ANALYSES OF DNA METHYLATION PATTERNS IN PROSTATE CANCER

INAUGURAL-DISSERTATION

to obtain the academic degree

Doctor rerum naturalium

(Dr. rer. nat.)

submitted to the Department of Biology, Chemistry and Pharmacy

of Freie Universität Berlin

by

Stefan Börno

from Berlin

2012

Erstellt unter der Leitung von Frau Dr. Dr. Michal-Ruth Schweiger am Max-Planck-Institut für Molekulare Genetik in Berlin im Zeitraum von Juli 2008 – Juli 2012.

1. Gutachter: Prof. Dr. Hans Lehrach, Max-Planck-Institut für Molekulare Genetik

2. Gutachter: Prof. Dr. Rupert Mutzel, Freie Universität Berlin

Tag der Disputation:

17.10.2012

ZUSAMMENFASSUNG

Die Entstehung von Prostatatumoren wird von weitreichenden Veränderungen der DNA-Methylierungs- und assoziierter Genexpressionsmuster begleitet. Zur Identifizierung tumor- und tumorsubklassenspezifischer Variationen der DNA-Methylierungsmuster habe ich methylierte DNA-Regionen von 53 Normal- und 51 Tumorproben über Immunopräzipitationen angereichert und mit dem SOLiD-System sequenziert (MeDIP-Seq). Dabei konnte ich signifikante Hypermethylierungen in Homöobox-, Tumorsuppressor- und Onkogenen, G-Protein-gekoppelten Rezeptoren, sowie in microRNA-Regionen feststellen. Hypomethylierungen betrafen hingegen vor allem repetitive extragenische Bereiche. Neben den in Prostatatumoren bereits als hypermethyliert beschriebenen Regionen wie *GSTP1* oder *RARB* habe ich hunderte weitere tumorspezifische Marker identifiziert und konnte mehr als 80 von ihnen unter Verwendung eines bisulfitspezifischen Massenspektrometrieansatzes validieren.

Für drei Marker habe ich in DNA, die ich aus Patientenurin isoliert habe, quantitative methylierungsspezifische PCRs durchgeführt und konnte so das Potential dieses nichtinvasiven Ansatzes für die Prostatakrebsdiagnose aufzeigen.

Die Integration von Gen- und microRNA-Expressionsdaten mit den Methylierungsdaten zeigte eine Abhängigkeit der Genexpression von DNA-Methylierungen insbesondere bei stark exprimierten Genen auf. Durch Luziferase-Methylierungsexperimente konnte ich eine methylierungsabhängige Herabregulierung der Expressionen von *DUOX1*, *miR23/24/27* und *miR26a* bestätigen.

Bei etwa 50% aller Prostatatumore läßt sich eine Fusion der androgenregulierten *transmembrane protease, serine 2* (*TMPRSS2*) und dem Onkogen *v-ets erythroblastosis virus E26 oncogene homolog* (*ERG*) nachweisen, die eine vermehrte Expression von *ERG* zur Folge hat. Vergleiche der Methylierungsmuster der *TMPRSS2:ERG*-fusionspositiven und -negativen Tumore zeigten eine deutlich ausgeprägtere differentielle Methylierung in letzteren („Methylator"-Phänotyp), während fusionspositive Tumore einen weniger stark veränderten Methylierungsphänotyp aufwiesen.

In weiterführenden Analysen konnte ich zeigen, daß *enhancer of zeste homolog 2* (*EZH2*) – eine Histonmethyltransferase und Bindeglied zwischen

Histonmodifikationen und DNA-Methylierung –in fusionsnegativen Proben signifikant stärker exprimiert vorliegt. Dies wird von einer genomischen Hypermethylierung der *miR26a*-Region begleitet, die zu einer verminderten Expression der *EZH2*-regulatorischen *miR26a* führt. Eine 5-Aza-2`-desoxycytidin-Behandlung fusionsnegativer PC3-Zellen führte zur Wiederherstellung der *miR26a*-Expression und damit zu einer Suppression von *EZH2*.

Basierend auf diesen Daten konnte ich ein Modell der Entstehung von fusionsnegativen und –positiven Tumoren ableiten: Bei Tumoren mit dem *TMPRSS2:ERG*-Fusionsgen führt die Überexpression des *ERG*-Onkogens zu einer direkten Aktivierung der Transkription von *EZH2* und weiteren Onkogenen. Bei fusionsnegativen Tumoren hingegen führt die genomische Methylierung der *miR26a* Region und die damiteinhergehende verminderte Expression von *miR26a* zu einer noch stärker ausgeprägten Überexpression von *EZH2*. Dieser sehr starke Anstieg von *EZH2* führt zu extensiven Veränderungen der DNA-Methylierungs- und Genexpressionsmuster in dieser Tumorsubgruppe.

Die in dieser Arbeit aufgezeigten Mechanismen haben weitreichende Konsequenzen für eine personalisierte Therapie von Patienten mit Prostatatumoren: Aufgrund des Pathomechanismus bieten sich bei fusionsgenpositiven Tumoren *PARP1*-Inhibitoren an, wohingegen Patienten ohne *TMPRSS2:ERG*-Genfusion von *EZH2*-Inhibitoren wie 3-deazaneplanocin A profitieren würden.

ABSTRACT

Prostate tumourigenesis is accompanied by extensive alterations in the DNA methylation patterns with presumably severe effects on gene expression. To identify tumour- and tumour subgroup- specific alterations in the DNA methylation pattern I applied an immunoprecipitation-based enrichment approach for methylated DNA regions followed by SOLiD sequencing (MeDIP-Seq) on 51 tumour and 53 normal prostate samples. I discovered significant hypermethylations in homeobox genes, tumour suppressor and oncogenes, G protein coupled receptors, as well as in microRNA regions; hypomethylations mainly occurred in repetitive extragenic regions. Beside regions that are already described as hypermethylated in prostate cancer like *GSTP1* and *RARB* I extracted hundreds of promising tumour specific markers and could validate more than 80 of them using bisulphite mass spectrometry analyses. Furthermore, I could show the diagnostic potential of three of the markers for the non-invasive detection of prostate cancer in urinary DNA by a quantitative methylation specific PCR approach.

Integration of gene and miRNA expression data with the methylation data revealed a dependency of gene expression from DNA methylation mainly in highly expressed genes. Using luciferase methylation assays I could prove methylation dependent downregulation of *DUOX1*, *miR23/24/27* and *miR26a*.

Approximately 50% of all prostate tumours show a fusion of the androgen regulated *transmembrane protease, serine 2* (*TMPRSS2*) and the *v-ets erythroblastosis virus E26 oncogene homolog* (*ERG*) oncogene resulting in an increased expression of *ERG*. Comparisons of the DNA methylation patterns of *TMPRSS2:ERG* fusion positive and fusion negative tumours revealed pronounced differential methylations in the latter (‘methylator’ phenotype) while fusion positive samples exhibited a less altered methylation phenotype.

In further analyses, I could show that enhancer of zeste homolog 2 (*EZH2*) – a histone methyltransferase linking histone modifications to DNA methylations – was significantly higher expressed in fusion negative samples. This went along with a hypermethylation of the *miR26a* genomic region leading to a decreased expression of the *EZH2* regulating *miR26a*. Treatment of fusion negative PC3 cells with 5-aza-2’-deoxycytidine could re-establish *miR26a* expression with subsequent *EZH2* suppression.

Based on these data, I deduced a model for the development of prostate cancer in fusion positive and fusion negative patients: In tumours bearing the *TMPRSS2:ERG* fusion gene the increased expression of *ERG* results in a direct activation of the transcription of *EZH2* and further oncogenes, while in fusion negative tumours the genomic methylation and suppression of *miR26a* causes an even more pronounced upregulation of *EZH2*. This strong increase of EZH2 leads to extensive changes in the DNA methylation and gene expression patterns.

The elucidated pathways can be exploited in personalised prostate cancer therapies: Based on the pathomechanism of fusion positive tumours PARP1 inhibitors could be used in this class, while patients without a *TMPRSS2:ERG* gene fusion might benefit from *EZH2* inhibitors like 3-deazaneplanocin A.

TABLE OF CONTENTS

ABBREVIATIONS

0CpG reads	UniqLocs after depletion of reads containing 0CpG in 200bp
5-hmC	5-hydroxymethylcytosine
5-meC	5-methylcytosine
6-DHT	6-dehydrotestosterone
AFMS	Anterior Fibromuscular Stroma
AR	Androgen receptor
ARE	Androgen response element
BC	Barcode
BMBF	Bundesministerium für Bildung und Forschung
bp	Base pair
BPH	Benign prostate hyperplasia
BS	Sodium bisulphite
BS-MS	Mass spectrometry of BS converted DNA
BS-Seq	Sequencing of BS converted DNA
cDNA	Complementary DNA
CGI	CpG island
CNA	Copy number alteration
Cor	Pearson correlation
CpG	Cytosine-phosphatidyl-guanine
CZ	Central zone
DAVID	Database for Annotation, Visualization and Integrated Discovery
DHT	Dihydrotestosterone
DMR	Differentially methylated region
DNMT	DNA methyltransferase
DRE	Digital rectal examination
DSB	double strand break
EB	Elution buffer (Qiagen)
ECM	Extracellular matrix
EMT	Epithelial to mesenchymal transition
ePCR	Emulsion PCR
ERG	*v-ets erythroblastosis virus E26 oncogene homolog*
ERSPC	European Randomised Study of Screening for PCa
ETS	E-twenty six transcription factor family
FCS	Fetal calf serum
FDR	False discovery rate
FPR	False positive rate
FUS-	*TMPRSS2:ERG* fusion negative PCa
FUS+	*TMPRSS2:ERG* fusion positive PCa
Gb	Gigabases
GPCR	G-protein coupled receptor
H3K27	Histone 3 Lysine 27
HGPIN	High grade PIN
HOX	homeobox
IP	Immunoprecipitation
LCM	Laser capture microdissection
LINE	Long interspersed element
LPA	Linear polyacrylamide
LTR	long terminal repeat
MALDI-TOF	Matrix assisted laser desorption/ionisation - time of flight
Mb	Megabase
MBD	Methyl CpG binding protein domain
MeDIP	Methylated DNA immunoprecipitation
MeDIP-Seq	MPS of MeDIP enriched DNA
MIR	Mammalian interspersed repetitive element
miRNA	MicroRNA
MO	Mappable output
MP	Multiplex
MPS	Multiparallel sequencing
MSP	Methylation specific PCR
NFW	Nuclease free water
NTC	No template control
OR	Odds ratio
p_{BH}	p_{MW} after Benjamini-Hochberg correction for multiple testing
PBS	Phosphate buffered saline
PCa	Prostate cancer
PCA	Principal component analysis
PEI	Polyethyleneimine
PIA	Proliferative inflammatory atrophy
PIN	Prostatic intraepithelial neoplasia
PLG	Phase lock gel
p_{MW}	Mann-Whitney p-value
PSA	Prostate specific antigen
PZ	Peripheral zone
qMSP	Quantitative real time MSP
qPCR	Quantitative real time PCR
Rho	Spearman correlation
RISC	RNA induced silencing complex
RO	Raw output
ROC	Receiver operating characteristic
RPE	Radical prostatectomy
rpm	Reads per million
RT	Room temperature
SAM	S-adenosylmethionine
SDS	Sodium dodecylsulphate
SINE	Short interspersed element
SNP	Single nucleotide polymorphism
SNV	Single nucleotide variation
SOLiD	Sequencing by Oligonucleotide Ligation and Detection
TAE	Tris/Acetic acid/EDTA buffer
TE	Tris/EDTA buffer
TMPRSS2	*Transmembrane protease serine 2*
TNM	Tumour node metastasis
TPR	True positive rate
TRUS	Transrectal ultrasound
TSS	Transcription start site
TZ	Transitional zone
UMO	Uniquely mappable output
UniqLocs	UMO after depletion of redundant reads
UTR	Untranslated region
WGA	Whole genome amplification
X_i	Inactive X chromosome

1 INTRODUCTION

1.1 The hallmarks of cancer according to the Hanahan-Weinberg tumour model

Tumours are formed in a multistep process that involves the shutdown or circumvention of several cellular control mechanisms. According to the 2011's revision of the Hanahan-Weinberg tumour model (HANAHAN 2011) genome instability and mutation as well as tumour-promoting inflammation lead to the emergence of several hallmarks (**Table 1**) that allow tumour growth, survival, invasion of surrounding tissue and formation of distant metastases. Genomic instability and mutations can be induced by a breakdown of the DNA repair machinery or protective mechanisms as well as excessive alterations in the DNA methylation pattern rendering the genome susceptible for further aberrations.

Table 1: Hallmarks of cancer according to Hanahan *et al.* (HANAHAN 2011).

Cancer hallmark	Examples of involved pathways
Sustain proliferative signalling	Activation of MAPK, *AKT1* signalling Loss of *RAS*, *PTEN*
Evade growth suppression	*RB* and *TP53* suppression
Resistance to cell death	Loss of *TP53* Anti-apoptotic signalling (*BCL2*) Survival factors (*IGF1/2)*
Replicative immortality	Telomerase activation
Reprogramming of energy metabolism	Increased glycolysis and lactic acid fermentation
Angiogenesis	*VEGF-A* and *TSP1* imbalance
Inflammatory signalling	Recruitment of immune cells and peritumoural inflammation
Evade immune destruction	Secretion of *TGF-β*
Invasion and metastasis	Loss of E-cadherin EMT

Through ligand independent, constitutive activation of mitogen activated protein kinase (MAPK) or *AKT1* signalling or the disruption of attenuating mechanisms like loss of *rat sarcoma* (*RAS*)-GTPase or *phosphatase and tensin homolog* (*PTEN*) activity cells can **sustain proliferative signalling.** Additionally, tumour cells have to **evade growth suppressors**. *Retinoblastoma* (*RB*) and *tumour protein p53* (*TP53*) are the central control nodes that are shut down in a majority of tumours. *RB* signalling transduces growth inhibitory signals from outside, while *TP53* integrates signals from inside the cell leading

to an arrest of cell proliferation or to apoptosis, the programmed cell death. In normal cells, imbalances in signalling as mentioned before, excessive DNA damage through hyperproliferation or exogenous noxae can induce an apoptotic caspase cascade that finally leads to cell disassembly and its consumption by neighbour cells or phagocytes. This induction of apoptosis is controlled by pro-apoptotic (*BCL2 associated X protein* (*BAX*), *BCL2 binding component 3* (*BBC3*) and *BCL2 antagonist/killer 1* (*BAK1*)) and anti-apoptotic proteins (*B-cell/CLL lymphoma 2* (*BCL2*) and its relatives) (CHOWDHURY 2006). Tumour cells acquire **resistance to cell death** by loss of *TP53*, increase in anti-apoptotic *BCL2*/*BCL2L1* (*BCL2 like 1*) or survival factor signalling (*insulin like growth factor 1/2* (*IGF1*/*2*) *or interleukin 3* (*IL3*)) or the downregulation of pro-apoptotic factors.

In non-disease state cells stop proliferation and enter senescence after a certain number of replication cycles caused by the shortening of telomeres – whose function is the protection of the DNA from end-to-end fusions – in each replication. Tumour cells evade this protective mechanism and **gain replicative immortality** by re-activation of the enzyme telomerase or by recombination based telomere maintenance mechanisms. Excessive telomere shortening prior to re-activation might even foster tumour growth by destabilising the genome leading to a massive acquisition of mutations.

Fast replication like in tumour or embryonic development requires large amounts of anabolic precursors. These are mainly derived from **reprogramming the energy metabolism** from oxidative phosphorylation (citric acid cycle) to glycolysis followed by lactic acid fermentation (Warburg effect). The increased glycolysis allows the use of glycolytic intermediates for biosynthesis.

The growing neoplasia needs to be supported by the formation of neovasculature (**angiogenesis**). In tumours, the imbalance of proangiogenic *vascular endothelial growth factor A* (*VEGF-A*), responsible for embryonic blood vessel growth that also can be induced by hypoxia and oncogene signalling, and *thrombospondin 1* (*TSP1*), a suppressor of angiogenesis, leads to an activation of the angiogenic switch resulting in formation of aberrant blood vessels. This can be supported by **inflammatory cell signalling**: Innate immune cells like macrophages, neutrophils, mast cells or myeloid progenitor cells can infiltrate the lesion and assemble at tumour margins creating a peritumoural inflammation. These cells can supply tumour promoting molecules like growth factors, survival factors, angiogenic factors, enzymes for the modification of the extracellular matrix facilitating invasion into surrounding tissue, and they can produce reactive oxygen species leading to increased mutations. Due to the downregulation of apoptosis, necrosis is the only way to

eliminate excessively aberrated cells in tumours. This controlled destruction leads to the release of proinflammatory signals further fostering inflammation. Thus, some tumours tolerate a certain degree of necrosis to recruit tumour supportive inflammatory cells.

On the other hand, tumour cells have to find ways to **evade immune destruction** by cytotoxic T lymphocytes and natural killer cells e.g. by secreting immunosuppressive *transforming growth factor beta* (*TGF-β*) or recruitment of regulatory T cells.

In tumour development, cells **gain the ability to invade surrounding tissue and metastasise**: Excessive angiogenesis and degradation of the extracellular matrix facilitates local invasion and the transit to blood and lymphatic vessels. The process is controlled by the epithelial to mesenchymal transition (EMT) programme, characterised by a loss of *E-cadherin,* a cell to cell adhesion molecule, and other cell-extracellular matrix (ECM) adhesion molecules. Adhesion molecules of the embryonic development programme on the other hand are upregulated. EMT results in a loss of adherens junctions, the conversion from an epithelial to a more fibroblastic morphology, the secretion of matrix degrading enzymes, an increased motility and resistance to apoptosis. Disseminated tumour cells form micrometastases in distant tissues that can grow into metastases after adaptation to the new environment (colonisation).

Altogether, tumour formation is a process requiring the shutdown of several cellular control mechanisms and a complex interplay between the tumour and its microenvironment.

1.2 Epidemiology of prostate cancer

Prostate cancer (PCa) accounts for more than 900,000 new cases and 250,000 deaths per year (2008) and is the second most common cancer among men worldwide (JEMAL 2011). Three quarters of cases occur in patients older than 65. With 2.5 deaths per 100,000 inhabitants Eastern Asia has the lowest mortality rate while in the Caribbean 26.3 per 100,000 inhabitants die of PCa. Additionally, strong interethnic differences even within one country can be observed with the highest mortality and most aggressive forms of PCa in people of African origin, who have a 1.7-1.8 fold increased risk compared to the white population, and the lowest risk in Asians with a 50% reduced mortality (PARKIN 2005, SHIBATA 1997). Nevertheless, survival after treatment with access to equal health care is comparable between the ethnicities (CULLEN 2011). The major part of the differences is likely to result from different allelic distributions of genes of the sexual hormone signalling

and metabolism pathways with a moderately elevated PCa risk between the populations containing polymorphisms in the *androgen receptor* (*AR*) or *steroid 5α–reductase* (*SRD5A2*) gene – an enzyme converting testosterone into its active form dihydrotestosterone (DHT) (SHIBATA 1997). These polymorphisms cause different activities of the enzymes leading to variable PCa risks.

Another factor influencing PCa incidence is the personal lifestyle: A soy or fruit based diet containing phytoestrogens like the isoflavone genistein has been shown to be beneficial for preventing prostate tumour formation (HORI 2011, SONODA 2004) by interfering with androgen and apoptotic signalling (BYLUND 2003, DE SOUZA 2010, PIHLAJAMAA 2011). Furthermore, there is evidence that soy isoflavones can even revert DNA hypermethylation in esophageal squamous cell carcinoma and prostate cancer cells by inhibition of the maintenance methyltransferase *DNMT1* (FANG 2005). Further positive effects are described for cessation of smoking (PLASKON 2003) and physical exercise (LIU 2011b).

1.3 Anatomy of the prostate

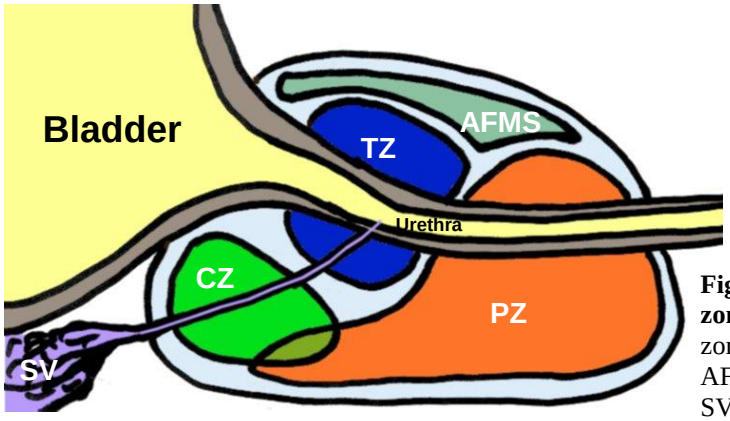


Figure 1. Lateral view of the prostate zones. PZ=Peripheral zone, CZ=Central zone, TZ=Transitional zone, AFMS=Anterior fibromuscular stroma, SV=Seminal vesicle.

The prostate is a walnut-sized exocrine gland surrounding the urethra posterior of the bladder. It has a ductal architecture and consists of several branches. The ducts are formed by a luminal secretory and a basal cell layer. The prostate secretes prostatic fluid containing prostate specific antigen (PSA), zinc and prostatic acid phosphatase liquefying the ejaculate and allowing a longer survival of the spermatozoa in the vaginal environment (CHRISTENSSON 1990). The prostate can be subdivided into four zones (**Figure 1**).

The transitional zone (TZ) is situated in the midgland laterally of the proximal urethra. Most cases of benign prostate hyperplasia (BPH) and around 20% of all PCa arise from this zone. The cone-shaped central zone (CZ) surrounds the ejaculatory ducts and gives rise to 1-5% of PCas. The largest part of the prostate is formed by the peripheral zone

surrounding the distal urethra and CZ. More than 70% of prostate tumours develop in this zone. The anterior part of the prostate is formed by muscles and fibrous tissue (anterior fibro-muscular stroma (AFMS)) (HAMMERICH 2008).

1.4 Development of prostate cancer: current model

An individual genetic background predisposes each patient differently towards inflammation or carcinogenic environmental agents. As described in the Hanahan-Weinberg cancer model inflammation is one of the initiating events of tumour formation (HANAHAN 2011). In PCa development it can be followed by the formation of epithelial confined prostatic intraepithelial neoplasias (PINs) (ANDREOIU 2010), lesions preceding PCa (BOSTWICK 2012, CAZARES 2011, MONTIRONI 2007). The majority of PINs develop in the peripheral zone – the cradle of most PCas. Low grade PINs (LGPINs) are characterised by secretory cells with enlarged heterogeneous nuclei and moderately increased chromatin content. High grade PINs (HGPINs) have large uniform nuclei with increased chromatin content and prominent nucleoli. Additionally, the basal cell layer might be disrupted. At these sites, stromal invasion, a first sign of PCa, might occur. HGPINs are found in more than 80% of cancerous prostatectomy specimens and are often directly adjacent to the tumour. Like PCa, HGPINs can be multifocal. Interestingly, HGPINs show similar chromosomal aberrations as PCa, like chromosome 8q-amplifications (QIAN 1995). Additionally, extensive shortening of telomeres and increased activity of telomerase can be observed in some HGPINs (KOENEMAN 1998). There are as well some epigenetic features common in HGPINs and PCa, e.g. *GSTP1* is hypermethylated in a large proportion of HGPINs (GOESSL 2001). In patients with HGPINs, PCa arises within 5-10 years but can be suppressed by hormone deprivation therapy (BOSTWICK 2012), revealing the androgen dependent nature of PINs.

Androgen action is mediated by the intracellular *androgen receptor* (*AR*) (TRAPMAN 1997). After entering the prostate cells, testosterone is converted to its active form dihydrotestosterone (DHT) by *steroid 5α-reductase* (*SRD5A*). DHT binds to the inactive *AR* causing dissociation of chaperones and subsequent transfer of the *AR* into the nucleus. The active *AR* dimerises and binds to the androgen response elements (AREs) of its target genes where it interacts with transcription inducing factors causing differentiation and proliferation. *AR* signalling is important for normal prostate cell growth and differentiation,

in tumour cells excessive *AR* signalling prevents apoptosis and results in excessive cell proliferation.

Early PCa is androgen sensitive; it relies on the activation of the *AR* through androgens. Blocking of this signalling through therapeutic approaches (inhibition of testosterone production by orchiectomy, anti-androgenic hormone therapy) in most cases leads to a regression of the tumour, but after 12-18 months aggressive, androgen insensitive (hormone refractory) PCa may reoccur. Recurrence is caused by alterations in a variety of pathways (DUTT 2009): *AR* might become extremely sensitive to small amounts of androgens, either due to point mutations in the ligand binding domain, to an amplification of this domain, or due to an increased *5α-reductase* activity resulting in an increased amount of the active DHT. The latter is often observed in people of African origin. Sensitivity might additionally increase by increased expression of co-activators. The *AR* pathway also might get activated by highly expressed growth factors like *IGF*, receptor tyrosine kinases or *interleukine 6*. Due to lost specificity of the *AR* also non-androgenic steroids and even anti-androgenic therapeutics can activate *AR* signalling. Finally, *AR* signalling can be bypassed by activation of anti-apoptotic *BCL2* expression, oncogenes or inactivation of tumour suppressors (DUTT 2009).

A large proportion (60-90%) of patients harbours a multifocal prostate cancer that in most cases is presumed to have arisen from independent lesions while in highly aggressive cancers the multifocality also might originate from a single lesion through intraglandular dissemination (ANDREOIU 2010). While tumours from the TZ are mostly well differentiated and confined to the TZ, those originating from the PZ show less differentiation and are likely to infiltrate surrounding tissue even at smaller sizes. Nevertheless, serum prostate specific antigen (PSA) values, a common marker of PCa, are higher in TZ tumours thus allowing earlier diagnosis but further impeding prognosis. Notwithstanding its multifocality, it has been shown that metastases forming circulating tumour cells mostly originate from a single focus that can be as small as 0.2mm (SCHMIDT 2006).

Although PCa becomes aggressive only in a small percentage of cases, a 1994 autopsy study showed that more than 60% of people older than 60 years who died of other reasons harboured a mostly clinically unobtrusive tumour (SAKR 1994, SOOS 2005). Despite of its high prevalence, the clinical management of PCa is limited by the low specificity and sensitivity of the existing diagnostic and prognostic tools and the lack of effective systemic therapeutic strategies.

1.5 Diagnosis of prostate cancer

Digital rectal examination (DRE) is the prevailing means of PCa diagnosis. The major drawbacks of this technique are the low acceptance in the population and the need for an experienced urologist to find small lesions. DRE can be complemented by transrectal ultrasonography (TRUS) and serum PSA level measurement. PSA level is a good tool to develop a first suspicion. PSA is a product of the *kallikrein 3 (KLK3*) gene and was first purified in 1979 by Wang *et al.* (WANG 1979). It liquefies the seminal fluid by proteolytic cleavage of gel forming seminal proteins like *seminogelin-1* (CHRISTENSSON 1990). PSA levels are increased in prostate tumours (CATALONA 1991, STAMEY 1987) and can be used to predict individual rescreening intervals if serum concentration is less than 1ng/ml (DAMBER 2008). However, this tool is not fine enough to differentiate between highly aggressive and mild forms of PCa and also bears a high false positive rate (ECKERSBERGER 2009, KILPELAINEN 2010, SHARIAT 2011). Elevated PSA levels do not need to result from a malignant transformation of the prostate but can also arise from inflammation, benign prostate hyperplasia (BPH) (SHARIAT 2011) or precedent TRUS (OESTERLING 1991). Additionally there are a lot of patients with low PSA harbouring a tumour that could be overseen by PSA screening alone using the current threshold of 4ng/ml. Thompson *et al.* have performed receiver operating characteristics (ROC) analyses and derived a sensitity of 20.5% and a specificity of 93.8% for this threshold (THOMPSON 2005). And even with lower thresholds many tumours are missed: For example, in 6.6% of men with PSA levels ≤0.5ng/ml a prostate tumour was detected (THOMPSON 2004, TOSOIAN 2010). Differentiation between free and complexed PSA bound to serum proteins like *α2-macroglobulin* improved the predictive value of PSA. A low ratio of free to total PSA indicates a higher probability of PCa (SHARIAT 2011).

After a suspicious finding, transrectal core biopsies are taken, extracting up to 10-14 cores from the prostate. This massive intervention can cause side effects like urinary-tract infection or sepsis (DAMBER 2008).

The prostate biopsies are graded according to the Gleason score (DELAHUNT 2012, GLEASON 1992). For that, the predominant and the second most common pattern of aberrant prostate cells are evaluated using a low magnification (40-100x). In the refined version of 1992 Gleason suggested five grades with up to three subgrades for classification **(Figure 2)** (GLEASON 1992). The patterns correlate with survival in follow up studies. Total Gleason scores (pattern 1 + pattern 2) of below 7 show a good and above 7 a bad

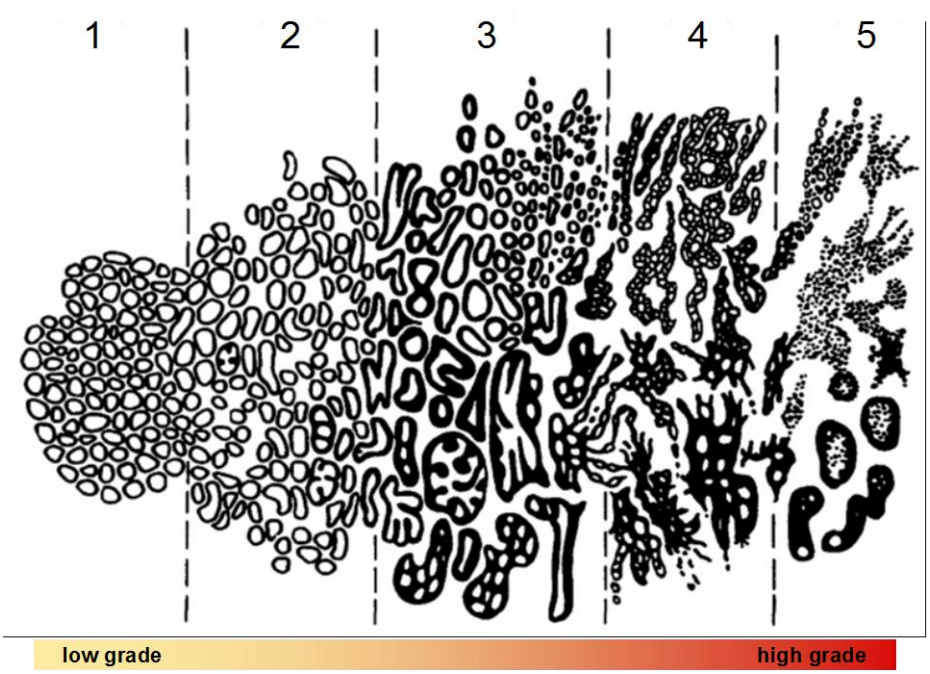


Figure 2. Histologic patterns of prostate cancers of different grades accoding to Gleason. Picture taken from Gleason (GLEASON 1992).

prognosis. Gleason score 7 can be subdivided into 3+4 and 4+3, where 4+3 has a worse prognosis. The exact reproducibility of the scores among trained urological pathologists lies within the range of 70% (DELAHUNT 2012). Scoring has been discussed and modified in 2005 resulting in the MG (modified Gleason) grading.

1.6 TNM classification of malignant tumours

Currently, PCa is classified following the 7th edition of the American Joint Committee on Cancer's (AJCC) tumour-node-metastasis (TNM) system of 2010 (EDGE 2010) (**Table 2**). In brief, this system describes the primary tumour extend (T), spreading to the lymph nodes (N) and whether distant metastases have been formed (M). T1 tumours are neither palpable nor detectable by imaging but can be found in biopsies. T2 prostate tumours are large enough for being found by DRE or TRUS and are confined within the prostate. T3 tumours break through the prostate 'capsule' and can invade the seminal vesicles. Tumours invading other organs like the pelvic wall or rectum are staged T4. If regional (pelvic) lymph nodes are affected the N classifier is 1, otherwise it is 0 (not affected) or X (not assessed). Distant metastases are described by M=1, if no metastases can be found M is 0. There are several subclassifications for the T, N and M groups, together with Gleason score and PSA value allowing a more precise description of the tumour and its prognosis (CHENG 2012).

Table 2: TNM classification according to the American Joint Committee on Cancer (EDGE 2010).

Tumour stage		
TX	Primary tumour not assessed	
T0	No primary tumour	
T1	Inapparent tumour, not palpable or visible	
	T1a	Incidental histologic finding in 5% or less of resected tissue
	T1b	Incidental histologic finding in more than 5% of resected tissue
	T1c	Identified by needle biopsy
T2	Prostate confined tumours	
	T2a	≤50% of one lobe affected
	T2b	>50% of one lobe affected
	T2c	Both lobes affected
T3	Tumour extends through prostate capsule	
	T3a	Extracapsular extension
	T3b	Invasion of seminal vesicles
T4	Invasion of adjacent structures other than seminal vesicles	
Lymph node stage		
NX	Regional lymph nodes not assessed	
N0	No regional lymph node metastasis	
N1	Regional lymph node metastasis	
Distant metastasis stage		
M0	No distant metastasis	
M1	Distant metastasis	
	M1a	Lymph nodes
	M1b	Bone
	M1c	Other sites

1.7 Management of prostate cancer

The clinical course of PCa is heterogeneous, spanning from indolent tumours requiring no therapy during lifetime to highly aggressive metastatic diseases (DAMBER 2008). Diagnosed at an early stage PCa is a curable disease. Nevertheless, because of its yet mainly unpredictable outcome PCa is overtreated mainly in elderly patients impairing their lives severely (DASKIVICH 2011). Patients with distant metastases show a poor prognosis with average survival rates of 24-48 months while for other patients no definite prognosis can be made. Patients with regional lymph node metastases show an average survival of 8 years. Furthermore, localised PCa can be grouped into risk groups of re-occurrence after treatment (D'AMICO 1998).

Prostate cancer can be treated by radical prostatectomy (RPE), radiation and hormone deprivation therapy and combinations of them (HEIDENREICH 2010), each showing severe side-effects: RPE often leads to erectile dysfunction and urinary incontinence, radiotherapy might show gastrointestinal and genito-urinary toxic effects and also might lead to incontinence and erectile dysfunction (DAMBER 2008). Hormonal therapies can be subdivided into hormonal castration, deprivation of androgens and estrogene therapy with

distinct effects like loss of libido, gynaecomastia, cognitive decline and other toxic effects. M1 patients are mostly treated with palliative hormone therapy (chemical castration) to slow down tumour growth. After treatment with a combination of RPE, pelvic lymphadenectomy and adjuvant hormonal therapy the average survival of patients with regional lymph node metastases can be extended to more than 10 years (KROEPFL 2006). Nevertheless, a review of the *European Randomised Study of Screening for Prostate Cancer* (*ERSPC*) states that out of 48 men who underwent a treatment only one would have died of PCa without therapy (SCHRODER 2009), making watchful waiting a good alternative to treatment as it shows the same efficiency in patients with short life expectancy (HEIDENREICH 2010).

This illustrates the great demand for molecular markers to distinguish between aggressive and mild forms and to increase specificity. Additionally, an increased sensitivity might lead to an earlier detection of PCa, thus giving a larger time frame for 'watchful waiting', allows a better prediction of tumour outcome, and eventually reduces unnecessary surgery.

1.8 Genomic alterations in prostate cancer

Familial aggregations of PCa imply a genetic cause, and indeed, a few germline risk mutations, i.e. mutations in parental germ cells that eventually can be found in every cell of the offspring, have been already discovered (DAMBER 2008). These include mutations in the genes *Rap guanine nucleotide exchange factor 4*, *ribonuclease L*, *macrophage scavenger receptor 1*, *checkpoint kinase 2*, *capping protein muscle Z-line beta*, *vitamin D receptor*, and *paraoxonase 1*. Nevertheless, only few PCa patients carry these mutations.

On the other hand, the occurrence of somatic, i.e. acquired non-germ line, protein altering mutations in prostate tumours (0.33/Mb) is significantly lower than in other cancers like breast (~1/Mb) or lung cancer (3.8/Mb) and lies within the lowest range of cancer mutations (BERGER 2011, KAN 2010). Thus, recurrent somatic mutations are rare. Nevertheless, a recently discovered mutation of the *speckle-type POZ protein* gene (*SPOP*) is described for 6-15% of PCa cases (BARBIERI 2012).

More common than somatic mutations are chromosomal anomalies like amplifications (chromosome 8q), deletions (chromosome 8p), inversions and translocations (QIAN 1995, TAYLOR 2010). Haplo-insufficiency of the tumour suppressor *PTEN* can be found in ~60% of all PCa and ~20-25% of all HGPIN (SIRCAR 2009, YOSHIMOTO 2006) and results in reduced repression of anti-apoptotic and proliferative *AKT* signalling (TESTA 2001).

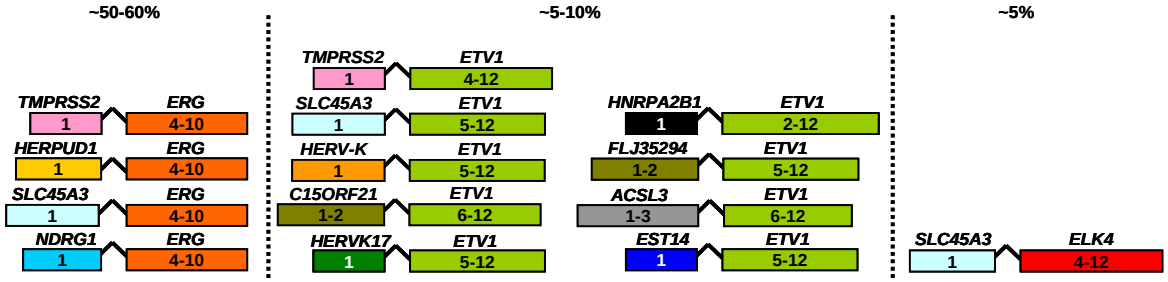


Figure 3. Common gene rearrangements involving members of the *E-twenty six* (*ETS*) gene family (modified from Rubin *et al.* (RUBIN 2011)). The percentages denote the estimated frequencies of the respective fusion class. The boxed numbers depict the exons involved in the fusions.

Another well known somatic alteration is the gene fusion between the androgen regulated *transmembrane protease serine 2* (*TMPRSS2*) and gene members of the oncogenic *E-twenty six* (*ETS*) family like *ETS variant 1* and *4* (*ETV1*, *ETV4*), *SRF accessory protein 1* (*ELK4*) or *v-ets erythroblastosis virus E26 oncogene homolog (ERG*) found in approximately 50% of all prostate tumours and detectable in HGPIN lesions, suggesting an early occurrence of these fusions (BRAUN 2011, RUBIN 2011). During recent years also other – but rare – androgen responsive 5' fusion partners of ETS genes have been identified, like for example *solute carrier family 45 member 3* (*SLC45A3*), *homocysteine-inducible endoplasmic reticulum stress-inducible ubiquitin-like domain member 1* (*HERPUD1*), and *N-myc downstream regulated 1* (*NDRG1*) (MEHRA 2007, PERNER 2006, RUBIN 2011, TOMLINS 2005) (**Figure 3**). Androgen induced expression of *ERG* directly results in increased expression of several oncogenic target genes like *MYC* (SUN 2008), genes of the WNT-signalling pathway (MOCHMANN 2011) and the *EZH2* histone methyltransferase (KUNDERFRANCO 2010).

Recent models for the generation of the gene fusions (HAFFNER 2010, LIN 2009) describe that androgen signalling leads to recruitment of *topoisomerase 2b* to transcribed genes, where it induces DNA double-strand breaks (DSB) to facilitate transcription. These DSBs are highly recombinogenic since they are often found within Alu repetitive elements (LIU 2007). During DSB repair the 3Mb interstitial region between *TMPRSS2* and *ERG* (or other ETS genes) might get homozygously or hemizygously deleted (in 60% of all fusion cases) or translocated resulting in gene fusions (BARRY 2007, HERMANS 2006, LIU 2007, PERNER 2006). Patients with a deletion type of *TMPRSS2*:*ERG* fusion may additionally suffer from loss of the genes of the interstitial region like *high mobility group nucleosome binding domain 1* (*HMGN1*) (PERNER 2006). Several breakpoints in *TMPRSS2* and *ERG* have been discovered so far involving, for example, the second or third intron of *ERG* and the first to fourth intron of *TMPRSS2*.

Clinically, *TMPRSS2:ERG* gene fusions are associated with severe forms of PCa in minimally treated population-based cohorts (DEMICHELIS 2007, RUBIN 2011). The overexpression of *ERG* is thought to be sufficient for the initiation of PIN lesions and seems to increase the invasive potential (KLEZOVITCH 2008). Nevertheless, ETS gene fusions alone are assumed to be only supportive and not causal for cancer formation (RUBIN 2011). Interestingly, the fusion state of different independently evolved foci can be heterogeneous while the intrafocal fusion state is homogeneous. This further indicates fusion events occuring early in PCa development (BARRY 2007). Thus, fusion positive and fusion negative tumours requiring different treatments may develop within one prostate. Overall, *TMPRSS2:ERG* is a promising marker for prostate cancer detection and monitoring of treatment success, since fusion gene products can easily be determined in blood or urine (LAXMAN 2006) and are – similar to other fusion genes – tumour unique drug targets.

Detection of recurrent gene fusions in half of all prostate cancer cases helped to elucidate the pathomechanisms underlying PCa development in this tumour subgroup. Nevertheless, for fusion negative samples these mechanisms are less clear. Given the low frequency of somatic mutations and the fact that recurrent mutations are rare (BARBIERI 2012, KRAWITZ 2010, TIMMERMANN 2010), epigenetic causes might be dominating in tumour development (COOLEN 2010, YEGNASUBRAMANIAN 2008).

1.9 Epigenetics: Histone modifications and DNA methylation

The term ‘epigenetics’ describes mechanisms causing heritable changes in phenotype that are not reflected by changes in the underlying DNA sequence. The two dominating mechanisms are histone modifications (JENUWEIN 2001) and DNA methylations (BIRD 1986). During histone modifications several amino acids in the histone tails can be covalently modified by methylation, acetylation, phosphorylation (JENUWEIN 2001), sumoylation (SHIIO 2003) and ubiquitination (JASON 2002). The resulting ‘histone code’ influences chromatin condensation leading to an altered access for the transcription machinery.

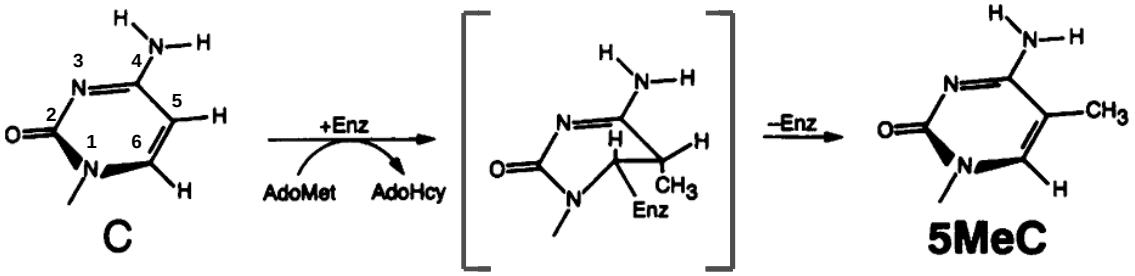


Figure 4. Molecular mechanism of cytosine methylation by DNA methyltransferases. The methylgroup donor S-adenosyl-methionine (AdoMet) is converted to S-adenosyl-homocystein (AdoHcy). The intermediate enzyme complex is shown in brackets. The enzyme is eliminated from the complex forming 5-methyl-cytosine. Figure modified from Smith *et al.* (SMITH 1992).

DNA methylation on the other hand mainly occurs in CpG-dinucleotides at the 5`-carbon of the cytosine. In human embryonic stem cells also non-CpG cytosine methylations are common (LISTER 2009b). Methylation of CpG-dinucleotides is mediated by DNA-methyltransferases (DNMTs) (JAENISCH 2003). DNMTs are mainly active during embryonic development and DNA replication: Shortly after fertilisation the parental methylation patterns are removed. Following implantation, the de-novo DNA methyltransferases *DNMT3A* and *DNMT3B* re-establish the DNA methylation pattern (OKANO 1999). During replication the DNA gets hemimethylated. The maintenance DNA-methyltransferase *DNMT1* (LEI 1996) recognises these hemimethylated sites and methylates the unmethylated strand (LEONHARDT 1992).

The enzymatic transfer of the methyl group to the cytosine was characterised by Smith *et al.* in 1992 (SMITH 1992): In a nucleophilic attack at the C6 of the cytosine the DNA methyltransferase allows the addition of a methyl group of S-adenosyl-methionine (SAM) (reacting to S-adenosyl-homocystein) at the activated C5 forming a dihydro-methylcytosine intermediate. Finally, the enzyme is eliminated from the complex and the double bond between C5 and C6 of the methyl-cytosine is regenerated (**Figure 4**).

CpG methylation has a critical function during development – knock outs of the DNMTs result in embryonic death (LI 1992), and mutations in DNMTs can cause severe syndromes (immunodeficiency, centromere instability and facial anomalies (ICF) syndrome, e.g.) (JAENISCH 2003). The central role of DNA methylation in cell homeostasis and development is further illustrated by the fact that, although methylated CpGs are prone to oxidative desamination resulting in transitional mutations, more than 80% of the CpGs in the human genome are methylated. This methylation spares high CpG content areas, called CpG-islands or Bird's islands, mainly found in genic or gene promoter regions (BIRD 1986,

CROSS 1995). According to an early definition, CpG islands have a minimum length of 200bp, a G+C content of >50% and an observed/expected CpG content of at least 0.6 (GARDINER-GARDEN 1987) in contrast to a genome-wide CpG content of 0.2 (observed/expected). Methylation of these islands and adjacent regions (CpG island shores) can be part of normal differentiation but is also repeatedly found in cancer (see below) and is thought to suppress gene expression. Methylated CpGs can be recognised by methyl-CpG-binding proteins that in turn recruit histone deacetylases which leads to a condensation of the chromatin and repression of the methylated genomic area. Mutations in these binding proteins can bring up severe phenotypes like the Rett syndrome caused by a mutation in the X-chromosomal *methyl CpG binding protein 2* (*MECP2*) gene (AMIR 1999).

Recently, an additional DNA modification and its epigenetic implications have been discovered – 5-hydroxymethylation of cytosines (5-hmC) (PASTOR 2011). 5-hmCs are formed by oxidation of 5-methylcytosines and potentially constitute a preliminary step in active DNA-demethylation (TAHILIANI 2009).

1.10 DNA methylation and tumourigenesis

In general, cancer epigenomes are characterised by genome-wide hypomethylation and hypermethylation of distinct CpG rich promoter sites as well as CpG island shores (EHRLICH 1982, FEINBERG 1983, FLATAU 1983, IRIZARRY 2009). Hypermethylation can promote tumour progression by causing repression of tumour suppressor and apoptotic genes, reactivation of embryonic developmental genes or inflicting changes in miRNA expression or promoter usage (EGGER 2004, FEINBERG 2006, JONES 2002). DNA hypomethylation on the other hand can support increased recombination at repetitive elements and reactivation of endogenous retroelements (WALSH 1998). Furthermore, due to aberrant DNA methylation imprinting information might get lost (NELSON 2009). It is noteworthy that during cancer development many developmental and differentiation inducing genes that are responsible for differential tissue development in normal tissue become aberrantly methylated (IRIZARRY 2009).

Several genes are described to be hypermethylated in a large variety of tumours: *adenomatous polyposis coli* (*APC*) (HILTUNEN 1997) in human colorectal carcinoma, *RB* in retinoblastoma (GREGER 1989), *cyclin-dependent kinase inhibitor 2A* (*CDKN2A*) in lung cancer (OTTERSON 1995), apoptotic genes like *BCL2* and *pyrin domain and caspase*

recruitment domain containing (*PYCARD*) in PCa (CARVALHO), DNA repair genes like *mutL homolog 1* (*MLH1*) in colorectal carcinoma (HERMAN 1998) or *breast cancer 1* (*BRCA1*) in breast cancer (DOBROVIC 1997), and genes controlling cell metabolism like *runt-related transcription factor 3* (*RUNX3*) in esophageal squamous cell carcinoma (LONG 2007).

DNA hypomethylation, on the other hand, is mainly described to occur in satellite DNA of centromeric regions, Alu regions and long interspersed elements (LINE)-1 (EHRLICH 2009). It is thought that demethylation of these elements as well as demethylation of retrotransposons might lead to chromosomal instability, increased recombination events or copy number variations (HOWARD 2008, RODRIGUEZ 2006, WALSH 1998). Less common is gene promoter hypomethylation resulting in increased transcription: It is described for the growth promoting genes *related RAS viral oncogene homolog* (*RRAS*), *serpin peptidase inhibitor clade B member 5* (*SERPINB5*), *S100 calcium binding protein A1* (*S100A1*), *melanoma antigen family* (*MAGE*), and *insulin like growth factor 2* (*IGF2*) (SHARMA 2009).

1.11 DNA methylation in PCa

Aberrant DNA methylation is proposed to be one of the earliest events in prostate tumourigenesis (FEINBERG 1983, GAMA-SOSA 1983, SCHULZ 2009, YEGNASUBRAMANIAN 2008); the most prominent one being a hypermethylation of *glutathione S-transferase pi 1* (*GSTP1*) exon 1 (LEE 1994, NAKAYAMA 2003). An assay for measuring the DNA methylation level of *GSTP1* is already included in clinical trials (HOPKINS 2007, LEE 1994). Other genes with changes in promoter methylation in some cases of PCa are summarised in the review of Li and Okino (LI 2004) and include *multi drug resistance protein 1* (*MDR1*), *O-6-methylguanine-DNA methyltransferase* (*MGMT*), *Ras association domain family member 1* (*RASSF1*), *retinoic acid receptor B* (*RARB*), *adenomatous polyposis coli* (*APC)*, *androgen receptor* (*AR*), *cyclin-dependent kinase inhibitor 2A* (*CDKN2A*), *cadherin 1* (*CDH1*), and *cluster of differentiation 44* (*CD44*). However, they show a wide range of methylation states over the studies performed and further analyses are required.

It has been suggested that during PCa progression hypermethylations are induced by the action of *enhancer of zeste homolog 2* (*EZH2*), a polycomb group protein (KARANIKOLAS 2010, YU 2007, YU 2010). *EZH2* is responsible for silencing of homeobox genes through

H3K27-methylation during tissue development (CAO 2002, HOBERT 1996, HOFFMANN 2007, KUZMICHEV 2004, PIRROTTA 1998). The consecutive hypermethylation of these target genes links histone modifications to DNA methylations (VIRE 2006). *EZH2* overexpression is associated with invasiveness and high malignancy of several tumours including PCa (ALAJEZ 2010, KARANIKOLAS 2010, SCHLESINGER 2007, VAN LEENDERS 2007, VARAMBALLY 2002, YANG 2009, YU 2007, YU 2010).

Unlike DNA hypermethylations, which occur early in PCa development, DNA hypomethylations seem to arise in late tumour phases mainly in metastatic stages. Yegnasubramanian *et al.* proposed a model in which the necessity to maintain promoter hypermethylation in tumour suppressor genes of tumour cells might lead to an overall demethylation of other regions that causes heterogeneity between the hypomethylation patterns of different metastases in a single patient while the hypermethylated regions are more uniform (YEGNASUBRAMANIAN 2004, YEGNASUBRAMANIAN 2008).

1.12 MicroRNAs

DNA methylation can inflict changes on the expression of genes but as well on short regulatory transcripts called microRNAs (miRNAs). This adds another level of cellular regulation and opens additional pathways for tumour formation.

Processed miRNAs are small (17-25nt) highly abundant single stranded RNAs that can regulate gene expression posttranscriptionally (BARTEL 2004, CAZARES 2011, LEE 2009). MiRNA genes are arranged in clusters or as single genes and are transcribed by *RNA polymerase II*. Primary transcripts of miRNA genes (pri-miRNA) are cleaved by nuclear *Drosha RNase III endonuclease* to a ~70nt stem loop intermediate (pre-miRNA). An alternative mechanism involves the usage of regular splicing machinery by miRNAs found in introns of genes. The pre-miRNA is transported to the cytoplasm where it is further processed by *Dicer* producing an imperfectly double stranded miRNA duplex. One strand is called the miRNA*, the other one the miRNA. One of the strands is loaded into the RNA induced silencing complex (RISC) containing Argonaute family proteins. Pairing of the 5'-end of the miRNA portion of the RISC with complementary sequences for example in the 3'UTR of target mRNAs leads to either degradation of the mRNA (highly specific pairing) or translational repression (imperfect pairing).

1.13 MiRNAs and methylation in cancer

Aberrant expression of miRNAs with functions in cancer pathways is a common feature of tumour development. For example, tumour suppressive *miR15a* and *miR16-1* target anti-apoptotic *BCL2* and are downregulated or deleted in chronic lymphatic leukaemia (CLL) as well as in PCa (BONCI 2008, CIMMINO 2005), while oncogenic androgen regulated *miR125b* suppresses pro-apoptotic *BAK1* and is found to be upregulated in PCa (SHI 2007). *MiR21* is upregulated in almost all kinds of cancers (LEE 2009) and targets the tumour suppressor *Ras homolog B* (*RHOB*) (LIU 2011a).

MiRNA expression can be regulated by DNA methylation (HAN 2007): For example, DNA hypermethylation causes downregulation of *miR124a* in colon cancer cell lines (LUJAMBIO 2007) and of *miR145* in prostate cancer (SUH 2011).

Interestingly, on the other side, miRNA expression can also influence DNA methylation levels, e.g. *miR29* targets *DNMT3A* and *3B* (FABBRI 2007). *MiR26a* regulates *EZH2* expression and has been described to be deregulated in nasopharyngeal carcinoma (ALAJEZ 2010, LU 2011), breast cancer (ZHANG 2011a) as well as in a murine lymphoma model (SANDER 2008). As mentioned above, the polycomb group protein *EZH2* is often found upregulated in PCa (HOFFMANN 2007) and its target genes have been shown to become subsequently methylated and silenced by DNA methyltransferases (YU 2007, YU 2010). Thus, *EZH2* links aberrant *miR26a* expression to DNA and histone methylation.

1.14 Technologies for high throughput DNA methylation analysis

Since previous studies mainly concentrated on assessing the methylation state of a few preselected regions or used small patient cohorts (FEBER 2011, HILL 2010, KIM 2011) there is a great demand for a comprehensive genome-wide analysis of DNA methylation in a large patient cohort since DNA methylation promises to be a valuable marker for prostate cancer early detection and diagnosis.

1.14.1 Multiparallel Sequencing

The emergence of multiparallel sequencing (MPS) systems like the ‘Sequencing by Oligonucleotide Ligation and Detection’ (SOLiD)-system of Applied Biosystems/Lifetech or Illumina’s Genome Analyzer has created the possibility of genome-wide DNA

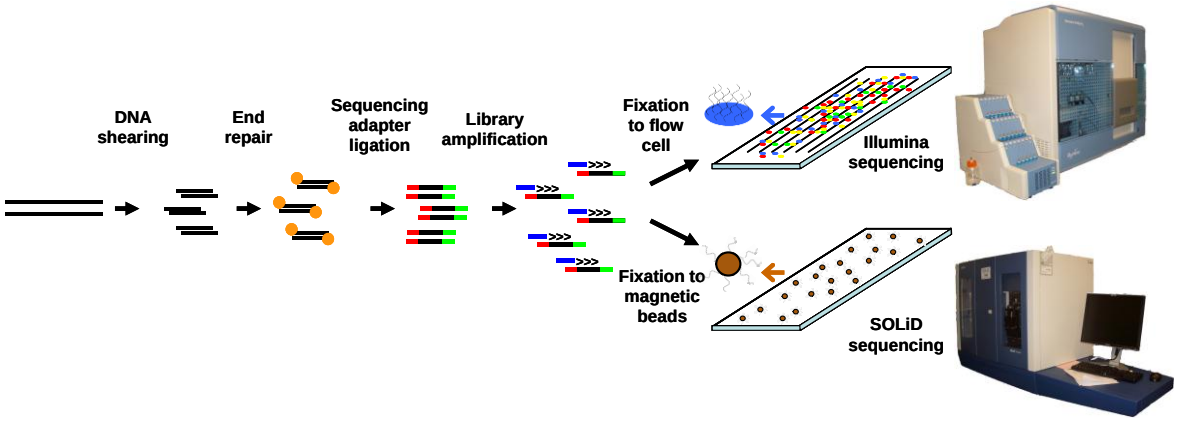


Figure 5. Workflow of MPS library preparation and sequencing. Genomic DNA is sheared, the ends are repaired and sequencing adaptors are ligated. In a PCR step the sequencing library is amplified. Following this, the library is either directly fixed to a flow cell through bridge PCR (cluster generation; Illumina) or to paramagnetic beads in an emulsion PCR reaction (SOLiD). Paramagnetic beads are subsequently deposited on a flow cell. The sequences of the monoclonal clusters or beads are read out during a polymerase (Illumina) or ligase (SOLiD) based second strand synthesis using fluorescence labelled nucleotides.

methylation analyses without any preselection of regions and allows the elucidation of yet unknown mechanisms and markers. The main idea of the MPS technologies is the fixation of millions to billions of short DNA fragments on a solid phase and their multiparallel read out using fluorescence labelled nucleotides/oligonucleotides (**Figure 5**).

For MPS, genomic DNA is sheared and universal sequencing adapters are ligated to the ends of the fragments. The resulting library is amplified using standard PCR. In Illumina's Genome Analyzer, the amplified library is applied to a flow cell containing fixed oligonucleotides complementary to the universal adapters of the library. During several rounds of PCR monoclonal clusters are generated from each fragment of the library. These clusters can be visualised (sequenced) by sequential annealing of the four fluorescence labelled nucleotides.

In SOLiD sequencing, the amplified library is mixed with more than one billion magnetic beads. In a water-in-oil emulsion PCR (ePCR) the library is amplified and thereby fixed to the beads in a way that each bead carries more than 30,000 monoclonal fragments. In a consecutive step these beads are enriched while beads without amplicons are depleted. Up to 700 million (SOLiD version 4) of the enriched beads are loaded onto a sequencing slide (DAHL 2010). Sequencing is performed by the sequential ligation of oligonucleotides. These MPS systems now have a throughput of >100Gb/run that allows for example the resequencing of the human genome with 30x coverage on one machine in one week.

1.14.2 DNA methylation analyses

DNA methylation analysis methods can be divided into direct and indirect methods (BOERNO 2010). The direct methods allow the assessment of the actual methylation state of every single CpG while the indirect methods interrogate the methylation state of whole regions instead of single CpGs and include an enrichment step.

The gold standard for DNA methylation analysis is the direct whole genome bisulphite sequencing (BS-Seq), because it provides digital information (methylated or unmethylated) for every single CpG (LISTER 2009a). The underlying principle of this technology is the selective conversion of unmethylated cytosine into uracil by sodium bisulphite (sodium hydrogensulphite) while methylated cytosines are unaffected (CLARK 1994) (**Figure 6**). This procedure conserves the epigenetic methylation information by converting it into a genetic alteration that is accessible by sequencing. The ratio of 'CG' to 'TG' reads at a former 'CG' position directly mirrors the amount of methylated to unmethylated cytosines at this position.

Because of the high resolution, whole genome BS-Seq is very time consuming and expensive and so far unsuitable for large sample cohorts. Additionally, the reduced complexity of the genome through the conversion (in most regions three instead of four bases) leads to difficulties in mapping and demands high bioinformatics skills. That is aggravated by partly incomplete conversion of unmethylated cytosines mainly in high GC content areas.

There are other methods published that restrict the size of the investigated genome that has to be sequenced and later on mapped. One example is reduced representation bisulphite sequencing (RRBS) (MEISSNER 2005) that involves an enzymatic digest of the DNA prior

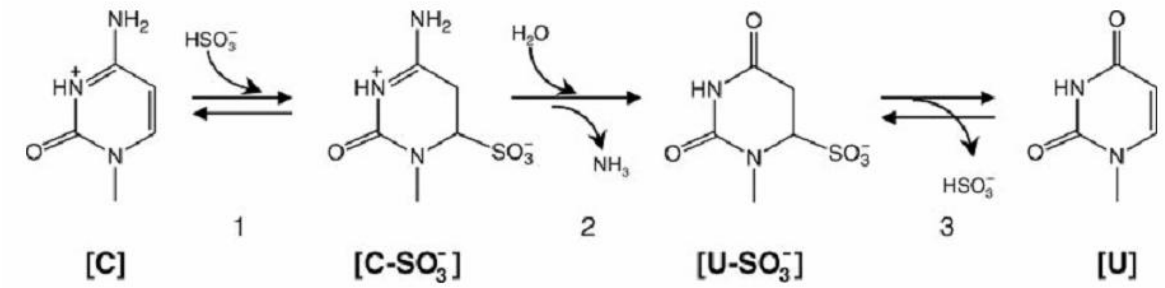


Figure 6. Molecular mechanism of bisulphite conversion of unmethylated cytosines. In a first step hydrogen sulphite is added to the C5-C6 double bond under acidic conditions forming 5,6-dihydrocytosine-6-sulphonate. Then, the product is hydrolytically desaminated giving rise to 5,6-dihydrouracil-6-sulphonate. In methylated cytosines the deamination reaction is two orders of magnitude slower than in unmethylated cytosines due to the electron donating methyl group at C5, leaving 5-methylcytosines nearly unaffected by BS-conversion. Under alcalic conditions the hydrogensulphite group is eliminated to form uracil (desulphonation reaction) that can be distinguished from cytosine in sequencing, mass spectrometry analysis or by a base specific restriction digest. Image taken from Hayatsu *et al.* (HAYATSU 2008).

to BS conversion to enrich high CpG areas. Further technologies are the PCR amplification and pyrosequencing of regions of interest following BS conversion (TOST 2007) or analyses of PCR fragments by mass spectrometry. The bisulphite specific padlock probe (BSPP) method of Ball *et al*. (BALL 2009) is another promising approach: Several thousand probes are designed to target every single CpG in a region of interest. Those probes are hybridised to the BS converted DNA in a horseshoe-like manner to finally amplify the targeted region and read out its methylation state. Nevertheless, all these technologies go hand in hand with loss of potentially important information. The methods have been reviewed in detail by Boerno *et al.* (BOERNO 2010).

For an almost unbiased cost effective high throughput screening, indirect immunoprecipitation methods like MeDIP-Seq (methylated DNA immunoprecipitation sequencing) or MBD-Seq (methyl CpG binding protein domain sequencing) are the methods of choice (KESHET 2006, LI 2010, WEBER 2005). After shearing of the DNA, methylated fragments are enriched using either an anti-5-meC antibody (MeDIP) or domains of the methyl CpG binding protein (MBD). The resulting DNA fraction ideally contains only methylated regions that – after sequencing – can be mapped to their genomic position. The resolution of these techniques is limited by the fragment size, thus no individual CpG read-out can be obtained. Another problem is that they are slightly biased towards high CpG content regions (LI 2010). Algorithms addressing this bias have been developed by Chavez *et al.* and Down *et al.* (CHAVEZ 2010, DOWN 2008).

Nevertheless, these technologies are very useful for screening large sample cohorts for interesting regions that later on may be BS sequenced to elucidate the actual methylation state of every single CpG.

With 3rd generation nanopore sequencing it might become possible to directly read out the methylation state of every cytosine without precedent BS conversion (WALLACE 2010). The principle of these technologies is a nanopore through which single DNA molecules are passed. During the passage the individual nucleotides induce different alterations in electric current that can be measured and directly converted into the underlying base sequence. Interestingly, methylated and hydroxymethylated cytosines show a different profile than unmodified cytosines and thus can be directly assessed.

1.15 The 'Interessengemeinschaft' Prostate Cancer (IGP)

Figure 7. The logo of the IGP consortium.

The IGP has been a BMBF funded consortium of urologists, pathologists, scientists, and representatives of the industry (**Figure 7**). One of the main goals was the identification and validation of markers for detection of early stage PCa in patient tissue, serum and urine. To achieve this, proteome, transcriptome, epigenome, miRNA, and genome data (including mutational analyses and copy number alterations) of more than one hundred prostate samples was generated exploiting novel sequencing and chip technologies. Furthermore, the integration of these various data levels should help to elucidate the molecular mechanisms underlying PCa formation and develop optimal therapeutic strategies for treatment. Our part of the work consisted of the acquisition of epigenome data and its integration with transcriptome and miRNA expression. Another scope was the identification and validation of epigenetic biomarkers.

1.16 Research Outline

In this work I present the results of a high throughput whole genome DNA methylation screen involving 51 tumour and 53 normal prostate samples. To identify tumour and tumour subgroup specific epigenetic alterations I used an immunoprecipitation-based enrichment approach for methylated DNA regions followed by SOLiD sequencing (MeDIP-Seq) (DOWN 2008, FEBER 2011, KESHET 2006, WEBER 2005). Therefor, I established the MeDIP-Seq protocol for the Applied Biosystems (Life) SOLiD sequencing system. I extracted hundreds of differentially methylated regions with the potential to serve as novel biomarkers for PCa detection. I performed small scale studies to evaluate the potential of these markers in additional patient cohorts. Furthermore, I integrated the MeDIP-Seq data with gene and miRNA expression data to gain further insight into the pathomechanisms of prostate cancer development.

2 METHODS

2.1 Routinely used laboratory materials and methods

Table 3: Buffers used for standard laboratory methods.

50x TAE	10x PBS (pH7.4)	TE (pH8)	Proteinase K buffer (pH8)
242g Tris	1.4M NaCl	10mM Tris	100mM NaCl
57.1ml glacial acetic acid	34mM KCl	1mM EDTA	10mM Tris-Cl
18.6g EDTA	0.1M Na_2HPO_4		25mM EDTA
ad 1l: bidestilled water	18mM KH_2PO_4		0.5% SDS

Table 4: Equipment and consumables used for standard laboratory methods.

Article	Catalog Number	Distributor
Benchtop centrifuge 5424	5424 000.410	Eppendorf
Centrifuge 5804	5805 000.017	Eppendorf
Centrifuge 5810R	5811 000.010	Eppendorf
Eppendorf Research 10µl		Eppendorf
Eppendorf Research 1ml		Eppendorf
Eppendorf Research 20µl		Eppendorf
Eppendorf Research 200µl		Eppendorf
Expression 1640 XL Scanner	E1640XL-GA	Epson
Gel imaging System Alpha Imager		Biozym
Mini Protean	165-8001	Biorad
Nanodrop	ND-1000 and ND-2000	Thermo Scientific
pH meter Toledo MP220	Z311286DK-1EA	Mettler
Pipetman 2µl		Gilson
PMR-30 Shaker	583-205	Grant-bio
Power supply unit EV202	55-EV202	Peqlab
Power supply unit PowerPac Basic	164-5050	Biorad
Rocky Shaker		Frobel Laborgeräte
Scales	EMB200-2	KERN
Tumbler		self-made
Acrylamide /Bis solution 37.5:1 (30%)	161-0154	Biorad
Adenosine-5´-triphosphate	P0756L	New England Biolabs
Agarose Standard	3810.2	Roth
Ammonium acetate	1.011.161.000	Merck
Ammoniumpersulfat (APS)	2300-25GM	Merck
Buffer QG Solubilization and Binding Buffer (with pH indicator)	19063	Qiagen
DNA LoBind Tubes 0,5ml	30.108.035	Eppendorf
DNA LoBind Tubes 1,5ml	30.108.051	Eppendorf
dNTP Set	BIO-39025	Bioline
EDTA Dinatriumsalz Dihydrat (Titrierkomplex III)	X986.2	Roth
Eppendorf Tips	30.077.024	epTIPS Filter
Ethanol vergällt	1.009.742.511	VWR
GenElute-LPA	56575-1ml	Sigma
GeneRuler 100bp DNA Ladder	SMO249	Fermentas
GeneRuler 1kb DNA Ladder	SMO319	Fermentas
H_2O bidest		self-made
H_2O millipore		self-made
HotStarTaq DNA Polymerase	203205	Qiagen
Isopropyl alcohol	19782-500ML	Sigma
Low Molecular Weight DNA Ladder 250µg	N3233L	New England Biolabs
$MgCl_2$ solution (50mM)	BIO-37026	Bioline
MinElute PCR Purification Kit	28006	Qiagen
Nuclease-free water	AM9938	Ambion
Orange DNA Loading Dye Solution (6x)	R0631	Fermentas

Article	Catalog Number	Distributor
PCR SoftStrips 0.2ml	710990	Biozym
PCR strips	30.124.359	Eppendorf
PCR strips	732-0545	VWR
PCR-05-C 0.5ml thin wall, clear, flat cap	732-0675	VWR
Pfu DNA Polymerase (recombinant) 2,5u/µl	EP0502	Fermentas
Phase lock Gel Tube Heavy 2ml	2302830	5Prime
Phase lock Gel Tube Light 2ml	2302820	5Prime
Phenol/Chloroform/Isoamylalcohol; 25/24/1	A0889.0100	Applichem
Platinum Taq DNA Polymerase	10966-034	Invitrogen
Polypropylene tubes (15ml)	430052	Corning
Polypropylene tubes (50ml)	430829	Corning
Proteinase K, recombinant, PCR Grade	3115852004	Roche
QIAquick Gel Extraction Kit	28706	Qiagen
QIAquick PCR Purification Kit	28106	Qiagen
Quant-iT dsDNA HS Assay Kit	Q32851	Invitrogen
Qubit assay tubes	Q32856	Invitrogen
SafeLock tubes (0.5ml)	0030 121.023	Eppendorf
SafeLock tubes (1.5ml)	0030 120.086	Eppendorf
SafeLock tubes (2ml)	0030 120.094	Eppendorf
Serological Pipettes (10ml)	4488	Costar
Serological Pipettes (25ml)	4489	Costar
Sodium acetate	S2889	Sigma
Taq DNA Polymerase, Recombinant	10342-020	Invitrogen
TEMED (Tetramethylethylendiamin)	15524-010	Invitrogen
TipOne RPT Filter Tips 10µl XL (sterile)	S1180-3810	STARLAB
TipOne RPT Tips 10µl XL	S1160-3800	STARLAB
TipOne RPT Tips 10µl XL, Refill	S1160-3700	STARLAB
TipOne RPT Tips 200µl,konisch, Refill	S1161-1700	STARLAB
TipOne RPT Tips 200µl, konisch	S1161-1800	STARLAB
TipOne XL RPT 10x96 Tips, 1250µl	S1161-1820	STARLAB
TipOne XL RPT 10x96 Tips, 1250µl, Refill	S1161-1720	STARLAB
Tween 20, 100ml	4974.01	Applichem
Water Molecular Biology grade, Nuclease and Protease free	W4502	Sigma

2.1.1 PCR for amplification of standard fragments

For amplification of standard fragments (e.g. for colony PCR) the following protocol was applied for 50µl total reaction volume: 5µl of 10x PCR buffer (650mM Tris-HCl, 166mM $(NH_4)_2SO_4$, 31mM $MgCl_2$, 0,1% (v/v) Tween 20, pH8.8), 1µl dNTPs (2.5mM each), 2µl of primermix (5mM each), 1µl of Taq polymerase (1U/µl), and nuclease free water (NFW) were mixed on ice with 10-100ng of DNA or a picked bacterial colony and thermal cycling was performed (initial denaturation: 5' 95°C; 40cycles of 30s 95°C, 30s annealing (T=(Melting temperature primer 1 + melting temperature primer 2)/2-2°C)), elongation at 72°C for 1min/1kb; 5' 72°C; forever 4°C). Fragments were analysed on agarose gels.

2.1.2 Phenol/chloroform/isoamylalcohol extraction of DNA and ethanol precipitation

Prior to use, phase lock gel (PLG)-tubes (**Table 4**) were centrifuged at 12,000g for 30s to pellet the gel. The samples containing cell lysate were transferred to the PLG Heavy tube and mixed with the same volume of phenol/chloroform/isoamylalcohol (25/24/1) without vortexing followed by a centrifugation step at 12,000g. The supernatant was mixed with the same volume of chloroform and centrifugation was repeated. Optionally, another volume of chloroform can be added followed by an additional centrifugation step. The supernatant was subjected to ethanol/ammonium acetate precipitation. Therefor, 3-10µl of linear polyacrylamide (LPA) were added to the sample. For a typical volume of 200µl of DNA solution 50µl (0.25 volumes) of a 10M ammonium acetate solution were added to obtain a concentration of 2M. 500µl (2.5 times the initial volume) of -20°C cold 100% ethanol (EtOH) were added (final concentration 66%) and the mixture was incubated for at least 30min at -20°C followed by 30min centrifugation at >20,000g (Eppendorf centrifuges). The pellet was washed with 500µl 95% EtOH and centrifuged for 20min at >20,000g. Washing and centrifugation might be repeated. The supernatant was discarded and the pellet was dried for 15min at RT to evaporate any residual EtOH. The pellet was resuspended in elution buffer (EB) from Qiagen's Qiaquick PCR purification kit.

2.1.3 DNA purification using Qiagen MinElute/PCR purification kits

DNA was mixed with 5 volumes of buffer PB from Qiagen's Qiaquick PCR purification kit/MinElute kit (**Table 4**). The pH value was adjusted using 3M sodium acetate until the indicator turns yellow. The mixture was transferred to Qiagen's Qiaquick MinElute/PCR purification kit columns and centrifuged for 30s at maximum speed. The flow-through was reloaded onto the column, centrifuged again and then discarded. Two rounds of washing with 500µl of PE buffer were performed. The flow-through was discarded; the columns were centrifuged for 30s at maximum speed and then transferred into a novel 2ml collection tube. The columns were centrifuged for 5min at maximum speed and transferred to 1.5ml DNA LoBind tubes. To ensure the evaporation of all remaining ethanol, the columns were incubated at 56°C with open lids for 5min. After addition of an appropriate volume of EB buffer the columns were incubated for 5min at room temperature to allow the complete elution of bound DNA and centrifuged for 1min at 12,000rpm. The DNA was stored at -20°C.

2.1.4 DNA fragment purification from agarose gels using Qiagen's GelExtraction kit

DNA fragments of the desired size were cut out from an agarose gel using a nuclease free scalpel on a UV table and transferred into a DNA LoBind tube. The gel slice was weighed, 3 volumes of buffer QG were added and the gel was solubilised by incubation on a thermal shaker. 1 volume of isopropanol was added and the mixture was transferred to Qiagen Qiaquick MinElute or Qiaquick PCR purification columns. Washing and elution were performed according to the standard Qiaquick PCR purification/MinElute kit protocol described above.

2.1.5 DNA fragment purification from polyacrylamide gels

For one gel (6%) (using the Bio-Rad Casting station with 1mm spacer) 4.16ml of water, 1.4ml 5xTBE, 1.4ml of 30% acrylamide/bis solution (37.5:1), 49µl APS and 2,45µl of TEMED were combined on ice and quickly filled into the prepared casting station. The gel was run in 1xTBE buffer using the MiniProtean Chamber and the PowerPac Basic (130V, ~0.5h). DNA fragments of the desired size were cut out and isolated from the gel using the crush and soak method (http://genepath.med.harvard.edu/~cepko/protocol/mike/D5.html): The polyacrylamide slice was crushed in a 1.5ml tube using a melted pipette tip as pestle and incubated overnight at 37°C in Crush and Soak Solution (500 mM ammonium acetate, 1% SDS, 0.1 mM EDTA). The next day, the tube was centrifuged for 10min at 14,000rpm, the supernatant was recovered and the pellet was mixed again with additional 500µl of Crush and Soak solution. After centrifugation, the recovered supernatants were combined and DNA was recovered by ethanol/ammonium acetate precipitation (see section 2.1.2).

2.1.6 Cell culture

Table 5: Cell lines and additional equipment and consumables used for cell culture.

Article	Catalog Number	Distributor
CaCO-2 cell line	HTB-37	ATCC
DU145 cell line	HTB-81	ATCC
LNCaP cell line	CRL-1740	ATCC
PC-3 cell line	CRL-1435	ATCC
RWPE-1 cell line	CRL-11609	ATCC
SW480 cell line	CCL-228	ATCC
SW620 cell line	CCL-227	ATCC
VCaP cell line	CRL-2876	ATCC
10ml Serological Pipettes	4488	Costar
25ml Serological Pipettes	4489	Costar
Cytoperm 2 Incubator	390-2430	Heraeus Instruments
Neubauer cell counting chamber	T728.1	Carl Roth GmbH
Pipetaid	61881	Drummond
Water bath WNB7	84198998	Memmert
Bad Stabil 100 mL	1-60953	NeoLab
cell culture dishes Φ 60x15mm	93060	TPP
cell culture flask 75ml	90076	TPP
cell culture plates 12well	92412	TPP
cell culture plates 6well	92406	TPP
DMEM (Medium)	F0415	Biochrom AG
DMSO	D8418-100ml	Sigma
Fetal bovine serum (FCS)	S01155.Lot611FF	Biochrom AG
Fetal bovine serum (FCS)	S0113.Lot0338T	Biochrom AG
FuGENE 6 Transfection Reagent	11815091001	Roche
HiPerfect Transfection Reagent	301704	Qiagen
L-Glutamine (Gln)	K0283	Biochrom AG
Lipofectamine 2000	11668019	Invitrogen
Lipofectamine RNAiMAX	13778150	Invitrogen
MEM (Medium)	F0325	Biochrom AG
PBS (- Ca, - Mg)	L1825	Biochrom AG
Penicillin/Streptomycin	A2213	Biochrom AG
Trypsin/EDTA	L2143	Biochrom AG

2.1.6.1 Thawing and seeding of cells

For culturing cells from a frozen stock, the frozen cells were slowly thawed and small amounts of freshly prepared growth medium were added stepwise. Cells were scraped off the frozen block and transferred into 5ml of medium until all cells of the stock were in solution. The suspension was topped with additional growth medium up to a volume of 10ml. Cells were centrifuged for 5min at 600rpm and the DMSO containing supernatant was discarded. Cells were resuspended in 7ml of medium containing additional fetal bovine serum (FCS) (e.g. for HEK293T cells: DMEM + 1% glutamine (Gln) + 1% Penicillin/Streptomycin + 20% FCS) following cell supplier's instructions and cultured in a T25 flask until reaching confluence (incubator settings: 37°C, 5% CO_2, 93% rH). After splitting they were grown in standard cell culture medium (e.g. for HEK cells: DMEM + 1% glutamine + 1% Penicillin/Streptomycin + 10% FCS) (**Table 5**).

2.1.6.2 Splitting cells

Before reaching confluence cells were split to allow continuous growth. Therefor, growth medium was aspirated and cells were washed with 10ml prewarmed PBS (37°C). After aspiration of PBS they were incubated with 1ml trypsin for several minutes at 37°C. Cell culture flasks should not be hit to avoid cell clumping. Cells were collected in 12ml growth medium supplemented with all necessary nutrients and diluted 1:3 – 1:4 in medium. 12ml of the diluted cells were seeded into a T75 cell culture flask.

2.1.6.3 Counting cells

After trypsinisation cells were collected in fresh cell culture medium. Three aliquots of 10µl were counted using a Neubauer cell counting chamber.

2.1.6.4 Isolation of genomic DNA from cultured human cells

T75 flasks of cultured human cells were washed with 10ml PBS and incubated with 1ml trypsin. Cells were collected with 10ml medium containing FCS to inactivate trypsin and centrifuged for 5min at 800rpm (~100g) to obtain a cell pellet.
10µl of proteinase K were freshly added to 2ml of proteinase K buffer (**Table 3**) and the pellet was incubated for 12h at 50°C in 300µl of this mixture. Nucleic acids were separated from the amino acids by phenol/chloroform/isoamylalcohol extraction and precipitated using ethanol and ammonium acetate. The resulting pellet was eluted in 70µl EB buffer and incubated with 1µl RNase for 60min at 37°C. The RNase digest was followed by phenol/chloroform/isoamylalcohol extraction and ethanol/ammonium acetate precipitation to yield purified genomic DNA.

2.1.7 RNA preparation techniques

Table 6: Additional consumables used for RNA isolation.

Article	Catalog Number	Distributor
AllPrep DNA/RNA/Protein Mini Kit	80004	Qiagen
RNase ZAP	AM9780	Ambion
RNase-free DNase set	79254	Qiagen
RNeasy MinElute Cleanup Kit	74204	Qiagen
TRIzol	15596-018	Invitrogen

Pipettes, workplace and gloves were treated with RNase inhibitors (RNaseZap) and experiments were performed on ice unless stated otherwise.

2.1.7.1 Isolation of RNA from cultured human cells using the RNeasy MinElute Cleanup Kit

For RNA isolation from a T75 flask of human cells, the growth medium was aspirated and the cells were washed with 10ml ice cold PBS. 7.5ml of Trizol (1ml/10cm^2) were added to each flask (**Table 6**). Cells were scraped off using a rubber policeman, transferred to a 50ml tube and kept on ice. Additional 7.5ml of Trizol were added and cells were homogenised by pipetting them up and down several times followed by 5min incubation at RT. 0.2 volumes of chloroform were added (here: 3ml), mixed by inverting the tube 4 times and incubated for another 2-3min at RT. Phases were separated by a 30min centrifugation step at 4000g at 4°C. The upper phase containing the RNA was transferred to a new tube containing 7.5ml isopropanol and 1µl LPA.
For RNA precipitation, the tube was incubated for 30min at -80°C and centrifuged for 45min at 5000g at 4°C. After discard of the supernatant the pellet was washed with 15ml of ice cold 75% ethanol and centrifuged at <7.500g at 4°C for 5min. The pellet was dried and dissolved in RNase free water. DNase digestion was performed by 10min incubation at RT: Up to 87.5µl of RNA solution were mixed with 10µl Buffer RRD and 2.5µl DNaseI stock solution (RNeasy MinElute Cleanup Kit) (**Table 6**) and RNase free water was added to 100µl. After DNase digest, the solution was mixed with 100µl chloroform and RNA was extracted using Phase Lock Gel tubes (Heavy) followed by a sodium acetate/EtOH precipitation. 1µl of LPA, 1/10th volume of 3M sodium acetate and 2 volumes of 100% ice cold EtOH were added (**Table 4**), the mixture was incubated for 40min at -80°C and subsequently centrifuged for 30min at 4°C (>20,000g). The pellet was washed on ice with 70% EtOH, dried and resuspended in 30µl RNase free water. The purified RNA was aliquoted and stored at -80°C.

2.1.7.2 Isolation of RNA from tissue using the Allprep DNA/RNA/Protein Mini Kit

For isolation of the RNA fraction from tumour tissue, the AllPrep DNA/RNA/Protein Mini Kit was used and RNA was prepared following the manufacturer's instructions (**Table 6**). In brief, up to 30mg frozen tissue were disrupted and homogenised in Buffer RLT. Cell

debris was pelleted by a 3min centrifugation step and the supernatant was transferred to an AllPrep DNA spin column and centrifuged. The DNA could later on be eluted from the column while the flow-through contained the RNA. 72/100th volumes of ethanol (100%) were added to the flow-through and transferred to an RNeasy spin column and centrifuged. The flow-through contains the protein and small RNA fraction that can be purified later on. The RNA on the column was treated with DNase: The column was washed with 350µl Buffer RW1 and 10µl DNaseI stock solution were mixed with 70µl Buffer RDD, added to the column and incubated for 15min at RT. After digestion one washing step with 350µl Buffer RW1 and two washings with 500µl Buffer RPE were performed. The column was dried by centrifugation and the RNA was eluted with 30-50µl RNase free water. The RNA was aliquoted and stored at -80°C.

2.1.7.3 Preparation of small RNAs for Small RNA Expression Kit (SREK)

Table 7: Additional consumables used for Small RNA expression kit library preparation.

Article	Catalog Number	Distributor
Small RNA Expression Kit (SREK)	4399434b	Applied Biosystems

For preparation of the small RNA fraction the flow-through of the Allprep RNeasy spin column containing the protein- and small RNA fraction (see section 2.1.7.2) was mixed with Buffer APP and centrifuged for 10min to precipitate the protein-fraction. 1 volume of 100% EtOH was added to the supernatant and the mixture was transferred to an RNeasy MinElute (Qiagen) spin column. The RNA was bound to the column by centrifugation and washed with 500µl Buffer RPE, followed by a washing step with 500µl 80% EtOH. The small RNA fraction was eluted into a new 1.5ml tube using 14µl of RNase free water and was used for Small RNA library preparation (SREK) after analysis on the Agilent Bioanalyzer.

2.1.8 cDNA synthesis

Table 8: Additional consumables used for cDNA synthesis.

Article	Catalog Number	Distributor
Recombinant RNasin Ribonuclease. Inhibitor	N2511	Promega
RNaseH, 2U/µl	18021-014	Invitrogen
Superscript III, reverse transcriptase	18080-044	Invitrogen

For first strand cDNA synthesis, 300ng of RNA, 250ng random hexamer primers (5'-NNNNNN-wobbles-3'; no of PTO bonds: 2; Position of PTO bonds (5'-3'): 4-6), 1µl

10mM dNTPs, and RNase free water were incubated in a 13µl reaction at 65°C for 5min and cooled immediately on ice.

4µl *first strand buffer* (5x), 1µl DTT (0,1M), 1µl RNasin (40 U/µl) and 1µl SuperScript III-reverse transcriptase were added and the following programme was run on a thermal cycler: 5' 25°C, 60' 50°C, 15' 70°C. Consecutively, an RNase H digest of the RNA was performed by adding 1µl of RNase H (2U/µl) followed by 20min incubation at 37°C and a 20min inactivation step at 65°C.

2.1.9 Second strand synthesis

Table 9: Additional consumables used for second strand synthesis.

Article	Catalog Number	Distributor
Klenow DNA Polymerase I	M021S	New England Biolabs
Superscript III,Double-Stranded cDNA Synthesis	11917-010	Invitrogen

For second strand synthesis 10µl of single stranded cDNA were mixed with 1µl random hexamers (100µM), 3µl dNTPs (40mM), 6µl Klenow fragment (5U/µl), 5µl NEB2 buffer, and 25µl NFW. Incubation was performed for 2h at 16°C and double stranded cDNA was purified using MinElute columns.

2.1.10 Sodium bisulphite conversion of DNA

Table 10: Additional consumables used for sodium bisulphite conversion.

Article	Catalog Number	Distributor
Epitect Bisulfite Kit	59104	Qiagen
EZ DNA Methylation Gold Kit	D5005	Zymo

Through treatment with sodium bisulphite (BS), unmethylated cytosines are converted into uracil while methylated cytosines remain unaltered. This leads to a conversion of the epigenetic methylation mark into a genetic feature that is stable through PCR amplification and can be assessed by Sanger and multiple parallel sequencing technologies, as well as by base specific restriction digest or mass spectrometry. For bisulphite conversion the EpiTect Bisulfite Kit (Qiagen) and EZ-DNA Methylation Gold kit (Zymo) were routinely used.

