

# **Unravelling Molecular Mechanisms Underlying Genetic and Epigenetic Instabilities in Colorectal Cancer**

**Inauguraldissertation**

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

von

Stefan Weis

aus

Rheinfelden, Deutschland

Basel, 2015

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät  
auf Antrag von

Prof. Dr. Primo Schär (Fakultätsverantwortlicher und Dissertationsleiter)

Prof. Dr. Giulio Spagnoli (Korreferent)

Basel, den 09.12.2014

Prof. Dr. Jörg Schibler  
Dekan der Philosophisch-Naturwissenschaftlichen Fakultät

## **Acknowledgements**

I would like to thank Primo Schär for supervising my thesis, for his motivating words at the right time and for his contagious enthusiasm.

Moreover, I want to express my gratitude to my PhD committee: Giulio Spagnoli, Giancarlo Marra, Kaspar Truninger and Gerhard Christophori for their critical evaluation of my work.

My thanks also go to Faiza Noreen for her ongoing support, readiness to help and critical reading of parts of this thesis. I would like to thank David Schürmann for the willingness to share his broad methodological knowledge with me and for reading parts of this thesis.

I would like to thank all past and present members of the group Schär for maintaining a really enjoyable work atmosphere. Thanks go to Annika Wirz, Claudia Krawczyk, Melissa Manser and Emina Gyenge Besic for their support and for being more than colleagues.

I thank my family, my parents Brigitte and Bodo Weis and my brothers Matthias and Daniel, who have supported me during my whole studies.

Last but not least, I wish to thank Simone for standing by me during the most challenging times, for encouraging me, for her trust in me and for her love.

*„Die Neugier steht immer an erster Stelle eines Problems, das gelöst werden will.“*

*-Galileo Galilei-*

# Table of Content

<b>1</b>	<b>Summary</b> .....	<b>8</b>
<b>2</b>	<b>Introduction</b> .....	<b>11</b>
2.1	The Instable Cancer Genome .....	11
2.1.1	Chromosomal Instability (CIN).....	11
2.1.2	Microsatellite Instability (MSI).....	15
2.1.3	DNA Sequence Mutability .....	17
2.2	Epigenetic Instability.....	18
2.2.1	Epigenetic Mechanisms .....	19
2.2.2	DNA methylation .....	24
2.2.3	DNA Demethylation.....	29
2.3	Colorectal Cancer .....	33
2.3.1	(Epi-) Genetic Instabilities in Colorectal Carcinogenesis.....	36
2.3.2	(Epi-) Genomic Instabilities in Colorectal Cancer Progression.....	39
<b>3</b>	<b>Aims of the Thesis</b> .....	<b>42</b>
<b>4</b>	<b>Results</b> .....	<b>43</b>
4.1	Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers (Appendix I).....	43
4.2	Modulation of Age- and Cancer-Associated DNA Methylation Change in Healthy Colon by Aspirin and Lifestyle (Appendix II).....	47
4.3	The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation (Appendix III) .....	50
4.4	Colorectal mucosa of healthy individuals displays part of the CpG island methylator phenotype signature (Appendix IV) .....	54
4.5	Supplementary Results .....	57
4.5.1	Association of Chromosomal and Epigenetic Alterations During Tumor Development.....	57
<b>5</b>	<b>Concluding Discussion and Outlook</b> .....	<b>67</b>
<b>6</b>	<b>References</b> .....	<b>72</b>



Appendix:

- I. Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers
- II. Modulation of Age- and Cancer-Associated DNA Methylation Change in Healthy Colon by Aspirin and Lifestyle
- III. The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation
- IV. Colorectal mucosa of healthy individuals displays part of the CpG island methylator phenotype signature

## Abbreviations

5-caC	5-carboxylcytosine
5-fC	5-formylcytosine
5-hmC	5-hydroxymethylcytosine
5-hmU	5-hydroxymethyluracil
5-mC	5-methylcytosine
A	adenine
BER	Base Excision Repair
C	cytosine
CGI	CpG island
ChIP	Chromatin IP
CIMP	CpG island methylator phenotype
CIN	Chromosomal instability
CRC	Colorectal cancer
DNA	Deoxyribonucleic acid
DNMT	DNA methyltransferase
DSB	Double strand break
EMT	Epithelial mesenchymal transition
ESC	Embryonic stem cell
FAP	Familial adenomatous polyposis
G	guanine
GCR	Gross chromosomal rearrangement
HNPCC	Hereditary non-polyposis colorectal cancer
HR	homologous recombination
ICR	Imprinting control region
IDL	Insertion deletion loop
IP	Immunoprecipitation
LOH	Loss of heterozygosity
LOI	Loss of imprinting
MeDIP	Methylated DNA IP
MET	Mesenchymal epithelial transition
MMR	Mismatch repair
MSI	Microsatellite instability

NHEJ	non-homologous end joining
OIS	Oncogene induced senescence
RNA	Ribonucleicacid
T	thymine
TIC	Tumor initiating cell
TSS	Transcription Start Site

# 1 Summary

Colorectal cancers (CRCs) show extensive genetic and epigenetic aberrations, including DNA sequence changes, chromosomal alterations and abnormal epigenetic chromatin modifications. These alterations often reflect defects in molecular mechanisms establishing and maintaining (epi-) genome integrity. Central to the maintenance of genetic stability is the cellular DNA damage response, including various systems of DNA repair, the malfunction of which increases the rate of mutation and, hence, drives tumor evolution. Epigenetic alterations in CRC are as prominent as genetic alterations and well documented, both, at the levels of histone modifications and DNA methylation (Kondo and Issa, 2004). Epigenetic mechanisms play critical roles in the regulation of DNA-templated processes such as transcription, DNA repair but also replication. Shaping higher order chromatin structure, they regulate DNA accessibility and, thereby, assure proper genome function. Epigenetic aberrations deregulate genomes and thus contribute to cell transformation in different ways. Although many epigenetic alterations in CRC and other cancers are described and, in specific cases, their role in carcinogenesis established, the origin of and the molecular mechanisms underlying epigenetic instability in CRC remained elusive (Goel and Boland, 2012).

The aims of my thesis were to elucidate the molecular basis of DNA methylation aberrations in CRC in a first part, and to investigate mechanisms by which CRC-associated DNA methylation alterations contribute to colorectal carcinogenesis in a second part.

A subset of human CRCs is characterized by genome-wide hypermethylation of promoter CpG islands (CGIs) and therefore referred to as displaying a CGI-methylator phenotype (CIMP). Although, overexpression of DNA methyltransferases (DNMTs) has been associated with CRC-CIMP, it has remained uncertain to what extent this is causal for CGI hypermethylation (Kanai et al., 2001; Nosho et al., 2009). Mutational inactivation and metabolic inhibition of the Ten Eleven Translocation (TET) DNA dioxygenases, key factors of active DNA demethylation, have been reported to be causal for DNA hypermethylation phenotypes in leukemias and gliomas (Figuroa et al., 2010; Letouze et al., 2013; Turcan et al., 2012). We investigated whether deregulated, TET-mediated DNA demethylation can account for CIMP in CRC. We discovered that TET1 is significantly downregulated, both at the mRNA and protein level, in CIMP-displaying CRCs. Knock-down of TET1 in the non-CIMP CRC cell line SW620 confirmed that the loss of TET1 results in genome-wide CGI hyper- and hypomethylation in cell culture, which, upon xenograft tumor formation, converted into a hypermethylation phenotype. Notably, the loss of TET1 in the CRC cell line resulted in increased promoter methylation and concomitant transcriptional downregulation of MLH1 and CDH1, a characteristic feature of CRC-CIMP. These results thus showed that depletion of TET1 and selection during tumor

growth can recapitulate features of CIMP in non-CIMP cancer cells. Thus, TET1 activity in colon epithelial cells contributes to maintenance of cell type specific CGI-methylation patterns and prevents the development of CRC through the CIMP-pathway.

Beside cancer-specific defects in the maintenance of CpG methylation, aging constitutes a major source of methylation alterations associated with CRC (Issa, 2014; Menigatti et al., 2009; Toyota et al., 1999). Lifestyle-related factors including overweight, aspirin use, hormone replacement therapy (HRT) and smoking have been reported to influence the risk of CRC over long periods of time. (Chan et al., 2007; Long et al., 2010; Parajuli et al., 2013; Robsahm et al., 2013). I contributed to a collaborative project, in which we investigated whether these factors exert their effects on cancer risk via modulation of age-related DNA methylation changes. Genome-wide and detailed target-wise methylation analysis in a cohort of 546 healthy women revealed that promoter-associated DNA methylation increases over age in specific patterns. We were able to determine and compare for the first time rates of age-dependent DNA methylation change and thereby observed that smoking and obesity promote whereas aspirin use and HRT reduce the genome-wide methylation drift. A significant fraction of promoter-associated CpG sites, whose age-related hypermethylation was modified by lifestyle factors, also occurred hypermethylated in CRC. Lifestyle factors affected the methylation drift of CRC-related CpGs concordant with the effects of these factors on CRC risk. These and further results suggested that the effect of the investigated lifestyle factors on CRC-risk can be accounted for by the modulation of age-related DNA methylation changes.

The promoter methylation of cancer-relevant genes detectable in the healthy colorectal mucosa, as reported for instance for the two DNA repair genes MLH1 and MGMT, allows the definition and monitoring of early epigenetic events in colorectal carcinogenesis (Menigatti et al., 2009). To explore this possibility, we selected five genes downregulated in colorectal adenomas (precancerous, benign tumors of the colorectum) and quantified their promoter methylation levels in CRC tissue as well as in the matched healthy appearing colonic mucosa. We included MLH1 and MGMT, two genes with established roles in colorectal carcinogenesis, in this analysis. The five marker genes; FOXF1, CA4, NPY1R, GREM1 and IFITM1 were previously not described as targets of aberrant hypermethylation in CRC. With different incidence all seven markers were found hypermethylated in 106 CRC samples tested. Interestingly, methylation of the two novel markers CA4 and GREM1 associated with CRC-CIMP. We found that these two markers are methylated at low levels also in cancer-associated mucosa as well as in colonic mucosa from healthy individuals, as it was reported previously for the CIMP marker MLH1 (Menigatti et al., 2009). The results of this collaborative efforts thus suggest that features of CIMP-specific DNA methylation are present and, hence, detectable already in the mucosa of healthy individuals.

DNA hypermethylation in gene promoters often interferes with transcription of the associated gene. Promoter methylation can exert its repressive function due to compaction of the chromatin structure but also by interfering with transcription factor binding (Kulis and Esteller, 2010). In a collaborative project, we identified SHOCA-2 as a regulator of Epidermal Growth Factor Receptor (EGFR) signaling and observed its loss of expression in advanced CRCs. The loss of SHOC-2 expression was associated with different genetic and epigenetic alterations. We detected chromosomal deletions affecting the SH2D4A gene (encoding for SHOCA-2) as well as mutations in the coding region of the gene. Finally, we found the cancer-specific hypermethylation of two CpG sites in the 5'-untranslated region of SH2D4A, one of which overlapped with a binding site for the transcription factor Sp1. I was able to show that methylation of this CpG site in the SH2D4A sequence context interferes with Sp1 binding *in vitro* and therefore might be responsible for SHOCA-2 expression loss in some CRCs.

Taken together, in collaboration with many colleagues from different laboratories, I was able to elucidate different mechanisms involved in the generation of CRC-related DNA methylation aberrations. First, I could show that impaired TET-mediated DNA demethylation contributes in the acquisition of methylation aberrations in CRCs characterized by a methylator phenotype. Second, together with colleagues I was able to provide new and fundamental insight into the relationships between aging and lifestyle and cancer-relevant DNA methylation change in a cohort of healthy individuals. Finally, by investigating promoter methylation of the newly established tumor suppressor SHOCA-2, I was able to describe a molecular mechanism by which site-specific DNA methylation may contribute to cancerous transformation.

## **2 Introduction**

### **2.1 The Instable Cancer Genome**

Human cancers show a large variety of genetic and epigenetic alterations when compared to the respective healthy tissue. These aberrations form the genetic basis of the malignant cell phenotype (Hanahan and Weinberg, 2011). Genomic changes range from subtle point mutations in the primary DNA sequence to chromosomal aberrations with structural and numerical changes of huge DNA sections. Cancer-associated changes in the epigenome include all layers of epigenetic information: DNA-methylation, histone modifications as well as the expression and function of regulatory RNA molecules (Esteller, 2011; Ting et al., 2006). In healthy human cells the genome integrity is maintained by a number of control and repair mechanisms which keep the rate and extent of mutations very low. However, as the normally low spontaneous mutation rate cannot account for the large number of mutations accumulating in human cancers, Loeb proposed that pre-cancerous cells elevate their mutation rate and acquire a so-called mutator phenotype (Loeb, 1991). Increased mutation can, for instance, result from replication stress induced by oncogene activation or by defects in processes responsible for the maintenance of genome integrity. This mutability of the cancer genome together with an epigenetic instability provides for a high genetic plasticity of cancer cells that facilitates fast diversification and high potential for adaptation to environmental changes like increased hypoxia and challenges by the immune system. The following chapters introduce the main forms of genetic and epigenetic instabilities found in human cancers as well as their underlying causes.

#### **2.1.1 Chromosomal Instability (CIN)**

Chromosomal aberrations have long been described as a hallmark of human cancers (Baudis, 2007; Beroukhim et al., 2010). These abnormalities can include large parts of chromosomes like chromosome arms or even whole chromosomes and were therefore originally identified as alterations in the karyotype. The occurrence of intra-tumor variations in chromosomal aberrations was referred to as chromosomal instability (CIN). The increased chromosomal plasticity in CIN cancers contributes to the phenotypic variation within the tumor and might facilitates adaption to environmental changes (D. J. Gordon et al., 2012). In line with this, CIN appears to influence cancer sensitivity to chemotherapy (McClelland et al., 2009). Chromosomal aberrations can affect genome functionality either by structural or numerical effects. Structural aberrations exert their effect due to the rearrangement of genetic information whereas numerical changes alter the abundance of genetic information. Translocations are most prominent type of structural aberrations in cancer (Nambiar et al., 2008). Chromosomal

translocation is an exchange of genetic material between two non-homologous chromosomes. This can happen either in a balanced manner with even exchange of material or unbalanced when the exchange is unequal resulting in gain or loss of chromosome regions. When involving two separate gene locations translocations can give rise to fusion genes which are commonly found in human cancers. Fusion genes can be oncogenic due to the deregulation of the original function or acquisition of a new function. The fusion between the histone methyltransferase MLL and the DNA hydroxylase TET1, for instance, which is frequent and causal in acute myeloid leukemia (AML) results from a translocation between chromosome 10q22 and chromosome 11q23 (Lorsbach et al., 2003). Translocations represent a form of gross chromosomal rearrangements (GCR), the origin of which is generally the misrepair of DNA double strand-breaks (DSB). Two major pathways are involved in the repair of DSBs: the homology dependent repair or homologous recombination (HR) system and the non-homologous end joining (NHEJ) pathway. Repair by the NHEJ-pathway, which directly re-joins two double-stranded DNA ends of a DSB has been linked to the formation of structural aberrations. This mechanism is error-prone because it uses only small sequence homologies at the single stranded ends of the DSB and can potentially link any two available chromosome ends (Kasperek and Humphrey, 2011; Natarajan et al., 2008). NHEJ represents the predominant DSB-repair mechanism in the cell cycle stage G1 before DNA is replicated and no sister chromatid is available for HR, which uses the homologous sister duplex as a template for DSB repair and, therefore, operates with higher fidelity. The tumor suppressors BRCA1 and BRCA2 are essential for DSB repair by HR. The functional loss of these or other HR factors due to mutation or transcriptional inactivation leads to a bias in DSB repair towards the error-prone NHEJ (Thompson et al., 2010). DSBs are relatively rare but deleterious lesions and originate from exposure of cells to ionizing radiation but also oxidative stress by reactive oxygen species (ROS). In most human cancers, however, replication-associated DSBs appear to be the major source of GCRs. They occur in response to “replication stress” when the DNA replication fork encounters DNA lesions or DNA secondary structures and becomes stalled. If unrepaired or uncorrected the stalled replication fork may collapse (the replication complex falls apart) and potentially generate a DSB (Zeman and Cimprich, 2014). Some DNA regions are intrinsically capable to produce replication stress and tend to break during replication, particularly under replication inhibitory conditions. Such fragile sites are characterized by the potential to form DNA secondary structures which interfere with replication fork progression and represent places of chromosome breakage and rearrangement in human cancers (Durkin and Glover, 2007). An example are G-quadruplexes (G4), intramolecular four-stranded structures which arise from guanine-rich sequences ( $G_3N_{1-7}G_3N_{1-7}G_3N_{1-7}G_3$ ) capable to obstruct polymerase movement. Many potential G4 sequences contain CpG sites in which cytosine can be methylated (see chapter 2.2.2 DNA methylation). Interestingly, it was shown

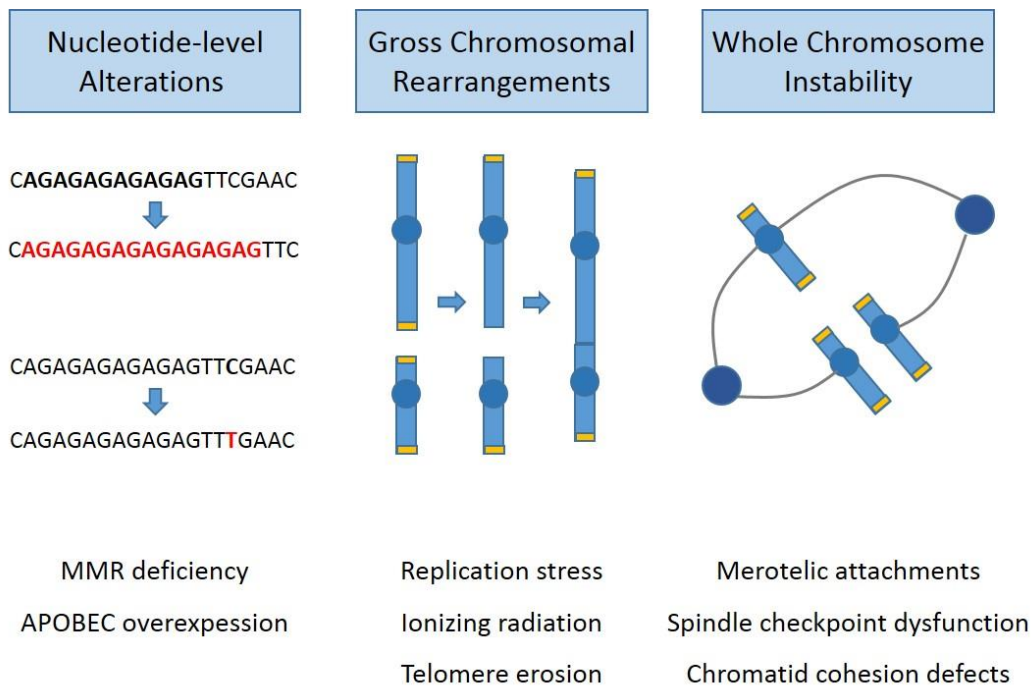


that G4 sequences, associated with chromosomal breakpoints are cancer-specifically hypomethylated (De and Michor, 2011). In addition to DSBs also telomere dysfunction can induce GCRs. In highly proliferative cancer cells the repetitive telomere-region at the chromosome ends becomes rapidly shortened due to the so-called “end-replication problem”- the incomplete synthesis of the 3' end of the DNA strand. When the telomere end protection is compromised chromosome ends are seen by the cells as DSBs and subsequent NHEJ can result in the fusion of two different chromosomes, generating a dicentric chromosome (two centromeres) (Gisselsson et al., 2000). In mitosis, the two chromatids involved may end up being attached to microtubules from opposite spindle poles, thus forming chromatin bridges in telophase. Breakage of these bridges that occurs not necessarily at the site where the chromosomes were fused can be the start of another round of breakage-fusion-bridge (BFB). Repeating BFB cycles contribute strongly to the intra-tumoral chromosomal instability and diversification (Gisselsson et al., 2000).

In addition to the structural changes, numerical chromosome aberrations, termed aneuploidy, are ubiquitous in human cancers. Aneuploidy usually includes either very short DNA segments whose length ranges from 1 kb (kilobase pair) to a few Mb (megabase pairs) (focal) or affects chromosome arms or whole chromosomes (arm-level). Small focal copy number variations are present in all humans and contribute to human evolution and genetic diversity. They may account for 13% of the human genome (Stankiewicz and Lupski, 2010). Accordingly, cancer cells are characterized by larger, arm-level somatic copy number alterations (SCNAs). It was shown that 25% of a typical human cancer genome is affected by arm-level SCNAs (Beroukhi et al., 2010). Aneuploidy can contribute differently to cancerous transformation depending on the genes affected. Aneuploidy can result in the deletion of tumor suppressor genes or the amplification of oncogenes. Linked to that SCNAs can cause loss of heterozygosity (LOH) which means the loss of the wildtype allele in heterozygous loss-of-function mutations. Eventually, aneuploidy could affect levels of DNA repair genes and thereby promote genomic instability (Janssen and Medema, 2013). However, aneuploidy has also potentially detrimental effects on cell viability and proliferation due to gene dosage effects. Cancer cells can adapt to tolerate an aneuploid genotype in several ways: they can increase the protein degradation to compensate for increased protein amounts, or they can buffer the effects of aneuploidy by facilitating further numerical whole-chromosome alterations. The gain or loss of a single chromosome is expected to have a larger impact in a diploid than in a polyploid cancer cell. Finally, cancer cells often have an impaired p53 signaling pathway, which limits the proliferation of aneuploid cells (Ganem and Pellman, 2007; D. J. Gordon et al., 2012; Thompson et al., 2010). Chromosome arm and whole chromosome numerical alterations originate from different underlying mechanisms. Whereas arm-sized deletions and amplifications are the result of DSB-initiated GCRs, the mechanisms underlying whole-

chromosome aneuploidy are more diverse. One mechanism includes the aberrant attachment of a single kinetochore to microtubules from both spindle poles, termed merotelic attachments. Merotelic attachments are common during mitosis due to stochastic interactions between the kinetochore and the spindle microtubules. The incidence of merotelic attachments is basically determined by the rate of their formation and the rate of their correction. An increased number of centrosomes often associated with an aberrant overexpression of the Aurora A kinase, elevates the rate of merotely. Defects in several kinetochore proteins like CENP-E can hyperstabilize kinetochore-microtubule attachments and thereby lower the correction rate of erroneous merotelic attachments (D. J. Gordon et al., 2012; Thompson et al., 2010). The second pathway of whole chromosome aneuploidy is characterized by defects in the spindle assembly checkpoint (SAC). SAC signaling arrests cells in metaphase until all chromosomes form proper bipolar attachments to the spindle microtubules. Yeast and mouse models have shown that impaired SAC increases the rate of chromosomal missegregation and can cause CIN. However, the role of SAC defects in human cancers is not clear. Mutations in SAC proteins appear extremely rare, but there is evidence of epigenetic inactivation of SAC genes (Cahill et al., 1999; Haruki et al., 2001; Myrie et al., 2000; Park et al., 2007; R. H. Wang et al., 2004). Finally, defects in chromosome cohesion were found to be associated with whole chromosome aberrations in cancer. Chromosome cohesion, mediated by the cohesin protein complex, prevents premature separation of sister chromatids during mitosis. Mutations in cohesin subunits were shown to cause aneuploidy in cancer (Solomon et al., 2011). Further, overexpression of separase, a protease which cleaves the cohesin complex during the onset of anaphase (before sister chromatids become separated) was found in breast cancer cells (N. Zhang et al., 2008). Excessive separase expression can result in premature sister chromatid dissociation and chromosome missegregation.

A special form of cancer-associated GCRs is characterized by massive structural and numerical (mostly deletions) aberrations occurring clustered in discrete sub-chromosomal regions. This phenomenon, termed chromothripsis, is most evident in a small fraction of solid tumors. The molecular mechanisms underlying chromothripsis are not clear, although some evidence links its occurrence to replication stress. Based on *in silico* simulations, chromothripsis and concomitant complex rearrangements were originally described as a result of a single catastrophic event. However, revision of initially applied statistical approaches indicates that they could also result from progressive chromosomal alterations (M. J. Jones and Jallepalli, 2012; Kinsella et al., 2014; Stephens et al., 2011).



**Figure 2-1: The spectrum of genomic instabilities in cancer.** Genetic alterations in cancer can basically be subdivided into three main classes. At the smallest scale, subtle sequence changes may affect single or a few adjacent nucleotides (left). Examples include changes in the length of microsatellites as a result of MMR deficiency or C to T transition mutations deriving from deregulation of DNA deaminases. At the intermediate scale, gross chromosomal rearrangements (GCRs) including translocations, deletions and amplifications arise mainly from double strand breaks and telomeric dysfunction (middle). At the largest scale, whole chromosome instability is the major cause of aneuploidy in cancer and can result from errors in virtually any aspect of mitosis (right). The depicted example shows the merotelic kinetochore-microtubule attachment (adapted from Jones and Jallepalli, 2012).

### 2.1.2 Microsatellite Instability (MSI)

Microsatellite instability (MSI) is a hallmark of a restricted group of human malignancies including colorectal, gastric and endometrial cancers. MSI is characterized by variability of the length of microsatellite repetitive DNA sequences. Microsatellites, also referred to as short tandem repeats (STR) or short sequence repeats (SSR), are sequences of 1-5 nucleotides repeated from ten to thousand times in tandem. Microsatellites account for about 3% of the human genome and are usually located in non-coding DNA regions. They can overlap with gene regulatory regions and MSI can therefore lead to downregulation of gene expression (Kim et al., 2013). A fraction of microsatellites resides within gene exons and the instability in the length of these microsatellites may induce frame shift mutations (Duval and Hamelin, 2002). For instance, TGF $\beta$ R2 contains a microsatellite of ten consecutive adenines (A<sub>10</sub>) in

exon 3. Single nucleotide deletions in this microsatellite are common in MSI colorectal cancers and render the gene product – a transforming growth factor receptor – non-functional (Markowitz et al., 1995).

Instability in the length of DNA microsatellites is caused by defects in the replication-coupled DNA mismatch repair system (MMR). Single nucleotide insertion- or deletion loops (IDL) occur occasionally during replication due to DNA polymerase slippage on templates with repetitive sequences. Microsatellite sequences may form secondary DNA structures like DNA hairpins or small stem-loop structures, which can pause the encountering DNA polymerase during replication. The stalled polymerase possibly dissociates and the newly synthesized DNA strand can separate from the template. Shifted re-annealing with another tandem repeat generates small loops either in the template or in the daughter strand (IDL). If unrepaired by MMR, subsequent reassembly of the polymerase and further replication would result in insertions or deletions in the daughter strand depending in which strand the IDLs occurred (Viguera et al., 2001). According to the role of MMR in correcting mis-incorporated single nucleotides and IDLs, the loss of functional MMR results in the accumulation of microsatellite-associated frameshift mutation but also in nucleotide substitutions and therefore generates a strong mutator phenotype. MMR deficient colorectal cancers, for instance, are characterized by their hypermutated genomes compared to microsatellite stable cancers (Cancer Genome Atlas, 2012). The core MMR machinery in mammals consists of seven proteins named according to their homology to the bacterial and yeast mutS, mutL and pms proteins, respectively. MutS homologues, including MSH2, MSH3 and MSH6 function as heterodimers in the recognition and binding to single nucleotide mismatches and IDLs. Upon MutS dimer binding to a DNA mismatch a MutL heterodimer is recruited. In mammals there are four MutL homologs: MLH1, MLH3, PMS1 and PMS2. All mutL heterodimers include MLH1, which therefore possesses a central role in mammalian mismatch repair. The MLH1-PMS2 dimer was shown to have an endonuclease activity which cleaves the newly synthesised DNA strand close to the lesion. The subsequently recruited exonuclease EXO1 and DNA polymerase  $\delta$  remove the incised DNA strand and resynthesize it. DNA ligase 1 seals the newly synthesized fragment and completes the repair process (Kadyrov et al., 2006; Pena-Diaz and Jiricny, 2012). Germline point mutations or deletions affecting mismatch repair genes, mainly MSH2 and MLH1, are associated with a severe hereditary predisposition to cancer, called Lynch syndrome (Boland and Goel, 2010; van der Klift et al., 2005). Lynch syndrome associated tumors, which predominantly develop in the colon but also in endometrium and stomach, are all characterized by MSI. Sporadic forms of MSI colon and endometrial cancers often show epigenetic inactivation of MLH1.

### 2.1.3 DNA Sequence Mutability

Genomes of cancer cells can harbor hundreds and thousands of DNA sequence mutations (Alexandrov et al., 2013). These include nucleotide substitutions as well as small insertions and deletions (indels). The prevalence of somatic mutations is the highest in cancers associated with chronic mutagen exposure like skin (melanoma, UV light) and lung cancers (tobacco smoking). This indicates an important role of exogenous mutagenic agents in the generation of somatic mutations in cancer (Alexandrov et al., 2013). Only in MMR deficient colorectal cancers the incidence of mutations is higher than in melanomas and lung cancers (Vogelstein et al., 2013). The most common form of mutations found in human cancers are C to T (C>T) substitutions in a cytosine-guanine (CpG) sequence context. Cytosines in a CpG dinucleotide can become methylated at the C-5 position (5mC). This modification does not alter base pairing but, in plants and vertebrates it serves as an epigenetic mark involved in the regulation of transcriptional activity. Methylated cytosines are hotspots of mutations in human cancers. Compared to unmethylated cytosines, 5mCs are more susceptible to spontaneous deamination resulting in the conversion of 5mC into thymine (C>T). Methylated CpGs also appear to be a preferential target for alkylation of guanine to generate 6-methyl-G, which then mispairs with T during replication, resulting in G>A mutations (Chen et al., 1998). Further, the apparently inefficient repair of T/G mismatches compared to U/T mismatches (resulting from cytosine deamination) and the higher vulnerability to UV light induced pyrimidine dimer formation may contribute to increased mutagenic potential of 5mC (Pfeifer et al., 2005). Interestingly, the extent of C>T mutations in a cancer genome correlates significantly with the age of the patient, indicating that these mutations are acquired continuously during life time rather than being the result of an increased cancer-associated mutability. However, other types of mutations appear to be associated with defects in DNA repair. As mentioned above small indels (mostly of 1bp) in short sequence repeats are characteristic for MSI cancers with defective MMR. Larger indels of up to 50bp with overlapping microhomology at the breakpoint junctions are likely to result from the functional loss of BRCA1 and BRCA2 and the concomitant bias towards NHEJ-mediated DSB repair (Alexandrov et al., 2013). Imperfect re-ligation of the overhanging DNA single strands at the sites of DSB during NHEJ is a plausible scenario explaining the generation of indels in this context. Also abnormalities in DNA modifying proteins contribute to DNA mutability. Members of the APOBEC protein family of cytidine deaminases, for instance, are highly expressed in several human cancers. APOBEC deaminase activity is involved in mRNA editing and contributes to innate immunity against retroviruses (Burns et al., 2013; Harris and Liddament, 2004; Roberts et al., 2013). Their overexpression gives rise to C>T substitutions in different genomic locations. Finally also aberrations in DNA polymerases can act as mutators. Mutational disruption of the proofreading exonuclease activity in the replicative DNA polymerases POLE and POLD1 has been associated with hypermutated,

MMR-proficient colorectal and endometrial cancers (Alexandrov et al., 2013; Cancer Genome Atlas, 2012; Palles et al., 2013).

Genetic mutations can be classified according to their contribution to cancer initiation and progression. Driver mutations are causally implicated in carcinogenesis and provide a selective advantage to a (pre-) cancerous cell. It is estimated that most human cancers have acquired 2-8 driver mutations. They can generate a gain of function in oncogenes, or they can inactivate a tumor suppressor gene, and therefore passively initiate or promote tumor formation. Very frequently mutated driver genes are the tumor suppressors TP53 and APC and the oncogene KRAS. p53 acts as a transcription factor, central in the transduction of a variety of stress induced signals to different anti-proliferative cellular endpoints. The DNA binding domain of p53 represents a hotspot of somatic mutations in human cancers. A significant fraction of these are C>T substitutions in a CpG sequence context, which probably derive from deamination of 5mC (Olivier et al., 2010; Zilfou and Lowe, 2009). Substitution mutations, which are not directly involved in carcinogenesis are called passengers. Compared to neutral mutations, passengers confer a selective advantage during cancer development. Classification into drivers and passengers requires detailed knowledge about the function of the affected gene, the genomic region or the precise mutation site and therefore proves to be difficult – also due to the fact that mutations not only affect protein functionality but also gene regulation. Such mutations are located in gene regulatory elements like the promoter region and alter the gene expression level. Promoter mutations are responsible for aberrant upregulation of TERT, the catalytic subunit of the telomerase, and a loss of SDHD (subunit D of the succinate dehydrogenase complex) expression in melanomas (F. W. Huang et al., 2013; Weinhold et al., 2014).

## **2.2 Epigenetic Instability**

In addition to genomic aberrations human cancer cells also display extensive alterations in the “epigenetic code”. The epigenetic code represents a second layer of genetic information that is clonally heritable and greatly contributes to the regulation of gene transcription and other DNA-templated processes including DNA repair and replication. Epigenetic mechanisms determine the accessibility of discrete chromosomal regions to enzymatic activities by regulating the structural compaction of the chromatin. Epigenetic information is laid down in the form of covalent chemical modifications at DNA bases and histone proteins. The epigenome represents a landscape of chromatin that shaped in response to intrinsic and extrinsic cues during embryonic development and cell differentiation to program cell-type specific gene expression and thus determine cell fate. Non-coding RNAs, including micro-RNAs (miRNAs), represent an additional epigenetic system with gene regulatory functions but are not subject of my PhD theses. Aberrations in all layers of epigenetic information are

widespread in human cancers and appear to occur already during early phases of carcinogenesis. The underlying molecular causes are diverse and often include transcriptional deregulation or mutational inactivation of enzymes involved in (i) the deposition or removal of epigenetic marks, (ii) “reading” of epigenetic marks or (iii) chromatin remodeling.

## 2.2.1 Epigenetic Mechanisms

### *Chromatin Structure*

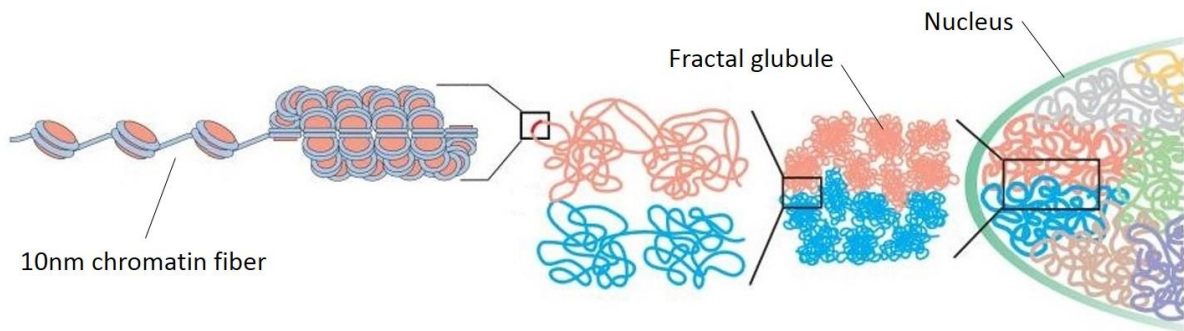
In the nucleus of eukaryotic cells, DNA exists in the form of a protein-DNA composite termed chromatin (Figure 2-2). The basic units of chromatin are the nucleosomes which consist of 145 base pairs of DNA wrapped around a histone octamer and the DNA linking these structures. The histone octamers are composed of two H3-H4 and two H2A-H2B histone dimers. The linker DNA is associated with the histone protein H1. The resulting chromatin fiber with a diameter of about 10nm, resembling “beads on a string”, can be further compacted. Results from mainly *in vitro* assays have suggested that the 10nm fiber becomes assembled in a regular helical structure containing 6-11 nucleosomes per turn generating a 30nm fiber. However, several observations made *in vivo*, using the chromosome conformation capture (3C) method, do not support the presence of a 30nm fiber but rather suggest that interphase chromatin exists in a dynamic, disordered and intertwined state with rather irregular spacing of nucleosomes (Hubner et al., 2013; Maeshima et al., 2010; Nishino et al., 2012). This form of chromatin organization is characterized by open and closed chromatin sections which appear as fractal globules at the chromosome level (see Figure 2-2). It enables dense chromatin compaction but still allows rapid conformational changes. In addition, the fractal globule chromatin structure describes interactions between genomic sites that are distant within a chromosome or located on different chromosomes. Interestingly, such interaction sites investigated by 3C were shown to be associated with chromosomal breakpoints in cancer (Fudenberg et al., 2011).

Closed and compacted chromatin regions are referred to as heterochromatin, in contrast to the open structure of the euchromatin. Heterochromatin is concentrated to centric, pericentric and telomeric regions and is enriched for repetitive DNA elements. It is usually highly DNA methylated and largely transcriptionally silent. Heterochromatic regions play an essential role in centromere function and chromosome end protection by telomeres and therefore contribute significantly to genome stability. Cancer cells are commonly characterized by the partial loss of heterochromatic structures. Aberrant opening of the chromatin structure results in the exposure of repetitive sequences, promoting erroneous recombination events, and thereby facilitating the accumulation of chromosomal aberrations. Furthermore, the loss of

heterochromatin can lead to the activation of transposons, mobile DNA elements that are capable to invade the genome and generate insertion mutations and chromosomal aberrations (Oberdoerffer and Sinclair, 2007; J. C. Peng and Karpen, 2008). Centric and telomeric heterochromatin is present throughout the cell cycle and in all cell types and is therefore termed constitutive heterochromatin. In contrast, facultative heterochromatin is established during embryonic development in a cell-type or gender specific manner, an example being the inactivated X chromosome in female genomes. Heterochromatic and euchromatic regions occupy different compartments in the nucleus. Heterochromatin clusters around the nuclear periphery and is associated with the nuclear lamina whereas euchromatin fills most of the internal nucleoplasm. Repressive heterochromatic and active euchromatic domains are demarcated by CTCF, an insulator protein that binds specific DNA sequences and prevents spreading of heterochromatin into neighboring euchromatic regions. Mutations and deletions affecting CTCF were found in breast and prostate cancers (Cuddapah et al., 2009; H. Shen and Laird, 2013).

The transition from a transcriptionally inactive and condensed chromatin state to an active and open chromatin state or vice versa requires the ATP-dependent activity of chromatin remodeling protein complexes. These enzyme complexes gate or close the access to the underlying DNA for transcription, replication and repair by reposition or eviction of nucleosomes or by replacement of canonical histones with histone variants (Petty and Pillus, 2013). In mammals, four major families of chromatin remodeling complexes can be distinguished according to their biochemical activity and subunit composition: SWI/SNF (switching defective/sucrose nonfermenting), ISWI (imitation SWI), NURD/Mi-2/CHD (nucleosome remodeling and deacetylation/chromodomain helicase DNA-binding) and the INO80 (inositol requiring 80) family. Mutations in chromatin remodeling proteins, especially in members of the SWI/SNF family, are frequently found in human cancers (Kadoch et al., 2013; Narlikar et al., 2013). However, little is known about the mechanisms by which these aberrations contribute to carcinogenesis (Dawson and Kouzarides, 2012).





**Figure 2-2: Chromatin organization in the mammalian nucleus according to the fractal globules model.** The 10nm chromatin fiber can form a dynamic and disordered fiber-shaped structure. This fiber is further organized as fractal globules at the chromosome scale. Fractal globules can intertwine and regions of chromosomal synapsis appear to present chromosomal breakpoints often observed in cancer. Individual chromosomes occupy different territories within the nucleus (Hubner et al., 2013).

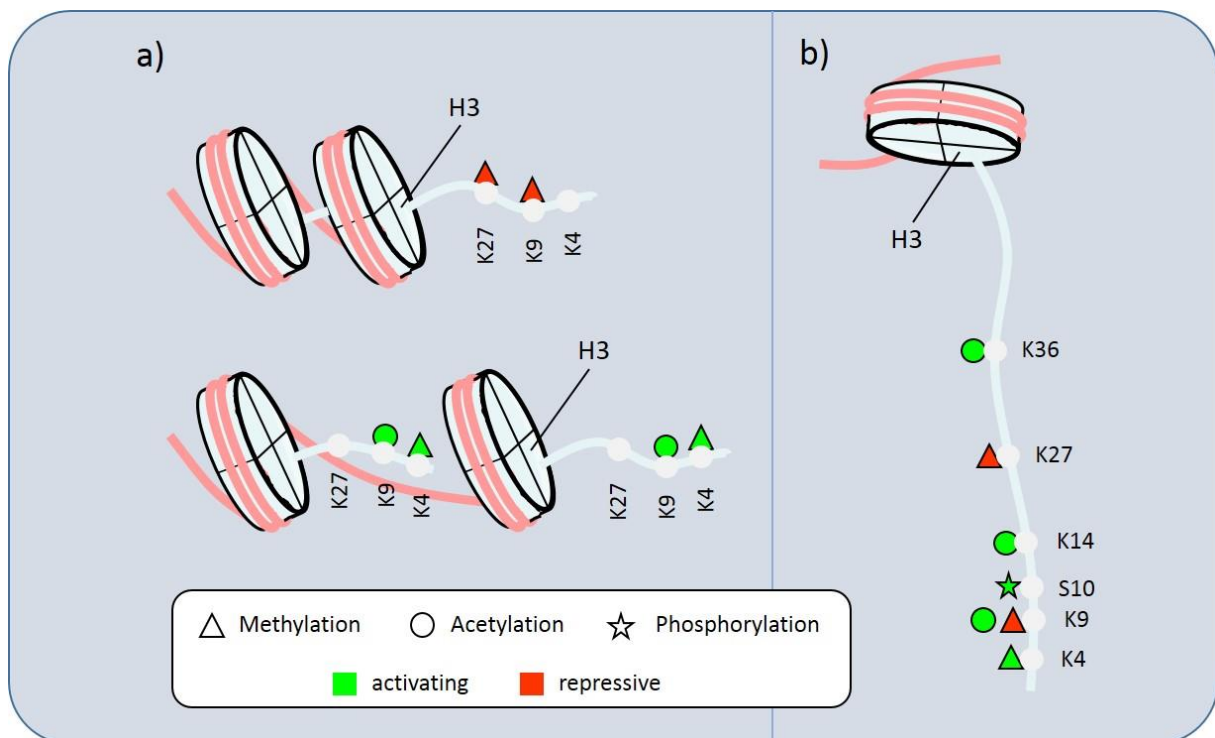
### *Histone Modifications*

Histones are largely globular proteins except for their unstructured N-termini which protrude outward from the nucleosomes and are available for interactions with other neighboring histones or non-histone proteins. Select amino acid residues within these N-terminal tails are subject to a variety of posttranslational modifications. These modifications can alter higher chromatin architecture either (i) by changing the charge of the modified amino acids in a way that affects the contacts between adjacent nucleosomes or the interaction between histones and DNA or (ii) by modulating the affinity of chromatin-associated proteins including chromatin remodeling proteins. Hence, histone modifications exert strong influence on the grade of chromatin condensation and transcriptional activity. Accordingly, hetero- and euchromatic regions are marked by different histone modifications. At least five different modifications are found in histones: lysine methylation and acetylation, serine, threonine and tyrosine phosphorylation, lysine ubiquitination and lysine sumoylation. (Kouzarides, 2007). Most prominent and best studied are modifications in the N-termini of the canonical histones H3 and H4. Acetylation of lysine residues, e.g. at H3K9 and H3K14, is dynamic and generally associated with an open chromatin structure and transcriptional activity. The role of lysine methylation appears to be more diverse. Methylation of H3K4 is associated with active transcription in euchromatic domains whereas H3K9 and H3K27 methylation is found in transcriptionally silenced regions. Lysine methylation can occur in mono-, di-, and trimethylated forms, which are associated with different chromatin states (Barski et al., 2007; Shilatifard, 2008). Histone acetylation and methylation are tightly controlled by the concerted

activity of histone acetyl/methyl transferases (HATs and HMTs) and histone deacetylases/demethylases (HDACs and HDMs). Histone modifications do occur and exert their effects on chromatin structure in combinations. For instance, phosphorylation of H3 serine 10 can be a marker for chromosomal condensation but, in combination with H3K14 acetylation it is found at actively transcribed gene loci (Prigent and Dimitrov, 2003; Strahl and Allis, 2000). Figure 2-3 summarizes well-established H3 modifications and their association with different chromatin states.

The combinatorial nature of histone modifications constitutes a “histone code”, which determines the final chromatin structural read-out. Chromatin remodeling complexes and chromatin architecture proteins like HP1 belong to the readers of this code (Yun et al., 2011). The heterochromatin protein HP1 recognizes and binds to methylated H3K9 in heterochromatic regions but can also be found in euchromatin where it contributes to gene regulation. HP1 contributes to the stability and functioning of telomeres and centromeres. Down regulation of HP1 expression is found in several human malignancies including leukemia and thyroid carcinomas and correlates with cancer progression, probably due to an increased chromosomal instability and alterations in gene expression (Dialynas et al., 2008). Similarly, histone modifying enzymes themselves are targeted to specific chromatin loci via specific pre-established histone modifications, further demonstrating the functional cooperation of different histone modifications. Polycomb group (PcG) and trithorax group (trxG) proteins are central to the maintenance of facultative hetero- and euchromatic domains, established during development. They form multimeric protein complexes which can contain histone readers and writers. The polycomb repressive complexes PRC1 and PRC2 recognize and deposit the repressive H3K27 trimethylation mark and remove activating H3K4 methylation (Lanzuolo and Orlando, 2012). On the opposite side, trxG proteins comprise H3K4 methyltransferases like MLL and chromatin remodeling factors (Schuettengruber et al., 2007). EZH2, the H3K27 methyltransferase component of PRC2 was originally thought to be oncogenic because both gain-of function mutations and overexpression were found in several human cancers. However, EZH2 loss-of-function mutations found in predominantly myeloid malignancies suggest that PRC2 can also act as a tumor suppressor (Gibbons, 2005; H. Shen and Laird, 2013).

In addition to mutation and transcriptional deregulation of histone modifying enzymes, translocation-generated gene fusions involving transcription factors, nuclear receptors or other histone writers can be found in cancers. Such alterations can result in mistargeting of the histone writing activity and, thus, in aberrant chromatin changes. The fusion between DNA-binding domain of the trithorax H3K4 methyltransferase MLL and histone acetyltransferase CBP, for instance, results in the aberrant activation of developmental HOX genes and thereby imparts stem cell-like properties (Gibbons, 2005; H. Shen and Laird, 2013).



**Figure 2-3: Activating and repressive histone marks in histone H3.** The N-terminus of histone H3 is subject to various post-translational modifications. a) Lysine acetylation generally marks active and open chromatin structure. Whereas, lysine methylation at K4 is associated with the active state, K9 and K27 methylation is present in transcriptionally silenced regions. b) Summarizing graph, illustrating well-established H3 modifications, their target amino acid residue and their function in transcription regulation.

### *Regulatory Non-coding RNAs*

Recent whole transcriptome sequencing approaches revealed that mammalian genomes are almost fully transcribed although only a very small proportion is in fact protein encoding (about 2%) and subsequently translated. The remaining “non-coding” RNAs (ncRNAs) are involved in numerous genomic processes including gene regulation, formation of heterochromatin and the maintenance of genomic stability (Amaral et al., 2008). Regulatory ncRNAs include the micro-RNAs (miRNAs) and the long non-coding RNAs (lncRNAs). The normally 22 nucleotides long miRNAs mediate post-transcriptional silencing by controlling the translation of mRNA. They are estimated to regulate the translation of more than 60% of the human genes. Accordingly, miRNAs are involved in the regulation of many biological processes like cell proliferation, differentiation and development. Some miRNAs regulate individual target genes,

whereas others act as master regulators implicated in expression regulation of hundreds of genes. miRNAs emerge either from separate transcriptional units or embedded within the introns of protein coding genes. Their biogenesis is multistep process in which the primary transcript becomes trimmed to eventually form the RNA-induced silencing complex (RISC) together with members of the Argonaut protein family which harbor an endonuclease activity. RISC is targeted to miRNA-complementary sequences in mRNAs usually located in their 3'UTR. Translation of the targeted mRNA is repressed either by mRNA degradation or inhibition of translation initiation (Jansson and Lund, 2012). A multitude of miRNAs are transcriptionally deregulated in human cancers due to promoter DNA methylation, mutation, aneuploidy or secondary to transcription factor dysregulation. Depending on their target genes miRNAs can function as tumor suppressors or oncogenes (Croce, 2009). The transcriptional silencing of miR-200 miRNA family members observed in lung, breast and bladder cancers results in a down-regulation of the cell adhesion molecule E-cadherin and consequently in epithelial-to-mesenchymal transition, a process which is essential for metastasis formation (Davalos et al., 2012). lncRNAs form a heterogeneous group of transcripts more than 200 nucleotides long. lncRNAs are involved in the regulation of chromatin state by the recruitment of chromatin remodelers and histone modification writers (Esteller, 2011). Hundreds of lncRNAs are expressed from the human HOX loci in a temporally and spatially concerted manner. HOTAIR is one of the best studied lncRNAs and emerges from the mammalian HOXC cluster. HOTAIR recruits PRC2 and the histone H3K4 demethylase LSD1 to the HOXD locus, resulting in its transcriptional repression. Overexpression of HOTAIR was observed in breast cancer leading to a genomewide re-targeting of PRC2 and altered H3K27 methylation and gene expression (R. A. Gupta et al., 2010).

### **2.2.2 DNA methylation**

The covalent modification of DNA by methylation of the C5 in cytosine bases (5mC) is the best studied mechanism of epigenetic chromatin regulation. Mammalian DNA methylation occurs almost exclusively in the CpG sequence context. Non-CpG methylation is common in plants but is found only in pluripotent stem cells in mammals. Due to the increased spontaneous deamination and mutation rate of methylated cytosine, evolution has depleted CpG dinucleotides from large parts of the mammalian genome (Colot and Rossignol, 1999). They are 5-10 fold underrepresented compared to what would be expected from the frequency of C and G nucleotides. The vast majority of CpG dinucleotides in mammals are methylated (~70% in humans) and associated with heterochromatic chromatin domains (Ehrlich et al., 1982). A small fraction of CpG sites (~7% in humans) occurs clustered forming CpG islands (CGI; observed versus expected ratio > 0.6). CGIs are found in promoters of about 60% of the human

genes including virtually all ubiquitously expressed housekeeping genes (Jeltsch, 2002). Although the great majority of CGIs appear to be unmethylated at all developmental stages and in all tissue types, methylation of a small but substantial fraction of CGIs associated with gene regulatory regions occurs in a tissue specific manner and effects downregulation of gene expression (Bird, 2002). Methylation of cytosine at the C5 position is catalyzed by DNA methyltransferases (DNMTs). In mammals there are three catalytically active DNMTs: DNMT1, DNMT3A and DNMT3B. All of them use the co-factor S-adenosylmethionin (SAM) as a donor for the methyl group. DNMT1 constitutes the key maintenance DNMT with a preference for hemimethylated CpG substrates. During DNA replication, the nascent newly synthesized DNA double-helix exists in a hemimethylated state, in which only the parental strand carries the methylation marks. DNMT1 faithfully copies the methylation pattern from the parental- to the daughter-strand to generate symmetrically methylated DNA strand, hence providing for the genetic inheritance of the parental methylation pattern. Interestingly, both biochemical and genetic data indicate that DNMT1 is activated to methylate unmodified CpG sites in the presence of fully methylated DNA and therefore might also play a role in *de novo* methylation (Fatemi et al., 2001). The stimulation of DNMT1 by methylated DNA indicates that heavily methylated regions might activate DNMT1 to promote spreading of methylation into neighboring unmethylated sequences (Turker, 1999). DNMT3A and DNMT3B exhibit no preference for hemimethylated DNA and are highly expressed during early embryonic development in the course of which they establish cell-type specific methylation patterns *de novo* but also may contribute to the maintenance of these patterns (P. A. Jones and Liang, 2009). The catalytically inactive DNMT3L interacts with DNMT3A and DNMT3B and might play a regulatory role in genomic imprinting (Hata et al., 2002).

### *Sites of DNA Methylation*

DNA methylation is generally associated with transcriptional repression and condensed chromatin states. In this capacity, DNA methylation contributes to the establishment and maintenance of tissue-specific gene expression, X-chromosome inactivation, suppression of transposable elements and the general landscape of heterochromatic structures. Accurate DNA methylation is thus essential for normal development and chromatin organization (Bird, 2002). Two mechanisms have been described by which DNA methylation mediates transcriptional repression. First, DNA methylation was proposed to generate a physical barrier for the binding of transcription factors such as E2F and CREB or the insulator protein CTCF, the DNA target sequence of which contain a CpG site (Bell and Felsenfeld, 2000; Griswold and Kim, 2001; Mancini et al., 1998). However, not all transcription factors prefer the unmethylated sequences for binding (Kulis and Esteller, 2010). The second mechanism

features the recruitment of methyl-CpG binding proteins (MBPs), DNA methylation readers that associate with several histone modifying enzymes. Five MBD proteins, characterized by the presence of methyl-CpG binding domain (MBD), have been identified: MeCP2, MBD1, MBD2, MBD3 and MBD4. Whereas MBD4 also has an active thymine DNA glycosylase domain and, thus, acts in the base excision repair (BER), MeCP2, MBD2 and MBD3 are associated with repressive histone modifying proteins like HDACs and H3K9 methyltransferases. Further, MBD2 and MBD3 are components of NURD/Mi-2 chromatin remodeling complexes which eventually mediate the repressive effect of underlying DNA methylation (Bogdanovic and Veenstra, 2009). Interestingly, it was found that HDAC1 and HDAC2 (components of the NURD/Mi-2 complex) and EZH2 (H3K27 methyltransferase in the polycomb repressive complex PRC2) interact with DNMTs. This bidirectional interdependence of histone modifications and DNA methylation is thought to assure transcriptional repression of a certain DNA region through replication and mitosis (D'Alessio and Szyf, 2006).

DNA methylation in the intergenic space can influence gene expression through enhancer regulation as for example in the context of genomic imprinting. Imprinted genes are expressed only from one parental allele. They usually reside within clusters that allow their organization by common imprinting control regions (ICRs). These ICRs have a high CpG density and their methylation state differs according to the parent-of-origin of the respective allele. At the H19/IGF2 imprinted locus, for instance, methylation of the ICR in the paternal allele which is located in between the non-coding RNA H19 and the growth factor IGF2, prevents binding of the insulator CTCF. In consequence, a downstream enhancer element can associate with the IGF2 promoter and drive its expression while H19 is silenced. In the maternal allele, however, the ICR is unmethylated and the bound CTCF blocks interaction between the enhancer and IGF2, allowing it to associate with the H19 promoter instead (Lim and Maher, 2010).

Finally, DNA methylation occurs within gene bodies where it has variable effects on gene expression and alternative splicing (Jjingo et al., 2012; Maunakea et al., 2013). Particularly interesting seems to be the effect on splicing as it involved the recruitment of MeCP2 and HDACs to sites of alternative splicing (Maunakea et al., 2013). The functions of intragenic methylation remains only poorly understood.

### *DNA Methylation and Aging*

It is well-documented for specific genomic loci that DNA methylation gradually changes over age (methylation drift). In general, methylation at promoter CGIs increases whereas intergenic methylation, including repetitive genomic elements declines with age. These age-dependent methylation changes are likely to reflect imperfect maintenance of methylation marks during

DNA replication (Issa, 2014). Accordingly, the highly proliferative population of tissue stem cells display an epigenetic mosaicism with some aged stem cells showing minor methylation changes while others having the highest methylation drift (Shibata, 2009). The stem cell-related methylation changes in aging tissues may affect the differentiation potential and epigenetic plasticity of these cells (Issa, 2014). It was shown that target regions of polycomb repressive complexes are prone to become hypermethylated over age (Rakyan et al., 2010). Interestingly, PRC2 target genes which include several tumorsuppressors are also predisposed to aberrant hypermethylation in cancer (Ohm et al., 2007; Schlesinger et al., 2007; Widschwendter et al., 2007). These observations indicate that methylation abnormalities in cancer may partially be targeted and that they occur in an age-dependent manner. Indeed, a large proportion of DNA methylation abnormalities in cancer can be traced to aging effects (Toyota et al., 1999). Furthermore, age represents the major risk factor for cancer. Remarkably, it was reported that lifestyle-related risk factors for colorectal cancer like smoking and overweight can influence the rate of methylation drift in the colon (Noreen et al., 2014). Chronic inflammation was found to be associated with substantial methylation drift in normal tissues from colon, stomach, lung and liver. Possibly, this is due to stimulation of stem cell proliferation for tissue regeneration and thereby acceleration of the drift (Issa, 2014).

### *Aberrant DNA Methylation and CIMP in Cancer*

Aberrant DNA hypermethylation of gene promoter associated CGIs is common in human cancers. Affected genes are usually downregulated, which can be of direct relevance for tumorigenesis if tumor suppressors are involved. For instance, CDKN2A (p16) and CDKN2B (p15), both encoding regulators of cell cycle progression, undergo aberrant methylation in several cancer types. Other cancer-specifically methylated genes control cell adhesion, cell survival or DNA repair. Most prominent are MLH1, essential for mismatch repair, BRCA1, involved in DSB repair and MGMT (O-6-methylguanine methyltransferase), which removes the methyl moiety from naturally occurring or mutagen-induced (alkylating agents) O6-methylated guanine (Kulis and Esteller, 2010). The functional loss of DNA repair enzymes can promote genetic instabilities like MSI, in the case of MLH1, or G:C to A:T mutability when MGMT is inactivated. Also promoter regions of ncRNAs can be target of cancer specific hypermethylation and transcriptional silencing. For instance members of the miR-200 family of miRNAs are frequently hypermethylated in colon, breast and lung cancers. Investigation of genome-wide methylation patterns revealed that a significant proportion of cancer-specifically hypermethylated CpG sites reside within regions adjacent to CGIs called CGI-shores (<2kb around CGIs). Cancer-specific methylation changes in these CGI-shores also affect

expression of associated genes and may result from the loss or shift of CGI-methylation boundaries (Hansen et al., 2011; Irizarry et al., 2009).

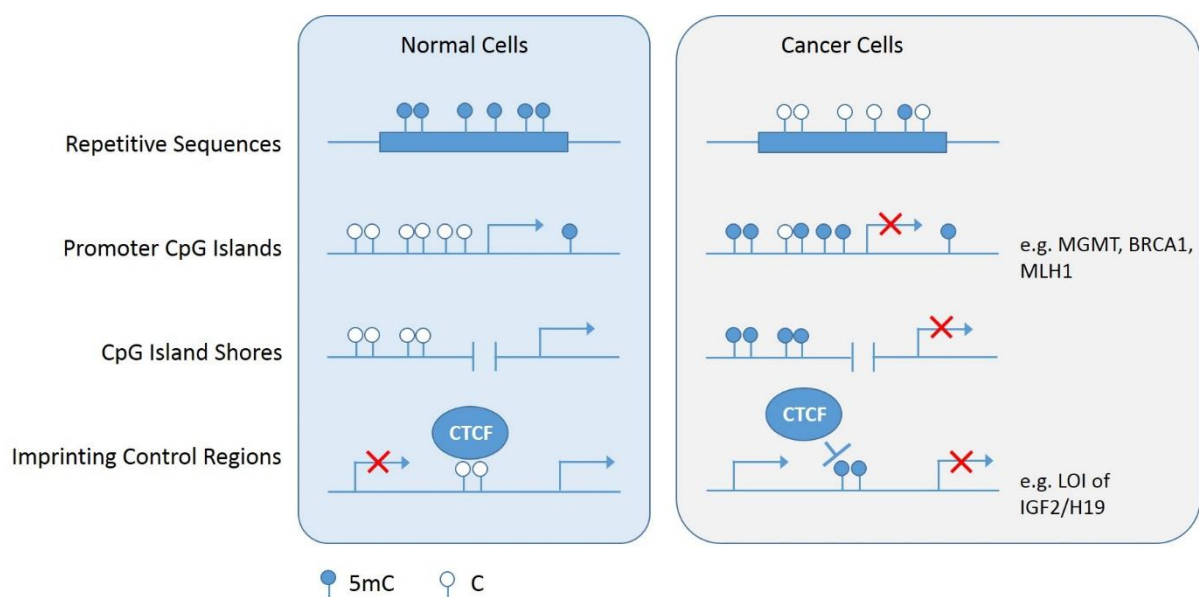
Subgroups of colon, liver, gastric, ovarian and lung cancers but also leukemias and gliomas feature an exceptionally high degree of gene promoter hypermethylation, and were therefore referred to as displaying a CGI-methylator phenotype (CIMP). CIMP accounts for about 10-20% of the respective cancers (Noushmehr et al., 2010; Teodoridis et al., 2008). The molecular mechanisms underlying this particular epigenetic phenotype are not fully understood and probably differ between types of CIMP cancers. For gliomas and leukemias, there is good evidence that the loss or inhibition of DNA-demethylation activity (see chapter 2.2.3 DNA demethylation) is causal for the hypermethylation phenotype (Figueroa et al., 2010; Turcan et al., 2012). The CGI methylator phenotype in colorectal cancer usually implicates the hypermethylation and transcriptional silencing of the MLH1 promoter. The functional loss of MLH1 leads to MSI and a concomitant characteristic mutator phenotype. This represents an illustrative example for the coincidence of cancer-related genetic instabilities. The different forms of genetic and epigenetic instabilities introduced in this chapter exist not mutually exclusive but rather appear to cooperate to increase the selectable variability.

In addition to focal CGI hypermethylation a global loss of DNA methylation is observed in a large variety of human cancers. This loss of methylation affects highly methylated and silenced heterochromatic regions in which satellite sequences, repetitive genomic sequences and transposable elements reside. Hypomethylation and increased accessibility of these elements have potential impact on chromosomal stability due to an increased rate of transposon-related mutagenesis and recombination events. DNMT1 knockdown experiments in animal models and studies of the molecular phenotype of ICF (Immunodeficiency, Centromere instability and Facial anomalies syndrome), a genetic syndrome caused by mutation of DNMT3B clearly show that the loss of DNMT activity is accompanied by the global loss of methylation, associated with increased chromosomal instability. However, deregulation of catalytically active DNMTs appears not to be causal for hypomethylation in cancer (Wilson et al., 2007). Still, catalytically inactive splice variants of DNMT3B, which negatively regulate methyltransferase activity have been linked to hypomethylation in leukemia and liver cancer (C. A. Gordon et al., 2013; Kulis and Esteller, 2010). There is evidence supporting that hypomethylation occurs secondary to altered histone modifications (e.g. H3K9 methylation) caused by deregulated histone modifying enzymes. Unlike global loss of methylation, hypomethylation of individual gene promoter CGIs is rather uncommon. Only few genes, some of which are oncogenes, are reported to be reactivated by aberrant hypomethylation of their promoter. For example, the CDH3 (P-cadherin) gene was reported to be overexpressed in colon and breast cancers leading to decreased cell polarity, promoting cell mobility and invasion. Focal hypomethylation in CGI-shores appeared to be more frequent in a series of human cancers. Genes overexpressed due



to CGI shore hypomethylation include several mitosis and cell-cycle-related genes (Hansen et al., 2011).

Aberrant hypo- and hypermethylation in cancer can be related to imprinting control regions (ICRs). Loss of imprinting (LOI) i.e. the biallelic expression or repression of imprinted genes is considered a hallmark of human cancers. A well-studied example is the H19/IGF2 locus where hypermethylation of the ICR results in the overexpression of IGF2 (see paragraph *Sites of DNA methylation*), which functions as an autocrine and paracrine growth factor in many cancers, and thereby promotes cell proliferation (Lim and Maher, 2010). Figure 2-4 summarizes the different forms of cancer-related aberrant methylation events.



**Figure 2-4: DNA methylation patterns in normal and cancer cells.** Repetitive sequences residing in heterochromatic regions are generally methylated in normal cells. Global loss of methylation in cancer cells leads to chromosomal instability and activation of transposable elements. Promoter CpG islands typically are unmethylated in normal cells whereas they can become hypermethylated in cancer cells, leading to transcriptional repression. Examples of genes affected are shown on the right. Similar patterns are seen in CpG island shores, located upstream of promoters. Loss of imprinting (LOI) due to aberrant methylation of imprinting control regions can block binding of the insulator CTCF which can result in the overexpression of oncogenes. A prominent example is LOI of the IGF2/H19 locus (Adapted from Varela-Rey et al., 2013).

### 2.2.3 DNA Demethylation

In terminally differentiated somatic cells, the cell type-specific methylation patterns established during embryogenesis are largely stable. Indeed, DNA methylation is thought to primarily

provide stability to the epigenome and thereby the somatic cell identity. However, particularly during early embryonic development genome-wide methylation patterns change profoundly and this includes the erasure of existent methylation marks. The methylation of the 5' position of cytosine has long been thought to be irreversible due to the strength of the established C-C bond. Methylation dilution by inhibition of methylation maintenance during DNA replication was considered the only way of demethylation (passive DNA demethylation). However, investigation of DNA methylation in the preimplantation mouse embryo suggested the existence of an active demethylation mechanism. Both the maternal as well as the paternal genome undergo global erasure of DNA methylation after fertilization in the context of restoring totipotency, the ability to generate any cell type. Whereas methylation levels of the maternal DNA decline gradually in a replication dependent manner (dilution), the paternal pronucleus loses global methylation rapidly within the first round of DNA replication (Jenkins and Carrell, 2012). This indicates the involvement of an enzymatic component actively removing the methyl mark. Numerous DNA repair pathways, DNA editing enzymes and even DNMTs were proposed to be involved in active DNA demethylation (Ooi and Bestor, 2008). However, for many mechanisms their role in demethylation seemed limited to the specific biological systems investigated. Eventually, the discovery of the TET family of DNA dioxygenases shed light on methylation dynamics during early development. TET proteins (TET1, TET2 and TET3) can oxidize 5mC to 5-hydroxymethyl cytosine (5hmC) which was shown to function as an intermediate of DNA demethylation (J. U. Guo, Su, et al., 2011; Ito et al., 2011; Tahiliani et al., 2009). 5hmC is not maintained by DNMT1 and lost with ongoing DNA replication (Valinluck and Sowers, 2007). Indeed, recent studies confirmed that such a semiactive mechanism mediated by TET3 contributes to the rapid loss of 5mC in the male pronucleus (Gu et al., 2011; Inoue and Zhang, 2011; Wossidlo et al., 2011). However, the observation that extensive demethylation occurs also in post-mitotic neurons suggests the existence of a truly active mechanism which completely restores the unmodified cytosine independent of replication (J. U. Guo, Su, et al., 2011). Two major pathways for the active replacement of 5hmC with unmodified C were proposed, both postulating the further chemical modification of 5hmC and the subsequent replacement of this modified base by the base excision repair (BER) pathway. The first pathway suggests deamination of 5hmC by AID/APOBEC cytidine deaminases, which results in the generation of 5-hydroxymethyl uracil (5hmU) (Cortellino et al., 2011). 5hmU-G base pairs are substrate for the DNA glycosylases TDG and SMUG which excise the base from the DNA by cleavage of its N-glycosidic bond, thus producing an abasic site (AP-site, apurinic/aprimidinic site). This represents the initial step of the BER pathway in the course of which the excised base becomes replaced with unmodified C. Several lines of evidence, however, challenge the concept of deamination-mediated active demethylation and favor a second pathway of active demethylation (Nabel et al., 2012; Rangam et al., 2012). This

pathway bases on the observation that TET proteins are capable to further oxidize 5hmC iteratively to 5-formyl cytosine (5fC) and 5-carboxy cytosine (5caC), two bases efficiently processed by TDG but apparently no other DNA glycosylase (Ito et al., 2011; Maiti and Drohat, 2011). Indeed, the loss of TDG leads to the accumulation of 5caC in embryonic stem cells (He et al., 2011). TDG was originally identified as an enzyme releasing thymine from G-T mismatches and thereby proposed to counteract C→T transition mutation by 5mC deamination (Cortazar et al., 2007). But in contrast to other DNA glycosylases and consistent with a non-redundant function in TET1 mediated active DNA demethylation, TDG is essential for embryonic development and mutant embryos feature epigenetic abnormalities including aberrations in DNA methylation (Cortazar et al., 2011; Cortellino et al., 2011). The oxidative pathway in which TDG acts on TET-generated 5hmC derivatives is currently the only biologically and biochemically validated mechanism for active DNA demethylation (Kohli and Zhang, 2013).

### *DNA Demethylation and Epigenetic Plasticity*

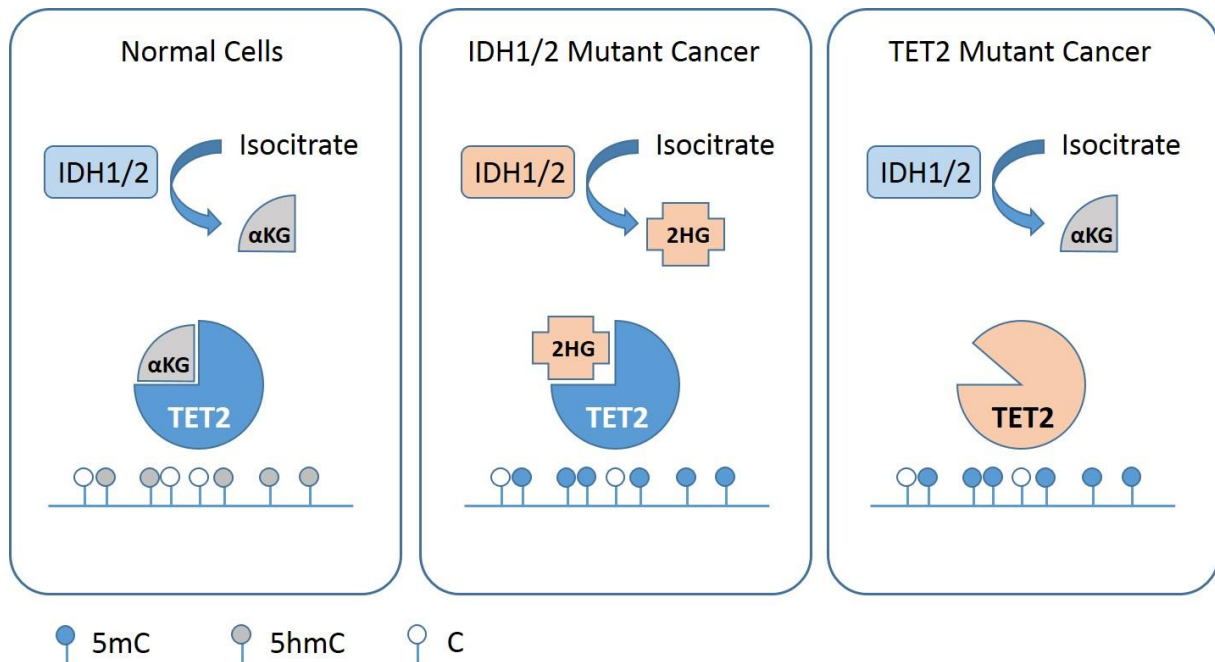
5hmC is enriched in embryonic stem cells (ESC) in which TET1 and TET2 are highly expressed and active. 5hmC in mouse ESCs is present mainly at transcription start sites (TSS) and in gene bodies. Whereas, TET2 appears to be associated with 5hmC in gene bodies, TET1 is rather related to hydroxymethylation at the TSS (Y. Huang et al., 2014; Williams et al., 2011). TET1 seems to bind preferentially near the TSS of CpG-rich gene promoters, including a significant fraction of polycomb target genes. During ESC differentiation the global 5mC content decreases concomitantly with a decline in TET1 and TET2 expression (Szwagierczak et al., 2010). It was reported that TET1 and TET2 transcription is regulated by the master regulator of ESC self-renewal, OCT4, and that they are themselves directly involved in the regulation of key ESC transcription factors like NANOG (Ito et al., 2010; Koh et al., 2011). These results suggest a major function of TET proteins in the ESC state. Koh and colleagues showed that TET1 depleted ESC display an altered differentiation potential *in vitro*, indicating a role of TET1 in maintenance of ESC pluripotency (Koh et al., 2011). However, TET1 and TET2 knockout mice develop normal and are fertile, an observation that is difficult to reconcile with a premature loss of pluripotency (Dawlaty et al., 2011). Although, the exact function of TET proteins in ESCs is poorly understood and requires further investigation, several studies with induced pluripotent stem cells (iPSCs, see Apostolou and Hochedlinger, 2013 for review) indicate a central role of TET proteins during cell reprogramming and suggest that TET function is associated with cell state plasticity (Gao et al., 2013). Interestingly, TET2 depletion in mice results in an augmented hematopoietic stem cell (HSC) population which, however, displays a skewed differentiation potential indicating that TET proteins exert important functions in

somatic stem cells as well (Ko and Rao, 2011; Nakajima and Kunimoto, 2014). Unlike most differentiated and post-mitotic somatic cells, neuronal cells in the adult hippocampus show a remarkable level of DNA methylation plasticity. Synaptic connections between neurons in the hippocampus are reorganized in the context of memory generation. Activity in these neurons can alter their DNA methylation profile *in vivo* including gain and loss of methylation. This activation-induced demethylation is likely to be active (independent of replication) and TET1-dependent (J. U. Guo, Ma, et al., 2011). In summary, TET-mediated active DNA demethylation appears to be associated with an increased epigenetic and phenotypic plasticity required during early embryonic development, in tissue homeostasis as well as in neuronal activity.

### *Defective DNA Demethylation in Cancer*

The first identified, cancer-related alteration involving TET proteins was the gene fusion between TET1 and the H3K4 methyltransferase MLL present in a subset of leukemias (Ono et al., 2002). Indeed, TET1 and the subsequently identified other members of the protein family were named with reference to this gene fusion which resulting from translocation between chromosome 10 q22 (TET1 locus) and chromosome 11 q23 (TET for Ten-Eleven-Translocation). Later, mutations and deletions affecting TET2 were found in ~25% of myeloid leukemia patients as well as in one third of patients with lymphomas (Delhommeau et al., 2009; Langemeijer et al., 2009; Quivoron et al., 2011). The functional loss of TET2 in leukemias is accompanied by decreased global 5hmC and the accumulation of 5mC, thus reflecting a hypermethylation phenotype (Figuroa et al., 2010; Ko et al., 2010). Interestingly, TET2 mutations in leukemia occur mutually exclusively with mutations in IDH1 and IDH2 (Isocitrate dehydrogenases). IDH enzymes normally catalyze the oxidative decarboxylation of isocitrate to  $\alpha$ -ketoglutarate ( $\alpha$ -KG) which, along with Fe(II), represents a cofactor for TET enzymes. Mutant IDH1/2 acquire a neomorphic enzymatic activity catalyzing the reduction of  $\alpha$ -KG to 2-hydroxyglutarate (2-HG) which directly inhibits TET2 (Figuroa et al., 2010; W. Xu et al., 2011). Similar to TET2-mutant leukemias, IDH-mutant leukemias are also characterized by genomewide hypermethylation, impaired hematopoietic differentiation and an enlarged stem cell compartment (Figuroa et al., 2010). Subsequently, it was shown that the same neomorphic mutation in IDH1 represents the molecular basis for the CGI-methylator phenotype in gliomas (G-CIMP) (Turcan et al., 2012). The TET2 and IDH1/2 mutation-related pathways of CGI-hypermethylation are illustrated in Figure 2-5. Irrespective of the association with specific hypermethylation phenotypes, reduced TET protein expression and the concomitant global loss of 5hmC appears to be a common feature of several human solid tumors (Yang et al., 2013).

Similar to impaired TET activity, the loss of TDG has potential impact on DNA methylation patterns. However, cancer-related aberrations affecting TDG are rare. Somatic mutation of TDG have so far only been described in a single rectal cancer which featured a C→T mutator phenotype (Vasovcak et al., 2012). Loss of TDG expression was reported in pancreatic cancers and myelomas and was associated with deletion of the TDG locus or the aberrant hypermethylation of its promoter, respectively (B. Peng et al., 2006; Yatsuoka et al., 1999).



