubMed PMID: 21878916.
33. Won KJ, Chepelev I, Ren B, Wang W. Prediction of regulatory elements in mammalian genomes using chromatin signatures. BMC bioinformatics. 2008;9:547. Epub 2008/12/20. doi: 10.1186/1471-2105-9-547. PubMed PMID: 19094206; PubMed Central PMCID: PMCPmc2657164.
34. Ernst J, Kellis M. Discovery and characterization of chromatin states for systematic annotation of the human genome. Nature biotechnology. 2010;28(8):817-25. Epub 2010/07/27. doi: 10.1038/nbt.1662. PubMed PMID: 20657582; PubMed Central PMCID: PMCPmc2919626.
35. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011;473(7345):43-9. Epub 2011/03/29. doi: 10.1038/nature09906. PubMed PMID: 21441907; PubMed Central PMCID: PMCPmc3088773.
36. Hoffman MM, Buske OJ, Wang J, Weng Z, Bilmes JA, Noble WS. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. Nature methods. 2012;9(5):473-6. Epub 2012/03/20. doi: 10.1038/nmeth.1937. PubMed PMID: 22426492; PubMed Central PMCID: PMCPmc3340533.
37. Ernst J, Kellis M. ChromHMM: automating chromatin-state discovery and characterization. Nature methods. 2012;9(3):215-6. Epub 2012/03/01. doi: 10.1038/nmeth.1906. PubMed PMID: 22373907; PubMed Central PMCID: PMCPmc3577932.
38. Hoffman MM, Ernst J, Wilder SP, Kundaje A, Harris RS, Libbrecht M, et al. Integrative annotation of chromatin elements from ENCODE data. Nucleic Acids Res. 2013;41(2):827-41. Epub 2012/12/12. doi: 10.1093/nar/gks1284. PubMed PMID: 23221638; PubMed Central PMCID: PMCPmc3553955.
39. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74. Epub 2012/09/08. doi: 10.1038/nature11247. PubMed PMID: 22955616; PubMed Central PMCID: PMCPmc3439153.
40. Yue F, Cheng Y, Breschi A, Vierstra J, Wu W, Ryba T, et al. A comparative encyclopedia of DNA elements in the mouse genome. Nature. 2014;515(7527):355-64. Epub 2014/11/21. doi: 10.1038/nature13992. PubMed PMID: 25409824; PubMed Central PMCID: PMCPmc4266106.
41. Fraga MF, Ballestar E, Villar-Garea A, Boix-Chornet M, Espada J, Schotta G, et al. Loss of acetylation at Lys16 and trimethylation at Lys20 of histone H4 is a common hallmark of human cancer. Nat Genet. 2005;37(4):391-400. Epub 2005/03/15. doi: 10.1038/ng1531. PubMed PMID: 15765097.
42. Hock H. A complex Polycomb issue: the two faces of EZH2 in cancer. Genes & development. 2012;26(8):751-5. Epub 2012/04/18. doi: 10.1101/gad.191163.112. PubMed PMID: 22508723; PubMed Central PMCID: PMCPmc3337450.
43. Barneda-Zahonero B, Parra M. Histone deacetylases and cancer. Molecular oncology. 2012;6(6):579-89. Epub 2012/09/12. doi: 10.1016/j.molonc.2012.07.003. PubMed PMID: 22963873.
44. Baer C, Claus R, Plass C. Genome-wide epigenetic regulation of miRNAs in cancer. Cancer Res. 2013;73(2):473-7. Epub 2013/01/15. doi: 10.1158/0008-5472.can-12-3731. PubMed PMID: 23316035.

45. Sahu A, Singhal U, Chinnaiyan AM. Long noncoding RNAs in cancer: from function to translation. Trends in cancer. 2015;1(2):93-109. Epub 2015/12/23. doi: 10.1016/j.trecan.2015.08.010. PubMed PMID: 26693181; PubMed Central PMCID: PMCPMC4672369.
46. Guil S, Esteller M. Cis-acting noncoding RNAs: friends and foes. Nature structural & molecular biology. 2012;19(11):1068-75. Epub 2012/11/08. doi: 10.1038/nsmb.2428. PubMed PMID: 23132386.
47. Hu X, Feng Y, Zhang D, Zhao Sihai D, Hu Z, Greshock J, et al. A Functional Genomic Approach Identifies FAL1 as an Oncogenic Long Noncoding RNA that Associates with BMI1 and Represses p21 Expression in Cancer. Cancer cell. 2014;26(3):344-57. doi: http://dx.doi.org/10.1016/j.ccr.2014.07.009. PubMed PMID: 25203321; PubMed Central PMCID: PMCPMC4159613.
48. Gendrel AV, Heard E. Noncoding RNAs and epigenetic mechanisms during X-chromosome inactivation. Annual review of cell and developmental biology. 2014;30:561-80. Epub 2014/07/09. doi: 10.1146/annurev-cellbio-101512-122415. PubMed PMID: 25000994.
49. Cai X, Cullen BR. The imprinted H19 noncoding RNA is a primary microRNA precursor. RNA (New York, NY). 2007;13(3):313-6. doi: 10.1261/rna.351707.
50. Kallen AN, Zhou XB, Xu J, Qiao C, Ma J, Yan L, et al. The imprinted H19 lncRNA antagonizes let-7 microRNAs. Molecular cell. 2013;52(1):101-12. Epub 2013/09/24. doi: 10.1016/j.molcel.2013.08.027. PubMed PMID: 24055342; PubMed Central PMCID: PMCPMC3843377.
51. Tsai MC, Manor O, Wan Y, Mosammaparast N, Wang JK, Lan F, et al. Long noncoding RNA as modular scaffold of histone modification complexes. Science. 2010;329(5992):689-93. Epub 2010/07/10. doi: 10.1126/science.1192002. PubMed PMID: 20616235; PubMed Central PMCID: PMCPMC2967777.
52. Mohammad F, Mondal T, Guseva N, Pandey GK, Kanduri C. Kcnq1ot1 noncoding RNA mediates transcriptional gene silencing by interacting with Dnmt1. Development (Cambridge, England). 2010;137(15):2493-9. Epub 2010/06/25. doi: 10.1242/dev.048181. PubMed PMID: 20573698.
53. Grote P, Wittler L, Hendrix D, Koch F, Wahrisch S, Beisaw A, et al. The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. Developmental cell. 2013;24(2):206-14. Epub 2013/02/02. doi: 10.1016/j.devcel.2012.12.012. PubMed PMID: 23369715; PubMed Central PMCID: PMCPMC4149175.
54. Arab K, Park YJ, Lindroth AM, Schafer A, Oakes C, Weichenhan D, et al. Long noncoding RNA TARID directs demethylation and activation of the tumor suppressor TCF21 via GADD45A. Molecular cell. 2014;55(4):604-14. Epub 2014/08/05. doi: 10.1016/j.molcel.2014.06.031. PubMed PMID: 25087872.
55. Grote P, Herrmann BG. The long non-coding RNA Fendrr links epigenetic control mechanisms to gene regulatory networks in mammalian embryogenesis. RNA biology. 2013;10(10):1579-85. Epub 2013/09/17. doi: 10.4161/rna.26165. PubMed PMID: 24036695; PubMed Central PMCID: PMCPMC3866236.
56. Di Ruscio A, Ebralidze AK, Benoukraf T, Amabile G, Goff LA, Terragni J, et al. DNMT1-interacting RNAs block gene-specific DNA methylation. Nature. 2013;503(7476):371-6. Epub 2013/10/11. doi: 10.1038/nature12598. PubMed PMID: 24107992; PubMed Central PMCID: PMCPMC3870304.
57. Lee C, Kikyo N. Strategies to identify long noncoding RNAs involved in gene regulation. Cell & bioscience. 2012;2(1):37. Epub 2012/11/07. doi: 10.1186/2045-3701-2-37. PubMed PMID: 23126680; PubMed Central PMCID: PMCPMC3499186.
58. Qin D, Xu C. Study strategies for long non-coding RNAs and their roles in regulating gene expression. Cellular & molecular biology letters. 2015;20(2):323-49. Epub 2015/07/24. doi: 10.1515/cmble-2015-0021. PubMed PMID: 26204411.
59. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals.

Nature. 2009;458(7235):223-7. Epub 2009/02/03. doi: 10.1038/nature07672. PubMed PMID: 19182780; PubMed Central PMCID: PMCPMC2754849.
60. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C, Rebelo M, Parkin DM, Forman D, Bray, F. Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 http://globocan.iarc.fr,: Lyon, France: International Agency for Research on Cancer; 2013; 2012 [cited 2015 16/01/].
61. Bertos NR, Park M. Breast cancer — one term, many entities? The Journal of Clinical Investigation. 2011;121(10):3789-96.
62. Shore AN, Rosen JM. Regulation of mammary epithelial cell homeostasis by lncRNAs. The international journal of biochemistry & cell biology. 2014;54:318-30. Epub 2014/04/01. doi: 10.1016/j.biocel.2014.03.012. PubMed PMID: 24680897; PubMed Central PMCID: PMCPmc4341896.
63. McDonnell DP, Wardell SE. The molecular mechanisms underlying the pharmacological actions of ER modulators: implications for new drug discovery in breast cancer. Current opinion in pharmacology. 2010;10(6):620-8. Epub 2010/10/12. doi: 10.1016/j.coph.2010.09.007. PubMed PMID: 20926342; PubMed Central PMCID: PMCPmc2981619.
64. Matsumoto A, Jinno H, Ando T, Fujii T, Nakamura T, Saito J, et al. Biological markers of invasive breast cancer. Japanese journal of clinical oncology. 2016;46(2):99-105. Epub 2015/10/22. doi: 10.1093/jjco/hyv153. PubMed PMID: 26486826.
65. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52. Epub 2000/08/30. doi: 10.1038/35021093. PubMed PMID: 10963602.
66. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, et al. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A. 2001;98(19):10869-74. Epub 2001/09/13. doi: 10.1073/pnas.191367098. PubMed PMID: 11553815; PubMed Central PMCID: PMCPmc58566.
67. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, Nobel A, et al. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A. 2003;100(14):8418-23. Epub 2003/06/28. doi: 10.1073/pnas.0932692100. PubMed PMID: 12829800; PubMed Central PMCID: PMCPmc166244.
68. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(15):2568-81. Epub 2008/05/20. doi: 10.1200/jco.2007.13.1748. PubMed PMID: 18487574.
69. Badve S, Dabbs DJ, Schnitt SJ, Baehner FL, Decker T, Eusebi V, et al. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2011;24(2):157-67. Epub 2010/11/16. doi: 10.1038/modpathol.2010.200. PubMed PMID: 21076464.
70. Network CGA. Comprehensive molecular portraits of human breast tumours. Nature. 2012;490(7418):61-70. Epub 2012/09/25. doi: 10.1038/nature11412. PubMed PMID: 23000897; PubMed Central PMCID: PMCPMC3465532.
71. Szyf M, Pakneshan P, Rabbani SA. DNA methylation and breast cancer. Biochemical pharmacology. 2004;68(6):1187-97. Epub 2004/08/18. doi: 10.1016/j.bcp.2004.04.030. PubMed PMID: 15313416.
72. Basse C, Arock M. The increasing roles of epigenetics in breast cancer: Implications for pathogenicity, biomarkers, prevention and treatment. International journal of cancer Journal international du cancer. 2015;137(12):2785-94. Epub 2014/11/21. doi: 10.1002/ijc.29347. PubMed PMID: 25410431.
73. Birgisdottir V, Stefansson OA, Bodvarsdottir SK, Hilmarsdottir H, Jonasson JG, Eyfjord JE. Epigenetic silencing and deletion of the BRCA1 gene in sporadic breast cancer. Breast cancer research : BCR. 2006;8(4):R38. Epub 2006/07/19. doi: 10.1186/bcr1522. PubMed PMID: 16846527; PubMed Central PMCID: PMCPmc1779478.

74. Sharma D, Saxena NK, Davidson NE, Vertino PM. Restoration of tamoxifen sensitivity in estrogen receptor-negative breast cancer cells: tamoxifen-bound reactivated ER recruits distinctive corepressor complexes. Cancer Res. 2006;66(12):6370-8. Epub 2006/06/17. doi: 10.1158/0008-5472.can-06-0402. PubMed PMID: 16778215; PubMed Central PMCID: PMCPmc2925469.
75. Yang X, Phillips DL, Ferguson AT, Nelson WG, Herman JG, Davidson NE. Synergistic activation of functional estrogen receptor (ER)-alpha by DNA methyltransferase and histone deacetylase inhibition in human ER-alpha-negative breast cancer cells. Cancer Res. 2001;61(19):7025-9. Epub 2001/10/05. PubMed PMID: 11585728.
76. Huang Y, Nayak S, Jankowitz R, Davidson NE, Oesterreich S. Epigenetics in breast cancer: what's new? Breast cancer research : BCR. 2011;13(6):225. Epub 2011/11/15. doi: 10.1186/bcr2925. PubMed PMID: 22078060; PubMed Central PMCID: PMCPmc3326545.
77. Dedeurwaerder S, Desmedt C, Calonne E, Singhal SK, Haibe-Kains B, Defrance M, et al. DNA methylation profiling reveals a predominant immune component in breast cancers. EMBO molecular medicine. 2011;3(12):726-41. Epub 2011/09/13. doi: 10.1002/emmm.201100801. PubMed PMID: 21910250; PubMed Central PMCID: PMCPmc3377115.
78. Holm K, Hegardt C, Staaf J, Vallon-Christersson J, Jonsson G, Olsson H, et al. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. Breast cancer research : BCR. 2010;12(3):R36. Epub 2010/06/23. doi: 10.1186/bcr2590. PubMed PMID: 20565864; PubMed Central PMCID: PMCPmc2917031.
79. Dalvai M, Bystricky K. The role of histone modifications and variants in regulating gene expression in breast cancer. Journal of mammary gland biology and neoplasia. 2010;15(1):19-33. Epub 2010/02/05. doi: 10.1007/s10911-010-9167-z. PubMed PMID: 20131086.
80. Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S, Zhang X, et al. Frequent switching of Polycomb repressive marks and DNA hypermethylation in the PC3 prostate cancer cell line. Proc Natl Acad Sci U S A. 2008;105(35):12979-84. Epub 2008/08/30. doi: 10.1073/pnas.0806437105. PubMed PMID: 18753622; PubMed Central PMCID: PMCPmc2529074.
81. Faryna M, Konermann C, Aulmann S, Bermejo JL, Brugger M, Diederichs S, et al. Genome-wide methylation screen in low-grade breast cancer identifies novel epigenetically altered genes as potential biomarkers for tumor diagnosis. FASEB journal : official publication of the Federation of American Societies for Experimental Biology. 2012;26(12):4937-50. Epub 2012/08/30. doi: 10.1096/fj.12-209502. PubMed PMID: 22930747.
82. O'Day E, Lal A. MicroRNAs and their target gene networks in breast cancer. Breast cancer research : BCR. 2010;12(2):201. Epub 2010/03/30. doi: 10.1186/bcr2484. PubMed PMID: 20346098; PubMed Central PMCID: PMCPmc2879559.
83. Wang G, Liu C, Deng S, Zhao Q, Li T, Qiao S, et al. Long noncoding RNAs in regulation of human breast cancer. Briefings in functional genomics. 2015. Epub 2015/11/20. doi: 10.1093/bfgp/elv049. PubMed PMID: 26582840.
84. Su X, Malouf GG, Chen Y, Zhang J, Yao H, Valero V, et al. Comprehensive analysis of long non-coding RNAs in human breast cancer clinical subtypes. Oncotarget. 2014;5(20):9864-76. Epub 2014/10/10. doi: 10.18632/oncotarget.2454. PubMed PMID: 25296969; PubMed Central PMCID: PMCPmc4259443.
85. Green JE, Shibata MA, Yoshidome K, Liu ML, Jorcyk C, Anver MR, et al. The C3(1)/SV40 T-antigen transgenic mouse model of mammary cancer: ductal epithelial cell targeting with multistage progression to carcinoma. Oncogene. 2000;19(8):1020-7. doi: 10.1038/sj.onc.1203280. PubMed PMID: WOS:000085796400007.
86. Maroulakou IG, Anver M, Garrett L, Green JE. Prostate and mammary adenocarcinoma in transgenic mice carrying a rat C3(1) simian virus 40 large tumor antigen fusion gene. Proc Natl Acad Sci U S A. 1994;91(23):11236-40. Epub 1994/11/08. PubMed PMID: 7972041; PubMed Central PMCID: PMCPMC45202.

87. Cheng J, DeCaprio JA, Fluck MM, Schaffhausen BS. Cellular transformation by Simian Virus 40 and Murine Polyoma Virus T antigens. Seminars in cancer biology. 2009;19(4):218-28. Epub 2009/06/10. doi: 10.1016/j.semcancer.2009.03.002. PubMed PMID: 19505649; PubMed Central PMCID: PMCPMC2694755.
88. Ahuja D, Saenz-Robles MT, Pipas JM. SV40 large T antigen targets multiple cellular pathways to elicit cellular transformation. Oncogene. 2005;24(52):7729-45. Epub 2005/11/22. doi: 10.1038/sj.onc.1209046. PubMed PMID: 16299533.
89. Yoshidome K, Shibata MA, Couldrey C, Korach KS, Green JE. Estrogen promotes mammary tumor development in C3(1)/SV40 large T-antigen transgenic mice: paradoxical loss of estrogen receptoralpha expression during tumor progression. Cancer Res. 2000;60(24):6901-10. Epub 2001/01/13. PubMed PMID: 11156389.
90. Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, et al. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biology. 2007;8(5). doi: 10.1186/gb-2007-8-5-r76. PubMed PMID: WOS:000246983100012.
91. Deeb KK, Michalowska AM, Yoon CY, Krummey SM, Hoenerhoff MJ, Kavanaugh C, et al. Identification of an integrated SV40 T/t-antigen cancer signature in aggressive human breast, prostate, and lung carcinomas with poor prognosis. Cancer Res. 2007;67(17):8065-80. doi: 10.1158/0008-5472.can-07-1515. PubMed PMID: WOS:000249406700022.
92. Zhu M, Yi M, Kim CH, Deng C, Li Y, Medina D, et al. Integrated miRNA and mRNA expression profiling of mouse mammary tumor models identifies miRNA signatures associated with mammary tumor lineage. Genome Biol. 2011;12(8):R77. Epub 2011/08/19. doi: 10.1186/gb-2011-12-8-r77. PubMed PMID: 21846369; PubMed Central PMCID: PMCPmc3245617.
93. Green JE, Shibata M-A, Shibata E, Moon RC, Anver MR, Kelloff G, et al. 2-Difluoromethylornithine and Dehydroepiandrosterone Inhibit Mammary Tumor Progression but not Mammary or Prostate Tumor Initiation in C3(1)/SV40 T/t-antigen Transgenic Mice. Cancer Res. 2001;61(20):7449-55.
94. Steiner JL, Davis JM, McClellan JL, Guglielmotti A, Murphy EA. Effects of the MCP-1 synthesis inhibitor bindarit on tumorigenesis and inflammatory markers in the C3(1)/SV40Tag mouse model of breast cancer. Cytokine. 2014;66(1):60-8. Epub 2014/02/20. doi: 10.1016/j.cyto.2013.12.011. PubMed PMID: 24548426; PubMed Central PMCID: PMCPMC4419732.
95. Wu K, Kim HT, Rodriquez JL, Hilsenbeck SG, Mohsin SK, Xu XC, et al. Suppression of mammary tumorigenesis in transgenic mice by the RXR-selective retinoid, LGD1069. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2002;11(5):467-74. Epub 2002/05/16. PubMed PMID: 12010861.
96. Mustafa A, Kruger WD. Suppression of tumor formation by a cyclooxygenase-2 inhibitor and a peroxisome proliferator-activated receptor gamma agonist in an in vivo mouse model of spontaneous breast cancer. Clinical Cancer Research. 2008;14(15):4935-42. doi: 10.1158/1078-0432.ccr-08-0958. PubMed PMID: WOS:000258217200036.
97. Leong H, Mathur PS, Greene GL. Inhibition of mammary tumorigenesis in the C3(1)/SV40 mouse model by green tea. Breast Cancer Research and Treatment. 2008;107(3):359-69. doi: 10.1007/s10549-007-9568-x. PubMed PMID: WOS:000252797200005.
98. Li Y, Meeran SM, Patel SN, Chen H, Hardy TM, Tollefsbol TO. Epigenetic reactivation of estrogen receptor-alpha (ERalpha) by genistein enhances hormonal therapy sensitivity in ERalpha-negative breast cancer. Molecular cancer. 2013;12:9. Epub 2013/02/06. doi: 10.1186/1476-4598-12-9. PubMed PMID: 23379261; PubMed Central PMCID: PMCPMC3577460.
99. Steiner J, Davis J, McClellan J, Enos R, Carson J, Fayad R, et al. Dose-dependent benefits of quercetin on tumorigenesis in the C3(1)/SV40Tag transgenic mouse model of breast cancer. Cancer biology & therapy. 2014;15(11):1456-67. Epub 2014/12/09. doi: 10.4161/15384047.2014.955444. PubMed PMID: 25482952.