2.1.10.1 EpiTect Bisulfite Kit (Qiagen)

The kit can be used for the BS conversion of 1ng-2µg of DNA. According to the protocol 800µl of water were added to one tube of bisulphite mix (sufficient for 8 reactions). In a 200µl PCR tube 20µl of DNA were mixed with 85µl of prepared BS mix and added with 35µl of DNA-Protect buffer from the kit. To convert the DNA the following thermal cycler protocol was used: 1. Denaturation: 99°C/5min; 2. Incubation 60°C/25min; 3. Denaturation 99°C/5min; 4. Incubation 60°C/85min; 5. Denaturation 99°C/5min; 6. Incubation 60°C/175min; 7. Hold 20°C. After cycling, the converted DNA was transferred to a new 1.5ml tube and 560µl Buffer BL containing 10µg/ml carrier RNA were added and transferred to an EpiTect Spin column. The columns were centrifuged for 1min at >14,000g and the flow-through was discarded. The columns were washed by adding 500µl BW buffer and centrifuged for 1min at >14,000g. The flow-through was discarded. The DNA was desulphonated by adding 500µl buffer BD followed by 15min incubation at room temperature (closed column lids). The columns were washed twice with 500µl BW buffer, transferred to a new 2ml tube and again centrifuged for 5min at >14,000g to get rid of any residual wash buffer. The columns were transferred to a new 1.5ml tube and the DNA was eluted from the column with at least 20µl EB buffer.

2.1.10.2 EZ-DNA Methylation Gold kit (Zymo)

The kit can be used for conversion of 500pg-2µg of DNA (200-500ng optimal). CT conversion reagent and M-Wash buffer were prepared according to the manufacturer's protocol. 130µl of the conversion reagent and 20µl of DNA were incubated in a thermal cycler for 10min at 98°C, 2.5h at 64°C and up to 20h at 4°C. The product and 600µl M-Binding buffer were added to a Zymo-Spin IC Column and centrifuged at >10,000g for 30s. The flow-through was discarded and the column was washed with 100µl of M-Wash buffer. After washing the DNA was desulphonated by 20min incubation with 200µl M-Desulphonation buffer at RT and washed twice with 200µl M-Wash buffer. The DNA was eluted with 10-20µl M-Elution buffer.

2.1.10.3 PCR for validating BS conversion success of genomic DNA

Table 11: Primers used for validation of BS conversion.

Name	Sequence	Purpose
HLA-A_FW	TGGGTTTTTAGAGAAGTTAATTAGTG	Validation of BS conversion success
HLA-A_RV	CCAAATCTAAATCAAAACCAAAAC	

To control for the success of bisulphite conversions the primers HLA-A_FW and HLA-A_RV were used in the following PCR to yield an amplicon of 162bp:
5µl 10xPCR buffer, 4µl dNTPs (2.5mM each), 2µl $MgCl_2$ (50mM), 1µl HLA-A_FW (10µM), 1µl HLA-A_RV (10µM), 50ng DNA, 1µl Platinum Taq and NFW ad 50µl were mixed and PCR was performed under the following conditions: 95°C 10min; 40 cycles: 95°C 30s, 57°C 1min, 72°C 30s; 72°C 5min; 4°C forever. Fragments were subsequently analysed on a 2% agarose gel.

2.1.11 Real time quantitative PCR (qPCR)

Table 12: Additional equipment and consumables used for qPCR.

Article	Catalog Number	Distributor
7900 HT Fast Real Time PCR System	4329001	Applied Biosystems
EvaGreen	31000	Biotium
Gotaq qPCR Mastermix (2x)	A6002	Promega
Maxima SYBR Green/Rox qPCR Master Mix (2x) 1000 Rxn	K0222	Fermentas
MicroAmp Optical 384-Well Reaction Plate	4309849	Applied Biosystems
MicroAmp Optical 96-Well Reaction Plate	4306737	Applied Biosystems
MicroAmp Optical Adhesive Covers	4311971	Applied Biosystems
SYBR Green PCR Master Mix	4309155	Applied Biosystems

Real time qPCR is a method for quantitation of DNA amplicons or cDNA. Like in standard PCR a DNA region is exponentially amplified by successive cycles of denaturation, primer annealing, and DNA elongation. SYBR Green or EvaGreen bind to the emerging double stranded DNA resulting in a fluorescence signal (ZIPPER 2004). The cycle number at which the signal intensity passes a predefined threshold (e.g. the inflexion point of the sigmoidal fluorescence curve) is called C_t value. The C_t value is inversely correlated to the logarithmic amount of input DNA.

2.1.11.1 Primer design for real time qPCR

Primers for real time qPCR were designed using the Primer3 software (http://frodo.wi.mit.edu/) with the following restrictions: primer size 20-22nt; melting

temperature 63-65°C, GC content 40-60%, maximal PolyX-sites 2-3, length of amplicon 70-120bp. For cDNA quantification the amplicon should span an exon-exon junction to exclude background genomic DNA signal.

2.1.11.2 qPCR using 2xqPCR mix (AB/Invitrogen)

Real time qPCRs were performed in triplicates in a 384 well plate. Each well contained 5µl PCR mix, 3.5µl DNA (at least 5ng genomic DNA) and 1.5µl primer mix (5mM each). An initial incubation of 95°C for 10min is required to activate the polymerase, followed by 40 cycles of a two step cycle programme (95°C denaturation for 15s + 60°C anneal and extend for 1min) on the Applied Biosystems 7500 real time PCR cycler. A subsequent dissociation step was used to assess the purity of the PCR amplicon. C_t values were extracted and the median was calculated for each triplicate and was used for further calculations (e.g. determining library concentrations, enrichment, gene expression, methylation values).

2.1.11.3 qPCR using EvaGreen

For cost efficient large scale real time PCRs $1/10^{th}$ volume of 10x Platinum Buffer, $1/200^{th}$ volume of dNTPs (10mM each), $1/100^{th}$ volume of primermix (50µM each), $3/100^{th}$ volume of $MgCl_2$ (50mM), $1/200^{th}$ volume of 100x EvaGreen (Biotium) and $1/200^{th}$ volume of Platinum Taq polymerase (Invitrogen) were mixed with template DNA and filled with NFW. SYBRGreen could not be used as it inhibited PCR amplification. Standard real time cycling was applied.

2.1.11.4 MicroRNA specific TaqMan assays

Table 13: Additional consumables used for TaqMan assays.

Article	Catalog Number	Distributor
miR26a TaqMan MicroRNA Assay RT/TM405	4427975ID000405	Applied Biosystems
RNU44 TaqMan MicroRNA Assay RT/TM001094	4427975ID001094	Applied Biosystems
TaqMan RNA Reverse Transcription Kit	4366596	Applied Biosystems
TaqMan Universal PCR Master Mix, no Amperase UNG	4324018	Applied Biosystems

To quantify *miR26a* and *RNU44* (control) expression in RNA isolated from various cell lines TaqMan MicroRNA Assays (Applied Biosystems) were used. The assay consists of a reverse transcription step with *miR26a* or *RNU44* specific primers followed by the

quantification step on the Applied Biosystems RealTime PCR system. This step involves Taq mediated cleavage of FAM labelled *miR26a* or *RNU44* specific probes specifically hybridised to the target cDNA. TaqMan probes rely on the quenching of the fluorescence of a dye (like FAM) that is covalently bound to the 5'-end of the probe by a nonfluorescent dye bound to the 3'-end through Förster resonance energy transfer (FRET). 5'-3'-exonuclease activity of the polymerase leads to digestion of the probe during strand elongation and thus to release of the fluorescent dye from the quencher's vicinity resulting in increased fluorescence.

To generate cDNA, for each assay 0.15µl 100mM dNTPs, 1µl MultiScribe Reverse Transcriptase (50U/µl), 1.5µl 10x Reverse Transcription Buffer, 0.19µl RNase Inhibitor (20U/µl), and 4.16µl NFW were mixed on ice and 5µl of total RNA (1-10ng) were added. After addition of 3µl target specific RT primermix (5x) the mixture was incubated on ice for 5min followed by thermal cycling: 30min 16°C; 30min 42°C; 5min 85°C; hold at 4°C. Quantification of the targets was performed in triplicates. For every qPCR reaction 0.5µl TaqMan MicroRNA Assay (target specific, 20x), 0.67 µl cDNA, 5µl TaqMan 2x Universal PCR Master Mix No AmpErase UNG, and 3.83µl NFW were mixed and standard two step thermocycling was performed on the Applied Biosystems 7900HT fast real time PCR system in 9600 emulation mode: 10min 95°C activation; 40 cycles of 15sec 95°C, 1min 60°C.

Relative concentrations were determined using the $\Delta\Delta C_t$ method comparing C_t values of *miR26a* and *RNU44* in treated and untreated cells.

2.1.12 Whole genome amplification with GenomePlex WGA kits

Table 14: Additional consumables for whole genome amplification.

Article	Catalog Number	Distributor
GenomePlex Complete Whole Genome Amplification (WGA) Kit WGA2	WGA2-50RXN	Sigma
GenomePlex WGA Reamplification Kit WGA3	WGA3-50RXN	Sigma

For whole genome amplification of small DNA amounts WGA2 and WGA3 kits (Sigma-Aldrich) were used. Genomic DNA (>10ng) was fragmented, a library was generated that was subsequently amplified to yield 5-10µg of ~400bp DNA.

For fragmentation, the DNA was diluted to 1ng/µl and 1µl of 10x fragmentation buffer was added to 10µl of diluted DNA sample followed by incubation at 95°C for exactly 4min. After fragmentation the DNA was kept on ice. 2µl of Library Preparation buffer and

1µl of Library Stabilization Solution were added to each sample, incubated for 2min at 95°C and cooled on ice. 1µl of Library Preparation Enzyme was added and the following incubation programme was performed in a thermal cycler: 16°C for 20min; 24°C for 20min; 37°C for 20min; 75°C for 5min; 4°C hold.
The resulting library was amplified within three days by adding 7.5µl of 10x Amplification Master Mix, 47.5µl of NFW and 5µl of WGA DNA Polymerase to the library. 14 cycles of amplification were performed on a thermal cycler using the following protocol: Initial denaturation 95°C for 3min; 14 cycles of 94°C for 15s, and 65°C for 5min.
After amplification the GenomePlex library was purified using Qiagen's PCR purification kit and could be reamplified using the WGA3 kit to obtain more DNA.
Therefor, 10µl of 1ng/µl preamplified GenomePlex library were supplied with 49.5µl NFW, 7.5µl of 10x Amplification Master Mix, 3.0µl of 10mM dNTP mix and 5.0µl of WGA DNA Polymerase. The reaction was incubated in a thermal cycler using the following protocol followed by a cleanup of the products: Initial Denaturation 95°C for 3min; Cycling programme (14 cycles): Denaturation 94°C for 15s, Anneal/Extend 65°C for 5min.

2.2 Statistical analyses

Statistical analyses were conducted using Microsoft Excel 2003 and R (version 2.9.2) (R_DEVELOPMENT_CORE_TEAM 2009).

2.3 Analyses of DNA methylation patterns in prostate tissues

2.3.1 Biological samples

Prostate tissue samples were obtained from the University Medical Centre Hamburg-Eppendorf. Approval for the study was obtained from the local ethics committee and all patients agreed to provide additional tissue sampling for scientific purposes. Tissue samples from 51 prostate cancer and 53 normal prostate tissues were included (**Supplementary Table 1**). None of the patients had been treated with neo-adjuvant radio-, cytotoxic, or endocrine therapy. To confirm the presence of tumour, all punches were sectioned, and tumour cell content was determined in every 10^{th} section. Only sections containing at least 70% tumour cells were included in the study. Normal prostate tissue samples were obtained from 53 patients who underwent radical prostatectomy for prostate

cancer. Only sections containing exclusively normal tissue material with epithelial cell content between 20 and 40% were included in the study. For Laser Capture Microdissection (LCM, Zeiss, Germany) of epithelial cells, 16µm tissue sections were mounted on special LCM slides and briefly stained with hematoxilin and eosin to facilitate localization of epithelial cells. Epithelial cells were collected by LCM from 10 tissue sections each. DNA was isolated using the DNA mini kit (Qiagen) according to the manufacturer's instructions. The *TMPRSS2:ERG* fusion status was determined with RT-PCR following protocols from Jhavar *et al.* (JHAVAR 2008) and Mertz *et al.* (MERTZ 2007).

2.3.2 Methylation profiling by MeDIP-Seq

Table 15: Additional equipment and consumables used in MeDIP-Seq.

Article	CatalogNumber	Distributor
Agilent 2100 Bioanalyser	G2938C	Agilent Technologies
Blue metal plate adaptor		Applied Biosystems
Covaris S2 System	4392718	Applied Biosystems
Dynal MPC Magnetic Particle Concentrator (Magnetic Rack)	120.20D	Invitrogen
FugeOne	PC-100	STARLAB
GeneAmp PCR System 9700 (96 well)	N8050200+4314443	Applied Biosystems
heated water bath: Julabo F26		Julabo
IKA ULTRA TURRAX Tube Drive	4400335	Applied Biosystems
Incubator: Heraeus Function Line		Heraeus Instruments
Repeater Xstream	22460811	Eppendorf
Serological Pipettes (10ml)	4488	Costar
Serological Pipettes (25ml)	4489	Costar
SOLiD System 2.0	4392719	Applied Biosystems
SOLiD version 3		Applied Biosystems
SOLiD version 3 HiDRA		Lifetech
SOLiD version 3 plus		Lifetech
SOLiD version 4		Lifetech
Thermomixer Compact	5350 000.013	Eppendorf
Vortex Genie 2	SI-0256	Scientific Industries
5-Methylcytidine antibody purified	BI-MECY-1000	Eurogentec
Agencourt AMPure 60ml Kit	000130	Agencourt
Agencourt AMPure 60ml Kit	A29152	Beckman Coulter
Agencourt AMPure XP 60ml Kit	A63881	Beckman Coulter
Agilent Dna 7500 Kit	5067-1506	Agilent
Butanol	4389770	Applied Biosystems
Combitips Plus (10ml)	22496123	Eppendorf
Combitips Plus (5ml)	22496107	Eppendorf
Covaris micro TUBE with AFA fiber and snap cap	520045	Covaris
DNA polymerase I (10U/µl)	M0209L	NEB
Dynabeads M280 sheep anti-mouse IgG paramagnetic beads (Invitrogen)	112-01D	Invitrogen
Dynabeads M280-Streptavidin	SKU112-05D	Invitrogen
End-Repair Mix (Low Concentration)	Y914-LC-L	Enzymatics
ePCR tubes 15ml (IKA)	4400401	Applied Biosystems
GenElute-LPA	56575-1ml	Sigma
High Fidelity Platinum Taq Polymerase	11305-029	Invitrogen
Mineral oil	M5904	Sigma
Phenol/Chloroform/Isoamylalcohol; 25:24:1	A0889.0100	Applichem
Polypropylene wide mouth jars, 15ml	2118-9050	Nalgene
Proteinase K, recombinant, PCR Grade	3115852004	Roche
Reagent reservoir	R9259	Sigma

Article	CatalogNumber	Distributor
SOLiD Bead Deposition Kit	4387895	Applied Biosystems
SOLiD Bead Deposition Kit	4387895	Applied Biosystems
SOLiD Bead Enrichment Kit	4387894	Applied Biosystems
SOLiD Buffer Kit	4387918	Applied Biosystems
SOLiD DH10B Fragment Library Controls	4391889	Applied Biosystems
SOLiD EZ Bead Emulsifier E80	4452722	Applied Biosystems
SOLiD Fragment Library Oligo Kit	4401151	Applied Biosystems
SOLiD Light Source	4388441	Applied Biosystems
SOLiD Opti Fragment Library Sequencing Master Mix 35	4442218	Applied Biosystems
SOLiD ToP Fragment Barcoded Sequencing MM50	4452697	Applied Biosystems
SOLiD ToP Instrument Buffer Kit	4452688	Applied Biosystems
SOLiD ToP Sequencing Kit - Barcode Tag	4449308	Applied Biosystems
SOLiD ToP Sequencing Kit, MM50	4449388	Applied Biosystems
SOLiD XD Bead Enrichment Kit	4453663	Applied Biosystems
SOLiD XD Slide & Deposition Kit v2	4456997	Applied Biosystems
SOLiD Opti Fragment Library Sequencing Kit 5 bp Barcode Set	4442261	Applied Biosystems
T4 DNA Ligase rapid (600U/µl)	L-6030-HC-L	Enzymatics

Table 16: Oligonucleotides used in MeDIP-Seq.

Name	Sequence	Purpose
Multiplex P1 Adaptor A	5'-ATCACCGACTGCCCATAGAGAGGTT-3'	Adaptor 1 part 1 for BC Libraries
Multiplex P1 Adaptor B	5'-CCTCTCTATGGGCAGTCGGTGAT-3'	Adaptor 1 part 2 for BC Libraries
Multiplex P2 Adaptor BC 001 A	5'-CGCCTTGGCCGTACAGCAGCCTCTTACACAGAGAA TGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC001
Multiplex P2 Adaptor BC 001 B	5'-CTGCCCCGGGTTCCTCATTCTCTGTGTAAGAGGCT GCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC001
Multiplex P2 Adaptor BC 002 A	5'-CGCCTTGGCCGTACAGCAGACCACTCCCTAGAGAA TGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC002
Multiplex P2 Adaptor BC 002 B	5'-CTGCCCCGGGTTCCTCATTCTCTAGGGAGTGGT CTGCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC002
Multiplex P2 Adaptor BC 003 A	5'-CGCCTTGGCCGTACAGCAGTATAACCTATAGAGAA TGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC003
Multiplex P2 Adaptor BC 003 B	5'-CTGCCCCGGGTTCCTCATTCTCTATAGGTTATACTG CTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC003
Multiplex P2 Adaptor BC 004 A	5'-CGCCTTGGCCGTACAGCAGGACCGCATCCAGA GAATGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC004
Multiplex P2 Adaptor BC 004 B	5'-CTGCCCCGGGTTCCTCATTCTCTGGATGCG GTCCTGCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC004
Multiplex P2 Adaptor BC 005 A	5'-CGCCTTGGCCGTACAGCAGCTTACACCACAGA GAATGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC005
Multiplex P2 Adaptor BC 005 B	5'-CTGCCCCGGGTTCCTCATTCTCTGTGGTG TAAGCTGCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC005
Multiplex P2 Adaptor BC 006 A	5'-CGCCTTGGCCGTACAGCAGTGTCCCTCGCAGA GAATGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC006
Multiplex P2 Adaptor BC 006 B	5'-CTGCCCCGGGTTCCTCATTCTCTGCGAGGGACAC TGCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC006
Multiplex P2 Adaptor BC 007 A	5'-CGCCTTGGCCGTACAGCAGGGCATAACCCAGA GAATGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC007
Multiplex P2 Adaptor BC 007 B	5'-CTGCCCCGGGTTCCTCATTCTCTGGGTTATGCCC TGCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC007
Multiplex P2 Adaptor BC 008 A	5'-CGCCTTGGCCGTACAGCAGATCCTCGCTCAGAGAA TGAGGAACCCGGGGCAGTT-3'	Adaptor 2 part 1 for BC libs BC008
Multiplex P2 Adaptor BC 008 B	5'-CTGCCCCGGGTTCCTCATTCTCTGAGCGAGGATC TGCTGTACGGCCAAGGCG-3'	Adaptor 2 part 2 for BC libs BC008
4994_F (BRD1)	5-GGGAATATAAGGAGCGCACA-3	Positive Control for MeDIP enrichment
4994_R (BRD1)	5-TCGGTTAAAACGGTCAGGTC-3	Positive Control for MeDIP enrichment
8804_F (COQ3)	5-CGAGGCGTGAGTTATTCCTG-3	Negative Control for MeDIP enrichment
8804_R (COQ3)	5-CTCTTGTGGCTGAGCTCCTT-3	Negative Control for MeDIP enrichment
Multiplex PCR Primer 1	5'-CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTC GGTGAT-3'	Amplification of barcoded libraries
Multiplex PCR Primer 2	5'-CTGCCCCGGGTTCCTCATTCT -3'	Amplification of barcoded libraries
Library-PCR-Primer1	5'-CCACTACGCCTCCGCTTTCCTCTCTAT-3'	Quantification of amplified libraries
Library-PCR-Primer2	5'-CTGCCCCGGGTTCCTCATTCT-3'	Quantification of amplified libraries

For genome-wide analyses of the DNA methylation patterns in prostate samples and cell line DNA a modified MeDIP protocol was integrated into the fragment library preparation protocol for the SOLiD MPS system (**Figure 8**). Briefly, genomic DNA was fragmented, end repaired and barcoded sequencing adapters were ligated. The library was subjected to MeDIP enrichment, PCR amplified and coupled to sequencing beads via emulsion PCR. The beads were sequenced on a SOLiD 3+ system.

2.3.2.1 Library preparation

2.5µg of genomic DNA in 100µl EB were fragmented to 100 to 200bp using the Covaris S2 system (10% duty cycle, Intensity 5, 200 cycles/burst, 7 cycles à 1min). Column purification was performed and DNA was eluted in 50µl EB while shearing was confirmed by gel electrophoresis. Purified DNA was subjected to end repair using Enzymatics End Repair Mix: The final volume was 25/19x the DNA volume. For example, 45µl of DNA were mixed with 5.9µl dNTPs (1mM; 1/10th of final volume), 5.9µl 10x End Repair buffer and 2.4µl End Repair Mix (1/25th of final volume) and incubated 30min at RT followed by an enzyme inactivation step for 20min at 75°C and Qiagen purification with elution in 50µl EB. Barcoded sequencing adaptors were ligated in a 30fold excess in a total reaction

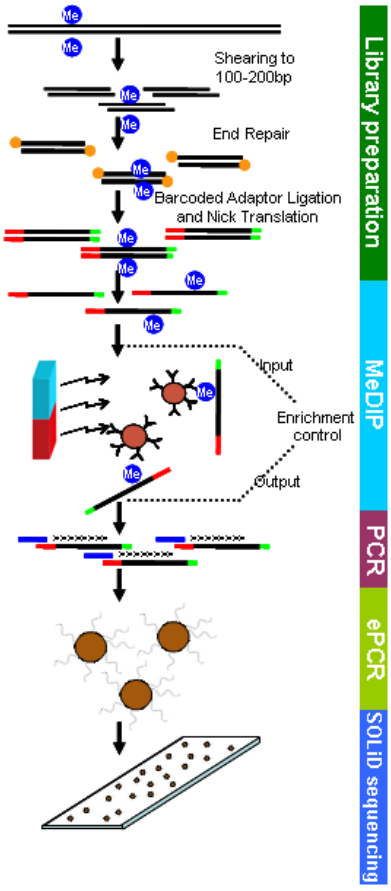


Figure 8. MeDIP-Seq protocol.
Genomic DNA of 51 tumour and 53 normal prostate samples was sheared to a size of 100-200bp using the Covaris S2 system and end repaired. Barcoded SOLiD sequencing adapters were ligated and DNA was heat denaturated. Single stranded DNA was enriched for methylated fragments using an anti-5-meC-antibody coupled to paramagnetic beads. Enrichment controls were performed. Enriched libraries were amplified in a library PCR reaction and quantified using qPCR. Equal molarities of each barcoded library were pooled into pools of eight different libraries that were subsequently bound to sequencing beads in an emulsion PCR reaction. Beads with bound DNA were enriched and loaded onto a SOLiD sequencing slide in a density of up to 500 million beads/slide. Sequencing was performed with 35bp+5bp sequencing chemistry on a SOLiD v3+ machine.

volume of 140µl. The amount (X in µl) of a 50µM adaptor stock was calculated using the following formula:

X= DNA in [µg] * 1,000,000/660 [g/(mol*bp)] * 30 /50[µmol/l] * 1/length in [bp]
(30x excess, 50µM adaptors; length~200bp)

45µl of DNA were mixed with X µl MP adaptor 1, X µl MP Barcode Adaptor, 70µl 2xRapid Ligation Buffer, 0.7µl T4 Ligase (600U/µl, Enzymatics) and added to 140µl with water. The reaction was incubated for 1h at RT and purified using AMPURE XP beads: 252µl of Ampure Beads were added and vortexed for 30s. After 5min of incubation at RT the supernatant was discarded using a magnetic rack. The beads were washed twice with 70% EtOH. After washing the bead were collected with the magnetic rack and the pellet was dried until all the ethanol evaporated. The DNA was eluted from the beads using 60µl of EB. 10µl of the library were kept for input DNA library analysis.
To close the nicks at the 3'-ends of the DNA resulting from the ligation of unphosphorylated adaptors to the 5' phosphorylated DNA, 48µl of the library were mixed with 4.5µl water, 6µl NEB2 buffer, 0.5µl dNTPs (40mM) and 1µl DNA polymerase I (10U/µl; NEB) and incubated for 15min at RT. Nick translated libraries were purified using 1.8 volumes AMPURE XP beads (108µl), washed twice with EtOH and eluted in 102µl EB. Then, a modified MeDIP protocol was applied to enrich methylated library fragments.

2.3.2.2 MeDIP enrichment

Briefly, 20µl of Dynabeads M280 sheep anti-mouse IgG paramagnetic beads (Invitrogen) were washed three times with 700µl PBS/0.5% BSA using a magnetic rack. After the last washing step beads were resuspended in 250µl PBS/0.5% BSA and 5µl of anti-5-methyl cytosine antibody (Eurogentec) were coupled to the beads by an overnight incubation at 4°C. After incubation the beads were washed twice with 1ml PBS/0.5%BSA, once with 1ml immunoprecipitation (IP) buffer (10mM Sodium phosphate buffer pH7/140mM NaCl/0.25%Triton X100) and were resuspended in 30µl of 1x IP buffer at 4°C. Meanwhile, 100µl of nick translated library were denaturated by 10min incubation at 95°C followed by an immediate cooling on ice. Denatured DNA and 100µl of 2x IP buffer (20mM sodium phosphate buffer pH7/280mM NaCl/0.5%Triton X100) were added to the prepared beads and incubated for 4h on a rotator at 4°C. After incubation, the beads were

washed three times with 1x IP buffer at 4°C using a magnetic rack. After the last washing step and removal of the IP puffer beads were pelleted by 3min centrifugation at 960g at 4°C. Any residual IP Puffer was removed. DNA was eluted with 210µl elution buffer (50mM Tris-HCl; pH 7.5; 10mM EDTA; 1% SDS) and 15min incubation at 65°C with continuous shaking. Beads were pelleted by 5min centrifugation at >10,000g at room temperature. 200µl of the solution containing the DNA were transferred to a new tube, 200µl TE buffer and 4µl proteinase K (20µg/µl) were added. Proteinase digest was performed for 2h at 55°C. Enriched methylated fragments were purified by phenol/chloroform/isoamylalcohol extraction using 5Prime Phase Lock Gel tubes (heavy) followed by ethanol/ammoniumacetate precipitation. The pellet was resuspended in 30µl EB and 2µl were used for real time qPCR enrichment analysis of methylated and unmethylated regions: To each 2µl library before and after MeDIP enrichment 19.5µl of qPCR mastermix (15µl SYBRGreen Mix (2x; Invitrogen) + 4.5µl H_2O) were added and 3.25µl were distributed into 6 wells of a qPCR plate. 1.75µl of primermix (2.14µM each forward and reverse: 8804 and 4994) were added according to the following loading scheme and a standard two step real time qPCR (10min 95°C, 40 cycles: 15s 95°C, 1min 60°C) with melting curve analysis was performed. Enrichment factors (F_E) were calculated according to the $\Delta\Delta C_t$ method using the median of the C_t values of each triplicate:

$$F_E = 2^{-((4994out-8804out)-(4994in-8804in))}$$

2.3.2.3 PCR for library amplification

Following successful MeDIP enrichments a first test amplification to find the optimal cycle number was performed using serial dilutions of the output libraries (1:2, 1:4, 1:8, and 1:16). For each library a mastermix was created containing 0.9µl of each primer MP-PCR-P1 (50µM) and MP-PCR-P2 (50µM), 4.5µl 10x Platinum PCR buffer, 1.35µl $MgCl_2$ (50mM), 26.82µl NFW, 0.45µl dNTPs (40mM), 0.9µl Platinum Taq Polymerase (5U/µl, Invitrogen) and 0.18µl Pfu Polymerase (2.5U/µl, Fermentas). 8µl of the mastermix were mixed with 2µl DNA of each dilution step. PCR was performed on a thermal cycler using the following temperature protocol: 5min 95°C, 17 rounds of (15s 95°C, 15s 62°C, 1min 70°C), 5min 70°C, and forever 4°C.

For input DNA libraries, a 20min nick translation step (72°C) was introduced prior to PCR cycling and only 10 rounds of PCR were performed.

Amplicons were checked on an agarose gel and the optimal cycle number was calculated (n = 17 - dilution step with minimal visible band) and used for the final amplification of the libraries: 1µl of each primer MP-PCR-P1 and MP-PCR-P2 (50µM) were mixed with 5µl 10x Platinum PCR buffer, 1.5µl $MgCl_2$ (50mM), 29.8µl NFW, 0.5µl dNTPs (40mM), 1µl Platinum Polymerase (5U/µl, Invitrogen), 0.2µl Pfu (2.5U/µl, Fermentas) and 10µl of output DNA library. Platinum Taq can be substituted by HIFI Platinum Taq polymerase (0.5µl + 0.5µl NFW; Invitrogen). Amplicons were purified with Qiagen's MinElute columns and eluted in 30µl EB. A 2% agarose gel was used for size selection of the library band (190-290bp, containing 103bp adaptor sequence). DNA was extracted from the gel using the Qiagen Gel Extraction Kit in combination with MinElute columns. Libraries were eluted with 15µl EB and quantified by real time qPCR using library PCR primers PCR-P1 and PCR-P2, thus only successfully amplified library was measured:

Dilutions of a prequantified SOLiD fragment library control (DH10B) were used to create the standard curve (100pg/µl, 10pg/µl, 1pg/µl, and 0.1pg/µl). Sample DNA was diluted 1:100 and 1µl was added to 1.5µl of primermix (5µM each), 2.5µl NFW, and 5µl 2xSYBRGreen mix. Triplicates were prepared for each sample. qPCR was performed using the following cycling protocol followed by melting curve analysis: 10min 95°C, 25 cycles of 15s 95°C and 1min 60°C.

The standard curve was created by using $\log_2$(amount of standard library) as ordinate and C_t values as abscissa. Library concentrations (c in pg/µl) were calculated using the slope m and the intercept k of the linear standard curve according to the following formula:

$$c[pg/\mu l] = 100 * 2^{(m * Ct\ library + k)}$$

Samples were diluted to 100pg/µl using 1x Low TE buffer (ABI) and qPCR was repeated. Identical amounts of up to 8 barcoded libraries were pooled.

2.3.2.4 Templated bead preparation for SOLiD sequencing

2.3.2.4.1 Emulsion PCR

Libraries were fixed to sequencing beads by emulsion PCR following the templated bead preparation protocol for SOLiD V3. During emulsion PCR microdroplets of an aqueous phase were created in an oil phase. Libraries were highly diluted in a way that every droplet contains maximally one DNA fragment. Furthermore, paramagnetic beads carrying single stranded adaptor 1 (P1 beads) were added. Statistically, in about 10-15% of the microdroplets one DNA molecule, one bead and amplification mix were present. During emulsion PCR (ePCR) the DNA molecules were amplified involving the bead coupled P1-adaptor as primer and thus were fixed to the bead surface. In subsequent enrichment beads without an amplicon were depleted. Microdroplets containing more than one type of library molecule yield polyclonal beads resulting in multifluorescence signal for the respective bead that cannot be analysed. For each pool of 8 samples ~ 4-5 ePCRs were needed to fill a full slide with up to 600 million beads. The oil phase was prepared by combining 1.8ml of Emulsion Stabilizer 1 and 400µl of Emulsion Stabilizer 2 with approximately 37.8ml mineral oil followed by vigorous vortexing. During degassing (>20min) of the oil, the aqueous phase and the SOLiD P1 beads were prepared. In a Nalgene wide-mouth jar 5µl of 100pg/µl library, 560µl 10x PCR buffer, 784µl dNTP mix (100mM), 140µl $MgCl_2$ (1M), 22.4µl ePCR Primer 1 (10µM), 3µM ePCR Primer 2 (500µM), 3295µl NFW, 600µl AmpliTaq Gold Polymerase (5U/µl) were mixed on ice.

P1 beads were suspended in 200µl Bead Block solution using a magnetic rack and sonicated with the Bead Block Declump programme on the Covaris S2 System (Cycle1 for 5s: duty cycle 1%, intensity 5, 50 cycles/burst; Cycle2 for 1min: duty cycle 5%, intensity 5, 100 cycles/burst). Following declumping, the supernatant was discarded (magnetic rack) and P1 beads were resuspended in 200µl 1x TEX buffer.

9ml of the degassed oil phase were transferred to a SOLiD ePCR tube on the ULTRA-TURRAX device (IKA). P1 beads were again sonicated with Covalent Declump 1 and 160µl of the beads were added to the aqueous phase. Using the Xstream pipettor (Dial Setting: Pip, Speed (aspirate UP): scale 5 (mid-range), Speed (dispense DOWN): scale 1 (slowest), Total volume: 5.60 ml) with a 10ml Combitip Plus tip the aqueous phase was aspirated and the ULTRA-TURRAX tube drive was started. The sample was dispensed into the oil phase through the center hole. After 5min of emulsifying the emulsion was

dispensed into a 96 well plate using cut 5ml Combitips Plus (Eppendorf) and sealed with optical adhesive covers.

The ePCR was performed on the GeneAmp PCR System 9700 (AB) with 60 cycles of amplification: 5min 95°C; 60 cycles of 15s 93°C, 30s 62°C, 75s 72°C; final extension 7min 72°C; 4°C forever (ramp speed: 9600, reaction volume: 50µl). In case the emulsion was broken in more than 3 wells after cycling (visible aqueous phase at the bottom of the wells) the ePCR has to be repeated. The emulsion was collected from the wells by 2min centrifugation at 550g using the SOLiD Emulsion Collection Tray. 10ml of isobutanol were added to the emulsion, mixed by vigorous pipetting and transferred to a new 50ml tube. The remaining beads were gathered with additional 6ml of isobutanol. To separate the beads from the oil phase the tube was centrifuged for 5min at 2000g. The liquid phase was discarded and the beads were dried upside down for 5 min. 600µL of 1x Bead Wash Buffer were used to resuspend the beads. Beads were transferred to a new 1.5ml LoBind tube and the 50ml tube was washed with additional 600µl of Bead Wash Buffer that were also transferred to the 1.5ml tube. Beads were collected by 1min centrifugation at minimum 14,000g. The oil phase was removed at the meniscus, and then the rest of the supernatant was removed with a new pipette tip. The pellet was resuspended in 150µl of 1x Bead Wash Buffer and transferred to a new 1.5ml tube. The original tube was rinsed with 150µl 1x Bead Wash Buffer that was also transferred to the new tube. 1ml of Bead Wash Buffer were added followed by a 1min centrifugation step at mininum 14,000g. The supernatant was discarded and beads were resuspended in 200µl TEX buffer. Using a magnetic rack, TEX buffer was discarded and beads were resuspended in 200µl TEX buffer.

After sonication with Covalent Declump 1, a 1:10 dilution of beads was quantitated by nanodrop (cell culture settings). An absorption value of 1 at 600nm equals 1 billion undiluted beads. These beads still contain about 80-90% beads without any amplicon. Thus, successfully loaded P1 beads have to be enriched.

2.3.2.4.2 Bead enrichment

650µl of Enrichment Beads were pelleted by 5min centrifugation at minimum 14,000g and washed twice with 900µl of 1x Bind & Wash Buffer. After the last centrifugation, Enrichment Beads were resuspended in 350µl Bind & Wash Buffer and 3.5µ of 1mM Enrichment Oligo were added and rotated at RT for 30min. Enrichment Beads were

collected by 5min centrifugation, washed twice with 900µl TEX buffer and resuspended in 150µl Low Salt Binding Buffer.

The templated P1 beads were prepared for enrichment by suspending them in 300µl of denaturing buffer solution (Denaturing Buffer + 10% denaturant, freshly prepared). After 1min incubation beads were collected with the magnetic rack and the denaturing step was repeated. Beads were washed twice with 300µl TEX Buffer, resuspended in 150µl TEX Buffer, and then transferred to a new 0.5-mL LoBind tube. Beads were declumped using the Covaris S2 with Declump 1. Then, the prepared enrichment beads were mixed with the templated beads and sonicated using the Covalent Declump 3 programme (3 cycles: 5s: duty cycle 2%, intensity 6, 100 cycles/burst, 30s: 5% duty cycle, intensity 9, 100 cycles/burst). The declumped bead mixture was incubated for 15min in a water bath at 61°C. Every 5 minutes the beads were vortexed and pulse-spinned. After the incubation beads were cooled on ice for 2min. To separate P2 beads with hybridised P1 beads from unloaded P1 beads 600µl freshly prepared 60% glycerol were added to a new 1.5ml LoBind tube and the beadmixture was carefully loaded on top of the glycerol followed by 3min centrifugation at minimum 14,000g. The top layer containing the P1 beads was transferred to the bottom of a new 2ml tube containing 1ml TEX buffer. Clean the pipette tip with some TEX buffer (‘No bead is left behind’). An additional 1ml of TEX was added to the tube and centrifuged for 1min at minimum 14,000g. If the beads were pelleted the supernatant was removed and beads were resuspended in 400µl of TEX. If the beads were not pelleted half of the volume was transferred to a new 2ml tube, 500µl TEX were added to each tube and centrifugation was repeated. Using 200µl of TEX each, enriched beads were resuspended and pooled.

To separate enriched P1 beads from the enrichment beads, beads were pelleted by 1min centrifugation, supernatant was discarded and 400µl freshly prepared denaturing buffer solution was incubated with the beads for 1min. Beads were collected with a magnetic rack and the supernatant was discarded. This step was repeated until the supernatant was clear and all white P2 beads were removed and the P1 beads were washed twice with 400µl TEX. Beads were resuspended in 200µl TEX and transferred to a new 1.5ml LoBind tube. The original tube was rinsed with additional 200µl TEX. Beads were declumped with Covalent Declump 1 and the supernatant was substituted by fresh 400µl TEX. Using TEX and a magnetic rack the beads were washed until the supernatant was clear. Finally, the beads were resuspended in 400µl TEX.

2.3.4.2.3 Bead deposition

To attach the enriched beads to the surface of a SOLiD sequencing slide, a bead linker was attached to the ends of the bead coupled DNA. The beads were therefor declumped with Covalent Declump 3 and the supernatant was removed using a magnetic rack. Beads were washed with 100µl Terminal Transferase solution (11% 10x Terminal Transferase Buffer and 11% 10x Cobalt Chloride in NFW) and resuspended in 178µl Terminal Transferase solution. 20µl of 1mM bead linker in Low TE buffer were added and beads were sonicated with Covalent Declump 3. 2µl of Terminal Transferase (20U/µl) were added and the reaction was rotated for 2h at 37°C. Then, the beads were washed with 400µl TEX buffer, resuspended in 400µl TEX buffer, declumped with Covalent Declump 1 and quantified using Nanodrop.

To achieve an approximate bead load of 500 million per slide, up to 700 million beads have to be deposited. Therefor, the appropriate number of beads was transferred to a 1.5ml LoBind tube. TEX was discarded and beads were washed three times with 400µl deposition buffer. Beads were resuspended in the volume of deposition buffer indicated on the SOLiD Deposition Chamber (each chamber was individually sized).

The SOLiD Deposition Chamber was assembled with a slide carrier and a SOLiD sequencing slide and incubated at 37°C. Meanwhile, the beads were sonicated twice using the Covalent Declump 3 programme. Immediately after sonication the beads were transferred to the Deposition Chamber and the portholes were sealed with adhesive tape. The Deposition Chamber was incubated at 37°C for 2h.

2.3.2.5 SOLiD sequencing

A multiplex run was configured on the SOLiD 3+ machine assigning all barcodes to the respective libraries and the sequencing chemistry was inserted into the machine: for barcode sequencing the SOLiD Opti Fragment Library Sequencing Kit - 5bp Barcode set and the SOLiD Opti Fragment Library Sequencing Kit 35 or version 2 chemistry adapted for version 3 were used according to the manufacturer's instructions. The camera was focused manually and the run was started. Sequencing of the slides including the barcode took about 5-7 days per slide yielding about 400-500 million reads of 35bp.

2.3.3 Alignment and data table

Alignment of sequencing reads was conducted by Axel Fischer. Total reads (raw output, RO) were aligned to the 2009 assembly of the human genome (GRCh37/HG19). Alignment was performed in colour space allowing a correction for single colour sequencing errors. Using Applied Biosystem's Bioscope Alignment module in seed and extend mode the first 25bp of the reads were taken as seeds allowing 2 mismatches and a mismatch penalty score of -2 for extension (→ mapped output, MO). From the MO, reads aligning to more than one genomic region were excluded (→ uniquely mapped output, UMO). The UMO reads were elongated to 200bp in a strand-oriented manner. Redundant reads and reads with no CpGs in the elongated sequence were excluded from further analyses. Primary sequencing data (read sequence files *.csfasta and read quality files *.qual) and secondary analysis data (bed files) for all prostate cancers are available under the GEO accession number GSE35342. Bed files contain only extended, 0-CpG and duplicate depleted reads. Next, the HG19 reference genome was split into adjacent 500bp bins and the number of reads per bin was counted. Reads were assigned to a bin when their centre was located within the bin. For sample-wise normalisation, binwise read counts were related to the total read count of each sample yielding a reads per million (rpm) value. A table was generated containing all genomic bins (6,191,344 rows) and the rpm values of all samples (104 columns). This table was used as a basis for all further analyses and was added with several descriptive values for each bin (see section 2.3.4) as further columns. Each genomic region could be conveniently addressed by an unique bin index (row number).

2.3.4 Annotations

Additional information of the following data bases was annotated to the 500bp bins for functional and topographical analyses: CpG-island (cpgIslandExt) and repeat masker data (rmsk) were extracted from the UCSC database version GRCh37/HG19. Bins containing at least one base of an annotated CGI were annotated as ‘CGI’ and the ±2kb surrounding bins as CGI-shore.

Genic, exonic and transcription start sites were used from Biomart Ensembl59, microRNA regions from miRBASE v16, homeobox genes from Holland *et al*. 2007 (HOLLAND 2007). The cancer gene census list was obtained from http://www.sanger.ac.uk/genetics

/CGP/Census. The list of G-protein coupled receptors was downloaded from http://www.iuphar-db.org. Regions of ±2kb around a TSS were annotated as 'promoter region'. Conserved regions were annotated as follows: The nucleotide-wise conservation scores of a set of 46 species were extracted from UCSC (HG19: http://hgdownload.cse.ucsc.edu/goldenPath/hg19/phyloP46way/) and the sums of these scores for 100bp sized consecutive bins were calculated. The highest scoring 1% quantile consisting of 281,450 100bp bins was assigned to 176,060 different genomic 500bp bins that were annotated as 'conserved'.

2.3.5 Detection of methylated regions

Statistical analyses were conducted using R (version 2.9.2) assuming a binomial distribution of reads as null hypothesis. Probability values for the mean tumour and mean normal read counts of each bin to appear solely by chance were assigned to each region:

P(bin) = dbinom(RCB, RC, PROB)

RC = mean total read count for specific tissue (tumour = 14,876,320; normal = 14,584,031)

RCB = mean tissue specific read count in this very bin (rounded)

PROB = probability for each read to map to this bin = 1/6,191,344 (total bin count)

Bins with p-values <0.05 in tumour or normal were called 'significantly methylated' and were used for principal component analyses.

To detect differences in the number of methylated bins between the tissue subgroups the calculations were repeated sample-wise using the sample's rpm values instead of mean values. Bins with p-values <0.001 were called 'significantly methylated' in this sample. Counts of significant bins were compared between normal, tumour, FUS+ and FUS- samples using T-test (two sided, unpaired).

2.3.6 Detection of differentially methylated regions (DMRs)

Significance of differences in rpm values (resembling methylation) between tissue subgroups (normal (n=54) vs. tumour (n=51), *TMPRSS2:ERG* fusion positive (n=17) vs. negative tumours (n=20)) was calculated for each bin using the nonparametric two-tailed Mann-Whitney test (MW) in R:

$p_{MW\_tum/norm}$ = wilcox.test(x[tumor],x[normal])[[3]]

$p_{MW\_FUS+/FUS-}$ = wilcox.test(x[FUS+],x[FUS-])[[3]]

Mann-Whitney p-values (p_{MW}) were corrected for multiple testing using the Benjamini-Hochberg (BH) approach (BENJAMINI 1995). Bins with adjusted p-values (p_{BH}) of <0.05 were called 'significantly differentially methylated'.

p_{BH} = p.adjust(array of p_{MW}-values, method="BH")

2.3.7 Principal component analyses (PCA) and hierarchical clustering

Principal component analyses (PCAs) can be used to get an overview of differences inherent in the samples. It reduces a variety of descriptive markers (in this case methylated bins) to those explaining the largest variance in the samples. Pairs of these principal components allow a two-dimensional separation of the analysed samples and elucidate the internal data structure like underlying sample groups or batch effects. PCAs were conducted with the *prcomp*-function in R using all bins significantly methylated in normal or tumour samples.

Hierarchical clustering on the other hand uses a small set of markers as input and groups samples with a similar pattern using all these markers. A hierarchy is established that can be visualised as dendrogram or heatmap. Hierarchical clusterings were performed with the heatmap.2-function included in the R package 'gplots' using the Euclidian distances and the agglomerative algorithms 'complete' and 'ward' or alternatively with the 'heatmap' function using the standard parameters.

2.3.8 Enrichment analyses

For each feature analysed (e.g. hypermethylation in TSS bins) the number of respective bins was counted for the average patient (tumour, normal, FUS+, and FUS-) and odds ratios were calculated.

$$OR = \frac{\text{feature in group 1 / feature in group2}}{\text{no feature in group 1 / no feature in group2}}$$

$$\text{e.g. } OR_{\text{hypermethylation in TSS in tumour}} = \frac{\text{hypermeth. } TSS_{tumour} \text{ / hypermeth. } TSS_{normal}}{\text{not hypermeth. } TSS_{tumour} \text{ / not hypermethylated } TSS_{normal}}$$

For calculation of odds ratios and confidence intervals the following R function was used:

structure(fisher.test(data.frame(c(feature$_1$,feature$_2$),c(notfeat$_1$,notfeat$_2$))))

2.3.9 Bisulphite mass spectrometry (BS-MS) with Sequenom EpiTYPER-Assay

For validation of the MeDIP-Seq results the Sequenom EpiTYPER-Assay (RADPOUR 2008) was used. It allows the determination of the methylation state of several CpG groups within DNA amplicons of up to 600bp with an accuracy of 5% and a resolution of up to one CpG.

2.3.9.1 BS-MS protocol

One microgram of DNA (patient sample DNA, cell line DNA, and *in vitro* methylated DNA) was bisulphite converted using the Epitect Bisulfite Conversion kit (Qiagen) and subsequently amplified with specific primer pairs carrying a T7-promoter that were designed using the Epidesigner tool (www.epidesigner.com) with standard criteria (amplicon length: 400-600bp). Amplification conditions used were as follows: 1µl of BS converted DNA was mixed with 0.5µl 10x HotStar buffer, 0.1µl dNTPs (10mM), 0.1µl

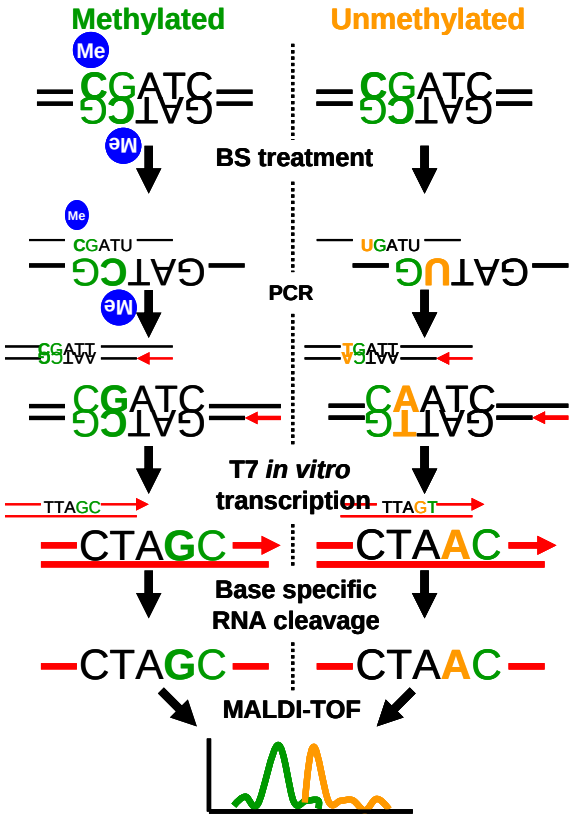


Figure 9. Schematics of BS conversion of methylated and unmethylated DNA and the BS-MS approach. Treatment of DNA with sodium bisulfite (BS) results in conversion of unmethylated cytosines into uracils and to partial denaturation of the DNA due to the introduced mismatches. During PCR amplification a T7 RNA polymerase is introduced. *In vitro* transcription generates the material for subsequent RNA cleavage and MALDI-TOF analysis of the reverse strand. Drawing modified from Sequenom AG.

primer mix (10µM each), 0.04µl HotStar Taq (5U/µl, Qiagen) and filled to 5µl with NFW. The reaction was incubated in a PCR cycler: 95°C 15min; 45 cycles of 94°C 30s, annealing temperature 30s, 72°C 1min; 72°C 5min, 4°C forever. Annealing temperatures were determined using gradient PCR with bisulphite converted cell line DNA prior to amplification of patient's DNA.

In vitro transcription and subsequent steps were performed in Heidelberg by Rainer Claus or at the Sequenom Facility in Hamburg. The transcripts were cleaved and analysed using MALDI-TOF mass spectrometry on a MassARRAY Analyser 4. The ratio of adenosine and guanosine determined for every CpG group mirrors the ratio of methylated and unmethylated cytosine at this specific site (**Figure 9**).

The BS-MS was performed three times:

1. **BS-MS 1**: Validation of MeDIP-Seq results of 48 heterogeneous genomic regions (**Table 17**) in six patient samples: three tumour (IGP28, IGP33, IGP41) and three normal (IGP122, IGP178, IGP179). The regions covered the *GSTP1* DMR, hypermethylated regions in developmental gene (n=8) and other promoters (n=6), hypermethylated regions outside of promoters (n=8), hypomethylated promoter (n=5) and non-promoter regions (n=8), differentially methylated miRNA associated regions (n=5), and regions with few reads or no significant difference between tumour and normal patients (n=8).

Table 17: Primers used to amplify BS converted DNA regions for BS-MS analysis 1. Column headings denote the primer name, the position of the amplicon in a TSS associated region, the genomic region containing the BS-MS target region, the size of amplified target region and the number of CpGs in amplified target region, the number of CpGs assayable by BS-MS, and the core sequence of the primers. To the 5'-end of each forward primer the sequence **aggaagagag** was added, and to the 5'-end of each reverse primer the sequence **cagtaatacgactcactatagggagaaggct** was added.

Amplicon Name	Promoter region	Genomic location	Size	CpG	CpG covered	Sequence (5'-3')
SEQ1_1_GSTP1_FW	GSTP1	chr11:67351401-67352100	556	37	32	GGGGGTTTAGGGGATTTAGGA
SEQ1_1_GSTP1_RV						CTCCTTCCAACTCTAACCCTAATCT
SEQ1_2_Olfactory Receptor_FW	PACSIN3	chr11: 47208401-47209100	498	25	24	TTTTTTTGGTTTTTAAGGTTTAGGG
SEQ1_2_Olfactory Receptor_RV						TTTCAAAAACTTACCCAAAATTTCC
SEQ1_3_Type1_FW	SIX6	chr14:60973901-60974600	528	25	20	GGTTTGTATAGATTTGGTAAATTAGAAGG
SEQ1_3_Type1_RV						AACAAAAAAATACTTTCCTCCCCTA
SEQ1_4_Type1_FW	ZNF154	chr19:58219901-58220600	510	42	30	GGTTATTTTAGTTTTTTTGAGGTGTG
SEQ1_4_Type1_RV						TTCCCTACAAAAAACAACTTTCC
SEQ1_5_Type1_FW	FALSE	chr3:62353401-62354100	485	27	21	TGTGATATGTTTGGGGTTATTAGGT
SEQ1_5_Type1_RV						ACATATTCCTCTAAATCCCAATCTC
SEQ1_6_Type1_FW	SCAND3	chr6:28556901-28557600	515	24	21	TGTTTTTATTATAGGAATTGGGTTTTT
SEQ1_6_Type1_RV						AACAATTACTCCTTCTTCAACAACAA
SEQ1_7_Type1_FW	ELAVL2	chr9:23823901-23824600	464	20	18	TTTTAGAGGATGGGTATAATAAATAGGG
SEQ1_7_Type1_RV						AAATAACAAAATAACAATACCCACCA

Amplicon Name	Promoter region	Genomic location	Size	CpG	CpG covered	Sequence (5'-3')
SEQ1_8_Type10_FW	NLRP8	chr19:56458901-56459600	533	11	8	TTTAGGTGATGGGTATGGTGTTAGT
SEQ1_8_Type10_RV						TTACTTACCTCTTCCCCAAAAAAAA
SEQ1_9_Type10_FW	TNFRSF10C	chr8:22939901-22940600	498	7	6	ATTAGTGAATTTTGGGTTTTTTTTG
SEQ1_9_Type10_RV						AACAATTCTAACTTTATCCCAACCA
SEQ1_10_Type11_FW	FALSE	chr11:9090401-9091100	497	8	7	TTTGTAAGGAGGTTTTAAGGGGTAG
SEQ1_10_Type11_RV						CATCAAAAAATTTAAACTATTCTCCAAA
SEQ1_11_Type11_FW	FALSE	chr16:3015901-3016600	462	21	20	TTAGGTTTAGGTTAGGGGTTATTGG
SEQ1_11_Type11_RV						ACAAACACAAAACCTTAAAAACCAC
SEQ1_12_Type2_FW	ALX4	chr11:44332901-44333600	517	26	20	GGTTTTTATTGGGTATGGAGTGG
SEQ1_12_Type2_RV						CTCTACTCAAACTAATCCCTCCACC
SEQ1_13_Type2_FW	TBX3	chr12:115123401-115124100	503	17	10	TTAGAGAAAGATAGATTGAAAGGGAAA
SEQ1_13_Type2_RV						CCCACAATTCTTAATTCTAAAAACC
SEQ1_14_Type2_FW	BMP4	chr14:54422401-54423100	539	42	31	AAATAAGAGAAGTTTTTGGTAGGGG
SEQ1_14_Type2_RV						AACCTAAAAAACCAAAAAAAACC
SEQ1_15_Type2_FW	GREM1	chr15:33008901-33009600	600	20	19	TTGGTTTGGTTTTGTTTTTTTAGTTA
SEQ1_15_Type2_RV						AAAAATTCTTTCAACCCAAATACC
SEQ1_16_Type2_FW	ONECUT1	chr15:53083401-53084100	560	34	34	AAGTAAATTTATAGTTTGAGTTTGGTTTA
SEQ1_16_Type2_RV						ATTACTCCCAAAACCCTCCTACTAA
SEQ1_17_Type2_FW	FOXC2	chr16:86599401-86600100	526	25	16	GTTTGTTAGAGGGAATTTTTATTTGG
SEQ1_17_Type2_RV						TCCCAAAAACCTTATAAATAACAATAAT
SEQ1_18_Type2_FW	COL18A1	chr21:46823901-46824600	538	18	17	TTGAGGTAGGTAGGGATGATGTTAG
SEQ1_18_Type2_RV						CCTCCCAAATAAAAACCCTATTTCT
SEQ1_19_Type2_FW	PHOX2B	chr4:41748901-41749600	501	30	29	TGGGGTTAGGTTAAAAGTATTTGAG
SEQ1_19_Type2_RV						CCCAACTCAAAAAACTAAAAAAAATC
SEQ1_20_Type3_FW	FALSE	chr12:65219901-65220600	504	19	18	TTTTGAGTTGGGTTTTAGGTTTTTA
SEQ1_20_Type3_RV						ACCAACTAACCCCTACTAAATTTCAC
SEQ1_21_Type3_FW	FALSE	chr19:17245901-17246600	511	21	18	TTAGAAGTGGGATTGTTAGATTTAGTGA
SEQ1_21_Type3_RV						CATACATAAAAACTACTCCCAAACCC
SEQ1_22_Type3_FW	FALSE	chr19:48983901-48984600	447	26	19	GGGAGATGAGGTTTTAGGGTAAATA
SEQ1_22_Type3_RV						TCTCAACAAATCCTTTTAACTCCTC
SEQ1_23_Type3_FW	FALSE	chr6:108439901-108440600	436	19	15	GATTTGGTAGAGGTAGAAGAGAGGTT
SEQ1_23_Type3_RV						AAAACAATTTAAAAAAACCCTTCTCA
SEQ1_24_Type3_FW	CAV2	chr7:116140401-116141100	567	16	14	TGTTTTTGGTTATTTTTTTGGTTTT
SEQ1_24_Type3_RV						TAAAACAACCCCTCCACACTACTAC
SEQ1_25_Type3_FW	FALSE	chr7:157360901-157361600	512	17	15	GTTGTTTGTTGTGTTTTTGTTGG
SEQ1_25_Type3_RV						CACAAACCTCAAATCTAAAACTCCC
SEQ1_26_Type3_FW	FALSE	chr9:37030401-37031100	436	12	8	GGGTTTTATTTTGATTTATAGTGATG
SEQ1_26_Type3_RV						CATACTAACAACATCCCTAATCTTTCC
SEQ1_27_Type3_FW	FALSE	chr9:77114401-77115100	437	19	14	TTTGGGTAATTGGAGAGTAGAGAATTT
SEQ1_27_Type3_RV						TTTCTCAAAATAAACATAAAAACCA
SEQ1_28_Type4_FW	FALSE	chr1:159409401-159410100	521	5	4	TTTTATTGGGATTATGTGATTTTTATG
SEQ1_28_Type4_RV						ACAATCAACCCAATACTACAAACCC
SEQ1_29_Type4_FW	FALSE	chr16:7222901-7223600	426	6	5	GATATTTGTTGTGGATGGTGGTATT
SEQ1_29_Type4_RV						AAACAATTAACTACAAAACTAAAATTTACC
SEQ1_30_Type4_FW	FALSE	chr18:37892901-37893600	450	10	7	TAATTGTTTTGTGGAAATTAAGAAA
SEQ1_30_Type4_RV						TTTTAACAAAATCTACAATAATACCCAA
SEQ1_31_Type4_FW	FALSE	chr20:56236901-56237600	583	23	22	GGGTTTGAGTTTTTAGTAGGTAGGG
SEQ1_31_Type4_RV						ACACTTTATACATCTATACCACCCTCAA
SEQ1_32_Type4_FW	FALSE	chr5:85072901-85073600	450	5	3	TTTATAGGTTAGGGATTAGGGAATTG
SEQ1_32_Type4_RV						AAACAACACAAATAAAACAACCTAA
SEQ1_33_Type4_FW	FALSE	chr8:32020901-32021600	402	4	4	TTTTTTATTAGTAGTTTTGGAAAGAGAG
SEQ1_33_Type4_RV						AACATCATCAAAATTTAAAACCACA
SEQ1_34_Type4_FW	FALSE	chr8:4609901-4610600	444	5	4	AATGATTTTTTTTGGTTTTGTTGTG
SEQ1_34_Type4_RV						AAAACTCAAAATTAAAACTCTATACCCAA
SEQ1_35_Type4_FW	FALSE	chrX:151073401-151074100	482	12	9	TTTTGATAGGGTTTTTGGATTTAGG
SEQ1_35_Type4_RV						TATAACCCACCCCAATTAAACATAA
SEQ1_36_Type5_FW	SFRS5	chr14:70235401-70236100	503	12	11	TTTTATAGGTGGAGTATTTTTGAGGAA
SEQ1_36_Type5_RV						AACCCTAAAACATAAAACAAACTAAATC
SEQ1_37_Type5_FW	EEF2	chr19:3982901-3983600	434	25	23	GGGTTTTAGTTGTAGTTTTAGTAGGG
SEQ1_37_Type5_RV						ACCAAATATATAACTCCTATCCCCC

Amplicon Name	Promoter region	Genomic location	Size	CpG	CpG covered	Sequence (5'-3')
SEQ1_38_Type5_FW	KLK4	chr19:51411401-51412100	491	29	23	GGAGTTAGATGGTGGAGGTTAGTTT
SEQ1_38_Type5_RV						ACCCTATATATCTCTATCTCCCCCTT
SEQ1_39_Type5_FW	ZNF649	chr19:52405901-52406600	518	10	6	TTTTTAGTTAAGAATTGTTTGAGAATGA
SEQ1_39_Type5_RV						TCAAATATACCCACATAATAAAACCAAC
SEQ1_40_Type5_FW	5S-RNA, AC022910.2	chr8:3557901-3558600	559	12	11	GTTTTTTGGTGAGTTTTGATTGTTT
SEQ1_40_Type5_RV						ACCATACCACCCTAAATACACCTAATC
SEQ1_41_Type6_FW	miR 9-1	chr1:156389901-156390600	505	36	25	GGTATTAGAAATTTTTTGGGTTTGG
SEQ1_41_Type6_RV						AAAATCTCTCTCCTCCTCTTATATCCT
SEQ1_42_Type6_FW	miR 34b	chr11:111385401-111386100	416	24	18	GGTAGGGGTTGGTATTAATTTGG
SEQ1_42_Type6_RV						CCATTACAATTTCTCCTAAAAAAAACTC
SEQ1_43_Type6_FW	miR 184	chr15:79501901-79502600	432	16	14	TTTGAATAAGTAAAGTGGAGATTTTATTG
SEQ1_43_Type6_RV						ACCAAAACAAAATAAAAAAACACAA
SEQ1_44_Type6_FW	miR 9-2	chr5:87963401-87964100	464	16	14	TTGAAAGAAGGAGAAAAGGGTTATT
SEQ1_44_Type6_RV						AAACAATAACAAAATACCTTCCATC
SEQ1_45_Type6_FW	miR 23b	chr9:97846401-97847100	423	16	14	AGGAATTATGTGTGTTAGGAAAGGG
SEQ1_45_Type6_RV						AAAAAACAAACTTACAACCCCTAAA
SEQ1_46_Type7_FW	snoU13, AL157818	chr13:95861401-95862100	492	4	4	TTTAGAATGGTTTTATTTGTGGTGG
SEQ1_46_Type7_RV						AAAACCCCCTTAACTAAACCTAAAA
SEQ1_47_Type8_FW	SFXN4	chr10:120925901-120926600	526	10	9	GAGTTTAGGATTGGGGGATTTAGTA
SEQ1_47_Type8_RV						ACAATCCAAAATAAAAAAACAAACC
SEQ1_48_Type9_FW	MRPL44	chr2:224822401-224823100	438	6	3	TTTTTATGGTTTTGTATATATGGGTTT
SEQ1_48_Type9_RV						TCCTCCTATATACTTTAAATCATCTCTAA

2. **BS-MS 2**: Comparison of methylation values of 38 differentially methylated regions (**Table 18**) in microdissected and macrodissected tissue of two additional patients (macro- and microdissected tumour and adjacent benign tissue). Additionally, the consistency of methylation values was analysed in one tumour sample (IGP33) already analysed in BS-MS 1.

Table 18: Primers used to amplify BS converted DNA regions for BS-MS analysis 2. Column headings denote the primer name, the position of the amplicon in a TSS associated region, the genomic region containing the BS-MS target region, the size of amplified target region and the number of CpGs in amplified target region, the number of CpGs assayable by BS-MS, and the core sequence of the primers. To the 5'-end of each forward primer the sequence **aggaagagag** was added, and to the 5'-end of each reverse primer the sequence **cagtaatacgactcactatagggagaaggct** was added.

Amplicon Name	Promoter region	Genomic location	Size	CpG	CpG covered	Sequence (5'-3')
SQ00001_SEQ2_FW	TMEM51	chr1:15480901-15481600	460	30	26	GGTTTTAGAGGGGAAAGTTTTTTA
SQ00001_SEQ2_RV						TTTCTTCATCTATAAAACTAACATAACAAC
SQ00010_SEQ2_FW	AL132857.1	chr14:36991901-36992600	484	34	29	TTGGAGGGAGAGAAAGGAGTTTTTT
SQ00010_SEQ2_RV						AAACATCTATATCTCCAAACTTATCCAAA
SQ00011_SEQ2_FW	SIX6	chr14:60973901-60974600	528	25	20	GGTTTGTATAGATTTGGTAAATTAGAAGG
SQ00011_SEQ2_RV						AACAAAAAAATACTTTCCTCCCCTA
SQ00012_SEQ2_FW	ANXA2	chr15:60690901-60691600	481	16	7	TTTTTTTTATGATTGTTGGGGTGT
SQ00012_SEQ2_RV						TCACATACAACAAATAAATATCTACCTACA
SQ00013_SEQ2_FW	KRT23	chr17:39093401-39094100	551	12	11	TGGTTTGGAATTAGAAAGGTTAGAA
SQ00013_SEQ2_RV						AAAACAAAATCTCAAATCCCAAAAT
SQ00014_SEQ2_FW	TMEM91	chr19:41882401-41883100	437	36	21	TTTGGAAGGTTTTTAAGTTTTTTG
SQ00014_SEQ2_RV						AAACTACAATTCTCCAAAAATTCCC
SQ00015_SEQ2_FW	AOX1, AC080164.1	chr2:201450401-201451100	475	39	26	AAAGTAATGTTTTAGGATTTATAATGTTTG
SQ00015_SEQ2_RV						ACAAAAACTCCTCTCTACCCC
SQ00016_SEQ2_FW	ARL4C	chr2:235404401-235405100	592	47	35	AAGGGTATTAGTTGTTATTTTTGGG
SQ00016_SEQ2_RV						ACTCCTTAAATCACCAATCCTACAT
SQ00018_SEQ2_FW	SLC2A2	chr3:170745901-170746600	518	30	27	TTGTTAGTGATATGATATTTTGTAGGAA
SQ00018_SEQ2_RV						AAACTAACTTCACCACAACCAAAAA

Amplicon Name	Promoter region	Genomic location	Size	CpG	CpG covered	Sequence (5'-3')
SQ00019_SEQ2_FW	GHSR	chr3:172165401-172166100	509	38	29	GGTTTGTTTGTGGTTTTGGTTTT
SQ00019_SEQ2_RV						TCTTCCTCTACATACCCCTAAACCT
SQ00002_SEQ2_FW	MPPED2	chr11:30605401-30606100	475	26	17	GGGATTTTGGAATTTGTTTGTTT
SQ00002_SEQ2_RV						TATATATAATTCATAAAACACCCCC
SQ00021_SEQ2_FW	AC092502.2	chr3:62353401-62354100	527	29	22	GTTTATTTTGGGGTTTTTTGGG
SQ00021_SEQ2_RV						TCACCAAACTTAATAAAACAAAAATTC
SQ00022_SEQ2_FW	DDIT4L	chr4:101111901-101112600	493	17	14	GATTTGTGTATTTTTTTGTTTTGGG
SQ00022_SEQ2_RV						CAAAAATCCTCCCAACAAAACTC
SQ00023_SEQ2_FW	KIT	chr4:55522901-55523600	417	18	16	AAGGATTTATTATGAGATTGTGGGT
SQ00023_SEQ2_RV						CAAACCTATTAAAAACTCCTAACACC
SQ00024_SEQ2_FW	AC095061.1	chr4:61527401-61528100	513	5	5	TGATATGGTTTGGTTGTGTTTTTAT
SQ00024_SEQ2_RV						TTCCTAACCACTACACCTAATCACC
SQ00025_SEQ2_FW	PDGFRB	chr5:149535401-149536100	400	14	12	GTTATTTTGGGGGAAAGGTTGTAG
SQ00025_SEQ2_RV						AACCACAAAAATAATCAAAATCTCC
SQ00027_SEQ2_FW	SEMA5A, SNORD123	chr5:9546401-9547100	551	35	26	TTGAGTATTTAGTAAGGTTTAAGGATTTTG
SQ00027_SEQ2_RV						CTAACTTTTAAAAACTCCCTCCCC
SQ00028_SEQ2_FW	HIST1H1A, HIST1H1PS2	chr6:26017401-26018100	432	22	18	GTGGTTTTTTTGAGTTTTTTAGATGT
SQ00028_SEQ2_RV						CACCATATCTAAAACAATACCTCCC
SQ00029_SEQ2_FW	HIST1H4F	chr6:26240401-26241100	497	35	30	TTTTTTTTAAGGTAATAGTGTAGGGGA
SQ00029_SEQ2_RV						AAATAAAACTCAACCACCAAAACC
SQ00003_SEQ2_FW	WIT1, WT1	chr11:32458901-32459600	523	16	12	GGTTGTAGGGAATTTTTTTTATTTTT
SQ00003_SEQ2_RV						TCACTACCAATTACCCTAAAAACCC
SQ00030_SEQ2_FW	ZSCAN12	chr6:28366901-28367600	452	24	21	AATTTGAAGTGAGGAAGGTTTAGGA
SQ00030_SEQ2_RV						ACCTTTATAAATCACTTTTAATAAATTCC
SQ00031_SEQ2_FW	ZSCAN12	chr6:28367401-28368100	516	31	26	GGAATTTATTAAAAGTGATTTATAAAGGT
SQ00031_SEQ2_RV						AATTTTCTAACATATAATCATACATCAACA
SQ00032_SEQ2_FW	SCAND3	chr6:28556901-28557600	515	24	21	TGTTTTTATTATAGGAATTGGGTTTTT
SQ00032_SEQ2_RV						AACAATTACTCCTTCTTCAACAACAA
SQ00033_SEQ2_FW	HLA-J, HCG4P3	chr6:29974401-29975100	453	38	29	GGAGATTTTGGTTTTAATTGAAGATG
SQ00033_SEQ2_RV						CCCACTCCATAAAATATTTCAACAC
SQ00034_SEQ2_FW	HMGCLL1	chr6:55443901-55444600	493	21	19	GGTGTATTGGTTGTGAGGATTAGAG
SQ00034_SEQ2_RV						AACTATCTTTTCCTTTTCATTAAACTATTA
SQ00035_SEQ2_FW	DST, BEND6	chr6:56818401-56819100	493	23	19	GGAAAGAGAATTTTTTTTGATTTTGA
SQ00035_SEQ2_RV						CAATCTCCCAAATTTAAATAACCAC
SQ00036_SEQ2_FW	AC069281.2	chr7:100106901-100107600	408	11	10	GTTTTGTGGATTAGGAATTTAGGGT
SQ00036_SEQ2_RV						ATTACAAACATAAACCACCACACCT
SQ00037_SEQ2_FW	CAV2	chr7:116140401-116141100	441	23	21	TTTGGGATAAGAGGAGAGTAGGGGGTA
SQ00037_SEQ2_RV						AACCAAAAATTCTTAACCAAATTCC
SQ00038_SEQ2_FW	AC084864.2	chr7:129424901-129425600	527	16	13	AGTTTTGAGGATTAGGTTTTAGGGT
SQ00038_SEQ2_RV						AACATTCTCATCCAAAAAAAACACA
SQ00039_SEQ2_FW	FALSE	chr7:157481401-157482100	460	41	33	GTTTGTTAGAATTTTTTTGTGGTA
SQ00039_SEQ2_RV						AAAAAACTAAAAACTCTTCCTTTCACC
SQ00004_SEQ2_FW	DTX4	chr11:58940401-58941100	596	29	23	AGGAAGTTTGGGGATAGTTTTTTTT
SQ00004_SEQ2_RV						TAATTTTACAAACCCAACTTCAACC
SQ00040_SEQ2_FW	CHN2	chr7:29185401-29186100	489	21	18	AAAAGTTTTTAGTATTAATAGGGGAGAAA
SQ00040_SEQ2_RV						CACTTCACAAAAAACTACTTCCACC
SQ00041_SEQ2_FW	NKX2-6	chr8:23563901-23564600	392	27	25	GGGTTGTATAGTAAGGATTTTATTTTTAAT
SQ00041_SEQ2_RV						CAAACAAACCCAAACTCTAAAACAAAC
SQ00005_SEQ2_FW	CHRM1	chr11:62690901-62691600	527	37	32	AATTGAGATTTAGAGAGGGGGTTTA
SQ00005_SEQ2_RV						CCCAAACTTAACTAACTCACCCTTA
SQ00006_SEQ2_FW	CHST11	chr12:104851901-104852600	541	35	31	TATTGTAGTTTTGGAGGAAATTTTGG
SQ00006_SEQ2_RV						AAAAACAAAAATAAAAAACACCAACC
SQ00007_SEQ2_FW	AL160397.2, AL160397.1	chr13:47530401-47531100	573	3	1	AGAATAGTTTTAGGTTTATAGTAAATTGAG
SQ00007_SEQ2_RV						AAAAACAACTAAAATACCTTCCAAT
SQ00008_SEQ2_FW	SALL2	chr14:22004401-22005100	459	12	11	TGGGGGTTTTTAAGTTTAAGTTTGT
SQ00008_SEQ2_RV						ACTCTCAATTCAACCCAACAATAAA
SQ00009_SEQ2_FW	COCH	chr14:31344401-31345100	500	24	22	TATTTGTTTTTTTTATGGTAGGGGG
SQ00009_SEQ2_RV						TAAAAAATACTTAACCCTCATTCCT

3. **BS-MS 3**: Analysis of methylation values in three promising prostate tumour marker regions (*TAC1*, *PTPRN2*, *GHSR*) in 94 prostate samples (+ *in vitro* methylated and unmethylated SW480-DNA) and analysis of two regions (TMP1, TMP2) distinguishing *TMPRSS2:ERG* fusion positive and negative samples in 46 tumour samples and fully methylated and fully unmethylated SW480 DNA (**Table 19**).

Table 19: Primers used to amplify BS converted DNA regions for BS-MS analysis 3. Column headings denote the primer name, the position of the amplicon in a TSS associated region, the genomic region containing the BS-MS target region, the size of amplified target region and the number of CpGs in amplified target region, the number of CpGs assayable by BS-MS, and the core sequence of the primers. To the 5'-end of each forward primer the sequence **aggaagagag** was added, and to the 5'-end of each reverse primer the sequence **cagtaatacgactcactatagggagaaggct** was added.

Amplicon Name	Promoter region	Genomic location	Size	CpG	CpG covered	Sequence (5'-3')
SEQ3_3_FW	TAC1	chr7:97360901-97361600	514	35	28	GGTGTTGAGTTTTTTTGGTTTTTT
SEQ3_3_RV						AATACCTACAATTATTTAACCATCCTAA
SEQ2_19_FW	GHSR	chr3:172165401-172166100	509	38	29	GGTTTGTTTGTGGTTTTGGTTTT
SEQ2_19_RV						TCTTCCTCTACATACCCCTAAACCT
SEQ2_39_FW	FALSE	chr7:157481401-157482100	460	41	33	GTTTGTTAGAATTTTTTTGTGGTA
SEQ2_39_RV						AAAAAACTAAAAACTCTTCCTTTCACC
TMP1_FW	FALSE	chr1:149032901-149033600	384	13	6	TTAATGAAGGGATTTGAATTGTGTTATT
TMP1_RV						TCATCCCCAATAATATTTCCTATTCTA
TMP2_FW	FALSE	chr16:46413901-46414600	294	12	5	GAATGGAATGGAATGGAATTATTTT
TMP2_RV						AATTTCCTTTACTTCCCTTTAATAATCA

2.3.9.2 BS-MS data analyses

For correlation analysis of BS-MS and MeDIP-Seq values the mean BS-MS value of a region was multiplied with the number of CpGs (seqval) in the respective region and compared to the respective MeDIP-Seq values (rpm) of that region. Spearman correlation coefficients were calculated using the following function in R:

```
cor.test(seqval, rpm, method="spearman")
```

Analysis of the tissue preparation methods was performed by CpG unit-wise comparison of matched micro- and macrodissected tissues.

BS-MS values were subjected to unsupervised hierarchical cluster analyses using the heatmap function in R.

CpG-wise methylation plots were used for selecting single CpG units suitable for urinary DNA methylation analysis with qPCR.

2.3.10 Determination of copy number alterations

Experiments were conducted by the group of Holger Sültmann. Data analyses was performed by Maria Fälth and Axel Fischer. Therefor, DNA from 41 tumor samples was hybridised on Affymetrix Genome-Wide SNP Arrays 6.0. The raw data was preprocessed and copy number alterations (CNAs) were estimated with CRMA v2 (BENGTSSON 2009). The median of all tumor samples was used as reference sample.

2.3.11 Gene expression analyses

2.3.11.1 Data generation

Gene expression values of the prostate samples were generated at the DKFZ in Heidelberg by the group of Holger Sültmann and primary data analyses and normalisation was performed by Maria Fälth. In brief, the Affymetrix GeneChip Whole Transcript Sense Target Labeling Assay was used to generate amplified and labeled sense DNA. 1µg of total RNA were initially used for rRNA reduction. Following manufactor's instructions, cDNA was hybridised to the Affymetrix 1.0 Human Exon ST arrays. The raw data files were preprocessed and normalised using Affymetrix powertools. The Affymetrix data was MIAMI-compliant submitted to GEO database (NCBI, GEO, GSE29079).

2.3.11.2 Data analysis

For analyses of differential expression nonparametric two-tailed Mann-Whitney tests with subsequent correction of the p-values for multiple testing (Benjamini-Hochberg approach (BENJAMINI 1995)) were applied. For correlation analyses only samples with gene expression and methylation data were considered (gene expression data: n_{normal}=48, n_{tumor}=47, n_{FUS+}=17, n_{FUS-}=20). The expression values of each gene in a subset of samples (normal and tumour samples, or normal and FUS+ samples, or normal and FUS- samples) were compared with the methylation values of these samples in all TSS associated bins of all respective gene transcripts (Ensembl59) using nonparametric Spearman correlation, yielding rho correlation values and significance p-values for each test. P-values were corrected for multiple testing using the Benjamini-Hochberg approach (BENJAMINI 1995), and bins with p_{BH}-values <0.05 were called ‘significantly correlated’. For analyses of

homeobox and *EZH2* target genes the genes described by Holland *et al.* (HOLLAND 2007) and Yu *et al.* (YU 2007) were used, respectively.
For enrichment analyses of functional terms the 'Database for Annotation, Visualization and Integrated Discovery' software (DAVID, http://david.abcc.ncifcrf.gov) was used. Gene sets were analysed using the standard parameters and significantly enriched functional terms were extracted.

2.3.12 miRNA expression analyses

2.3.12.1 Data generation

Total RNA including miRNAs was extracted from tissue sections and quantitative real time PCR amplification of miRNAs were performed by Jan-Christoph Brase *et al.* using low-density TaqMan arrays v2.0 and the TaqMan MicroRNA Reverse Transcription Kit and TaqMan MicroRNA Assays (Applied Biosystems) (BRASE 2011).

2.3.12.2 Data analysis

Differential expression was assessed using the Mann-Whitney approach with subsequent correction of p-values (Benjamini-Hochberg correction (BENJAMINI 1995)). MiRNAs with $p_{BH}<0.05$ were called differentially expressed.
Spearman correlation analyses of expression values and MeDIP values of all bins in an area of ±2kb around the miRNA genes were performed in several patient subgroups ($n_{normal}=48$, $n_{tumor}=51$, $n_{FUS+}=17$, $n_{FUS-}=20$). Correlations with Benjamini-Hochberg corrected p-values of <0.05 were called 'significant'.
Correlations of miRNA expression data and gene expression data were determined similarly ($n_{normal}=48$, $n_{tumor}=47$, $n_{FUS+}=17$, $n_{FUS-}=20$).

2.3.13 Experiments in prostate cancer cell lines

Table 20: Additional cell lines and consumables used for experiments with prostate cell lines.

Article	Catalog Number	Distributor
DU145 cell line	HTB-81	ATCC
LNCaP cell line	CRL-1740	ATCC
PC-3 cell line	CRL-1435	ATCC
RWPE-1 cell line	CRL-11609	ATCC
VCaP cell line	CRL-2876	ATCC
5-aza-2`-deoxycytidine	A3656	Sigma
ERG custom siRNAs: CGACAUCCUUCUCUCACA duplexes with UU overhang		Thermo Scientific
miRIDIAN mimic human hsa-miR-26a	C-300499	Thermo Scientific
OnTarget Plus siRNA control pool	D-001810-10	Thermo Scientific
OnTarget Plus SMARTpool human EZH2	L-004218	Thermo Scientific

Table 21: Primers used for qPCR quantification of cell line cDNA.

Name	Sequence	Purpose
EZH2_FW	TGTGGATACTCCTCCAAGGAA	EZH2 expression analysis
EZH2_RV	GAGGAGCCGTCCTTTTTCA	
ERG_FW	AAGTAGCCGCCTTGCAAAT	ERG expression analysis
ERG_RV	CAGCTGGAGTTGGAGCTGT	
GAPDH_FW	GCTCTCTGCTCCTCCTGTTC	GAPDH expression analysis (control)
GAPDH_RV	GCGGCCGGGTCCGTTGGCTC	
TBP_FW	CGGCTGTTTAACTTCGCTTC	TBP expression analysis (control)
TBP_RV	CACACGCCAAGAAACAGTGA	

VCaP, LNCaP, PC-3, DU145 and RWPE-1 cells were obtained from ATCC (American Type Culture Collection). All cells were maintained and propagated according to the recommendations of ATCC. DNA and RNA were isolated following standard protocols.

2.3.13.1 Knock down of *EZH2* in DU145 cells and transfection of DU145 cells with *miR26a* mimics

Knock down of *EZH2* in DU145 cells and transfection of DU145 cells with *miR26a* mimics were performed by Behnam Sayanjali: 200,000 DU145 cells were seeded into each well of a 6well plate and grown to 70% confluency in MEM supplemented with 1% L-glutamine, 10% FCS and penicillin/streptomycin. Before transfection, the medium was changed to antibiotics free medium. 100pmol of the RNA (non-targeting OnTarget Plus siRNA control pool, *EZH2*-siRNA: OnTarget Plus SMARTpool human *EZH2*, or *miR26a* mimics: miRIDIAN mimic human hsa-miR-26a (Dharmacon), respectively) in 250µl MEM and 250µl of MEM containing 5µl Lipofectamine 2000 were mixed, incubated at RT for 20min and carefully pipetted to the cells. After 72h of incubation DNA and RNA were isolated and subjected to further analyses. Expressions of *EZH2* and *TBP* were assessed by real time qPCR.

2.3.13.2 Knock down of *ERG* in VCaP cells

ERG knock down experiments were performed by Mark Laible. For the knock down of *ERG*, VCaP cells were transfected with 50nM of OnTarget Plus siRNA control pool or *ERG* custom siRNA (SUN 2008) (CGACAUCCUUCUCUCACA duplexes with UU overhang) (both from Dharmacon) using Lipofectamine RNAiMAX (Invitrogen) as recommended by the supplier. Cells were harvested for DNA and RNA extraction 96h after transfection. Expressions of *EZH2*, *ERG* and *TBP*, as well as *miR26a* expression were assessed by real time qPCR.

DNA of two VCaP *ERG* knock down experiments and VCaP untransfected and non-targeting control cells that was obtained from Mark Laible (DKFZ) was subjected to MeDIP-Seq analysis. Sequencing was performed on a full slide with SOLiD version 4. Mapping and filtering of reads was performed as described above. Reads were assigned to 500bp non-overlapping genomic bins and for each bin a sample-wise MeDIP-Seq value (rpm value) was calculated. Assuming a binomial distribution of reads as null hypothesis, significantly methylated regions were identified. Bins with p-values below 0.001 in both the control experiments or in both the *ERG* knock down experiments were called 'significantly methylated'. Log_2(ratios) of the MeDIP-Seq values of each *ERG* knock down to each control experiment were calculated and significantly methylated bins with values of <(-0.5) or >0.5 were called 'significantly differentially methylated'.

2.3.13.3 5-aza-2`-deoxycytidine treatment of PC-3 cells

5-aza-2`-deoxycytidine (Sigma) experiments were performed by Mark Laible. PC-3 cells were incubated for 96h with 0.5μM, 2 μM or a corresponding amount of DMSO (vehicle control). The medium was changed every 24h. Cells were then harvested for DNA and RNA extraction.

2.3.14 In vitro methylation and luciferase reporter assay

Table 22: Additional equipment, consumables and cell lines used for *in vitro* methylation and luciferase reporter assays.

Article	Catalog Number	Distributor
HEK293T cells	CRL-11268	ATCC
GloMax Multi Detection System	E7031	Promega
Neubauer cell counting chamber	T728.1	Carl Roth GmbH
Water bath WNB7	84198998	Memmert
96-Well Clear-Bottom Plates, Tissue culture treated; White	07-200-566	Fischer Scientific
Ampicillin 5g	A6140	Sigma
ApaL1 restriction enzyme	R0507S	New England Biolabs
BglII restriction enzyme	R0144S	New England Biolabs
CpG Methyltransferase (M.SssI)	M0226S	New England Biolabs
Dual-Luciferase Reporter Assay System /100 assays	E1910	Promega
EndoFree Plasmid Maxi Kit (10)	12362	Qiagen
HindIII restriction enzyme	R0104S	New England Biolabs
LB agar		self-made
LB broth		self-made
pGl3 basic vector (Firefly)	E1751	Promega
Polyethyleneimine		
pRL-TK vector (Renilla)	E2241	Promega
QIAprep Spin Miniprep Kit (250)	27106	Qiagen
Quick Ligation Kit	M2200S	New England Biolabs
Quick T4 DNA Ligase	M2200S	New England Biolabs
T4 DNA Ligase	M0202T	New England Biolabs
T4 DNA Ligase rapid (600U/µl)	L-6030-HC-L	Enzymatics

Table 23: Oligonucleotides used for the luciferase methylation assay.

Name	Sequence	Purpose
miR26a_FW	TTTAAGCTTGGGTGACAGGAGAGGAGACA	Amplification of miR26a region
miR26a_RV	TTTAGATCTTGGTCATTGAGGGGAAAAAG	
DUOX1_short_FW	TTTAGATCTGCACCGACGGAACATCTCTA	Amplification of DUOX1 promoter region
DUOX1_short_RV	TTTAAGCTTCTCTCGTCCGGTGCCTCT	
miR23b_long_FW	TTTAGATCTTGCTTATGAAAACCATTTCTGTG	Amplification of long miR23b region
miR23b_long_RV	TTTAAGCTTCAGCCAGAGCACCTGAGAG	
miR23b_short_FW	TTTAGATCTCAGCATCTTCGATCCTGTCC	Amplification of short miR23b region
miR23b_short_RV	TTTAAGCTTCAGCCAGAGCACCTGAGAG	
miR26meth_check_FW	GGGAAGATGTGGTGTTAATTTTT	Sanger Sequencing of BS converted miR26a construct
miR26meth_check_RV	ATAATCCAACAACTCCCTTTTAATC	

Luciferase reporter assays were developed to assess the effect of methylation of a miRNA associated region on gene expression. Target regions (*miR23b*, *miR26a*, *DUOX*) containing BglII and HindIII restriction sites were amplified from PC3 cell line DNA.

For each reaction 5µl of 10x PCR Buffer, 2µl of 10mM dNTPs, 2µl of Taq polymerase, 1.25µl of primer mix (10µM each), 1µl PC3 DNA, 0.2µl Pfu polymerase (2.5U/µl), and 38.55µl of NFW were mixed. Amplification conditions were as follows: 2min 95ºC; 40 cycles of 15s 95ºC, 30s Tm, 1.15min 72°C; 5min 72°C; forever 4°C.