**Figure 2-5: Defects in TET-mediated DNA demethylation in cancer.** (Left) In normal cells the IDH1 and IDH2 enzymes catalyze the oxidative decarboxylation of Isocitrate to  $\alpha$ -ketoglutarate ( $\alpha$ KG) which serves as a cofactor for TET DNA dioxygenases. TET proteins prevent aberrant methylation events by hydroxylation of 5mC to 5hmC which represents a demethylation intermediate. (Middle) Mutations in IDH1/2 result in a neomorphic catalytic activity reducing  $\alpha$ KG to 2-hydroxyglutarate (2HG) which inhibits TET activity. This leads to an accumulation of 5mC and genome-wide hypermethylation. (Right) Loss-of-function mutations in TET2 found in myeloid leukemia and in lymphoma similarly generate a methylator phenotype. (Adapted from Cimmino et al., 2011)

### 2.3 Colorectal Cancer

Colorectal cancer (CRC) is the third most common form of cancer, accounting for approximately 10% of all cancer cases (World Cancer Report 2014. World Health Organization., 2014). CRCs are usually adenocarcinomas with glandular origin. Numerous factors have been shown to be associated with increased risk of CRC. These include heritable predisposition, age, chronic inflammation, and several lifestyle factors like diet, smoking and

alcohol intake. Most CRCs arise sporadically without any familial background. Only 10-15% of all CRCs are etiologically linked to a hereditary predisposition. The most significant forms of familial CRCs are the hereditary non-polyposis colorectal cancer (HNPCC, Lynch-syndrome) and the familial adenomatous polyposis (FAP) which together account for 3-5% of all CRCs. Most HNPCC patients harbor heterozygous germline mutations in mismatch repair genes including MLH1 (33%), MSH2 (31%), MSH6, PMS1 and PMS2. Accordingly, HNPCC related tumors are usually characterized by MSI. A hallmark of FAP is the development of myriad adenomatous polyps in the colon from which some will inevitably progress and form carcinomas. FAP is caused by germline mutations in the APC tumor suppressor gene. APC (adenomatous polyposis coli) is a negative regulator of the canonical Wnt signaling pathway.

Wnt-signaling is one of the fundamental pathways regulating cell proliferation, cell polarity and cell fate determination in embryonic development and tissue homeostasis. Canonical Wnt-signaling functions by regulating the amount of the transcriptional coactivator  $\beta$ -catenin, which controls developmental gene expression programs. In this pathway secreted Wnt ligands bind to a frizzled family receptor and its coreceptor LRP5 or LRP6 resulting in the activation of Dishevelled (DVL). In its active form DVL inhibits the  $\beta$ -catenin destruction protein complex consisting of APC, Axin, protein phosphatase 2A (PP2A), glycogen synthase kinase 3 (GSK3) and casein kinase 1 $\alpha$  (CK1 $\alpha$ ).  $\beta$ -catenin, normally targeted to ubiquitination by the destruction complex and consequently degraded by proteasomes, accumulates and translocates to the nucleus where it interacts with transcription factors of the TCF/LEF family. The TCF/ $\beta$ -catenin complex regulates the expression of a multitude of Wnt-responsive genes by recruiting further co-activators or co-repressors, respectively (MacDonald et al., 2009). TCF/ $\beta$ -catenin promotes the expression of several oncogenes including MYC and genes involved in cell cycle regulation like Cyclin D. Also expression of the pluripotency factors NANOG and OCT4 can be induced by canonical Wnt signaling (Cole et al., 2008). Activity of the Wnt/ $\beta$ -catenin pathway is highly associated with the adult colonic stem cell compartment and is essential for its maintenance also in colorectal cancer. LGR5, a target gene of TCF/ $\beta$ -catenin is widely used as a reliable marker for intestinal stem cells (Becker et al., 2008; Schepers and Clevers, 2012; van de Wetering et al., 2002).

The molecular mechanisms underlying HNPCC and FAP demonstrate the central roles of mismatch repair and the Wnt-signaling pathway in the carcinogenesis of human CRC. In sporadic CRCs both pathways are commonly affected also in consequence of (epi-) genomic instabilities.

Another significant cellular signaling pathway in colorectal carcinogenesis is the extracellular signal-regulated kinase (ERK) pathway involved in the regulation of fundamental cellular processes such as growth, proliferation, differentiation, migration and apoptosis. The ERK

pathway belongs to the mitogen-activated protein kinase (MAPK) pathways and constitutes the best studied member of this extracellular signal-regulated kinase pathways. The ERK signaling pathway can be triggered by activation of transmembrane receptors tyrosine kinases (RTKs) which include growth factor receptors like EGFR (epidermal growth factor receptor) and FGFR (fibroblast growth factor receptor). Binding of the growth factors to the extracellular domain of their receptors results in the phosphorylation of the intracellular domains. The activated RTKs trigger the GTP-loading to Ras GTPases, which phosphorylate the Raf kinases. Activated Raf proteins in turn induce the phosphorylation of Mek kinases which eventually phosphorylate the ERK kinase (Katz et al., 2007). ERK phosphorylates a multitude of cytoplasmic proteins and after its translocation to the nucleus it also activates various transcription factors including the oncogenes MYC, FOS and JUN. The aberrant activation of the ERK pathway can be considered as a hallmark of human cancers and cancer-related abnormalities occur across the whole signaling cascade. Activating mutations in Ras genes (KRAS and NRAS), which impair their GTPase activity and render Ras in active, GTP bound state are found in a large variety of human cancers with an exceptionally high frequency. Mutations of Raf genes are mainly found in melanomas and CRCs and affect almost exclusively the B-RAF family member (Dhillon et al., 2007). The most common mutation of BRAF is a V600E change that implicates the constitutive activation of its kinase activity. KRAS mutations are more common in CRC than BRAF mutations and they occur mutually exclusive which is indicative of their common function in activating ERK signaling during cancer development. However, the BRAF V600E mutation in CRC is highly associated with the CGI-methylator phenotype (see chapter 2.3.1), whereas KRAS is not significantly correlated with epigenetic aberrations. This suggests that the oncogenic potential of BRAF and KRAS mutations is not exclusively related to activation of ERK. The over-activation of the ERK pathway by BRAF and KRAS mutations and the consequent hyper-proliferation trigger a cellular tumor suppressive response mechanism resulting in the senescence of the cells, the ceasing of cell division. The mechanism of oncogene-induced senescence (OIS) depends on the activation of p53. (Epi-) genetic aberrations affecting p53 function itself or affecting cofactors of p53 confer resistance to OIS and are common in CRC.

In 1990, Fearon and Vogelstein proposed a multistep molecular model for colorectal tumorigenesis. According to this model, different genetic driver aberrations accompany the development of discrete intermediates in the progression to cancer (Fearon and Vogelstein, 1990). These precancerous lesions include the aberrant crypt foci (ACF), the earliest detectable tissue anomalies, and different stages of adenomas, benign tumors of colon epithelium. Mutations and LOH of APC occur early in this ACF-adenoma-carcinoma sequence followed by mutations in KRAS (Haigis et al., 2008). Subsequent malignant transformation is driven by additional aberrations affecting the p53 tumor suppressor, the transforming growth

factor (TGF)- $\beta$  pathway and the phosphatidylinositol bisphosphate kinase (PIK3CA). PIK3CA acts in the PI3K/AKT signaling pathway mediating cell growth and proliferation signals from several growth factors including insulin-like growth factors (IGFs). Activating mutations in PIK3CA lead to the constitutive activation of this pathway and increased cell proliferation, survival and motility (Karakas et al., 2006). TGF- $\beta$  signaling exerts a dual role in tumorigenesis and cancer progression. As a potent inhibitor of cell proliferation TGF- $\beta$  signaling activity is blocked in early phases of colorectal tumorigenesis, for instance due to mutation of TGF- $\beta$  receptor II (TGF $\beta$ 2R) in MSI tumors or mutation of the SMAD TGF- $\beta$  signaling molecules. In later stages when tumor cells became resistant to TGF- $\beta$ -mediated growth inhibition they often upregulate their TGF- $\beta$  expression and secretion, which triggers increased tumor cell mobility and invasiveness and promotes angiogenesis (Blobe et al., 2000). Due to TGF $\beta$ 2R mutations, MSI tumors are less susceptible to the late metastasis promoting effects of TGF- $\beta$  and thus have usually a better prognosis (Pino et al., 2010). The multistep model for colorectal carcinogenesis, even with ongoing refinements due to the identification of other carcinogenic pathways, became a paradigm for solid tumor progression. It indicates that only a relatively small number of (epi-) genetic aberrations are required for cancerous transformation. Genomic and epigenomic instabilities are likely to be involved in the acquisition of these driver events.

### **2.3.1 (Epi-) Genetic Instabilities in Colorectal Carcinogenesis**

Based on the integrative analysis of genetic and epigenetic aberrations two molecular pathways of colorectal carcinogenesis can be distinguished: The CIN pathway and the CIMP/MSI pathway. Besides the different underlying genetic instabilities, tumors arising from these pathways are characterized by a different set of associated somatic mutations, by different degree of malignancy and also by different precursor lesions.

CIN is the most frequent form of genomic instability in CRC and is observed in 65%-70% of sporadic cancers. CIN occurs at very early stages of tumorigenesis. More than 90% of adenomas show allelic imbalances of individual chromosomes (Shih et al., 2001). However, the molecular basis of CRC-CIN is not well understood, yet. To identify gene mutations underlying colorectal CIN, Wang and colleagues searched for CRC-related mutations in human gene homologues corresponding to genes involved in CIN in *Drosophila melanogaster* and in budding yeast. They identified mutations in MRE11, an important factor of DSB repair by NHEJ, in DING, required for proper chromosome disjunction in yeast, and in a number of factors of a kinetochor-associated protein complex. Together, these mutations accounted for ~10% of the analyzed CIN cancers (Z. Wang et al., 2004). However, functional prove of the causal relationship between these mutations and CIN is still missing. Interestingly, mutations in APC and concomitant activation of the Wnt/ $\beta$ -catenin signaling pathway were reported to



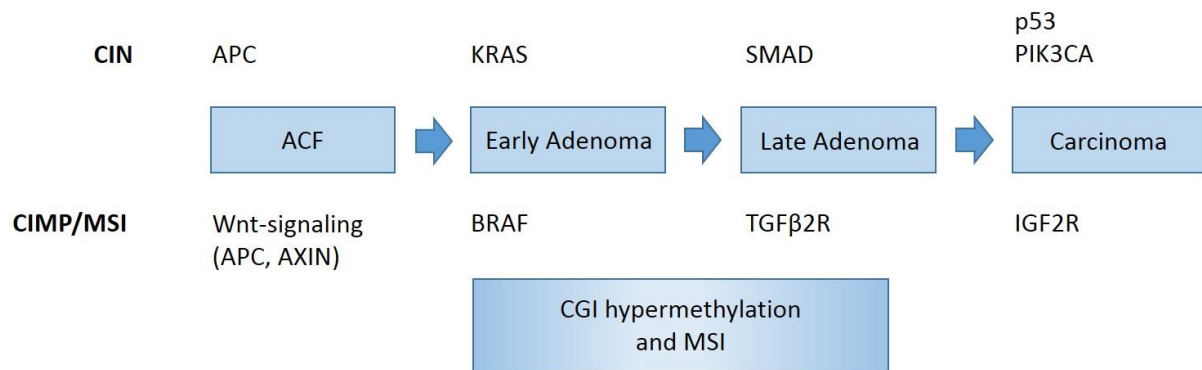
induce CIN *in vivo*, maybe due to upregulation of Axin, a component of the  $\beta$ -catenin destruction complex. Transcription of Axin is induced by TCF/ $\beta$ -catenin in the context of a negative regulatory circuit. Overexpression of Axin compromises the mitotic spindle assembly checkpoint and is thereby capable to induce CIN (Aoki et al., 2007; Hadjihannas et al., 2006). A characteristic set of chromosomal aberrations was recurrently observed in CRC-CIN. It includes several arm-level deletions like chromosome 5q encompassing the APC and the Mutated in Colorectal Cancer (MCC) gene loci. MCC is a cell cycle regulatory protein that induces cell cycle arrest in response to DNA damage and also acts as a repressor of  $\beta$ -catenin-dependent transcription (Fukuyama et al., 2008). Deletions of chromosome 17p segment which includes the p53 locus and chromosome 18q containing many candidate tumor suppressors are found in approximately 70% of CIN CRCs. These examples again illustrate the potential of CIN in inducing LOH, the loss of the functional allele in heterozygous mutations. On the other hand amplifications of genomic loci containing oncogenes like MYC at chromosome 8q are frequent in colorectal cancer (Sheffer et al., 2009). Although, CIN-CRCs are primarily characterized by genomic instability, it is important to note that epigenetic alterations are not absent in CIN-CRCs and include, for instance, focal hyper- and hypomethylation events (Sakai et al., 2014).

Microsatellite instability in sporadic CRCs is basically caused by aberrant hypermethylation and transcriptional silencing of the MLH1 promoter in the context of a genomewide CGI methylator phenotype (CIMP). Therefore, MSI and CIMP are mechanistically linked in CRC and form the second major pathway of sporadic colorectal carcinogenesis, accounting for 10%-15% of the cancers. The functional loss of MLH1 generates a characteristic mutator phenotype. TGF $\beta$ 2R as well as the mismatch repair genes MSH3 and MSH6 harbor single nucleotide repeats in their coding regions which are targets of MSI induced mutations. CRC-CIMP was first described 1999 by Toyota and colleagues, who found that a subgroup of CRCs shows simultaneous cancer-specific methylation of a small set of promoter-associated CGIs (Toyota et al., 1999). Since their initial marker panel included MLH1, they concluded that CIMP may be causal for MSI in CRC. Recent genome-wide methylation studies, however, revealed that CIMP is not restricted to a small number of CGIs but rather represents a genomewide phenomenon with a multitude of aberrantly hypermethylated CGIs (Ang et al., 2010; Hinoue et al., 2012; L. Shen et al., 2007; Yagi et al., 2010). These studies also introduced a third DNA methylation-related phenotype characterized by an intermediate degree of CGI methylation referred to as CIMP-low (CIMP-L) or CIMP2 in contrast to CIMP-high (CIMP-H) or CIMP1, respectively. Whereas, CIMP-H tumors were found to be highly associated with V600E BRAF mutations and MSI, the intermediate CIMP state was characterized by mutations in KRAS and rare MSI. In contrast to the CIMP-negative condition, both CIMP forms correlated negatively with p53 mutations. Since, CIMP-L methylation profiles occur also in CIN tumors, this

methylation phenotype appears not as distinctive as the CIMP-H/MSI pathway. It is currently unclear whether CIMP-L represents a discrete pathway of carcinogenesis, although its unique association with KRAS mutations suggests a separate underlying oncogenic mechanism. The molecular processes leading to the aberrant hypermethylation in CRC-CIMP are largely ambiguous. Overexpression of DNMT1 and DNMT3B was proposed to be involved in the generation of CIMP (Kanai et al., 2001; Noshro et al., 2009). However, a causal link between DNMT expression and CIMP-related hypermethylation has not been shown, yet. It was found that BRAF V600E mutations can induce MSI *in vitro* and *in vivo* indicating a mechanistic link between the BRAF V600E mutation and CIMP/MSI (Oikonomou et al., 2009; Rad et al., 2013). Results from our group indicate that a loss of TET1 expression is involved in aberrant CGI methylation in CRC-CIMP (see 4. Results). The observation that the BRAF V600E-induced transformation of mouse fibroblast cells leads to the downregulation of Tet1 expression substantiates the hypothesis that impaired demethylation rather than aberrantly active methylation is responsible for 5mC accumulation in CRC-CIMP (Kudo et al., 2012). This would coincide with different reports indicating that mutation or inhibition of TET proteins is causal for methylator phenotypes in leukemias and gliomas.

By and large, CIN and CIMP/MSI features of genomic instability arise in a mutually exclusive manner in CRC (Cancer Genome Atlas, 2012; Cheng et al., 2008), implicating separate underlying pathways of carcinogenesis, differing in specific molecular, clinical and histological aspects. For instance, due to MMR-deficiency CIMP/MSI tumors appear hypermutated compared to CIN cancers. Moreover, both pathways are associated with different driver mutation profiles: KRAS and p53 mutations are hardly found in tumors derived from CIMP/MSI which often harbor BRAF mutations, instead. However, in accordance to the critical role of the Wnt-signaling pathways in colorectal carcinogenesis, APC mutations are found in tumors from both pathways with a high frequency. Patients diagnosed for CIMP/MSI derived cancers are usually older than patients with CIN cancers. In addition, CIMP/MSI tumors are associated with the female gender and are mainly located at the right side of the colon (proximal colon). Finally, CIN- and CIMP/MSI-related tumors arise from different precursor lesions. CIN tumors develop from conventional adenomatous polyps and ACFs in which mutations in APC, KRAS and p53 can be already found. Tumors from the CIMP/MSI pathway arise from serrated adenomas, referring to the sawtooth-shaped infoldings of the surface of these polyps. Mutation of BRAF has been observed in serrated adenomas and even serrated ACFs and therefore represents a very early event in the CIMP/MSI pathway. Aberrant hypermethylation events can also be detected in serrated polyps such as methylation of CDKN2A (p16), a p53 stabilizing factor, and MCC (Leggett and Whitehall, 2010). Figure 2-6 illustrates the different molecular events associated with the CIN and CIMP/MSI oncogenic pathways.

In early stages of tumorigenesis genetic and epigenetic instabilities of both pathways contribute to the generation of a persistent hyper-proliferative phenotype. Further tumor progression and the transition to malignant stages i.e. the capability to invade other organs to establish metastasis require the induction of angiogenesis and acquisition of a reversible migratory cell behavior.



**Figure 2-6: The CIN and the CIMP/MSI pathway of colorectal carcinogenesis.** The two oncogenic pathways of colorectal cancer differ regarding the key (epi-) genetic driver aberrations. The CIN pathway (top) is initialized by mutations in APC and the concomitant over-activation of the Wnt/ $\beta$ -catenin signaling. Also in the CIMP/MSI pathway (bottom) activation of canonical Wnt signaling represents an early event and is associated with mutations in APC and AXIN. Whereas, further cancer development in the CIN pathway is related to mutations in KRAS, p53 and the SMAD TGF $\beta$  signaling mediators, the CIMP/MSI pathway is characterized by the BRAF V600E mutation which might be causally linked to the CGI-methylator phenotype. Hypermethylation of the MLH1 promoter results in MSI which induces frameshift mutations in TGF $\beta$ 2R and IGF2R (Adapted from Walther et al., 2009).

### 2.3.2 (Epi-) Genomic Instabilities in Colorectal Cancer Progression

CRCs derive from the adenomatous epithelium in colon and rectum. Normal colonic epithelial cells typically show an apical (facing the colonic lumen) versus basolateral (attached to the subjacent basal lamina) polarity. In addition, cells of an epithelial layer are strongly attached to each other due to tight junctions and adherence junctions anchoring epithelial cells in this structure. However, in the course of cancer progression tumor cells break out of the epithelial cell layer and invade other organs leading to the establishment of distant metastases. For this purpose cancer cells use a mechanism which is normally functions during embryonic development where certain differentiated epithelial cells lose their polarity and cell-cell junctions and acquire a migratory mesenchymal phenotype. Upon this epithelial-to-

mesenchymal transition (EMT) the cells become recruited to specific sites in the developing embryo wherein they can undergo the reverse process of mesenchymal-to-epithelial transition (MET) to form epithelial tissues at distal locations. Accordingly, cancer cells which underwent EMT possess migratory and invasive properties which are lost again by MET after initiation of secondary tumor growth (Ocana et al., 2012; Tsai et al., 2012). TGF- $\beta$  signaling is a potent inducer of EMT. In consequence to TGF- $\beta$  signaling, the E-cadherin repressors SNAI1, SNAI2, ZEB1 and TWIST are upregulated. Reduced of E-cadherin expression leads to the breakdown of adherence junctions and, along with other signaling events, loss of cell polarity ensues. In addition to its function in the canonical Wnt-pathway, cytoplasmic  $\beta$ -catenin is a component of adherence junctions. The collapse of these cell adhesion complexes results in the accumulation of  $\beta$ -catenin and in its nuclear translocation triggering expression of Wnt target genes. These include several E-cadherin repressors like SNAI1/2, ZEB1 and TWIST (Heuberger and Birchmeier, 2010; Sanchez-Tillo et al., 2011). Hence, also the canonical Wnt-signaling pathway can induce EMT and, moreover, can reinforce the mesenchymal phenotype. It was shown that EMT generates cells with stem cell properties e.g. expression of the stem cell marker CD44, a TCF/ $\beta$ -catenin target (Fan et al., 2013; Mani et al., 2008). Vermeulen and colleagues reported that the hepatocyte growth factor (HGF), a potent inducer of EMT secreted by the tumor-associated stroma, can activate stem cell-related Wnt/ $\beta$ -catenin signaling also *in vivo* (Vermeulen et al., 2010). Together these findings indicate that cancer cells can dedifferentiate to acquire stem cell-like features in a process resembling developmental EMT. Such stem cell features are the capacity of self-renewal (maintain the undifferentiated state over cell division) and the potential to differentiate to other cell types. In numerous human cancers, including CRC, a small cell population was identified, which is capable to give rise to tumors when xenografted into immunosuppressed mice. In contrast, most of tumor cells were not able to induce tumor growth (Ricci-Vitiani et al., 2007). The tumors derived from these tumor-initiating cells (TICs) resembled the cellular heterogeneity of the original human cancer and contained again a small TIC population. This suggested that TICs are able to differentiate and to self-renew and thus have stem cell character, a reason why they were also referred to as cancer stem cells (CSCs). Compared to normal colonic stem cells in healthy mucosa TICs occurred with much higher frequency in cancers. Several protein markers were proposed for the identification of TICs in CRC such as the surface proteins CD133, CD166, CD44 and as well as LGR5. The expression of CD44 and LGR5 is induced by Wnt/ $\beta$ -catenin signaling, suggesting that the Wnt-pathway also plays a role generation of CRC-TICs. *In vitro* data suggest that TICs and non-TICs don't follow a hierarchical model in which the stem-like TICs could differentiate to non-TIC cells but not *vice versa*. Rather, they appear to be linked by dynamic and stochastic inter-conversions. For instance, LGR5 negative cells can reversible convert into LGR positive under selective pressure (Kobayashi et al., 2012). Separation and

subsequent culturing of the differentiated cell fraction from two breast cancer cell lines re-established the original heterogeneous population including stem-like cells (P. B. Gupta et al., 2011). Whether these dynamic state transitions are linked to EMT and MET processes remains elusive, particularly in the light of reports showing that different types of TICs, possibly associated with different mesenchymal and epithelial features, are likely to co-exist in the same tumor (Liu et al., 2014).

EMT and MET are related to extensive gene expression reprogramming accompanied by reversible epigenetic changes at the levels of histone modifications and DNA methylation (Carmona et al., 2014; Ke et al., 2010). The EMT-inducing factors SNAI1 and SNAI2 were also shown to be regulated by DNA methylation during EMT (Y. Chen et al., 2013). The reversible nature of EMT and MET in cancer indicates therefore that underlying epigenetic mechanisms exhibit a certain plasticity or instability, respectively. However, it is currently not known, whether and how CRC-related deregulation of epigenetic mechanisms, which is particularly pronounced in cancers from the CIMP/MSI type, is involved in the generation of the implicated epigenetic plasticity.

### **3 Aims of the Thesis**

DNA methylation aberrations are a hallmark of human CRCs. Focal hypermethylation of promoter CpG islands results in the transcriptional silencing of tumor suppressor genes, whereas the global loss of methylation may trigger chromosomal instability. Although, CRC-specific methylation changes are well documented (Goel and Boland, 2012), the underlying molecular mechanisms are poorly understood.

**The first aim of my PhD thesis was to elucidate causes of aberrant hypermethylation in CRC.**

Towards this aim, I first investigated the effect of aging on the accumulation of methylation aberrations associated with CRC. To this end, I – in collaboration with others determined DNA methylation changes in the aging colon in a cohort of 546 healthy women by analyzing target-wise and genome-scale CpG methylation (**Appendix II**).

In a second project, I addressed the molecular basis of the CpG island methylator phenotype (CIMP) in CRC. Together with colleagues, I identified DNA methylation-regulating enzymes which are transcriptionally deregulated in CRC-CIMP, and investigated the effect of their deregulation on genome-wide CGI methylation and the cell phenotype (**Appendix I**).

DNA methylation modulates the accessibility of DNA for transcription, recombination and, presumably, replication. Accordingly, methylation aberrations in cancer contribute to cancerous transformation in different ways. Although, a multitude of methylation alterations and their function in colorectal carcinogenesis have been described, many aspects of CRC-related aberrant methylation remain undiscovered (Goel and Boland, 2012).

**The second aim of my PhD thesis was to explore consequences of aberrant methylation in CRC.**

Towards this aim, I investigated the role of aberrant methylation in the formation of numerical chromosomal aberrations during tumor development. In this pilot approach, I in collaboration with colleagues from different laboratories, characterized and correlated chromosomal breakpoint patterns and genome-wide DNA methylation patterns in a model system for secondary tumor formation (**Supplementary Results 4.5.1**).

In a second collaborative project, I addressed the role of aberrant methylation in downregulation of the tumor suppressor SHOCA-2. To this end, I analyzed DNA methylation in the promoter CGI of SHOCA-2 and investigated the mechanisms by which DNA methylation regulates SHOCA-2 expression (**Appendix III**).

## 4 Results

The following section summarizes the results presented in the manuscripts provided in the appendix as well as supplementary results not included in the manuscripts.

### 4.1 Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers (Appendix I)

The CIMP pathway of colorectal carcinogenesis is characterized by extensive hypermethylation of promoter-associated CGIs. Although, CRC-CIMP has been associated with overexpression of DNMT1 and DNMT3B, a direct link between overabundant DNA methyltransferase activity and CGI hypermethylation has not been established and the molecular basis of the aberrant methylation remains elusive (Kanai et al., 2001; Nosho et al., 2009). Recent studies showed that the glioma- and paraganglioma-associated competitive inhibition of  $\alpha$ -KG-dependent dioxygenases, which include the TET family of DNA dioxygenases, is sufficient to generate hypermethylation phenotypes in cell culture *in vitro* (Letouze et al., 2013; Turcan et al., 2012). These studies therefore suggest a plausible scenario whereby impaired TET-mediated DNA demethylation is involved in the development of hypermethylated tumors. To elucidate the role of impaired DNA demethylation in CRC-CIMP we analyzed mRNA expression levels of the three TET family members, TET1, TET2 and TET3 in CIMP and non-CIMP CRC cell lines. We found that TET1 is significantly downregulated in CIMP cell lines compared to non-CIMP cell lines (Appendix I, Figure 1a-c). Reduced TET1 expression coincides with reduced hydroxymethylation in the promoter CGIs of several CIMP marker genes, already indicating that TET proteins are involved in methylation regulation of these loci (Appendix I, Figure 1d). To test the association between CRC-CIMP and TET1 loss in primary human CRCs, we analyzed mRNA expression of TET1-3 in a series of 23 CIMP and non-CIMP cancers as well as in 143 primary colon cancers for which TET1 expression data and genome-wide methylation data were available from the The Cancer Genome Atlas (TCGA) database. In both cancer cohorts, TET1 expression was significantly downregulated in CIMP-associated cancers compared to non-CIMP cancers (Appendix I, Figure 2a and 2b). Remarkably, we observed that the loss of TET1 is frequently associated with hypermethylation of its own promoter region. Treatment of the CIMP cell line SW48 with the demethylating agent 5-Aza-cytidine then indeed confirmed that TET1 promoter methylation causes its transcriptional repression and hence that TET1 loss in CIMP can be explained by

epigenetic silencing (Appendix Figure 2c and 2d). This raised the question whether TET1 downregulation is cause or consequence of genome-wide hypermethylation in CIMP cancers. To address this question, we performed a sh-RNA mediated knockdown of TET1 in the non-CIMP cell line SW620 (TET1-kd) and analyzed the genome-wide CGI methylation patterns in two most efficient TET1-kd clones using the Illumina Infinium human methylation (HM) 450 platform. In these clones TET1 depletion resulted in hyper- and hypomethylation of CGI-associated CpG sites (Appendix I, Supplementary Figure 3d). Promoter hypermethylation was observed in MLH1 and targets included in the CIMP five-marker panel (Weisenberger et al., 2006). Increased levels of MLH1 promoter methylation were accompanied by reduction of its mRNA expression (Appendix I, Supplementary Figure 4a-c). Eventually, individual CpG sites in the TET1 promoter also became hypermethylated upon TET1-kd demonstrating that TET1 promoter methylation can result from its transcriptional repression (Appendix I, Figure 3d). We then wondered how altered methylation pattern in TET1-kd cells would further evolve in the course of tumor formation. To address this question, we engrafted cells from the two TET1-kd clones subcutaneously into immunosuppressed mice (NOD/SCID) and analyzed CGI methylation at genome-scale (HM 450) in the emerged tumors. Xenograft methylation patterns from both TET1-kd clones showed a significant shift towards hypermethylation compared to the engrafted cells (Appendix I, Figure 3c).

Tet1 expression has been shown to be critical for embryonic stem cell identity in mouse (Freudenberg et al., 2012; Ito et al., 2010). To address the role of TET1 in stem cell phenotype in CRC cells, we investigated the expression of LGR5, which has been previously identified as a marker of CRC-TICs but also of normal epithelial stem cells of the colon (Dame et al., 2014; Forster et al., 2014; Hirsch et al., 2014). TET1-kd cells showed significantly reduced TET1 expression and the respective xenografts reduced numbers of LGR5 expressing cells in flowcytometric analyses when compared to the relevant controls. Similarly, the number of cells expressing CD44 and CD166, two additional putative markers for CRC-TICs was also reduced in TET1-kd xenografts (Appendix I, Figure 4).

Cancer cell lines and primary tumors represent a heterogeneous cell population regarding the degree of differentiation. They contain cells which appear fully differentiated and cells exhibiting stem cell character. It has been previously shown that these cell states are interconvertible in cell lines (P. B. Gupta et al., 2011). Since, such a cell state plasticity corresponds to and requires epigenetic plasticity we wondered how TET1 depletion would affect the behavior of cancer cell populations. We found that populations of TET1-kd cells have an altered cell composition regarding the expression of the two putative CRC-TIC markers CD24 and CD44 (Appendix I, Figure 5a). This composition was either shifted to the CD24/CD44 double positive state or the CD24/CD44 double negative state. We then isolated CD24/CD44 double-positive and CD24/CD44 double-negative cell subpopulations from TET1-kd and



control cells and expanded them in culture. We found that sorted TET1-kd cells remained in their cell state whereas control cells re-established the original cellular heterogeneity (Appendix I, Figure 5b). We concluded that TET1 depletion results in a reduced DNA methylation plasticity and TET1 depleted cells are locked in the cell state at the time of TET1 depletion.

Analysis of mRNA expression data from TCGA revealed that CIMP-CRCs feature a significant loss of the epithelial cell marker E-cadherin (Appendix I, Figure 6c). We wondered whether the loss of E-cadherin in CIMP cancers can be attributed to TET1 depletion and the ensuing methylation aberrations. We found that E-cadherin is downregulated also in TET1-kd cells and that this is indeed associated with increased promoter methylation (Appendix I, Figure 6b). In addition, TET1-kd cells showed elevated expression of the mesenchymal markers fibronectin and vimentin, and acquired a migratory cell behavior (Appendix I, Figure 6d and 6e). Therefore, the loss of E-cadherin, associated with hypermethylation of its promoter generates a mesenchymal cell behavior, a process which resembles an epithelial-to-mesenchymal (EMT) in some aspects.

In conclusion, we show that the CIMP pathway of colorectal carcinogenesis is characterized by the loss of TET1 expression. Further, we show that TET1 depletion results in genome-wide methylation aberrations which are associated with the acquisition of CRC-CIMP-specific features, including the transcriptional repression of MLH1 and E-cadherin. Whereas TET1-depleted cells showed both, CpG hyper- and hypomethylation, mouse xenograft tumors derived from these cells displayed a significant shift towards hypermethylation, suggesting that the selection associated with tumor formation shapes cancer specific CpG hypermethylation patterns. Furthermore, our results show that the loss of TET1 expression results in a decreased cell plasticity, most likely due to the loss of an essential pathway providing DNA methylation dynamics and, hence, epigenetic plasticity. Eventually, we showed that TET1 depletion is associated with a loss of epithelial cell character and the acquisition of mesenchymal cell features, resembling an EMT. This has potential clinical implications.

**Contribution:** I characterized CRC cell lines as well as the tester set of primary CRCs regarding their CIMP status by analyzing the methylation status of the five marker panel. I further investigated mRNA and protein expression levels of TETs, DNMTs and TDG in CRC cell lines and in the tester set of primary CRCs. Moreover, I established and validated the sh-RNA mediated knockdown of TET1 in the SW620 cell line. In TET1 knockdown cells I performed all methylation analysis (except for the pre-processing of Illumina Infinium HM 450

data) and hydroxymethylation analysis by 5hmC dot blot and I analyzed mRNA or protein expression of TET1, TET2, TET3, FN1, VIM and MLH1. I initiated and coordinated collaborations with colleagues from different laboratories. Eventually, I designed and wrote the manuscript provided in Appendix I.

## **4.2 Modulation of Age- and Cancer-Associated DNA Methylation Change in Healthy Colon by Aspirin and Lifestyle (Appendix II)**

DNA methylation patterns change over age (methylation drift) in various human tissues including the colonic mucosa (Issa, 2014). Age-related de novo methylation arises also in promoter CGIs of cancer-relevant genes like MLH1 and MGMT (Menigatti et al., 2009). We addressed the question to what extent epidemiologically implicated, lifestyle-related CRC-risk factors, like smoking and overweight, but also protective factors, including hormone replacement therapy (HRT) and aspirin use, influence the methylation drift of these two genes (Chan et al., 2007; Long et al., 2010; Parajuli et al., 2013; Robsahm et al., 2013). Using the very sensitive locus-normalized quantitative methylation specific PCR method (In-qMSP) we quantified methylation levels of the promoter-associated CGIs of MLH1 and MGMT in a total of 1092 biopsies obtained from the cecum (proximal, right colon) and the sigmoid (distal, left) colon of 546 healthy women. Multivariable regression analysis showed that MLH1 promoter methylation increases significantly with age in the proximal colon (Appendix II, Figure 1A). MGMT promoter methylation levels were not significantly age-related unless lifestyle factors were taken into account: Long-term aspirin use, defined as two or more tablets per week for more than 2 years, significantly lowered the methylation rate of MGMT (Appendix II, Figure 1B). Long-term aspirin use also reduced the risk of developing serrated polyps (hyperplastic polyps and sessile serrated adenomas) and consistently, women with no polyps had a significantly lower rate of MGMT methylation compared to women with serrated polyps (Appendix II Figure 1C and 1D). In contrast to the protective effect of aspirin, long-term smoking, defined as 1 or more cigarettes per day for more than 20 years increased the risk of serrated polyps (Appendix II Figure 1D).

To extend our investigation to a genome-wide level we further analyzed DNA methylation in 178 normal colon samples from our cohort, using the Illumina Infinium human methylation (HM) 27 platform. We analyzed methylation of 20,025 promoter-associated CpG sites with multivariable regression adjusted for colonic location, smoking, aspirin use and also for received HRT and the body mass index (BMI). We found that methylation of promoter-associated CpG sites mainly increases over age (1713 CpGs gaining and 343 CpGs losing methylation). Further, we found that age-related methylation changes are predominantly associated with the distal part of the colon (Appendix II, Figure 2A). Regarding the lifestyle factors, HRT and aspirin use reduced the number of age-dependent differentially methylated CpG sites. However, a high BMI (>30) promoted an age-related hypermethylation (Appendix II, Figure 2B). Smoking promoted age-related hypermethylation in proximal colon but suppressed age-dependent methylation changes in the distal colon (Appendix II,

Supplementary Figure 2A and 2B). Next, we investigated the rate of age-dependent hypermethylation regarding the colonic location and lifestyle factors. The rate of methylation changes was significantly higher in the distal colon compared to the proximal colon but it was significantly lower in long-term aspirin users (compared to non-users) and women who reported HRT (compared to no-HRT) (Appendix II, Figure 2C and D). Long-term smoking was associated with an increased rate of hypermethylation in the proximal colon but a decreased rate in the distal colon (Appendix II, Supplementary Figure 2A and B).

It has been shown previously, that sites of age-dependent methylation changes in blood are associated with the H3K27 trimethylation (H3K27me3) mark established by polycomb-group proteins (PcG) during embryonic development (Rakyan et al., 2010). We wondered whether sites of age-related methylation changes in the colon are also associated with polycomb targets and classified HM27 CpG sites according to coincidence of H3K27me3 in human embryonic stem cells (GSE13084; (Ku et al., 2008). Regions (500bp) around age-dependent hypermethylated CpG sites were significantly enriched for H3K27me3 compared to total H3K27me3-marked CpG on the HM27 array. In addition, the rate of hypermethylation at H3K27me3-associated CpG sites was significantly higher compared to the median rate of hypermethylation. Aspirin use suppressed the hypermethylation of PcG target regions, whereas long-term smoking and a high BMI promoted methylation of these sites (Appendix II, Figure 2F).

To address the cancer relevance of the age-related and lifestyle-modulated methylation changes in the healthy colon, we compared genome-wide methylation data of 59 primary CRCs (female patients) with those from our healthy colonic mucosa samples (Hinoue et al., 2012). Half of the CpG sites hypermethylated in cancer samples showed age-dependent hypermethylation. These sites were enriched for H3K27me3 compared to sites only age-dependent or only cancer-related. In addition, CpG sites methylated also in cancer displayed a higher rate of age-related methylation change than sites showing age-dependent methylation change only (Appendix II, Figure 3B). We used gene expression data available for colonic adenomas (GSE8671; (Sabates-Bellver et al., 2007) and found a significant overlap between CpG sites associated with age-dependent hypermethylation and transcriptional repression of the corresponding genes in adenomas (Appendix II, Figure 3C). Cancer-relevant methylation changes in the normal colonic mucosa were influenced by lifestyle factors. A statistically significant fraction of promoters showing either aspirin- or HRT-suppressed methylation or high BMI- or smoking-promoted methylation in the aging colon coincided with CpG sites hypermethylated in CRC (Appendix II, Figure 4A). These included several genes with established roles in carcinogenesis like CDKN2A and CDH1. In this context, we further investigated the effect of lifestyle factors on the age-dependent methylation of CpG sites located in the promoter regions of 664 annotated tumor-associated genes (J. S. Chen et al.,

2013). The rate of methylation change at these loci was significantly reduced in long-term aspirin users compared to non-users and women reporting HRT compared to no-HRT but increased in women with high BMI vs normal BMI and long-term smokers vs non-smokers (Appendix II, Figure 4B).

In conclusion, we were able to show that lifestyle factors which have been reported as modulators of CRC risk have genome-wide influence on DNA methylation stability in the colonic mucosa. Age-associated gene promoter methylation increase was suppressed by HRT and regular long-term aspirin use but was promoted by long-term smoking and obesity, which is concordant with the effects of these factors on CRC risk (Appendix II, Figure 4C). Together with the finding that a significant fraction of targets of age- and cancer-related hypermethylation coincided, our results provide a functional link between age-related hypermethylation, CRC and lifestyle-related CRC risk factors.

**Contribution:** I established the methods for quantitation of MGMT promoter methylation by In-qMSP. Further, I analysed MGMT promoter methylation in the genomic DNA from 510 healthy colonic mucosa samples.

### **4.3 The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation (Appendix III)**

Allelic loss in short arm of chromosome 8 (8p) is a frequent chromosomal abnormality in CRC and linked to chromosome breakage at fragile sites located at 8p12 and 8p22 (Chughtai et al., 1999; Emi et al., 1992). However, a tumor suppressor gene at this locus has not been identified. We found, that the SH2D4A gene which encodes for a SH2-domain containing protein named SHOCA-2 (SH2-domain containing adaptor protein), maps to human chromosome 8p21.3. SH2D4A was originally identified as a paralogue to SH2D4B (encoding SHOCA-1) which is differentially expressed during maturation of epithelial cells in the embryo. Both genes are evolutionary conserved and their gene products display a 65% amino acid sequence identity (Appendix III, Figure 1A), neither of them has been functionally characterized. We analysed SHOCA-2 expression in 400 sporadic CRCs by immunohistochemistry and found significant differences in detectable SHOCA-2, correlating with clinical tumor stage (I-IV). SHOCA-2 expression was reduced in advanced CRC stages (Appendix III, Figure 1B and 1C). Patients with low or absent SHOCA-2 had a significantly poorer clinical outcome compared to patients with the highest measured SHOCA-2 expression (Appendix III, Figure 1D). We found that SHOCA-1 mRNA expression is concurrently lost in half of the samples showing low or no SHOCA-2 expression (Appendix III, Figure S1F).

To determine the molecular basis of SHOCA-2 loss we selected 27 CRCs from a cohort of 70 CRC patients, in which SH2D4A alleles could be distinguished by 3 microsatellite markers and 6 gene specific single nucleotide polymorphisms (SNPs). 17 of these primary tumors had lost or diminished SHOCA-2 expression. We detected LOH of SH2D4A in 7 of the 17 tumors. Gene dosage quantification of chromosome 8p indicated mono- and biallelic loss in 6 and 1 of the LOH tumors, respectively (Appendix III, Figure 2A). In the matched healthy mucosa, 4 of the 6 patients with a monoallelic loss of SH2D4A were heterozygous for an intronic SNP (rs17128221, c.342-5T>C) existing in a T (wildtype) or C allele. The C mutated allele leads to exon 4 skipping causing a frame shift mutation and premature translation termination in exon 5 (Appendix III, Figure 2B). The corresponding cancer samples of heterozygote patients uniformly demonstrated the selective loss of the T allele whereas homozygosity for the C allele was never observed in 83 normal mucosa samples. In 3 of the 10 primary tumors in which loss of SH2D4A expression was not associated with LOH we found missense mutations and deletions. These alterations either produced amino acid changes or frame shift mutations (Appendix III, Figure 2C). To further elucidate the causes for SH2D4A loss in CRC we investigated the promoter methylation status of SH2D4A by pyrosequencing of bisulfite converted genomic DNA. Two neighboring CpG sites, located downstream of the transcription

start site, were hypermethylated compared to the matched healthy mucosa in 4 patients with SH2D4A loss, that did not show LOH or mutations (Appendix III, Figure 2D). One of these CpG sites coincided with a canonical Sp1 transcription factor binding motif and its methylation significantly reduced Sp1 affinity (Appendix III, Figure S2B). Hence, we found that distinct genetic but also epigenetic abnormalities contribute the loss of SH2D4A expression in CRC.

The function of SHOCA-2 was uncharacterized so far. A proteomic study of EGFR-driven phosphor-tyrosine signaling in lung cancer cells identified a target peptide with sequence homology to SHOCA-2 (A. Guo et al., 2008). We confirmed that EGFR activation triggers tyrosine phosphorylation of SHOCA-2 in HeLa cells (Appendix III, Figure 3A). EGF stimulation of HeLa cells overexpressing SHOCA-2 decreased phosphorylation of the transducer and activator of transcription 3 (STAT3) at tyrosine (Tyr) 705. Two other EGFR signal transduction pathways, PI3K/Akt and MAPK/ERK were activated or unaffected, respectively, in response to SHOCA-2 overexpression (Appendix III, Figure S3A). To further investigate the function of SHOCA-2 in EGFR signaling we performed a sh-RNA mediated knockdown (kd) of SHOCA-2 in HeLa cells which reduced SHOCA-2 levels by 90%. When stimulating kd-cells with EGF, STAT3 Tyr 705 phosphorylation was increased compared to control cells (Appendix III, Figure 3C). In contrast, phosphorylation of Akt was decreased while that of ERK1/2 was unchanged (Appendix III, Figure S3C). These results suggest an involvement of SHOCA-2 in the regulation of STAT3-mediated EGFR-signaling. To prove this hypothesis, we used a STAT3-driven reporter construct in SHOCA-2 overexpressing and SHOCA-2-kd HeLa cells. EGF treatment in overexpressing cells reduced reporter expression, however, in kd-cells it resulted in a robust transcriptional activity (Appendix III, Figure 3D). Using immunoprecipitation assays we found that SHOCA-2 binds to EGFR as well as STAT3 independent of its SH2 domain (Appendix III, Figure 3E and Figure S3F). These results suggest that SHOCA-2 functions as an inhibitor of EGF-induced STAT3 activation.

Using tandem affinity purification (TAP) we identified the serine/threonine protein phosphatase-1 beta (PP1 $\beta$ ) as an additional interaction partner of SHOCA-1 and SHOCA-2 (Appendix III, Figure 4A). The interaction between SHOCA-2, STAT3 and PP1 $\beta$  increased upon EGF treatment in HeLa cells (Appendix III, Figure 4D). We wondered whether the EGFR-mediated phosphorylation of SHOCA-2 is important for the formation of the SHOCA-2/EGFR/STAT3/PP1 $\beta$  complex and mutated the SHOCA-2 phosphorylation site (Tyr 131). Although, the point mutation (Y131A) did not disturb the ability to interact with EGFR, STAT3 and PP1 $\beta$ , it resulted in a constitutive phosphorylation of other SHOCA-2 tyrosine residues and in the abolishment of SHOCA-2-mediated inhibition of STAT3 phosphorylation (Appendix III, Figure S4H and S4I). These results show that the STAT3-inhibitory function of SHOCA-2 depends on its phosphorylation by EGFR. In addition, it depends on the association of SHOCA-

2 with PP1 $\beta$  as mutation of the PP1 $\beta$ -binding motif in SHOCA-2 abolishes its ability to inhibit STAT3 activation upon EGF stimulation (Appendix III, Figure 4E).

To address the contribution of SHOCA-2 to the regulation of cell proliferation we assessed the expression of STAT3-controlled genes in SHOCA-2-kd HeLa cells. Upon stimulation with EGF, protein expression of c-Myc, Cyclin D1 and Jun B was increased in SHOCA-2-kd cells compared to control cells. Consistently, the fraction of cells in G2/M stage of the cell cycle was significantly higher in EGF-stimulated and unstimulated SHOCA-2-kd cells (Appendix III, Figure 5A). Increased cell proliferation marked by Ki-67 inversely correlated with SHOCA-2 expression also in tissue sections from primary CRCs (Appendix III, Figure 5B). To address whether the inhibition of STAT3 signaling by SHOCA-2 is responsible for the increased cell proliferation in kd-cells we overexpressed STAT3 in the CRC cell line SW480 and co-expressed either wildtype or mutant SHOCA-2. Overexpression of STAT3 resulted in an increased potential of colony formation coinciding with high levels of STAT3 phosphorylation. Co-transfection with wildtype SHOCA-2 significantly reduced the colony forming potential of SW480 cells as well as STAT3 phosphorylation. This growth suppressive effect was less pronounced when cells were co-transfected with either the SHOCA-2 Y131A (phosphorylation site) mutant or the PP1 $\beta$ -binding site mutated form of SHOCA-2 (Appendix III, Figure 5C and 5D). Furthermore, SHOCA-2 overexpression in the CRC cell line SW620 in which STAT3 is constitutively phosphorylated suppressed cell growth (Appendix III, Figure 5 E).

Further, we found that the sh-RNA-mediated knockdown of SHOCA-2 in SW480 cells changes the cell shape towards a spindle-like appearance. This was accompanied by the loss of E-cadherin expression and the increased expression of mesenchymal markers including N-cadherin, vimentin, ZEB-1 and SNAI1; features characteristic for an epithelial-to-mesenchymal transition (EMT) (Appendix III, Figure 6A and 6B).

To investigate the potential tumor-suppressive role of SHOCA-2 *in vivo* we engrafted SW480 cells and SW480 SHOCA-2-kd cells into nude mice. The tumor forming potential of SHOCA-2-depleted cells was higher compared to normal SW480 cell, in terms of tumor incidence as well as tumor growth rate (Appendix III, Figure 6 E). In addition, higher levels of phosphorylated nuclear STAT3 were detected in tumors derived from SHOCA-2-kd cells.

In conclusion, our results establish SHOCA-2 as a tumor suppressor due to the inhibition of STAT3-mediated EGFR signaling which impedes apoptosis and promotes cell proliferation. Inactivation of SHOCA-2 in colorectal cancers included genetic (chromosomal aberrations and mutations) and epigenetic mechanisms. Further, our data indicate that the tumor-suppressive role of SHOCA-2 takes effect primarily during cancer progression (advanced cancer stages) as the loss of SHOCA-2 is particularly pronounced at later cancer stages and induces EMT which is essential for metastasis formation. Eventually, SH2D4A represents the first identified



tumor suppressor located within chromosome 8p21-23, the deletion of which constitutes a characteristic of CRCs but its contribution to CRC development has remained unclear until now.

**Contribution:** I quantified mRNA expression levels of SHOCA-1 in cancer-associated matched mucosa samples as well as in primary CRCs using qRT-PCR. Furthermore, I quantified SH2D4A promoter methylation levels by pyrosequencing of bisulfite converted genomic DNA. This included the assay design as well as data analysis. In addition, I determined the affinity of the Sp1 transcription factor to the methylated and unmethylated SH2D4A-associated Sp1 binding site using an electrophoretic mobility shift assay. This included the assay design, execution and data analysis.

#### **4.4 Colorectal mucosa of healthy individuals displays part of the CpG island methylator phenotype signature (Appendix IV)**

Aberrant promoter hypermethylation is a hallmark of human CRCs and is particularly pronounced in cancers of the CIMP oncogenic pathway. However, little is known about the precise timing of CIMP during tumor development. We had previously shown that promoter methylation of MLH1, which is associated with CIMP, can be detected at low levels in colonic mucosa from healthy individuals (Menigatti et al., 2009). This observation suggested that CIMP cancer-related methylation signatures are partially reflected in the healthy colon. To further investigate the extent of cancer-specific methylation in normal colonic tissues, we identified gene promoters affected by aberrant DNA methylation early in colorectal carcinogenesis. To this end, we explored gene expression data from adenomas and matched normal mucosa samples (GSE8671, (Sabates-Bellver et al., 2007)) and selected five gene candidates, showing reduced mRNA expression compared to normal mucosa and harboring a promoter CpG island; Forkhead box F1 (FOXF1), Carbonic anhydrase IV (CA4), Neuropeptide Y receptor Y1 (NPY1R), Gremlin 1 (GREM1) and Interferon-induced transmembrane protein 1 (IFITM1). We quantified promoter methylation levels of these genes as well as MLH1 and MGMT in genomic DNA of matched primary colorectal cancer tissue (CRC), CRC-associated normal mucosa (CAM) and blood samples from 106 unselected patients, using a highly sensitive, locus-normalized methylation-specific quantitative PCR assay (In-qMSP). By comparing methylation levels in these three different tissues, we established a methylation threshold for each target gene, allowing us to discriminate CRC from CAM with  $\geq 99\%$  specificity. These thresholds ranged from 4% (FOXF1) to 7% (IFITM1). Tumor samples showing methylation levels above the threshold were classified as hypermethylated. (Appendix IV, Figure 1A). FOXF1 represented the most frequently hypermethylated target gene (66% of all cancer samples), whereas GREM1 was only hypermethylated in 13% of the cancer samples. When applying the criterion that at least one of the target genes must be hypermethylated, our seven-marker panel separated CRC from CAM in 82% of the patients (87/106) with a specificity of 99%. The coverage reached 94% (30/32) when the panel was applied to CRCs located at the proximal (right) colon (Appendix IV, Figure 1B and 1C). Treatment of CRC cell lines, in which the seven marker panel was highly methylated and transcriptional repressed, with the demethylating agent 5-Aza-Cytidine and/or the histone deacetylase inhibitor Trichostatin A partially restored mRNA expression of the genes and showed that promoter methylation of the genes is directly involved in their transcription regulation (Appendix IV, Supplementary Figure 3). We correlated the methylation patterns of our seven-marker panel with clinical features including gender, tumor location, and cancer

stage. Tumors hypermethylated in IFITM1 occurred more frequently in men whereas FOXF1, MGMT, NPY1R, GREM1, and MLH1 were more frequently hypermethylated in women. All markers, except IFITM1, were more frequently hypermethylated in proximal than in distal (left) CRCs. Further, MGMT was found hypermethylated more frequently at advanced cancer stages (Appendix IV, Figure 1B and 2A and Supplementary Figure 4A and 4B).

To address whether methylation of our seven-marker panel is associated with an overall methylator phenotype as proposed for CIMP, we correlated the methylation status of all markers in 106 CRCs. We observed concomitant hypermethylation of CA4, NPY1R and GREM1 with the CIMP associated marker MLH1 (Appendix IV, Figure 2B and 3C). To ascertain the possible association of CA4, NPY1R and GREM1 hypermethylation with CIMP, we analyzed genome methylation patterns of 3 CIMP and 4 non-CIMP (classified according to the BRAF V600E mutation status) CRCs as well as 8 healthy colonic biopsies (HM) and 1 blood sample using the Illumina Infinium human methylation 27 platform. Methylation levels of CA4, GREM1 and MLH1 were significantly higher in CIMP-associated CRCs compared to non-CIMP cancers. For NPY1R this trend did not reach statistical significance (Appendix IV, Figure 4). Hence, our results suggest that GREM1 and CA4 represent novel markers for CRC-CIMP. All CRCs hypermethylated for CA4, GREM1, MLH1 but also NPY1R located to the proximal colon coinciding with the high incidence of CIMP cancers in this colonic region (Appendix IV, Figure 2A).

Finally, we asked whether methylation of our seven-marker panel is detectable in healthy colonic mucosa as it has already been reported for MLH1 (Menigatti et al., 2009). Therefore, we quantified their methylation levels in 106 CAM samples and in tissue samples from 32 healthy subjects (HM). We found concomitant low-level (below cancer-specific thresholds) methylation of CA4, NPY1R and GREM1 in CAM and HM samples (Appendix IV, Figure 3A and 3B). Genome-wide methylation data from 8 HM samples substantiated this observation (Appendix IV, Figure 4). Hence, in addition to MLH1, also for the novel CIMP markers CA4 and GREM1 low levels of methylation are detectable in healthy colonic mucosa.

In conclusion, our data show that CRC-specific methylation aberrations occur in a cancer stage, gender and tumor location specific manner, suggesting different roles in colorectal carcinogenesis. A subset (CA4, GREM1 and NPY1R) of investigated methylation target genes displayed association with CRC-CIMP and, as previously reported for the CIMP marker MLH1, showed low levels of methylation in healthy colonic mucosa. These observations suggest that CIMP arises very early in colorectal carcinogenesis and then progress throughout tumor development. The extended CIMP marker panel represents a potential diagnostic tool for cancer-risk assessment with 99% specificity and a sensitivity of 83% (94% for proximal CRCs).

**Contribution:** I quantified promoter methylation of CA4, NPY1R and IFITM1 in 106 CRC samples, 106 CAM samples as well as in 32 healthy mucosa samples using In-qMSP. Further, I quantified promoter methylation of CA4, NPY1R and IFITM1 as well as the mRNA expression levels in 5-Aza-cytidine and Trichostatin A-treated CRC cell lines CO115 and SW48. Eventually, I analyzed promoter methylation of FOXF1, GREM1, CA4, NPY1R and IFITM1 by sequencing of bisulfite-converted genomic DNA from two CRCs each, using a TA-cloning approach. My contribution included design and execution of experiments such as qPCR, cloning, DNA sequencing and data analysis.

## 4.5 Supplementary Results

### 4.5.1 Association of Chromosomal and Epigenetic Alterations During Tumor Development

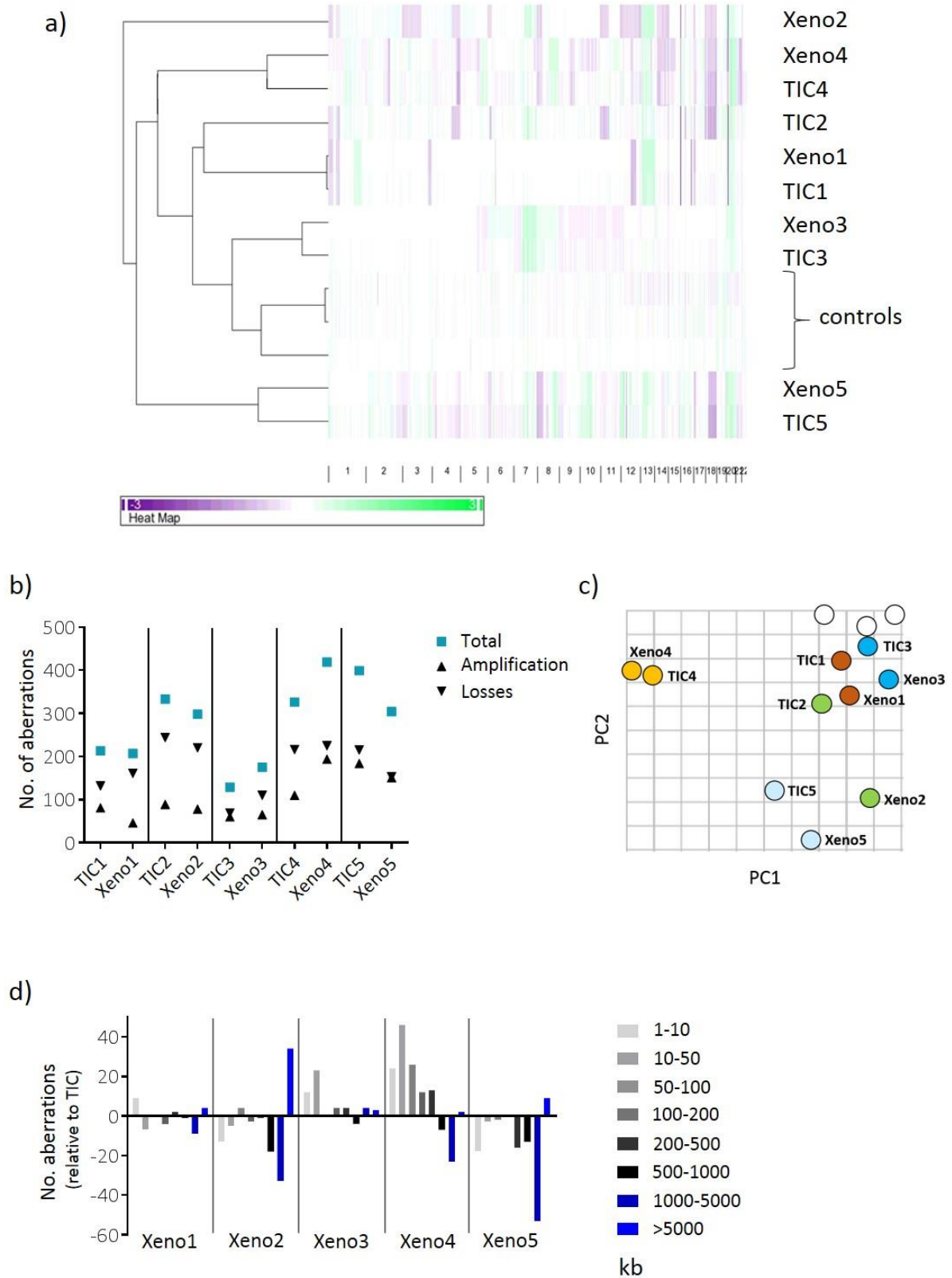
Chromosomal instability (CIN) constitutes a frequent form of genetic instability in CRC and is present in 65%-70% of sporadic CRCs. CIN is characterized by structural and numerical (aneuploidy) chromosome alterations and is therefore a major source of gains or losses of genetic information, including Loss of Heterozygosity (LOH). Interestingly, somatic DNA Copy number alterations (SCNAs) in CRC-CIN preferentially affect a certain genomic loci. Large (chromosome-arm level) amplifications are mainly found at 7p and q, 8q, 13q and 20q whereas arm-level deletions primarily affect 1p, 5q, 8p, 15q, 17p, 18p and q (Al-Sohaily et al., 2012; Cancer Genome Atlas, 2012). Sites of chromosomal breakage and rearrangements, also referred to as fragile sites, are often capable to obstruct DNA polymerase progression during replication due to the formation of secondary structures. Stalled replication forks represent a potential source of DNA double strand breaks (DSBs), which are believed to give rise to chromosomal rearrangements (Zeman and Cimprich, 2014). Guanine-rich G4 sequences can form intramolecular four-stranded structures (G4 quadruplexes) that interfere with DNA polymerase movement, and the formation of which is negatively affected by CpG methylation. It has been shown that G4 sequences associated with chromosomal breakpoints are cancer-specifically hypomethylated in colorectal cancer (De and Michor, 2011). This observation suggests a mechanistic relationship between epigenetic aberrations and the emergence of chromosomal instability. It was reported that patterns of SCNAs change during metastasis formation in colorectal cancer and we wondered whether these changes are triggered or accompanied by epigenetic alterations (Xie et al., 2014).

To address this question we used mouse xenograft tumors derived from stem cell-like tumor-initiating cells (TICs) as a model for secondary tumor formation. We investigated genome-wide patterns of SCNAs and DNA methylation in 5 pairs of TICs and xenografts. TIC cell lines were established from primary colorectal tumor tissues on the basis of their self-renewal capacity and expression of the putative stem cell markers CD133 and LGR5. These cells were shown to have potential to differentiate and to be tumorigenic in immunosuppressed mice (Ricci-Vitiani et al., 2007). Samples of TICs and xenograft tissues, which were generated by subcutaneous inoculation of TICs into SCID mice, were kindly provided by Lucia Ricci-Vitiani from the Istituto Superiore di Sanità in Rome, Italy. Clinical characteristics of the cancer

patients and the molecular features of TICs (analyzed in the lab of Ricci-Vitiani) are given in Table 4-1.

Genome-wide SCNA-patterns in TICs and xenografts were analyzed in collaboration with Tobias Heckel at Roche, Basel, Switzerland, using a comparative genome hybridization (CGH) microarray from NimbleGen in the 2.1M format. Two human ES cell lines and a healthy peripheral blood sample served as normal controls. Using the segMNT algorithm with the conservative criteria of a  $\pm 0.5$  shift in fluorescence signal log<sub>2</sub> ratio (sample vs. reference), we identified high confidence SCNA-segments. TIC2, TIC4 and TIC5 showed a high number of SCNAs, spanning entire chromosomes and chromosome arms. TIC1 and TIC3 displayed a lower level of copy number alterations with segments usually smaller than 50kb (Figure 4-1a and 4-1b). Such small numerical aberrations with a size below 100kb commonly present in human cells and also observed in our normal controls (Figure 4-1a) (Conrad et al., 2010). Several chromosomal aberrations observed were previously described in CRC, including deletions of 1p, 8p and 18q and amplifications of 13q and 20q. Hierarchical clustering and principal component analysis revealed that SCNA-patterns of TICs and corresponding xenografts are highly correlated, except for TIC2. TIC2 showed unstable SCNA-pattern with clear changes occurring during xenograft formation (Figure 4-1a and 4-1c). Notably, in the TIC2 derived xenograft, short chromosomal aberrations appeared to be merged into or superimposed by larger (>5Mb) numerical alterations, affecting chromosome arms and even whole chromosomes, e.g. 3p, 14q and chromosome 12 (Figure 4-1a and 4-1d). Similar to the TIC2 xenograft also the TIC5 xenograft displayed increased occurrence of very large SCNAs (>5Mb) and a concomitant loss of shorter alterations (Figure 4-1a and 4-1d). Xenografts from TIC1, TIC3 and TIC4 acquired additional, relatively short copy number alterations shorter than 50kb (TIC1 and TIC3) and 500kb (TIC4), respectively. According to the extent of SCNAs and their stability during xenograft formation, we distinguished three classes of TICs: Chromosomal stable (TIC1, TIC3), chromosomal degenerated (TIC4, TIC5) and chromosomal instable (TIC2). Chromosomal stable TICs displayed the lowest level of chromosomal aberrations which appeared stable during xenograft formation (Figure 4-1b and 4-1d). Chromosomal degenerated and chromosomal instable TICs featured a high number of chromosomal aberrations which appeared largely stable during xenograft formation in degenerated TICs, but changed drastically in instable TIC (Figure 4-1b and 4-1c).

Figure 4-1



**Figure 4-1: SCNA profiles of CRC-TICs and derived mouse xenografts.** a) Hierarchical clustering of SCNA profiles of TICs and xenografts (Xeno). Log<sub>2</sub> converted ratios of fluorescence signals from samples compared to reference genome is indicated by the color key from purple (deletion) to green (amplification). Autosome number is indicated at the bottom. b) Number of numerical aberrations in TICs and xenografts (Xeno). Total number of SCNAs, number of copy number losses and number of copy number amplifications are depicted. c) Principle component analysis (PCA) of SCNA-profiles in TICs and xenografts (Xeno). Corresponding TIC-xenograft pairs are depicted in the same color. Control samples are depicted in white color. d) Changes in the number of SCNAs in mouse xenografts (Xeno) compared to corresponding TICs, according to the size of aberrations.

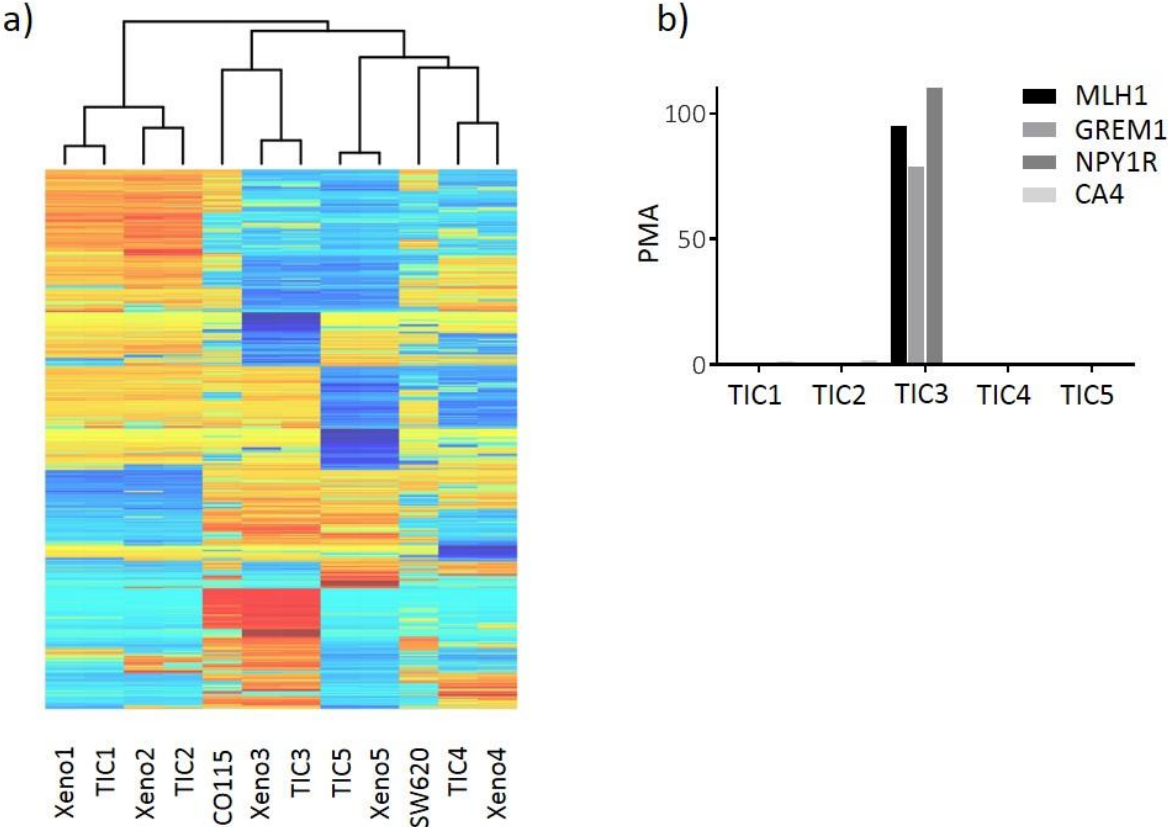
Genome-wide DNA methylation profiles of TICs and Xenografts were analyzed using the Illumina Infinium human methylation (HM) 450 platform. The CIMP CRC cell line CO115 and the non-CIMP cell line SW620 were included in this analysis as references for methylator and non-methylator phenotype CRC samples. RPMM-based unsupervised clustering using the 25% most variably methylated CpG sites (6571 CpG sites) revealed that genome-wide methylation patterns of TICs and corresponding xenografts are closely related (Figure 4-2a). In contrast to the SCNA patterns, methylation profiles of TIC2 were largely stable during xenograft formation. Methylation pattern of TIC3 was closely related to that of the CIMP cell line CO115. TIC3 harbors an activating KRAS G12V mutation (Table 4-1), which has been shown to be associated with an intermediate methylator phenotype referred to as CIMP-low or CIMP-2 (Hinoue et al., 2012). In addition, TIC3 displayed microsatellite instability (MSI; Table 4-1) associated with CRC-CIMP as well. We analyzed promoter methylation levels of the CIMP marker MLH1 and the CIMP-associated marker genes CA4, NPY1R and GREM1 (see chapter 4.4) using the sensitive In-qMSP assay. Promoter regions of MLH1, GREM1 and NPY1R were highly methylated in TIC3, further indicating that TIC3 derived from a tumor which arose by a CIMP-related oncogenic pathway (Figure 4-2b). In accordance with previous studies, showing that CIN and CIMP occur mutually exclusive in CRC, TIC3 showed the lowest level of chromosomal aberrations (Cheng et al., 2008). Interestingly, TIC3 showed a monoallelic amplification of a 3.6 Mb segment in chromosome 7, containing the BRAF gene. Activating mutations of BRAF are highly associated with CRC-CIMP. A link between CIMP and the amplification of the BRAF locus has not been described but may possible, as indicated by this observation.

To perform an initial analysis into the relationship between DNA methylation and chromosomal(in)stability, we examined the correlation between CpG methylation levels and the presence of chromosomal breakpoints. To this end, we defined sections of 1kb in size (according to the median probe spacing of 1169bp in CGH array) around identified chromosome breakpoints (500bp upstream and downstream). Chromosome breakpoints were



determined by the segMNT algorithm as endpoints and startpoints of SCNA segments. Subsequently, we identified CpG sites in the breakpoint sections that were represented on the HM450 array. 9% of the breakpoint-sections in TICs and xenografts did overlay with CpG sites represented on the HM 450 array. Breakpoint-associated CpG sites were usually highly methylated (median methylation  $\beta$ -values in TICs and xenografts ranging from 0.63 to 0.84). To correlate xenograft-specific changes in the SCNA-profiles with possible changes in CpG methylation, we compared methylation  $\beta$ -values of CpG sites located within breakpoint-sections specific for xenografts to the methylation value of corresponding section in the TICs.

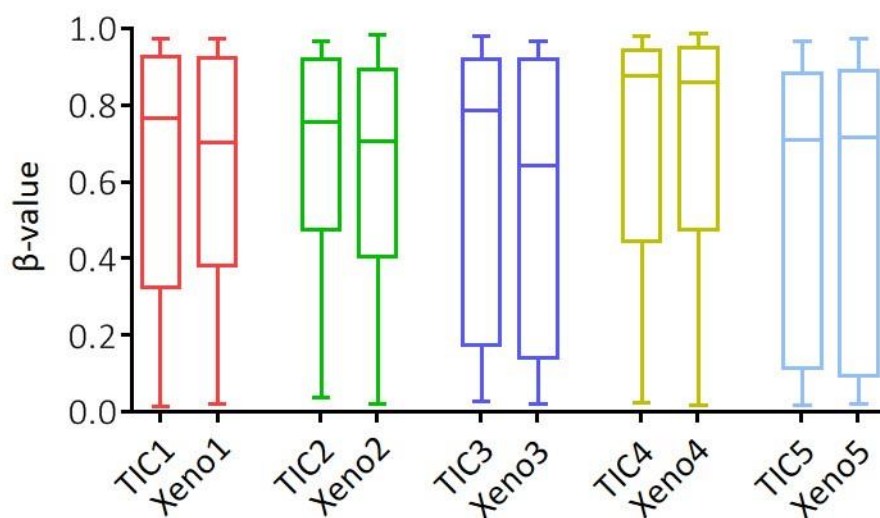
Figure 4-2



**Figure 4-2: Characterization of TICs and xenografts by genome-wide and target-wise methylation analysis.** a) Illumina Infinium HumanMethylation450 data of TIC and xenografts. Heat map illustrating methylation levels ( $\beta$ -values) of the 25% most variable CpG sites (6571 sites; blue, low methylation; red, high methylation). Unsupervised clustering is indicated at the top. b) Methylation analysis of the CIMP markers MLH1, GREM1, NPY1R and CA4 in 5 TICs using In-qMSP. PMA, percentage methylated alleles.

Xenografts from TIC2 and TIC3 showed a trend of reduced methylation levels at xenograft-specific breakpoint sections (Figure 4-3). However, this trend did not reach statistical significance (xenograft 2:  $p=0.113$ ; xenograft 3:  $p=0.233$ ; Wilcoxon matched-pairs test). Other TIC/xenograft pairs did not show a correlation between the occurrence of chromosomal breakpoints and methylation levels.

Figure 4-3



**Figure 4-3: Methylation changes of breakpoint-associated CpG during xenograft formation.** Depicted are the methylation  $\beta$ -values of CpGs in chromosome sections associated with breakpoints specific for xenografts (Xeno). Methylation  $\beta$ -values of the corresponding chromosome section in TICs (TIC) are also depicted. Box plots indicate the median value and the first and third quartile. Whiskers indicate the minimum and maximum values.

In summary, our preliminary analysis of SCNA-profiles in TICs and mouse xenografts derived from them, distinguishes cancer cells not only according to the extent of chromosomal aberrations but also with regard to the inherent chromosomal instability. TIC2, TIC4 and TIC5 displayed a high level of numerical chromosomal alterations but only in TIC2 the copy number profile was unstable and changed during xenograft formation. This reflects different molecular mechanisms and deficiencies associated with the generation of chromosomal aberrations.

Further, the currently existing results show that in some TIC/xenograft pairs (TIC2 and TIC3) the loss of CpG methylation at breakpoint regions might be associated with emerging chromosomal aberrations. However, only for a small fraction of chromosomal breakpoint regions were covered by methylation data from the HM450 array and further investigation will require a higher density of methylation data (genome-wide bisulfite DNA sequencing) and a detailed high resolution analysis of methylation changes between TICs and xenografts. Interestingly, the indication for an association between DNA hypomethylation and the emergence of chromosome breakage was strongest for TIC2, which displayed the highest instability of SCNA patterns during xenograft formation, and for TIC3, which is associated with a methylator phenotype.

## **Methods**

### **Colorectal Cancer Initiating Cells**

Established cell lines from colorectal cancer initiating cells (CR-CICs) were kindly provided by Lucia Ricci-Vitiani from the Istituto Superiore di Sanità in Rome, Italy. CIC cell lines were established as described in (Cammareri et al., 2008). Briefly, surgical specimens were washed in PBS and incubated overnight in DMEM/F12 medium containing antibiotics and antimycotics at high concentrations. Afterwards, tissues were mechanically and enzymatically disrupted, filtered and seeded onto ultra-low adhesion flasks with serumfree DMEM/F12 medium containing EGF and FGF. Under these conditions self-renewing stem like cells could be continuously cultured whereas differentiated cells progressively died out. The propagating cell population had cancer-initiating potential in immunosuppressed mice and showed expression of putative colorectal cancer stem cell markers including CD133. Genomic DNA from CR-CICs was extracted with the QIAamp DNA mini kit (Qiagen) according to the manufacturer's instructions including the RNase-treatment step.

### **Mouse Xenografts**

Cancer initiating cells were resuspended in a 1:1 mixture of phosphate-buffered saline and growth factor-reduced Matrigel matrix (BD Biosciences) and inoculated subcutaneously into the flank of recipient mice. After 8-10 weeks mice were sacrificed, tumors were removed and genomic DNA was extracted as described for CR-CICs.