100. Sonnet M, Baer C, Rehli M, Weichenhan D, Plass C. Enrichment of methylated DNA by methyl-CpG immunoprecipitation. Methods in molecular biology (Clifton, NJ). 2013;971:201-12. Epub 2013/01/09. doi: 10.1007/978-1-62703-269-8_11. PubMed PMID: 23296965.
101. Schilling E, Rehli M. Global, comparative analysis of tissue-specific promoter CpG methylation. Genomics. 2007;90(3):314-23. doi: 10.1016/j.ygeno.2007.04.011. PubMed PMID: WOS:000248715900004.
102. Lefrancois P, Euskirchen GM, Auerbach RK, Rozowsky J, Gibson T, Yellman CM, et al. Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. BMC genomics. 2009;10:37. Epub 2009/01/23. doi: 10.1186/1471-2164-10-37. PubMed PMID: 19159457; PubMed Central PMCID: PMCPmc2656530.
103. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009;25(14):1754-60. Epub 2009/05/20. doi: 10.1093/bioinformatics/btp324. PubMed PMID: 19451168; PubMed Central PMCID: PMCPMC2705234.
104. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics (Oxford, England). 2009;25(16):2078-9. Epub 2009/06/10. doi: 10.1093/bioinformatics/btp352. PubMed PMID: 19505943; PubMed Central PMCID: PMCPMC2723002.
105. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics (Oxford, England). 2014;30(2):284-6. Epub 2013/11/15. doi: 10.1093/bioinformatics/btt650. PubMed PMID: 24227674; PubMed Central PMCID: PMCPMC3892689.
106. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Molecular cell. 2010;38(4):576-89. Epub 2010/06/02. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PubMed Central PMCID: PMCPMC2898526.
107. Karolchik D, Hinrichs AS, Furey TS, Roskin KM, Sugnet CW, Haussler D, et al. The UCSC Table Browser data retrieval tool. Nucleic Acids Res. 2004;32(suppl 1):D493-D6. doi: 10.1093/nar/gkh103.
108. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics (Oxford, England). 2010;26(6):841-2. Epub 2010/01/30. doi: 10.1093/bioinformatics/btq033. PubMed PMID: 20110278; PubMed Central PMCID: PMCPMC2832824.
109. Ehrich M, Nelson MR, Stanssens P, Zabeau M, Liloglou T, Xinarianos G, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci U S A. 2005;102(44):15785-90. doi: 10.1073/pnas.0507816102. PubMed PMID: 16243968; PubMed Central PMCID: PMCPMC1276092.
110. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods (San Diego, Calif). 2001;25(4):402-8. Epub 2002/02/16. doi: 10.1006/meth.2001.1262. PubMed PMID: 11846609.
111. Holzer RG, MacDougall C, Cortright G, Atwood K, Green JE, Jorcyk CL. Development and Characterization of a Progressive Series of Mammary Adenocarcinoma Cell Lines Derived from the C3(1)/SV40 Large T-antigen Transgenic Mouse Model. Breast Cancer Research and Treatment. 2003;77(1):65-76. doi: 10.1023/a:1021175931177.
112. Green H, Meuth M. An established pre-adipose cell line and its differentiation in culture. Cell. 1974;3(2):127-33. Epub 1974/10/01. PubMed PMID: 4426090.
113. Morita S, Kojima T, Kitamura T. Plat-E: an efficient and stable system for transient packaging of retroviruses. Gene therapy. 2000;7(12):1063-6. Epub 2000/06/29. doi: 10.1038/sj.gt.3301206. PubMed PMID: 10871756.

114. Deckers M, van Dinther M, Buijs J, Que I, Löwik C, van der Pluijm G, et al. The Tumor Suppressor Smad4 Is Required for Transforming Growth Factor β–Induced Epithelial to Mesenchymal Transition and Bone Metastasis of Breast Cancer Cells. Cancer Res. 2006;66(4):2202-9. doi: 10.1158/0008-5472.can-05-3560.
115. Darlington GJ. Liver cell lines. Methods in enzymology. 1987;151:19-38. Epub 1987/01/01. PubMed PMID: 3431441.
116. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. Bioinformatics (Oxford, England). 2011;27(12):1739-40. Epub 2011/05/07. doi: 10.1093/bioinformatics/btr260. PubMed PMID: 21546393; PubMed Central PMCID: PMCPmc3106198.
117. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50. Epub 2005/10/04. doi: 10.1073/pnas.0506580102. PubMed PMID: 16199517; PubMed Central PMCID: PMCPmc1239896.
118. Sinjab A. Investigating the roles of HIPK2 and Slug in the mammary stem and progenitor cell state. 2015.
119. Diagenode. AUTO Chip Kit Manual Version 8. Diagenode Resources. 2013.
120. Arner E, Daub CO, Vitting-Seerup K, Andersson R, Lilje B, Drabløs F, et al. Transcribed enhancers lead waves of coordinated transcription in transitioning mammalian cells. Science. 2015;347(6225):1010-4. doi: 10.1126/science.1259418.
121. Shen Y, Yue F, McCleary DF, Ye Z, Edsall L, Kuan S, et al. A map of the cis-regulatory sequences in the mouse genome. Nature. 2012;488(7409):116-20. Epub 2012/07/06. doi: 10.1038/nature11243. PubMed PMID: 22763441; PubMed Central PMCID: PMCPMC4041622.
122. Lemay DG, Pollard KS, Martin WF, Freeman Zadrowski C, Hernandez J, Korf I, et al. From genes to milk: genomic organization and epigenetic regulation of the mammary transcriptome. PloS one. 2013;8(9):e75030. Epub 2013/10/03. doi: 10.1371/journal.pone.0075030. PubMed PMID: 24086428; PubMed Central PMCID: PMCPMC3784412.
123. Warzecha CC, Sato TK, Nabet B, Hogenesch JB, Carstens RP. ESRP1 and ESRP2 are epithelial cell-type-specific regulators of FGFR2 splicing. Molecular cell. 2009;33(5):591-601. Epub 2009/03/17. doi: 10.1016/j.molcel.2009.01.025. PubMed PMID: 19285943; PubMed Central PMCID: PMCPMC2702247.
124. Roring M, Herr R, Fiala GJ, Heilmann K, Braun S, Eisenhardt AE, et al. Distinct requirement for an intact dimer interface in wild-type, V600E and kinase-dead B-Raf signalling. The EMBO journal. 2012;31(11):2629-47. Epub 2012/04/19. doi: 10.1038/emboj.2012.100. PubMed PMID: 22510884; PubMed Central PMCID: PMCPMC3365413.
125. Baer C, Oakes CC, Ruppert AS, Claus R, Kim-Wanner S-Z, Mertens D, et al. Epigenetic silencing of miR-708 enhances NF-κB signaling in chronic lymphocytic leukemia. International Journal of Cancer. 2015;137(6):1352-61. doi: 10.1002/ijc.29491.
126. Cheng Y, Ma Z, Kim BH, Wu W, Cayting P, Boyle AP, et al. Principles of regulatory information conservation between mouse and human. Nature. 2014;515(7527):371-5. Epub 2014/11/21. doi: 10.1038/nature13985. PubMed PMID: 25409826; PubMed Central PMCID: PMCPmc4343047.
127. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, Giannoukos G, et al. Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature. 2007;448(7153):553-60. Epub 2007/07/03. doi: 10.1038/nature06008. PubMed PMID: 17603471; PubMed Central PMCID: PMCPMC2921165.
128. McBryan J, Howlin J, Kenny PA, Shioda T, Martin F. ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene. 2007;26(44):6406-19. Epub 2007/05/09. doi: 10.1038/sj.onc.1210468. PubMed PMID: 17486082.

129. Smid M, Wang Y, Zhang Y, Sieuwerts AM, Yu J, Klijn JG, et al. Subtypes of breast cancer show preferential site of relapse. Cancer Res. 2008;68(9):3108-14. Epub 2008/05/03. doi: 10.1158/0008-5472.can-07-5644. PubMed PMID: 18451135.
130. Rijnkels M, Freeman-Zadrowski C, Hernandez J, Potluri V, Wang L, Li W, et al. Epigenetic modifications unlock the milk protein gene loci during mouse mammary gland development and differentiation. PloS one. 2013;8(1):e53270. Epub 2013/01/10. doi: 10.1371/journal.pone.0053270. PubMed PMID: 23301053; PubMed Central PMCID: PMCPMC3534698.
131. Kang K, Yamaji D, Yoo KH, Robinson GW, Hennighausen L. Mammary-specific gene activation is defined by progressive recruitment of STAT5 during pregnancy and the establishment of H3K4me3 marks. Molecular and cellular biology. 2014;34(3):464-73. Epub 2013/11/28. doi: 10.1128/mcb.00988-13. PubMed PMID: 24277936; PubMed Central PMCID: PMCPmc3911501.
132. Li R, Campos J, Iida J. A Gene Regulatory Program in Human Breast Cancer. Genetics. 2015;201(4):1341-8. Epub 2015/10/30. doi: 10.1534/genetics.115.180125. PubMed PMID: 26510790; PubMed Central PMCID: PMCPmc4676531.
133. Tiemann F, Deppert W. Stabilization of the tumor suppressor p53 during cellular transformation by simian virus 40: influence of viral and cellular factors and biological consequences. Journal of virology. 1994;68(5):2869-78. Epub 1994/05/01. PubMed PMID: 8151757; PubMed Central PMCID: PMCPmc236775.
134. Jovanovic B, Beeler JS, Pickup MW, Chytil A, Gorska AE, Ashby WJ, et al. Transforming growth factor beta receptor type III is a tumor promoter in mesenchymal-stem like triple negative breast cancer. Breast cancer research : BCR. 2014;16(4):R69. Epub 2014/07/06. doi: 10.1186/bcr3684. PubMed PMID: 24985072; PubMed Central PMCID: PMCPmc4095685.
135. Dong M, How T, Kirkbride KC, Gordon KJ, Lee JD, Hempel N, et al. The type III TGF-beta receptor suppresses breast cancer progression. J Clin Invest. 2007;117(1):206-17. Epub 2006/12/13. doi: 10.1172/jci29293. PubMed PMID: 17160136; PubMed Central PMCID: PMCPmc1679965.
136. Chau YM, Pando S, Taylor HS. HOXA11 silencing and endogenous HOXA11 antisense ribonucleic acid in the uterine endometrium. The Journal of clinical endocrinology and metabolism. 2002;87(6):2674-80. Epub 2002/06/07. doi: 10.1210/jcem.87.6.8527. PubMed PMID: 12050232.
137. Yu H, Lindsay J, Feng ZP, Frankenberg S, Hu Y, Carone D, et al. Evolution of coding and non-coding genes in HOX clusters of a marsupial. BMC genomics. 2012;13:251. Epub 2012/06/20. doi: 10.1186/1471-2164-13-251. PubMed PMID: 22708672; PubMed Central PMCID: PMCPMC3541083.
138. Yarmishyn A, Batagov A, Tan J, Sundaram G, Sampath P, Kuznetsov V, et al. HOXD-AS1 is a novel lncRNA encoded in HOXD cluster and a marker of neuroblastoma progression revealed via integrative analysis of noncoding transcriptome. BMC genomics. 2014;15(Suppl 9):S7. PubMed PMID: doi:10.1186/1471-2164-15-S9-S7.
139. Meissner A, Mikkelsen TS, Gu H, Wernig M, Hanna J, Sivachenko A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008;454(7205):766-70. Epub 2008/07/05. doi: 10.1038/nature07107. PubMed PMID: 18600261; PubMed Central PMCID: PMCPMC2896277.
140. Ben-Porath I, Thomson MW, Carey VJ, Ge R, Bell GW, Regev A, et al. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet. 2008;40(5):499-507. Epub 2008/04/30. doi: 10.1038/ng.127. PubMed PMID: 18443585; PubMed Central PMCID: PMCPMC2912221.
141. Nikolsky Y, Sviridov E, Yao J, Dosymbekov D, Ustyansky V, Kaznacheev V, et al. Genome-wide functional synergy between amplified and mutated genes in human breast cancer. Cancer Res. 2008;68(22):9532-40. Epub 2008/11/18. doi: 10.1158/0008-5472.can-08-3082. PubMed PMID: 19010930.
142. Wang W, Huper G, Guo Y, Murphy SK, Olson JA, Jr., Marks JR. Analysis of methylation-sensitive transcriptome identifies GADD45a as a frequently methylated gene in

breast cancer. Oncogene. 2005;24(16):2705-14. Epub 2005/03/01. doi: 10.1038/sj.onc.1208464. PubMed PMID: 15735726.
143. Jaeger J, Koczan D, Thiesen HJ, Ibrahim SM, Gross G, Spang R, et al. Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. Clinical cancer research : an official journal of the American Association for Cancer Research. 2007;13(3):806-15. Epub 2007/02/10. doi: 10.1158/1078-0432.ccr-06-1820. PubMed PMID: 17289871.
144. Shapiro IM, Cheng AW, Flytzanis NC, Balsamo M, Condeelis JS, Oktay MH, et al. An EMT-driven alternative splicing program occurs in human breast cancer and modulates cellular phenotype. PLoS genetics. 2011;7(8):e1002218. Epub 2011/08/31. doi: 10.1371/journal.pgen.1002218. PubMed PMID: 21876675; PubMed Central PMCID: PMCPMC3158048.
145. Warzecha CC, Jiang P, Amirikian K, Dittmar KA, Lu H, Shen S, et al. An ESRP-regulated splicing programme is abrogated during the epithelial-mesenchymal transition. The EMBO journal. 2010;29(19):3286-300. Epub 2010/08/17. doi: 10.1038/emboj.2010.195. PubMed PMID: 20711167; PubMed Central PMCID: PMCPMC2957203.
146. Warzecha CC, Shen S, Xing Y, Carstens RP. The epithelial splicing factors ESRP1 and ESRP2 positively and negatively regulate diverse types of alternative splicing events. RNA biology. 2009;6(5):546-62. Epub 2009/10/16. PubMed PMID: 19829082; PubMed Central PMCID: PMCPMC2823379.
147. Li AM, Tian AX, Zhang RX, Ge J, Sun X, Cao XC. Protocadherin-7 induces bone metastasis of breast cancer. Biochemical and biophysical research communications. 2013;436(3):486-90. Epub 2013/06/12. doi: 10.1016/j.bbrc.2013.05.131. PubMed PMID: 23751349.
148. Iwakiri M, Mizukami K, Ikonomovic MD, Ishikawa M, Abrahamson EE, DeKosky ST, et al. An immunohistochemical study of GABA A receptor gamma subunits in Alzheimer's disease hippocampus: relationship to neurofibrillary tangle progression. Neuropathology : official journal of the Japanese Society of Neuropathology. 2009;29(3):263-9. Epub 2008/11/21. doi: 10.1111/j.1440-1789.2008.00978.x. PubMed PMID: 19019179; PubMed Central PMCID: PMCPMC3078755.
149. Xu Z, Pei L, Wang L, Zhang F, Hu X, Gui Y. Snail1-dependent transcriptional repression of Cezanne2 in hepatocellular carcinoma. Oncogene. 2014;33(22):2836-45. Epub 2013/06/25. doi: 10.1038/onc.2013.243. PubMed PMID: 23792447.
150. Bai Y, Fang N, Gu T, Kang Y, Wu J, Yang D, et al. HOXA11 gene is hypermethylation and aberrant expression in gastric cancer. Cancer cell international. 2014;14:79. Epub 2014/01/01. doi: 10.1186/s12935-014-0079-7. PubMed PMID: 25788862; PubMed Central PMCID: PMCPMC4364045.
151. Hwang JA, Lee BB, Kim Y, Park SE, Heo K, Hong SH, et al. HOXA11 hypermethylation is associated with progression of non-small cell lung cancer. Oncotarget. 2013;4(12):2317-25. Epub 2013/11/22. doi: 10.18632/oncotarget.1464. PubMed PMID: 24259349; PubMed Central PMCID: PMCPMC3926829.
152. Masuda S, Oda Y, Sasaki H, Ikenouchi J, Higashi T, Akashi M, et al. LSR defines cell corners for tricellular tight junction formation in epithelial cells. Journal of cell science. 2011;124(Pt 4):548-55. Epub 2011/01/20. doi: 10.1242/jcs.072058. PubMed PMID: 21245199.
153. Forrest AR, Kawaji H, Rehli M, Baillie JK, de Hoon MJ, Haberle V, et al. A promoter-level mammalian expression atlas. Nature. 2014;507(7493):462-70. Epub 2014/03/29. doi: 10.1038/nature13182. PubMed PMID: 24670764; PubMed Central PMCID: PMCPMC4529748.
154. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome Biol. 2015;16:22. Epub 2015/02/28. doi: 10.1186/s13059-014-0560-6. PubMed PMID: 25723102; PubMed Central PMCID: PMCPMC4310165.
155. Severin J, Lizio M, Harshbarger J, Kawaji H, Daub CO, Hayashizaki Y, et al. Interactive visualization and analysis of large-scale sequencing datasets using ZENBU.