Amplicons were purified with Qiagen's PCR purification kit, eluted in 30µl EB and digested with BglII and HindIII (NEB) restriction enzymes: 20µl of DNA, 3µl NEB2 buffer, 0.5µl HindIII, 1µl BglII, and 5.5µl NFW were mixed and incubated for 1h at 37°C.

Digested amplicons were purified by gel electrophoresis and extracted from the gel with Qiagen's gel extraction kit using MinElute columns and 20µl EB for elution.

Fragments were ligated into the BglII/HindIII digested pGL3_basic vector (Promega) in 5fold excess, thus, for each nanogram of vector 5*insert size/vector size of insert were used. In 20µl total reaction volume the appropriate amount of insert and NFW, 10µl of 2x Quick Ligation Buffer (NEB), and 1.5µl of T4 ligase (3U/µl; NEB) were used and incubated for 1h at RT. Ligated vectors were purified with MinElute columns and eluted in 30µl EB.

Six nanograms of plasmid (max. 10µl) were used to transform 30µl of chemocompetent *E. coli* DH5α cells for massive plasmid amplification by 30min incubation on ice followed by a 45s heat shock and immediate suspension in 800µl LB buffer. After 1h of incubation at 37°C cells were pelleted by 5min centrifugation at 5,000rpm and the pellet was streaked on a LB agar plate containing ampicillin (50µg/ml) and incubated overnight at 37°C. The next day colony PCRs were performed using 1µl of 10x PCR buffer, 0.4µl dNTPs (10mM), 0.4µl Taq polymerase, 0.25 primer mix (10µM each), and 7.95µl NFW and some scratched cells of the colony. Colonies showing the target were incubated overnight at 37°C in 3ml LB/amp medium and the next day plasmids were isolated using the Qiagen MiniPrep kit and a restriction with HindIII/BglII was performed to validate the plasmid.

To obtain unmethylated DNA the plasmids were cloned into the DNA methylase negative *E. coli* GM2929 strain: 20µl of chemocompetent cells were incubated for 30min on ice with 1µl isolated plasmid and heat shocked for 45s at 42°C. After 1h incubation at 37°C in 800µl LB, 20µl of the medium was streaked on LB/amp plates and incubated overnight at 37°C. Colony PCRs were performed and plasmids were prepared of the successfully amplifying colonies using Qiagen MiniPrep kit. Restriction digest was performed of the isolated plasmids to check for the right insert size. 100ml of LB/amp were inoculated with successfully transformed colonies and incubated overnight at 37°C. Cells were pelleted and plasmids were isolated using the EndoFree Plasmid Maxi kit (Qiagen). One microgram of isolated plasmids was subjected to *in vitro* DNA methylation with SssI-CpG-Methyltransferase (NEB) following the manufacturer's instruction: up to 5µl of DNA were mixed with 5µl NEB2 buffer, 2µl SAM (1600µM), 1µl SssI-CpG-Methyltransferase and NFW to yield 50µl, and incubated 1h at 37°C. The enzyme was deactivated by 20min incubation at 65°C.

Methylated DNA was purified with MinElute columns and eluted in 30µl EB. Methylation was confirmed by methylation sensitive restriction enzyme digest (ApaL1, NEB). MiR26a

plasmid methylation was also confirmed by bisulphite conversion and subsequent amplification and Sanger sequencing of the amplicon using the following bisulphite converted DNA specific primers miR26meth_check_FW and RV.

Artificially methylated and unmethylated plasmids were transfected into HEK293T cells: ~10,000 cells/well were seeded into a 96well plate and grown overnight in 100µl DMEM supplied with 10% FCS, 2mM glutamine and penicillin/streptomycin. After 24h (40-60% confluence) medium was changed to antibiotics free medium. Plasmids were mixed 1:1 with pRL-TK-Renilla-plasmid as background control to a final concentration of 50ng/µl. 4µl of this mix were incubated with 13µl DMEM (without FCS, penicllin/streptomycin, glutamine) and 0,4µl polyethyleneimine (PEI; 1mg/ml stock). A Dual-Luciferase Reporter assay (Promega) was performed 24h after transfection according to the manufacturer's instructions. Luciferase signals were detected using the Promega Multi Detection System. Ratios of firefly luciferase signal (expressed pGL3_basic plasmid) and Renilla luciferase signal (expressed pRL-TK) were compared to determine the effect of methylation on promoter activity.

2.4 Detection of prostate tumour associated DNA methylation in urinary DNA

Table 24: Additional consumables used for isolation of urinary DNA.

Article	Catalog Number	Distributor
Urine (Exfoliated Cell) DNA Purification Kit	22300	Norgen Biotek
Urine Collection and Preservation Tubes	18111	Norgen Biotek
QiaAmp DNA Mini Kit (50)	51304	Qiagen
RNase A (17,500 U) 2.5 ml (100 mg/ml; 7000 units/ml, solution)	19101	Qiagen

Table 25: Primers used for DNA methylation analyses of urinary DNA.

Name	Sequence	Purpose
ZSCAN12_AMPLI_FW	TGGAATTGCTGCAGAGACTG	Generation of ZSCAN12 fragment prior to *in vitro* methylation and BS conversion
ZSCAN12_AMPLI_RV	CGGGAGATTGCTGTATTCCA	Generation of ZSCAN12 fragment prior to *in vitro* methylation and BS conversion
ZSCAN12_TAQ_FW	GTATTTTGGGGAGGGAGAGTTT	Amplification and quantification of BS converted ZSCAN12 fragment
ZSCAN12_TAQ_RV	CAACTACTAAACCCCTCAAACC	Amplification and quantification of BS converted ZSCAN12 fragment
ZSCAN12_MSP_UNMETH_RV	CCAATACAAAATAAACCCCAAT	Together with ZSCAN12_TAQ_FW quantification of BS converted unmethylated ZSCAN12
ZSCAN12_MSP_METH_RV	GATACAAAATAAACCCCGAT	Together with ZSCAN12_TAQ_FW quantification of BS converted methylated ZSCAN12
PTPRN2_AMPLI_FW	CTCTGTGGCACGACTTTGAG	Generation of PTPRN2 fragment prior to *in vitro* methylation and BS conversion
PTPRN2_AMPLI_RV	GCGATCGAACAAGAGAGGAA	Generation of PTPRN2 fragment prior to *in vitro* methylation and BS conversion
PTPRN2_QUANT_FW	TATTTTTATAGGAGTTATAAAGATTAGGTT	Amplification and quantification of BS converted PTPRN2 fragment
PTPRN2_QUANT_RV	TAATTAAAAACCTTAATCCTCCCTC	Amplification and quantification of BS converted PTPRN2 fragment

Name	Sequence	Purpose
PTPRN2_UNMETH_BS_FW	AGGTTAAGGGTTTTTGGGAGTT	Together with PTPRN2_QUANT_RV quantification of BS converted unmethylated PTPRN2
PTPRN2_METH_BS_FW	GTTAAGGGTTTTCGGGAGTC	Together with PTPRN2_QUANT_RV quantification of BS converted methylated PTPRN2
CAV2_BBST_FW	TGATTTCAAAGAGGGCATGG	Generation of CAV2 fragment prior to *in vitro* methylation and BS conversion
CAV2_BBST_RV	CCGTCCCACCATTGCTAC	Generation of CAV2 fragment prior to *in vitro* methylation and BS conversion
CAV2_AMPLI_FW	TAGATTTATATTTTGTTAAAGGAGTG	Amplification and quantification of BS converted CAV2 fragment
CAV2_AMPLI_RV	AAAAATTCTTAACCAAATTCCTATAC	Amplification and quantification of BS converted CAV2 fragment
CAV2_UNMETH_RV	AAAAATTCTTAACCAAATTCCTATACA	Together with CAV2_AMPLI_FW quantification of BS converted unmethylated CAV2
CAV2_METH_RV	AATTCTTAACCAAATTCCTATACG	Together with CAV2_AMPLI_FW quantification of BS converted methylated CAV2

Quantitative methylation specific PCR (qMSP) assays were set up to assess the value CpG methylation for non-invasive detection of prostate cancer in urinary DNA. Methylation levels of single CpG dinucleotides were determined using PCR primers specific for formerly methylated bisulphite converted DNA and primers for formerly unmethylated bisulphite converted DNA in parallel. Three independent sets of patient's DNA/exprimate urine were obtained with patient's informed consent at the Martini clinic in Hamburg (set 1 and 3) and at the Charité Berlin (set 2).

2.4.1 Isolation of DNA from urine cells

Urine from 20 patients for DNA methylation analyses was obtained with patient's informed consent after prostatic massage during a digital rectal examination (DRE) at the Charité Berlin. All patients were biopsised because of suspicious findings. Patients were classified as tumour or healthy patients according to the results of at least 10 biopsies taken (maximum 19). 5ml of exprimate urine were centrifuged for 15min at 1000g at 4°C to pellet the cells. The supernatant was stored at -20°C for further analysis. Cells were resuspended in 200µl PBE.

To isolate DNA from the cells the QiaAmp DNA Mini Kit (Qiagen) was used. 20µl of proteinase K and 4µl of RNase A (100mg/ml) were added to the cells and mixed well. 200µl Buffer AL were added, and digest was performed for 10min at 56°C. After incubation, 200µl of 100% EtOH were mixed with the sample and the mixture was applied to a QIAamp Mini spin column followed by centrifugation. The column was washed with 500µl Buffer AW1 and transferred to another 2ml collection tube. Another washing was performed using 500µl buffer AW2 (3min centrifugation). An additional centrifugation

step with a new 2ml collection tube was conducted. The column was transferred to a new 1.5ml collection tube and was eluted with 50µl of buffer AE (elution buffer) for at least 5min and DNA concentration was measured using the Nanodrop ND-2000 system. DNA was subjected to bisulphite conversion using Qiagen's EpiTect Bisulfite Conversion kit and analysed with methylation specific real time qPCR assays.

In another approach (urine set 3 part 1), urinary DNA of 20 tumour patients was isolated from the cellular content of up to 50ml exprimate urine collected at the Martini clinic in Hamburg. Post collection, urine samples were added with Norgen's Urine Preservative for conservation. Cells were pelleted by 5min centrifugation at 650g and counted using a cell counting chamber. For DNA isolation the Urine (Exfoliated Cells) DNA Isolation Kit of Norgen Biotek was used: After addition of 300µl Lysis Solution cells were incubated at 55°C for 30min. 60µl of Binding Solution were mixed with the samples that were afterwards applied to the kit's spin columns. Columns were centrifuged for 3min at 5,200g and washed with 500µl Wash Solution I. After centrifugation, the columns were washed with 500µl of Wash Solution II followed by another centrifugation step. 450µl of 100% EtOH were added to each column, centrifuged and the columns were washed with another 450µl of 100% EtOH. After discarding of the flow-through columns were dried at 55°C for 3min.
DNA was eluted with 100µl Elution Buffer followed by another elution step with 100µl Elution Buffer. The liquid was evaporated in a heated vacuum centrifuge and samples were subjected to BS conversion using Zymo's DNA Methylation Gold Kit.
The second part of urine set 3 was isolated independently from the first part. It consisted of 10 urine samples from patients with negative prostate biopsies but elevated PSA values (3.4-8.59ng/ml). Isolation of DNA from urine was again performed using Norgen's Urine DNA Isolation Kit. Isolated DNA was precipitated with ammonium acetate and ethanol, and was BS converted using Qiagen's EpiTect BS conversion kit.

2.4.2 Assay Design

To develop methylation specific real time qPCR assays for promising marker regions the CpG-wise methylation levels obtained by BS-MS were analysed, because this approach provides a higher resolution than the MeDIP-Seq approach. For single CpGs showing the highest discriminatory potential in the *ZSCAN12*, the *PTPRN2* and *CAV2* gene bisulphite

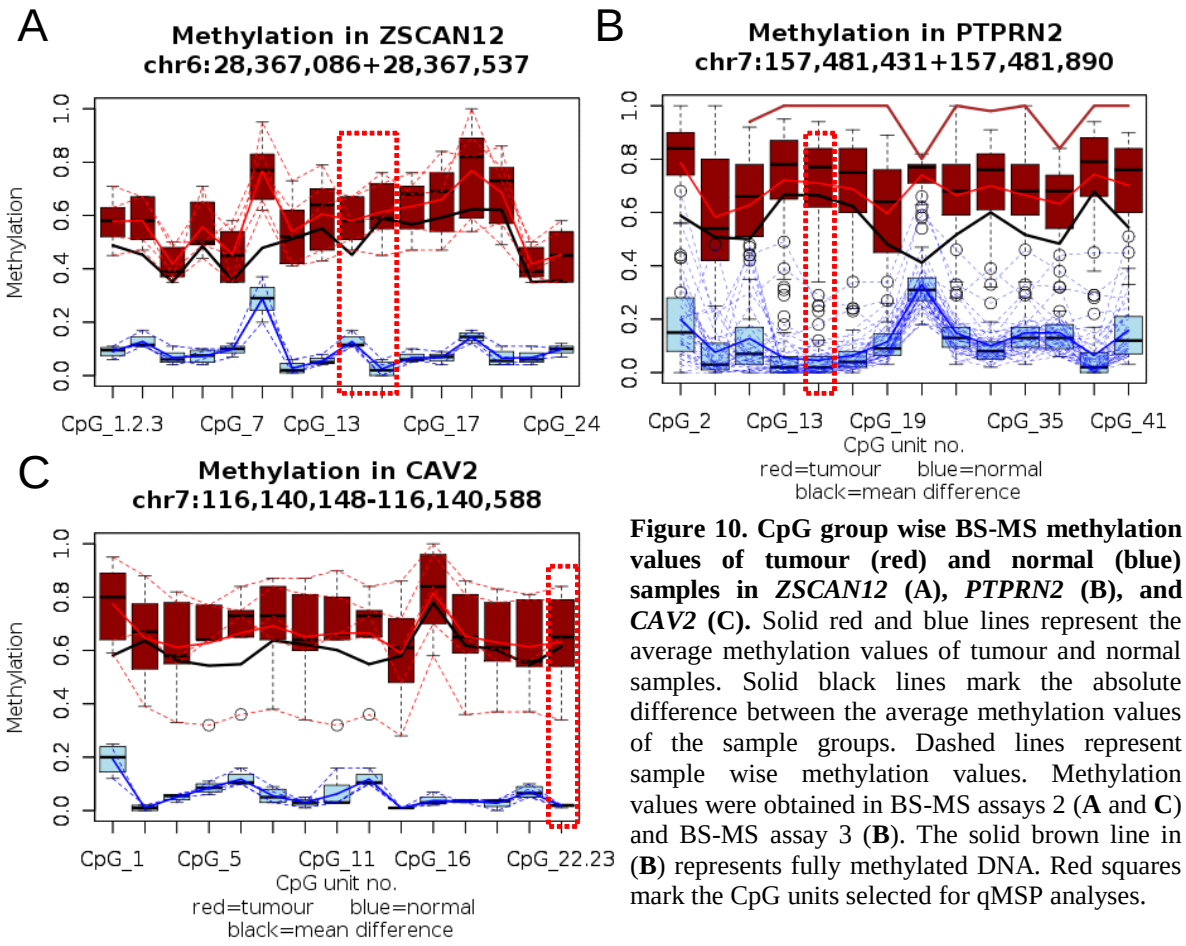


Figure 10. CpG group wise BS-MS methylation values of tumour (red) and normal (blue) samples in *ZSCAN12* (A), *PTPRN2* (B), and *CAV2* (C). Solid red and blue lines represent the average methylation values of tumour and normal samples. Solid black lines mark the absolute difference between the average methylation values of the sample groups. Dashed lines represent sample wise methylation values. Methylation values were obtained in BS-MS assays 2 (**A** and **C**) and BS-MS assay 3 (**B**). The solid brown line in (**B**) represents fully methylated DNA. Red squares mark the CpG units selected for qMSP analyses.

specific primers were designed (**Figure 10**). Selected CpGs were located at chr6:28,367,331-28,367,332 (CG14 *ZSCAN12*), chr6:28,367,348-28,367,349 (CG15 *ZSCAN12*), chr7:157,481,628-157,481,629 (CG14 *PTPRN2*) and chr7:116,140,177-116,140,178 (CG23 *CAV2*). The underlying principle of the assay is that the amounts of formerly methylated and unmethylated cytosines (i.e. the amount of cytosines and thymines) at a specific CpG group were determined in a qPCR assay after bisulphite conversion of the patient's DNA using primer sets carrying either a 'G' or an 'A' at the 3'-end. For quantitation of the DNA an additional bisulphite specific primer set that is not influenced by former DNA methylation was used.

Assay conditions were optimised using artificially methylated and unmethylated DNA templates. Therefor, a region containing the target CpG was amplified from unconverted cell line DNA using standard PCR conditions with 0.2µl of additional Pfu DNA polymerase added per 50µl resulting in an unmethylated amplicon. Half of the DNA was methylated *in vitro* with SssI methyltransferase. Methylated and unmethylated fragments were BS converted and amplified using a BS specific primer set enclosing the target region. In real time qPCR the optimal amplification protocol was established by

minimising the false positive rate of the methylation specific primer set (falsely detecting unmethylated DNA) and the primer set specific for unmethylated DNA (falsely detecting methylated DNA). Three different regions were examined: *ZSCAN12*, *PTPRN2*, and *CAV2*.

ZSCAN12

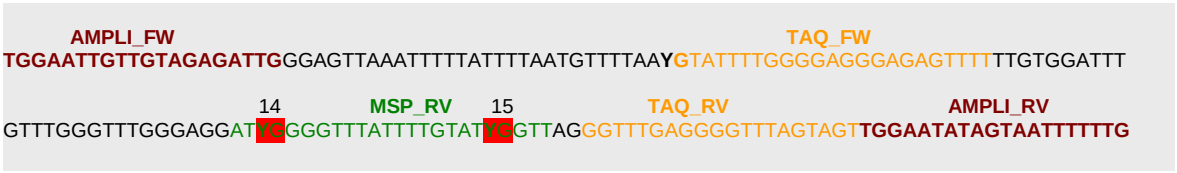


Figure 11. Localisations of primers used for the *ZSCAN12* assay in BS converted DNA. Marked in red are the CpG groups whose methylation levels were assessed. Region: chr6:28367230+28367394 (plus strand).

To generate fragments containing the CpGs 14 and 15 of the *ZSCAN12* BS-MS assay (SEQ2_30) for *in vitro* methylation AMPLI_FW and AMPLI_RV primers were used. Following BS conversion the fragments could be quantified using primers TAQ_FW and TAQ_RV in qPCR analysis. Primers TAQ_FW and MSP_METH_RV, and TAQ_FW and MSP_UNMETH_RV were used to determine the methylated and unmethylated fractions of the interrogated CpGs, respectively (**Figure 11**). The *ZSCAN12* assay was used to probe urine set 1 and set 2.

PTPRN2

To generate fragments containing CpG 14 of the *PTPRN2* BS-MS assay (SEQ2_39) for *in vitro* methylation AMPLI_FW and AMPLI_RV primers were used. Following BS conversion the fragments could be quantified using primers QUANT_FW and QUANT_RV in qPCR analysis. Primers METH_BS_FW and QUANT_RV, and UNMETH_BS_FW and QUANT_RV were used to determine the methylated and unmethylated fractions of the interrogated CpG, respectively (**Figure 12**). The *PTPRN2* assay was used to probe urine set 2 and set 3.

Figure 12. Localisations of primers used for the *PTPRN2* assay in BS converted DNA. Marked in red is the CpG group whose methylation levels were assessed. Region: chr7:157481445+157481929 (plus strand).

CAV2

Figure 13. Localisations of primers used for the *CAV2* assay in BS converted DNA. Marked in red is the CpG group whose methylation levels were assessed. Region: chr7:116140061-116140401 (minus strand).

To generate fragments containing CpG 23 of the *CAV2* BS-MS assay (SEQ2_37) for *in vitro* methylation BBST_FW and BBST_RV primers were used. Following BS conversion the fragments could be quantified using primers AMPLI_FW and AMPLI_RV in qPCR analysis. Primers METH_RV and AMPLI_FW, and UNMETH_RV and AMPLI_FW were used to determine the methylated and unmethylated fractions of the interrogated CpG, respectively (**Figure 13**). The *CAV2* assay was used to probe urine set 2 and set 3.

2.4.3 Analyses of the methylation state of marker CpGs in urinary DNA

For analysis of clinical samples, DNA was isolated from exfoliated urinary cells and bisulphite converted. Optionally, the target regions were amplified prior to qPCR measurement using methylation unspecific primers. A two step qPCR protocol was performed with methylation specific primers and primers specific for unmethylated CpGs using the following optimised primer annealing/elongation temperatures: *ZSCAN12* assay 60°C, *PTPRN2* assay 60°C, *CAV2* assay 62°C. The ratio of methylated and unmethylated DNA was used to calculate the total methylation value at this distinct position:

$$\text{Methylation in \%} = 100 * 2^{(Ctunmeth-Ctmeth)}/(1+2^{(Ctunmeth-Ctmeth)})$$

2.4.4 Determination of Receiver Operator Characteristics (ROCs)

To find optimal methylation thresholds for the markers used in urinary DNA methylation analyses Receiver Operator Characteristics (ROCs) were determined as follows. Methylation cutoffs were increased stepwise by 0.5% from 0 to 100% and the true positive (TPR) and false positive rates (FPR) were plotted for each cutoff in a two-dimensional graphic. For ROC analyses of two markers simultaneously, all combinations of

methylation thresholds were tested using steps of 0.5% methylation. Samples were assigned to the tumour group whenever their methylation values exceeded the respective thresholds in both markers.

TPR = count of correctly assigned tumour samples/all tumour samples

FPR = count of falsely assigned normal samples/all normal samples

The methylation value with a maximum ROC value was used as methylation threshold.

ROC value = TPR + 1-FPR

To minimise false positive detection a penalty factor of 2 was added to the ROC value equation:

ROC_{pen} = TPR + 1-2*FPR.

3 RESULTS

The goal of my thesis was to identify alterations of DNA methylation patterns during prostate tumourigenesis. Furthermore, the underlying mechanisms of PCa development should be elucidated integrating gene and microRNA expression data.

3.1 Patient statistics

To determine genome-wide methylation levels we performed 5-methylcytosine antibody based enrichment of methylated DNA regions followed by whole genome sequencing (MeDIP-Seq) for 53 normal and 51 tumour prostate tissues. All tumours selected for this study were staged pT2-pT4. For each patient the following parameters were obtained during tissue resection: Patient age, postoperative Gleason score, TNM grading, *TMPRSS2:ERG* fusion state and the preoperative PSA level (**Supplementary Table 1, Figure 14**). The *TMPRSS2:ERG* fusion was present in 46% of the tumours assessed (n_{FUS+}=17, n_{FUS-}=20). For 14 tumour samples the fusion state could not be determined by qPCR. We did not find significant differences in the distribution of the clinical parameters between tumour and normal as well as between *TMPRSS2:ERG* fusion positive (FUS+)

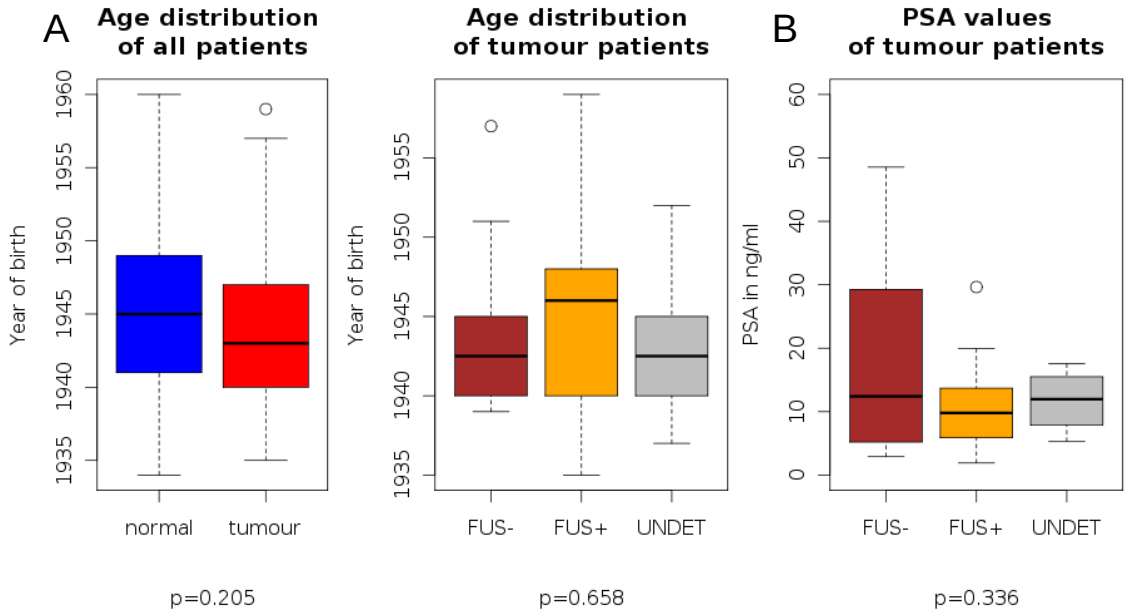


Figure 14. Patient cohort statistics. (**A**) The distribution of the years of birth for each patient group is shown in a boxplot comparing normal and tumour (left) and *TMPRSS2:ERG* fusion negative (FUS-), positive (FUS+) and undetermined (UNDET) cohorts (right). The significance of the observed differences is given below as p-value of the Mann-Whitney test. (**B**) Distribution of preoperative PSA levels (in ng/ml) for each tumour patient cohort. Given below is the significance value for the differences between FUS- and FUS+ samples (Mann-Whitney test).

and negative (FUS-) samples. Fusion negative (FUS-) samples showed a not significantly elevated PSA level.

3.2 MeDIP enrichment and sequencing statistics

We used an immunoprecipitation based enrichment approach with an antibody directed against 5-methylcytosines followed by multiparallel sequencing (MPS) on a SOLiD3+ machine to determine tumour specific alterations in DNA methylation compared to normal prostate tissue samples (MeDIP-Seq) (**Figure 8**). MeDIP-Seq is a technology assessing the whole genome. However, due to the enrichment step only a small (the methylated) fraction of the genome has to be sequenced. Thus, the required sequencing capacity is significantly reduced as compared to BS-Seq where the whole genome has to be sequenced with high coverage to allow determination of CpG specific methylation values. To monitor the performance of this crucial enrichment step – after fragmentation of the DNA, end repair, and ligation of sequencing adapters – an aliquot of the library was spared (input) while the majority of the library was enriched for methylated regions with anti-5meC-antibodies coupled to paramagnetic beads. Next, we determined the fraction of a methylated and thus enriched region (*bromodomain containing 1, BRD1*) compared to an unmethylated and thus depleted control region (*coenzyme Q3 homolog, COQ3*) for each – enriched and input – library by qPCR before subjecting the library to SOLiD sequencing. We aimed for an enrichment of 30x (BUTCHER 2010, WEBER 2005). The average enrichment factor was 89 (median 71) (**Figure 15**).

Due to the high enrichment during the MeDIP step we were able to sequence multiple libraries in parallel to acquire a through put of 15-20 million uniquely mappable reads per library. We sequenced 19 full slides with eight barcoded library samples per slide on the

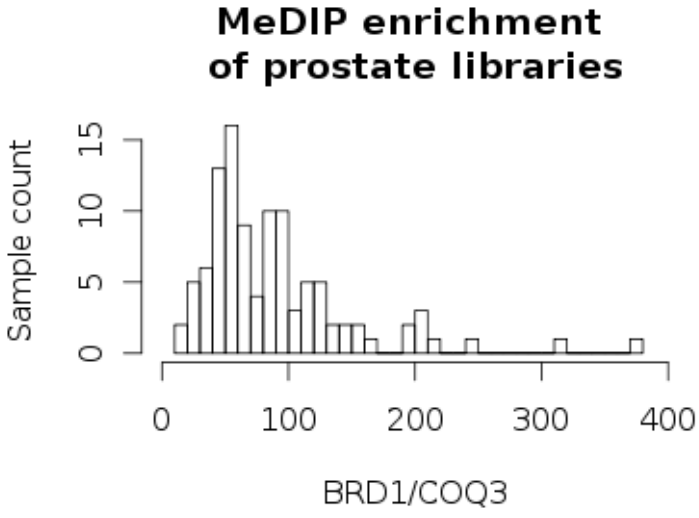


Figure 15. Enrichment of methylated regions. Shown is a histogram of the MeDIP enrichment scores determined by comparing the amounts of methylated *BRD1* and unmethylated *COQ3* in the sequencing libraries before and after MeDIP enrichment. A median enrichment of 71fold was determined.

SOLiD 3+ MPS system. On average, we obtained 57 million reads/sample (raw output, RO) of which 35.5 million were mappable (63%) (mappable output, MO) and 20.7 million were uniquely mappable (36.4%) (uniquely mappable output, UMO) (**Figure 16, Supplementary Table 2**). After *in silico* depletion of all reads without any CpG (see below) and all reads with the same start and end position that potentially resulted from clonal amplifications, we received 14.7 million reads per sample with no significant differences in the sequencing depth of the sample subgroups.

We found a slightly but significantly increased ratio of uniquely mappable reads/mappable reads in tumour samples (Mann-Whitney p-value (p_{MW})=0.00015) indicating a lower fraction of reads originating from repetitive regions that are mapping to multiple genomic locations. This difference was mainly caused by the FUS- samples, since we did not find a significant difference between the UMO/MO ratio of FUS+ and normal samples (p_{MW}=0.23). Comparing FUS+ and FUS- tumour samples we found a significantly higher UMO/MO ratio in FUS- samples (p_{MW}=0.0025). Since MeDIP-Seq enriches GC rich regions, we investigated the GC content to identify potential differences between the sample groups. Indeed, the GC content in tumour tissues was elevated compared to normal tissues (mean: 45.5% and 44.7%, respectively, p_{MW}=0.044). Again, this higher GC content was solely caused by the FUS- samples ($p_{MW\ FUS+/FUS-}$=0.027, $p_{MW\ FUS+/normal}$=0.91, $p_{MW\ FUS-/normal}$=0.0057) (**Figure 16B,C**).

As a quality control, we estimated the genome-wide enrichment of potentially methylated

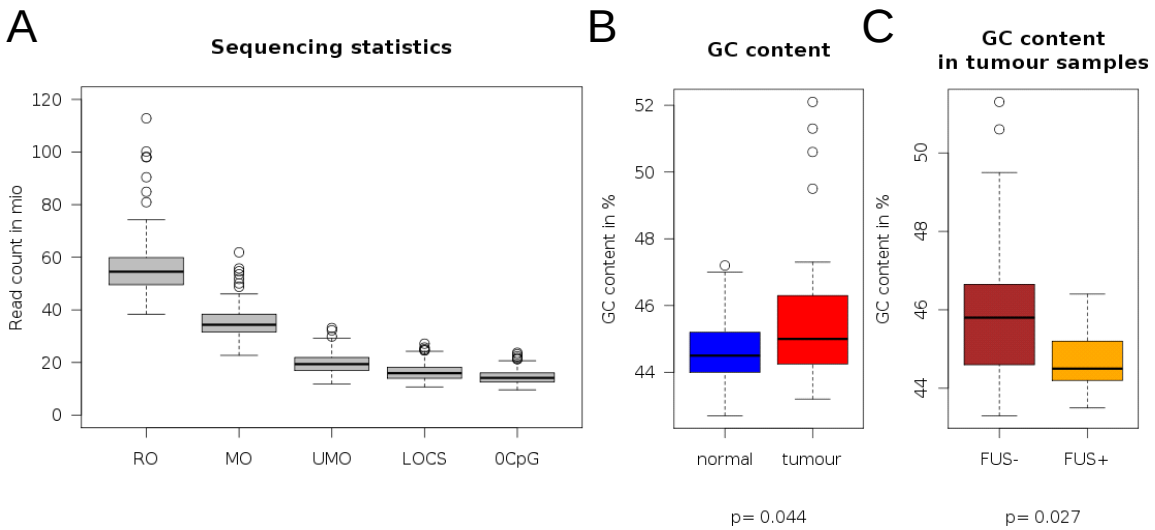


Figure 16. Sequencing statistics. (**A**) The distribution of the numbers of reads among all sequenced samples is given in boxplots for RO = raw output (all reads obtained for the sample during sequencing), MO = mappable output (all reads mapping at least once to the human reference genome), UMO = uniquely mappable output (all reads mapping at only one position in the human reference genome), LOCS = number of UMO reads after depletion of redundant reads, 0CpG = number of LOCS after depletion of reads containing no CpG in the 200bp elongated sequence. (**B**) and (**C**) show the cohort-wise GC content for normal (blue) and tumour (red), and FUS- (brown) and FUS+ (orange) samples. Given below is the Mann-Whitney p-value.

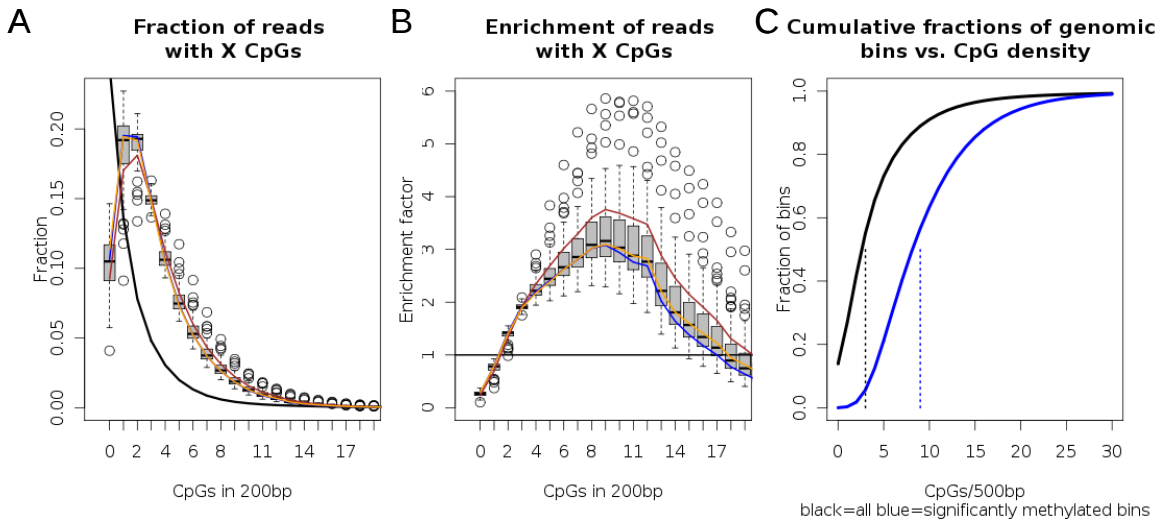


Figure 17. CpG content after MeDIP-Seq. (**A**) After MeDIP-Seq the fractions of reads with a certain number of CpGs in the 200bp elongated reads were counted for each sample. Boxplots show the distribution among the samples. The solid black line shows the background distribution in the human genome. Blue, brown and orange lines represent the median distributions in the normal, the FUS- and the FUS+ cohort, respectively. (**B**) shows the enrichments of reads with a certain number of CpGs among all samples (boxes) compared to the background genomic distribution. Blue, brown and orange lines represent the median enrichments in the normal, the FUS- and the FUS+ cohort, respectively. (**C**) Cumulative fractions of genomic bins with up to x CpGs in the whole genome (black) and in all bins with significant methylation in normal samples (blue). Dotted lines mark the medians of CpGs/500bp in the bin sets.

regions after the sequencing. For this purpose, we counted the numbers of CpGs in the mapped and elongated reads (200bp) of each library. We compared the results to the CpG distribution of the human genome. On average, 11.0% of uniquely mapped reads (median 11.1%) contained no CpG while the fraction in the human genome is 39.3%, meaning that we indeed found a depletion of fragments without any CpG. On the other hand we found an enrichment of fragments with 3-17 CpGs in our samples (**Figure 17A,B)**. Interestingly, we determined significantly higher fractions of reads with more than three CpGs in FUS- samples than in FUS+ or normal samples. Reads containing no CpGs potentially resulted from unspecific binding of DNA to the antibody or the beads and thus do not contain methylation information since non-CpG methylation is a feature of embryonic stem cells (LISTER 2009b). We depleted these reads prior to further analyses. In the human genome, 50% of all 500bp bins contained three and less CpGs while the majority of reads in our enriched normal samples was found in regions with more than seven CpGs/500bp (5^{th} pentile of CpG density in all bins) (**Figure 17C**). Nevertheless, 120,455 out of 566,894 bins with significant read counts in normal samples were located in regions with up to five CpGs/500bp (= 1CpG/100bp).

3.3 Methylation in normal and tumour prostates

Within our enriched fragments we detected at least 10% without any CpG group (**Figure 17A**). Either these regions are background noise resulting from unspecific enrichment or, even though we did not analyse stem cells, harbour non-CpG methylations. As we could not exclude either of the two possibilities we calculated a threshold of minimal read counts per region above which the false discovery rate is below 0.1%: We assumed a binomial distribution of the reads on the genome as null hypothesis and determined p-values for the sample-wise read count of every 500bp bin in order to identify regions that are likely to be enriched because of an underlying methylation and not just by chance. We determined the sample-wise counts of bins with significant methylation and found significantly more regions methylated in tumour than in normal samples (p_{MW}=0.005) (**Figure 18A, Table 26**). Furthermore, comparing FUS+ and FUS- samples we detected a higher number of methylated regions in FUS- tumours (p_{MW}=0.036), with FUS+ specimens exhibiting similar counts as normal samples (**Figure 18C**). To exclude that this difference is caused by differences in DNA copy number, we restricted the analyses to regions not affected by copy number variations and yielded a similar picture – the tumour samples showed significantly more methylated bins than the normal (p_{MW}=0.002) while FUS+ samples showed fewer methylated bins than FUS- samples (p_{MW}=0.03) (**Figure 18B,D**).

Table 26: Sample-wise counts of significantly methylated bins using a p-value cutoff of the binomial distribution of 0.001.

Sample	Significantly methylated bins	Class	Sample	Significantly methylated bins	Class	Sample	Significantly methylated bins	Class	Sample	Significantly methylated bins	Class
IGP84	117505	NORM	IGP103	207544	NORM	IGP124	395024	NORM	IGP60	257258	FUS+
IGP91	141173	NORM	IGP92	211019	NORM	IGP30	174207	UNDET	IGP61	258772	FUS+
IGP186	146240	NORM	IGP180	213921	NORM	IGP54	179522	UNDET	IGP62	260639	FUS+
IGP127	151554	NORM	IGP20	217822	NORM	IGP27	205135	UNDET	IGP80	288798	FUS+
IGP176	151682	NORM	IGP145	218470	NORM	IGP113	206289	UNDET	IGP25	306160	FUS+
IGP142	159754	NORM	IGP188	219862	NORM	IGP59	219740	UNDET	IGP117	321621	FUS+
IGP138	159924	NORM	IGP78	220848	NORM	IGP53	224165	UNDET	IGP55	141547	FUS-
IGP144	169355	NORM	IGP17	226979	NORM	IGP39	226113	UNDET	IGP57	145097	FUS-
IGP182	171464	NORM	IGP75	230684	NORM	IGP63	241027	UNDET	IGP43	155132	FUS-
IGP143	173561	NORM	IGP18	234575	NORM	IGP42	264027	UNDET	IGP32	203914	FUS-
IGP88	175985	NORM	IGP16	236956	NORM	IGP190	269972	UNDET	IGP118	211599	FUS-
IGP187	176271	NORM	IGP76	237369	NORM	IGP111	290376	UNDET	IGP110	221935	FUS-
IGP89	177992	NORM	IGP85	237516	NORM	IGP45	310577	UNDET	IGP31	249398	FUS-
IGP175	179990	NORM	IGP140	240256	NORM	IGP34	319200	UNDET	IGP29	259085	FUS-
IGP90	181042	NORM	IGP79	242131	NORM	IGP82	465238	UNDET	IGP26	279380	FUS-
IGP179	182437	NORM	IGP177	250136	NORM	IGP64	119707	FUS+	IGP33	286158	FUS-
IGP139	183580	NORM	IGP119	253280	NORM	IGP37	135275	FUS+	IGP35	291706	FUS-
IGP185	187247	NORM	IGP174	256845	NORM	IGP105	166100	FUS+	IGP116	292212	FUS-
IGP86	189906	NORM	IGP141	267314	NORM	IGP41	185631	FUS+	IGP23	308684	FUS-
IGP137	196751	NORM	IGP123	281402	NORM	IGP115	189046	FUS+	IGP36	312013	FUS-
IGP120	197285	NORM	IGP189	301950	NORM	IGP44	207846	FUS+	IGP21	326762	FUS-
IGP173	200270	NORM	IGP184	323482	NORM	IGP49	215306	FUS+	IGP38	376301	FUS-
IGP126	202714	NORM	IGP87	335333	NORM	IGP28	230148	FUS+	IGP52	404068	FUS-
IGP102	204073	NORM	IGP122	344261	NORM	IGP40	230553	FUS+	IGP81	410812	FUS-
IGP104	205543	NORM	IGP183	372192	NORM	IGP48	241508	FUS+	IGP51	517029	FUS-
IGP178	206465	NORM	IGP181	390394	NORM	IGP47	250996	FUS+	IGP83	545164	FUS-

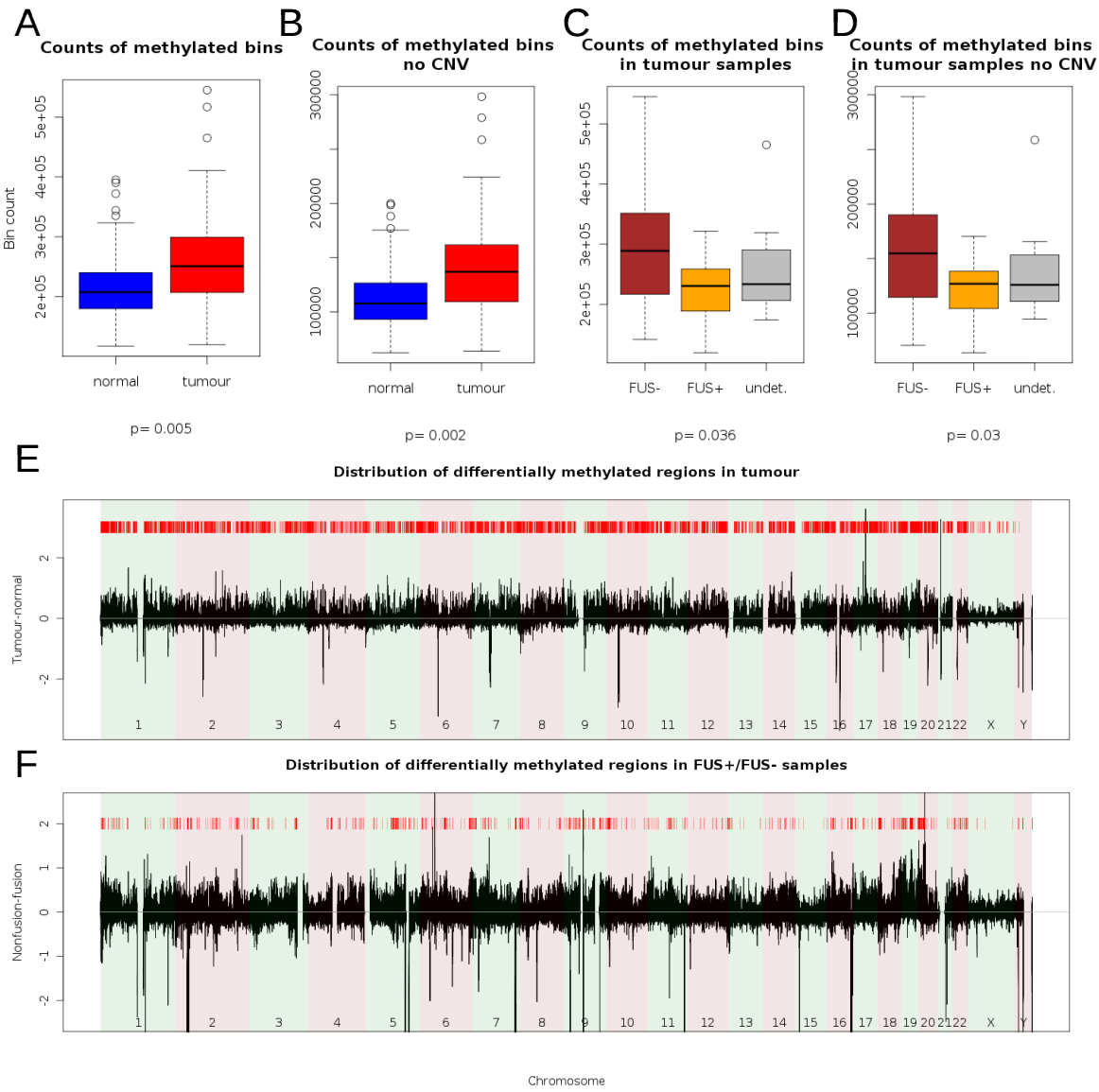


Figure 18. Methylation in prostate samples. The number of methylated bins (i.e. bins with a probability value of <0.001 of their read count in a binomial distribution) was determined for each sample and the distribution of the results is given for normal (blue) and tumour (red) samples in all bins (**A**) and in bins that are outside of regions effected by copy number alterations (**B**). Mann-Whitney p-values indicate the significance of the differences observed. (**C**) and (**D**) show the distributions in subclasses of tumour samples in all regions and non copy number altered regions, respectively. (**E**) shows the distribution of hyper- and hypomethylated regions on the genome as differences of the average tumour and average normal methylation values (Δrpm values). The red marks represent regions with highly significant differences in methylation ($p_{BH}<10^{-7}$). (**F**) shows the differences of the average methylation values of FUS- and FUS+ samples (Δrpm values). The red marks represent regions with highly significant differences in methylation ($p_{BH}<10^{-2}$).

3.4 The average methylation pattern of normal and tumour tissues

To further illustrate differences in the methylation patterns, we determined the average methylation pattern in normal, tumour, FUS- and FUS+ patients. We determined bin-wise average read counts for the sample groups and found the genome significantly

hypermethylated in the tumour compared to the normal tissues (**Figure 18E**). Furthermore, the FUS- genome was hypermethylated compared to the FUS+ genome (**Figure 18F**).
To determine significantly methylated regions over all patients we applied a less stringent cutoff of $p<0.05$ for the binomial distribution since the numbers of significantly methylated bins varied between 100,000 and 550,000 between the patients using a cutoff of 0.001 and thus, we would have missed regions that are significantly methylated in only a subgroup of patients and that are potentially important for this subgroup. We could extract 682,510 bins that were significantly methylated according to their average methylation values in the tumour or normal samples ($p<0.05$). These regions were subjected to principal component analyses. To allow unbiased analyses despite of varying sequencing depths of the samples, we normalised the bin-wise read counts to the total read count of a sample to obtain 'reads per million' (rpm) values that were used for further analyses.

3.5 Principal component analyses

Principal component analyses (PCAs) allow the unbiased identification of underlying patterns in a large set of parameters. As input parameters we used the sample-wise methylation values (rpm) of 682,510 bins found to be methylated (see section 3.4). Methylation patterns clearly distinguished tumour and normal tissues (**Figure 19A**). Furthermore, we also found FUS- and FUS+ samples separated (**Figure 19C**). We observed that FUS+ samples were more similar to normal than to FUS- samples implying that the DNA methylation patterns in FUS- PCa are considerably more altered. In line with previous reports on gene expression (SBONER 2010, SETLUR 2008), the PCA analyses suggested that FUS+ are more homogeneous than FUS- samples. MeDIP enrichment might be biased by DNA copy number alterations (CNAs). To account for that we restricted the analyses to bins not affected by copy number alterations in any of the samples and yielded the same results (**Figure 19B,D**).

3.6 Identification of differentially methylated regions (DMRs)

Besides global differences in DNA methylation we were also interested in the identification of site specific differential methylation events. Thus, we performed bin-wise comparisons of the normalised MeDIP-Seq values (rpm) between tumour and normal samples using nonparametric Mann-Whitney tests. After correction of the resulting p_{MW}-

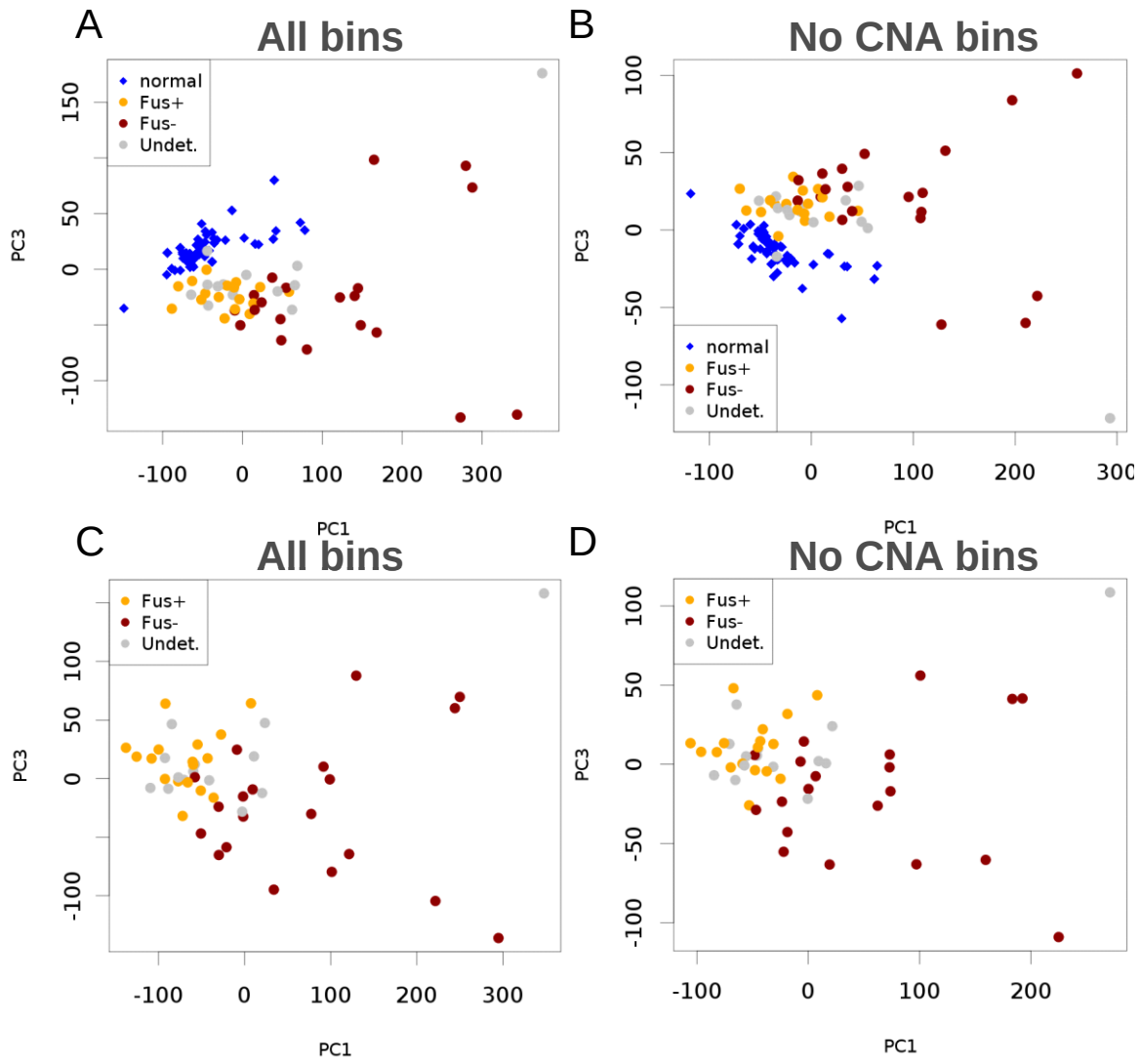


Figure 19. Principal component analyses of methylated regions. (**A**) shows the results of the principal component analyses (PCA) using the MeDIP-Seq values of 682,510 methylated bins of all samples. Here, principal component 1 plotted against principal component 3 allows a separation of normal (blue) and tumour (brown=FUS-, orange=FUS+, gray=undetermined) samples. (**B**) Using only bins unaffected by copy number alterations (CNA) also allowed separation of the cohorts. PCAs of the tumour cohort only revealed inherent separation of FUS+ from FUS- samples by using all methylated bins (**C**) or only bins outside of CNAs (**D**). (Courtesy of Axel Fischer).

values for multiple testing we identified 85,406 regions that were significantly hyper- and 61,308 regions that were significantly hypomethylated in prostate cancer (Benjamini-Hochberg corrected Mann-Whitney-test $p_{BH}<0.05$) (**Figure 18E, Figure 20A**).

When looking at the chromosome-wise distributions of DMRs we found that the chromosomes X and Y exhibited a low density of DMRs while the chromosomes 8, 16, and 19 showed an elevated density of DMRs (**Figure 20A**). The 110 most significantly differentially methylated regions are depicted in **Supplementary Table 3.** As an example the well known hypermethylated region in *GSTP1* (HOPKINS 2007, NAKAYAMA 2003) is visualised in **Figure 20B.**

Hypermethylated regions in the tumour samples were more homogeneous than hypomethylated regions as concluded from the lower p-values of hypermethylated regions (**Table 27**). This might indicate a specific and site-directed methylation process as compared to a more unspecific global loss of methylation in cancer.

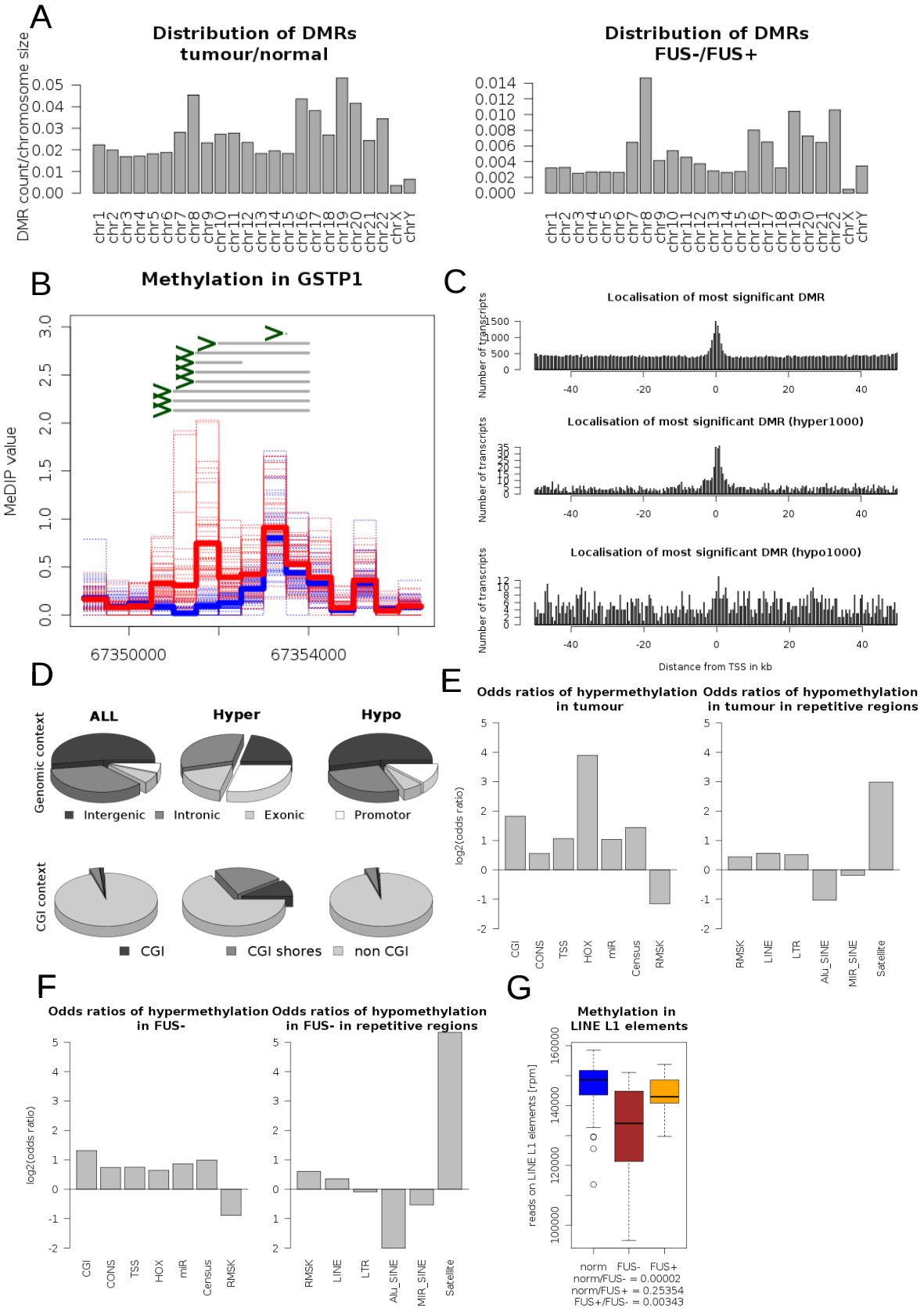
Table 27: Counts of hyper- and hypomethylated regions in tumour below a given p_{BH}-value cutoff.

p_{BH}-value	Counts of hypermethylated bins	Counts of hypomethylated bins
5,E-02	85406	61308
1,E-02	48794	36414
1,E-03	28457	17625
1,E-04	19535	8284
1,E-05	14224	3597
1,E-06	10652	1535
1,E-07	8004	581

We analysed the surrounding regions of the transcription start sites (TSS) of all annotated transcripts (Ensembl59) and found an enrichment of differentially methylated regions within ±2kb around the TSS (IRIZARRY 2009, JONES 1999). Next, we analysed the localisations of the most significant DMRs relative to the transcription start sites: We restricted the analysis to the 1,000 transcripts with their TSS within an area of ±50kb around the most significantly hyper- and hypomethylated bins, respectively. Indeed, hypermethylations were concentrated at a region ±2kb around the TSS while hypomethylation did not show a regional preference among these transcripts (**Figure 20C**). Thus, for further analyses we annotated bins within ±2kb of a transcription start site as 'promoter bins'.

Next, we annotated each bin to investigate the enrichment of specific features among the hyper- and hypomethylated regions (see section 2.3.4). While most hypomethylated bins were detected in intergenic regions (63%), only a quarter of the hypermethylated bins was located outside of genes and promoters. Hypermethylations were highly enriched in CpG islands and CpG island shores (**Figure 20D**).

Additionally, we detected that hypermethylations were enriched within conserved, homeobox gene (HOLLAND 2007), cancer gene (FUTREAL 2004), and microRNA regions, whereas hypomethylations of these sites were less frequent than expected. Hypomethylations on the other hand were predominantly found within repetitive regions (DASKALOS 2009, FEBER 2011, GAUDET 2003), mainly within long interspersed elements (LINEs), long terminal repeats (LTRs), and satellite regions. Only Alu-elements showed fewer hypomethylations than expected (**Figure 20E**, **Table 28**).


A
Distribution of DMRs tumour/normal
DMR count/chromosome size
Distribution of DMRs FUS-/FUS+
B
Methylation in GSTP1
MeDIP value
67350000
67354000
C
Localisation of most significant DMR
Localisation of most significant DMR (hyper1000)
Localisation of most significant DMR (hypo1000)
Number of transcripts
Distance from TSS in kb
D
ALL
Hyper
Hypo
Genomic context
Intergenic
Intronic
Exonic
Promotor
CGI context
CGI
CGI shores
non CGI
E
Odds ratios of hypermethylation in tumour
Odds ratios of hypomethylation in tumour in repetitive regions
log2(odds ratio)
CGI
CONS
TSS
HOX
miR
Census
RMSK
LINE
LTR
Alu_SINE
MIR_SINE
Satellite
F
Odds ratios of hypermethylation in FUS-
Odds ratios of hypomethylation in FUS- in repetitive regions
G
Methylation in LINE L1 elements
reads on LINE L1 elements [rpm]
norm
FUS-
FUS+
norm/FUS- = 0.00002
norm/FUS+ = 0.25354
FUS+/FUS- = 0.00343

Since we have found significant differences in global methylation between FUS- and FUS+ tumours (**Figure 18C**, **Figure 19**) we were also interested in the identification of site specific differential methylations between these tumour subgroups. Using bin-wise Mann-Whitney testing with Benjamini-Hochberg correction of p_{MW}-values of the 17 FUS+ and 20 FUS- samples we detected significant differential methylation in 27,488 regions (17,971 bins hyper-, 9,517 bins hypomethylated in FUS-; $p_{BH}<0.05$). The 110 most significantly differentially methylated regions are shown in **Supplementary Table 4**.

Here, again chromosome 8 bears a comparably high density of DMRs, and again chromosome X is only slightly differentially methylated. Chromosome Y on the other hand contains an average density of DMRs in contrast to the comparisons of tumour and normal samples (**Figure 20A**).

The differentially methylated regions could be assigned to the TSS of 4,053 unique genes. We again calculated which regions were most affected by differential methylations. We determined an enrichment of FUS- sample hypermethylations in conserved regions, cancer census genes (FUTREAL 2004), homeobox (HOLLAND 2007) and microRNA genes (**Figure 20F, Table 29**).

Figure 20. Differential methylations in prostate cancer. (**A**) Chromosome-wise densities of DMRs comparing tumour and normal (left) and FUS- and FUS+ samples (right). The chromosome-wise DMR counts were normalised to the chromosome size (number of total bins on chromosome). (**B**) Genomic region plot of the vicinity of the *GSTP1* gene. Depicted are the bin-wise MeDIP-Seq values of all prostate tumour (red dashed lines) and normal prostate samples (blue dashed lines) in this region. Solid blue and red lines represent the bin-wise means of normal and tumour MeDIP-Seq values (rpm), respectively. The gray lines represent all *GSTP1* transcripts according to Ensembl59, the TSS are marked by green chevrons. (**C**) Histograms of the localisations of the most significantly differentially methylated regions in reference to the TSS of all transcripts (upper diagram), of the 1,000 transcripts in the ±50kB vicinity of the best hypermethylated bins (middle) and the 1,000 transcripts associated with the best hypomethylated bins. (**D**) Regional distribution of the differentially methylated regions in tumour samples. Top: Intergenic, intronic, exonic and promoter regions, bottom: CpG-islands (CGI), CGI shores and non-CGI regions. ALL: reference distribution in the human genome (hg19), Hyper: regions with significant hypermethylations; Hypo: regions with significant hypomethylations in tumour samples. (**E**) Enrichments of hypermethylated regions in genomic areas (left) and enrichment of hypomethylated regions in repetitive genomic regions (right) in tumours. Depicted are $\log_2$(odds ratios) of regions associated with CpG islands and shores (CGI), conserved regions (CONS), promoter regions (TSS: 2kb up- and downstream of TSS), homeobox genes (HOX), miRNAs (miR: 2kb up- and downstream of miR), cancer census genes (Census), repeat regions (RMSK), long interspersed repetitive elements (LINE), long terminal repeats (LTR), Alu elements (Alu_SINE), mammalian interspersed elements (MIR_SINE), and satellite DNA (Satellite). (**F**) Enrichments of hypermethylated regions in genomic areas in FUS- compared to FUS+ (left) and enrichment of hypomethylated regions in repetitive genomic regions in FUS- (right). Abbreviations similar to (**E**). (**G**) Methylation in LINE L1 repetitive elements in normal, FUS- and FUS+ samples. Shown are the sums of all MeDIP-Seq values in LINE L1 bins as boxplots of all samples belonging to each cohort. Mann-Whitney p-values of the significance of the differences are given below.

Table 28: Differential methylations between tumour and normal tissues are found in specific genomic contexts. Total numbers of bins located on specific genomic regions are indicated in column (**B**). Significantly methylated bins are given in columns (**C**) and split into hypermethylated (**D**) and hypomethylated (**E**) bins. Differentially methylated bins (p_{BH}<0.05) between both tissue types are indicated in column (**F**), and distributions of hypermethylated and hypomethylated (in tumour vs. normal or FUS- vs. FUS+ samples) are indicated in columns (**H**) and (**J**), respectively. In addition, for each genomic feature odds ratios (OR) are calculated; for enrichments of DMRs (**G**), hypermethylated (**I**) and hypomethylated (**K**) regions, and the predominance of hypermethylated regions in all DMRs (**L**). For homeobox and cancer census OR calculations of gene names, the gene was counted as affected by the feature when it contained at least one significant bin within the TSS region.

Feature (A)	total number (B)	Methy-lated (C)	hyper (D)	hypo (E)	DMR (F)	OR DMR (G)	hyper DMR (H)	OR hyper DMR (I)	hypo DMR (J)	OR hypo DMR (K)	OR hyper/hypo DMR (L)
total	6191344	682510	420849	261661	146714		85406		61308		
Non repetitive	1566920	189798	131886	57912	50727	1,51	36897	2,21	13830	0,74	2,61
CGI + CGI shore	261291	76561	62479	14082	24535	1,87	22270	**3,53**	2265	0,28	9,19
Conserved	176060	55890	41834	14056	11764	0,97	9403	**1,47**	2361	0,42	3,09
TSS bins	595703	107661	80531	27130	26701	1,25	22126	**2,09**	4575	0,41	4,34
Gene names	44386	30818	24864	16022	14352		11433		3586		
Homeobox (HOX) bins	3283	1024	956	68	705	8,11	695	14,88	10	0,10	50,29
HOX gene names	231	211	203	40	177	6,03	175	**8,35**	6	0,22	9,27
Cancer census bins	11238	2293	1887	406	697	1,60	640	2,72	57	0,26	8,11
Cancer census gene names	440	369	348	173	224	1,78	210	**2,26**	33	0,74	2,01
miRNA bins	7951	1759	1299	460	501	1,46	398	**2,05**	103	0,63	2,78
Repeats	4624424	492712	288963	203749	95987	0,66	48509	0,45	47478	**1,36**	0,38
DNA	357921	36597	21375	15222	7340	0,91	3757	0,79	3583	1,11	0,74
LINE	1688553	146834	76775	70059	28836	0,87	11655	0,54	17181	**1,48**	0,41
LTR	719517	71187	38855	32332	14350	0,91	5855	0,60	8495	**1,43**	0,46
Alu SINE	932706	112974	77938	35036	16531	0,58	10868	0,71	5663	0,49	1,43
MIR SINE	458738	62240	40212	22028	13918	1,06	8877	1,18	5041	0,88	1,29
Satellite	26788	5687	893	4794	2551	3,01	106	0,13	2445	**7,92**	0,03

We also detected more hypomethylated regions in FUS- samples than in FUS+ with strong enrichments of satellite regions. For LINE L1 elements Kim *et al*. (KIM 2011) have shown significant hypomethylations in FUS- samples compared to FUS+ and normal samples, and indeed, we found significantly fewer methylated LINE L1 element associated bins in FUS- than in FUS+ or normal samples (**Figure 20G**). We investigated the CpG content dependency of differential methylation in order to describe tumour subgroup specific alterations in more detail. As mentioned before, after MeDIP we found the most significant enrichments in regions with 3-17 CpGs (in 200bp) while fragments with 0-1 CpGs were depleted. Interestingly, we determined significantly higher fractions of reads with more than three CpGs/200bp in FUS- samples than in FUS+ or normal samples (**Figure 17C**). This reflects the hypermethylation of high CpG promoter regions and CpG islands in FUS-

samples. The majority of significantly methylated regions in normal samples was found in bins with more than five CpGs (3rd tertile of CpG density). The increased amount of significantly methylated bins that we determined in FUS- compared to FUS+ and normal samples (**Figure 18C**) was also found in this last tertile, i.e. we found more bins with more than five CpGs significantly methylated in FUS- than in FUS+ and normal samples, while in the first tertile of the CpG distribution (0-2 CpG/500bp) FUS- samples showed fewer methylated bins (not significant) (**Figure 21A**).

Table 29: Differential methylations between FUS- and FUS+ samples are found in specific genomic contexts. Total numbers of bins located on specific genomic regions are indicated in column (**B**). Significantly methylated bins are given in columns (**C**) and split into hypermethylated (**D**) and hypomethylated (**E**) bins. Differentially methylated bins (p_{BH}<0.05) between both tissue types are indicated in column (**F**), and distributions of hypermethylated and hypomethylated (in tumour vs. normal or FUS- vs. FUS+ samples) are indicated in columns (**H**) and (**J**), respectively. In addition, for each genomic feature odds ratios (OR) are calculated; for enrichments of DMRs (**G**), hypermethylated (**I**) and hypomethylated (**K**) regions and the predominance of hypermethylated regions in all DMRs. For homeobox and cancer census OR calculations of gene names, the gene was counted as affected by the feature when it contained at least one significant bin within the TSS region.

Feature (A)	total number (B)	methy-lated (C)	hyper (D)	hypo (E)	DMR (F)	OR DMR (G)	hyper DMR (H)	OR hyper DMR (I)	hypo DMR (J)	OR hypo DMR (K)	OR hyper/hypo DMR (L)
total	6191344	656919	447534	209333	27488		17971		9517		
Non repetitive	1566920	185919	137772	48133	9457	1,35	7485	1,84	1972	0,66	2,73
CGI + CGI shore	261291	80411	69218	11193	4784	1,54	4515	**2,49**	269	0,21	11,54
Con-served	176060	55732	44995	10734	2660	1,16	2370	1,67	290	0,34	4,83
TSS bins	595703	110253	87700	22547	5154	1,15	4498	1,68	656	0,36	4,51
gene names	44386	30458	25551	14451	4053		3488		579		
HOX bins	3283	1120	936	183	55	1,18	47	1,56	8	0,49	3,12
HOX gene names	231	214	197	91	44	1,69	38	**1,68**	6	1,49	1,05
Cancer census bins	11238	2401	1977	424	137	1,39	127	1,99	10	0,28	6,77
Cancer census gene names	440	371	347	182	83	1,90	75	**1,98**	8	1,14	1,57
miRNA bins	7951	1807	1465	342	98	1,31	88	**1,82**	10	0,38	4,68
repeats	4624424	471000	309762	161200	18031	0,74	10486	0,54	7545	**1,52**	0,37
DNA	357921	34397	22282	12110	1285	0,88	798	0,84	487	0,98	0,86
LINE	1688553	134313	77940	56363	4625	0,78	2270	0,56	2355	1,28	0,44
LTR	719517	65302	42260	23035	2237	0,80	1344	0,73	893	0,94	0,78
Alu SINE	932706	117049	92404	24636	3750	0,72	3264	1,02	486	0,25	4,12
MIR SINE	458738	59828	42168	17657	2419	0,96	1799	1,11	620	0,69	1,60
Satellite	26788	5273	1147	4125	1751	12,09	20	0,13	1731	**40,41**	0,01

Since we have seen significant differences of LINE L1 methylation between tumour and normal as well as between FUS- and FUS+ samples we also performed CpG density stratification of these regions. We compared the methylation of low CpG, LINE L1 containing bins (first tertile) with high CpG, LINE L1 containing bins (third tertile)

(**Figure 21B**) and found that the exclusive reduction in L1 methylation in FUS- samples predominantly resulted from hypomethylation of low CpG content LINE L1 bins, while reduction of methylation of high CpG content LINE L1 bins seems to be a tumour associated feature found in both tumour subgroups and is more subtle than the reduction in low CpG LINE L1 associated bins.

In general, examining only differentially methylated regions between tumour and normal, we found that hypomethylated regions had lower CpG contents (peak at 5 CpGs/500bp) than hypermethylated regions (peak at 10 CpGs/500bp). Furthermore, hypermethylation of FUS- samples compared to FUS+ samples affected bins with even higher CpG content: We detected a broad peak between 10 and 13 CpGs/500bp (**Figure 21C**).

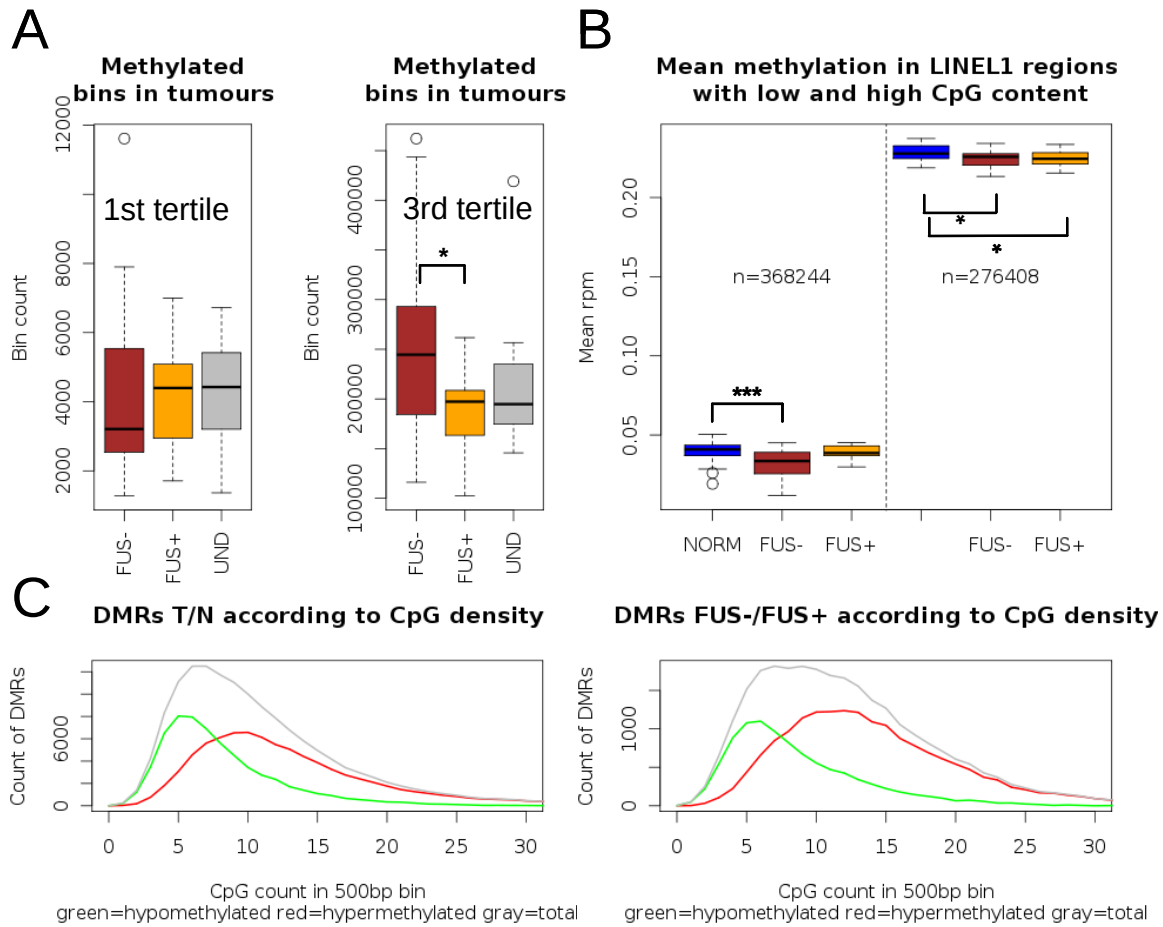


Figure 21. Stratifications of differential methylations according to CpG content. (A) Boxplots of sample-wise counts of significantly methylated bins containing 0-2 CpG (1st tertile of CpG distribution) or more than 5 CpGs (3rd tertile) in tumour samples. Undetermined tumour samples are depicted in gray (UND). (**B**) Boxplots of sample-wise mean rpm values in LINE L1 element containing bins with 0-1 CpG (1st tertile of LINE L1 elements, left side of plot) and mean rpm values of LINE L1 element containing bins with more than 3 CpGs (3rd tertile, right side of plot) (left plot). (**C**) Histograms of counts of differentially methylated regions between tumour and normal (left) and FUS- and FUS+ samples (right) according to CpG density (gray line). Hypermethylated regions are depicted in red, hypomethylated regions in green.
Significances of differences are marked with asterisks (p_{MW}<0.05 (*), <0.01 (**), <0.001 (***)).

3.7 Bisulphite mass spectrometry validation of MeDIP-Seq data

We used Sequenom EpiTYPER bisulphite mass spectrometry (BS-MS) analyses to validate our MeDIP-Seq results with an independent technology (RADPOUR 2008). BS-MS allows the determination of the CpG methylation level after BS conversion with single CpG resolution. We assessed the methylation levels of 48 heterogeneous regions including DMRs, promoters, miRNA genes, extragenic regions, and low and high CpG regions in three tumour and three normal samples (**Table 17**). As controls we used an *in vitro* methylated DNA and a water control (BS-MS1).

After adjusting the mean BS-MS values of a region for CpG density (see section 2.3.9.2), we determined a correlation of Rho=0.78 with the MeDIP-Seq values (**Figure 22A**), proving MeDIP-Seq to be a valid technology to read out the actual methylation. In the fully methylated DNA we found methylation values of 75-100% with a few outliers, suggesting an incomplete *in vitro* methylation. In comparison, in tumour tissues most CpG dinucleotides of hypermethylated regions displayed methylation values of 60-90% while in hypomethylated regions the CpG methylation values in tumour dropped to 10-30% (**Figure 22B**). The methylation within a region was homogeneous in each sample because the absolute variances showed values between 0 to 10% methylation with a few outliers with more than 20% absolute methylation variance (**Figure 22C**).

Next, to assess the consistency of BS-MS, we re-analysed two regions (chr14: 60,973,901-60,974,600 and chr6: 28,556,901-28,557,600) in one prostate tumour sample containing 17 and 20 CpG units with BS-MS data, respectively, and found a high correlation (Rho=0.69) of the CpG-wise methylation BS-MS values between the experiments (**Figure 22D**) with two outlier CpGs.

We performed an additional BS-MS analysis of 36 differentially methylated regions on microdissected and matched macrodissected tumour and adjacent benign prostate samples to exclude differential methylation being an artefact of different cellular compositions of tumour and normal samples (BS-MS2, **Table 18**). We yielded a correlation of 0.92 between the BS-MS values of the micro- and macrodissected tissues in these regions (**Figure 22E**).

Unsupervised cluster analyses of the BS-MS values could clearly separate tumour and normal samples (**Figure 22F**), showing that differential methylations are tumour cell specific.

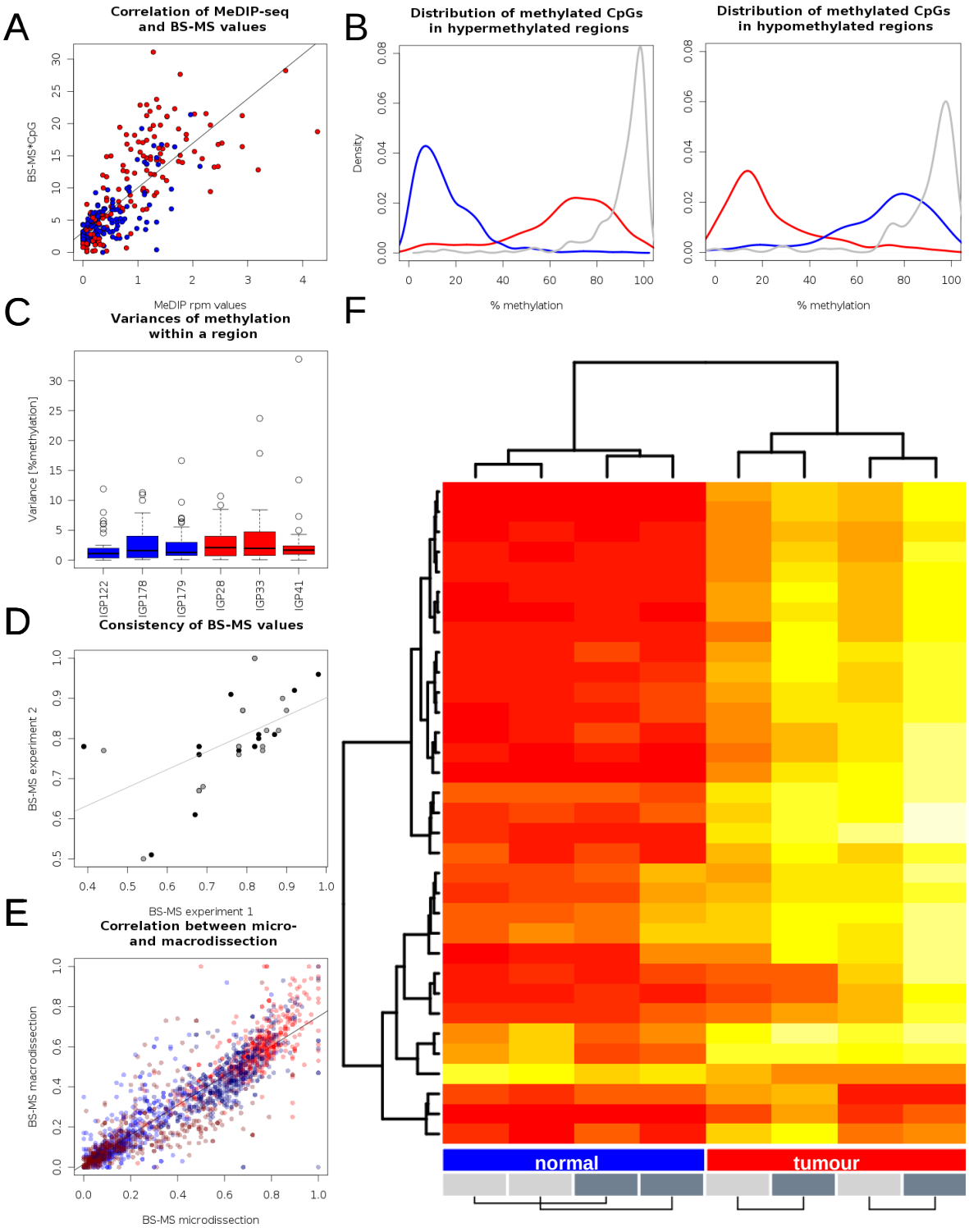


Figure 22. Validation of MeDIP-Seq data with bisulphite mass spectrometry analyses (BS-MS). (**A**) Correlation between MeDIP rpm values and adjusted BS-MS values of 79 regions analysed in six samples. Rho = 0.78. (**B**) Distributions of the methylation values of single CpGs in hypermethylated (left) and hypomethylated (right) regions showed bimodal patterns of methylation. Blue=normal, red=tumour, gray=fully methylated control. (**C**) For each region analysed the intraregional variance of the methylation values was calculated and depicted as boxplot for each sample. (**D**) Two regions (black: chr14:60973901-60974600; gray: chr6:28556901-28557600) were repeatedly analysed to assess BS-MS consistency. Rho=0.69 including the two outlier CpGs in the upper left region of the plot. (**E**) Correlation analyses of CpG-wise BS-MS methylation values between four microdissected and four matched macrodissected samples in 36 differentially methylated regions. Rho=0.92. Blue: normal, red: tumour samples. (**F**) Unsupervised cluster analyses using the mean BS-MS values of 35 DMRs showed a perfect separation of tumour (red) and normal samples (blue). Light gray=macrodissected, dark gray bar=microdissected samples. Samples originating from the same patient are connected by brackets.

3.8 Marker detection

Early detection of prostate cancer is crucial for choosing the best possible treatment. Li *et al.* have reviewed commonly described prostate cancer markers (LI 2004) (**Table 30**). After Benjamini-Hochberg correction for multiple testing and testing for sufficient read counts (see section 2.3.5) one marker (*CDH1*) was found not significantly differentially methylated while six marker regions did not contain enough reads to be above our threshold value (NA). The best region (*RARB*) ranks on position 36 in our MeDIP-Seq data set, implying that we have found 35 other more potent candidate markers.

Table 30: Markers described to be hypermethylated in PCa as reviewed in (LI 2004). Column headings are as follows: Gene name, Rank of the best gene associated DMR, chromosomal position of the DMR, p_{MW}-value of differential methylation between tumour and normal samples in the best DMR, BH corrected p_{MW}-values. NA indicates that methylation in this bin was not significant.

Gene	Rank of DMR	Position of DMR	p_{MW}	p_{BH}
RARB	36	chr3:25469501-25470000	4,4E-18	6,8E-14
GSTP1	108	chr11:67351501-67352000	1,2E-17	7,6E-14
ABCB1	405	chr7:87229501-87230000	1,1E-16	1,9E-13
DAB2IP	724	chr9:124463001-124463500	5,5E-16	5,2E-13
CDKN2A	869	chr9:21970501-21971000	9,1E-16	7,2E-13
PTGS2	1861	chr1:186650001-186650500	2,0E-14	7,6E-12
RARRES1	2283	chr3:158450501-158451000	5,7E-14	1,8E-11
APC	2535	chr5:112073501-112074000	9,8E-14	2,8E-11
LAMA3	4770	chr18:21270001-21270500	5,2E-12	8,1E-10
RASSF1	7826	chr3:50378501-50379000	1,7E-10	1,7E-08
CD44	7911	chr11:35160001-35160500	1,8E-10	1,8E-08
CAV1	10882	chr7:116164501-116165000	1,7E-09	NA
ESR1	11213	chr6:152130001-152130500	2,1E-09	1,6E-07
CCND2	12958	chr12:4380501-4381000	5,7E-09	NA
LAMB3	16339	chr1:209825501-209826000	2,5E-08	1,3E-06
HIC1	79488	chr17:1956501-1957000	5,3E-05	NA
MGMT	109170	chr10:131266001-131266500	1,7E-04	2,1E-03
ESR2	208841	chr14:64761001-64761500	1,3E-03	NA
AR	230754	chrX:66865001-66865500	1,7E-03	NA
LAMC2	393532	chr1:183214001-183214500	6,9E-03	NA
CDH1	762928	chr16:68835501-68836000	3,1E-02	1,1E-01

Further on, we evaluated the potential of the 110 most significantly differentially methylated regions to separate tumour from normal samples (**Supplementary Table 3**) (patent pending: MI4294).

The MeDIP-Seq values of single regions were not sufficient for a complete separation. Thus, we used linear combinations of two markers and calculated whether the smallest sum of the methylation values of the two evaluated bins in tumour samples was larger than the largest sum of the methylation values of the two evaluated bins in normal samples. 887 out of 5995 combinations of two markers strictly separated the sample cohorts (**Figure 23A**).

FUS+ and FUS- samples could not be separated by using a single differentially methylated region either. Combinations of two of the best 110 differentially methylated regions

between FUS- and FUS+ **(Supplementary Table 4)** yielded 96 pairs that could separate the tumour subgroups (**Figure 23B**).

We validated our results with BS-MS of five differentially methylated regions in 46 tumour and 48 normal prostate samples as well as *in vitro* methylated and not methylated SW480-DNA as controls (BS-MS3, **Table 19**). Again, we compared the BS-MS and MeDIP-Seq values of the investigated regions and determined correlations between 0.66 and 0.85 (**Figure 23C**).

Combinations of two of the three tumour specific markers *PTPRN2* (*protein tyrosine phosphatase, receptor type N polypeptide 2*), *TAC1* (*tachykinin precursor 1*), and *GHSR* (*growth hormone secretagogue receptor*) could clearly separate tumour and normal samples in unsupervised clustering using the MeDIP-Seq as well as the BS-MS values (**Figure 23D**). Linear combinations of these markers also allowed a clear separation. Furthermore, the average methylation in the *GHSR* region alone was sufficient to part tumour from normal samples in BS-MS analysis.