## **Copy Number Variation Analysis**

Genomic DNA from CICs, xenograft tissues, two human embryonic stem cell lines (SA001 and SA167) and a normal human colonic epithelium specimen (HT2010) was isolated as described above. DNA quality and fragmentation assessment was performed using 50 ng of genomic DNA loaded onto a microfluidic Agilent DNA 12000 chip in combination with an Agilent 2100 bioanalyzer. 0.5 µg of DNA samples were labeled with random-nonamer Cy3 primer and Promega reference DNA samples (Blood DNA, Promega) were labeled with random-nonamer Cy5 primer using isothermal PCR according to the manufacturer's instructions (Roche-Nimblegen, Madison, Wisconsin, USA). Cy3 and Cy5 labeled DNA was co-hybridized for ~70 h at 42°C to the NimbleGen human Whole Genome Tiling Array (design ID: 100718\_HG18\_WG\_CGH\_v2D\_HX1) that contains 2.1 million isothermal ~60mer oligonucleotide probes with a median probe spacing of 1169 bp (Roche-Nimblegen, Madison, Wisconsin, USA). After washing and drying microarray data was collected by confocal scanning using the Roche NimbleGen MS200 Microarray Scanner at 2 µm pixel resolution (Roche NimbleGen). Using the software NimbleScan version 2.6 (Roche NimbleGen), microarray probe intensities were subjected LOESS Spatial Correction to adjust signal intensities based on probe position (Smyth and Speed, 2003), qspline fit normalization to compensate for inherent differences in signal between the Cy3 and Cy5 dyes (Workman et al., 2002) and copy number variation analysis to detect log<sub>2</sub> (test/reference) values that deviate significantly from baseline using the SegMNT algorithm. A set of conservative criteria (five or more probes in a segment, mean amplitude of log<sub>2</sub> shift across segment +/- 0.5) were used to define the final set of high confidence CNV calls.

## **Genomewide methylation analysis**

Array-based gene-specific DNA methylation analysis was performed using Infinium HumanMethylation450 bead chip technology (Illumina, San Diego, CA, USA). Genomic DNA (500 ng) from each sample was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research Corporation). Bisulfite treated genomic DNA was whole-genome amplified and hybridized to HumanMethylation450 BeadChips. The methylation status of a specific cytosine was indicated by average beta (AVB) values where 1 corresponds to full methylation and 0 to no methylation. Raw array data were quantile normalised as described previously (Touleimat and Tost, 2012) and p-values for comparisons between different datasets were statistically adjusted using the Benjamini-Hochberg correction. For the data analysis we masked data point as "NA" for the following conditions 1) Probes containing single-nucleotide polymorphisms (SNPs), 2) those that overlap with a repetitive element that covers the targeted CpG dinucleotide, 3) those that overlap with regions of insertions and deletions in the human

genome, 4) probes that are targeting X or Y chromosome. Beta values with detection p value greater than 0.05 were also replaced as "NA". Furthermore, probes containing at least one "NA" across any sample were also masked as "NA". Methylation analysis was performed using R package limma.

### Locus-normalized quantitative methylation-specific PCR

All qMSP reactions were carried out on the Rotor-Gene 3000 real time thermal cycler (Qiagen) using the QuantiTect® SYBR® Green Kit (Qiagen) according to the manufacturer's instructions. 15µl reactions were performed containing 0.5µM of each primer as well as 1.5µl bisulfite-converted DNA. DNA from a peripheral blood sample was treated with M.SSSI methyltransferase to obtain full methylation within the promoter regions of all target genes (methylation positive control). For each target gene standard curves for the M- (specific for bisulfite-unconverted sequence) and S- (standardizer; methylation independent) primers were established by amplifying serial 1:10 dilutions (100% - 0.1%) of the bisulphite treated positive control DNA. Methylation of target genes (expressed as percentage of methylated alleles [PMA]) in sample DNA was quantified by relating M-amplification rate to the S-amplification rate both of which were previously determined using the recorded standard curves. DNA from peripheral blood showed no or only minor fractions of methylated alleles for the target genes and was therefore used as a methylation negative control. For In-qMSP primer sequences see Table 4-2.

**Table 4-1: Clinical characteristics of patients and molecular features of TICs**

TIC	Gender	Age	Tumor grade	APC	TP53	KRAS	PIK3CA	Micro-satellite
1	m	68	G3	mut	mut	wt	mut	stable
2	m	68	G2	mut	wt	wt	mut	stable
3	f	51	G2	mut	mut	mut	wt	instable
4	f	63	G2	mut	mut	wt	wt	stable
5	m	n/a	G3	mut	mut	wt	wt	stable

wt, wiltype; mut, mutated

**Table 4-2: In-qMSP primers**

Name	Orientation	Sequence (5´-3´)
MLH1 M	F	ACGCAGACGCTCCACCAGGGCCGC
MLH1 M	R	TCCGGAGGGCGATGGGGCCCTGTGC

MLH1 S	F	AGTTTGGAGTGGTAGGTTTTTAGAGG
MLH1 S	R	CATCCCCAAATTCTTACTCCTTCTTTC
CA4 M	F	TTTCGGGTTGCGTGTGCGTC
CA4 M	R	CTCGTCTCGCCGAACCACG
CA4 S	F	GGGTATTGGTTGAAGATAAAGGGGAAT
CA4 S	R	TTCTCCCACTCTTACAATACACCCTACC
NPY1R M	F	GCGCGTCGTTTAGCGTCGTAC
NPY1R M	R	CGCGACTCGATTACAACCGAAAAACG
NPY1R S	F	GGATAATTTTTAGGTTTTGGTTTGAGGGAG
NPY1R S	R	TTACATATACCCAACCCTCAATCTTTACC
GREM1 M	F	CGTCGGTATTTAAACGGGAGACG
GREM1 M	R	CAAACCGCCGAAACTCGACG
GREM1 S	F	GTGTGAATATGGTGGTTTGTGGTATG
GREM1 S	R	CACAATAAATAACAAACATCCACTCCTAC

M, Methylation-specific primer; S, Standardizer primer; F, forward; R, reverse

## 5 Concluding Discussion and Outlook

Colorectal cancer features a multitude of epigenetic aberrations, and although these aberrations are well documented at the histone and DNA modification levels, little is known about their origin and the underlying molecular defects. During my PhD studies, I addressed, on the one hand, the molecular processes underlying the wide-spread changes of DNA methylation in CRC and, on the other hand, the consequences of malfunction of these mechanisms in colorectal carcinogenesis.

Methylation alterations in CRC are particularly pronounced in cancers developing through the CIMP pathway, which is characterized by the hypermethylation of multiple gene promoters. Previous studies linked hypermethylation phenotypes in glioma and leukemia with impaired TET-mediated DNA demethylation (Figuroa et al., 2010; Letouze et al., 2013; Turcan et al., 2012). Together my colleagues, I discovered that the loss of TET1 expression is characteristic of CIMP-associated CRCs (**Appendix I**). We showed that TET1 depletion in the non-CIMP CRC cell line SW620 results in genome-wide CpG island-associated methylation changes which included hyper- as well as hypomethylation events. These initial results indicate that the loss of TET protein function does not primarily generate a hypermethylation phenotype. This is in line with previous studies showing that Tet1 loss in mouse embryonic stem cells results in hyper- and hypomethylation (Y. Xu et al., 2011). Tumors derived from engrafting of TET1 depleted cells into immunosuppressed mice then displayed a shift in the methylation patterns towards hypermethylation. This may indicate the genetic consequences of CGI hypermethylation are directly relevant for cancer development and thus provide a selective advantage in tumor evolution. Indeed, we observed transcriptional repression and hypermethylation of MLH1 in TET1 depleted cells, a hallmark of CRC-CIMP. Further, we found that TET1 depletion results in the downregulation of LGR5, an established marker of epithelial intestinal stem cells as well as of colorectal cancer initiating cells (Koo and Clevers, 2014). E-cadherin (CDH1) is an essential component of intercellular adherence junctions and therefore an important marker for epithelial cells. CDH1 expression was downregulated in TET1 depleted cells whereas the mesenchymal markers vimentin and fibronectin were upregulated. In addition TET1 depleted cells showed an increased migratory behavior. These are features of an epithelial-to-mesenchymal transition (EMT), which has been associated a poor cancer prognosis as increased cell migration and reduced cell-to-cell contacts are linked to metastasis formation (Findlay et al., 2014). Together, these results show that the loss of TET1 not only generates a DNA hypermethylation phenotype and a loss of tissue specific stem cell features, but also to the concomitant acquisition of mesenchymal cell properties. The latter might suggest that the loss of TET1 is associated with poor cancer prognosis. In line with, this a

recently published study indicates that CIMP-CRCs have a worse prognosis compared to non-CIMP cancers (Juo et al., 2014). Impaired DNA demethylation due to TET1 depletion and the ensuing restriction of DNA methylation plasticity is likely to affect cell state plasticity. We isolated and expanded differentially CD24/CD44–marked subpopulations of TET1 depleted cells and control cells and found that TET1 depleted cells can hardly re-establish the original heterogeneity. We propose that TET1 depleted cells are locked in the cell state at the time TET1 was lost and fail to memorize differentiation cues from the environment. Notably, it has been recently reported that the BRAF V600E mutation, which is highly associated with CRC-CIMP, results in the downregulation of TET1 (Kudo et al., 2012). This makes oncogenic BRAF a promising upstream factor of TET1 downregulation in CRC-CIMP. Overall my results provide novel insight into the molecular mechanisms underlying the CIMP oncogenic pathway in CRC. In future studies it will be interesting to address the molecular processes by which oncogenic BRAF affects the expression of TET1.

A significant fraction of methylation aberrations described in CRC can be linked to aging (Toyota et al., 1999). It has been reported previously that environmental factors may influence age-dependent methylation changes in the colorectum (Tapp et al., 2013). To further investigate the effect of lifestyle-related factors in the dynamics of methylation changes in the aging colon we quantified promoter methylation levels target-wise and genome-wide in colonic mucosa specimens from a large cohort of 546 healthy women (**Appendix II**). We correlated methylation data with CRC-risk modulating lifestyle factors smoking, overweight, hormone replacement therapy (HRT) and aspirin use (Chan et al., 2007; Long et al., 2010; Parajuli et al., 2013; Robsahm et al., 2013). The regular use of aspirin and HRT reduced the number of age-hypermethylated CpG sites as well as the rate of their methylation increase. Obesity, however, promoted age-related hypermethylation. Interestingly, long-term smoking was associated with increased methylation changes in the proximal (right) colon but suppressed age-dependent hypermethylation at the distal (left) colonic side. A significant fraction of CpGs, the age-related hypermethylation of which was modulated by lifestyle factors, is hypermethylated in CRC. The identification of this age- and cancer-related subgroup is of particular interest. First, it provides insight into the molecular mechanisms by which lifestyle-related factors may exert their effect on CRC risk and, second, the hypermethylation at these CpG sites implies a risk for CRC. Hence, the identification and discrimination of a full spectrum of age- and cancer-related hypermethylation targets allows the establishment of biomarkers for subtype specific CRC prediction, i.e. risk assessment, diagnosis and monitoring of treatment response. Since, our studies focused on women for technical reasons, the findings will have to be validated by studies including men.



DNA methylation is involved in the regulation of numerous cellular processes including transcription, DNA repair and replication. Therefore the consequences of methylation aberrations in cancer are multifaceted. I contributed to an original study, which identified SHOCA-2 as a tumor suppressor in CRC. We were able to show that SHOCA-2 inhibits EGFR signaling by suppressing STAT3 activation (**Appendix III**). The loss of SHOCA-2 expression correlated with advanced CRC stages, a poor clinical outcome and was associated with different genetic and epigenetic alterations. In a large fraction of SHOCA-2-depleted cancers the expression loss was associated with chromosomal deletion affecting the SH2D4A locus (encoding for SHOCA-2) resulting in a loss of heterozygosity (LOH) for a point mutation causing premature termination. Additional DNA sequence alterations, resulting in frame shift or missense mutations, were not associated with LOH. In addition, within the scope of this collaboration, I showed that cancer-specific hypermethylation of two CpG sites, part of a CGI located in the 5'-UTR of the SH2D4A gene, is associated with SHOCA-2 downregulation in a small fraction of SHOCA-2-depleted cancers. Interestingly, other CpGs in the CGI did not show cancer-specific methylation changes. The specific hypermethylation of these two CpG sites suggests an underlying selective process, which promotes and preserves accumulation of methylation at these sites. One of the two hypermethylated CpG sites overlapped with a consensus binding motif for the transcription factor Sp1. In an electrophoretic mobility shift assay (EMSA) I could show that methylation at this CpG significantly reduced the affinity of Sp1 to its binding motif *in vitro*. This could indeed be sufficient to explain a loss of SHOCA-2 expression. However, different aspects need further investigation. For instance, it is unknown whether or to what extent Sp1 indeed contributes to transcriptional activation of SHOCA-2. Furthermore, my findings do not imply that methylation of the respective CpG site does also interfere with Sp1 binding *in vivo*. There is some disagreement in the literature over the effect of methylation on Sp1 binding to its consensus sequence, although several recent studies confirm the methylation sensitivity of Sp1 binding (Harrington et al., 1988; Reynard et al., 2014; X. Zhang et al., 2014). Despite these uncertainties, our results introduce a pathway by which well-defined, aberrant methylation events can contribute to colorectal carcinogenesis.

It has been reported that hypomethylation of fragile sites in the genome is associated chromosome breakage (De and Michor, 2011). However, it is unknown whether hypomethylation is causally linked to the emergence of chromosome breakage. To address this question, I and colleagues from different laboratories at Roche, Basel, analyzed somatic copy number alterations (SCNAs) profiles and genome-wide CpG methylation patterns in five pairs of tumor-initiating cells (TICs) and mouse xenografts derived from these cells (**Supplementary Results 4.5.1**). SCNA analysis revealed that cancer cells differed in the extent of numerical chromosomal aberrations as well as in the stability of these aberrations. Whereas for most TICs SCNA patterns remained largely unchanged in the course of xenograft

formation, in one TIC the patterns were unstable and xenografts acquired additional arm-level SCNAs. The difference in the stability of SCNA profiles among different CRC samples have not been reported, yet, and could reflect different molecular mechanisms underlying chromosomal aberrations. Consistent with previous findings we found that cancer cells featuring a CGI hypermethylation phenotype show a lower level of chromosomal aberrations (Cancer Genome Atlas, 2012; Cheng et al., 2008). To address whether SCNA changes during xenograft formation are associated with methylation changes, we correlated methylation levels of CpG sites located in regions of chromosomal breakpoints specific for xenografts and compared them to the levels in corresponding region in TICs, where these sites were not associated with breakpoints. Two TIC/xenograft pairs indeed displayed a trend for reduced methylation levels at chromosomal breakpoints. The differences in the association of differential methylation and chromosomal breakage may again reflect different mechanisms underlying the accumulation of SCNAs. Notably, the TICs in which we observed an association between hypomethylation and the emergence of chromosomal breakpoints displayed continuous instability of SCNAs or were associated with a hypermethylation phenotype, respectively. Although, the preliminary results from our TIC/xenograft model introduce several interesting aspects regarding the development of genetic and epigenetic profiles during tumor formation, the validity of our findings is limited due to the small number of TIC/xenograft pairs. Further detailed analysis of TICs and xenografts will be required to address mechanistic aspects of SCNA-profile instability which, furthermore, needs to be addressed in primary CRCs by future studies.

Aberrant methylation occurs early in colorectal carcinogenesis and is partially related to the aging process as discussed before (Menigatti et al., 2009; Noreen et al., 2014). This allows for the establishment of specific biomarkers for diagnosis and CRC risk assessment. In collaboration with my colleagues, I contributed to the identification of seven genes (CA4, NPY1R, IFITM1, GREM1, FOXF1, MLH1 and MGMT) downregulated in precancerous benign adenomas compared to matched normal colonic mucosa and to the analysis of their promoter methylation in matched samples from primary human CRCs, healthy mucosa and peripheral blood (**Appendix IV**). This marker panel could discriminate between CRC and matched normal mucosa with a sensitivity of 82% and a specificity of 99%, when the criterion of at least one methylated marker was applied. The sensitivity was even higher when the panel was applied to CRC located at the proximal side of the colon. The seven marker panel therefore provides a potential tool for cancer diagnosis and CRC risk assessment with a high specificity and sensitivity. As it was reported previously for MLH1, we detected low levels of CA4, NPY1R, GREM1 promoter methylation also the mucosa from healthy individuals suggesting that the level rather than the presence of methylation in the promoter of the marker genes distinguishes healthy from cancerous tissues (Menigatti et al., 2009). This has implications for the clinical

utilization of methylation biomarkers, as quantitative rather than qualitative approaches are required to determine pathologic DNA methylation changes.

In the course of my PhD thesis I produced data that may contribute to the understanding of how DNA methylation aberrations in human CRC occur and how they can promote cancerous transformation.

## 6 References

- Al-Sohaily, S., Biankin, A., et al. (2012). Molecular pathways in colorectal cancer. *J Gastroenterol Hepatol*, 27(9), 1423-1431. doi: 10.1111/j.1440-1746.2012.07200.x
- Alexandrov, L. B., Nik-Zainal, S., et al. (2013). Signatures of mutational processes in human cancer. *Nature*, 500(7463), 415-421. doi: 10.1038/nature12477
- Amaral, P. P., Dinger, M. E., et al. (2008). The eukaryotic genome as an RNA machine. *Science*, 319(5871), 1787-1789. doi: 10.1126/science.1155472
- Ang, P. W., Loh, M., et al. (2010). Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer*, 10, 227. doi: 10.1186/1471-2407-10-227
- Aoki, K., Aoki, M., et al. (2007). Chromosomal instability by beta-catenin/TCF transcription in APC or beta-catenin mutant cells. *Oncogene*, 26(24), 3511-3520. doi: 10.1038/sj.onc.1210141
- Barski, A., Cuddapah, S., et al. (2007). High-resolution profiling of histone methylations in the human genome. *Cell*, 129(4), 823-837. doi: 10.1016/j.cell.2007.05.009
- Baudis, M. (2007). Genomic imbalances in 5918 malignant epithelial tumors: an explorative meta-analysis of chromosomal CGH data. *BMC Cancer*, 7, 226. doi: 10.1186/1471-2407-7-226
- Becker, L., Huang, Q., et al. (2008). Immunostaining of Lgr5, an intestinal stem cell marker, in normal and premalignant human gastrointestinal tissue. *ScientificWorldJournal*, 8, 1168-1176. doi: 10.1100/tsw.2008.148
- Bell, A. C., & Felsenfeld, G. (2000). Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature*, 405(6785), 482-485. doi: 10.1038/35013100
- Beroukhim, R., Mermel, C. H., et al. (2010). The landscape of somatic copy-number alteration across human cancers. *Nature*, 463(7283), 899-905. doi: 10.1038/nature08822
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev*, 16(1), 6-21. doi: 10.1101/gad.947102
- Blobe, G. C., Schieman, W. P., et al. (2000). Role of transforming growth factor beta in human disease. *N Engl J Med*, 342(18), 1350-1358. doi: 10.1056/NEJM200005043421807
- Bogdanovic, O., & Veenstra, G. J. (2009). DNA methylation and methyl-CpG binding proteins: developmental requirements and function. *Chromosoma*, 118(5), 549-565. doi: 10.1007/s00412-009-0221-9
- Boland, C. R., & Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology*, 138(6), 2073-2087 e2073. doi: 10.1053/j.gastro.2009.12.064
- Burns, M. B., Temiz, N. A., et al. (2013). Evidence for APOBEC3B mutagenesis in multiple human cancers. *Nat Genet*, 45(9), 977-983. doi: 10.1038/ng.2701
- Cahill, D. P., da Costa, L. T., et al. (1999). Characterization of MAD2B and other mitotic spindle checkpoint genes. *Genomics*, 58(2), 181-187. doi: 10.1006/geno.1999.5831
- Cammareri, P., Lombardo, Y., et al. (2008). Isolation and culture of colon cancer stem cells. *Methods Cell Biol*, 86, 311-324. doi: 10.1016/S0091-679X(08)00014-9
- Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330-337. doi: 10.1038/nature11252
- Carmona, F. J., Davalos, V., et al. (2014). A Comprehensive DNA Methylation Profile of Epithelial-to-Mesenchymal Transition. *Cancer Res*, 74(19), 5608-5619. doi: 10.1158/0008-5472.CAN-13-3659
- Chan, A. T., Ogino, S., et al. (2007). Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. *N Engl J Med*, 356(21), 2131-2142. doi: 10.1056/NEJMoa067208
- Chen, J. S., Hung, W. S., et al. (2013). In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. *Bioinformatics*, 29(4), 420-427. doi: 10.1093/bioinformatics/bts715

- Chen, J. X., Zheng, Y., et al. (1998). Carcinogens preferentially bind at methylated CpG in the p53 mutational hot spots. *Cancer Res*, 58(10), 2070-2075.
- Chen, Y., Wang, K., et al. (2013). DNA methylation is associated with transcription of Snail and Slug genes. *Biochem Biophys Res Commun*, 430(3), 1083-1090. doi: 10.1016/j.bbrc.2012.12.034
- Cheng, Y. W., Pincas, H., et al. (2008). CpG island methylator phenotype associates with low-degree chromosomal abnormalities in colorectal cancer. *Clin Cancer Res*, 14(19), 6005-6013. doi: 10.1158/1078-0432.CCR-08-0216
- Chughtai, S. A., Crundwell, M. C., et al. (1999). Two novel regions of interstitial deletion on chromosome 8p in colorectal cancer. *Oncogene*, 18(3), 657-665. doi: 10.1038/sj.onc.1202340
- Cimmino, L., Abdel-Wahab, O., et al. (2011). TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell*, 9(3), 193-204. doi: 10.1016/j.stem.2011.08.007
- Cole, M. F., Johnstone, S. E., et al. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev*, 22(6), 746-755. doi: 10.1101/gad.1642408
- Colot, V., & Rossignol, J. L. (1999). Eukaryotic DNA methylation as an evolutionary device. *Bioessays*, 21(5), 402-411. doi: 10.1002/(SICI)1521-1878(199905)21:5<402::AID-BIES7>3.0.CO;2-B
- Conrad, D. F., Pinto, D., et al. (2010). Origins and functional impact of copy number variation in the human genome. *Nature*, 464(7289), 704-712. doi: 10.1038/nature08516
- Cortazar, D., Kunz, C., et al. (2007). The enigmatic thymine DNA glycosylase. *DNA Repair (Amst)*, 6(4), 489-504. doi: DOI 10.1016/j.dnarep.2006.10.013
- Cortazar, D., Kunz, C., et al. (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. *Nature*, 470(7334), 419-U210. doi: Doi 10.1038/Nature09672
- Cortellino, S., Xu, J., et al. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell*, 146(1), 67-79. doi: 10.1016/j.cell.2011.06.020
- Croce, C. M. (2009). Causes and consequences of microRNA dysregulation in cancer. *Nat Rev Genet*, 10(10), 704-714. doi: 10.1038/nrg2634
- Cuddapah, S., Jothi, R., et al. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. *Genome Res*, 19(1), 24-32. doi: 10.1101/gr.082800.108
- D'Alessio, A. C., & Szyf, M. (2006). Epigenetic tete-a-tete: the bilateral relationship between chromatin modifications and DNA methylation. *Biochem Cell Biol*, 84(4), 463-476. doi: 10.1139/o06-090
- Dame, M. K., Jiang, Y., et al. (2014). Human colonic crypts in culture: segregation of immunochemical markers in normal versus adenoma-derived. *Lab Invest*, 94(2), 222-234. doi: 10.1038/labinvest.2013.145
- Davalos, V., Moutinho, C., et al. (2012). Dynamic epigenetic regulation of the microRNA-200 family mediates epithelial and mesenchymal transitions in human tumorigenesis. *Oncogene*, 31(16), 2062-2074. doi: 10.1038/onc.2011.383
- Dawlaty, M. M., Ganz, K., et al. (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. *Cell Stem Cell*, 9(2), 166-175. doi: 10.1016/j.stem.2011.07.010
- Dawson, M. A., & Kouzarides, T. (2012). Cancer epigenetics: from mechanism to therapy. *Cell*, 150(1), 12-27. doi: 10.1016/j.cell.2012.06.013
- De, S., & Michor, F. (2011). DNA secondary structures and epigenetic determinants of cancer genome evolution. *Nat Struct Mol Biol*, 18(8), 950-955. doi: 10.1038/nsmb.2089
- Delhommeau, F., Dupont, S., et al. (2009). Mutation in TET2 in myeloid cancers. *N Engl J Med*, 360(22), 2289-2301. doi: 10.1056/NEJMoa0810069
- Dhillon, A. S., Hagan, S., et al. (2007). MAP kinase signalling pathways in cancer. *Oncogene*, 26(22), 3279-3290. doi: 10.1038/sj.onc.1210421
- Dialynas, G. K., Vitalini, M. W., et al. (2008). Linking Heterochromatin Protein 1 (HP1) to cancer progression. *Mutation Research-Fundamental and Molecular Mechanisms of Mutagenesis*, 647(1-2), 13-20. doi: DOI 10.1016/j.mrfmmm.2008.09.007
- Durkin, S. G., & Glover, T. W. (2007). Chromosome fragile sites. *Annu Rev Genet*, 41, 169-192. doi: 10.1146/annurev.genet.41.042007.165900

- Duval, A., & Hamelin, R. (2002). Genetic instability in human mismatch repair deficient cancers. *Ann Genet*, 45(2), 71-75.
- Ehrlich, M., Gama-Sosa, M. A., et al. (1982). Amount and distribution of 5-methylcytosine in human DNA from different types of tissues of cells. *Nucleic Acids Res*, 10(8), 2709-2721.
- Emi, M., Fujiwara, Y., et al. (1992). Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. *Cancer Res*, 52(19), 5368-5372.
- Esteller, M. (2011). Non-coding RNAs in human disease. *Nat Rev Genet*, 12(12), 861-874. doi: 10.1038/nrg3074
- Fan, H., Zhao, X., et al. (2013). Function of focal adhesion kinase scaffolding to mediate endophilin A2 phosphorylation promotes epithelial-mesenchymal transition and mammary cancer stem cell activities in vivo. *J Biol Chem*, 288(5), 3322-3333. doi: 10.1074/jbc.M112.420497
- Fatemi, M., Hermann, A., et al. (2001). The activity of the murine DNA methyltransferase Dnmt1 is controlled by interaction of the catalytic domain with the N-terminal part of the enzyme leading to an allosteric activation of the enzyme after binding to methylated DNA. *J Mol Biol*, 309(5), 1189-1199. doi: 10.1006/jmbi.2001.4709
- Fearon, E. R., & Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell*, 61(5), 759-767.
- Figueroa, M. E., Abdel-Wahab, O., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*, 18(6), 553-567. doi: 10.1016/j.ccr.2010.11.015
- Findlay, V. J., Wang, C., et al. (2014). Epithelial-to-mesenchymal transition and the cancer stem cell phenotype: insights from cancer biology with therapeutic implications for colorectal cancer. *Cancer Gene Ther*, 21(5), 181-187. doi: 10.1038/cgt.2014.15
- Forster, R., Chiba, K., et al. (2014). Human intestinal tissue with adult stem cell properties derived from pluripotent stem cells. *Stem Cell Reports*, 2(6), 838-852. doi: 10.1016/j.stemcr.2014.05.001
- Freudenberg, J. M., Ghosh, S., et al. (2012). Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res*, 40(8), 3364-3377. doi: 10.1093/nar/gkr1253
- Fudenberg, G., Getz, G., et al. (2011). High order chromatin architecture shapes the landscape of chromosomal alterations in cancer. *Nat Biotechnol*, 29(12), 1109-1113. doi: 10.1038/nbt.2049
- Fukuyama, R., Nicolaita, R., et al. (2008). Mutated in colorectal cancer, a putative tumor suppressor for serrated colorectal cancer, selectively represses beta-catenin-dependent transcription. *Oncogene*, 27(46), 6044-6055. doi: 10.1038/onc.2008.204
- Ganem, N. J., & Pellman, D. (2007). Limiting the proliferation of polyploid cells. *Cell*, 131(3), 437-440. doi: 10.1016/j.cell.2007.10.024
- Gao, Y., Chen, J., et al. (2013). Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. *Cell Stem Cell*, 12(4), 453-469. doi: 10.1016/j.stem.2013.02.005
- Gibbons, R. J. (2005). Histone modifying and chromatin remodelling enzymes in cancer and dysplastic syndromes. *Hum Mol Genet*, 14 Spec No 1, R85-92. doi: 10.1093/hmg/ddi106
- Gisselsson, D., Pettersson, L., et al. (2000). Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity. *Proc Natl Acad Sci U S A*, 97(10), 5357-5362. doi: 10.1073/pnas.090013497
- Goel, A., & Boland, C. R. (2012). Epigenetics of colorectal cancer. *Gastroenterology*, 143(6), 1442-1460 e1441. doi: 10.1053/j.gastro.2012.09.032
- Gordon, C. A., Hartono, S. R., et al. (2013). Inactive DNMT3B splice variants modulate de novo DNA methylation. *PLoS One*, 8(7), e69486. doi: 10.1371/journal.pone.0069486
- Gordon, D. J., Resio, B., et al. (2012). Causes and consequences of aneuploidy in cancer. *Nat Rev Genet*, 13(3), 189-203. doi: 10.1038/nrg3123
- Griswold, M. D., & Kim, J. S. (2001). Site-specific methylation of the promoter alters deoxyribonucleic acid-protein interactions and prevents follicle-stimulating hormone receptor gene transcription. *Biol Reprod*, 64(2), 602-610.

- Gu, T. P., Guo, F., et al. (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature*, *477*(7366), 606-610. doi: 10.1038/nature10443
- Guo, A., Villen, J., et al. (2008). Signaling networks assembled by oncogenic EGFR and c-Met. *Proc Natl Acad Sci U S A*, *105*(2), 692-697. doi: 10.1073/pnas.0707270105
- Guo, J. U., Ma, D. K., et al. (2011). Neuronal activity modifies the DNA methylation landscape in the adult brain. *Nat Neurosci*, *14*(10), 1345-1351. doi: 10.1038/nn.2900
- Guo, J. U., Su, Y., et al. (2011). Emerging roles of TET proteins and 5-hydroxymethylcytosines in active DNA demethylation and beyond. *Cell Cycle*, *10*(16), 2662-2668.
- Gupta, P. B., Fillmore, C. M., et al. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, *146*(4), 633-644. doi: 10.1016/j.cell.2011.07.026
- Gupta, R. A., Shah, N., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature*, *464*(7291), 1071-1076. doi: 10.1038/nature08975
- Hadjihannas, M. V., Bruckner, M., et al. (2006). Aberrant Wnt/beta-catenin signaling can induce chromosomal instability in colon cancer. *Proc Natl Acad Sci U S A*, *103*(28), 10747-10752. doi: 10.1073/pnas.0604206103
- Haigis, K. M., Kendall, K. R., et al. (2008). Differential effects of oncogenic K-Ras and N-Ras on proliferation, differentiation and tumor progression in the colon. *Nat Genet*, *40*(5), 600-608. doi: 10.1038/ng.115
- Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. *Cell*, *144*(5), 646-674. doi: 10.1016/j.cell.2011.02.013
- Hansen, K. D., Timp, W., et al. (2011). Increased methylation variation in epigenetic domains across cancer types. *Nat Genet*, *43*(8), 768-775. doi: 10.1038/ng.865
- Harrington, M. A., Jones, P. A., et al. (1988). Cytosine methylation does not affect binding of transcription factor Sp1. *Proc Natl Acad Sci U S A*, *85*(7), 2066-2070.
- Harris, R. S., & Liddament, M. T. (2004). Retroviral restriction by APOBEC proteins. *Nat Rev Immunol*, *4*(11), 868-877. doi: 10.1038/nri1489
- Haruki, N., Saito, H., et al. (2001). Molecular analysis of the mitotic checkpoint genes BUB1, BUBR1 and BUB3 in human lung cancers. *Cancer Lett*, *162*(2), 201-205.
- Hata, K., Okano, M., et al. (2002). Dnmt3L cooperates with the Dnmt3 family of de novo DNA methyltransferases to establish maternal imprints in mice. *Development*, *129*(8), 1983-1993.
- He, Y. F., Li, B. Z., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, *333*(6047), 1303-1307. doi: 10.1126/science.1210944
- Heuberger, J., & Birchmeier, W. (2010). Interplay of cadherin-mediated cell adhesion and canonical Wnt signaling. *Cold Spring Harb Perspect Biol*, *2*(2), a002915. doi: 10.1101/cshperspect.a002915
- Hinoue, T., Weisenberger, D. J., et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*, *22*(2), 271-282. doi: 10.1101/gr.117523.110
- Hirsch, D., Barker, N., et al. (2014). LGR5 positivity defines stem-like cells in colorectal cancer. *Carcinogenesis*, *35*(4), 849-858. doi: 10.1093/carcin/bgt377
- Huang, F. W., Hodis, E., et al. (2013). Highly recurrent TERT promoter mutations in human melanoma. *Science*, *339*(6122), 957-959. doi: 10.1126/science.1229259
- Huang, Y., Chavez, L., et al. (2014). Distinct roles of the methylcytosine oxidases Tet1 and Tet2 in mouse embryonic stem cells. *Proc Natl Acad Sci U S A*, *111*(4), 1361-1366. doi: 10.1073/pnas.1322921111
- Hubner, M. R., Eckersley-Maslin, M. A., et al. (2013). Chromatin organization and transcriptional regulation. *Curr Opin Genet Dev*, *23*(2), 89-95. doi: 10.1016/j.cde.2012.11.006
- Inoue, A., & Zhang, Y. (2011). Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science*, *334*(6053), 194. doi: 10.1126/science.1212483
- Irizarry, R. A., Ladd-Acosta, C., et al. (2009). The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet*, *41*(2), 178-186. doi: 10.1038/ng.298

- Issa, J. P. (2014). Aging and epigenetic drift: a vicious cycle. *J Clin Invest*, *124*(1), 24-29. doi: 10.1172/JCI69735
- Ito, S., D'Alessio, A. C., et al. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, *466*(7310), 1129-1133. doi: 10.1038/nature09303
- Ito, S., Shen, L., et al. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science*, *333*(6047), 1300-1303. doi: 10.1126/science.1210597
- Janssen, A., & Medema, R. H. (2013). Genetic instability: tipping the balance. *Oncogene*, *32*(38), 4459-4470. doi: 10.1038/onc.2012.576
- Jansson, M. D., & Lund, A. H. (2012). MicroRNA and cancer. *Mol Oncol*, *6*(6), 590-610. doi: 10.1016/j.molonc.2012.09.006
- Jeltsch, A. (2002). Beyond Watson and Crick: DNA methylation and molecular enzymology of DNA methyltransferases. *Chembiochem*, *3*(4), 274-293.
- Jenkins, T. G., & Carrell, D. T. (2012). Dynamic alterations in the paternal epigenetic landscape following fertilization. *Front Genet*, *3*, 143. doi: 10.3389/fgene.2012.00143
- Jjingo, D., Conley, A. B., et al. (2012). On the presence and role of human gene-body DNA methylation. *Oncotarget*, *3*(4), 462-474.
- Jones, M. J., & Jallepalli, P. V. (2012). Chromothripsis: chromosomes in crisis. *Dev Cell*, *23*(5), 908-917. doi: 10.1016/j.devcel.2012.10.010
- Jones, P. A., & Liang, G. (2009). Rethinking how DNA methylation patterns are maintained. *Nat Rev Genet*, *10*(11), 805-811. doi: 10.1038/nrg2651
- Juo, Y. Y., Johnston, F. M., et al. (2014). Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. *Ann Oncol*. doi: 10.1093/annonc/mdu149
- Kadoch, C., Hargreaves, D. C., et al. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat Genet*, *45*(6), 592-601. doi: 10.1038/ng.2628
- Kadyrov, F. A., Dzantiev, L., et al. (2006). Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, *126*(2), 297-308. doi: 10.1016/j.cell.2006.05.039
- Kanai, Y., Ushijima, S., et al. (2001). DNA methyltransferase expression and DNA methylation of CPG islands and peri-centromeric satellite regions in human colorectal and stomach cancers. *Int J Cancer*, *91*(2), 205-212.
- Karakas, B., Bachman, K. E., et al. (2006). Mutation of the PIK3CA oncogene in human cancers. *Br J Cancer*, *94*(4), 455-459. doi: 10.1038/sj.bjc.6602970
- Kasperek, T. R., & Humphrey, T. C. (2011). DNA double-strand break repair pathways, chromosomal rearrangements and cancer. *Semin Cell Dev Biol*, *22*(8), 886-897. doi: 10.1016/j.semcdb.2011.10.007
- Katz, M., Amit, I., et al. (2007). Regulation of MAPKs by growth factors and receptor tyrosine kinases. *Biochim Biophys Acta*, *1773*(8), 1161-1176. doi: 10.1016/j.bbamcr.2007.01.002
- Ke, X. S., Qu, Y., et al. (2010). Global profiling of histone and DNA methylation reveals epigenetic-based regulation of gene expression during epithelial to mesenchymal transition in prostate cells. *BMC Genomics*, *11*, 669. doi: 10.1186/1471-2164-11-669
- Kim, T. M., Laird, P. W., et al. (2013). The landscape of microsatellite instability in colorectal and endometrial cancer genomes. *Cell*, *155*(4), 858-868. doi: 10.1016/j.cell.2013.10.015
- Kinsella, M., Patel, A., et al. (2014). The elusive evidence for chromothripsis. *Nucleic Acids Res*, *42*(13), 8231-8242. doi: 10.1093/nar/gku525
- Ko, M., Huang, Y., et al. (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. *Nature*, *468*(7325), 839-843. doi: 10.1038/nature09586
- Ko, M., & Rao, A. (2011). TET2: epigenetic safeguard for HSC. *Blood*, *118*(17), 4501-4503. doi: 10.1182/blood-2011-08-373357
- Kobayashi, S., Yamada-Okabe, H., et al. (2012). LGR5-positive colon cancer stem cells interconvert with drug-resistant LGR5-negative cells and are capable of tumor reconstitution. *Stem Cells*, *30*(12), 2631-2644. doi: 10.1002/stem.1257



- Koh, K. P., Yabuuchi, A., et al. (2011). Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell*, 8(2), 200-213. doi: 10.1016/j.stem.2011.01.008
- Kohli, R. M., & Zhang, Y. (2013). TET enzymes, TDG and the dynamics of DNA demethylation. *Nature*, 502(7472), 472-479. doi: 10.1038/nature12750
- Kondo, Y., & Issa, J. P. (2004). Epigenetic changes in colorectal cancer. *Cancer Metastasis Rev*, 23(1-2), 29-39.
- Koo, B. K., & Clevers, H. (2014). Stem cells marked by the R-spondin receptor LGR5. *Gastroenterology*, 147(2), 289-302. doi: 10.1053/j.gastro.2014.05.007
- Kouzarides, T. (2007). Chromatin modifications and their function. *Cell*, 128(4), 693-705. doi: 10.1016/j.cell.2007.02.005
- Ku, M., Koche, R. P., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*, 4(10), e1000242. doi: 10.1371/journal.pgen.1000242
- Kudo, Y., Tateishi, K., et al. (2012). Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci*, 103(4), 670-676. doi: 10.1111/j.1349-7006.2012.02213.x
- Kulis, M., & Esteller, M. (2010). DNA methylation and cancer. *Adv Genet*, 70, 27-56. doi: 10.1016/B978-0-12-380866-0.60002-2
- Langemeijer, S. M., Kuiper, R. P., et al. (2009). Acquired mutations in TET2 are common in myelodysplastic syndromes. *Nat Genet*, 41(7), 838-842. doi: 10.1038/ng.391
- Lanzuolo, C., & Orlando, V. (2012). Memories from the Polycomb Group Proteins. *Annual Review of Genetics*, Vol 46, 46, 561-589. doi: DOI 10.1146/annurev-genet-110711-155603
- Leggett, B., & Whitehall, V. (2010). Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology*, 138(6), 2088-2100. doi: 10.1053/j.gastro.2009.12.066
- Letouze, E., Martinelli, C., et al. (2013). SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 23(6), 739-752. doi: 10.1016/j.ccr.2013.04.018
- Lim, D. H., & Maher, E. R. (2010). Genomic imprinting syndromes and cancer. *Adv Genet*, 70, 145-175. doi: 10.1016/B978-0-12-380866-0.60006-X
- Liu, S., Cong, Y., et al. (2014). Breast cancer stem cells transition between epithelial and mesenchymal states reflective of their normal counterparts. *Stem Cell Reports*, 2(1), 78-91. doi: 10.1016/j.stemcr.2013.11.009
- Loeb, L. A. (1991). Mutator phenotype may be required for multistage carcinogenesis. *Cancer Res*, 51(12), 3075-3079.
- Long, M. D., Martin, C. F., et al. (2010). Hormone replacement therapy, oral contraceptive use, and distal large bowel cancer: a population-based case-control study. *Am J Gastroenterol*, 105(8), 1843-1850. doi: 10.1038/ajg.2010.123
- Lorsbach, R. B., Moore, J., et al. (2003). TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). *Leukemia*, 17(3), 637-641. doi: 10.1038/sj.leu.2402834
- MacDonald, B. T., Tamai, K., et al. (2009). Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell*, 17(1), 9-26. doi: 10.1016/j.devcel.2009.06.016
- Maeshima, K., Hihara, S., et al. (2010). New insight into the mitotic chromosome structure: irregular folding of nucleosome fibers without 30-nm chromatin structure. *Cold Spring Harb Symp Quant Biol*, 75, 439-444. doi: 10.1101/sqb.2010.75.034
- Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*, 286(41), 35334-35338. doi: 10.1074/jbc.C111.284620
- Mancini, D. N., Rodenhiser, D. I., et al. (1998). CpG methylation within the 5' regulatory region of the BRCA1 gene is tumor specific and includes a putative CREB binding site. *Oncogene*, 16(9), 1161-1169. doi: 10.1038/sj.onc.1201630
- Mani, S. A., Guo, W., et al. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell*, 133(4), 704-715. doi: 10.1016/j.cell.2008.03.027
- Markowitz, S., Wang, J., et al. (1995). Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science*, 268(5215), 1336-1338.

- Maunakea, A. K., Chepelev, I., et al. (2013). Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res*, 23(11), 1256-1269. doi: 10.1038/cr.2013.110
- McClelland, S. E., Burrell, R. A., et al. (2009). Chromosomal instability: a composite phenotype that influences sensitivity to chemotherapy. *Cell Cycle*, 8(20), 3262-3266.
- Menigatti, M., Truninger, K., et al. (2009). Normal colorectal mucosa exhibits sex- and segment-specific susceptibility to DNA methylation at the hMLH1 and MGMT promoters. *Oncogene*, 28(6), 899-909. doi: 10.1038/onc.2008.444
- Myrie, K. A., Percy, M. J., et al. (2000). Mutation and expression analysis of human BUB1 and BUB1B in aneuploid breast cancer cell lines. *Cancer Lett*, 152(2), 193-199.
- Nabel, C. S., Jia, H. J., et al. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nature Chemical Biology*, 8(9), 751-758. doi: Doi 10.1038/Nchembio.1042
- Nakajima, H., & Kunimoto, H. (2014). TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Sci*, 105(9), 1093-1099. doi: 10.1111/cas.12484
- Nambiar, M., Kari, V., et al. (2008). Chromosomal translocations in cancer. *Biochim Biophys Acta*, 1786(2), 139-152. doi: 10.1016/j.bbcan.2008.07.005
- Narlikar, G. J., Sundaramoorthy, R., et al. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell*, 154(3), 490-503. doi: 10.1016/j.cell.2013.07.011
- Natarajan, A. T., Berni, A., et al. (2008). The type and yield of ionising radiation induced chromosomal aberrations depend on the efficiency of different DSB repair pathways in mammalian cells. *Mutat Res*, 642(1-2), 80-85. doi: 10.1016/j.mrfmmm.2008.05.002
- Nishino, Y., Eltsov, M., et al. (2012). Human mitotic chromosomes consist predominantly of irregularly folded nucleosome fibres without a 30-nm chromatin structure. *EMBO J*, 31(7), 1644-1653. doi: 10.1038/emboj.2012.35
- Noreen, F., Roosli, M., et al. (2014). Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. *J Natl Cancer Inst*, 106(7). doi: 10.1093/jnci/dju161
- Nosho, K., Shima, K., et al. (2009). DNMT3B expression might contribute to CpG island methylator phenotype in colorectal cancer. *Clin Cancer Res*, 15(11), 3663-3671. doi: 10.1158/1078-0432.CCR-08-2383
- Noushmehr, H., Weisenberger, D. J., et al. (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*, 17(5), 510-522. doi: 10.1016/j.ccr.2010.03.017
- Oberdoerffer, P., & Sinclair, D. A. (2007). The role of nuclear architecture in genomic instability and ageing. *Nat Rev Mol Cell Biol*, 8(9), 692-702. doi: 10.1038/nrm2238
- Ocana, O. H., Corcoles, R., et al. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell*, 22(6), 709-724. doi: 10.1016/j.ccr.2012.10.012
- Ohm, J. E., McGarvey, K. M., et al. (2007). A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nat Genet*, 39(2), 237-242. doi: 10.1038/ng1972
- Oikonomou, E., Makrodoouli, E., et al. (2009). BRAF(V600E) efficient transformation and induction of microsatellite instability versus KRAS(G12V) induction of senescence markers in human colon cancer cells. *Neoplasia*, 11(11), 1116-1131.
- Olivier, M., Hollstein, M., et al. (2010). TP53 mutations in human cancers: origins, consequences, and clinical use. *Cold Spring Harb Perspect Biol*, 2(1), a001008. doi: 10.1101/cshperspect.a001008
- Ono, R., Taki, T., et al. (2002). LCX, leukemia-associated protein with a CXXC domain, is fused to MLL in acute myeloid leukemia with trilineage dysplasia having t(10;11)(q22;q23). *Cancer Res*, 62(14), 4075-4080.
- Ooi, S. K., & Bestor, T. H. (2008). The colorful history of active DNA demethylation. *Cell*, 133(7), 1145-1148. doi: 10.1016/j.cell.2008.06.009
- Palles, C., Cazier, J. B., et al. (2013). Germline mutations affecting the proofreading domains of POLE and POLD1 predispose to colorectal adenomas and carcinomas. *Nat Genet*, 45(2), 136-144. doi: 10.1038/ng.2503

- Parajuli, R., Bjerkaas, E., et al. (2013). The increased risk of colon cancer due to cigarette smoking may be greater in women than men. *Cancer Epidemiol Biomarkers Prev*, 22(5), 862-871. doi: 10.1158/1055-9965.EPI-12-1351
- Park, H. Y., Jeon, Y. K., et al. (2007). Differential promoter methylation may be a key molecular mechanism in regulating BubR1 expression in cancer cells. *Exp Mol Med*, 39(2), 195-204. doi: 10.1038/emmm.2007.22
- Pena-Diaz, J., & Jiricny, J. (2012). Mammalian mismatch repair: error-free or error-prone? *Trends Biochem Sci*, 37(5), 206-214. doi: 10.1016/j.tibs.2012.03.001
- Peng, B., Hurt, E. M., et al. (2006). DNA hypermethylation and partial gene silencing of human thymine-DNA glycosylase in multiple myeloma cell lines. *Epigenetics*, 1(3), 138-145.
- Peng, J. C., & Karpen, G. H. (2008). Epigenetic regulation of heterochromatic DNA stability. *Curr Opin Genet Dev*, 18(2), 204-211. doi: 10.1016/j.gde.2008.01.021
- Petty, E., & Pillus, L. (2013). Balancing chromatin remodeling and histone modifications in transcription. *Trends Genet*, 29(11), 621-629. doi: 10.1016/j.tig.2013.06.006
- Pfeifer, G. P., You, Y. H., et al. (2005). Mutations induced by ultraviolet light. *Mutat Res*, 571(1-2), 19-31. doi: 10.1016/j.mrfmmm.2004.06.057
- Pino, M. S., Kikuchi, H., et al. (2010). Epithelial to mesenchymal transition is impaired in colon cancer cells with microsatellite instability. *Gastroenterology*, 138(4), 1406-1417. doi: 10.1053/j.gastro.2009.12.010
- Prigent, C., & Dimitrov, S. (2003). Phosphorylation of serine 10 in histone H3, what for? *J Cell Sci*, 116(Pt 18), 3677-3685. doi: 10.1242/jcs.00735
- Quivoron, C., Couronne, L., et al. (2011). TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. *Cancer Cell*, 20(1), 25-38. doi: 10.1016/j.ccr.2011.06.003
- Rad, R., Cadinanos, J., et al. (2013). A genetic progression model of Braf(V600E)-induced intestinal tumorigenesis reveals targets for therapeutic intervention. *Cancer Cell*, 24(1), 15-29. doi: 10.1016/j.ccr.2013.05.014
- Rakyan, V. K., Down, T. A., et al. (2010). Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*, 20(4), 434-439. doi: 10.1101/gr.103101.109
- Rangam, G., Schmitz, K. M., et al. (2012). AID Enzymatic Activity Is Inversely Proportional to the Size of Cytosine C5 Orbital Cloud. *PLoS One*, 7(8). doi: ARTN e43279
- DOI 10.1371/journal.pone.0043279
- Reynard, L. N., Bui, C., et al. (2014). CpG methylation regulates allelic expression of GDF5 by modulating binding of SP1 and SP3 repressor proteins to the osteoarthritis susceptibility SNP rs143383. *Hum Genet*, 133(8), 1059-1073. doi: 10.1007/s00439-014-1447-z
- Ricci-Vitiani, L., Lombardi, D. G., et al. (2007). Identification and expansion of human colon-cancer-initiating cells. *Nature*, 445(7123), 111-115. doi: 10.1038/nature05384
- Roberts, S. A., Lawrence, M. S., et al. (2013). An APOBEC cytidine deaminase mutagenesis pattern is widespread in human cancers. *Nat Genet*, 45(9), 970-976. doi: 10.1038/ng.2702
- Robsahm, T. E., Aagnes, B., et al. (2013). Body mass index, physical activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis of cohort studies. *Eur J Cancer Prev*, 22(6), 492-505. doi: 10.1097/CEJ.0b013e328360f434
- Sabates-Bellver, J., Van der Flier, L. G., et al. (2007). Transcriptome profile of human colorectal adenomas. *Mol Cancer Res*, 5(12), 1263-1275. doi: 10.1158/1541-7786.MCR-07-0267
- Sakai, E., Nakajima, A., et al. (2014). Accumulation of aberrant DNA methylation during colorectal cancer development. *World J Gastroenterol*, 20(4), 978-987. doi: 10.3748/wjg.v20.i4.978
- Sanchez-Tillo, E., de Barrios, O., et al. (2011). beta-catenin/TCF4 complex induces the epithelial-to-mesenchymal transition (EMT)-activator ZEB1 to regulate tumor invasiveness. *Proc Natl Acad Sci U S A*, 108(48), 19204-19209. doi: 10.1073/pnas.1108977108
- Schepers, A., & Clevers, H. (2012). Wnt signaling, stem cells, and cancer of the gastrointestinal tract. *Cold Spring Harb Perspect Biol*, 4(4), a007989. doi: 10.1101/cshperspect.a007989

- Schlesinger, Y., Straussman, R., et al. (2007). Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for de novo methylation in cancer. *Nat Genet*, 39(2), 232-236. doi: 10.1038/ng1950
- Schuettengruber, B., Chourrout, D., et al. (2007). Genome regulation by polycomb and trithorax proteins. *Cell*, 128(4), 735-745. doi: DOI 10.1016/j.cell.2007.02.009
- Sheffer, M., Bacolod, M. D., et al. (2009). Association of survival and disease progression with chromosomal instability: a genomic exploration of colorectal cancer. *Proc Natl Acad Sci U S A*, 106(17), 7131-7136. doi: 10.1073/pnas.0902232106
- Shen, H., & Laird, P. W. (2013). Interplay between the cancer genome and epigenome. *Cell*, 153(1), 38-55. doi: 10.1016/j.cell.2013.03.008
- Shen, L., Toyota, M., et al. (2007). Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A*, 104(47), 18654-18659. doi: 10.1073/pnas.0704652104
- Shibata, D. (2009). Inferring human stem cell behaviour from epigenetic drift. *J Pathol*, 217(2), 199-205. doi: 10.1002/path.2461
- Shih, I. M., Zhou, W., et al. (2001). Evidence that genetic instability occurs at an early stage of colorectal tumorigenesis. *Cancer Res*, 61(3), 818-822.
- Shilatifard, A. (2008). Molecular implementation and physiological roles for histone H3 lysine 4 (H3K4) methylation. *Curr Opin Cell Biol*, 20(3), 341-348. doi: 10.1016/j.ceb.2008.03.019
- Smyth, G. K., & Speed, T. (2003). Normalization of cDNA microarray data. *Methods*, 31(4), 265-273.
- Solomon, D. A., Kim, T., et al. (2011). Mutational inactivation of STAG2 causes aneuploidy in human cancer. *Science*, 333(6045), 1039-1043. doi: 10.1126/science.1203619
- Stankiewicz, P., & Lupski, J. R. (2010). Structural variation in the human genome and its role in disease. *Annu Rev Med*, 61, 437-455. doi: 10.1146/annurev-med-100708-204735
- Stephens, P. J., Greenman, C. D., et al. (2011). Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell*, 144(1), 27-40. doi: 10.1016/j.cell.2010.11.055
- Strahl, B. D., & Allis, C. D. (2000). The language of covalent histone modifications. *Nature*, 403(6765), 41-45.
- Szwagierczak, A., Bultmann, S., et al. (2010). Sensitive enzymatic quantification of 5-hydroxymethylcytosine in genomic DNA. *Nucleic Acids Res*, 38(19), e181. doi: 10.1093/nar/gkq684
- Tahiliani, M., Koh, K. P., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929), 930-935. doi: 1170116 [pii] 10.1126/science.1170116
- Tapp, H. S., Commane, D. M., et al. (2013). Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. *Aging Cell*, 12(1), 148-155. doi: 10.1111/ace1.12030
- Teodoridis, J. M., Hardie, C., et al. (2008). CpG island methylator phenotype (CIMP) in cancer: causes and implications. *Cancer Lett*, 268(2), 177-186. doi: 10.1016/j.canlet.2008.03.022
- Thompson, S. L., Bakhoun, S. F., et al. (2010). Mechanisms of chromosomal instability. *Curr Biol*, 20(6), R285-295. doi: 10.1016/j.cub.2010.01.034
- Ting, A. H., McGarvey, K. M., et al. (2006). The cancer epigenome--components and functional correlates. *Genes Dev*, 20(23), 3215-3231. doi: 10.1101/gad.1464906
- Touleimat, N., & Tost, J. (2012). Complete pipeline for Infinium((R)) Human Methylation 450K BeadChip data processing using subset quantile normalization for accurate DNA methylation estimation. *Epigenomics*, 4(3), 325-341. doi: 10.2217/epi.12.21
- Toyota, M., Ahuja, N., et al. (1999). CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A*, 96(15), 8681-8686.
- Tsai, J. H., Donaher, J. L., et al. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell*, 22(6), 725-736. doi: 10.1016/j.ccr.2012.09.022

- Turcan, S., Rohle, D., et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, 483(7390), 479-483. doi: 10.1038/nature10866
- Turker, M. S. (1999). The establishment and maintenance of DNA methylation patterns in mouse somatic cells. *Semin Cancer Biol*, 9(5), 329-337. doi: 10.1006/scbi.1999.0133
- Valinluck, V., & Sowers, L. C. (2007). Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. *Cancer Res*, 67(3), 946-950. doi: 10.1158/0008-5472.CAN-06-3123
- van de Wetering, M., Sancho, E., et al. (2002). The beta-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell*, 111(2), 241-250.
- van der Klift, H., Wijnen, J., et al. (2005). Molecular characterization of the spectrum of genomic deletions in the mismatch repair genes MSH2, MLH1, MSH6, and PMS2 responsible for hereditary nonpolyposis colorectal cancer (HNPCC). *Genes Chromosomes Cancer*, 44(2), 123-138. doi: 10.1002/gcc.20219
- Varela-Rey, M., Woodhoo, A., et al. (2013). Alcohol, DNA methylation, and cancer. *Alcohol Res*, 35(1), 25-35.
- Vasovcak, P., Krepelova, A., et al. (2012). Unique mutational profile associated with a loss of TDG expression in the rectal cancer of a patient with a constitutional PMS2 deficiency. *DNA Repair (Amst)*, 11(7), 616-623. doi: 10.1016/j.dnarep.2012.04.004
- Vermeulen, L., De Sousa, E. M. F., et al. (2010). Wnt activity defines colon cancer stem cells and is regulated by the microenvironment. *Nat Cell Biol*, 12(5), 468-476. doi: 10.1038/ncb2048
- Viguera, E., Canceill, D., et al. (2001). Replication slippage involves DNA polymerase pausing and dissociation. *EMBO J*, 20(10), 2587-2595. doi: 10.1093/emboj/20.10.2587
- Vogelstein, B., Papadopoulos, N., et al. (2013). Cancer genome landscapes. *Science*, 339(6127), 1546-1558. doi: 10.1126/science.1235122
- Walther, A., Johnstone, E., et al. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer*, 9(7), 489-499. doi: 10.1038/nrc2645
- Wang, R. H., Yu, H., et al. (2004). A requirement for breast-cancer-associated gene 1 (BRCA1) in the spindle checkpoint. *Proc Natl Acad Sci U S A*, 101(49), 17108-17113. doi: 10.1073/pnas.0407585101
- Wang, Z., Cummins, J. M., et al. (2004). Three classes of genes mutated in colorectal cancers with chromosomal instability. *Cancer Res*, 64(9), 2998-3001.
- Weinhold, N., Jacobsen, A., et al. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. *Nat Genet*, 46(11), 1160-1165. doi: 10.1038/ng.3101
- Weisenberger, D. J., Siegmund, K. D., et al. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*, 38(7), 787-793. doi: 10.1038/ng1834
- Widschwendter, M., Fiegl, H., et al. (2007). Epigenetic stem cell signature in cancer. *Nat Genet*, 39(2), 157-158. doi: 10.1038/ng1941
- Williams, K., Christensen, J., et al. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. *Nature*, 473(7347), 343-348. doi: 10.1038/nature10066
- Wilson, A. S., Power, B. E., et al. (2007). DNA hypomethylation and human diseases. *Biochim Biophys Acta*, 1775(1), 138-162. doi: 10.1016/j.bbcan.2006.08.007
- Workman, C., Jensen, L. J., et al. (2002). A new non-linear normalization method for reducing variability in DNA microarray experiments. *Genome Biol*, 3(9), research0048.
- Wossidlo, M., Nakamura, T., et al. (2011). 5-Hydroxymethylcytosine in the mammalian zygote is linked with epigenetic reprogramming. *Nat Commun*, 2, 241. doi: 10.1038/ncomms1240
- Xie, T., Cho, Y. B., et al. (2014). Patterns of somatic alterations between matched primary and metastatic colorectal tumors characterized by whole-genome sequencing. *Genomics*. doi: 10.1016/j.ygeno.2014.07.012
- Xu, W., Yang, H., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell*, 19(1), 17-30. doi: 10.1016/j.ccr.2010.12.014

- Xu, Y., Wu, F., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell*, 42(4), 451-464. doi: 10.1016/j.molcel.2011.04.005
- Yagi, K., Akagi, K., et al. (2010). Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res*, 16(1), 21-33. doi: 10.1158/1078-0432.CCR-09-2006
- Yang, H., Liu, Y., et al. (2013). Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene*, 32(5), 663-669. doi: 10.1038/onc.2012.67
- Yatsuoka, T., Furukawa, T., et al. (1999). Genomic analysis of the thymine-DNA glycosylase (TDG) gene on 12q22-q24.1 in human pancreatic ductal adenocarcinoma. *Int J Pancreatol*, 25(2), 97-102. doi: 10.1385/IJGC:25:2:97
- Yun, M. Y., Wu, J., et al. (2011). Readers of histone modifications. *Cell Res*, 21(4), 564-578. doi: 10.1038/Cr.2011.42
- Zeman, M. K., & Cimprich, K. A. (2014). Causes and consequences of replication stress. *Nat Cell Biol*, 16(1), 2-9. doi: 10.1038/ncb2897
- Zhang, N., Ge, G., et al. (2008). Overexpression of Separase induces aneuploidy and mammary tumorigenesis. *Proc Natl Acad Sci U S A*, 105(35), 13033-13038. doi: 10.1073/pnas.0801610105
- Zhang, X., Yang, R., et al. (2014). Hypermethylation of Sp1 binding site suppresses hypothalamic POMC in neonates and may contribute to metabolic disorders in adults: impact of maternal dietary CLAs. *Diabetes*, 63(5), 1475-1487. doi: 10.2337/db13-1221
- Zilfou, J. T., & Lowe, S. W. (2009). Tumor suppressive functions of p53. *Cold Spring Harb Perspect Biol*, 1(5), a001883. doi: 10.1101/cshperspect.a001883

## Appendix:

- I. Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers
- II. Modulation of Age- and Cancer-Associated DNA Methylation Change in Healthy Colon by Aspirin and Lifestyle
- III. The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation
- IV. Colorectal mucosa of healthy individuals displays part of the CpG island methylator phenotype signature

# Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers

Weis, S.<sup>1</sup>, Muraro, M.G.<sup>1</sup>, Noreen, F.<sup>1</sup>, Mele, V.<sup>1</sup>, Wirz, A.<sup>1</sup>, Truninger, K.<sup>2</sup>, Schär, P.<sup>1</sup>

<sup>1</sup> Department of Biomedicine, University of Basel, Switzerland; <sup>2</sup> FMH Gastroenterology and Internal Medicine, Langenthal, Switzerland

Correspondence: primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566



## **Abstract**

Colorectal cancers and other human malignancies displaying a CpG island methylator phenotype (CIMP) constitute a well-defined subtype of tumors characterized by extensive epigenetic aberrations and a distinct etiology. Although, epigenetic as well as genetic alterations specific for colorectal cancer CIMP (CRC-CIMP) are well described its molecular basis remains unresolved. Here we report that expression of the DNA hydroxylase TET1 is significantly reduced in CRC-CIMP, associated with an aberrant hypermethylation of the TET1 promoter. Accordingly, in CIMP cell lines levels of 5-hydroxymethylcytosine (5hmC) were reduced at CIMP-specifically hypermethylated loci. The sh-RNA mediated depletion of TET1 in colorectal cancer (CRC) cells generated genome-wide CpG island-associated methylation aberrations with equal dimensions of hyper- and hypomethylation. TET1-depleted cells acquired CRC-CIMP specific characteristics including promoter hypermethylation and transcriptional repression of MLH1 and E-cadherin. Furthermore, the loss of TET1 was accompanied by the loss of cell phenotype plasticity and reduced expression of stem cell markers. The altered methylome of TET1 depleted cells evolved towards a methylator phenotype in mouse xenograft tumors. Together, our data therefore provide evidence that the loss of TET1 expression is causal for genome-wide hypermethylation observed in CRC-CIMP and involved in changes in the cellular behavior with potential clinical implications.

## Introduction

Aberrant hypermethylation of promoter-associated CpG islands (CGIs) is prevalent in many human cancers and is usually associated with transcriptional repression of the respective genes (Ting et al., 2006). The phenomenon was first described in colorectal cancers (CRC), based on the finding that a subgroup of tumors showed strong and widespread CGI methylation, a feature that was named CpG island methylator phenotype (CIMP) (Toyota et al., 1999). Subsequent studies identified similar methylator subgroups also in other human cancers including gastric, lung, liver, ovarian and also in gliomas and leukemias (Noushmehr et al., 2010; Teodoridis et al., 2008). The existence of distinct methylation phenotypes indicates that aberrant methylation does not occur randomly in the genomes of cancer cells but rather is associated with a specific pathway of carcinogenesis, characterized by epigenetic instability. CRCs showing CIMP are associated with BRAF V600E mutation and are often microsatellite instable (MSI), due to hypermethylation and concomitant silencing of the MLH1 promoter (Ang et al., 2010; Shen et al., 2007; Yagi et al., 2010). In addition, CIMP CRCs are mainly found in women and on the right side of the colon. Patients with CIMP-associated CRC are older at diagnosis and several studies indicate that they have worse prognosis compared to those with non-CIMP cancers. Histologically, CIMP CRCs are characterized by poor cellular differentiation (Nazemalhosseini Mojarad et al., 2013). Recently, it has been shown that CIMP-CRCs also arise from a distinct class of precursor lesions described as the serrated pathway of colorectal carcinogenesis. The significant association of the defined group of serrated adenomas with MSI, BRAF V600E mutation and with CRC-CIMP further substantiates the evidence for a discrete CIMP-associated pathway of colorectal carcinogenesis (Leggett and Whitehall, 2010).

The molecular mechanisms underlying CGI hypermethylation in CIMP cancers are poorly understood and remain speculative. Overexpression of DNA methyltransferases (DNMTs), enzymes that establish and maintain CpG methylation, has been implicated in CGI hypermethylation in cancer. Increased levels of DNMT1, the maintenance DNA methyltransferase, have been associated with CGI hypermethylation in colorectal, pancreatic and gastric human cancers (Etoh et al., 2004; Kanai et al., 2001; Peng et al., 2006), while overexpression of the de novo methyltransferase DNMT3B has been described in breast and colorectal cancers with a hypermethylator phenotype (Nosho et al., 2009; Van der Auwera et al., 2010). However, no study has directly shown that DNMT overexpression is sufficient to generate a CGI hypermethylation phenotype. More recently, several studies have implicated impaired DNA demethylation in the development of CIMP cancers, through inhibition or genetic inactivation of ten-eleven-translocation 5-methylcytosine dioxygenases family of enzymes.

The three family members, TET1, TET2 and TET3 catalyze the iterative oxidation of 5-methylcytosine (5mC) to 5-hydroxymethyl cytosine (5hmC), 5-formylcytosine (5fC) and 5carboxycytosien (5caC) (Ito et al., 2010; Tahiliani et al., 2009). 5fC and 5caC are substrates for the Thymine DNA glycosylase (TDG), which excises these bases to give way for the de novo incorporation of an unmethylated cytosine by DNA repair. (He et al., 2011; Maiti and Drohat, 2011).

Loss-of-function mutations of TET2 are associated with a hypermethylation phenotype in acute myeloid leukemia (AML) (Nakajima and Kunimoto, 2014). These occur in a mutually exclusive manner with mutations in isocitrate dehydrogenases (IDH1 and IDH2), which give rise similar methylation aberrations (Figueroa et al., 2010). This is rationalized by the neomorphic nature of the respective IDH1 and IDH2 mutations, conferring an  $\alpha$ -KG reducing activity to these enzymes, which then generate the oncometabolite 2-Hydroxyglutarate (2-HG), a competitive inhibitor of  $\alpha$ -KG-dependend dioxygenases like TET proteins (W. Xu et al., 2011). Such a neomorphic mutation in IDH1 was also suggested to constitute the molecular basis for glioma-CIMP (Turcan et al., 2012). IDH1 mutations, however, are rare in most other solid cancers including CRC and appear to play a role only in a small group of CIMP cancers (Bleeker et al., 2009; Borger et al., 2012; Kang et al., 2009). Similarly, loss-of-function mutations in Succinate dehydrogenase (SDH) were associated with a hypermethylation phenotype in paragangliomas and gastrointestinal stromal tumors. These mutations result in an accumulation of succinate (Killian et al., 2013; Letouze et al., 2013), an intermediate of the Krebs cycle and, the same as fumarate, a competitive inhibitor of  $\alpha$ -KG-dependend dioxygenases including histone demethylases and TET family proteins (Xiao et al., 2012). These studies implicate a central role to impaired TET-mediated DNA demethylation in the development of CIMP cancers in humans.

In this study, we addressed to what extent deregulation of TET activity may explain the origin, phenotype and biology of CIMP cancers in the colon. We report the consistent loss of TET1 expression in CIMP CRC cell lines and primary tumors and describe epigenetic and cellular changes upon the knockdown of TET1 in the non-CIMP CRC cell line SW620 which mirror characteristics of CIMP-CRCs. Thereby, our results also provide insight into molecular mechanisms which are associated with cancer progression.

## Results

### Promoter hypermethylation in CRC cell lines is associated with low TET1 expression

We first used a set of eight CRC cell lines (SW48, SW620, COLO320, COLO205, HCT-116, CACO2, LOVO and DLD-1) to investigate the role of the oxidative DNA demethylation pathway in the CRC associated methylator phenotype. To arrange the cell lines according to their methylation patterns in an unbiased manner, we performed unsupervised cluster analysis on publicly available genome-wide CGI methylation data from Illumina Infinium HumanMethylation27 (HM27) arrays; GEO accession number GSE35573 (Easwaran et al., 2012). We then validated the CIMP relevance in the resulting clusters by examining the methylation status of conventional CIMP markers (5 marker panel; CACNA1G, NEUROG1, RUNX3, IGF2 and SOCS2 (Weisenberger et al., 2006) as well as the microsatellite instability (MSI) status. The methylation levels of the CIMP marker panel were quantified by In-qMSP; cell lines showing methylation of at least three of these markers were considered to reflect CIMP positivity. The MSI status in these cell lines was previously determined (Ahmed et al., 2013).

In the analysis of the CpG methylation array data, we excluded probes that are overlapping with repetitive elements, regions of insertions and deletions and single-nucleotide polymorphisms and probes designed for sequences on sex chromosomes (Hinoue et al., 2012). In the remaining 19,773 probes, we then identified CpG sites showing top 25% highest variation in methylation, from which we further extracted those associated with promoter CpG islands (3713 CpG sites) for an unsupervised clustering analysis (Figure 1a). This grouped cell lines along a gradient of increasing methylation ranging from low methylation in COLO320 to high levels of gene promoter methylation in HCT116 and SW48. Higher levels of promoter methylation by genome-wide analysis correlated with the presence of hypermethylation at the CIMP five marker panel as well as MSI. Based on this associations, we classified COLO205, LOVO, DLD-1, HCT116 and SW48 as CIMP cell lines and COLO320, SW620 and CACO2 as non-CIMP cell lines. We then quantified the mRNA expression of TET1, TET2 and TET3 by RT-qPCR in all cell lines. Averaged methylation levels of the 3713 differentially methylated CGI-associated CpG sites were inversely associated with the mRNA expression of TET1 but not with the expression levels of TET2 or TET3 (Figure 1b). The correlation applied also for TET1 protein levels (Figure 1c).

Next, we asked whether TETs expression correlates with the occurrence of 5hmC, the product of TET catalyzed oxidation of 5mC, in the cell line genomes. Global levels of 5hmC were

quantified by LC-MS/MS. We did not find any correlation between expression of TET1, TET2 or TET3 and global 5hmC levels (Supplementary Figure 1). In addition, we assessed levels of locus-specific 5hmC and 5mC modifications at nine CGI containing gene promoters (PROK2, NPPC, DLX5, LHX6, POU3F2, WNT10A, SLC30A, TRIM9, CDK5R2) by immunoprecipitation of modified DNA (MeDIP and GLIP) and subsequent quantification by qPCR. It has been previously shown that these genes are specifically hypermethylated in CRC-CIMP (Hinoue et al., 2012; Y. Xu et al., 2012) and targeted by TET1 in mouse ESCs (Williams et al., 2011). We found significantly higher levels of promoter methylation at all nine targets in CIMP cell lines (SW48 and HCT116) compared to non-CIMP cell lines (CACO2 and SW620) (Figure 1d). Since, 5mC constitutes the substrate for TET proteins in the generation of 5hmC, and 5mC-rich regions are therefore more likely to possess higher levels of 5hmC, we normalized the levels of 5hmC against the levels of 5mC in the respective promoter. Normalized 5hmC levels were significantly lower in CIMP cell lines versus non-CIMP cell lines (Figure 1d).

We thus conclude that CRC cell lines displaying a methylator phenotype have low levels of TET1 expression, which coincides with lower 5hmC levels at promoter regions known to be hypermethylated in CRC-CIMP.

### **TET1 expression correlates with methylation profiles in primary human colorectal cancers**

To test relevance of these findings for primary CRCs, we examined methylation levels of the CIMP five-marker panel in a consecutively collected set of 23 primary CRCs, 18 matched cancer-associated normal mucosa samples (CAM) and 8 colonic mucosa biopsies from healthy individuals (HM). Cancers were classified as CIMP (n=7) when  $\geq 3$  markers were methylated, or else as non-CIMP (n=15). TET1 mRNA expression was significantly lower in CIMP compared to non-CIMP cancers, HM and CAM. No difference was apparent in TET2 and TET3 expression between CIMP and non-CIMP cancers (Figure 2a).

We validated these results in a larger set of colon adenocarcinoma samples (n=143) using genome-wide DNA methylation and mRNA expression data available from The Cancer Genome Atlas (TCGA). 25 colon cancers were classified previously as CIMP using Recursively Portioned Mixture Model (RPMM) (Cancer Genome Atlas, 2012). We observed significantly lower TET1 and TET3 expression in CIMP cancers compared to non-CIMP (Figure 2b). In contrast to the findings in our test set of CRCs, TET2 expression was significantly higher in CIMP cancers from the validation set (Figure 2b). We also investigated the expression of other

factors involved in regulation of CpG methylation. Expression of DNMT3A and DNMT3B was significantly lower in CIMP cancers compared to non-CIMP (Supplementary Figure 2a). No significant difference was found in the expression of DNMT1, TDG, IDH1 and IDH2 between CIMP and non-CIMP tumors (Supplementary Figure 2).

To investigate whether the loss of TET1 expression in CIMP-CRCs is associated itself with hypermethylation of the TET1 promoter, we quantified the promoter methylation levels of TET1 in 22 primary CRCs from our test set as well as in 8 CRC cell lines. Using In-qMSP, we found TET1 promoter hypermethylated in several CIMP tumors (5/7) and CIMP cell lines (3/4) (Figure 2c). In accordance with this, when we treated CIMP CRC cell line SW48 with the demethylating agent 5-azacytidine (5-aza-C) TET1 mRNA was re-expressed upon the reduction of promoter methylation levels (Figure 2d). This observation raised the question whether the loss of TET1 expression occurs before genome-wide CGI methylation in CIMP or as a result of it.

### **TET1 knockdown cells generate hypermethylated tumors**

To address the causal relationship between reduced TET1 expression and promoter hypermethylation, we depleted TET1 in the non-CIMP cell line SW620 by transduction of a lentiviral short-hairpin-RNA (sh-RNA) expression construct. A non-targeting sh-RNA (sh-ctrl) was included as a control. The two most efficient knockdown (kd) clones shTET1-1 (60% knockdown) and shTET1-2 (68% knockdown) derived from two different sh-RNAs were subjected to further analysis (Figure 3a). Notably, although sh-RNA sequences were specifically designed for TET1, we also found reduced TET2 mRNA levels in both knockdown clones (Figure 3a). We cultured shTET1 knockdown clones for 6 passages after viral transduction before assessing global levels of 5mC and 5hmC by dot blot analysis with specific antibodies and LC-MS/MS. 5hmC was clearly reduced in both knockdown clones whereas total 5mC levels were unaffected, as expected (Figure 3b, Supplementary Figure 3a/b). To address whether and how the loss of TET1 affects CpG methylation we performed genome-scale methylation analysis using the Illumina Infinium HumanMethylation450 (HM450) platform. Focusing the analysis on CGI-associated CpG sites (252,129 CpGs), we found both hypermethylation and hypomethylation resulting from TET1 depletion (Supplementary Figure 3d). A total of 1698 CpG sites were differentially methylated (FDR-adjusted  $p < 0.0001$ , Fold change in methylation M-values  $> 4$ ), commonly in both TET1 kd-clones compared to sh-ctrl, comprising 859 hypermethylated and 839 hypomethylated CpG sites (Figure 3d). To follow up knockdown methylation patterns in the process of tumor development *in vivo*, sh-ctrl and

shTET1 kd cells were xeno-transplanted subcutaneously into NOD/SCID mice. Genome-wide methylation patterns (HM450) were analysed when tumors reached a diameter of 10 mm (5-7 weeks). We verified TET1 knockdown in mouse xenografts (Supplementary Figure 3c). A total of 1674 CpG sites were differentially methylated, commonly in xenografts from both TET1-kd clones, 1069 of which were hypermethylated and 605 were hypomethylated (Figure 3d, Supplementary Figure 3d). Hence, in the process of xenograft-tumor formation the ratio of hyper- to hypomethylated sites increased significantly ( $p < 0.0001$ , Chi-square), with CpGs showing hyper- and hypomethylation overlapping significantly between cell lines and xenografts (each with  $p < 0.0001$ , Chi-square) (Figure 3d). To characterize the CGI-associated CpGs, differentially methylated upon the depletion of TET1, we further subdivided them into island, shore (0-2kb from island) and shelf (2-4kb from island). In TET depleted cells, hyper- and hypomethylated CpGs ranked randomly among the three subcategories. However, in xenografts hypermethylated CpGs were enriched in the island category compared to hypomethylated CpGs ( $p = 0.036$ , fisher's exact).