Nature biotechnology. 2014;32(3):217-9. Epub 2014/04/15. doi: 10.1038/nbt.2840. PubMed PMID: 24727769.
156. Hon GC, Rajagopal N, Shen Y, McCleary DF, Yue F, Dang MD, et al. Epigenetic memory at embryonic enhancers identified in DNA methylation maps from adult mouse tissues. Nat Genet. 2013;45(10):1198-206. doi: 10.1038/ng.2746.
157. Song Q, Decato B, Hong EE, Zhou M, Fang F, Qu J, et al. A Reference Methylome Database and Analysis Pipeline to Facilitate Integrative and Comparative Epigenomics. PloS one. 2013;8(12):e81148. doi: 10.1371/journal.pone.0081148.
158. Parkhomchuk D, Borodina T, Amstislavskiy V, Banaru M, Hallen L, Krobitsch S, et al. Transcriptome analysis by strand-specific sequencing of complementary DNA. Nucleic Acids Res. 2009;37(18):e123. doi: 10.1093/nar/gkp596.
159. Jiang L, Schlesinger F, Davis CA, Zhang Y, Li R, Salit M, et al. Synthetic spike-in standards for RNA-seq experiments. Genome Research. 2011;21(9):1543-51. doi: 10.1101/gr.121095.111.
160. Madden SF, Clarke C, Gaule P, Aherne ST, O'Donovan N, Clynes M, et al. BreastMark: an integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome. Breast cancer research : BCR. 2013;15(4):R52. Epub 2013/07/04. doi: 10.1186/bcr3444. PubMed PMID: 23820017; PubMed Central PMCID: PMCPMC3978487.
161. Conerly M, Grady WM. Insights into the role of DNA methylation in disease through the use of mouse models. Disease models & mechanisms. 2010;3(5-6):290-7. Epub 2010/04/30. doi: 10.1242/dmm.004812. PubMed PMID: 20427558; PubMed Central PMCID: PMCPmc2860849.
162. Pathania R, Ramachandran S, Elangovan S, Padia R, Yang P, Cinghu S, et al. DNMT1 is essential for mammary and cancer stem cell maintenance and tumorigenesis. Nature communications. 2015;6:6910. Epub 2015/04/25. doi: 10.1038/ncomms7910. PubMed PMID: 25908435; PubMed Central PMCID: PMCPMC4410389.
163. Chen M, Shabashvili D, Nawab A, Yang SX, Dyer LM, Brown KD, et al. DNA methyltransferase inhibitor, zebularine, delays tumor growth and induces apoptosis in a genetically engineered mouse model of breast cancer. Mol Cancer Ther. 2012;11(2):370-82. Epub 2011/12/29. doi: 10.1158/1535-7163.mct-11-0458. PubMed PMID: 22203734.
164. Ulahannan N, Greally JM. Genome-wide assays that identify and quantify modified cytosines in human disease studies. Epigenetics & chromatin. 2015;8:5. Epub 2015/03/20. doi: 10.1186/1756-8935-8-5. PubMed PMID: 25788985; PubMed Central PMCID: PMCPmc4363328.
165. Wong NC, Ng J, Hall NE, Lunke S, Salmanidis M, Brumatti G, et al. Exploring the utility of human DNA methylation arrays for profiling mouse genomic DNA. Genomics. 2013;102(1):38-46. Epub 2013/05/04. doi: 10.1016/j.ygeno.2013.04.014. PubMed PMID: 23639479.
166. Sonnet M, Claus R, Becker N, Zucknick M, Petersen J, Lipka DB, et al. Early aberrant DNA methylation events in a mouse model of acute myeloid leukemia. Genome medicine. 2014;6(4):34. Epub 2014/06/20. doi: 10.1186/gm551. PubMed PMID: 24944583; PubMed Central PMCID: PMCPMC4062060.
167. Fujiwara K, Ghosh S, Liang P, Morien E, Soma M, Nagase H. Genome-wide screening of aberrant DNA methylation which associated with gene expression in mouse skin cancers. Molecular carcinogenesis. 2015;54(3):178-88. Epub 2013/10/12. doi: 10.1002/mc.22085. PubMed PMID: 24115114.
168. Grimm C, Chavez L, Vilardell M, Farrall AL, Tierling S, Bohm JW, et al. DNA-methylome analysis of mouse intestinal adenoma identifies a tumour-specific signature that is partly conserved in human colon cancer. PLoS genetics. 2013;9(2):e1003250. Epub 2013/02/15. doi: 10.1371/journal.pgen.1003250. PubMed PMID: 23408899; PubMed Central PMCID: PMCPmc3567140.
169. Guo Y, Lee JH, Shu L, Huang Y, Li W, Zhang C, et al. Association of aberrant DNA methylation in Apc(min/+) mice with the epithelial-mesenchymal transition and Wnt/beta-catenin pathways: genome-wide analysis using MeDIP-seq. Cell & bioscience. 2015;5:24.

Epub 2015/06/24. doi: 10.1186/s13578-015-0013-2. PubMed PMID: 26101583; PubMed Central PMCID: PMCPmc4476183.
170. Yang AY, Lee JH, Shu L, Zhang C, Su ZY, Lu Y, et al. Genome-wide analysis of DNA methylation in UVB- and DMBA/TPA-induced mouse skin cancer models. Life sciences. 2014;113(1-2):45-54. Epub 2014/08/06. doi: 10.1016/j.lfs.2014.07.031. PubMed PMID: 25093921.
171. Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, et al. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. Nature. 2011;479(7371):74-9. Epub 2011/10/04. doi: 10.1038/nature10442. PubMed PMID: 21964334.
172. Medvedeva YA, Khamis AM, Kulakovskiy IV, Ba-Alawi W, Bhuyan MS, Kawaji H, et al. Effects of cytosine methylation on transcription factor binding sites. BMC genomics. 2014;15:119. Epub 2014/03/29. doi: 10.1186/1471-2164-15-119. PubMed PMID: 24669864; PubMed Central PMCID: PMCPmc3986887.
173. Oliver JR, Kushwah R, Hu J. Multiple roles of the epithelium-specific ETS transcription factor, ESE-1, in development and disease. Laboratory investigation; a journal of technical methods and pathology. 2012;92(3):320-30. Epub 2011/12/14. doi: 10.1038/labinvest.2011.186. PubMed PMID: 22157719.
174. Bonne S, van Hengel J, Nollet F, Kools P, van Roy F. Plakophilin-3, a novel armadillo-like protein present in nuclei and desmosomes of epithelial cells. Journal of cell science. 1999;112 (Pt 14):2265-76. Epub 1999/06/25. PubMed PMID: 10381383.
175. Swisshelm K, Macek R, Kubbies M. Role of claudins in tumorigenesis. Advanced drug delivery reviews. 2005;57(6):919-28. Epub 2005/04/12. doi: 10.1016/j.addr.2005.01.006. PubMed PMID: 15820559.
176. Crowley MR, Head KL, Kwiatkowski DJ, Asch HL, Asch BB. The mouse mammary gland requires the actin-binding protein gelsolin for proper ductal morphogenesis. Developmental biology. 2000;225(2):407-23. Epub 2000/09/14. doi: 10.1006/dbio.2000.9844. PubMed PMID: 10985859.
177. Bach LA, Fu P, Yang Z. Insulin-like growth factor-binding protein-6 and cancer. Clinical science (London, England : 1979). 2013;124(4):215-29. Epub 2012/11/07. doi: 10.1042/cs20120343. PubMed PMID: 23126425.
178. Huh SJ, Clement K, Jee D, Merlini A, Choudhury S, Maruyama R, et al. Age- and pregnancy-associated DNA methylation changes in mammary epithelial cells. Stem cell reports. 2015;4(2):297-311. Epub 2015/01/27. doi: 10.1016/j.stemcr.2014.12.009. PubMed PMID: 25619437; PubMed Central PMCID: PMCPmc4325231.
179. Dos Santos CO, Dolzhenko E, Hodges E, Smith AD, Hannon GJ. An epigenetic memory of pregnancy in the mouse mammary gland. Cell Rep. 2015;11(7):1102-9. Epub 2015/05/12. doi: 10.1016/j.celrep.2015.04.015. PubMed PMID: 25959817; PubMed Central PMCID: PMCPmc4439279.
180. Mesquita B, Lopes P, Rodrigues A, Pereira D, Afonso M, Leal C, et al. Frequent copy number gains at 1q21 and 1q32 are associated with overexpression of the ETS transcription factors ETV3 and ELF3 in breast cancer irrespective of molecular subtypes. Breast Cancer Res Treat. 2013;138(1):37-45. Epub 2013/01/19. doi: 10.1007/s10549-013-2408-2. PubMed PMID: 23329352.
181. Demirag GG, Sullu Y, Yucel I. Expression of Plakophilins (PKP1, PKP2, and PKP3) in breast cancers. Medical oncology (Northwood, London, England). 2012;29(3):1518-22. Epub 2011/09/29. doi: 10.1007/s12032-011-0071-1. PubMed PMID: 21947748.
182. Abba MC, Hu Y, Levy CC, Gaddis S, Kittrell FS, Hill J, et al. Identification of modulated genes by three classes of chemopreventive agents at preneoplastic stages in a p53-null mouse mammary tumor model. Cancer prevention research (Philadelphia, Pa). 2009;2(2):175-84. Epub 2009/01/29. doi: 10.1158/1940-6207.capr-08-0104. PubMed PMID: 19174580; PubMed Central PMCID: PMCPmc3023963.
183. Yang Y, Sheng M, Huang F, Bu D, Liu X, Yao Y, et al. Downregulation of Insulin-like growth factor binding protein 6 is associated with ACTH-secreting pituitary adenoma growth.

Pituitary. 2014;17(6):505-13. Epub 2014/01/01. doi: 10.1007/s11102-013-0535-8. PubMed PMID: 24379119.
184. Blanchard AA, Skliris GP, Watson PH, Murphy LC, Penner C, Tomes L, et al. Claudins 1, 3, and 4 protein expression in ER negative breast cancer correlates with markers of the basal phenotype. Virchows Archiv : an international journal of pathology. 2009;454(6):647-56. Epub 2009/04/24. doi: 10.1007/s00428-009-0770-6. PubMed PMID: 19387682.
185. Lee HJ, Hinshelwood RA, Bouras T, Gallego-Ortega D, Valdes-Mora F, Blazek K, et al. Lineage specific methylation of the Elf5 promoter in mammary epithelial cells. Stem cells (Dayton, Ohio). 2011;29(10):1611-9. Epub 2011/08/09. doi: 10.1002/stem.706. PubMed PMID: 21823211.
186. Diede SJ, Yao Z, Keyes CC, Tyler AE, Dey J, Hackett CS, et al. Fundamental differences in promoter CpG island DNA hypermethylation between human cancer and genetically engineered mouse models of cancer. Epigenetics. 2013;8(12):1254-60. Epub 2013/10/11. doi: 10.4161/epi.26486. PubMed PMID: 24107773; PubMed Central PMCID: PMCPmc3933486.
187. Houseman EA, Accomando WP, Koestler DC, Christensen BC, Marsit CJ, Nelson HH, et al. DNA methylation arrays as surrogate measures of cell mixture distribution. BMC bioinformatics. 2012;13:86. Epub 2012/05/10. doi: 10.1186/1471-2105-13-86. PubMed PMID: 22568884; PubMed Central PMCID: PMCPmc3532182.
188. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Molecular oncology. 2015;9(3):555-68. Epub 2014/12/04. doi: 10.1016/j.molonc.2014.10.012. PubMed PMID: 25468711.
189. Locke WJ, Zotenko E, Stirzaker C, Robinson MD, Hinshelwood RA, Stone A, et al. Coordinated epigenetic remodelling of transcriptional networks occurs during early breast carcinogenesis. Clinical epigenetics. 2015;7(1):52. Epub 2015/05/12. doi: 10.1186/s13148-015-0086-0. PubMed PMID: 25960784; PubMed Central PMCID: PMCPmc4424562.
190. Stirzaker C, Zotenko E, Song JZ, Qu W, Nair SS, Locke WJ, et al. Methylome sequencing in triple-negative breast cancer reveals distinct methylation clusters with prognostic value. Nature communications. 2015;6:5899. Epub 2015/02/03. doi: 10.1038/ncomms6899. PubMed PMID: 25641231.
191. Holm K, Staaf J, Lauss M, Aine M, Lindgren D, Bendahl PO, et al. An integrated genomics analysis of epigenetic subtypes in human breast tumors links DNA methylation patterns to chromatin states in normal mammary cells. Breast cancer research : BCR. 2016;18(1):27. Epub 2016/03/01. doi: 10.1186/s13058-016-0685-5. PubMed PMID: 26923702; PubMed Central PMCID: PMCPmc4770527.
192. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692-702. Epub 2011/05/20. PubMed PMID: 21593595.
193. Visvader JE. Cells of origin in cancer. Nature. 2011;469(7330):314-22.
194. Oakes CC, Seifert M, Assenov Y, Gu L, Przekopowitz M, Ruppert AS, et al. DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia. Nat Genet. 2016;48(3):253-64. Epub 2016/01/19. doi: 10.1038/ng.3488. PubMed PMID: 26780610.
195. Van Keymeulen A, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, et al. Distinct stem cells contribute to mammary gland development and maintenance. Nature. 2011;479(7372):189-93. doi: http://www.nature.com/nature/journal/v479/n7372/abs/nature10573.html#supplementary-information.
196. Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, et al. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med. 2009;15. doi: 10.1038/nm.2000.

197. Shehata M, Teschendorff A, Sharp G, Novcic N, Russell IA, Avril S, et al. Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. Breast Cancer Research. 2012;14(5):1-19. doi: 10.1186/bcr3334.
198. Drobysheva D, Smith BA, McDowell M, Guillen KP, Ekiz HA, Welm BE. Transformation of enriched mammary cell populations with polyomavirus middle T antigen influences tumor subtype and metastatic potential. Breast cancer research : BCR. 2015;17(1):132. Epub 2015/10/03. doi: 10.1186/s13058-015-0641-9. PubMed PMID: 26429062; PubMed Central PMCID: PMCPmc4589945.
199. Visvader JE, Stingl J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes & development. 2014;28(11):1143-58. Epub 2014/06/04. doi: 10.1101/gad.242511.114. PubMed PMID: 24888586; PubMed Central PMCID: PMCPmc4052761.
200. Kasowski M, Kyriazopoulou-Panagiotopoulou S, Grubert F, Zaugg JB, Kundaje A, Liu Y, et al. Extensive variation in chromatin states across humans. Science. 2013;342(6159):750-2. Epub 2013/10/19. doi: 10.1126/science.1242510. PubMed PMID: 24136358; PubMed Central PMCID: PMCPmc4075767.
201. Bogu GK, Vizan P, Stanton LW, Beato M, Di Croce L, Marti-Renom MA. Chromatin and RNA Maps Reveal Regulatory Long Noncoding RNAs in Mouse. Molecular and cellular biology. 2015;36(5):809-19. Epub 2015/12/30. doi: 10.1128/mcb.00955-15. PubMed PMID: 26711262; PubMed Central PMCID: PMCPmc4760218.
202. Sugathan A, Waxman DJ. Genome-wide analysis of chromatin states reveals distinct mechanisms of sex-dependent gene regulation in male and female mouse liver. Molecular and cellular biology. 2013;33(18):3594-610. Epub 2013/07/10. doi: 10.1128/mcb.00280-13. PubMed PMID: 23836885; PubMed Central PMCID: PMCPMC3753870.
203. Rhie SK, Hazelett DJ, Coetzee SG, Yan C, Noushmehr H, Coetzee GA. Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. BMC genomics. 2014;15:331. Epub 2014/06/03. doi: 10.1186/1471-2164-15-331. PubMed PMID: 24885402; PubMed Central PMCID: PMCPmc4035062.
204. Zhang T, Cooper S, Brockdorff N. The interplay of histone modifications - writers that read. EMBO reports. 2015;16(11):1467-81. Epub 2015/10/18. doi: 10.15252/embr.201540945. PubMed PMID: 26474904; PubMed Central PMCID: PMCPmc4641500.
205. Kim S, Yu NK, Kaang BK. CTCF as a multifunctional protein in genome regulation and gene expression. Experimental & molecular medicine. 2015;47:e166. Epub 2015/06/06. doi: 10.1038/emm.2015.33. PubMed PMID: 26045254; PubMed Central PMCID: PMCPmc4491725.
206. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, et al. The accessible chromatin landscape of the human genome. Nature. 2012;489(7414):75-82. doi: http://www.nature.com/nature/journal/v489/n7414/abs/nature11232.html#supplementary-information.
207. Boyle AP, Davis S, Shulha HP, Meltzer P, Margulies EH, Weng Z, et al. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. Cell. 2008;132(2):311-22. doi: http://dx.doi.org/10.1016/j.cell.2007.12.014.
208. Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM, Harrison DJ, et al. Transcriptionally repressed genes become aberrantly methylated and distinguish tumors of different lineages in breast cancer. Proc Natl Acad Sci U S A. 2011;108(11):4364-9. Epub 2011/03/04. doi: 10.1073/pnas.1013224108. PubMed PMID: 21368160; PubMed Central PMCID: PMCPmc3060255.
209. Hon GC, Hawkins RD, Caballero OL, Lo C, Lister R, Pelizzola M, et al. Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. Genome Res. 2012;22(2):246-58. Epub 2011/12/14. doi: 10.1101/gr.125872.111. PubMed PMID: 22156296; PubMed Central PMCID: PMCPmc3266032.