FUS+ and FUS- samples were as well separable using the BS-MS values of the top two marker regions 'TMP1' and 'TMP2' with three outliers: two FUS+ samples were classified as FUS- and one FUS- sample was classified as FUS+. Using the respective MeDIP-Seq values for clustering we found only a single outlier: one FUS- sample was classified as FUS+ (**Figure 23E**).

3.9 DNA methylation assays in urine DNA

Diagnosis of PCa through digital rectal examination strongly depends on the experience of the urologist. Furthermore, it has been shown that even unpalpable tumours as small as 0.2mm can disseminate and form metastases (SCHMIDT 2006). PSA as commonly used non-invasive marker of PCa has a low sensitivity and specificity (ECKERSBERGER 2009, KILPELAINEN 2010, SHARIAT 2011, THOMPSON 2005): Using a threshold of 4ng/ml the sensitivity was 20.5% (93.8% specificity), a threshold of 1ng/ml yielded a sensitivity of 83.4% (38.9% specificity) (THOMPSON 2005). Thus, a reliable tumour specific marker which might detect PCa in early stages is desirable. This implicates that the marker should be detectable in urine or blood, both rather easy to collect body fluids. As we have found hundreds of regions that are almost exclusively hypermethylated in PCa we developed methylation specific qPCR based assays (qMSP) to test the diagnostic potential in urine DNA.

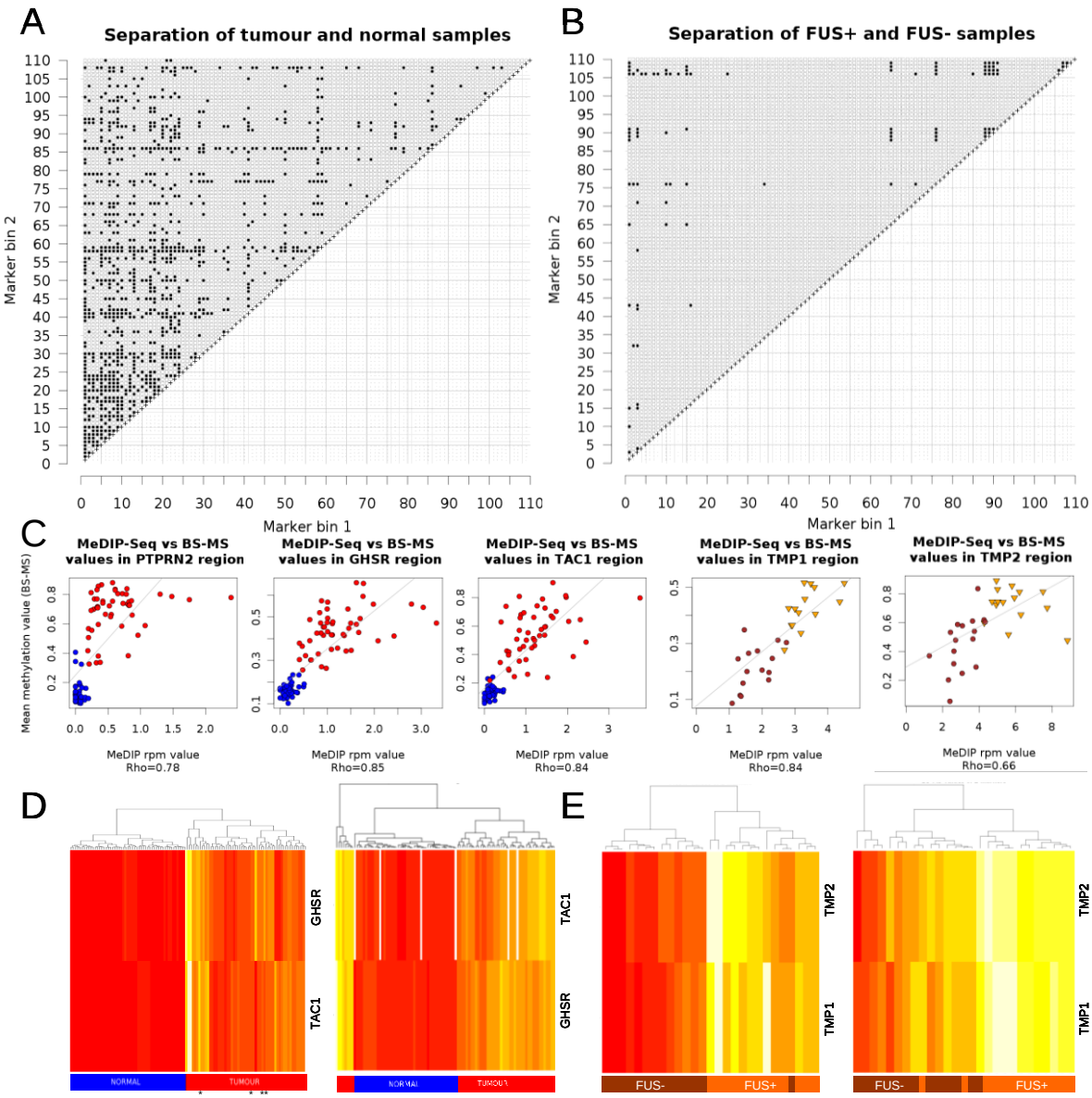


Figure 23. Detection of prostate tumour markers. (A) The 110 most significantly differentially methylated regions were used for analyses of the discriminatory potential of linear combinations of them. Marker pairs separating all tumour from all normal samples are coloured black. Out of 5995 possible pairs 887 allowed a perfect discrimination. (**B**) Analyses of the separation potential of the best 110 DMRs between FUS+ and FUS- samples yielded 96 combinations (black) that allowed a perfect separation of the two tumour classes. (**C**) Correlation analyses of the BS-MS and MeDIP-Seq values of the *PTPRN2*, *GHSR*, *TAC1*, "TMP1" and "TMP2" regions yielded Spearman's correlation coefficients of 0.78, 0.85, 0.84, 0.84, and 0.66, respectively. Tumour samples are depicted in red, normal samples in blue, FUS- samples are represented by brown circles, and FUS+ samples by orange triangles. (**D**) Unsupervised cluster analyses of the MeDIP-Seq (left) and BS-MS values (right) of the two marker regions *GHSR* and *TAC1* allowed perfect separation of tumour (red) and normal (blue) samples. The asterisks (*) mark the prostate cancer cell lines PC3, LNCaP, VCaP, and DU145, that were all correctly classified as "tumour" samples. (**E**) Unsupervised cluster analyses of the MeDIP-Seq (left) and BS-MS values (right) of the two regions "TMP1" and "TMP2" differentially methylated between FUS- (brown) and FUS+ (orange) samples.

The CpG-wise methylation values derived from BS-MS analyses allowed us to design quantitative methylation specific PCR (qMSP) assays for single CpGs with the highest discriminatory potential in the marker regions (**Table 25**, **Figure 10**). We generated three different assays covering CpGs in the vicinity of *ZSCAN12* (chr6:28,367,331-28,367,332 and chr6:28,367,348-28,367,349), *PTRPN2* (chr7:157,481,628-157,481,629) and *CAV2* (chr7:116,140,177-116,140,178).

To set up the reaction conditions and assess the specificity of each assay we used DNA amplicons containing these regions that were either methylated *in vitro* using *SssI* DNA methyltransferase or unmethylated prior to BS conversion. For each region we used two primer pairs: One that detected methylated DNA and one that detected unmethylated DNA. The primer specific for formerly methylated DNA of *ZSCAN12* had a false discovery rate (detection of methylated DNA in completely unmethylated DNA; FDR) of 0.1-0.6%, the FDR of the primer specific for unmethylated DNA was determined as 0.1%. *PTPRN2* primers had FDRs of 0.7% (methylated DNA primer) and 2-18% (not methylated DNA primer). The rather high values of the primer specific for unmethylated DNA may result from an incomplete *in vitro* methylation of this CpG rich region. FDRs of *CAV2* primers could not be assessed due to problems in generating methylated and unmethylated *CAV2* fragments.

The assays were used to probe three different sample sets of urine DNA. **Set 1** contained two circulating and five cellular urine DNA samples from five PCa patients and circulating and cellular DNA from urine of two female patients as negative controls and was used to assess whether circulating or cellular DNA was more suitable for methylation analyses. **Set 2** was obtained from the Charité in Berlin and contained nine non-tumour, ten tumour and one sample with unknown disease state. S**et 3** contained 20 tumour samples and 10 samples from patients with negative prostate biopsies.

Analyses of sample **set 1** with the *ZSCAN12*-primers revealed that the tested CpG was completely unmethylated in female samples. The maximum methylation value was 1.1% which is close to the FDR of the methylation primer. The minimum methylation value of the male samples was 1.3% (**Figure 24A**). One male sample did not show any results. Thus, male and female samples could be perfectly separated

The limited amount of isolated DNA (approximately 40ng prior to BS conversion) required an amplification step after the BS conversion using methylation independent primers. We used this amplified DNA for a repetition of the *ZSCAN12* assay and for the *PTPRN2* assay. Three male samples (one circulating DNA, two cellular DNA samples) could not be

amplified using the *ZSCAN12* primers. The maximum methylation value for female samples was 6.3% – the minimum methylation in male samples was 1% (**Figure 24B**). Thus, no clear separation of the two sample classes could be achieved anymore.
For the *PTPRN2* assay only one male sample of cellular DNA failed the amplification. The *PTPRN2* assay allowed a perfect separation of male cellular DNA samples from female samples (**Figure 24C**). Male circulating DNA did not show methylation in the *PTPRN2* region. These results suggested that DNA of exfoliated cells is more suitable for DNA methylation analyses in urine than circulating DNA. In further experiments only cellular DNA was prepared from urine.

For the next analyses we purified DNA from 20 exprimate urine samples obtained from the Charité Berlin (**set 2**). Since DNA amounts were very low (median 150ng, minimum 25ng before BS conversion) we amplified the BS converted DNA with methylation independent primer sets prior to methylation analyses. Three normal and two tumour samples could not be amplified in the *ZSCAN12* region. Furthermore, methylation of *ZSCAN12* alone was not sufficient to distinguish between non-tumour and tumour samples. Receiver operating characteristics (ROC) analyses revealed a minimum methylation cutoff of 3% for optimal ROC values (true positive rate (TPR): 71.4%; false positive rate (FPR): 42.9%) and of 5.5% for a minimal FPR (ROC_{pen}, FPR=0%, TPR=28.6%) (**Figure 24D**). The *PTPRN2* assay was more robust and yielded methylation values for all samples. With a methylation cutoff of 1.5% (optimal ROC value) we could correctly classify 90% of the tumour samples (TPR=90%) with four normal samples being falsely classified as tumour (FPR=44.4%). Using a cutoff value of 14% to minimise the FPR (optimal ROC_{pen} value) we correctly classified five tumour samples and misclassified two normal samples (TPR=50%, FPR=22%) (**Figure 24E**). To achieve a better performance we combined the two assays. ROC analyses revealed cutoff values of 1.5% minimum methylation in *PTPRN2* and 3.5% minimum methylation in *ZSCAN12* as optimal (**Figure 24F**). We yielded the same cutoffs calculating the ROC_{pen} scores for minimal FPR. Using these cutoffs we correctly classified four out of seven tumour samples that showed values in *PTPRN2* and *ZSCAN12* analysis (TPR=57.1%) without any false positive (**Figure 24G**). However, we still misclassified three tumour samples as normal.

Analyses of the same samples using the Methylight® *GSTP1* assay yielded methylation results for only one tumour sample (22% methylation). The other samples did not contain enough DNA for analysis.

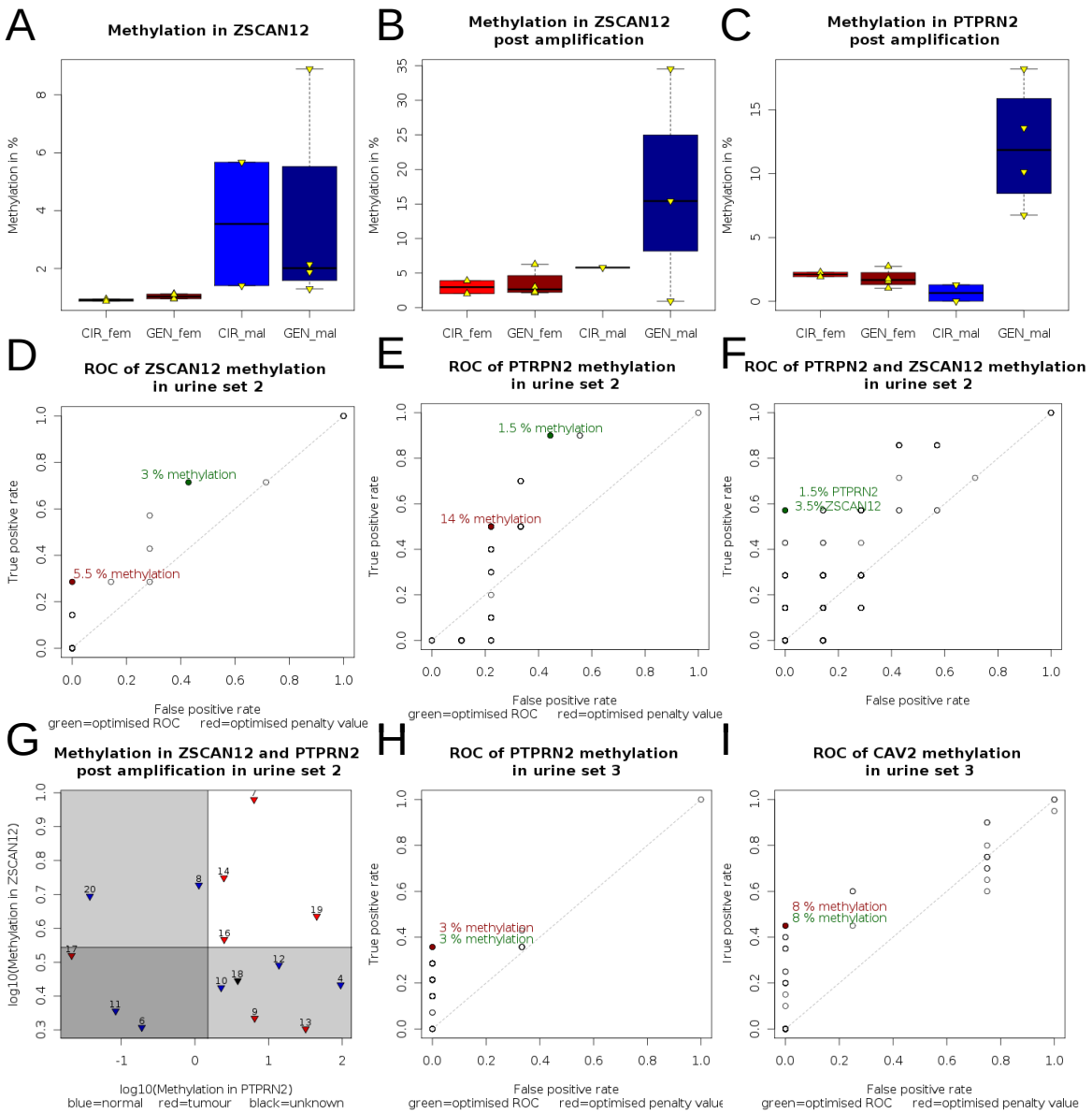


Figure 24. QMSP analyses of methylation in urinary DNA. (CIR=circulating/free DNA, GEN=DNA from cell pellets, fem=female samples, mal=male samples) (**A**) Boxplots of methylation in the four patient cohorts in *ZSCAN12* in urine set 1. (**B**) Boxplots of methylation in the four patient cohorts in *ZSCAN12* in urine set 1 after methylation unspecific amplification of the *ZSCAN12* region. (**C**) Boxplots of methylation in the four patient cohorts in *PTPRN2* in urine set 1 after methylation unspecific amplification of the *PTPRN2* region. (**D**) Receiver operating characteristics (ROC) curve of methylation in *ZSCAN12* after amplification in urine set 2. The fraction of normal samples falsely classified as tumour (abscissa) is plotted against the fraction of tumour samples correctly classified as tumour (ordinate) for every methylation cutoff value from 0 to 100% (in 0.5% steps). Optimal points are marked in green (optimised ROC value) and red (optimal ROC_{pen} value). (**E**) ROC curve of methylation in *PTPRN2* after amplification in urine set 2. (**F**) ROC curve of combined methylation in *ZSCAN12* and *PTPRN2* after amplification in urine set 2. The data points were obtained for every combination of methylation cutoff values from 0 to 100% (in 0.5% steps) in *ZSCAN12* and *PTPRN2*. (**G**) Plot of the methylation values of *PTPRN2* and *ZSCAN12* of the samples of urine set 2 including the cutoffs derived from ROC curve analyses. Samples in the upper right quarter are classified as "tumour" by the assay. Red triangles mark tumour, blue triangles represent normal samples. Numbers denominate sample IDs. (**H** and **I**) Individual ROC curves of methylation in *PTPRN2* and *CAV2* after amplification in urine set 3.

To further investigate the potential of the methylation assays, we obtained an additional set of 20 urine samples from male tumour patients and of 10 urine samples from men with negative biopsies collected at the Martini Clinics in Hamburg (**set 3**). Again, the amounts of isolated cellular DNA were low (minimum amount 50ng before BS conversion), so prior to assaying the samples in qMSP we performed 20 cycles of methylation unspecific PCR using the *PTPRN2* amplification primers. The qMSP analysis of *PTPRN2* failed for six tumour and seven normal samples, i.e. methylation C_t values were in the NTC range. With a methylation cutoff of 3% (optimal ROC and ROC_{pen} values) we could correctly classify five of the tumour samples without any normal sample being falsely classified as tumour (**Figure 24H**).

To further analyse the samples we set up a novel assay interrogating a CpG in the *CAV2* promoter. Due to problems in generating fully methylated and unmethylated *CAV2* fragments we could not assess the false positive rate of this assay. Again, we performed 20 cycles of methylation unspecific PCR using the *CAV2* amplification primers on BS converted urine DNA. All tumour samples were assayable by *CAV2* qMSP while six normal samples failed. Using a methylation cutoff value of 8% (ROC and ROC_{pen} optimisation) all normal and nine tumour samples were correctly classified (45% sensitivity, 100% specificity) (**Figure 24I**). A combination of the *CAV2* and *PTPRN2* assays revealed the same classification as the *CAV2* assay alone.

3.10 Functional analyses

Futreal *et al.* published a list of genes associated to cancer development – the 'cancer gene census' list (FUTREAL 2004). Interestingly, 224 of these 440 genes were significantly differentially methylated between tumour and normal samples (OR=1.8, **Table 28**), suggesting an influence of differential methylation of these genes on tumour development. Furthermore, comparing FUS+ and FUS- samples we found 83 of them to be differentially methylated (OR=1.9, **Supplementary Table 5, Table 29**).

To identify functional pathways most significantly affected by differential methylation, the top 1,000 hyper- and hypomethylated genes were subjected to DAVID (Database for Annotation, Visualization, and Integrated Discovery) analyses using standard parameters. Out of 1,000 genes hypermethylated in tumour DAVID could annotate 746; of the 1,000 hypomethylated genes 497 were found. Among the most significantly enriched features in

hypermethylated regions we determined genes of developmental, homeobox associated, and transcription regulatory pathways as top candidates (**Table 31**).

Table 31: DAVID pathway analyses of DMRs. The 1,000 most significantly differentially methylated unique genes were subjected to DAVID analyses and the most significantly enriched categories are shown for the analysis of hypermethylated and hypomethylated genes in tumour compared to normal, and the analyses of hypermethylated and hypomethylated genes in FUS- compared to FUS+ samples. Column headings are as follows: Term = enriched functional association term, Count = number of genes associated with this term, % = percentage of genes of input list associated to this term, P-value = p-value of enrichment of this term.

Term	Count	% of genes	P-value
Hypermethylated in tumour			
Homeobox	73	9,8	4,71E-46
GO:0043565~sequence-specific DNA binding	107	14,3	1,97E-41
developmental protein	118	15,8	2,56E-41
IPR012287:Homeodomain-related	69	9,2	5,08E-41
GO:0003700~transcription factor activity	134	18,0	3,89E-40
GO:0030528~transcription regulator activity	145	19,4	3,81E-26
Hypomethylated in tumour			
IPR000725:Olfactory receptor	33	6,6	1,17E-10
GO:0007606~sensory perception of chemical stimulus	35	7,0	6,32E-10
IPR017452:GPCR, rhodopsin-like superfamily	42	8,5	1,46E-09
topological domain:Extracellular	98	19,7	1,52E-09
sensory transduction	37	7,4	2,09E-09
PIRSF003152:G protein-coupled olfactory receptor, class II	29	5,8	8,23E-09
Hypermethylated in FUS-			
alternative splicing	334	45,6	2,70E-10
phosphoprotein	294	40,2	1,37E-04
GO:0002028~regulation of sodium ion transport	5	0,7	1,09E-03
GO:0008047~enzyme activator activity	24	3,3	1,25E-03
IPR011511:Variant SH3	9	1,2	1,87E-03
Hypomethylated in FUS-			
glycosylation site:N-linked (GlcNAc...)	102	23,5	5,77E-06
GO:0050877~neurological system process	42	9,7	1,72E-05
GO:0007214~gamma-aminobutyric acid signaling pathway	6	1,4	1,79E-05
sequence variant	230	53,0	3,86E-05
cell junction	19	4,4	1,03E-04

Since homeobox genes are among the top candidates we investigated this group in more detail. Out of a list of 231 homeobox genes (HOLLAND 2007) we identified 177 with a significant TSS associated differential methylation (OR=6, **Table 28**). Among the best 10,000 DMRs that are associated to 3,064 unique genes we found 124 genes of the homeobox list. Furthermore, among the best 1,000 bins which are associated to 546 unique genes we found 48 homeobox genes (OR=23.0). Thus, aberrant methylation of homeobox genes seems to be crucial in prostate tumour development (**Supplementary Table 6**).

The hypomethylated genes were enriched for G-protein coupled receptor (mainly olfactory) and signal transduction pathways. Out of the list of G-protein coupled receptors

(GPCR) containing 363 receptor genes (http://www.iuphar-db.org/DATABASE), 68 were significantly hypomethylated (BH<0.05). However, in 256 GPCR genes the most significantly differentially methylated region was hypermethylated (**Supplementary Table 7**). Nevertheless, 117 of these hypermethylated GPCRs contained also at least one significantly hypomethylated bin in their TSS vicinity.

Comparing the top 1,000 hyper- and hypomethylated genes in FUS- and FUS+ samples, we found hypermethylations (732 genes annotated) predominantly in genes affected by differential splicing events, while promoter associated hypomethylation (434 genes found) occurred mainly in signalling (GABA signalling) and extracellular protein genes (**Table 31**). Interestingly, the comparison of FUS+ and FUS- samples also yielded homeobox genes as an enriched group in differentially methylated regions: 44 out of 231 homeobox genes were significantly differentially methylated (OR=1.7, **Table 29**).

3.11 Integration of DNA methylation and gene expression data

To determine the influence of promoter methylation on gene expression we compared the MeDIP-Seq data with expression array data from the same samples we have used for methylation analyses. The array data comprises 17,163 unique genes and was generated in the group of Holger Sültmann (DKFZ).

3.11.1 Differential gene expression

Of these genes, 10,797 were significantly differentially expressed between tumour and normal samples (p_{BH}<0.05), with 5,839 of them upregulated. Since in the pathway analyses of differential methylation we had found homeobox, cancer gene census, and G-protein coupled receptor genes as most significantly affected pathways we were particularly interested in the gene expression data of 201 homeobox, 431 cancer census, and 345 GPCR genes. Interestingly, 123 of the homeobox genes (61%) showed a reduced expression in tumour; only 30 genes (15%) were upregulated. On the other hand, 234 of the cancer census genes (54%) were upregulated and 77 of them (18%) had a reduced expression in tumour. The cancer gene census list contains 91 potential tumour suppressor genes, for 87 of them expression data was available. Focussing on these genes we found a comparable situation: 53 of the tumour suppressor genes (61%) exhibited an elevated expression in tumour samples, and only 11 (13%) showed a reduced expression. Of 334 oncogenes 175

(52%) were increased and 65 (19%) downregulated. In the GPCR set 134 genes (39%) were downregulated, complemented by 39 upregulated genes (11%) (**Table 32**).

Table 32: Differential gene expression in prostate cancer (p_{BH}<0.05).

	Total	Upregulated in PCa	Downregulated in PCa
All genes	17,163	5,839 (34%)	5,038 (29%)
HOX genes	201	30 (15%)	123 (61%)
Cancer census genes	431	234 (54%)	77 (18%)
Tumour suppressor genes	87	53 (61%)	11 (13%)
Oncogenes	334	175 (52%)	65 (19%)
GPCR genes	345	39 (11%)	134 (39%)

Comparing FUS+ and FUS- samples, we found 1,598 genes differentially expressed. Of these, 409 had a higher expression in FUS- and 1,190 genes had a higher expression in FUS+ samples. Expression of 696 genes was exclusively altered in FUS+ and of 388 exclusively in FUS- samples compared to normal gene expression (**Table 33**). Interestingly, the majority of FUS+ specific genes was upregulated (514) while the majority of FUS- specific genes (258) was downregulated.

Next, we determined the tumour subgroup specific differential expression of genes: In FUS+ samples we found four homeobox (*HNF1B*, *IRX3*, *SATB1*, *SATB2*) and 14 cancer census genes (*CBLC*, *CDH1*, *CDH11*, *ERC1*, *ERG*, *ETV6*, *EXT2*, *GOPC*, *GPC3*, *SEPT9*, *RABEP1*, *SEPT6*, *SETD2*, *TFEB*) exclusively upregulated and four homeobox (*ARX*, *HOXC10*, *HOXC11*, *HOXC12)* and four cancer census genes (*ALDH2*, *CEBPA*, *HOXC11*, *HRAS*) exclusively downregulated compared to the normal gene expression. Interestingly, five of the upregulated (*CBLC*, *CDH1*, *EXT2*, *GPC3*, *SETD2*) and none of the downregulated census genes are putative tumour suppressor genes.

Table 33: Genes exclusively differentially expressed in one tumour subgroup compared to normal samples (p_{BH}<0.05).

	Upregulated only in FUS+	Downregulated only in FUS+	Upregulated only in FUS-	Downregulated only in FUS-
All genes	514	182	130	258
HOX genes	4	4	1	6
Cancer census genes	14	4	3	8
Oncogenes	9	4	3	7
Tumour suppressor genes	5	0	0	1

In FUS- samples we found a divergent picture: Here, one homeobox gene (*OTX1*) and three census genes (*CASC5*, *SLC45A3*, *TMPRSS2*) were upregulated while six homeobox (*HDX*, *ISL1*, *MEIS1*, *MEIS2*, *PRRX1*, *ZEB1*) and eight census genes (*FIP1L1*, *HLF*, *KDM6A*, *MYH11*, *NCOA4*, *PDGFRA*, *PRRX1*, *ZNF331*) were exclusively suppressed.

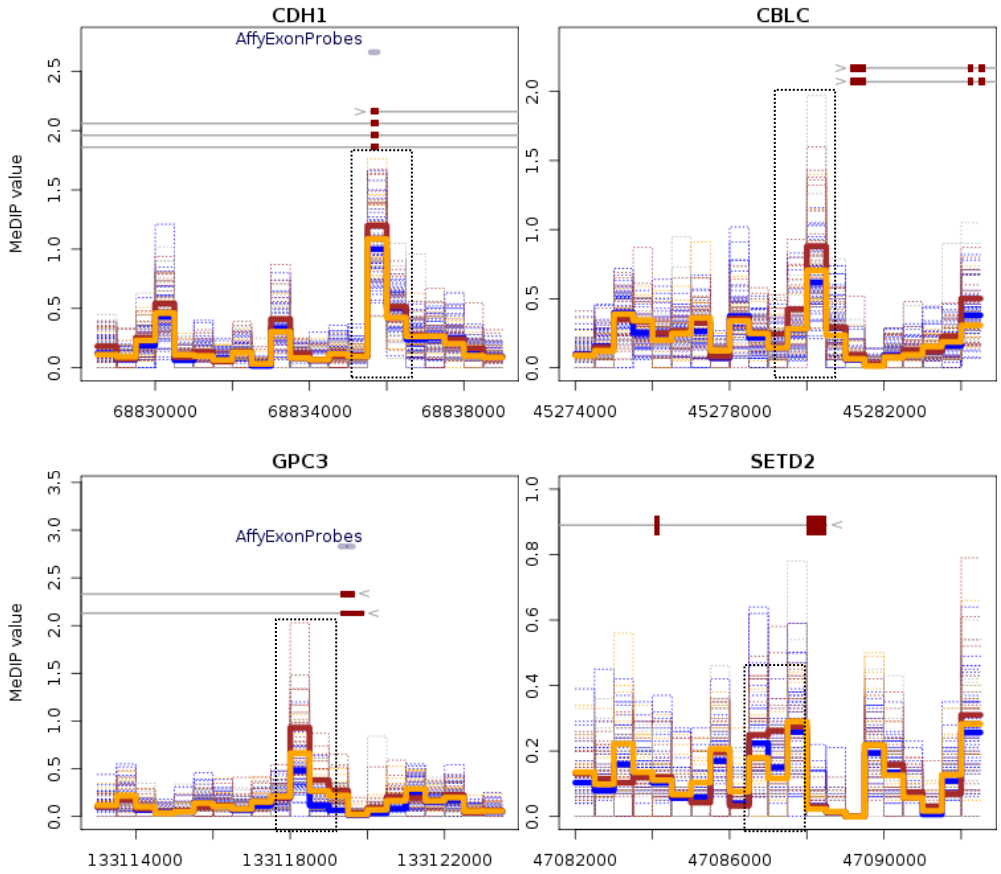


Figure 25. Methylation plots of the four tumour suppressor genes found upregulated in FUS+ samples and hypermethylated and not increased in FUS- samples. Shown are differentially methylated TSS associated bins and their vicinity along the chromosomal positions (x-axis). Blue lines mark normal, orange lines FUS+ and brown lines FUS- samples. Solid lines represent the mean values of the respective sample group. Black dotted rectangles enclose the areas of FUS- hypermethylation.

Here, we found one tumour suppressor gene (*KDM6A*) among the down- and no tumour suppressor gene among the upregulated census genes.

Interestingly, three of the five potential tumour suppressor genes found exclusively upregulated in FUS+ samples exhibited hypermethylations of the promoter region in FUS- but not in FUS+ samples: *CBLC* (chr19:45,280,001-45,280,500: $p_{MW\ FUS-/NORM}$=0.004, $p_{MW\ FUS+/NORM}$=0.33), *CDH1* (chr16:68,835,501-68,836,000: $p_{MW\ FUS-/NORM}$=0.012, $p_{MW\ FUS+/NORM}$=0.30), and *SETD2* (chr3:47,087,001-47,087,500: $p_{MW\ FUS-/NORM}$=0.0005, $p_{MW\ FUS+/NORM}$=0.18). *GPC3* was hypermethylated in both tumour subgroups, but hypermethylation in FUS- was significantly more pronounced (chrX:133,118,001-133,118,500: $p_{MW\ FUS-/NORM}=3.2*10^{-6}$, $p_{MW\ FUS+/NORM}$=0.006) (**Figure 25**).

These numbers mirror the observation of a general increased gene expression in FUS+ samples while in FUS- samples gene repression prevailed. Whether this suppression might be caused by an elevated methylation – a 'methylator' phenotype – as it is suggested by the elevated methylation levels in FUS- samples will be examined below.

3.11.2 Correlation of gene expression and MeDIP-Seq data

In the literature DNA methylation is associated with gene suppression. We thus correlated the normalised logarithmic expression data of each gene with the corresponding MeDIP-Seq values (rpm) of all promoter associated bins using Spearman correlations. We integrated 95 samples for which expression and methylation data were available in these analyses (48 normal and 47 tumour samples of which 17 were FUS+ and 20 were FUS-). We extracted the best correlating bin showing the lowest p-value for each gene. After correction of the p-values for multiple testing (Benjamini-Hochberg) we found the expression of 1,986 unique genes significantly correlated with methylation in at least one of their TSS associated bins (**Figure 26A**). For 839 of them the best correlation was positive: we found elevated expression and elevated methylation and *vice versa*, for 1,143 the correlation was negative. Since the gene expression data set contained multiple expression values for some genes (multiple array cluster IDs/gene) we found four genes with significant positive correlations as well as significant negative correlations of corresponding array cluster IDs.

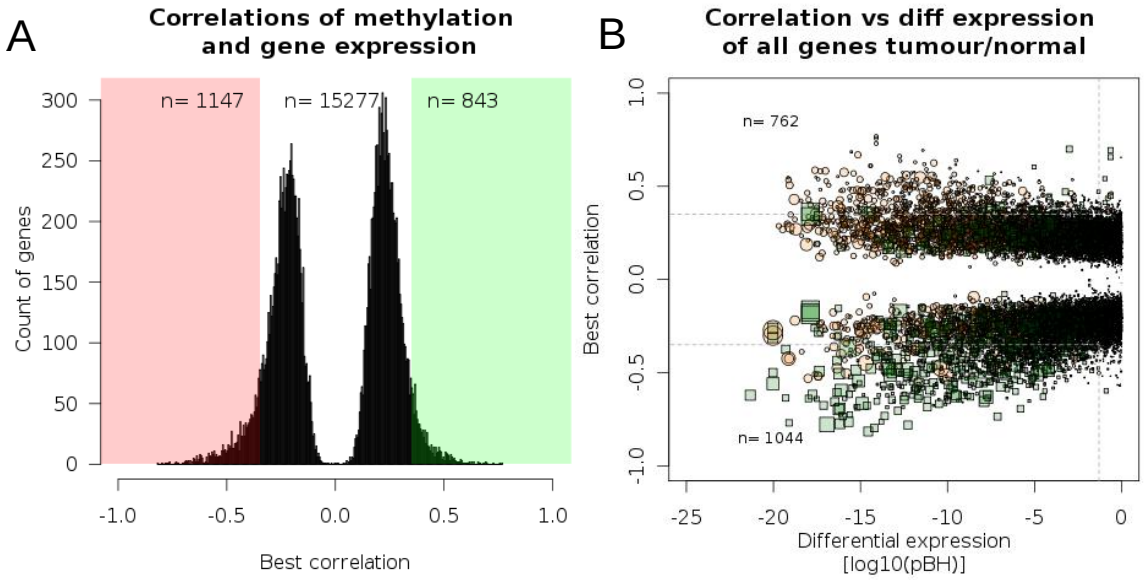


Figure 26. Correlations between gene expression and methylation. (**A**) Histogram of the correlations between DNA methylation and gene expression of all genes. Genes in red and green areas exhibit significant correlations of DNA methylation and gene expression after correction of the p-values for multiple testing (p_{BH}<0.05). "n" denominates the number of genes in the respective region. (**B**) Scatter plot integrating the significance of differential expression (abscissa) and the correlation of the methylation values to the expression of the respective gene (ordinate). Dotted lines enclose areas of significance (p_{BH}<0.05 for differential expression of the gene and p_{BH}<0.05 for best correlation with methylation). The size of the symbols is equivalent to the expression difference, the larger the symbols the larger the difference. Green squares mark downregulated, red circles upregulated genes. Green squares in the lower left significance area represent hypermethylated and suppressed, red circles in this area hypomethylated and upregulated genes. The numbers "n" denominate the counts of unique genes in the respective significance area.

Next, we determined whether significantly correlating genes were additionally significantly differentially expressed between tumour and normal samples. Integrating significance values of differential expression of each gene (p_{BH}<0.05) 762 unique genes were significantly positively correlated and 1,044 genes significantly negatively correlated (**Figure 26B**). These numbers substantiate the notion of DNA methylation mediated gene suppression.

We investigated the differentially expressed genes in more detail: 5,038 were suppressed and 5,839 were upregulated in tumour. With 953 genes (18.9%) with a significant correlation of expression and methylation (p_{BH}<0.05, Rho<-0.35 or Rho>0.35), we found a slightly higher fraction of genes in the suppressed than in the upregulated genes (n=853, 14.6%). Interestingly, 92% (878) of the significantly correlating and suppressed genes exhibited a negative correlation, i.e. they were downregulated and hypermethylated. In the upregulated genes, on the other hand, only 19.5% (166) of the correlated genes were negatively correlated (**Figure 27A**). These findings suggest that the majority of genes is presumably not deregulated by aberrant DNA methylation, provided a negative connection between DNA methylation and gene expression.

Next, we stratified the genes according to their expression values in the normal samples into normally highly and lowly expressed genes (first quartile: expression values below 5.68, fourth quartile: expression values above 7.45) to investigate potential differences in methylation dependent regulation of these groups (**Figure 27B**). In general, normally lowly expressed genes were suppressed (1,944 suppressed, 770 upregulated), while highly expressed genes were upregulated in tumour (1,943 increased, 821 suppressed). The total amount of negatively correlating genes was comparable in both quartiles (239 and 265, respectively) (**Figure 27B left**), while in the positively correlating fraction we found for highly expressed genes twice the number of genes as in the lowly expressed fraction (97 genes in 1st, 223 genes in 4th quartile). This difference was mainly caused by genes that were upregulated despite of a hypermethylation (68 genes in the 1st and 206 genes in the 4th quartile) (**Figure 27B right**). Although the fractions of positively correlated genes in the upregulated genes were equal in both quartiles (10%), the fraction of negatively correlated genes was higher in highly expressed genes (4.2% in 4th, 1.8% in 1st quartile). The suppressed genes on the other hand were dominated by negative correlations with DNA methylation in both quartiles and showed a twice as large fraction of negative correlations in the 4th quartile (22.3% vs. 11.6%).

There are fewer highly than lowly expressed genes downregulated in tumour but a DNA methylation mediated downregulation seems to be more likely in genes of the 4th quartile (**Figure 27B** middle).

Taken together, lowly expressed genes are mainly downregulated and highly expressed genes are mainly upregulated in tumour. Assuming a negative connection between DNA methylation and gene expression the highly expressed genes contain a larger subgroup of

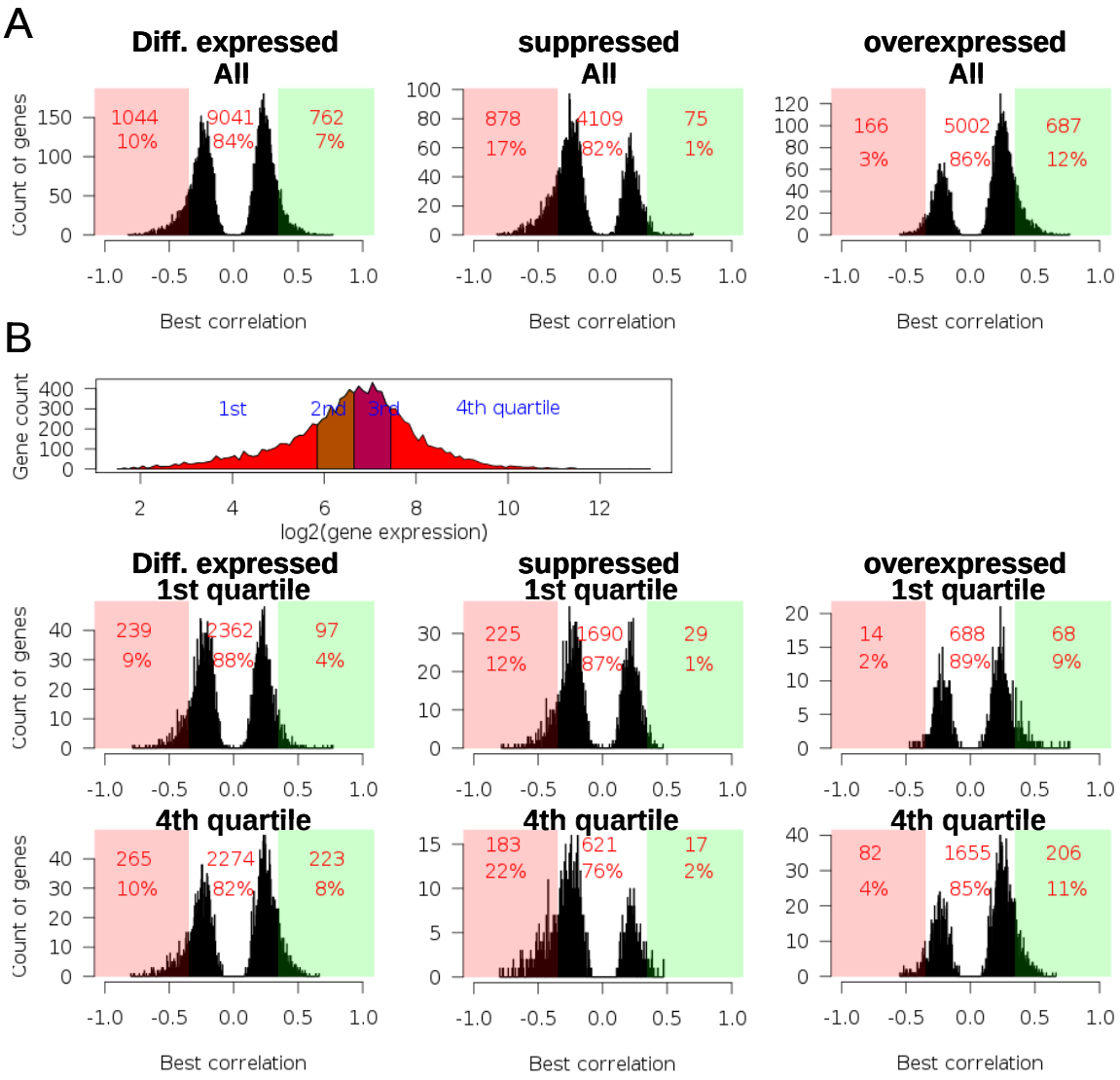


Figure 27. Stratification of differentially expressed genes (p_{BH}<0.05) according to their expression values in normal tissues and their expression changes in tumour tissues. (**A**) Histograms of the correlations between DNA methylation and gene expression in all differentially expressed genes. The left histogram gives information on the correlation values of all differentially expressed, the middle histogram on all suppressed, and the right histogram on all upregulated genes. (**B**) Top: Histogram of normal gene expression in differentially expressed genes. Below: Stratification of differentially expressed genes into normally lowly expressed genes (upper row, 1st quartile of expression) and normally highly expressed genes (lower row, 4th quartile of expression). The left histograms give information on the correlation values of all differentially expressed, the middle histogram on all suppressed, and the right histogram on all overexpressed genes. Genes in red and green areas exhibit significant correlations of DNA methylation and expression after correction of the p-values of the correlations for multiple testing (p_{BH}<0.05), "n" denominates the number of genes in the respective region.

genes following this model than the lowly expressed genes. Here, highly expressed and suppressed genes exhibit the highest fraction of genes where DNA methylation is likely to cause the suppression (negative correlation).

An example illustrating this finding is *caveolin 2* (*CAV2*), a significantly differentially methylated tumour marker. *CAV2* which is highly expressed in normal tissues exhibited a highly significant downregulation ($p_{BH}<3*10^{-14}$) and high negative correlation (Rho=-0.73) between methylation and expression in tumours. The same holds true for *DUOX1* (*dual oxidase 1*) ($p_{BH}<4*10^{-14}$, Rho=-0.69) (**Figure 28**). *In vitro* methylation luciferase assays using the *DUOX1* region are described below and demonstrate the high probability of aberrant DNA methylation of this region influencing gene expression in tumours. *GSTP1* that has been described to be hypermethylated and downregulated in PCa also showed a strong negative correlation between promoter methylation and expression and was indeed significantly downregulated in our samples ($pBH<7*10^{-15}$, Rho=-0.8).

Aldehyde oxidase 1 (*AOX1*), an enzyme involved in the biosynthesis of retinoic acid from retinal (HUANG 1994, SIGRUENER 2007), is another key enzyme found among the genes with the highest negative correlation between gene expression and DNA methylation ($pBH<2*10^{-17}$, Rho=-0.78) (**Figure 28**).

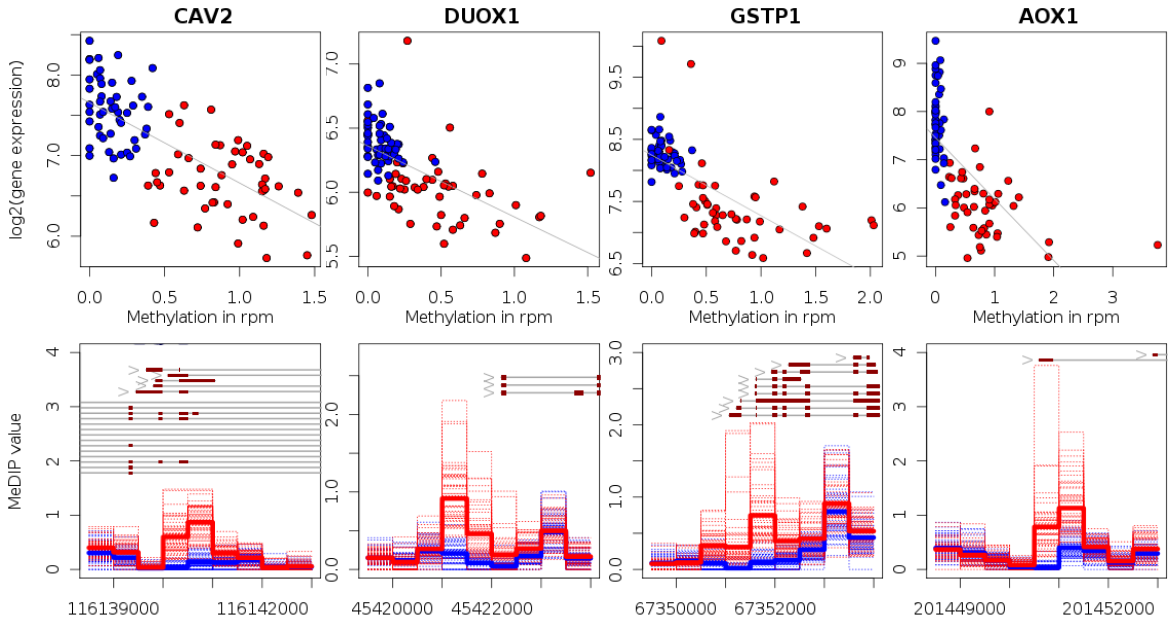


Figure 28. Correlations between gene expression and methylation of selected genes.
Correlation plots of the DNA methylation values of the best correlating bin and the gene expression of *CAV2*, *DUOX1*, *GSTP1*, and *AOX1* show a highly significant correlation for these genes (Rho_{CAV2} =-0.73, Rho_{DUOX1}=-0.69, Rho_{GSTP1}=-0.80, Rho_{AOX1}=-0,78) (top) and genomic region plots of the differentially methylated promoter regions (bottom). Depicted are the bin-wise MeDIP-Seq values for the indicated chromosomal positions (abscissa) of all prostate tumour (red dashed lines) and normal prostate samples (blue dashed lines) in this region. Solid blue and red lines represent the bin-wise means of normal and tumour MeDIP-Seq values, respectively. The gray lines represent all gene transcripts according to Ensembl59, the TSS are marked by chevrons.

Looking at gene groups instead of single genes it became obvious that the DNA-methylation-expression axis is differently employed between the different gene groups. Comparing the correlation patterns of all genes to those of the homeobox genes we could show that homeobox genes have a higher fraction of significantly correlating genes (OR=3.4, $p=5.7*10^{-13}$) and furthermore a significantly higher fraction of genes with negative correlation (OR=5.1, $p=1.0*10^{-6}$), suggesting methylation as predominant mechanism of homeobox gene regulation (**Figure 29A,B; Table 34**).

In cancer census genes we found a less significant enrichment of correlating genes than in homeobox genes (OR=1.5, p=0.006) and less negative correlations than in the whole gene set (OR=0.56, p=0.06) (**Figure 29A,C; Table 34**).

G-protein coupled receptor genes showed marginal less correlation than the complete gene set (OR=0.81, p=0.27) but a slightly higher fraction of negatively correlating genes (OR=1.98, p=0.11), also suggesting methylation dependent gene regulation (**Figure 29A,D; Table 34**).

Table 34: Correlation of gene expression and DNA methylation in different gene groups. Given are the numbers of genes with significant positive or negative correlations ($p_{BH}<0.05$), and of genes showing no correlation. The odds ratios (OR) show enrichments of correlation or enrichment of negative correlation in the respective gene group.

gene class	negative correlation	positive correlation	no correlation	OR correlation/no correlation	OR negative correlation/positive correlation
All genes	1147	843	15277		
HOX genes	53	8	141	3,4	5,1
Cancer census genes	32	37	365	1,5	0,6
GPCR genes	24	9	312	0,8	2,0

GPCR pathway associated genes were enriched among the hypomethylated genes as determined by DAVID pathway analysis (see section 3.10.). Despite the domination of hypermethylation and gene suppression in this class of genes, the olfactory receptor *OR51E2* is an example of a hypomethylated and overexpressed gene: On average it showed a 5fold higher expression in tumour ($p_{BH}<4.2*10^{-11}$) and a correlation between gene expression and methylation of Rho=-0.47 (**Figure 30A**). The *OR51E2* receptor can be activated by 6-dehydrotestosterone (6-DHT) (NEUHAUS 2009), a testosterone metabolite synthesised by cytochrome oxidases of the CYP3A class in the liver (HALVORSON 1990). *CYP3A4, 5, 7,* and *43* were downregulated in tumour and furthermore *CYP3A5* and *CYP3A7* were hypermethylated. For these two we found significant correlations between gene expression and methylation ((Rho_{CYP3A5}=-0.42, $p_{BH\ CYP3A5}<0.005$; Rho_{CYP3A7}=-0.36, $p_{BH\ CYP3A7}<0.05$) (**Figure 30B**).

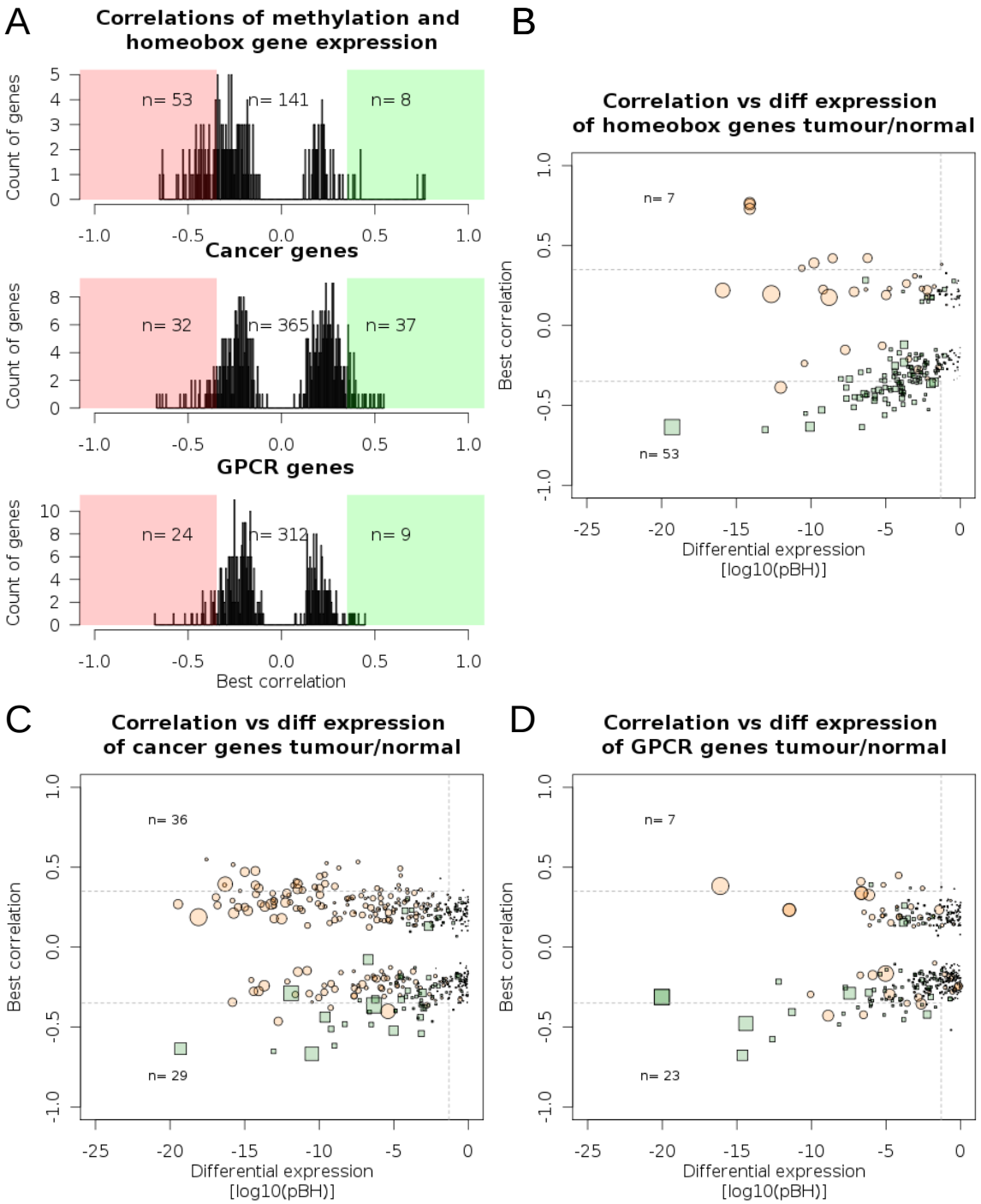


Figure 29. Correlations between gene expression and methylation in PCa associated gene sets.
(**A**) Histogram of the most significant correlations between DNA methylation and gene expression of homeobox genes (top), cancer census genes (middle), and GPCR genes (bottom). Genes in red and green areas exhibit significant correlations of DNA methylation and expression after correction of the p-values of the correlations for multiple testing (p_{BH}<0.05), "n" denominates the number of genes in the respective region. (**B**-**D**) Scatter plots integrating the significance of differential expression of each gene (abscissa) and the correlation of the methylation values to the expression of the respective gene (ordinate) in homeobox (**B**), cancer census (**C**), and GPCR genes (**D**). Dotted lines enclose areas of significance (p_{BH}<0.05 for differential expression of the gene and p_{BH}<0.05 for best correlation with methylation). The size of the symbols is equivalent to the expression difference, the larger the symbols the larger the difference. Green squares mark downregulated, red circles upregulated genes. Green squares in the lower left significance area represent hypermethylated and suppressed, red circles in this area hypomethylated and overexpressed genes. The numbers "n" denominate the gene count in the respective significance area.

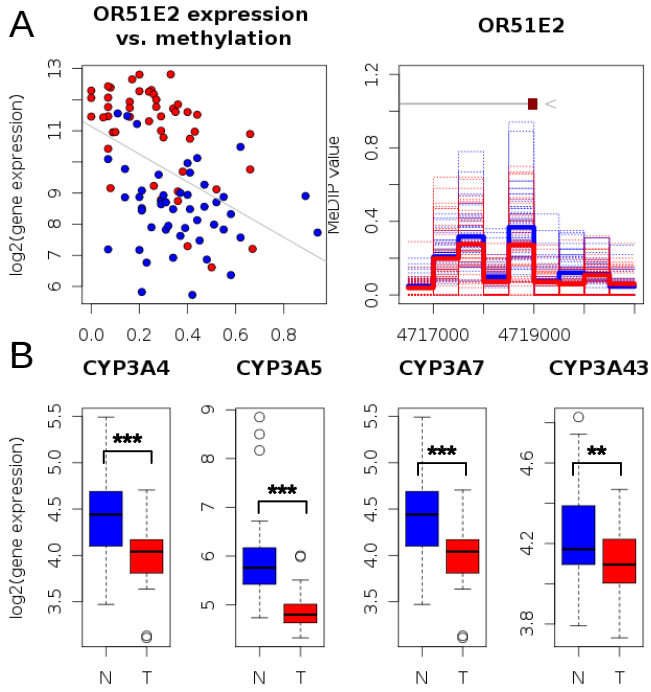


Figure 30. Methylation and expression of *OR51E2* and CYP3A genes. (**A**) Correlation plot of promoter methylation and expression of *OR51E2* (left). Tumour (red) and normal samples (blue) are highly differentially expressed and differentially methylated. The right methylation plot of the upregulated gene *OR51E2* reveals hypomethylation of promoter associated bins in tumour. Blue lines mark normal, red lines tumour samples. Solid lines represent the mean values of the respective sample group. (**B**) Boxplots of the expressions of *CYP3A4*, *5*, *7*, and *43* reveal downregulation of the CYP3A genes in tumour. Asterisks mark significance of differential expression: $^{**}p_{BH}<0.01$, $^{***}p_{BH}<0.001$.

3.11.3 Correlation of miRNA expression and DNA methylation data

Besides hypermethylation of gene promoter regions we also determined significant alterations in the methylation patterns of miRNA genes implying tumour relevant functional consequences on miRNA expression (**Figure 20E**). Thus, we assessed miRNA expression differences between the sample groups. Expression of 483 different miRNAs had been determined by Jan Brase in the group of Prof. Holger Sültmann (DKFZ) investigating the same samples that we have screened in our methylation study ($n_{tumour}=51$, $n_{normal}=48$). We used this data and found a strong deregulation in prostate tumours with 318 miRNAs differentially expressed ($p_{BH}<0.05$), 127 were up- and 191 were downregulated in tumour. Between FUS- and normal samples we found 284 (120 up- and 164 downregulated), and between FUS+ and normal samples we found 236 differentially expressed miRNAs (97 up- and 139 downregulated). Of these, 98 miRNAs were exclusively differentially expressed in FUS- (58 down- and 40 upregulated) and 50 exclusively in FUS+ (33 down-, 17 upregulated), i.e. expression was only altered in one tumour subgroup – the other showed normal expression values. We did not find any differentially expressed miRNA between FUS- and FUS+ samples using a p_{BH}-value cutoff of 0.05. Nevertheless, applying a less stringent cutoff without correction for multiple testing we determined 81 miRNAs with p_{MW}-values of differential expression of $p_{MW}<0.05$ between the tumour subgroups, 29 with a higher and 52 with a lower expression in FUS- than in FUS+ (**Supplementary Table 8**).

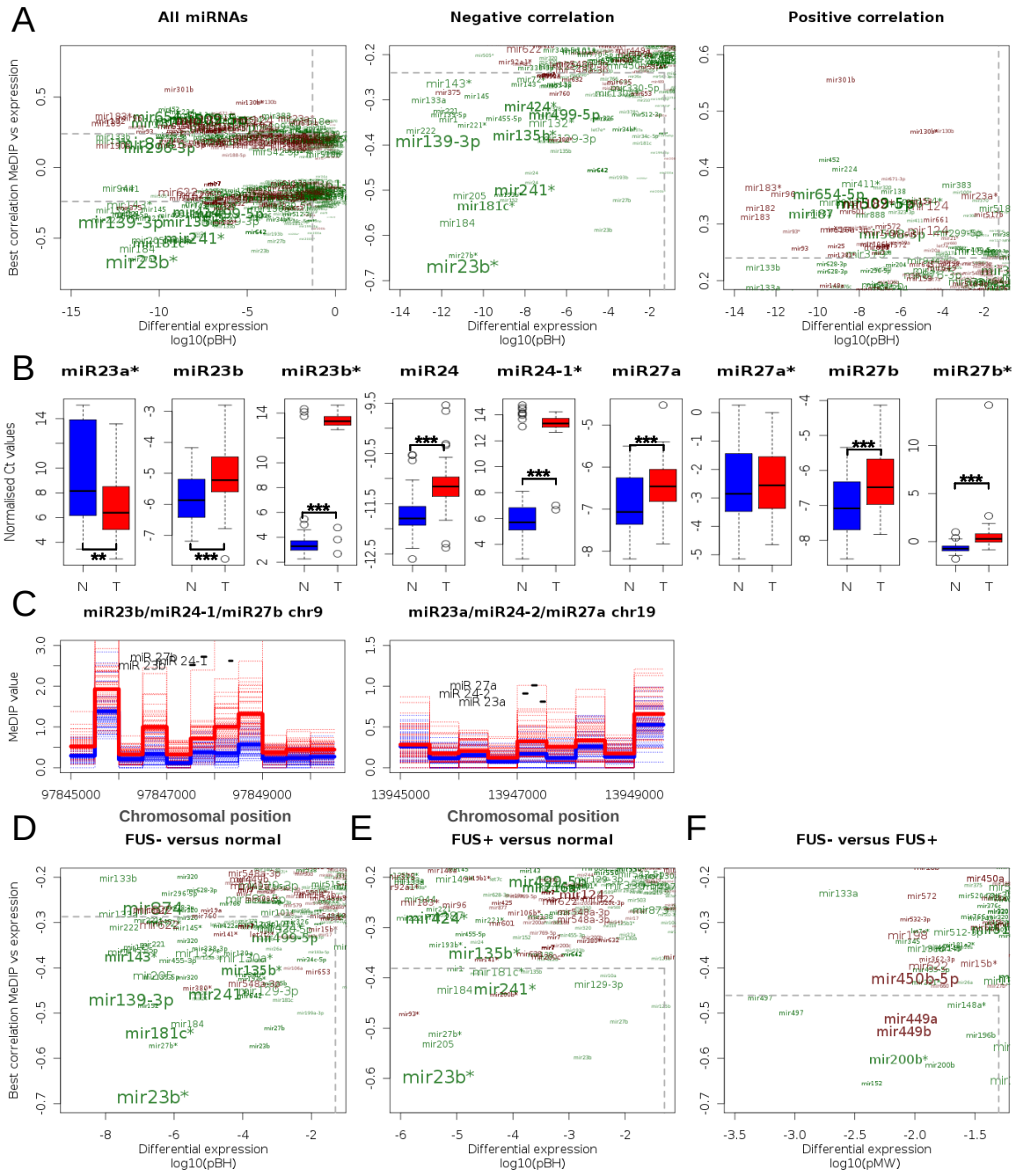


Figure 31. Methylation and expression of miRNAs. (**A**) The plots show the p_{BH}-values of differential expression versus the correlation value of DNA methylation and miRNA expression for each miRNA in an overview of all miRNAs (left), the miRNAs with negative correlations (middle) and miRNAs with positive correlations (right). Dotted lines enclose areas of significance of differential expression (p_{BH}<0.05) and of correlation (p_{BH}<0.05). The size of the symbols is equivalent to the logarithmic expression difference, the larger the symbols the larger the difference. Green symbols mark downregulated, red symbols upregulated miRNAs. Green symbols in the lower left significance area represent hypermethylated and suppressed, red symbols in this area hypomethylated and overexpressed miRNAs. (**B**) Expression of each *miR23/24/27* associated miRNA is shown for all tumour (red) and all normal (blue) samples normalised C_t values. Asterisks denote the significance of the expression differences: **p_{MW}<0.01, ***p_{MW}<0.001. (**C**) Genomic region plots of the vicinity of the *miR23/24/27* cluster on chromosome 9 (left) and chromosome 19 (right). Depicted are the bin-wise MeDIP-Seq values of all prostate tumour (red dotted lines) and normal prostate samples (blue dotted lines) in this region. Solid blue and red lines represent the bin-wise means of normal and tumour MeDIP-Seq values, respectively. The black lines represent miRNA genes. (**D-F**) The plots show significance of the differential expression vs. the correlation value of DNA methylation and miRNA expression for all miRNAs with a negative correlation in FUS- and normal samples (**D**), in FUS+ and normal samples (**E**), and in FUS+ and FUS- samples (**F**).

Correlation analyses of miRNA expression and MeDIP-Seq data of the corresponding samples in a region of ±2kb around each miRNA revealed a significant negative correlation for 107 miRNAs in the whole sample set (**Figure 31A,B**). Of these, 82 miRNAs were additionally significantly differentially expressed: 65 downregulated and 17 upregulated. For example, *miR184* was significantly suppressed and hypermethylated but most interestingly, the cluster of *miR23b*, *miR27b,* and *miR24* on chromosome 9 showed the most significant negative correlation (Rho=-0.48..-0.67) and was highly significantly downregulated. Furthermore, we found at least one hypermethylated bin in the vicinity of each *miR23/24/27* cluster and a reduced expression of all cluster associated miRNAs except *miR23a** and *miR27a** (**Figure 31B,C**).
Restricting correlation analyses to FUS- and normal samples yielded 74 miRNAs (**Figure 31D**), and to FUS+ and normal samples only 16 miRNAs with significant negative correlations and significant differential expression (**Figure 31E**). These numbers suggest a deregulation of miRNA expression through methylation predominantly in FUS- tumours. Both analyses again revealed *miR23b* as one of the most significantly deregulated miRNAs, independent of the fusion status. Nine miRNAs exhibited significant negative correlations and differential expressions comparing FUS- and FUS+ samples (**Figure 31F**).

To further investigate the impact of DNA methylation on the *miR23b* promoter activity, we performed **luciferase reporter assays**. Therefor, a long (982bp, chr9:97,846,525-97,847,506) and a short fragment (501bp, chr9:97847006-97847506) upstream of the *miR23b* were cloned into the pGL3 luciferase reporter plasmid. As control we used the promoter region of *DUOX1* (220bp, chr15:45422000-45422219). The plasmids were *in vitro* methylated with SssI and methylation was confirmed by methylation sensitive restriction digest with ApaL1 (**Figure 32A**). HEK cells were transfected either with a methylated or an unmethylated plasmid. After 24h the cells were harvested and the luciferase activity was determined. Two independent experiments in triplicates showed a downregulation of *miR23b* and *DUOX1* construct luciferase signalling following *in vitro* methylation (**Figure 32B**).

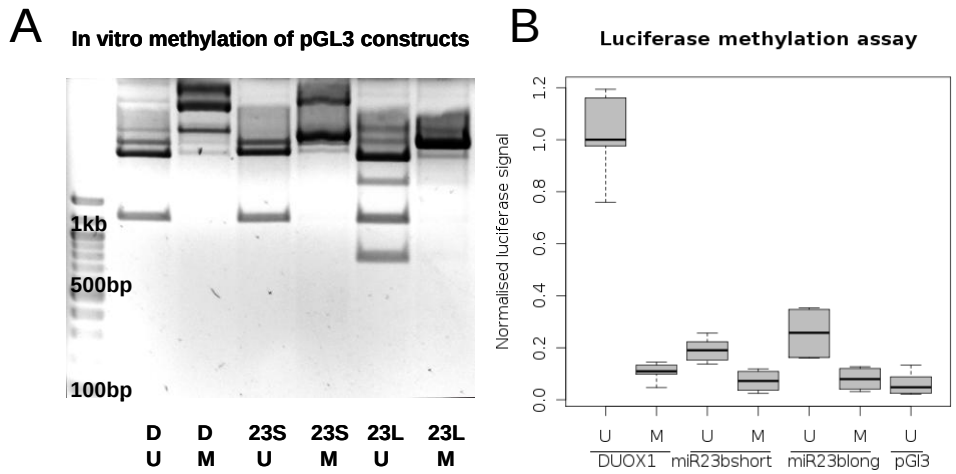


Figure 32. miR23b luciferase methylation assay. (A) DNA agarose gel of enzymatic restriction digests of unmethylated (U) and *in vitro* methylated pGL3 vector containing the DUOX1 (D), miR23b-short (23S) or the miR23b-long (23L) promoter constructs. Restriction has been performed with methylation sensitive ApaL1 enzyme leaving *in vitro* methylated vector constructs uncut at methylated positions (M) while restriction in unmethylated constructs was not obstructed. (**B**) Luciferase assay of methylated (M) and unmethylated (U) reporter constructs of DUOX1, miR23b-short and miR23b-long compared to unmethylated pGL3 empty vector.

3.12 Causes of differential methylation

So far, we have shown that the DNA methylation patterns of tumour samples exhibit extensive alterations, mainly comprising of hypermethylations of promoters and CpG islands of homeobox and miRNA genes. Furthermore, these alterations were more pronounced in FUS- samples. To identify the causes for increased DNA methylation we investigated the gene expression levels of major DNA and histone methyltransferases (*DNMT1, DNMT3A, DNMT3B, EZH1, EZH2*) in the gene expression data of Holger Sültmann (DKFZ). *DNMT1, DNMT3A*, and *EZH2* expression levels were increased in tumour samples. Furthermore, *EZH2* was higher expressed in FUS- than in FUS+ tissues (p_{MW}=0.013) (**Figure 33**).

DNMT1 is responsible for methylation of hemimethylated CpGs and thus for maintaining DNA methylation during cell division (LEONHARDT 1992), while *DNMT3A* and *B* mediate *de novo* methylation (JAENISCH 2003, OKANO 1999). Tumour associated upregulation of *DNMT3A* might be responsible for *de novo* promoter methylation and methylation of CpG islands. *DNMT3A* expression can be regulated by miRNAs of the *miR29*-group (FABBRI 2007). Thus, we examined the *miR29* expression and found *miR29a* and *b* downregulated in the tumour samples (**Figure 34A**), suggesting that this downregulation contributed to *DNMT3A*'s increased expression. We examined the DNA methylation levels of the *miR29*

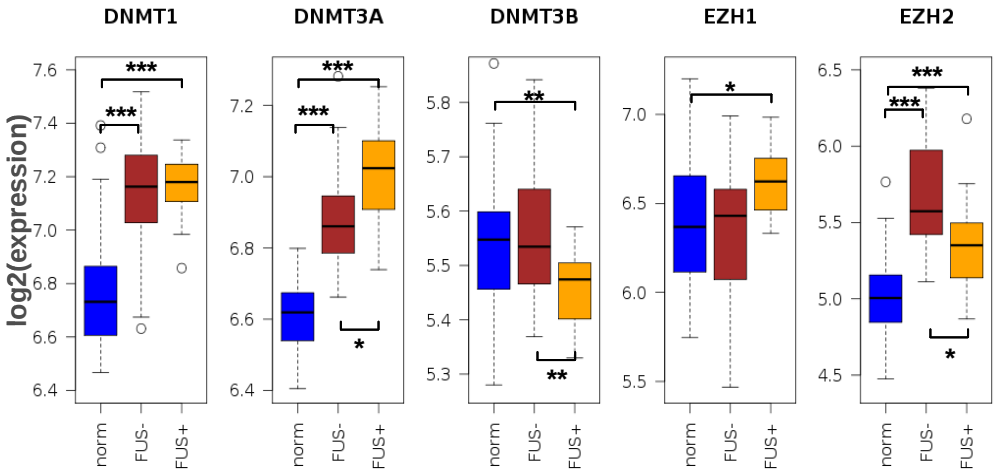


Figure 33. Gene expression levels of *DNMT1*, *DNMT3A*, *DNMT3B*, *EZH1* and *EZH2* in normal (blue), FUS- (brown) and FUS+ (orange) prostate tissues are given as log_2 values of the array signal. Asterisks mark significances of differential expressions. ***p_{MW}<0.001, **p_{MW}<0.01, *p_{MW}<0.05.

region and found only a not significant increase in the miRNA associated bins (**Figure 34B**), making a DNA methylation mediated *miR29* regulation implausible.

In contrast to the globally acting DNMTs, *EZH2*, a polycomb group protein with histone H3K27-methyltransferase activity, causes DNA hypermethylation of distinct target genes during cancer progression and tissue development (VARAMBALLY 2002, VIRE 2006, YU 2007). In general, it functions as a transcriptional repressor and links histone and DNA methylation (CAO 2002, KUZMICHEV 2004, VARAMBALLY 2002, VIRE 2006). As we found developmental genes as a prominent group among the hypermethylated genes, *EZH2* overexpression might be causative for the hypermethylations observed. Furthermore, since we have found *EZH2* even higher expressed in FUS- compared to FUS+ samples, *EZH2* becomes an interesting candidate to explain the differences observed in the methylation patterns of these tumour subgroups.

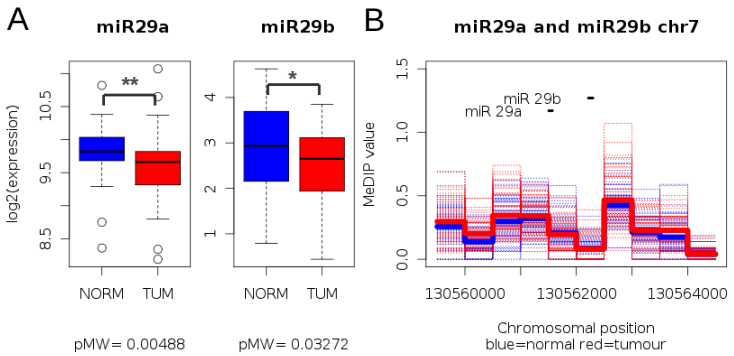


Figure 34. Expression and methylation of *miR29*. (**A**) Expression values of *miR29a* and *miR29b* show significant downregulations of these miRNAs in tumour. (**B**) The methylation plot of the genomic region of the *miR29a/b* cluster on chromosome 7 did not show differences in methylation between tumour (red) and normal samples (blue).

3.13 Methylation and expression of *EZH2* target genes

Yu *et al.* have described 84 polycomb group protein target genes that are downregulated in metastatic PCa (YU 2007). In our data we found that fifty of these 84 genes (60%) contained at least one TSS associated bin significantly hypermethylated in tumour (p_{BH}<0.05). On the other hand, for only four (5%) of the target genes we found at least one hypomethylated TSS associated bin and no hypermethylation in the examined regions. Seven *EZH2* target genes showed significant hyper- as well as hypomethylation (**Supplementary Table 9**).

The finding that FUS- samples had an elevated expression of *EZH2* compared to FUS+ samples (**Figure 33**) was reflected by the methylation of the *EZH2* target gene promoters: In 18 genes at least one promoter associated region was significantly hypermethylated compared to FUS+ samples, but only in three genes a significant hypomethylation could be detected in FUS- samples (**Supplementary Table 9**).

EZH2 is described to cause hypermethylation of its target genes and corresponding downregulation of the gene expression. Indeed, for 88% of the *EZH2* target genes (74 of 84) we confirmed a significant positive correlation (p_{BH}<0.05) between the expression of *EZH2* and their promoter methylation. Interestingly, this effect was again dominated by FUS- samples: Using only FUS- and normal samples for correlation analysis, we found 85% of the genes as significantly positively correlated (p_{BH}<0.05). Performing the same analyses with only FUS+ and normal samples yielded 11% of significantly positively correlated genes (**Supplementary Table 9**).

Next, we correlated the expression of *EZH2* with the expression of its targets to validate the function of *EZH2* as transcriptional suppressor. Of the target genes 36 showed a negative correlation (43%). Restricting the analysis to normal and FUS- samples yielded a negative correlation for 35 of the target genes, while in the analysis of FUS+ and normal samples 16 genes were negatively correlated with the expression of *EZH2*. Sample-wise methylation and expression patterns of *EZH2* target genes presented in **Figure 35** revealed significantly more alterations in the FUS- tumour group.

An analysis of the 231 homeobox genes, the primary targets of *EZH2* mediated cellular effects, revealed a positive correlation for 200 (87%) of them between *EZH2* expression and homeobox gene promoter methylation. Taking only FUS- and normal samples into account, methylation of 192 homeobox genes was significantly correlated, whereas FUS+

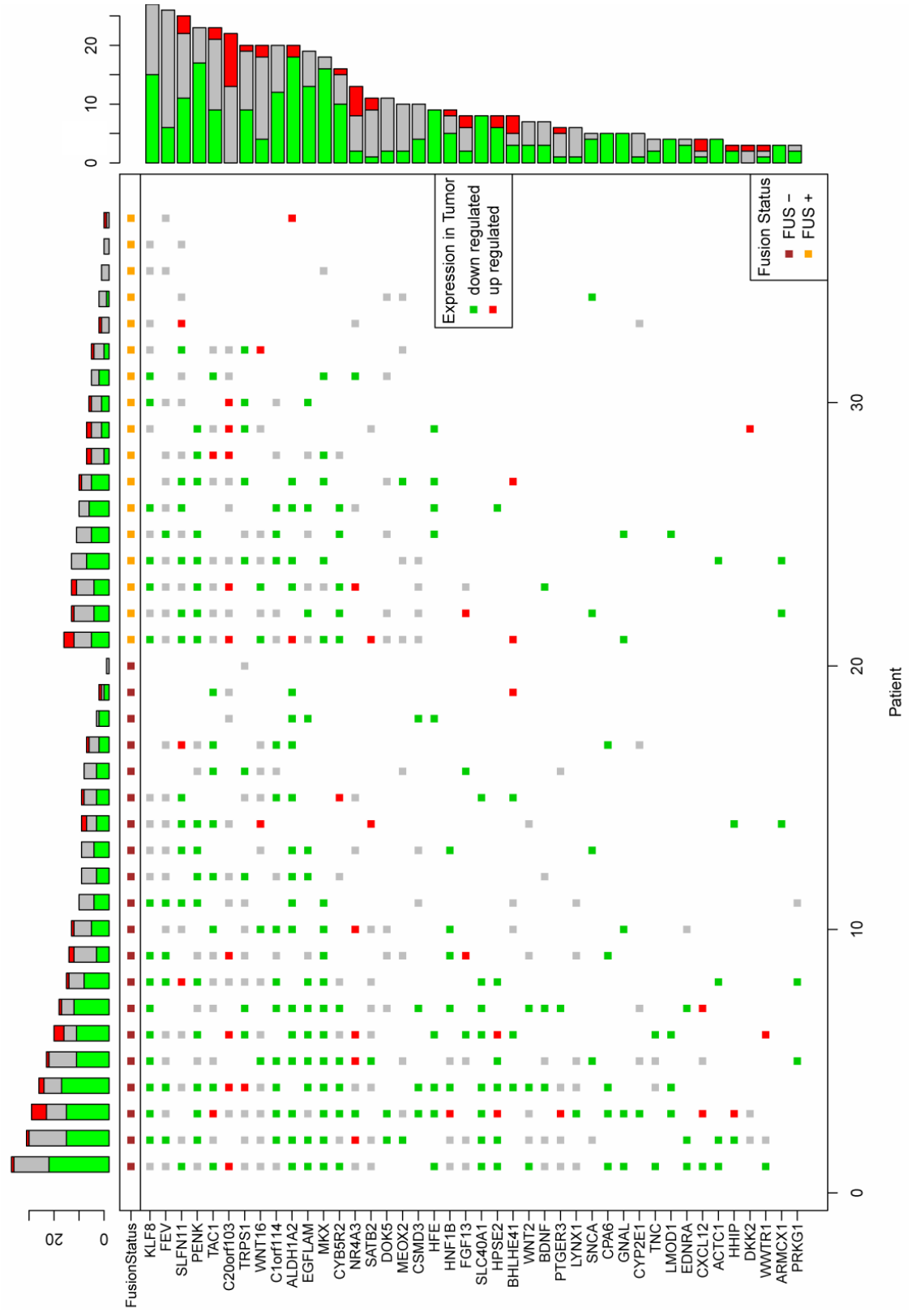


Figure 35. **Patient-wise patterns of promoter hypermethylation and corresponding gene expression of *EZH2* target genes according to Yu *et al.* (YU 2007).** *EZH2* target genes had to be differentially methylated in at least 3 patients to be included in the plots. Patients are ordered according to the fusion state (brown=FUS-, orange=FUS+) and the sum of affected genes. Each visible square in the matrix marks a hypermethylated promoter. Colour coding of squares depicts expression changes (red: upregulated in tumour, green: downregulated in tumour, gray: no differential expression). The histograms on both sides summarise the data patient-wise (top) and gene-wise (side) (courtesy of Dr. Martin Kerick).

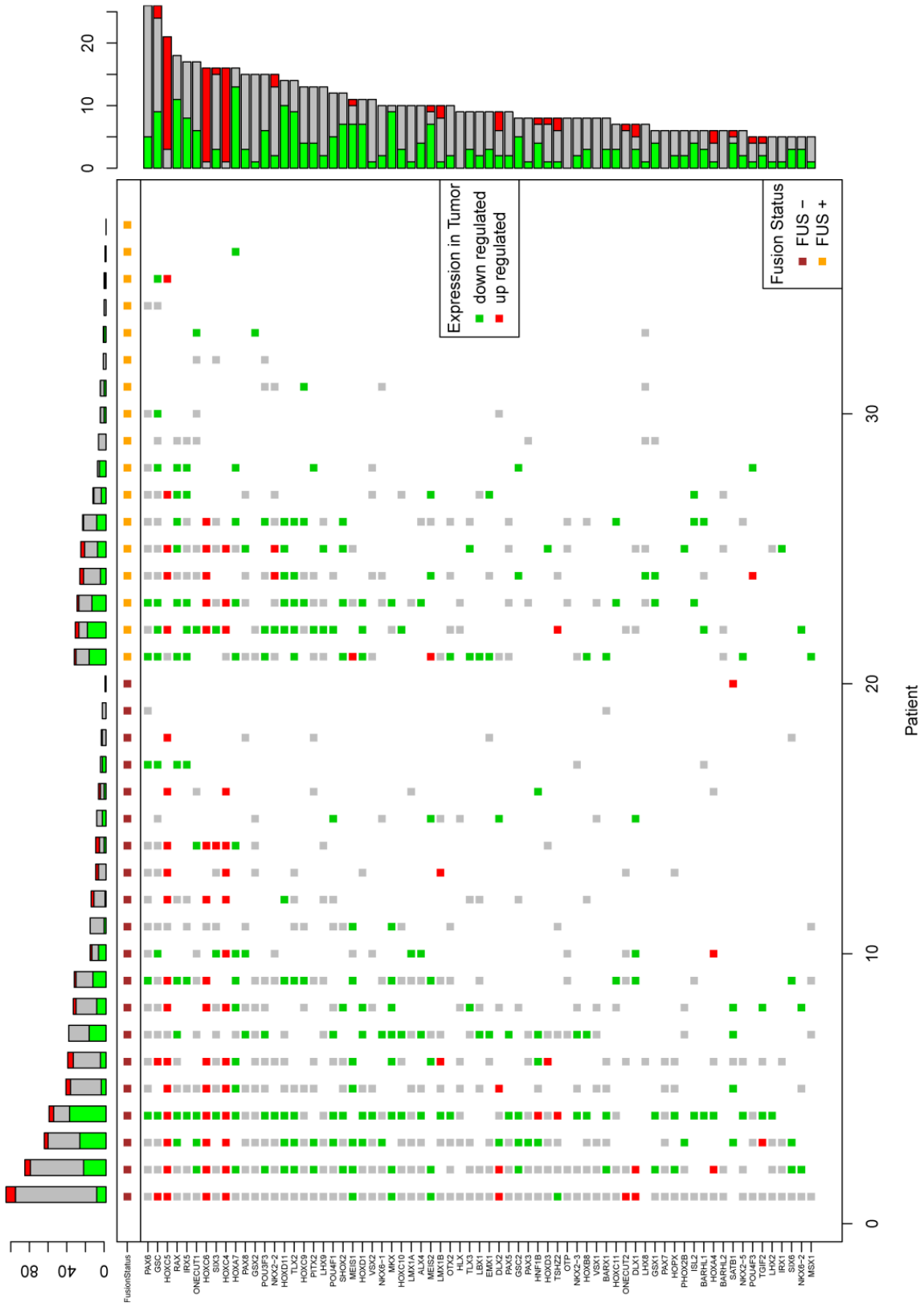


Figure 36. **Expression and methylation of homeobox genes in *TMPRSS:ERG* fusion positive and negative tumors.** Patient-wise patterns of promoter hypermethylation and corresponding gene expression of homeobox genes. Homeobox genes had to be differentially methylated in at least 5 patients to be included in the plot. Patients are ordered according to fusion state (brown=FUS-, orange=FUS+) and sum of affected genes. Each visible square in the matrix marks a hypermethylated promoter. Colour coding of squares depicts expression changes (red: upregulated in tumour, green: downregulated in tumour, grey: no differential expression). The histograms on both sides summarise the data patient-wise (top) and gene-wise (side) (courtesy of Dr. Martin Kerick).

and normal samples showed a correlation for 65 homeobox genes. Most of these 65 genes were also contained in the significantly correlated genes in FUS- samples. Again, sample-wise methylation and expression patterns of homeobox genes presented in **Figure 36** revealed significantly more alterations in the FUS- tumour group.

3.14 Causes for increased *EZH2* expression

Based on our analyses *EZH2* seems to be a key regulator of PCa associated DNA methylations, particularly in FUS- samples. Next, we investigated possible causes of its increased expression in the tumour subgroups. *EZH2* is a known target of the *ERG* transcription factor (KUNDERFRANCO 2010, YU 2010). *ERG* was 12fold overexpressed (**Figure 37A**) and correlated with *EZH2* expression in FUS+ samples (Rho=0.52), but not in FUS- samples (Rho=0.24) (**Figure 37B**). Consistent with this finding, a knock down experiment of *ERG* in fusion positive VCaP prostate cancer cells performed by Mark Laible (DKFZ) caused a downregulation of *EZH2* as determined by qPCR (**Figure 37C**).
In FUS- samples on the other hand, we found normal levels of *ERG*. However, the expression of *EZH2* was even higher than in FUS+ samples (**Figure 33**). Thus, we investigated the *EZH2* promoter methylation, but found no association of enhanced expression of *EZH2* with promoter methylation in FUS- samples; rather the methylation levels were comparable in both tumour subsets (minimal p-value after BH correction = 0.29). Since we had identified miRNA genes as major sites of differential methylation in our MeDIP-Seq experiments (see section 3.6), we were interested to know whether a deregulated miRNA could be the reason for the increased *EZH2* expression in FUS- samples.

Of six suggested regulators of *EZH2* (*miR26a* (ALAJEZ 2010, LU 2011), *miR101* (VARAMBALLY 2008), *miR138* (KISLIOUK 2011), *miR124* (ZHENG 2011), *miR214* (JUAN 2009), and *let-7b (TZATSOS 2011)*), three miRNAs were significantly differentially expressed between normal and tumour tissues (*miRNA-124*, *-214*, and *-138*). However, these miRNAs did not show differential expression between FUS- and FUS+ samples. *MiR26a* on the other hand exhibited significant expression differences between FUS+ and FUS- samples (p_{MW}=0.026) (**Figure 37D**).