To address whether methylation changes upon TET1 depletion mirror to some extent methylation patterns of CIMP CRCs, we assessed methylation levels of the CIMP five-marker panel and hMLH1 by In-qMSP. For several marker genes (IGF2, NEUROG1 and RUNX3) methylation levels in promoter CGIs increased (Supplementary Figure 4a/b). The increase of MLH1 promoter methylation in TET1-kd cells was accompanied by reduced MLH1 mRNA expression (Supplementary Figure 4c). We also asked whether loss of TET1 expression affects methylation of its own promoter and investigated the TET1 promoter CGI methylation in shTET1-kd cells by pyrosequencing of bisulfite-converted genomic DNA. This identified three CpG located at the 3' end of the CGI as highly methylated in shTET1-1 but not in shTET1-2 cells. Data from the HM450 confirmed the increased methylation of individual CpG sites within the TET1 promoter in shTET1-kd cells and mouse xenografts compared to control samples (data not shown).

From these results we conclude that TET1 depletion in SW620 CRC cells gives rise to both hyper- and hypomethylation at CGIs, including known CIMP targets as well as the TET1 promoter itself. Xenograft tumor formation from these TET1 knockdown cells, however, is accompanied with a significant increase in the fraction of hypermethylated CpGs compared to the respective knockdown cell line, indicating that tumor growth generates a bias towards hypermethylation in CIMP.

## **TET1 knockdown cells and xenografts show a reduced expression of the colon cancer initiating cell marker LGR5**

In mouse embryonic stem cells Tet1 is critical for stem cell identity (Freudenberg et al., 2012; Ito et al., 2010). To address the role of TET1 in the stem cell phenotype of CRC cells we quantified mRNA levels of LGR5, a generally accepted marker for CRC-initiating cells but also normal epithelial intestinal stem cells (Dame et al., 2014; Forster et al., 2014; Hirsch et al., 2014). LGR5 mRNA expression was consistently reduced in shTET1-kd cells compared to sh-ctrl cells (Figure 4a). Similarly, xenografts derived from shTET1-kd cells showed significantly reduced number of LGR5-positive cells. Also the fraction of xenograft cells positive for CD44 and CD166, two additional putative markers for colon cancer initiating cells was reduced in TET1-kd tumors compared to sh-ctrl (Dalerba et al., 2007) (Figure 4b). We therefore conclude that TET1 expression is associated with a stem cell-like phenotype in human CRC cells.

## **TET1 knockdown cells display reduced cell state plasticity**

In cultures, the SW620 CRC cell line constitutes a heterogeneous population of cells showing stem-like cells features or a more differentiated phenotype, as inferred from the heterogeneous expression of putative cancer-initiating cell markers CD24 and CD44 (Muraro et al., 2012). It was reported previously that such cell states are interconvertible, most likely as a result of epigenetic plasticity involving DNA demethylation (Carmona et al., 2014; Gupta et al., 2011). We thus wondered whether the loss of TET1 would affect the cell-state plasticity. Flowcytometric analysis showed that the two shTET1-kd clones behave differently regarding the expression of CD24 and CD44. Whereas, shTET1-1 displayed a higher fraction of cells negative for both markers (double negative) compared to sh-ctrl, shTET1-2 was enriched for cells positive for both markers (double positive) (Figure 5a and 5b). Using fluorescence-activated cell sorting (FACS), we isolated cells positive for CD24 and CD44 (CD24<sup>hi</sup>CD44<sup>hi</sup>, double positive) and negative for both (CD24<sup>neg</sup>CD44<sup>neg</sup>, double negative) from sh-ctrl and shTET1-kd cells (resulting in subpopulations which were at least 96% pure based on reanalysis immediately after sorting). We then allowed the separated cells to expand in culture for eight days before we analysed phenotypic composition again by flow cytometry. Sh-ctrl cells from both subpopulations showed progression towards establishing the original population heterogeneity. However, the double negative subpopulation of shTET1-1 cells largely remained in their cell state, whereas the double positive cells showed an enhanced transition



to the double negative state compared to sh-ctrl (Figure 5b). The behaviour of shTET1-2 diametrically opposite: Double positive cells remained largely in this state whereas double negative cells showed a more efficient conversion to double positive state compared to sh-ctrl (Figure 5b). Since we did not observe any difference in cell proliferation, the differences in the progression rate were unlikely the result from different growth rates.

From these results we conclude that depletion of TET1 in SW620 cells reduced their capacity to change cell states and establish a stable heterogeneity in the population. Cells appear to be blocked in their differentiation state at the time of TET1 depletion and fail to memorize differentiation cues from the environment.

### **TET1-knockdown cells lose epithelial cell characteristics and acquire mesenchymal features**

shTET1-kd cell cultures showed morphological changes, they formed colonies larger than control cells (sh-ctrl) with a dispersed composition in which individual cells were identifiable (Figure 6a). This as well as the observed loss of LGR5 expression could be interpreted as a loss of the epithelial phenotype. Therefore, we quantified mRNA expression of E-cadherin (CDH1) which constitutes an essential component of adherence junctions mediating cell-to-cell contacts and which expression is characteristic of epithelial cells (van Roy and Berx, 2008). CDH1 mRNA levels were drastically reduced in both shTET1-kd clones and this was accompanied with increased methylation in the CDH1 promoter (Figure 6b). Consistent with this observation, CDH1 mRNA levels are reduced in CIMP-CRCs from the TCGA validation set (Figure 6c; Cancer Genome Atlas; 2012). By contrast, mRNA levels of vimentin (VIM) and fibronectin (FN1), two markers of mesenchymal cells, were increased in TET1-kd cells (Figure 6c). Consistently with the acquisition of a mesenchymal phenotype, both SW620 TET1-kd clones showed an increased level of cell migration compared to sh-ctrl cells (Figure 6e).

ZEB1, SNAI1 and SNAI2 are EMT-promoting transcription factors which directly inhibit expression of E-cadherin (J. Xu et al., 2009). Although mRNA levels of these factors were affected by depletion of TET1, only ZEB1 was consistently up-regulated in both kd-clones (Supplementary Figure 5a). We thus conclude that the loss of TET1 induces a change in cell phenotype that shows aspects of but does not fully recapitulate EMT.

## Discussion

The molecular pathway resulting in the CpG island methylator phenotype (CIMP) in human CRC has remained elusive. This work shows that CRC-CIMP is associated with a downregulation of the DNA hydroxylase TET1. Reduced TET1 expression results in CGI-methylation alterations, which evolve to a methylator phenotype during tumor growth, accompanied by the acquisition of molecular characteristics of CRC-CIMP, including the transcriptional repression of MLH1 and CDH1. These findings are in line with recent reports suggesting that impaired TET-mediated DNA demethylation is causal for the development of hypermethylation phenotypes in several other human cancers (Figueroa et al., 2010; Letouze et al., 2013; Turcan et al., 2012). Furthermore, we provide evidence that reduced TET1 expression affects the cell phenotype and the cell state plasticity of CRC cells with potential clinical implications.

In contrast to previous reports, which associated CRC-CIMP with elevated expression of DNA methyltransferases, we observed significantly lower DNMT expression levels in CIMP-associated CRCs (Kanai et al., 2001; Nosho et al., 2009). These discrepancies can be explained by (i) the different methods used for DNMT expression quantitation (IHC versus mRNA based methods), (ii) the different sample sizes but also (iii) the large variety of DNMT transcripts and enzyme isoforms. Several DNMT isoforms display no methyltransferase activity but have regulatory functions which can even interfere with DNA methylation (Gordon et al., 2013; Weisenberger et al., 2004). The observation that 5hmC levels are significantly reduced in CGIs of CIMP marker genes further indicate that the molecular basis of CRC-CIMP is rather related to TET function than to a deregulation of DNMTs.

The sh-RNA mediated TET1 knockdown in the CRC cell line SW620 resulted in genome-wide CGI methylation changes with almost equal dimensions of hyper- and hypo-methylation. Previous studies showed that the loss of Tet1 expression in mESCs does also lead to CpG hyper- and hypomethylation (Y. Xu et al., 2011). Based on its function in DNA demethylation the loss of TET1 is expected to result exclusively in *de novo* methylation at its target regions. The possible causes for hypomethylation linked to TET1 depletion remain elusive, and may be related to the dynamics of methylation regulation (Shen et al., 2013) or occur secondary to hypermethylation-associated transcriptional deregulation of TET1 target genes. We investigated the further development of the aberrant methylation in the context of tumor formation by engrafting TET1-kd cells into immunosuppressed mice, thus allowing the effects of more physiological tumor growth conditions (nutrient supply, tumor-stroma interactions, hypoxia) on the shaping of DNA methylation patterns to be examined. Compared to the TET1-kd cells, which displayed a general perturbation of DNA methylation in both directions,

hypermethylation was clearly predominant in their xenografts. Interestingly, several known CIMP marker genes became consistently hypermethylated upon the knockdown of TET1 and might thus represent a distinct subset of TET1 targets. One of these genes is the DNA mismatch repair gene MLH1 the hypermethylation of which and concomitant transcriptional silencing are hallmarks of CIMP-associated CRCs (Weisenberger et al., 2006). The resulting mismatch repair deficiency leads to instability in the length of microsatellite DNA sequences (MSI) due to a failure to correct polymerase slippage events at those repetitive elements. We found that MLH1 hypermethylation in TET1-kd cells is accompanied by transcriptional repression of MLH1. However, the only moderate loss of MLH1 mRNA expression in TET1-kd cells and xenografts was not sufficient to cause microsatellite instability (data not shown).

In addition, we present evidence that the transcriptional repression of TET1 and the hypermethylation of its promoter can generate a negative feedback loop. We found that the loss of TET1 expression results in an accumulation of 5mC in its own promoter region, which then reinforces and presumably stabilizes its transcriptional repression. In consequence, transcriptional repression of the TET1 promoter followed by its hypermethylation could constitute an epigenetic driver event, which promotes subsequent accumulation of 5mC in TET1 targeted promoter CGIs resulting in a methylator phenotype. Interestingly, induced BRAF V600E expression in the mouse fibroblast cell line NIH3T3 has been shown to reduce expression of all three TETs most prominently TET1. This effect was apparently independent from MAPK-signaling activation (Kudo et al., 2012). The BRAF V600E mutation, which is highly associated with CIMP in colorectal cancer, might therefore be a promising candidate for an upstream factor.

Furthermore, we showed that reduced TET1 expression in the SW620 cell line results in the loss of LGR5, an marker for epithelial colonic stem cells, and the putative CRC-initiating cell markers CD44 and CD166 (Dalerba et al., 2007; Koo and Clevers, 2014). Together with the loss of stem cell markers, we found that the depletion of TET1 results in a reduction of cell state plasticity. TET1 depletion shifted the cell population equilibrium state regarding the expression of the cancer-initiating cell markers CD24 and CD44. Isolated subpopulations of CD24<sup>hi</sup>/CD44<sup>hi</sup> and CD24<sup>low</sup>/CD44<sup>low</sup> TET1-kd cells reverted more efficiently to the initial cellular heterogeneity than control cells. We propose, that impaired DNA demethylation due to the loss of TET1, restricts epigenetic plasticity and locks the cells in the current cell state. External signals may modulate the phenotype of TET1 deficient cells, but the cells cannot epigenetically memorize these changes and preferentially revert to their locked state in the absence of these signals.

Finally, we showed that the loss of TET1 expression promotes EMT-like processes and enhances cellular migration. This coincides with promoter hypermethylation and concomitant

downregulation of E-cadherin (CDH1) and the upregulation of the mesenchymal markers fibronectin (FN1) and vimentin (VIM). EMT and the associated migratory and invasive cell behavior are involved in metastasis formation (Ocana et al., 2012; Thiery et al., 2009; Tsai et al., 2012). Both, the loss of E-cadherin as well as CIMP have been previously associated with a poor prognosis in colorectal cancer (Bruun et al., 2014; Juo et al., 2014). Our results provide a mechanistic link between the prognostic features of CIMP and loss of E-cadherin in CRC. Notably, in hepatocellular and gastric cancers the loss of TET1 expression was reported to be associated with an unfavorable prognosis (Liu et al., 2013; Yang et al., 2013).

Overall our study provides novel insight into the molecular mechanisms involved in the development of CIMP-associated CRC and identifies TET1 as a factor involved in the maintenance of CGI methylation patterns in colonic epithelial cells.

## **Methods**

### **Patients, Tissue Samples and Cell Lines**

Tissue specimens from colorectal cancer patients were obtained from the Department of Surgery of the Kantonsspital Aarau, Switzerland and the Department of Gastroenterology of the Inselspital in Berne, Switzerland and were randomly collected on the occasion of surgical therapy. From each patient two biopsies were taken, one from the primary tumor and one from the adjacent healthy appearing mucosa (excised approximately 7cm aside the primary tumor). Tissue samples from healthy individuals were randomly collected by the Department of Gastroenterology of the Spital Region Oberaargau in Langenthal, Switzerland during routinely performed colonoscopy examinations. All patients gave their informed consent for the use of their tissue specimens for research purposes. Immediately after removal all tissue samples were submerged in RNeasy (Qiagen) and stored at -80°C until further usage. For patients clinical data see Supplementary Table 1.

Cell lines used in this study and the composition of their culture media are given in Supplementary Table 2. All cell lines were cultured at 37°C with 100% humidity and 5% CO<sub>2</sub>.

### **DNA extraction and Bisulfite Conversion**

For the preparation of genomic DNA from cell lines 10<sup>6</sup>-10<sup>7</sup> cells were collected in 300µl Tris-EDTA buffer (10mM Tris-HCl pH 8.0, 1mM EDTA) and 300µl lysis buffer (20mM Tris-HCl pH 8.0, 4mM EDTA, 20mM NaCl, 1% SDS, 1mg/ml proteinase K) were added. After incubation at 55°C for 5h and subsequent phenol/chloroform extraction DNA was Na-acetate/ethanol precipitated. DNA pellets were resuspended in 10mM Tris-HCl pH 8.0. RNA was removed by incubation with 20µg/ml RNase A (Qiagen) at 37°C for 30min followed by phenol/chloroform extraction and Na-acetate/ethanol precipitation. DNA pellets were resuspended in 10mM Tris-HCl pH 8.0 and concentration was measured by absorbance at 260nm.

Genomic DNA from RNeasy-preserved colorectal cancer tissues was extracted with the QIAamp DNA mini kit (Qiagen) according to the manufacturer's instructions including the RNase-treatment step.

1µg of genomic DNA was treated with sodium bisulfite using the EZ DNA Methylation Kit (Zymo Research) according to manufacturer's instructions.

### **Locus-normalized Quantitative Methylation Specific PCR (In-qMSP)**

In-qMSP was performed as described previously (Noreen et al., 2014). Reactions were carried out on the Rotor-Gene 3000 real time thermal cycler (Qiagen) using the QuantiTect® SYBR® Green Kit (Qiagen) according to the manufacturer's instructions. Each reaction was carried out in duplicates in a final volume of 15µl containing 0.5µM of each primer and 1.5µl of bisulfite-converted DNA. DNA from a peripheral blood sample was treated with M.SssI methyltransferase to obtain full methylation within the promoter regions of all target genes (methylation positive control). For each target gene standard curves for the M- (specific for bisulfite-unconverted sequence) and S- (standardizer; methylation independent) primers were established by amplifying serial 1:10 dilutions (100% - 0.1%) of the bisulphite-treated positive control DNA. Methylation of target genes (expressed as percentage of methylated alleles [PMA]) in sample DNA was quantified by relating M-amplification rate to the S-amplification rate both of which were previously determined using the recorded standard curves. DNA from peripheral blood showed no or only minor fractions of methylated alleles for the target genes and was therefore used as a methylation negative control. For In-qMSP primer sequences see Supplementary Table 3.

### **Pyrosequencing of Bisulfite-Converted DNA**

Gene promoter regions were PCR amplified using primers annealing methylation-independently to the (oxidative) bisulfite-converted sequence. Reverse primers were either directly Biotin-labeled (5'-end of reverse primer) or extended at the 5'-end by an 11 nucleotide long universal primer target sequence. In the latter case PCR amplifications were performed in 20µl reaction volume, PCR products were diluted 1:10<sup>5</sup> with 10mM Tris-HCl pH 8.0 and again PCR amplified with a universal biotin-labeled reverse primer in 50µl reactions. The biotin-labeled PCR products were purified using the NucleoSpin Gel and PCR clean-up kit (Macherey-Nagel). In case of direct biotin-labeling PCR reactions were performed in 50µl and were used for pyrosequencing without further processing. All PCR reactions were performed with 0.03units TrueStart Hot Start Taq DNA polymerase (Thermo) per µl reaction volume,

0.5 $\mu$ M primers and 2mM MgCl<sub>2</sub>. Biotin-labeled PCR products were immobilized on streptavidin-coated sepharose beads (GE Healthcare) and sequenced with a PyroMark Q24 pyrosequencing system (Qiagen) using methylation independent sequencing primers according to manufacturer's instructions. Sequence of PCR and sequencing primers are given in Supplementary Table 4.

### **hmC Dot Blot**

3 $\mu$ l of Rnase-treated genomic DNA was blotted onto an H-bond N+ nylon membrane (Amshersham) and dried for 15min. Membrane-bound DNA was denatured in 400mM NaOH for 4h. Membrane was washed twice with SSC buffer pH7 (300mM NaCl, 34mM Na-citrate) and blocked with 10% milk in TBST (20mM Tris-HCl pH7.5, 150mM NaCl, 0.1% Tween 20) for 1hr at RT. After incubation membrane was washed three times with TBST. Anti-5hmC antibody (Active Motif) was diluted 1:10<sup>4</sup>. Primary antibody was incubated overnight at 4°C. After washing three times for 10min in TBST HRP-linked  $\alpha$ -rabbit secondary antibody (GE Healthcare) was diluted 1:10<sup>4</sup> in 10% milk (in TBST) and incubated for 1hr at RT. After again washing three times with TBST for 10min 5hmC signals were detected on Amersham Hyperfilm ECL (GE Healthcare) using the Pierce ECL Western Blotting Substrate (Thermo). In order to prove equal sample loading the membrane was subsequently stained with 0.02% methylene blue in 300mM Na-citrate pH5.2 for 3min.

### **MeDIP (Methylated DNA Immunoprecipitation) and GLIP (glucosylation, periodate oxidation and biotinylation)**

MeDIP detection of 5-mC was performed essentially as described in (Weber et al., 2005). DNA was sonicated to yield fragments of 100-500bp followed by NaCl/ethanol precipitation in the presence of glycogen-carrier. 1 $\mu$ g fragmented DNA in TE buffer (10mM Tris-HCl pH 8.0, 1mM EDTA) was denatured (95°C for 10min) and incubated with 0,3 $\mu$ g of monoclonal anti-5-methylcytidine (Diagenode, clone 33D3) at 4°C for 2 h in 1x immunoprecipitation (IP) buffer (10mM sodium phosphate pH 7.0, 140 mM NaCl, 0.05% Triton X-100). Immuno-complexes were precipitated by the addition of 20  $\mu$ l M-280 sheep anti-mouse IgG antibody coupled

Dynabeads (Invitrogen) and incubation at 4°C for 2 h followed by three washes in IP buffer. Bound material was treated with 250 µl proteinase K digestion buffer (50 mM Tris-HCl pH 8.0, 10 mM EDTA, 0.5% SDS and 0.25mg/ml proteinase K) at 50°C for 3 h. Immunoprecipitated methylated DNA was purified by phenol/chloroform extraction followed by NaCl/ethanol precipitation and re-suspended in 10 mM Tris-HCl pH 8.0. 5-hmC containing DNA fragments were captured with the Hydroxymethyl Collector kit from Active Motif as described in the manufacturer's instructions. qPCR analysis of sonicated genomic input DNA and MeDIP/GLIB DNA with target specific primers was performed using Quantitect SYBR Green (Qiagen) with a Rotor-Gene 3000 thermocycler (Qiagen). Sequences of primers used in MeDIP and GLIB are given in Supplementary Table 5.

### **LC-MS/MS (High Performance Liquid Chromatography Coupled with Tandem Mass Spectrometry)**

Genomic DNA was enzymatically hydrolyzed to nucleosides as described (Crain, 1990), followed by addition of 3 volumes of methanol and centrifugation (16,000g, 30 min, 4°C). The supernatants were dried and dissolved in 50 µl 5% methanol in water (v/v) for LCMSMS analysis of the deoxynucleosides 5-hm(dC), 5-f(dC), 5-ca(dC), and 5-hm(dU). A portion of each sample was diluted for the quantification of 5-m(dC) and unmodified deoxynucleosides (dA, dC, dG, and dT). Chromatographic separation was performed on a Shimadzu Prominence HPLC system with a Zorbax SB-C18 2.1x150 mm i.d. (3.5 µm) column equipped with an Eclipse XDB-C8 2.1x12.5 mm i.d. (5 µm) guard column (Agilent Technologies). The mobile phase consisted of water and methanol (both supplemented with 0.1% formic acid), for 5-m(dC) and 5-hm(dC) starting with a 5 min gradient of 5-60% methanol, followed by 6 min re-equilibration with 5% methanol, and for unmodified nucleosides maintained isocratically with 85% methanol. Mass spectrometry detection was performed using an MDS Sciex API5000 triple quadrupole (Applied Biosystems) operating in positive electrospray ionization mode for the mass transitions 258.1/ 142.1 (5-hm(dC)), 242.1/ 126.1 (5-m(dC)), 252.1/136.1 (dA), 228.1/112.1 (dC), 268.1/152.1 (dG), and 243.1/127.1 (dT).



## **Array-based DNA Methylation Profiling and Data Analysis**

Array-based gene-specific DNA methylation analysis was performed using Infinium HumanMethylation450 bead chip technology (Illumina, San Diego, CA, USA). Genomic DNA (500 ng) from each sample was bisulfite converted using the EZ-96 DNA Methylation Kit (Zymo Research Corporation). Bisulfite treated genomic DNA was whole-genome amplified and hybridized to HumanMethylation450 BeadChips. The methylation status of a specific cytosine was indicated by average beta (AVB) values where 1 corresponds to full methylation and 0 to no methylation. Raw array data were quantile normalised as described previously (Nizar Touleimat 2012) and p-values for comparisons between different datasets were statistically adjusted using the Benjamini-Hochberg correction. For the data analysis we masked data point as "NA" for the following conditions 1) Probes containing single-nucleotide polymorphisms (SNPs), 2) those that overlap with a repetitive element that covers the targeted CpG dinucleotide, 3) those that overlap with regions of insertions and deletions in the human genome, 4) probes that are targeting X or Y chromosome. Beta values with detection p value greater than 0.05 were also replaced as "NA". Furthermore, probes containing at least one "NA" across any sample were also masked as "NA". After data filtering, total of 413,594 probes were kept for differential methylation analysis using R package limma. Individual loci were scored as differentially methylated if fold change difference was more than 4 (FC>4) and p value <0.001.

Genome-wide methylation data of 8 CRC cell lines was obtained from GEO (GSE35573), which was done on Infinium HumanMethylation27 Beadchip arrays interrogating methylation at 27,578 CpGs distributed in the promoters of 14,475 coding genes. Data was normalized using lumi and methylumi packages in R/Bioconductor as described previously (Du P 2008). After quality control and data filtering, a total of 19,773 probes were further analysed.

### **5-Azacytidine Treatment**

SW48 CRC cells were cultured in standard medium containing 1 $\mu$ M 5-Azacytidine (5-AzaC, Sigma) for 5 days. Medium containing fresh 5-AzaC was renewed every 24 h. Genomic DNA (see DNA extraction paragraph) and total RNA (see qRT-PCR paragraph) were extracted from treated and untreated cells.

## **Quantitative Reverse Transcription PCR (qRT-PCR)**

Total RNA was extracted from tissue specimens and CRC cell lines using the Tri-Reagent (Sigma #93289) according to manufacturer's instructions. For tissue samples, 50-80mg of tissues were homogenized directly in 1ml Tri-Reagent with the MP-FastPrep-24 system together with Zirconia beads (BioSpec). For cell lines,  $10^6$ - $10^7$  cultured cells were resuspended in 1ml Tri-Reagent. Extracted RNA was Dnase treated (New England Biolabs) and cDNA was synthesized using either the RevertAid H Minus First Strand cDNA Synthesis Kit (Thermo Scientific) or the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer's manuals. All qRT-PCR reactions were carried out on the Rotor-Gene 3000 real time thermal cycler (Qiagen) using the QuantiTect<sup>®</sup> SYBR<sup>®</sup> Green Kit (Qiagen). 15 $\mu$ l reactions were performed containing 0.5 $\mu$ M of each primer as well as 1 $\mu$ l cDNA template. mRNA levels of target genes were expressed as the averaged expression relative to ACTB and GAPDH. For qRT-PCR primer sequences see Supplementary Table 5.

## **Western Blot**

Cells lysed (50mM Na-P buffer pH 8, 125mM NaCl, 1% NP-40, 500 $\mu$ M EDTA, 1mM PMSF, 1x cOmplete EDTA free [Roche], 1mM DTT) on ice for 30min. After centrifugation at 21,000g for 15min at 4°C, supernatant was collected, aliquoted and snap-frozen in liquid nitrogen prior use. Total protein concentration in lysate was determined using the Protein Assay Dye Reagent Concentrate (Bio-Rad). 50 $\mu$ g total protein were boiled in SDS loading buffer (45mM Tris-HCl pH 8, 10% glycerol, 1% SDS, 0.01% bromphenol blue, 50mM DTT) at 95°C for 2min and afterwards loaded onto 7% and 10% PAA (amresco) gels, respectively, together with 5 $\mu$ l of Precision Plus Dual Color protein standard (Bio-Rad). Proteins were separated at 100V and transferred onto a nitrocellulose membrane (Whatman, GE Healthcare) in transfer buffer (25mM Tris, 192mM glycine, 10% methanol, 0.005% SDS) at 30V and 4°C o/n. Unspecific binding sites were blocked by incubation with 10% milk in TBST (20mM Tris-HCl pH7.5, 150mM NaCl, 0.1% Tween 20) for 1hr at RT and then sequentially probed with 1:1000 (5% milk in TBST) anti-Tet1 antibody (polyclonal rabbit, Abiocode #R1084-4) for 4h at RT and HRP-linked anti-rabbit antibody (GE Healthcare). For loading control blots were probed with 1:10,000 (5% milk in TBST) anti-Actin antibody (polyclonal mouse, Abcam) and HRP-linked

anti-mouse antibody (GE Healthcare). Blots were washed with TBST and subsequently incubated with Pierce ECL Western Blotting Substrate (Thermo) for 1min at RT. Chemiluminescence signals were detected on an Amersham Hyperfilm ECL (GE Healthcare).

### **Generation of TET1 knockdown cell line**

A set of five pLKO.1-puro vectors encoding for short hairpin RNA constructs against TET1 mRNA (TRCN0000075023, TRCN0000075024, TRCN0000075025, TRCN0000075026, TRCN0000075027) as well as a pLKO.1-puro non-mammalian target shRNA control vector were obtained from Sigma-Aldrich as bacterial glycerol stocks (#SHCLNG-NM\_030625 and #SHC002). Sh-vectors were purified from o/n cultures in LB medium containing 100µg/ml Ampicillin using the QIAfilter Plasmid Midi Kit (Qiagen) together with the EndoFree Plasmid Buffer Set (Qiagen). Identity of purified sh-vectors was checked by restriction digest with BamHI and XhoI. Using the JetPEI transfection reagent (Polyplus transfection) with a N/P ratio of 5 HEK293T cells were cotransfected with 0.5µg sh-vector and 5µl packaging mix (Sigma-Aldrich) containing the two lentiviral packaging vectors. Lentivirus-containing supernatant medium was collected 48h and 96h after transfection, pooled and concentrated using the Lenti-X Concentrator (Clontech). Lentiviral particles (LPs) were quantified by means of the Lenti-X p24 Rapid Titer ELISA Kit (Clontech # 632200). Aliquots of LPs were frozen at -80°C before use.

1.2-1.4x10<sup>5</sup> cells of SW620 were transduced with lentiviral particles in 48-well plates at MOI's (multiplicity of infection) ranging from 0.5 to 8 using 8µg/ml Hexadimethrine bromide. Medium was replaced 24h post-transduction and 48h post-transduction with standard medium containing 5µg/ml Puromycin (Sigma). Cells were passaged three times in selective medium before individual cell clones were picked. Knockdown efficiency was analysed in monoclonal knockdown cells by qRT-PCR.

### **Mouse xenografts**

In vivo experiments were approved by the Basel Cantonal Veterinary Office. NOD/SCID mice, that were initially obtained from Charles River Laboratories (Germany), and bred and

maintained under pathogen free conditions in the animal facility of the Department of Biomedicine of the University of Basel. Eight to ten week old mice were used for xenografting experiments.

10<sup>5</sup> TET1-knockdown SW620 cells were resuspended in a 1:1 mixture of phosphate-buffered saline and growth factor-reduced Matrigel matrix (BD Biosciences) and inoculated subcutaneously into the flank of recipient mice. Tumor development was monitored by palpation. Time to onset of a palpable tumor was recorded and the tumor size was measured weekly by a dial caliper. Mice were sacrificed when tumors reached a maximum diameter of 10 mm.

### **Flow cytometric analysis and fluorescence activated cell sorting (FACS)**

Xenograft tissues were enzymatically digested in serum-free DMEM (GIBCO), containing 1mg/ml Collagenase IV (Sigma Aldrich) and 50mg/ml DNase I (Sigma-Aldrich), as well as several antibiotics, including Kanamycin (GIBCO), Amphotericin B (Sigma-Aldrich), 200mg/ml Metronidazol (Sigma-Aldrich) and 6mg/ml Cefuroxim (Sigma-Aldrich). After digestion, xenograft tissues were ground inside a 100 µm sterile cell filter with the blunt end of a Luer syringe while continually rinsing with DMEM medium. The filtrate was filtered again through a 70 µm sterile cell filter and then centrifuged at 300g for 5min. The supernatant was removed and cells were resuspended in 1 mL of DMEM medium.

For flow cytometric analysis the following antibodies were used: AlexaFluor488 (AF488) labeled anti-EpCAM (1:200, clone 9C4, Biolegend), Phycoerythrin (PE) labeled anti-CD166 (1:25, clone 3A6, BD Biosciences), Peridinin Chlorophyll Protein Complex- (PerCP) labeled anti-CD24 (1:100, clone ML5, BD Biosciences), Allophycocyanin (APC) labeled anti-CD133 (1:25, clone AC133/1, MiltenyiBiotec), Allophycocyanin cyanine 7 analog (APC-H7) labeled anti-CD44 (1:200, clone G44-26, BD Biosciences), Brilliant Violet 421 (BV421) labeled anti-Lgr5 (1:200, clone 8F2, BD Biosciences). LIVE/DEAD Fixable Yellow Dead Cell Stain Kit (Molecular Probes) was used according to manufacturer's specifications. Relative fluorescence intensities were measured using a BD LSRFortessa cell analyzer (BD Biosciences), following exclusion of dead cells based on high Yellow Dead signal. Analysis was performed using FlowJo software (Tree Star).

SW620 TET1-kd cells and sh-ctrl. cells were sorted using the above specified anti-CD24 and anti-CD44 antibodies. FACS was performed using a BD FACSAria IIu (BD Biosciences).

## **Cell migration / invasion assay**

The migratory behavior and the invasive potential of SW620 TET1 knock down cells was assessed by an in vitro chemoinvasion assay as previously described (Albini and Benelli, 2007). Briefly, SW620 cells were seeded into upper chambers of transwell plates onto uncoated (cell migration) or matrigel-coated (invasiveness) membranes (8 $\mu$ m pore size, BD Biocoat Tumor invasion assay, BD Biosciences). Lower chambers contained standard culture medium supplemented with 5% FBS. Plates were incubated at 37°C for 20 h. Inserts were then removed and the cell numbers in the lower chambers were quantified by CyQUANT Cell Proliferation Assay Kit (Invitrogen).

## References

- Albini, A., & Benelli, R. (2007). The chemoinvasion assay: a method to assess tumor and endothelial cell invasion and its modulation. *Nat Protoc*, 2(3), 504-511. doi: 10.1038/nprot.2006.466
- Ang, P. W., Loh, M., et al. (2010). Comprehensive profiling of DNA methylation in colorectal cancer reveals subgroups with distinct clinicopathological and molecular features. *BMC Cancer*, 10, 227. doi: 10.1186/1471-2407-10-227
- Bleeker, F. E., Lamba, S., et al. (2009). IDH1 mutations at residue p.R132 (IDH1(R132)) occur frequently in high-grade gliomas but not in other solid tumors. *Hum Mutat*, 30(1), 7-11. doi: 10.1002/humu.20937
- Borger, D. R., Tanabe, K. K., et al. (2012). Frequent mutation of isocitrate dehydrogenase (IDH)1 and IDH2 in cholangiocarcinoma identified through broad-based tumor genotyping. *Oncologist*, 17(1), 72-79. doi: 10.1634/theoncologist.2011-0386
- Bruun, J., Kolberg, M., et al. (2014). Prognostic Significance of beta-Catenin, E-Cadherin, and SOX9 in Colorectal Cancer: Results from a Large Population-Representative Series. *Front Oncol*, 4, 118. doi: 10.3389/fonc.2014.00118
- Cancer Genome Atlas, N. (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature*, 487(7407), 330-337. doi: 10.1038/nature11252
- Carmona, F. J., Davalos, V., et al. (2014). A Comprehensive DNA Methylation Profile of Epithelial-to-Mesenchymal Transition. *Cancer Res*, 74(19), 5608-5619. doi: 10.1158/0008-5472.CAN-13-3659
- Crain, P. F. (1990). Preparation and enzymatic hydrolysis of DNA and RNA for mass spectrometry. *Methods Enzymol*, 193, 782-790.
- Dalerba, P., Dylla, S. J., et al. (2007). Phenotypic characterization of human colorectal cancer stem cells. *Proc Natl Acad Sci U S A*, 104(24), 10158-10163. doi: 10.1073/pnas.0703478104
- Dame, M. K., Jiang, Y., et al. (2014). Human colonic crypts in culture: segregation of immunochemical markers in normal versus adenoma-derived. *Lab Invest*, 94(2), 222-234. doi: 10.1038/labinvest.2013.145
- Easwaran, H., Johnstone, S. E., et al. (2012). A DNA hypermethylation module for the stem/progenitor cell signature of cancer. *Genome Res*, 22(5), 837-849. doi: 10.1101/gr.131169.111
- Etoh, T., Kanai, Y., et al. (2004). Increased DNA methyltransferase 1 (DNMT1) protein expression correlates significantly with poorer tumor differentiation and frequent DNA hypermethylation of multiple CpG islands in gastric cancers. *Am J Pathol*, 164(2), 689-699. doi: 10.1016/S0002-9440(10)63156-2
- Figuroa, M. E., Abdel-Wahab, O., et al. (2010). Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. *Cancer Cell*, 18(6), 553-567. doi: 10.1016/j.ccr.2010.11.015
- Forster, R., Chiba, K., et al. (2014). Human intestinal tissue with adult stem cell properties derived from pluripotent stem cells. *Stem Cell Reports*, 2(6), 838-852. doi: 10.1016/j.stemcr.2014.05.001
- Freudenberg, J. M., Ghosh, S., et al. (2012). Acute depletion of Tet1-dependent 5-hydroxymethylcytosine levels impairs LIF/Stat3 signaling and results in loss of embryonic stem cell identity. *Nucleic Acids Res*, 40(8), 3364-3377. doi: 10.1093/nar/gkr1253
- Gordon, C. A., Hartono, S. R., et al. (2013). Inactive DNMT3B splice variants modulate de novo DNA methylation. *PLoS One*, 8(7), e69486. doi: 10.1371/journal.pone.0069486
- Gupta, P. B., Fillmore, C. M., et al. (2011). Stochastic state transitions give rise to phenotypic equilibrium in populations of cancer cells. *Cell*, 146(4), 633-644. doi: 10.1016/j.cell.2011.07.026
- He, Y. F., Li, B. Z., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science*, 333(6047), 1303-1307. doi: 10.1126/science.1210944
- Hinoue, T., Weisenberger, D. J., et al. (2012). Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*, 22(2), 271-282. doi: 10.1101/gr.117523.110

- Hirsch, D., Barker, N., et al. (2014). LGR5 positivity defines stem-like cells in colorectal cancer. *Carcinogenesis*, 35(4), 849-858. doi: 10.1093/carcin/bgt377
- Ito, S., D'Alessio, A. C., et al. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310), 1129-1133. doi: nature09303 [pii]
- 10.1038/nature09303
- Ito, S., D'Alessio, A. C., et al. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature*, 466(7310), 1129-1133. doi: 10.1038/nature09303
- Juo, Y. Y., Johnston, F. M., et al. (2014). Prognostic value of CpG island methylator phenotype among colorectal cancer patients: a systematic review and meta-analysis. *Ann Oncol*. doi: 10.1093/annonc/mdu149
- Kanai, Y., Ushijima, S., et al. (2001). DNA methyltransferase expression and DNA methylation of CPG islands and peri-centromeric satellite regions in human colorectal and stomach cancers. *Int J Cancer*, 91(2), 205-212.
- Kang, M. R., Kim, M. S., et al. (2009). Mutational analysis of IDH1 codon 132 in glioblastomas and other common cancers. *Int J Cancer*, 125(2), 353-355. doi: 10.1002/ijc.24379
- Killian, J. K., Kim, S. Y., et al. (2013). Succinate dehydrogenase mutation underlies global epigenomic divergence in gastrointestinal stromal tumor. *Cancer Discov*, 3(6), 648-657. doi: 10.1158/2159-8290.CD-13-0092
- Koo, B. K., & Clevers, H. (2014). Stem cells marked by the R-spondin receptor LGR5. *Gastroenterology*, 147(2), 289-302. doi: 10.1053/j.gastro.2014.05.007
- Kudo, Y., Tateishi, K., et al. (2012). Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. *Cancer Sci*, 103(4), 670-676. doi: 10.1111/j.1349-7006.2012.02213.x
- Leggett, B., & Whitehall, V. (2010). Role of the serrated pathway in colorectal cancer pathogenesis. *Gastroenterology*, 138(6), 2088-2100. doi: 10.1053/j.gastro.2009.12.066
- Letouze, E., Martinelli, C., et al. (2013). SDH mutations establish a hypermethylator phenotype in paraganglioma. *Cancer Cell*, 23(6), 739-752. doi: 10.1016/j.ccr.2013.04.018
- Liu, C., Liu, L., et al. (2013). Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. *PLoS One*, 8(5), e62828. doi: 10.1371/journal.pone.0062828
- Maiti, A., & Drohat, A. C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J Biol Chem*, 286(41), 35334-35338. doi: 10.1074/jbc.C111.284620
- Muraro, M. G., Mele, V., et al. (2012). CD133+, CD166+CD44+, and CD24+CD44+ phenotypes fail to reliably identify cell populations with cancer stem cell functional features in established human colorectal cancer cell lines. *Stem Cells Transl Med*, 1(8), 592-603. doi: 10.5966/sctm.2012-0003
- Nakajima, H., & Kunimoto, H. (2014). TET2 as an epigenetic master regulator for normal and malignant hematopoiesis. *Cancer Sci*, 105(9), 1093-1099. doi: 10.1111/cas.12484
- Nazemalhosseini Mojarad, E., Kuppen, P. J., et al. (2013). The CpG island methylator phenotype (CIMP) in colorectal cancer. *Gastroenterol Hepatol Bed Bench*, 6(3), 120-128.
- Noreen, F., Roosli, M., et al. (2014). Modulation of age- and cancer-associated DNA methylation change in the healthy colon by aspirin and lifestyle. *J Natl Cancer Inst*, 106(7). doi: 10.1093/jnci/dju161
- Nosho, K., Shima, K., et al. (2009). DNMT3B expression might contribute to CpG island methylator phenotype in colorectal cancer. *Clin Cancer Res*, 15(11), 3663-3671. doi: 10.1158/1078-0432.CCR-08-2383
- Noushmehr, H., Weisenberger, D. J., et al. (2010). Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. *Cancer Cell*, 17(5), 510-522. doi: 10.1016/j.ccr.2010.03.017
- Ocana, O. H., Corcoles, R., et al. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrx1. *Cancer Cell*, 22(6), 709-724. doi: 10.1016/j.ccr.2012.10.012

- Peng, D. F., Kanai, Y., et al. (2006). DNA methylation of multiple tumor-related genes in association with overexpression of DNA methyltransferase 1 (DNMT1) during multistage carcinogenesis of the pancreas. *Carcinogenesis*, 27(6), 1160-1168. doi: 10.1093/carcin/bgi361
- Shen, L., Toyota, M., et al. (2007). Integrated genetic and epigenetic analysis identifies three different subclasses of colon cancer. *Proc Natl Acad Sci U S A*, 104(47), 18654-18659. doi: 10.1073/pnas.0704652104
- Shen, L., Wu, H., et al. (2013). Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell*, 153(3), 692-706. doi: 10.1016/j.cell.2013.04.002
- Tahiliani, M., Koh, K. P., et al. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science*, 324(5929), 930-935. doi: 1170116 [pii] 10.1126/science.1170116
- Teodoridis, J. M., Hardie, C., et al. (2008). CpG island methylator phenotype (CIMP) in cancer: causes and implications. *Cancer Lett*, 268(2), 177-186. doi: 10.1016/j.canlet.2008.03.022
- Thiery, J. P., Acloque, H., et al. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell*, 139(5), 871-890. doi: 10.1016/j.cell.2009.11.007
- Ting, A. H., McGarvey, K. M., et al. (2006). The cancer epigenome--components and functional correlates. *Genes Dev*, 20(23), 3215-3231. doi: 10.1101/gad.1464906
- Toyota, M., Ahuja, N., et al. (1999). CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A*, 96(15), 8681-8686.
- Tsai, J. H., Donaher, J. L., et al. (2012). Spatiotemporal regulation of epithelial-mesenchymal transition is essential for squamous cell carcinoma metastasis. *Cancer Cell*, 22(6), 725-736. doi: 10.1016/j.ccr.2012.09.022
- Turcan, S., Rohle, D., et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature*, 483(7390), 479-483. doi: 10.1038/nature10866
- Van der Auwera, I., Yu, W., et al. (2010). Array-based DNA methylation profiling for breast cancer subtype discrimination. *PLoS One*, 5(9), e12616. doi: 10.1371/journal.pone.0012616
- van Roy, F., & Berx, G. (2008). The cell-cell adhesion molecule E-cadherin. *Cell Mol Life Sci*, 65(23), 3756-3788. doi: 10.1007/s00018-008-8281-1
- Weber, M., Davies, J. J., et al. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet*, 37(8), 853-862. doi: 10.1038/ng1598
- Weisenberger, D. J., Siegmund, K. D., et al. (2006). CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*, 38(7), 787-793. doi: 10.1038/ng1834
- Weisenberger, D. J., Velicescu, M., et al. (2004). Role of the DNA methyltransferase variant DNMT3b3 in DNA methylation. *Mol Cancer Res*, 2(1), 62-72.
- Xiao, M., Yang, H., et al. (2012). Inhibition of alpha-KG-dependent histone and DNA demethylases by fumarate and succinate that are accumulated in mutations of FH and SDH tumor suppressors. *Genes Dev*, 26(12), 1326-1338. doi: 10.1101/gad.191056.112
- Xu, J., Lamouille, S., et al. (2009). TGF-beta-induced epithelial to mesenchymal transition. *Cell Res*, 19(2), 156-172. doi: 10.1038/cr.2009.5
- Xu, W., Yang, H., et al. (2011). Oncometabolite 2-hydroxyglutarate is a competitive inhibitor of alpha-ketoglutarate-dependent dioxygenases. *Cancer Cell*, 19(1), 17-30. doi: 10.1016/j.ccr.2010.12.014
- Xu, Y., Hu, B., et al. (2012). Unique DNA methylome profiles in CpG island methylator phenotype colon cancers. *Genome Res*, 22(2), 283-291. doi: 10.1101/gr.122788.111
- Xu, Y., Wu, F., et al. (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. *Mol Cell*, 42(4), 451-464. doi: 10.1016/j.molcel.2011.04.005
- Yagi, K., Akagi, K., et al. (2010). Three DNA methylation epigenotypes in human colorectal cancer. *Clin Cancer Res*, 16(1), 21-33. doi: 10.1158/1078-0432.CCR-09-2006



Yang, Q., Wu, K., et al. (2013). Decreased 5-hydroxymethylcytosine (5-hmC) is an independent poor prognostic factor in gastric cancer patients. *J Biomed Nanotechnol*, 9(9), 1607-1616.

## Figure legends

**Figure 1: Correlation of TET expression with genomewide promoter methylation in CRC cell lines.** a) Illumina Infinium HumanMethylation27 data of eight human CRC cell lines. Shown is the unsupervised clustering based on the 25% of the probes with the highest methylation variability across the eight cancer cell lines (4945 CpG sites). Methylation M-values are represented in a colour scale from green (low methylation) to red (high methylation). Black bars indicate microsatellite instable (MSI) and CIMP-positive cell lines, respectively. CIMP status was determined by methylation analysis (In-qMSP) of the five marker panel. b) Linear correlation of log<sub>2</sub> converted TET1, TET2 and TET3 mRNA expression levels (relative to the averaged expression of ACTB and GAPDH) in eight CRC cell lines with methylation M-value means of the 3713 most variable CpG-island-associated Infinium methylation probes. Correlation coefficients (R) and p-values (p) are specified for each correlation. c) TET1 and  $\beta$ -Actin protein expression analysis by western blotting in eight human CRC cell lines. d) 5mC and 5hmC levels quantified by MeDIP and GLIP in the promoter regions of PROK2, NPPC, DLX5, LHX6, POU3F2, WNT10A, SLC30A, TRIM9, CDK5R2 in the CRC cell lines CACO2, SW620, HCT116 and SW48. 5mC and 5hmC levels are given as enrichment relative to ACTB (RE vs. ACTB). \*\*,  $p \leq 0.01$ ; n.s., not statistical significant (Wilcoxon matched pairs signed rank test)

**Figure 2: TET expression in CIMP and non-CIMP primary colorectal cancers.** a) mRNA expression (analysed by qRT-PCR) of TET1, TET2 and TET3 in CIMP (n=7, colored), non-CIMP (n=15) primary human colorectal cancers (CIMP status according to the five marker panel), 18 matched cancer-associated normal mucosa samples (CAM) and 8 colonic mucosa biopsies from healthy individuals (HM). Expression was normalized against expression of ACTB and GAPDH (mean). Boxplots indicate the first and the third quartile plus the median, Whiskers indicate the minimum and maximum values. b) mRNA expression (Agilent expression array G4502A) of TET1, TET2 and TET3 in CIMP colorectal cancers (n=25) compared to non-CIMP cancers (n=118) (Cancer Genome Atlas, 2012). Expression levels are given as log<sub>2</sub> converted fold-change relative to reference (log<sub>2</sub> FC). Boxplots indicate the first and the third quartile plus the median, Whiskers indicate the minimum and maximum values. \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$  (Mann-Whitney test). c) TET1 promoter methylation analysis by In-qMSP in 22 primary colorectal cancers, classified in CIMP and non-CIMP according to the methylation status of the five marker panel, and in 8 CRC cell lines. PMA, percentage

methylated alleles. d) 5-AzaC treatment of CRC cell line SW48. TET1 mRNA expression quantified by qRT-PCR (left panel) and promoter methylation analysed by In-MSP (right panel) in untreated cells and 5-AzaC treated cells. TET1 mRNA expression is relative to expression of GAPDH and ACTB (mean). PMA, percentage methylated alleles.

**Figure 3: Knockdown of TET1 in non-CIMP CRC cell line SW620.** a) TET1 knockdown (TET1-kd) efficiency determined by qRT-PCR and western blotting and TET2 and TET3 mRNA expression in SW620 TET1-kd cells. TET1, TET2 and TET3 mRNA expression levels (normalized to ACTB and GAPDH) in TET1-kd cell clones (shTET1-1 and shTET1-2) are depicted relative to control vector-transduced cells (sh-ctrl.). Error bars show the standard deviation from two cDNA replicates. b) Dot Blot analysis of global 5hmC levels in SW620 sh-ctrl and knockdown clones shTET1-1 and shTET1-2. Left panel: Chemiluminescence signals of 5hmC-antibody-incubated membrane. Loaded amount of genomic DNA is indicated at the top. Right panel: Loading control by Bromophenol blue staining. c) Venn diagram illustrating number of commonly hyper- and hypo-methylated Infinium HM450 array probes in shTET1-1 and shTET1-2 cells compared to sh-ctrl (clear circles) as well as array probes commonly hyper- or hypo-methylated in shTET1-1 and shTET1-2 derived xenografts compared to sh-ctrl derived xenografts (grey circles). Circle size is proportional to number of probes. Overlaps indicate commonly hyper- or hypo-methylated probes in cells and xenografts. These overlaps were statistical significant (each with  $p < 0.0001$ , Chi square). The increase in the fraction of hypermethylated probes from cell lines to xenografts is statistical significant,  $p < 0.0001$  (Chi-square). d) TET1 promoter methylation levels in shTET1-1 and shTET1-2 (relative to sh-ctrl.) analysed by pyrosequencing of bisulfite-converted genomic DNA (from position +603 to +728 relative to transcription start site). Each bar represents a single CpG site within the TET1 promoter CGI ordered in 3'-5' direction. PMA, percentage methylated alleles.

**Figure 4: TET1 depletion results in the loss of cancer-initiating cell markers.** a) LGR5 mRNA expression in SW620 sh-ctrl and TET1-kd clones shTET1-1 and shTET1-2 analysed by qRT-PCR. LGR5 expression is given relative to the expression of GAPDH and ACTB (mean). Error bars indicate the standard deviation from three RNA preparations.\*\*,  $p \leq 0.01$ ; \*\*\*,  $p \leq 0.005$ , t-test. b) The fraction of LGR-, CD44- and CD166-positive cells in mouse xenografts derived from sh-ctrl and TET1-kd clones shTET1-1 and shTET1-2 determined by flow-cytometry. Fraction of marker-positive cells is represented as the mean percentage of

viable cells. Error bars indicate the standard deviation from four xenografts. \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$ ; \*\*\*,  $p \leq 0.005$ , t-test

**Figure 5: TET1 depletion decreases cell state plasticity.** a) Dot blots showing viable CD24<sup>hi</sup>CD44<sup>hi</sup> and CD24<sup>neg</sup>CD44<sup>neg</sup> cells in sh-ctrl and the TET1-kd clones shTET1-1 and shTET1-2 analysed by flow-cytometry. Dot plots indicate the gates used for quantitation. b) Dot plots illustrating the cell population distribution of sh-ctrl, shTET1-1 and shTET1-2 cells after sorting and expansion of CD24<sup>hi</sup>CD44<sup>hi</sup> (CD24<sup>+</sup>/CD44<sup>+</sup>) and CD24<sup>neg</sup>CD44<sup>neg</sup> (CD24<sup>-</sup>/CD44<sup>-</sup>) subpopulations. Staining for CD24 and CD44 was analysed by flow-cytometry. Bar graphs indicate the fraction of CD24<sup>+</sup>/CD44<sup>+</sup> and CD24<sup>-</sup>/CD44<sup>-</sup> in sh-ctrl, shTET1-1 and shTET1-2 cells, before sorting (US, unsorted), after sorting ( $t_0$ ) and 8 days after expansion in culture ( $t_8$ ), normalized to the total number of cells comprised in the gates depicted in a). Error bars indicate the range of values from two experiments.

**Figure 6: TET1 depletion induces EMT-like process.** a) Light microscopy pictures of cell colonies from sh-ctrl and TET1 knockdown clones shTET1-1 and shTET1-2 scale bar = 200 $\mu$ m). b) CDH1 (E-cadherin) mRNA expression (left panel) and promoter methylation (right panel) in sh-ctrl cells and TET1 knockdown clones shTET1-1 and shTET1-2 analysed by qRT-PCR and pyrosequencing of bisulfite converted genomic DNA, respectively. CDH1 expression is given relative to the expression of ACTB and GAPDH (mean). Error bars indicate the standard deviation from three RNA preparations. \*,  $p \leq 0.05$ , t-test. Blue bars indicate the median methylation value. PMA, percentage methylated alleles; \*,  $p \leq 0.05$ , Wilcoxon matched-pairs test. c) CDH1 mRNA expression data from TCGA (Cancer Genome Atlas, 2012). Expression levels are shown as log<sub>2</sub> converted fold change relative to reference (log<sub>2</sub> FC). Boxplots indicate the first and the third quartile plus the median, Whiskers indicate the minimum and maximum values. \*\*\*\*,  $p < 0.0001$ , Mann-Whitney test. d) FN1 (Fibronectin) and VIM (Vimentin) mRNA expression in sh-ctrl and the TET1-kd clones shTET1-1 and shTET1-2 analysed by qRT-PCR. Expression levels are normalized to the expression of GAPDH and ACTB (mean). Error bars indicate the standard deviation from three RNA preparations. \*,  $p \leq 0.05$ ; \*\*\*,  $p \leq 0.005$ , t-test. e) Cell migration assay with SW620 sh-ctrl cells and TET1 kd clones shTET1-1 and shTET1-2. Illustrated is the mean fluorescence signal from five experiments. Error bars indicate the standard deviation. \*\*,  $p \leq 0.01$ , t-test.

Figure1

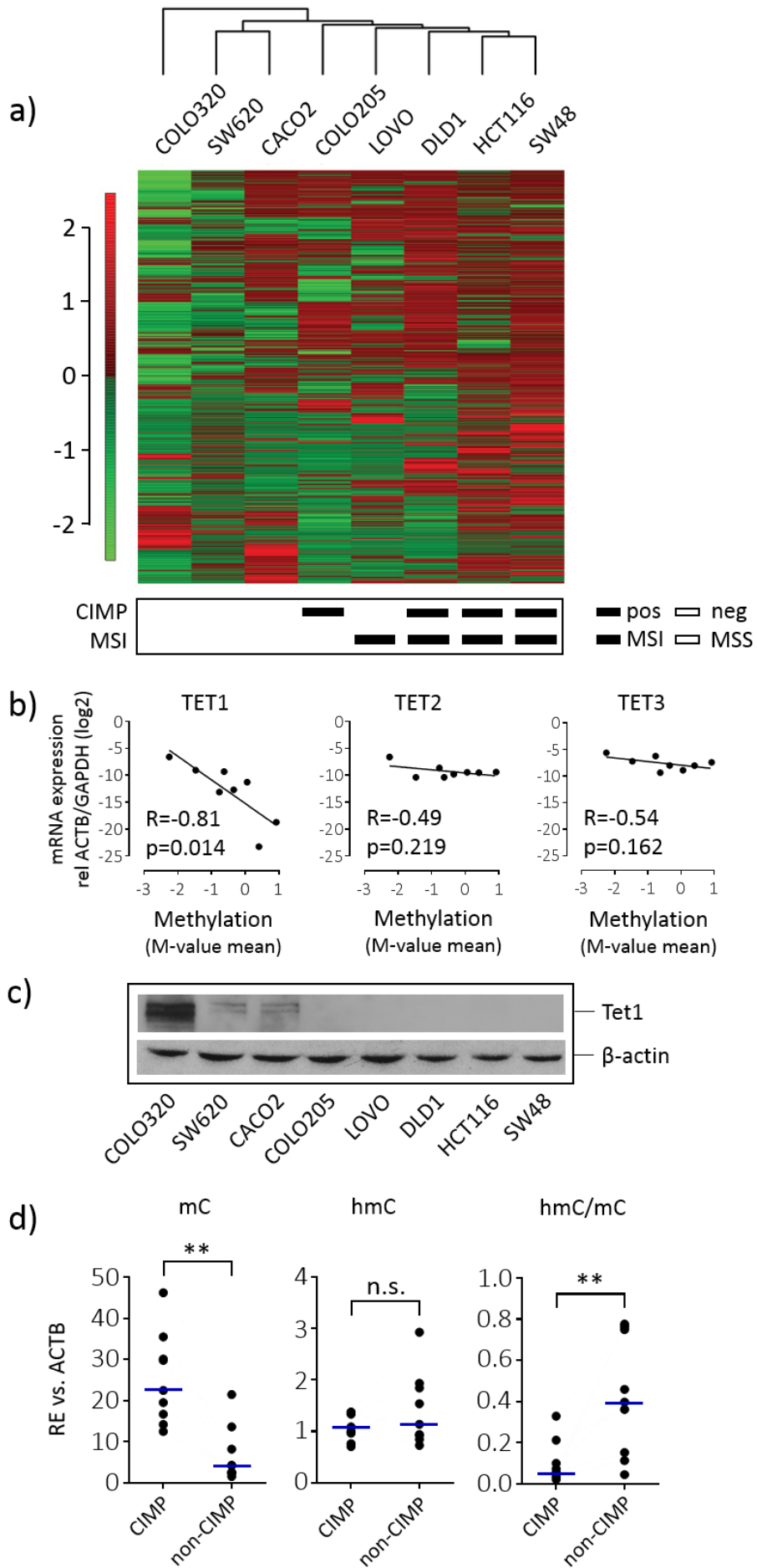


Figure2

Primary Tumors

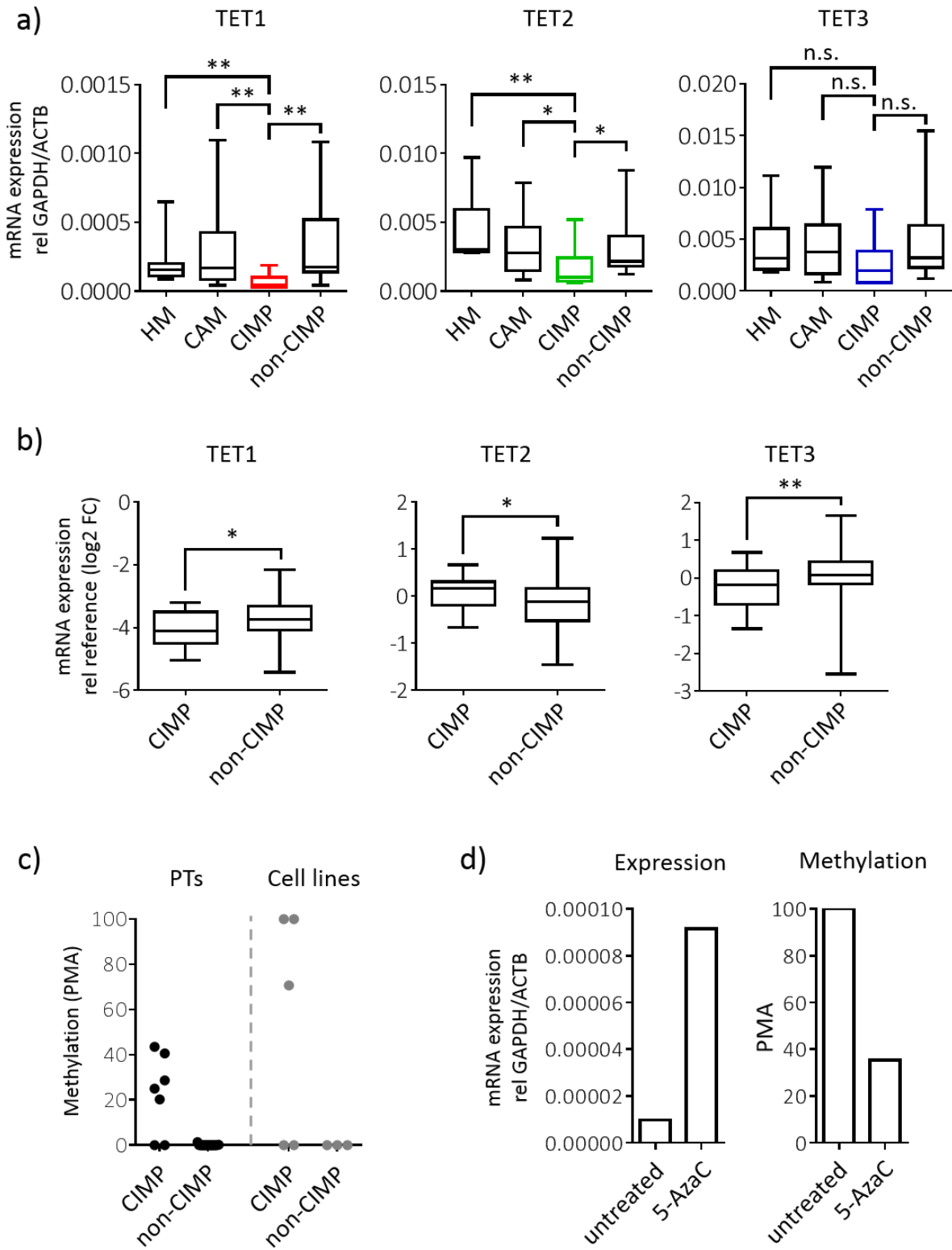


Figure3

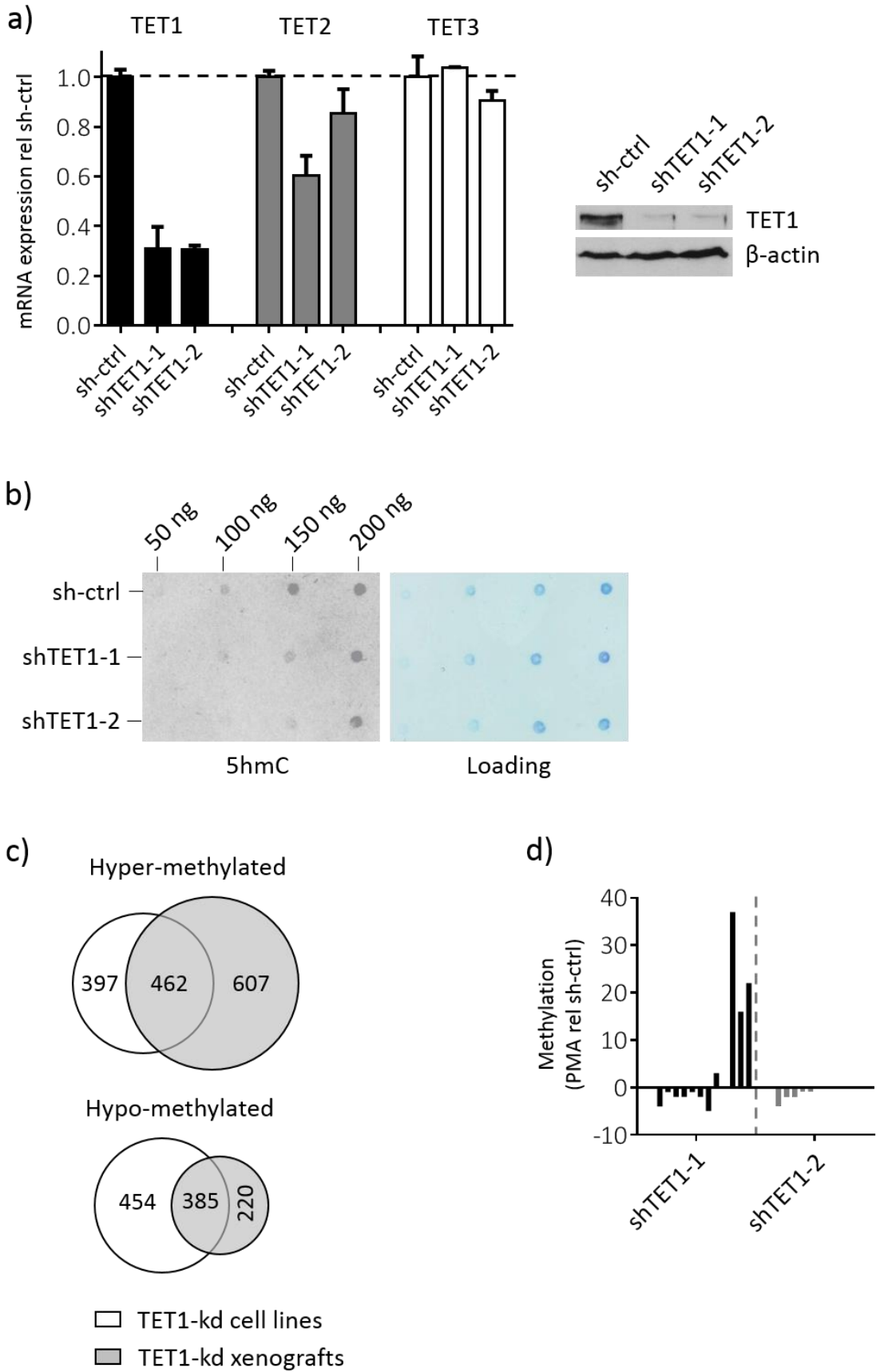


Figure4

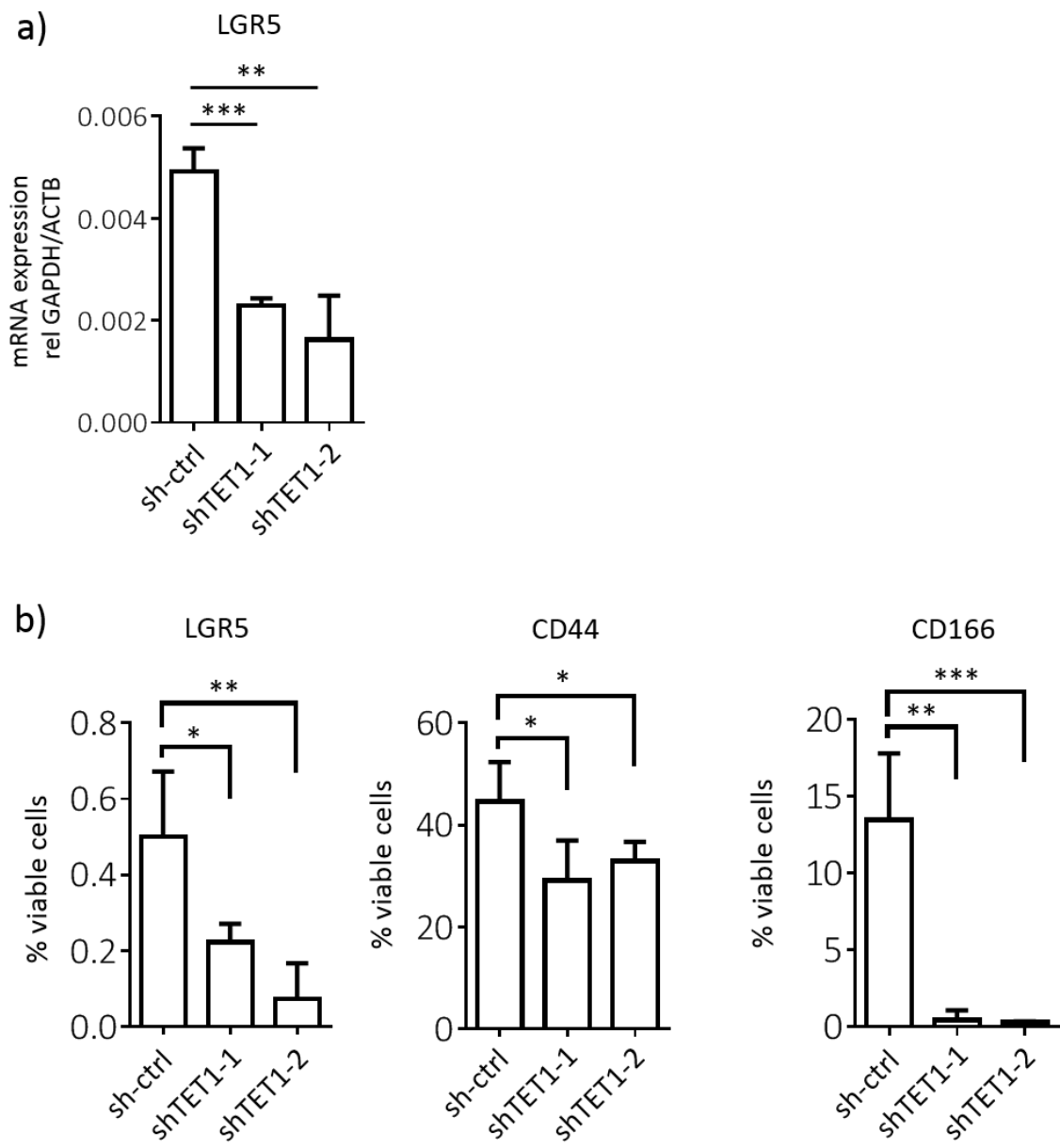
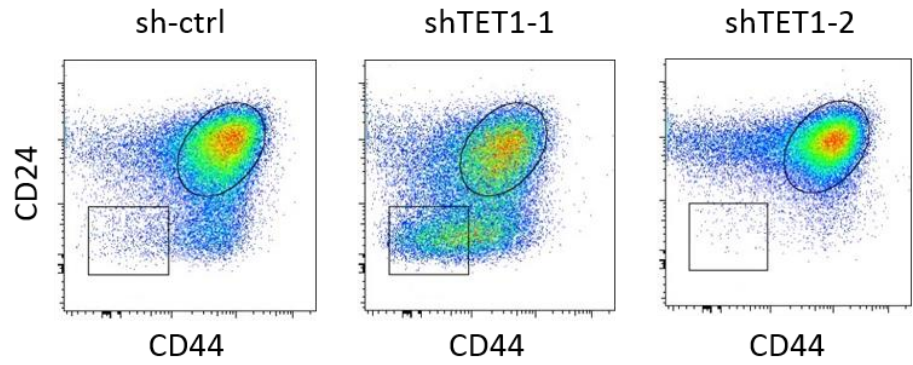




Figure5 a)



b)

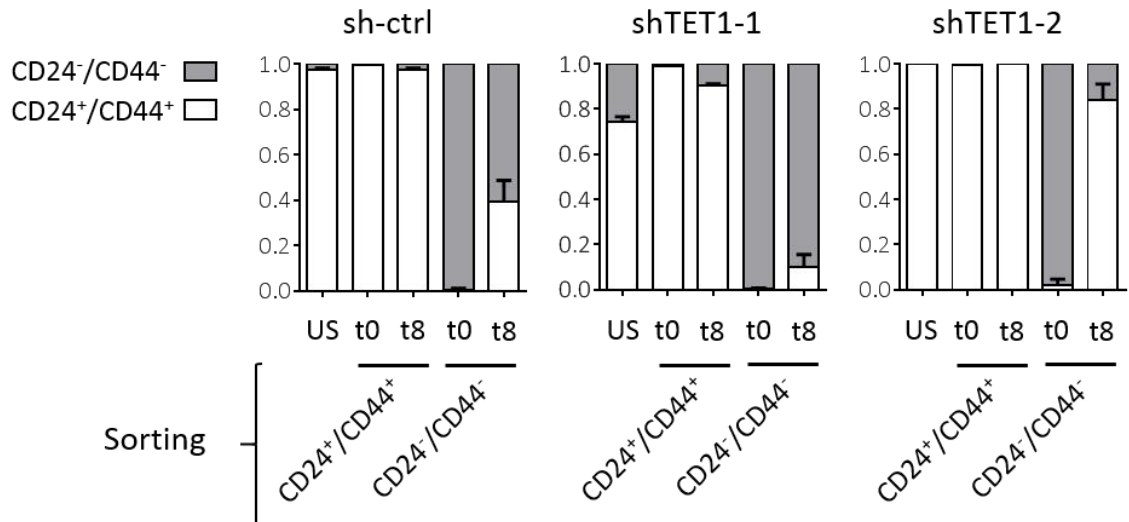
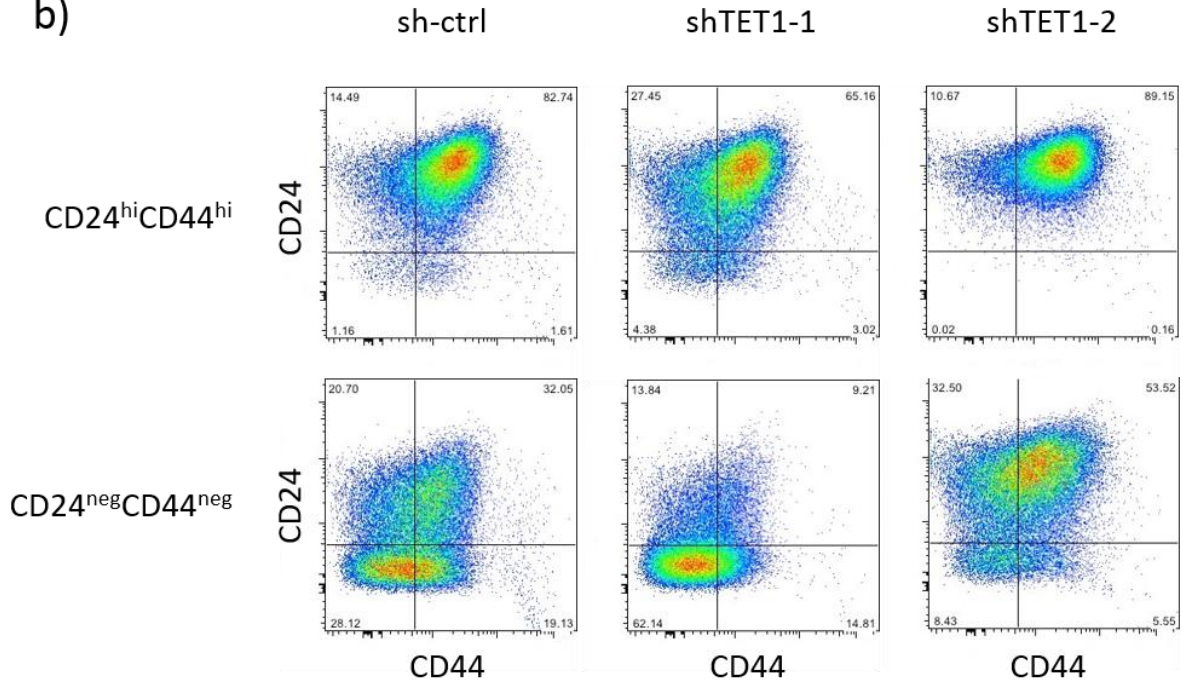
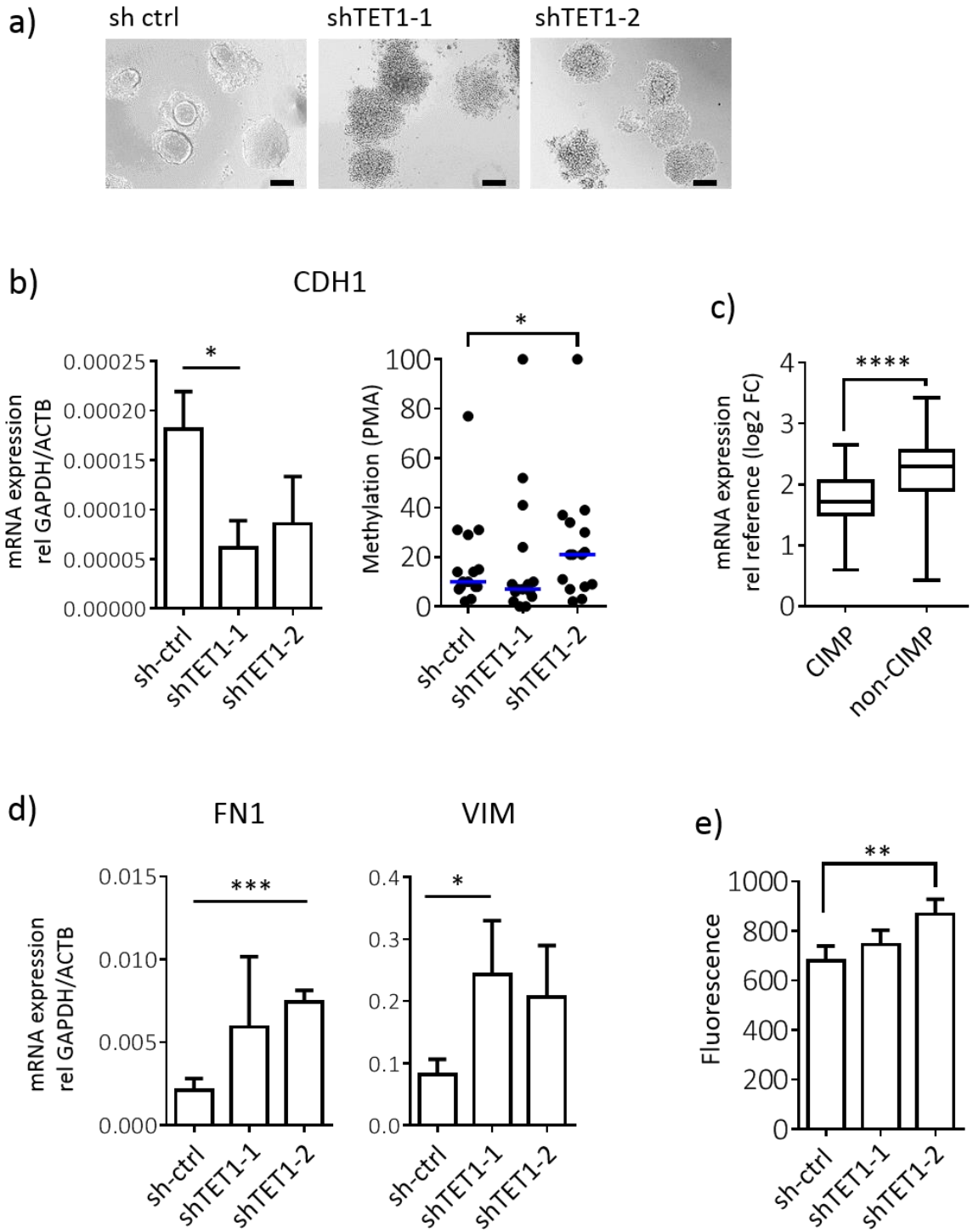


Figure6



## Supplementary Material

Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers

Weis, S.<sup>1</sup>, Muraro, M.G.<sup>1</sup>, Noreen, F.<sup>1</sup>, Mele, V.<sup>1</sup>, Wirz, A.<sup>1</sup>, Truninger, K.<sup>2</sup>, Schär, P.<sup>1</sup>

<sup>1</sup> Department of Biomedicine, University of Basel, Switzerland; <sup>2</sup> FMH Gastroenterology and Internal Medicine, Langenthal, Switzerland

Correspondence: primo.schaer@unibas.ch; Tel. +41 61 267 3561; Fax: +41 61 267 3566

## Supplementary Figure Legends

**Figure S1: Correlation of DNMTs and TDG with genomewide methylation and TETs with global hydroxymethylation in CRC cell lines.** Linear correlation of log<sub>2</sub> converted TET1, TET2 and TET3 mRNA expression levels in eight CRC cell lines with log<sub>2</sub> converted global 5hmC levels analysed by LC-MS/MS. mRNA expression values were determined by qRT-PCR relative to the combined expression of GAPDH and ACTB. Correlation coefficients (R) and p-values (p) are specified for each correlation.