210. Walter M, Teissandier A, Perez-Palacios R, Bourc'his D. An epigenetic switch ensures transposon repression upon dynamic loss of DNA methylation in embryonic stem cells. eLife. 2016;5. Epub 2016/01/28. doi: 10.7554/eLife.11418. PubMed PMID: 26814573; PubMed Central PMCID: PMCPmc4769179.
211. Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome Res. 2012;22(6):1128-38. Epub 2012/04/03. doi: 10.1101/gr.133728.111. PubMed PMID: 22466170; PubMed Central PMCID: PMCPmc3371717.
212. Muratani M, Deng N, Ooi WF, Lin SJ, Xing M, Xu C, et al. Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements. Nature communications. 2014;5:4361. Epub 2014/07/11. doi: 10.1038/ncomms5361. PubMed PMID: 25008978.
213. Zhang B, Zhou Y, Lin N, Lowdon RF, Hong C, Nagarajan RP, et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. Genome Res. 2013;23(9):1522-40. Epub 2013/06/28. doi: 10.1101/gr.156539.113. PubMed PMID: 23804400; PubMed Central PMCID: PMCPmc3759728.
214. Aran D, Sabato S, Hellman A. DNA methylation of distal regulatory sites characterizes dysregulation of cancer genes. Genome Biology. 2013;14(3):R21. PubMed PMID: doi:10.1186/gb-2013-14-3-r21.
215. Zhang B, Xing X, Li J, Lowdon RF, Zhou Y, Lin N, et al. Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. BMC genomics. 2014;15:868. Epub 2014/10/08. doi: 10.1186/1471-2164-15-868. PubMed PMID: 25286960; PubMed Central PMCID: PMCPmc4198682.
216. Taberlay PC, Statham AL, Kelly TK, Clark SJ, Jones PA. Reconfiguration of nucleosome-depleted regions at distal regulatory elements accompanies DNA methylation of enhancers and insulators in cancer. Genome Res. 2014;24(9):1421-32. Epub 2014/06/12. doi: 10.1101/gr.163485.113. PubMed PMID: 24916973; PubMed Central PMCID: PMCPmc4158760.
217. Heyn H, Vidal E, Ferreira HJ, Vizoso M, Sayols S, Gomez A, et al. Epigenomic analysis detects aberrant super-enhancer DNA methylation in human cancer. Genome Biol. 2016;17(1):11. Epub 2016/01/28. doi: 10.1186/s13059-016-0879-2. PubMed PMID: 26813288; PubMed Central PMCID: PMCPmc4728783.
218. Stadler MB, Murr R, Burger L, Ivanek R, Lienert F, Scholer A, et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature. 2011;480(7378):490-5. Epub 2011/12/16. doi: 10.1038/nature10716. PubMed PMID: 22170606.
219. Rijnkels M, Kabotyanski E, Montazer-Torbati MB, Hue Beauvais C, Vassetzky Y, Rosen JM, et al. The epigenetic landscape of mammary gland development and functional differentiation. Journal of mammary gland biology and neoplasia. 2010;15(1):85-100. Epub 2010/02/17. doi: 10.1007/s10911-010-9170-4. PubMed PMID: 20157770; PubMed Central PMCID: PMCPmc3006238.
220. Kurpios NA, MacNeil L, Shepherd TG, Gludish DW, Giacomelli AO, Hassell JA. The Pea3 Ets transcription factor regulates differentiation of multipotent progenitor cells during mammary gland development. Developmental biology. 2009;325(1):106-21. Epub 2008/11/04. doi: 10.1016/j.ydbio.2008.09.033. PubMed PMID: 18977342.
221. Kurpios NA, Sabolic NA, Shepherd TG, Fidalgo GM, Hassell JA. Function of PEA3 Ets transcription factors in mammary gland development and oncogenesis. Journal of mammary gland biology and neoplasia. 2003;8(2):177-90. Epub 2003/11/26. PubMed PMID: 14635793.
222. Yuan ZY, Dai T, Wang SS, Peng RJ, Li XH, Qin T, et al. Overexpression of ETV4 protein in triple-negative breast cancer is associated with a higher risk of distant metastasis.

OncoTargets and therapy. 2014;7:1733-42. Epub 2014/10/21. doi: 10.2147/ott.s66692. PubMed PMID: 25328406; PubMed Central PMCID: PMCPmc4196788.
223. Shepherd T, Hassell JA. Role of Ets transcription factors in mammary gland development and oncogenesis. Journal of mammary gland biology and neoplasia. 2001;6(1):129-40. Epub 2001/07/27. PubMed PMID: 11467448.
224. Shepherd TG, Kockeritz L, Szrajber MR, Muller WJ, Hassell JA. The pea3 subfamily ets genes are required for HER2/Neu-mediated mammary oncogenesis. Current biology : CB. 2001;11(22):1739-48. Epub 2001/11/24. PubMed PMID: 11719215.
225. Ito Y. Oncogenic potential of the RUNX gene family: 'overview'. Oncogene. 2004;23(24):4198-208. Epub 2004/05/25. doi: 10.1038/sj.onc.1207755. PubMed PMID: 15156173.
226. Ferrari N, Mohammed ZM, Nixon C, Mason SM, Mallon E, McMillan DC, et al. Expression of RUNX1 correlates with poor patient prognosis in triple negative breast cancer. PloS one. 2014;9(6):e100759. Epub 2014/06/27. doi: 10.1371/journal.pone.0100759. PubMed PMID: 24967588; PubMed Central PMCID: PMCPmc4072705.
227. Sokol ES, Sanduja S, Jin DX, Miller DH, Mathis RA, Gupta PB. Perturbation-expression analysis identifies RUNX1 as a regulator of human mammary stem cell differentiation. PLoS computational biology. 2015;11(4):e1004161. Epub 2015/04/22. doi: 10.1371/journal.pcbi.1004161. PubMed PMID: 25894653; PubMed Central PMCID: PMCPmc4404314.
228. Gilbert PM, Mouw JK, Unger MA, Lakins JN, Gbegnon MK, Clemmer VB, et al. HOXA9 regulates BRCA1 expression to modulate human breast tumor phenotype. J Clin Invest. 2010;120(5):1535-50. Epub 2010/04/15. doi: 10.1172/jci39534. PubMed PMID: 20389018; PubMed Central PMCID: PMCPmc2860938.
229. Li D, Ilnytskyy Y, Kovalchuk A, Khachigian LM, Bronson RT, Wang B, et al. Crucial role for early growth response-1 in the transcriptional regulation of miR-20b in breast cancer. Oncotarget. 2013;4(9):1373-87. Epub 2013/08/16. doi: 10.18632/oncotarget.1165. PubMed PMID: 23945289; PubMed Central PMCID: PMCPmc3824527.
230. Baker KM, Wei G, Schaffner AE, Ostrowski MC. Ets-2 and components of mammalian SWI/SNF form a repressor complex that negatively regulates the BRCA1 promoter. The Journal of biological chemistry. 2003;278(20):17876-84. Epub 2003/03/15. doi: 10.1074/jbc.M209480200. PubMed PMID: 12637547.
231. Shimizu Y, Luk H, Horio D, Miron P, Griswold M, Iglehart D, et al. BRCA1-IRIS overexpression promotes formation of aggressive breast cancers. PloS one. 2012;7(4):e34102. Epub 2012/04/19. doi: 10.1371/journal.pone.0034102. PubMed PMID: 22511931; PubMed Central PMCID: PMCPmc3325250.
232. ElShamy WM, Livingston DM. Identification of BRCA1-IRIS, a BRCA1 locus product. Nature cell biology. 2004;6(10):954-67. Epub 2004/09/28. doi: 10.1038/ncb1171. PubMed PMID: 15448696.
233. Shimizu Y, Mullins N, Blanchard Z, Elshamy WM. BRCA1/p220 loss triggers BRCA1-IRIS overexpression via mRNA stabilization in breast cancer cells. Oncotarget. 2012;3(3):299-313. Epub 2012/03/21. doi: 10.18632/oncotarget.462. PubMed PMID: 22431556; PubMed Central PMCID: PMCPmc3359886.
234. Hermannstadter A, Ziegler C, Kuhl M, Deppert W, Tolstonog GV. Wild-type p53 enhances efficiency of simian virus 40 large-T-antigen-induced cellular transformation. Journal of virology. 2009;83(19):10106-18. Epub 2009/07/25. doi: 10.1128/jvi.00174-09. PubMed PMID: 19625393; PubMed Central PMCID: PMCPmc2748005.
235. Bocchetta M, Eliasz S, De Marco MA, Rudzinski J, Zhang L, Carbone M. The SV40 large T antigen-p53 complexes bind and activate the insulin-like growth factor-I promoter stimulating cell growth. Cancer Res. 2008;68(4):1022-9. Epub 2008/02/19. doi: 10.1158/0008-5472.can-07-5203. PubMed PMID: 18281476.
236. Reihsaus E, Kohler M, Kraiss S, Oren M, Montenarh M. Regulation of the level of the oncoprotein p53 in non-transformed and transformed cells. Oncogene. 1990;5(1):137-45. Epub 1990/01/01. PubMed PMID: 2157179.

237. Nguyen TA, Menendez D, Resnick MA, Anderson CW. Mutant TP53 posttranslational modifications: challenges and opportunities. Human mutation. 2014;35(6):738-55. Epub 2014/01/08. doi: 10.1002/humu.22506. PubMed PMID: 24395704; PubMed Central PMCID: PMCPmc4074372.
238. Zhu J, Sammons MA, Donahue G, Dou Z, Vedadi M, Getlik M, et al. Gain-of-function p53 mutants co-opt chromatin pathways to drive cancer growth. Nature. 2015;525(7568):206-11. Epub 2015/09/04. doi: 10.1038/nature15251. PubMed PMID: 26331536; PubMed Central PMCID: PMCPmc4568559.
239. Tzeng YJ, Zimmermann C, Guhl E, Berg B, Avantaggiati ML, Graessmann A. SV40 T/t-antigen induces premature mammary gland involution by apoptosis and selects for p53 missense mutation in mammary tumors. Oncogene. 1998;16(16):2103-14. Epub 1998/05/08. doi: 10.1038/sj.onc.1201733. PubMed PMID: 9572491.
240. Li Y, Zhang Y, Li S, Lu J, Chen J, Wang Y, et al. Genome-wide DNA methylome analysis reveals epigenetically dysregulated non-coding RNAs in human breast cancer. Scientific reports. 2015;5:8790. Epub 2015/03/06. doi: 10.1038/srep08790. PubMed PMID: 25739977; PubMed Central PMCID: PMCPMC4350105.
241. Deeb KK, Michalowska AM, Yoon CY, Krummey SM, Hoenerhoff MJ, Kavanaugh C, et al. Identification of an integrated SV40 T/t-antigen cancer signature in aggressive human breast, prostate, and lung carcinomas with poor prognosis. Cancer Res. 2007;67(17):8065-80. doi: 10.1158/0008-5472.CAN-07-1515. PubMed PMID: 17804718.
242. Gebhard C, Schwarzfischer L, Pham TH, Schilling E, Klug M, Andreesen R, et al. Genome-wide profiling of CpG methylation identifies novel targets of aberrant hypermethylation in myeloid leukemia. Cancer Res. 2006;66(12):6118-28. Epub 2006/06/17. doi: 66/12/6118 [pii]

10.1158/0008-5472.CAN-06-0376. PubMed PMID: 16778185.
243. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, et al. PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. Nat Genet. 2003;34(3):267-73. Epub 2003/06/17. doi: 10.1038/ng1180. PubMed PMID: 12808457.
244. Abate-Shen C. Deregulated homeobox gene expression in cancer: cause or consequence? Nat Rev Cancer. 2002;2(10):777-85. Epub 2002/10/03. doi: 10.1038/nrc907. PubMed PMID: 12360280.
245. Tommasi S, Karm DL, Wu X, Yen Y, Pfeifer GP. Methylation of homeobox genes is a frequent and early epigenetic event in breast cancer. Breast cancer research : BCR. 2009;11(1):R14. Epub 2009/03/03. doi: 10.1186/bcr2233. PubMed PMID: 19250546; PubMed Central PMCID: PMCPmc2687719.
246. Reaves DK, Fagan-Solis KD, Dunphy K, Oliver SD, Scott DW, Fleming JM. The role of lipolysis stimulated lipoprotein receptor in breast cancer and directing breast cancer cell behavior. PLoS One. 2014;9(3):e91747. doi: 10.1371/journal.pone.0091747. PubMed PMID: 24637461; PubMed Central PMCID: PMCPMC3956714.
247. Horiguchi K, Sakamoto K, Koinuma D, Semba K, Inoue A, Inoue S, et al. TGF-beta drives epithelial-mesenchymal transition through deltaEF1-mediated downregulation of ESRP. Oncogene. 2012;31(26):3190-201. Epub 2011/11/01. doi: 10.1038/onc.2011.493. PubMed PMID: 22037216; PubMed Central PMCID: PMCPMC3391666.
248. De Craene B, Berx G. Regulatory networks defining EMT during cancer initiation and progression. Nat Rev Cancer. 2013;13(2):97-110. Epub 2013/01/25. doi: 10.1038/nrc3447. PubMed PMID: 23344542.
249. Bock C, Beerman I, Lien WH, Smith ZD, Gu H, Boyle P, et al. DNA methylation dynamics during in vivo differentiation of blood and skin stem cells. Molecular cell. 2012;47(4):633-47. Epub 2012/07/31. doi: 10.1016/j.molcel.2012.06.019. PubMed PMID: 22841485; PubMed Central PMCID: PMCPMC3428428.
250. Barutcu AR, Fritz AJ, Zaidi SK, van Wijnen AJ, Lian JB, Stein JL, et al. C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization. Journal of cellular physiology. 2016;231(1):31-5. Epub

2015/06/11. doi: 10.1002/jcp.25062. PubMed PMID: 26059817; PubMed Central PMCID: PMCPMC4586368.
251. Sigova AA, Abraham BJ, Ji X, Molinie B, Hannett NM, Guo YE, et al. Transcription factor trapping by RNA in gene regulatory elements. Science. 2015;350(6263):978-81. Epub 2015/10/31. doi: 10.1126/science.aad3346. PubMed PMID: 26516199.
252. Sessa L, Breiling A, Lavorgna G, Silvestri L, Casari G, Orlando V. Noncoding RNA synthesis and loss of Polycomb group repression accompanies the colinear activation of the human HOXA cluster. RNA (New York, NY). 2007;13(2):223-39. Epub 2006/12/23. doi: 10.1261/rna.266707. PubMed PMID: 17185360; PubMed Central PMCID: PMCPMC1781374.
253. Musahl AS, Huang X, Rusakiewicz S, Ntini E, Marsico A, Kroemer G, et al. A long non-coding RNA links calreticulin-mediated immunogenic cell removal to RB1 transcription. Oncogene. 2015;34(39):5046-54. doi: 10.1038/onc.2014.424.
254. Ishii H, Saitoh M, Sakamoto K, Kondo T, Katoh R, Tanaka S, et al. Epithelial splicing regulatory proteins 1 (ESRP1) and 2 (ESRP2) suppress cancer cell motility via different mechanisms. The Journal of biological chemistry. 2014;289(40):27386-99. Epub 2014/08/22. doi: 10.1074/jbc.M114.589432. PubMed PMID: 25143390; PubMed Central PMCID: PMCPMC4183779.

Appendix

Supplementary Table 1 Samples and quality control for MClp-seq

Sample ID	Genotype	Age	# of unique reads	Saturation correlation
kinetics 41	tg	12w	2469523	0.86
kinetics 44	tg	12w	4062985	0.86
kinetics 78	tg	12w	4632929	0.92
kinetics 86	wt	12w	3556016	0.9
kinetics 34	wt	12w	4129539	0.89
kinetics 35	wt	12w	2651980	0.86
kinetics 40	tg	16w	2974597	0.86
kinetics 46	tg	16w	3645077	0.89
kinetics 37	tg	16w	2422712	0.85
kinetics 63	wt	16w	3006720	0.85
kinetics 39	wt	16w	3775421	0.92
kinetics 60	wt	16w	3912173	0.9
kinetics 28	wt	20w	2254463	0.8
kinetics 33	wt	20w	3694733	0.75
kinetics 27	wt	20w	4849901	0.9
kinetics 8	tg	21w	2178260	0.85
kinetics 47	tg	21w	2660666	0.79
kinetics 7	tg	21w	2942252	0.85
kinetics 15	tg	23w	3276032	0.75
kinetics 16	tg	23w	3317800	0.87
kinetics 4	tg	24w	2017508	0.85
kinetics 57	wt	24w	3555475	0.91
kinetics 58	wt	24w	2328101	0.89
kinetics 56	wt	24w	3290014	0.79
kinetics 29	tg	8w	2918066	0.86
kinetics 80	tg	8w	2137183	0.88
kinetics 79	tg	8w	3543092	0.91
kinetics 59	wt	8w	3242237	0.83
kinetics 42	wt	8w	3990270	0.92
kinetics 45	wt	8w	4072307	0.91

Supplementary Table 2 Genomic Regions for technical and biological validation

Genomic location for EpiTYPER amplicon	Amplicon name	Spearman r	p-value	R^2	DMR
Techincal validation					
chr2:179895645-179896133	*Cdh4*	-0.49	0.01	0.29	Hypomethylation
chr2:32751370-32751845	*Tor2a*	0.77	< 0.0001	0.60	Hypomethylation
chr2:54436305-54436765	*Galnt13*	0.69	< 0.0001	0.63	Hypermethylation
chr3:55783149-55783648	*Mab21l1*	0.86	< 0.0001	0.76	Hypermethylation
chr4:152126124-152126565	*Espin*	0.62	0.0004	0.67	Hypermethylation
chr5:115446949-115447324	*Msi1*	0.46	0.01	0.35	Hypomethylation
chr6:125145282-125145767	*Iffo1*	0.71	< 0.0001	0.78	Hypermethylation
chr6:42261839-42262170	*Tmem139*	0.47	0.01	0.32	Hypomethylation
chr6:52204531-52204798	*Hoxa5*	0.68	< 0.0001	0.52	Hypomethylation
chr6:90462107-90462398	*Klf15*	0.70	< 0.0001	0.70	Hypermethylation
chr7:98144903-98145367	*Omp*	0.59	0.0006	0.41	Hypomethylation
chr11:119044877-119045298	*Cbx8_intergenic*	0.57	0.001	0.73	Hypermethylation
chr16:7982658-7983035	*Chr16_intergenic*	0.80	< 0.0001	0.56	Hypomethylation
chr17:23679488-23679927	*Cldn6*	0.51	0.004	0.78	Hypermethylation
Biological validation					
chr1:135258389-135258793	*Elf3*	0.76	< 0.0001	0.51	Hypomethylation
chr1:24160814-24160833	*Unc5b*	0.59	0.007	0.53	Hypomethylation
chr2:35282286-35282572	*Gsn*	0.56	0.02	0.83	Hypermethylation
chr5:134946186-134946577	*Cldn4*	0.82	< 0.0001	0.62	Hypomethylation
chr7:141080004-141080320	*Pkp3*	0.75	< 0.0001	0.45	Hypomethylation
chr18:26501830-26501849	*Igfbp6*	0.57	0.001	0.67	Hypermethylation