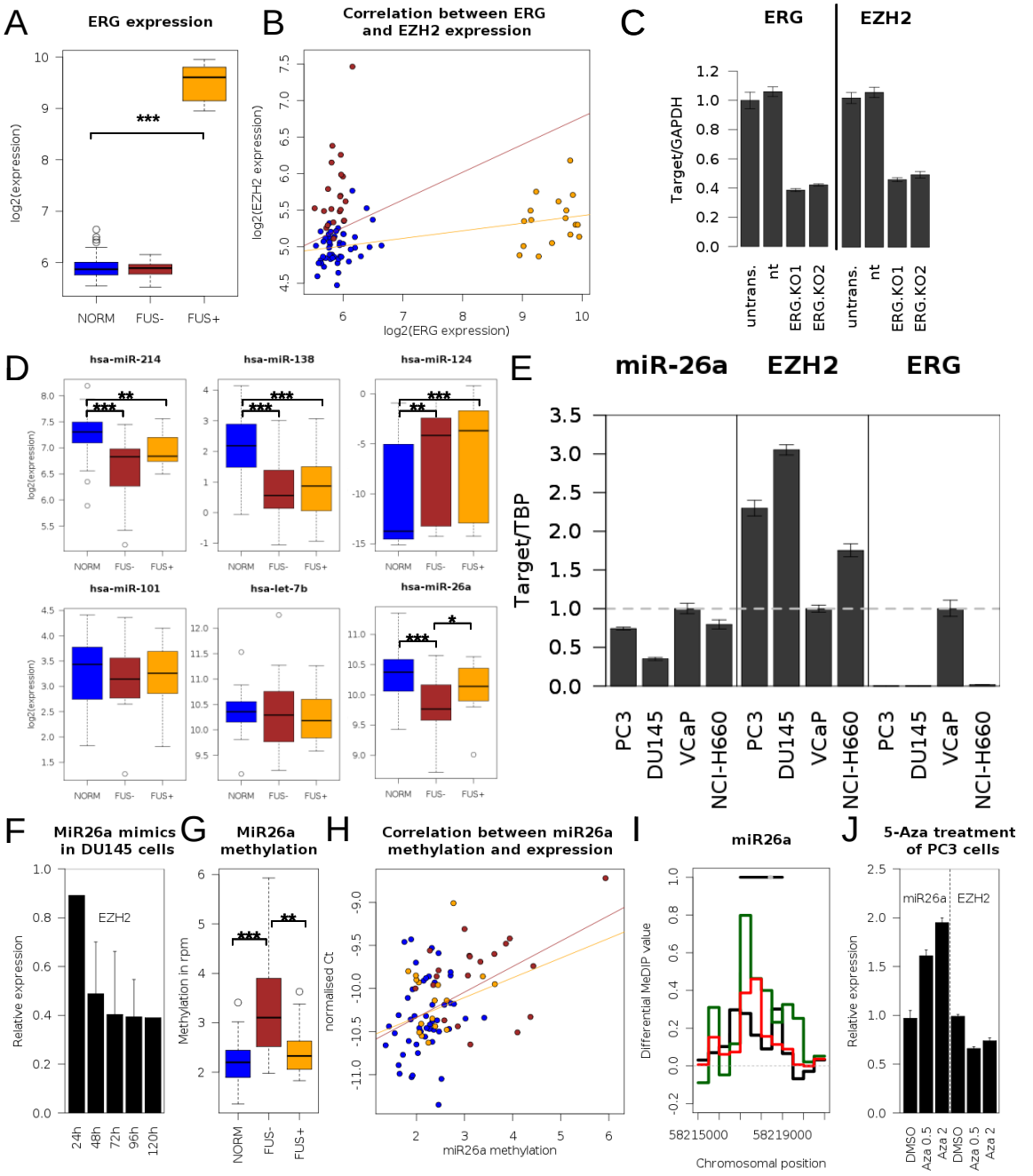


Figure 37. Causes of *EZH2* overexpression. (**A**) Gene expression analysis of *ERG* from microarray data: Expression of *ERG* is comparable in normal and FUS-, and 12fold increased in FUS+ samples. (**B**) *EZH2* expression correlates with *ERG* expression in FUS+ (orange, cor=0.52) but not in FUS- samples (brown, cor=0.24). (**C**) Expression of *ERG* and *EZH2* in VCaP control samples (untrans.=untransfected, nt=non targeting control) and VCaP cell lines after knock down of *ERG* (ERG.KO1 and 2) measured by qPCR (courtesy of Mark Laible, DKFZ). (**D**) Boxplots of the expressions of six *EZH2* regulating miRNAs derived from qPCR analyses. Asterisks denote differential expressions: *pMW<0.05, **pMW<0.01,***pMW<0.001. (**E**) Normalised qPCR results of the expressions of *miR26a*, *EZH2* and *ERG* in prostate cancer cell lines (courtesy of Mark Laible, DKFZ). (**F**) QPCR time course of *EZH2* expression in DU145 cells after treatment with *miR26a* mimics (courtesy of Behnam Sayanjali). (**G**) Boxplot of methylation values in the *miR26a* associated region on chr12:58,217,001-58,219,000. (**H**) Correlation plot of miR26a expression (C_t values) and methylation of region chr12:58,217,001-58,219,000 in FUS- (brown) and normal (blue) and FUS+ (orange) and normal samples. Pearson's correlation values are -0.47 for FUS- and normal, and -0.23 for FUS+ and normal samples. (**I**) Differential methylation in DU145 (black), PC3 cells (green), and FUS- samples (red) compared to normal samples in the DMR of *miR26a* (chr12:58,217,001-58,219,000) (black line). (**J**) Relative expression of *miR26a* and *EZH2* in PC3 cells after treatment with 5-aza-2'-deoxycytidine (0.5µM or 2µM).

Furthermore, *miR26a* expression was negatively correlated to *EZH2* expression in FUS-, but not FUS+ samples (FUS-: cor=-0.37, p<0.003; FUS+: cor=0.09, p=0.48). Quantitative PCR analyses of *miR26a* and *EZH2* expression in the FUS- cell lines DU145 and PC3 as well as in the FUS+ cell lines VCaP and NCI-H660 recapitulated the results of the tumour tissue analyses: *miR26a* was lower and *EZH2* was higher expressed in the FUS- cell lines than in FUS+ cell lines (**Figure 37E**). In contrast FUS+ cell lines had a higher expression of *ERG* (VCaP: 2.400x elevated, NCI-H660 40x elevated).

To further elaborate these findings, we transfected fusion negative DU145 prostate cancer cells with *miR26a* mimics and found a more than twofold decrease of *EZH2* gene expression level after 48h (**Figure 37F**).
We next investigated if an increased DNA methylation of the *miR26a* locus could cause the decreased *miR26a* expression in FUS- tumours. Indeed, we identified a hypermethylated region of 2kb (chr12:58,217,001-58,219,000) in FUS- samples (**Figure 37G**) that significantly negatively correlated with *miR26a* expression (cor=-0.47; $p<4.8*10^{-5}$) (**Figure 37H**).
Since we have found the *miR26a* region methylated in fusion negative DU145 and PC3 cell lines, too (**Figure 37I**), we tested if a treatment of PC3 cells with 5-aza-2`-deoxycytidine – blocking DNA methyltransferase activity – resulted in an increased *miR26a* level. Indeed, we determined an increased expression of *miR26a* and accompanying decrease in *EZH2* after treatment (**Figure 37J**). In FUS+ samples we did not find this region to be hypermethylated and did not detect a correlation between methylation and *miR26a* expression (**Figure 37G,H**).

3.15 Luciferase Reporter Assays of *miR26a* methylation

To assess whether methylation of the *miR26a* region caused a reduced *miR26a* expression, we established a methylation sensitive luciferase reporter assay. Therefor, a 1.5kb large fragment of the *miR26a* region (chr12:58,217,030-58,218,580) was cloned into the pGL3 luciferase reporter plasmid and methylated *in vitro*. As control we used the *DUOX1* promoter region (chr15:45,422,000-45,422,219). *In vitro* methylation of the plasmids was confirmed by methylation sensitive restriction digest with ApaL1 as well as BS sequencing of the plasmid using the mir26_meth_check primers.

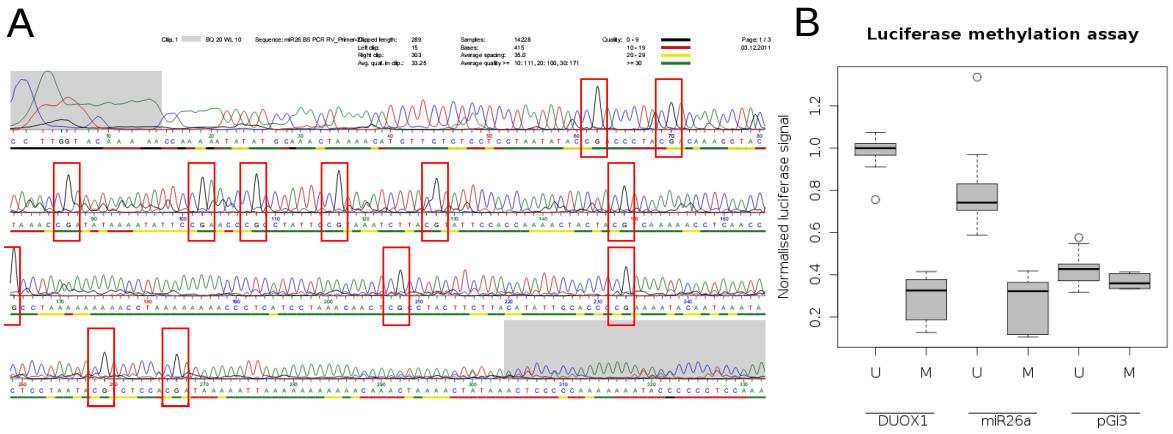


Figure 38. Luciferase methylation assay of the miR26a region.
(**A**) BS Sanger sequencing results of the reverse complement of the *miR26a* region cloned into pGL3. Framed are methylated CpG groups. (**B**) Luciferase methylation assay of methylated (M) and unmethylated (U) reporter constructs of *DUOX1* and the *miR26a* region (chr12:58,217,030-58,218,580) compared to unmethylated and methylated pGL3 empty vector showed a downregulation of the luciferase signal induced by *in vitro* methylation.

Three independent experiments using triplicates yielded the same results: *In vitro* methylation led to a significant reduction of luciferase signal in the *miR26a* as well as in the *DUOX1* construct (**Figure 38**).

3.16 Knock down of ERG in VCaP cells followed by MeDIP-Seq

In FUS+ samples the increased expression of *ERG* leads to an increase in *EZH2* expression. This probably results in an elevated methylation of *EZH2* target and homeobox genes. To further analyse this relation Mark Laible (DKFZ) performed siRNA mediated knock downs of *ERG* in FUS+ VCaP cells. Quantitative PCR analyses of *ERG* confirmed a knock down efficiency of 60% accompanied by a reduction of *EZH2* expression of more than 50% (KUNDERFRANCO 2010) (**Figure 37C**).

To assess the effect of *ERG* and resulting *EZH2* downregulation on the DNA methylation patterns, we performed MeDIP-Seq experiments on VCaP cell line DNA after 96h *ERG* knock down in duplicates and compared it with untransfected VCaP cells and VCaP cells transfected with a non-targeting siRNA pool. Overall, we detected 18,010 regions with altered methylation values (cutoff: x<71% or x>141%) (**Figure 39**). TSS associated regions were enriched among the DMRs (OR=1.60, $p<2.2*10^{-16}$). Among these TSS associated DMRs, we found a subtle enrichment of *EZH2* target gene (OR=1.69, p=0.067) and of cancer census gene promoter regions (OR=1.27, p=0.071) (**Table 35**).

Within the DMRs hypermethylation prevailed. We found 11,656 hyper- and 6,354 hypomethylated regions. MiRNA genes were most significantly enriched among the hypermethylated genes (OR=2.95, p=0.012).

Although *EZH2* target gene and homeobox gene promoters were enriched in the hypomethylated gene fraction, due to the small number of differentially methylated genes of these two classes this enrichment was statistically not significant (*EZH2* target gene hypomethylation: OR=1.77, p=0.374; homeobox gene hypomethylation: OR=1.29, p=0.768, respectively) (**Table 35**). Altogether, homeobox and *EZH2* target genes were

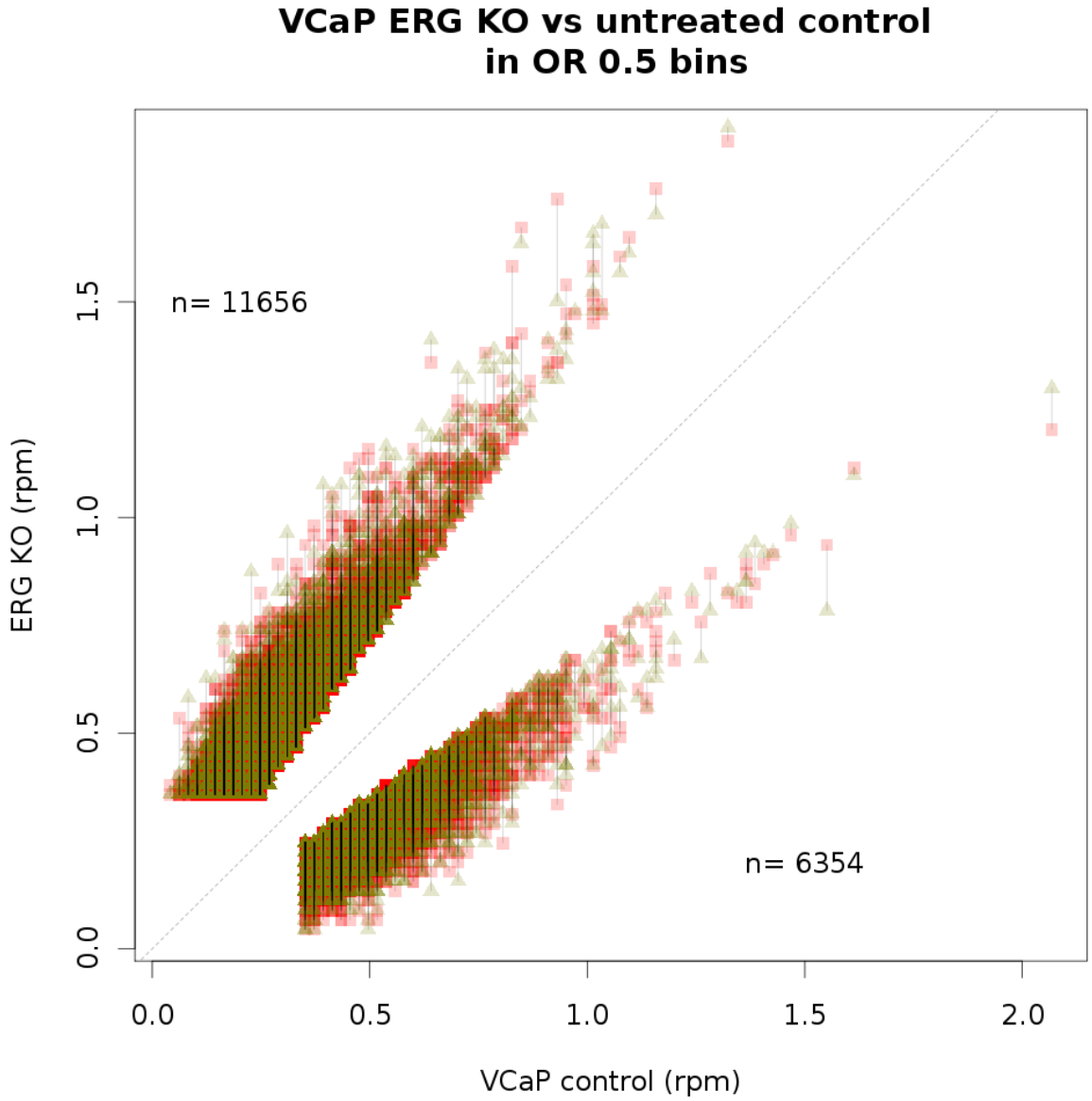


Figure 39. *ERG* knock down experiments in VCaP cells led to significant changes in methylation in 18,010 regions as determined by MeDIP-Seq analyses. Shown are the methylation values (rpm) of significantly methylated bins with $\log_2$(ratios) of <(-0.5) or >0.5 of knock down to control cells. The methylation values of the *ERG* knock down experiments are plotted against the methylation values of the untransfected control experiment. Bin-wise results of knock down experiment 1 (red squares) and 2 (green triangles) are connected by black lines. Bins below the diagonal line are hypomethylated, bins above the line are hypermethylated in *ERG* knock down cells.

barely affected by differential methylation after *ERG* knock down and hyper- and hypomethylation occurred in almost equal amounts in these genes despite a global hypermethylation in *ERG* knock down samples.

Table 35: Differential methylations induced by *ERG* knock down in VCaP cells. Total numbers of bins located in all bins and promoter regions of gene sets are indicated in column (A). Differentially methylated bins are given in column (B). Odds ratios of enrichment of the respective regions in DMRs are shown in column (C), the respective p-values in column (D). The number of hypermethylated bins is given in (E). The odds ratios of enrichment of the regions in the hypermethylated bins after *ERG* knock down with corresponding p-values are depicted in columns (F-G). The number of hypomethylated bins after *ERG* knock down is given in (H). The odds ratios of enrichment of the regions in the hypermethylated bins with corresponding p-values are depicted in columns (I-J).

	Total bin count (A)	All DMRs (B)	OR DMR (C)	OR DMR p-value (D)	hyper (E)	OR hyper (F)	OR hyper p-value (G)	hypo (H)	OR hypo (I)	OR hypo p-value (J)
All bins	6191344	18010			11656			6354		
All promoters	595703	2613	1,60	2,2E-16	1760	1,15	2,2E-03	853	0,87	2,2E-03
Census genes	11238	62	1,27	7,1E-02	42	1,02	1	20	0,98	1
MiRNAs	7951	42	1,21	2,3E-01	36	2,95	1,2E-02	6	0,34	1,2E-02
EZH2 target genes	1759	13	1,69	6,7E-02	7	0,56	3,7E-01	6	1,77	3,7E-01
Hox genes	3283	13	0,90	8,9E-01	8	0,77	7,7E-01	5	1,29	7,7E-01

We subjected all unique gene IDs associated with hyper- or hypomethylated bins to DAVID pathway analysis and observed a significant overlap of the top categories of the genes: Of 2,267 unique hypermethylated genes 1,556, and of 1,095 hypomethylated genes 751 could be analysed by DAVID. Cytoskeleton associated gene sets were enriched in both cohorts (**Table 36**).

Table 36: DAVID analyses of hyper- and hypomethylated genes in VCaP *ERG* knock down cells. Only enriched terms with Benjamini corrected p-values<0.05 are listed. Column headings denote the enriched functional association term, the number of genes associated to this term, the percentage of genes of the input list associated to this term, the p-value of enrichment of this term, and the fold enrichment in input list.

Term	Count	%	p-value	Fold Enrichment
Hypermethylated genes				
alternative splicing	665	44,81	1,26E-24	1,35
cytoskeleton	81	5,46	1,48E-08	1,94
actin-binding	33	2,22	1,84E-04	2,03
alternative promoter usage	12	0,81	1,83E-04	3,88
atp-binding	119	8,02	4,01E-04	1,36
chromosomal rearrangement	35	2,36	3,76E-04	1,91
Proto-oncogene	30	2,02	5,65E-04	1,98
chromatin regulator	28	1,89	7,93E-04	2,00
Hypomethylated genes				
alternative splicing	330	45,45	2,53E-11	1,33
GO:0008092~cytoskeletal protein binding	38	5,23	7,71E-06	2,22
GO:0005856~cytoskeleton	77	10,61	1,40E-05	1,63

Furthermore, both were significantly enriched for genes with multiple transcripts due to alternative splicing, while the term 'alternative promoter usage' was uniquely enriched in hypermethylated genes. Between the hyper- and hypomethylated gene sets we found only a small overlap of 110 genes, i.e. genes that have hyper- and hypomethylated regions implying that splicing and cytoskeleton associated hyper- and hypomethylated gene sets strongly differ.

MeDIP-Seq analyses of FUS+ and FUS- samples had revealed 27,488 differentially methylated regions (17,971 with higher methylation in FUS- and 9,517 with higher methylation in FUS+). These DMRs were associated with 4,933 unique genes. Of these differentially methylated genes 668 were as well differentially methylated in the *ERG* knock down experiment. We found an enrichment of alternatively spliced and cytoskeleton genes, as well as nucleotide exchange factor associated genes in DAVID analysis (**Table 37**).

Table 37: DAVID analysis of genes found differentially methylated between FUS+ and FUS- as well as between VCaP *ERG* knock down cells and control cells. Only enriched terms with Benjamini corrected p-values<0.05 are listed. Column headings denote the enriched functional association term, the number of genes associated to this term, the percentage of genes of the input list associated to this term, the p-value of enrichment of this term.

Term	Count	%	p-value
alternative splicing	289	52,9	7,80E-18
cytoskeletal protein binding	38	7	4,90E-08
actin-binding	21	3,8	8,00E-06
sh3 domain	19	3,5	1,00E-05
Ras guanyl-nucleotide exchange factor activity	13	2,4	4,90E-06
Rho guanyl-nucleotide exchange factor activity	11	2	3,30E-05
guanyl-nucleotide exchange factor activity	16	2,9	1,90E-05
chromosomal rearrangement	19	3,5	4,00E-04
Dbl homology (DH) domain	10	1,8	1,10E-04
SPEC	7	1,3	3,60E-04
Src homology-3 domain	18	3,3	9,10E-05
alternative initiation	8	1,5	1,30E-03

The comparison of genes with differential methylation found in VCaP *ERG* knock down cells with the 14,352 genes found to be differentially methylated in tumour compared to normal samples yielded an overlap of 1,740 genes. Again, DAVID analysis showed an enrichment of 'alternative splicing' and 'cytoskeleton' terms (**Table 38**).

Table 38: DAVID analysis of genes found differentially methylated between tumour and normal as well as between VCaP *ERG* knock down cells and control cells. Only enriched terms with Benjamini corrected p-values<0.05 are listed. Column headings denote the enriched functional association term, the number of genes associated to this term, the percentage of genes of the input list associated to this term, the p-value of enrichment of this term.

Term	Count	%	p-value
alternative splicing	663	49,4	3,80E-33
cytoskeleton	86	6,4	1,50E-11
actin-binding	40	3	6,80E-08
cell projection	78	5,8	2,70E-06
chromosomal rearrangement	36	2,7	5,80E-05
cell junction	46	3,4	7,70E-05
Ras guanyl-nucleotide exchange factor activity	19	1,4	1,60E-05
calcium	78	5,8	9,10E-05
Immunoglobulin domain	52	3,9	7,30E-05
alternative promoter usage	12	0,9	1,10E-04
sh3 domain	29	2,2	1,10E-04
plasma membrane	295	22	1,20E-04
Rho guanyl-nucleotide exchange factor activity	16	1,2	9,00E-05
guanyl-nucleotide exchange factor activity	25	1,9	6,50E-05
Proto-oncogene	29	2,2	5,00E-04
disease mutation	132	9,8	4,80E-04

However, with 29 genes we also observed an enrichment of 'Proto-oncogenes' (**Table 39**). Independent of the DAVID analyses, we had found a strong enrichment of hypermethylations in miRNA genes in PCa tissue as well as in VCaP *ERG* knock down cells. 18 miRNAs were hypermethylated in both experiments: *miR1290*, *miR125b-1*, *miR337*, *miR665*, *miR4310*, *miR549*, *miR657*, *miR3065*, *miR338*, *miR133a-2*, *miR185*, *miR1280*, *miR4285*, *miR1205*, *miR1234*, *and miR24-2*, *miR27a*, *miR23a*. Methylation values of the *miR23a/24-2/27a* cluster are depicted in **Figure 40**.

Interestingly, for *miR23* we had already found a significant negative correlation between gene expression and methylation in the patient samples (see section 3.11.3), thus indicating that the *miR23/miR24/miR27* clusters seem to play a major role in prostate cancer.

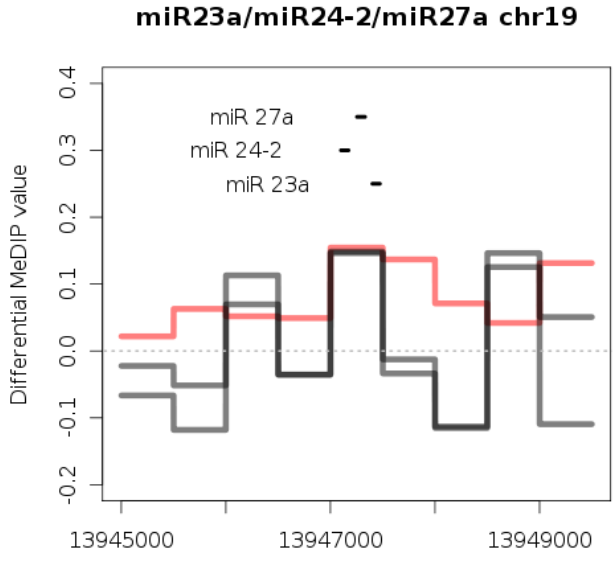


Figure 40. Genomic region plot of the miR23a/miR24-2/miR27a region on chromosome 19. Depicted are the differential methylation values (Δrpm) of tumour and normal samples (red), and of VCaP *ERG* knock down and VCaP cells (black).

Table 39: Genes of the 'Proto-oncogene' class found to be enriched in DAVID pathway analysis of genes differentially methylated between tumour and normal as well as between VCaP cells and VCaP cells after knock down of *ERG*.

ID	Gene Name	Full name
ENSG00000091436	AC013461.1	sterile alpha motif and leucine zipper containing kinase AZK
ENSG00000078061	ARAF	v-raf murine sarcoma 3611 viral oncogene homolog
ENSG00000145819	ARHGAP26	Rho GTPase activating protein 26
ENSG00000069399	BCL3	B-cell CLL/lymphoma 3
ENSG00000186716	BCR	breakpoint cluster region
ENSG00000110619	CARS	cysteinyl-tRNA synthetase
ENSG00000182511	FES	feline sarcoma oncogene
ENSG00000184481	FOXO4	forkhead box O4
ENSG00000010810	FYN	FYN oncogene related to SRC, FGR, YES
ENSG00000087460	GNAS	GNAS complex locus
ENSG00000078399	HOXA9	homeobox A9
ENSG00000107968	MAP3K8	mitogen-activated protein kinase kinase kinase 8
ENSG00000126217	MCF2L	MCF.2 cell line derived transforming sequence-like
ENSG00000119950	MXI1	MAX interactor 1
ENSG00000213672	NCKIPSD	NCK interacting protein with SH3 domain
ENSG00000119508	NR4A3	nuclear receptor subfamily 4, group A, member 3
ENSG00000198400	NTRK1	neurotrophic tyrosine kinase, receptor, type 1
ENSG00000126883	NUP214	nucleoporin 214kDa
ENSG00000009709	PAX7	paired box 7
ENSG00000078674	PCM1	pericentriolar material 1
ENSG00000181690	PLAG1	pleiomorphic adenoma gene 1
ENSG00000143294	PRCC	papillary renal cell carcinoma
ENSG00000169855	ROBO1	roundabout, axon guidance receptor, homolog 1
ENSG00000159216	RUNX1	runt-related transcription factor 1
ENSG00000157933	SKI	v-ski sarcoma viral oncogene homolog (avian)
ENSG00000128487	SPECC1	cytospin B
ENSG00000109685	WHSC1	Wolf-Hirschhorn syndrome candidate 1
ENSG00000104415	WISP1	WNT1 inducible signaling pathway protein 1
ENSG00000109906	ZBTB16	zinc finger and BTB domain containing 16

4 DISCUSSION

We determined alterations in the DNA methylation patterns during prostate tumourigenesis and investigated their consequences on gene and miRNA expression. For these studies we were the first group to combine the MeDIP-Seq technology with SOLiD sequencing. This allowed a genome-wide assessment of DNA methylation patterns with high coverage in multiple samples in parallel, because at this time SOLiD V3+ machines provided an unrivalled sequencing depth with relatively low sequencing costs compared to their competitors Illumina GA2 and Roche 454. We set up a protocol allowing high throughput preparation and sequencing of up to eight barcoded libraries in parallel. Due to the limited amount of patient's DNA (only up to 4µg/sample) we introduced modifications to the original protocol that allowed 500ng-2µg of input DNA in comparison to 4µg used in Weber's original publication (WEBER 2005). We assessed genome-wide DNA methylation patterns of 104 prostate tissues and generated a yet unparalleled amount of methylation data for a single disease. Typical numbers in publications so far are for prostate cancer 32 tissue samples analysed by MethylPlex-next-generation sequencing (KIM 2011), for Lupus erythematodes twelve pairs of samples analysed by BS arrays (LIN 2012), for peripheral nerve sheath tumours 26 samples analysed by MeDIP-Seq (FEBER 2011), and for sporadic amyotrophic lateral sclerosis 20 samples analysed by MeDIP-arrays (MORAHAN 2009). This large sample number allowed extensive statistical analyses. We were able to detect candidate biomarkers for prostate cancer detection and started with validation experiments in independent sample cohorts where we tried to establish a urine-based detection method. On a functional side, we detected more than 147,000 changes in the DNA methylation patterns which potentially influence the expression of genes of major cancer and developmental pathways. Moreover, we determined a 'methylator'-like phenotype in *TMPRSS2:ERG* fusion negative tumour samples resulting in increased hypermethylations of miRNA and developmental gene pathways, and thereby provided an explanation for tumour formation in the fusion negative group of prostate cancers. Integration of miRNA and gene expression data allowed further dissection of key pathways and led to the identification of *miR26a* and *EZH2*, a polycomb group protein, as central deregulated nodes involved in PCa formation.

4.1 Discussion of the experimental setup

We analysed 53 normal and 51 prostate tumour samples including 17 cases with a *TMPRSS2:ERG* translocation (FUS+) and 20 cases without (FUS-). All tumours were selected to contain at least 70% tumour cells. Nevertheless, in principal component analyses and in sample-wise analyses of differentially methylated regions we identified the tumour sample IGP34 to show a pattern more resembling that of normal prostate tissues. Reclassification of this sample by trained pathologists revealed a tumour cell content of only 40%. Nevertheless, we left IGP34 in the data set to identify markers that are also useful to identify samples with low tumour cell content.

We analysed the sequencing statistics of the samples to detect potential sources of bias that might have introduced artefacts into our results. We could not detect any differences regarding sequencing depths between the different subgroups analysed in this study (minimal $p_{MW}>0.35$). Nevertheless, we found a significantly increased ratio of uniquely mappable reads/mappable reads (UMO/MO) in tumour samples ($p_{MW}<0.0002$). This might be due to the finding that a lower fraction of reads mapped to repetitive regions in tumours. At this point a sampling bias could be excluded because during library preparation the samples of the different groups were pooled randomly. The observed tendency was mainly caused by FUS- samples which exhibited significantly higher UMO/MO ratios ($p_{MW}<0.003$). These findings could be explained by hypomethylation of repetitive elements in tumour and especially FUS- samples and are supported by analyses of LINE L1 elements that were covered by significantly less reads in FUS- samples (**Figure 20G**).

We also detected slightly elevated GC contents of the sequenced fragments in FUS- samples (**Supplementary Table 2**). With 37.6% the GC content in bins annotated as LINE L1 elements is significantly lower than the GC content in promoter associated bins (GC content = 40.8%). Thus, an enrichment of promoter and a depletion of LINE L1 regions due to underlying hyper- or hypomethylation in FUS- consequently led to a higher total GC content in sequencing statistics.

FUS- and FUS+ samples in our study showed the same distributions of age, Gleason score, TNM grading, and PSA level (**Figure 14**). Quantitative PCR analyses of a methylated (*BRD1*) and an unmethylated control region (*COQ3*) revealed no differences in immunoprecipitation enrichments between the sample subgroups. Thus, differences found between these tumour subgroups are most likely due to different mechanisms underlying tumour formation and are not an artefact of sampling or library preparation.

On average, 11% of the uniquely mapped MeDIP-Seq reads did not contain any CpG within 200bp. This was potentially caused by i) an unspecific binding of DNA to the enrichment beads, or ii) – as libraries were size selected on a gel prior to sequencing – by fragments slightly larger than 200bp with at least one methylated CpG in the distal region, or iii) results from methylated and enriched cytosines in a non-CpG context, which, however, has not been described for tumours so far. If the first point was true, we would have to expect approximately 11% false positive reads per bin (=genomic region). To account for this bias, we introduced individual threshold values for each sample: We applied a binomial testing approach and called a bin 'significantly methylated' only if it contained more sequenced fragments than the threshold value. Analyses of differential methylation between the different sample groups was only employed on regions with at least one group showing significant read counts above the threshold level.

Through sample-wise normalisations of the read counts per bin by the total sample read counts we yielded 'reads per million' (rpm) values for each bin in each sample. The rpm values depend on both, the number of CpGs contained within a region and the degree of methylation of each CpG: Highly methylated CpG poor regions thus might yield lower rpm values than poorly methylated high CpG regions. This is also reflected by the strong enrichment of regions with 3-17 CpGs in our MeDIP-Seq data (**Figure 17A,B**), although in normal samples CpG poor regions are highly methylated while CpG rich regions like functional promoters are almost completely unmethylated (ECKHARDT 2006). Although the majority of significantly methylated bins in normal samples was found in regions with more than seven CpGs/500bp we also detected more than 31,000 significantly methylated bins in regions with up to three CpGs/500bp in normal samples (**Figure 17C**), thus showing that we were also able to analyse low CpG regions.

The MeDIP-Seq approach itself preferentially enriches higher CpG content areas compared to low CpG content areas. This preference is not as strong as in the Methyl-Binding-Domain (MBD)-Seq approach (HARRIS 2010). Bioinformatic based corrections of the CpG enrichment bias using software packages like those developed by Chavez *et al.* (CHAVEZ 2010) are mainly required for sample-wise investigations, e.g. for comparisons of the methylation state of different regions within one sample. We did not apply these algorithms since we analysed differential methylations between samples within the same genomic region, implying an identical CpG background of each bin and thus an identical probability for enrichment of this bin in each group – variations should thus only be caused by differences in DNA methylation.

Nevertheless, we accounted for this bias when we compared MeDIP-Seq with BS-MS values: Since we found in the BS-MS analyses that methylation values of the CpGs within a region were relatively homogeneous (**Figure 22C**) we multiplied the mean methylation values of assayed CpGs within a region with the absolute number of CpGs in the respective 500bp bin. This approach we used for correlation analyses with MeDIP-Seq values and thus also assessed regions were we did not yield BS-MS values for every single CpG but that also contributed to MeDIP enrichment. Our approach is also supported by Eckhardt *et al.* who previously showed strong correlations (>0.8) between the methylation levels of neighboured CpGs within distances below 1,000bp that rapidly declined with increasing distance (ECKHARDT 2006). Integrating the CpG density in such a way, we could achieve correlation values of ~0.8 between the MeDIP rpm and BS-MS values. This shows that MeDIP-Seq is a valid technology to assess the actual DNA methylation level. Harris *et al.* (HARRIS 2010) corroborated this notion by achieving concordance rates of >96% between MeDIP-Seq and an Infinium bead-array, and concordances of >99% between MeDIP-Seq, MBD-Seq and MethylC-Seq.

Interestingly, assessment of the actual methylation value with BS-MS also revealed that artificially methylated DNA which should be completely methylated showed methylation values of down to 75%, implying an incomplete *in vitro* methylation. SssI DNA-methyltransferase that we had used in this experiment as well as in the DNA methylation luciferase assays to generate fully methylated DNA has been described to exhibit reduced activity on DNA in a nucleosomal context (FATEMI 2005). This may explain the incomplete methylation observed in some CpGs in the BS-MS experiments. Here, methylation of control DNA was performed prior to PCR amplification with BS specific primers. In the luciferase experiments on the other hand, *in vitro* methylation was performed on plasmid DNA that does not contain nucleosomal structures, thus incomplete methylation should not have occurred in the luciferase constructs.

Using BS-MS analyses we could recapitulate the bimodal distribution of DNA methylation as described by Eckhardt *et al.* (ECKHARDT 2006): Most CpG dinucleotides of hypermethylated regions had a methylation value of 60-90%, while in hypomethylated regions the CpG methylation values in tumour dropped to 10-30% (**Figure 22B**). Additionally, BS-MS analyses of micro- and macrodissected samples of the same patients also demonstrated the high quality of our samples since the methylation values in our macrodissected samples reflect the methylation values of samples prepared by microdissection (**Figure 22E**).

4.2 Distribution of hyper- and hypomethylations

Hypomethylation of repetitive elements has been described as an epigenetic feature of tumours and potentially leads to chromosomal instability through reactivation of transposable elements (DASKALOS 2009, RODRIGUEZ 2006). On the other hand, hypomethylation has also been described as a feature of late and metastatic prostate cancer and – unlike hypermethylation that occurs at specific CpG rich sites early in tumourigenesis (YEGNASUBRAMANIAN 2004) – seems to be a more stochastic process (YEGNASUBRAMANIAN 2008). This is reflected in our data: Comparing tumour to normal samples we found that hypermethylated regions showed lower p-values than hypomethylated ones, indicating concerted hypermethylations affecting all tumour samples as compared to more unspecific demethylation processes (**Table 27**). Most hypomethylated regions were detected in intergenic repetitive areas. This is in accordance with the literature, because in normal tissues CpGs in repetive, extragenic regions are methylated, while CpG islands in promoters or genes are more or less free of methylation (CROSS 1995, EHRLICH 1982). This pattern was reversed in hypermethylated tumour regions: We found only a quarter of hypermethylated bins located outside of genes and promoters (**Figure 20D**).

Interestingly, these effects were even more pronounced in FUS- samples. Comparing FUS- samples to FUS+ and normal samples we detected an enrichment of reads with more than three CpGs/200bp, reflecting hypermethylation of high CpG regions like gene promoters in FUS- samples. Furthermore, the largest difference between FUS- and FUS+ samples in the number of significantly methylated bins was found in regions with more than six CpGs/500bp, also indicating a specific targeting of high CpG regions as a potentially tumourigenic mechanism in FUS- samples. This was further illustrated by a reduction of methylation of low CpG LINE L1 elements that occurred exclusively in FUS- samples. Hypomethylation of high CpG LINE L1 elements was found in both tumour groups.

Hypomethylation in prostate cancer has been described to occur due to a shortage of the methyl-group donor S-adenosylmethionine (SAM) caused by an increased cell division rate and accompanying targeted hypermethylations (SCHULZ 2006). The probability of a CpG rich LINE L1 to be affected by unspecific hypomethylations in at least one CpG is higher compared to a low CpG LINE L1. Thus, hypomethylation of high CpG content LINE L1 elements can be observed in both tumour classes. Furthermore, FUS- samples exhibiting an even higher number of specific hypermethylations than FUS+ samples

consequently will be more affected by SAM shortage additionally leading to hypomethylation of low CpG LINE L1 elements. Indeed, this is the case in our data set: In addition to more hypermethylations in FUS- we also find more hypomethylations. Thus, low CpG LINE L1 hypomethylation *per se* could be used as an indicator of this phenotype. Further hypomethylations were found in other repeat regions (DASKALOS 2009, FEBER 2011, GAUDET 2003), mainly within LTRs and satellite regions. Alu-elements showed fewer hypomethylations than expected (**Figure 20E, Table 28**).

We found many hypermethylated regions to be common to all tumour samples. This could be explained by events happening early in prostate tumour development – such as *GSTP1* hypermethylation that can be found in PIN and even PIA (proliferative inflammatory atrophy) lesions (NAKAYAMA 2003) – or by common mechanisms which lead to target specific hypermethylations. These common alterations were responsible for the separation of tumour and normal samples in principal component analyses (PCA). Furthermore, our data also revealed that in addition to general tumour alterations there seems to exist a second mechanism causing separation of FUS+ and FUS- samples in PCA. In line with previous reports on gene expression (SBONER 2010, SETLUR 2008), the PCA analyses suggested that methylation patterns in FUS+ are more homogeneous than in FUS- samples. Common to the FUS- group was the extensive global hypermethylation as compared to FUS+ samples suggesting a 'methylator'-like phenotype, and the hypomethylation of LINE L1 elements (**Figure 18**).

These results are in contrast to Kim *et al*. (KIM 2011). They also showed hypomethylation of LINE L1 elements in FUS-, but not in FUS+, but their data also implicated lower overall numbers of DMRs in ETS-negative tumours than in fusion positive tumours. The discrepancy to our findings is most likely due to different technologies used. The study of Kim *et al.* is based on a MethylPlex-next generation sequencing (M-NGS) approach concentrating on high GC content regions. A comparison of M-NGS and MeDIP-Seq resulted in concordance rates of 89% in CpG island regions and 62% outside of CpG islands with more than three times more DMRs detected in MeDIP-Seq experiments (3,928 in MeDIP-Seq and 1,274 in M-NGS in CpG containing promoter regions) (KIM 2011). This low number of detected DMRs within M-NGS might be due to a restriction to regions targetable by the enzymes 'Mrse1' and 'Mrse2'. Furthermore, the M-NGS protocol involves a specific GC enrichment step resulting in a loss of low CpG density DMRs: In our analyses we found 51% of hypomethylated regions and 26% of all DMRs between

FUS- and FUS+ in regions with a CpG content of <8CpG/500bp – regions which might have been missed by Kim *et al*. Finally, we have used a more than five times higher number of samples: Kim *et al.* used only nine tumour and eight normal/benign samples while we compared 53 tumour and 51 normal samples.

Among the hypermethylated promoter regions we found miRNA, homeobox (HOLLAND 2007), and cancer census genes (FUTREAL 2004) as main targets in tumours. The hypermethylation was pronounced in FUS- samples in comparison to FUS+ samples.

MiRNAs are regulators of gene expression (BARTEL 2009) and are commonly deregulated in human cancer (LEE 2009). Methylation of miRNA promoters can impair miRNA expression (HAN 2007), resulting in an increased expression of miRNA target genes.

Homeobox genes are involved in tissue development and cell differentiation processes where they act as regulators of complete gene sets (SCOTT 1989). Hypermethylation of these genes causing aberrant expression might be involved in de-differentiation of the affected cells. Hypermethylation of *HOXB13* (SCHULZ 2006), an AR signalling repressor with tumour suppressor properties, and *HOXD3* (KRON 2009), a gene associated with TGFβ signalling, have been described in prostate tumours causing a suppression of these genes. Homeobox genes are also often found hypermethylated in other tumours like breast cancer (TOMMASI 2009), lung cancer (RAUCH 2007), or cholangiocarcinoma (SHU 2011), indicating the central role of this mechanism in tumour formation.

A compendium of genes associated with cancer initiation or progression was created by Futreal *et al.* (FUTREAL 2004). This cancer gene census list contains oncogenes as well as tumour suppressor genes which we found to be extensively differentially methylated and expressed in our data. Hypermethylation of tumour suppressors can lead to tumour initiation and evasion of cell death since tumour suppressor genes involve cell cycle check point or pro-apoptotic proteins.

4.3 Marker detection

One of the aims of our study was the identification of differentially methylated genomic regions that can be used for early PCa detection in clinical practice, potentially even for diagnosis of PCa in urine or blood. Therefor, the markers should be reliably significantly differentially methylated in tumour tissues, i.e. that all tumour samples have higher or lower methylation values than all normal samples.

Of a list of markers described to be differentially methylated in PCa – reviewed by Li *et al.* (LI 2004) – we could validate 15 of 21 (**Table 30**). The ranks of these markers in **Table 30** show that we found a large number of additional markers that perform equally well or even better in separating normal from tumour.

For our further analyses we concentrated on the 110 most significantly differentially methylated regions in our data set. Interestingly, they were all hypermethylated in PCa (**Supplementary Table 3**). Two markers of Li's list were found among these best 110 markers: *glutathione-S-transferase P1* (*GSTP1*, on position 108) and *retinoic acid receptor β* (*RARB*, position 36). These markers have already been described in a variety of studies (HOPKINS 2007, LEE 1994, NAKAYAMA 2003, NAKAYAMA 2001, YAMANAKA 2003). *GSTP1* is an enzyme that catalyses the conjugation of glutathione to toxic electrophilic compounds. Furthermore, it is involved in central apoptotic and cell cycle pathways. It is upregulated in several types of cancer and cancer cell lines, potentially involved in resistance to chemotherapies and inhibition of the *JNK* signal transduction pathway (ELSBY 2003, LABORDE 2010). Nevertheless, in prostate cancer the suppression of *GSTP1* through hypermethylation is one of the earliest events, already occurring in proliferative inflammatory atrophies (PIAs) (NAKAYAMA 2003).

The other marker already described in the literature, *RARB*, is a nuclear receptor that forms dimers upon activation through retinoic acid (DAFTARY 2006), translocates to the nucleus and modulates transcription of genes carrying a retinoic acid response element (RARE). Interestingly, *RARB* itself is regulated by a RARE (DE THE 1990, HOFFMANN 1990). Retinoic acid, the natural ligand of *RARB*, plays a crucial role during embryonic development and is supposed to regulate homeobox genes including *HOXA1*, *HOXB1*, and *HOXD4* (DAFTARY 2006). These genes were also found to be hypermethylated in our tumour samples. It is noteworthy that we have also found *aldehyde oxidase 1* (*AOX1*) catalysing the biosynthesis of retinoic acid from retinal (HUANG 1994, SIGRUENER 2007) among the most significantly suppressed genes (position 38 in most significantly differentially methylated bins; correlation between gene expression and methylation = -0.78; **Figure 28**), further supporting a decreased *RARB* function. This might in turn enhance the homeobox gene deregulation and indicates that deregulation of the *RARB*/homeobox gene axis might play an important role in PCa development.

We performed BS-MS analyses (EHRICH 2008, RADPOUR 2008) of 23 of the best 110 potential marker regions (**Supplementary Table 3**) and were able to clearly separate the

tumour from the normal samples. This further demonstrates the applicability of these regions as biomarkers. Nevertheless, the proposed markers need further validation in independent sample cohorts, ideally in a double blinded study.

Although BS-MS is a robust assay with most variance resulting from the BS conversion step and region specific PCRs (EHRICH 2007) it has shown 4% background methylation in a completely unmethylated region (*GHSR*, marker 29). Additionally, Ehrich *et al.* have shown variances in methylation of at least 5% in MALDI-TOF analyses of the BS converted fragments (EHRICH 2007). Thus, to detect small amounts of methylated DNA in a background of unmethylated DNA it might be not sensitive enough. Here, methylation specific quantitative real time PCR (qMSP) provides a more sensitive method for the detection of traces of tumour DNA in body fluids like urine (HOQUE 2005, PAYNE 2009). We designed primers for qMSP assays to assess the percentage of methylation in single informative CpGs: Using the BS-MS results we selected single CpG groups in three promising marker regions (marker 1 (*PTPRN2*), marker 2 (*CAV2*), and marker 74 (*ZSCAN12*)) showing highest absolute differential methylation between tumour and normal samples, i.e. methylation in tumour samples was >60% while methylation in normal samples was near 0%. Using completely unmethylated DNA we could determine low false discovery rates of below 1% for the primer sets specific for methylated DNA. Assuming 50% methylation in the target CpG in tumour and 1% methylation in normal DNA these assays allow detection of as few as 5 % tumour content in the DNA.

Despite the low number of samples in sample set 1 we could show that only cellular DNA was suitable for the assay while circulating DNA failed to separate the groups. This might be due to the higher tumour cell content in cellular DNA because cells exfoliated during the prostate massage step most likely originate from prostatic epithelial tumours or lesions, while cell free DNA originates from all epithelial cells of the urinary tract and is most likely released from apoptotic cells. Furthermore, circulating DNA is mainly fragmented to sizes of 150-400bp (BRYZGUNOVA 2006). Further fragmentation during BS conversion steps might decrease the amount of complete template DNA fragments for PCR amplification to a non-determinable minimum. Most studies in the literature assay methylation values in urine or prostatic fluid with either total (PAYNE 2009) or cellular DNA (CROCITTO 2004, GOESSL 2001, HOQUE 2005, ROUPRET 2007). Only Bryzgunova *et al.* used DNA from cell free supernatant (BRYZGUNOVA 2006).

Due to the overall small number of samples we have used for the assays the results are preliminary, but already reveal the high potential of our assays. Using ROC optimisation of our thresholds we could achieve specificities of 100% and sensitivities of up to 57%. In comparison, Goessl *et al.* (GOESSL 2001) achieved a sensitivity of 73% and a specificity of 98% assaying the methylation of *GSTP1* in urinary DNA.

One crucial step in the protocol that needs further improvement is the maximisation of the DNA amount isolated from the exfoliated cells to circumvent the potentially biased PCR amplification step preceding qMSP. Alternatively, it is desirable to increase the tumour cell content of the samples. Increasing the specificity of the primers by sequence modification, more precise determination of the annealing temperature or alterations in the buffer concentrations also would allow a better determination of 'true' tumour methylation against a strong normal background. Furthermore, during a screening study further marker combinations that have shown their potential in separating tumour and normal samples in linear combinations (**Figure 23**) should be tested. Further experiments using larger sample cohorts in a double blinded study will show whether the markers can be used for detection of tumour DNA in urine and thus provide a non-invasive means for diagnosis of PCa.

4.4 Gene expression and DNA methylation

4.4.1 Differential expression: tumour vs. normal

Besides our methylation data we had access to microarray expression data (from Holger Sültmann, DKFZ) of the same tumour and normal samples that we had already used for the methylation analyses of more than 17,000 genes. In this data set we found more than 60% of the genes altered in PCa with equal numbers of genes being suppressed and upregulated. We focussed our analyses on gene sets which we had identified to be most severely affected by differential methylations: Homeobox, cancer census and G-protein coupled receptor (GPCR) genes. The majority (61%) of homeobox genes was repressed in PCa. This finding again indicates the central role of homeobox genes in cancer development and is in line with the finding of a de-differentiated phenotype in tumour cells. In the cancer census genes we found a totally different picture: the majority of genes was upregulated. We did not observe a difference in the numbers of increased and suppressed genes between the oncogene and the tumour suppressor classes.

Although genes of the G-protein coupled receptor pathways appeared among the top hypomethylated features we found less increased expression than expected in the GPCR

group. Furthermore, more than two thirds of the GPCR genes were hypermethylated. Alterations in the GPCR signalling pathways play a principal role in the transition of androgen dependent PCa to androgen independent growth (LEE 2001). Among the hypomethylated GPCRs we found *follicle stimulating hormone receptor* (*FSHR*, $p_{BH}<1.2*10^{-4}$). It is described to be overexpressed in PCa and in the androgen refractory cell lines DU145 and PC3, and has been associated with androgen independent growth (BEN-JOSEF 1999). In our samples it did not show an increased expression in tumours. Nevertheless, hypomethylation of GPCR genes might precede their overexpression in late PCa. Additionally, other epigenetic modifications like histone deacetylations or methylations or the absence of the adequate transcription factors potentially abrogate the effects of DNA hypomethylation on gene re-expression (SPROUL 2011).

Further evidence for the importance of olfactory receptors in PCa was given by Neuhaus *et al.* (NEUHAUS 2009), who showed that they can be activated by steroids in the prostate. They described an upregulation of the olfactory receptor *OR51E2* in PCa. Interestingly, activation of this receptor by steroids like 6-dehydrotestosterone (6-DHT) – a testosterone metabolite – leads to growth inhibition (NEUHAUS 2009), thus the receptor shows tumour suppressor properties. We could recapitulate an upregulation of *OR51E2* in our expression data set ($p_{MW}<2.9*10^{-12}$). Furthermore, we found the promoter bin slightly hypomethylated ($p_{MW}<7*10^{-3}$), suggesting an epigenetic cause for the increased expression (**Figure 30A**). The question now is why the upregulation of *OR51E2* does not inhibit PCa growth. A possibility might be that for the tumour suppressive activity also 6-DHT is required. The cytochrome oxidases of the CYP3A family have been described to catalyse the oxidation of testosterone to 6-DHT (HALVORSON 1990). Downregulation of *CYP3A4* has been correlated with poor clinical outcome and Gleason score (FUJIMURA 2009). In our data set we found all members of the CYP3A family – with available expression data (*CYP3A4*, *5*, *7*, and *43*) – downregulated in tumour. For *CYP3A5* and *CYP3A7* we additionally determined significant correlations between promoter hypermethylation and downregulation (**Figure 30B**). Thus, it is tempting to speculate that – despite the increased expression of growth inhibitory *OR51E2* – the cell cycle of PCa is not affected due to a decreased level of the metabolite 6-DHT. This regulation is a good example of the cancer hallmark 'Evading growth suppressors' (HANAHAN 2011). Nevertheless, the *OR51E2* pathway could be exploited by treatment with 6-DHT or other *OR51E2* ligands (NEUHAUS 2009).

4.4.2 Differential expression: FUS- vs. FUS+

While the fractions of upregulated and suppressed genes in tumour are almost equal (5,038 down- and 5,839 upregulated), we found almost two times more genes with a higher expression in FUS+ than in FUS- samples (1,190 vs. 409). Furthermore – comparing the gene expressions of the tumour subclasses to normal gene expressions – we found that almost 75% of the genes that were differentially expressed only in FUS+ samples exhibited an increased expression. In FUS- samples the picture was converse. Here, more than 66% of the genes exclusively differentially expressed in this class were suppressed. Additionally, FUS+ samples exhibited nearly twice the number of exclusively differentially expressed genes as FUS- samples (696 vs. 388).

These findings might be explained by the following model: Tumour progression in FUS+ samples relies on the elevated expression of the transcription factor *ERG* leading to an increased expression of further genes like genes of the cancer census list. In FUS- tumours on the other hand widespread hypermethylations occur (‘methylator’ phenotype) which may induce permanent silencing of already not expressed genes, thus making them unavailable for the emergency responses or for proper cell differentiation and function.

Among the cancer census genes we found nine oncogenes and five tumour suppressors with an increased expression and four genes that were suppressed in FUS+ samples compared to FUS- and normal samples. Upregulation of the tumour suppressors might be used as a cell defence mechanism counteracting tumour formation in FUS+. In FUS- samples on the other hand we found higher promoter methylation levels in four of the five tumour suppressor genes overexpressed in FUS+ (*CBLC*, *CDH1*, *GPC3*, *SETD2*) potentially suppressing the expression of those genes (**Figure 25**). This would be in accordance with the model of epigenetically blocked transcription of antitumourigenic pathways.

4.4.3 Correlation of methylation and gene expression

A negative correlation between promoter methylation and gene expression has been described in many publications assessing DNA methylation and gene expression for single genes (CARVALHO 2010, DOBROVIC 1997, HERMAN 1998, HIBI 2008, NAKAYAMA 2001). Promoter methylation can interfere with transcription factor binding and thus can suppress gene expression (WATT 1988). Another mechanism of methylation mediated

transcriptional silencing involves binding of methyl-DNA binding proteins to methylated DNA. These proteins either can recruit histone deacetylases leading to condensed chromatin or can prevent transcription factor binding as well (reviewed in (BIRD 1999)). Furthermore, chromosome-wide DNA methylation plays a role in X chromosomal inactivation in female mammals. However, here DNA methylation itself is not causative for primary gene repression on the inactive X chromosome (X_i). Repression is mediated on one of the X chromosomes through cis-acting *Xist* RNA during uterine implantation of the embryo. Methylation patterns involving hypermethylation of CpG islands are then established in the course of X_i silencing (RIGGS 1992, WUTZ 2000). These established repression patterns are stable during lifetime. 5-aza-2`-deoxycytidine treatment leads to a reactivation of suppressed genes suggesting synergistic suppressive mechanisms of DNA methylation and the *Xist* gene product (CSANKOVSZKI 2001, JAENISCH 2003).

Additionally, DNA methylation is involved in repression of imprinted genes on autosomes during embryogenesis (IDERAABDULLAH 2008). In normal tissue the promoter regions of most genes are unmethylated with the exception of a few promoter associated CpG islands which become methylated in early embryonic cells. Here again, methylation of these genes is now seen to be a result of, and not a reason for, downregulation – i.e. non-expressed genes are subjected to DNA hypermethylation (BIRD 2002). Nevertheless, transcriptional regulation does not necessarily imply changes in the DNA methylation pattern in normal cells: genes that are not expressed do not have to be methylated (BIRD 2002). Taken together, in normal cells methylation seems to contribute to an irreversibility of transcriptional silencing of already not expressed genes (WUTZ 2000). Interestingly, in a few cases hypermethylation can also provoke the expression of a gene: In case of *IGF2* imprinting, methylation of a *CTCF* binding site represses binding of *CTCF* resulting in the restoration of the interaction of an enhancer element with the promoter, with the consequence that the methylated *IGF2* becomes expressed (BELL 2000).

Cancer cells on the other hand show significantly altered patterns of methylation, and cancer specific hypermethylations of several genes have been associated with transcriptional downregulation (EGGER 2004, FEINBERG 2006, JONES 2002). Furthermore, genes with repressive chromatin marks like the H3K27me3 Polycomb mark can be hypermethylated during tumourigenesis, reducing their transcriptional plasticity by locking them in the repressed state (GAL-YAM 2008, VIRE 2006). Thus, cancer exploits embryonic mechanisms of gene silencing through hypermethylation (see above) (WIDSCHWENDTER 2007, WUTZ 2000). This is in line with the model of tumour cells recapitulating features of

early embryonic cells. Nevertheless, Raynal *et al.* could re-express silenced genes using histone deacetylase inhibitors despite of persistent DNA hypermethylations of the genes, indicating that a temporal change of histone marks can momentarily outweigh repressive effects of DNA methylation (RAYNAL 2012). However, two weeks after treatment the re-expressed genes all had returned to their silent state. Thus, it is thought that DNA methylation serves as a long-term memory of gene silencing whereas histone modifications can alter short-term gene expression levels.

In our tumour samples we found an upregulation of the *de novo* DNA methyltransferase 3A (*DNMT3A*), a DNMT usually expressed in embryonic tissues (OKANO 1999), and of the maintenance DNA methyltransferase *DNMT1* (LEONHARDT 1992) (**Figure 33**).

Overexpression of *DNMT3A* might be responsible for methylation of normally unmethylated promoters and explains the observed enrichment of promoters and CpG islands among the hypermethylated features. *DNMT3A* expression can be regulated by miRNAs of the *miR29*-group (FABBRI 2007). Indeed, we found *miR29a* and *miR29b* downregulated in tumour samples **(Figure 34A),** suggesting this mechanism of *DNMT3A* control being disturbed in our samples.

A further chromatin modifying enzyme associated with DNA methylation, the histone methyltransferase *EZH2*, was also found overexpressed in our tumour samples. *EZH2*-affected genes can be subjected to DNA methylation for permanent silencing (VIRE 2006, WIDSCHWENDTER 2007). In this line we found exclusive hypermethylation in 60% of cancer associated *EZH2* target genes (YU 2007). *EZH2* has been associated with high grade PCa (KARANIKOLAS 2010, VAN LEENDERS 2007, YU 2010) and is also found increased in many other tumor types (ALAJEZ 2010, ZHANG 2011a).

Up to this part of the work we have examined potential connections between differential methylation and altered gene expression only for a few single genes. To elucidate global mechanisms of gene regulation by DNA methylation in prostate tumours we performed correlation analyses integrating expression and methylation values of 17,163 genes. This is one of the first times that this kind of analyses has been performed on a large cohort of samples in PCa. Recently published studies integrated epigenetic and gene expression data of 47 primary breast tumours (SPROUL 2011), of 59 primary neuroblastoma tumours (CAREN 2011), of 55 glioblastoma and five non-tumour brain samples (ETCHEVERRY 2010), or of nine renal cell carcinomas and three normal kidney samples (MORRIS 2011).

In our analyses, we have shown that hypermethylation mainly affected a region of ±2kb around the TSS (**Figure 20C**). In highly expressed genes this region normally is only

weakly methylated while the gene body is extensively methylated (BALL 2009). Thus, we concentrated on the 4kb region around the TSS of each transcript to perform correlation analyses. We found that 12% (1,986) of all genes showed a correlation between gene expression and gene methylation and 91% (1,806) of significantly correlating genes were differentially expressed, suggesting mutual effects. In the downregulated fraction we found 92% (878 of 953) of correlated genes with hypermethylation (negative correlation), while 80.5% of upregulated genes (687 of 853) were also hypermethylated in the best correlating bin (**Figure 27A**). We found a higher fraction of hypermethylations in downregulated genes. Thus, assuming a negative correlation between gene expression and DNA methylation, a subset of genes might indeed be suppressed by DNA hypermethylation. Sproul *et al.* have demonstrated in breast cancer cells that about 16% of methylated genes could be derepressed by 5-aza-2`-deoxycytidine treatment and thus are directly controlled by DNA hypermethylations (SPROUL 2011).

Stratification of the differentially expressed genes showed that weakly expressed genes are mainly suppressed and highly expressed genes are mainly upregulated in tumour. Nevertheless, both groups are more likely to be suppressed when hypermethylated. This holds particularly true for normally highly expressed genes (**Figure 27B**).

However, expression of most genes is independent of DNA methylation i.e. that although being hypermethylated a gene can become highly expressed (RAYNAL 2012). Furthermore, hypomethylation does not necessarily lead to re-expression of a silent gene if the adequate transcription factors are missing (SPROUL 2011). Additionally, Coolen *et al.* have found long regions in PCa that consist of consecutive, silenced genes (long-range epigenetic silencing (LRES) regions) (COOLEN 2010). Thirty percent of these regions contain cancer associated genes and miRNAs – genes that are strongly affected by differential methylation and expression in our data. Genes of these regions either are hypermethylated or carry repressive chromatin marks. Thus, for a subgroup of downregulated genes in our data the downregulation could be explained by their localisation in LRES regions despite of the absence of a direct hypermethylation.

Possible explanations for the positive correlations between upregulation of genes and hypermethylation might be that DNA methylation has to reach a certain density before becoming inhibitory on gene transcription (TATE 1993), that many genes are expressed despite of being methylated unless they are carrying repressive histone marks (RAYNAL 2012), or there might even exist transcriptional repressors that are excluded from DNA

binding through DNA methylation. Thus, for a comprehensive picture of epigenetic regulation of gene expression it is essential to include analyses of the histone marks.
In addition to our analyses on the influence of DNA methylation on gene expression, integration of transcript-wise gene expression data will further refine the model. Here, a transcript-wise analysis was not possible since the expression data was condensed to gene-wise expression values due to the analysis strategy of microarray data. Using the 4kb region of all transcripts of one gene for correlation analyses the investigated region can stretch from nine to several hundred bins. Using expression values of the whole gene and extracting only the most significant correlation values we might have missed a negatively correlating TSS bin that showed a slightly less significant correlation than the positively correlating bin of another gene associated transcript. Furthermore, we also might have missed a DNA methylation induced alternative promoter usage which might not have shown an impact on gene expression values in our correlation analyses.

For single selected genes like *DUOX1, CAV2, AOX1* and *GSTP1*, we demonstrated highly significant negative correlations between gene expression and DNA methylation (**Figure 28**). Since we were especially interested in long term regulatory effects underlying PCa in general, we will stick to this model of inverse correlation for further discussions.
Of the gene sets analysed homeobox genes were especially impacted by DNA methylation and gene expression changes. These genes showed a fivefold enrichment of genes with negative correlations to DNA methylation suggesting hypermethylation as predominant mechanism of homeobox gene suppression (**Table 34**). At least a subfraction of GPCR genes was also potentially regulated through hypermethylation. For cancer census genes in general, a less strong connection between hypermethylation and downregulation was visible. This implies a gene regulation through transcription factors rather than through differential methylation.
However, using this kind of analyses it is not possible to differentiate between cause and effect, nevertheless, they can give clues to potential regulatory mechanisms. Following the hypothesis that mainly not expressed genes are affected by methylation and thus can be permanently shut down in embryonic cells (WUTZ 2000) or exhibit a reduced transcriptional plasticity (COOLEN 2010, GAL-YAM 2008), the following model of promoter hypermethylation is reasonable in tumour cells: Tumour suppressor genes like genes of the caspase cascade (CHOWDHURY 2006) are not active but unmethylated in normal cells. Following an apoptotic stimulus pro-apoptotic genes like *BBC3* (better

known as *PUMA*: *p53-upregulated modulator of apoptosis*) are upregulated (WANG 2007). Selective mechanisms lead to hypermethylation of the gene promoters in tumour cells hampering a normal transcriptional response – as it has been described for genes of LRES regions (COOLEN 2010) – and thus inhibiting apoptosis. Indeed, we found tumour associated hypermethylations in a variety of apoptosis related genes like the *BBC3,* the *PYCARD* or the *harakiri* (*HRK*) gene (**Table 40**) (CHOWDHURY 2006, INOHARA 1997). Even more, gene expression was slightly but significantly downregulated in these three genes (**Figure 41**).

Table 40: Differential methylations of apoptosis associated genes. Column headings are as follows: "TSS bins" denotes the number of TSS associated bins, "DMRs" gives the number of differentially methylated TSS bins and "DMR HYPER" and "DMR HYPO" denote, how many of these are hyper- or hypomethylated in tumour, respectively. "p_{MW} EXP" shows the p_{MW}-value for differential expression of this gene. If "Downregulated" = TRUE, the respective gene is significantly downregulated in tumour. Hypermethylated genes are highlighted in gray.

Name	TSS bins	DMRs	DMR HYPER	DMR HYPO	p_{MW} EXP	Downregulated
AIM2	34	3	0	3	6,9E-02	FALSE
APAF1	9	0	0	0	8,1E-03	FALSE
BAD	20	0	0	0	5,3E-02	FALSE
BAK1	9	0	0	0	2,9E-01	FALSE
BAK1	9	0	0	0	2,9E-01	FALSE
BAX	12	1	1	0	9,6E-08	FALSE
BBC3	**13**	**4**	**4**	**0**	**4,4E-06**	**TRUE**
BCL2L11	27	7	7	0	9,9E-09	FALSE
BIK	9	1	1	0	3,5E-12	FALSE
BMF	15	0	0	0	3,1E-02	FALSE
BOK	9	0	0	0	2,6E-02	FALSE
CASP1	18	1	0	1	5,8E-05	TRUE
CASP2	22	0	0	0	8,0E-14	FALSE
CASP3	20	0	0	0	3,3E-01	FALSE
CASP4	21	0	0	0	8,5E-01	FALSE
CASP5	18	1	0	1	9,1E-01	FALSE
CASP6	25	1	1	0	4,7E-01	FALSE
CASP7	42	1	1	0	4,6E-01	FALSE
CASP8	46	3	2	1	3,5E-09	FALSE
CASP9	23	0	0	0	5,8E-05	FALSE
CASP10	18	0	0	0	2,4E-01	FALSE
CASP12	9	0	0	0	7,2E-02	FALSE
CASP14	15	4	0	4	1,4E-02	TRUE
DFFA	18	0	0	0	1,2E-10	FALSE
DIABLO	26	0	0	0	6,3E-06	FALSE
FADD	9	1	1	0	6,9E-01	FALSE
HRK	**9**	**2**	**2**	**0**	**7,2E-03**	**TRUE**
PARP1	42	4	4	0	5,6E-09	FALSE
PMAIP1	9	0	0	0	1,1E-02	TRUE
PYCARD	**9**	**2**	**2**	**0**	**6,1E-12**	**TRUE**
TRADD	16	3	3	0	7,6E-01	FALSE

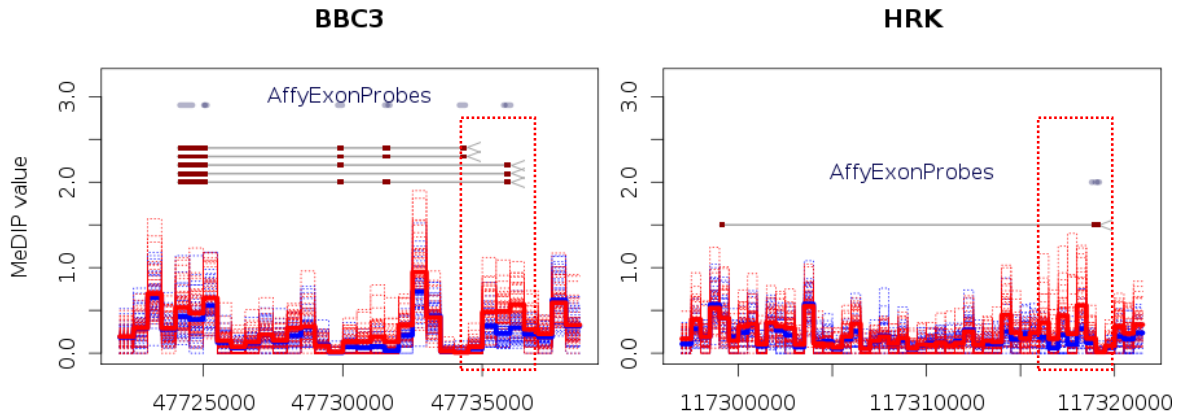


Figure 41. Methylation plots of the downregulated pro-apoptotic genes *BBC3* (*PUMA*) and *HRK* (*harakiri*) reveal extensive hypermethylation of promoter associated bins in tumour. Blue lines mark normal, red lines tumour samples. Solid lines represent the mean values of the respective sample group.

4.5 Correlation analyses of methylation and miRNA expression

For the correlation analyses of miRNA expression and DNA methylation we did not have to include different transcript isoforms. This and the fact that we thus only had to use nine associated bins for each miRNA (miRNA ±2kb) made our analyses easier than correlations of gene expression and methylation (see section 4.4).

Unlike differential gene expressions with nearly the same numbers of up- and downregulations, 191 differentially expressed miRNA genes (60%) were downregulated compared to 127 upregulated miRNA genes. We had found hypermethylations to be enriched in miRNA genes in tumours and indeed, we determined significant negative correlations between expression and methylation for 82 differentially expressed miRNAs: 65 were downregulated and hypermethylated, while 17 were upregulated and hypomethylated.

Among these genes, *miR184* and the *miR23b/miR24/miR27b* cluster showed the most significant negative correlations with significant downregulation in tumour samples (**Figure 31**). Pro-apoptotic functions of *miR184* have been described in neuroblastoma cell lines (CHEN 2007, TIVNAN 2010), its upregulation, on the other hand, is associated with squamous cell carcinoma progression (WONG 2008). Interestingly, Liu *et al.* have demonstrated the involvement of *miR184* in cell differentiation processes and even showed an epigenetic regulation of this miRNA. They also found a proliferative function in adult neuronal stem cells (LIU 2010a). Furthermore, Chen *et al.* could show that induction of differentiation using all-trans retinoic acid led to overexpression of *miR184* (CHEN 2007).

Since we found *miR184* being suppressed and hypermethylated in our data, this further underscores the importance of alterations in differentiation processes in PCa formation.
For *miR23b/miR24/miR27b* the literature data is equally controversial: *MiR23b* is generally downregulated and has inhibitory effects on cell migration, invasion, growth, and angiogenesis in colon cancer (ZHANG 2011b), while in the retina *miR23* and *miR27* are associated with angiogenesis and vascularisation (ZHOU 2011). In renal carcinoma upregulation of *miR23* has been described to target tumour suppressor genes (LIU 2010b). Thus, it appears that deregulation of miRNA patterns can cause converse effects depending on the tissue background. Nevertheless, in our samples we found a highly significant downregulation and correlating hypermethylation of the *miR23b/miR24/miR27b* cluster. Using a luciferase system we could show that hypermethylation is indeed functionally relevant for suppression of this cluster. Experiments with prostate cancer cell lines using *miR23b/miR24/miR27b* mimics or the complementary experiment using *miR23b/miR24/miR27b* antagomirs in a normal prostate cell line would help to elucidate the function of these miRNAs in prostate cancer formation.

Comparing FUS- and FUS+ samples we determined a higher number of downregulated miRNAs in FUS- (164 vs. 139). Furthermore, we found that FUS- samples contained 98 exclusively differentially expressed miRNA genes in comparison to 50 in FUS+ samples, and almost four times more negative correlations to DNA methylation (74 vs. 16).
These results resemble the differences in methylation patterns in the tumour subgroups where FUS- samples showed significantly more differential methylations than FUS+ samples. Furthermore, we have observed miRNAs as enriched targets of FUS- specific hypermethylation suggesting differential methylation and expression of miRNAs being another key pathway for tumour formation in fusion negative prostates.

4.6 A model for prostate cancer formation

4.6.1 The central role of *EZH2*

As mentioned above we had found *DNMT3A* and *DNMT1* upregulated in tumour samples – similarly for FUS+ and FUS- tumours, resembling the embryonic expression pattern of these genes. We also found *EZH2* upregulated in tumours, but this increase was more

pronounced in FUS- samples, thus making *EZH2* a candidate responsible for the pronounced differential DNA methylation in FUS-.

EZH2 is a H3K27-methyltransferase with highest activity in embryogenesis. It links repression of developmental genes through histone methylation with DNA methylation (CAO 2002, KUZMICHEV 2004, VARAMBALLY 2002, VIRE 2006). We thus investigated whether differential *EZH2* expression was reflected in DNA expression and methylation patterns of its target genes. Developmental genes were enriched among the hypermethylated genes in our data set, suggesting that *EZH2* overexpression might be causative for the hypermethylations observed. We analysed expression and methylation of a set of *EZH2* target genes described to be hypermethylated in metastatic prostate cancer (YU 2007). We found 60% of the target genes hypermethylated. Additional 8% were hypermethylated but also showed signs of hypomethylation. These numbers indicate a direct influence of increased *EZH2* in tumours on the methylation state of this group of genes. Furthermore, FUS- samples – which showed an even more elevated expression of *EZH2* – also exhibited higher methylation values of *EZH2* target genes. These findings are underscored by an almost eightfold higher number of *EZH2* target genes with a significant correlation between *EZH2* expression and their promoter methylation in FUS- samples than in FUS+ samples. Since *EZH2* expression is described to cause a downregulation of the expression of its target genes we further investigated if we could find differences in the expression profiles of *EZH2* target genes in FUS- and FUS+ samples. Indeed, the expression of *EZH2* target genes was significantly more altered in FUS- samples than in FUS+ samples.

The same observations were made in homeobox genes – the main targets of *EZH2* mediated cellular effects: We found two times more genes with significant methylation correlating to *EZH2* expression in the FUS- class than in FUS+ (192 vs. 65).

These results support the notion that deregulation of methylation and gene expression patterns in prostate tumours might be severely influenced by aberrant *EZH2* expression. Furthermore, differences in *EZH2* expression between FUS+ and FUS- samples might even explain the differences in DNA methylation and gene expression between these two classes of PCa and hint at the underlying mechanisms of tumour formation.

4.6.2 Regulation of *EZH2*

Next, we dissected the potential causes of *EZH2* upregulation in both tumour subgroups. Increased *ERG* expression causes *EZH2* upregulation in FUS+ samples (KUNDERFRANCO 2010, YU 2010), and knocking down *ERG* in VCaP fusion positive cells consequently led to a reduction in *EZH2* mRNA levels (**Figure 37**). In FUS- samples we did not observe elevated *ERG* levels, thus another mechanism underlies *EZH2* upregulation. We found no changes in the DNA methylation pattern – the promoter was unmethylated in normal as well as in both tumour groups. Next, we turned our attention to regulatory miRNAs.

Six miRNAs have been suggested to regulate *EZH2* expression: *miR26a* (ALAJEZ 2010, LU 2011), *miR101* (VARAMBALLY 2008), *miR138* (KISLIOUK 2011), *miR124* (ZHENG 2011), *miR214* (JUAN 2009), and *let-7b* (TZATSOS 2011). *MiR214* and *miR138* were suppressed in general in tumours; this may contribute to the general upregulation of *EZH2* in tumour. *MiR124* was upregulated in tumours, thus potentially suppressing *EZH2*. However, this miRNA was more than 30fold less expressed than *miR138* and more than 4,000fold less than *miR214*. Thus, the relative expression changes of *miR124* are high, but are not reflected by a high absolute change in expression.

We found *miR26a* as the only *EZH2* targeting miRNA that exhibited significant expression differences between FUS+ and FUS- samples with an exclusive downregulation in FUS-. *MiR26a* was also found suppressed in other tumours like breast cancer (ZHANG 2011a), nasopharyngeal carcinoma (HOFFMANN 2007, LU 2011) and lymphomas (SANDER 2008), revealing its central role in tumour development. Correlation analyses of *miR26a* and *EZH2* expression as well as functional experiments with *miR26a* mimics further validated the regulation of *EZH2* by *miR26a*. These results were reflected in the *miR26a* and *EZH2* expression levels of fusion positive and negative cell lines (**Figure 37**). Furthermore, using a luciferase methylation assay we could prove that hypermethylation of a 2kb region in the vicinity of *miR26a* can induce *miR26a* downregulation. This region was indeed found to be hypermethylated in FUS- but not in FUS+ samples, which showed a normal *miR26a* expression level. Furthermore, 5-aza-2`-deoxycytidine treatment of PC3 cells which show the highest levels of hypermethylation in this region could reconstitute *miR26a* expression and consecutively suppress *EZH2*.

We derived a model of *EZH2* regulation in FUS+ and FUS- samples that integrates our findings (**Figure 42**): In FUS+ tumours the *ERG* upregulation leads to an upregulation of *EZH2*. In FUS- tumours *miR26a* is methylated and thus suppressed. The suppression of

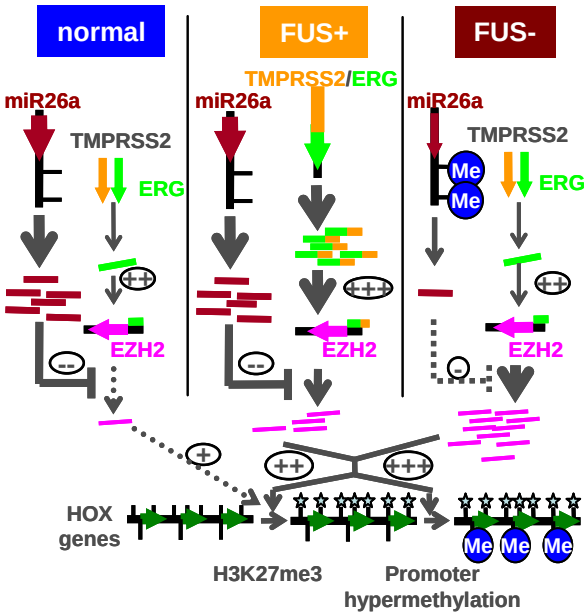


Figure 42. Model for *EZH2* expression and consecutive homeobox gene promoter methylation in normal, FUS- and FUS+ prostate cancers. In normal samples *EZH2* expression is controlled by *miR26a*. Fusion positive samples show an overexpression of *ERG* targeting the *EZH2* promoter and thus leading to an *EZH2* increase. In fusion negative samples on the other hand a region near *miR26a-2* is hypermethylated and *miR26a* expression is decreased resulting in an increase of *EZH2*. *EZH2* in turn suppresses homeobox gene expression by methylation of H3K27 and finally even manifests this silencing by DNA hypermethylation of homeobox gene promoters. This effect is stronger in fusion negative samples as these show a higher expression of *EZH2* and a stronger methylation of homeobox genes.

miR26a consequently leads to increased *EZH2*. In both tumour groups, *EZH2* upregulation inflicts hypermethylations of developmental and homeobox genes which is more pronounced in FUS- samples.

Nevertheless, deregulation of *miR26a* is not necessarily the only miRNA associated mechanism in FUS- tumour development,which is demonstrated by the higher number of deregulated miRNAs in FUS- compared to FUS+ samples (**Supplementary Table 8**).

4.6.3 *ERG* knock down experiments in VCaP cells

Besides this mechanism in FUS- tumours we also wanted to dissect the mechanism of *ERG* regulation of *EZH2*. For this purpose, we performed – together with Holger Sültmanns group at the DKFZ – *ERG* knock down experiments in fusion positive VCaP cells followed by MeDIP-Seq analysis.

Tumour formation in FUS+ tumours is to a large part driven by overexpression of *ERG* leading to an upregulation of further oncogenes. Nevertheless, we also found homeobox and polycomb group target genes (HOLLAND 2007, YU 2007) hypermethylated in FUS+ tumours due to *ERG* mediated increase of *EZH2* (KUNDERFRANCO 2010), although not to such a large extend as in FUS- samples. Knock down of *ERG* thus impacted both potential causes for tumour formation: primary *ERG* overexpression and *EZH2* upregulation (**Figure 37C**).

Interestingly, despite the *EZH2* downregulation after *ERG* knock down, we detected twice as many hyper- than hypomethylations, indicating a shift to the ‘methylator’ phenotype of FUS- samples. TSS regions and therein *EZH2* target, cancer census and miRNA genes

were again enriched among the regions with altered methylation after *ERG* knock down. The last group showed the highest enrichment among the hypermethylated features, indicating that miRNA hypermethylations might be essential for tumour formation in prostate cancer cells devoid of ETS gene fusions. This assumption is further supported by detection of the *miR23a/miR24-2/miR27a* cluster among the hypermethylated regions, a cluster that also was found hypermethylated in our tumour samples.

VCaP cells have a DNA methylation pattern that probably cannot be converted to a FUS- pattern within 96 hours of *ERG* knock down. This is reflected by the decreased expression of *EZH2* which is opposed to the situation in FUS- prostate tissues bearing increased *EZH2* levels, probably caused by *miR26a* hypermethylation. Overall, *ERG* knock down might push the cells into the direction of a normal gene expression phenotype. Thus, to maintain replication potential in cell culture, cells might have obtained general tumour DNA methylation features. To further investigate this hypothesis, we examined the overlap between genes differentially methylated in prostate tumour samples and genes showing a differential methylation in the *ERG* knock down experiment. In DAVID pathway analysis 'alternative splicing' and 'alternative promoter usage' terms were dominating (**Table 38**). This might be due to the enrichment of these terms in the differentially methylated genes in the VCaP *ERG* knock down experiment. However, we also discovered a significant enrichment of genes of the 'Proto-oncogene' class (**Table 39**). These genes can induce tumour growth when mutated or overexpressed. At this point, integration of gene expression data is essential to determine if the observed alterations in the methylation pattern could be associated with changes in proto-oncogene expression or with changes in transcription initiation and splicing.

4.6.4 Further experiments

To gain further insights into the proposed mechanisms of PCa development and to further validate the findings of this thesis the following experiments would be interesting: An overexpression of an artificial *TMPRSS2:ERG* construct in the normal prostate cell line RWPE-1 would simulate the FUS+ state, which was characterised by moderate alterations in the DNA methylation patterns and upregulation of *ERG* and its oncogenic targets. Gene expression analyses would show the success of this overexpression, while DNA methylation analyses could reveal hypermethylations that are essential in FUS+ tumours additional to the overexpression of oncogenes.

For the simulation of the FUS- state that was characterised by extensive hypermethylations of *EZH2* target and homeobox genes as well as by hypomethylations of LINE L1 elements, *EZH2* could be overexpressed in RWPE-1 cells. Furthermore, *miR26a* suppression experiments using antagomirs should produce the same results. MeDIP-Seq and gene expression analyses will reveal the value of our model.

Ideally, time course analyses are performed in the above mentioned model systems to monitor the progression of alterations in the DNA methylation pattern.

Since we have found the miRNAs of the miR23/24/27 clusters to be affected by hypermethylations and consecutive downregulations further experiments investigating the functions of these miRNAs, like transfection of normal cell lines with miR23/24/27 antagomirs or rescue experiments in prostate cancer cells using miR23/24/27 mimics followed by gene expression analyses, might be of interest and probably reveal therapeutic targets.

4.7 Conclusion

Integrating MeDIP-Seq and gene expression data of almost one hundred prostate tissue samples we found different mechanisms of tumour formation in FUS- and FUS+ PCa (**Figure 43**). Both classes show common features of early embryonic development like increased expression of *de novo* methyltransferase *DNMT3A* (OKANO 1999) and consecutively common hypermethylations of *GSTP1*, *CAV2*, *PTPRN2*, and retinoic acid receptor signalling genes like *RARB*, potentially acquired early in tumourigenesis. Furthermore, hypermethylations block emergency genes of the apoptotic pathway like *BBC3*, *HRK* or *PYCARD* (CHOWDHURY 2006, WANG 2007). Differential methylations in miRNA genes like *miR23*/*miR24*/*miR27* probably support deregulation of apoptotic, cellular repair pathways or pathways controlling genomic integrity.

Approximately 50% of the prostate lesions acquire gene fusions of *TMPRSS2* and *ERG*, which brings *ERG* under an androgen regulated promoter leading to a significant increase of *ERG* expression (BASTUS 2010, TOMLINS 2005). This induces an increased expression of oncogenes like *MYC* in FUS+ lesions (SUN 2008). *ERG* mediated induction of *EZH2* unleashes a second wave of methylations leading to moderate hypermethylations of homeobox and polycomb group target genes (RAUCH 2007, TOMMASI 2009). In embryonic stem cells, suppression of homeobox genes blocks differentiation such that they remain

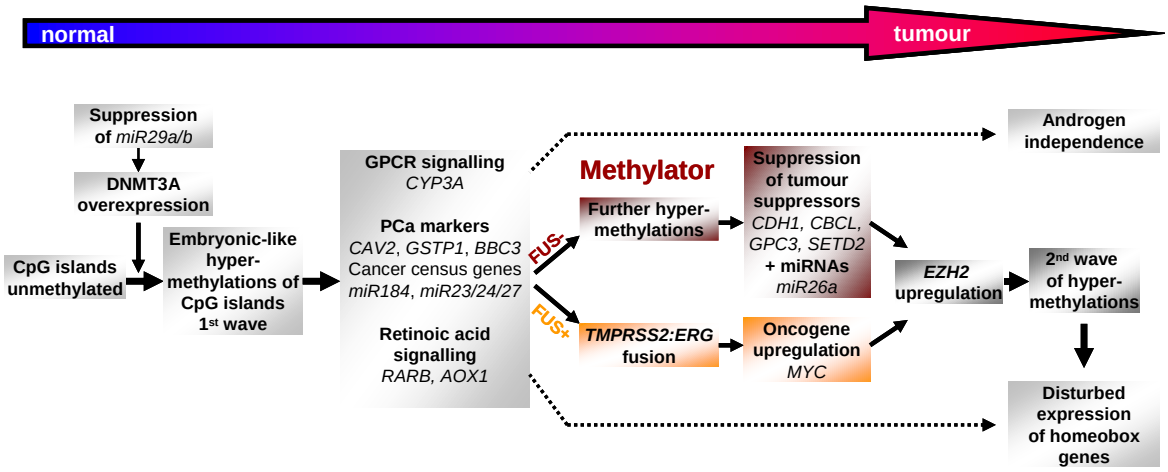


Figure 43. Model of PCa development. Proposed model of PCa development derived from our integrated gene expression and methylation data. Genes in italic font are examples of genes involved in the pathways and discussed in the text.

pluripotent (VALK-LINGBEEK 2004). This mechanism might also be exploited by tumour cells through homeobox gene hypermethylation.

FUS- lesions on the other hand acquire additional hypermethylations accompanied by demethylations of repetitive elements, resulting in a more severe differential methylation pattern compared to FUS+ lesions ('methylator' phenotype). Here, hypermethylation of tumour suppressor genes might additionally contribute to the deregulation of cancer pathways.

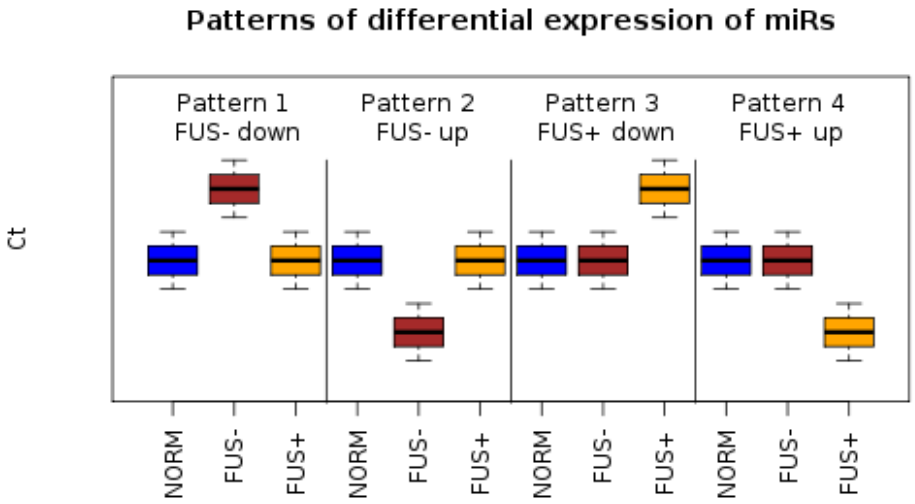
Another group of genes specifically deregulated by hypermethylations in FUS- tumours are the miRNA genes: Here, *miR26a* plays a prominent role because hypermethylation of its genomic region – resulting in its decreased expression – impairs the repression of *EZH2*, leading to an increased expression of this histone methyltransferase (LU 2011, ZHANG 2011a). The *EZH2* increase is even stronger than in FUS+ lesions and the wave of hypermethylations exhibits even more pronounced effects on methylation and expression of homeobox and polycomb group target genes than in FUS+ lesions.