**Figure S2: DNMT and TDG expression in CIMP and non-CIMP primary CRCs.** mRNA expression (Agilent expression array G4502A) of DNMT1, DNMT3A, DNMT3B, TDG, IDH1 and IDH2 in colorectal cancers from the CIMP-H cluster (n=25) compared to non-CIMP-H clusters (CIMP-L, 3, 4; n=118) (Cancer Genome Atlas, 2012). Expression levels are given as log<sub>2</sub> converted fold-change relative to reference (log<sub>2</sub> FC). Boxplots indicate the first and the third quartile plus the median, Whiskers indicate the minimum and maximum values. \*, p ≤ 0.05; \*\*, p ≤ 0.01; \*\*\*, p ≤ 0.005; n.s., not statistical significant (Mann-Whitney test).

**Figure S3: Global 5hmC levels and genome-wide methylation in TET1 knockdown cells**  
a) Specificity test for hmC-antibody. Shown are the chemiluminescence signals of the 5hmC-antibody incubated membrane. Loaded amounts of hmC- and mC-containing control oligomers are indicated at the top. b) Global 5mC and 5hmC levels in TET1 knockdown cells quantified by LC-MS/MS and shown as number of modified cytosines in one million nucleotides. Depicted are two sh-controls (sh-ctrl) and four monoclonal TET1 knockdown cell lines (shTET1). Blue bars indicate the median value. c) TET1 mRNA expression in mouse xenografts derived from sh-ctrl and TET1-kd clones shTET1-1 and shTET1-2. Shown is the mean expression value from four xenografts together with the standard deviation. Expression was determined relative to the expression of GAPDH and ACTB (mean). d) Volcano blots depicting differential methylation of TET1-kd cells and xenografts shTET1-1 (shTET1-xeno1) and shTET1-2 (shTET1-xeno2) compared to sh-ctrl cells and xenografts analysed using the Illumina Infinium 450k methylation platform. Plotted are the log<sub>2</sub> fold changes (FC) in methylation with the negative log<sub>10</sub> of false discovery rate (FDR)-adjusted p-values. Numbers indicate the fraction of significantly (log<sub>2</sub>FC>2, p-value<0.0001) differently methylated CpG methylation probes.

**Figure S4: CIMP five marker panel in TET1-kd cells** a) Methylation levels of CIMP five-marker panel (Weisenberger et al. 2006) quantified by In-qMSP in sh-ctrl and TET1-kd clones shTET1-1 and shTET1-2. Shown is the mean value plus standard deviation of two PCR replicates. PMA, percentage methylated alleles. b) Agarose gel electrophoresis of In-qMSP products (after 45 PCR cycles) from methylation quantitation of the five-marker panel. Genomic DNA from human peripheral blood treated with the CpG-methyltransferase M.SssI served as positive control (Pb SSSI). c) Left panel: Promoter methylation level of MLH1 analysed by pyrosequencing of bisulfite-converted genomic DNA from sh-ctrl and the TET1-kd clones shTET1-1 and shTET1-2. Each dot represents a single CpG site, blue bars indicate the median methylation value. \*,  $p \leq 0.05$ ; \*\*,  $p \leq 0.01$  (Wilcoxon matched-pairs test). Right panel: MLH1 mRNA expression in sh-ctrl and TET1-kd clones shTET1-1 and shTET1-2 analysed by qRT-PCR. Error bars indicate the standard deviation from three RNA preparations. Expression values are shown relative to the expression of GAPDH and ACTB (mean). \*,  $p \leq 0.05$  (t-test).

**Figure S5: EMT marker expression in TET1-kd cells.** SNAI1, SNAI2 and ZEB1 mRNA expression quantified by qRT-PCR in sh-ctrl and the TET1-kd clones shTET1-1 and shTET1-2. Expression levels are shown relative to the expression levels of GAPDH.

**Figure S6: TET1 Western Blots.** a) TET1 Western Blot with whole cell extracts from human CRC cell lines. Chemiluminescence signals of anti-Tet1 antibody ( $\alpha$ -TET1, left panel) and anti-ACTB antibody ( $\alpha$ -ACTB) incubated membranes. Molecular weights are shown in kilo Daltons (kDa) on the left. b) TET1 Western Blot with whole cell extracts from monoclonal SW620 TET1-kd cell lines (clone1 and clone2) and sh-control cells. Chemiluminescence signals of anti-Tet1 antibody ( $\alpha$ -TET1, left panel) and anti-ACTB antibody ( $\alpha$ -ACTB) incubated membranes. Molecular weights are shown in kilo Daltons (kDa) on the left. c) TET1 Western Blot with whole cell extracts from CACO2 TET1-kd cell lines (vec1 and vec2) and sh-control cells. Chemiluminescence signals of anti-Tet1 antibody ( $\alpha$ -TET1, left panel) and anti-ACTB antibody ( $\alpha$ -ACTB) incubated membranes. Molecular weights are shown in kilo Daltons (kDa) on the left.

Figure S1

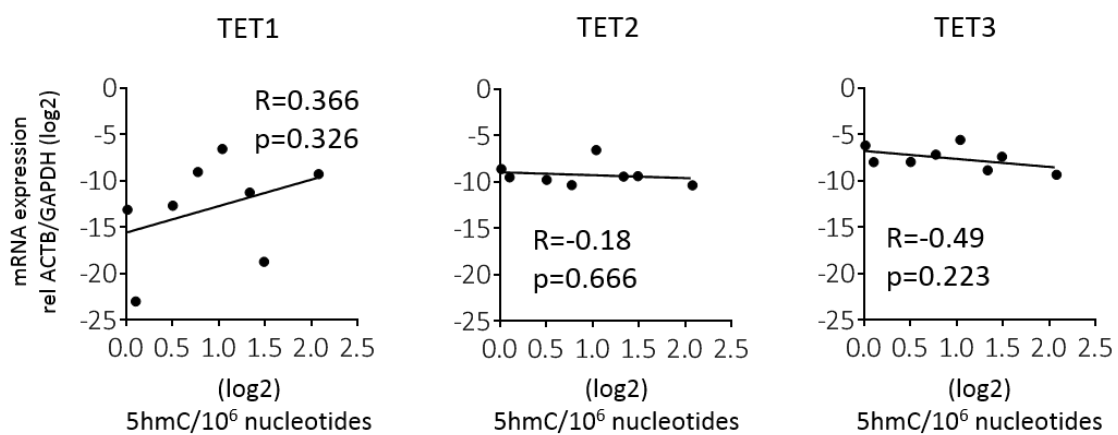


Figure S2

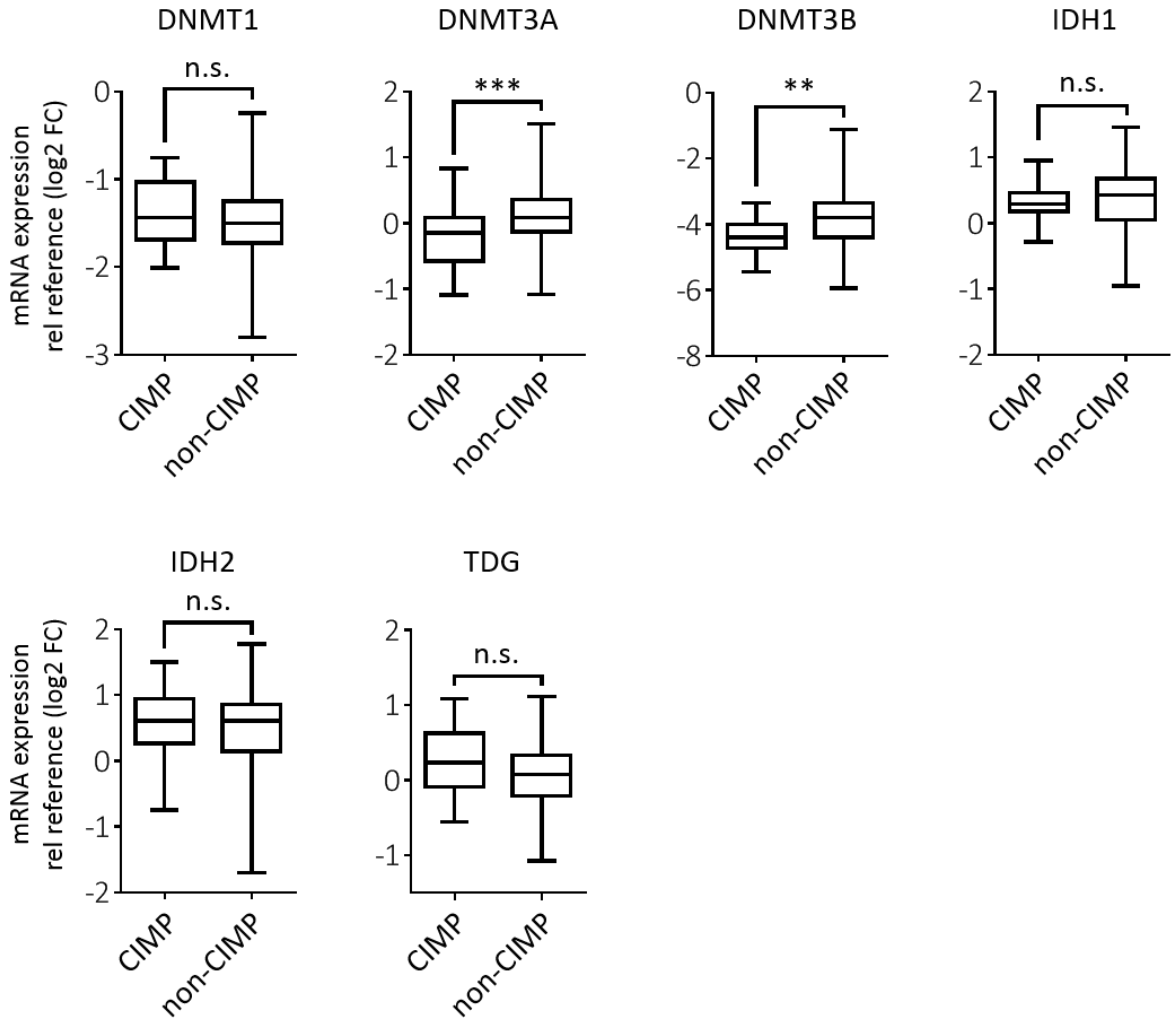


Figure S3

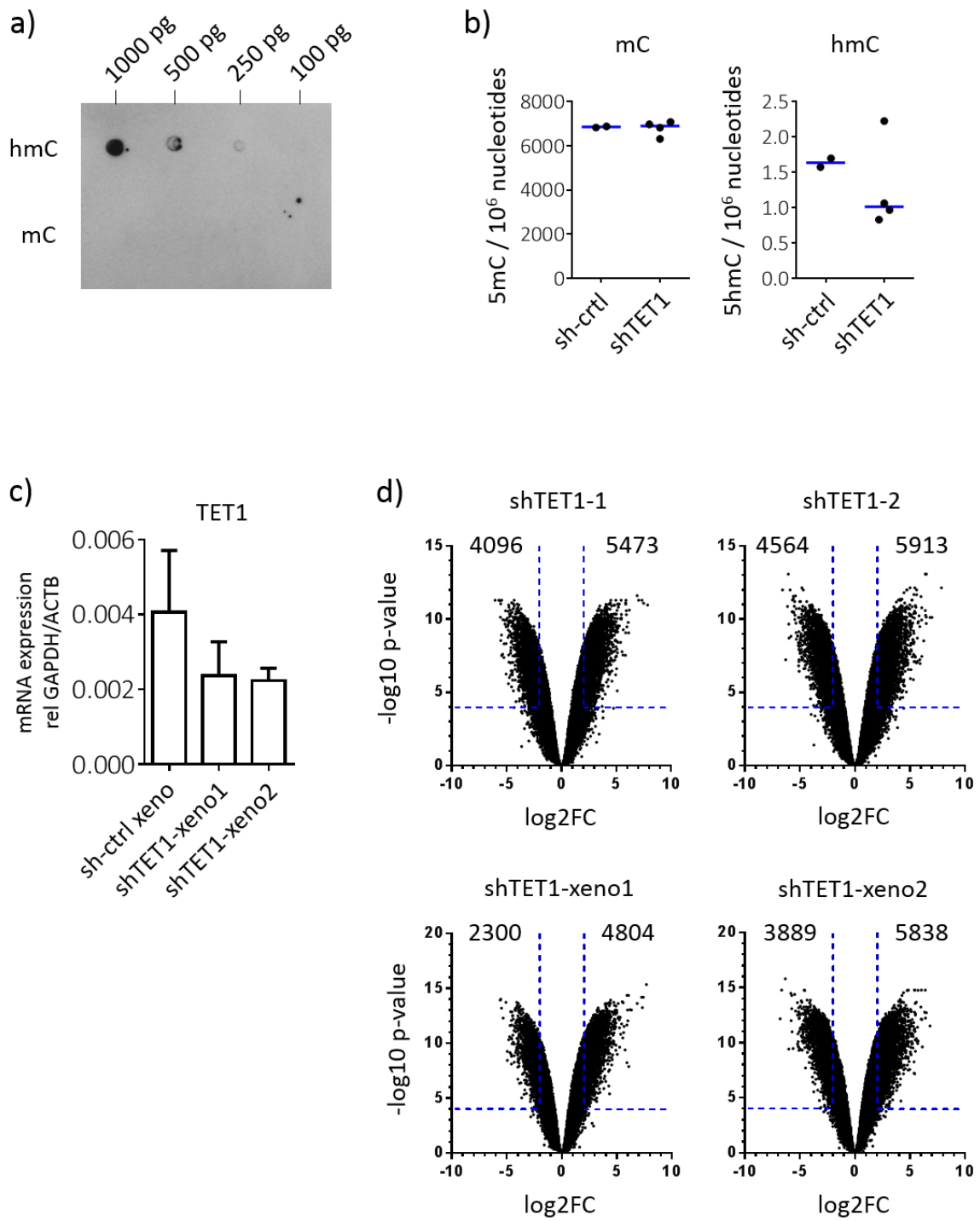




Figure S4

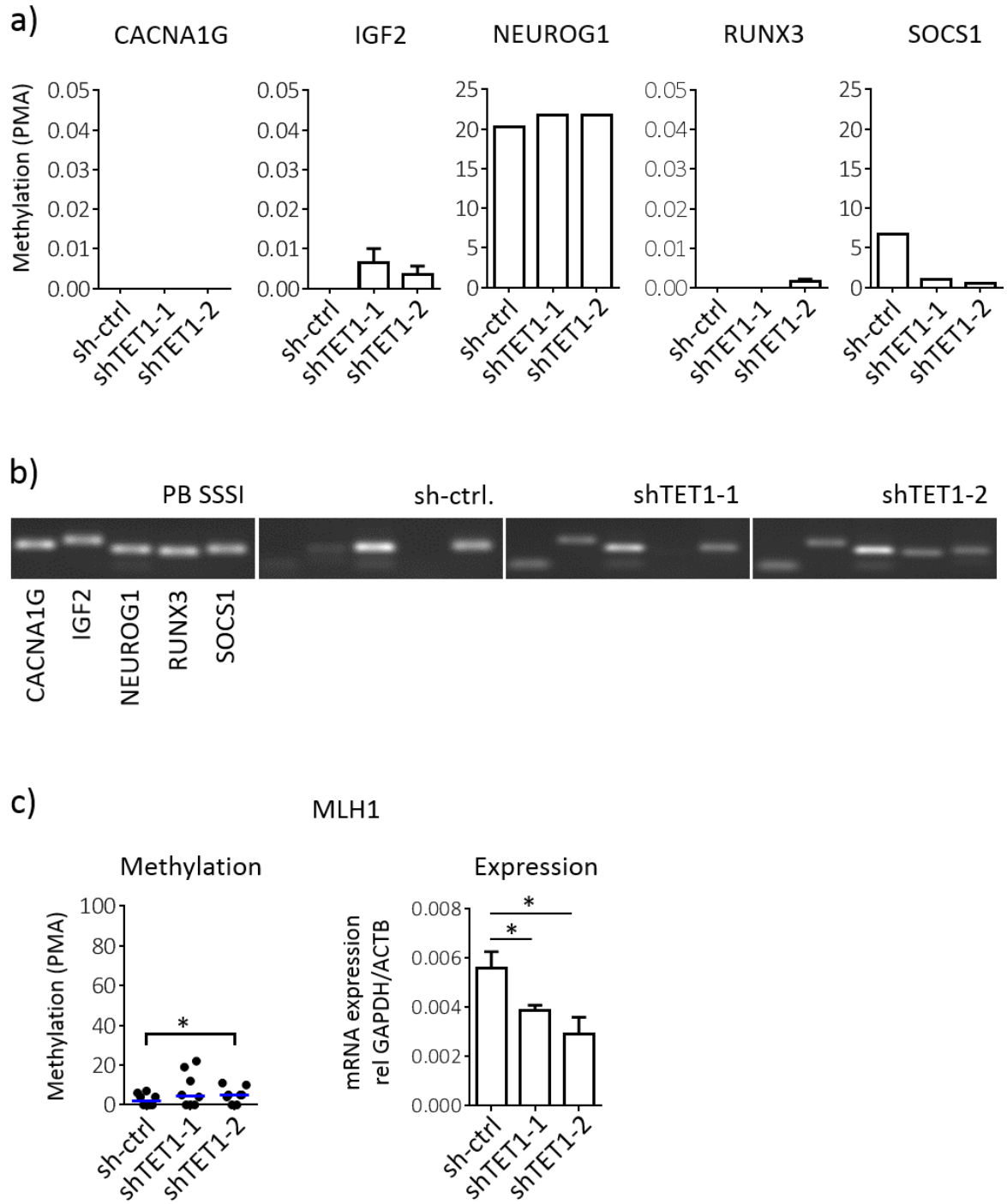


Figure S5

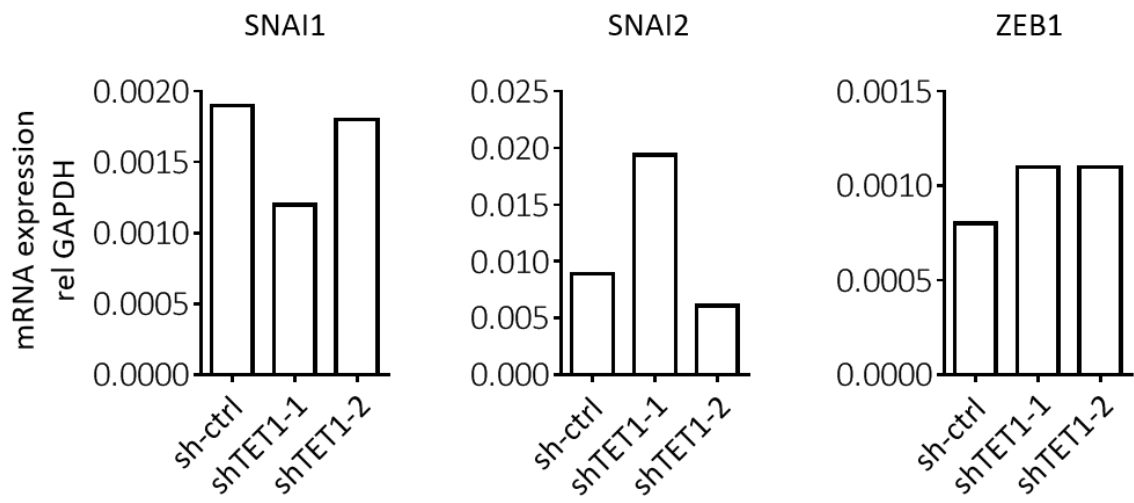
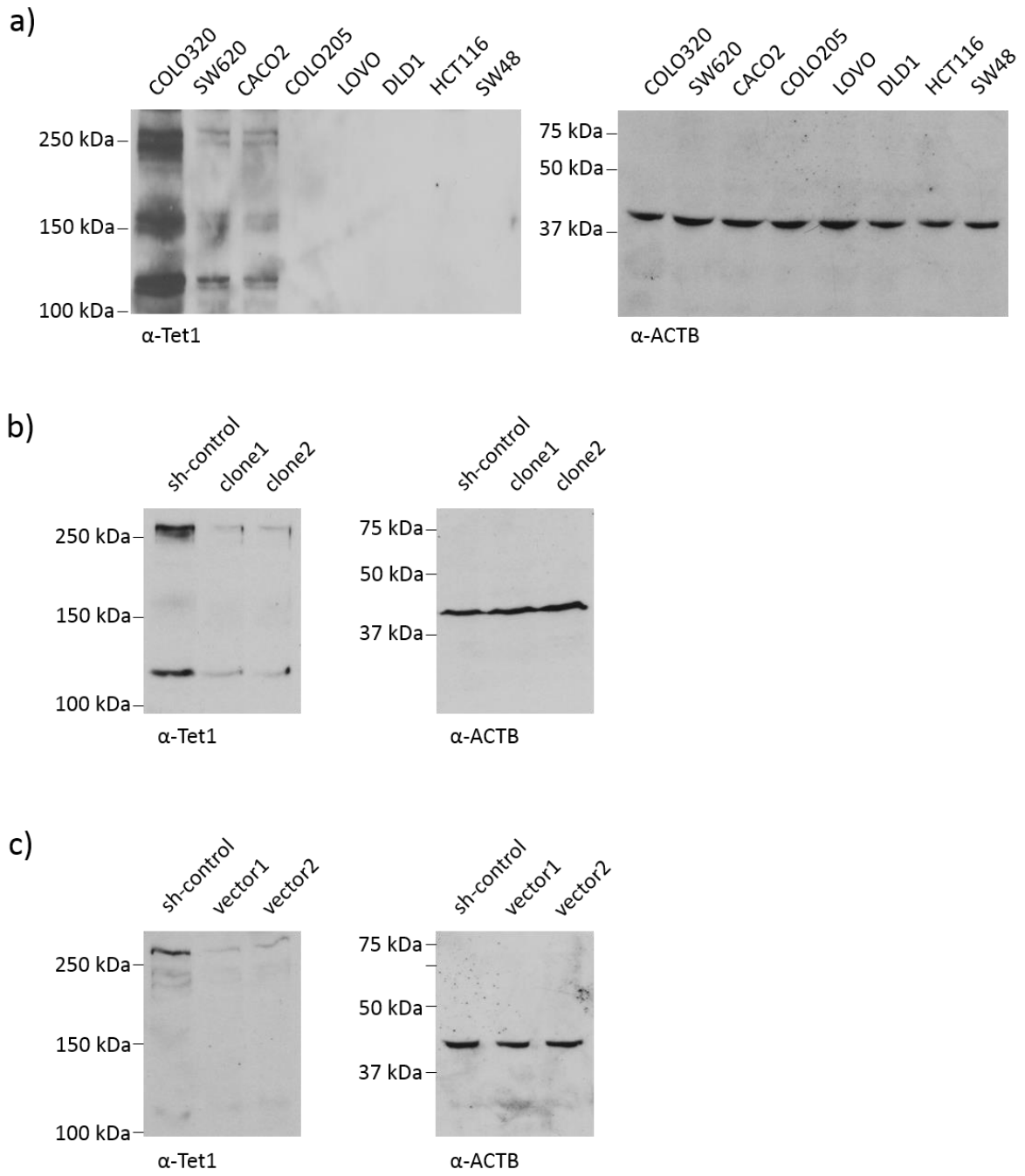


Figure S6



Supplementary Table 1: Clinical data of CRC patients

Patient #	Gender	Age	Tumor location	Colonic site	Tumor stage	CIMP	CAM
1	f	65	Caecum	Right	T3 N0 M1	negative	no
2	f	67	Rectosigmoid	Left	T2 N0 MX	negative	no
3	f	75	Rectum	Left	T3 N2 M1	negative	yes
4	m	61	Rectum	Left	T3 N1 M0	negative	yes
5	m	53	C. sigmoideum	Left	T4 N1 M1	negative	yes
6	f	46	Rectum	Left	T3 N2 M0	negative	yes
7	m	76	C. ascendens	Right	T4 N1 M1	negative	yes
8	m	75	C. sigmoideum	Left	T3 N0 M0	negative	yes
9	f	66	C. sigmoideum	Left	T3 N1 M1	negative	yes
10	m	82	C. sigmoideum	Left	T3 N0 MX	negative	yes
11	m	67	Rectum	Left	T2 N0 M0	negative	yes
12	f	85	C. transversum	Right	T4 N1 M0	negative	no
13	f	57	Rectum	Left	T1 N0 MX	negative	yes
14	m	69	Rectum	Left	T3 N0 M0	negative	yes
15	m	81	Rectosigmoid	Left	T4 N2 M1	negative	yes
16	m	65	C. ascendens	Right	T3 N2 MX	positive	yes
17	m	70	C. sigmoideum	Left	T3 N0 M0	positive	yes
18	m	74	C. ascendens	Right	T2 N0 M1	positive	yes
19	f	82	C. ascendens	Right	T3 N0 M0	positive	yes
20	f	85	n/a	n/a	T3 N0 M0	positive	no
21	f	80	C. ascendens	Right	T3 N0 MX	positive	no
22	m	73	C. ascendens	Right	T3 N0 M0	positive	yes
23	f	70	C. ascendens	Right	T3 N2 M0	positive	yes

Supplementary Table 2: Cell lines used in this study

Name	Species	Source	Standard medium
HEK293T	human	Embryonic kidney	DMEM + 10% FCS + 2mM L-Glutamin
CACO2	human	Colon adenocarcinoma	EMEM + 20% FCS
CO115	human	Colon adenocarcinoma	DMEM + 10% FCS + 2mM L-Glut
COLO205	human	Colon adenocarcinoma	RPMI 1640 + 10% FCS + 2mM L-Glut
COLO320	human	Colon adenocarcinoma	RPMI 1640 + 10% FCS
COLO741	human	Colon adenocarcinoma	RPMI 1640 + 10% FCS + 2mM L-Glut
DLD-1	human	Colon adenocarcinoma	RPMI 1640 + 10% FCS
GP5-d	human	Colon adenocarcinoma	DMEM + 10% FCS + 2mM L-Glutamin
HCT116	human	Colon	McCoy 5a + 10% FCS
LOVO	human	Colon adenocarcinoma	F-12 + 10% FCS + 2mM L-Glut
SW48	human	Colon adenocarcinoma	DMEM + 10% FCS + 2mM L-Glut
SW620	human	Colon adenocarcinoma	DMEM + 10% FCS + 2mM L-Glut

Supplementary Table 3: Primers used for In-qMSP

Name	Orientation	Sequence (5' - 3')	Length
MLH1 qMSP M-primer	F	ACGCAGACGCTCCACCAGGGCCGC	24
MLH1 qMSP M-primer	R	TCCGGAGGGCGATGGGGCCCTGTGC	25
MLH1 qMSP S-primer	F	AGTTTGGAGTGGTAGGTTTTAGAGG	26
MLH1 qMSP S-primer	R	CATCCCCAAATTCTTACTCCTTCTTTC	27
TET1 qMSP M-primer	F	TATTCGGGACGTTCCGGAC	20
TET1 qMSP M-primer	R	CGCTATAACGAAACCTCCGAAACG	24
TET1 qMSP S-primer	F	TGTAGAGGGAATTTGAGTTAGGTTTGG	27
TET1 qMSP S-primer	R	CCCCATTTTCTTCCCCAAATAACAAC	26
NEUROG1 qMSP M-primer	F	GGGCGTATTATTGTTGCGTC	20
NEUROG1 qMSP M-primer	R	CGATCTCCTAAATAATATCGCCGACG	26
NEUROG1 qMSP S-primer	F	GATGTTTGGGTTTTTGTGGTTTTTGG	26
NEUROG1 qMSP S-primer	R	CCTATTCCAACAAATCTTCCCTCTACATATC	31
CACNA1G qMSP M-primer	F	GATTCGGTTTTAGTTGCGTCGC	22
CACNA1G qMSP M-primer	R	AATTCGACGCTACCTTCGACG	21
CACNA1G qMSP S-primer	F	GGTTGTGAGGGAGGGGTAAAG	21
CACNA1G qMSP S-primer	R	ACCTCAATTTCTCAACAAAACACTAAACC	30
RUNX3 qMSP M-primer	F	GATTACGTAGGCGAGTTCGTGC	22
RUNX3 qMSP M-primer	R	CTCACCTTAAAACGACGAACAACG	25
RUNX3 qMSP S-primer	F	GATTGTGGTTTTGGGAGATGTAAGAGG	27
RUNX3 qMSP S-primer	R	TCCTCCTCAAATCACTTCTTCTACCTC	25
SOCS1 qMSP M-primer	F	CGTGGTTATCGTGGGTCGC	19
SOCS1 qMSP M-primer	R	GTACACGACGAACGCTACCG	20
SOCS1 qMSP S-primer	F	TATTAGGGGGTGAGGTGAGGTTTG	24
SOCS1 qMSP S-primer	R	CTACCCCAACCCAAAACCACTAC	23
IGF2 qMSP M-primer	F	GTAGTTCGGTTCGTAGGTCGTC	22
IGF2 qMSP M-primer	R	ACGATAACGACACCGAAACCG	21
IGF2 qMSP S-primer	F	GGGAGGTTGTAGAAGGTTGGTG	22
IGF2 qMSP S-primer	R	ATCTACATCAATCACCCCATATCCC	26

Supplementary Table 4: Primers used for Pyrosequencing

Name	Orientation	Sequence (5' - 3')	Length
MLH1 Pyro amp	F	TTTTGTGATATTTGGAGATAAG	22
MLH1 Pyro amp	R	BIO-ATAAAACCCTATACCTAATCTAT	23
MLH1 Pyro seq 1		GTAATATTTTTATGTATTGGTATATAAAG	29
MLH1 Pyro seq 2		TAGGAGTGAAGGAGGT	16
CDH1 Pyro amp	F	GAATTGTAAAGTATTTGTGAGTTTG	25
CDH1 Pyro amp	R	GCCCCGCCCCCCCACTCCCATCACTAA	28
CDH1 Pyro seq 1		GTTGTTGTTGTAGGTAT	17
CDH1 Pyro seq 2		TTTTTTTTTTTTAAGAAAG	19
Pyro Biotin universal	F	BIO-GCCCCGCCCC	11

Target sequences for the Biotin-labeled universal primer are marked in red

Supplementary Table 5: Primers used for MeDIP and GLIP

Name	Orientation	Sequence (5' - 3')	Length
DLX5	F	GGAGGAAGAGGAGGAGGAGA	20
DLX5	R	GGCCAATAGAACCAGATCCA	20
NPPC	F	GAGGGAGAGCAGCGTGAG	18
NPPC	R	TATAAAGGCGCGAGCAGAG	19
PROK2	F	CCTCTAGCCTGCCCTTCAG	19
PROK2	R	CCCACTCCTGCTCCTCTTG	19
LHX6	F	GCGTTCTCATGCTTCCAGTA	20
LHX6	R	AGCCTCTGCTCCCACTGC	18
POU3F2	F	GCAGCGTCTAACCCTACAGC	21
POU3F2	R	GGGTGTCCGTTGCTCTGC	18
WNT10A	F	AAAGGGAAGGCGGCAGAG	18
WNT10A	R	GGTGCCTCTCAAGGGTAAGG	20
SLC30A	F	TCTGTCTACCTCTTCAAACG	21
SLC30A	R	CAGTTCGCTCCACTCTGGT	19
TRIM9	F	CCTACCTGCCCTCCTAGTCC	20
TRIM9	R	CGAAGACTGCCCTATCCTCA	20
CDK5R2	F	CGCAGCTCAGGATTAGAGCA	20
CDK5R2	R	CAGGCGAAAGAGACAGCAC	19

Supplementary Table 6: Primers used for qRT-PCR

Name	Orientation	Sequence (5' - 3')	Length
TET1 RT-primer	F	GTTGTTGTGCCTCTGGAGTTATAAAGG	28
TET1 RT-primer	R	TCTGACTTGGGGCCATTTACTGG	23
TET2 RT-primer	F	GAAGAGCAGTAAGGGACTGAGC	22
TET2 RT-primer	R	GGACTTAGTCTGTTGCCCTCAAC	23
TET3 RT-primer	F	CCCACCTGCGATTGCGTC	18
TET3 RT-primer	R	CTCTCCATACCGCTCCTCCATG	22
DNMT1 RT-primer	F	CCAGAGAACGAGTTGCTAGACC	22
DNMT1 RT-primer	R	TTCTTTGGTTTGACTTCGGAGTCTC	25
DNMT3A RT-primer	F	CAGTGGAAAATGGCTGCTGC	20
DNMT3A RT-primer	R	AGCCCTCCATTTTCATGGATTCCG	23
DNMT3B RT-primer	F	GCTACACACAGGACTTGACAGG	22
DNMT3B RT-primer	R	GGCTGGAGACACTGTTGTTATTTCCG	25
TDG RT-primer	F	CGGAAGTGCAGCACAGAAAG	20
TDG RT-primer	R	CTGAAGGGAATAGCTGCCCG	20
LGR5 RT-primer	F	GCTGCCTTCCAACCTCAGC	19
LGR5 RT-primer	R	CCGCAAGACGTAACCTCCTCCAG	22
ACTB RT-primer	F	AGCCTCGCCTTTGCCGA	17
ACTB RT-primer	R	CTGGTGCCTGGGGCG	15
GAPDH RT-primer	F	ATCTTCTTTTGCCTCGCCAG	20
GAPDH RT-primer	R	AATCCGTTGACTCCGACCTTC	21

# Modulation of Age- and Cancer-Associated DNA Methylation Change in the Healthy Colon by Aspirin and Lifestyle

Faiza Noreen, Martin Rösli, Pawel Gaj, Jakub Pietrzak, Stefan Weis, Patric Urfer, Jaroslaw Regula, Primo Schär, Kaspar Truninger

Manuscript received September 2, 2013; revised April 23, 2014; accepted May 12, 2014.

**Correspondence to:** Primo Schär, PhD, Department of Biomedicine, University of Basel, Mattenstrasse 28, 4058 Basel, Switzerland (e-mail: [primo.schaer@unibas.ch](mailto:primo.schaer@unibas.ch)) and Kaspar Truninger, MD, FMH Gastroenterology and Internal Medicine, Seilereistrasse 1, 4900 Langenthal, Switzerland ([k.truninger@gastroenterologie-oberaargau.ch](mailto:k.truninger@gastroenterologie-oberaargau.ch)).

**Background** Aberrant DNA methylation in gene promoters is associated with aging and cancer, but the circumstances determining methylation change are unknown. We investigated the impact of lifestyle modulators of colorectal cancer (CRC) risk on the stability of gene promoter methylation in the colonic mucosa.

**Methods** We measured genome-wide promoter CpG methylation in normal colon biopsies (n = 1092) from a female screening cohort, investigated the interaction of lifestyle factors with age-dependent increase in methylation with log-linear multivariable regression, and related their modifying effect to hypermethylation in CRC. All statistical tests were two-sided.

**Results** Of 20025 promoter-associated CpGs analyzed, 1713 showed statistically significant age-dependent methylation gains. Fewer CpGs acquired methylation in users of aspirin ( $\geq 2$  years) and hormonal replacement therapy (HRT age  $\geq 50$  years) compared with nonusers (43 vs 1355; 1 vs 1377, respectively), whereas more CpGs were affected in smokers ( $\geq 20$  years) and individuals with a body mass index (BMI) of 25 kg/m<sup>2</sup> and greater compared with control groups (180 vs 39; 554 vs 144, respectively). Fifty percent of the CpGs showing age-dependent methylation were found hypermethylated in CRC (odds ratio [OR] = 20; 95% confidence interval [CI] = 18 to 23;  $P < 2 \times 10^{-16}$ ). These loci gained methylation with a higher median rate compared with age-only methylated sites ( $P = 2 \times 10^{-76}$ ) and were enriched for polycomb regions (OR = 3.67). Importantly, aspirin ( $P < .001$ ) and HRT use ( $P < .001$ ) reduced the methylation rate at these cancer-related genes, whereas smoking ( $P < .001$ ) and high BMI ( $P = .004$ ) increased it.

**Conclusions** Lifestyle, including aspirin use, modulates age-associated DNA methylation change in the colonic epithelium and thereby impacts the evolution of cancer methylomes.

JNCI J Natl Cancer Inst (2014) 106(7): dju161 doi:10.1093/jnci/dju161

Aberrations in DNA methylation deregulate the genome and contribute to the loss of tissue homeostasis observed in aging and diseases such as cancer (1,2). They can constitute both driver and passenger events of tumorigenesis, as evident from widespread changes in CpG methylation observed in cancers. Among those, the hypermethylation of promoter-associated CpG islands is a well-documented phenomenon (3), often associated with the loss of expression of the respective gene (4). Exactly why and how such epigenetic change arises is currently unknown.

Hypermethylation at promoters of tumor suppressor genes is a hallmark of colorectal cancer (CRC), for which age is a major risk factor (5,6). An increasing body of evidence indicates that aging is, at the same time, associated with an accumulation of aberrations in DNA methylation in human tissues, including colon (7–10). We and others have shown that de novo methylation at CpG island promoters of cancer-relevant genes arise in the aging mucosa of healthy individuals with no evidence of intestinal disease (11,12). Notably, these methylation changes appeared in patterns implicating a contribution of

physiological and environmental factors (11,13,14). Understanding what controls the stability of DNA methylation in aging tissue is of both biological and clinical importance. Herein we address to what extent age-associated methylation change in the human colon is influenced by lifestyle and whether the ensuing epigenetic aberrations have relevance for neoplastic transformation. For this purpose, we determined CpG methylation levels in normal colon biopsies from a large female screening cohort and related these to the exposure of well-accepted CRC risk modulators (15–19): aspirin use, hormone replacement therapy (HRT), BMI, and smoking habits.

## Methods

### Study Participants and Data Collection

The study included 546 healthy women randomly recruited from a national colonoscopy screening for CRC in Poland (20). Ethical approval was obtained for all samples. From each participant, biopsies of normal mucosa from the cecum and sigmoid colon were

collected in RNA<sub>later</sub> (Applied Biosystems, Lucerne, Switzerland) and stored at  $-80^{\circ}\text{C}$ . Information on the use of aspirin, HRT, and cigarette smoking was collected using a self-administered questionnaire under the assistance of a study nurse (Table 1; Supplementary Materials, available online).

### Quantification of DNA Methylation

**Locus-Specific DNA Methylation.** Bisulfite-converted genomic DNA of 1092 samples was used to measure human MutL homolog 1 (*bMLH1*) and O<sup>6</sup>-methylguanine DNA methyltransferase (*MGMT*) promoter methylation levels (percentage of methylated alleles) by locus normalized quantitative methylation specific polymerase chain reaction (ln-qMSP). Full details are available in the Supplementary Materials (available online). Primer sequences are provided in the Supplementary Table 1.

**Genome-Wide DNA Methylation.** Genome-wide assessment of DNA methylation was done on Infinium HumanMethylation27 Beadchip arrays (Illumina, San Diego, CA), interrogating methylation at 27 578 CpGs distributed in the promoters of 14 475 coding genes. Details on quantitation of methylation levels are available in the Supplementary Materials (available online). Microarray data were validated by bisulfite pyrosequencing (Supplementary Materials, available online).

All primary data generated in the study were deposited in the NCBI Gene Expression Omnibus (GEO; <http://www.ncbi.nlm.nih.gov/geo/>; accession No. GSE48988). Methylation data of 59 female CRC samples were obtained from GEO (accession No. GSE25062) (21).

### Gene Expression Analysis

Expression data for 32 colon adenomas/normal tissues were obtained from GEO (accession No. GSE8671) (22). The data were median normalized, log<sub>2</sub> transformed, and analyzed using the R/Bioconductor limma package.

### H3K27me3 Status in Human Embryonic Stem Cells

H3K27me3 is a polycomb repressive mark of interest because its presence in stem cells has been shown to predispose genomic sequences to aberrant DNA methylation during aging and carcinogenesis (8,23,24). An H3K27me3 profile in human embryonic stem cells was obtained from GSE13084 (25). Five hundred base-pair windows symmetrically positioned around the Illumina Infinium HumanMethylation27 CpGs were analyzed for enrichment of H3K27me3.

### Statistical Analyses

**Locus-Specific DNA Methylation Analysis.** Log-linear multivariable regression models were applied to investigate the interaction of age-related DNA methylation changes with lifestyle factors by considering age as a linear predictor and the four lifestyle factors of aspirin use, HRT use, BMI, and smoking as categorical predictors. To analyze methylation interaction with age, corresponding interaction terms were introduced in the model. To investigate the interaction of age with polyps, separate multivariable regression models (adjusted for the interaction of the four lifestyle factors with age) were done by considering age as a linear predictor and polyps as categorical predictors (either as one group [any polyp] or classified as tubular adenomas

**Table 1.** Clinical characteristics of the study population

Characteristic	No. of subjects (%)
Age, y	
<60	275 (50)
≥ 60	271 (50)
Aspirin regular use*	
Nonuser†	391 (71)
Short-term user (< 2y)‡	86 (16)
Long-term user (≥ 2y)§	69 (19)
Hormone replacement therapy	
Nonuser¶	245 (45)
Aged <50 y#	175 (32)
Aged ≥ 50 y**	126 (23)
Body mass index, kg/m <sup>2</sup> ††	
Normal, 18.5–25	227 (42)
Overweight, 26–30	224 (41)
Obese, >30	95 (17)
Cigarette smoking‡‡	
Nonsmoker§§	272 (50)
Short-term smoker, <20 y	122 (22)
Long-term smoker, ≥20 y¶¶	152 (28)
Polyps	
No polyps	350 (64)
Yes	196 (36)
Proximal colon##	
Tubular adenoma	49 (25)
Serrated lesion***	25 (13)
Hyperplastic polyps	22 (11)
Sessile serrated adenoma	3 (2)
Distal colon†††	
Tubular adenoma	41 (21)
Serrated lesion	81 (41)
Hyperplastic polyps	69 (35)
Sessile serrated adenoma	12 (6)

\* Regular use defined as two or more tablets per week for 1 or more months.

† Non-user: women who indicated that they did not use two or more aspirin tablets per week for 1 or more months (minimum level).

‡ Short-term user: women who indicated that they used two or more aspirin tablets per week for less than 2 years.

§ Long-term user: women who indicated that they used two or more aspirin tablets per week for 2 or more years.

|| Hormone replacement therapy defined as estrogen therapy and/or oral contraceptive for 1 or more years.

¶ Non-user: women who indicated that they did not have hormone replacement therapy for 1 or more years (minimum level).

# Aged <50: women who indicated that they did had hormone replacement therapy for 1 or more years before age 50 years.

\*\* Aged ≥50 years: women who indicated that they did have hormone replacement therapy for 1 or more years after the age of 50 years.

†† Body mass index: height (cm) and weight (kg) were self-reported, and body mass index was calculated (kg/m<sup>2</sup>) from these variables.

‡‡ Cigarette smoking was defined as one or more cigarettes per day for one or more years.

§§ Nonsmoker: women who indicated that they did not smoke one or more cigarettes per day for 1 or more years (minimum level).

||| Short-term smoker: women who indicated that they smoked one or more cigarettes per day for less than 20 years.

¶¶ Long-term smoker: women who indicated that they smoked one or more cigarettes per day for 20 or more years.

## Proximal colon: cecum.

\*\*\* Serrated lesion: any serrated polyp including hyperplastic and sessile serrated adenoma.

††† Distal colon: sigmoid.



and serrated lesions, including both hyperplastic polyps and sessile serrated adenomas). Separate calculations were done for each colonic location (proximal and distal colon) and gene. For data presentation, log-linear regression coefficients were back transformed and used to calculate rates of change in methylation per 10-year increase in age in respective lifestyle strata. The fold difference in the rate of change in methylation between two categories was defined as the methylation rate ratio (MRR). Full details and analyses are given in the [Supplementary Materials](#) (available online) and [Supplementary Tables 2–4](#) (available online). The association between lifestyle factors and occurrence of polyps was examined by means of logistic multivariable regression models adjusted for age, aspirin use, HRT use, BMI, and smoking but without age interaction ([Supplementary Table 5](#), available online). Regression coefficients are expressed as odds ratios (ORs). Median regression analysis was conducted as supplementary analysis to evaluate the impact of extreme values on the results. The impact of methylation values below the detection limit was tested with tobit regression. Neither supplementary analysis changed the results, and thus they are not reported.

**Genome-Wide DNA Methylation Analysis.** The differential analysis of methylation data was performed on the *M* value (log<sub>2</sub> ratios of methylated and unmethylated probes) using routines implemented in the R/Bioconductor *limma* package (26). Full details are given in the [Supplementary Materials](#) (available online). False-discovery rate (FDR)-adjusted *P* values for multiple comparison were calculated using the Benjamini and Hochberg approach. CpGs were considered statistically significantly differentially methylated with age when displaying an FDR-adjusted *P* value less than .05. CpGs were considered statistically significantly hypermethylated in colorectal cancer vs normal when displaying FDR-adjusted *P* value less than .0001 and fold change difference in *M* values greater than 2. *P* values for the difference in median rate of DNA methylation change were calculated by Wilcoxon rank sum test. *P* values and ORs for overlap were calculated using Fisher exact test. Cancer relevance of identified age-related hypermethylation targets was addressed using the tumor-associated genes (TAG) database (27). Pathway analyses were performed with KEGGprofile in R/Bioconductor. All statistical tests were two-sided.

## Results

We investigated a cohort of 546 healthy women aged 50 to 80 years for relationships between gene promoter methylation in the colonic mucosa and the reported use of aspirin, reported HRT, BMI, and smoking ([Table 1](#)).

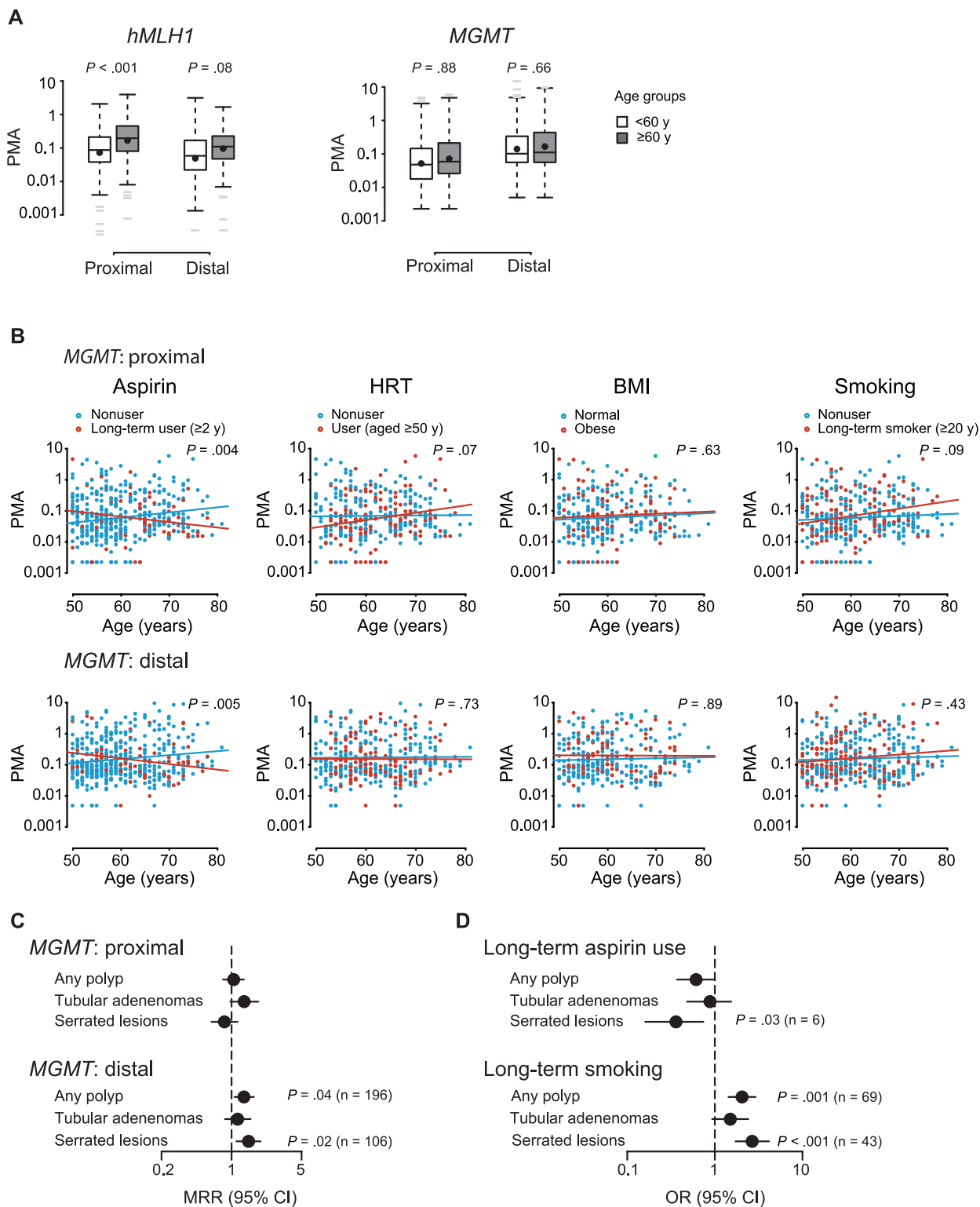
### Age-Related DNA Methylation and Association With Lifestyle and Polyp Occurrence

We first measured *bMLH1* and *MGMT* promoter methylation as percentage of methylated alleles by ln-qMSP in a total of 1092 normal biopsies obtained from cecum (proximal) and sigmoid (distal) colon. Log-linear multivariable regression analysis showed that methylation at the *bMLH1* promoter statistically increased with age in the proximal colon (rate per 10 years of age = 2.1%; *P* < .001) ([Figure 1A](#); [Supplementary Table 2](#), available online). By contrast, the *MGMT* promoter did not show an overall age-dependent methylation change ([Figure 1A](#); [Supplementary Table 3](#), available online) unless lifestyle

factors were taken into account ([Figure 1B](#); [Supplementary Table 3](#), available online); long-term aspirin use was associated with a more than 50% suppressed rate of methylation when compared with non-use (proximal colon: MRR = 0.44, 95% CI = 0.27 to 0.70, *P* = .004; distal colon: MRR = 0.47, 95% CI = 0.31 to 0.73, *P* = .005), whereas HRT after the age of 50 years and long-term smoking slightly stimulated *MGMT* methylation in the proximal colon (HRT: MRR = 1.63, 95% CI = 1.05 to 2.53, *P* = .07; smoking: MRR = 1.51, 95% CI = 1.01 to 2.26, *P* = .09). Notably, women with serrated lesions had a 48% higher rate of *MGMT* promoter methylation in the distal colon compared with women with no polyps (MRR = 1.48; 95% CI = 1.12 to 1.95; *P* = .02) ([Figure 1C](#); [Supplementary Table 4](#), available online), and the risk of serrated lesions was reduced in long-term aspirin users (OR = 0.36; 95% CI = 0.16 to 0.74; *P* = .03) but increased in long-term smokers (OR = 2.67; 95% CI = 1.72 to 4.15; *P* < .001) ([Figure 1D](#); [Supplementary Table 5](#), available online). No association with lifestyle and/or the occurrence of polyps was found for *bMLH1* promoter methylation ([Supplementary Table 2](#), available online).

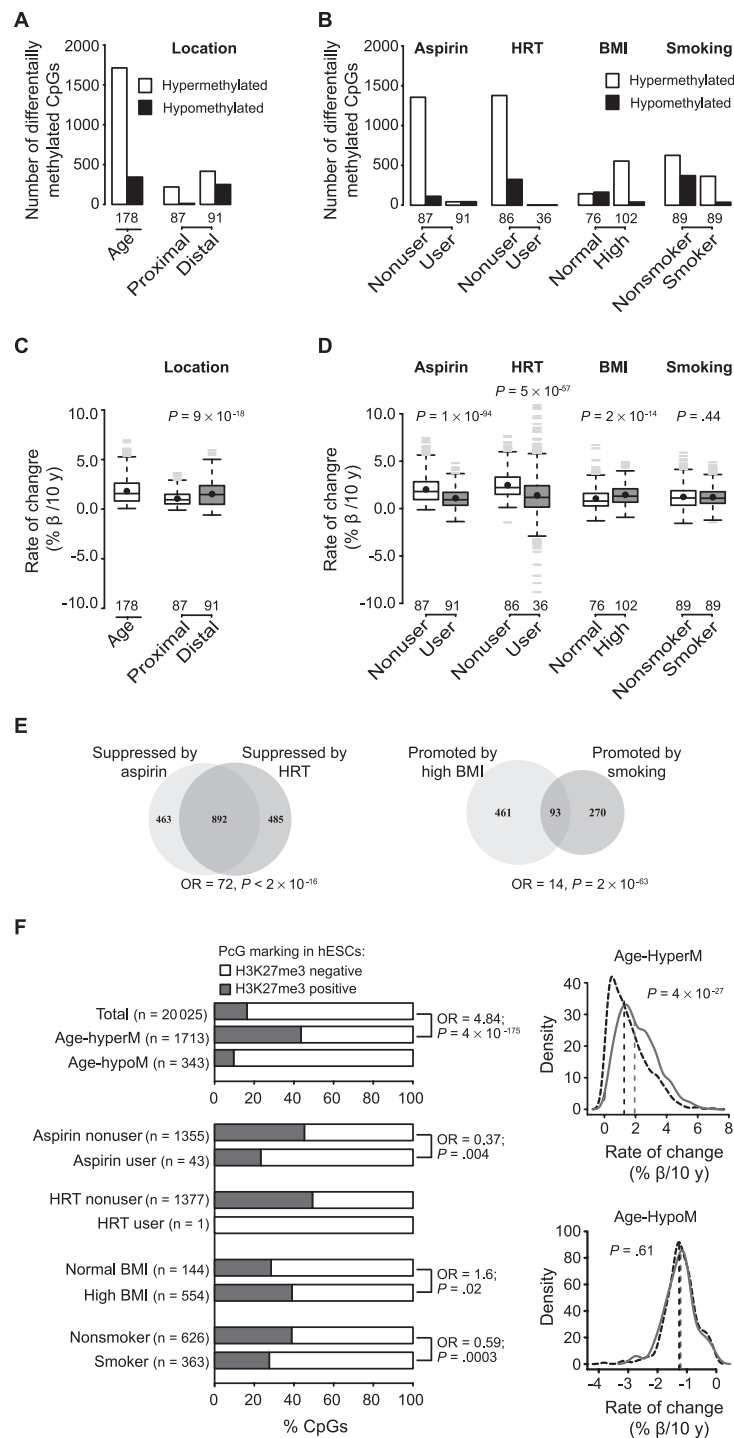
Next, we analyzed genome-wide DNA methylation in 178 normal colon samples from our cohort (Illumina Infinium HumanMethylation27 BeadChip). We analyzed 20025 promoter-associated CpGs with multivariable regression adjusted for colonic location, aspirin use, HRT, BMI, and smoking and identified 1713 CpGs gaining and 343 losing methylation with age; 219 CpGs acquired methylation exclusively in the proximal, and 416 acquired methylation exclusively in the distal colon ([Figure 2A](#)). When stratified for lifestyle factors, age-dependent hypermethylation was suppressed in aspirin users (43 CpGs in users vs 1355 in nonusers) and/or in women reporting HRT (1 CpG in users vs 1377 in nonusers) but promoted in individuals with a high BMI (554 CpGs in high BMI vs 144 in normal BMI) ([Figure 2B](#)). Whereas aspirin use, HRT, and BMI modulated age-dependent DNA methylation changes along the entire colon, the effect of smoking was confined to the proximal colon (180 CpGs in smokers vs 39 in nonsmokers) ([Supplementary Figure 2](#), A and B, available online). The median rate of age-dependent methylation change at hypermethylated CpGs (median rate = 1.6%; range = 0.07%–6.97%) was 56% higher in the distal colon than in the proximal colon (MRR = 1.56; *P* =  $9 \times 10^{-18}$ ) ([Figure 2C](#)). The rate was 48% lower in aspirin users and 47% lower in women reporting HRT compared with non-users (aspirin: MRR = 0.52, *P* =  $1 \times 10^{-94}$ ; HRT: MRR = 0.53; *P* =  $5 \times 10^{-37}$ ) ([Figure 2D](#)), whereas a high BMI was associated with a 27% increased methylation rate in the distal colon (MRR = 1.27; *P* =  $5 \times 10^{-9}$ ), and smoking was associated with a 400% increased rate in the proximal colon (MRR = 4.67; *P* =  $1 \times 10^{-28}$ ) but a 33% decreased rate in the distal colon (MRR = 0.66; *P* =  $8 \times 10^{-46}$ ) ([Supplementary Figure 2](#), A and B, available online).

Examining potential interactions between these lifestyle effects, we found statistically significant overlaps between CpGs showing suppression of age-related methylation by aspirin use or HRT (OR = 72; 95% CI = 62 to 83; *P* <  $2 \times 10^{-16}$ ) ([Figure 2E](#)) and, less pronounced, between CpGs showing BMI- or smoking-stimulated hypermethylation (OR = 14; 95% CI = 11 to 19; *P* =  $2 \times 10^{-63}$ ). Although the methylation-retarding effect of aspirin use was apparent in the entire study population, it appeared more effective in individuals with a high BMI (52%; MRR = 0.48; *P* =  $2 \times 10^{-31}$ ) compared with individuals with a normal BMI (22%; MRR = 0.78; *P* = .0001)



**Figure 1.** Association of human MutL homolog 1 (*hMLH1*) and O<sup>6</sup>-methylguanine DNA methyltransferase (*MGMT*) promoter methylation with lifestyle factors and polyps. **A** *hMLH1* and *MGMT* promoter methylation in proximal (cecum) and distal (sigmoid) colon as percentages of methylated alleles (PMAs) determined by locus normalized quantitative methylation specific polymerase chain reaction (ln-qMSP). For presentation only, age is shown in two groups as indicated. Shown are median (lines) and mean (black circles) PMAs with interquartile ranges (boxes), 1.5 times the interquartile ranges (whiskers), and extreme values (gray lines). **B** Association between age-dependent *MGMT* promoter methylation and lifestyle factors. Each point represents one

biopsy. *P* values are derived from log-linear multivariable regression analysis (Supplementary Table 3, available online) representing significance of the difference in two regression lines. **C** Association of *MGMT* promoter methylation with the occurrence of polyps. Methylation rate ratios (MRRs) and *P* values are derived from log-linear multivariable regression (Supplementary Table 4, available online). **D** Association of lifestyle parameters with the occurrence of polyps. Odds ratios (ORs) and *P* values are derived from logistic multivariable regression analysis (Supplementary Table 5, available online). BMI = body mass index; CI = confidence interval; HRT = hormone replacement therapy. All statistical tests were two-sided.



**Figure 2.** Genome-wide DNA methylation and its association with lifestyle factors. **A** and **B**) Numbers of age-associated differentially methylated CpGs in all samples or when stratified by colon location (proximal [cecum] vs distal [sigmoid]) and lifestyle factors (aspirin: nonuser vs user [long-term]; hormone replacement therapy [HRT]: nonuser vs user [aged  $\geq 50$  years]; body mass index [BMI]: normal vs high [ $>25 \text{ kg/m}^2$ ]; smoking: nonsmoker vs smoker [long-term]). Numbers of samples tested in each category are indicated at the bottom of each bar. **C** and **D**) Ten-year rates of DNA methylation change for CpGs showing age-associated hypermethylation in all samples or when stratified by colon location and lifestyle factors. Shown are median (lines) and mean (black circles) rates with interquartile ranges (boxes), 1.5 times the interquartile ranges (whiskers), and extreme values (gray lines).  $P$  values according to the Wilcoxon rank sum test. **E**) Concordance of probes showing suppression of age-associated

methylation by aspirin-use and HRT or promotion of age-associated methylation by a high BMI and long-term smoking.  $P$  values and odds ratios (ORs) according to the Fisher exact test. **F**) Enrichment of age-associated differentially methylated sites marked by histone 3 lysine 27 tri-methylation (H3K27me3) in human embryonic stem cells (hESCs). Barplots indicate percentages of age-associated hypermethylated (Age-hyperM) and hypomethylated (Age-hypoM) CpGs either marked by H3K27me3 (positive) or not (negative), or the enrichment of CpGs marked by H3K27me3 in lifestyle modulated age-related hypermethylation (bottom). Odds ratios and  $P$  values according to Fisher exact test. Density plots on the right show the rate of change in Age-hyperM and Age-hypoM probes at H3K27me3 positive (gray line) and negative (black, dashed line) CpGs. Dashed vertical lines indicate median rate of change per 10 years of age. PcG = polycomb group. All statistical tests were two-sided.

or in smokers (79%; MRR = 0.21;  $P = 3 \times 10^{-20}$ ) compared with non-smokers (31%; MRR = 0.69;  $P = 2 \times 10^{-84}$ ) (Supplementary Figure 2C, available online). HRT use retarded methylation in individuals with a high BMI only (70%; MRR = 0.30;  $P = 9 \times 10^{-44}$ ), whereas its effect on smoking remained unclear because of limited statistical power.

### Association of Lifestyle-Modulated DNA Methylation With Genomic Regions Disposed to Hypermethylation in Cancer

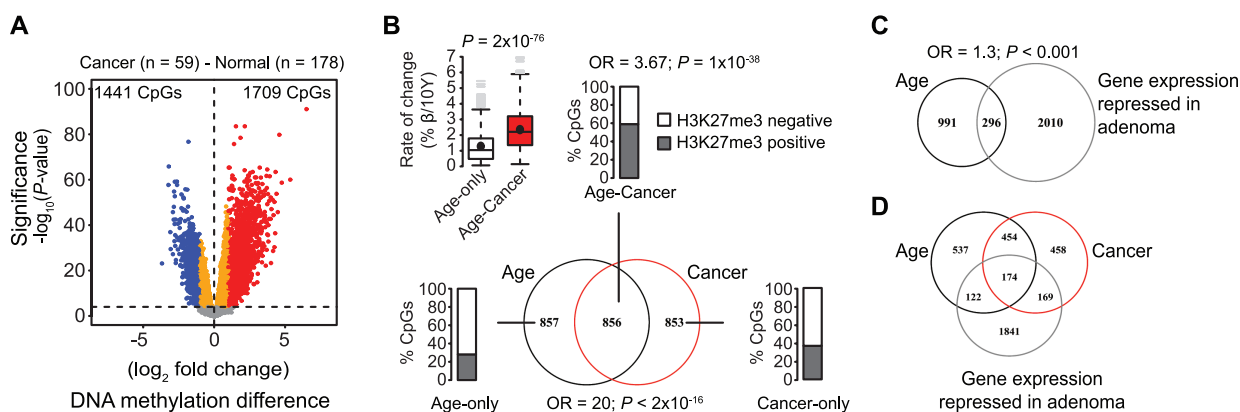
To investigate whether CpGs prone to age-dependent methylation are distinguished by specific genomic features, we classified array CpG probes with regard to coincidence of H3K27me3 in human embryonic stem cells. Regions surrounding age-hypermethylated CpGs were statistically significantly enriched for domains marked by H3K27me3 in stem cells compared with total H3K27me3 marked CpGs on the array (OR = 4.84; 95% CI = 4 to 5;  $P = 4 \times 10^{-175}$ ) (Figure 2F), and the median rate of age-dependent methylation was 53% higher at these sites compared with sites depleted for H3K27me3 (MRR = 1.53;  $P = 4 \times 10^{-27}$ ). Aspirin use suppressed hypermethylation selectively at H3K27me3 marked sequences (user vs nonuser: OR = 0.37; 95% CI = 0.2 to 0.8;  $P = .004$ ), whereas a high BMI promoted methylation at these sites (high vs normal BMI: OR = 1.6; 95% CI = 1.1 to 2.5;  $P = .02$ ). The association of hypermethylation with H3K27me3 modification was less in smokers (smoker vs nonsmokers: OR = 0.59; 95% CI = 0.5 to 0.8;  $P = .0003$ ), indicating that smoking promotes methylation in CpG-rich regions other than the typical polycomb targets.

### Association of Lifestyle-Modulated DNA Methylation with CRC

To address the cancer relevance of these age- and lifestyle-modulated DNA methylation changes in the healthy colon,

we compared methylation data of 59 CRCs (female patients) included in a genome-wide methylation study (21) with those of our healthy mucosa samples. We identified 1709 CpGs statistically significantly hypermethylated and 1441 CpGs hypomethylated in the CRCs (Figure 3A). Half of the CpGs showing age-dependent hypermethylation in the healthy mucosa ( $n = 856$  of 1713; OR = 20; 95% CI = 18 to 23;  $P < 2 \times 10^{-16}$ ) coincided with sites hypermethylated in CRC (Figure 3B; Age-Cancer). At these loci, the median rate of methylation gain per 10 years was doubled (MRR = 2.21;  $P = 2 \times 10^{-76}$ ) compared with age-methylated sites only (Figure 3B; Age-only), and they were most highly enriched for polycomb regions (OR = 3.67; 95% CI = 3.0 to 4.5;  $P = 1 \times 10^{-38}$ ). Age-hypermethylated CpGs also coincided partially with genes transcriptionally downregulated ( $n = 296$  of 1287) (Figure 3C) in colonic adenomas (22), and 174 genes showed all features (ie, age- and CRC-dependent hypermethylation and transcriptional downregulation in adenomas) (Figure 3D).

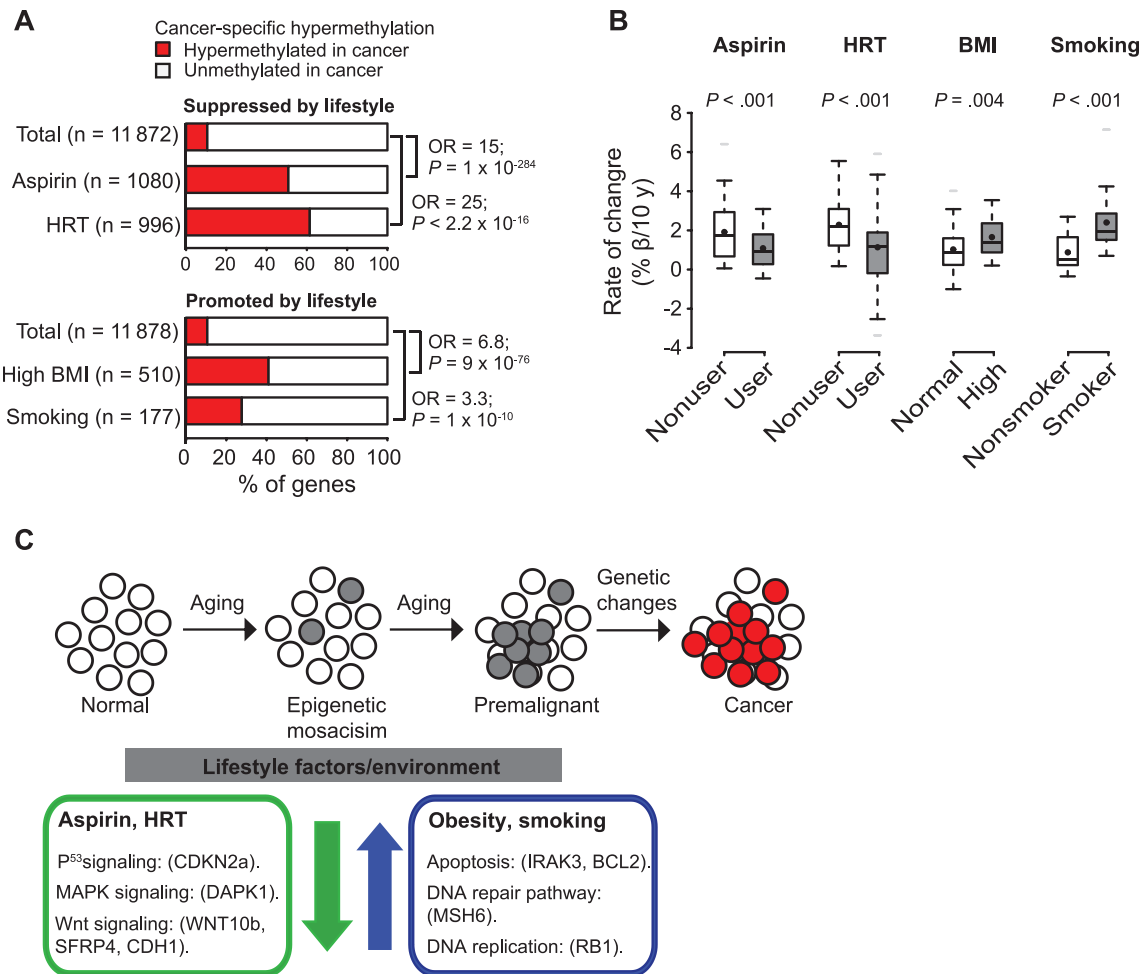
Importantly, cancer-relevant methylation change in the normal colonic mucosa was influenced by lifestyle. A statistically significant fraction of promoters showing either aspirin- or HRT-suppressed methylation (aspirin:  $n = 549$  of 1080, OR = 15, 95% CI = 13 to 17,  $P = 1 \times 10^{-284}$ ; HRT:  $n = 612$  of 996, OR = 25, 95% CI = 22 to 30,  $P < 2.2 \times 10^{-16}$ ) or BMI- or smoking-promoted methylation (high BMI:  $n = 209$  of 510, OR = 6.8, 95% CI = 6 to 8,  $P = 9 \times 10^{-76}$ ; smoking in proximal colon:  $n = 49$  of 177, OR = 3.3, 95% CI = 2 to 7,  $P = 1 \times 10^{-10}$ ) in the aging colon coincided with sites hypermethylated in CRC (Figure 4A). These included genes controlling key aspects of carcinogenesis (Supplementary Figures 3–6, available online) such as cell cycle regulation (*CDKN2A*), DNA repair (*MGMT*), apoptosis (*DAPK1*), cell invasion (*CDH1*), and Wnt (*WNT16*) and RAS signaling (*RASSF1*) (2,28–30). Moreover, the rates of methylation change at promoters of established



**Figure 3.** Enrichment of colorectal cancer (CRC)-associated hypermethylation in age-related methylated genes. **A**) Differences in DNA methylation between 59 CRC samples (21) of female patients and 178 normal biopsies. Plotted are difference in log<sub>2</sub>-fold change (FC) in DNA methylation on the x-axis with false discovery rate (FDR)-adjusted  $P$  values (calculated by moderated  $t$  statistics;  $-1 \times \log_{10}$  scale) on the y-axis. CpGs statistically significantly hypermethylated in CRC are highlighted in red ( $n = 1709$ ; FDR-adjusted  $P < .0001$ ; FC > 2). **B**) Intersection between age-related hypermethylated CpGs in healthy mucosa and CpGs hypermethylated in tumor samples. Barplots indicate percentages of CpGs marked by histone 3 lysine 27 tri-methylation (H3K27me3) in each intersection (Age-only, Age-Cancer, Age-only, Cancer-only). The boxplot shows median rates of DNA methylation change per 10 years

for age-tumor vs age-only hypermethylated loci. Shown are median (lines) and mean (black circles) rates with interquartile ranges (boxes), 1.5 times the interquartile ranges (whiskers), and extreme values (gray lines). **C**) Intersection between 1287 age-related hypermethylated genes ( $n = 1713$  CpGs) in the normal colon mucosa and genes downregulated (FDR-adjusted  $P \leq .05$ ) in colon adenomas (22). **D**) Intersection between genes statistically significantly hypermethylated over age in the normal colon mucosa, genes hypermethylated in CRC samples, and genes downregulated in colon adenomas. Odds ratios (ORs) and associated  $P$  values were calculated according to the Fisher exact test.  $P$  values for the difference in median rates of DNA methylation change were calculated according to the Wilcoxon rank sum test. All statistical tests were two-sided.





**Figure 4.** Modulation of cancer-associated hypermethylation by lifestyle factors. **A**) Concordance between genes showing modulation of age-associated methylation by lifestyle parameters and hypermethylation in colorectal cancer (CRC). Enrichment in each category is calculated over the percentage of 11 872 genes ( $n = 20\,025$  CpGs) present on the Illumina array (Total). Odds ratios (ORs) and  $P$  values were calculated according to the Fisher exact test.  $n$  = number of genes. **B**) Median rates of DNA methylation on 664 annotated tumor-associated genes (TAG database) (27).  $n$  = number of tumor-associated genes hypermethylated in each category.  $P$  values are calculated by Wilcoxon rank sum test. Shown are median (lines) and mean (black circles) rates with interquartile ranges

(boxes), 1.5 times the interquartile ranges (whiskers), and extreme values (gray lines). **C**) Mode for the modulation of CRC risk by lifestyle factors (41). Inherent DNA methylation instability generates epigenetic mosaicism (gray cells) in the aging colonic epithelium. Some methylation changes will affect transcription of genes controlling carcinogenesis, eventually contributing to the evolution of premalignant (gray cell cluster) and cancer cells (red cell cluster). Lifestyle factors are capable of negatively (blue arrow) or positively (green arrow) influencing progressive age-related methylation change, thereby connecting lifestyle with cancer risk. BMI = body mass index; HRT = hormone replacement therapy; OR = odds ratio. All statistical tests were two-sided.

tumor-associated genes (27) were statistically significantly reduced in aspirin users or in women reporting HRT compared with nonusers (aspirin: MRR = 0.53,  $P < .001$ ; HRT: MRR = 0.54,  $P < .001$ ) (Figure 4B) but increased in smokers vs nonsmokers or in individuals with high BMI vs normal BMI (smoking in proximal colon: MRR = 4.0,  $P < .001$ ; high BMI: MRR = 1.57,  $P = .004$ ).

## Discussion

We report that lifestyle factors that are known modulators of CRC risk have widespread effects on the stability of gene promoter methylation in the aging colonic mucosa. We found age-associated DNA hypermethylation to be suppressed by regular aspirin use and HRT use but accelerated by long-term smoking and a high BMI, which is concordant with the effects of these factors on CRC risk. Statistically significant subsets of affected promoters were

associated with genes controlling key aspects of carcinogenesis and with chromatin features known to predispose to hypermethylation in cancer, accentuating the cancer relevance of the lifestyle-modulated DNA methylation change in the normal colon. These findings provide an important resource for understanding the interplay between lifestyle exposure and aging in the modulation of epigenetic (in)stability in colonic epithelial cells and, thereby, in the evolution of CRC methylomes.