Supplementary Table 3 RefSeq genes with hypermethylated promoter

Genomic location of DMR	Gene ID	RefSeqID	Expression difference (mouse) [a)]	p-value [b)]	expression difference (human) [c)]
chr2:35282311-35282561	*Gsn*	NM_146120	-2.95	0.00017	-1.64
chr15:102144667-102144917	*Igfbp6*	NM_008344	-2.11	0.00002	-2.08
chr9:65631078-65631328	*Rbpms2*	NM_028030	-1.82	0.00056	-0.79
chr2:91931686-91931936	*Mdk*	NM_001012335	-1.78	0.00158	0.46
chr4:141010703-141010953	*Mfap2*	NM_001161799	-1.76	0.00234	1.58
chr9:37527495-37527745	*Esam*	NM_027102	-1.51	0.00039	0.05
chr6:125145567-125145817	*Iffo1*	NM_001039669	-1.50	0.00011	
chr1:174921625-174921875	*Grem2*	NM_011825	-1.42	0.0028	0.07
chr11:75167750-75168000	*Hic1*	NM_001098203	-1.34	0.00000	
chrX:23364574-23364824	*Klhl13*	NM_026167	-1.32	0.05	-0.89
chr11:75468058-75468308	*Tlcd2*	NM_027249	-1.30	0.00236	
chr4:136247311-136247561	*Tcea3*	NM_011542	-1.27	0.015	-0.76
chr6:145121043-145121293	*Lrmp*	NM_008511	-1.15	0.019	-0.03
chr16:18588802-18589052	*Tbx1*	NM_011532	-1.12	0.05	-0.01
chr15:75893856-75894106	*Naprt1*	NM_172607	-1.12	0.00120	0.59
chr11:95841993-95842243	*Gngt2*	NM_001038664	-1.01	0.00022	0.06
chr14:63944058-63944308	*Sox7*	NM_011446	-1.00	0.005	
chr14:70627074-70627324	*Epb4.9*	NM_001252662	-0.91	0.00637	
chr8:125669863-125670113	*4933403G14Rik*	NM_028908	-0.82	0.00557	
chr11:83849410-83849660	*Hnf1b*	NM_009330	-0.79	0.03	-0.05
chr7:3301951-3302201	*Prkcg*	NM_011102	-0.72	0.00695	0.05
chr7:3666993-3667243	*Leng1*	NM_027203	-0.67	0.02	-0.24
chr6:65381235-65381485	*C130060K24Rik*	NM_175524	-0.67	0.00137	
chr7:130692536-130692786	*Tacc2*	NM_021314	-0.66	0.13	-0.14
chr7:57591335-57591585	*Gabrb3*	NM_008071	-0.66	0.02	-0.07
chr2:74682346-74682596	*Hoxd11*	NM_008273	-0.64	0.10	0.05
chr7:79515616-79515866	*AI854517*	NR_040312	-0.62	0.20	
chrX:81070354-81070604	*Tmem47*	NM_138751	-0.57	0.15	0.22
chr2:54436319-54436569	*Galnt13*	NM_173030	-0.54	0.09	0.84
chr18:37737159-37737409	*Pcdhga9*	NM_033592	-0.51	0.15	
chr7:40899050-40899300	*Vstm2b*	NM_021387	-0.49	0.13	
chr8:18741159-18741409	*Angpt2*	NM_007426	-0.47	0.06	1.28
chr11:45979956-45980206	*Sox30*	NM_173384	-0.46	0.09	0.30
chr7:57386882-57387132	*Gabrg3*	NM_008074	-0.45	0.04	
chr7:57509401-57509651	*Gabra5*	NM_176942	-0.44	0.23	
chr16:31428016-31428266	*Bdh1*	NM_001122683	-0.41	0.29	0.19
chr17:34629479-34629729	*Prrt1*	NM_030890	-0.41	0.11	-0.07
chr6:31565131-31565381	*Podxl*	NM_013723	-0.41	0.27	0.07
chr3:68572420-68572670	*Schip1*	NM_013928	-0.38	0.34	0.72
chrX:155216273-155216523	*Sat1*	NM_009121	-0.36	0.17	
chr8:29219472-29219722	*Unc5d*	NM_153135	-0.35	0.26	-0.20
chr16:4418965-4419215	*Adcy9*	NM_009624	-0.34	0.30	-0.14
chr4:91371538-91371788	*Elavl2*	NM_010486	-0.34	0.30	0.53
chrX:120290484-120290734	*Pcdh11x*	NM_001081385	-0.33	0.16	-0.10
chr16:41533262-41533512	*Lsamp*	NM_175548	-0.32	0.26	
chr5:3845061-3845311	*4932412H11Rik*	NM_172879	-0.32	0.08	
chr9:21990096-21990346	*Rgl3*	NM_023622	-0.30	0.27	
chr11:95841993-95842243	*Abi3*	NM_001163464	-0.28	0.26	0.23
chr19:6926058-6926308	*1700019N12Rik*	NM_001039494	-0.28	0.34	
chr2:152754023-152754273	*Cox4i2*	NM_053091	-0.27	0.27	-0.05
chr1:171360360-171360610	*Klhdc9*	NM_001033039	-0.26	0.33	
chr17:55445750-55446000	*St6gal2*	NM_172829	-0.24	0.16	0.21
chr18:37662063-37662313	*Pcdhga1*	NM_033584	-0.24	0.21	
chr13:34132346-34132596	*Tubb2b*	NM_023716	-0.23	0.44	
chr11:69120344-69120594	*Hes7*	NM_033041	-0.23	0.10	0.00
chr19:31083182-31083432	*Cstf2t*	NM_031249	-0.22	0.33	-0.02
chrX:74270554-74270804	*Rpl10*	NM_052835	-0.20	0.48	-1.29
chr11:120048259-120048509	*Aatk*	NM_001198785	-0.20	0.51	-0.19
chr6:6864858-6865108	*Dlx6as2*	NR_002839	-0.20	0.51	
chr11:11684966-11685216	*Ikzf1*	NM_009578	-0.18	0.30	
chr4:15881210-15881460	*Calb1*	NM_009788	-0.16	0.69	-0.51
chr14:39473342-39473592	*Nrg3*	NM_001190187	-0.15	0.59	0.45
chr3:82876256-82876506	*4930564K09Rik*	NR_040382	-0.14	0.58	
chr8:12947852-12948102	*Mcf2l*	NM_001159485	-0.14	0.57	

chr1:135584935-135585185	*Nav1*	NM_173437	-0.13	0.61	0.10
chr5:138280393-138280643	*Gpc2*	NM_172412	-0.13	0.60	0.52
chr2:74763148-74763398	*Hoxd1*	NM_010467	-0.10	0.55	-0.29
chr13:72816469-72816719	*D730050B12Rik*	NR_046196	-0.10	0.68	
chr9:56865005-56865255	*Cspg4*	NM_139001	-0.09	0.82	-0.16
chr14:15438659-15438909	*Lrrc3b*	NM_146052	-0.08	0.71	-0.17
chr6:113469662-113469912	*Il17rc*	NM_134159	-0.08	0.61	0.12
chr1:32172648-32172898	*Khdrbs2*	NM_133235	-0.05	0.77	
chr1:187607657-187607907	*Esrrg*	NM_001243792	-0.05	0.90	-0.04
chr5:119669000-119669250	*Tbx3*	NM_011535	-0.04	0.87	-1.50
chr7:101818524-101818774	*Phox2a*	NM_008887	-0.03	0.93	-0.13
chr8:84701610-84701860	*Lyl1*	NM_008535	-0.02	0.95	-0.22
chr17:50293046-50293296	*Dazl*	NM_010021	-0.01	0.95	-0.52
chr11:72961653-72961903	*Atp2a3*	NM_001163336	-0.01	0.97	0.05
chr3:86548149-86548399	*Mab21l2*	NM_011839	0.00	0.97	0.03
chr8:99416030-99416280	*Cdh8*	NM_001039154	0.01	0.97	-0.02
chrX:73967794-73968044	*Hcfc1*	NM_008224	0.03	0.86	-0.15
chr18:37747221-37747471	*Pcdhga10*	NM_033593	0.03	0.92	
chr5:128432816-128433066	*Tmem132d*	NM_172885	0.04	0.89	
chr4:120748378-120748628	*Kcnq4*	NM_001081142	0.05	0.93	-0.15
chr6:90462100-90462350	*Klf15*	NM_023184	0.05	0.71	0.00
chrX:66649445-66649695	*Slitrk2*	NM_001161431	0.06	0.81	0.13
chr14:49525613-49525863	*Slc35f4*	NM_029238	0.07	0.73	-0.03
chr11:97995908-97996158	*Arl5c*	NM_207231	0.08	0.79	
chr6:110645837-110646087	*Grm7*	NM_177328	0.08	0.60	-0.07
chr15:85131982-85132232	*Ribc2*	NM_026357	0.08	0.38	0.46
chr1:109982763-109983013	*Cdh7*	NM_172853	0.09	0.56	-0.19
chr1:69826920-69827170	*Spag16*	NM_029160	0.10	0.62	0.25
chr15:85131982-85132232	*Smc1b*	NM_080470	0.10	0.73	1.11
chr15:103011834-103012084	*Hoxc5*	NM_175730	0.11	0.47	0.07
chr3:53486672-53486922	*Stoml3*	NM_153156	0.14	0.59	0.15
chr1:55363275-55363525	*Boll*	NM_001113367	0.19	0.06	0.23
chr1:90202930-90203180	*Cxcr7*	NM_007722	0.22	0.50	-0.32
chr18:37685448-37685698	*Pcdhga4*	NM_033587	0.23	0.24	
chr17:23679308-23679558	*Cldn6*	NM_018777	0.25	0.40	-0.30
chr19:50678214-50678464	*Sorcs1*	NM_001252501	0.31	0.21	-0.21
chr11:96365809-96366059	*Hoxb1*	NM_008266	0.32	0.31	-0.09
chr14:121738378-121738628	*Dock9*	NM_001128307	0.32	0.37	0.20
chr4:123105426-123105676	*Bmp8b*	NM_007559	0.33	0.42	
chrX:36874776-36875026	*Ube2a*	NM_019668	0.42	0.14	0.27
chr11:98349233-98349483	*Ppp1r1b*	NM_144828	0.43	0.49	-0.30
chrX:12160255-12160505	*Bcor*	NM_001168321	0.44	0.18	-0.09
chrX:12761920-12762170	*Med14*	NM_012005	0.45	0.21	-0.22
chr3:66219976-66220226	*Ptx3*	NM_008987	0.48	0.45	0.70
chr11:98386638-98386888	*Pnmt*	NM_008890	0.55	0.05	-1.42
chr11:120608620-120608870	*Npb*	NM_153288	0.63	0.00207	
chr4:155345209-155345459	*Prkcz*	NM_001039079	0.66	0.00006	0.51
chrX:142390725-142390975	*Acsl4*	NM_001033600	0.79	0.07	0.13
chr8:105374538-105374788	*Plekhg4*	NM_001081333	0.88	0.04	
chr17:24206640-24206890	*Tbc1d24*	NM_001163847	0.91	0.01	
chr17:15376986-15377236	*Dll1*	NM_007865	0.94	0.09	-0.99
chr3:84220856-84221106	*Trim2*	NM_030706	0.97	0.03	0.18
chr7:112158999-112159249	*Dkk3*	NM_015814	0.99	0.00904	-1.69
chr5:138280393-138280643	*Stag3*	NM_016964	1.01	0.09	0.50
chr17:23741629-23741879	*Paqr4*	NM_023824	1.35	0.00137	0.13
chr16:89818004-89818254	*Tiam1*	NM_001145887	1.56	0.00003	-0.03
chr7:63444459-63444709	*Otud7a*	NM_130880	2.36	0.00009	-0.37
chr17:23679308-23679558	*Tnfrsf12a*	NM_001161746	2.60	0.00005	0.63
chr1:39720817-39721067	*Rfx8*	NM_001145660	Not covered		
chr1:180332539-180332789	*Gm5069*	NR_003623	Not covered		
chr1:92848494-92848744	*Mir149*	NR_029559	Not covered		
chr3:89183647-89183897	*Fam189b*	NM_001014995	Not covered		
chr3:95240913-95241163	*Gm128*	NM_001024841	Not covered		
chr3:82876256-82876506	*Rbm46*	NM_001146328	Not covered		
chr3:55780359-55780609	*Mab21l1*	NM_010750	Not covered		-0.15
chr3:37896911-37897161	*LOC626410*	NR_040559	Not covered		
chr4:43653502-43653752	*Spag8*	NM_001007463	Not covered		0.25
chr4:98726869-98727119	*L1td1*	NM_001081202	Not covered		-0.66
chr4:145463882-145464132	*Smarca5-ps*	NR_002888	Not covered		
chr4:114907183-114907433	*9130206I24Rik*	NR_030721	Not covered		
chr5:138820004-138820254	*Gm5294*	NM_001195128	Not covered		

chr5:144545803-144546053	*Nptx2*	NM_016789	Not covered		0.40
chr5:57719494-57719744	*4932441J04Rik*	NR_015588	Not covered		
chr6:134981757-134982007	*Apold1*	NM_001109914	Not covered		-0.64
chr6:47943239-47943489	*Zfp783*	NR_027963	Not covered		
chr6:52315726-52315976	*5730457N03Rik*	NR_038163	Not covered		
chr7:144284384-144284634	*Shank2*	NM_001081370	Not covered		
chr7:57386882-57387132	*Gm9962*	NR_033504	Not covered		
chr7:40899050-40899300	*A230077H06Rik*	NR_040329	Not covered		
chr7:14622792-14623042	*Gm18756*	NR_045119	Not covered		
chr8:121085019-121085269	*1110050K14Rik*	NR_045471	Not covered		
chr9:76014949-76015199	*Hmgcll1*	NM_173731	Not covered		
chr10:59223216-59223466	*Sept10*	NM_001024910	Not covered		
chr10:53751281-53751531	*Fam184a*	NM_001081428	Not covered		
chr11:120013446-120013696	*Mir3065*	NR_037225	Not covered		
chr15:36140308-36140558	*Rgs22*	NM_001195748	Not covered		-0.43
chr18:89769226-89769476	*Dok6*	NM_001039173	Not covered		
chr18:37751821-37752071	*Pcdhgb7*	NM_033579	Not covered		
chr18:37694503-37694753	*Pcdhga5*	NM_033588	Not covered		
chr18:37725808-37726058	*Pcdhga8*	NM_033591	Not covered		
chr18:37756018-37756268	*Pcdhga11*	NM_033594	Not covered		
chrUn_JH584304:59822-60072	*Pisd-ps3*	NR_003518	Not covered		
chrX:74270554-74270804	*Snora70*	NR_002899	Not covered		
chrX:12160255-12160505	*2900008C10Rik*	NR_045434	Not covered		
chrX:12761920-12762170	*Gm14634*	NR_045852	Not covered		

[a] Expression data are taken from [90]. Values are reported as the differences between tumor and normal samples in log2 array intensities relative to a pool of whole mouse embryo RNA

[b] Student's t-test, two-sided.

[c] Expression data are taken from [90]. Values are reported as the differences between tumor and normal samples in log2 array intensities relative to a pool of Stratagene universal reference RNA mixed with 1/10 added MCF7 and ME16C RNAs.

Cells with gray background are significantly (p-value < 0.05) downregulated.

Supplementary Table 4 RefSeq genes with hypomethylated promoter

Genomic location of DMR	Gene ID	RefSeqID	Expression difference (mouse) [a]	p-value [b]	Expression difference (human) [c]
chr7:24636377-24636564	*Lypd3*	NM_133743	3.57	0.00002	-0.98
chr13:100616541-100616728	*Marveld2*	NM_001038602	3.12	0.002	0.08
chr15:76457331-76458226	*Scx*	NM_198885	2.97	0.00004	
chr4:127355683-127355994	*Gjb4*	NM_008127	2.72	0.00000003	-0.12
chr11:100260949-100261138	*Krt17*	NM_010663	2.60	0.002	-1.12
chr15:101412163-101412892	*Krt7*	NM_033073	2.56	0.00000054	1.40
chr11:69965053-69965203	*Cldn7*	NM_001193619	2.39	0.00001	-0.10
chr4:140774047-140774234	*Padi4*	NM_011061	2.33	0.00003	-0.19
chr7:141079944-141080170	*Pkp3*	NM_019762	2.26	0.0001	0.11
chr5:31251572-31252211	*Krtcap3*	NM_027221	2.25	0.0003	1.10
chr6:83135716-83136125	*Rtkn*	NM_001136227	2.11	0.00002	0.28
chr11:70323238-70323425	*Alox12e*	NM_145684	1.96	0.03	
chr1:135258308-135258540	*Elf3*	NM_001163131	1.88	0.000005	0.08
chr11:5263062-5263287	*Kremen1*	NM_032396	1.80	0.04	-0.23
chr9:21312083-21312591	*Ap1m2*	NM_001110300	1.75	0.009	-0.68
chr11:121519255-121519443	*Zfp750*	NM_178763	1.62	0.02	
chr16:96235671-96236094	*B3galt5*	NM_001122993	1.62	0.03	-0.06
chr16:4640000-4640413	*Vasn*	NM_139307	1.58	0.00004	0.61
chr15:102003991-102004357	*Krt8*	NM_031170	1.55	0.0004	-0.65
chr9:43311200-43311641	*Trim29*	NM_023655	1.48	0.03	-0.08
chr10:60832976-60833163	*Unc5b*	NM_029770	1.46	0.006	2.10
chr6:113458903-113459053	*Il17re*	NM_001034029	1.38	0.004	-1.48
chr11:99492865-99493102	*Krt23*	NM_033373	1.34	0.17	1.45
chr6:86628202-86628467	*Asprv1*	NM_026414	1.31	0.12	
chr11:106267316-106267466	*Smarcd2*	NM_031878	1.26	0.004	-0.15
chr7:31115151-31115352	*Hpn*	NM_001110252	1.26	0.02	-0.64
chr15:89174219-89174369	*Plxnb2*	NM_001159521	1.25	0.0003	-0.76
chr5:134946095-134946730	*Cldn4*	NM_009903	1.22	0.02	0.29
chr15:76616535-76616685	*Slc39a4*	NM_028064	1.20	0.03	0.94
chr17:25273274-25273424	*Ube2i*	NM_001177609	1.19	0.0003	-0.42
chr4:133601497-133601987	*Sfn*	NM_018754	1.10	0.005	0.01
chr11:120949514-120949701	*Slc16a3*	NM_001038653	1.09	0.02	1.74
chr3:88364477-88364665	*Paqr6*	NM_198410	1.09	0.006	-0.26
chr17:46728667-46729254	*Gnmt*	NM_010321	1.03	0.18	-0.63
chr14:87141776-87142066	*Diap3*	NM_019670	1.02	0.005	
chr19:53040296-53040495	*Xpnpep1*	NM_133216	0.92	0.01	0.27
chr15:75910613-75910913	*Eef1d*	NM_023240	0.85	0.03	-0.36
chr2:167931553-167931703	*Ptpn1*	NM_011201	0.83	0.010	0.16
chr4:118527171-118527416	*2610528J11Rik*	NM_025572	0.82	0.04	
chr19:8897036-8897281	*Ganab*	NM_008060	0.82	0.004	0.67
chr14:55576389-55576578	*Psme1*	NM_011189	0.80	0.005	-0.35
chr17:45733385-45733574	*1600014C23Rik*	NM_028164	0.80	0.06	
chr17:33918366-33918632	*Tapbp*	NM_001025313	0.80	0.01	0.35
chr17:63937835-63938145	*Fert2*	NM_008000	0.71	0.007	
chr9:119562480-119562959	*Scn5a*	NM_001253860	0.70	0.06	0.20
chr5:24427536-24427763	*Slc4a2*	NM_001253892	0.68	0.0001	-0.19
chr18:36663949-36664099	*Eif4ebp3*	NM_201256	0.66	0.02	
chr7:110626431-110626581	*Adm*	NM_009627	0.64	0.26	1.24
chr2:77171675-77172145	*Ccdc141*	NM_001025576	0.63	0.009	
chr11:3540494-3540768	*Smtn*	NM_001159284	0.60	0.09	-0.43
chr18:42053900-42054199	*Sh3rf2*	NM_001146299	0.58	0.10	1.40
chr9:75561161-75561349	*Tmod3*	NM_016963	0.57	0.12	-0.22
chr14:70429027-70429310	*Piwil2*	NM_021308	0.55	0.09	0.07
chrX:72274358-72274583	*Gabre*	NM_017369	0.53	0.24	0.09
chr11:69858652-69858877	*Tnk1*	NM_031880	0.53	0.006	0.15
chr8:9977186-9977374	*Abhd13*	NM_001081119	0.52	0.05	0.32
chr8:24674648-24674873	*Adam18*	NM_010084	0.50	0.01	0.07
chr14:26533368-26533518	*Slmap*	NM_032008	0.49	0.12	-0.64
chr14:21991252-21991402	*Zfp503*	NM_145459	0.45	0.29	
chr7:98145038-98145486	*Omp*	NM_011010	0.45	0.008	-0.21
chr15:76307199-76307558	*Oplah*	NM_153122	0.45	0.05	0.91
chr16:29544864-29545014	*Atp13a4*	NM_001164613	0.44	0.25	-0.33
chr5:139812997-139813184	*Tmem184a*	NM_144914	0.42	0.15	
chr17:47926035-47926224	*Foxp4*	NM_001110824	0.40	0.04	0.20
chr4:133582444-133582631	*Gpn2*	NM_133884	0.40	0.03	0.36
chr8:68276613-68276763	*Sh2d4a*	NM_028182	0.40	0.17	0.07
chr15:76329097-76329362	*Gpaa1*	NM_010331	0.39	0.20	0.13