Elucidation of abnormal processes underlying tumour formation and stratification of patients into tumour subgroups opens the way for personalised targeted therapeutics and for novel drug developments. FUS+ patients might benefit from *PARP1* inhibitors as suggested by Brenner *et al.* (BRENNER 2011) who showed that *PARP1* is essential for *TMPRSS2:ERG* target gene transcription and directly interacts with ETS genes. DNA methyltransferase inhibitors like the DNMT inhibitor 5-aza-2`-desoxycytidine (SAMLOWSKI 2005) or the direct *DNMT1* inhibitor procaine (VILLAR-GAREA 2003) might preferentially be useful as treatment for patients with FUS- prostate tumors (PERRY 2010).

Given the central role of *EZH2* in prostate cancer, a combination therapy with the *EZH2* inhibitor 3-deazaneplanocin A (CHIBA 2011) might be considered. Further benefits can be expected from retinoic acid treatment (NOY 2010). Retinoic acid is described to promote expression of homeobox genes (DAFTARY 2006) and is strongly involved in embryonic development (DUESTER 2008, MARSHALL 1996). In our data set we find several points of evidence that endogenous retinoic acid production is severely disturbed: We have found the promoter of *aldehyde oxidase 1* (*AOX1*) – an enzyme catalysing the biosynthesis of retinoic acid from retinal (HUANG 1994, SIGRUENER 2007) – hypermethylated, and observed a subsequent reduction of gene expression. Furthermore, we also observed that the main receptor of the retinoic acid pathway – *RARB* – was also hypermethylated and suppressed (Rho= -0.34) in tumour samples. The *RARB* gene is controlled by retinoic acid receptor elements (DE THE 1990, HOFFMANN 1990). Thus, exogenous retinoic acid could help to re-express *RARB* which might then abrogate alterations in homeobox gene expression and re-establish a differentiated phenotype.

Preceding treatment, the prostate tumour has to be detected and ideally its aggressiveness assessed. We have demonstrated the potential of three markers for non-invasive detection of PCa in urine and filed a patent application for PCa markers (MI4294). Non-invasive screening methods would reduce patient's timidity of routine prostate cancer screening and thus could lead to an earlier detection of small, maybe even not palpable tumours. Early detection on the other hand would also allow earlier, potentially less aggressive treatment. One might even envision markers which could identify patients which might benefit from a treatment and those for whom 'watchful waiting' is a good alternative. This would be beneficial for the patient and the health system – for the patient because he would not have to cope with severe side effects of castration, prostatectomy, chemo- or hormone therapy, and for the health system because nowadays 48 PCa patients are treated to save one patient's life (SCHRODER 2009) and an accurate detection would significantly reduce the number of treated patients.

5 SUPPLEMENT



Supplementary Figure 1. Expression patterns of miRNAs indicated in Supplementary Table 8. Pattern 1 and 2 symbolise significant differences between FUS- and FUS+, and between FUS- and normal samples but not between FUS+ and normal samples, representing an exclusive expression change in the FUS- subgroup. Pattern 3 and 4 symbolise significant differences between FUS- and FUS+, and between FUS+ and normal samples but not between FUS- and normal samples, representing an exclusive expression change in the FUS+ subgroup.

Supplementary Table 1: Patient statistics. Shown are the clinical data of all patients subjected to MeDIP-Seq analyses. Column headings denote the sample name, the tissue type: normal or tumour, the patient's year of birth, the pT tumour stage: for normal tissues the pT value corresponds to the tumour extracted from the patients at an independent site, the pN lymph node status, the Gleason score of tumour determined after prostatectomy (RRP Gleason), the *TMPRSS2:ERG* fusion status, the preoperative PSA level in ng/ml, and whether gene expression or miRNA data sets are available.

sample ID	tissue type	year of birth	pT	pN	RRP Gleason	TMPRSS2:ERG fusion	PSA	Gene expression data	miRNA expression data
IGP102	normal	1947	pT2c	N0	3+3	undet.	5,2	TRUE	TRUE
IGP103	normal	1948	pT2c	N0	3+3	undet.	6,3	FALSE	TRUE
IGP104	normal	1944	pT2c	N0	4+3	undet.	9,7	TRUE	TRUE
IGP119	normal	1948	pT3a	N0	3+4	undet.	1,6	TRUE	TRUE
IGP120	normal	1946	pT2c	N0	3+4	undet.	4,2	TRUE	TRUE
IGP122	normal	1948	pT3a	N0	3+4	undet.	14,2	TRUE	TRUE
IGP123	normal	1942	pT3a	N0	3+4	undet.	12,8	TRUE	TRUE
IGP124	normal	1940	pT2c	NX	3+3	undet.	6,7	TRUE	TRUE
IGP126	normal	1953	pT2a	NX	3+3	undet.	1,7	TRUE	TRUE
IGP127	normal	1940	pT2c	NX	3+4	undet.	13,9	TRUE	TRUE
IGP137	normal	1941	pT2c	N0	3+3	undet.	8,4	TRUE	TRUE
IGP138	normal	1958	pT2c	NX	3+4	undet.	NA	TRUE	TRUE
IGP139	normal	1938	pT3a	NX	3+4	undet.	6,3	TRUE	TRUE
IGP140	normal	1941	pT2a	NX	3+4	undet.	3,2	FALSE	TRUE
IGP141	normal	1943	pT2a	NX	3+3	undet.	3,8	TRUE	TRUE
IGP142	normal	1952	pT2c	NX	3+3	undet.	4,4	TRUE	TRUE
IGP143	normal	1948	pT3a	NX	3+4	undet.	9,1	TRUE	TRUE
IGP144	normal	1938	pT2c	NX	3+4	undet.	5,3	TRUE	TRUE
IGP145	normal	1944	pT2c	N0	3+3	undet.	8,5	TRUE	TRUE
IGP16	normal	1955	pT2a	NX	3+3	undet.	4,2	TRUE	TRUE
IGP17	normal	1947	pT2a	NX	3+3	undet.	4,6	TRUE	TRUE
IGP173	normal	1946	pT2a	NX	3+3	undet.	8,5	FALSE	TRUE
IGP174	normal	1950	pT2c	NX	3+3	undet.	6,2	TRUE	TRUE
IGP175	normal	1942	pT2c	N0	3+4	undet.	6,6	TRUE	TRUE
IGP176	normal	1940	pT2c	NX	3+3	undet.	1,8	TRUE	TRUE
IGP177	normal	1938	pT3a	N0	3+4	undet.	0,1	TRUE	TRUE
IGP178	normal	1949	pT2c	NX	3+3	undet.	7,5	TRUE	TRUE
IGP179	normal	1949	pT2a	N0	4+5	undet.	3,4	TRUE	TRUE
IGP18	normal	1937	pT2a	NX	3+3	undet.	11,4	TRUE	TRUE
IGP180	normal	1945	pT2c	NX	3+4	undet.	5,3	TRUE	TRUE
IGP181	normal	1946	pT2c	NX	3+3	undet.	0,9	TRUE	TRUE
IGP182	normal	1939	pT2c	NX	3+4	undet.	4,1	TRUE	TRUE
IGP183	normal	1940	pT2a	N0	3+3	undet.	4,3	TRUE	TRUE
IGP184	normal	1937	pT3a	N0	3+4	undet.	4,5	TRUE	TRUE
IGP185	normal	1951	pT2c	NX	3+3	undet.	5,6	TRUE	TRUE
IGP186	normal	1945	pT2c	NX	3+3	undet.	5,5	TRUE	FALSE
IGP187	normal	1949	pT2a	N0	3+4	undet.	4,3	FALSE	FALSE
IGP188	normal	1949	pT2c	NX	3+4	undet.	6,5	TRUE	FALSE
IGP189	normal	1941	pT2c	N0	3+3	undet.	8,0	FALSE	FALSE
IGP20	normal	1952	pT2a	NX	3+3	undet.	9,5	TRUE	FALSE
IGP75	normal	1941	pT2a	NX	3+4	undet.	2,3	TRUE	TRUE
IGP76	normal	1950	pT2c	N0	3+3	undet.	8,5	TRUE	TRUE
IGP78	normal	1939	pT2c	NX	3+3	undet.	7,8	TRUE	TRUE
IGP79	normal	1958	pT2c	NX	3+4	undet.	5,0	TRUE	TRUE
IGP84	normal	1958	pT2a	NX	3+3	undet.	10,0	TRUE	TRUE
IGP85	normal	1942	pT2c	NX	4+3	undet.	8,6	TRUE	TRUE
IGP86	normal	1944	pT2a	N0	3+4	undet.	4,5	TRUE	TRUE
IGP87	normal	1936	pT3a	NX	3+4	undet.	8,0	TRUE	TRUE
IGP88	normal	1941	pT2c	N0	4+3	undet.	7,2	TRUE	TRUE
IGP89	normal	1960	pT2c	NX	3+3	undet.	5,7	TRUE	TRUE
IGP90	normal	1953	pT2a	NX	3+3	undet.	1,7	TRUE	TRUE
IGP91	normal	1951	pT3a	N0	3+4	undet.	8,3	TRUE	TRUE
IGP92	normal	1934	pT2c	N0	3+4	undet.	11,4	TRUE	TRUE
IGP110	tumour	1951	pT2c	N0	4+3	no	37,5	TRUE	TRUE

sample ID	tissue type	year of birth	pT	pN	RRP Gleason	TMPRSS2:ERG fusion	PSA	Gene expression data	miRNA expression data
IGP116	tumour	1942	pT3a	N0	3+4	no	6,5	TRUE	TRUE
IGP118	tumour	1947	pT3a	N0	3+4	no	16,8	TRUE	TRUE
IGP21	tumour	1940	pT3b	N0	4+5	no	28,0	TRUE	TRUE
IGP23	tumour	1957	pT2c	N0	4+3	no	8,0	TRUE	TRUE
IGP26	tumour	1940	pT4a	N1	5+4	no	4,0	TRUE	TRUE
IGP29	tumour	1939	pT3a	N0	4+3	no	6,0	TRUE	TRUE
IGP31	tumour	1942	pT3b	N0	4+4	no	4,4	TRUE	TRUE
IGP32	tumour	1939	pT3b	N0	5+4	no	2,9	TRUE	TRUE
IGP33	tumour	1940	pT3a	N1	4+5	no	40,0	TRUE	TRUE
IGP35	tumour	1945	pT3b	N1	5+4	no	22,9	TRUE	TRUE
IGP36	tumour	1941	pT2c	N0	4+3	no	16,7	TRUE	TRUE
IGP38	tumour	1942	pT3b	N1	4+5	no	48,6	TRUE	TRUE
IGP43	tumour	1951	pT2c	N0	3+4	no	7,5	TRUE	TRUE
IGP51	tumour	1944	pT3b	N0	4+3	no	47,0	TRUE	TRUE
IGP52	tumour	1939	pT2c	N0	4+3	no	4,1	TRUE	TRUE
IGP55	tumour	1943	pT3b	N0	4+3	no	30,5	TRUE	TRUE
IGP57	tumour	1945	pT2c	N1	4+3	no	8,1	TRUE	TRUE
IGP81	tumour	1943	pT3a	N0	4+3	no	23,0	TRUE	TRUE
IGP83	tumour	1943	pT2c	N0	4+3	no	4,2	TRUE	TRUE
IGP111	tumour	1944	pT2a	N0	3+4	undet.	7,8	TRUE	TRUE
IGP113	tumour	1940	pT3a	N1	4+5	undet.	16,4	TRUE	TRUE
IGP190	tumour	1944	pT3b	N0	3+4	undet.	7,3	TRUE	TRUE
IGP27	tumour	1941	pT3b	N0	4+5	undet.	5,3	FALSE	TRUE
IGP30	tumour	1943	pT2c	N0	4+3	undet.	7,9	TRUE	TRUE
IGP34	tumour	1938	pT3a	N0	4+5	undet.	9,3	FALSE	TRUE
IGP39	tumour	1945	pT2c	N0	4+3	undet.	13,4	TRUE	TRUE
IGP42	tumour	1952	pT4	N1	4+5	undet.	100,0	TRUE	TRUE
IGP45	tumour	1937	pT3a	N0	4+3	undet.	15,5	TRUE	TRUE
IGP53	tumour	1950	pT3a	N0	4+3	undet.	14,7	FALSE	TRUE
IGP54	tumour	1939	pT3a	N1	4+3	undet.	13,6	TRUE	TRUE
IGP59	tumour	1952	pT3a	N0	3+4	undet.	10,5	TRUE	TRUE
IGP63	tumour	1940	pT3a	N0	4+3	undet.	8,2	FALSE	TRUE
IGP82	tumour	1942	pT3a	N0	4+5	undet.	17,5	TRUE	TRUE
IGP105	tumour	1938	pT2c	NX	3+3	yes	6,4	TRUE	TRUE
IGP115	tumour	1957	pT2c	NX	3+3	yes	NA	TRUE	TRUE
IGP117	tumour	1938	pT2a	NX	3+3	yes	5,5	TRUE	TRUE
IGP25	tumour	1946	pT3a	N1	4+5	yes	16,5	TRUE	TRUE
IGP28	tumour	1943	pT2c	NX	3+3	yes	4,6	TRUE	TRUE
IGP37	tumour	1959	pT2c	N0	4+3	yes	14,0	TRUE	TRUE
IGP40	tumour	1940	pT3a	N1	4+3	yes	12,7	TRUE	TRUE
IGP41	tumour	1941	pT3b	N1	4+3	yes	12,0	TRUE	TRUE
IGP44	tumour	1940	pT3a	N1	4+3	yes	13,4	TRUE	TRUE
IGP47	tumour	1947	pT3a	N0	3+4	yes	7,8	TRUE	TRUE
IGP48	tumour	1948	pT3a	N0	3+4	yes	6,3	TRUE	TRUE
IGP49	tumour	1935	pT3b	N1	4+3	yes	1,9	TRUE	TRUE
IGP60	tumour	1938	pT3b	N0	3+4	yes	11,7	TRUE	TRUE
IGP61	tumour	1950	pT3b	N0	3+4	yes	19,9	TRUE	TRUE
IGP62	tumour	1946	pT3b	N0	4+3	yes	29,6	TRUE	TRUE
IGP64	tumour	1947	pT2c	N0	3+4	yes	5,4	TRUE	TRUE
IGP80	tumour	1957	pT3a	N0	3+4	yes	7,0	TRUE	TRUE

Supplementary Table 2: Sample-wise sequencing and enrichment statistics. Column headings are as follows: (sample ID) Sample name, (status) tumour or normal sample, (Fusion state) *TMPRSS2:ERG* fusion state, (RO) raw output, all reads obtained for the sample during sequencing, (MO) mappable output, all reads mapping at least once to the human reference genome, (UMO) uniquely mappable output, all reads mapping at only one position in the human reference genome, (Uniq Locs) number of UMO reads after depletion of redundant reads, (0CpG depleted count) number of UniqLocs after depletion of reads containing no CpG in the 200bp elongated sequence, (%) Ratio of 0CpG depleted read count and UniqLocs in %, (GC %) GC content of the curated reads, (Enrichment factor) MeDIP enrichment factor, (*BRD1*/*COQ3*), (log2) $\log_2$(enrichment factor).

sample ID	status	Fusion state	RO	MO	UMO	Uniq Locs	0CpG depleted count	[%]	GC [%]	Enrich-ment factor	log2
IGP102	normal	undet.	68456560	45475526	25975279	20382937	17707823	86,9	43,3	14	3,8
IGP103	normal	undet.	74318146	42675806	23804014	19035568	16819487	88,4	44,6	151	7,2
IGP104	normal	undet.	55675343	35269796	20426022	16383403	14142442	86,3	43,1	26	4,7
IGP119	normal	undet.	50908740	34767127	19795621	15827541	14136461	89,3	44,0	191	7,6
IGP120	normal	undet.	80921862	42644806	24552994	21250761	17480756	82,3	44,5	51	5,7
IGP122	normal	undet.	1E+08	54670328	29877288	25692140	22330658	86,9	44,9	90	6,5
IGP123	normal	undet.	90420071	48786362	27016764	23620634	20638761	87,4	44,8	87	6,4
IGP124	normal	undet.	98006476	53521608	28601765	24568274	21280487	86,6	44,9	93	6,5
IGP126	normal	undet.	52892173	32581227	18364009	15516115	13497418	87,0	44,2	63	6,0
IGP127	normal	undet.	47412111	27201661	14939017	13012755	10854956	83,4	43,7	41	5,4
IGP137	normal	undet.	66848601	39838366	23219159	20203441	17132370	84,8	44,3	87	6,4
IGP138	normal	undet.	63082047	35906937	20629183	18301446	15871125	86,7	44,3	56	5,8
IGP139	normal	undet.	39616752	27403901	15382690	12492637	11021051	88,2	43,5	203	7,7
IGP140	normal	undet.	49529363	34479687	18890379	14813613	13489695	91,1	44,5	213	7,7
IGP141	normal	undet.	51224936	34724388	19537152	15484722	14106891	91,1	44,6	94	6,6
IGP142	normal	undet.	43963091	28080583	15677236	13617611	11902552	87,4	44,7	65	6,0
IGP143	normal	undet.	52982668	33310383	18963566	16336437	14449267	88,4	44,0	56	5,8
IGP144	normal	undet.	61124190	38731710	21487037	18111379	16029908	88,5	44,0	45	5,5
IGP145	normal	undet.	52783688	33159361	19424898	16911923	14612053	86,4	43,4	25	4,6
IGP16	normal	undet.	51057239	28675421	15608230	12594775	11469588	91,1	45,2	71	6,1
IGP17	normal	undet.	56740988	40400203	22487915	15272669	14004292	91,7	45,6	43	5,4
IGP173	normal	undet.	51417886	28641639	14416723	13112089	11904231	90,8	46,4	53	5,7
IGP174	normal	undet.	59041613	33276094	16809555	15178229	13720054	90,4	45,8	35	5,1
IGP175	normal	undet.	51925065	33823597	19395047	16634262	14676473	88,2	44,4	47	5,6
IGP176	normal	undet.	57284530	36356697	21324415	18562919	16102239	86,7	43,5	40	5,3
IGP177	normal	undet.	52058091	33443364	19041749	16663862	14793051	88,8	43,8	45	5,5
IGP178	normal	undet.	51104131	34297202	18490495	14371889	13172807	91,7	47,2	201	7,7
IGP179	normal	undet.	54128701	34045189	17444721	12288238	11208990	91,2	46,2	204	7,7
IGP18	normal	undet.	45650698	33620042	17841523	13543504	12449207	91,9	45,6	52	5,7
IGP180	normal	undet.	57847675	33161266	17223682	14592705	13043858	89,4	44,6	83	6,4
IGP181	normal	undet.	98172123	55734105	29165848	25066973	21712342	86,6	44,3	129	7,0
IGP182	normal	undet.	57099861	33629121	18519053	15992475	13835738	86,5	44,4	95	6,6
IGP183	normal	undet.	1,13E+08	61876870	32202440	27137153	23707799	87,4	44,3	56	5,8
IGP184	normal	undet.	61981298	35899582	20009549	17352422	15372704	88,6	45,0	92	6,5
IGP185	normal	undet.	53649442	30436056	16071428	13973848	12221449	87,5	44,7	131	7,0
IGP186	normal	undet.	54515436	31708116	16676050	14420409	12767104	88,5	44,5	112	6,8
IGP187	normal	undet.	58085890	33333127	17114023	15129622	13340683	88,2	44,9	105	6,7
IGP188	normal	undet.	56646174	35557086	20122646	17913535	15952275	89,1	44,3	45	5,5
IGP189	normal	undet.	57222448	35833353	20084794	17232187	15379151	89,2	44,7	149	7,2
IGP20	normal	undet.	48206419	27433261	15659087	13362436	11781988	88,2	44,0	68	6,1
IGP75	normal	undet.	84880133	50069760	26992346	19036888	16932613	88,9	45,2	63	6,0
IGP76	normal	undet.	48034724	31885241	17100196	13942540	12694730	91,1	46,3	100	6,6
IGP78	normal	undet.	45909128	26150283	14655147	12016485	10862203	90,4	45,4	98	6,6
IGP79	normal	undet.	69981957	36066238	20179838	16632938	14994177	90,1	45,7	93	6,5
IGP84	normal	undet.	39600397	22704940	11868919	11214603	10214972	91,1	45,6	90	6,5
IGP85	normal	undet.	43193831	24174166	11995768	10707426	9593343	89,6	46,0	120	6,9
IGP86	normal	undet.	54865647	32496585	16933259	14397938	13186951	91,6	46,4	91	6,5
IGP87	normal	undet.	57567293	34677735	17904425	15246219	14138399	92,7	47,0	76	6,2
IGP88	normal	undet.	47261785	27497227	15369466	13354588	11493515	86,1	43,3	46	5,5
IGP89	normal	undet.	55766773	32728549	17862831	15224091	13117240	86,2	43,8	35	5,1
IGP90	normal	undet.	56862178	34155660	19029453	16687643	14749862	88,4	44,5	66	6,0

sample ID	status	Fusion state	RO	MO	UMO	Uniq Locs	0CpG depleted count	[%]	GC [%]	Enrich-ment factor	log2
IGP91	normal	undet.	53506642	35333967	20688965	18190022	16067935	88,3	44,1	59	5,9
IGP92	normal	undet.	53182799	34943004	21225080	17246366	14787537	85,7	42,7	74	6,2
IGP110	tumor	no	54499098	29666880	15948545	13759478	12415995	90,2	45,9	50	5,6
IGP116	tumor	no	61533753	40654229	23742239	18999044	16962097	89,3	44,1	121	6,9
IGP118	tumor	no	43971881	30633533	17216250	13962289	12634478	90,5	44,7	154	7,3
IGP21	tumor	no	65234680	37238055	22084069	18552547	16774311	90,4	45,0	150	7,2
IGP23	tumor	no	45803513	33463519	19346383	15083941	13852518	91,8	46,2	118	6,9
IGP26	tumor	no	49686845	36497324	20836011	16021287	14693917	91,7	46,4	123	6,9
IGP29	tumor	no	55061939	37051789	21720788	16620700	14893959	89,6	46,8	378	8,6
IGP31	tumor	no	56439838	40500979	23792758	18295961	16583315	90,6	45,5	200	7,6
IGP32	tumor	no	43187431	32277055	18906642	15480491	13993192	90,4	45,1	104	6,7
IGP33	tumor	no	62232985	45203064	27020748	14865355	13812647	92,9	47,3	139	7,1
IGP35	tumor	no	47772233	28070845	16179288	13221765	11998935	90,8	45,7	84	6,4
IGP36	tumor	no	52738090	38848297	22776596	12342626	11472811	93,0	46,5	13	3,7
IGP38	tumor	no	63429557	36451086	20675251	18957329	17291411	91,2	46,4	70	6,1
IGP43	tumor	no	49619013	28250864	16655105	14385435	12262929	85,2	43,6	25	4,6
IGP51	tumor	no	52914337	41100616	26975755	21913181	20413996	93,2	51,3	60	5,9
IGP52	tumor	no	52880082	42708302	27453350	22246891	20589666	92,6	49,5	36	5,2
IGP55	tumor	no	43393093	29509776	18612224	15315383	13364200	87,3	43,3	65	6,0
IGP57	tumor	no	44415765	30606321	18575564	15048413	13155715	87,4	43,4	52	5,7
IGP81	tumor	no	53579451	43437432	29215621	24267508	21262595	87,6	44,5	87	6,4
IGP83	tumor	no	65581870	51785995	33099484	24623023	23115679	93,9	50,6	35	5,1
IGP111	tumor	undet.	66531650	36239709	19377696	16517972	14754578	89,3	45,4	50	5,6
IGP113	tumor	undet.	65487752	33364449	17982737	15020912	13124558	87,4	45,2	102	6,7
IGP190	tumor	undet.	51152759	32785278	17908432	13584580	12153433	89,5	46,7	320	8,3
IGP27	tumor	undet.	55137370	34092347	18860054	16189399	14460163	89,3	44,7	127	7,0
IGP30	tumor	undet.	38367718	26118854	14359187	12157295	11056813	90,9	46,8	59	5,9
IGP34	tumor	undet.	71158068	37459691	20314521	16713681	15076801	90,2	45,4	44	5,5
IGP39	tumor	undet.	58393579	31414142	16857301	15687192	14219347	90,6	46,5	127	7,0
IGP42	tumor	undet.	54908905	32513285	18627612	15648939	13920499	89,0	44,9	90	6,5
IGP45	tumor	undet.	54524821	34398989	19937909	17187437	15306848	89,1	44,7	41	5,4
IGP53	tumor	undet.	54881316	35281615	20604272	16510097	14308278	86,7	43,2	33	5,0
IGP54	tumor	undet.	51552904	35445122	21088617	17107025	14945132	87,4	43,4	51	5,7
IGP59	tumor	undet.	45599115	28307008	15604491	13303304	11636238	87,5	44,3	96	6,6
IGP63	tumor	undet.	50649851	32002122	18243799	15892695	14121610	88,9	44,2	89	6,5
IGP82	tumor	undet.	45675137	36922926	23116221	18602576	17796170	95,7	52,1	54	5,8
IGP105	tumor	yes	59046782	39955939	22627403	17939913	15854076	88,4	43,5	44	5,5
IGP115	tumor	yes	60501526	33910719	17919489	15450186	13781030	89,2	45,2	82	6,4
IGP117	tumor	yes	56230109	37960650	21264249	16751183	15051356	89,9	44,7	54	5,8
IGP25	tumor	yes	44791527	29955186	15696357	13286411	12155376	91,5	46,2	241	7,9
IGP28	tumor	yes	62740332	39437264	21309174	18206591	16362222	89,9	44,9	29	4,9
IGP37	tumor	yes	43026265	24602998	12253351	11251583	10217100	90,8	45,6	63	6,0
IGP40	tumor	yes	51633965	29778137	15274268	13967881	12656385	90,6	46,4	72	6,2
IGP41	tumor	yes	55523762	32957526	18449676	15667805	13794937	88,0	44,4	55	5,8
IGP44	tumor	yes	46499316	28432968	16217866	13931309	11946365	85,8	43,6	24	4,6
IGP47	tumor	yes	60957035	37825490	20728441	17958198	15904768	88,6	44,6	169	7,4
IGP48	tumor	yes	58140900	46074499	29065211	23516332	20664309	87,9	43,8	41	5,4
IGP49	tumor	yes	49148262	39685619	24679291	19782744	17529461	88,6	44,2	114	6,8
IGP60	tumor	yes	59253857	36556557	21143264	17355076	14793353	85,2	44,3	51	5,7
IGP61	tumor	yes	45595176	26838936	15746162	13826874	12121973	87,7	45,2	96	6,6
IGP62	tumor	yes	67815258	40346421	22650006	19671682	16758331	85,2	44,5	119	6,9
IGP64	tumor	yes	38347329	25134042	14007683	12026531	10625711	88,4	44,2	53	5,7
IGP80	tumor	yes	55965783	44433499	27755540	22842781	20050726	87,8	43,9	68	6,1

Supplementary Table 3: Top 110 differentially methylated regions between tumour and normal samples. Column headings are as follows: (LOCATION) genomic location, (tumour mean) average rpm value of all tumour samples in this region, (normal mean) average rpm value of all normal samples in this region, (p_{MW}) Mann-Whitney p-value of DMR, (TSS within 2kb) TSS that is located within 2kb range, (Rank) Rank number in ordered DMR list, (BS-MS) Marker validated in BS-MS experiment.

LOCATION	tumour mean	normal mean	p_{MW}	TSS within 2kb	Rank	BS-MS
chr7:157481501-157482000	0,61	0,04	7,9E-19	FALSE	1	BS-MS 2 and 3
chr7:116140001-116140500	0,60	0,05	9,7E-19	CAV2|AC002066.1	2	
chr14:31344501-31345000	0,66	0,06	1,3E-18	COCH	3	BS-MS 2
chr9:37002501-37003000	0,91	0,07	1,3E-18	FALSE	4	
chr8:70947001-70947500	0,74	0,05	1,5E-18	FALSE	5	
chr7:157484001-157484500	1,27	0,11	1,6E-18	FALSE	6	
chr6:26017501-26018000	1,11	0,15	1,7E-18	HIST1H1A|HIST1H1PS2	7	BS-MS 2
chr9:112810501-112811000	1,09	0,14	1,7E-18	AKAP2	8	
chr12:65220001-65220500	1,14	0,09	1,8E-18	TBC1D30	9	BS-MS 1
chr12:54441001-54441500	0,91	0,09	1,9E-18	FALSE	10	
chr6:29974501-29975000	1,21	0,09	2,0E-18	HCG4P3|HLA-J	11	BS-MS 2
chr4:185937001-185937500	0,91	0,18	2,0E-18	FALSE	12	
chr6:56818501-56819000	1,19	0,20	2,4E-18	DST|BEND6	13	BS-MS 2
chr11:58940501-58941000	1,19	0,14	2,4E-18	DTX4	14	BS-MS 2
chr9:126774501-126775000	0,55	0,03	2,5E-18	LHX2|RP11-85O21.4	15	
chr12:104852001-104852500	0,89	0,07	2,5E-18	CHST11	16	BS-MS 2
chr4:85414501-85415000	0,77	0,09	2,5E-18	FALSE	17	
chr7:143579501-143580000	0,79	0,09	2,5E-18	FAM115A	18	
chr3:170746001-170746500	1,09	0,10	2,6E-18	SLC2A2	19	BS-MS 2
chr2:235404501-235405000	1,01	0,10	2,6E-18	ARL4C	20	BS-MS 2
chr4:85402001-85402500	1,04	0,19	2,7E-18	FALSE	21	
chr7:29185501-29186000	1,32	0,15	2,9E-18	CPVL|CHN2	22	BS-MS 2
chr4:41882501-41883000	1,16	0,20	2,9E-18	RP11-457P14.4	23	
chr3:138154001-138154500	0,75	0,06	2,9E-18	ESYT3	24	
chr17:43974501-43975000	1,10	0,10	3,0E-18	FALSE	25	
chr1:15481001-15481500	0,83	0,06	3,2E-18	TMEM51	26	BS-MS 2
chr1:203598501-203599000	0,90	0,06	3,2E-18	FALSE	27	
chr13:100641501-100642000	1,10	0,16	3,3E-18	FALSE	28	
chr3:172165501-172166000	1,23	0,18	3,3E-18	GHSR	29	BS-MS 2 and 3
chr4:41868001-41868500	0,88	0,14	3,4E-18	FALSE	30	
chr11:3181501-3182000	1,00	0,07	3,6E-18	FALSE	31	
chr11:62691001-62691500	0,92	0,09	3,7E-18	CHRM1	32	BS-MS 2
chr7:116140501-116141000	0,87	0,14	3,7E-18	CAV2|AC002066.1	33	BS-MS 1 and 2
chr19:17246001-17246500	1,30	0,24	3,8E-18	FALSE	34	BS-MS 1
chr14:36992001-36992500	1,25	0,21	3,9E-18	FALSE	35	BS-MS 2
chr3:25469501-25470000	0,68	0,09	4,4E-18	RARB|AC098477.2	36	
chr1:119528501-119529000	1,18	0,35	4,4E-18	TBX15	37	
chr2:201450501-201451000	0,78	0,03	4,4E-18	AOX1|AC080164.1	38	BS-MS 2
chr20:50721001-50721500	1,17	0,15	4,5E-18	ZFP64	39	
chr7:127808001-127808500	0,64	0,07	4,5E-18	FALSE	40	
chr1:197889001-197889500	0,79	0,12	4,7E-18	FALSE	41	
chr9:126777501-126778000	0,94	0,14	4,8E-18	LHX2	42	
chr6:150286001-150286500	0,78	0,09	4,9E-18	ULBP1	43	
chr1:24648501-24649000	0,94	0,07	5,0E-18	GRHL3|RP11-10N16.3	44	
chr2:237079501-237080000	1,02	0,20	5,0E-18	FALSE	45	
chr1:119527001-119527500	0,87	0,09	5,1E-18	FALSE	46	
chr20:37356001-37356500	1,03	0,16	5,1E-18	FALSE	47	
chr4:85402501-85403000	0,76	0,12	5,2E-18	FALSE	48	
chr19:16436501-16437000	0,55	0,05	5,3E-18	KLF2	49	
chr6:127836001-127836500	0,99	0,14	5,4E-18	AL096711.1	50	
chr1:58716001-58716500	0,82	0,12	5,5E-18	DAB1	51	
chr1:146551501-146552000	0,75	0,07	5,6E-18	U1|AL596177.2	52	
chr3:48632001-48632500	0,91	0,16	5,8E-18	COL7A1	53	
chr15:90039501-90040000	0,69	0,06	5,8E-18	RHCG	54	
chr20:20345501-20346000	1,29	0,27	5,8E-18	FALSE	55	
chr14:29254501-29255000	0,81	0,11	5,8E-18	FALSE	56	

LOCATION	tumour mean	normal mean	p_{MW}	TSS within 2kb	Rank	BS-MS
chr10:94822001-94822500	0,68	0,08	5,9E-18	CYP26C1|RP11-348J12.2	57	
chr18:56939501-56940000	0,49	0,05	6,2E-18	RAX	58	
chr7:151107501-151108000	0,73	0,08	6,2E-18	WDR86|RP4-555L14.5	59	
chr1:119543001-119543500	1,02	0,12	6,4E-18	RP4-712E4.1|RP4-712E4.2	60	
chr20:50721501-50722000	0,45	0,03	6,4E-18	ZFP64	61	
chr19:48983501-48984000	1,13	0,13	6,4E-18	AC008403.1	62	
chr3:125899501-125900000	0,63	0,08	6,4E-18	ALDH1L1|RP11-124N2.3	63	
chr7:129422001-129422500	1,53	0,20	6,7E-18	FALSE	64	
chr2:27529501-27530000	0,70	0,05	6,7E-18	TRIM54|UCN	65	
chr6:28367501-28368000	0,86	0,12	6,8E-18	ZSCAN12	66	BS-MS 2
chr5:140810501-140811000	1,59	0,41	6,9E-18	PCDHGA12	67	
chr11:20619001-20619500	0,67	0,14	7,0E-18	SLC6A5	68	
chr9:135620501-135621000	1,69	0,47	7,1E-18	FALSE	69	
chr7:19146001-19146500	1,16	0,12	7,3E-18	FALSE	70	
chr7:45613501-45614000	0,58	0,07	7,6E-18	ADCY1	71	
chr2:73147501-73148000	0,98	0,10	7,8E-18	FALSE	72	
chr4:11429501-11430000	1,06	0,13	7,9E-18	HS3ST1	73	
chr6:28367001-28367500	1,19	0,11	7,9E-18	ZSCAN12	74	BS-MS 2
chr10:102896001-102896500	0,82	0,12	8,1E-18	FALSE	75	
chr1:146556501-146557000	1,03	0,16	8,1E-18	RP11-325P15.2	76	
chr14:85998001-85998500	0,99	0,07	8,1E-18	AL049775.1|FLRT2	77	
chr6:137809501-137810000	1,00	0,12	8,2E-18	FALSE	78	
chr14:85997501-85998000	0,53	0,05	8,3E-18	AL049775.1|FLRT2	79	
chr9:135462001-135462500	1,14	0,16	8,4E-18	FALSE	80	
chr7:96632001-96632500	0,96	0,20	8,4E-18	DLX6AS	81	
chr1:70035001-70035500	0,91	0,09	8,4E-18	LRRC7	82	
chr7:157478501-157479000	0,52	0,06	8,4E-18	FALSE	83	
chr12:54440501-54441000	1,11	0,15	8,5E-18	FALSE	84	
chr17:41363501-41364000	0,92	0,10	8,6E-18	TMEM106A	85	
chr7:97361001-97361500	1,26	0,14	8,7E-18	TAC1	86	BS-MS 3
chr2:87016001-87016500	0,83	0,15	8,8E-18	CD8A	87	
chr12:54447501-54448000	0,84	0,12	8,9E-18	HOXC4	88	
chr5:77268001-77268500	0,86	0,19	9,0E-18	FALSE	89	
chr3:68980501-68981000	1,03	0,15	9,1E-18	FAM19A4	90	
chr12:122016501-122017000	0,75	0,07	9,1E-18	KDM2B	91	
chr4:16085501-16086000	1,01	0,24	9,4E-18	PROM1	92	
chr19:46916501-46917000	0,50	0,05	9,7E-18	CCDC8	93	
chr6:29974001-29974500	0,81	0,15	9,8E-18	HCG4P3|HLA-J	94	
chr3:154145501-154146000	1,13	0,15	1,0E-17	GPR149	95	
chr10:94822501-94823000	0,44	0,03	1,0E-17	CYP26C1	96	
chr2:220117501-220118000	0,96	0,12	1,0E-17	TUBA4A|TUBA4B	97	
chr7:32981501-32982000	0,95	0,16	1,0E-17	RP9P	98	
chr2:45169501-45170000	0,84	0,11	1,1E-17	SIX3|AC012354.5	99	
chr10:102905501-102906000	0,58	0,08	1,1E-17	FALSE	100	
chr16:54971501-54972000	1,34	0,24	1,1E-17	FALSE	101	
chr17:78806501-78807000	0,63	0,08	1,1E-17	FALSE	102	
chr7:128337501-128338000	0,77	0,12	1,1E-17	5S_rRNA	103	
chr3:129024501-129025000	1,14	0,22	1,1E-17	FALSE	104	
chr5:140892501-140893000	1,48	0,13	1,2E-17	AC005618.1	105	
chr2:162284001-162284500	1,09	0,18	1,2E-17	FALSE	106	
chr4:54975501-54976000	0,92	0,19	1,2E-17	FALSE	107	
chr11:67351501-67352000	0,75	0,10	1,2E-17	GSTP1	108	BS-MS 1
chr19:58220001-58220500	1,34	0,14	1,2E-17	ZNF154|AC004017.1	109	BS-MS 1
chr10:112838001-112838500	1,34	0,29	1,2E-17	ADRA2A	110	

Supplementary Table 4: Top 110 differentially methylated regions between FUS- and FUS+ samples. Column headings are as follows: (LOCATION) genomic location, (FUS- mean) average rpm value of all FUS- samples in this region, (FUS+ mean) average rpm value of all FUS+ samples in this region, (normal mean) average rpm value of all normal samples in this region, (p_{MW}) Mann-Whitney p-value, (TSS within 2kb) TSS that is located within 2kb range, (Rank) Rank number in ordered DMR list, (BS-MS) Marker validated in BS-MS experiment.

LOCATION	FUS- mean	FUS+ mean	normal mean	p_{MW}	TSS within 2kb	Rank	BS-MS
chr1:149033001-149033500	1,69	3,44	3,40	2,5E-10	FALSE	1	BS-MS 3
chr16:46414001-46414500	3,16	5,98	5,47	8,8E-10	FALSE	2	BS-MS 3
chr16:33879501-33880000	3,02	5,77	5,51	2,4E-09	FALSE	3	
chr16:33898001-33898500	4,02	7,72	7,49	5,7E-09	FALSE	4	
chr22:16854501-16855000	3,29	6,73	6,40	8,4E-09	FALSE	5	
chr7:61791001-61791500	2,79	6,42	5,95	8,4E-09	FALSE	6	
chr16:46449001-46449500	3,56	6,91	6,99	1,2E-08	FALSE	7	
chr16:46455501-46456000	2,83	5,62	5,15	1,2E-08	FALSE	8	
chr7:61743001-61743500	3,08	6,01	5,84	1,2E-08	FALSE	9	
chr10:131503001-131503500	1,44	0,67	0,97	1,7E-08	FALSE	10	
chr16:33890501-33891000	1,93	3,97	3,76	1,7E-08	FALSE	11	
chr22:16853001-16853500	3,49	6,70	6,18	1,7E-08	FALSE	12	
chr7:61796501-61797000	3,37	6,66	6,85	1,7E-08	FALSE	13	
chr7:61798501-61799000	2,74	6,13	5,63	1,7E-08	FALSE	14	
chr1:61892001-61892500	1,26	0,63	0,73	2,5E-08	FALSE	15	
chrY:28801001-28801500	6,22	8,15	7,64	2,5E-08	FALSE	16	
chr10:38876501-38877000	3,00	5,76	5,52	3,4E-08	FALSE	17	
chr16:34183001-34183500	4,42	7,88	7,96	3,4E-08	FALSE	18	
chr16:46437501-46438000	4,10	8,91	9,09	3,4E-08	FALSE	19	
chr16:46439001-46439500	4,23	8,10	7,74	3,4E-08	FALSE	20	
chr16:46498001-46498500	3,88	8,01	7,91	3,4E-08	FALSE	21	
chr7:61749501-61750000	5,13	9,39	9,55	3,4E-08	FALSE	22	
chrY:13834501-13835000	5,71	8,59	8,61	3,4E-08	FALSE	23	
chr16:46453001-46453500	3,19	6,59	6,52	4,7E-08	FALSE	24	
chr16:46497501-46498000	2,40	4,84	4,77	4,7E-08	FALSE	25	
chr16:33877501-33878000	2,40	4,98	5,15	6,4E-08	FALSE	26	
chr16:33884501-33885000	4,41	7,48	7,53	6,4E-08	FALSE	27	
chr16:46454501-46455000	3,29	6,46	6,51	6,4E-08	FALSE	28	
chr2:89833501-89834000	3,52	6,45	6,86	6,4E-08	FALSE	29	
chr7:61793501-61794000	2,58	5,10	5,17	6,4E-08	FALSE	30	
chrY:13802501-13803000	3,82	5,83	5,62	6,4E-08	FALSE	31	
chr10:42815501-42816000	3,32	5,62	5,60	8,6E-08	FALSE	32	
chr16:33871001-33871500	3,95	7,11	7,07	8,6E-08	FALSE	33	
chr16:33882501-33883000	2,08	4,24	4,24	8,6E-08	FALSE	34	
chr16:34195001-34195500	3,78	7,09	6,97	8,6E-08	FALSE	35	
chr16:34196501-34197000	3,21	5,76	5,75	8,6E-08	FALSE	36	
chr16:46444501-46445000	1,86	3,97	3,74	8,6E-08	FALSE	37	
chr2:89831001-89831500	1,59	3,24	3,46	8,6E-08	FALSE	38	
chr20:29811501-29812000	3,81	6,21	6,28	8,6E-08	FALSE	39	
chr22:16854001-16854500	4,09	7,69	7,49	8,6E-08	FALSE	40	
chr22:16855501-16856000	3,22	5,96	6,15	8,6E-08	FALSE	41	
chr10:39105501-39106000	5,11	8,04	7,93	1,2E-07	FALSE	42	
chr10:39145501-39146000	5,12	7,60	7,57	1,2E-07	AL590623.1	43	
chr10:42666001-42666500	2,41	4,68	4,81	1,2E-07	FALSE	44	
chr16:33881501-33882000	3,64	6,30	6,33	1,2E-07	FALSE	45	
chr16:46396501-46397000	5,41	9,71	9,50	1,2E-07	FALSE	46	
chr16:46425001-46425500	2,86	5,71	5,00	1,2E-07	FALSE	47	
chr16:46443501-46444000	4,54	7,70	7,47	1,2E-07	FALSE	48	
chr16:46450001-46450500	4,19	8,00	8,22	1,2E-07	FALSE	49	
chr22:16860501-16861000	3,78	6,71	6,65	1,2E-07	FALSE	50	
chr22:16861001-16861500	3,91	7,34	7,40	1,2E-07	FALSE	51	
chr7:61750001-61750500	3,58	6,69	6,59	1,2E-07	FALSE	52	
chr7:61787001-61787500	3,97	7,82	7,96	1,2E-07	FALSE	53	
chr10:42383501-42384000	8,22	14,41	13,78	1,5E-07	FALSE	54	
chr16:33891501-33892000	3,17	6,71	6,69	1,5E-07	FALSE	55	

LOCATION	FUS- mean	FUS+ mean	normal mean	p_{MW}	TSS within 2kb	Rank	BS-MS
chr16:34190501-34191000	2,95	5,57	5,23	1,5E-07	FALSE	56	
chr16:46452501-46453000	3,00	6,52	6,46	1,5E-07	FALSE	57	
chr20:29807501-29808000	3,53	5,75	5,63	1,5E-07	FALSE	58	
chr20:29811001-29811500	4,62	7,20	6,93	1,5E-07	FALSE	59	
chr21:10837501-10838000	4,51	6,96	7,06	1,5E-07	FALSE	60	
chr7:61749001-61749500	4,15	7,47	7,61	1,5E-07	FALSE	61	
chr7:61791501-61792000	4,27	7,84	7,76	1,5E-07	FALSE	62	
chr7:61800501-61801000	2,68	5,25	5,15	1,5E-07	FALSE	63	
chrY:13649501-13650000	3,99	6,32	6,01	1,5E-07	FALSE	64	
chr3:53529501-53530000	0,80	0,02	0,05	1,5E-07	CACNA1D	65	
chr10:38785001-38785500	4,49	6,78	6,73	2,0E-07	FALSE	66	
chr10:42598001-42598500	11,45	17,52	17,05	2,0E-07	FALSE	67	
chr16:46432001-46432500	1,32	3,05	2,97	2,0E-07	FALSE	68	
chr16:46438001-46438500	4,27	8,17	8,34	2,0E-07	FALSE	69	
chr16:46449501-46450000	3,67	6,99	6,53	2,0E-07	FALSE	70	
chr21:10814501-10815000	4,88	7,66	7,61	2,0E-07	FALSE	71	
chr7:61745501-61746000	3,59	6,86	6,95	2,0E-07	FALSE	72	
chr7:61752001-61752500	4,21	7,58	7,81	2,0E-07	FALSE	73	
chr7:61754001-61754500	3,38	6,64	6,93	2,0E-07	FALSE	74	
chr7:61802001-61802500	3,56	6,86	7,25	2,0E-07	FALSE	75	
chr15:33011501-33012000	1,10	0,11	0,13	2,4E-07	GREM1	76	
chr10:39142501-39143000	5,75	8,76	8,26	2,6E-07	AL590623.1	77	
chr10:39153501-39154000	5,68	8,98	8,96	2,6E-07	FALSE	78	
chr10:42383001-42383500	6,22	9,52	9,52	2,6E-07	FALSE	79	
chr16:33895501-33896000	4,08	7,38	6,99	2,6E-07	FALSE	80	
chr16:46412001-46412500	3,46	6,27	5,99	2,6E-07	FALSE	81	
chr16:46433501-46434000	1,80	4,10	3,66	2,6E-07	FALSE	82	
chr16:46452001-46452500	4,13	7,72	8,05	2,6E-07	FALSE	83	
chr2:92267501-92268000	3,53	5,91	5,90	2,6E-07	FALSE	84	
chr21:10833001-10833500	3,88	6,11	5,97	2,6E-07	FALSE	85	
chr22:16852001-16852500	3,94	7,25	7,36	2,6E-07	FALSE	86	
chrY:13845001-13845500	3,83	6,36	6,06	2,6E-07	FALSE	87	
chr5:88675501-88676000	0,03	0,21	0,10	2,8E-07	FALSE	88	
chr1:47882001-47882500	0,22	0,01	0,01	3,2E-07	FOXE3	89	
chr3:53530001-53530500	0,63	0,09	0,16	3,2E-07	CACNA1D	90	
chr8:144950501-144951000	0,78	0,21	0,37	3,3E-07	FALSE	91	
chr10:42598501-42599000	5,56	8,84	8,25	3,4E-07	FALSE	92	
chr16:33885001-33885500	3,20	5,54	6,07	3,4E-07	FALSE	93	
chr16:34182501-34183000	3,95	7,60	7,61	3,4E-07	FALSE	94	
chr16:46401001-46401500	4,84	9,50	9,36	3,4E-07	FALSE	95	
chr16:46409501-46410000	4,69	8,62	7,88	3,4E-07	FALSE	96	
chr16:46425501-46426000	4,89	9,11	8,09	3,4E-07	FALSE	97	
chr16:46429501-46430000	3,77	7,09	6,45	3,4E-07	FALSE	98	
chr2:89849001-89849500	3,94	6,89	6,97	3,4E-07	FALSE	99	
chr20:29823501-29824000	2,77	4,64	4,45	3,4E-07	FALSE	100	
chr20:29824501-29825000	6,33	9,50	9,30	3,4E-07	FALSE	101	
chr20:29826001-29826500	4,47	7,09	6,57	3,4E-07	FALSE	102	
chr7:61055501-61056000	1,59	3,22	3,34	3,4E-07	FALSE	103	
chr7:61748501-61749000	3,85	7,35	7,56	3,4E-07	FALSE	104	
chrY:13855001-13855500	4,79	7,55	7,29	3,4E-07	FALSE	105	
chr10:42358501-42359000	6,34	9,34	8,98	3,6E-07	FALSE	106	
chr9:136844501-136845000	0,84	0,12	0,46	3,8E-07	FALSE	107	
chr10:121130001-121130500	0,67	0,16	0,17	3,8E-07	FALSE	108	
chr4:10683501-10684000	0,10	0,32	0,22	4,4E-07	FALSE	109	
chr16:33886501-33887000	2,95	5,55	5,27	4,4E-07	FALSE	110	

Supplementary Table 5: Genes with at least one differentially methylated region (BH < 0.05) between tumour and normal or between FUS- and FUS+ samples from the cancer genes census list (FUTREAL 2004). For each gene the most significant bin (according to p-values) of all TSS associated bins is shown. Column headings are as follows: (Name) Gene name, (best DMR tumour) Chromosomal location of best tumour DMR (start), (T) mean MeDIP rpm value in tumour samples, (N) mean MeDIP rpm value of normal samples, (p_{MW} T/N) Mann-Whitney p-value tumour vs. normal, (p_{BH} T/N) p-value after Benjamini-Hochberg correction (over all significantly methylated bins), (Rank) rank of bin in ordered list of all genomic bins (according to p_{BH}-values), (best DMR FUS) Chromosomal location of best FUS- DMR (start), (Fus-) mean MeDIP rpm value of FUS- samples, (Fus+) mean MeDIP rpm value of FUS+ samples, (N) mean MeDIP rpm value of normal samples, (p_{MW} Fus -/+) Mann-Whitney p-value FUS- vs. FUS+, (p_{BH} Fus -/+) p-value after Benjamini-Hochberg correction (over all significantly methylated bins), (Rank) rank of bin in ordered list of all genomic bins (according to p_{BH}-values), (DMR T/N) DMR significant in tumour, (DMR Fus-/+) DMR significant in FUS-.

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
ACSL6	chr5:131346501	0,6	0,4	1,E-04	2,E-03	52217	chr5:131310501	1,1	0,7	0,9	1,E-04	2,E-02	4565	X	X
AFF1	chr4:87968501	0,5	0,3	5,E-05	7,E-04	42841	chr4:88035501	0,6	0,3	0,4	7,E-04	3,E-02	13445	X	X
AFF3	chr2:100722001	0,4	0,2	3,E-06	8,E-05	26496	chr2:100193501	0,5	0,3	0,3	1,E-02	1,E-01	81653	X	
AFF4	chr5:132270501	0,4	0,3	3,E-01	5,E-01	416240	chr5:132272501	0,5	0,3	0,4	3,E-04	2,E-02	7201		X
AKAP9	chr7:91568001	0,4	0,3	9,E-03	4,E-02	138938	chr7:91726001	1,0	0,8	0,9	9,E-03	9,E-02	66100	X	
AKT2	chr19:40784501	0,8	0,5	2,E-09	1,E-07	9011	chr19:40747501	0,8	0,4	0,3	1,E-04	2,E-02	3879	X	X
ALDH2	chr12:112205001	0,6	0,1	1,E-12	2,E-10	3411	chr12:112205001	0,4	0,9	0,1	2,E-03	5,E-02	29431	X	
ALK	chr2:29550001	0,7	0,5	1,E-05	2,E-04	32856	chr2:29446001	0,6	0,5	0,5	1,E-01	3,E-01	273450	X	
APC	chr5:112073501	0,7	0,1	1,E-13	3,E-11	2424	chr5:112160501	0,5	0,6	0,5	1,E-01	3,E-01	270788	X	
ARHGAP26	chr5:142281501	1,1	0,9	3,E-04	4,E-03	63924	chr5:142253501	0,4	0,3	0,3	2,E-03	5,E-02	24751	X	X
ARID1A	chr1:27101001	1,1	0,9	9,E-04	8,E-03	78931	chr1:27099501	0,8	0,6	0,6	2,E-02	1,E-01	94904	X	
ATM	chr11:108094501	0,4	0,3	1,E-03	9,E-03	82595	chr11:108094501	0,4	0,6	0,3	4,E-02	2,E-01	136137	X	
BAP1	chr3:52440001	0,4	0,2	1,E-04	2,E-03	52326	chr3:52440001	0,5	0,3	0,2	2,E-03	5,E-02	23809	X	X
BCL10	chr1:85743001	0,3	0,4	9,E-05	1,E-03	48447	chr1:85745001	0,4	0,4	0,3	1,E+00	1,E+00	647144	X	
BCL11A	chr2:60757501	0,5	0,3	9,E-06	2,E-04	31578	chr2:60774001	0,3	0,5	0,2	2,E-03	4,E-02	22169	X	X
BCL2	chr18:60985501	0,4	0,1	2,E-14	7,E-12	1761	chr18:60985001	1,0	0,8	0,4	3,E-01	5,E-01	357364	X	
BCL3	chr19:45261001	1,2	0,7	4,E-05	6,E-04	41506	chr19:45263501	0,4	0,3	0,3	7,E-02	2,E-01	186770	X	
BCL6	chr3:187455001	0,8	0,3	3,E-10	2,E-08	7076	chr3:187447501	0,9	0,6	0,7	1,E-03	4,E-02	20874	X	X
BCR	chr22:23524001	1,4	0,8	2,E-09	2,E-07	9341	chr22:23558001	0,6	0,2	0,2	1,E-06	3,E-03	225	X	X
BRD3	chr9:136922001	0,9	0,7	5,E-04	5,E-03	69659	chr9:136918501	1,3	0,7	0,8	3,E-04	3,E-02	8869	X	X
BRD4	chr19:15389501	0,4	0,3	1,E-01	3,E-01	302688	chr19:15389501	0,5	0,3	0,3	3,E-04	2,E-02	7413		X
C15orf55	chr15:34634001	0,9	1,0	5,E-02	1,E-01	224685	chr15:34634001	1,1	0,7	1,0	4,E-04	3,E-02	10001		X
C16orf75	chr16:11409501	0,5	0,2	1,E-09	9,E-08	8456	chr16:11409501	0,6	0,4	0,2	1,E-03	4,E-02	21344	X	X
CANT1	chr17:76993001	0,6	0,9	3,E-06	7,E-05	25895	chr17:76993501	0,7	0,5	0,9	3,E-02	1,E-01	113986	X	
CARD11	chr7:3083001	0,4	0,2	1,E-05	3,E-04	33578	chr7:2968001	1,5	1,1	1,0	2,E-02	1,E-01	90228	X	
CARS	chr11:3051501	0,5	0,3	6,E-07	2,E-05	20287	chr11:3075001	0,5	0,3	0,4	5,E-05	1,E-02	2524	X	X
CBLB	chr3:105586001	0,4	0,1	1,E-11	1,E-09	4696	chr3:105586001	0,6	0,4	0,1	3,E-02	1,E-01	114670	X	
CBLC	chr19:45280001	0,8	0,6	5,E-03	3,E-02	118784	chr19:45279501	0,4	0,3	0,3	1,E-01	3,E-01	251631	X	
CCNB1IP1	chr14:20798001	0,6	0,5	2,E-02	8,E-02	175296	chr14:20786001	0,9	0,6	0,6	1,E-03	4,E-02	19253		X
CCND3	chr6:41908001	0,8	0,3	3,E-15	2,E-12	1181	chr6:42017001	0,7	0,4	0,5	2,E-04	2,E-02	6463	X	X
CD74	chr5:149792001	0,4	0,1	2,E-11	2,E-09	5049	chr5:149792501	0,4	0,3	0,1	7,E-02	2,E-01	194034	X	
CD79A	chr19:42381501	0,5	0,4	5,E-04	5,E-03	70501	chr19:42383001	1,4	1,2	1,3	1,E-01	3,E-01	246374	X	
CD79B	chr17:62007501	0,5	0,3	1,E-06	4,E-05	22640	chr17:62008501	0,7	0,4	0,3	4,E-03	6,E-02	37720	X	
CDC73	chr1:193090001	0,5	0,4	2,E-03	1,E-02	91349	chr1:193090001	0,6	0,5	0,4	2,E-01	4,E-01	290517	X	
CDK6	chr7:92355001	0,5	0,4	1,E-01	2,E-01	292489	chr7:92355001	0,7	0,3	0,4	8,E-05	1,E-02	3488		X
CDKN2A	chr9:21970501	0,8	0,2	9,E-16	7,E-13	857	chr9:21995501	0,4	0,1	0,1	1,E-03	4,E-02	21827	X	X
CDX2	chr13:28544501	0,7	0,2	6,E-14	2,E-11	2201	chr13:28543501	0,4	0,3	0,2	6,E-02	2,E-01	176879	X	
CEP110	chr9:123914001	0,4	0,3	6,E-05	9,E-04	45244	chr9:123907501	0,4	0,2	0,3	3,E-03	6,E-02	37307	X	
CHEK2	chr22:29129501	0,7	0,6	2,E-01	3,E-01	336852	chr22:29136001	0,6	0,4	0,5	2,E-04	2,E-02	6927		X
CIC	chr19:42790001	0,5	0,2	6,E-11	7,E-09	5971	chr19:42789001	0,5	0,3	0,3	4,E-02	2,E-01	139595	X	
CIITA	chr16:10971001	0,4	0,3	2,E-05	3,E-04	35067	chr16:10971001	0,5	0,4	0,3	1,E+00	1,E+00	644441	X	
CLTCL1	chr22:19280001	0,6	0,4	9,E-05	1,E-03	48497	chr22:19172501	0,5	0,3	0,2	2,E-02	1,E-01	96074	X	
COL1A1	chr17:48262001	0,6	0,4	1,E-04	1,E-03	50120	chr17:48267501	0,8	0,5	0,6	3,E-03	5,E-02	31544	X	
CREB1	chr2:208393001	0,4	0,5	5,E-03	3,E-02	117146	chr2:208392501	0,6	0,4	0,6	8,E-02	3,E-01	206602	X	
CREBBP	chr16:3931501	1,5	1,0	9,E-06	2,E-04	31538	chr16:3931501	1,8	1,2	1,0	1,E-02	9,E-02	66908	X	
CRTC3	chr15:91072001	0,9	0,8	1,E-02	5,E-02	146574	chr15:91072001	1,0	0,8	0,8	1,E-01	3,E-01	258259	X	
CXCR7	chr2:237477001	0,7	0,3	1,E-09	1,E-07	8634	chr2:237475501	0,6	0,3	0,5	3,E-03	6,E-02	32669	X	

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
CYTSB	chr17:19990001	0,7	0,5	1,E-03	1,E-02	87963	chr17:19990001	0,6	0,7	0,5	9,E-02	3,E-01	223101	X	
DAXX	chr6:33288501	1,0	0,8	2,E-03	2,E-02	97854	chr6:33289001	0,7	0,6	0,6	8,E-03	8,E-02	59492	X	
DDIT3	chr12:57912001	0,1	0,1	3,E-01	NA	2E+06	chr12:57915501	0,4	0,2	0,1	1,E-03	4,E-02	17428		X
DNMT3A	chr2:25562501	0,4	0,2	8,E-10	7,E-08	8104	chr2:25475501	0,6	0,3	0,2	7,E-04	3,E-02	13346	X	X
EBF1	chr5:158524001	1,4	0,3	3,E-17	1,E-13	201	chr5:158525001	0,6	0,7	0,4	1,E-01	3,E-01	263530	X	
EGFR	chr7:55085501	0,5	0,2	7,E-12	1,E-09	4480	chr7:55259501	0,6	0,4	0,4	3,E-03	6,E-02	33985	X	
EIF4A2	chr3:186506001	0,2	0,4	8,E-07	3,E-05	21276	chr3:186498501	0,2	0,2	0,2	4,E-01	NA	4E+06	X	
ELF4	chrX:129243501	0,4	0,0	1,E-15	9,E-13	938	chrX:129246001	0,4	0,2	0,3	4,E-03	6,E-02	38471	X	
ELN	chr7:73443501	0,7	0,3	3,E-09	2,E-07	9624	chr7:73457001	0,7	0,3	0,3	1,E-04	2,E-02	4961	X	X
EPS15	chr1:51986001	0,4	0,3	6,E-03	3,E-02	126051	chr1:51986001	0,4	0,3	0,3	2,E-01	4,E-01	290410	X	
ERC1	chr12:1137001	1,2	1,3	4,E-01	6,E-01	484508	chr12:1101001	0,5	0,3	0,2	2,E-03	4,E-02	21986		X
ERCC2	chr19:45872001	0,8	0,6	1,E-04	2,E-03	53802	chr19:45872001	0,9	0,7	0,6	3,E-02	2,E-01	134247	X	
ERCC4	chr16:14029501	0,9	0,6	9,E-06	2,E-04	32096	chr16:14014501	0,8	0,5	0,5	4,E-02	2,E-01	147581	X	
ERG	chr21:40033501	0,6	0,1	1,E-11	2,E-09	4943	chr21:39817001	0,9	0,6	0,9	5,E-04	3,E-02	11684	X	X
ETV1	chr7:14031501	0,5	0,4	4,E-04	4,E-03	67538	chr7:13945501	0,4	0,5	0,4	1,E-01	3,E-01	242274	X	
ETV5	chr3:185828501	0,7	0,4	3,E-07	1,E-05	18025	chr3:185828501	0,8	0,7	0,4	1,E-01	3,E-01	277847	X	
EXT1	chr8:119122001	0,7	0,1	6,E-15	3,E-12	1440	chr8:119122501	0,6	0,3	0,1	4,E-03	7,E-02	43070	X	
FANCC	chr9:97886501	0,7	0,4	6,E-08	3,E-06	14348	chr9:98080501	0,7	0,4	0,4	2,E-02	1,E-01	99098	X	
FCRL4	chr1:157558501	0,2	0,4	1,E-08	7,E-07	11440	chr1:157548001	0,1	0,1	0,1	4,E-01	NA	937997	X	
FEV	chr2:219847001	0,9	0,1	4,E-17	1,E-13	228	chr2:219850001	0,5	0,2	0,2	8,E-04	3,E-02	14356	X	X
FGFR1	chr8:38323001	0,7	0,1	3,E-15	2,E-12	1193	chr8:38282001	0,8	0,5	0,7	6,E-04	3,E-02	12662	X	X
FGFR2	chr10:123355501	0,7	0,2	3,E-16	3,E-13	572	chr10:123355501	1,0	0,6	0,2	7,E-03	8,E-02	55473	X	
FGFR3	chr4:1809001	0,5	0,4	7,E-03	4,E-02	131450	chr4:1802001	1,1	0,7	0,9	7,E-04	3,E-02	13912	X	X
FLCN	chr17:17141501	0,5	0,4	6,E-05	9,E-04	45116	chr17:17131001	1,4	1,0	1,0	4,E-03	7,E-02	41224	X	
FLI1	chr11:128565001	0,5	0,2	4,E-07	2,E-05	19353	chr11:128563501	0,4	0,2	0,2	1,E-02	1,E-01	82987	X	
FLT3	chr13:28674001	0,4	0,1	1,E-11	2,E-09	4896	chr13:28673501	0,3	0,5	0,2	1,E-02	1,E-01	83083	X	
FNBP1	chr9:132758001	0,5	0,3	4,E-04	4,E-03	65349	chr9:132806001	0,4	0,2	0,2	2,E-02	1,E-01	90616	X	
FOXL2	chr3:138663501	0,7	0,1	6,E-14	2,E-11	2181	chr3:138666001	0,5	0,1	0,1	3,E-03	6,E-02	36248	X	
FOXO1	chr13:41137501	0,6	0,5	3,E-02	9,E-02	188385	chr13:41137001	0,4	0,2	0,3	3,E-04	2,E-02	7361		X
FOXO4	chrX:70320501	0,6	0,4	3,E-03	2,E-02	104272	chrX:70320501	0,7	0,6	0,4	1,E-01	3,E-01	264320	X	
FOXP1	chr3:71295001	0,4	0,3	1,E-03	1,E-02	85955	chr3:71357001	0,9	0,5	0,6	1,E-03	4,E-02	18860	X	X
GAS7	chr17:10102501	0,4	0,2	1,E-03	8,E-03	80425	chr17:9938001	0,4	0,6	0,6	2,E-03	5,E-02	25280	X	X
GATA2	chr3:128200501	0,6	0,3	4,E-05	6,E-04	41580	chr3:128199501	0,6	0,3	0,3	6,E-03	7,E-02	50474	X	
GATA3	chr10:8097501	0,8	0,1	2,E-17	8,E-14	137	chr10:8100001	1,1	0,7	0,5	2,E-03	5,E-02	22623	X	X
GMPS	chr3:155612001	0,4	0,3	1,E-02	5,E-02	146498	chr3:155612001	0,5	0,4	0,3	4,E-01	6,E-01	446507	X	
GNA11	chr19:3092501	0,6	0,4	2,E-07	8,E-06	17221	chr19:3092501	0,8	0,5	0,4	1,E-03	4,E-02	20720	X	X
GNAS	chr20:57471001	0,5	0,2	1,E-07	6,E-06	16325	chr20:57479501	0,7	0,5	0,4	3,E-04	2,E-02	7743	X	X
GPC3	chrX:133118001	0,8	0,5	9,E-07	3,E-05	21543	chrX:133118001	0,9	0,7	0,5	4,E-02	2,E-01	138298	X	
HIP1	chr7:75186501	0,9	0,7	4,E-04	4,E-03	67843	chr7:75192001	0,4	0,2	0,3	9,E-04	4,E-02	16618	X	X
HIST1H4I	chr6:27107001	1,0	0,3	3,E-12	5,E-10	3975	chr6:27107001	1,2	0,8	0,3	3,E-02	2,E-01	127263	X	
HLF	chr17:53342501	0,4	0,1	2,E-15	1,E-12	1058	chr17:53341501	0,8	0,5	0,3	4,E-02	2,E-01	144903	X	
HMGA1	chr6:34203501	0,5	0,1	2,E-15	1,E-12	1065	chr6:34202501	0,7	0,4	0,4	6,E-02	2,E-01	181734	X	
HMGA2	chr12:66220001	0,4	0,1	2,E-09	1,E-07	9102	chr12:66220001	0,5	0,3	0,1	8,E-03	9,E-02	61807	X	
HNF1A	chr12:121418501	0,6	0,4	2,E-04	2,E-03	54755	chr12:121416501	0,6	0,5	0,4	3,E-01	5,E-01	360453	X	
HOOK3	chr8:42750001	0,3	0,2	3,E-01	NA	5E+06	chr8:42750001	0,4	0,2	0,2	4,E-04	3,E-02	9470		X
HOXA11	chr7:27222501	0,9	0,5	1,E-05	3,E-04	34348	chr7:27222501	1,0	0,7	0,5	6,E-02	2,E-01	181874	X	
HOXA9	chr7:27219001	0,5	0,2	2,E-03	2,E-02	99668	chr7:27206501	0,4	0,1	0,2	1,E-03	4,E-02	19596	X	X
HOXC11	chr12:54366501	0,8	0,2	5,E-15	3,E-12	1388	chr12:54367001	1,2	0,7	0,2	2,E-01	4,E-01	313750	X	
HOXC13	chr12:54330501	0,5	0,2	1,E-05	2,E-04	32972	chr12:54334001	0,4	0,3	0,2	5,E-02	2,E-01	170672	X	
HOXD11	chr2:176969001	0,5	0,1	2,E-15	1,E-12	1076	chr2:176973501	0,9	0,7	0,4	2,E-01	4,E-01	295984	X	
HOXD13	chr2:176956501	1,1	0,7	8,E-09	5,E-07	10898	chr2:176956501	1,3	1,0	0,7	9,E-02	3,E-01	220160	X	
HRAS	chr11:533501	1,3	1,0	1,E-03	9,E-03	81888	chr11:533501	1,6	1,1	1,0	4,E-03	7,E-02	42099	X	
HSP90AA1	chr14:102554501	0,5	0,3	6,E-06	1,E-04	29433	chr14:102555001	0,6	0,4	0,3	8,E-02	3,E-01	212521	X	
HSP90AB1	chr6:44213001	0,6	0,5	2,E-03	2,E-02	98812	chr6:44213001	0,7	0,5	0,5	2,E-03	4,E-02	22357	X	X
IKZF1	chr7:50343001	0,6	0,2	5,E-09	3,E-07	10354	chr7:50435501	0,4	0,8	0,7	1,E-04	2,E-02	4199	X	X
IKZF4	chr12:56415001	1,2	0,7	1,E-09	1,E-07	8598	chr12:56416501	0,5	0,3	0,4	2,E-01	4,E-01	302229	X	
IRF4	chr6:392501	0,4	0,1	5,E-12	8,E-10	4308	chr6:392501	0,5	0,3	0,1	2,E-02	1,E-01	94316	X	
JAK1	chr1:65312001	1,0	0,8	4,E-03	2,E-02	114451	chr1:65306501	1,0	0,7	0,8	4,E-03	6,E-02	38524	X	
JAK3	chr19:17958501	0,5	0,1	1,E-12	2,E-10	3444	chr19:17957001	0,6	0,4	0,5	1,E-02	1,E-01	74125	X	
JAZF1	chr7:28218501	0,4	0,1	1,E-08	7,E-07	11562	chr7:27903501	1,0	0,5	0,7	3,E-04	2,E-02	7523	X	X

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
KDM5C	chrX:53228001	0,8	0,6	9,E-05	1,E-03	48471	chrX:53246501	0,4	0,3	0,4	2,E-01	4,E-01	297666	X	
KDR	chr4:55990501	1,0	0,4	3,E-14	1,E-11	1925	chr4:55990001	0,5	0,5	0,3	3,E-01	5,E-01	390254	X	
KIAA1549	chr7:138667001	0,6	0,3	2,E-08	1,E-06	12550	chr7:138667001	0,8	0,4	0,3	8,E-04	4,E-02	15526	X	X
KIT	chr4:55523501	0,5	0,1	4,E-15	2,E-12	1312	chr4:55523501	0,7	0,5	0,1	2,E-01	4,E-01	289216	X	
KLF6	chr10:3824001	0,7	0,6	1,E-03	1,E-02	86701	chr10:3824001	0,8	0,6	0,6	4,E-03	6,E-02	38567	X	
KLK2	chr19:51378001	0,3	0,5	2,E-06	4,E-05	23655	chr19:51374501	0,7	0,8	0,7	2,E-01	4,E-01	295746	X	
KRAS	chr12:25405001	0,6	0,5	5,E-05	8,E-04	44182	chr12:25400001	0,5	0,6	0,4	2,E-01	4,E-01	313725	X	
LASP1	chr17:37024501	0,5	0,3	2,E-05	3,E-04	35914	chr17:37031001	0,4	0,4	0,2	4,E-01	6,E-01	409093	X	
LCK	chr1:32739001	0,4	0,3	3,E-03	2,E-02	100531	chr1:32717501	0,7	0,5	0,5	6,E-03	8,E-02	51026	X	
LCP1	chr13:46716501	0,5	0,7	3,E-04	3,E-03	60869	chr13:46726001	0,4	0,6	0,6	5,E-04	3,E-02	10723	X	X
LIFR	chr5:38559001	0,4	0,3	4,E-03	2,E-02	111046	chr5:38595001	0,4	0,3	0,3	8,E-03	9,E-02	61037	X	
LMO1	chr11:8251501	0,9	0,6	1,E-05	3,E-04	33710	chr11:8290501	1,0	0,6	0,7	6,E-03	7,E-02	48435	X	
LMO2	chr11:33890001	0,4	0,2	4,E-08	2,E-06	13699	chr11:33892001	0,4	0,2	0,2	6,E-03	8,E-02	52612	X	
LPP	chr3:187958501	0,7	0,5	8,E-05	1,E-03	48223	chr3:188477501	0,7	0,5	0,5	4,E-02	2,E-01	137425	X	
MAFB	chr20:39319001	0,9	0,4	6,E-10	6,E-08	7910	chr20:39319501	1,1	0,7	0,6	8,E-04	3,E-02	14370	X	X
MAML2	chr11:96074501	0,5	0,1	2,E-15	1,E-12	1040	chr11:95825501	0,6	0,5	0,5	6,E-01	7,E-01	500699	X	
MDM4	chr1:204484501	0,6	0,5	4,E-03	2,E-02	109605	chr1:204505001	0,6	0,3	0,4	3,E-03	6,E-02	36317	X	
MEN1	chr11:64573501	0,8	0,7	1,E-02	5,E-02	143127	chr11:64573501	0,9	0,6	0,7	8,E-03	8,E-02	58577	X	
MET	chr7:116312501	0,4	0,2	1,E-09	1,E-07	8620	chr7:116313001	0,4	0,3	0,2	4,E-02	2,E-01	140692	X	
MITF	chr3:69789501	0,5	0,3	2,E-05	4,E-04	36942	chr3:69789501	0,7	0,3	0,3	5,E-03	7,E-02	43976	X	
MKL1	chr22:40811501	0,8	0,5	6,E-05	9,E-04	45222	chr22:40812001	1,1	0,8	0,7	3,E-03	6,E-02	33776	X	
MLF1	chr3:158289501	0,7	0,4	5,E-04	5,E-03	68911	chr3:158290001	2,1	1,6	1,6	4,E-02	2,E-01	140164	X	
MLL3	chr7:152007001	0,3	0,4	2,E-03	1,E-02	88998	chr7:151876001	0,5	0,4	0,4	1,E-01	3,E-01	263944	X	
MLLT1	chr19:6278001	0,4	0,3	1,E-02	6,E-02	153733	chr19:6278001	0,5	0,3	0,3	1,E-04	2,E-02	4313		X
MLLT4	chr6:168226001	0,6	0,2	4,E-14	1,E-11	2055	chr6:168333001	0,8	0,4	0,5	3,E-03	6,E-02	37066	X	
MLLT6	chr17:36866501	0,6	0,4	3,E-04	4,E-03	64040	chr17:36875501	0,6	0,4	0,4	1,E-03	4,E-02	20002	X	X
MN1	chr22:28198001	0,7	0,0	2,E-14	8,E-12	1860	chr22:28191001	1,5	1,0	1,0	1,E-03	4,E-02	20830	X	X
MNX1	chr7:156805001	0,5	0,3	5,E-08	2,E-06	13818	chr7:156801001	0,6	0,3	0,3	2,E-04	2,E-02	6486	X	X
MSH2	chr2:47629001	0,7	0,5	8,E-07	2,E-05	21087	chr2:47631501	0,9	0,7	0,7	2,E-02	1,E-01	111910	X	
MSH6	chr2:48011501	0,5	0,3	1,E-07	6,E-06	16152	chr2:48012501	0,4	0,5	0,5	6,E-02	2,E-01	180810	X	
MSI2	chr17:55362501	0,4	0,3	7,E-03	4,E-02	132674	chr17:55365001	0,6	0,4	0,4	4,E-02	2,E-01	136655	X	
MUC1	chr1:155158001	0,7	0,5	1,E-03	1,E-02	87019	chr1:155158001	0,8	0,7	0,5	2,E-01	4,E-01	282943	X	
MUTYH	chr1:45799501	0,5	0,4	2,E-02	6,E-02	161761	chr1:45797001	0,6	0,3	0,5	3,E-04	3,E-02	8549		X
MYC	chr8:128752501	0,7	0,6	7,E-01	8,E-01	589609	chr8:128752501	0,9	0,4	0,6	1,E-04	2,E-02	3923		X
MYCN	chr2:16083001	0,7	0,4	3,E-09	2,E-07	9467	chr2:16083001	0,9	0,7	0,4	1,E-01	3,E-01	248139	X	
MYH9	chr22:36745001	0,5	0,2	6,E-09	4,E-07	10617	chr22:36737501	0,5	0,3	0,3	5,E-04	3,E-02	10471	X	X
MYST4	chr10:76587001	0,9	0,6	4,E-06	9,E-05	27467	chr10:76597501	0,5	0,4	0,4	1,E-01	3,E-01	272051	X	
NBN	chr8:90988501	0,5	0,3	2,E-04	2,E-03	57169	chr8:90949501	0,7	0,5	0,4	9,E-02	3,E-01	218024	X	
NCKIPSD	chr3:48720001	0,9	0,6	2,E-04	2,E-03	54565	chr3:48725501	0,4	0,3	0,3	4,E-03	6,E-02	38026	X	
NCOA1	chr2:24713001	1,1	0,9	3,E-05	5,E-04	39539	chr2:24712501	0,5	0,4	0,4	1,E-02	1,E-01	74201	X	
NF1	chr17:29421001	0,6	0,4	2,E-05	4,E-04	37574	chr17:29684001	0,4	0,5	0,4	2,E-01	4,E-01	299173	X	
NFE2L2	chr2:178109501	0,4	0,2	9,E-07	3,E-05	21843	chr2:178113501	0,4	0,3	0,3	2,E-03	5,E-02	24568	X	X
NFIB	chr9:14308001	0,4	0,3	2,E-04	3,E-03	58305	chr9:14306501	0,7	0,5	0,5	2,E-02	1,E-01	90547	X	
NIN	chr14:51259501	1,0	0,8	2,E-03	1,E-02	89062	chr14:51296001	0,5	0,4	0,3	2,E-02	1,E-01	108973	X	
NKX2-1	chr14:36991001	0,5	0,1	7,E-15	3,E-12	1473	chr14:36987001	0,5	0,1	0,0	6,E-04	3,E-02	12760	X	X
NOTCH1	chr9:139420001	0,8	0,5	8,E-06	2,E-04	31147	chr9:139417001	0,9	0,5	0,5	2,E-03	5,E-02	25798	X	X
NR4A3	chr9:102587501	0,8	0,3	4,E-11	5,E-09	5695	chr9:102587501	1,0	0,6	0,3	9,E-04	4,E-02	16732	X	X
NSD1	chr5:176693001	1,5	1,2	3,E-03	2,E-02	106355	chr5:176558501	0,8	0,5	0,5	6,E-03	7,E-02	49257	X	
NTRK1	chr1:156784001	0,5	0,3	8,E-06	2,E-04	31498	chr1:156785501	0,7	0,4	0,4	8,E-04	4,E-02	15141	X	X
NTRK3	chr15:88797501	0,2	0,1	5,E-01	NA	2E+06	chr15:88798001	0,3	0,5	0,3	2,E-03	5,E-02	25216		X
NUMA1	chr11:71792501	0,4	0,5	3,E-04	3,E-03	60466	chr11:71789501	0,2	0,1	0,1	3,E-01	NA	1E+06	X	
NUP214	chr9:134104001	0,6	0,5	1,E-02	5,E-02	145917	chr9:134002001	0,5	0,3	0,4	3,E-03	6,E-02	37320	X	
NUP98	chr11:3720001	0,4	0,5	3,E-03	2,E-02	103816	chr11:3721001	0,4	0,3	0,5	4,E-02	2,E-01	152714	X	
OLIG2	chr21:34397501	0,5	0,1	1,E-13	3,E-11	2511	chr21:34398501	0,8	0,6	0,2	1,E-01	3,E-01	248503	X	
OMD	chr9:95188501	0,5	0,4	2,E-04	2,E-03	54591	chr9:95188501	0,6	0,5	0,4	7,E-02	2,E-01	191434	X	
PALB2	chr16:23653501	0,3	0,4	2,E-03	1,E-02	90787	chr16:23651001	1,2	1,0	0,8	3,E-01	5,E-01	376743	X	
PATZ1	chr22:31737001	0,2	0,1	3,E-03	NA	4E+06	chr22:31740501	0,5	0,1	0,1	3,E-04	2,E-02	7463		X
PAX3	chr2:223161501	1,0	0,3	1,E-16	2,E-13	389	chr2:223161501	1,1	1,1	0,3	1,E-01	3,E-01	277334	X	
PAX5	chr9:37036001	1,7	0,9	5,E-13	1,E-10	3077	chr9:37032001	0,6	0,5	0,3	7,E-02	2,E-01	194640	X	
PAX7	chr1:18959001	0,6	0,1	2,E-15	1,E-12	1132	chr1:18956001	1,1	0,9	0,6	1,E-01	3,E-01	268106	X	

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
PAX8	chr2:114034501	0,6	0,1	4,E-16	4,E-13	645	chr2:114036501	0,5	0,3	0,4	7,E-03	8,E-02	54561	X	
PBX1	chr1:164595001	0,4	0,2	5,E-04	5,E-03	69539	chr1:164599001	0,5	0,3	0,3	4,E-02	2,E-01	146833	X	
PCM1	chr8:17823501	0,6	0,9	8,E-05	1,E-03	47325	chr8:17779001	0,4	0,4	0,5	3,E-01	6,E-01	407213	X	
PDE4DIP	chr1:145039001	1,4	0,4	5,E-15	2,E-12	1345	chr1:145075001	1,7	1,3	1,0	4,E-03	7,E-02	40863	X	
PDGFB	chr22:39641501	0,9	0,6	1,E-06	4,E-05	23547	chr22:39635001	0,4	0,3	0,3	7,E-02	2,E-01	190384	X	
PDGFRA	chr4:55093501	0,7	0,2	5,E-13	1,E-10	3055	chr4:55097501	0,8	0,4	0,1	6,E-03	8,E-02	51926	X	
PDGFRB	chr5:149535501	0,9	0,2	8,E-17	2,E-13	361	chr5:149537001	0,5	0,4	0,3	3,E-02	2,E-01	129804	X	
PER1	chr17:8056001	0,7	0,4	2,E-04	3,E-03	59706	chr17:8053501	0,5	0,3	0,4	3,E-03	6,E-02	34541	X	
PHOX2B	chr4:41749001	0,9	0,2	6,E-17	1,E-13	289	chr4:41748501	0,5	0,5	0,1	3,E-01	5,E-01	358471	X	
PIK3R1	chr5:67521501	0,5	0,4	2,E-04	2,E-03	55086	chr5:67522501	0,6	0,5	0,5	3,E-01	5,E-01	366474	X	
PLAG1	chr8:57125001	0,4	0,2	5,E-04	5,E-03	68273	chr8:57125001	0,5	0,3	0,2	2,E-01	4,E-01	312600	X	
POU2AF1	chr11:111249501	0,5	0,2	5,E-12	8,E-10	4321	chr11:111251001	0,5	0,4	0,4	4,E-02	2,E-01	138786	X	
POU5F1	chr6:31139001	0,4	0,3	1,E-04	1,E-03	50192	chr6:31132501	0,6	0,4	0,4	5,E-03	7,E-02	45363	X	
PPARG	chr3:12328501	0,6	0,4	1,E-05	2,E-04	32310	chr3:12332501	0,5	0,7	0,6	3,E-02	2,E-01	121837	X	
PPP2R1A	chr19:52703001	0,4	0,3	6,E-04	5,E-03	71441	chr19:52714501	0,9	0,5	0,7	1,E-04	2,E-02	4713	X	X
PRCC	chr1:156756501	0,6	0,5	1,E-04	2,E-03	52802	chr1:156756501	0,7	0,6	0,5	1,E-01	3,E-01	260969	X	
PRDM16	chr1:3159501	0,4	0,6	3,E-06	7,E-05	25813	chr1:3154001	0,5	0,4	0,4	1,E-02	9,E-02	68180	X	
PRRX1	chr1:170631001	0,4	0,1	4,E-08	2,E-06	13583	chr1:170699501	0,2	0,5	0,4	3,E-04	2,E-02	7319	X	X
PSIP1	chr9:15508501	0,6	0,4	2,E-03	2,E-02	99899	chr9:15512001	0,4	0,5	0,4	4,E-02	2,E-01	149365	X	
PTCH1	chr9:98279501	0,5	0,2	6,E-05	9,E-04	45508	chr9:98279501	0,6	0,3	0,2	1,E-02	1,E-01	71472	X	
RAD51L1	chr14:68284001	0,7	0,5	7,E-03	4,E-02	128888	chr14:68285501	0,7	0,5	0,6	2,E-02	1,E-01	89056	X	
RALGDS	chr9:135998501	0,6	0,4	3,E-06	7,E-05	25971	chr9:135994501	0,6	0,2	0,2	5,E-05	1,E-02	2795	X	X
RAP1GDS1	chr4:99184001	0,5	0,4	5,E-03	3,E-02	121577	chr4:99184001	0,4	0,6	0,4	7,E-02	2,E-01	187502	X	
RARA	chr17:38465001	0,4	0,1	4,E-13	9,E-11	2939	chr17:38508001	0,7	0,5	0,4	8,E-03	8,E-02	58908	X	
RECQL4	chr8:145741501	0,6	0,4	8,E-03	4,E-02	133967	chr8:145741001	1,4	0,7	0,7	2,E-05	9,E-03	1709	X	X
RET	chr10:43600501	0,4	0,1	5,E-15	2,E-12	1331	chr10:43601501	1,2	0,9	1,0	1,E-02	1,E-01	80953	X	
RHOH	chr4:40198501	0,7	0,5	2,E-03	1,E-02	92489	chr4:40191001	0,9	0,6	0,7	9,E-05	2,E-02	3770	X	X
RNF213	chr17:78354501	1,0	0,8	5,E-05	8,E-04	43590	chr17:78233501	0,9	0,6	0,6	5,E-03	7,E-02	46097	X	
RPL22	chr1:6268501	0,7	0,4	1,E-05	3,E-04	34626	chr1:6267001	0,6	0,3	0,3	2,E-03	5,E-02	25022	X	X
RUNDC2A	chr16:12071001	0,4	0,3	3,E-03	2,E-02	105630	chr16:12069001	0,5	0,3	0,3	2,E-02	1,E-01	111456	X	
RUNX1	chr21:36263001	0,8	0,4	2,E-08	1,E-06	12233	chr21:36422501	0,7	0,4	0,5	1,E-04	2,E-02	4926	X	X
RUNX1T1	chr8:93074501	0,8	0,4	4,E-09	3,E-07	10007	chr8:93107501	0,8	0,5	0,4	9,E-04	4,E-02	16664	X	X
SDHC	chr1:161293501	0,4	0,4	4,E-01	6,E-01	478033	chr1:161283501	0,6	0,3	0,5	4,E-04	3,E-02	9564		X
SEPT5	chr22:19706501	0,5	0,3	5,E-08	2,E-06	13851	chr22:19701001	0,7	0,4	0,6	8,E-04	3,E-02	14890	X	X
SEPT6	chrX:118829001	0,6	0,4	2,E-03	1,E-02	94695	chrX:118829001	0,7	0,4	0,4	3,E-03	6,E-02	35212	X	
SEPT9	chr17:75370001	1,1	0,1	4,E-16	4,E-13	633	chr17:75282001	1,0	0,3	0,4	5,E-07	3,E-03	130	X	X
SET	chr9:131450501	0,4	0,2	9,E-04	7,E-03	78144	chr9:131448501	0,3	0,5	0,3	6,E-03	8,E-02	52360	X	
SFPQ	chr1:35654501	0,3	0,4	5,E-04	5,E-03	69013	chr1:35652501	0,6	0,5	0,5	1,E-02	9,E-02	68217	X	
SLC45A3	chr1:205632001	0,3	0,5	6,E-07	2,E-05	20317	chr1:205629501	0,2	0,2	0,2	5,E-01	NA	1E+06	X	
SMARCA4	chr19:11096501	0,6	0,5	6,E-02	2,E-01	238662	chr19:11095501	1,2	0,7	0,8	2,E-04	2,E-02	5898		X
SMARCB1	chr22:24133501	0,6	0,3	2,E-11	2,E-09	5021	chr22:24133501	0,7	0,5	0,3	2,E-02	1,E-01	87770	X	
SMO	chr7:128829501	0,6	0,4	8,E-05	1,E-03	48079	chr7:128851001	0,7	0,3	0,4	6,E-04	3,E-02	13080	X	X
SOCS1	chr16:11348501	0,6	0,2	3,E-11	4,E-09	5452	chr16:11351501	0,5	0,3	0,5	3,E-03	6,E-02	32472	X	
SOX2	chr3:181428501	1,1	0,6	2,E-07	9,E-06	17590	chr3:181429001	0,4	0,2	0,2	1,E-02	1,E-01	76374	X	
SRGAP3	chr3:9289501	0,6	0,2	2,E-11	3,E-09	5289	chr3:9258001	1,3	0,9	1,0	1,E-03	4,E-02	20157	X	X
SS18L1	chr20:60732001	0,4	0,3	3,E-03	2,E-02	103828	chr20:60731501	0,4	0,2	0,2	4,E-04	3,E-02	9062	X	X
SYK	chr9:93591001	0,4	0,3	5,E-03	3,E-02	119345	chr9:93588001	0,5	0,4	0,4	3,E-01	5,E-01	391186	X	
TAL1	chr1:47696501	0,8	0,1	9,E-17	2,E-13	378	chr1:47695001	1,1	0,7	0,4	2,E-02	1,E-01	97076	X	
TCEA1	chr8:54936001	0,6	0,5	8,E-03	4,E-02	134299	chr8:54935501	1,0	0,6	0,6	5,E-03	7,E-02	46726	X	
TCF3	chr19:1654001	0,5	0,3	3,E-04	3,E-03	61831	chr19:1649001	1,1	0,8	0,8	4,E-03	7,E-02	41292	X	
TCL6	chr14:96136001	0,4	0,5	4,E-02	1,E-01	221167	chr14:96137501	0,3	0,5	0,5	2,E-03	5,E-02	25194		X
TET1	chr10:70321501	1,1	0,3	8,E-16	6,E-13	799	chr10:70321501	1,4	0,9	0,3	2,E-02	1,E-01	90927	X	
TET2	chr4:106067001	0,9	0,1	7,E-17	1,E-13	321	chr4:106067001	1,1	0,7	0,1	9,E-02	3,E-01	217324	X	
TFEB	chr6:41671501	0,6	0,4	8,E-04	7,E-03	76638	chr6:41691001	0,8	0,4	0,5	1,E-04	2,E-02	4197	X	X
TFPT	chr19:54617501	0,3	0,5	3,E-04	3,E-03	61949	chr19:54620501	0,5	0,4	0,4	2,E-02	1,E-01	93740	X	
TFRC	chr3:195798501	0,5	0,4	3,E-03	2,E-02	107856	chr3:195784501	0,4	0,2	0,3	4,E-03	6,E-02	39202	X	
TLX1	chr10:102892501	0,9	0,3	4,E-14	1,E-11	2025	chr10:102891001	0,4	0,1	0,0	6,E-04	3,E-02	11902	X	X
TLX3	chr5:170737501	0,9	0,3	4,E-17	1,E-13	248	chr5:170735001	1,2	0,8	0,2	3,E-01	5,E-01	362654	X	
TMPRSS2	chr21:42866001	0,7	1,1	9,E-07	3,E-05	21624	chr21:42841001	1,0	0,8	1,0	7,E-02	2,E-01	193522	X	
TNFRSF14	chr1:2494501	0,8	0,6	4,E-03	2,E-02	113844	chr1:2494501	1,1	0,7	0,6	5,E-04	3,E-02	11032	X	X

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
TNFRSF17	chr16:12057501	0,4	0,3	2,E-03	2,E-02	98701	chr16:12060001	0,4	0,5	0,4	7,E-01	9,E-01	569401	X	
TOP1	chr20:39656001	0,5	0,4	2,E-01	4,E-01	364387	chr20:39655501	0,4	0,3	0,4	2,E-03	5,E-02	24591		X
TPM4	chr19:16186501	0,7	0,2	2,E-14	8,E-12	1809	chr19:16177001	0,5	0,3	0,3	1,E-03	4,E-02	18744	X	X
TRIM24	chr7:138143501	0,8	0,6	2,E-03	1,E-02	93932	chr7:138238501	0,4	0,3	0,3	3,E-02	2,E-01	130102	X	
TSC1	chr9:135821001	0,5	0,4	5,E-03	3,E-02	119349	chr9:135797001	1,2	0,5	0,7	1,E-05	7,E-03	1170	X	X
TSC2	chr16:2122001	0,5	0,3	3,E-04	3,E-03	59838	chr16:2105001	1,3	0,7	1,1	2,E-04	2,E-02	5271	X	X
USP6	chr17:5019501	0,9	0,2	5,E-15	3,E-12	1375	chr17:5019001	0,5	0,3	0,1	8,E-02	2,E-01	202711	X	
VHL	chr3:10184501	0,8	0,9	7,E-02	2,E-01	251805	chr3:10184501	1,0	0,5	0,9	8,E-04	3,E-02	14911		X
WHSC1	chr4:1958001	0,5	0,3	1,E-03	9,E-03	82753	chr4:1979001	0,9	0,6	0,7	8,E-04	3,E-02	14427	X	X
WIF1	chr12:65515501	2,0	1,5	2,E-03	1,E-02	95996	chr12:65515501	2,5	1,8	1,5	6,E-02	2,E-01	173714	X	
WRN	chr8:30889501	0,7	0,9	2,E-03	2,E-02	98203	chr8:30889001	0,3	0,4	0,4	9,E-02	3,E-01	224874	X	
WT1	chr11:32455001	1,0	0,2	3,E-15	2,E-12	1202	chr11:32459001	0,4	0,7	0,2	3,E-04	3,E-02	8592	X	X
XPA	chr9:100461001	0,8	0,5	7,E-03	4,E-02	131880	chr9:100461001	0,8	0,7	0,5	1,E-01	3,E-01	264214	X	
XPC	chr3:14189001	0,8	0,6	4,E-03	3,E-02	116648	chr3:14189001	0,9	0,7	0,6	2,E-02	1,E-01	107461	X	
ZBTB16	chr11:113929001	0,9	0,5	3,E-11	4,E-09	5409	chr11:113929001	1,0	0,8	0,5	1,E-02	1,E-01	77361	X	
ZNF331	chr19:54059001	0,5	0,3	2,E-06	6,E-05	24702	chr19:54025001	0,5	0,2	0,3	1,E-03	4,E-02	18760	X	X

Supplementary Table 6: Genes with at least one differentially methylated region (BH < 0.05) between tumour and normal or between FUS- and FUS+ samples from the homeobox gene list (HOLLAND 2007). For each gene the most significant bin (according to p-values) of all TSS associated bins is shown. Column headings are as follows: (Name) Gene name, (best DMR tumour) Chromosomal location of best tumour DMR, (T) mean MeDIP rpm value in tumour samples, (N) mean MeDIP rpm value of normal samples, (p_{MW} T/N) Mann-Whitney p-value tumour vs. normal, (p_{BH} T/N) p_{MW}-value after Benjamini-Hochberg correction (over all significantly methylated bins), (Rank) rank of bin in ordered list of all genomic bins (according to p_{BH}-values), (best DMR FUS) Chromosomal location of best FUS- DMR, (Fus-) mean MeDIP rpm value of FUS- samples, (Fus+) mean MeDIP rpm value of FUS+ samples, (N) mean MeDIP rpm value of normal samples, (p_{MW} Fus-/+) Mann-Whitney p-value FUS- vs. FUS+, (p_{BH} Fus-/+) p-value after Benjamini-Hochberg correction (over all significantly methylated bins), (Rank) rank of bin in ordered list of all genomic bins (according to p_{BH}-values), (DMR T/N) DMR significant in tumour, (DMR Fus-/+) DMR significant in FUS-.