Our data extend the well-established relationships between aging and gene promoter hypermethylation (9,31,32) by providing a functional link to cancer. Comparing genome-wide methylation data of CRC (21) with those of normal aging mucosa allowed us to discriminate between age-only, cancer-only and age-cancer hypermethylation. Importantly, the rate of methylation change at age-cancer hypermethylated sites was statistically significantly higher than at loci showing age-only methylation, and aspirin use and

HRT use suppressed whereas high BMI and smoking promoted methylation in a large fraction of these promoters. The identification of this specific subgroup of gene promoters, displaying age-associated and cancer-associated hypermethylation modulated by lifestyle factors, is important for several reasons. First, the targets provide insight into molecular pathogenesis of CRC and how lifestyle factors may exert their effects on CRC risk. Second, age-associated hypermethylation at such targets in the normal colonic mucosa implies a risk for CRC. Hence, age-cancer hypermethylated CpGs may serve as biomarkers for risk prediction and disease progression because their methylation can be monitored in the healthy mucosa. Third, unlike genetic mutations, epigenetic alterations are reversible and the identification of gene promoters at which hypermethylation can be modulated by lifestyle will help develop tailored methylation and thus cancer-preventive strategies. Future investigations will thus have to be directed toward establishing biomarker panels with methylation thresholds predicting cancer risk in healthy individuals and to evaluate the clinical potential of methylation-suppressive medication.

Numerous studies have shown that aspirin use is associated with a lower incidence of colonic neoplasia, metastatic CRC, and death due to CRC (33–36). Although the genetic makeup of a tumor is likely to play a role in this context (37), our methylation profiling of the normal colonic mucosa adds to the understanding of this cancer protective effect, showing that regular aspirin use stabilizes DNA methylation at promoters of genes controlling critical cancer pathways. This is likely to retard carcinogenesis at different stages, including metastasis, in the colon and other tissues. This insight will thus be instrumental in identifying both patients and healthy individuals who will benefit from aspirin use.

Our study has several strengths. First, we collected normal mucosa samples from an unselected screening cohort, providing for an unbiased analysis of methylation profiles. Second, the age-related and lifestyle-modulated methylation changes observed in the promoters of the two DNA-repair and tumor-suppressor genes *hMLH1* and *MGMT* are an important proof of concept because the pathogenic consequence of their epigenetic silencing is well known. Third, the comparison of our genome-wide methylation data in normal mucosa of a female screening cohort with those of women with CRC enabled us not only to exclude sex-associated methylation variance but also to identify the gene promoters at which age-related hypermethylation is relevant for CRC development.

There are also limitations to our study. Residual confounding by measured or unmeasured additional factors cannot be excluded. To limit biological variation, we investigated only women and, therefore, our findings will have to be validated by independent studies including men. Genome-wide methylation array is largely restricted to CpG islands, thereby limiting the detection of methylation globally.

We conclude that the aging colonic mucosa undergoes changes in gene promoter methylation with rates modulated by lifestyle. The highly variable methylation rates at individual CpGs are consistent with an underlying inherent instability of DNA methylation rather than an outgrowth of subpopulations of tissue cells; a concept supported by many other studies (38–40). We propose that this epigenetic instability contributes to the aging of the colonic epithelium (41), generating opportunities for the evolution of early

cancer-initiating cells (Figure 4C) showing patterns of aberrant DNA methylation typically observed in malignant tissues (2). The finding that lifestyle factors specifically modulate the rate of the underlying DNA methylation change provides a novel paradigm for how the environment modulates cancer risk.

## References

1. You JS, Jones PA. Cancer genetics and epigenetics: two sides of the same coin? *Cancer Cell*. 2012;22(1):9–20.
2. De Carvalho DD, Sharma S, You JS, et al. DNA methylation screening identifies driver epigenetic events of cancer cell survival. *Cancer Cell*. 2012;21(5):655–667.
3. Sandoval J, Esteller M. Cancer epigenomics: beyond genomics. *Curr Opin Genet Dev*. 2012;22(1):50–55.
4. Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. *Nat Rev Cancer*. 2011;11(10):726–734.
5. Hoeijmakers JH. DNA damage, aging, and cancer. *N Engl J Med*. 2009;361(15):1475–1485.
6. Balducci L, Ershler WB. Cancer and ageing: a nexus at several levels. *Nat Rev Cancer*. 2005;5(8):655–662.
7. Belshaw NJ, Pal N, Tapp HS, et al. Patterns of DNA methylation in individual colonic crypts reveal aging and cancer-related field defects in the morphologically normal mucosa. *Carcinogenesis*. 2010;31(6):1158–1163.
8. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. *Genome Res*. 2010;20(4):440–446.
9. Bell JT, Tsai PC, Yang TP, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. *PLoS Genet*. 2012;8(4):e1002629.
10. Landan G, Cohen NM, Mukamel Z, et al. Epigenetic polymorphism and the stochastic formation of differentially methylated regions in normal and cancerous tissues. *Nat Genet*. 2012;44(11):1207–1214.
11. Menigatti M, Truninger K, Gebbers JO, et al. Normal colorectal mucosa exhibits sex- and segment-specific susceptibility to DNA methylation at the *hMLH1* and *MGMT* promoters. *Oncogene*. 2009;28(6):899–909.
12. Wallace K, Grau MV, Levine AJ, et al. Association between folate levels and CpG island hypermethylation in normal colorectal mucosa. *Cancer Prev Res*. 2010;3(12):1552–1564.
13. Tapp HS, Commane DM, Bradburn DM, et al. Nutritional factors and gender influence age-related DNA methylation in the human rectal mucosa. *Ageing Cell*. 2013;12(1):148–155.
14. Horii J, Hiraoka S, Kato J, et al. Age-related methylation in normal colon mucosa differs between the proximal and distal colon in patients who underwent colonoscopy. *Clin Biochem*. 2008;41(18):1440–1448.
15. Chan AT, Ogino S, Fuchs CS. Aspirin and the risk of colorectal cancer in relation to the expression of COX-2. *N Engl J Med*. 2007;356(21):2131–2142.
16. Long MD, Martin CF, Galanko JA, et al. Hormone replacement therapy, oral contraceptive use, and distal large bowel cancer: a population-based case-control study. *Am J Gastroenterol*. 2010;105(8):1843–1850.
17. Robsahm TE, Aagnes B, Hjartaker A, et al. Body mass index, physical activity, and colorectal cancer by anatomical subsites: a systematic review and meta-analysis of cohort studies. *Eur J Cancer Prev*. 2013;22(6):492–505.
18. Parajuli R, Bjerkaas E, Tverdal A, et al. The increased risk of colon cancer due to cigarette smoking may be greater in women than men. *Cancer Epidemiol Biomarkers*. 2013;22(5):862–871.
19. Sandler RS, Halabi S, Baron JA, et al. A randomized trial of aspirin to prevent colorectal adenomas in patients with previous colorectal cancer. *N Engl J Med*. 2003;348(10):883–890.
20. Regula J, Rupinski M, Kraszewska E, et al. Colonoscopy in colorectal-cancer screening for detection of advanced neoplasia. *N Engl J Med*. 2006;355(18):1863–1872.
21. Hinoue T, Weisenberger DJ, Lange CP, et al. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*. 2012;22(2):271–282.
22. Sabates-Bellver J, Van der Flier LG, de Palo M, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res*. 2007;5(12):1263–1275.

23. Watson CT, Disanto G, Sandve GK, et al. Age-associated hyper-methylated regions in the human brain overlap with bivalent chromatin domains. *PLoS One*. 2012;7(9):e43840.
24. Rakyan VK, Down TA, Maslau S, et al. Human aging-associated DNA hypermethylation occurs preferentially at bivalent chromatin domains. *Genome Res*. 2010;20(4):434–439.
25. Ku M, Koche RP, Rheinbay E, et al. Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet*. 2008;4(10):e1000242.
26. Du P, Zhang X, Huang CC, et al. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.
27. Chen JS, Hung WS, Chan HH, et al. In silico identification of oncogenic potential of fyn-related kinase in hepatocellular carcinoma. *Bioinformatics*. 2013;29(4):420–427.
28. Richter AM, Pfeifer GP, Dammann RH. The RASSF proteins in cancer; from epigenetic silencing to functional characterization. *Bba-Rev Cancer*. 2009;1796(2):114–128.
29. Gonzalo V, Lozano JJ, Munoz J, et al. Aberrant gene promoter methylation associated with sporadic multiple colorectal cancer. *PLoS One*. 2010;5(1):e8777.
30. Pehlivan S, Artac M, Sever T, et al. Gene methylation of SFRP2, P16, DAPK1, HIC1, and MGMT and KRAS mutations in sporadic colorectal cancer. *Cancer Genet Cytogen*. 2010;201(2):128–132.
31. Koch CM, Suschek CV, Lin Q, et al. Specific age-associated DNA methylation changes in human dermal fibroblasts. *PLoS One*. 2011;6(2):e16679.
32. Alisch RS, Barwick BG, Chopra P, et al. Age-associated DNA methylation in pediatric populations. *Genome Res*. 2012;22(4):623–632.
33. Bosetti C, Rosato V, Gallus S, et al. Aspirin and cancer risk: a quantitative review to 2011. *Ann Oncol*. 2012;23(6):1403–1415.
34. Algra AM, Rothwell PM. Effects of regular aspirin on long-term cancer incidence and metastasis: a systematic comparison of evidence from observational studies versus randomised trials. *Lancet Oncol*. 2012;13(5):518–527.
35. Maxwell PH. Tumor strengths and frailties: aspiring to prevent colon cancer. *Nat Med*. 2012;18(1):32–33.
36. Thun MJ, Jacobs EJ, Patrono C. The role of aspirin in cancer prevention. *Nat Rev Clin Oncol*. 2012;9(5):259–267.
37. Liao XY, Lochhead P, Nishihara R, et al. Aspirin use, tumor PIK3CA mutation, and colorectal-cancer survival. *New Engl J Med*. 2012;367(17):1596–1606.
38. West J, Beck S, Wang X, et al. An integrative network algorithm identifies age-associated differential methylation interactome hotspots targeting stem-cell differentiation pathways. *Sci Rep*. 2013;3:1630.
39. Maegawa S, Hinkal G, Kim HS, et al. Widespread and tissue specific age-related DNA methylation changes in mice. *Genome Res*. 2010;20(3):332–340.
40. Teschendorff AE, Menon U, Gentry-Maharaj A, et al. An epigenetic signature in peripheral blood predicts active ovarian cancer. *PLoS One*. 2009;4(12):e8274.
41. Estecio MR, Issa JP. Dissecting DNA hypermethylation in cancer. *FEBS Lett*. 2011;585(13):2078–2086.

## Funding

This work was supported by the Swiss Cancer League; Association for International Cancer Research (AICR-11-0247); Stanley Thomas Johnson foundation; and OPO Foundation.

## Notes

We wish to thank Robert Ivanek for critically reviewing the bioinformatics involved in the global methylation analyses, Giancarlo Marra for his contributions to gene expression analyses, and members of the Department of Biomedicine for input and critical discussions throughout the course of the study.

**Affiliations of authors:** Department of Biomedicine, University of Basel, Basel, Switzerland (FN, SW, PU, PS, KT); Department of Epidemiology & Public Health, Swiss Tropical and Public Health Institute, Basel, Switzerland (MR); Department of Gastroenterology, Medical Center for Postgraduate Education, Maria Skłodowska-Curie Memorial Cancer Center, Warsaw, Poland (PG, JP, JR); FMH Gastroenterology and Internal Medicine, Langenthal, Switzerland (KT).

## Supplementary Materials

---

### Supplementary Methods

#### DNA Isolation and bisulfite conversion

Genomic DNA was extracted from freshly collected biopsies of normal mucosa stored in RNA*later* using QIAamp DNA mini kit (Qiagen, Hilden, Germany) including RNAse A treatment according to the manufacturer protocol. 2 µg genomic DNA from each sample was bisulfite-treated using the EZ DNA Methylation kit (Zymo Research, Orange, CA), according to the manufacturer's protocol and stored at -80°C until for quantitative methylation-specific PCR.

#### Study participants and data collection

Aspirin use was determined as never versus (vs.) ever ( $\geq$  twice per week for >1 month) and medication duration was scaled as short-term (< 2 years) vs. long-term ( $\geq$  2 years). Hormonal medication use (estrogen therapy and/or oral contraceptives) was determined as never vs. ever ( $\geq$  one year) and whether hormone replacement therapy (HRT) was started before or after age 50 years. Height (cm) and weight (kg) were self-reported and BMI was calculated ( $\text{kg}/\text{m}^2$ ) from these variables. Smoking status was determined as never vs. ever ( $\geq$  one cigarette daily for  $\geq$  1 year), while smoking duration was determined as short-term (< 20 years) vs. long-term ( $\geq$  20 years). Polyp status was determined as no polyps vs. yes (any type), while polyp type was classified as tubular adenoma vs. serrated lesion (hyperplastic, sessile serrated adenoma). Subjects with missing information in each category were coded as not being exposed to these factors.

#### Quantification of DNA methylation

**Locus-specific DNA methylation:** *hMLH1* and *MGMT* promoter methylation was measured by locus normalized quantitative methylation specific PCR (ln-qMSP). Bisulfite-modified DNA was used as a template using primers pair specific for either methylated (M) or modified unmethylated (U) sequence. Amplification was normalized for total DNA input using a set of standardized primers (S). Primer sequences are provided in Supplementary Table 2. These primers were located in areas without CpG



nucleotides, thus amplifying the modified DNA independently of methylation status of CpG nucleotides. In order to ensure product specificity, a melting point analysis was performed after each run. PCR reactions were performed with QuantiTect SYBR Green PCR Kit (Qiagen, Hilden, Germany) using Rotor-Gene 3000 real time thermo cycler (Corbett life science, Sydney, Australia). Each reaction was carried out in duplicates in a final volume of 15µl containing 1.5µl of bisulfite-modified DNA and 0.5µM concentration of each forward and reverse primer. PCR conditions were as followed: activation of polymerase at 95°C for 10 sec; primer annealing at 58° for 25 sec and extension at 62°C for 20 sec. To quantify methylation, blood DNA from healthy individual was *in vitro* methylated with SssI methyltransferase (New England Biolabs, Hitching, UK) and further bisulfite-modified to generate positive-control for each PCR reaction. Serial dilutions (from 100 to 0.1%) of positive-control DNA were used to generate standard curves for each gene using M- and S-primers. The quality of positive-control DNA was proven by unsuccessful amplification with U-primers under optimized conditions. To determine the methylation levels in each sample, the value of amplification with M-primers was compared to the value of amplification with S-primers and this ratio was then multiplied by 100 to give percentage of methylated alleles (PMA) ( $[M\text{-primers amplification levels}/S\text{-primers amplification level}] \times 100$ ). A sample was considered methylation positive, if the assay performed in duplicate resulted in a PMA above 0.0001% (reflecting the detection limit for reproducible measurements). At this threshold, the frequency of individuals showing methylation positivity ranged from 85% to 98%. To conduct *log* transformation of the methylation results for data analysis, all non-detects were replaced by half of the lowest observed value. This was 0.00015% for right colon *hMLH1* and 0.00035% left colon *hMLH1*, 0.0023% for right colon *MGMT* and 0.005% left colon *MGMT*.

***Genome-wide DNA methylation:***

***Quality control and data normalization:*** Total of 184 normal colon biopsies from our female cohort were processed on 16 chips (12 samples per chip). Samples were run together in a randomized order to avoid confounding batch effects within the population. Mean non-background corrected signal intensities of the methylated (M) and

unmethylated (U) probes were extracted using the Illumina BeadStudio software v3.2 and then exported to lumi and methylumi packages (1) using R/Bioconductor (<http://www.bioconductor.org>). Samples were filtered according to CpG coverage, using *P* values of detection of signal above background, demanding at least 95% coverage per sample. After this quality control step, 178 samples remained. Using same formula applied in GenomeStudio, Beta ( $\beta$ ) values ( $M/[M+U+100]$ ) and M-value ( $\log_2$  ratios of methylated and unmethylated probes) were calculated after performing smoothQuantileNormalization for color balance adjustment followed by background level correction and quantile based probe level normalization according to the routine implemented in lumi and methylumi packages in R/Bioconductor (2).

**Data Filtering:** For the data analysis we masked data point as “NA” for the following conditions as described previously:[3] 1) Probes containing single-nucleotide polymorphisms (SNPs), 2) those that overlap with a repetitive element that covers the targeted CpG dinucleotide, 3) those that overlap with regions of insertions and deletions in the human genome, 4) probes that are targeting X or Y chromosome. Beta values with detection p value greater than 0.01 were also replaced as "NA". Furthermore, probes containing at least one “NA” across the normal sample set were masked as “NA”. In total, 20,025 probes corresponding to 11,872 genes were kept for further analysis.

### **Validation of microarray data by pyrosequencing**

Five genes were selected for validation analysis of microarray results in 8 healthy tissue samples by bisulfite/pyrosequencing (Supplementary Figure 7). PCR was performed in two reactions. The first PCR used a forward and a reverse primer with one primer contained an 11-base tag sequence, 5'-GCCCCCGCCCG-3'. The second PCR contained one of the same primer as before and second universal primer was an 11-base tag sequence, 5'-(biotin)GCCCCCGCCCG-3', which had been biotin labeled at the 5' end and HPLC-purified (Microsynth AG, Switzerland).

The first PCR was carried out in a volume of 25  $\mu$ l using the FastStart Taq DNA polymerase (Roche Diagnostics,USA). After 6 min of initial denaturation at 95°C, the cycling conditions of 35 cycles consisted of denaturation at 95°C for 30 s, annealing at

56°C (PYCARD), 58°C (TLE1, RAC2), 59°C (ZBTB9) or 62°C (CHD5) for 30 s and elongation at 72°C for 45 s. The amplified fragment was diluted 1000-fold and used as template in the second PCR. The second PCR was carried out in a volume of 50 µl, which after initial denaturation at 95°C for 5 min comprised 20 cycles with denaturation at 95°C for 30 s, annealing at 52°C for 30 s and elongation at 72°C for 45 s. The PCR products were purified using NucleoSpin Extract II kit (Macherey-Nagel GmbH, Germany) and were stored at 4°C until ready for pyrosequencing. 500ng of PCR product was used for pyrosequencing using the PyroMark Q24 Pyrosequencing system according to manufacturer's instructions (Qiagen). The primers used in the PCR runs and pyrosequencing reactions are shown in Supplementary Table 2.

### **Statistical analysis**

**Genome-wide methylation analysis:** To assess differences in CpG methylation over 10 years of age, we fitted linear regression models on M-values [2] adjusted for important covariates, including aspirin use, use of HRT, BMI and smoking. By doing stratified multivariable regression analyses, we further evaluated potential age-related differential methylation modification for aspirin use (user  $\geq$  2 years vs. non-user), HRT (user after age 50 vs. non-user), BMI (high  $>$  25kg/m<sup>2</sup> vs. normal  $\leq$  25kg/m<sup>2</sup>) and smoking status (smoker  $\geq$  20 years vs. nonsmoker). All above models were also done separately for each colonic location. Similarly, stratified multivariable regression analyses were done for Aspirin/BMI (non-user/normal BMI, user/normal BMI, non-user/high BMI, user/high BMI), Aspirin/smoking (non-user/non-smoker, user/non-smoker, non-user/smoker, user/smoker), HRT/BMI (non-user/normal BMI, user/normal BMI, non-user/high BMI, user/high BMI) and HRT/smoking (non-user/non-smoker, user/non-smoker, non-user/smoker, user/smoker). All above models were also performed on  $\beta$  values and coefficients of multivariable regression models were used to determine the rate of DNA methylation change / 10 years in each group.

An equivalent approach was used to assess differences in CpG methylation between 59 tumors tissues (3) versus 178 healthy tissues. We first assessed the inherent

batch effect between two data sets. We compared these methylation profiles with those obtained from 4 CRC samples processed together with the healthy mucosa samples in our laboratory. This showed a significant correlation ( $r^2 = 0.95$ ;  $P < .001$ ), suggesting that batch-effect in the data sets are minimal.

## References

1. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*. 2008;24(13):1547-1548.
2. Du P, Zhang X, Huang CC, *et al*. Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics*. 2010;11:587.
3. Hinoue T, Weisenberger DJ, Lange CP, *et al*. Genome-scale analysis of aberrant DNA methylation in colorectal cancer. *Genome Res*. 2012;22(2):271-82.

**Supplementary Table 1. Primers used for In-qMSP and pyrosequencing**

Primer name	Sequences
<b>Locus normalized quantitative methylation specific PCR (In-qMSP) primers</b>	
<i>hMLH1-M</i> <sup>†</sup>	F: 5'-CGG GTA AGT CGT TTT GAC GTA GAC-3' R: 5'-ACG CCA CTA CGA AAC TAA ACA CG-3'
<i>hMLH1-U</i> <sup>‡</sup>	F: 5'-GAT GTA GAT GTT TTA TTA GGG TTG TGT-3' R: 5'-TATCAC CAC CTC ATC ATA ACT ACC CA-3'
<i>hMLH1-S</i> <sup>¶</sup>	F: 5'-AGT TTG GAG TGG TAG GTT TTT AGA GG-3' R: 5'-CAT CCC CAA ATT CTT ACT CCT TCT TTC-3'
<i>MGMT-M</i> <sup>†</sup>	F: 5'-AGG CGT CT TCG GTT TGT ATC G-3' R: 5'-CGC CGA ACC TAA AAC AAT CTA CG-3'
<i>MGMT-U</i> <sup>‡</sup>	F: 5'-GGT GGA AGT TGG GAA GGT GTT G-3' R: 5'-CAA CAC ATA CCC AAT ACA ACA ACA CCA-3'
<i>MGMT-S</i> <sup>¶</sup>	F: 5'-GTG TTT GGA GGA AAA GGG GTT-3' R: 5'-CTA CCC CTC ACT AAA CAA CCA-3'
<b>Pyrosequencing primers</b>	
<i>PYCARD</i>	F: 5'-GTT ATT TTG GAT GAG TTG GAG AAT TTG AT-3' R: 5'- <u>GCC CCC GCC CGA</u> CCA TCT CCT ACA AAC CCA TAT C-3' seq:5'-GAG TTT AAG AAG TTT AAG TTG A-3'
<i>TLE1</i>	F: 5'- <u>GCC CCC GCC CGG</u> GGT TTG TGT GGG GTT AG-3' R: 5'-CCA AAC CAA CTA CTT TTC TTT ACT CTT CTC-3' seq:5'-CTT TTC TTT ACT CTT CTC C-3'
<i>RAC2</i>	F: 5'-GGG GAG AGT AGT TTG AGT AAG TTT-3' R: 5'- <u>GCC CCC GCC CGC</u> CAC CAC ACA CTT AAT AAC CTA CA-3' seq:5'-GTT TTA TTA GTT TTT ATA TTT TTT G-3'
<i>CHD5</i>	F: 5'-TTA GTT TTG TTT TAG AGG GTT TTG GTT T-3' R: 5'- <u>GCC CCC GCC CGC</u> CAA AAA ACC CCT CCA AAC T-3' seq:5'-TTG TTT TGA GAT AGG TGT-3'
<i>ZBTB9</i>	F: 5'-AGG ATA TTT TGG GTT TGT GTT TTT TTA TTA-3' R: 5'- <u>GCC CCC GCC CGA</u> CAA TAC CCA CAA AAA TCT ACT CC-3' seq:5'-GAG TTG TAG AGG TAT TGT AA-3'
Universal primer	5'-GCC CCC GCC CG-3'

<sup>†</sup>Methylated primer. <sup>‡</sup>Unmethylated primer. <sup>¶</sup>Standard primer. seq: sequencing primer. Sequence underlined represents the 11-base tag.

**Supplementary Table 2.** Association of *hMLH1* promoter methylation and various lifestyle factors

Variables	Proximal colon		Distal colon	
	Ratio (95% CI) <sup>*</sup>	P <sup>†</sup>	Ratio (95% CI) <sup>*</sup>	P <sup>†</sup>
Age (for every 10 years)	2.06 (1.49-2.85)	0.0002	1.41 (1.02-1.96)	0.08
<b>Age-aspirin use interaction</b>				
Non-user <sup>c</sup>	1 (ref)	-	1 (ref)	-
Short-term user (<2y)	1.06 (0.65-1.74)	0.84	0.73 (0.44-1.20)	0.29
Long-term user (≥2y)	1.01 (0.64-1.61)	0.97	1.06 (0.66-1.69)	0.84
<b>Age-HRT interaction</b>				
Non-user <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
< age 50	1.25 (0.85-1.83)	0.35	1.33 (0.90-1.96)	0.22
≥ age 50	1.24 (0.80-1.92)	0.41	0.97 (0.63-1.51)	0.92
<b>Age-BMI interaction</b>				
Normal (18.5-25) <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
Overweight (26-30)	0.79 (0.56-1.11)	0.25	1.12 (0.79-1.59)	0.59
Obese (>30)	1.05 (0.65-1.69)	0.88	0.69 (0.43-1.12)	0.20
<b>Age-smoking interaction</b>				
Non-smoker <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
Short-term smoker (<20y)	0.88 (0.57-1.35)	0.62	1.06 (0.68-1.64)	0.84
Long-term smoker (≥20y)	1.06 (0.71-1.58)	0.82	1.26 (0.84-1.89)	0.34

<sup>\*</sup> Log-linear regression coefficients from multivariable regression model were back transformed and referred as rate ratio (RR) with respect to reference category in each comparison.. In bracket is 95% confidence interval. Separate models were done for each location and gene. Explanatory variables of primary interest were age, regular aspirin use, cigarette smoking, HRT and BMI. To test whether age-related increase of methylation is influenced by any of the lifestyle variable, corresponding interaction terms were used in the models. <sup>†</sup>P derived from multivariable regression model, <sup>ψ</sup>Reference category.

**Supplementary Table 3.** Association of *MGMT* promoter methylation and various lifestyle factors

Variables	Proximal colon		Distal colon	
	RR (95% CI) <sup>‡</sup>	P <sup>§</sup>	RR (95% CI) <sup>‡</sup>	P <sup>§</sup>
Age (for every 10 years)	1.03 (0.74-1.43)	0.89	1.08 (0.80-1.46)	0.67
<b>Age-aspirin use interaction</b>				
Non-user <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
Short-term user (<2y)	0.82 (0.50-1.35)	0.51	0.70 (0.44-1.11)	0.21
Long-term user (≥2y)	0.44 (0.27-0.70)	0.004	0.47 (0.31-0.73)	0.005
<b>Age-HRT interaction</b>				
Non-user <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
< age 50	1.25 (0.85-1.84)	0.34	1.31 (0.92-1.87)	0.21
≥ age 50	1.63 (1.05-2.53)	0.07	0.92 (0.61-1.38)	0.73
<b>Age-BMI interaction</b>				
Normal (18.5-25) <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
Overweight (26-30)	1.20 (0.84-1.70)	0.39	1.22 (0.89-1.69)	0.30
Obese (>30)	1.15 (0.71-1.87)	0.63	1.04 (0.67-1.62)	0.89
<b>Age-smoking interaction</b>				
Non-smoker <sup>ψ</sup>	1 (ref)	-	1 (ref)	-
Short-term smoker (<20y)	1.01 (0.65-1.56)	0.98	1.01 (0.67-1.51)	0.10
Long-term smoker (≥20y)	1.51 (1.01-2.26)	0.09	1.20 (0.83-1.73)	0.43

<sup>‡</sup>Log-linear regression coefficients from multivariable regression model were back transformed and referred as rate ratio (RR) with respect to reference category in each comparison. In bracket is 95% confidence interval. Separate models were done for each location and gene. Explanatory variables of primary interest were age, regular aspirin use, cigarette smoking, HRT and BMI. To test whether age-related increase of methylation is influenced by any of the lifestyle variable, corresponding interaction terms were used in the models. <sup>§</sup>P derived from multivariable regression model, <sup>ψ</sup>Reference category.

**Supplementary Table 4.** Association of promoter methylation with occurrence of polyps

	Proximal colon			Distal colon		
	Cases	RR <sup>‡</sup> (95% CI)	P <sup>§</sup>	Cases	RR <sup>‡</sup> (95% CI)	P <sup>§</sup>
<b><i>hMLH1</i></b>						
No polyps <sup>ψ</sup>	350	1.00		350	1.00	
Any polyp	74	0.92 (0.72-1.17)	0.57	122	1.01 (0.79-1.29)	0.96
Tubular adenoma	49	0.96 (0.70-1.32)	0.84	41	0.86 (0.62-1.19)	0.45
Serrated lesion	25	0.91 (0.67-1.22)	0.59	81	1.15 (0.85-1.55)	0.46
<b><i>MGMT</i></b>						
No polyps <sup>ψ</sup>	350	1.00		350	1.00	
Any polyp	74	1.05 (0.82-1.33)	0.77	122	1.30 (1.00-1.69)	0.04
Tubular adenoma	49	1.34 (0.57-1.21)	0.13	41	1.15 (0.86-1.55)	0.44
Serrated lesion	25	0.85 (0.27-1.49)	0.37	81	1.48 (1.12-1.95)	0.02

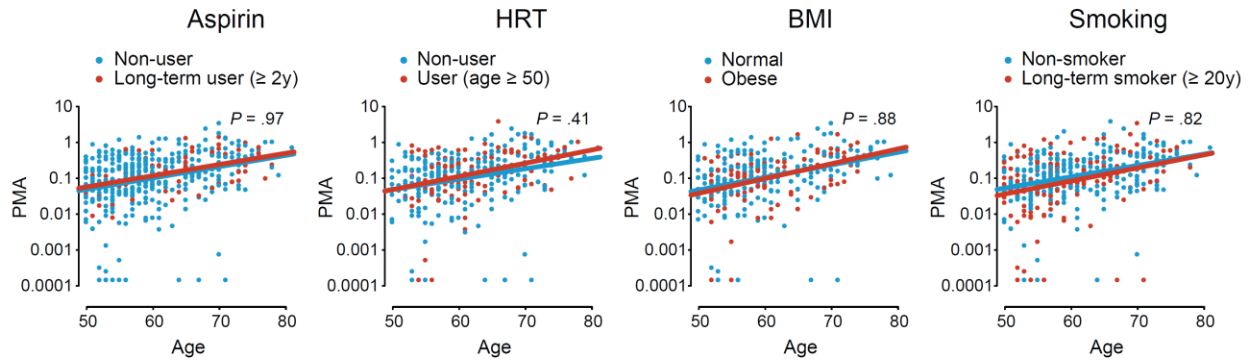
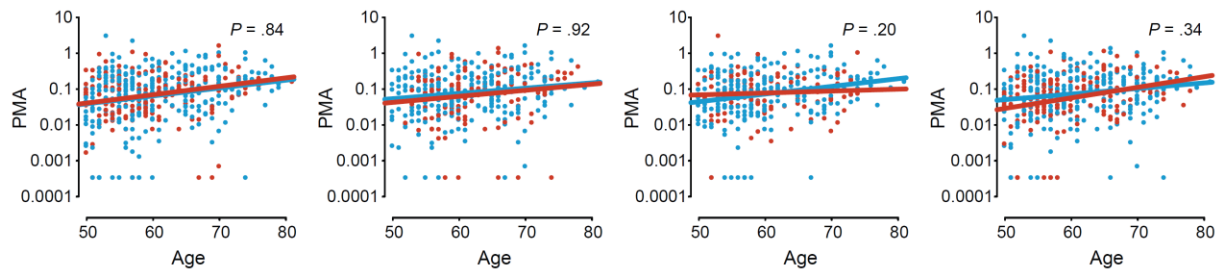
<sup>‡</sup>Log-linear regression coefficients from multivariable regression model were back transformed and referred as rate ratio (RR) with respect to reference category in each comparison. In bracket is 95% confidence interval. Separate models were done for each location and gene. <sup>§</sup>P derived from multivariable regression model, <sup>ψ</sup>Reference category.



**Supplementary Table 5. Association of polyp formation with lifestyle factors**

Variables	Any			Tubular adenoma			Serrated lesion		
	Cases	OR* (95% CI)	P	Cases	OR* (95% CI)	P	Cases	OR* (95% CI)	P
<b>Aspirin regular use</b>									
Non-user <sup>¶</sup>	145	1		65	1		80	1	
Short-term user (<2y)	31	0.96 (0.63-1.46)	0.88	11	0.71 (0.38-1.27)	0.35	20	1.27 (0.76-2.09)	0.43
Long-term user (≥2y)	20	0.61 (0.37-0.98)	0.1	14	0.88 (0.48-1.54)	0.71	6	0.36 (0.16-0.74)	0.03
<b>Hormone replacement therapy (HRT)</b>									
Non-user <sup>¶</sup>	85	1		38	1		47	1	
< age 50	58	0.84 (0.59-1.21)	0.44	25	0.89 (0.54-1.43)	0.68	33	0.81 (0.51-1.25)	0.42
≥ age 50	53	1.32 (0.90-1.93)	0.23	27	1.45 (0.89-2.36)	0.21	26	1.19 (0.73-1.92)	0.55
<b>Body mass index (BMI)</b>									
Normal (18.5-25 kg/m <sup>2</sup> ) <sup>¶</sup>	82	1		32	1		50	1	
Overweight (26-30 kg/m <sup>2</sup> )	84	1.08 (0.78-1.51)	0.69	44	1.41 (0.91-2.19)	0.20	40	0.83 (0.55-1.26)	0.47
Obese (>30 kg/m <sup>2</sup> )	30	0.90 (0.57-1.39)	0.68	14	1.04 (0.57-1.87)	0.91	16	0.78 (0.44-1.35)	0.47
<b>Cigarette smoking</b>									
Non-smoker <sup>¶</sup>	81	1		43	1		38	1	
Short-term smoker (<20y)	46	1.45 (0.98-2.15)	0.11	21	1.24 (0.74-2.06)	0.49	25	1.74 (1.05-2.85)	0.07
Long-term smoker (≥20y)	69	2.04 (1.43-2.92)	0.001	26	1.50 (0.93-2.41)	0.16	43	2.67 (1.72-4.15)	0.0002

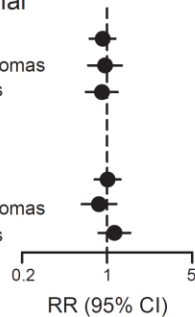
\*Odds ratio derived from coefficients of logistic multivariable regression. In bracket is 95% confidence interval. <sup>¶</sup>P derived from logistic multivariable regression model, <sup>¶¶</sup>Reference category.

**A***hMLH1*: proximal*hMLH1*: distal**B***hMLH1*: proximal

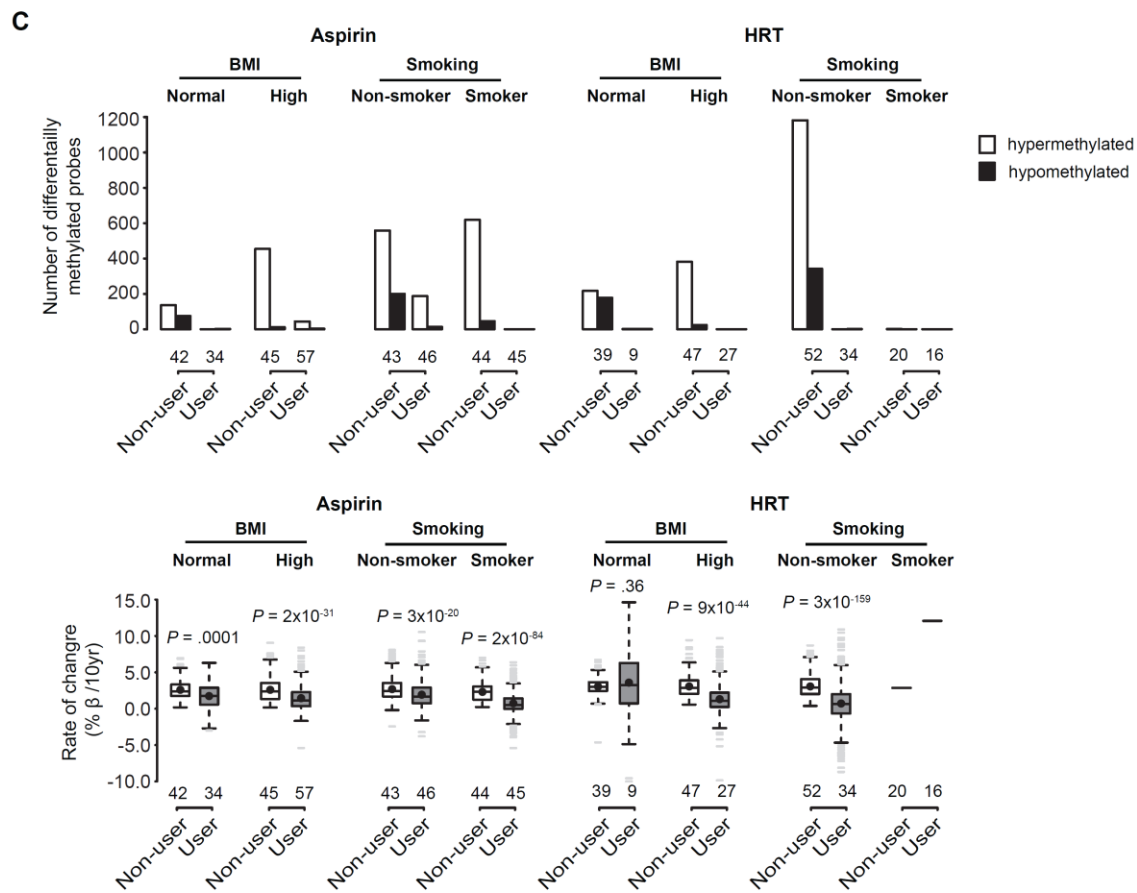
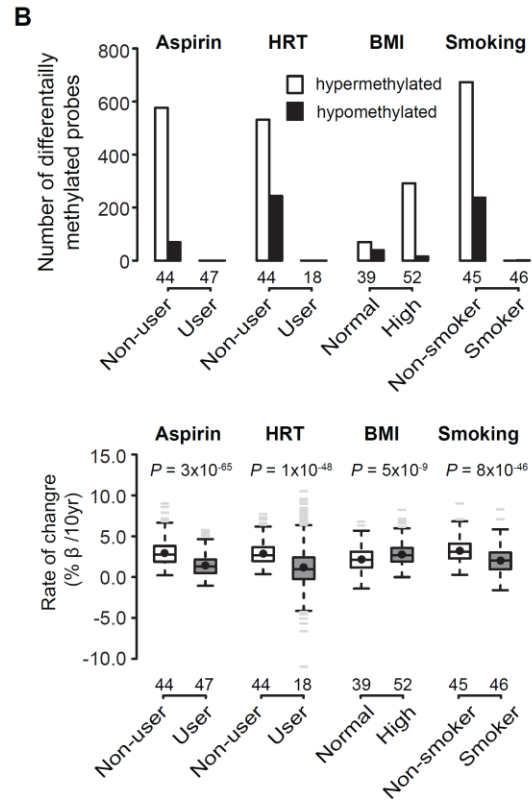
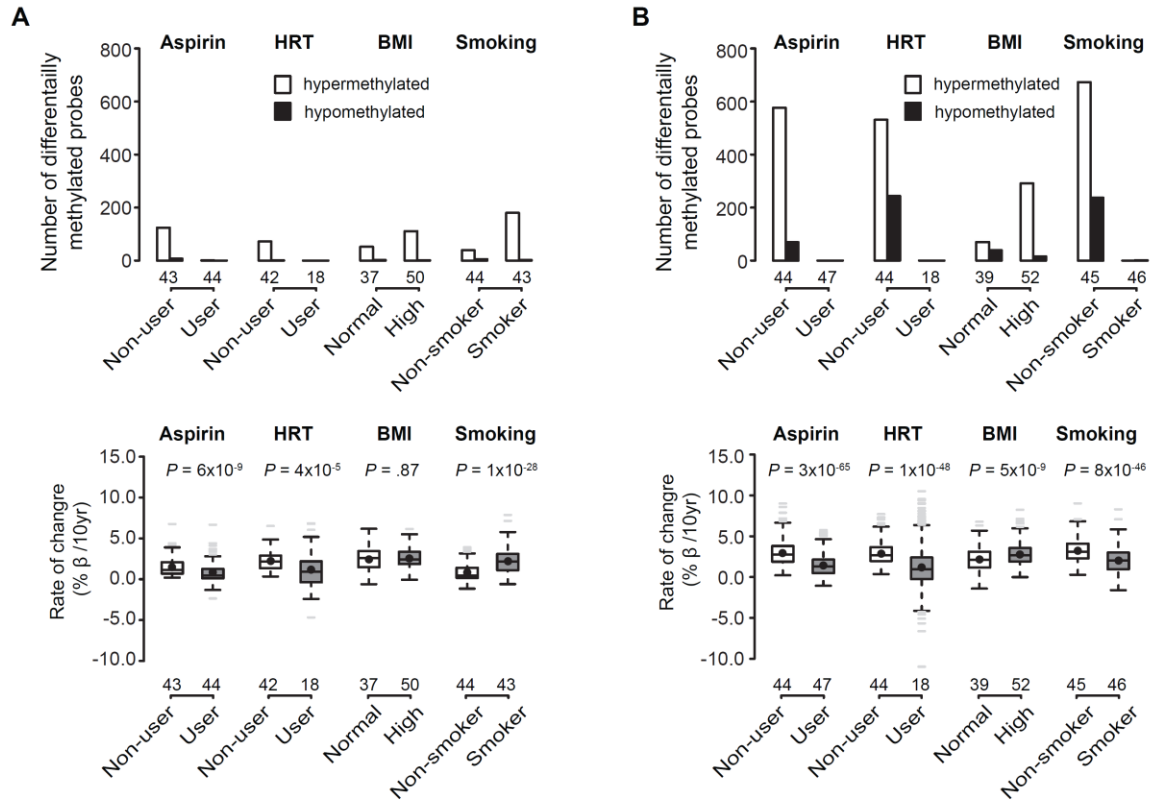
Any polyp  
 Tubular adenomas  
 Serrated lesions

*hMLH1*: distal

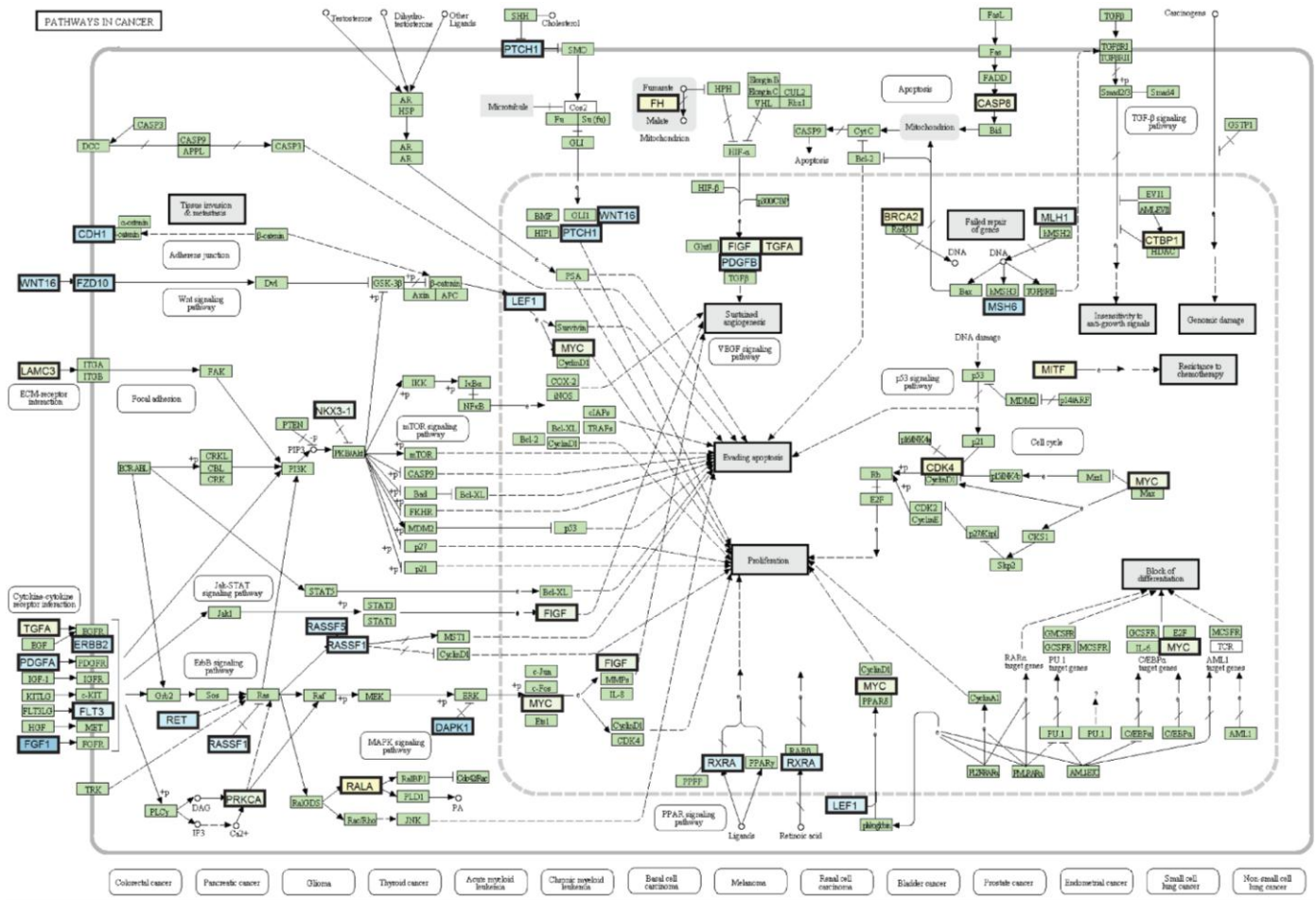
Any polyp  
 Tubular adenomas  
 Serrated lesions



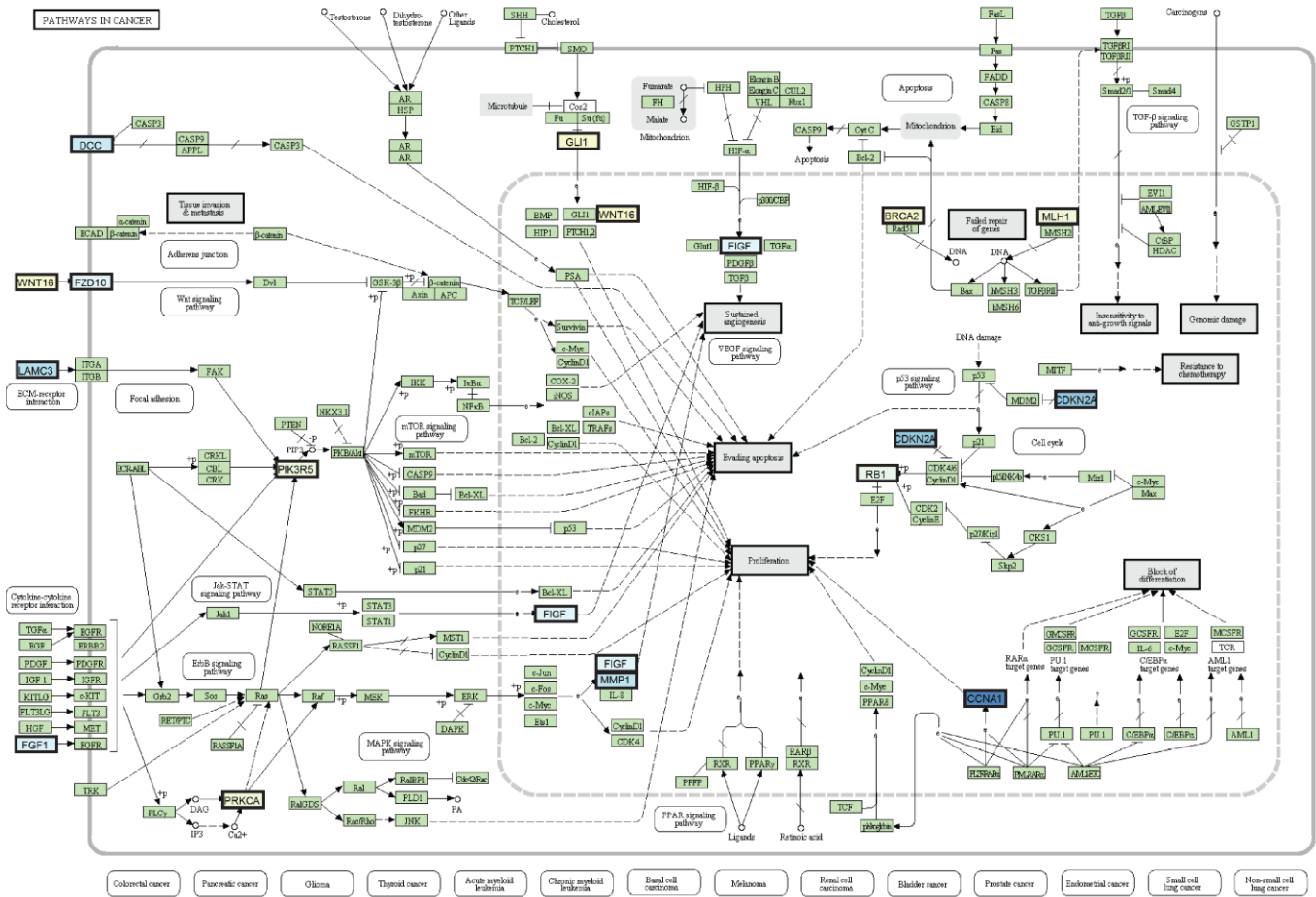
**Supplementary Figure 1.** Association of age-specific *hMLH1* promoter methylation with various lifestyle factors. **A)** Association between age-dependent *hMLH1* promoter methylation and lifestyle factors. Each point represents one biopsy.  $P$  values for the difference are derived from log-linear multivariable regression models (Supplementary Table 2). **B)** Association of *MGMT* promoter methylation with the occurrence of polyps.  $P$  values are derived from log-linear multivariable regression (Supplementary Table 4).



**Supplementary Figure 2.** Genome-wide association of age-specific DNA methylation with various lifestyle factors when stratified by colon location. Lifestyle modulated age-associated differentially methylated CpGs when stratified by colon location (**A**) proximal (**B**) distal. Numbers of samples tested in each category are indicated at the bottom of each bar. **C**) Effect of aspirin and HRT on number of DNA methylation changes in individuals with a normal BMI versus a high BMI and non-smokers versus smoker. Numbers of samples are indicated at the bottom of each bar. FDR adjusted  $P < .05$  was used as cutoff to identify significant differential methylated probes.  $P$  values to test differences in median rate of methylation change are calculated by Wilcoxon rank sum test.

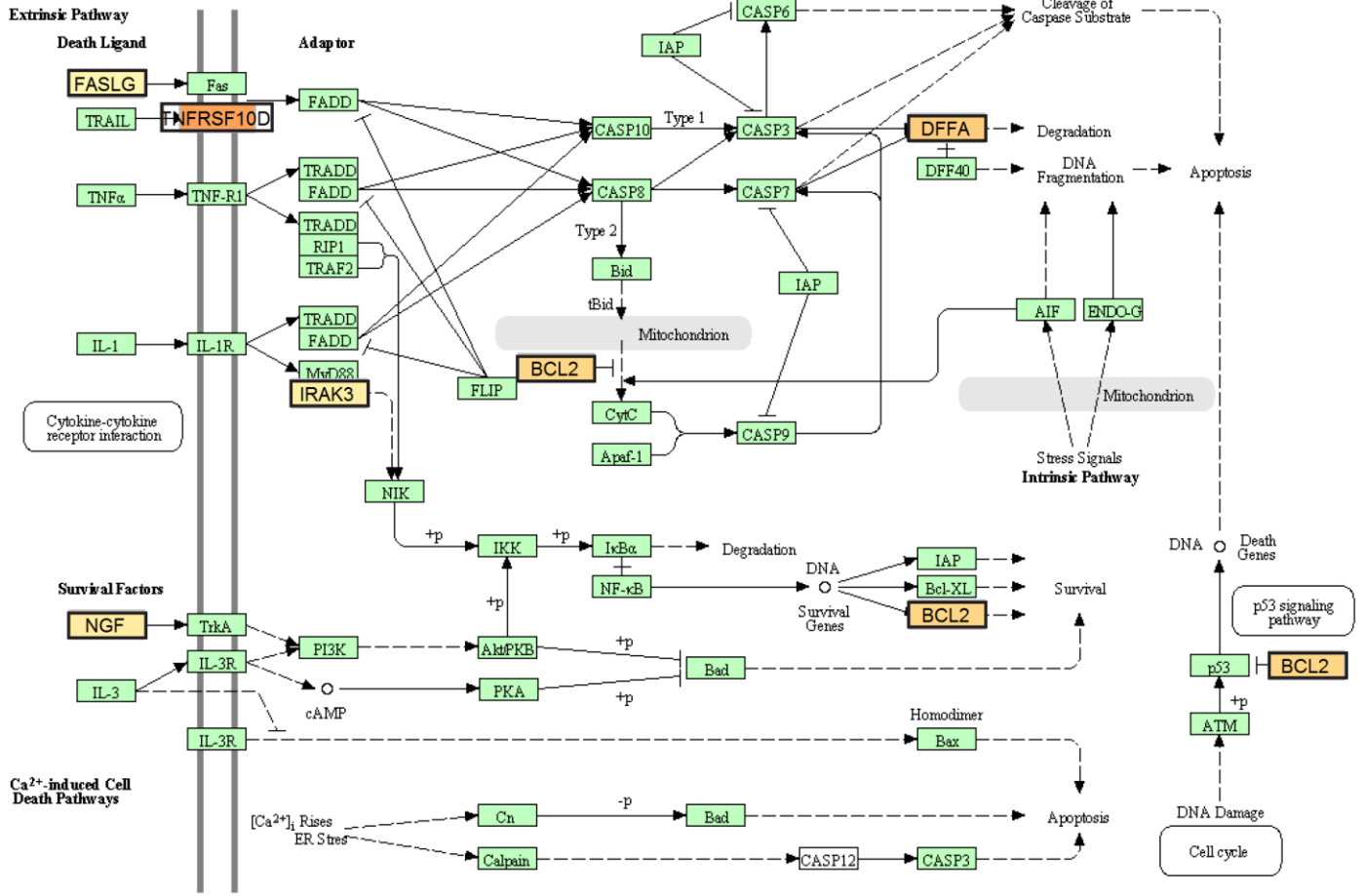


**Supplementary Figure 3.** KEGG cancer reference pathways associated with genes showed repressed methylation in aspirin user. Entries mapped to the pathways are outlined in bold. Color gradient (yellow to blue) for mapped entries represents degree of repressed methylation compared to aspirin non-user.



**Supplementary Figure 4.** KEGG cancer reference pathways associated with genes showed repressed methylation in HRT user. Entries mapped to the pathways are outlined in bold. Color gradient (yellow to blue) for mapped entries represents degree of repressed methylation compared to HRT non-users.

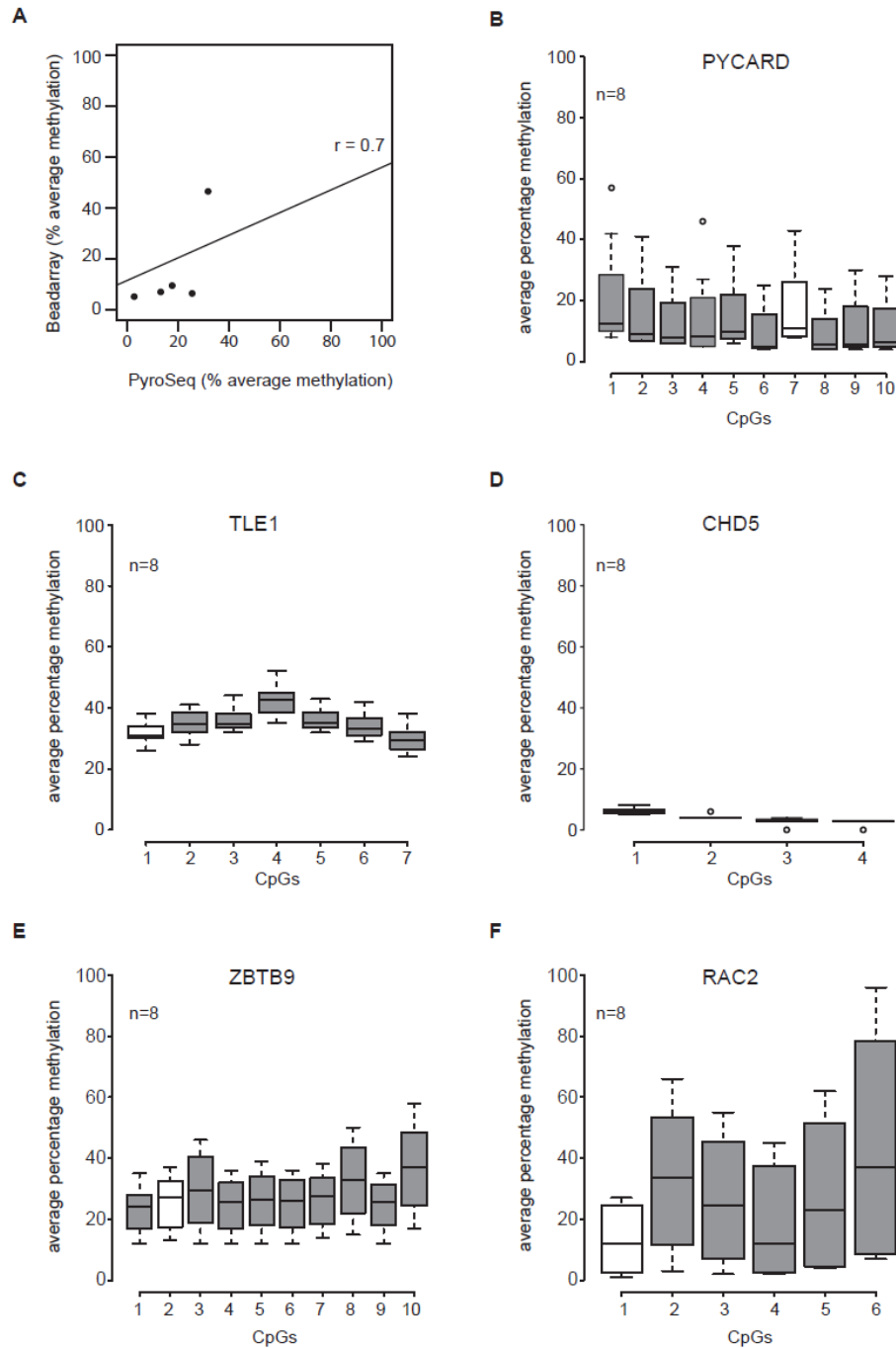
**APOPTOSIS**



**Supplementary Figure 5.** KEGG cancer reference pathways associated with genes showed hypermethylation in high BMI. Entries mapped to the pathways are outlined in bold. Color gradient (yellow to red) for mapped entries represents degree of hypermethylation over normal BMI.







**Supplementary Figure 7.** Pyrosequencing validation of DNA methylation measured by Illumina Infinium HumanMethylation27 BeadChip. **A)** Comparison between methylation values measured by array and pyrosequencing in 8 healthy tissue samples.  $r$  = Pearson correlation coefficient. **B-F)** Average percentage methylation of all CpG sites for each pyrosequencing assay. White boxes show the CpG of the pyrosequencing target that corresponds to the one that was measured by array.

**The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation**

Sébastien Loeffler<sup>1,2</sup>, Michal Kovac<sup>1,2</sup>, Stefan Weis<sup>1</sup>, Salvatore Piscuoglio<sup>1,2</sup>, Saulius Zuklys<sup>1,2</sup>, Marcel Keller<sup>1,2</sup>, Katrin Hafen<sup>1,2</sup>, Daniel Hess<sup>3</sup>, Kaspar Truninger<sup>1</sup>, Inti Zlobec<sup>4</sup>, Luigi Terracciano<sup>4</sup>, Karl Heinimann<sup>1,2</sup>, Primo Schär<sup>1</sup> and Georg A. Holländer<sup>1,2, 5</sup>

Affiliations

1: Department of Biomedicine - Mattenstrasse, University of Basel, 4058 Basel, Switzerland

2: Basel University Children's Hospital, 4058 Basel, Switzerland

3: Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland

4: Institute of Pathology, University of Basel, 4031 Basel, Switzerland

5: The Department of Paediatrics and The Weatherall Institute of Molecular Medicine, University of Oxford, Oxford, United Kingdom

6: Current address: Actelion Pharmaceuticals Ltd, Allschwil, Switzerland

Corresponding author: Georg A. Holländer ([georg-a.hollaender@unibas.ch](mailto:georg-a.hollaender@unibas.ch))

Tel: +41 61 695 30 69 / +44 1865234 225

## Highlights

- SHOCA-2 is a newly identified tumor suppressor in colorectal cancer (CRC)
- SHOCA-2 acts as a novel adaptor protein that binds PP1 $\beta$  to inactivate STAT3
- SHOCA-2 expression is lost in CRC due to genetic and epigenetic silencing
- Low SHOCA-2 expression alters CRC proliferation and correlates with poor prognosis

## Summary

Chromosomal deletions at 8p21.3 have been associated with cancers of various tissues, including the colon, but the tumor suppressor function affected has not been described. Here, we identify *SH2D4A* as a suppressor of colorectal cancer (CRC) mapping to 8p21.3. *SH2D4a* encodes SHOCA-2, a SH2 domain-containing adapter protein that, upon EGF signaling, interacts with the serine/threonine phosphatase PP1 $\beta$  and consequently inhibits STAT3 activity via dephosphorylation. Conversely, knock-down of SHOCA-2 stimulates c-Myc, Cyclin D1 and Jun B expression and accelerates CRC growth through unopposed STAT3 activity. Loss of SHOCA-2 expression in CRC correlates with advanced disease stages and poor prognosis and is caused by chromosomal rearrangements, point mutations and epigenetic changes, Thus, SHOCA-2 functions as an inhibitor of the EGFR pathway and its absence impacts on CRC development and progression.

## **Significance**

Colorectal cancer (CRC) yearly claims more than 655,000 lives worldwide and represents the third most frequent cause of death in the Western world. Though a 8p21-23 chromosomal deletion has been linked to CRC development and progression, the molecular mechanisms related to this genomic region have not been recognized to date. Encoded on 8p21.3, the SH2 domain-containing SHOCA-2 protein represents a hitherto unrecognized CRC tumor suppressor that limits cell proliferation via a negative feedback loop controlling EGFR signaling. Consequently, a loss of SHOCA expression results in the oncogenic activation of STAT3, a hallmark of many CRC. Since drugs inhibiting EGFR and STAT3 activity offer only limited efficiency, enhancing SHOCA-2 function constitutes a novel therapeutic target for the treatment of CRC.

## **Introduction**

Oncogene activation and loss of tumor suppressor activity are responsible for cancer initiation and progression (Vogelstein and Kinzler, 2004). Overexpression of the epithelial cell growth factor receptor (EGFR) is frequently observed in colorectal cancer (CRC), and correlates with a poor clinical outcome (Saif, 2010). EGFR is receptor tyrosine-kinase that relays its activation through the signal transducer and activator of transcription 3 (STAT3) to the nucleus controlling the transcription of genes involved in cell proliferation and survival (Aggarwal et al., 2009). Constitutive activation of the STAT3 signaling pathway as a consequence of genetic mutation is oncogenic, and a hallmark of many CRC (Bromberg et al., 1999; Klampfer, 2008).

Genetic and epigenetic alterations in the expression of tumor suppressor genes are typical features of neoplastic transformation and contribute to both initiation and progression of tumor formation (Weinberg, 1991; Vogelstein and Kinzler, 2004). The loss of heterozygosity (LOH) across variable parts of chromosome 8p has been identified as a frequent abnormality associated with CRC (Emi et al., 1992) and linked to DNA breakage at fragile sites located at 8p12 and 8p22 (Birnbaum et al., 2003). Although predicted from chromosome transfer studies (Tanaka et al., 1996), specific tumor suppressor genes have to date not been identified in this critical chromosomal region. It is thus unknown how genetic alterations in 8p contribute to CRC development and progression.

## Results

### Identification of SHOCA-2 and its loss of expression in colorectal cancer

cDNA Representational Difference Analysis (cDNA-RDA) comparing epithelial cells at two sequential stages of embryonic maturation identified transcripts of a previously uncharacterized gene, designated *SH2D4B* (NCBI gene ID 387694; S.Z. and G.H., unpublished data). Homology searches identified *SH2D4A* as a paralogue (NCBI gene ID 63898; Figure 1A). Both genes are evolutionary conserved (Figure S1A) and ubiquitously expressed, with *SH2D4A* being significantly more abundant than *SH2D4B* (Figure S1B). The proteins encoded by *SH2D4A* and *SH2D4B* display 65% amino acid sequence identity at their C- and N-termini and share several structural motifs including three coiled-coil domains in the middle of the molecule and a Src homology (SH2) domain at the C-terminal end (Figure 1A; Figure S1C). It is for these motifs and the putative function as an adaptor that we named this family of proteins SHOCA, an acronym for SH2-domain containing adaptor protein.

*SH2D4A* maps to human chromosome 8p21.3 (Figure 1A) and encodes SHOCA-2. Because the chromosomal region surrounding *SH2D4A* (8p21-23) is frequently lost in CRC and other epithelial tumors (Emi et al., 1992), we determined by immunohistochemistry the expression of SHOCA-2 in 400 consecutive, sporadic CRCs (Figure S1D). The revealed significant differences in detectable SHOCA-2 protein, correlating with clinical tumor stage (I-IV) (Figures 1B and 1C; Table S1A and S1B); SHOCA-2 expression was reduced or a lost of with advanced CRC pathology. Patients with tumors displaying either low or absent SHOCA-2 expression had a poorer outcome (log-rank test  $p=0.0118$ ; Figure 1D; Table S1C) though SHOCA-2 expression, when stratified by disease stage, did not serve on its own as an independent predictor of survival. An identical conclusion was drawn from a Cox regression analysis ( $p=0.0134$  with HR (95%CI)=0.56 (0.35-0.89) and a Wilcoxon test ( $p=0.0133$ ) (Table S2). Gene expression data available in public databases for carcinomas of different origin including bladder, breast, head and neck, esophagus, liver and ovary suggested

similar correlations between disease stage and SHOCA-2 expression (Figure S1E). Finally, further analysis of SHOCA-1 expression (mRNA levels) in CRC samples with absent or reduced SHOCA-2 expression revealed a concurrent loss of SHOCA-1 in approximately half of the samples whereas SHOCA-1 loss was neither observed in healthy mucosa nor in CRC tissues proficient for SHOCA-2 (Figure S1F).

### **Genetic and epigenetic alterations of the *SH2D4A* locus in CRC**

To determine the molecular basis for decreased SHOCA-2 expression, we chose from an unselected cohort of 70 CRC patients 27 subjects with *SH2D4A* alleles that could be distinguished by microsatellite markers and SNPs (Figure S1D). In 17 of these (63%), the primary tumor had lost or diminished SHOCA-2 expression whereas only 3 samples (11%) displayed an *EGFR* gene amplification. A loss of the *SHOCA-2* gene and a simultaneous amplification of *EGFR* were, however, noted in 2 out of these 3 individuals. This result is in accordance with the observation that *EGFR* over-expression in CRC is rarely caused by *EGFR* gene amplification (Saif, 2010). Using 3 flanking microsatellite markers and 6 gene-specific SNPs (Figure S2A), we detected LOH at *SH2D4A* in 7 of the 17 tumors (41%). Gene dosage quantification of the short arm of chromosome 8 indicated mono- and biallelic deletions in 6 and 1 of these LOH tumors, respectively (Figure 2A). When normal mucosa was examined, four of the six patients with a monoallelic *SH2D4A* loss in tumor tissue were heterozygous for the intronic SNP rs17128221 (c.342-5T>C). This SNP is located in a mRNA splice acceptor site and causes non-canonical exon 4 skipping, which in turn causes a translational frame shift and premature termination in exon 5 (Figure 2B). Consistent with a pathological effect, the cancer tissue of patients heterozygous for the C allele uniformly demonstrated a selective loss of the T allele (Chi square  $p=0.006$ ) whereas homozygosity for the C allele was never observed in mucosa tissue samples of 83 healthy Caucasian individuals (Figure 2B).

Missense mutations and deletions in the SH2 domain of SHOCA-2 were observed in three of the 10 *SH2D4A* biallelic CRC samples characterized by a low or absent SHOCA-2 expression. These alterations either produced single amino acid changes (p.Arg324Trp and p.Ser430Phe) with predicted structural effects (Polyphen-2 score of >0.95, SIFT score<0.02), or caused a translational frame shift (p.Ile378fsX15) giving rise to a loss of amino acids 377 to 454 and reduced protein expression (Figure 2C).

To determine causes for low or absent SHOCA-2 expression other than LOH and gene mutations, we next analyzed primary CRC tissue and surrounding healthy mucosa for DNA cytosine methylation in the 5' untranslated region (5'UTR) of *SH2D4A* (Figure 2D). Pyrosequencing of bisulfite-converted genomic DNA (Bettstetter et al., 2007) revealed two specific CpG dinucleotides showing increased methylation in cancer tissue when compared to matched normal mucosa. One of these CpGs is within a canonical Sp1 transcription factor recognition motif and its methylation significantly reduced the Sp1 binding affinity (Figure S2B). Epigenetic changes at the DNA level may therefore represent another cause for reduced *SH2D4A* gene expression in CRC. Taken together, in 15 of 17 CRCs investigated in molecular detail (Figure 2E), distinct genetic and epigenetic alterations explaining the loss or reduction in SHOCA-2 expression could be identified; chromosomal deletions at 8p21.3, genetic mutations in the *SH2D4A* gene, a mRNA splicing defect associated with a specific SNP, and site-specific changes in promoter DNA methylation.

### **SHOCA-2 negatively regulates EGF-induced STAT3 phosphorylation and transcriptional activity**

A proteomic analysis of phospho-tyrosine signaling in non-small-cell lung cancer cell lines driven by EGFR-activating mutations identified a target peptide with sequence homology to SHOCA-2 (Guo et al., 2008). We confirmed that EGFR activation triggered tyrosine phosphorylation of SHOCA-2 in HeLa cells (Figure 3A). To investigating a potential involvement of SHOCA-2 in the activation of the



downstream EGFR effectors STAT3, ERK and PI3K-Akt, we stimulated HeLa cells with EGF. The ectopic expression of SHOCA-2 in these cells specifically decreased the STAT3 phosphorylation at Tyr705 (Figure 3B). The phosphorylation of this residue constitutes an important post-translational modification of STAT3 as it triggers its dimerization, nuclear translocation and DNA binding (Heinrich et al., 2003). By contrast, the relative level of Akt phosphorylation was increased whereas the MAP kinase pathway remained unaffected in EGF simulated HeLa cells (Figure S3A). Although SHOCA-2 had been previously suggested to be associated with the ER alpha/PKC signaling pathway (Li et al., 2009), we were unable in several independent attempts to observe in HeLa cells overexpressing SHOCA-2 a change in PKC phosphorylation (Figure S3B).

To analyze EGF signaling in cells with reduced SHOCA-2 levels, we generated short hairpin RNA (shRNA) expressing HeLa cell clones in which SHOCA-2 protein was knocked down by 90% (SHOCA-2 KD; Figure 3C). When stimulated with EGF, the extent of STAT3 phosphorylation at tyrosine 705 (Tyr<sup>705</sup>) and serine 727 (Ser<sup>727</sup>) was inversely correlated with the SHOCA-2 protein level (Figure 3C). In keeping with a significant increase in phosphorylated STAT3, the transcription of the c-MYC and Cyclin D1 was enhanced (see below). In contrast, the phosphorylation of AKT was decreased while that of ERK 1/2 remained unchanged in SHOCA-2 KD cells (Figure S3C). These results further corroborated the specific involvement of SHOCA-2 in STAT3-mediated signaling following EGF stimulation.

Using a STAT3-driven luciferase reporter assay (Kreis et al., 2007), we next examined in HeLa cells whether STAT3 transcriptional activity was altered in the presence of decreased or increased SHOCA-2 protein levels. As demonstrated in Figure 3D, EGF treatment of HeLa cells with a significant loss of SHOCA-2 expression (SHOCA-2 KD cells) resulted in a robust transcriptional activity. In contrast, the transcription of the luciferase reporter was significantly reduced in HeLa cells overexpressing SHOCA-2 and exposed to EGF (Figure 3D; Figure

S3D). To investigate the nature of SHOCA-2's influence on STAT3, we also tested porcine aortic endothelial cells (PAE) that lack endogenous EGFR expression and mouse embryonic fibroblasts (MEF) rendered deficient for STAT3 expression (Huang et al., 2007; Costa-Pereira et al., 2002). In these cells, SHOCA-2 bound to STAT3 and EGFR and this interaction was always independent of the missing binding partner normally present in wild type cells (Figure 3E, Figure S3E). Moreover, the binding of SHOCA-2 to EGFR and STAT3 was unrelated to its SH2 domain (Figure S3F) and isoform-specific because SHOCA-1 did not physically bind to STAT3 (Figure S3G). Taken together, these results suggest that SHOCA-2 inhibits EGF-induced STAT3 activation through its physical association with both EGFR and STAT3 independently of its SH2 domain.

### **SHOCA paralogues bind PP1 $\beta$**

Using tandem affinity purification (TAP) and reciprocal immunoprecipitation, we and others (Ewing et al., 2007) identified the serine/threonine Protein Phosphatase-1 beta (PP1 $\beta$ ) as an interaction partner of SHOCA-1 and SHOCA-2 (Figure 4A; Figures S4A-C; Table S2). The eukaryotic PP1 protein family is composed of the 3 isoforms  $\alpha$ ,  $\beta$  and  $\gamma$ , whose functionality relies on the association with different regulatory proteins (Ceulemans and Bollen, 2004). Both SHOCA paralogues bound PP1 $\beta$  but neither PP1 $\alpha$  or PP1 $\gamma$ , demonstrating an isoform-specific interaction (Figure 4B). Confocal immunofluorescence microscopy and immunoprecipitation of PP1 $\beta$  from cytoplasmic and nuclear protein fractions detected the SHOCA-2-PP1 $\beta$  interaction in both subcellular compartments further confirming their physical association (Figure S4D).

The [RK]-X(0,1)-[VI]-{P}-[FW] sequence has previously been identified as a PP1-interacting motif (Hendrickx et al., 2009). We therefore tested whether the KXILF and KX[VI][QH]W sequences located in the N-terminus of SHOCA-1 and -2 may mediate such binding (Figure 4C left panel; Figure S4E). Site-directed mutagenesis was used to alter in the SHOCA isoforms the first motif from KXILF to AXALA (designated PP1<sub>Mut1</sub>) and the second sequence from KX[VI][QH]W to

AXA[QH]A (PP1<sub>Mut2</sub>; Figure 4C). Overexpression of SHOCA mutant proteins in HEK293 cells demonstrated that only the PP1<sub>Mut2</sub> proteins failed to interact with PP1 $\beta$  identifying the KX[VI][QH]W sequence as the relevant docking site (Figure 4C, middle and right panels).

### **STAT3 inhibition by SHOCA-2 requires EGF-induced phosphorylation of Tyr131 and the recruitment of PP1 $\beta$**

We next investigated whether EGFR and/or STAT3 could pre-associate with SHOCA-2 independently of any EGFR stimulation, or, alternatively, whether these interactions required an activation-induced phosphorylation of the individual binding partners. Immunoprecipitation experiments showed that SHOCA-2, STAT3 and PP1 $\beta$  co-precipitated in serum starved, unstimulated HeLa cells that was independent of prior EGF stimulation. To further characterize this interaction, unstimulated HeLa cells were transfected with mutant forms of either EGFR or STAT3. The EGFR mutant had its Tyr<sup>1068</sup> and Tyr<sup>1086</sup> replaced by phenylalanine (EGFR Y1068/1086F) and hence lost a motif known for classical STAT3 recruitment (Shao et al., 2003), whereas Tyr<sup>705</sup> was replaced in the STAT3 mutant to a phenylalanine thus eliminating several essential functions (Heinrich et al., 2003). Neither the association of mutant EGFR with wild type STAT3 nor the interaction of altered STAT3 with EGFR were affected by these changes (Figure S4F), thus demonstrating that EGFR's Tyr<sup>1068</sup>/Tyr<sup>1086</sup> and STAT3's Tyr<sup>705</sup> were not critically important for the interaction with each other (Figure S4F). The latter finding is in accordance with a pTyr<sup>705</sup> independent recruitment of STAT3 to the IL-22 receptor, an interaction that however requires the coiled-coil domain of STAT3 (Dumoutier et al., 2009). Following EGF activation, the association of SHOCA-2, STAT3 and PP1 $\beta$  significantly increased in HeLa cells (Figure 4D) demonstrating the formation of additional complexes with a possibly different composition of their signaling components. Although STAT3 can be indirectly activated through c-SRC (Quesnelle et al., 2007), c-SRC was not associated with the SHOCA-2, STAT3 and PP1 $\beta$  complex excluding a role of c-SRC in the inhibition of STAT3 by SHOCA-2 (Figure S4G).