chr5:139423338-139423565	*Gper1*	NM_029771	0.39	0.06	-0.51
chr9:21337847-21338109	*Slc44a2*	NM_152808	0.39	0.31	-0.12
chr11:78324412-78324562	*Aldoc*	NM_009657	0.36	0.40	-0.17
chr15:101940014-101940239	*Krt79*	NM_146063	0.34	0.23	
chr6:42261973-42262160	*Tmem139*	NM_175408	0.33	0.008	1.23
chr9:106562149-106562339	*Iqcf3*	NM_026645	0.32	0.23	
chr8:3392847-3393246	*Arhgef18*	NM_133962	0.31	0.24	0.17
chr8:9977186-9977374	*Lig4*	NM_176953	0.30	0.50	0.13
chr15:76205758-76206103	*Plec*	NM_001163540	0.26	0.10	0.33
chr7:142899883-142900310	*Th*	NM_009377	0.25	0.10	0.03
chr11:98936204-98936561	*Rara*	NM_001177302	0.24	0.32	0.05
chr9:110624418-110624693	*Nradd*	NM_026012	0.24	0.13	
chr2:168601930-168602163	*Nfatc2*	NM_001037178	0.23	0.20	0.27
chr13:40287758-40287945	*Ofcc1*	NM_172143	0.23	0.40	-0.27
chr15:102103467-102103655	*Tenc1*	NM_153533	0.23	0.62	-0.61
chr11:118442724-118442874	*C1qtnf1*	NM_001204130	0.21	0.32	0.09
chr7:30233312-30233698	*Tbcb*	NM_025548	0.20	0.40	0.21
chr11:96269651-96269801	*Hoxb9*	NM_008270	0.18	0.51	-0.68
chr14:65968928-65969117	*Clu*	NM_013492	0.18	0.63	-1.95
chr8:71688023-71688210	*Insl3*	NM_013564	0.17	0.49	0.14
chr5:32353688-32353838	*Plb1*	NM_172147	0.17	0.35	0.35
chr18:34580715-34580910	*Nme5*	NM_080637	0.17	0.51	-1.19
chr4:149675411-149675607	*Pik3cd*	NM_001164051	0.17	0.09	-0.03
chr5:135000839-135001026	*Wbscr25*	NR_026907	0.17	0.24	
chr1:9601670-9601857	*3110035E14Rik*	NM_178399	0.14	0.53	
chr11:78385885-78386073	*Foxn1*	NM_008238	0.13	0.48	0.20
chr12:59095638-59095825	*Mia2*	NM_177321	0.13	0.59	0.07
chr2:14873795-14873983	*Cacnb2*	NR_045533	0.12	0.60	0.12
chr8:105290443-105290630	*4931428F04Rik*	NM_001166394	0.11	0.72	
chr2:32741419-32742591	*6330409D20Rik*	NM_027529	0.10	0.66	
chr5:135980336-135980607	*Zp3*	NM_011776	0.10	0.50	0.16
chr12:58210486-58210636	*Sstr1*	NM_009216	0.09	0.59	0.14
chr7:46861140-46861328	*Ldhc*	NM_013580	0.08	0.47	-0.56
chr15:100936150-100936624	*Scn8a*	NM_011323	0.07	0.82	-0.19
chr4:43706208-43706395	*Olfr71*	NM_019486	0.06	0.64	
chr5:33104465-33104653	*Slc5a1*	NM_019810	0.05	0.93	0.08
chr5:149411715-149412325	*Medag*	NM_027519	0.05	0.93	
chr11:104186807-104186994	*Sppl2c*	NM_001082535	0.05	0.82	
chr2:120032977-120033127	*Pla2g4b*	NM_145378	0.03	0.86	0.36
chr7:30306593-30306780	*Alkbh6*	NM_198027	0.03	0.88	-0.43
chr8:84044760-84045022	*Rln3*	NM_173184	0.00	1.00	0.04
chr9:108796235-108796385	*Ip6k2*	NM_029634	-0.01	0.98	-0.88
chr7:28691213-28691363	*Fbxo27*	NM_001163702	-0.03	0.90	0.11
chr2:163397212-163397658	*Jph2*	NM_001205076	-0.03	0.82	-0.05
chr16:88562344-88562839	*Cldn8*	NM_018778	-0.04	0.90	-0.84
chr11:99089291-99089478	*Tns4*	NM_172564	-0.04	0.83	-2.14
chr5:114970817-114971005	*Hnf1a*	NM_009327	-0.05	0.84	-0.22
chr9:58741420-58741570	*2410076I21Rik*	NM_028598	-0.06	0.88	
chr11:100269700-100269850	*Krt42*	NM_212483	-0.06	0.79	
chr4:52989285-52989472	*Nipsnap3a*	NM_028529	-0.08	0.76	0.04
chr3:116968340-116968490	*Palmd*	NM_023245	-0.09	0.84	-0.78
chr14:54935173-54935434	*Cmtm5*	NM_026066	-0.10	0.75	0.10
chr6:145048444-145048899	*Bcat1*	NM_001024468	-0.11	0.46	-0.22
chr8:48099294-48099481	*Dctd*	NM_001161516	-0.11	0.53	-0.34
chr11:69686883-69687033	*Tnfsf13*	NM_001159505	-0.13	0.73	-0.24
chr17:31198230-31198488	*Tmprss3*	NM_001163776	-0.17	0.35	0.20
chr5:112391522-112391985	*Asphd2*	NM_028386	-0.18	0.28	-0.04
chr17:57087408-57087783	*Tubb4a*	NM_009451	-0.18	0.68	0.01
chr11:4949263-4949505	*Nefh*	NM_010904	-0.19	0.20	-0.03
chr9:107579125-107579275	*Nat6*	NM_019750	-0.19	0.29	-0.29
chr1:89932541-89932953	*Gbx2*	NM_010262	-0.20	0.26	0.36
chr11:82181445-82181595	*Ccl1*	NM_011329	-0.20	0.47	0.27
chr7:28540939-28541313	*Sycn*	NM_026716	-0.21	0.49	-0.11
chrX:38600594-38600744	*Mcts1*	NM_026902	-0.22	0.47	0.31
chr6:127015494-127015750	*Fgf6*	NM_010204	-0.23	0.29	0.12
chr11:34207857-34208088	*Foxi1*	NM_023907	-0.24	0.46	0.23
chr7:45667807-45668003	*Fut2*	NM_018876	-0.24	0.50	-0.18
chr6:113501774-113502020	*Prrt3*	NM_172487	-0.25	0.18	-0.43
chr5:77114817-77115086	*Hopx*	NM_175606	-0.26	0.44	
chr11:121160590-121160960	*Uts2r*	NM_145440	-0.27	0.22	-1.19
chr4:156228136-156228332	*Plekhn1*	NM_001008233	-0.27	0.70	0.37

chr13:38149954-38150142	*Dsp*	NM_023842	-0.28	0.21	-0.04
chr7:19092583-19092733	*Six5*	NM_011383	-0.30	0.24	0.60
chr5:119832520-119832711	*Tbx5*	NM_011537	-0.30	0.21	0.26
chr15:101491270-101491532	*Krt85*	NM_016879	-0.30	0.13	-0.05
chr15:10952359-10952509	*C1qtnf3*	NM_001204134	-0.30	0.26	-0.04
chr17:34038638-34039065	*Col11a2*	NM_009926	-0.31	0.28	0.40
chr14:118922461-118922950	*Dzip1*	NM_025943	-0.33	0.30	0.61
chr9:115908341-115908491	*Gadl1*	NM_028638	-0.34	0.12	0.11
chr11:96282039-96282189	*Hoxb8*	NM_010461	-0.37	0.18	0.44
chr19:5426400-5426588	*Drap1*	NM_024176	-0.37	0.10	-0.21
chr17:35074114-35074264	*Ly6g6d*	NM_033478	-0.37	0.43	
chr6:125690481-125690722	*Ano2*	NM_153589	-0.38	0.41	-0.08
chr11:78341746-78341937	*Unc119*	NM_011676	-0.38	0.12	-0.15
chr6:52165691-52166054	*Hoxa2*	NM_010451	-0.39	0.28	
chr7:45638724-45639041	*Mamstr*	NM_172418	-0.40	0.34	
chr7:27257234-27257539	*Numbl*	NM_010950	-0.41	0.33	0.38
chr2:170601772-170602085	*4930470P17Rik*	NR_027825	-0.43	0.03	
chr6:138583226-138583413	*Lmo3*	NM_207222	-0.44	0.17	0.25
chr14:70207897-70208047	*Sorbs3*	NM_011366	-0.45	0.03	-0.43
chr17:25471983-25472133	*Tekt4*	NM_027951	-0.47	0.16	0.19
chr1:192189769-192189957	*Kcnh1*	NM_001038607	-0.48	0.17	0.02
chr11:49793227-49793377	*Gfpt2*	NM_013529	-0.48	0.01	0.82
chr10:127077624-127077774	*Agap2*	NM_001033263	-0.49	0.13	-0.29
chr7:45420078-45420313	*Lhb*	NM_008497	-0.50	0.06	-0.11
chr7:126624820-126625007	*Nupr1*	NM_019738	-0.51	0.30	-0.59
chr5:24445657-24445859	*Fastk*	NM_023229	-0.52	0.009	0.09
chr11:99245094-99245283	*Krt222*	NM_172946	-0.58	0.09	
chr7:45240824-45241284	*Cd37*	NM_007645	-0.58	0.002	0.74
chr2:155945167-155945370	*Gdf5*	NM_008109	-0.61	0.09	0.10
chr4:133220894-133221081	*Cd164l2*	NM_027152	-0.64	0.06	
chrX:73659919-73660108	*Pnck*	NM_001199351	-0.65	0.0005	0.00
chr10:41886559-41886709	*Sesn1*	NM_001013370	-0.66	0.05	-1.04
chr15:101924561-101924828	*Krt4*	NM_008475	-0.67	0.03	-0.29
chr12:8598791-8598941	*Slc7a15*	NM_177802	-0.70	0.09	
chr10:78466662-78466812	*Pdxk*	NM_172134	-0.70	0.11	-0.13
chr11:96341911-96342400	*Hoxb3*	NM_010458	-0.73	0.003	0.17
chr9:50727354-50727544	*Dixdc1*	NM_178118	-0.77	0.003	
chr4:155655249-155655478	*Mmp23*	NM_011985	-0.78	0.002	
chr19:4229814-4229964	*Pold4*	NM_027196	-0.79	0.002	-0.32
chr11:112711264-112711414	*BC006965*	NR_024085	-0.81	0.19	
chr7:24480418-24480610	*Cadm4*	NM_153112	-0.83	0.002	
chr7:45725346-45725496	*Spaca4*	NM_027055	-0.90	0.15	0.27
chr9:50752151-50752338	*Hspb2*	NM_001164708	-1.05	0.0009	0.13
chr10:116177244-116177394	*Ptprr*	NM_001161838	-1.06	0.0010	-0.17
chr11:69397660-69397887	*Tmem88*	NM_025915	-1.16	0.00003	-0.36
chr11:96301942-96302131	*Hoxb5*	NM_008268	-1.20	0.002	0.04
chr3:94371712-94371862	*Rorc*	NM_011281	-1.21	0.002	-0.07
chr5:30920635-30921209	*Khk*	NM_008439	-1.26	0.00009	0.32
chr2:113828891-113829191	*Scg5*	NM_009162	-1.34	0.001	-0.69
chr11:99374660-99374965	*Krt28*	NM_027574	-1.45	0.04	
chr6:52206332-52206737	*Hoxa5*	NM_010453	-1.54	0.006	-1.66
chr15:76724917-76725202	*C030006K11Rik*	NM_145472	-1.65	0.003	
chr9:50751281-50751786	*Cryab*	NM_009964	-1.65	0.001	-0.31
chr2:30904811-30904961	*Ptges*	NM_022415	-1.71	0.006	0.38
chr7:44533648-44533798	*Spib*	NM_019866	-1.77	0.04	-0.27
chr15:76696854-76697090	*Gpt*	NM_182805	-1.83	0.0003	0.08
chr14:67231822-67232168	*Ebf2*	NM_010095	-1.91	0.00008	-0.28
chr1:74713575-74713913	*Cyp27a1*	NM_024264	-2.41	0.00010	0.23
chr7:137315682-137315832	*Ebf3*	NM_001113414	-2.44	0.0002	
chr5:116422343-116422531	*Hspb8*	NM_030704	-2.47	0.005	-0.59
chr13:113663203-113663442	*Hspb3*	NM_019960	-3.10	0.00006	0.16
chr7:44525122-44525353	*Mybpc2*	NM_146189	-3.60	0.004	0.40
chr16:95929025-95929215	*1600002D24Rik*	NR_040484	Not covered		
chr19:53076090-53076287	*1700054A03Rik*	NR_045320	Not covered		
chr9:40333366-40333516	*1700110K17Rik*	NR_040728	Not covered		
chr5:66191007-66191326	*1700126H18Rik*	NR_040695	Not covered		
chr8:106133567-106133754	*1810019D21Rik*	NR_040344	Not covered		
chr8:70774662-70774971	*2010320M18Rik*	NR_029440	Not covered		
chr7:29247728-29248246	*2200002D01Rik*	NM_028179	Not covered		
chr3:116968340-116968490	*4930455H04Rik*	NR_040596	Not covered		
chr2:68656660-68656810	*4932414N04Rik*	NM_183113	Not covered		

chr8:104794354-104794504	*4932416K20Rik*	NM_001002775	Not covered		
chr15:101491270-101491532	*5430421N21Rik*	NM_001201323	Not covered		
chr5:64044947-64045097	*5830416I19Rik*	NR_045384	Not covered		
chr8:11005662-11005916	*9530052E02Rik*	NR_046017	Not covered		
chr13:44216615-44216802	*A330076C08Rik*	NR_045088	Not covered		
chr9:51875614-51875967	*Arhgap20os*	NR_033560	Not covered		
chr2:158511087-158511317	*Arhgap40*	NM_001145015	Not covered		
chr10:80370475-80370662	*Bc1*	NR_038088	Not covered		
chr17:87194749-87194936	*C330024C12Rik*	NR_046016	Not covered		
chr6:90430160-90430310	*Ccdc37*	NM_173775	Not covered		
chr12:112721582-112721732	*Cep170b*	NM_001024602	Not covered		
chr7:66743292-66743442	*Cers3*	NM_001164201	Not covered		
chr11:113709429-113709616	*Cpsf4l*	NM_001164532	Not covered		
chr5:143758213-143758402	*D130017N08Rik*	NR_015486	Not covered		
chr15:103147963-103148113	*D930007P13Rik*	NR_045743	Not covered		
chr1:153414109-153414337	*E330020D12Rik*	NR_033736	Not covered		
chr7:141459454-141459702	*Efcab4a*	NM_001025103	Not covered		
chr1:91322312-91322500	*Espnl*	NM_001033292	Not covered		
chr2:151973248-151973547	*Fam110a*	NM_028666	Not covered		
chr7:30971965-30972190	*Fam187b*	NR_038860	Not covered		
chr7:44494816-44494966	*Fam71e1*	NM_028169	Not covered		
chr15:57985957-57986154	*Fam83a*	NM_173862	Not covered		0.27
chr2:155834558-155834783	*Fam83c*	NM_027788	Not covered		0.18
chr11:61684726-61685209	*Fam83g*	NM_178618	Not covered		
chr4:141982150-141982300	*Fhad1os1*	NR_040672	Not covered		
chr15:101053918-101054399	*Fignl2*	NM_001214911	Not covered		
chr11:83429160-83429400	*Gas2l2*	NM_001013759	Not covered		-0.09
chr4:57171555-57171743	*Gm12530*	NR_040669	Not covered		
chr7:127589512-127589706	*Gm166*	NM_001033040	Not covered		
chr1:136696240-136696465	*Gm16880*	NR_037986	Not covered		
chr13:63296519-63296709	*Gm16907*	NR_045794	Not covered		
chr17:25273274-25273424	*Gm17801*	NR_027452	Not covered		
chr19:41747378-41747634	*Gm19424*	NR_040320	Not covered		
chr1:133270030-133270218	*Gm19461*	NR_037984	Not covered		
chrX:102933791-102934335	*Gm5126*	NR_026596	Not covered		
chr3:37724358-37724589	*Gm5148*	NM_198657	Not covered		
chrX:56549472-56549622	*Gm648*	NM_001033372	Not covered		
chr1:71888225-71888412	*Gm8883*	NR_027658	Not covered		
chrX:37211404-37211695	*Gm9*	NM_001033234	Not covered		
chr19:4042222-4042420	*Gstp2*	NM_181796	Not covered		
chr6:52243420-52243652	*Hoxa11os*	NR_015348	Not covered		
chr19:47012639-47012826	*Ina*	NM_146100	Not covered		0.25
chr7:44707895-44708095	*Izumo2*	NM_029317	Not covered		
chr7:143296298-143296487	*Kcnq1ot1*	NR_001461	Not covered		
chr11:99627769-99627959	*Krtap4-1*	NM_001048196	Not covered		
chr2:74725490-74725679	*Mir10b*	NR_029566	Not covered		
chr8:107550028-107550215	*Mir140*	NR_029553	Not covered		
chr6:124718604-124718946	*Mir141*	NR_029554	Not covered		
chr11:96264243-96264474	*Mir196a-1*	NR_029721	Not covered		
chr2:180389141-180389291	*Mir1a-1*	NR_029528	Not covered		
chr6:124718604-124718946	*Mir200c*	NR_029792	Not covered		
chr2:25440727-25440877	*Mir3087*	NR_037270	Not covered		
chr8:23144626-23144981	*Mir3107*	NR_037293	Not covered		
chr7:142655625-142655775	*Mir483*	NR_030251	Not covered		
chr1:34432869-34433098	*Mir5103*	NR_039562	Not covered		
chr12:109563627-109563889	*Mir770*	NR_030427	Not covered		
chr16:93369678-93369957	*Mir802*	NR_030429	Not covered		
chr6:52215862-52216088	*Mira*	NR_045199	Not covered		
chr10:110000033-110000183	*Nav3*	NM_001081035	Not covered		-0.36
chr15:98497651-98497878	*Olfr279*	NM_001001807	Not covered		
chr7:140176391-140176726	*Olfr523*	NM_146518	Not covered		
chr11:66912454-66912645	*Pirt*	NM_178656	Not covered		
chr12:76417536-76417964	*Ppp1r36*	NM_001163103	Not covered		
chr11:59377440-59377627	*Prss38*	NM_001045521	Not covered		
chr1:34560630-34561039	*Prss40*	NM_009356	Not covered		
chr2:90481104-90481254	*Ptprj*	NM_001135657	Not covered		0.27
chr7:139247438-139247666	*Pwwp2b*	NM_001033206	Not covered		
chr19:42752289-42752439	*Pyroxd2*	NM_029011	Not covered		
chr4:156130517-156131098	*Rnf223*	NM_001220499	Not covered		
chr1:157525749-157525936	*Sec16b*	NM_001159986	Not covered		
chr2:151999850-152000093	*Slc52a3*	NM_001164819	Not covered		

chr2:25846642-25846792	*Sohlh1*	NM_001001714	Not covered		
chr8:124896195-124896561	*Sprtn*	NM_001111141	Not covered		
chr16:10395025-10395275	*Tekt5*	NM_001099275	Not covered		
chr11:69879894-69880186	*Tmem95*	NM_001195710	Not covered		
chr7:128004523-128004931	*Trim72*	NM_001079932	Not covered		
chr7:140835481-140835669	*Urah*	NM_029821	Not covered		
chr12:86988267-86988455	*Zdhhc22*	NM_001080943	Not covered		-2.29
chr14:55071861-55072060	*Zfhx2os*	NR_004444	Not covered		
chr2:77703137-77703428	*Zfp385b*	NM_001113399	Not covered		
chr7:39449748-39449898	*Zfp939*	NM_001243021	Not covered		
chr2:165858239-165858468	*Zmynd8*	NM_001252585	Not covered		