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
ALX1	chr12:85673001	1,1	0,2	2,9E-17	9,8E-14	198	chr12:85672001	0,5	0,4	0,2	7,7E-02	2,5E-01	202170	x	
ALX3	chr1:110611001	1,4	0,4	2,1E-15	1,3E-12	1110	chr1:110612501	1,1	0,6	0,3	6,7E-03	7,9E-02	53921	x	
ALX4	chr11:44333001	0,9	0,2	3,8E-16	4,2E-13	625	chr11:44330501	0,4	0,1	0,1	5,5E-03	7,3E-02	48449	x	
ARGFX	chr3:121291001	0,4	0,4	5,1E-01	6,7E-01	513160	chr3:121291001	0,5	0,3	0,4	5,3E-04	3,0E-02	11709		x
ARX	chrX:25033501	0,4	0,1	2,9E-09	2,1E-07	9563	chrX:25032001	0,5	0,7	0,5	6,7E-02	2,3E-01	191531	x	
BARHL1	chr9:135456001	0,5	0,1	4,7E-14	1,5E-11	2109	chr9:135456501	0,4	0,4	0,1	6,9E-01	8,2E-01	551812	x	
BARHL2	chr1:91182501	0,7	0,2	1,0E-14	4,4E-12	1587	chr1:91182001	0,2	0,4	0,1	2,2E-03	5,1E-02	27527	x	
BARX1	chr9:96715001	0,6	0,1	6,4E-16	5,8E-13	753	chr9:96716501	0,8	0,5	0,3	1,1E-01	3,0E-01	246191	x	
BARX2	chr11:129244001	1,4	0,5	1,2E-13	3,3E-11	2519	chr11:129244501	0,7	0,5	0,3	7,4E-02	2,4E-01	198745	x	
BSX	chr11:122850001	0,4	0,2	2,0E-11	2,7E-09	5172	chr11:122853001	0,5	0,6	0,5	4,0E-01	6,1E-01	432571	x	
CDX2	chr13:28544501	0,7	0,2	5,8E-14	1,8E-11	2201	chr13:28543501	0,4	0,3	0,2	5,9E-02	2,2E-01	176879	x	
CDX4	chrX:72667001	0,5	0,2	2,5E-11	3,2E-09	5330	chrX:72667001	0,4	0,6	0,2	5,1E-02	2,0E-01	167013	x	
CRX	chr19:48323001	0,4	0,3	1,1E-03	9,1E-03	82788	chr19:48326001	0,4	0,3	0,4	2,2E-02	1,3E-01	107058	x	
CUX1	chr7:101462001	0,5	0,3	1,1E-07	4,8E-06	15732	chr7:101746001	1,5	1,0	1,1	4,8E-04	2,9E-02	10930	x	x
CUX2	chr12:111536001	0,7	0,4	3,1E-08	1,6E-06	13144	chr12:111536001	0,9	0,6	0,4	3,3E-03	5,9E-02	35412	x	
DBX1	chr11:20181001	0,6	0,2	1,5E-14	6,0E-12	1695	chr11:20181001	0,5	0,7	0,2	2,9E-02	1,5E-01	123133	x	
DBX2	chr12:45444001	0,4	0,2	1,7E-06	4,9E-05	24033	chr12:45444001	0,5	0,4	0,2	6,3E-01	7,8E-01	524637	x	
DLX1	chr2:172948001	0,7	0,2	1,0E-11	1,5E-09	4710	chr2:172951501	0,4	0,3	0,1	4,5E-02	1,9E-01	156682	x	
DLX2	chr2:172966001	0,7	0,1	8,0E-16	6,7E-13	815	chr2:172965501	0,9	0,6	0,2	2,0E-01	4,1E-01	318834	x	
DLX3	chr17:48071501	0,9	0,4	8,1E-14	2,4E-11	2336	chr17:48072001	0,6	0,3	0,2	1,2E-02	1,0E-01	75792	x	
DLX4	chr17:48050001	0,9	0,5	4,0E-08	2,0E-06	13599	chr17:48050001	1,1	0,7	0,5	4,1E-02	1,8E-01	147754	x	
DLX5	chr7:96652001	1,0	0,4	4,5E-14	1,5E-11	2090	chr7:96654501	0,4	0,2	0,2	1,8E-03	4,7E-02	24807	x	x
DLX6	chr7:96632501	1,2	0,3	6,3E-17	1,4E-13	304	chr7:96636001	1,0	0,5	0,3	2,0E-04	2,1E-02	6215	x	x
DRGX	chr10:50604501	0,5	0,2	6,7E-12	1,0E-09	4477	chr10:50597501	0,3	0,6	0,5	1,6E-04	1,9E-02	5226	x	x
EMX1	chr2:73147001	0,6	0,1	4,8E-15	2,4E-12	1352	chr2:73152501	0,5	0,3	0,1	2,2E-02	1,3E-01	107139	x	
EMX2	chr10:119301001	0,7	0,2	3,8E-16	4,1E-13	624	chr10:119303001	0,5	0,2	0,1	1,8E-02	1,2E-01	95087	x	
EN1	chr2:119607001	0,6	0,2	6,8E-13	1,4E-10	3214	chr2:119607001	0,5	0,7	0,2	1,1E-01	3,0E-01	241143	x	
EN2	chr7:155252001	0,4	0,1	2,3E-10	2,3E-08	6992	chr7:155252001	0,5	0,4	0,1	8,2E-01	9,0E-01	595518	x	
EVX1	chr7:27283001	0,8	0,3	1,2E-12	2,3E-10	3515	chr7:27283001	1,1	0,5	0,3	2,1E-05	8,8E-03	1559	x	x
EVX2	chr2:176948001	0,7	0,1	1,7E-16	2,4E-13	482	chr2:176949001	0,5	0,7	0,3	5,7E-02	2,1E-01	174682	x	
GBX1	chr7:150871501	0,5	0,3	1,6E-05	3,1E-04	35312	chr7:150870001	0,7	0,6	0,9	2,6E-02	1,5E-01	117472	x	
GBX2	chr2:237078501	0,8	0,2	1,8E-17	7,9E-14	150	chr2:237074501	0,8	0,5	0,2	2,0E-02	1,3E-01	100424	x	
GSC	chr14:95235001	1,1	0,1	1,1E-16	1,8E-13	399	chr14:95234001	0,6	0,3	0,2	4,6E-03	6,8E-02	43523	x	
GSC2	chr22:19136501	0,4	0,1	4,6E-13	1,0E-10	3037	chr22:19135501	0,8	0,5	0,6	5,5E-03	7,3E-02	49024	x	
GSX1	chr13:28367501	0,7	0,1	7,8E-17	1,6E-13	339	chr13:28367501	0,9	0,6	0,1	2,5E-02	1,4E-01	113615	x	
GSX2	chr4:54967501	0,8	0,1	1,7E-14	6,6E-12	1732	chr4:54965501	0,9	0,4	0,1	3,8E-03	6,3E-02	39224	x	
HHEX	chr10:94448501	0,7	0,2	6,0E-05	9,1E-04	45002	chr10:94448501	1,2	0,4	0,2	2,5E-03	5,4E-02	30298	x	
HLX	chr1:221054501	1,2	0,4	3,7E-15	2,0E-12	1268	chr1:221054501	1,4	0,9	0,4	4,4E-03	6,6E-02	42041	x	
HMX1	chr4:8872501	0,5	0,3	2,8E-04	3,1E-03	61320	chr4:8874501	0,4	0,3	0,1	3,8E-01	5,8E-01	422542	x	
HMX2	chr10:124908501	0,4	0,2	5,6E-07	1,9E-05	20034	chr10:124909501	0,4	0,2	0,1	4,3E-02	1,8E-01	149880	x	
HMX3	chr10:124896501	0,5	0,1	9,8E-10	8,1E-08	8328	chr10:124896501	0,9	0,3	0,1	4,6E-03	6,8E-02	43361	x	
HNF1A	chr12:121418501	0,6	0,4	1,6E-04	2,0E-03	54755	chr12:121416501	0,6	0,5	0,4	2,7E-01	4,8E-01	360453	x	
HNF1B	chr17:36106001	0,6	0,2	1,3E-12	2,5E-10	3560	chr17:36103501	0,4	0,1	0,1	2,6E-04	2,3E-02	7308	x	x
HOPX	chr4:57522501	0,7	0,2	1,2E-15	9,0E-13	945	chr4:57521001	0,5	0,2	0,1	6,1E-03	7,6E-02	51928	x	
HOXA1	chr7:27135001	0,5	0,1	1,7E-13	4,3E-11	2643	chr7:27134001	1,8	1,3	1,4	2,3E-03	5,2E-02	29051	x	
HOXA11	chr7:27222501	0,9	0,5	1,4E-05	2,8E-04	34348	chr7:27222501	1,0	0,7	0,5	6,1E-02	2,2E-01	181874	x	
HOXA2	chr7:27142001	1,1	0,8	3,9E-06	9,6E-05	27572	chr7:27141501	0,6	0,3	0,3	5,3E-04	3,0E-02	11776	x	x

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
HOXA3	chr7:27155001	1,2	0,4	5,4E-14	1,7E-11	2174	chr7:27155001	1,6	1,0	0,4	8,8E-03	8,9E-02	64458	x	
HOXA4	chr7:27191501	0,6	0,3	1,2E-05	2,4E-04	33269	chr7:27194001	1,1	0,9	0,9	3,6E-03	6,2E-02	38220	x	
HOXA5	chr7:27183001	1,0	0,5	3,8E-08	1,9E-06	13467	chr7:27184001	1,7	1,2	1,1	2,0E-03	4,9E-02	26424	x	x
HOXA6	chr7:27187001	1,2	0,6	6,5E-08	3,1E-06	14567	chr7:27185001	1,9	1,3	1,2	5,5E-03	7,3E-02	49619	x	
HOXA7	chr7:27196001	0,7	0,1	1,8E-16	2,5E-13	486	chr7:27198001	1,6	1,0	1,1	2,0E-04	2,1E-02	6206	x	x
HOXA9	chr7:27219001	0,5	0,2	2,4E-03	1,7E-02	99668	chr7:27206501	0,4	0,1	0,2	1,2E-03	4,1E-02	19596	x	x
HOXB1	chr17:46607001	0,5	0,4	1,6E-03	1,2E-02	89930	chr17:46607001	0,6	0,4	0,4	3,4E-03	6,0E-02	36616	x	
HOXB2	chr17:46619001	0,6	0,3	1,2E-08	7,2E-07	11559	chr17:46622001	0,8	0,5	0,3	1,6E-02	1,2E-01	89294	x	
HOXB3	chr17:46656001	1,0	0,3	1,7E-14	6,6E-12	1734	chr17:46629001	1,1	0,9	0,9	6,1E-03	7,6E-02	51444	x	
HOXB4	chr17:46656001	1,0	0,3	1,7E-14	6,6E-12	1734	chr17:46657501	0,7	0,6	0,3	4,7E-02	1,9E-01	159099	x	
HOXB5	chr17:46669501	1,1	0,7	1,7E-06	4,8E-05	24030	chr17:46672501	0,9	0,6	0,6	2,7E-03	5,5E-02	31542	x	
HOXB6	chr17:46673501	1,5	0,9	5,1E-08	2,5E-06	14098	chr17:46672501	0,9	0,6	0,6	2,7E-03	5,5E-02	31542	x	
HOXB7	chr17:46711001	0,5	0,1	1,6E-11	2,1E-09	5001	chr17:46711001	0,7	0,4	0,1	1,7E-02	1,2E-01	91426	x	
HOXB8	chr17:46690501	0,5	0,1	5,1E-11	6,0E-09	5798	chr17:46688501	0,5	0,3	0,2	9,1E-02	2,7E-01	219666	x	
HOXB9	chr17:46699001	1,5	0,8	6,3E-12	9,6E-10	4441	chr17:46699001	1,7	1,4	0,8	5,9E-02	2,2E-01	177366	x	
HOXC10	chr12:54379001	0,8	0,2	5,9E-16	5,5E-13	728	chr12:54379501	0,7	0,4	0,1	1,0E-02	9,4E-02	68431	x	
HOXC11	chr12:54366501	0,8	0,2	5,5E-15	2,7E-12	1388	chr12:54367001	1,2	0,7	0,2	2,0E-01	4,0E-01	313750	x	
HOXC12	chr12:54350001	0,6	0,2	4,7E-10	4,2E-08	7634	chr12:54350001	0,7	0,5	0,2	1,5E-01	3,5E-01	276070	x	
HOXC13	chr12:54330501	0,5	0,2	1,1E-05	2,3E-04	32972	chr12:54334001	0,4	0,3	0,2	5,5E-02	2,1E-01	170672	x	
HOXC4	chr12:54447501	0,8	0,1	8,9E-18	6,8E-14	20	chr12:54409001	0,5	0,2	0,1	1,3E-03	4,2E-02	19911	x	x
HOXC5	chr12:54425001	1,1	0,1	6,1E-17	1,4E-13	299	chr12:54409001	0,5	0,2	0,1	1,3E-03	4,2E-02	19911	x	x
HOXC6	chr12:54423501	1,4	0,4	1,7E-17	7,9E-14	147	chr12:54409001	0,5	0,2	0,1	1,3E-03	4,2E-02	19911	x	x
HOXC8	chr12:54400501	0,4	0,1	1,6E-11	2,2E-09	5037	chr12:54400501	0,5	0,4	0,1	3,2E-01	5,3E-01	392245	x	
HOXC9	chr12:54388001	0,5	0,1	3,6E-14	1,2E-11	2013	chr12:54394501	0,4	0,2	0,1	2,6E-02	1,5E-01	115907	x	
HOXD1	chr2:177054001	1,2	0,1	1,1E-15	8,1E-13	920	chr2:177054001	1,4	1,1	0,1	4,6E-01	6,6E-01	462049	x	
HOXD10	chr2:176974001	0,7	0,4	2,0E-09	1,5E-07	9130	chr2:176979001	0,7	0,5	0,6	2,3E-03	5,2E-02	28810	x	
HOXD11	chr2:176969001	0,5	0,1	1,9E-15	1,2E-12	1076	chr2:176973501	0,9	0,7	0,4	1,7E-01	3,8E-01	295984	x	
HOXD13	chr2:176956501	1,1	0,7	7,9E-09	4,9E-07	10898	chr2:176956501	1,3	1,0	0,7	9,1E-02	2,7E-01	220160	x	
HOXD3	chr2:177029501	0,9	0,1	2,0E-15	1,3E-12	1087	chr2:177017001	1,0	0,6	0,6	4,8E-03	6,9E-02	45046	x	
HOXD4	chr2:177016001	0,8	0,5	4,1E-06	1,0E-04	27806	chr2:177017001	1,0	0,6	0,6	4,8E-03	6,9E-02	45046	x	
HOXD8	chr2:176993501	1,0	0,4	5,2E-11	6,1E-09	5816	chr2:176993001	1,6	1,1	1,0	3,5E-02	1,7E-01	137044	x	
HOXD9	chr2:176987001	0,6	0,4	7,9E-04	7,0E-03	76837	chr2:176989501	0,4	0,3	0,3	4,0E-02	1,8E-01	145329	x	
IRX1	chr5:3596501	0,4	0,1	1,8E-11	2,4E-09	5106	chr5:3597501	0,7	0,9	0,5	7,6E-03	8,4E-02	59428	x	
IRX2	chr5:2753001	0,7	0,3	4,0E-09	2,8E-07	10003	chr5:2754501	0,7	0,6	0,3	3,8E-01	5,8E-01	422663	x	
IRX3	chr16:54321501	0,6	0,2	5,8E-10	5,1E-08	7806	chr16:54322001	1,2	0,9	0,9	4,9E-02	2,0E-01	162036	x	
IRX4	chr5:1879001	0,6	0,3	5,9E-08	2,8E-06	14366	chr5:1880501	0,6	0,4	0,4	2,5E-03	5,4E-02	30867	x	
IRX5	chr16:54966501	0,8	0,1	1,2E-13	3,3E-11	2506	chr16:54966501	0,7	1,0	0,1	7,0E-02	2,4E-01	192635	x	
IRX6	chr16:55356501	0,3	0,4	1,5E-03	1,1E-02	88564	chr16:55356501	0,3	0,4	0,4	2,0E-02	1,3E-01	99958	x	
ISL1	chr5:50683001	0,7	0,3	1,7E-09	1,3E-07	8974	chr5:50677501	0,3	0,6	0,2	1,7E-04	2,0E-02	5554	x	x
ISL2	chr15:76628001	0,8	0,3	2,4E-15	1,4E-12	1135	chr15:76627001	0,9	0,5	0,6	1,4E-03	4,3E-02	20621	x	x
LBX1	chr10:102987001	0,7	0,1	1,3E-14	5,3E-12	1655	chr10:102987001	1,0	0,5	0,1	1,9E-02	1,3E-01	97232	x	
LBX2	chr2:74726501	0,5	0,1	1,2E-15	8,9E-13	941	chr2:74730501	1,3	0,9	0,5	7,3E-03	8,2E-02	57537	x	
LEUTX	chr19:40265001	0,4	0,6	1,2E-02	5,2E-02	149536	chr19:40268001	0,4	0,6	0,6	1,8E-03	4,7E-02	24508		x
LHX1	chr17:35292001	0,4	0,1	6,1E-14	1,9E-11	2227	chr17:35294001	0,4	0,3	0,1	2,0E-02	1,3E-01	100024	x	
LHX2	chr9:126774501	0,6	0,0	2,5E-18	6,8E-14	88	chr9:126779001	1,0	0,7	0,2	1,4E-02	1,1E-01	82583	x	
LHX3	chr9:139094001	0,5	0,2	1,8E-11	2,4E-09	5093	chr9:139096001	0,5	0,4	0,2	2,9E-01	5,0E-01	375307	x	
LHX5	chr12:113910501	0,5	0,2	2,3E-06	6,3E-05	25252	chr12:113911001	0,6	0,6	0,4	8,0E-01	8,9E-01	585077	x	
LHX6	chr9:124988501	1,3	0,4	9,4E-16	7,4E-13	865	chr9:124991501	0,7	0,4	0,2	2,7E-03	5,5E-02	32135	x	
LHX8	chr1:75598501	0,5	0,2	1,5E-06	4,3E-05	23559	chr1:75598501	0,3	0,7	0,2	1,2E-03	4,0E-02	18509	x	x
LHX9	chr1:197879501	1,2	0,2	1,3E-17	7,6E-14	106	chr1:197882001	1,0	0,6	0,2	6,4E-03	7,8E-02	52500	x	
LMX1A	chr1:165323501	1,1	0,2	1,7E-16	2,4E-13	474	chr1:165326001	0,9	0,5	0,3	1,4E-03	4,3E-02	21181	x	x
LMX1B	chr9:129377001	0,5	0,1	4,0E-11	4,9E-09	5648	chr9:129377001	0,8	0,4	0,1	2,5E-03	5,4E-02	31181	x	
MEIS1	chr2:66666501	0,7	0,1	7,4E-16	6,4E-13	785	chr2:66670001	0,4	0,5	0,4	5,5E-02	2,1E-01	171624	x	
MEIS2	chr15:37390001	0,8	0,2	4,9E-16	4,9E-13	676	chr15:37389501	0,4	0,5	0,1	1,7E-01	3,8E-01	298867	x	
MEIS3	chr19:47921501	0,6	0,2	4,0E-09	2,8E-07	9991	chr19:47921001	0,7	0,3	0,2	6,0E-04	3,1E-02	12502	x	x
MEIS3P1	chr17:15690001	0,4	0,3	2,7E-03	1,8E-02	101938	chr17:15690001	0,5	0,3	0,3	3,7E-02	1,7E-01	139313	x	
MEOX2	chr7:15725001	0,6	0,1	1,5E-14	6,0E-12	1701	chr7:15724501	0,5	0,6	0,3	1,4E-01	3,4E-01	271097	x	
MKX	chr10:28035501	0,7	0,1	2,1E-16	2,8E-13	513	chr10:28035001	0,7	0,4	0,1	2,6E-02	1,5E-01	115711	x	
MNX1	chr7:156805001	0,5	0,3	4,5E-08	2,2E-06	13818	chr7:156801001	0,6	0,3	0,3	2,1E-04	2,1E-02	6486	x	x
MSX1	chr4:4862501	0,6	0,1	5,1E-16	5,0E-13	694	chr4:4860001	0,9	0,6	0,4	5,0E-03	7,1E-02	46480	x	

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
MSX2	chr5:174152501	0,5	0,3	1,9E-04	2,3E-03	56407	chr5:174152501	0,7	0,5	0,3	1,4E-01	3,4E-01	270847	x	
NKX1-1	chr4:1399001	0,5	0,2	6,4E-14	1,9E-11	2242	chr4:1399001	0,5	0,6	0,2	1,1E-01	2,9E-01	238214	x	
NKX1-2	chr10:126137001	0,5	0,3	5,4E-07	1,8E-05	19918	chr10:126136501	0,5	0,3	0,3	1,7E-01	3,7E-01	290729	x	
NKX2-1	chr14:36991001	0,5	0,1	7,3E-15	3,4E-12	1473	chr14:36987001	0,5	0,1	0,0	6,2E-04	3,2E-02	12760	x	x
NKX2-2	chr20:21494001	0,4	0,1	7,1E-15	3,3E-12	1464	chr20:21496501	0,6	0,6	0,2	2,5E-01	4,7E-01	354028	x	
NKX2-3	chr10:101293501	1,2	0,3	1,6E-17	7,8E-14	138	chr10:101293001	0,6	0,3	0,1	7,5E-04	3,4E-02	14187	x	x
NKX2-4	chr20:21378501	0,6	0,2	1,1E-12	2,2E-10	3469	chr20:21378001	0,6	0,4	0,2	4,9E-02	2,0E-01	162735	x	
NKX2-5	chr5:172663501	1,3	0,3	1,4E-17	7,7E-14	124	chr5:172660001	0,4	0,1	0,1	2,8E-05	9,9E-03	1874	x	x
NKX2-6	chr8:23564001	0,9	0,1	7,9E-17	1,6E-13	351	chr8:23562501	0,9	0,8	0,2	2,2E-01	4,4E-01	336097	x	
NKX2-8	chr14:37049501	0,6	0,2	6,5E-13	1,4E-10	3196	chr14:37049501	0,8	0,5	0,2	4,2E-03	6,5E-02	41103	x	
NKX3-2	chr4:13544001	0,9	0,3	1,0E-12	2,0E-10	3428	chr4:13544001	1,1	0,7	0,3	8,5E-02	2,6E-01	213947	x	
NKX6-1	chr4:85417501	1,0	0,2	1,9E-17	8,3E-14	164	chr4:85421501	0,5	0,6	0,2	2,1E-01	4,2E-01	323363	x	
NKX6-2	chr10:134598501	0,4	0,1	1,0E-15	7,7E-13	893	chr10:134598501	0,6	0,3	0,1	9,1E-03	9,1E-02	64984	x	
NKX6-3	chr8:41504501	0,5	0,4	1,9E-02	7,6E-02	173412	chr8:41508001	0,8	0,4	0,5	1,2E-04	1,7E-02	4390		x
NOBOX	chr7:144105001	0,4	0,2	1,7E-08	9,8E-07	12124	chr7:144105501	0,6	0,4	0,4	3,2E-04	2,5E-02	8449	x	x
NOTO	chr2:73427001	0,1	0,2	9,8E-02	NA	3059808	chr2:73430501	0,5	0,2	0,1	3,2E-04	2,5E-02	8347		x
ONECUT1	chr15:53083001	0,5	0,1	1,0E-15	7,7E-13	890	chr15:53083001	0,7	0,4	0,1	5,5E-02	2,1E-01	170942	x	
ONECUT2	chr18:55103501	0,6	0,1	6,2E-13	1,3E-10	3178	chr18:55103501	0,8	0,6	0,1	2,6E-01	4,7E-01	357353	x	
OTP	chr5:76934001	0,4	0,1	8,5E-14	2,5E-11	2362	chr5:76931001	0,3	0,5	0,2	3,4E-02	1,7E-01	135007	x	
OTX1	chr2:63275001	1,0	0,4	6,1E-13	1,3E-10	3168	chr2:63275501	0,8	0,8	0,2	2,1E-01	4,2E-01	322551	x	
OTX2	chr14:57275001	0,9	0,2	3,2E-16	3,6E-13	597	chr14:57279001	0,9	0,6	0,1	1,5E-02	1,1E-01	87051	x	
PAX2	chr10:102506001	0,5	0,3	1,0E-06	3,1E-05	22060	chr10:102568001	0,5	0,4	0,4	9,3E-02	2,7E-01	222219	x	
PAX3	chr2:223161501	1,0	0,3	1,0E-16	1,8E-13	389	chr2:223161501	1,1	1,1	0,3	1,5E-01	3,5E-01	277334	x	
PAX5	chr9:37036001	1,7	0,9	5,0E-13	1,1E-10	3077	chr9:37032001	0,6	0,5	0,3	7,0E-02	2,4E-01	194640	x	
PAX6	chr11:31841501	0,6	0,1	2,3E-17	8,6E-14	178	chr11:31838001	0,6	0,4	0,1	1,2E-02	1,0E-01	77310	x	
PAX7	chr1:18959001	0,6	0,1	2,3E-15	1,4E-12	1132	chr1:18956001	1,1	0,9	0,6	1,4E-01	3,4E-01	268106	x	
PAX8	chr2:114034501	0,6	0,1	4,2E-16	4,4E-13	645	chr2:114036501	0,5	0,3	0,4	6,6E-03	7,9E-02	54561	x	
PBX1	chr1:164595001	0,4	0,2	5,0E-04	4,9E-03	69539	chr1:164599001	0,5	0,3	0,3	4,1E-02	1,8E-01	146833	x	
PBX2	chr6:32156501	0,7	0,5	2,1E-07	8,4E-06	17310	chr6:32156501	0,9	0,6	0,5	1,3E-03	4,2E-02	20249	x	x
PBX4	chr19:19729001	0,6	0,3	9,9E-11	1,1E-08	6304	chr19:19731501	0,4	0,3	0,3	1,2E-02	1,0E-01	75930	x	
PDX1	chr13:28492501	0,5	0,1	2,0E-12	3,7E-10	3813	chr13:28496001	0,4	0,3	0,1	3,1E-01	5,2E-01	384345	x	
PHOX2A	chr11:71955501	1,9	1,1	3,2E-11	4,0E-09	5501	chr11:71956501	0,6	0,4	0,3	4,0E-02	1,8E-01	144338	x	
PHOX2B	chr4:41749001	0,9	0,2	5,7E-17	1,4E-13	289	chr4:41748501	0,5	0,5	0,1	2,6E-01	4,7E-01	358471	x	
PITX1	chr5:134370501	0,5	0,2	3,1E-11	3,9E-09	5468	chr5:134371001	0,6	0,5	0,4	7,9E-02	2,5E-01	207395	x	
PITX2	chr4:111543501	0,4	0,1	1,0E-11	1,5E-09	4721	chr4:111560501	0,3	0,5	0,2	8,0E-03	8,6E-02	60996	x	
PITX3	chr10:104000501	0,5	0,1	4,6E-15	2,3E-12	1333	chr10:104000001	0,7	0,5	0,4	1,7E-01	3,7E-01	290707	x	
PKNOX1	chr21:44442001	0,9	0,7	4,2E-03	2,5E-02	114432	chr21:44393501	0,6	0,4	0,4	7,0E-03	8,1E-02	56165	x	
PKNOX2	chr11:125036001	0,7	0,5	4,2E-06	1,0E-04	27952	chr11:125036501	1,1	0,9	0,7	5,8E-01	7,5E-01	508682	x	
POU2F2	chr19:42638001	0,8	0,6	6,5E-03	3,5E-02	128428	chr19:42634501	0,7	0,5	0,5	3,5E-02	1,7E-01	136886	x	
POU2F3	chr11:120111001	0,6	0,3	2,8E-03	1,8E-02	103033	chr11:120109501	0,8	0,6	0,8	9,6E-03	9,2E-02	66714	x	
POU3F1	chr1:38510001	0,5	0,2	9,1E-12	1,3E-09	4656	chr1:38513501	0,5	0,3	0,1	1,5E-02	1,1E-01	84675	x	
POU3F3	chr2:105470001	0,4	0,1	2,0E-16	2,7E-13	505	chr2:105473001	0,7	0,3	0,1	1,1E-02	9,8E-02	72427	x	
POU4F1	chr13:79175501	0,6	0,1	2,1E-13	5,3E-11	2713	chr13:79177001	0,6	0,2	0,0	1,1E-04	1,7E-02	4287	x	x
POU4F2	chr4:147561501	0,8	0,1	7,1E-17	1,5E-13	325	chr4:147559501	0,4	0,3	0,1	2,9E-02	1,5E-01	124533	x	
POU4F3	chr5:145719501	0,8	0,1	2,2E-17	8,4E-14	173	chr5:145718501	0,5	0,4	0,1	2,9E-01	5,1E-01	378370	x	
POU5F1	chr6:31139001	0,4	0,3	1,0E-04	1,4E-03	50192	chr6:31132501	0,6	0,4	0,4	4,8E-03	6,9E-02	45363	x	
POU5F1P1	chr8:128428001	0,7	0,5	2,1E-03	1,5E-02	96746	chr8:128428501	0,5	0,3	0,3	2,2E-03	5,1E-02	28258	x	
POU6F1	chr12:51612501	0,5	0,3	1,2E-04	1,6E-03	51959	chr12:51609001	0,4	0,3	0,3	4,0E-02	1,8E-01	144421	x	
POU6F2	chr7:39074001	0,4	0,3	1,1E-02	5,0E-02	147377	chr7:39073501	0,6	0,9	0,7	2,2E-04	2,2E-02	6726		x
PROP1	chr5:177423001	0,5	0,3	5,2E-07	1,8E-05	19832	chr5:177424001	0,5	0,3	0,4	6,4E-04	3,2E-02	13035	x	x
PRRX1	chr1:170631001	0,4	0,1	4,0E-08	2,0E-06	13583	chr1:170699501	0,2	0,5	0,4	2,7E-04	2,3E-02	7319	x	x
PRRX2	chr9:132427001	0,7	0,3	5,5E-10	4,8E-08	7759	chr9:132427001	0,8	0,5	0,3	6,1E-03	7,6E-02	52362	x	
RAX	chr18:56939501	0,5	0,0	6,2E-18	6,8E-14	33	chr18:56939501	0,5	0,5	0,0	6,1E-01	7,7E-01	521547	x	
RAX2	chr19:3770001	0,9	0,7	3,5E-05	5,8E-04	40730	chr19:3770501	0,4	0,2	0,2	4,1E-03	6,5E-02	41299	x	
SATB1	chr3:18391001	0,7	0,6	1,2E-01	2,7E-01	309392	chr3:18485001	0,6	0,1	0,1	2,1E-04	2,1E-02	6423		x
SATB2	chr2:200329001	0,5	0,1	1,2E-13	3,4E-11	2520	chr2:200326501	1,4	0,9	0,7	1,8E-03	4,7E-02	24571	x	x
SHOX2	chr3:157820501	1,1	0,3	1,9E-15	1,2E-12	1072	chr3:157823001	0,5	0,1	0,1	6,6E-05	1,4E-02	3144	x	x
SIX2	chr2:45237501	0,6	0,2	1,6E-11	2,2E-09	5032	chr2:45235501	1,3	0,9	0,4	7,2E-02	2,4E-01	196456	x	
SIX3	chr2:45169501	0,8	0,1	1,1E-17	7,3E-14	100	chr2:45169501	1,2	0,6	0,1	6,6E-03	7,9E-02	54534	x	
SIX4	chr14:61188501	0,0	0,0	9,1E-01	NA	2162993	chr14:61192501	0,4	0,2	0,2	7,1E-04	3,3E-02	13737		x

Name	best DMR tumour	T	N	p_{MW} T/N	p_{BH} T/N	Rank	best DMR FUS	Fus-	Fus+	N	p_{MW} Fus -/+	p_{BH} Fus -/+	Rank	DMR T/N	DMR Fus -/+
SIX6	chr14:60973501	0,9	0,2	1,3E-16	2,0E-13	423	chr14:60975501	0,6	0,2	0,1	8,7E-03	8,9E-02	63578	x	
TGIF2	chr20:35200001	0,5	0,3	1,8E-05	3,4E-04	36054	chr20:35201001	0,4	0,2	0,2	2,3E-03	5,2E-02	28833	x	
TLX1	chr10:102892501	0,9	0,3	3,7E-14	1,3E-11	2025	chr10:102891001	0,4	0,1	0,0	5,5E-04	3,0E-02	11902	x	x
TLX2	chr2:74742501	0,6	0,1	6,1E-16	5,7E-13	738	chr2:74743501	1,3	1,1	0,4	1,0E-01	2,8E-01	230538	x	
TLX3	chr5:170737501	0,9	0,3	4,3E-17	1,2E-13	248	chr5:170735001	1,2	0,8	0,2	2,7E-01	4,8E-01	362654	x	
TPRXL	chr3:13976501	0,3	0,4	3,0E-03	2,0E-02	105653	chr3:14060501	0,4	0,4	0,5	2,1E-01	4,2E-01	326978	x	
TSHZ1	chr18:72996001	0,5	0,6	3,0E-02	1,0E-01	196614	chr18:72995001	0,8	0,5	0,5	1,2E-03	4,0E-02	18717		x
TSHZ2	chr20:51590001	0,7	0,2	4,9E-17	1,2E-13	266	chr20:51872001	1,7	1,1	1,4	1,9E-04	2,1E-02	5918	x	x
TSHZ3	chr19:31839001	0,6	0,2	3,0E-14	1,0E-11	1948	chr19:31839001	0,7	0,5	0,2	1,7E-01	3,7E-01	292005	x	
UNCX	chr7:1270501	1,5	0,3	1,6E-17	7,8E-14	140	chr7:1274001	0,4	0,6	0,3	1,9E-02	1,3E-01	98750	x	
VAX1	chr10:118899501	0,5	0,1	3,2E-15	1,8E-12	1232	chr10:118897501	0,5	0,3	0,2	6,1E-02	2,2E-01	179663	x	
VAX2	chr2:71126501	1,0	0,3	9,9E-13	2,0E-10	3424	chr2:71128501	0,7	0,4	0,2	2,0E-03	4,9E-02	26164	x	x
VSX1	chr20:25061501	1,1	0,4	3,8E-16	4,1E-13	623	chr20:25063001	0,6	0,4	0,3	6,5E-02	2,3E-01	187017	x	
VSX2	chr14:74706001	0,4	0,1	5,4E-14	1,7E-11	2171	chr14:74705501	0,5	0,4	0,3	3,8E-01	5,8E-01	420854	x	
ZEB2	chr2:145273001	0,5	0,2	3,1E-07	1,2E-05	18351	chr2:145186501	0,3	0,4	0,4	3,1E-03	5,8E-02	34681	x	
ZFHX2	chr14:24019501	0,9	0,5	6,7E-08	3,1E-06	14625	chr14:23994501	0,5	0,3	0,3	3,8E-04	2,6E-02	9277	x	x
ZFHX4	chr8:77595001	0,5	0,2	4,2E-08	2,1E-06	13663	chr8:77595501	0,4	0,5	0,3	4,3E-01	6,3E-01	447459	x	
ZHX1	chr8:124284501	0,3	0,2	2,5E-03	NA	5485308	chr8:124288501	0,5	0,3	0,3	4,5E-04	2,8E-02	10577		x
ZHX2	chr8:123791501	0,5	0,4	3,0E-03	1,9E-02	104863	chr8:123792001	0,5	0,2	0,2	1,4E-03	4,3E-02	21033	x	x
ZHX3	chr20:39831501	1,1	0,9	8,4E-02	2,1E-01	272007	chr20:39831001	1,1	0,7	0,8	2,1E-03	5,0E-02	27053		x

Supplementary Table 7: Differentially methylated G-protein coupled receptor (GPCR) genes (http://www.iuphar-db.org). For each gene the values of the most significant bin (according to p-values) of all TSS associated bins are used shown. Column headings denote the gene name, the receptor family and class, the Benjamini-Hochberg corrected Mann-Whitney p-value of differential methylation between tumour and normal samples (p_{BH}), the ype of methylation in the most differentially methylated TSS associated bin, and the rank of the DMR in the list of ordered hypomethylated bins. NA values indicate that the respective gene does not contain any promoter associated hypomethylated bin.

Gene name	Family	CLASS	p_{BH}	Methylation in tumour	Rank (hypo)
HTR1A	5-Hydroxytryptamine receptors	A	1,05E-14	HYPER	1216424
HTR1B	5-Hydroxytryptamine receptors	A	5,17E-13	HYPER	589695
HTR1D	5-Hydroxytryptamine receptors	A	1,76E-05	HYPER	341866
HTR2A	5-Hydroxytryptamine receptors	A	2,70E-05	HYPO	21426
HTR2B	5-Hydroxytryptamine receptors	A	1,43E-02	HYPER	1746588
HTR2C	5-Hydroxytryptamine receptors	A	2,77E-13	HYPER	1679164
HTR4	5-Hydroxytryptamine receptors	A	5,16E-03	HYPER	456831
HTR5A	5-Hydroxytryptamine receptors	A	7,94E-06	HYPER	15747
HTR6	5-Hydroxytryptamine receptors	A	6,94E-08	HYPER	NA
HTR7	5-Hydroxytryptamine receptors	A	6,47E-08	HYPER	1128956
CHRM1	Acetylcholine receptors (muscarinic)	A	6,68E-16	HYPER	791138
CHRM2	Acetylcholine receptors (muscarinic)	A	2,63E-11	HYPER	1677228
CHRM3	Acetylcholine receptors (muscarinic)	A	4,94E-05	HYPO	27307
CHRM4	Acetylcholine receptors (muscarinic)	A	9,10E-04	HYPER	2203903
CHRM5	Acetylcholine receptors (muscarinic)	A	1,21E-02	HYPO	280733
ADORA1	Adenosine receptors	A	2,68E-05	HYPER	996452
ADORA2A	Adenosine receptors	A	1,06E-09	HYPER	1079344
ADORA2B	Adenosine receptors	A	3,58E-08	HYPER	2285306
ADORA3	Adenosine receptors	A	1,14E-03	HYPER	225472
ADRA1A	Adrenoceptors	A	1,09E-10	HYPER	1814988
ADRA1B	Adrenoceptors	A	2,92E-04	HYPER	2084077
ADRA1D	Adrenoceptors	A	5,33E-07	HYPER	937876
ADRA2A	Adrenoceptors	A	1,13E-15	HYPER	952221
ADRA2B	Adrenoceptors	A	2,81E-06	HYPER	NA
ADRA2C	Adrenoceptors	A	1,16E-07	HYPER	1968503
ADRB1	Adrenoceptors	A	4,79E-04	HYPO	66996
ADRB2	Adrenoceptors	A	3,10E-09	HYPO	467
ADRB3	Adrenoceptors	A	5,41E-06	HYPER	1303880
C5AR1	Anaphylatoxin receptors	A	2,15E-02	HYPER	NA
GPR77	Anaphylatoxin receptors	A	1,53E-02	HYPER	2067027
AGTR1	Angiotensin receptors	A	7,65E-05	HYPER	1167714
AGTR2	Angiotensin receptors	A	5,76E-03	HYPO	196473
APLNR	Apelin receptor	A	2,39E-04	HYPER	687582
GPBAR1	Bile acid receptor	A	8,05E-06	HYPER	NA
NMBR	Bombesin receptors	A	3,24E-11	HYPER	230429
GRPR	Bombesin receptors	A	3,05E-03	HYPO	147503
BRS3	Bombesin receptors	A	8,46E-04	HYPO	85052
BDKRB1	Bradykinin receptors	A	8,02E-04	HYPER	213812
BDKRB2	Bradykinin receptors	A	5,95E-03	HYPER	2421448
CNR1	Cannabinoid receptors	A	6,41E-13	HYPER	22476
CNR2	Cannabinoid receptors	A	3,58E-02	HYPO	479164
CCR1	Chemokine receptors	A	1,81E-02	HYPER	896878
CCR2	Chemokine receptors	A	2,78E-06	HYPO	8727
CCR3	Chemokine receptors	A	1,74E-02	HYPER	744470
CCR4	Chemokine receptors	A	2,90E-02	HYPER	1500164
CCR5	Chemokine receptors	A	2,37E-03	HYPO	131507
CCR6	Chemokine receptors	A	1,47E-04	HYPER	188288
CCR7	Chemokine receptors	A	6,99E-03	HYPER	259543
CCR9	Chemokine receptors	A	1,69E-03	HYPER	1181453
CCR10	Chemokine receptors	A	3,11E-04	HYPER	357168
CXCR1	Chemokine receptors	A	4,94E-02	HYPO	561294
CXCR2	Chemokine receptors	A	6,70E-04	HYPER	680829
CXCR3	Chemokine receptors	A	2,67E-03	HYPER	2199235
CXCR4	Chemokine receptors	A	2,70E-07	HYPER	1529303

Gene name	Family	CLASS	p_{BH}	Methylation in tumour	Rank (hypo)
CXCR6	Chemokine receptors	A	1,08E-02	HYPER	1782762
CXCR7	Chemokine receptors	A	7,07E-09	HYPER	243183
CX3CR1	Chemokine receptors	A	2,02E-02	HYPER	857364
CCKAR	Cholecystokinin receptors	A	8,52E-04	HYPER	2156769
CCKBR	Cholecystokinin receptors	A	1,15E-10	HYPER	1044911
DRD1	Dopamine receptors	A	6,87E-06	HYPER	397234
DRD2	Dopamine receptors	A	3,08E-04	HYPER	181816
DRD3	Dopamine receptors	A	6,99E-03	HYPO	215357
DRD4	Dopamine receptors	A	2,38E-03	HYPER	NA
DRD5	Dopamine receptors	A	6,07E-10	HYPER	129852
EDNRA	Endothelin receptors	A	8,88E-07	HYPER	711028
EDNRB	Endothelin receptors	A	2,36E-07	HYPER	136634
GPER	Estrogen (G protein coupled) receptor	A	1,05E-05	HYPER	2468366
FPR1	Formylpeptide receptors	A	5,38E-03	HYPO	190012
FPR2	Formylpeptide receptors	A	1,45E-02	HYPO	306684
FPR3	Formylpeptide receptors	A	9,75E-03	HYPER	1642154
FFAR1	Free fatty acid receptors	A	1,26E-03	HYPER	1366557
FFAR2	Free fatty acid receptors	A	4,34E-02	HYPER	1109570
FFAR3	Free fatty acid receptors	A	2,90E-02	HYPER	NA
GALR1	Galanin receptors	A	4,96E-09	HYPER	NA
GALR2	Galanin receptors	A	2,78E-04	HYPER	1473740
GALR3	Galanin receptors	A	3,72E-05	HYPER	2667344
GHSR	Ghrelin receptor	A	6,68E-16	HYPER	2585198
FSHR	Glycoprotein hormone receptors	A	1,18E-04	HYPO	37947
LHCGR	Glycoprotein hormone receptors	A	1,21E-05	HYPER	19553
TSHR	Glycoprotein hormone receptors	A	1,01E-02	HYPO	257966
GNRHR	Gonadotrophin-releasing hormone receptors	A	2,38E-02	HYPER	982942
HRH1	Histamine receptors	A	1,12E-03	HYPER	205782
HRH2	Histamine receptors	A	4,93E-11	HYPER	32156
HRH3	Histamine receptors	A	8,76E-07	HYPER	2790812
LTB4R	Leukotriene receptors	A	5,09E-07	HYPER	1166750
LTB4R2	Leukotriene receptors	A	1,70E-02	HYPER	2033502
CYSLTR1	Leukotriene receptors	A	3,01E-02	HYPER	570241
CYSLTR2	Leukotriene receptors	A	8,53E-07	HYPO	5247
OXER1	Leukotriene receptors	A	1,44E-02	HYPER	1267171
FPR2	Leukotriene receptors	A	1,45E-02	HYPO	306684
LPAR1	Lysophospholipid (LPA) receptors	A	2,81E-15	HYPER	258670
LPAR2	Lysophospholipid (LPA) receptors	A	8,77E-03	HYPER	NA
LPAR4	Lysophospholipid (LPA) receptors	A	1,81E-02	HYPO	342313
LPAR5	Lysophospholipid (LPA) receptors	A	8,60E-09	HYPER	62812
S1PR1	Lysophospholipid (S1P) receptors	A	1,25E-06	HYPER	2871348
S1PR2	Lysophospholipid (S1P) receptors	A	2,42E-08	HYPER	2869029
S1PR3	Lysophospholipid (S1P) receptors	A	1,39E-12	HYPER	2668102
S1PR4	Lysophospholipid (S1P) receptors	A	5,63E-04	HYPER	1347098
S1PR5	Lysophospholipid (S1P) receptors	A	5,95E-07	HYPER	2014253
MCHR1	Melanin-concentrating hormone receptors	A	1,73E-06	HYPER	NA
MCHR2	Melanin-concentrating hormone receptors	A	5,50E-10	HYPER	769923
MC2R	Melanocortin receptors	A	4,18E-05	HYPER	42670
MC3R	Melanocortin receptors	A	4,24E-04	HYPO	63729
MC4R	Melanocortin receptors	A	1,21E-05	HYPO	15710
MC5R	Melanocortin receptors	A	3,73E-05	HYPO	24261
MTNR1A	Melatonin receptors	A	1,41E-02	HYPO	301201
MTNR1B	Melatonin receptors	A	1,43E-09	HYPER	1344913
MLNR	Motilin receptor	A	2,17E-07	HYPER	2160613
NMUR1	Neuromedin U receptors	A	1,48E-06	HYPER	NA
NMUR2	Neuromedin U receptors	A	4,57E-03	HYPO	175742
NPFFR1	Neuropeptide FF/neuropeptide AF receptors	A	1,82E-02	HYPO	344158
NPFFR2	Neuropeptide FF/neuropeptide AF receptors	A	4,68E-07	HYPER	311169
NPSR1	Neuropeptide S receptor	A	9,64E-06	HYPO	14333
NPBWR1	Neuropeptide W/neuropeptide B receptors	A	7,11E-08	HYPER	858852
NPBWR2	Neuropeptide W/neuropeptide B receptors	A	8,41E-06	HYPER	2220049

Gene name	Family	CLASS	p_{BH}	Methylation in tumour	Rank (hypo)
NPY1R	Neuropeptide Y receptors	A	1,25E-06	HYPER	393154
NPY2R	Neuropeptide Y receptors	A	2,70E-03	HYPER	592435
PPYR1	Neuropeptide Y receptors	A	5,24E-03	HYPER	NA
NPY5R	Neuropeptide Y receptors	A	1,25E-06	HYPER	70985
NPY6R	Neuropeptide Y receptors	A	4,77E-02	HYPER	836023
NTSR1	Neurotensin receptors	A	2,29E-04	HYPO	48885
NTSR2	Neurotensin receptors	A	7,95E-06	HYPER	2120090
OPRD1	Opioid receptors	A	4,96E-02	HYPER	1123880
OPRK1	Opioid receptors	A	1,72E-02	HYPER	NA
OPRM1	Opioid receptors	A	1,20E-08	HYPER	12351
OPRL1	Opioid receptors	A	4,50E-05	HYPER	64835
HCRTR1	Orexin receptors	A	4,77E-06	HYPER	771210
HCRTR2	Orexin receptors	A	4,14E-05	HYPER	51277
P2RY1	P2Y receptors	A	4,09E-04	HYPER	485303
P2RY2	P2Y receptors	A	8,89E-05	HYPER	2022046
P2RY4	P2Y receptors	A	8,54E-03	HYPER	1468875
P2RY6	P2Y receptors	A	5,47E-03	HYPER	1934246
P2RY11	P2Y receptors	A	7,90E-04	HYPER	NA
P2RY12	P2Y receptors	A	2,38E-03	HYPER	1273716
P2RY13	P2Y receptors	A	8,95E-03	HYPER	2060584
P2RY14	P2Y receptors	A	1,46E-02	HYPER	1238285
QRFPR	Peptide P518 receptor	A	1,05E-12	HYPER	64964
PTAFR	Platelet-activating factor receptor	A	7,89E-09	HYPER	2428190
PROKR1	Prokineticin receptors	A	1,88E-03	HYPER	452922
PROKR2	Prokineticin receptors	A	1,05E-12	HYPER	41306
PRLHR	Prolactin-releasing peptide receptor	A	3,84E-14	HYPER	278751
PTGDR	Prostanoid receptors	A	1,80E-08	HYPER	888703
PTGER1	Prostanoid receptors	A	1,35E-08	HYPER	NA
PTGER2	Prostanoid receptors	A	8,07E-13	HYPER	NA
PTGER3	Prostanoid receptors	A	2,59E-11	HYPER	1369137
PTGER4	Prostanoid receptors	A	8,06E-14	HYPER	869271
PTGFR	Prostanoid receptors	A	3,36E-04	HYPER	219213
PTGIR	Prostanoid receptors	A	7,78E-05	HYPER	1818459
TBXA2R	Prostanoid receptors	A	4,60E-06	HYPER	839026
F2R	Protease-activated receptors	A	3,27E-07	HYPER	1314949
F2RL2	Protease-activated receptors	A	1,24E-04	HYPER	1113492
F2RL3	Protease-activated receptors	A	2,84E-08	HYPER	2093374
RXFP1	Relaxin family peptide receptors	A	1,47E-08	HYPO	948
RXFP2	Relaxin family peptide receptors	A	7,10E-03	HYPO	216953
RXFP3	Relaxin family peptide receptors	A	4,59E-11	HYPER	399547
RXFP4	Relaxin family peptide receptors	A	1,29E-10	HYPER	NA
SSTR1	Somatostatin receptors	A	1,52E-15	HYPER	1763572
SSTR2	Somatostatin receptors	A	2,44E-02	HYPER	543869
SSTR3	Somatostatin receptors	A	2,45E-06	HYPER	1450475
SSTR4	Somatostatin receptors	A	7,56E-05	HYPER	1135853
SSTR5	Somatostatin receptors	A	2,04E-03	HYPER	649911
TACR1	Tachykinin receptors	A	2,07E-05	HYPER	705985
TACR2	Tachykinin receptors	A	1,02E-06	HYPER	1987230
TACR3	Tachykinin receptors	A	1,25E-11	HYPER	502435
TRHR	Thyrotropin-releasing hormone receptors	A	4,93E-04	HYPER	1941183
UTS2R	Urotensin receptor	A	1,16E-05	HYPER	1844537
AVPR1A	Vasopressin and oxytocin receptors	A	3,56E-07	HYPER	1356558
AVPR1B	Vasopressin and oxytocin receptors	A	7,48E-04	HYPER	1890953
AVPR2	Vasopressin and oxytocin receptors	A	3,09E-03	HYPER	2395285
OXTR	Vasopressin and oxytocin receptors	A	1,33E-06	HYPER	1163360
CCRL2	Class A Orphans	A	3,67E-04	HYPER	1920901
CMKLR1	Class A Orphans	A	1,17E-06	HYPO	5990
GPR1	Class A Orphans	A	8,91E-03	HYPO	243795
GPR3	Class A Orphans	A	1,80E-04	HYPER	1754107
GPR4	Class A Orphans	A	5,66E-03	HYPER	1768277
GPR6	Class A Orphans	A	1,63E-07	HYPER	18128

Gene name	Family	CLASS	p_{BH}	Methylation in tumour	Rank (hypo)
GPR12	Class A Orphans	A	1,84E-03	HYPER	1195999
GPR15	Class A Orphans	A	3,20E-02	HYPO	452575
GPR17	Class A Orphans	A	1,79E-03	HYPER	715462
GPR18	Class A Orphans	A	8,84E-04	HYPER	674624
GPR19	Class A Orphans	A	3,31E-02	HYPO	460865
GPR20	Class A Orphans	A	2,43E-06	HYPER	2563203
GPR21	Class A Orphans	A	1,33E-06	HYPER	2268525
GPR25	Class A Orphans	A	2,84E-13	HYPER	2549273
GPR26	Class A Orphans	A	7,73E-09	HYPO	720
GPR27	Class A Orphans	A	9,52E-04	HYPER	1450120
GPR34	Class A Orphans	A	8,07E-03	HYPER	1929400
GPR35	Class A Orphans	A	2,70E-07	HYPER	1117056
GPR37	Class A Orphans	A	5,71E-10	HYPER	1437840
GPR37L1	Class A Orphans	A	5,38E-03	HYPER	1709139
GPR39	Class A Orphans	A	1,59E-02	HYPER	456290
GPR45	Class A Orphans	A	1,53E-02	HYPER	1556294
GPR50	Class A Orphans	A	1,43E-09	HYPER	1849785
GPR52	Class A Orphans	A	1,19E-02	HYPER	1593078
GPR55	Class A Orphans	A	8,67E-04	HYPER	1334939
GPR62	Class A Orphans	A	4,86E-09	HYPER	2040015
GPR68	Class A Orphans	A	7,37E-04	HYPER	NA
GPR75	Class A Orphans	A	2,16E-09	HYPER	807596
GPR78	Class A Orphans	A	3,85E-05	HYPER	235383
GPR79	Class A Orphans	A	3,81E-03	HYPER	2104131
GPR83	Class A Orphans	A	1,37E-09	HYPER	1564314
GPR84	Class A Orphans	A	3,20E-02	HYPO	452918
GPR85	Class A Orphans	A	7,54E-03	HYPER	738603
GPR87	Class A Orphans	A	1,77E-09	HYPER	1111304
GPR88	Class A Orphans	A	2,26E-11	HYPER	2752493
GPR101	Class A Orphans	A	7,43E-09	HYPER	2124738
GPR119	Class A Orphans	A	2,76E-02	HYPER	2215264
GPR132	Class A Orphans	A	5,96E-04	HYPER	1104793
GPR135	Class A Orphans	A	2,80E-05	HYPER	1940516
GPR139	Class A Orphans	A	3,17E-07	HYPER	2158486
GPR141	Class A Orphans	A	5,58E-03	HYPO	193361
GPR146	Class A Orphans	A	4,93E-13	HYPER	322997
GPR148	Class A Orphans	A	2,03E-07	HYPER	889340
GPR149	Class A Orphans	A	1,13E-15	HYPER	1835299
GPR150	Class A Orphans	A	5,48E-10	HYPER	2838184
GPR151	Class A Orphans	A	8,04E-03	HYPER	2267420
GPR152	Class A Orphans	A	4,14E-05	HYPER	2101582
GPR153	Class A Orphans	A	3,50E-03	HYPER	1739045
GPR160	Class A Orphans	A	4,61E-07	HYPO	4100
GPR161	Class A Orphans	A	5,89E-07	HYPER	308922
GPR162	Class A Orphans	A	1,44E-03	HYPO	106837
GPR171	Class A Orphans	A	4,46E-02	HYPO	533436
GPR173	Class A Orphans	A	1,43E-05	HYPER	1395399
GPR174	Class A Orphans	A	3,58E-02	HYPO	479221
GPR176	Class A Orphans	A	2,60E-09	HYPER	398257
GPR182	Class A Orphans	A	1,50E-03	HYPER	NA
GPR183	Class A Orphans	A	8,04E-03	HYPER	1913346
LGR4	Class A Orphans	A	2,80E-03	HYPER	585883
LGR5	Class A Orphans	A	4,25E-08	HYPER	1969761
LGR6	Class A Orphans	A	1,57E-06	HYPER	687936
LPAR6	Class A Orphans	A	1,23E-02	HYPO	282968
MAS1	Class A Orphans	A	4,24E-04	HYPO	63845
MAS1L	Class A Orphans	A	3,19E-02	HYPO	451184
MRGPRD	Class A Orphans	A	4,78E-05	HYPER	1770711
MRGPRE	Class A Orphans	A	1,57E-05	HYPER	676029
MRGPRF	Class A Orphans	A	3,84E-07	HYPER	NA
MRGPRX1	Class A Orphans	A	1,32E-03	HYPO	102972

Gene name	Family	CLASS	p_{BH}	Methylation in tumour	Rank (hypo)
MRGPRX2	Class A Orphans	A	8,65E-04	HYPO	86065
MRGPRX4	Class A Orphans	A	7,69E-03	HYPO	225555
OPN3	Class A Orphans	A	1,30E-02	HYPO	290229
OPN5	Class A Orphans	A	8,24E-03	HYPO	235019
OXGR1	Class A Orphans	A	4,04E-10	HYPER	294040
P2RY10	Class A Orphans	A	7,69E-03	HYPO	225679
SUCNR1	Class A Orphans	A	1,05E-03	HYPER	2211819
TAAR2	Class A Orphans	A	1,96E-03	HYPO	121761
TAAR3	Class A Orphans	A	5,84E-06	HYPO	11607
TAAR4P	Class A Orphans	A	2,33E-04	HYPO	49365
TAAR5	Class A Orphans	A	7,40E-04	HYPO	80066
TAAR6	Class A Orphans	A	8,34E-04	HYPO	84484
TAAR8	Class A Orphans	A	3,98E-04	HYPO	61834
TAAR9	Class A Orphans	A	2,37E-04	HYPO	49850
CCBP2	Non-signalling 7TM chemokine-binding proteins	A	4,87E-03	HYPER	697634
CCRL1	Non-signalling 7TM chemokine-binding proteins	A	5,26E-03	HYPO	187503
DARC	Non-signalling 7TM chemokine-binding proteins	A	1,48E-04	HYPO	41517
CALCR	Calcitonin receptors	B	1,84E-07	HYPER	72152
CRHR1	Corticotropin-releasing factor receptors	B	1,34E-06	HYPER	503511
CRHR2	Corticotropin-releasing factor receptors	B	5,31E-11	HYPER	327280
GHRHR	Glucagon receptor family	B	6,34E-08	HYPER	328354
GLP1R	Glucagon receptor family	B	5,55E-07	HYPER	2116425
GLP2R	Glucagon receptor family	B	3,76E-03	HYPO	161368
GCGR	Glucagon receptor family	B	2,93E-04	HYPER	NA
SCTR	Glucagon receptor family	B	5,32E-06	HYPER	278758
PTH1R	Parathyroid hormone receptors	B	1,48E-06	HYPER	1299470
PTH2R	Parathyroid hormone receptors	B	3,51E-07	HYPER	398699
ADCYAP1R1	VIP and PACAP receptors	B	1,67E-11	HYPER	89537
VIPR1	VIP and PACAP receptors	B	5,48E-10	HYPER	68606
VIPR2	VIP and PACAP receptors	B	1,61E-15	HYPER	2667174
BAI1	Class B Orphans	B	1,64E-05	HYPER	NA
BAI2	Class B Orphans	B	1,65E-06	HYPER	1279189
BAI3	Class B Orphans	B	8,04E-03	HYPER	321679
CD97	Class B Orphans	B	3,41E-04	HYPER	1829585
CELSR1	Class B Orphans	B	8,07E-13	HYPER	2087325
CELSR2	Class B Orphans	B	2,17E-07	HYPER	236802
CELSR3	Class B Orphans	B	7,07E-09	HYPER	1949481
ELTD1	Class B Orphans	B	8,07E-13	HYPER	741573
EMR1	Class B Orphans	B	5,26E-03	HYPO	187730
EMR2	Class B Orphans	B	2,85E-02	HYPO	425184
EMR3	Class B Orphans	B	3,67E-04	HYPO	59859
EMR4P	Class B Orphans	B	6,20E-03	HYPO	203214
GPR56	Class B Orphans	B	1,37E-09	HYPER	1545331
GPR64	Class B Orphans	B	1,20E-02	HYPER	454954
GPR97	Class B Orphans	B	2,78E-03	HYPER	2597245
GPR98	Class B Orphans	B	3,46E-04	HYPO	58377
GPR110	Class B Orphans	B	9,92E-04	HYPO	91611
GPR111	Class B Orphans	B	2,90E-02	HYPER	656463
GPR112	Class B Orphans	B	6,79E-03	HYPO	211713
GPR113	Class B Orphans	B	4,05E-03	HYPO	167090
GPR114	Class B Orphans	B	2,48E-03	HYPER	939436
GPR115	Class B Orphans	B	8,49E-06	HYPER	519962
GPR116	Class B Orphans	B	6,91E-05	HYPER	499102
GPR123	Class B Orphans	B	4,86E-06	HYPO	10805
GPR124	Class B Orphans	B	5,02E-09	HYPER	917212
GPR125	Class B Orphans	B	2,54E-05	HYPO	20859
GPR126	Class B Orphans	B	1,07E-03	HYPO	94538
GPR128	Class B Orphans	B	7,19E-03	HYPO	218457
GPR133	Class B Orphans	B	2,37E-04	HYPO	49830
GPR144	Class B Orphans	B	3,57E-11	HYPER	NA
GPR157	Class B Orphans	B	2,88E-03	HYPER	236026

Gene name	Family	CLASS	p_{BH}	Methylation in tumour	Rank (hypo)
LPHN1	Class B Orphans	B	6,92E-03	HYPER	1127217
LPHN2	Class B Orphans	B	1,20E-02	HYPO	279193
LPHN3	Class B Orphans	B	3,64E-07	HYPER	34194
CASR	Calcium-sensing receptors	C	4,59E-10	HYPER	209366
GPRC6A	Calcium-sensing receptors	C	1,75E-02	HYPO	337139
GABBR1	GABA_B receptors	C	1,13E-12	HYPER	1060835
GABBR2	GABA_B receptors	C	8,92E-06	HYPER	147405
GRM1	Metabotropic glutamate receptors	C	1,32E-11	HYPER	426638
GRM2	Metabotropic glutamate receptors	C	7,78E-05	HYPER	2908804
GRM3	Metabotropic glutamate receptors	C	6,77E-08	HYPER	62770
GRM4	Metabotropic glutamate receptors	C	1,84E-05	HYPER	690902
GRM5	Metabotropic glutamate receptors	C	4,24E-04	HYPO	63621
GRM6	Metabotropic glutamate receptors	C	1,42E-09	HYPER	139947
GRM7	Metabotropic glutamate receptors	C	1,13E-12	HYPER	47108
GRM8	Metabotropic glutamate receptors	C	7,73E-09	HYPER	4699
GPR156	Class C Orphans	C	3,62E-11	HYPER	641429
GPR158	Class C Orphans	C	1,14E-13	HYPER	96763
GPR179	Class C Orphans	C	8,04E-03	HYPER	2708171
GPRC5B	Class C Orphans	C	8,06E-14	HYPER	600449
GPRC5C	Class C Orphans	C	2,04E-06	HYPER	1013292
GPRC5D	Class C Orphans	C	4,06E-02	HYPO	509437
TAS1R1	Taste 1 receptors	C	6,22E-03	HYPER	919691
TAS1R2	Taste 1 receptors	C	3,26E-03	HYPER	564739
FZD1	Frizzled receptors	FRIZZLED	1,14E-13	HYPER	1283476
FZD2	Frizzled receptors	FRIZZLED	7,43E-09	HYPER	1123069
FZD3	Frizzled receptors	FRIZZLED	8,04E-03	HYPO	231163
FZD5	Frizzled receptors	FRIZZLED	2,11E-12	HYPER	2931537
FZD6	Frizzled receptors	FRIZZLED	9,92E-04	HYPER	2302665
FZD7	Frizzled receptors	FRIZZLED	3,96E-13	HYPER	1243951
FZD8	Frizzled receptors	FRIZZLED	2,31E-09	HYPER	1234237
FZD9	Frizzled receptors	FRIZZLED	6,20E-04	HYPER	2919357
FZD10	Frizzled receptors	FRIZZLED	3,10E-13	HYPER	NA
SMO	Frizzled receptors	FRIZZLED	1,70E-04	HYPER	2224992
GPR107	Other 7TM proteins	OTHER7TM	3,78E-03	HYPER	756570
GPR137	Other 7TM proteins	OTHER7TM	8,04E-03	HYPER	1277882
GPR143	Other 7TM proteins	OTHER7TM	1,43E-09	HYPER	945448

Supplementary Table 8: List of miRNAs showing differential expression in tumour subgroups. Column headings are as follows. (shortname) miRNA name, (Identifier) Identifier in miRNA expression list, (p_{BH} EXP FUS- NORM) BH corrected p_{MW}-value of differential expression between FUS- and normal samples, (p_{BH} EXP FUS+ NORM) BH corrected p_{MW}-value of differential expression between FUS+ and normal samples, (p_{MW} EXP FUS- FUS+) Mann-Whitney p-value of differential expression between FUS- and FUS+ samples, (Ct normal) Ct value of mean expression in normal samples, (Ct FUS-) Ct value of mean expression in FUS- samples, (Ct FUS+) Ct value of mean expression in FUS+ samples, (Excl. FUS- NORM) differential expression between FUS- and normal samples and (Excl. FUS+ NORM) differential expression between FUS+ and normal samples. FALSE indicates that the sample is not or not exclusively differentially expressed between the tumour subgroup and the normal samples, DOWN indicates a downregulation in the tumour subgroup, UP an upregulation. (FUS- vs FUS+) Differential expression between FUS- and FUS+. DOWN represents a downregulation in FUS- compared to FUS+, UP an upregulation in FUS- compared to FUS+. (Pattern) Expression pattern according to **Supplementary Figure 1**.

shortname	Identifier	p_{BH} EXP FUS- NORM	p_{BH} EXP FUS+ NORM	p_{MW} EXP FUS- FUS+	Ct normal	Ct FUS-	Ct FUS+	Excl. FUS- NORM	Excl. FUS+ NORM	FUS- vs FUS+	Pattern
let7e*	4395518	1,93E-05	6,59E-02	1,09E-02	2,4	3,2	2,7	DOWN	FALSE	DOWN	1
mir766	4395177.1	4,20E-05	6,35E-02	3,05E-02	-1,6	-0,6	-1,3	DOWN	FALSE	DOWN	1
mir100*	4395253	4,39E-05	2,02E-01	5,04E-03	1,8	2,8	1,9	DOWN	FALSE	DOWN	1
mir136*	4395211	4,39E-05	2,56E-01	7,31E-03	-1,3	-0,5	-1,0	DOWN	FALSE	DOWN	1
mir582-5p	4395175	2,40E-04	8,32E-01	8,01E-04	7,0	11,4	7,1	DOWN	FALSE	DOWN	1
mir512-3p	4381034	2,43E-04	7,79E-01	2,32E-02	2,3	3,8	2,5	DOWN	FALSE	DOWN	1
mir181a2*	4395428	3,30E-04	3,84E-01	2,31E-02	-1,1	-0,5	-1,0	DOWN	FALSE	DOWN	1
mir519d	4395514	4,65E-04	8,24E-01	3,42E-02	5,0	7,1	5,6	DOWN	FALSE	DOWN	1
mir26a	4395166	8,19E-04	1,60E-01	2,60E-02	-10,3	-9,8	-10,1	DOWN	FALSE	DOWN	1
mir485-3p	4378095	1,01E-03	2,76E-01	3,68E-02	0,1	0,7	0,2	DOWN	FALSE	DOWN	1
mir409-5p	4395442	2,13E-03	2,64E-01	4,76E-02	4,3	6,6	4,9	DOWN	FALSE	DOWN	1
mir654-3p	4395350	3,36E-03	3,45E-01	4,43E-02	2,9	3,6	3,1	DOWN	FALSE	DOWN	1
mir504	4395195	3,50E-03	5,57E-01	5,05E-03	-0,8	0,3	-0,8	DOWN	FALSE	DOWN	1
mir655	4381015	3,78E-03	6,35E-01	3,82E-02	1,0	1,6	1,1	DOWN	FALSE	DOWN	1
mir758	4395180	4,49E-03	5,92E-01	1,89E-02	2,7	3,5	2,9	DOWN	FALSE	DOWN	1
mir370	4395386	4,52E-03	2,64E-01	3,82E-02	-0,1	1,0	0,1	DOWN	FALSE	DOWN	1
mir130a*	4395242	1,22E-04	5,05E-02	1,61E-01	6,7	10,7	8,6	DOWN	FALSE	FALSE	other
mir34b*	4373037	1,59E-04	2,70E-01	5,61E-02	4,2	5,7	4,5	DOWN	FALSE	FALSE	other
mir377*	4395239	1,61E-04	6,59E-02	1,16E-01	6,3	10,1	8,1	DOWN	FALSE	FALSE	other
mir766	4395177.2	1,65E-04	1,13E-01	9,67E-02	-1,7	-0,8	-1,3	DOWN	FALSE	FALSE	other
mir517a	4395513	1,85E-04	2,66E-01	7,71E-02	2,7	3,8	3,1	DOWN	FALSE	FALSE	other
mir299-5p	4373188	2,63E-04	6,16E-02	1,25E-01	8,0	11,7	10,0	DOWN	FALSE	FALSE	other
mir379	4373349	5,90E-04	6,52E-02	9,98E-02	-1,2	-0,6	-1,0	DOWN	FALSE	FALSE	other
mir134	4373299	1,17E-03	2,39E-01	7,97E-02	0,4	1,2	0,7	DOWN	FALSE	FALSE	other
let7i*	4395283	1,30E-03	5,80E-02	2,06E-01	3,1	3,7	3,4	DOWN	FALSE	FALSE	other
mir181c	4373115	1,48E-03	7,87E-02	1,35E-01	1,3	2,1	1,8	DOWN	FALSE	FALSE	other
mir517c	4373264	3,66E-03	5,92E-01	7,46E-02	3,1	4,2	3,4	DOWN	FALSE	FALSE	other
mir127-3p	4373147	4,22E-03	2,16E-01	2,11E-01	-5,4	-4,9	-5,1	DOWN	FALSE	FALSE	other
mir656	4380920.2	4,97E-03	7,24E-01	7,21E-02	0,6	1,4	0,8	DOWN	FALSE	FALSE	other
mir154	4373270	6,39E-03	5,95E-02	7,84E-01	7,8	11,3	10,5	DOWN	FALSE	FALSE	other
mir382	4373019	7,13E-03	2,83E-01	1,10E-01	-1,2	-0,2	-0,9	DOWN	FALSE	FALSE	other
mir889	4395313	7,71E-03	2,56E-01	1,35E-01	2,1	2,7	2,3	DOWN	FALSE	FALSE	other
mir518e	4395506	9,38E-03	8,29E-01	8,23E-02	8,0	11,4	9,0	DOWN	FALSE	FALSE	other
mir199a-3p	4395415	9,38E-03	1,68E-01	1,20E-01	-8,6	-8,0	-8,4	DOWN	FALSE	FALSE	other
mir181a*	4373086	9,51E-03	3,14E-01	2,44E-01	1,3	1,8	1,6	DOWN	FALSE	FALSE	other
mir34c-5p	4373036	1,01E-02	8,67E-02	4,02E-01	0,8	1,8	1,4	DOWN	FALSE	FALSE	other
mir432	4373280.2	1,10E-02	4,74E-01	9,36E-02	-0,3	0,4	-0,3	DOWN	FALSE	FALSE	other
mir409-3p	4395443.2	1,30E-02	4,42E-01	5,29E-02	-4,1	-3,5	-4,0	DOWN	FALSE	FALSE	other
mir378*	4373024	1,56E-02	7,15E-02	4,19E-01	1,0	1,6	1,3	DOWN	FALSE	FALSE	other
mir432	4373280.1	1,58E-02	7,11E-01	7,96E-02	-0,3	0,3	-0,2	DOWN	FALSE	FALSE	other
mir9*	4395342	1,59E-02	9,26E-01	1,17E-01	-1,1	-0,8	-1,2	DOWN	FALSE	FALSE	other
mir193a-5p	4395392	1,59E-02	6,64E-01	2,09E-01	-4,9	-4,6	-4,8	DOWN	FALSE	FALSE	other
mir656	4380920.1	2,00E-02	9,88E-01	6,09E-02	0,7	1,3	0,8	DOWN	FALSE	FALSE	other
mir409-3p	4395443.1	2,13E-02	7,22E-01	6,08E-02	-4,1	-3,6	-4,0	DOWN	FALSE	FALSE	other
mir885-5p	4395407	2,13E-02	3,48E-01	3,37E-01	0,8	1,3	0,9	DOWN	FALSE	FALSE	other
mir329	4373191	2,27E-02	7,15E-02	4,64E-01	5,6	6,6	6,0	DOWN	FALSE	FALSE	other
mir381	4373020	2,30E-02	2,94E-01	2,73E-01	4,1	5,9	4,8	DOWN	FALSE	FALSE	other

shortname	Identifier	p_{BH} EXP FUS- NORM	p_{BH} EXP FUS+ NORM	p_{MW} EXP FUS- FUS+	Ct normal	Ct FUS-	Ct FUS+	Excl. FUS- NORM	Excl. FUS+ NORM	FUS- vs FUS+	Pattern
mir628-5p	4395544	2,66E-02	7,96E-02	7,15E-01	-1,6	-1,3	-1,4	DOWN	FALSE	FALSE	other
mir195	4373105	2,74E-02	7,41E-01	9,98E-02	-8,9	-8,5	-8,9	DOWN	FALSE	FALSE	other
mir125a-5p	4395309	2,83E-02	9,78E-01	1,20E-01	-5,4	-5,0	-5,4	DOWN	FALSE	FALSE	other
mir487a	4378097	3,70E-02	3,02E-01	7,52E-01	6,7	9,6	8,6	DOWN	FALSE	FALSE	other
mir137	4373301	4,15E-02	8,24E-01	5,29E-02	2,2	2,8	1,5	DOWN	FALSE	FALSE	other
mir337-5p	4395267	4,15E-02	4,34E-01	4,02E-01	0,7	1,2	0,9	DOWN	FALSE	FALSE	other
mir372	4373029	4,15E-02	1,64E-01	8,45E-01	5,7	6,9	6,9	DOWN	FALSE	FALSE	other
mir519b-3p	4395495.1	4,25E-02	5,01E-01	6,09E-02	7,1	9,7	7,4	DOWN	FALSE	FALSE	other
mir515-5p	4373242	4,25E-02	9,26E-01	1,10E-01	9,9	11,8	10,4	DOWN	FALSE	FALSE	other
mir433	4373205	4,78E-02	6,35E-01	2,06E-01	0,4	1,0	0,5	DOWN	FALSE	FALSE	other
mir29b2*	4395277	4,91E-02	9,67E-01	7,97E-02	-0,5	-0,2	-0,5	DOWN	FALSE	FALSE	other
mir572	4381017.1	3,46E-06	7,61E-02	1,14E-02	5,3	3,8	4,7	UP	FALSE	UP	2
mir15b	4373122	5,43E-05	6,89E-02	4,43E-02	-5,7	-6,5	-6,0	UP	FALSE	UP	2
mir198	4395384	1,12E-04	3,04E-01	8,76E-03	3,8	2,2	3,7	UP	FALSE	UP	2
mir449b	4381011	1,57E-04	8,29E-01	8,01E-03	3,0	0,1	2,6	UP	FALSE	UP	2
mir449a	4373207	1,59E-04	7,11E-01	9,16E-03	2,3	-0,6	1,8	UP	FALSE	UP	2
mir660	4380925	1,59E-04	2,60E-01	1,60E-02	-5,1	-5,7	-5,3	UP	FALSE	UP	2
mir532-3p	4395466	1,73E-04	3,09E-01	9,99E-03	-5,3	-5,9	-5,4	UP	FALSE	UP	2
mir30b	4373290	5,78E-04	2,33E-01	1,41E-02	-10,4	-10,9	-10,6	UP	FALSE	UP	2
mir362-3p	4395228	1,73E-03	8,32E-01	1,54E-02	-0,2	-0,9	-0,3	UP	FALSE	UP	2
mir18b	4395328	3,32E-03	4,34E-01	1,24E-02	1,6	0,7	1,4	UP	FALSE	UP	2
mir185	4395382	4,41E-03	9,63E-01	7,65E-03	-1,8	-2,1	-1,8	UP	FALSE	UP	2
mir618	4380996	4,77E-03	3,41E-01	3,82E-02	5,3	4,1	5,0	UP	FALSE	UP	2
mir532-5p	4380928	5,17E-03	8,19E-01	1,41E-02	-6,1	-6,5	-6,1	UP	FALSE	UP	2
mir362-5p	4378092	1,10E-02	7,58E-01	1,41E-02	-1,5	-1,8	-1,4	UP	FALSE	UP	2
mir15b*	4395284	2,13E-02	9,42E-01	3,42E-02	1,3	0,0	1,1	UP	FALSE	UP	2
mir483-5p	4395449	2,42E-02	6,17E-01	1,35E-02	0,3	-0,6	0,5	UP	FALSE	UP	2
mir548d-3p	4381008	3,61E-02	5,29E-01	7,26E-03	9,1	7,7	10,2	UP	FALSE	UP	2
mir629*	4380969	4,80E-04	9,18E-02	8,78E-02	0,4	-0,1	0,1	UP	FALSE	FALSE	other
mir548d-5p	4395348	1,75E-03	3,04E-01	1,24E-01	6,7	5,3	6,6	UP	FALSE	FALSE	other
mir148b	4373129	3,78E-03	5,95E-02	3,93E-01	-1,1	-1,8	-1,6	UP	FALSE	FALSE	other
mir30d	4373059.1	5,99E-03	5,80E-02	1,10E-01	-6,0	-6,7	-6,2	UP	FALSE	FALSE	other
mir551b*	4395457	9,38E-03	3,18E-01	2,93E-01	10,4	8,8	10,1	UP	FALSE	FALSE	other
mir548c-3p	4380993	1,05E-02	1,38E-01	3,45E-01	11,7	10,2	11,2	UP	FALSE	FALSE	other
mir30d	4373059.2	1,21E-02	6,96E-02	1,80E-01	-5,9	-6,6	-6,2	UP	FALSE	FALSE	other
mir635	4380982.2	1,33E-02	8,03E-02	5,03E-01	7,5	6,9	6,3	UP	FALSE	FALSE	other
mir625*	4395543	1,37E-02	1,87E-01	2,66E-01	-2,5	-2,9	-2,7	UP	FALSE	FALSE	other
mir147b	4395373	1,48E-02	6,13E-01	1,39E-01	10,0	7,7	10,2	UP	FALSE	FALSE	other
mir653	4395403	1,54E-02	1,33E-01	3,85E-01	11,9	10,6	11,8	UP	FALSE	FALSE	other
mir649	4381005.1	1,58E-02	7,96E-02	3,26E-01	11,2	9,3	10,0	UP	FALSE	FALSE	other
mir513-3p	4395202	1,61E-02	2,70E-01	3,93E-01	8,6	7,4	7,8	UP	FALSE	FALSE	other
mir635	4380982.1	1,73E-02	1,68E-01	7,29E-01	7,1	6,1	6,8	UP	FALSE	FALSE	other
mir639	4380987.1	1,79E-02	3,21E-01	1,95E-01	3,7	3,0	3,3	UP	FALSE	FALSE	other
mir651	4381007	2,11E-02	3,96E-01	3,57E-01	7,9	6,7	8,1	UP	FALSE	FALSE	other
mir18a*	4395534	2,42E-02	7,63E-01	9,67E-02	5,0	4,6	5,1	UP	FALSE	FALSE	other
mir517b	4373244	2,58E-02	8,03E-02	6,37E-01	10,5	8,8	9,3	UP	FALSE	FALSE	other
mir489	4395469	2,58E-02	2,31E-01	9,76E-01	1,0	0,6	0,6	UP	FALSE	FALSE	other
mir30a	4373061	2,74E-02	2,42E-01	2,17E-01	-8,7	-9,2	-8,8	UP	FALSE	FALSE	other
mir892b	4395325	3,21E-02	1,40E-01	4,55E-01	12,1	11,5	12,2	UP	FALSE	FALSE	other
mir544	4395376	3,70E-02	4,23E-01	2,11E-01	11,2	10,1	11,6	UP	FALSE	FALSE	other
mir526b*	4395494	4,91E-02	1,87E-01	4,37E-01	7,1	6,2	6,4	UP	FALSE	FALSE	other
let7c	4373167	6,13E-01	5,06E-05	1,48E-02	-8,7	-8,7	-8,1	FALSE	DOWN	UP	3
mir450b-5p	4395318	7,07E-01	2,23E-03	1,00E-02	3,3	3,8	6,7	FALSE	DOWN	UP	3
mir450a	4395414	9,67E-01	3,42E-02	3,86E-02	3,1	2,9	4,0	FALSE	DOWN	UP	3
mir589*	4380953	5,19E-02	4,69E-03	7,84E-01	1,5	1,9	1,9	FALSE	DOWN	FALSE	other
mir210	4373089	6,67E-02	1,88E-02	5,83E-01	-2,7	-2,2	-2,2	FALSE	DOWN	FALSE	other
mir30a*	4373062	6,86E-02	5,02E-03	8,31E-01	-8,3	-7,9	-7,9	FALSE	DOWN	FALSE	other
mir873	4395467	8,39E-02	2,86E-02	3,37E-01	9,3	11,8	12,4	FALSE	DOWN	FALSE	other
mir99a*	4395252	9,44E-02	1,96E-06	1,20E-01	-3,2	-3,0	-2,4	FALSE	DOWN	FALSE	other
mir365	4373194	9,63E-02	5,11E-04	2,72E-01	-6,8	-6,4	-6,2	FALSE	DOWN	FALSE	other