Because EGF-stimulation resulted in Tyr<sup>131</sup> phosphorylation of SHOCA-2 (Figure 3A and (Guo et al., 2008)), we next examined whether this modification is essential for the formation of the SHOCA-2/EGFR/STAT3/PP1 $\beta$  complex and its function. Though changing the tyrosine residue to alanine (Y131A) did not disturb SHOCA-2's ability to complex with EGFR, STAT3 and PP1 $\beta$  (Figure S4H), the point mutation resulted in a constitutive phosphorylation of other SHOCA-2 tyrosine residues and an insensitivity to EGFR signal-mediated inhibition of STAT3 tyrosine phosphorylation (Figure S4I). The regulation of STAT3 activity by EGFR is therefore critically dependent on the phosphorylation of SHOCA-2's Tyr 131.

We next characterized in further detail the conditions under which SHOCA-2 can bind to its partners. Testing a SHOCA-2 mutant unable to associate with PP1 $\beta$  (SHOCA-2-PP1<sub>Mut2</sub>), did not require an association with PP1 $\beta$  □□ SHOCA-2 bind to the EGFR/STAT3 complex (Figure S4H). However, the inactivation of STAT3 was dependent on SHOCA-2's ability to associate with PP1 $\beta$  as the overexpression of SHOCA-2 PP1<sub>Mut2</sub> failed to alter STAT3 Tyr<sup>705</sup> phosphorylation following EGF stimulation (Figure 4E). Consequently, STAT3-dependent transcription in a reporter assay was not diminished in HeLa cells overexpressing SHOCA-2 PP1<sub>Mut2</sub> that had been treated with EGF (Figure 4F). SHOCA-2 thus binds to the EGFR/STAT3 complex independently of PP1 $\beta$ , but requires the association with the phosphatase to modulate STAT3 activity.

### **SHOCA-2 requires PP1 $\beta$ to inhibit STAT3-dependent tumor cell growth**

To address the contribution of SHOCA-2 to cellular growth control, we assessed the expression of STAT3-controlled cell proliferation genes (Aggarwal et al., 2009; Trenerry et al., 2007) in HeLa cells expressing reduced levels of SHOCA-2 (SHOCA-2 KD). The increase of c-Myc, Cyclin D1 and Jun B transcripts in these cells upon EGF stimulation inversely correlated with a robust reduction in SHOCA-2 expression (Figure 5A, left panel). Consequently, the fraction of cells in G2/M of the cell cycle was significantly higher in both unstimulated (8.8% vs. 5% G2/M cells

[ $p < 0.0002$ ]) and EGF activated SHOCA-2 KD HeLa cells (21.2% vs. 9.4% G2/M cells, [ $p < 0.004$ ]) when compared to wild type controls (Figure 5A, right panel). A significantly increased cell proliferation (Ki-67 positivity) and higher levels of c-Myc, Jun B, and Cyclin D1 transcripts were also detected in tissue sections of clinically advanced CRC further confirming an inverse correlation between SHOCA-2 expression and STAT3 mediated cellular responses (Figure 5B; Figure S5A). A decrease in SHOCA-2 expression therefore promotes the transcription of cell proliferation factors that in term drive cell cycle progression.

To determine the molecular mechanism by which SHOCA-2 suppresses cell proliferation, we next transfected the colorectal cancer cell line SW480 to over-express STAT3 alone or in combination with either wild-type or mutant SHOCA-2. In a colony formation assay, overexpression of only STAT3 significantly increased the number of colony forming units (Figure 5C), a response that correlated with a higher degree of STAT3 phosphorylation (Figure 5D). Overexpression of STAT3 together with wildtype SHOCA-2 significantly reduced the colony forming potential of SW480 cells as well as STAT3 phosphorylation. This growth suppressive effect was less pronounced when SHOCA-2 mutants deficient in either PP1 $\beta$  interaction (SHOCA-2 PP1<sub>Mut2</sub>) or Tyr<sup>131</sup> (Y131A) phosphorylation were co-overexpressed instead (Figure 5C). The extent to which the number of colonies could be decreased as a consequence of SHOCA-2 PP1<sub>Mut2</sub> expression was, however, limited by the continued presence of endogenous (i.e. wild type) SHOCA-2 thus effecting a competitive condition. Comparable results using the overexpression of wild type and mutant SHOCA-2 were obtained in the adenocarcinoma cell line H1975 which expresses an EGFR mutation constitutively activating STAT3 (Figure S5B) (Lu et al., 2007). Consistent with a role of SHOCA-2 in suppressing STAT3-mediated cell growth, SW480 cells in which SHOCA-2 expression was knocked down by shRNA diminished their proliferation rate upon exposure to pharmacological inhibition of STAT3 by S31-201, a chemical probe blocking STAT3-STAT3 complex formation and STAT3-DNA binding (Figure S5C). Finally, overexpression of SHOCA-2 in a CRC cell line (SW620) with constitutive STAT3

phosphorylation (Maa et al., 2007) and spontaneously low endogenous SHOCA-2 expression (Figure S5D) suppressed anchorage-independent cell growth (Figure 5E). Taken together, these results demonstrate under different experimental conditions that SHOCA-2 acts as a suppressor of STAT3-driven cancer cell proliferation.

### **Downregulation of SHOCA-2 alters the cell phenotype and promotes *in vivo* tumor growth**

The cobble-stone shape of SW480 cells, which usually express high endogenous SHOCA-2 levels, changed to a spindle-like morphology following the knock-down of SHOCA-2 transcripts (Figure 6A, Figures S5D and S6A). This change in cell shape was paralleled by features of epithelial-mesenchymal transition (EMT) (Christofori, 2006) such as a reduction in E-cadherin and an increase in N-cadherin, vimentin, SNAIL and ZEB1 expression (Figure 6B). These alterations were, however blocked by the inhibition of EGFR using Tyrphostin (AG1478; Figure S6B). Moreover, SHOCA-2 deficient SW480 cells proliferated at a greater rate and displayed a higher sensitivity to EGFR inhibition when compared to mock transfected controls (Figure 6C). Re-establishing mouse Shoca-2 expression in human SHOCA-2 deficient SW480 cells however decreased their proliferation rate (Figure 6D), thus demonstrating again that SHOCA-2 controls STAT3-mediated tumor growth.

To establish an *in vivo* role for SHOCA-2 as a tumor suppressor, SW480 cells that had their SHOCA-2 expression either knocked-down or left unchanged were transplanted into nude mice. Grafts with cells where SHOCA-2 was reduced displayed a significant increase both in tumor incidence and size when compared to control transplants (Figure 6E). Moreover, significant levels of phosphorylated, nuclear STAT3 were detected in tumors that had emerged from grafted cells lacking regular SHOCA-2 expression (Figure 6F, upper left panel). Consistent with this experimental finding, increased STAT3 phosphorylation at residues Tyr<sup>705</sup> and Ser<sup>727</sup> (Figure 6F, upper right and lower panels) were also detected in the majority

of CRC biopsies with low SHOCA-2 expression (H-score <100). Taken together, experimental *in vivo* tumor models using CRC cell lines and the detailed analysis of biopsies taken from CRC patients identified the inhibitory role of SHOCA-2 on STAT3 activation.

## Discussion

Genomic deletions within the 8p21-23 region constitute a characteristic cytogenetic feature of CRC (Emi et al., 1992) and have been linked to carcinogenesis (Macartney-Coxson et al., 2008). How deletions in this chromosomal region contribute to CRC development has remained unclear since tumor suppressor genes could so far not be allocated to 8p21-23 (Macartney-Coxson et al., 2008). With the assignment of the SHOCA-2 encoding *SH2D4A* gene to 8p21.3, we have linked genomic instability at this locus with the loss of a specific tumor suppressor activity in CRC. We provide experimental evidence and clinical data that a loss of SHOCA-2 expression results in unopposed STAT3 activation following EGFR stimulation and marks in patients with colorectal cancer accelerated tumor growth as well as poor prognosis. Our findings of a role for SHOCA-2 in the control of tumor growth also extends to malignancies where the mechanism of tumor development and progression are still only poorly understood (Roessler et al., 2011).

EGFR is an upstream activator of multiple pathways involved in cell proliferation, apoptosis and carcinogenesis. Overexpression of EGFR and constitutive activation of its major downstream effector, STAT3, have been linked to cancer progression, a higher risk for metastasis and reduced survival (Quesnelle et al., 2007; Klampfer, 2008; Saif, 2010). Since STAT3 regulates cell growth and tissue homeostasis (Aggarwal et al., 2009), its activation must be under stringent control. We found that SHOCA-2 cooperates with PP1 $\beta$  in modulating the EGFR-induced STAT3 activation, thereby limiting cell proliferation. In view of the specific genetic and epigenetic alterations of *SH2D4A*, that particularly emerge in late stage CRCs, and given the correlation between poor clinical outcome and the loss of SHOCA-2 expression, we conclude that SHOCA-2 is a novel tumor suppressor modulating the activation of the EGFR signaling pathway (Walther et al., 2009). However, in the cohort studied here, the SHOCA-2 status was not a reliable predictor of survival in patients stratified by disease stage. This result is not surprising as it is consistent



with the predictive potential of other tumor suppressors. For example, p53, APC, and DCC also fail to serve as prognostic markers though their loss of expression in colorectal cancer is correlated with increased tumor aggressiveness (Roth, 1999; Walther et al., 2009).

Based on the evidence that SHOCA-2 binds to both EGFR and phosphatase PP1 $\beta$  and is itself phosphorylated upon EGF signaling, we propose that SHOCA-2 controls STAT3 activity through the catalytic activity of PP1 $\beta$ . SHOCA-2 and PP1 $\beta$  form a complex with STAT3 in that SHOCA-2 acts as an adaptor to bring the PP1 $\beta$  phosphatase into physical proximity of STAT3, which in turn effects the dephosphorylation of activated STAT3. Indeed, phosphorylation/dephosphorylation of Ser<sup>727</sup> plays an important role in modulating STAT3's transcriptional activity (Levy and Darnell, 2002; Wen et al., 1995; Shen et al., 2004), though the molecular mechanisms underlying Ser<sup>727</sup> dephosphorylation have not yet been fully delineated. Though the serine/threonine phosphatases PP1 and PP2A have already been identified to dephosphorylate Ser<sup>727</sup> (Lütticken et al., 1995; Woetmann et al., 1999; Togi et al., 2009; Haridas et al., 2009), the regulatory subunit that confers selectivity, specificity, and subcellular localization of PP1 towards STAT3 has so far and contrary to PP2A (Togi et al., 2009) remained unidentified. For the EGF pathway, we now demonstrate that Ser<sup>727</sup> phosphorylation of STAT3 decreases on the condition that PP1 $\beta$  associates with activated SHOCA-2.

Dephosphorylation of Tyr<sup>705</sup> constitutes another critical step in the events leading to STAT3 deactivation (Heinrich et al., 2003). One of the phosphatases involved in this event is the receptor protein tyrosine phosphatase delta (PTPRD), which has been shown to act as a tumor suppressor in both colon cancer and other neoplasms (Zhang et al., 2007; Veeriah et al., 2009). Though PP1 $\beta$  constitutes yet another phosphatase involved in STAT3 inactivation, it remains unknown whether it acts indirectly via a tyrosine-specific phosphatase (e.g. SHP (Neviani et al., 2005)), or directly through target promiscuity (MacKintosh et al., 1996; Shi, 2009).

A dual specificity of PP1 $\beta$  catalytic activity is not unparalleled as a comparable mechanism has been described for the protein tyrosine phosphatase SHP-2 which catalyzes the dephosphorylation of both Tyr<sup>701</sup> and Tyr<sup>727</sup> residues in STAT1 (Wu et al., 2002). Alternatively, the enhanced physical interaction between PP1 $\beta$ , SHOCA-2 and STAT3 following EGF binding to its receptor may be sufficient to inactivate STAT3 secondary to conformational changes independent of any catalytic activity (Lee et al., 2009). The precise molecular mechanism by which Tyr<sup>705</sup> is de-phosphorylated remains to date unidentified. Irrespective of this lack of knowledge, inappropriate STAT3 regulation is the noticeable consequence of aberrant SHOCA-2 expression and results in an up-regulation of genes involved in cell proliferation.

The reduction of SHOCA-2 expression in the CRC cell line SW480 also produced phenotypic features characteristic of EMT. Such a transformation has been implicated in the conversion of early stage malignancies towards more aggressive tumors showing invasive growth and the formation of metastasis (Thiery et al., 2009). Consistent with a role in EMT, LOH at *SH2D4A* was mostly detected in patients with a disease stage typically associated with metastasis. It is thus conceivable that the loss of SHOCA-2 plays a part in tumor progression not only through its impact on tumor cell growth but possibly also via its contribution to EMT.

EGFR and STAT3 constitute promising drug targets in the treatment of cancer including CRC (Quesnelle et al., 2007). However, therapeutic responses are observed in only 10 to 20% of patients treated with EGFR antagonists (Ciardiello and Tortora, 2008) and none of the available STAT3 inhibitors can to date be considered as viable drug candidates because of their limited efficiency in disrupting STAT3 homodimerization (Yue and Turkson, 2009). SHOCA-2 may therefore constitute a novel therapeutic target for the design of strategies to interfere rationally with a STAT3-mediated activation in cancer cells.

## **Experimental Procedures**

A detailed description of the methods used can be found in Supplemental Experimental Procedures.

### **Plasmids, antibodies, cells and reagents.**

Antibodies, expression plasmids, cells and reagents were either gifts from investigators or bought commercially, as specified in Extended Experimental Procedures.

### **Bioinformatic analysis.**

The SHOCA-1 (*SH2D4B*) and SHOCA-2 (*SH2D4A*) sequences were analyzed using NCBI, ENSEMBL (Hubbard et al., 2009), the kinBase database (<http://www.kinase.com/kinbase/>), the TCOFFEE software (Notredame et al., 2000), and OncoPrint (Rhodes et al., 2007).

### **TAP purification and mass spectrometry.**

The TAP purification (Chen and Gingras, 2007) and mass spectrometry (Hess et al., 2008) were carried out as previously reported.

### **Western Blot, immunoprecipitation and cell fractionation.**

Cells lysates were either separated on a 8%-12% SDS-PAGE or immunoprecipitated with protein G Plus-Sepharose beads coupled to specified antibodies and analyzed by western blotting. Cell fractionation was achieved using the CellLytic Nuclear extraction kit (Sigma).

### **Luciferase reporter gene assays.**

HeLa cells were transfected with a firefly luciferase reporter SIE (sis-inducible element) plasmid plus a Renilla luciferase plasmid (both Promega) and plasmids that encode either STAT3 or SHOCA-2. Cell extracts were tested using the Dual

Luciferase Assay System (Promega) and analyzed using a luminometer. Firefly luciferase activity was normalized to Renilla luciferase activity.

### **Cell cycle analysis.**

Cells were incubated for 15 min with ice-cold hypotonic Propidium Iodide staining solution and subsequently analyzed by flow cytometry (FACSCalibur, Becton Dickinson) using the FlowJo software (Tree Star, Oregon Corporation). G1, S and G2/M phases were defined using the mathematical Watson Pragmatic model.

### **Colony formation and anchorage-independent growth assays.**

For colony formation assay, SW480 cells co-transfected with a control vector or plasmids that encode STAT3, SHOCA-2, SHOCA-2 PP1<sup>Mut2</sup> or SHOCA-2 Y131A were cultured for 15 days under neomycin selection (G418: 1 mg/ml) and then analyzed for colony frequency following fixation and crystal violet staining of the cells. SW620 cells transfected with a control vector or plasmids that encode SHOCA-2, SHOCA-2 PP1<sup>Mut2</sup> or SHOCA-2 Y131A, cultured under neomycin selection (G418, 1 mg/ml) in a soft agar (CytoSelect, Cell Biolabs) for 15 days and then analyzed for colony frequency.

### ***In vivo* analyses of tumor formation.**

SW480 cells ( $5 \times 10^6$ ) were grafted to the flank of 6 week-old female BALB/c nude mice and the size of the tumor was measured weekly. Mice were sacrificed after 5 weeks and tumors were removed for further analysis. The animal experiments were carried out in accordance with the legal requirements of the Swiss veterinary authorities.

### **Human tissues and immunohistochemistry.**

For the genetic and epigenetic study, specimens from 70 CRC patients were obtained from the Department of Surgery of the Kantonsspital Aarau, Switzerland and from of the Department of Gastroenterology of the Inselspital in Berne, Switzerland, with approval of the local medical ethics boards and written consent

from patients. Tissues were stored in RNA*later* reagent at -70°C. Single-punch tumor samples from 501 patients (Institute of Pathology, Stadtspital Triemli, Zürich, Switzerland) were analyzed by tissue microarrays as described previously (Zlobec et al., 2010). SHOCA-2 expression and STAT3 phosphorylation were quantified using the H score (Hirsch et al., 2003).

### **Mutational screening, loss of heterozygosity and gene dosage analysis.**

The complete coding sequence including flanking intronic regions of the *SH2D4A* gene was analyzed in all cancers lacking SHOCA-2 protein expression and/or that were positive for LOH. LOH for the *SH2D4A* locus was determined using a set of 3 flanking microsatellite markers and 6 gene-specific single nucleotide polymorphisms (SNP) (for details see Extended Experimental Procedures). Gene dosage was determined for each cancer using a chromosome 8-specific multiplex ligation-dependent probe amplification (MLPA) assay (SALSA P014-1A, MRC Holland) and data were processed by the GeneMarker software package (Softgenetics).

### **Methylation analysis.**

Genomic DNA was converted by bisulfite treatment using the EZ DNA Methylation Kit™ according to the manufacturer's instructions (Zymo Research). CpG island regions were amplified by PCR from bisulfite treated DNA and sequenced with the PyroMark Q24 pyrosequencing system (Qiagen) allowing the quantitation of methylated CpG sites (PCR conditions and primer sequences are given in Extended Experimental Procedures). Bisulfite treated genomic DNA, previously *in vitro* methylated with M.SssI methyltransferase (New England BioLabs), served as the positive control.

### **Statistical analyses.**

Student's t test, Likelihood ratio test, Chi-square test, Gehran-Wilcoxon test, Fisher's Exact test, binomial test for equal proportions, Log-rank test, multivariate

Cox regression analysis and Kaplan-Meier method were used for statistical analyzes, as indicated.

## Acknowledgements

We would like to thank Annick Peter, Dorothea Maass and Ragna Sack for expert technical help; Dr. Giancarlo Marra for microarray data; Dr. Fabrizio Bianchi for statistical analyzes; Dr. Bérengère Fayard and Dr. Laura Trinkle-Mulcahy for valuable discussions; Drs. Nancy Hynes, Thomas Barthlott, Poul Sørensen and Gerhard Christofori for critical reading of the manuscript; Sabrina Harris for secretarial assistance.

This work was supported by grants from the Swiss National Science Foundation (to G.A.H.), the Krebsliga beider Basel (to S.L. and G.A.H.) and the Oxford Biomedical Research Centre (to G.A.H.).

S.L. and G.A.H. designed the experiments, analyzed the data and composed together with P.S. the manuscript; S.L. performed experiments including Western Blot, immunoprecipitation, luciferase assays, mutagenesis, TAP assay, phosphatase assay, bioinformatic analyses, colony formation and anchorage-independent growth assays, and *in vivo* experiments; S.Z., K.H. and M.Ke. participated in the identification and cloning of *SH2D4B* and *SH2D4A* genes; D.H. performed the mass spectrometry analysis; supervised by P.S. and K.H., M.Ko. and S.W. performed the human genetic and epigenetic studies; immunohistochemistry was done by S.P.; K.T. provided human CRC samples; I.Z. performed statistical analyzes; L.T. supplied and analyzed CRC tissue microarray. The authors declare no competing financial interests.

## References

- Aggarwal, B.B., Kunnumakkara, A.B., Harikumar, K.B., Gupta, S.R., Tharakan, S.T., Koca, C., Dey, S., and Sung, B. (2009). Signal transducer and activator of transcription-3, inflammation, and cancer: how intimate is the relationship? *Ann N Y Acad Sci* 1171, 59-76.
- Bettstetter, M., Dechant, S., Ruemmele, P., Grabowski, M., Keller, G., Holinski-Feder, E., Hartmann, A., Hofstaedter, F., and Dietmaier, W. (2007). Distinction of hereditary nonpolyposis colorectal cancer and sporadic microsatellite-unstable colorectal cancer through quantification of MLH1 methylation by real-time PCR. *Clin Cancer Res* 13, 3221-28.
- Birnbaum, D., Adélaïde, J., Popovici, C., Charafe-Jauffret, E., Mozziconacci, M.J., and Chaffanet, M. (2003). Chromosome arm 8p and cancer: a fragile hypothesis. *Lancet Oncol* 4, 639-642.
- Bromberg, J.F., Wrzeszczynska, M.H., Devgan, G., Zhao, Y., Pestell, R.G., Albanese, C., and Darnell, J.E. (1999). Stat3 as an oncogene. *Cell* 98, 295-303.
- Ceulemans, H., and Bollen, M. (2004). Functional diversity of protein phosphatase-1, a cellular economizer and reset button. *Physiol Rev* 84, 1-39.
- Chen, G.I., and Gingras, A.C. (2007). Affinity-purification mass spectrometry (AP-MS) of serine/threonine phosphatases. *Methods* 42, 298-305.
- Christofori, G. (2006). New signals from the invasive front. *Nature* 441, 444-450.
- Ciardiello, F., and Tortora, G. (2008). EGFR antagonists in cancer treatment. *N Engl J Med* 358, 1160-174.
- Costa-Pereira, A.P., Tininini, S., Strobl, B., Alonzi, T., Schlaak, J.F., Is'harc, H., Gesualdo, I., Newman, S.J., Kerr, I.M., and Poli, V. (2002). Mutational switch of an IL-6 response to an interferon-gamma-like response. *Proc Natl Acad Sci U S A* 99, 8043-47.
- Dumoutier, L., de Meester, C., Tavernier, J., and Renauld, J.C. (2009). A new activation modus of STAT3: A tyrosine-less region of the IL-22 receptor recruits STAT3 by interacting with its coiled-coil domain. *J Biol Chem*
- Emi, M., Fujiwara, Y., Nakajima, T., Tsuchiya, E., Tsuda, H., Hirohashi, S., Maeda, Y., Tsuruta, K., Miyaki, M., and Nakamura, Y. (1992). Frequent loss of heterozygosity for loci on chromosome 8p in hepatocellular carcinoma, colorectal cancer, and lung cancer. *Cancer Res* 52, 5368-372.



Ewing, R.M., Chu, P., Elisma, F., Li, H., Taylor, P., Climie, S., McBroom-Cerajewski, L., Robinson, M.D., O'Connor, L., et al. (2007). Large-scale mapping of human protein-protein interactions by mass spectrometry. *Mol Syst Biol* 3, 89.

Guo, A., Villén, J., Kornhauser, J., Lee, K.A., Stokes, M.P., Rikova, K., Possemato, A., Nardone, J., Innocenti, G., et al. (2008). Signaling networks assembled by oncogenic EGFR and c-Met. *Proc Natl Acad Sci U S A* 105, 692-97.

Haridas, V., Nishimura, G., Xu, Z.X., Connolly, F., Hanausek, M., Walaszek, Z., Zoltaszek, R., and Gutterman, J.U. (2009). Avicin D: a protein reactive plant isoprenoid dephosphorylates Stat 3 by regulating both kinase and phosphatase activities. *PLoS One* 4, e5578.

Heinrich, P.C., Behrmann, I., Haan, S., Hermanns, H.M., Müller-Newen, G., and Schaper, F. (2003). Principles of interleukin (IL)-6-type cytokine signalling and its regulation. *Biochem J* 374, 1-20.

Hendrickx, A., Beullens, M., Ceulemans, H., Abt, T.D., Van Eynde, A., Nicolaescu, E., Lesage, B., and Bollen, M. (2009). Docking motif-guided mapping of the interactome of protein phosphatase-1. *Chem Biol* 16, 365-371.

Hess, D., Keusch, J.J., Oberstein, S.A., Hennekam, R.C., and Hofsteenge, J. (2008). Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective Omicron-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* 283, 7354-360.

Hirsch, F.R., Varella-Garcia, M., Bunn, P.A., Di Maria, M.V., Veve, R., Bremmes, R.M., Barón, A.E., Zeng, C., and Franklin, W.A. (2003). Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* 21, 3798-3807.

Huang, F., Goh, L.K., and Sorkin, A. (2007). EGF receptor ubiquitination is not necessary for its internalization. *Proc Natl Acad Sci U S A* 104, 16904-09.

Hubbard, T.J., Aken, B.L., Ayling, S., Ballester, B., Beal, K., Bragin, E., Brent, S., Chen, Y., Clapham, P., et al. (2009). Ensembl 2009. *Nucleic Acids Res* 37, D690-97.

Klampfer, L. (2008). The role of signal transducers and activators of transcription in colon cancer. *Front Biosci* 13, 2888-899.

Kreis, S., Munz, G.A., Haan, S., Heinrich, P.C., and Behrmann, I. (2007). Cell density dependent increase of constitutive signal transducers and activators of transcription 3 activity in melanoma cells is mediated by Janus kinases. *Mol Cancer Res* 5, 1331-341.

Levy, D.E., and Darnell, J.E. (2002). Stats: transcriptional control and biological impact. *Nat Rev Mol Cell Biol* 3, 651-662.

Li, T., Li, W., Lu, J., Liu, H., Li, Y., and Zhao, Y. (2009). SH2D4A regulates cell proliferation via the ERalpha/PLC-gamma/PKC pathway. *BMB Rep* 42, 516-522.

Lu, Y., Liang, K., Li, X., and Fan, Z. (2007). Responses of cancer cells with wild-type or tyrosine kinase domain-mutated epidermal growth factor receptor (EGFR) to EGFR-targeted therapy are linked to downregulation of hypoxia-inducible factor-1alpha. *Mol Cancer* 6, 63.

Lütticken, C., Coffey, P., Yuan, J., Schwartz, C., Caldenhoven, E., Schindler, C., Kruijer, W., Heinrich, P.C., and Horn, F. (1995). Interleukin-6-induced serine phosphorylation of transcription factor APRF: evidence for a role in interleukin-6 target gene induction. *FEBS Lett* 360, 137-143.

Maa, M.C., Lee, J.C., Chen, Y.J., Chen, Y.J., Lee, Y.C., Wang, S.T., Huang, C.C., Chow, N.H., and Leu, T.H. (2007). Eps8 facilitates cellular growth and motility of colon cancer cells by increasing the expression and activity of focal adhesion kinase. *J Biol Chem* 282, 19399-9409.

Macartney-Coxson, D.P., Hood, K.A., Shi, H.J., Ward, T., Wiles, A., O'Connor, R., Hall, D.A., Lea, R.A., Royds, J.A., et al. (2008). Metastatic susceptibility locus, an 8p hot-spot for tumour progression disrupted in colorectal liver metastases: 13 candidate genes examined at the DNA, mRNA and protein level. *BMC Cancer* 8, 187.

MacKintosh, C., Garton, A.J., McDonnell, A., Barford, D., Cohen, P.T., Tonks, N.K., and Cohen, P. (1996). Further evidence that inhibitor-2 acts like a chaperone to fold PP1 into its native conformation. *FEBS Lett* 397, 235-38.

Neviani, P., Santhanam, R., Trotta, R., Notari, M., Blaser, B.W., Liu, S., Mao, H., Chang, J.S., Galletta, A., et al. (2005). The tumor suppressor PP2A is functionally inactivated in blast crisis CML through the inhibitory activity of the BCR/ABL-regulated SET protein. *Cancer Cell* 8, 355-368.

Notredame, C., Higgins, D.G., and Heringa, J. (2000). T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* 302, 205-217.

Quesnelle, K.M., Boehm, A.L., and Grandis, J.R. (2007). STAT-mediated EGFR signaling in cancer. *J Cell Biochem* 102, 311-19.

Rhodes, D.R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., Yu, J., Briggs, B.B., Barrette, T.R., Anstet, M.J., Kincaid-Beal, C., et al. (2007). Oncomine 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* 9, 166-180.

Roessler, S., Long, E.L., Budhu, A., Chen, Y., Zhao, X., Ji, J., Walker, R., Jia, H.L., Ye, Q.H., et al. (2011). Integrative Genomic Identification of Genes on 8p Associated with Hepatocellular Carcinoma Progression and Patient Survival. *Gastroenterology*

Roth J. A. (1999). *Clin Cancer Res*. In *p53 prognostication: paradigm or paradox?*. UNITED STATES: [UNKNOWN REFERENCE TYPE]

Saif, M.W. (2010). Colorectal cancer in review: the role of the EGFR pathway. *Expert Opin Investig Drugs* 19, 357-369.

Shao, H., Cheng, H.Y., Cook, R.G., and Tweardy, D.J. (2003). Identification and characterization of signal transducer and activator of transcription 3 recruitment sites within the epidermal growth factor receptor. *Cancer Res* 63, 3923-930.

Shen, Y., Schlessinger, K., Zhu, X., Meffre, E., Quimby, F., Levy, D.E., and Darnell, J.E. (2004). Essential role of STAT3 in postnatal survival and growth revealed by mice lacking STAT3 serine 727 phosphorylation. *Mol Cell Biol* 24, 407-419.

Shi, Y. (2009). Serine/threonine phosphatases: mechanism through structure. *Cell* 139, 468-484.

Tanaka, K., Kikuchi-Yanoshita, R., Muraoka, M., Konishi, M., Oshimura, M., and Miyaki, M. (1996). Suppression of tumorigenicity and invasiveness of colon carcinoma cells by introduction of normal chromosome 8p12-pter. *Oncogene* 12, 405-410.

Thiery, J.P., Aclouque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* 139, 871-890.

Togi, S., Kamitani, S., Kawakami, S., Ikeda, O., Muromoto, R., Nanbo, A., and Matsuda, T. (2009). HDAC3 influences phosphorylation of STAT3 at serine 727 by interacting with PP2A. *Biochem Biophys Res Commun* 379, 616-620.

Trenerry, M.K., Carey, K.A., Ward, A.C., and Cameron-Smith, D. (2007). STAT3 signaling is activated in human skeletal muscle following acute resistance exercise. *J Appl Physiol* 102, 1483-89.

Veeriah, S., Brennan, C., Meng, S., Singh, B., Fagin, J.A., Solit, D.B., Paty, P.B., Rohle, D., Vivanco, I., et al. (2009). The tyrosine phosphatase PTPRD is a tumor suppressor that is frequently inactivated and mutated in glioblastoma and other human cancers. *Proc Natl Acad Sci U S A* 106, 9435-440.

Vogelstein, B., and Kinzler, K.W. (2004). Cancer genes and the pathways they control. *Nat Med* 10, 789-799.

Walther, A., Johnstone, E., Swanton, C., Midgley, R., Tomlinson, I., and Kerr, D. (2009). Genetic prognostic and predictive markers in colorectal cancer. *Nat Rev Cancer* 9, 489-499.

Weinberg, R.A. (1991). Tumor suppressor genes. *Science* 254, 1138-146.

Wen, Z., Zhong, Z., and Darnell, J.E. (1995). Maximal activation of transcription by Stat1 and Stat3 requires both tyrosine and serine phosphorylation. *Cell* 82, 241-250.

Woetmann, A., Nielsen, M., Christensen, S.T., Brockdorff, J., Kaltoft, K., Engel, A.M., Skov, S., Brender, C., Geisler, C., et al. (1999). Inhibition of protein phosphatase 2A induces serine/threonine phosphorylation, subcellular redistribution, and functional inhibition of STAT3. *Proc Natl Acad Sci U S A* 96, 10620-25.

Wu, T.R., Hong, Y.K., Wang, X.D., Ling, M.Y., Dragoi, A.M., Chung, A.S., Campbell, A.G., Han, Z.Y., Feng, G.S., and Chin, Y.E. (2002). SHP-2 is a dual-specificity phosphatase involved in Stat1 dephosphorylation at both tyrosine and serine residues in nuclei. *J Biol Chem* 277, 47572-580.

Yue, P., and Turkson, J. (2009). Targeting STAT3 in cancer: how successful are we? *Expert Opin Investig Drugs* 18, 45-56.

Zhang, X., Guo, A., Yu, J., Possemato, A., Chen, Y., Zheng, W., Polakiewicz, R.D., Kinzler, K.W., Vogelstein, B., et al. (2007). Identification of STAT3 as a substrate of receptor protein tyrosine phosphatase T. *Proc Natl Acad Sci U S A* 104, 4060-64.

Zlobec, I., Molinari, F., Kovac, M., Bihl, M.P., Altermatt, H.J., Diebold, J., Frick, H., Germer, M., Horcic, M., et al. (2010). Prognostic and predictive value of TOPK stratified by KRAS and BRAF gene alterations in sporadic, hereditary and metastatic colorectal cancer patients. *Br J Cancer* 102, 151-161.

## Figure legends

### Figure 1. Reduced SHOCA-2 expression in human colorectal cancers correlates with advanced disease stage.

(A) Left, location of SHOCA-2 on human chromosome (huChr) 8p. Middle, degree of conservation between the two human SHOCA sequences with higher degrees of conservation specified by darker shades of grey; box, a consensus MyPhoNE sequence typically present in myosin homologues (Hendrickx et al., 2009). Right, predicted domain organization of SHOCA; the amino acid identity between SHOCA-1 and -2 is given for the MyPhoNE and SH2 domains. Of note, SHOCA-2 protein contains only one coiled-coil domain (CC), namely CC2.

(B) Immunohistochemical analysis of a single tissue microarray containing 400 samples from unselected, untreated patients with sporadic CRC. SHOCA-2 expression was quantified using the H-score that determines the percentage of positive tumor cells multiplied by their staining intensity generating individual groups with scores of >200 (n=275), defining a strong expression (green); 100-200 (n=87), identifying a moderate expression (blue); <100 (n=38), characterizing a low or absent expression (red).

(C) Representative immunohistochemical analysis of tumor tissue of distinct CRC stages depicting either diffuse (>90% cells stained; left), patchy (<30%; middle) or no SHOCA-2 expression (right). T = tumor tissue, N = normal tissue. Scale bar, 200  $\mu$ m.

(D) Kaplan-Meier survival analysis of patients with high (green), moderate (blue) and low SHOCA-2 expression (red) for which sufficient clinical information was available (log-rank test  $p = 0.0118$  when comparing high to low SHOCA-2 expression).

### Figure 2. Genetic and epigenetic silencing of *SH2D4A* encoding SHOCA-2.

(A) HuChr8p copy number variation profiles in SHOCA-2-positive and SHOCA-2-negative CRC. Values <0.75 indicate chromosomal deletions (Del) whereas values >1.25 designate chromosomal duplications (Dup). Red dashed lines, threshold.

(B) Top, comparison of the *SH2D4A* c.342-5T>C (rs17128221) SNP frequency between 83 unrelated healthy Caucasian controls and 16 CRC patients (Chi square p value = 0.006). Bottom, SNP rs17128221 RT-PCR (right) and sequencing analysis (left), the diagram depicts exon usage and the arrows indicate the annealing sites for the primers.

(C) i: top, alignment of amino acid sequences for the SH2 domain (AA 320-454) of human SHOCA-2 (AAH82982) with homologues of *mus musculus* (AAI16683), *rattus norvegicus* (NP\_001012048), and *gallus gallus* (XP\_420452); bottom, sequence analysis of patient tumors. ii: immunoblot analysis of HeLa cells transfected to express Flag-fusion proteins of the indicated human SHOCA-2 mutations.

(D) DNA methylation analysis of CpG island located in the 5'UTR of *SH2D4A* comparing normal and tumor tissue.

(E) Summary of all genetic and epigenetic changes detected in a cohort of 27 informative CRC samples; CNV = copy number variation, SNP = single nucleotide polymorphism. Green = normal finding, red = pathological result, n.d. = not done, roman numerals = number of alleles affected.

### **Figure 3. SHOCA-2 is phosphorylated following EGFR activation and inhibits STAT3 activity.**

(A) SHOCA-2-Flag expressing, serum-starved HeLa cells were stimulated as indicated with EGF (100 ng/ml). Cell lysates were immunoprecipitated with anti-phosphotyrosine (pTyr) antibody or rabbit IgG and immunoblotted to detect Flag and pTyr.

(B) Serum-starved HeLa cells were transiently transfected with either a control vector or a vector encoding the human SHOCA-2-Flag and stimulated as indicated with EGF (100 ng/ml); immunoblotting for the detection of the indicated proteins. The ratio of phosphoprotein to total protein (designated P/Total) was determined by densitometry.

(C) Serum-starved HeLa cells that express normal (control, 100%) or 10% of SHOCA-2 (SHOCA-2 KD) were stimulated for 5, 10 and 15 min with EGF (100

ng/ml). The indicated proteins were detected by immunoblotting. Densitometry was used to determine the ratio of phosphoprotein to total protein (designated P/Total).

(D) Serum-starved HeLa cells and SHOCA-2 KD HeLa cells transiently co-transfected with a SIE firefly luciferase reporter, a renilla luciferase plasmid and a SHOCA-2 expression plasmid were stimulated for 6 hrs with EGF (100 ng/ml; filled bar) or left untreated (open bar). Luciferase assay was performed on lysates. Renilla luciferase activity was used to normalize transfection efficiency. Data represent the mean  $\pm$  SD of three independent, triplicate assays (Student's t test, \*\*p < 0.01; \*\*\*p < 0.005).

(E) Immunoblotting of HeLa, PAE and MEF lysates immunoprecipitated with anti-EGFR, anti-SHOCA-2, anti-Flag or mouse IgG antibodies for the detection of EGFR, STAT3, SHOCA-2 and Flag.

**Figure 4. SHOCA-2 controls STAT3 activity in a PP1 $\beta$ -dependent fashion.**

(A) Lysates from HEK293 cells (bottom) and HEK293 cells expressing SHOCA-2-Flag (top) were immunoprecipitated with anti-PP1 $\beta$  antibody or IgG and immunoblotted for PP1 $\beta$ , and Flag (top) or SHOCA-2 (bottom). Human SHOCA-2 is detected as isoforms due to exon 5 skipping.

(B) Lysates from HEK293 cells expressing CTAP fusion proteins and EGFP-tagged PP1 isoforms were immunoprecipitated with IgG sepharose beads (recognizing CTAP fusion proteins) and immunoblotted for the detection of CTAP and EGFP fusion proteins.

(C) Left, diagram representing wild-type and mutant SHOCA altered PP1 $\beta$  -binding sites (boxes) are marked by crosses. Middle, CTAP fusion proteins and PP1 $\beta$  were detected in total and immunoprecipitated lysates from transfected HEK293 cells. Right, SHOCA-2, Flag and PP1 $\beta$ , respectively, were detected in total lysates and anti-Flag immunoprecipitates from HEK293 cells expressing wild type or mutant SHOCA fusion proteins.

(D) Left, serum-starved HeLa cell expressing the human SHOCA-2-Flag fusion protein were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were

analyzed unmanipulated or were immunoprecipitated with anti-EGFR antibody for the detection of EGFR, STAT3, Flag and PP1 $\beta$  by immunoblotting. Right, ratio of immunoprecipitated protein to total protein (designated IP/Total) as determined by densitometry.

(E) Serum-starved HeLa cell co-expressing the indicated proteins were stimulated with EGF (100 ng/ml) for the specified times. Lysates were immunoblotted for the detection of phospho-EGFR (Tyr845), phospho-STAT3 (Tyr705/Ser727), STAT3, Flag, and GAPDH.

(F) Serum-starved HeLa cells transiently co-transfected with a SIE firefly luciferase reporter, a renilla luciferase plasmid and plasmids encoding either STAT3, the human SHOCA-2-Flag fusion protein or a mutant human SHOCA-2-Flag fusion protein unable to associate with PP1 $\beta$  (PP1<sub>Mut2</sub>) were stimulated for 6 hrs with EGF (100 ng/ml; filled bar) or left untreated (open bar). Luciferase assay was performed on lysates. Renilla luciferase activity was used to normalize transfection efficiency. Data represent the mean  $\pm$  SD of three independent, triplicate assays (Student's t test, \*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.005).

**Figure 5. SHOCA-2 and PP1 $\beta$  arrest tumor cell proliferation sustained by STAT3.**

(A) Left panel: Serum-starved HeLa cells that expressed wild type (control) or reduced (SHOCA-2 KD) SHOCA-2 levels were either left untreated or stimulated with EGF (100 ng/ml) for 6 and 12 hrs. Cell lysates were immunoblotted for the detection of phospho-STAT3 (Tyr705), STAT3, c-Myc, Jun B, Cyclin D1, SHOCA-2 and GAPDH. Right panel: After 24 hrs of serum starvation, control and SHOCA-2 KD HeLa cells were stimulated with EGF (100 ng/ml) for 12 hrs. The cell cycle profile was evaluated by flow cytometry using propidium iodide staining (PI). The bar graph shows the percentage of cells in the indicated phases of the cell cycle. The data is representative of 3 separate experiments, each performed in triplicate (\*p < 0.05; \*\*\*p < 0.005 using Student's t test).

(B) CRC tissue sections from a cohort of 27 informative CRC patients (see Figure 2E) were analyzed for the proportion of tumor cells that stained for SHOCA-2 and



Ki-67 (Wilcoxon test, \* $p < 0.05$ ; \*\*\* $p < 0.005$ ).

(C) SW480 cells transfected to express the indicated empty (pYN3218, Flag) and recombinant vectors were grown for 15 days under neomycin selection and then scored for the number (top) and appearance of colonies (bottom). The graph is representative of two independent experiments performed in triplicates (Student's t test, \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.005$ ).

(D) SW480 cells were transiently transfected with a control vector or recombinant vectors as indicated, serum-starved, and then stimulated with EGF (100 ng/ml) for 10 and 15 min. Lysates were immunoblotted to detect phospho-STAT3 (Tyr705 and Ser727), STAT3, phospho-Akt (Ser473), Akt, phospho-ERK1/2 (Thr202/Tyr204), Flag and GAPDH.

(E) SW620 cells expressing wild-type human SHOCA-2 or a SHOCA-2 mutant (PP1<sub>Mut2</sub>, Y131A) were grown for 15 days under neomycin selection in soft agar and then scored for the number (top) and appearance of colonies (bottom). The graph is representative of two independent experiments performed in triplicates (Student's t test, \*\*\* $p < 0.005$ ).

**Figure 6. SHOCA-2 is required for maintaining epithelial morphology and its down-regulation promotes tumorigenesis *in vivo***

(A) Phase-contrast images of control SHOCA-2 expressing SW480 cells and polyclonal SHOCA-2 knockdown (KD) SW480 cells grown at different densities. Scale bar, 100  $\mu\text{m}$ .

(B) Immunoblot analysis of E-cadherin, N-cadherin, Vimentin, ZEB-1, SNAIL, SHOCA-2 and GAPDH protein expressions in lysates from SW480 stable transfectants (shRNA control, SHOCA-2 KD).

(C) Proliferation of SW480 cells that have been stably transfected with shRNA for the knock-down of SHOCA-2. The cells have been grown in the presence or absence of EGFR inhibitor (AG1478). Data represent the mean  $\pm$  SEM of two independent experiments.

(D) SW480 cells with or without a knock-down of SHOCA-2 were transfected to express either a control plasmid (designated Flag) or a murine (m) SHOCA-2-Flag

fusion protein that was not targeted by the shRNA used. Transfected cells selected with neomycin and scored for the number of colonies detected (top). Cell lysates were immunoblotted for SHOCA-2 and GAPDH protein expression (bottom). The data are representative of two independent experiments.

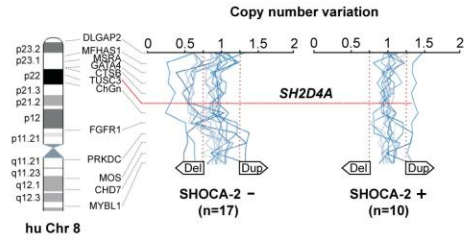
(E) Monitoring tumor incidence and size in nude mice (n=5 mice per group) grafted with either SHOCA-2 wild type (shRNA CT) or knock-down SW480 cells (SHOCA-2 KD). Data represent the mean  $\pm$  SEM of two independent experiments.

(F) Immunohistochemistry of grafted SW480 tumor cells either wild type or knock-down for SHOCA-2 expression (upper left panel; scale bar, 100  $\mu$ m) and of human CRC tissues (right; scale bar, 50  $\mu$ m). Left upper panel: Immunohistochemical analysis of in vivo grafted SW480 cells for STAT3 phosphorylated at Tyr<sup>705</sup>; Left lower panel: H-score analysis for Tyr<sup>705</sup> and Ser<sup>727</sup> STAT3 phosphorylation in CRC samples with low to absent SHOCA-2 expression, displayed as a relative percentage of the biopsies investigated. Right: analysis of CRC tissues on consecutive section for the detection of SHOCA-2 and STAT3 phosphorylated either at Tyr<sup>705</sup> or Ser<sup>727</sup>.

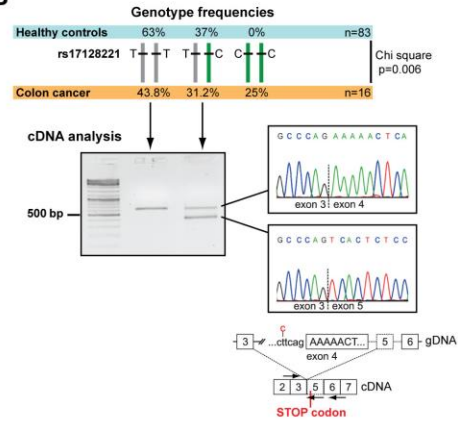


**Figure 2**

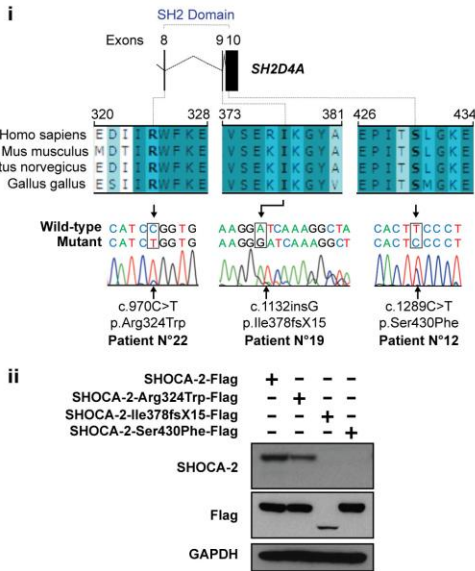
**A**



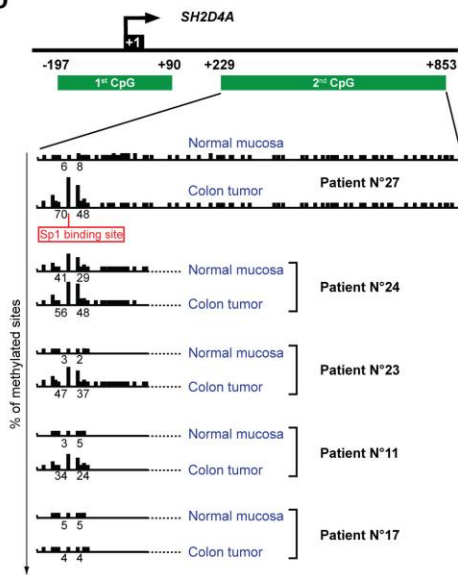
**B**



**C**



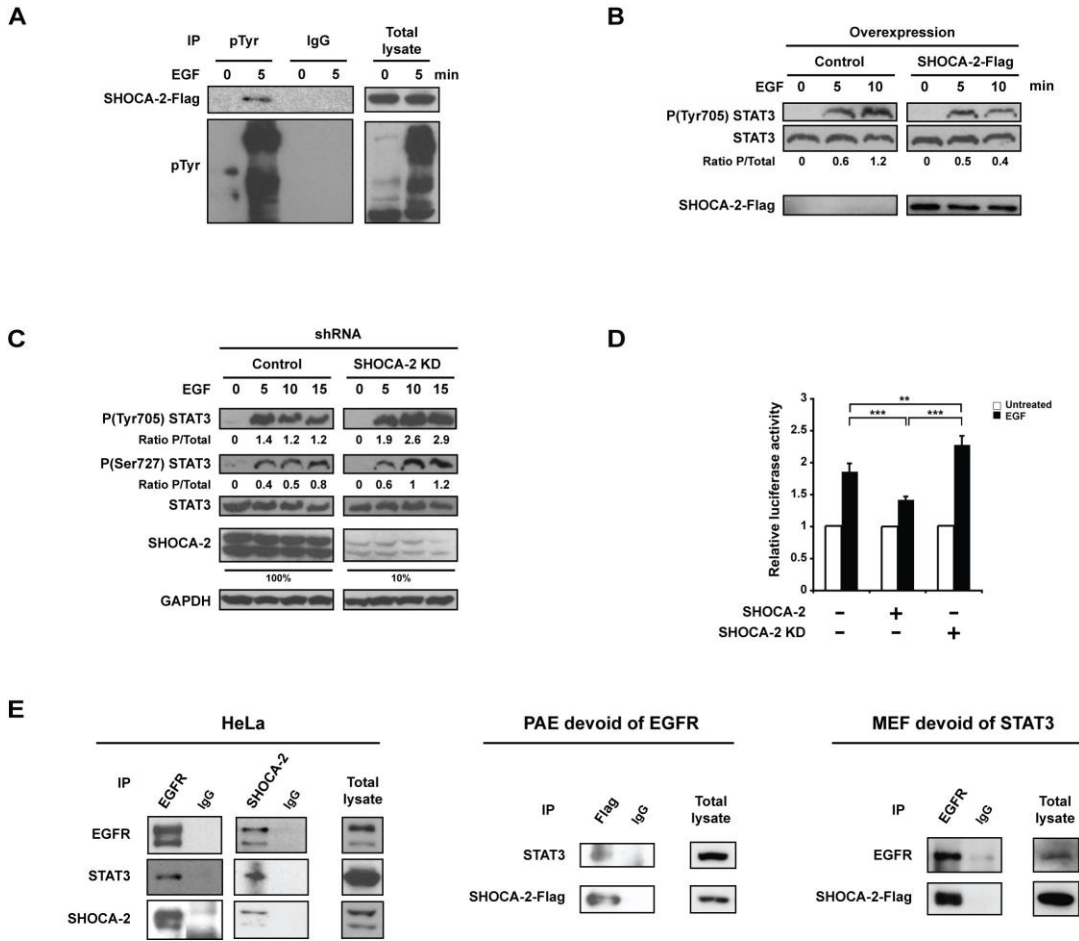
**D**



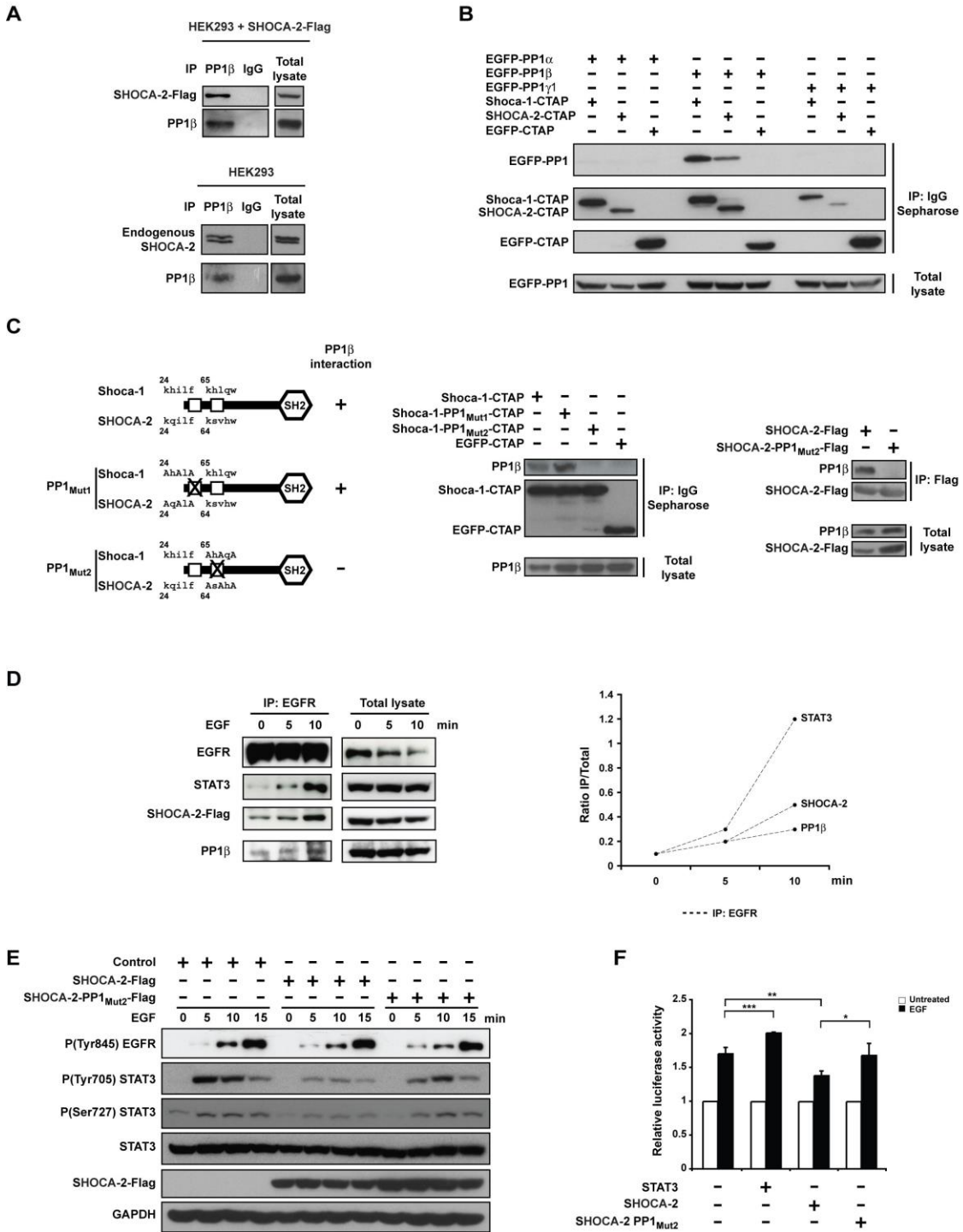
**E**

Tumor stage	I-II									III						IV											
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
Patient N°																											
IHC SHOCA-2	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green
CNV	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Green	Loss I	Loss I	Loss II	Loss I	Gain I	Green	Loss I	Green	Green	Green	Loss I	Loss I	Loss I	Green
Mutation	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	5430P I	Green	Green	Green	Green	Green	Green	Green	1378X I	Green	Green	Green	R324W II	Green	Green	Green	
SNP	n.d.	Green	Green	I	n.d.	n.d.	n.d.	Green	n.d.	Green	I	Green	I	I	I	I	I	I	I	I	I	I	I	I	I	I	
Methylation	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	Green	Green	n.d.	Green	n.d.	Green	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	

**Figure 3**

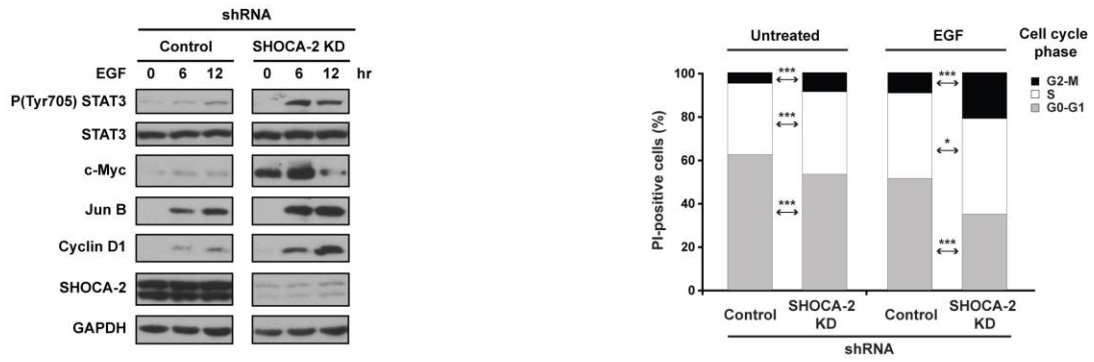


**Figure 4**

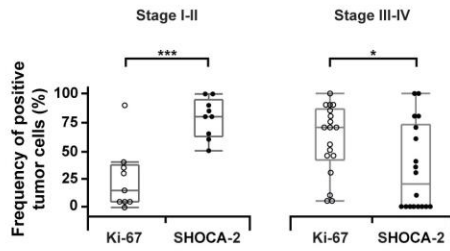


**Figure 5**

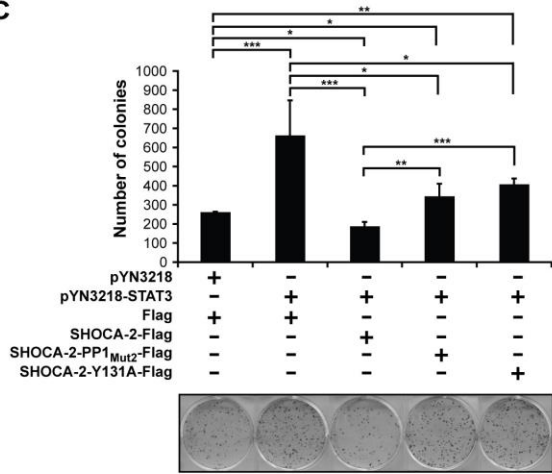
**A**



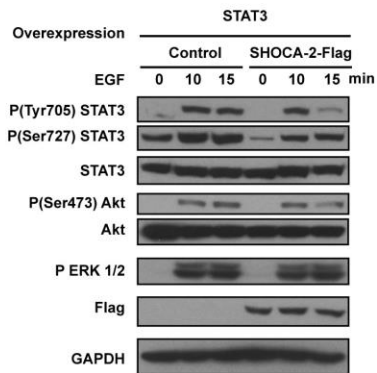
**B**



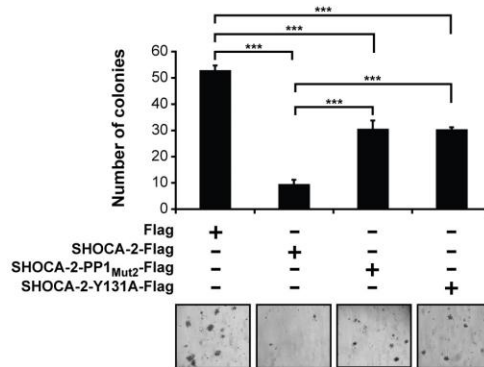
**C**



**D**

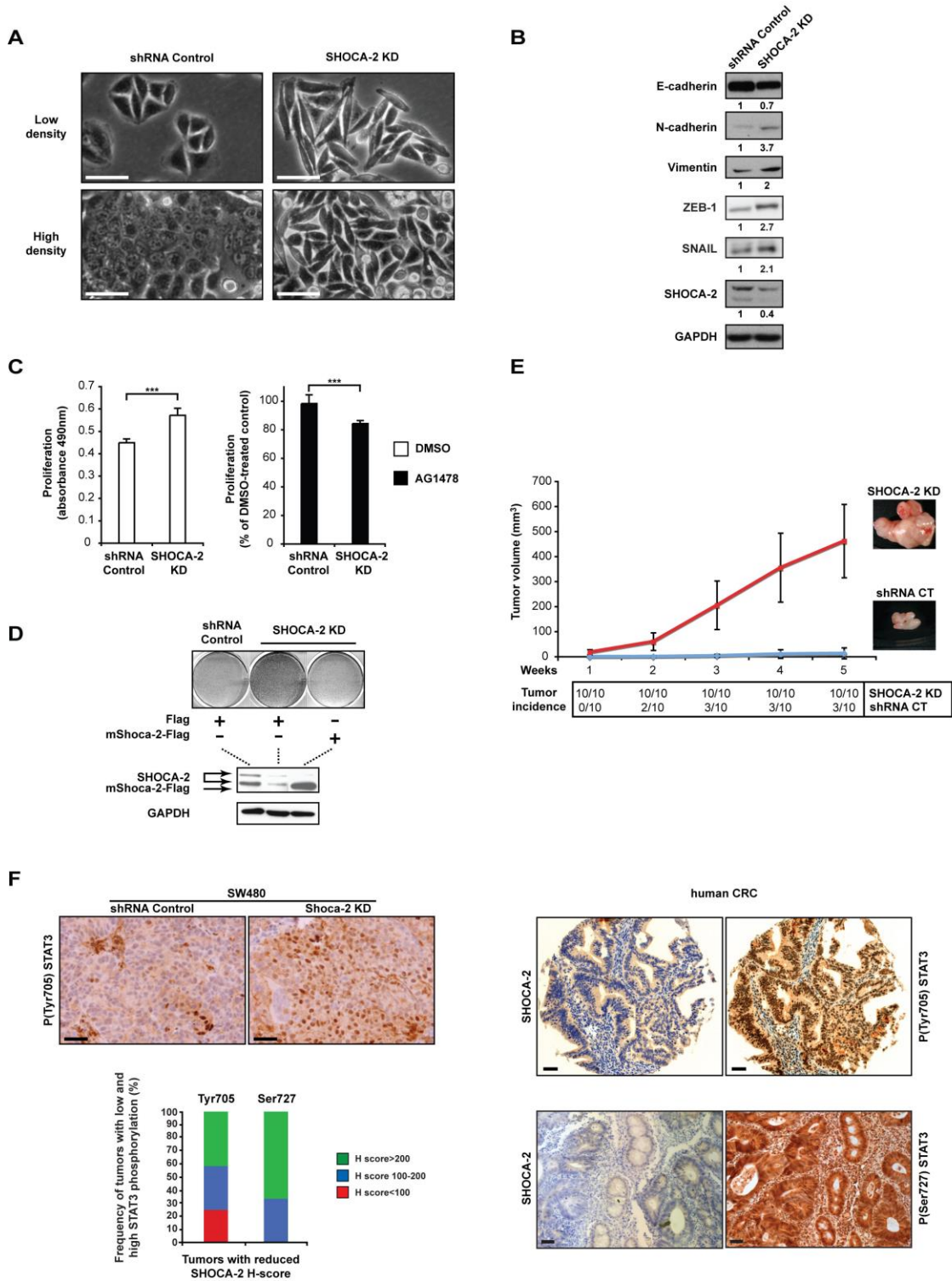


**E**





**Figure 6**





## **Supplementary Information**

**The 8p21.3 encoded SHOCA-2 represses STAT3 activation and acts as a tumor suppressor in colorectal cancer**

Sébastien Loeffler, Michal Kovac, Stefan Weis, Salvatore Piscuoglio, Saulius Zuklys, Marcel Keller, Katrin Hafen, Daniel Hess, Kaspar Truninger, Inti Zlobec, Luigi Terracciano, Karl Heinimann, Primo Schär and Georg A. Holländer

### **Inventory of Supplemental Information**

**Figure S1, related to Figure 1**

**Table S1, related to Figure 1**

**Figure S2, related to Figure 2**

**Figure S3, related to Figure 3**

**Figure S4, related to Figure 4**

**Table S2, related to Figure 4**

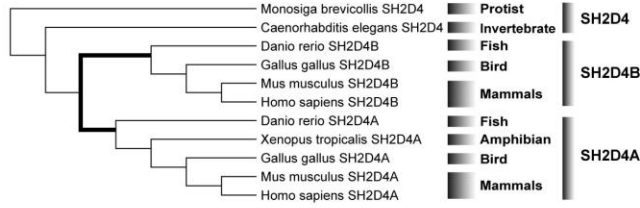
**Figure S5, related to Figure 5**

**Figure S6, related to Figure 6**

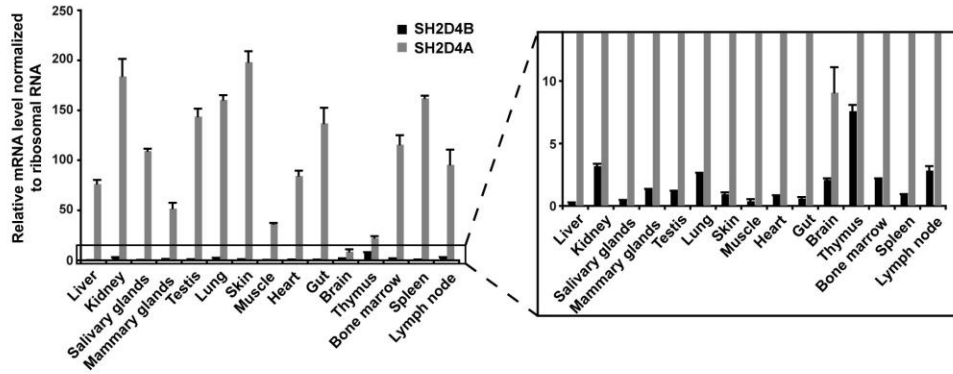
**Supplemental Experimental Procedures**

# Supplementary Data

**A**



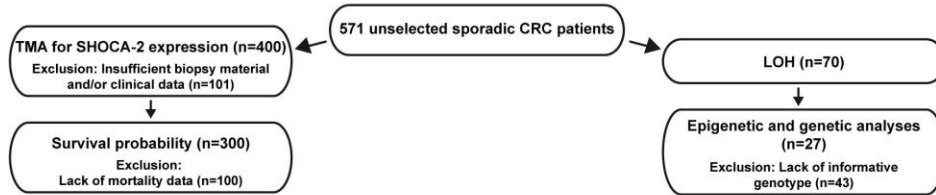
**B**



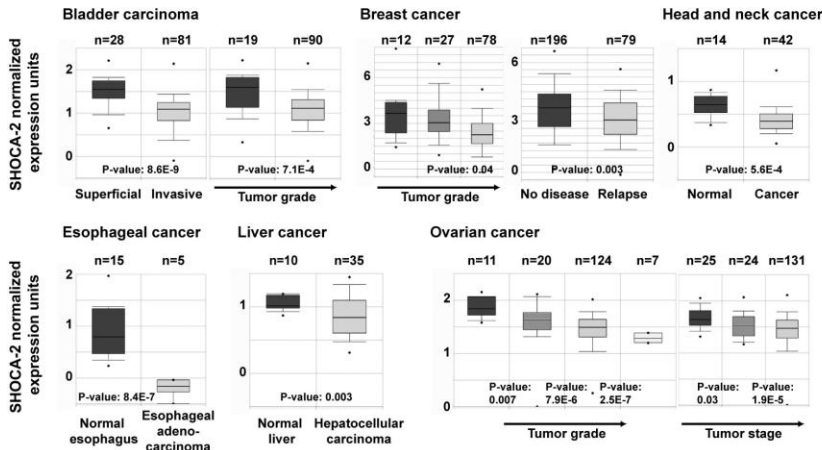
**C**



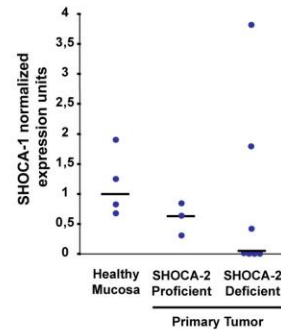
**D**



**E**



**F**



**Figure S1. Related to Figure 1. SHOCA sequence analyses and expression profiling in mouse tissue and in tumors.** (A) The figure represents an unrooted evolutionary tree analysis of *SH2D4* paralogues using the Unweighted Pair Group Method with Arithmetic mean (UPGMA) method. Concurrent with the development of vertebrates (thick line), two distinct *SH2D4* family members developed. (B) Tissues from 10 week-old Balb/c mice were analyzed for *SH2D4B* and *SH2D4A* expression using real time RT-PCR. The results are normalized to 18S rRNA levels. The black bars correspond to *SH2D4B* mRNA levels and the gray bars correspond to *SH2D4A* mRNA levels. (C) The N-terminal amino acid sequences (AA1-51) of human SHOCA-1 and human SHOCA-2 are aligned with human myosin homologues: MYO16 (Myosin-XVI), MYPT1 (Myosin phosphatase-targeting subunit 1, a.k.a. PPP1R12A), MYPT2 (Myosin phosphatase-targeting subunit 2, a.k.a. PPP1R12B), PPP1R12C (Protein phosphatase 1 regulatory subunit 12C), MYPT3 (Myosin phosphatase-targeting subunit 3, a.k.a. PPP1R16A) and ANKRD4 (Ankyrin repeat domain-containing protein 4, a.k.a. PPP1R16B). The box identifies the MyPhone consensus sequence typically present in myosin homologues. (D) From a cohort of 571 unselected patients with sporadic CRC, paraffin-embedded tissue for tissue microarrays (TMA) was available for 501 patients. SHOCA-2 expression analysis was carried out on 400 samples with sufficient tumor tissue (>50%/biopsy) and documented disease stage. Given an informative immunohistochemical analysis and clinical data, survival probability could be calculated for 300 of these patients. LOH was investigated on fresh frozen tissues for 70 of the 571 patients, and epigenetic and genetic analyses were carried out on 27 cases informative for LOH. (E) SHOCA-2 mRNA levels were analyzed by using 9 public gene expression datasets available for meta-analysis (Oncomine)<sup>1</sup>. n designates the number of cases included in the analysis. The Y-axis represents expression units that have been normalized as described elsewhere (<http://www.compendiablo.com/support/normal.htm>). Shaded boxes represent the interquartile range (i.e. the 25th-75th percentile); whiskers define the 10th-90th percent range; bars indicate the median value; and closed circles identify outliers. The p-value was calculated using Student's t-test. (F) Tissues from human healthy mucosa (n= 4) and tumors (n=10) were analyzed for SHOCA-1 expression using real time RT-PCR. The results are normalized to GAPDH levels. The dots correspond to the relative SHOCA-1 expression of each sample and the bars represent the group's median value. The p-value was calculated using Mann-Whitney test.

**A**

H-score		P-value
<b>&lt;100</b>	Stage I vs II	<0.0001
	Stage I vs III	0.0013
	Stage I vs IV	0.0588
	Stage II vs III	0.3692
	Stage II vs IV	0.0143
	Stage III vs IV	0.1083
<b>100-200</b>	Stage I vs II	0.0039
	Stage I vs III	<0.0001
	Stage I vs IV	0.0003
	Stage II vs III	0.1025
	Stage II vs IV	0.3865
	Stage III vs IV	0.4386
<b>&gt;200</b>	Stage I vs II	<0.0001
	Stage I vs III	0.0005
	Stage I vs IV	0.7532
	Stage II vs III	0.4610
	Stage II vs IV	<0.0001
	Stage III vs IV	0.0002

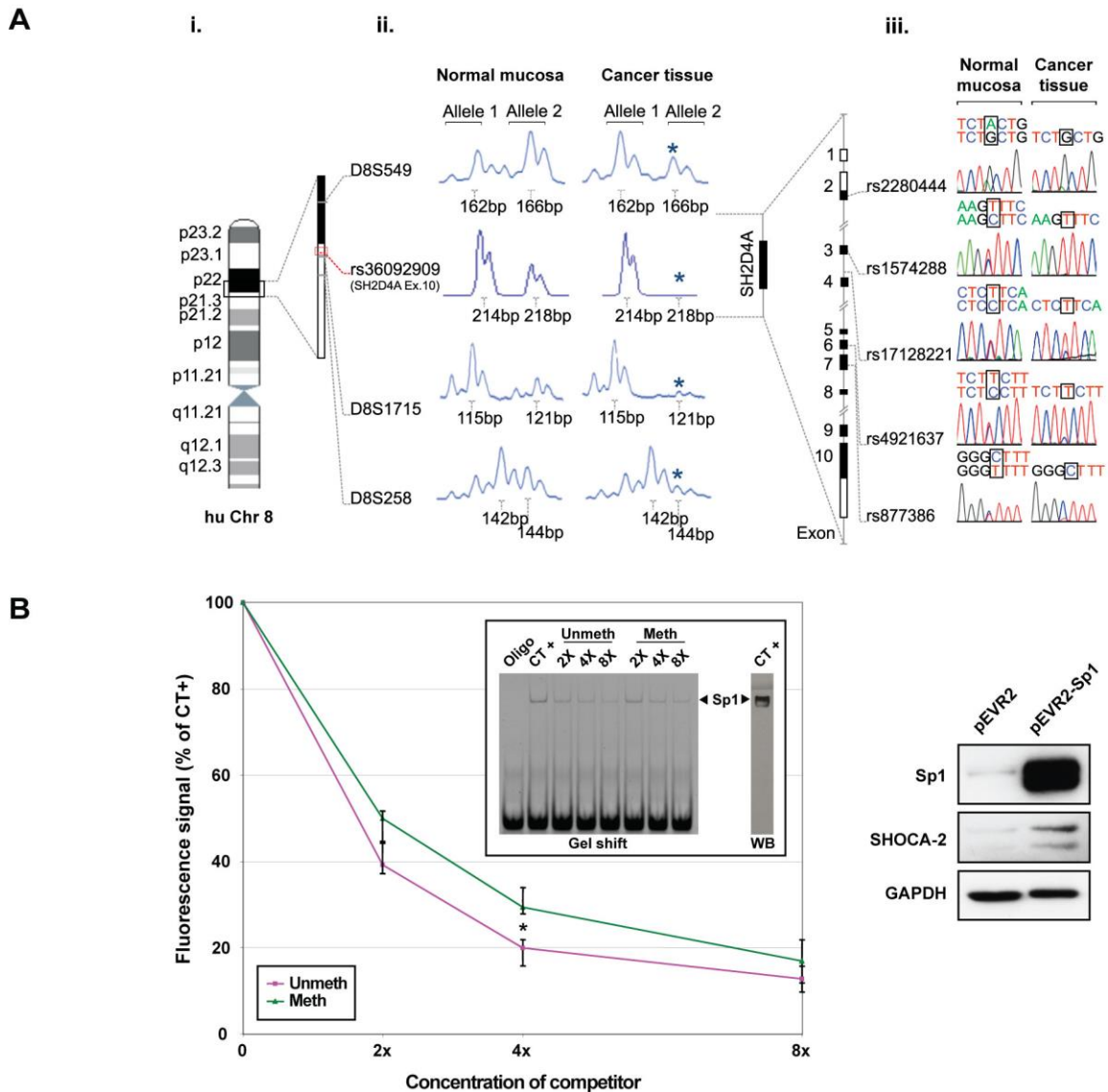
**B**

H-score		P-value
<b>100-200 v s &gt;200</b>	Stage I vs II	0.2828
	Stage I vs III	0.0189
	Stage I vs IV	0.0009
	Stage II vs III	0.0739
	Stage II vs IV	0.002
	Stage III vs IV	0.1298
<b>&lt;100 vs &gt;200</b>	Stage I vs II	0.0139
	Stage I vs III	0.0368
	Stage I vs IV	0.112
	Stage II vs III	0.5808
	Stage II vs IV	0.5409
	Stage III vs IV	0.8622

**C**

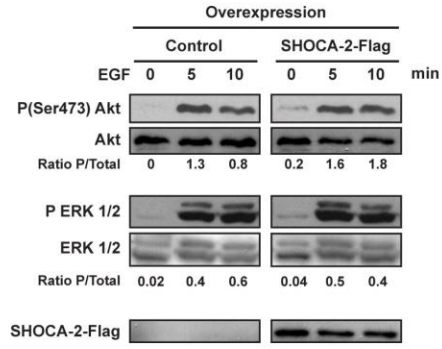
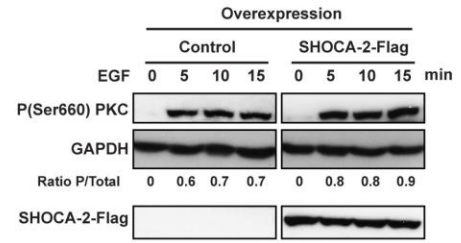
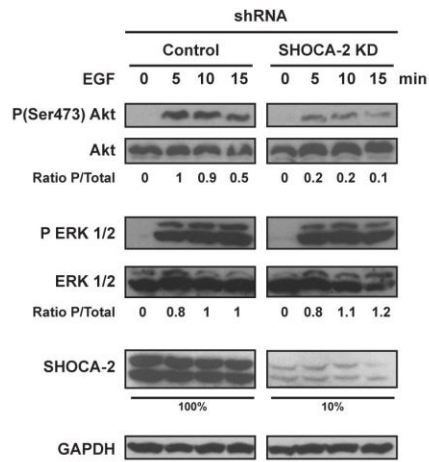
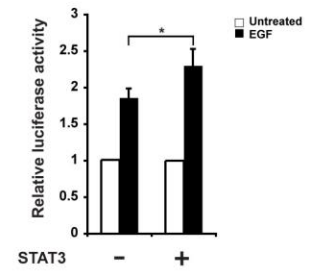
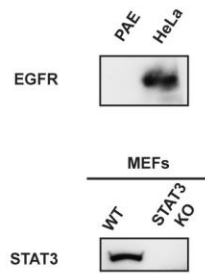
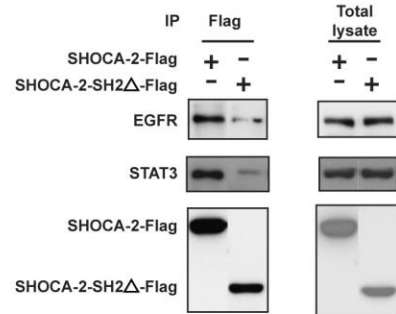
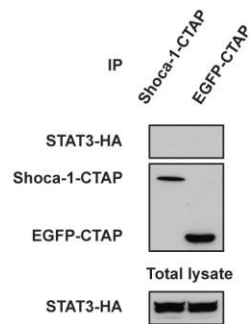
	Comparison	Log-rank (p-value)	Wilcoxon (p-value)	Cox (p-value)	HR (95% CI)
<b>H-scores</b>	100-200 v s >200	0.2312	0.2164	0.2365	0.81 (0.57-1.15)
<b>Stage</b>	Stage I (no events)	-	-	-	
	Stage II	0.146	0.096	0.159	0.67 (0.38-1.17)
	Stage III	0.582	0.75	0.586	0.82 (0.39-1.7)
	Stage IV	0.843	0.824	0.846	0.94 (0.51-1.74)
<b>H scores</b>	<100 v s >200	0.0118	0.0133	0.0134	0.56 (0.35-0.89)
<b>Stage</b>	Stage I (no events)	-	-	-	
	Stage II	0.355	0.553	0.36	0.62 (0.22-1.73)
	Stage III	0.548	0.551	0.55	0.78 (0.34-1.77)
	Stage IV	0.569	0.609	0.58	0.82 (0.41-1.64)

**Table S1. Related to Figure 1. Statistical analyses for association of SHOCA-2 H-score, stage and survival time** (A) Differences in the stage distribution among each H-score category (binomial test for equal proportions). (B) Unadjusted p-values for association of H-scores and stage (Chi-Square or Fisher's Exact test). (C) P-values for survival time differences overall and stratified by stage using three different statistical tests.