[a] Expression data are taken from [90]. Values are reported as the differences between tumor and normal samples in log2 array intensities relative to a pool of whole mouse embryo RNA

[b] Student's t-test, two-sided.

[c] Expression data are taken from [90]. Values are reported as the differences between tumor and normal samples in log2 array intensities relative to a pool of Stratagene universal reference RNA mixed with 1/10 added MCF7 and ME16C RNAs.

Cells with gray background are significantly (p-value < 0.05) upregulated.

Supplementary Table 5 Samples and quality control for ChIP-seq data

Sample	Antibody	# of unique reads
Tumor 1	H3K4me1	23795321
Tumor 1	H3K4me3	22796348
Tumor 1	H3K27ac	23374206
Tumor 1	H3K27me3	26473994
Tumor 1	Input (5%)	22487101
Tumor 2	H3K4me1	20677813
Tumor 2	H3K4me3	17440855
Tumor 2	H3K27ac	23725681
Tumor 2	H3K27me3	25312350
Tumor 2	Input (5%)	24670676
Tumor 3	H3K4me1	29927526
Tumor 3	H3K4me3	17876458
Tumor 3	H3K27ac	27201814
Tumor 3	H3K27me3	22750426
Tumor 3	Input (5%)	23184949
MECs	H3K4me1	25823216
MECs	H3K4me3	21928903
MECs	H3K27ac	29116445
MECs	H3K27me3	28295966
MECs	Input (5%)	18823822

Supplementary Table 6 Gene set enrichment for genes with hypomethylated tumor-specific active promoter states

Gene ID	MCBRYAN_PUBERTAL_BREAST_4_5WK_UP	MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	LANDIS_ERBB2_BREAST_TUMORS_324_UP	GRAESSMANN_APOPTOSIS_BY_DOXORUBICIN_DN	WONG_ADULT_TISSUE_STEM_MODULE	KOINUMA_TARGETS_OF_SMAD2_OR_SMAD3	CUI_TCF21_TARGETS_2_DN	GROSS_HYPOXIA_VIA_ELK3_AND_HIF1A_UP	PEREZ_TP63_TARGETS	DACOSTA_UV_RESPONSE_VIA_ERCC3_DN	LANDIS_BREAST_CANCER_PROGRESSION_UP	PEREZ_TP53_TARGETS	LIM_MAMMARY_STEM_CELL_DN	LIM_MAMMARY_STEM_CELL_UP	Genes without enrichment
SHROOM3	x	x	x				x				x				PLXNB2
CD9	x	x	x												SGMS2
EGR2	x	x			x				x			x		x	CYP2D6
LAD1	x	x				x									IP6K2
SOX9	x	x								x					AXIN2
ST14	x	x											x		LRG1
IRX1	x	x													GNS
FAM134B	x		x	x	x		x				x	x			PGLYRP1
DUSP6	x		x			x	x							x	PFDN1
CLDN7	x		x			x					x		x		VIL1
TFAP2C	x		x			x					x				SPEF1
KRT8	x		x			x						x	x		SLC25A29
CLDN3	x		x								x		x		IFITM10
SPINT1	x		x								x		x		TMEM154
SLC44A2	x			x											LNX1
GATA3	x				x	x	x	x				x	x		PWWP2B
THBS1	x				x	x	x							x	DIXDC1
COL7A1	x					x								x	TRPM1
KRT14	x					x								x	FAM179A
SLC2A1	x							x							ATP13A4
PTX3	x								x	x					MYH4
KRT18	x												x		NUFIP1
CLDN8	x												x		SRRM2
HEBP2	x														RALBP1
VDR		x	x		x				x		x	x			CD37
PPAP2B		x		x	x		x	x	x	x				x	GAS2L1
FOXC1		x		x	x		x					x			PICK1
NFIB		x		x	x					x					FANCE
IGFBP5		x		x	x										LPHN2
NFIX		x		x	x										TNFRSF25
KCTD1		x		x		x									BCL11A
TCF7L2		x		x						x					NUMBL

GAB1		x		x								x			TOR2A
DLX5		x		x											PTPLB
GLI2		x		x											ZBP1
CTGF		x			x	x	x	x		x				x	FAM83H
KLF4		x			x		x	x							LOC646862
MYCN		x			x										FZD9
CMTM8		x			x										SMARCD2
AFAP1L2		x				x	x								B3GALT5
SMAD6		x				x			x	x		x			ETV4
CASZ1		x				x			x				x		KHK
SIGIRR		x				x							x		SORBS3
MMP23B		x					x								NRBF2
KCTD11		x						x							CREBL2
SPRY4		x						x							RPS6KA4
UNC5B		x							x			x			COMMD10
ARHGEF17		x							x			x			S1PR2
BMP7		x										x		x	SDF2L1
WIF1		x												x	GPR179
AMIGO2		x													PAFAH2
SHOX2		x													GDF5
CDC42EP1		x													TRIM46
ACE		x													MYBPC2
SLC16A3		x													RIOK1
DDX6			x	x											PARK7
FRRS1			x	x											KCNS1
LRRFIP1			x		x		x			x	x				TCEB3
AIM1			x		x						x				SLC9A4
TIAM1			x							x					MGAT4A
STK39			x							x					FADS6
ITPR2			x												SLC45A1
PRKCDBP				x	x			x							TNFRSF11A
NRIP1				x	x					x					MESP1
RHOU				x	x										SIX5
PBX1				x	x										MPO
RAB4A				x		x									GPR37L1
PBX2				x			x		x						PITHD1
RREB1				x			x						x		HCCS
BTBD3				x			x								MEN1
TANC1				x			x								CBR4
VWA1				x				x				x			CCDC94

TRAK1				x						x		x		
AHDC1				x						x		x		
ZFHX3				x						x				
NR3C1				x						x				
MBD2				x						x				
LGALS7				x										x
SLC4A3				x										x
FAM3C				x										
PPP2R1B				x										
DIP2B				x										
AP3M2				x										
LRRC58				x										
LTA4H				x										
SNX9				x										
SRPRB				x										
CYB5B				x										
EXOC6				x										
IQGAP3				x										
D2HGDH				x										
COMMD2				x										
TINAGL1					x	x						x		x
DAB2IP					x	x						x		
LTBP2					x	x								
VLDLR					x					x				
BASP1					x					x				
LRP4					x									x
NUPR1					x									
KAZALD1					x									
RIN2						x	x			x				
RBPJ						x		x		x				
GPRC5A						x		x						
TFAP2A						x							x	
ZFP36L1						x								
ABHD2						x								
FAM83A						x								
BCL9L						x								
CD109						x								
MYH9						x								
CRYAB							x		x			x		x
PTPRJ							x		x			x		

CD80
SKOR1
RPLP0
CIT
CATSPER1
C8orf82
HNRNPA3
SLC35B3
FAN1
TMEM180
LONRF1
DHX8
MRPS23
TNFRSF13B
RAD9B
BLOC1S3
KRTCAP3
TMEM17
TRIM72
NCKAP5L
KRTAP11-1
SLC15A5
LMOD3
OFCC1
MIA2
A3GALT2P
MSL3P1
HSBP1L1

SCHIP1							x			x				x
LMO7							x					x		
MYO6							x					x		
RAPGEF5							x					x		
PHF17							x							
SEC11C							x							
WWC2							x							
LPHN3							x							
TPI1								x						
XPNPEP1								x						
SYNJ2									x	x				
LUZP1									x			x		x
WNT7B									x			x		
HS6ST1									x			x		
ARID1B									x			x		
RAB11FIP3									x					
WNT5B									x					
GAPVD1										x				
RYBP										x				
PLCB4										x				
MED13L										x				
GNMT												x		
SPATA18												x		
SMAD5												x		
CELSR2												x		
FAM83G												x		
OPLAH													x	
LYPD3													x	
TCIRG1													x	
NEU1													x	
SLC28A3													x	
PYROXD2													x	
ZDHHC12													x	
TSHZ2														x
HSPB2														x
SMTN														x
PRICKLE1														x

Supplementary Table 7 Gene set enrichment for genes with hypermethylated MEC-specific inactive/poised promoter states

Gene ID	BENPORATH_SUZ12_TARGETS	BENPORATH_EED_TARGETS	MEISSNER_BRAIN_HCP_WITH_H3K4ME3_AND_H3K27ME3	BENPORATH_PRC2_TARGETS	MEISSNER_NPC_HCP_WITH_H3K4ME2_AND_H3K27ME3
CASZ1	X	X	X	X	X
ITPKA	X	X	X	X	
SHOX2	X	X	X	X	
LHX2	X	X	X	X	
SOX7	X	X	X	X	
ESPN	X	X		X	
ADCYAP1	X	X		X	
GRM7	X	X		X	
EPHA5	X	X		X	
GABRA2	X	X		X	
NRG1	X	X		X	
HOXC4	X	X		X	
DOK6	X	X		X	
OSR1	X	X		X	
SORCS1	X	X		X	
FAM84A	X	X		X	
POU3F4	X	X		X	
LONRF3	X	X		X	
MAB21L2	X	X		X	
SLC18A3	X	X			X
HOXD11	X	X			X
CDH11	X				X
LRRC3B	X				
PCDH11X	X				
SLITRK5	X				
SLITRK2	X				
FOXC1		X	X		
ELAVL2		X			
MORF4L2		X			
DPP10			X		X
RAB11FIP1			X		X
VIPR1			X		X
PLD5			X		X

Genes without enrichment	
FGF12	SAT1
OTUD7A	MYH14
AATK	TCFL5
RAPGEFL1	EXTL1
ANK1	PRSS16
GRIK2	USP10
TMEM132D	UBQLN2
MARCH11	DCTN1
VAV3	TXLNG
PNMT	ARMC2
ACTN2	RGS22
CLDN6	RIBC2
WNT7B	MID1IP1
NETO1	SLC35F4
L1TD1	ATG4A
AJAP1	ZMYM2
NFAM1	KHDRBS2
MPP3	UPF3B
SLC6A8	MAGEE1
SHANK2	WDR44
MECP2	FAM19A5
DLGAP1	UBOX5
GABRB3	SGCZ
IL10	RBM46
SSTR4	
ACSL4	
NALCN	
HRASLS5	
DKK3	
IGFBP6	
SPATA20	
RPL10	
HUWE1	

NRG3			X			NKRF
ALPL			X			LRRN1
FAM46B			X			MED14
CPXM1			X			ZFHX4
PRKCZ			X			PTPRD
CBLN2			X			SP6
VSTM2B			X			UBE2T
TUBB2B			X			ARTN
PODXL			X			TJP1
NR4A2			X			UBE2A
CDKN2A			X			CLDN9
PPP1R1B			X			VMA21
FLT1			X			WIPF3
CSPG4			X			EPHB2
ST6GAL2					X	AXIN2
FAM155A					X	TIAM1
PRR16					X	TMEM132C
GALNT13					X	TNFRSF18
KCTD8					X	TMIE
GABRB1					X	ADAM33
UNC5D					X	COX4I2

Supplementary Table 8 Gene set enrichment for genes with hypomethylated tumor-specific enhancer states

Gene ID	MARTINEZ_RB1_AND_TP53_TARGETS_DN	MARTINEZ_TP53_ TARGETS_DN	MEISSNER_BRAIN_HCP_WITH _H3K4ME3_AND_H3K27ME3	BLALOCK_ALZHEIMERS_ DISEASE_UP	MARTINEZ_RB1_ TARGETS_DN	JAEGER_METASTASIS_DN	WONG_ADULT_TISSUE_STEM_ MODULE	DODD_NASOPHARYNGEAL_ CARCINOMA_UP	PEREZ_TP63_TARGETS	SMID_BREAST_CANCER_ BASAL_UP	MCBRYAN_PUBERTAL_ BREAST_4_5WK_UP
SOX9	X	X	X	X	X					X	X
MYO10	X	X	X	X	X					X	
ST14	X	X	X		X	X		X			X
VDR	X	X	X		X		X		X		
FZD7	X	X	X		X					X	
GJB2	X	X	X		X						
LTBP2	X	X		X	X		X				
PTTG1IP	X	X		X	X		X				
RPL37A	X	X		X	X						X
PTPRF	X	X		X	X						
NAV2	X	X		X							
TCF20	X	X		X							
KRT23	X	X			X	X		X		X	
CTNNBIP1	X	X			X	X					
CXXC5	X	X			X		X	X			
PMEPA1	X	X			X		X				
DOPEY2	X	X			X			X			
IFFO2	X	X			X			X			
ELN	X	X			X					X	
RAB4A	X	X			X						
LAMA5	X	X			X						
POLG	X	X			X						
CELSR2	X	X			X						
ACVR2B	X	X			X						
SCYL1	X	X			X						
HOPX	X	X				X		X			
KRT17	X	X				X				X	
ERRFI1	X	X					X				X

Genes without enrichment	
TRIO	SLC45A1
DCLK1	HHAT
PLEKHA1	SLC25A31
SLC25A42	DBX1
SH3BP4	GBX1
SLC41A2	YWHAG
MGAT3	BPGM
TBC1D16	LAPTM4A
CDC42BPA	VPS8
SEC14L1	ARL5C
AHDC1	GPR179
PEX14	ANO9
FBXW11	TM4SF20
KIAA0182	TMEM44
PTPN1	CARS2
TIAM1	GINS2
CHSY1	RIOK1
CTIF	PSMB8
MED13L	SNORA7A
STK39	BCL7A
ARHGEF10L	CRYBB1
NRM	TMEM185B
PARD6B	EXOG
GSR	CHRDL2
TANC1	BCL2L14
VAV3	ARHGEF19
IQGAP3	C1orf198
POLD2	GSG1

WWC1	X	X						X			X
OVOL1	X	X							X		
NHSL1	X	X									X
ENC1	X	X									
KRT79	X	X									
ERO1L	X	X									
SRRM1	X	X									
CRYBA4	X	X									
LTBP1	X				X		X			X	
TNNT2	X				X						
FGFR3			X	X		X					
ITGB5			X	X			X				
TCF7L2			X	X							
TNS3			X	X							
KIT			X			X	X		X	X	
KLF4			X			X	X				
ASS1			X			X		X	X		
CDH1			X			X					X
CMTM8			X				X	X			
GRHL3			X					X	X		
CASZ1			X					X	X		
EZR			X					X			X
RUNX1			X					X			X
ADAMTS8			X					X			
TRNP1			X					X			
CRISPLD2			X						X		
FOXP4			X						X		
FOSL1			X							X	
CSPG4			X							X	
PTPN14			X							X	
SHROOM3			X								X
ROR1			X								
DYSF			X								
TESC			X								
WNT4			X								
KIAA1804			X								
AFAP1L2			X								
NFATC2			X								
LFNG			X								
ACE			X								

FARP1	WNT8A
LTA4H	YBX2
POLM	GFI1B
WDSUB1	NDUFB9
PSAPL1	DERL3
DNASE1L2	MIR34A
ETV4	FCHSD1
MFI2	GALNT9
CMIP	CILP2
ITGB6	CCDC24
TNFRSF25	TEX22
LDB3	SNORD12
ZBP1	
DAXX	
PDXP	
EIF2B2	
XAB2	
ZDHHC7	
SYT7	
MRPL38	
MEN1	
SYT8	
DMRTB1	
LGI4	
TMEM174	
NKX6-3	
ACOT7	
GRAMD4	
GAA	
AMN	
BST1	
SOAT2	
ETV6	
ASPRV1	
KIF26A	
APEH	
PAQR6	
PLEKHJ1	
PAFAH2	
BMP8B	

SCN5A			X								
MAP3K6			X								
E2F2			X								
GLIS1			X								
SCUBE3			X								
SFN				X		X				X	
COL7A1				X		X					X
MYO6				X		X					
ITGB4				X		X					
LRRFIP1				X			X	X			
SOCS3				X			X		X	X	
ARPC1B				X			X				
LRP4				X			X				
RHOQ				X			X				
WWP2				X				X			
WNT7B				X					X		
ARHGAP26				X					X		
TCF7L1				X						X	
TGIF1				X							X
ZFHX3				X							
SPSB1				X							
BIN3				X							
TFEB				X							
RYBP				X							
MYO1E				X							
ACVR1B				X							
FDFT1				X							
SEMA3B				X							
MTMR11				X							
SLC2A9				X							
SCARB1				X							
RBM15				X							
AP3S2				X							
ADCY3				X							
TMBIM6				X							
CSNK1E				X							
VGLL4				X							
UNG				X							
DUT				X							
FANCE				X							