shortname	Identifier	p_{BH} EXP FUS- NORM	p_{BH} EXP FUS+ NORM	p_{MW} EXP FUS- FUS+	Ct normal	Ct FUS-	Ct FUS+	Excl. FUS- NORM	Excl. FUS+ NORM	FUS- vs FUS+	Pattern
mir545*	4395377	1,20E-01	4,63E-02	7,37E-01	4,5	5,0	4,9	FALSE	DOWN	FALSE	other
mir99a	4373008	1,20E-01	1,70E-02	8,07E-01	-9,1	-8,8	-8,6	FALSE	DOWN	FALSE	other
mir193b	4395478	1,33E-01	1,27E-03	3,85E-01	-8,2	-7,8	-7,6	FALSE	DOWN	FALSE	other
mir100	4373160	1,35E-01	2,86E-02	6,47E-01	-9,5	-9,2	-9,0	FALSE	DOWN	FALSE	other
mir545	4395378	1,56E-01	6,84E-03	3,45E-01	1,5	1,8	2,0	FALSE	DOWN	FALSE	other
mir486-5p	4378096	1,94E-01	1,61E-02	4,78E-01	4,8	8,2	8,7	FALSE	DOWN	FALSE	other
mir29a	4395223	1,96E-01	1,37E-03	2,47E-01	-9,8	-9,6	-9,5	FALSE	DOWN	FALSE	other
mir551b	4380945	2,06E-01	1,59E-02	5,52E-01	5,8	7,7	7,7	FALSE	DOWN	FALSE	other
mir217	4395448	2,07E-01	2,67E-03	1,06E-01	6,0	7,2	8,5	FALSE	DOWN	FALSE	other
mir624*	4380964	2,25E-01	4,59E-03	1,39E-01	5,3	6,2	7,3	FALSE	DOWN	FALSE	other
mir331-5p	4395344	2,43E-01	4,42E-02	6,62E-01	1,5	1,9	2,0	FALSE	DOWN	FALSE	other
mir363*	4380917	2,74E-01	3,32E-02	7,29E-01	0,1	1,3	0,7	FALSE	DOWN	FALSE	other
mir29b	4373288	3,23E-01	1,61E-02	2,06E-01	-2,9	-2,6	-2,2	FALSE	DOWN	FALSE	other
mir542-3p	4378101	3,23E-01	3,27E-02	3,85E-01	3,3	3,6	4,0	FALSE	DOWN	FALSE	other
mir509-3p	4395347	3,41E-01	3,17E-02	1,31E-01	2,2	2,9	2,8	FALSE	DOWN	FALSE	other
mir29c*	4381131	3,80E-01	3,04E-02	3,29E-01	-2,6	-2,6	-2,4	FALSE	DOWN	FALSE	other
mir125b	4373148	4,62E-01	4,31E-02	1,65E-01	-10,1	-10,0	-9,7	FALSE	DOWN	FALSE	other
mir10a	4373153	5,66E-01	5,02E-03	1,75E-01	-3,5	-3,4	-2,9	FALSE	DOWN	FALSE	other
mir149	4395366	6,48E-01	2,26E-02	1,58E-01	-7,3	-7,1	-6,5	FALSE	DOWN	FALSE	other
mir503	4373228	6,52E-01	1,59E-02	8,51E-02	4,7	4,8	5,4	FALSE	DOWN	FALSE	other
mir500	4395539	6,67E-01	6,34E-03	7,45E-02	-1,5	-1,4	-1,2	FALSE	DOWN	FALSE	other
mir500*	4373225	7,03E-01	1,82E-03	8,51E-02	2,4	3,0	3,2	FALSE	DOWN	FALSE	other
mir219-5p	4373080	7,03E-01	3,66E-02	3,41E-01	5,5	5,9	6,2	FALSE	DOWN	FALSE	other
mir550*	4380954	7,28E-01	1,25E-02	1,28E-01	6,1	6,7	7,2	FALSE	DOWN	FALSE	other
mir148a*	4395245	6,33E-02	5,50E-06	2,71E-02	4,0	3,3	2,4	FALSE	UP	DOWN	4
mir345	4395297	1,63E-01	2,13E-02	8,75E-03	-5,1	-4,9	-5,5	FALSE	UP	DOWN	4
mir520c-3p	4395511	3,28E-01	4,78E-03	3,41E-02	5,4	5,2	4,5	FALSE	UP	DOWN	4
mir200b*	4395385	5,17E-01	6,79E-05	7,32E-03	-4,1	-2,7	-4,9	FALSE	UP	DOWN	4
mir196b	4395326	6,17E-01	3,27E-02	3,55E-02	-5,5	-5,2	-5,9	FALSE	UP	DOWN	4
mir95	4373011	6,52E-01	1,79E-03	2,13E-02	-2,9	-3,0	-3,4	FALSE	UP	DOWN	4
mir200b	4395362	6,87E-01	2,54E-03	1,61E-02	-8,9	-8,9	-9,7	FALSE	UP	DOWN	4
mir103	4373158	8,28E-02	3,04E-02	7,26E-01	-5,0	-5,2	-5,2	FALSE	UP	FALSE	other
mir335*	4395296	1,26E-01	3,66E-02	6,41E-01	-2,1	-2,3	-2,4	FALSE	UP	FALSE	other
let7e	4395517	1,79E-01	1,38E-02	1,43E-01	-10,7	-10,9	-11,3	FALSE	UP	FALSE	other
mir708	4395452	2,26E-01	4,27E-03	1,06E-01	-1,6	-1,8	-2,1	FALSE	UP	FALSE	other
mir335	4373045	2,96E-01	9,13E-03	1,75E-01	-4,5	-4,7	-5,0	FALSE	UP	FALSE	other
mir17*	4395532	2,96E-01	3,59E-03	3,61E-01	1,5	1,4	0,9	FALSE	UP	FALSE	other
mir200a	4378069	4,25E-01	1,79E-03	6,74E-02	-6,8	-5,5	-7,3	FALSE	UP	FALSE	other
mir200a*	4373273	6,95E-01	2,50E-04	7,46E-02	-1,4	-0,3	-2,1	FALSE	UP	FALSE	other
let7g*	4395229	7,11E-01	6,70E-03	7,44E-02	2,8	2,7	2,4	FALSE	UP	FALSE	other
mir429	4373203	8,68E-01	2,60E-04	5,61E-02	-5,5	-4,4	-6,3	FALSE	UP	FALSE	other
mir1	4395333	2,41E-08	9,35E-06	1,60E-03	-5,6	-3,0	-4,2	FALSE	FALSE	DOWN	other
mir133a	4395357	2,41E-08	1,40E-06	2,30E-03	-9,0	-6,3	-7,4	FALSE	FALSE	DOWN	other
mir133b	4395358	2,41E-08	1,01E-06	1,24E-02	-4,7	-2,3	-3,3	FALSE	FALSE	DOWN	other
mir143*	4395257	5,30E-08	9,12E-06	3,43E-03	-4,6	0,8	-2,1	FALSE	FALSE	DOWN	other
mir145	4395389	7,49E-08	1,03E-05	1,89E-02	-14,0	-12,2	-12,9	FALSE	FALSE	DOWN	other
mir455-5p	4378098	9,32E-08	1,73E-05	1,35E-02	-3,4	-1,7	-2,5	FALSE	FALSE	DOWN	other
mir152	4395170	1,70E-07	1,44E-04	4,36E-03	-6,1	-5,0	-5,5	FALSE	FALSE	DOWN	other
mir221*	4395207	1,70E-07	3,80E-05	1,24E-02	2,1	3,8	3,2	FALSE	FALSE	DOWN	other
mir891a	4395302	2,44E-07	7,04E-04	3,61E-03	3,0	7,4	4,5	FALSE	FALSE	DOWN	other
mir190b	4395374	2,76E-07	5,09E-07	1,00E-02	1,6	0,0	-0,9	FALSE	FALSE	DOWN	other
mir143	4395360	3,64E-07	2,02E-04	3,55E-02	-10,8	-9,1	-9,8	FALSE	FALSE	DOWN	other
mir181c*	4395444	3,84E-07	5,56E-05	4,43E-02	1,9	8,7	5,3	FALSE	FALSE	DOWN	other
mir376c	4395233	4,40E-07	2,04E-04	4,11E-02	-4,0	-2,9	-3,3	FALSE	FALSE	DOWN	other
mir411*	4395349	5,11E-07	6,34E-03	3,30E-02	4,7	8,9	6,2	FALSE	FALSE	DOWN	other
mir72*	4395425	5,36E-07	1,76E-02	2,96E-03	3,4	6,0	4,2	FALSE	FALSE	DOWN	other
mir375	4373027	6,32E-07	1,01E-06	3,04E-02	-10,4	-11,7	-12,1	FALSE	FALSE	DOWN	other
mir543	4395487	7,28E-07	9,76E-03	3,59E-03	0,5	2,3	1,1	FALSE	FALSE	DOWN	other
mir455-3p	4395355	1,25E-06	1,11E-03	1,74E-02	-3,5	-2,0	-2,7	FALSE	FALSE	DOWN	other
mir145*	4395260	2,41E-06	6,70E-03	3,79E-03	-5,9	-4,7	-5,5	FALSE	FALSE	DOWN	other

shortname	Identifier	p_{BH} EXP FUS- NORM	p_{BH} EXP FUS+ NORM	p_{MW} EXP FUS- FUS+	Ct normal	Ct FUS-	Ct FUS+	Excl. FUS- NORM	Excl. FUS+ NORM	FUS- vs FUS+	Pattern
mir320	4395388	2,41E-06	1,79E-03	4,93E-02	-8,2	-7,2	-7,6	FALSE	FALSE	DOWN	other
mir125b1*	4395489	2,04E-05	8,78E-03	4,75E-02	2,2	3,7	2,9	FALSE	FALSE	DOWN	other
mir411	4381013	3,14E-05	3,72E-02	1,47E-02	-2,5	-1,6	-2,1	FALSE	FALSE	DOWN	other
mir214*	4395404	4,20E-05	4,42E-02	1,14E-02	-2,8	-2,2	-2,6	FALSE	FALSE	DOWN	other
mir495	4381078	1,24E-04	3,42E-02	3,97E-02	-2,6	-1,8	-2,3	FALSE	FALSE	DOWN	other
mir497	4373222.1	8,81E-02	1,06E-01	5,11E-04	-4,5	-4,1	-4,8	FALSE	FALSE	DOWN	other
mir376a*	4395238	9,63E-02	6,96E-02	2,19E-03	4,5	5,3	4,1	FALSE	FALSE	DOWN	other
mir140-5p	4373374	9,75E-02	3,61E-01	1,54E-02	-6,8	-6,6	-6,9	FALSE	FALSE	DOWN	other
mir497	4373222.2	1,26E-01	6,67E-02	9,37E-04	-4,4	-4,1	-4,8	FALSE	FALSE	DOWN	other
mir302b	4378071	5,70E-01	1,99E-01	3,82E-02	10,6	12,2	9,7	FALSE	FALSE	DOWN	other
mir1381*	4395273	2,27E-07	6,44E-05	2,82E-02	0,9	-0,4	0,0	FALSE	FALSE	UP	other
mir622	4380961.2	2,44E-07	7,03E-04	2,05E-02	6,5	2,5	3,7	FALSE	FALSE	UP	other
mir622	4380961.1	3,84E-07	3,34E-03	1,30E-02	6,3	2,7	4,4	FALSE	FALSE	UP	other
mir27b*	4395285	4,85E-07	6,25E-06	4,93E-02	-0,7	0,6	1,0	FALSE	FALSE	UP	other
mir18a	4395533	1,89E-05	7,03E-04	4,59E-02	-1,5	-2,9	-2,3	FALSE	FALSE	UP	other
mir629	4395547	9,41E-05	3,09E-02	4,76E-02	2,1	1,4	1,8	FALSE	FALSE	UP	other
mir590-5p	4395176	2,28E-01	8,27E-02	6,66E-03	-4,4	-4,5	-4,2	FALSE	FALSE	UP	other
mir570	4395458	2,85E-01	1,87E-01	3,68E-02	5,5	5,4	6,3	FALSE	FALSE	UP	other
mir339-5p	4395368	4,40E-01	1,10E-01	3,68E-02	-2,1	-2,3	-1,6	FALSE	FALSE	UP	other

Supplementary Table 9: Methylation of polycomb group protein target genes (according to (YU 2007)) and correlation of target gene methylation to *EZH2* expression. Column headings are as follows: (Gene name) Gene name; (p_{BH} hyper T/N) BH corrected Mann-Whitney p-value of most significant TSS associated DMR hypermethylated in tumour compared to normal samples; (p_{BH} hypo T/N) BH corrected Mann-Whitney p-value of most significant TSS associated DMR hypomethylated in tumour compared to normal samples; (p_{BH} hyper FUS -/+) BH corrected Mann-Whitney p-value of most significant TSS associated DMR hypermethylated in FUS- compared to FUS+; (p_{BH} hypo FUS -/+) BH corrected Mann-Whitney p-value of most significant TSS associated DMR hypomethylated in FUS- compared to FUS+; (Rho T/N) Spearman correlation of *EZH2* expression and gene promoter methylation in all samples; (p_{BH} Rho T/N) BH corrected p-value of correlation; (Rho FUS-/N) Spearman correlation of *EZH2* expression and gene promoter methylation in FUS- and normal samples; (p_{BH} Rho FUS-/N) BH corrected p-value of correlation; (Rho FUS+/N) Spearman correlation of *EZH2* expression and gene promoter methylation in FUS+ and normal samples; (p_{BH} Rho FUS+/N) BH corrected p-value of correlation. Significant values are highlighted in gray.

	Differential methylation				Correlation of EZH2 expression and gene methylation					
Gene name	p_{BH} hyper T/N	p_{BH} hypo T/N	p_{BH} hyper FUS -/+	p_{BH} hypo FUS -/+	Rho T/N	p_{BH} Rho T/N	Rho FUS-/N	p_{BH} Rho FUS-/N	Rho FUS+/N	p_{BH} Rho FUS+/N
ABP1	5,6E-04	2,5E-01	1,3E-01	4,8E-02	0,27	4,5E-02	0,19	2,3E-01	0,26	3,6E-01
ACTC1	3,9E-12	NA	2,7E-01	3,5E-01	0,57	1,6E-07	0,57	1,6E-07	0,46	3,5E-02
ADAMTS3	6,4E-02	NA	3,3E-01	NA	0,35	4,4E-03	0,35	4,4E-03	0,30	2,6E-01
ALDH1A2	6,2E-13	2,2E-01	2,8E-02	1,3E-01	0,53	1,0E-06	0,53	1,0E-06	0,33	1,7E-01
ARHGAP20	6,8E-02	2,3E-01	7,5E-01	1,8E-01	0,41	4,2E-04	0,41	4,2E-04	0,32	2,0E-01
ARMCX1	5,0E-04	NA	2,3E-01	NA	0,50	6,8E-06	0,50	6,8E-06	0,36	1,3E-01
BDNF	5,4E-07	NA	2,4E-01	8,5E-01	0,49	9,1E-06	0,46	4,9E-05	0,37	1,1E-01
BMPER	9,4E-08	NA	4,3E-02	6,1E-01	0,63	4,0E-09	0,63	4,0E-09	0,29	2,7E-01
C1orf114	7,6E-14	NA	6,1E-01	8,6E-01	0,57	1,5E-07	0,48	2,2E-05	0,39	8,4E-02
C20orf103	1,5E-13	NA	7,1E-03	3,8E-01	0,55	3,2E-07	0,55	3,2E-07	0,46	3,5E-02
CACNB2	3,6E-12	5,5E-01	7,9E-02	5,3E-01	0,58	7,9E-08	0,58	7,9E-08	0,43	6,5E-02
CAMK1G	2,5E-04	NA	2,0E-01	NA	0,26	5,8E-02	0,26	5,8E-02	0,18	6,3E-01
CD40	9,2E-08	4,6E-02	3,3E-01	7,6E-01	0,51	4,4E-06	0,51	4,4E-06	0,32	2,0E-01
CHI3L1	3,9E-05	7,9E-01	4,5E-02	NA	0,37	2,0E-03	0,35	4,3E-03	0,33	1,7E-01
CLIC5	2,2E-06	1,9E-01	1,3E-01	1,9E-01	0,34	7,4E-03	0,34	7,4E-03	0,34	1,7E-01
COLEC12	7,8E-03	NA	2,9E-02	NA	0,36	3,4E-03	0,36	3,4E-03	0,14	7,3E-01
CPA6	7,2E-12	NA	3,3E-02	NA	0,61	1,9E-08	0,61	1,9E-08	0,42	6,5E-02
CPAMD8	8,8E-02	NA	8,6E-02	NA	0,36	3,1E-03	0,36	3,1E-03	0,13	7,6E-01
CSMD3	4,0E-07	1,3E-05	8,8E-01	4,3E-02	0,49	9,1E-06	0,49	9,1E-06	0,34	1,6E-01
CTSC	NA	NA	NA	1,6E-02	0,42	3,0E-04	0,42	3,0E-04	0,30	2,7E-01
CTSG	7,7E-01	6,0E-01	NA	3,1E-01	0,11	5,7E-01	0,11	5,7E-01	0,19	6,1E-01
CXCL12	1,1E-07	6,2E-01	1,0E-01	7,6E-01	0,37	2,1E-03	0,37	2,1E-03	0,18	6,3E-01
CYB5R2	7,6E-13	NA	5,8E-02	NA	0,56	2,6E-07	0,56	2,6E-07	0,32	2,2E-01
CYP2E1	3,0E-02	6,9E-02	1,1E-01	1,5E-01	0,25	7,3E-02	0,25	7,3E-02	0,23	4,8E-01
DARC	NA	1,1E-03	NA	NA	0,01	9,8E-01	0,01	9,8E-01	0,03	9,5E-01
DKK2	1,6E-05	1,8E-01	NA	7,9E-01	0,50	8,0E-06	0,50	8,0E-06	0,40	6,5E-02
DNAJB4	NA	1,6E-03	NA	NA	0,16	3,6E-01	0,16	3,6E-01	0,24	4,5E-01
DOK5	2,1E-11	5,2E-01	NA	1,9E-01	0,58	9,1E-08	0,58	9,1E-08	0,41	6,5E-02
DPP4	2,7E-02	NA	9,6E-02	8,4E-02	0,21	1,7E-01	0,17	2,9E-01	0,30	2,5E-01
EDNRA	1,2E-03	NA	2,5E-01	5,4E-01	0,42	3,0E-04	0,42	3,0E-04	0,27	3,1E-01
EPHB6	1,3E-10	2,7E-05	7,2E-02	NA	0,64	4,0E-09	0,64	4,0E-09	0,48	2,6E-02
FAM49A	2,9E-04	8,0E-01	2,2E-01	9,4E-01	0,35	4,8E-03	0,31	1,7E-02	0,37	9,7E-02
FEV	1,1E-13	NA	3,4E-02	NA	0,58	8,9E-08	0,58	8,9E-08	0,43	5,8E-02
FGF13	7,1E-01	NA	5,6E-01	NA	0,49	9,2E-06	0,49	9,2E-06	0,44	4,5E-02
EGFLAM	7,2E-13	8,5E-01	1,6E-01	4,7E-01	0,54	7,9E-07	0,49	1,0E-05	0,46	3,5E-02
FST	8,5E-12	8,8E-01	7,4E-01	3,2E-01	0,53	1,7E-06	0,53	1,7E-06	0,36	1,2E-01
GNAL	3,0E-04	7,7E-02	NA	4,7E-01	0,46	5,5E-05	0,46	5,5E-05	0,38	8,4E-02
HFE	1,2E-09	9,0E-01	1,9E-01	NA	0,47	3,2E-05	0,47	3,2E-05	0,39	7,7E-02
HHIP	7,1E-09	1,3E-02	2,7E-01	3,3E-01	0,58	1,0E-07	0,58	1,0E-07	0,38	8,4E-02
HPSE2	1,1E-13	NA	3,3E-01	8,5E-01	0,62	8,7E-09	0,62	8,7E-09	0,34	1,7E-01
ITGAL	8,1E-06	NA	8,2E-02	8,8E-01	0,54	6,4E-07	0,54	6,4E-07	0,42	6,5E-02
ITGB2	3,5E-04	2,7E-01	4,2E-02	4,9E-01	0,36	3,6E-03	0,36	3,6E-03	0,20	5,8E-01
ITGB8	5,2E-01	5,2E-01	NA	5,8E-01	0,44	1,4E-04	0,44	1,4E-04	0,20	5,8E-01
KLF8	1,6E-13	NA	1,1E-01	NA	0,58	7,9E-08	0,56	2,6E-07	0,49	2,6E-02

	Differential methylation				Correlation of EZH2 expression and gene methylation					
KRT17	6,9E-10	NA	4,7E-03	NA	0,50	5,4E-06	0,45	7,1E-05	0,33	1,9E-01
LMOD1	NA	NA	NA	NA	0,35	3,9E-03	0,35	3,9E-03	0,41	6,5E-02
LYNX1	6,6E-07	NA	2,1E-02	NA	0,55	3,6E-07	0,55	3,6E-07	0,34	1,7E-01
MEOX2	6,0E-12	NA	6,1E-01	3,4E-01	0,60	1,9E-08	0,60	1,9E-08	0,34	1,6E-01
MKX	2,8E-13	NA	1,5E-01	NA	0,60	1,9E-08	0,60	1,9E-08	0,47	3,5E-02
MMEL1	5,7E-06	5,9E-01	3,6E-02	4,2E-01	0,47	2,4E-05	0,46	4,9E-05	0,31	2,4E-01
MPPED2	9,2E-13	4,3E-01	7,3E-01	NA	0,62	5,9E-09	0,62	5,9E-09	0,43	6,2E-02
MYH11	2,3E-01	NA	1,7E-01	NA	0,38	1,6E-03	0,38	1,6E-03	0,27	3,3E-01
MYH7	2,2E-03	3,0E-02	1,6E-01	8,9E-01	0,33	7,9E-03	0,33	7,9E-03	0,24	4,3E-01
MYL4	4,9E-01	4,7E-01	7,1E-01	2,8E-01	0,28	3,8E-02	0,28	3,8E-02	0,22	5,0E-01
MYO6	NA	7,7E-01	9,1E-01	6,6E-01	0,13	4,6E-01	0,11	5,6E-01	0,24	4,2E-01
MYOM2	2,5E-04	1,1E-02	1,7E-01	3,6E-01	0,41	5,1E-04	0,41	5,1E-04	0,28	2,7E-01
MYRIP	6,0E-09	NA	5,5E-02	2,4E-01	0,53	1,3E-06	0,53	1,3E-06	0,32	2,1E-01
NCKAP1L	3,2E-01	NA	NA	6,1E-01	0,15	3,8E-01	0,15	3,8E-01	0,18	6,3E-01
NR4A3	5,2E-09	9,2E-01	3,7E-02	7,5E-01	0,56	2,6E-07	0,56	2,6E-07	0,30	2,5E-01
OSMR	6,3E-12	6,7E-01	6,2E-02	NA	0,63	4,0E-09	0,63	4,0E-09	0,40	6,5E-02
PAGE4	5,5E-02	NA	4,3E-02	NA	0,45	9,1E-05	0,45	9,1E-05	0,35	1,5E-01
PCDHB5	3,4E-01	NA	6,4E-01	NA	0,41	4,4E-04	0,41	4,4E-04	0,25	3,8E-01
PENK	3,2E-12	7,1E-01	3,4E-01	3,0E-01	0,57	1,1E-07	0,57	1,1E-07	0,41	6,5E-02
PMP22	8,7E-01	5,7E-03	3,4E-01	2,2E-01	0,49	1,3E-05	0,49	1,3E-05	0,31	2,5E-01
PRKG1	8,9E-01	3,2E-04	NA	1,8E-01	0,32	1,3E-02	0,32	1,3E-02	0,14	7,5E-01
PTGER3	4,2E-10	NA	4,1E-02	4,5E-01	0,57	1,4E-07	0,57	1,4E-07	0,46	3,5E-02
RLN1	NA	NA	NA	NA	0,12	5,1E-01	0,12	5,1E-01	0,17	6,5E-01
SATB2	3,4E-11	1,1E-02	4,7E-02	1,7E-01	0,61	1,9E-08	0,61	1,9E-08	0,41	6,5E-02
SLC1A1	6,1E-01	6,3E-01	4,3E-02	1,3E-01	0,32	1,2E-02	0,23	1,2E-01	0,30	2,5E-01
SLC40A1	7,2E-01	NA	1,0E-01	6,6E-01	0,49	1,2E-05	0,49	1,2E-05	0,25	4,0E-01
SLFN11	6,3E-12	NA	NA	2,2E-01	0,44	1,2E-04	0,44	1,2E-04	0,29	2,7E-01
SNCA	8,2E-09	NA	5,3E-01	1,7E-01	0,48	2,0E-05	0,48	2,0E-05	0,27	3,1E-01
SOCS2	NA	NA	NA	NA	0,24	9,2E-02	0,21	1,5E-01	0,11	8,3E-01
SP8	4,4E-10	NA	3,1E-01	3,6E-01	0,38	1,2E-03	0,37	2,5E-03	0,23	4,8E-01
SRPX	NA	NA	NA	NA	0,33	8,6E-03	0,33	8,6E-03	0,20	5,9E-01
TAC1	6,8E-14	NA	6,0E-01	4,5E-01	0,57	1,3E-07	0,57	1,3E-07	0,35	1,4E-01
TGFB1	5,1E-02	NA	1,2E-01	NA	0,28	3,3E-02	0,25	7,2E-02	0,09	8,5E-01
TNC	4,0E-03	8,7E-01	5,3E-02	4,9E-01	0,39	1,0E-03	0,37	2,1E-03	0,31	2,5E-01
TNFRSF19	9,2E-02	NA	1,5E-01	NA	0,29	2,4E-02	0,29	2,4E-02	0,30	2,5E-01
TRPS1	5,4E-08	NA	1,4E-01	NA	0,43	1,8E-04	0,37	2,1E-03	0,45	3,5E-02
WDR18	1,3E-01	NA	4,5E-02	5,4E-01	0,34	5,5E-03	0,34	5,5E-03	0,17	6,7E-01
WNT16	3,4E-11	NA	1,2E-01	NA	0,58	9,1E-08	0,58	1,1E-07	0,39	7,6E-02
WNT2	3,5E-01	NA	5,1E-02	3,1E-01	0,50	5,1E-06	0,50	5,1E-06	0,16	6,8E-01
WWTR1	8,3E-05	NA	1,2E-02	NA	0,46	5,4E-05	0,46	5,4E-05	0,29	2,7E-01
Significant genes	**57**	**11**	**18**	**3**		**74**		**71**		**9**
Overlap	**7**		**0**					**9**		

6 LITERATURE

Alajez NM, Shi W, Hui AB, Bruce J, Lenarduzzi M *et al*: **Enhancer of Zeste homolog 2 (EZH2) is overexpressed in recurrent nasopharyngeal carcinoma and is regulated by miR-26a, miR-101, and miR-98**. *Cell Death Dis* 2010, **1**:e85.

Amir RE, Van den Veyver IB, Wan M, Tran CQ, Francke U *et al*: **Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2**. *Nat Genet* 1999, **23**(2):185-188.

Andreoiu M, Cheng L: **Multifocal prostate cancer: biologic, prognostic, and therapeutic implications**. *Hum Pathol* 2010, **41**(6):781-793.

Ball MP, Li JB, Gao Y, Lee JH, LeProust EM *et al*: **Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells**. *Nat Biotechnol* 2009, **27**(4):361-368.

Barbieri CE, Baca SC, Lawrence MS, Demichelis F, Blattner M *et al*: **Exome sequencing identifies recurrent SPOP, FOXA1 and MED12 mutations in prostate cancer**. *Nat Genet* 2012.

Barry M, Perner S, Demichelis F, Rubin MA: **TMPRSS2-ERG fusion heterogeneity in multifocal prostate cancer: clinical and biologic implications**. *Urology* 2007, **70**(4):630-633.

Bartel DP: **MicroRNAs: genomics, biogenesis, mechanism, and function**. *Cell* 2004, **116**(2):281-297.

Bartel DP: **MicroRNAs: target recognition and regulatory functions**. *Cell* 2009, **136**(2):215-233.

Bastus NC, Boyd LK, Mao X, Stankiewicz E, Kudahetti SC *et al*: **Androgen-induced TMPRSS2:ERG fusion in nonmalignant prostate epithelial cells**. *Cancer Res* 2010, **70**(23):9544-9548.

Bell AC, Felsenfeld G: **Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene**. *Nature* 2000, **405**(6785):482-485.

Ben-Josef E, Yang SY, Ji TH, Bidart JM, Garde SV *et al*: **Hormone-refractory prostate cancer cells express functional follicle-stimulating hormone receptor (FSHR)**. *J Urol* 1999, **161**(3):970-976.

Bengtsson H, Wirapati P, Speed TP: **A single-array preprocessing method for estimating full-resolution raw copy numbers from all Affymetrix genotyping arrays including GenomeWideSNP 5 & 6**. *Bioinformatics* 2009, **25**(17):2149-2156.

Benjamini Y, Hochberg Y: **Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing**. *Journal of the Royal Statistical Society Series B-Methodological* 1995, **57**(1):289-300.

Berger MF, Lawrence MS, Demichelis F, Drier Y, Cibulskis K *et al*: **The genomic complexity of primary human prostate cancer**. *Nature* 2011, **470**(7333):214-220.

Bird A: **DNA methylation patterns and epigenetic memory**. *Genes Dev* 2002, **16**(1):6-21.

Bird AP: **CpG-rich islands and the function of DNA methylation**. *Nature* 1986, **321**(6067):209-213.

Bird AP, Wolffe AP: **Methylation-induced repression--belts, braces, and chromatin**. *Cell* 1999, **99**(5):451-454.

Boerno ST, Grimm C, Lehrach H, Schweiger MR: **Next-generation sequencing technologies for DNA methylation analyses in cancer genomics**. *Epigenomics* 2010, **2**(2):199-207.

Bonci D, Coppola V, Musumeci M, Addario A, Giuffrida R *et al*: **The miR-15a-miR-16-1 cluster controls prostate cancer by targeting multiple oncogenic activities**. *Nat Med* 2008, **14**(11):1271-1277.

Bostwick DG, Cheng L: **Precursors of prostate cancer**. *Histopathology* 2012, **60**(1):4-27.

Brase JC, Johannes M, Schlomm T, Falth M, Haese A *et al*: **Circulating miRNAs are correlated with tumor progression in prostate cancer**. *Int J Cancer* 2011, **128**(3):608-616.

Braun M, Scheble VJ, Menon R, Scharf G, Wilbertz T *et al*: **Relevance of cohort design for studying the frequency of the ERG rearrangement in prostate cancer**. *Histopathology* 2011, **58**(7):1028-1036.

Brenner JC, Ateeq B, Li Y, Yocum AK, Cao Q *et al*: **Mechanistic rationale for inhibition of poly(ADP-ribose) polymerase in ETS gene fusion-positive prostate cancer**. *Cancer Cell* 2011, **19**(5):664-678.

Bryzgunova OE, Skvortsova TE, Kolesnikova EV, Starikov AV, Rykova EY *et al*: **Isolation and comparative study of cell-free nucleic acids from human urine**. *Ann N Y Acad Sci* 2006, **1075**:334-340.

Butcher LM, Beck S: **AutoMeDIP-seq: a high-throughput, whole genome, DNA methylation assay**. *Methods* 2010, **52**(3):223-231.

Bylund A, Lundin E, Zhang JX, Nordin A, Kaaks R *et al*: **Randomised controlled short-term intervention pilot study on rye bran bread in prostate cancer**. *Eur J Cancer Prev* 2003, **12**(5):407-415.

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H *et al*: **Role of histone H3 lysine 27 methylation in Polycomb-group silencing**. *Science* 2002, **298**(5595):1039-1043.

Caren H, Djos A, Nethander M, Sjoberg RM, Kogner P *et al*: **Identification of epigenetically regulated genes that predict patient outcome in neuroblastoma**. *Bmc Cancer* 2011, **11**:66.

Carvalho JR, Filipe L, Costa VL, Ribeiro FR, Martins AT *et al*: **Detailed analysis of expression and promoter methylation status of apoptosis-related genes in prostate cancer**. *Apoptosis* 2010, **15**(8):956-965.

Catalona WJ, Smith DS, Ratliff TL, Dodds KM, Coplen DE *et al*: **Measurement of prostate-specific antigen in serum as a screening test for prostate cancer**. *N Engl J Med* 1991, **324**(17):1156-1161.

Cazares LH, Drake RR, Esquela-Kirscher A, Lance RS, Semmes OJ *et al*: **Molecular pathology of prostate cancer**. *Cancer Biomark* 2011, **9**(1-6):441-459.

Chavez L, Jozefczuk J, Grimm C, Dietrich J, Timmermann B *et al*: **Computational analysis of genome-wide DNA methylation during the differentiation of human embryonic stem cells along the endodermal lineage**. *Genome Res* 2010, **20**(10):1441-1450.

Chen Y, Stallings RL: **Differential patterns of microRNA expression in neuroblastoma are correlated with prognosis, differentiation, and apoptosis**. *Cancer Res* 2007, **67**(3):976-983.

Cheng L, Montironi R, Bostwick DG, Lopez-Beltran A, Berney DM: **Staging of prostate cancer**. *Histopathology* 2012, **60**(1):87-117.

Chiba T, Suzuki E, Negishi M, Saraya A, Miyagi S *et al*: **3-deazaneplanocin A is a promising therapeutic agent for the eradication of tumor-initiating hepatocellular carcinoma cells**. *Int J Cancer* 2011.

Chowdhury I, Tharakan B, Bhat GK: **Current concepts in apoptosis: the physiological suicide program revisited**. *Cell Mol Biol Lett* 2006, **11**(4):506-525.

Christensson A, Laurell CB, Lilja H: **Enzymatic activity of prostate-specific antigen and its reactions with extracellular serine proteinase inhibitors**. *Eur J Biochem* 1990, **194**(3):755-763.

Cimmino A, Calin GA, Fabbri M, Iorio MV, Ferracin M *et al*: **miR-15 and miR-16 induce apoptosis by targeting BCL2**. *Proc Natl Acad Sci U S A* 2005, **102**(39):13944-13949.

Clark SJ, Harrison J, Paul CL, Frommer M: **High sensitivity mapping of methylated cytosines**. *Nucleic Acids Res* 1994, **22**(15):2990-2997.

Coolen MW, Stirzaker C, Song JZ, Statham AL, Kassir Z *et al*: **Consolidation of the cancer genome into domains of repressive chromatin by long-range epigenetic silencing (LRES) reduces transcriptional plasticity**. *Nat Cell Biol* 2010, **12**(3):235-246.

Crocitto LE, Korns D, Kretzner L, Shevchuk T, Blair SL *et al*: **Prostate cancer molecular markers GSTP1 and hTERT in expressed prostatic secretions as predictors of biopsy results**. *Urology* 2004, **64**(4):821-825.

Cross SH, Bird AP: **CpG islands and genes**. *Curr Opin Genet Dev* 1995, **5**(3):309-314.

Csankovszki G, Nagy A, Jaenisch R: **Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation**. *J Cell Biol* 2001, **153**(4):773-784.

Cullen J, Brassell SA, Chen Y, Porter C, L'Esperance J *et al*: **Racial/Ethnic patterns in prostate cancer outcomes in an active surveillance cohort**. *Prostate Cancer* 2011, **2011**:234519.

D'Amico AV, Whittington R, Malkowicz SB, Schultz D, Blank K *et al*: **Biochemical outcome after radical prostatectomy, external beam radiation therapy, or interstitial radiation therapy for clinically localized prostate cancer**. *JAMA* 1998, **280**(11):969-974.

Daftary GS, Taylor HS: **Endocrine regulation of HOX genes**. *Endocr Rev* 2006, **27**(4):331-355.

Dahl A, Mertes F, Marchfelder U, Boerno ST, Fischer A *et al*: **Bewertung der SOLiD-Sequenzierungsplattform aus Nutzerperspektive**. *Laborwelt* 2010, **1/2010**:8-10.

Damber JE, Aus G: **Prostate cancer**. *Lancet* 2008, **371**(9625):1710-1721.

Daskalos A, Nikolaidis G, Xinarianos G, Savvari P, Cassidy A *et al*: **Hypomethylation of retrotransposable elements correlates with genomic instability in non-small cell lung cancer**. *Int J Cancer* 2009, **124**(1):81-87.

Daskivich TJ, Chamie K, Kwan L, Labo J, Palvolgyi R *et al*: **Overtreatment of men with low-risk prostate cancer and significant comorbidity**. *Cancer* 2011, **117**(10):2058-2066.

de Souza PL, Russell PJ, Kearsley JH, Howes LG: **Clinical pharmacology of isoflavones and its relevance for potential prevention of prostate cancer**. *Nutr Rev* 2010, **68**(9):542-555.

de The H, Vivanco-Ruiz MM, Tiollais P, Stunnenberg H, Dejean A: **Identification of a retinoic acid responsive element in the retinoic acid receptor beta gene**. *Nature* 1990, **343**(6254):177-180.

Delahunt B, Miller RJ, Srigley JR, Evans AJ, Samaratunga H: **Gleason grading: past, present and future**. *Histopathology* 2012, **60**(1):75-86.

Demichelis F, Fall K, Perner S, Andren O, Schmidt F *et al*: **TMPRSS2:ERG gene fusion associated with lethal prostate cancer in a watchful waiting cohort**. *Oncogene* 2007, **26**(31):4596-4599.

Dobrovic A, Simpfendorfer D: **Methylation of the BRCA1 gene in sporadic breast cancer**. *Cancer Res* 1997, **57**(16):3347-3350.

Down TA, Rakyan VK, Turner DJ, Flicek P, Li H *et al*: **A Bayesian deconvolution strategy for immunoprecipitation-based DNA methylome analysis**. *Nat Biotechnol* 2008, **26**(7):779-785.

Duester G: **Retinoic acid synthesis and signaling during early organogenesis**. *Cell* 2008, **134**(6):921-931.

Dutt SS, Gao AC: **Molecular mechanisms of castration-resistant prostate cancer progression**. *Future Oncol* 2009, **5**(9):1403-1413.

Eckersberger E, Finkelstein J, Sadri H, Margreiter M, Taneja SS *et al*: **Screening for Prostate Cancer: A Review of the ERSPC and PLCO Trials**. *Rev Urol* 2009, **11**(3):127-133.
Eckhardt F, Lewin J, Cortese R, Rakyan VK, Attwood J *et al*: **DNA methylation profiling of human chromosomes 6, 20 and 22**. *Nat Genet* 2006, **38**(12):1378-1385.
Edge SB, American Joint Committee on Cancer.: **AJCC cancer staging manual**, 7th edn. New York: Springer; 2010.
Egger G, Liang G, Aparicio A, Jones PA: **Epigenetics in human disease and prospects for epigenetic therapy**. *Nature* 2004, **429**(6990):457-463.
Ehrich M, Zoll S, Sur S, van den Boom D: **A new method for accurate assessment of DNA quality after bisulfite treatment**. *Nucleic Acids Res* 2007, **35**(5):e29.
Ehrich M, Turner J, Gibbs P, Lipton L, Giovanneti M *et al*: **Cytosine methylation profiling of cancer cell lines**. *Proc Natl Acad Sci U S A* 2008, **105**(12):4844-4849.
Ehrlich M, Gama-Sosa MA, Huang LH, Midgett RM, Kuo KC *et al*: **Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells**. *Nucleic Acids Res* 1982, **10**(8):2709-2721.
Ehrlich M: **DNA hypomethylation in cancer cells**. *Epigenomics* 2009, **1**(2):239-259.
Elsby R, Kitteringham NR, Goldring CE, Lovatt CA, Chamberlain M *et al*: **Increased constitutive c-Jun N-terminal kinase signaling in mice lacking glutathione S-transferase Pi**. *J Biol Chem* 2003, **278**(25):22243-22249.
Etcheverry A, Aubry M, de Tayrac M, Vauleon E, Boniface R *et al*: **DNA methylation in glioblastoma: impact on gene expression and clinical outcome**. *BMC Genomics* 2010, **11**:701.
Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N *et al*: **MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B**. *Proc Natl Acad Sci U S A* 2007, **104**(40):15805-15810.
Fang MZ, Chen D, Sun Y, Jin Z, Christman JK *et al*: **Reversal of hypermethylation and reactivation of p16INK4a, RARbeta, and MGMT genes by genistein and other isoflavones from soy**. *Clin Cancer Res* 2005, **11**(19 Pt 1):7033-7041.
Fatemi M, Pao MM, Jeong S, Gal-Yam EN, Egger G *et al*: **Footprinting of mammalian promoters: use of a CpG DNA methyltransferase revealing nucleosome positions at a single molecule level**. *Nucleic Acids Res* 2005, **33**(20):e176.
Feber A, Wilson G, Zhang L, Presneau N, Idowu B *et al*: **Comparative methylome analysis of benign and malignant peripheral nerve sheath tumours**. *Genome Res* 2011.
Feinberg AP, Vogelstein B: **Hypomethylation distinguishes genes of some human cancers from their normal counterparts**. *Nature* 1983, **301**(5895):89-92.
Feinberg AP, Ohlsson R, Henikoff S: **The epigenetic progenitor origin of human cancer**. *Nat Rev Genet* 2006, **7**(1):21-33.
Flatau E, Bogenmann E, Jones PA: **Variable 5-methylcytosine levels in human tumor cell lines and fresh pediatric tumor explants**. *Cancer Res* 1983, **43**(10):4901-4905.
Fujimura T, Takahashi S, Urano T, Kumagai J, Murata T *et al*: **Expression of cytochrome P450 3A4 and its clinical significance in human prostate cancer**. *Urology* 2009, **74**(2):391-397.
Futreal PA, Coin L, Marshall M, Down T, Hubbard T *et al*: **A census of human cancer genes**. *Nat Rev Cancer* 2004, **4**(3):177-183.
Gal-Yam EN, Egger G, Iniguez L, Holster H, Einarsson S *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):12979-12984.
Gama-Sosa MA, Wang RY, Kuo KC, Gehrke CW, Ehrlich M: **The 5-methylcytosine content of highly repeated sequences in human DNA**. *Nucleic Acids Res* 1983, **11**(10):3087-3095.
Gardiner-Garden M, Frommer M: **CpG islands in vertebrate genomes**. *J Mol Biol* 1987, **196**(2):261-282.
Gaudet F, Hodgson JG, Eden A, Jackson-Grusby L, Dausman J *et al*: **Induction of tumors in mice by genomic hypomethylation**. *Science* 2003, **300**(5618):489-492.
Gleason DF: **Histologic grading of prostate cancer: a perspective**. *Hum Pathol* 1992, **23**(3):273-279.
Goessl C, Muller M, Heicappell R, Krause H, Straub B *et al*: **DNA-based detection of prostate cancer in urine after prostatic massage**. *Urology* 2001, **58**(3):335-338.
Greger V, Passarge E, Hopping W, Messmer E, Horsthemke B: **Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma**. *Hum Genet* 1989, **83**(2):155-158.
Haffner MC, Aryee MJ, Toubaji A, Esopi DM, Albadine R *et al*: **Androgen-induced TOP2B-mediated double-strand breaks and prostate cancer gene rearrangements**. *Nat Genet* 2010, **42**(8):668-675.
Halvorson M, Greenway D, Eberhart D, Fitzgerald K, Parkinson A: **Reconstitution of testosterone oxidation by purified rat cytochrome P450p (IIIA1)**. *Arch Biochem Biophys* 1990, **277**(1):166-180.

Hammerich GEA, and Thomas M. Wheeler: **Anatomy of the prostate gland and surgical pathology of prostate cancer** In: *Prostate Cancer* Edited by Scardino HHaP, 2009 edn: Cambridge University Press 2008.

Han L, Witmer PD, Casey E, Valle D, Sukumar S: **DNA methylation regulates MicroRNA expression**. *Cancer Biology & Therapy* 2007, **6**(8):1284-1288.

Hanahan D, Weinberg RA: **Hallmarks of cancer: the next generation**. *Cell* 2011, **144**(5):646-674.

Harris RA, Wang T, Coarfa C, Nagarajan RP, Hong C *et al*: **Comparison of sequencing-based methods to profile DNA methylation and identification of monoallelic epigenetic modifications**. *Nat Biotechnol* 2010, **28**(10):1097-1105.

Hayatsu H, Shiraishi M, Negishi K: **Bisulfite modification for analysis of DNA methylation**. *Curr Protoc Nucleic Acid Chem* 2008, **Chapter 6**:Unit 6 10.

Heidenreich A, Bellmunt J, Bolla M, Joniau S, Mason M *et al*: **EAU Guidelines on Prostate Cancer. Part 1: Screening, Diagnosis, and Treatment of Clinically Localised Disease**. *Eur Urol* 2010.

Herman JG, Umar A, Polyak K, Graff JR, Ahuja N *et al*: **Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma**. *Proc Natl Acad Sci U S A* 1998, **95**(12):6870-6875.

Hermans KG, van Marion R, van Dekken H, Jenster G, van Weerden WM *et al*: **TMPRSS2:ERG fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer**. *Cancer Res* 2006, **66**(22):10658-10663.

Hibi K, Sakata M, Sakuraba K, Shirahata A, Goto T *et al*: **Aberrant methylation of the HACE1 gene is frequently detected in advanced colorectal cancer**. *Anticancer Res* 2008, **28**(3A):1581-1584.

Hill VK, Hesson LB, Dansranjavin T, Dallol A, Bieche I *et al*: **Identification of 5 novel genes methylated in breast and other epithelial cancers**. *Mol Cancer* 2010, **9**:51.

Hiltunen MO, Alhonen L, Koistinaho J, Myohanen S, Paakkonen M *et al*: **Hypermethylation of the APC (adenomatous polyposis coli) gene promoter region in human colorectal carcinoma**. *Int J Cancer* 1997, **70**(6):644-648.

Hobert O, Jallal B, Ullrich A: **Interaction of Vav with ENX-1, a putative transcriptional regulator of homeobox gene expression**. *Mol Cell Biol* 1996, **16**(6):3066-3073.

Hoffmann B, Lehmann JM, Zhang XK, Hermann T, Husmann M *et al*: **A retinoic acid receptor-specific element controls the retinoic acid receptor-beta promoter**. *Mol Endocrinol* 1990, **4**(11):1727-1736.

Hoffmann MJ, Engers R, Florl AR, Otte AP, Muller M *et al*: **Expression changes in EZH2, but not in BMI-1, SIRT1, DNMT1 or DNMT3B are associated with DNA methylation changes in prostate cancer**. *Cancer Biol Ther* 2007, **6**(9):1403-1412.

Holland PW, Booth HA, Bruford EA: **Classification and nomenclature of all human homeobox genes**. *BMC Biol* 2007, **5**:47.

Hopkins TG, Burns PA, Routledge MN: **DNA methylation of GSTP1 as biomarker in diagnosis of prostate cancer**. *Urology* 2007, **69**(1):11-16.

Hoque MO, Topaloglu O, Begum S, Henrique R, Rosenbaum E *et al*: **Quantitative methylation-specific polymerase chain reaction gene patterns in urine sediment distinguish prostate cancer patients from control subjects**. *J Clin Oncol* 2005, **23**(27):6569-6575.

Hori S, Butler E, McLoughlin J: **Prostate cancer and diet: food for thought?** *BJU Int* 2011, **107**(9):1348-1359.

Howard G, Eiges R, Gaudet F, Jaenisch R, Eden A: **Activation and transposition of endogenous retroviral elements in hypomethylation induced tumors in mice**. *Oncogene* 2008, **27**(3):404-408.

Huang DY, Ichikawa Y: **Two different enzymes are primarily responsible for retinoic acid synthesis in rabbit liver cytosol**. *Biochem Biophys Res Commun* 1994, **205**(2):1278-1283.

Ideraabdullah FY, Vigneau S, Bartolomei MS: **Genomic imprinting mechanisms in mammals**. *Mutat Res* 2008, **647**(1-2):77-85.

Inohara N, Ding L, Chen S, Nunez G: **harakiri, a novel regulator of cell death, encodes a protein that activates apoptosis and interacts selectively with survival-promoting proteins Bcl-2 and Bcl-X(L)**. *EMBO J* 1997, **16**(7):1686-1694.

Irizarry RA, Ladd-Acosta C, Wen B, Wu Z, Montano C *et al*: **The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores**. *Nat Genet* 2009, **41**(2):178-186.

Jaenisch R, Bird A: **Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals**. *Nat Genet* 2003, **33 Suppl**:245-254.

Jason LJ, Moore SC, Lewis JD, Lindsey G, Ausio J: **Histone ubiquitination: a tagging tail unfolds?** *Bioessays* 2002, **24**(2):166-174.

Jemal A, Bray F, Center MM, Ferlay J, Ward E *et al*: **Global cancer statistics**. *CA Cancer J Clin* 2011, **61**(2):69-90.

Jenuwein T, Allis CD: **Translating the histone code**. *Science* 2001, **293**(5532):1074-1080.
Jhavar S, Reid A, Clark J, Kote-Jarai Z, Christmas T *et al*: **Detection of TMPRSS2-ERG translocations in human prostate cancer by expression profiling using GeneChip Human Exon 1.0 ST arrays**. *J Mol Diagn* 2008, **10**(1):50-57.
Jones PA: **The DNA methylation paradox**. *Trends Genet* 1999, **15**(1):34-37.
Jones PA, Baylin SB: **The fundamental role of epigenetic events in cancer**. *Nat Rev Genet* 2002, **3**(6):415-428.
Juan AH, Kumar RM, Marx JG, Young RA, Sartorelli V: **Mir-214-dependent regulation of the polycomb protein Ezh2 in skeletal muscle and embryonic stem cells**. *Mol Cell* 2009, **36**(1):61-74.
Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D *et al*: **Diverse somatic mutation patterns and pathway alterations in human cancers**. *Nature* 2010, **466**(7308):869-873.
Karanikolas BD, Figueiredo ML, Wu L: **Comprehensive evaluation of the role of EZH2 in the growth, invasion, and aggression of a panel of prostate cancer cell lines**. *Prostate* 2010, **70**(6):675-688.
Keshet I, Schlesinger Y, Farkash S, Rand E, Hecht M *et al*: **Evidence for an instructive mechanism of de novo methylation in cancer cells**. *Nat Genet* 2006, **38**(2):149-153.
Kilpelainen TP, Tammela TL, Maattanen L, Kujala P, Stenman UH *et al*: **False-positive screening results in the Finnish prostate cancer screening trial**. *Br J Cancer* 2010, **102**(3):469-474.
Kim JH, Dhanasekaran SM, Prensner JR, Cao X, Robinson D *et al*: **Deep sequencing reveals distinct patterns of DNA methylation in prostate cancer**. *Genome Res* 2011, **21**(7):1028-1041.
Kisliouk T, Yosefi S, Meiri N: **MiR-138 inhibits EZH2 methyltransferase expression and methylation of histone H3 at lysine 27, and affects thermotolerance acquisition**. *Eur J Neurosci* 2011, **33**(2):224-235.
Klezovitch O, Risk M, Coleman I, Lucas JM, Null M *et al*: **A causal role for ERG in neoplastic transformation of prostate epithelium**. *Proc Natl Acad Sci U S A* 2008, **105**(6):2105-2110.
Koeneman KS, Pan CX, Jin JK, Pyle JM, 3rd, Flanigan RC *et al*: **Telomerase activity, telomere length, and DNA ploidy in prostatic intraepithelial neoplasia (PIN)**. *J Urol* 1998, **160**(4):1533-1539.
Krawitz PM, Schweiger MR, Rodelsperger C, Marcelis C, Kolsch U *et al*: **Identity-by-descent filtering of exome sequence data identifies PIGV mutations in hyperphosphatasia mental retardation syndrome**. *Nat Genet* 2010, **42**(10):827-829.
Kroepfl D, Loewen H, Roggenbuck U, Musch M, Klevecka V: **Disease progression and survival in patients with prostate carcinoma and positive lymph nodes after radical retropubic prostatectomy**. *BJU Int* 2006, **97**(5):985-991.
Kron K, Pethe V, Briollais L, Sadikovic B, Ozcelik H *et al*: **Discovery of novel hypermethylated genes in prostate cancer using genomic CpG island microarrays**. *PLoS One* 2009, **4**(3):e4830.
Kunderfranco P, Mello-Grand M, Cangemi R, Pellini S, Mensah A *et al*: **ETS transcription factors control transcription of EZH2 and epigenetic silencing of the tumor suppressor gene Nkx3.1 in prostate cancer**. *PLoS One* 2010, **5**(5):e10547.
Kuzmichev A, Jenuwein T, Tempst P, Reinberg D: **Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3**. *Mol Cell* 2004, **14**(2):183-193.
Laborde E: **Glutathione transferases as mediators of signaling pathways involved in cell proliferation and cell death**. *Cell Death Differ* 2010, **17**(9):1373-1380.
Laxman B, Tomlins SA, Mehra R, Morris DS, Wang L *et al*: **Noninvasive detection of TMPRSS2:ERG fusion transcripts in the urine of men with prostate cancer**. *Neoplasia* 2006, **8**(10):885-888.
Lee LF, Guan J, Qiu Y, Kung HJ: **Neuropeptide-induced androgen independence in prostate cancer cells: roles of nonreceptor tyrosine kinases Etk/Bmx, Src, and focal adhesion kinase**. *Mol Cell Biol* 2001, **21**(24):8385-8397.
Lee WH, Morton RA, Epstein JI, Brooks JD, Campbell PA *et al*: **Cytidine methylation of regulatory sequences near the pi-class glutathione S-transferase gene accompanies human prostatic carcinogenesis**. *Proc Natl Acad Sci U S A* 1994, **91**(24):11733-11737.
Lee YS, Dutta A: **MicroRNAs in cancer**. *Annu Rev Pathol* 2009, **4**:199-227.
Lei H, Oh SP, Okano M, Juttermann R, Goss KA *et al*: **De novo DNA cytosine methyltransferase activities in mouse embryonic stem cells**. *Development* 1996, **122**(10):3195-3205.
Leonhardt H, Page AW, Weier HU, Bestor TH: **A targeting sequence directs DNA methyltransferase to sites of DNA replication in mammalian nuclei**. *Cell* 1992, **71**(5):865-873.
Li E, Bestor TH, Jaenisch R: **Targeted mutation of the DNA methyltransferase gene results in embryonic lethality**. *Cell* 1992, **69**(6):915-926.
Li LC, Okino ST, Dahiya R: **DNA methylation in prostate cancer**. *Biochim Biophys Acta* 2004, **1704**(2):87-102.
Li N, Ye M, Li Y, Yan Z, Butcher LM *et al*: **Whole genome DNA methylation analysis based on high throughput sequencing technology**. *Methods* 2010, **52**(3):203-212.
Lin C, Yang L, Tanasa B, Hutt K, Ju BG *et al*: **Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer**. *Cell* 2009, **139**(6):1069-1083.

Lin SY, Hsieh SC, Lin YC, Lee CN, Tsai MH *et al*: **A whole genome methylation analysis of systemic lupus erythematosus: hypomethylation of the IL10 and IL1R2 promoters is associated with disease activity**. *Genes Immun* 2012, **13**(3):214-220.

Lister R, Ecker JR: **Finding the fifth base: genome-wide sequencing of cytosine methylation**. *Genome Res* 2009a, **19**(6):959-966.

Lister R, Pelizzola M, Dowen RH, Hawkins RD, Hon G *et al*: **Human DNA methylomes at base resolution show widespread epigenomic differences**. *Nature* 2009b, **462**(7271):315-322.

Liu C, Teng ZQ, Santistevan NJ, Szulwach KE, Guo W *et al*: **Epigenetic regulation of miR-184 by MBD1 governs neural stem cell proliferation and differentiation**. *Cell Stem Cell* 2010a, **6**(5):433-444.

Liu M, Tang Q, Qiu M, Lang N, Li M *et al*: **miR-21 targets the tumor suppressor RhoB and regulates proliferation, invasion and apoptosis in colorectal cancer cells**. *FEBS Lett* 2011a, **585**(19):2998-3005.

Liu W, Ewing CM, Chang BL, Li T, Sun J *et al*: **Multiple genomic alterations on 21q22 predict various TMPRSS2/ERG fusion transcripts in human prostate cancers**. *Genes Chromosomes Cancer* 2007, **46**(11):972-980.

Liu W, Zabirnyk O, Wang H, Shiao YH, Nickerson ML *et al*: **miR-23b targets proline oxidase, a novel tumor suppressor protein in renal cancer**. *Oncogene* 2010b, **29**(35):4914-4924.

Liu Y, Hu F, Li D, Wang F, Zhu L *et al*: **Does physical activity reduce the risk of prostate cancer? A systematic review and meta-analysis**. *Eur Urol* 2011b, **60**(5):1029-1044.

Long C, Yin B, Lu Q, Zhou X, Hu J *et al*: **Promoter hypermethylation of the RUNX3 gene in esophageal squamous cell carcinoma**. *Cancer Invest* 2007, **25**(8):685-690.

Lu J, He ML, Wang L, Chen Y, Liu X *et al*: **MiR-26a inhibits cell growth and tumorigenesis of nasopharyngeal carcinoma through repression of EZH2**. *Cancer Res* 2011, **71**(1):225-233.

Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C *et al*: **Genetic unmasking of an epigenetically silenced microRNA in human cancer cells**. *Cancer Res* 2007, **67**(4):1424-1429.

Marshall H, Morrison A, Studer M, Popperl H, Krumlauf R: **Retinoids and Hox genes**. *FASEB J* 1996, **10**(9):969-978.

Mehra R, Tomlins SA, Shen R, Nadeem O, Wang L *et al*: **Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer**. *Mod Pathol* 2007, **20**(5):538-544.

Meissner A, Gnirke A, Bell GW, Ramsahoye B, Lander ES *et al*: **Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis**. *Nucleic Acids Res* 2005, **33**(18):5868-5877.

Mertz KD, Setlur SR, Dhanasekaran SM, Demichelis F, Perner S *et al*: **Molecular characterization of TMPRSS2-ERG gene fusion in the NCI-H660 prostate cancer cell line: a new perspective for an old model**. *Neoplasia* 2007, **9**(3):200-206.

Mochmann LH, Bock J, Ortiz-Tanchez J, Schlee C, Bohne A *et al*: **Genome-wide screen reveals WNT11, a non-canonical WNT gene, as a direct target of ETS transcription factor ERG**. *Oncogene* 2011, **30**(17):2044-2056.

Montironi R, Mazzucchelli R, Lopez-Beltran A, Cheng L, Scarpelli M: **Mechanisms of disease: high-grade prostatic intraepithelial neoplasia and other proposed preneoplastic lesions in the prostate**. *Nat Clin Pract Urol* 2007, **4**(6):321-332.

Morahan JM, Yu B, Trent RJ, Pamphlett R: **A genome-wide analysis of brain DNA methylation identifies new candidate genes for sporadic amyotrophic lateral sclerosis**. *Amyotroph Lateral Sc* 2009, **10**(5-6):418-U247.

Morris MR, Ricketts CJ, Gentle D, McRonald F, Carli N *et al*: **Genome-wide methylation analysis identifies epigenetically inactivated candidate tumour suppressor genes in renal cell carcinoma**. *Oncogene* 2011, **30**(12):1390-1401.

Nakayama M, Bennett CJ, Hicks JL, Epstein JI, Platz EA *et al*: **Hypermethylation of the human glutathione S-transferase-pi gene (GSTP1) CpG island is present in a subset of proliferative inflammatory atrophy lesions but not in normal or hyperplastic epithelium of the prostate: a detailed study using laser-capture microdissection**. *Am J Pathol* 2003, **163**(3):923-933.

Nakayama T, Watanabe M, Yamanaka M, Hirokawa Y, Suzuki H *et al*: **The role of epigenetic modifications in retinoic acid receptor beta2 gene expression in human prostate cancers**. *Lab Invest* 2001, **81**(7):1049-1057.

Nelson WG, De Marzo AM, Yegnasubramanian S: **Epigenetic alterations in human prostate cancers**. *Endocrinology* 2009, **150**(9):3991-4002.

Neuhaus EM, Zhang W, Gelis L, Deng Y, Noldus J *et al*: **Activation of an olfactory receptor inhibits proliferation of prostate cancer cells**. *J Biol Chem* 2009, **284**(24):16218-16225.

Noy N: **Between death and survival: retinoic acid in regulation of apoptosis**. *Annu Rev Nutr* 2010, **30**:201-217.

Oesterling JE: **Prostate specific antigen: a critical assessment of the most useful tumor marker for adenocarcinoma of the prostate**. *J Urol* 1991, **145**(5):907-923.

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):247-257.

Otterson GA, Khleif SN, Chen W, Coxon AB, Kaye FJ: **CDKN2 gene silencing in lung cancer by DNA hypermethylation and kinetics of p16INK4 protein induction by 5-aza 2'deoxycytidine**. *Oncogene* 1995, **11**(6):1211-1216.

Parkin DM, Bray F, Ferlay J, Pisani P: **Global cancer statistics, 2002**. *CA Cancer J Clin* 2005, **55**(2):74-108.

Pastor WA, Pape UJ, Huang Y, Henderson HR, Lister R *et al*: **Genome-wide mapping of 5-hydroxymethylcytosine in embryonic stem cells**. *Nature* 2011, **473**(7347):394-397.

Payne SR, Serth J, Schostak M, Kamradt J, Strauss A *et al*: **DNA methylation biomarkers of prostate cancer: confirmation of candidates and evidence urine is the most sensitive body fluid for non-invasive detection**. *Prostate* 2009, **69**(12):1257-1269.

Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM *et al*: **TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer**. *Cancer Res* 2006, **66**(17):8337-8341.

Perry AS, Watson RWG, Lawler M, Hollywood D: **The epigenome as a therapeutic target in prostate cancer**. *Nat Rev Urol* 2010, **7**(12):668-680.

Pihlajamaa P, Zhang FP, Saarinen L, Mikkonen L, Hautaniemi S *et al*: **The phytoestrogen genistein is a tissue-specific androgen receptor modulator**. *Endocrinology* 2011, **152**(11):4395-4405.

Pirrotta V: **Polycombing the genome: PcG, trxG, and chromatin silencing**. *Cell* 1998, **93**(3):333-336.

Plaskon LA, Penson DF, Vaughan TL, Stanford JL: **Cigarette smoking and risk of prostate cancer in middle-aged men**. *Cancer Epidemiol Biomarkers Prev* 2003, **12**(7):604-609.

Qian J, Bostwick DG, Takahashi S, Borell TJ, Herath JF *et al*: **Chromosomal anomalies in prostatic intraepithelial neoplasia and carcinoma detected by fluorescence in situ hybridization**. *Cancer Res* 1995, **55**(22):5408-5414.

R_Development_Core_Team: **R: A language and environment for statistical computing.** . Vienna, Austria: R Foundation for Statistical Computing; 2009.

Radpour R, Haghighi MM, Fan AX, Torbati PM, Hahn S *et al*: **High-throughput hacking of the methylation patterns in breast cancer by in vitro transcription and thymidine-specific cleavage mass array on MALDI-TOF silico-chip**. *Mol Cancer Res* 2008, **6**(11):1702-1709.

Rauch T, Wang Z, Zhang X, Zhong X, Wu X *et al*: **Homeobox gene methylation in lung cancer studied by genome-wide analysis with a microarray-based methylated CpG island recovery assay**. *Proc Natl Acad Sci U S A* 2007, **104**(13):5527-5532.

Raynal NJ, Si J, Taby RF, Gharibyan V, Ahmed S *et al*: **DNA methylation does not stably lock gene expression but instead serves as a molecular mark for gene silencing memory**. *Cancer Res* 2012, **72**(5):1170-1181.

Riggs AD, Pfeifer GP: **X-chromosome inactivation and cell memory**. *Trends Genet* 1992, **8**(5):169-174.

Rodriguez J, Frigola J, Vendrell E, Risques RA, Fraga MF *et al*: **Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers**. *Cancer Res* 2006, **66**(17):8462-9468.

Roupret M, Hupertan V, Yates DR, Catto JW, Rehman I *et al*: **Molecular detection of localized prostate cancer using quantitative methylation-specific PCR on urinary cells obtained following prostate massage**. *Clin Cancer Res* 2007, **13**(6):1720-1725.

Rubin MA, Maher CA, Chinnaiyan AM: **Common gene rearrangements in prostate cancer**. *J Clin Oncol* 2011, **29**(27):3659-3668.

Sakr WA, Grignon DJ, Crissman JD, Heilbrun LK, Cassin BJ *et al*: **High grade prostatic intraepithelial neoplasia (HGPIN) and prostatic adenocarcinoma between the ages of 20-69: an autopsy study of 249 cases**. *In Vivo* 1994, **8**(3):439-443.

Samlowski WE, Leachman SA, Wade M, Cassidy P, Porter-Gill P *et al*: **Evaluation of a 7-day continuous intravenous infusion of decitabine: inhibition of promoter-specific and global genomic DNA methylation**. *J Clin Oncol* 2005, **23**(17):3897-3905.

Sander S, Bullinger L, Klapproth K, Fiedler K, Kestler HA *et al*: **MYC stimulates EZH2 expression by repression of its negative regulator miR-26a**. *Blood* 2008, **112**(10):4202-4212.

Sboner A, Demichelis F, Calza S, Pawitan Y, Setlur SR *et al*: **Molecular sampling of prostate cancer: a dilemma for predicting disease progression**. *BMC Med Genomics* 2010, **3**:8.

Schlesinger Y, Straussman R, Keshet I, Farkash S, Hecht M *et al*: **Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer**. *Nat Genet* 2007, **39**(2):232-236.

Schmidt H, DeAngelis G, Eltze E, Gockel I, Semjonow A *et al*: **Asynchronous growth of prostate cancer is reflected by circulating tumor cells delivered from distinct, even small foci, harboring loss of heterozygosity of the PTEN gene**. *Cancer Res* 2006, **66**(18):8959-8965.

Schroder FH, Hugosson J, Roobol MJ, Tammela TL, Ciatto S *et al*: **Screening and prostate-cancer mortality in a randomized European study**. *N Engl J Med* 2009, **360**(13):1320-1328.

Schulz WA, Hatina J: **Epigenetics of prostate cancer: beyond DNA methylation**. *J Cell Mol Med* 2006, **10**(1):100-125.

Schulz WA, Hoffmann MJ: **Epigenetic mechanisms in the biology of prostate cancer**. *Semin Cancer Biol* 2009, **19**(3):172-180.

Scott MP, Tamkun JW, Hartzell GW, 3rd: **The structure and function of the homeodomain**. *Biochim Biophys Acta* 1989, **989**(1):25-48.

Setlur SR, Mertz KD, Hoshida Y, Demichelis F, Lupien M *et al*: **Estrogen-dependent signaling in a molecularly distinct subclass of aggressive prostate cancer**. *J Natl Cancer Inst* 2008, **100**(11):815-825.

Shariat SF, Semjonow A, Lilja H, Savage C, Vickers AJ *et al*: **Tumor markers in prostate cancer I: blood-based markers**. *Acta Oncol* 2011, **50 Suppl 1**:61-75.

Sharma S, Kelly TK, Jones PA: **Epigenetics in cancer**. *Carcinogenesis* 2009, **31**(1):27-36.

Shi XB, Xue L, Yang J, Ma AH, Zhao J *et al*: **An androgen-regulated miRNA suppresses Bak1 expression and induces androgen-independent growth of prostate cancer cells**. *Proc Natl Acad Sci U S A* 2007, **104**(50):19983-19988.

Shibata A, Whittemore AS: **Genetic predisposition to prostate cancer: possible explanations for ethnic differences in risk**. *Prostate* 1997, **32**(1):65-72.

Shiio Y, Eisenman RN: **Histone sumoylation is associated with transcriptional repression**. *Proc Natl Acad Sci U S A* 2003, **100**(23):13225-13230.

Shu Y, Wang B, Wang J, Wang JM, Zou SQ: **Identification of methylation profile of HOX genes in extrahepatic cholangiocarcinoma**. *World J Gastroentero* 2011, **17**(29):3407-3419.

Sigruener A, Buechler C, Orso E, Hartmann A, Wild PJ *et al*: **Human aldehyde oxidase 1 interacts with ATP-binding cassette transporter-1 and modulates its activity in hepatocytes**. *Horm Metab Res* 2007, **39**(11):781-789.

Sircar K, Yoshimoto M, Monzon FA, Koumakpayi IH, Katz RL *et al*: **PTEN genomic deletion is associated with p-Akt and AR signalling in poorer outcome, hormone refractory prostate cancer**. *J Pathol* 2009, **218**(4):505-513.

Smith SS, Kaplan BE, Sowers LC, Newman EM: **Mechanism of human methyl-directed DNA methyltransferase and the fidelity of cytosine methylation**. *Proc Natl Acad Sci U S A* 1992, **89**(10):4744-4748.

Sonoda T, Nagata Y, Mori M, Miyanaga N, Takashima N *et al*: **A case-control study of diet and prostate cancer in Japan: possible protective effect of traditional Japanese diet**. *Cancer Sci* 2004, **95**(3):238-242.

Soos G, Tsakiris I, Szanto J, Turzo C, Haas PG *et al*: **The prevalence of prostate carcinoma and its precursor in Hungary: an autopsy study**. *Eur Urol* 2005, **48**(5):739-744.

Sproul D, Nestor C, Culley J, Dickson JH, Dixon JM *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):4364-4369.

Stamey TA, Yang N, Hay AR, McNeal JE, Freiha FS *et al*: **Prostate-specific antigen as a serum marker for adenocarcinoma of the prostate**. *N Engl J Med* 1987, **317**(15):909-916.

Suh SO, Chen Y, Zaman MS, Hirata H, Yamamura S *et al*: **MicroRNA-145 is regulated by DNA methylation and p53 gene mutation in prostate cancer**. *Carcinogenesis* 2011, **32**(5):772-778.

Sun C, Dobi A, Mohamed A, Li H, Thangapazham RL *et al*: **TMPRSS2-ERG fusion, a common genomic alteration in prostate cancer activates C-MYC and abrogates prostate epithelial differentiation**. *Oncogene* 2008, **27**(40):5348-5353.

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H *et al*: **Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1**. *Science* 2009, **324**(5929):930-935.

Tate PH, Bird AP: **Effects of DNA methylation on DNA-binding proteins and gene expression**. *Curr Opin Genet Dev* 1993, **3**(2):226-231.

Taylor BS, Schultz N, Hieronymus H, Gopalan A, Xiao Y *et al*: **Integrative genomic profiling of human prostate cancer**. *Cancer Cell* 2010, **18**(1):11-22.

Testa JR, Bellacosa A: **AKT plays a central role in tumorigenesis**. *Proc Natl Acad Sci U S A* 2001, **98**(20):10983-10985.

Thompson IM, Pauler DK, Goodman PJ, Tangen CM, Lucia MS *et al*: **Prevalence of prostate cancer among men with a prostate-specific antigen level < or =4.0 ng per milliliter**. *N Engl J Med* 2004, **350**(22):2239-2246.

Thompson IM, Ankerst DP, Chi C, Lucia MS, Goodman PJ *et al*: **Operating characteristics of prostate-specific antigen in men with an initial PSA level of 3.0 ng/ml or lower**. *JAMA* 2005, **294**(1):66-70.

Timmermann B, Kerick M, Roehr C, Fischer A, Isau M *et al*: **Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis**. *PLoS One* 2010, **5**(12):e15661.

Tivnan A, Foley NH, Tracey L, Davidoff AM, Stallings RL: **MicroRNA-184-mediated inhibition of tumour growth in an orthotopic murine model of neuroblastoma**. *Anticancer Res* 2010, **30**(11):4391-4395.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R *et al*: **Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer**. *Science* 2005, **310**(5748):644-648.

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 Res* 2009, **11**(1):R14.

Tosoian J, Loeb S: **PSA and beyond: the past, present, and future of investigative biomarkers for prostate cancer**. *ScientificWorldJournal* 2010, **10**:1919-1931.

Tost J, Gut IG: **DNA methylation analysis by pyrosequencing**. *Nat Protoc* 2007, **2**(9):2265-2275.

Trapman J, Cleutjens KB: **Androgen-regulated gene expression in prostate cancer**. *Semin Cancer Biol* 1997, **8**(1):29-36.

Tzatsos A, Paskaleva P, Lymperi S, Contino G, Stoykova S *et al*: **A Lysine (K)-specific demethylase 2B (KDM2B)-let-7-Enhancer of Zester Homolog 2 (EZH2) pathway regulates cell cycle progression and senescence in primary cells**. *J Biol Chem* 2011.

Valk-Lingbeek ME, Bruggeman SW, van Lohuizen M: **Stem cells and cancer; the polycomb connection**. *Cell* 2004, **118**(4):409-418.

van Leenders GJ, Dukers D, Hessels D, van den Kieboom SW, Hulsbergen CA *et al*: **Polycomb-group oncogenes EZH2, BMI1, and RING1 are overexpressed in prostate cancer with adverse pathologic and clinical features**. *Eur Urol* 2007, **52**(2):455-463.

Varambally S, Dhanasekaran SM, Zhou M, Barrette TR, Kumar-Sinha C *et al*: **The polycomb group protein EZH2 is involved in progression of prostate cancer**. *Nature* 2002, **419**(6907):624-629.

Varambally S, Cao Q, Mani RS, Shankar S, Wang X *et al*: **Genomic loss of microRNA-101 leads to overexpression of histone methyltransferase EZH2 in cancer**. *Science* 2008, **322**(5908):1695-1699.

Villar-Garea A, Fraga MF, Espada J, Esteller M: **Procaine is a DNA-demethylating agent with growth-inhibitory effects in human cancer cells**. *Cancer Res* 2003, **63**(16):4984-4989.

Vire E, Brenner C, Deplus R, Blanchon L, Fraga M *et al*: **The Polycomb group protein EZH2 directly controls DNA methylation**. *Nature* 2006, **439**(7078):871-874.

Wallace EV, Stoddart D, Heron AJ, Mikhailova E, Maglia G *et al*: **Identification of epigenetic DNA modifications with a protein nanopore**. *Chem Commun (Camb)* 2010, **46**(43):8195-8197.

Walsh CP, Chaillet JR, Bestor TH: **Transcription of IAP endogenous retroviruses is constrained by cytosine methylation**. *Nat Genet* 1998, **20**(2):116-117.

Wang MC, Valenzuela LA, Murphy GP, Chu TM: **Purification of a human prostate specific antigen**. *Invest Urol* 1979, **17**(2):159-163.

Wang P, Yu J, Zhang L: **The nuclear function of p53 is required for PUMA-mediated apoptosis induced by DNA damage**. *Proc Natl Acad Sci U S A* 2007, **104**(10):4054-4059.

Watt F, Molloy PL: **Cytosine methylation prevents binding to DNA of a HeLa cell transcription factor required for optimal expression of the adenovirus major late promoter**. *Genes Dev* 1988, **2**(9):1136-1143.

Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M *et al*: **Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells**. *Nat Genet* 2005, **37**(8):853-862.

Widschwendter M, Fiegl H, Egle D, Mueller-Holzner E, Spizzo G *et al*: **Epigenetic stem cell signature in cancer**. *Nat Genet* 2007, **39**(2):157-158.

Wong TS, Liu XB, Wong BY, Ng RW, Yuen AP *et al*: **Mature miR-184 as Potential Oncogenic microRNA of Squamous Cell Carcinoma of Tongue**. *Clin Cancer Res* 2008, **14**(9):2588-2592.

Wutz A, Jaenisch R: **A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation**. *Mol Cell* 2000, **5**(4):695-705.

Yamanaka M, Watanabe M, Yamada Y, Takagi A, Murata T *et al*: **Altered methylation of multiple genes in carcinogenesis of the prostate**. *Int J Cancer* 2003, **106**(3):382-387.

Yang X, Karuturi RK, Sun F, Aau M, Yu K *et al*: **CDKN1C (p57) is a direct target of EZH2 and suppressed by multiple epigenetic mechanisms in breast cancer cells**. *PLoS One* 2009, **4**(4):e5011.

Yegnasubramanian S, Kowalski J, Gonzalgo ML, Zahurak M, Piantadosi S *et al*: **Hypermethylation of CpG islands in primary and metastatic human prostate cancer**. *Cancer Res* 2004, **64**(6):1975-1986.

Yegnasubramanian S, Haffner MC, Zhang Y, Gurel B, Cornish TC *et al*: **DNA hypomethylation arises later in prostate cancer progression than CpG island hypermethylation and contributes to metastatic tumor heterogeneity**. *Cancer Res* 2008, **68**(21):8954-8967.

Yoshimoto M, Cutz JC, Nuin PA, Joshua AM, Bayani J *et al*: **Interphase FISH analysis of PTEN in histologic sections shows genomic deletions in 68% of primary prostate cancer and 23% of high-grade prostatic intra-epithelial neoplasias**. *Cancer Genet Cytogenet* 2006, **169**(2):128-137.

Yu J, Rhodes DR, Tomlins SA, Cao X, Chen G *et al*: **A polycomb repression signature in metastatic prostate cancer predicts cancer outcome**. *Cancer Res* 2007, **67**(22):10657-10663.

Yu J, Mani RS, Cao Q, Brenner CJ, Cao X *et al*: **An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression**. *Cancer Cell* 2010, **17**(5):443-454.

Zhang B, Liu XX, He JR, Zhou CX, Guo M *et al*: **Pathologically decreased miR-26a antagonizes apoptosis and facilitates carcinogenesis by targeting MTDH and EZH2 in breast cancer**. *Carcinogenesis* 2011a, **32**(1):2-9.

Zhang H, Hao Y, Yang J, Zhou Y, Li J *et al*: **Genome-wide functional screening of miR-23b as a pleiotropic modulator suppressing cancer metastasis**. *Nat Commun* 2011b, **2**:554.

Zheng F, Liao YJ, Cai MY, Liu YH, Liu TH *et al*: **The putative tumour suppressor microRNA-124 modulates hepatocellular carcinoma cell aggressiveness by repressing ROCK2 and EZH2**. *Gut* 2011.

Zhou Q, Gallagher R, Ufret-Vincenty R, Li X, Olson EN *et al*: **Regulation of angiogenesis and choroidal neovascularization by members of microRNA-23~27~24 clusters**. *Proc Natl Acad Sci U S A* 2011, **108**(20):8287-8292.

Zipper H, Brunner H, Bernhagen J, Vitzthum F: **Investigations on DNA intercalation and surface binding by SYBR Green I, its structure determination and methodological implications**. *Nucleic Acids Res* 2004, **32**(12):e103.

EIGENSTÄNDIGKEITSERKLÄRUNG

Hiermit versichere ich, daß ich die vorliegende Dissertation mit dem Titel *"ANALYSES OF DNA METHYLATION PATTERNS IN PROSTATE CANCER"* selbständig verfaßt und keine weiteren als die angegebenen Hilfsmittel verwendet habe. In der Arbeit verwendete Aussagen anderer Autoren habe ich durch Quellenangabe kenntlich gemacht. Die vorliegende Arbeit wurde in keinem früheren Promotionsverfahren eingereicht oder als ungenügend beurteilt.

Berlin, den 24.07.2012

Stefan Börno

CURRICULUM VITAE

Der Lebenslauf ist in der Online-Version aus Gründen des Datenschutzes nicht enthalten.

PUBLICATION RECORD

Kacprzyk LA, Laible M, Andrasiuk T, **Boerno ST**, Brase JC, Fälth M, Kuner R, Lehrach H, Schweiger MR, Sültmann H: **Epigenetic activation of TDRD1 in TMPRSS2:ERG-positive prostate cancer**. 2012, in preparation.

Roehr C, Kerick M, Fischer A, Li J, Kashofer K, Timmermann B, Meinel T, Drichel D, **Boerno ST**, Nowka A, Krobitsch S, Becker T, Wunderlich A, Barmeyer C, Viertler C, Zatloukal K, Wierling C, Lehrach H, Schweiger M: **Identification of miRNA-1 as tumor suppressor and modeling of its therapeutic functions.** 2012, in preparation.

Boerno ST, Fischer A, Kerick M, Fälth M, Laible M, Brase J, Kuner R, Dahl A, Grimm C, Sayanjali B, Isau M, Roehr C, Wunderlich A, Timmermann B, Claus R, Plass C, Graefen M, Simon R, Demichelis F, Rubin M et al: **Genome-wide DNA methylation events in TMPRSS2:ERG fusion negative prostate cancers implicate an EZH2 dependent mechanism with miRNA-26a hypermethylation.** *Cancer Discovery* 2012, accepted.

Haupt A, Joberty G, Bantscheff M, Frohlich H, Stehr H, Schweiger MR, Fischer A, Kerick M, **Boerno ST**, Dahl A, Lappe M, Lehrach H, Gonzalez C, Drewes G, Lange BM: **Hsp90 inhibition differentially destabilises MAP kinase and TGF-beta signalling components in cancer cells revealed by kinase-targeted chemoproteomics.** *BMC Cancer* 2012, 12:38.

Boerno ST, Grimm C, Lehrach H, Schweiger MR: **Next-generation sequencing technologies for DNA methylation analyses in cancer genomics.** *Epigenomics* 2010, 2(2):199-207.

Dahl A, Mertes F, Marchfelder U, **Boerno ST**, Fischer A, Schweiger MR, Lehrach H: **Bewertung der SOLiD-Sequenzierungsplattform aus Nutzerperspektive.** *Laborwelt* 2010, 1/2010:8-10.

Timmermann B, Kerick M, Roehr C, Fischer A, Isau M, **Boerno ST**, Wunderlich A, Barmeyer C, Seemann P, Koenig J, Lappe M, Kuss AW, Garshasbi M, Bertram L, Trappe K, Werber M, Herrmann BG, Zatloukal K, Lehrach H, Schweiger MR: **Somatic mutation profiles of MSI and MSS colorectal cancer identified by whole exome next generation sequencing and bioinformatics analysis.** *PLoS One* 2010, 5(12):e15661.

ACKNOWLEDGEMENTS

I am grateful for the opportunity to thank all the people who contributed to this thesis and enriched my life during my time as PhD student.

First of all I want to thank Professor Dr. Hans Lehrach for having given me the chance to work in this extraordinary and inspiring environment. I also want to thank Professor Dr. Rupert Mutzel for taking the time of being my second supervisor.

I am especially indebted to my supervisor Dr. Dr. Michal-Ruth Schweiger who made this whole endeavour possible and whose advice I always appreciated. I am glad that she had so much confidence in me to entrust me with the investigations in this interesting and challenging field of research. Furthermore, I want to thank her for her patience and the time she spent in teaching me the principles of independent scientific work. I am eminently thankful for the multiple possibilities she gave me to attend conferences and meetings in Hamburg, Heidelberg, Berlin, Antwerpes, Darmstadt and Potsdam, to present my data and get into contact with other scientists.

On this occasion I want to thank our collaborators from the Universitätsklinikum Hamburg Eppendorf and the Martini-Klinik, Professor Dr. Guido Sauter, Dr. Thorsten Schlomm and Dr. Ronald Simon for providing me with the prostate and urine samples, the excellent organisation of the hanseatic conferences and the interesting insights into clinical research.

I also feel fortunate that I could work together with Professor Dr. Holger Sültmann, Dr. Maria Fälth, Dr. Jan Brase, Dr. Ruprecht Kuner and Mark Laible from the DKFZ who gave me access to the gene and microRNA expression data. I always had a very pleasant time during our meetings. I appreciate the help of Professor Dr. Christoph Plass and Dr. Rainer Claus from the DKFZ with the BS-MS experiments. Furthermore, I want to thank Dr. Hans Krause from the Charité in Berlin for providing patients' urine specimens for methylation analyses.

I sing the praises of the dedicated technicians Nada Kumer, Anna Kosiura, Uta Marchfelder and Michael Zinke who were a great help in handling the vast amount of experiments and whose precise work was a guarantee for high quality data. I furthermore want to thank Dr. Christina Grimm who helped me during my first encounters with the MeDIP protocol. Together with Dr. Andreas Dahl I fought through the challenges and obstacles of early next generation sequencing and I am grateful for the numerous fruitful discussions and advice for protocol optimisations. I want to thank Axel Fischer for 'digesting' the output of the SOLiDs into a manageable format. Furthermore, I thank him

and Dr. Martin Kerick for their patience in introducing me in R programming and GUI independent computer work.
Next I want to thank my colleagues from the lab in order of appearance: Andrea Wunderlich, Christina Röhr, Melanie Isau, and Michelle Hussong who all contributed to a large part to making our lab a happy place. I always appreciated their help with the challenges of laboratory work and always found an expert for a certain question in at least one of them.
I am thankful for the opportunity to share my knowledge with the undergraduate students Behnam Sayanjali and Jana-Carina Elias, a very talented and skilful biologist.

Christina Röhr, Meryem Ralser and Thomas Bergmann became good friends and enriched my life also outside the lab. At this point I want to express my regret that Mirjam Blattner left Berlin for the U.S. and want to thank her for her hospitality in the Tuscany and in NY. Dr. Bernd Timmermann and his wonderful technicians are appreciated for organising several in-house meetings to discuss advantages and disadvantages of the various massive parallel sequencing technologies.

My deepest debt is to my family, my parents and my brother. They always helped me and formed me into the person I am now. Without their support and confidence I never would have made it here. I also want to include Anke Gründel, Anika Dreilich and Björn Stelbrink who helped me through the hardest part of my studies.
I want to thank Petra, Hartmut and Jan Zumkowski who cordially welcomed me into the novel half of my family.

There are no words to express my gratitude to my wonderful and beloved wife Maria. She always believed in me and made my life complete. I enjoy every day she is around.