CRTC2
PGLYRP1
FBL
RPL19
PYROXD2

HLCS				X							
ENGASE				X							
FOS					X		X				X
TRIB1					X						
NDST2					X						
ETS2						X	X				
EVPL						X		X	X		
TRIM29						X		X		X	
AIM1L						X		X			
DNASE1L3						X		X			
LYPD3						X					
MSI2							X	X			
EPHB3							X		X	X	
ZFPM1							X		X		
ALDOC							X				X
IGF2R							X				
EPHB4							X				
RILPL1							X				
IL1B							X				
NEDD9							X				
PADI4							X				
SLC41A1							X				
EYA2							X				
MICAL3								X			
CLMN								X			
MYOF								X			
TRAK1								X			
OCEL1								X			
MAPK3								X			
TMEM121								X			
ELF3								X			
PCDH1								X			
FAM83A								X			
TTC39B								X			
SLC25A29								X			
SLC27A1								X			
C22orf23								X			
ACTL7B								X			
DTX4								X			
CROCC								X			

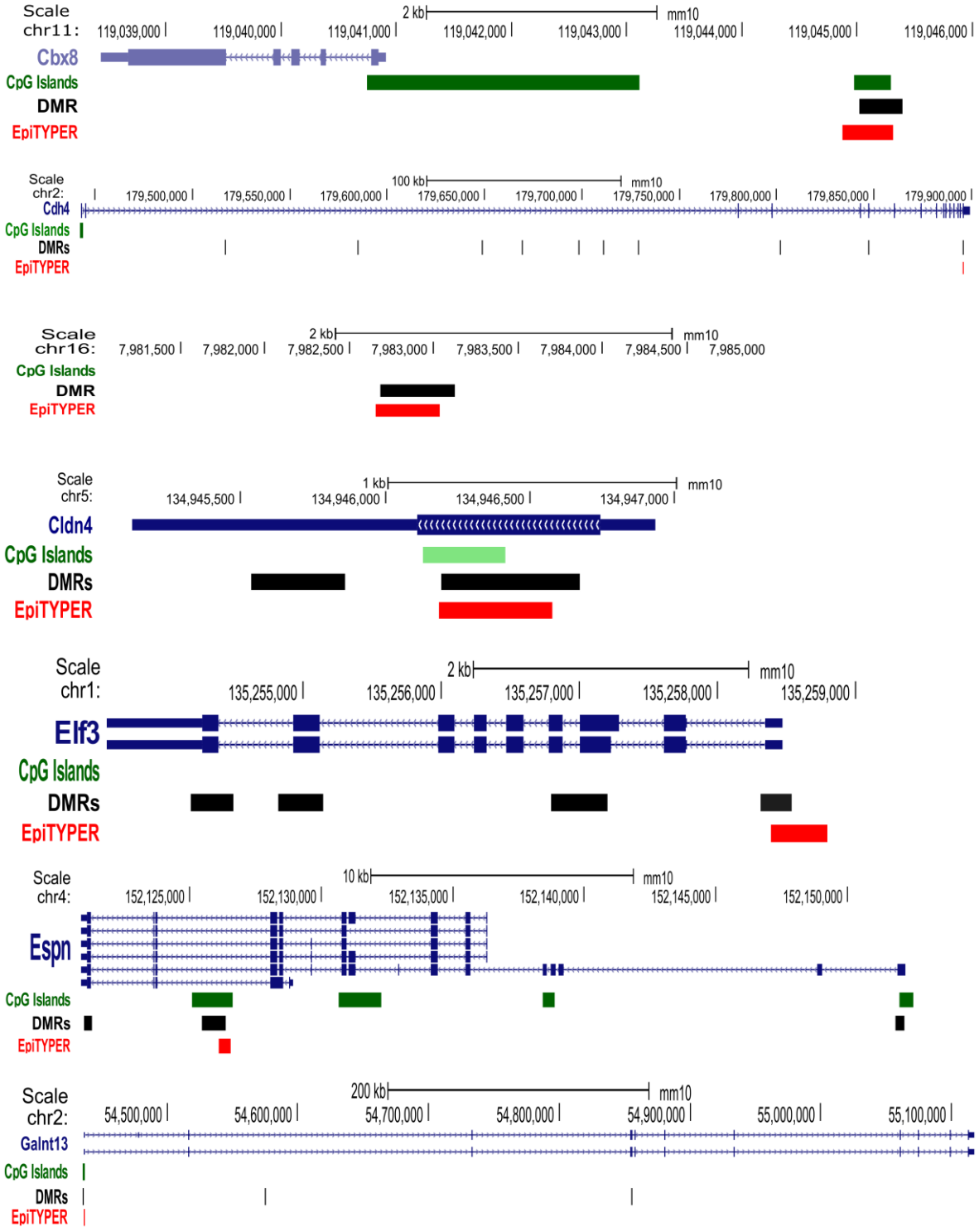
TGM1								X			
ST3GAL3								X			
TMEM154								X			
C2orf54								X			
KATNB1								X			
FAM129C								X			
ATP13A4								X			
RUNX3									X	X	
WNT5B									X	X	
SOX8									X		
PLCD1									X		
FAM43A									X		
ABCB9									X		
ZCCHC24									X		
CD101									X		
NFIL3										X	X
COL11A2										X	
NRTN										X	
SCHIP1										X	
ARTN										X	
FZD9										X	
KIF1A										X	
MFAP2										X	
EDN2										X	
NEK2										X	
TGFB3											X
COL16A1											X
ATP1A1											X
BSPRY											X

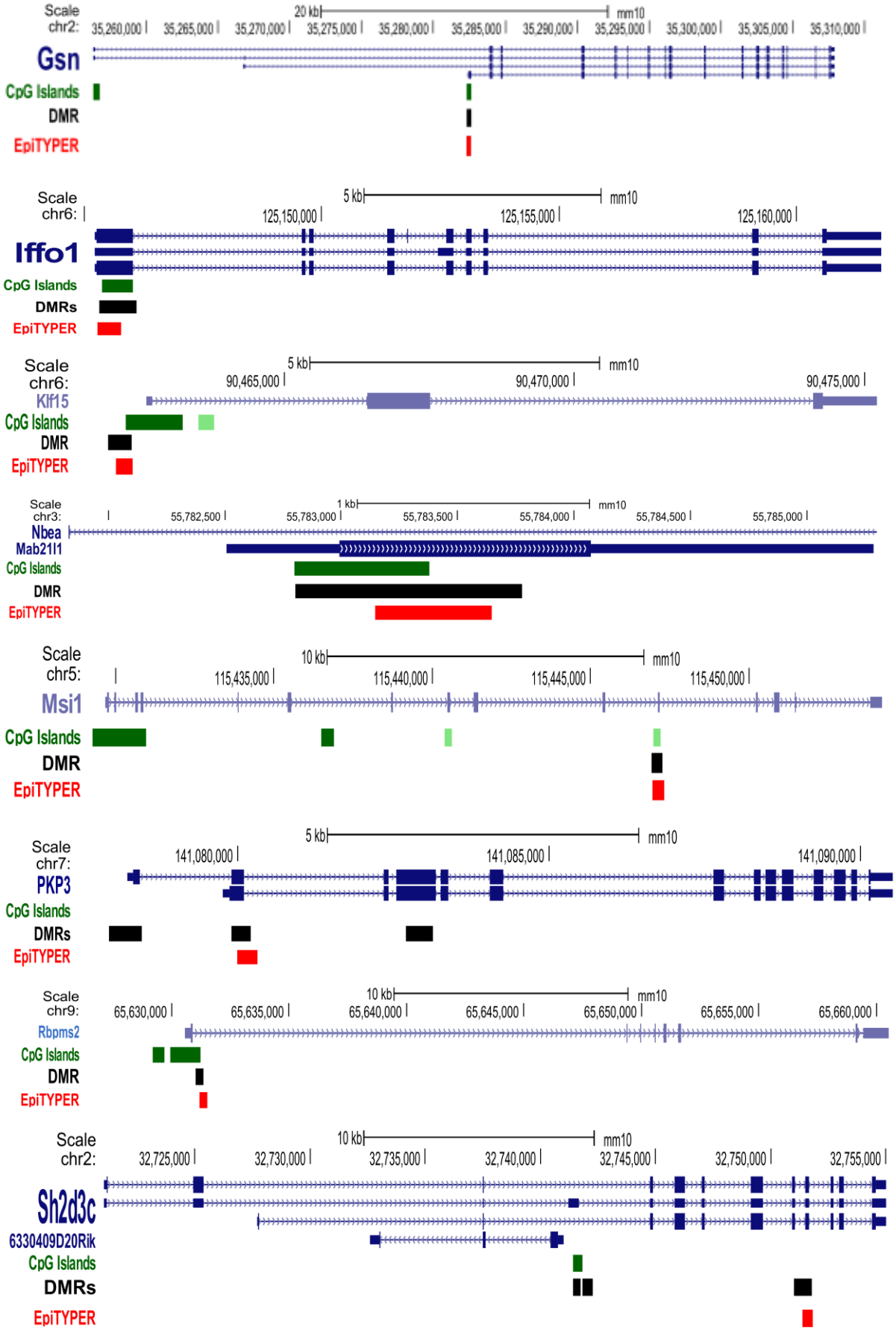
Supplementary Table 9 Gene set enrichment for genes with MEC-specific hypermethylated enhancer states

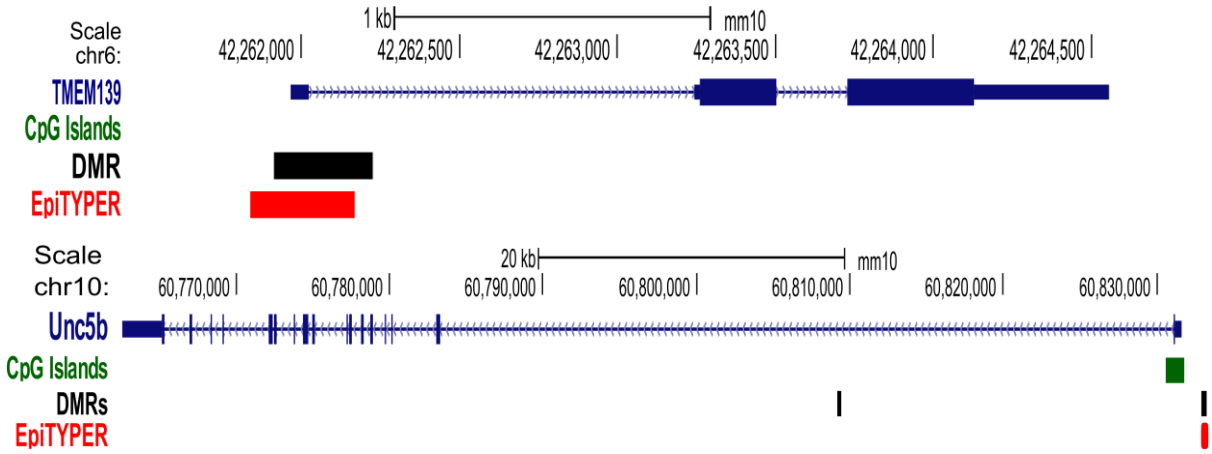
Gene ID	CHYLA_CBFA2T3_TARGETS_UP	RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_DN	LIM_MAMMARY_STEM_CELL_UP	RODRIGUES_THYROID_CARCINOMA_ANAPLASTIC_DN	WEST_ADRENOCORTICAL_TUMOR_DN	MULLIGHAN_MLL_SIGNATURE_2_UP	DACOSTA_UV_RESPONSE_VIA_ERCC3_DN	MULLIGHAN_MLL_SIGNATURE_1_UP	HAN_SATB1_TARGETS_UP	BOYLAN_MULTIPLE_MYELOMA_CLUSTER_DN	JOHNSTONE_PARVB_TARGETS_3_UP	ONKEN_UVEAL_MELANOMA_UP	XU_CREBBP_TARGETS_DN
RBPMS	X	X	X	X	X		X						
ARHGEF3	X	X			X	X		X					
ID1	X			X			X						
MYO7A	X				X								
MGST2	X					X		X				X	
DNMBP	X						X				X		
TRERF1	X								X				
ESAM	X										X		
RHOBTB1	X												
STOX2	X												
TMEM2	X												
GSN		X	X	X					X	X	X		X
TCF4		X	X	X					X				
TGFBR3		X	X				X						
GNAS		X		X								X	
NSD1		X		X									
NR2F2		X			X		X				X		
CDC42EP4		X			X						X		
CYTH1		X					X						X
TSC22D1		X							X			X	
MBP		X											
KCTD10		X											
LRP1			X		X						X	X	
TCF7L1			X		X								
WIPF1			X						X			X	
DLL1			X										
LIMS2			X										
POPDC2			X										
FZD1				X	X								
RXRA				X		X		X					
CREB3L2				X			X						
TACC2				X								X	
NPR1					X								
RNF39					X								
CCND3						X		X		X			
LAT2						X		X					X
LRRC20						X		X					
NINJ2						X		X					
CAMKK2						X				X			
ALDH3B1						X					X		
MBD2							X					X	
NCOA3							X						
SPRED2							X						
CDS2							X						
RBBP8								X					
STX2									X	X			
PLAU									X		X		
ATMIN									X			X	
ASAP1												X	
IFFO1												X	
ANP32A													X

Genes without enrichment
ZCCHC14
SMYD2
PITPNM2
PRDM5
PVRL2
CSTF2T
CAPZB
STRAP
FOXK1
PML
NDUFB7
DGKZ
CCDC68
PLVAP
SOX7
RBPMS2
TLCD2
ALDH4A1
GNGT2
ARL5C
KLHL35
PRKCG
LENG1
ANKRD33
SMPD5
RFX8

Supplementary Figure 1: Genomic location for MCIp DMRs and EpiTYPER amplicons




Scale
chr2:
20 kb
mm10
35,260,000
35,265,000
35,270,000
35,275,000
35,280,000
35,285,000
35,290,000
35,295,000
35,300,000
35,305,000
35,310,000
Gsn
CpG Islands
DMR
EpiTYPER
Scale
chr6:
5 kb
mm10
125,150,000
125,155,000
125,160,000
Iffo1
CpG Islands
DMRs
EpiTYPER
Scale
chr6:
5 kb
mm10
90,465,000
90,470,000
90,475,000
Klf15
CpG Islands
DMR
EpiTYPER
Scale
chr3:
1 kb
mm10
55,782,500
55,783,000
55,783,500
55,784,000
55,784,500
55,785,000
Nbea
Mab21l1
CpG Islands
DMR
EpiTYPER
Scale
chr5:
10 kb
mm10
115,435,000
115,440,000
115,445,000
115,450,000
Msi1
CpG Islands
DMR
EpiTYPER
Scale
chr7:
5 kb
mm10
141,080,000
141,085,000
141,090,000
PKP3
CpG Islands
DMRs
EpiTYPER
Scale
chr9:
10 kb
mm10
65,630,000
65,635,000
65,640,000
65,645,000
65,650,000
65,655,000
65,660,000
Rbpms2
CpG Islands
DMR
EpiTYPER
Scale
chr2:
10 kb
mm10
32,725,000
32,730,000
32,735,000
32,740,000
32,745,000
32,750,000
32,755,000
Sh2d3c
6330409D20Rik
CpG Islands
DMRs
EpiTYPER



Supplementary Figure 1 Genomic location of DMRs and EpiTYPER amplicons used in technical and biological validation assays. RefSeq gene locations are depicted in blue, CpG islands in green, DMRs in black and amplicons for EpiTYPER MassARRAY in red.

Acknowledgements

I am not the person who likes to give long speeches, but I am glad that I finally reached the stage of this thesis to say thank you to all the people that helped me on this journey.

First of all to Dr. Clarissa Gerhäuser who directly swept over me with her enthusiastic project description in the PhD selection and made it easy for me to decide whose division I want to join. Thanks for all the lively discussions about the data and the drive to dig deeper even in the most complicated stages of the project, since nothing is ever simple. Thanks a lot Clarissa for this time and all the opportunities to learn new things. You still have to teach me the recipe for your famous raspberry cake!

I'd like to thank the members of my TAC PD Dr. Karin Müller-Decker, PD Dr. Dieter Weichenhan, Dr. Carolin Mogler and Prof. Dr. Christoph Plass for the helpful discussions and ideas about the project that helped to complete this thesis. Furthermore, I would like to thank Prof. Dr. Frauke Melchior, and Prof. Dr. Stefan Wiemann for agreeing to be part of my thesis examination committee.

I want to thank our technicians Monika Helf and Jana Petersen, who introduced me to the work with mice and helped me to collect all the samples from the little fury objects of my work.

Furthermore, thanks to Dr. Daniel Lipka, Dr. Yassen Assenov, Dr. Reka Toth and David Brocks who helped me at different stages of bioinformatic analyses, be it the analysis of MCIp-seq or ChIP-seq data, 450k arrays or just that R did not like me at a particular day. Thanks for your patience and the last minute rescues.

I am especially grateful to my proof reading team. Your eagle eyes spotted a lot of inconsistencies and typos that reduced the amount of errors to a minimum. Even if it was at the last minute, you were really there for me and I owe you one Clarissa, Maria, and Cornelia!

I want to thank our technician Karin Klimo for her assistance in any kinds of experiments, technical suggestions, ordering supplies or whatever came up, so that we could focus on our projects. I really enjoyed the hours of laughter and will always remember the tasty Alfredo. Thunderweather that really overrushed me!

Not to forget my other office mate and fellow PhD student Maria Pudenz. I really miss our common times in the office discussing about problems with the bioinformatics, bench work, writing reports and the thesis, the last handball game or triathlon or just life in general. You really helped me a lot, also to decide when it was time to think about a holiday break!

Furthermore, a big thank you to the C010-2 group members Karin, Annette, Anna-Lena, Clarissa, Marvin, Maria, Melanie and David for a good time in our group meetings and an easy going atmosphere in the lab that allowed for organized working. Sorry for all the changes on last notice about the Friday seminar schedule especially in the last months!

I also want to mention the other fellow students, be it PhD, Master or bachelor, who made time in and outside of the lab, lunch or coffee/tea breaks a lot of fun! Thanks to you Maria, Cornelia, Clarissa, Anna-Lena and Kevin for a great time!

Not to forget all the other present and former members of the C010 division. I want to thank you for support with experiments, answering stupid questions, and in general for all the discussions about scientific and non scientific topics. It was a great experience and I am glad that I had the chance to meet you.

Zum Schluss geht ein ganz besonderer Dank an meine Familie, die mich schon von Beginn meines Studiums an in meiner Berufswahl unterstützt hat. Danke, dass ihr immer ein offenes Ohr hattet, wenn es schwierig war, ich genervt war und ich keine Zeit hatte mich bei euch zu melden. Auch ganz liebe Grüße an unseren kleinen Sonnenschein und Frechdachs Heidi.