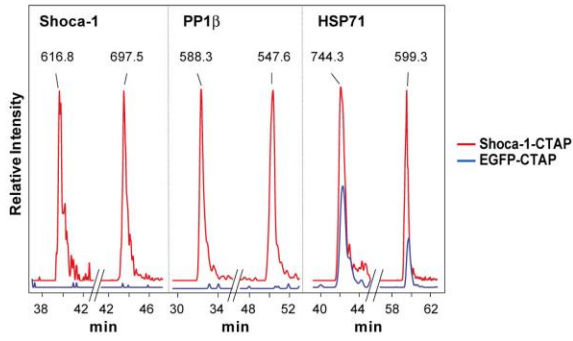
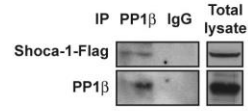
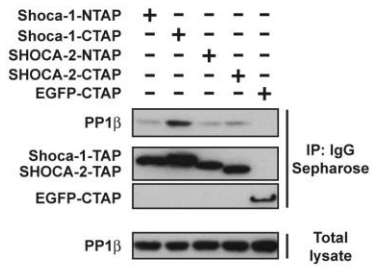
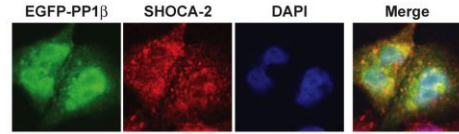
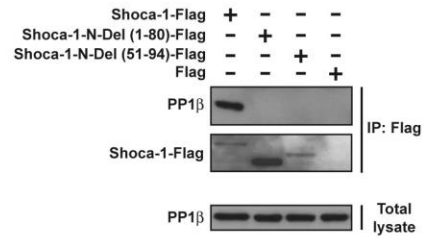
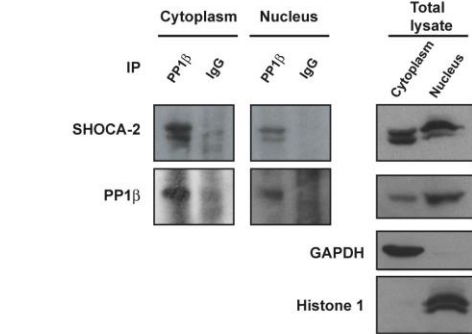
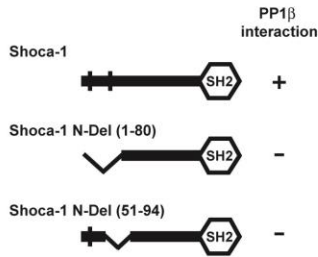
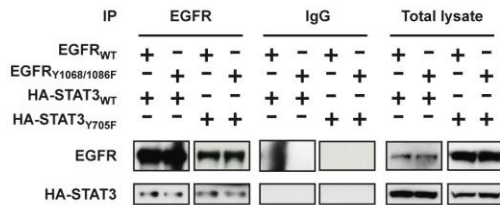
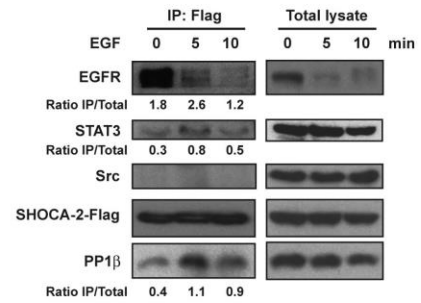
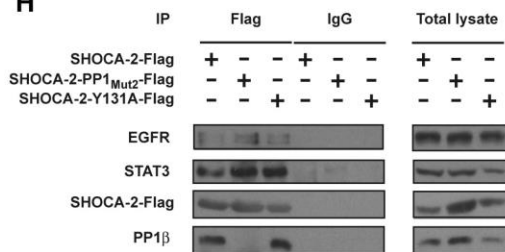
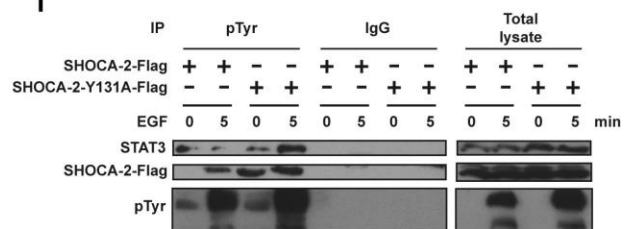
**Figure S2. Related to Figure 2. Epigenetic and genetic analyses on *SH2D4A* gene.** (A) Loss of Heterozygosity (LOH) analysis for *SH2D4A*. Loss of heterozygosity analysis for the *SH2D4A* locus. Panels **i** and **ii** depict 3 flanking and 1 gene-specific microsatellite marker, respectively, that cover the region (approximately 7Mb) from 8p21.3 to the telomeric boundary of 8p22. Panel **iii** demonstrate for cancer sample number 17 the analysis of 5 gene-specific SNPs and reveal a selective loss of a single SHOCA-2 allele. (B) Left, Sp1 binding affinity differs between methylated and unmethylated *SH2D4A* DNA oligomers. The Sp1 specific gel-shifts were carried out using a fluorescein-labeled oligomer binding Sp1 (CT+). CT+ and SW48 nuclear extract were incubated in a competitive manner in the absence or presence of increasing concentrations of an unlabeled oligomers. This oligomer sequence, which represents part of the genomic *SH2D4A* sequence, had either an unmethylated (Unmeth) or a methylated (Meth) Sp1 binding site. As a negative control (designated Oligo), CT+ oligomer was incubated in the absence of nuclear extract. Data from four independent experiments are shown and represent the mean and the interquartile range (i.e. the 25th-75th percentile) of the percentage of the fluorescence signal obtained with experimental samples in comparison to control (CT+; \*p < 0.05 using Mann-Whitney test). The presence of Sp1 in the shifted bands was confirmed by western blotting (WB). Right, HeLa cells were transiently

transfected with either a control vector or a vector encoding Sp1; immunoblotting for the detection of the indicated proteins.

**A****B****C****D****E****F****G**



**Figure S3. Related to Figure 3. Effect of SHOCA-2 on EGFR-mediated signaling.** (A,B) Serum-starved HeLa cells were transiently transfected with either a control vector or a vector encoding the human SHOCA-2-Flag and stimulated as indicated with EGF (100 ng/ml); immunoblotting for the detection of the indicated proteins. The ratio of phosphoprotein to total protein (designated P/Total) was determined by densitometry. (C) HeLa cells were stably transfected either with a plasmid encoding an shRNA specific for SHOCA-2 or with plasmid devoid of an insert. Following serum starvation for 24 hrs, the cells were stimulated with EGF (100 ng/ml) as indicated. Western blotting for the indicated proteins were performed on aliquots of transfected cell lysates. The efficiency of SHOCA-2 silencing is indicated as a percentage of control transfected cells. Densitometry was used to determine the ratio of phosphoprotein to total protein signal (designated P/Total). (D) A SIE firefly luciferase reporter construct and a renilla luciferase plasmid were transiently transfected together with a STAT3 expression vector into unmanipulated HeLa cells. Transfected cells were serum-starved for 24 hrs and then either stimulated with EGF (100 ng/ml; filled bar) for 6 hrs or left untreated (open bar). The renilla luciferase activity was used to normalize for the transfection efficiency. Data are representative of 3 experiments, each performed in triplicate (\*p < 0.05; \*\*p < 0.01; \*\*\*p < 0.005 using Student's t test). (E) Immunoblotting of lysates from PAE cells, HeLa, and wild type and Stat3 deficient MEFs for the detection of EGFR and Stat3, respectively. (F) HeLa cells were transfected to express a Flag-fusion protein of either the wild-type form of human SHOCA-2 or a mutant form of human SHOCA-2 that lacks the SH2 domain (designated SHOCA-2-SH2 $\Delta$ ). Cell lysates were immunoprecipitated with anti-Flag antibodies and subsequently analyzed by immunoblotting using antibodies specific for Flag, STAT3 and EGFR (left panel). Total lysates were used as loading controls (right panel). (G) HEK293 cells were transfected to express STAT3-HA and a C-terminal TAP-tag fusion proteins of SHOCA-1 or EGFP. Cell lysates were immunoprecipitated with IgG beads, and immunoblotted with either antibodies specific for HA-tag or with Rabbit IgG recognizing the protein A component of the TAP-tag.

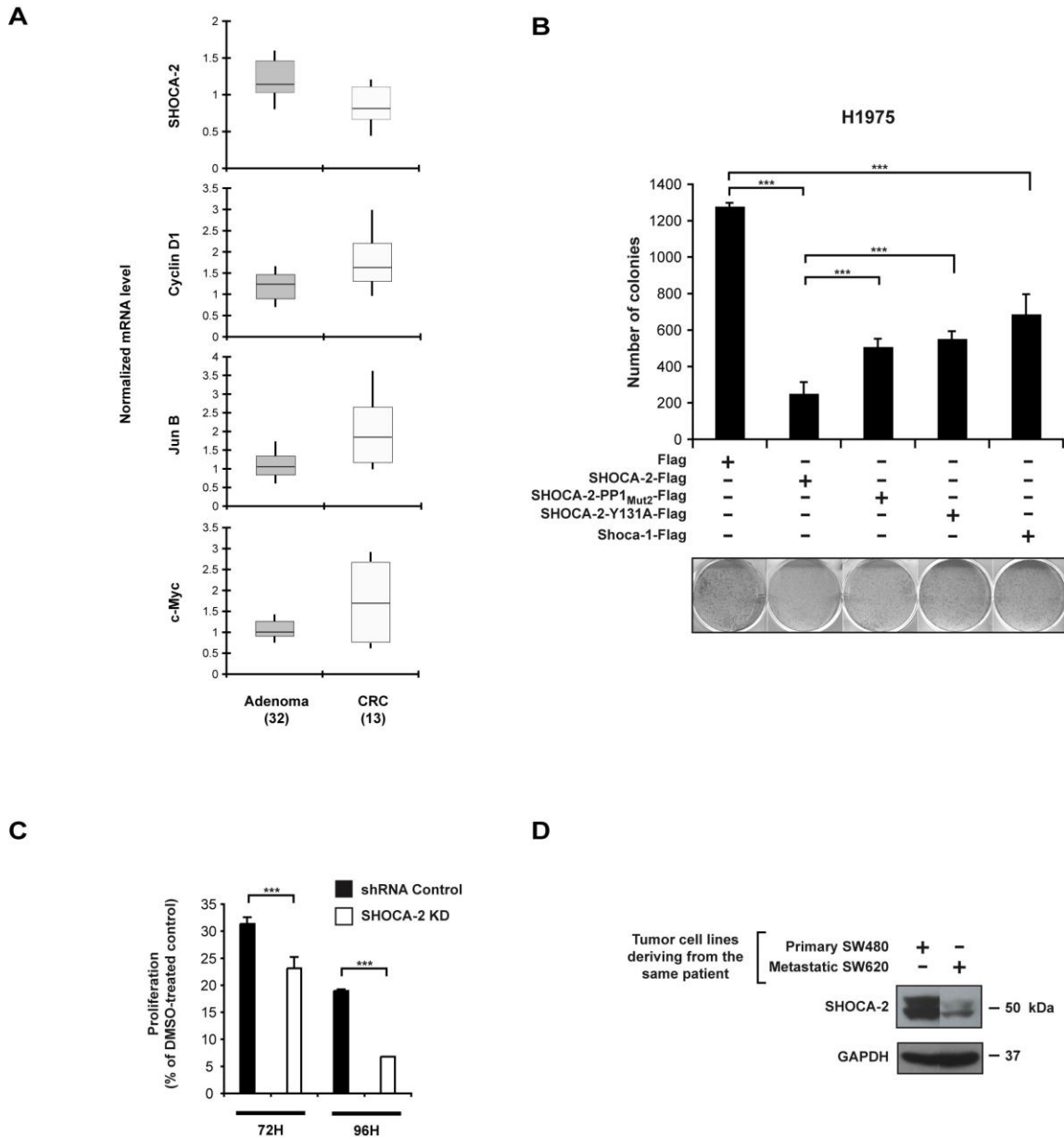
**A****B****C****D****E****F****G****H****I**

**Figure S4. Related to Figure 4. Interaction of SHOCA-2 with the EGFR/STAT3/PP1 $\beta$  complex.**

(A) Multiple reaction monitoring (MRM) comparison of two selected peptides from SHOCA-1, PP1 $\beta$  and HSP71, respectively. Red and blue lines represent peptides retrieved from the SHOCA-1-CTAP and control EGFP-CTAP pull-downs, respectively. The transition signals are labeled with the mass (m/z) of the peptides. Peptides derived from non-specific interactions (e.g. heat shock protein HSP71) were identified in both eluates whereas SHOCA-1 and PP1 $\beta$  peptides were only found in the SHOCA-1-CTAP immunoprecipitation. (B) Cell lysates from HEK293 cells transfected to express SHOCA-1-Flag fusion protein were immunoprecipitated either with an antibody specific for PP1 $\beta$  or with goat IgG and then immunoblotted with antibodies specific for Flag and PP1 $\beta$ . (C) HEK293 cells were transfected to express the indicated TAP-tag fusion proteins. The cell lysates were immunoprecipitated with IgG beads, and immunoblotted with either antibody specific for PP1 $\beta$  (loading control) or with Rabbit IgG recognizing the protein A component of the TAP-tag. (D) Top, HeLa cells expressing an EGFP-PP1 $\beta$  fusion protein were stained with anti-SHOCA-2 antibody and DAPI to detect colocalization by confocal microscopy. Bottom, endogenous SHOCA-2 was detected in cytoplasmic (GAPDH+) and nuclear (Histone H1+) HeLa cell lysates immunoprecipitated with anti-PP1 $\beta$ . (E) HEK293 cells were transfected to express either a wild-type SHOCA-1-Flag fusion protein or one of two deletion mutants AA1-80 or AA51-94, fused to Flag (left panel; the alleged PP1-binding sites are indicated as vertical bars). Lysates from transfected cells were immunoprecipitated with anti-Flag antibodies and immunoblotted with antibodies directed against Flag or PP1 $\beta$  (right panel). Interaction of wild type and mutant SHOCA-2 with EGFR and STAT3 (F) Lysates of HeLa cells expressing HA-tagged wild-type (WT) STAT3 or the Y705F STAT3 mutant and wild-type (WT) EGFR or the Y1068/1086F EGFR mutant were immunoprecipitated with an EGFR specific antibody. EGFR and HA were detected by immunoblotting. (G) Serum-starved HeLa cell expressing the human SHOCA-2-Flag fusion protein were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were immunoprecipitated with anti-Flag antibody. EGFR, STAT3, c-SRC, Flag, PP1 $\beta$  was detected by immunoblotting. (H) HeLa cells were transfected with an expression plasmid encoding a Flag-fusion protein of either the wild-type (WT), the PP1<sub>Mut2</sub> or the Y131A mutant form of human SHOCA-2. Cell lysates were immunoprecipitated with anti-Flag antibodies and subsequently analysed by immunoblotting using antibodies specific for Flag, STAT3, EGFR and PP1 $\beta$  (left panel). Total lysates were used as loading controls (right panel). (I) Serum-starved HeLa cells expressing human wild-type SHOCA-2-Flag or the Y131A mutant were stimulated with EGF (100 ng/ml) for the indicated times. Lysates were immunoprecipitated with a phosphotyrosine (pTyr) specific antibody. STAT3, Flag and pTyr were detected by immunoblotting.

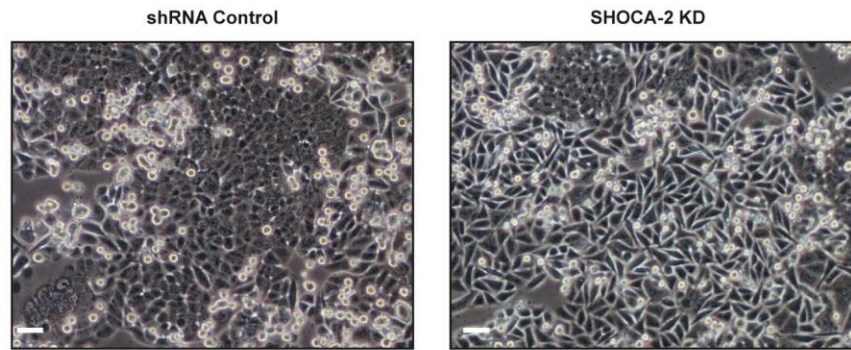
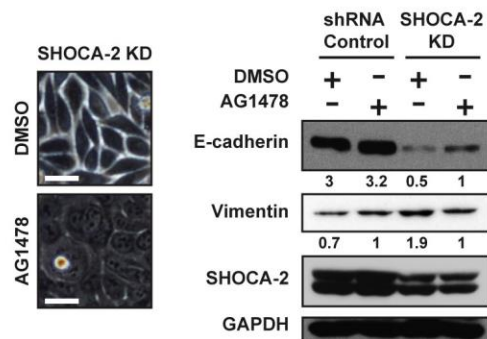
Uniprot	Description	Mass	Shoca1-CTAP		EGFP-CTAP	
			Score	Spectral count	Score	Spectral count
P07437	Tubulin beta chain	49671	1226	56	360	11
B2RSQ7	SH2 domain containing 4B (Shoca1)	51120	926	61	0	0
P62140	Serine/threonine-protein phosphatase PP1-beta catalytic subunit	37187	747	35	0	0
P81947	Tubulin alpha-1B chain	50152	673	51	265	10
P08107	Heat shock 70 kDa protein 1	70052	343	17	88	2
P04517	Genome polyprotein	346164	279	13	182	9
P68104	Elongation factor 1-alpha	50141	228	10	81	3
P08238	Heat shock protein HSP 90-beta	83264	219	5	168	4
P62158	Calmodulin	16838	118	4	266	13
Q9H1R3	Myosin light chain kinase 2	64685	57	6	181	13
P42212	Green fluorescent protein	26886	0	0	357	23

**Table S2. Related to Figure 4. LC-MS/MS analysis: comparison of peptides eluted following SHOCA-1- and EGFP-TAP purification.** Proteins were identified using 2 or more unique peptides with a MASCOT score above 20. To distinguish between correct and incorrect peptide assignments, protein spectra with less than 4 peptides were manually evaluated. Spectral count: number of spectra identified for each protein.



**Figure S5. Related to Figure 5. SHOCA-2 influences cell cycle progression and proliferation in STAT3-dependent fashion.** (A) mRNA expression of SHOCA-2, Cyclin D1, Jun B, c-Myc, was assessed in 32 adenoma and 13 sporadic colorectal cancers by means of Affymetrix GeneChip U133 expression arrays. Shaded boxes represent interquartile range (i.e. the 25th-75th percentile); whiskers indicate the 10th-90th percent range; bars provide the median value. (B) H1975 cells were transfected to express either a vector without insertion (designated Flag) or a vector encoding Flag-fusion proteins of either wild type SHOCA-2, a SHOCA-2 mutant unable to bind to PP1 $\beta$  (SHOCA-2-PP1<sup>Mut2</sup>), or a SHOCA-2 mutant in which the tyrosine at position 131 was exchanged for an alanine (SHOCA-2-Y131A), or wild type Shoca-1. The cells were grown for 15 days under neomycin selection and then scored for the number (bar graphs) and appearance of colonies (microphotographs). The data is representative of two independent experiments performed in triplicates (Student's t test, \* $p < 0.05$ ; \*\* $p < 0.01$ ; \*\*\* $p < 0.005$ ). (C) Proliferation of SW480 cells that have been stably transfected with a control (shRNA control) vector or a recombinant vector to knock-down SHOCA-2 expression (SHOCA-2 KD). Successfully transduced cells were grown

either in the presence or in the absence of the STAT3 inhibitor S3I-201 diluted in DMSO (100 $\mu$ M). Data represent the mean  $\pm$  SD of two independent experiments carried out in triplicates. (D) SHOCA-2 protein expression in the CRC cell lines SW480 (primary tumor) and SW620 (metastasis from the same patient) was assessed using Western-blotting.

**A****B****Figure S6. Related to Figure 6. SHOCA-2 knock-down SW480 cells display features of EMT.**

(A) Low power view of SW480 cells expressing either a control vector (shRNA control, left panel) and or a vector to knock-down SHOCA-2 expression (SHOCA-2 KD; right panel). Scale bar, 100  $\mu$ m. (B) Left, phase-contrast images of Shoca-2 KD SW480 cells grown in the presence or absence of EGFR inhibitor AG1478 diluted in DMSO (10 $\mu$ M; scale bar: 50  $\mu$ m). Right, immunoblot analysis of E-cadherin, Vimentin, Shoca-2 and GAPDH protein expressions in lysates from shRNA control and Shoca-2 KD SW480 cells cultured in the in the presence or absence of EGFR inhibitor AG1478 diluted in DMSO (10 $\mu$ M).

## **Supplemental Experimental Procedures**

### **Plasmids, antibodies and reagents**

Mouse and human SHOCA-1 and SHOCA-2 sequences were amplified from thymus cDNA libraries by PCR, subcloned into the expression vectors pcDNA-Flag 3.1, CTAP and NTAP (a kind gift from Dr. A.C. Gingras, Samuel Lunenfeld Research Institute at Mount Sinai Hospital, Toronto, Canada) and sequenced. Site-directed mutagenesis (Qiagen) was used to introduce specific mutations. The human EGFP-PP1 $\alpha$ ,  $\beta$ ,  $\gamma$ 1 constructs were generously provided by Dr. L. Trinkle-Mulcahy (Wellcome Trust Centre for Gene Regulation and Expression, University of Dundee, Scotland, UK), the EGFP-CTAP plasmid was a gift from Dr. S. Meloche (Institute of Research in Immunology and Cancer, Université de Montréal, Canada), the SIE firefly luciferase reporter plasmid was kindly made available by Dr. I. Behrmann (Life Sciences Research Unit, University of Luxembourg, Luxembourg), the EGFR wild-type and Y1068/1086F mutant constructs were generously provided by Dr. S. Sigismund (IFOM, Milano, Italy) and the STAT3 expression vector was obtained courtesy of Dr. J.Y. Yoo (Pohang University of Science and Technology, Pohang, Republic of Korea). Antibodies used were directed against PP2A (Upstate); Flag, ERK 1/2, anti-vimentin (all from Sigma); Histone H1, Jun B, c-Myc, cyclin D1, c-SRC and ZEB-1 (all from Santa-Cruz); GAPDH, SNAIL and PP1 $\beta$  (all from Abcam); SH2D4A (Abnova); EGF receptor, Tyr845-phosphorylated EGFR, Ser660-phosphorylated PKC, phospho-tyrosine (designated pTyr), Tyr705/Ser727-phosphorylated STAT3, Akt, Ser473-phospho-Akt, Thr202/Tyr204-phospho-ERK 1/2 (all from Cell Signaling); GFP (Roche); anti-N-cadherin (Takara Biomedicals); anti-E-cadherin and STAT3 (both from BD Biosciences). EGF was obtained from Peprotech and used at a concentration of 100 ng/ml. AG1478 was purchased from Axxora and the STAT3 inhibitor (S3I-201) was obtained from Merck.

### **Cell culture and transfection**



HEK293, HeLa, SW480, SW620 cells (a kind gift from Dr. G. Marra, Institute of Molecular Cancer Research, University of Zurich, Switzerland), Murine embryonic fibroblasts (MEF) either wild type (fl/fl) or deficient for Stat3 (MEF $\Delta/\Delta$ ), a generous gifts of Dr. Valeria Poli, University of Turin, Torino, Italy, were grown in Dulbecco's modified Eagle's medium (DMEM) supplemented with Glutamax (4 mM), and 10% fetal calf serum (Invitrogen-Gibco). Porcine aortic endothelial (PAE) cells (kindly made available by Dr. P. Berger, Paul Scherrer Institut, Villigen, Switzerland) were grown in F-12 Nutrient Mixture (Ham) supplemented with Glutamax (4 mM), and 10% fetal calf serum (Invitrogen-Gibco). Transient transfections were accomplished using Lipofectamine 2000 (Invitrogen) for HEK293 and Fugene HD (Roche) for HeLa, SW480, SW620 cells according to the manufacturer's recommendations. For stable pShoca-1-CTAP or pEGFP-CTAP transfectants, HEK293 cells were co-transfected with the puromycin resistance plasmid, pBabe-puro (Cell Biolabs). Cells were first plated onto 100-mm plates and allowed to grow to 80% confluence, then rinsed twice with phosphate-buffered saline (PBS) and finally exposed to the plasmids/Lipofectamine-DMEM mixture. After overnight incubation, the cells were washed, split into a 150-mm plate and for selection grown in DMEM containing puromycin (10.8  $\mu\text{g/ml}$  for the first 48 hrs and then 3.6  $\mu\text{g/ml}$ ). Several puromycin resistant clones were established. Cells expressing comparable amounts of either Shoca-1-CTAP or EGFP-CTAP protein (as analyzed by western-blotting) were selected for further analysis. For the stable knock-down, HeLa and SW480 cells were transfected using different human SHOCA-2 (*SH2D4A*)-specific shRNA constructs (Origene).

### **Bioinformatic analysis**

Sequences from SHOCA-1 (SH2D4B) and SHOCA-2 (SH2D4A) were retrieved by homology search with ENSEMBL<sup>2</sup>: *C. elegans* SHOCA, F13B12.6; *Danio rerio* SHOCA-1, ENSDARG00000015144; *Gallus gallus* SHOCA-1, ENSGALG00000002370; *Mus musculus* SHOCA-1, ENSMUSG00000037833; *H. sapiens* SHOCA-1, ENSG00000178217; *Danio rerio* SHOCA-2, ENSDARG00000041378; *Xenopus laevis* SHOCA-2, ENSXETG00000004464;

Gallus gallus SHOCA-2, ENSGALG00000010122; Mus musculus SHOCA-2, ENSMUSG00000053886; H. sapiens SHOCA-2, ENSG00000104611. The common SHOCA precursor (Sh2d4) sequence of Monosiga brevicollis (30425.AA) was identified in the KinBase database (<http://www.kinase.com/kinbase/>). Sequences were aligned with the TCoffee software (<http://tcoffee.vital-it.ch>)<sup>3</sup>. An unrooted evolutionary tree of SHOCA homologues was generated using the Unweighted Pair Group Method with Arithmetic mean (UPGMA). The gene microarray analysis tool made available by Oncomine<sup>1</sup> (Compendia Bioscience, Ann Arbor, MI, USA: <http://www.oncomine.org>) was used to determine SHOCA-2 mRNA expression in various cancers.

### **Quantification of SHOCA-2 expression and STAT3 phosphorylation on tissue microarrays & whole-tissue sections**

The H score determines the percentage of positive tumor cells multiplied by their staining intensity<sup>4</sup>. Given the size of the cohort of patients analyzed, three ranges of scores were considered: >200, identifying strong expression or phosphorylation; 100-200, classifying moderate expression or phosphorylation; <100 characterizing a low or absent expression or phosphorylation.

### **Statistical analyses**

The likelihood ratio test and binomial test for differences in proportions were used to compare H-scores between and among disease stages, respectively<sup>5</sup>. Survival time analysis was performed using three different methods: the non-parametric log-rank test, the non-parametric Gehran-Wilcoxon tests and the Cox regression analysis which was carried out after verification of the assumption of proportional hazards; survival time differences were represented using the Kaplan-Meier method<sup>6</sup>. Hazard ratios (HR) and 95% confidence intervals (CI) were obtained to measure effect size with values <1.0 indicating more favorable survival time with higher protein expression. P-values were two-sided and considered statistically significant when  $p < 0.05$ .

### **Real-time RT-PCR**

Using TRI reagent (Molecular Research Center), total RNA was prepared from various tissues freshly isolated from 10 week-old Balb/c mice. Oligo(dT20) or random-N6-primed cDNA was generated with M-MLV Reverse Transcriptase (Invitrogen-Gibco) employing standard protocols. For quantitative real-time PCR, the SYBR Green-based method was used (Sensimix, Quantace); primers are available upon request. PCR specificity was controlled by analyzing the melting curves and agarose gel electrophoresis of the PCR products. Amounts of specific mRNA were normalized to levels of 18S rRNA transcripts.

### **Mass spectrometry**

TCA precipitated and acetone washed or Chloroform/Methanol treated protein pellets were reduced with TCEP (tris(2-carboxyethyl)phosphine), alkylated with Iodoacetamide and digested with Trypsin. The generated peptides were analyzed by NanoLC-MS/MS on a 4000Q Trap (MDS Sciex) as described<sup>7</sup>. The proteins were identified with Mascot (Matrix Science) searching the Uniprot database<sup>8</sup>. MRM-buddy, a software developed at the Friedrich Miescher Institut (A. di Cara, R. Sack and R. Portmann, unpublished results) was used to extract the MRM-relevant information for the quantification of selected proteins. MRM analysis was done as described<sup>7</sup>.

### **Western Blot and immunoprecipitation**

Total protein (40µg as quantified by the Bradford method; Biorad) was separated on a 8%-12% SDS-PAGE and transferred to nitrocellulose membranes (Biorad). Membranes were blocked with 5% milk (Migros) in TBS-Tween and subsequently incubated overnight with primary antibody using concentrations that have either been previously established or used according to the manufacturer's recommendations. After a one hour incubation with a secondary antibody conjugated to horseradish peroxidase (Southern Biotech), the reaction was visualized by ECL (Pierce). For immunoprecipitation, cell lysates containing equal amounts of total protein were precleared for 1 hr with 40 µl of 50% (wt/vol) protein

G Plus-Sepharose beads (Amersham) and incubated for 2-3 hrs at 4°C with primary or control (IgG) antibodies. The beads were added for 1 hr and then washed extensively with lysis buffer; bound proteins were fractionated on a 8%-12% SDS-PAGE, and analyzed by western blotting as detailed above. Immunoprecipitations were performed using RIPA buffer with the notable exception of analyses shown in Figures 3e and 4d which were carried out using the JS lysis buffer (50 mM HEPES at pH 7.5, 1% glycerol, 50 mM NaCl, 1% Triton X-100, 1.5 mM MgCl<sub>2</sub>, 5 mM EGTA).

### **Confocal microscopy**

HeLa cells grown plated on glass coverslips were transiently transfected to express the EGFP-PP1 $\beta$  fusion protein. Twenty-four hours later, cells were washed in PBS and fixed for 15 min using 4% paraformaldehyde at room temperature (RT). Following three washes with PBS, cells were permeabilized on ice for 15 min using cold 0.1% Triton X-100 (v/v) in PBS. All staining reactions were performed at RT. To prevent unspecific binding, the cells were exposed for 30 min to 5% normal goat serum diluted in PBS. Next, the cells were incubated for 1 hr with the primary antibody, then washed three times in PBS and finally exposed for 30 min to the secondary antibody, an Alexa 555-conjugated goat anti-mouse antibody (Invitrogen). Prior to mounting, nuclei were stained with DAPI (4',6'-diamidino-2-phenylindole, diluted 1:10,000 in PBS from a 1 mg/ml stock) and washed twice with PBS. Coverslips were mounted with fluorescent mounting medium (Dako). Images were acquired using an LSM 510 Meta confocal microscope and LSM510 imaging software (Carl Zeiss).

### **Cell cycle analysis**

HeLa cells stably transfected with shRNA plasmid specific for the knock-down of SHOCA-2 or with a control plasmid were serum-starved for 24 hrs prior to a 12 hrs stimulation with EGF (100 ng/ml). After trypsinization and centrifugation,  $5 \times 10^5$  cells were resuspended in 0.5 ml of cold hypotonic PI (Propidium Iodide) staining solution (50  $\mu$ g/ml PI (Sigma), 0.1 mg/ml RNase A, 0.1% Triton X-100, 0.1%

sodium citrate). Following vigorous resuspension, the cells were incubated (15 min) on ice and then analyzed by flow cytometry (FACSCalibur, Becton Dickinson) using the FlowJo software (Tree Star, Oregon Corporation). G1, S and G2/M phases of cell cycle were defined using the mathematical Watson Pragmatic model.

### **Cell proliferation**

SW480 cells ( $5 \times 10^4$  cells/well) were cultured in 96-well plates in the absence or presence of AG1478 (10  $\mu$ M) or S3I-201 (100 $\mu$ M) in 10% serum condition. After 48 hours of treatment, cell proliferation was analyzed using CellTiter 96 aqueous nonradioactive cell proliferation assay kit (Promega, Madison, WI). The cell proliferation index was calculated as a percentage of the absorbance in relation to the untreated control cells.

### **Human tissues and DNA extraction**

The tissues from primary colorectal carcinomas and the macroscopically healthy-appearing neighboring mucosa (7cm proximal of tumor) were resected on the occasion of surgical cancer therapy. All the tissue samples were immediately submerged into RNA/later<sup>®</sup> RNA stabilization reagent (Qiagen) and kept at room temperature for several hours to guarantee a complete penetration by the reagent. The vials were then stored at -80°C until the tissue was further processed. DNA from approximate 20 mg of tissue was extracted with the QIAamp DNA Mini Kit (Qiagen) according to the manufacturer's specifications.

### **LOH and mutational analysis of the SH2D4A gene**

All canonical exons, the flanking intronic sequences as well as 5' and 3' UTR regions were subjected to PCR amplification (Kapa<sup>®</sup>HiFi™ system ; Kapa Biosystems, USA) and sequenced (BigDye terminator kit v1.1 on ABI 3130xl; both Applied Biosystems, USA) using the set of primers described in the supplementary table. Putative mutations were confirmed by bi-directional sequencing of a second independent PCR product. *In Silico* mutation impact prediction scores (SIFT:

<http://sift.jcvi.org> and PolyPhen-2: <http://genetics.bwh.harvard.edu/pph2>) were calculated for all somatic point alterations found. Loss of heterozygosity was assessed using 3 flanking microsatellite markers (D8S549, D8S1715 and D8S258 labeled at 5' end with 6-Carboxyfluorescein (6-FAM) dye) as well as 6 single nucleotide polymorphisms refSNP rs2280444, rs877386, rs1574288, rs17128221, rs4921637, rs36092909) located within the *SH2D4A* gene itself. The relative proportions of alleles 1 and 2 in non-tumor and tumor samples corresponds to the ratio of the heights of the corresponding peaks. An LOH index was calculated by dividing the ratio of allele 2 to allele 1 in the non-tumor DNA by the corresponding ratio in the tumor DNA. LOH positivity was defined as an LOH index of <0.5 (reflecting a substantial loss of allele 1 in the tumor sample) or >1.5 (indicative of substantial loss of allele 2).

#### **Identification and characterization of exon 4 skipping variant of the human *SH2D4A* gene by RT-PCR**

First strand cDNA was synthesized from 500ng of total RNA using random-primed reverse transcription (Verso cDNA; Thermo Scientific). PCR amplification was next carried out on a cDNA template, products were separated on a 2% agarose gel and visualized with ethidium bromide. Individual bands were excised from the gel, eluted (Quiaquick gel extraction kit; Quiagen, Germany), reamplified and sequenced.

#### **Conditions and primers for sequence, LOH and cDNA analyses**

For the *SH2D4A* sequence analysis the following PCR conditions were used: initial denaturation of 2min at 94°C followed by 35 cycles of 10sec at 94°C, of 15sec at 56°C and of 15sec at 72°C and a final elongation step of 3min at 72°C. When using the ABI Prism® TrueAllele® premix, the initial denaturation step was extended to 10 minutes. The following primers were used:

<b>Mutational Analysis</b>	<b>DNA sequence 5'→3'</b>
SH2D4A_x1_For	CGATTGCGCCCCGCCAGTCA
SH2D4A_x1_Rev	CTCCCATGCACCCATGCAAC

SH2D4A_x2_For	GGAAC TTTTGCCACAAGTAT
SH2D4A_x2_Rev	CTTGCACCTGCTTTGCTTG
SH2D4A_x3_For	ATTGCTACCTACAGATGTTC
SH2D4A_x3_Rev	AAGTGACTTTTGTGACCCTC
SH2D4A_x4_For	ACGTCTTTACACACCTAGC
SH2D4A_x4_Rev	AAATTCCGCACGTCTACAG
SH2D4A_x5_For	TCACATGCCCGTTGCTGT
SH2D4A_x5_Rev	ATGCCTAGGTATGCAGCCC
SH2D4A_x6_For	AGGAAACTGCTGGATTTGCT
SH2D4A_x6_Rev	ATCAGCTGAGAGCCTGGAAG
SH2D4A_x7_For	ATAAGCAGACAACAACTG
SH2D4A_x7_Rev	CTCATACAATTCCCCTTTAA
SH2D4A_x8_For	TATATGACTTTTGAGGGC
SH2D4A_x8_Rev	TTCCTGGTAGATTAGCACAG
SH2D4A_x9_For	TGCTCTTATTGTGTCTTATTTAGGC
SH2D4A_x9_Rev	CTCAGAAGGCCAAATTAAGTT
SH2D4A_x10_For	GGTCACATGTAGATAAGAAA
SH2D4A_x10_Rev	AAAACAAACTCAGCGATGT
<b>LOH Analysis</b>	
D8S1715_For	CAGGTGATGTCCCAGAGG
D8S1715_Rev	CGAACATGAATTAGAAATCCAGTG
D8S258_For	CTGCCAGGAATCAACTGAG
D8S258_Rev	TTGACAGGGACCCACG
D8S549_For	AAATGAATCTCTGATTAGCCAAC
D8S549_Rev	TGAGAGCCAACCTATTTCTACC
rs36092909_For	ACAAATGCCACTGCAACATT
rs36092909_Rev	TACATCATTAAGGTCATAC
<b>cDNA Analysis</b>	
SH2D4A_cx2.1_For	CAGACCAAAGAAAGAGAATGGCA
SH2D4A_cx2_Rev	ATCTGCCAACATTTGTTGAA
SH2D4A_cx2.1_Rev	CTTTGGATTTTCGCAGAGATG

### **Multiplex-ligation dependent-probe-amplification (MLPA)**

Chromosome 8 copy-number variation was investigated by multiplex ligation-dependent probe amplification assay (kit P014-1A, MRC Holland) according to the manufacturer's protocols. Calculations were performed using GeneMarker software (SoftGenetics). Gene dosage with relative value of 1 is expected for two copies, whereas values of 0.5 and 1.5 indicate loss and gain of one copy, respectively. MLPA results indicative of a somatic deletion were independently confirmed in at least one additional experiment.

## Methylation analysis

Genomic DNA from primary tumor samples and adjacent matched healthy appearing mucosa, respectively, was extracted using the QIAamp DNA Mini Kit (Qiagen), and bisulfite converted using the EZ DNA Methylation Kit™ (Zymo Research). Promoter associated CpG-island regions of *SH2D4A* were amplified by PCR and biotin labeled either directly using biotin-labelled reverse primers or during a second PCR step using a biotinylated universal primer annealing to a sequence introduced by the reverse primer used in the first PCR. PCR conditions and primer sequences are provided on supplementary table. Labelled PCR products were purified (QIAquick® PCR Purification Kit, Qiagen), bound to streptavidin-covered sepharose beads and analyzed by pyrosequencing using the PyroMark Q24 pyrosequencing system (Qiagen). Sequencing-assay design as well as quantification of methylation was performed with the PyroMark Q24 1.0.10 software. Peripheral blood DNA treated with M.SssI methyltransferase (New England BioLabs) and subsequently bisulfite treated served as positive control.

## Conditions and primers for methylation analysis

PCR Amplification of CpG islands	
1 <sup>st</sup> CpG island forward	5'-GTT TTA TGA ATT AAG AGA GGA GAG GAT AAG TTT ATT T-3'
1 <sup>st</sup> CpG island reverse	5'-GCC CCC GCC CGC CCC CAA CAC CTA CCA ATA C-3'
Biotinylated primer	5'-Bio-GCC CCC GCC CG-3'
2 <sup>nd</sup> CpG island forward	5'-AG GTT GTA TGG GTG TAT GGG AG-3'
2 <sup>nd</sup> CpG island biotinylated reverse	5'-Bio-CAA CTC AAA AAA CCT ACA AAC TAT AAC TCC TTC-3'
<u>PCR conditions:</u>	
1 <sup>st</sup> CpG island 1 <sup>st</sup> PCR	Enzyme activation: 95°C for 2min 40 cycles of Denaturation: 95°C for 30 sec Annealing: 61°C for 25 sec Extension: 72°C for 30 sec



1 <sup>st</sup> CpG island 2 <sup>nd</sup> PCR	3.5mM MgCl <sub>2</sub> , 0.5μM primer  Enzyme activation: 95°C for 2min 20 cycles of Denaturation: 95°C for 30 sec Annealing: 61°C for 30 sec Extension: 72°C for 30 sec
2 <sup>nd</sup> CpG island	2.5mM MgCl <sub>2</sub> , 0.5μM primer  Enzyme activation: 95°C for 4min 35 cycles of Denaturation: 95°C for 30 sec Annealing: 61°C for 30 sec Extension: 72°C for 55 sec  2.5mM MgCl <sub>2</sub> , 0.25μM primer
<b>Pyrosequencing primers</b>	
<b>1<sup>st</sup> CpG island</b>	
Primer 1	5'- GGA GGA GAG GAT AAG T -3'
Primer 2	5'- GTT TTG GAG AGT TTT TAG -3'
Primer 3	5'- GTT ATA AGA GTT GTT TTG -3'
Primer 4	5'- GGT TAA GTG GAT GTG -3'
<b>2<sup>nd</sup> CpG island</b>	
Primer 1	5'- TAT GGG TGT ATG GGA G -3'
Primer 2	5'- GGG TTT AGG TG -3'
Primer 3	5'- GTT GGT TAG GTG AG -3'
Primer 4	5'- AGG GAG AGG GTT TG -3'
Primer 5	5'- GGG GAG TGT AGG TTT -3'
Primer 6	5'- TAT GGA AAT TTT TTA GGA GG -3'

Primer 7	5'- TGG GGT TTG GGA G -3'
Primer 8	5'- GGG TTG GTT TAG ATT TAG -3'

### **Gel-Shift, Shift-Western Blot, and Western Blot**

The gel-shift method employed was adopted from Harrington et al<sup>9</sup>. A 28 bp long, fluorescein-labeled dsDNA probe containing a Sp1 recognition site was incubated for 15 min at RT with SW48 nuclear extract in the presence of increasing concentrations of an unlabeled 28mer oligomer. The competitor had either a methylated or an unmethylated Sp1 recognition site. The mixture was then run in a 6% native polyacrylamide gel and fluorescence signals were subsequently detected using a biomolecular imager (Typhoon 9400, Amersham Biosciences). Fluorescence intensity of shifted bands was normalized against the sum of signals provided by both unbound oligomers and shifted bands. Experiments were performed in quadruplicates.

Bands from the polyacrylamide gel were transferred onto a nitrocellulose membrane using the wet blotting method and Sp1 was detected via chemiluminescence with anti-human Sp1 primary antibody (polyclonal rabbit, Millipore) and a HRP-labeled anti-rabbit secondary antibody (GE healthcare).

Sequences of oligomers:

Positive control: Forward: 5'-ACG TTG CAG CCG GGG CGG GGC TTC TGC A-3', Reverse: 5'-TGC AGA AGC CCC GCC CCG GCT GCA ACG T-3'

Methylated competitor: Forward: 5'-CCT CAC CCC CGC CTC CAC CCC TTC GCG G-3', Reverse: 5'-F-CCG CGA AGG GCT GGA GGC<sup>m</sup> GGG GGT GAG G-3'

Unmethylated competitor: Forward: 5'-CCT CAC CCC CGC CTC CAC CCC TTC GCG G-3', Reverse: 5'-F-CCG CGA AGG GCT GGA GGC GGG GGT GAG G-3'

## Supplemental References

1. Rhodes, D. R., Kalyana-Sundaram, S., Mahavisno, V., Varambally, R., *et al.* OncoPrint 3.0: genes, pathways, and networks in a collection of 18,000 cancer gene expression profiles. *Neoplasia* **9**, 166-180 (2007).
2. Hubbard, T. J., Aken, B. L., Ayling, S., Ballester, B., *et al.* Ensembl 2009. *Nucleic Acids Res* **37**, D690-D697 (2009).
3. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: A novel method for fast and accurate multiple sequence alignment. *J Mol Biol* **302**, 205-217 (2000).
4. Hirsch, F. R., Varella-Garcia, M., Bunn, P. A., Di Maria, M. V., *et al.* Epidermal growth factor receptor in non-small-cell lung carcinomas: correlation between gene copy number and protein expression and impact on prognosis. *J Clin Oncol* **21**, 3798-3807 (2003).
5. Bewick, V., Cheek, L. & Ball, J. Statistics review 8: Qualitative data - tests of association. *Crit Care* **8**, 46-53 (2004).
6. Bewick, V., Cheek, L. & Ball, J. Statistics review 12: survival analysis. *Crit Care* **8**, 389-394 (2004).
7. Hess, D., Keusch, J. J., Oberstein, S. A., Hennekam, R. C. & Hofsteenge, J. Peters Plus syndrome is a new congenital disorder of glycosylation and involves defective Omicron-glycosylation of thrombospondin type 1 repeats. *J Biol Chem* **283**, 7354-7360 (2008).
8. Perkins, D. N., Pappin, D. J., Creasy, D. M. & Cottrell, J. S. Probability-based protein identification by searching sequence databases using mass spectrometry data. *Electrophoresis* **20**, 3551-3567 (1999).
9. Harrington, M. A., Jones, P. A., Imagawa, M. & Karin, M. Cytosine methylation does not affect binding of transcription factor Sp1. *Proc Natl Acad Sci U S A* **85**, 2066-2070 (1988).

# **Colorectal mucosa of healthy individuals displays part of the CpG island methylator phenotype signature**

Patric Urfer<sup>1</sup>, Faiza Noreen<sup>1</sup>, Stefan Weis<sup>1</sup>, Mirco Menigatti<sup>1</sup>, Fredy Siegrist<sup>2</sup>, Corina Kim-Fuchs<sup>3</sup>, Oliver Tschaler<sup>3</sup>, Rolf Schlumpf<sup>3</sup>, Ulli Certa<sup>2</sup>, Primo Schär<sup>1</sup>, Kaspar Truninger<sup>1,4</sup>

<sup>1</sup> Department of Biomedicine, University of Basel, Basel, Switzerland

<sup>2</sup> Roche, Basel, Switzerland

<sup>3</sup> Department of Surgery, Cantonal Hospital Aarau, Aarau, Switzerland

<sup>4</sup> Gastroenterologie Oberaargau, Langenthal, Switzerland

**Corresponding author** Kaspar Truninger, Gastroenterologie Oberaargau, St. Urbanstrasse 67, CH 4901 Langenthal, Switzerland. Phone: +41-62-916 3164; Fax +41-62-916 4148; E-mail: [k.truninger@sro.ch](mailto:k.truninger@sro.ch)

## Abstract

Gene silencing by aberrant DNA methylation is a hallmark of colorectal cancer (CRC). However, little is known about its precise timing during tumor development. We aimed to identify novel targets of aberrant methylation early in colorectal carcinogenesis. In addition to *MGMT* and *hMLH1*, we determined the methylation in five gene promoters in matched tumor tissue, cancer-associated mucosa (CAM) and blood samples from 106 unselected CRC patients, and in 62 normal colorectal mucosa (NM) samples from 31 healthy individuals, using locus-normalized quantitative methylation-specific PCR (ln-qMSP). We identified *FOXF1*, *CA4*, *NPY1R*, *IFITM1* and *GREM1* as novel targets of aberrant methylation in CRC. The frequency of tumors showing > 99% cancer-specific hypermethylation was 66% for *FOXF1*, 36% for *MGMT*, 21% for *CA4*, 19% for *NPY1R*, 15% for *IFITM1*, 14% for *hMLH1* and 13% for *GREM1*. This seven-marker panel separated CRC from CAM in 82% of all and in 94% of right-sided tumors with a specificity of 99%. We observed concomitant methylation of *CA4*, *NPY1R* and *GREM1* with *hMLH1* in CRC, suggesting them to represent novel CIMP markers. Using highly sensitive ln-qMSP and genome-wide methylation profiling, we also found synchronous low level methylation in *CA4*, *NPY1R*, *GREM1* and several validated CIMP markers, but not in *hMLH1*, in virtually all CAM and NM from healthy individuals. Thus, concomitant low-level methylation of some, but not all CIMP markers is readily detectable in the colorectum of healthy individuals, suggesting that part of the CIMP signature reflects a physiological condition.

## Introduction

Gene silencing by DNA hypermethylation in promoter CpG islands (CGIs) is a hallmark of colorectal cancer (CRC). Yet, little is known about the underlying causes and the precise timing of these epigenetic changes during neoplastic transformation. A key question is whether aberrant DNA methylation observed in cancers originates from a random instability in DNA methylation and selection during carcinogenesis, or from specific defects in the methylation machinery. The identification of a subgroup of CRC displaying widespread hypermethylation at gene promoters, a phenomenon referred to as the CGI methylator phenotype (CIMP), argues for a specific pathway driving aberrant methylation in CIMP-positive cancers (1). CIMP-positive cancers indeed present with distinct clinical features, such as proximal location, older age at diagnosis and female gender, and they appear to be defined by specific patterns of gene promoter hypermethylation (1-6). The cellular and molecular events underlying the formation of CIMP-positive as opposed to CIMP-negative CRC, however, have not been identified.

Some genes undergo progressive methylation in their promoter CGIs in an age-dependent manner in normal colorectal mucosa. The causes and mechanisms underlying age-related methylation changes are largely unknown. While many targets displaying hypermethylation in CRC are thought to result from aberrant *de novo* methylation in normal colonic cells during aging, others exhibit cancer-specific methylation as observed predominantly in CIMP-positive tumors (1). Only few studies, however, investigated normal colorectal mucosa to dissect methylation changes occurring early in colorectal carcinogenesis. Probably the majority of methylation changes accumulated during aging and observed in CIMP-positive tumors is not causally involved in colorectal carcinogenesis and accompany rather than initiate tumorigenesis. By contrast, the pathogenetic consequence of inactivation of the mismatch repair gene *hMLH1* by CGI hypermethylation is well established and provides the molecular basis for sporadic microsatellite (MS) unstable CRC, which occur mainly in the proximal colon of elderly females (7, 8). Although *hMLH1* is a robust CIMP marker, i.e. cancer-specific, we have recently demonstrated that methylation in its promoter is detectable at a low level in the normal appearing colorectal mucosa of healthy individuals (9, 10). Notably, the epidemiological pattern of sporadic MS-unstable CRC was already established in healthy individuals, i.e., the age-related methylation increase in *hMLH1* was observed in the proximal colon of healthy females. Thus, targets formerly being proposed to be cancer-specifically methylated are in fact also methylated in an age-dependent manner.

In the present study, we aimed to (i) identify gene promoter targets affected by aberrant promoter methylation early in colorectal carcinogenesis; (ii) establish cancer-specific methylation thresholds for these targets (iii) compile a highly cancer-specific marker panel that covers the majority of CRC; and (iv) investigate their DNA methylation patterns in normal colorectal mucosa from healthy individuals. Towards these aims, we determined methylation levels of seven gene promoters in tumor tissue, matched adjacent normal mucosa and blood samples of 106 unselected CRC patients and in normal colorectal mucosa biopsies of 31 healthy individuals using sensitive locus-normalized qMSP (ln-qMSP).

## **Material and Methods**

### **Recruitment of patients and healthy individuals and tissue sampling**

Tissue and blood samples from 106 unselected colorectal cancer patients undergoing surgical therapy at the Department of Surgery of the cantonal hospital of Aarau, Switzerland, were collected. Informed consent was received from all patients and the study protocol was approved by the local medical ethical committee. From each patient, a tissue sample of the primary tumor and of apparently normal mucosa 7 cm proximal to the primary tumor (cancer associated mucosa, CAM) were collected in the operating theatre immediately after surgical removal of colorectal segment. All samples were submerged in RNAlater and stored at -80°C until further processing. All primary tumors were histopathologically confirmed adenocarcinomas. Blood samples were drawn from each patient and stored in ethylenediaminetetraacetate (EDTA) at -80°C. From 31 healthy individuals without colorectal polyps, we obtained biopsies of normal mucosa from the cecum and the sigmoid colon at their screening colonoscopy at the cantonal hospital of Uri, Switzerland (9). The stage of disease refers to (11). Tumors from the cecum to the splenic flexure and from the splenic flexure to the rectum were termed right- and left-sided, respectively.

### **Selection of gene promoters for methylation analysis**

To identify novel gene promoters affected by aberrant methylation early in colorectal carcinogenesis, we explored data of a previously published transcriptome analysis of 32 matched adenoma and normal colorectal tissue samples (12). Genes fulfilling the following criteria were selected for further examination: (i) mRNA expression downregulated (<90%) in at least 26 adenomas compared to matched normal tissue, (ii) presence of a promoter CpG island ( $\geq 200$  bp length, GC percentage > 50%, observed/expected CpG ratio > 60%), (iii) not previously reported as a target of aberrant methylation in CRC, (iv) cancer-specific methylation (specificity), and (v) sufficient diagnostic sensitivity in a pilot study performed in matched cancer, CAM and blood samples from 20 patients. More detailed information about the reanalysis of the mRNA expression chips is given in the *supplemental data*.

### **DNA extraction and bisulfite conversion**

Genomic DNA from colorectal tissue and colorectal cancer cell lines was extracted with the QIAamp DNA mini kit (Qiagen, Hilden, Germany). An RNase A (Qiagen) treatment was included in the protocol according to the manufacturer's protocol. Genomic DNA from blood



samples was extracted with the NucleoSpin Blood L Kit (Macherey-Nagel, Dueren, Germany). For methylation analysis, we converted 2 µg of purified DNA with sodium bisulfite using the EZ DNA Methylation Kit<sup>®</sup> (Zymo Research, Orange, USA).

### **Methylation analysis, locus-normalized quantitative methylation-specific PCR and bisulfite sequencing**

Methylation of the selected *target* promoters was analyzed by locus-normalized quantitative methylation specific PCR (ln-qMSP), which is a modification of the quantitative MSP assay developed and validated in our laboratory (9). To achieve accurate locus normalization, we introduced standardizer primer pairs for each target, amplifying input DNA after bisulfite conversion at the locus examined, thereby minimizing copy number biases. Methylation at target sequences was then expressed as percentage of methylated alleles (PMA), ranging from 0% (unmethylated) to 100% (fully methylated). ln-qMSP results were confirmed by bisulfite sequencing. Sequences of primers used for qMSP and bisulfite sequencing, the regions and CpG sites examined by both methods as well as detailed information on the assays are given in the *Supplemental data*.

### **Genome-wide promoter methylation analysis**

1 µg of genomic DNA was sodium bisulfite converted with the Zymo EZ DNA Methylation kit (zymo Research,D5002) by following the manufacturer's protocol with modifications for the illumina Infinium methylation assay. Briefly, after the denaturation step, DNA was incubated in a thermocycler for 16 cycles at 95°C for 30 seconds and 50°C for 60 minutes. DNA was eluted from the column in 12 µl elution buffer. Concentration of the eluted samples were measured using Nanodrop-1000. The methylome state of bisulfite-converted genomic DNA was analyzed using illumina's Infinium Human Methylation27 beadchip kit (WG-311-1202), interrogating 27,578 CpG loci covering more than 14,000 human RefSeq genes at single nucleotide resolution. Chip process and data analysis were performed by using the kit reagents provided according to the manufacturer's manual. Raw signals were normalized using background subtraction in the BeadStudio v3.0 software. Methylation values for each CpG locus are expressed as beta (β) - value representing a continuous measurement from 0 (unmethylated) to 1 (fully methylated). For each CpG site, the β value is calculated by dividing the signal intensity of methylation-detection probe (M) by the sum of signals for both methylation-detection probe and non-methylation detection probe (U):  $\beta = \text{Max}(M,0) / [\text{Max}(M,0) + \text{Max}(U,0) + 100]$ .

## **mRNA expression in CRC and CAM tissue**

Matched CRC and CAM samples from 10 patients with extensively methylated tumors were selected to analyze the correlation between *FOXF1* methylation and gene expression. Simultaneous purification of DNA and RNA from tissue samples was performed with the All-Prep DNA/RNA/Protein Mini Kit<sup>®</sup> (Qiagen, Germany). For protocol modifications, cDNA synthesis and qRT-PCR conditions see *Supplemental data*.

## **Promoter derepression by TSA and 5-Aza-cytidine treatment**

To correlate the promoter methylation levels with gene activity, the human colorectal cancer cell lines CO115, SW48 and GP5d, showing full methylation in the gene promoters of interest, were treated with the demethylating agent 5-Aza-cytidine and/or the histone-deacetylase inhibitor Trichostatin A (TSA). After total RNA extraction and cDNA synthesis mRNA levels were analyzed by quantitative RT-PCR. For detailed information on the drug treatment, the qRT-PCR conditions, and cell culture conditions see *Supplemental data*.

## **Statistical analyses**

To analyze the correlation of *FOXF1* gene methylation and RNA expression in patient samples, the probability for the number of matched sample pairs which show an RNA expression higher in the poorly methylated CAM than in the extensively methylated CRC was calculated by a binomial probability function. For the analysis of concomitant target methylation, PMA values of all targets in NM and CAM samples were clustered by means of unsupervised hierarchical clustering (Euclidian distance) with the MeV software Version 4.3 (<http://mev.tm4.org/>). Methylation levels of all targets were subjected to a repeated measures ANOVA analysis for NM (normally distributed), and to a Friedman non-parametric test for CAM samples (not normally distributed). Both statistical analyses were followed by a Tukey's multiple comparison test and a Dunn's multiple comparison test, respectively. Resulting P-values were considered an inverse measure for concomitant target methylation with  $P > 0.05$  or  $\leq 0.05$  representing concomitant or independent, respectively. Cancer PMA values were categorized in either "hypermethylated" or "not hypermethylated" according to the target-specific threshold established and contingency tables were created for pairs of all targets. The probabilities of the resulting contingency tables were calculated by a Fisher's exact test.

## Results

### Identification of methylation targets of early colorectal carcinogenesis

To identify gene promoters affected by aberrant cytosine methylation early in colorectal carcinogenesis, we explored previously published gene expression profiles of matched adenoma and normal colorectal tissue samples (15). We screened the data for genes with reduced mRNA expression in adenomas that have a CGI in their promoter not previously associated with aberrant methylation in CRC. We analyzed five candidate genes in detail for cancer-specific DNA methylation: *Forkhead box F1 (FOXF1)*, *Carbonic anhydrase IV (CA4)*, *Neuropeptide Y receptor Y1 (NPY1R)*, *Gremlin 1 (GREM1)* and *Interferon-induced transmembrane protein 1 (IFITM1)*.

We determined CpG methylation as percentage of methylated alleles (PMA) in the promoters of these five genes along with two known targets of aberrant methylation in CRC, the DNA repair genes *human MutL Homolog 1 (hMLH1)* and *O-6-methylguanine-DNA methyltransferase (MGMT)*. We performed these measurements in DNA of matched colorectal cancer tissue (CRC), CRC-associated mucosa (CAM) and blood samples from 106 unselected patients (Table 1), using a highly sensitive, locus-normalized methylation-specific quantitative PCR assay (In-qMSP). The mean PMA at all targets increased from blood to CAM by approximately two-fold and from CAM to tumor tissue by one order of magnitude (Fig. 1A, Supplementary Table S1). Comparing the methylation levels in these three tissues allowed us to establish a methylation threshold for each target, discriminating cancer from CAM with  $\geq 99\%$  specificity. These thresholds ranged from 4% methylated alleles for *FOXF1* to 7% for *IFITM1* (Fig. 1A, Table 2). Tumor samples showing methylation levels above this cut-off were classified as hypermethylated. Accordingly, the frequency of CRCs hypermethylated in these promoters were 66% for *FOXF1*, 36% for *MGMT*, 21% for *CA4*, 19% for *NPY1R*, 15% for *IFITM1*, 14% for *hMLH1* and 13% for *GREM1* (Fig. 1B, Table2). According to the criterion that one of these targets must be methylated above threshold, this seven-marker panel separated cancer from CAM in 87 of the 106 unselected CRC cases (82%) with a specificity of 99%. When applied to right-sided CRCs, the coverage reached 94% (30/32) at the same level of specificity (Fig. 1C).

To address the effect of promoter methylation on the expression of the newly identified hypermethylation target genes, we measured the respective mRNA levels in matched cancer tissue and CAM (*FOXF1*) or in colorectal cancer cell lines (*FOXF1*, *CA4*, *NPY1R*, *IFITM1* and *GREM1*) before and after treatment with the demethylating agents 5'-Aza-Cytidine and/or the

histone deacetylase inhibitor Trichostatin A (TSA). Consistent with a gene silencing effect of the promoter methylation, *FOXF1* expression was significantly reduced in hypermethylated cancers ( $p = 0.04$ ) (Supplementary Fig. S2) and promoter demethylation partly restored RNA expression for all genes, except *GREM1*, in CRC cell lines (Supplementary Fig. S3).

We then investigated the association of promoter hypermethylation patterns in CRCs with clinical parameters (Table 1), including gender, tumor location and stage of disease. Tumors hypermethylated in *IFITM1* occurred more frequently in men (20.9% vs. 5.2%,  $P = 0.046$ ), while *FOXF1*, *MGMT*, *NPY1R*, *GREM1*, and *hMLH1* were more frequently hypermethylated in women, with the difference for *hMLH1* being statistically significant (25.6% vs. 7.5%,  $P = 0.018$ ) (Supplementary Table S2, Supplementary Fig. S4A). All targets, except *IFITM1*, were more frequently hypermethylated in proximal than in distal CRC (*FOXF1* 91% vs. 55%,  $P = 0.0003$ ; *CA4* 59% vs. 4%,  $P < 0.0001$ ; *NPY1R* 34% vs. 12%,  $P = 0.01$ ; *GREM1* 31% vs. 5%,  $P = 0.0008$ ; *hMLH1* 34% vs. 5%,  $P = 0.0003$ ) (Fig. 1B, Supplementary Table S3). *MGMT* was found hypermethylated more frequently at advanced tumor stages ( $P = 0.049$ ), whereas none of the other targets showed an association between tumor stage and hypermethylation (Supplementary Table S4, Supplementary Fig. S4B).

### **Association of the Newly Identified Methylation Targets with CIMP**

We then asked whether the methylation targets identified in this study acquire aberrant methylation as a consequence of an overall methylator phenotype as proposed for CIMP. The analysis of all CRC samples according to their methylation status at all promoter targets (hypermethylated vs. not hypermethylated), revealed concomitant hypermethylation of *CA4*, *NPY1R* and *GREM1* with the established CIMP marker *hMLH1*, while *MGMT* and *IFITM1* appear to be methylated independently (Fig. 3C, Supplementary Table S5), and this was confirmed by an unsupervised hierarchical clustering (Fig. 2B). Finally, we performed a genome-wide promoter methylation analysis in three CIMP-positive and four CIMP-negative cancers, classified according to their *BRAF*<sup>V600E</sup> mutation status. As controls, eight normal colorectal mucosa biopsies and one blood sample from healthy subjects were included in the analysis. The clinical characteristics of the cancer patients and healthy probands are shown in Supplementary Table S6. We used single CpG site-resolving bead chip technology (Infinium HumanMethylation27, Illumina), which simultaneously captures the methylation status of 27'578 CpG sites. The methylation levels of *CA4*, *GREM1* and the validated CIMP markers *CACNA1G*, *NEUROG1* and *hMLH1* were significantly higher in CIMP-positive compared to CIMP-

negative tumors, while the difference was clearly notable but not statistically significant for *NPY1R* (Fig. 4). Hence, our data suggests that *CA4* and *GREM1* represent novel CIMP markers while *NPY1R* appears to be loosely associated with CIMP as well. Consistently, all CRCs hypermethylated in *CA4*, *NPY1R*, *GREM1* and *hMLH1* (n = 8) located exclusively to the right-sided colon, which is a typical clinical feature of CIMP ( $P = 0.00003$ ) (Fig 2A). All CIMP-positive tumors, as defined by *hMLH1* hypermethylation, were also hypermethylated in *FOXF1*, suggesting that these cancers represent a subgroup of *FOXF1*-hypermethylated CRC.

### **CIMP Signatures in Non-Cancerous Mucosa and Healthy Individuals**

We reported that methylation at the *hMLH1* promoter, a bona fide CIMP marker, are detectable in the normal colorectal mucosa of healthy individuals (9). We investigated the methylation levels of all seven targets in CAM of the 106 CRC patients and, except the two DNA repair genes, in normal mucosa of 31 healthy subjects. We designated methylation levels above  $\geq 0.01\%$  (detection limit), but below cancer-specific thresholds as low-level methylation.

In CAM, the frequency of samples with low-level methylation ranged from 95.9% (*GREM1*) to 100% (*FOXF1*, *CA4*, *NPY1R* and *IFITM1*) with mean PMA ranging from 0.99% (*FOXF1*) to 2.49% (*IFITM1*) (Supplementary Table S7). Regarding *hMLH1* and *MGMT*, both the frequency of low-level methylation and mean PMA were significantly lower (*hMLH1* 36.8%, 0.16%; *MGMT* 55.6%, 0.28%). Unsupervised hierarchical clustering in CAM revealed concomitant methylation of *CA4*, *NPY1R* and *GREM1*, but not with *hMLH1* (data not shown), which was confirmed by non-parametric repeated measures multiple comparison testing (Friedman and Dunn's test) (Fig. 3B, Supplementary Table S8). Concomitant methylation of these three targets was independent of the anatomic location along the colorectum (Fig. 3B).

In the colorectal mucosa of healthy individuals, the frequency of low-level methylation ranged from 90.3% (*FOXF1*) to 100% (*NPY1R*, *IFITM1*) (Supplementary Table S9), which was clearly higher than the frequencies measured for *hMLH1* (70%) and *MGMT* (44%) as reported previously (9). Unsupervised hierarchical clustering and multiple comparison ANOVA testing revealed concordant methylation of *CA4*, *NPY1R* and *GREM1*, again irrespective of the anatomic location (Fig. 3A, Supplementary Table S10). Thus, the only three targets being concordantly methylated in cancer tissue, CAM and normal mucosa from healthy individuals were *CA4*, *NPY1R* and *GREM1*, as indicated by the gray area in Fig. 3A-C. We investigated each four samples of normal mucosa from healthy individuals aged 50 and 75, respectively, for genome-wide promoter methylation, using the aforementioned technology. The clinical characteristics

of the healthy subjects are shown in Supplementary Table S6. This assay showed low-level methylation of *CA4*, *NPY1R* and *GREM1* as well as in the established CIMP markers *CACNA1G*, *NEUROG1*, *IGF2*, *CDKN2A* and *hMLH1* in all samples (Fig. 4). Hence, methylation in validated and in the newly identified CIMP markers is readily detectable at low levels in healthy individuals.

## Discussion

Our effort was to identify genes that show cancer-specific aberrant promoter methylation early in colorectal carcinogenesis. We could establish  $\geq 99\%$  cancer-specific methylation thresholds for seven gene promoters and identified *FOXF1*, *IFITM1*, *CA4*, *NPY1R* and *GREM1* as novel potent targets of aberrant methylation in CRC, the later three targets representing novel CIMP markers. Though all investigated targets displayed highly cancer-specific methylation thresholds, they showed marked methylation differences in CAM and healthy individuals, suggesting different functional roles early in colorectal carcinogenesis.

Detection of aberrant DNA methylation in normal mucosa of healthy individuals may serve as a biomarker for CRC risk. Particularly, monitoring increasing methylation levels of targets with cancer-specific methylation thresholds might identify individuals at risk to develop CRC. For optimal sensitivity, there is a need for specific markers that cover the multiple pathways leading to CRC. The identification of a highly cancer-specific marker panel, which has a sensitivity of 83% for all and 94% for right-sided CRC, offers the possibility to monitor early methylation patterns in the colorectal carcinogenesis. The single most sensitive marker *FOXF1* was hypermethylated in 66% of all and in 91% of proximal tumors, a clearly higher frequency compared to other targets that are hypermethylated in CRC (12). *FOXF1* has previously been described as a target of aberrant methylation in the SW48 colorectal cancer cell line, but was not systematically investigated in CRC (13).

Several lines of evidence demonstrate that the genes investigated here become differentially methylated early in colorectal carcinogenesis. In CAM, both the frequency and levels of low-level methylation were higher in the newly identified targets compared to *hMLH1* and *MGMT*. This difference is also reflected by two different patterns of methylation distribution observed in CAM, i.e., a narrow and discontinuous distribution of the novel targets and a more wide and continuous pattern in the two DNA repair genes (Fig. 1A). Similarly in normal mucosa of healthy individuals, low-level methylation of *FOXF1*, *CA4*, *NPY1R*, *IFITM1*, and *GREM1* was detected in  $> 90\%$  of samples compared to 70% of *hMLH1* and 44% of *MGMT*, as reported in our previous study (9). These data suggest low-level methylation in the newly identified targets to be constitutive and their cancer-specific methylation to be characterized not by the presence or absence of methylation itself, but rather by levels. Cancer-specific methylation in these promoters might therefore derive from clonal expansion of cells harboring physiological low-level methylation in these targets. Hence, methylation of *FOXF1*, *CA4*, *NPY1R*, *IFITM1*,

and *GREM1* is rather consequence than cause of neoplastic transformation in CRC development (*constitutive* type methylation markers). By contrast, any methylation in *hMLH1* and *MGMT* in normal mucosa probably occurs *de novo* in aging mucosa and is of pathologic functional consequence (*oncogenic* type methylation markers).

We addressed the question of whether the newly identified targets represent markers of CIMP, designating a subgroup of tumors with widespread CpG islands methylation. Our analyses, including a genome-wide promoter methylation assay, showed concordant methylation of *CA4*, *NPY1R* and *GREM1* with *hMLH1* and other validated CIMP markers (e.g. *CACNA1G*, *NEUROG1*) in cancer tissue and we propose that these genes represent newly identified CIMP targets. In addition to the molecular data, tumors hypermethylated in *CA4*, *NPY1R* and *GREM1* showed the typical right-sided location of CIMP-positive tumors. Notably, not all CIMP loci become simultaneously aberrantly methylated during the neoplastic process. We observed concomitant methylation of *hMLH1* with *CA4*, *NPY1R* and *GREM1* in tumor tissue, while the later three targets were not methylated together with the DNA repair gene in CAM and normal mucosa of healthy individuals. By contrast, *CA4*, *NPY1R* and *GREM1* also showed concomitant methylation in CAM and healthy individuals (Fig 3). Our genome-wide analysis revealed methylation not only in our newly identified targets, but also in several established CIMP markers in all samples of normal mucosa from healthy individuals. Hence, low-level methylation of CIMP markers is not rare, but readily detectable in healthy individuals (14). These data therefore support our previous finding that targets being proposed to be cancer-specifically methylated are in fact also methylated in healthy individuals. However, methylation changes in CIMP markers develop not uniformly in early colorectal carcinogenesis, although these markers ultimately display concomitant hypermethylation in a specific CRC group.

CIMP-positive cancers present mainly in the proximal colon, however, concomitant methylation of *CA4*, *NPY1R* and *GREM1* occurred independently of the anatomic location along the colorectum in CAM and normal mucosa of healthy individuals (Fig 3). This suggests that low-level methylation in these promoters derives from the same cell population in the differentiated colonic epithelium along the colorectum. Hence, the mechanism underlying the simultaneous methylation of *CA4*, *NPY1R* and *GREM1* is not carcinogenic *per se* and not initiated during cancer initiation, but rather reflects a physiological condition in healthy individuals. A location-specific trigger might drive the aberrant proliferation of the cell population in the right-sided colon, thereby increasing methylation above cancer-specific thresholds and establish the CIMP phenotype.



Sporadic MS-unstable CRC develops mainly in the right-sided colon of elderly females, indicating that that epigenetic instability in *hMLHI* is both location- and gender-specific. In our previous study we showed that the age-related increase of *hMLHI* methylation is most pronounced in the proximal colon of elderly females. Thus, the methylation signature underlying sporadic MS-unstable CRC is yet recognizable in the normal mucosa of healthy individuals. Integrating our results, we propose the following model for the development of CIMP-positive cancer (Fig. 5): a cell type in the normal mucosa along the colorectum of aging individuals harbors constitutive methylation in a group of gene promoters, including our novel and several established CIMP markers, but not in *hMLHI*. Critical genetic and/or epigenetic events, such as activating *BRAF*<sup>V600E</sup> mutation and/or aberrant *hMLHI* methylation, might occur in this cell population, thereby transforming them into a CIMP cancer stem cell. Clonal expansion of this cell type is associated with progressive methylation at selected genes and results in the establishment of the CIMP-positive phenotype. *hMLHI* therefore not only represent a genetic, but also an epigenetic gatekeeper.

Taken together, the methylation patterns of genes displaying highly cancer-specific thresholds are not uniform in healthy individuals. The occurrence of aberrant methylation in CIMP markers is not homogeneous early in colorectal carcinogenesis and parts of the CIMP methylation signature, characterizing a distinct cancer phenotype, are already established in the normal mucosa of healthy individuals. Our data indicate that only some genes initiate neoplastic transformation, while others achieve cancer-specific methylation and accompany the CIMP phenotype by clonal expansion of cells harboring constitutive methylation at various loci.

## References

1. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. *Proc Natl Acad Sci U S A*. 1999;96:8681-6.
2. Nosho K, Irahara N, Shima K, Kure S, Kirkner GJ, Schernhammer ES, et al. Comprehensive biostatistical analysis of CpG island methylator phenotype in colorectal cancer using a large population-based sample. *PLoS One*. 2008;3:e3698.
3. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet*. 2006;38:787-93.
4. Ferracin M, Gafa R, Miotto E, Veronese A, Pultrone C, Sabbioni S, et al. The methylator phenotype in microsatellite stable colorectal cancers is characterized by a distinct gene expression profile. *J Pathol*. 2008;214:594-602.
5. Ogino S, Kawasaki T, Kirkner GJ, Loda M, Fuchs CS. CpG island methylator phenotype-low (CIMP-low) in colorectal cancer: possible associations with male sex and KRAS mutations. *J Mol Diagn*. 2006;8:582-8.
6. Ogino S, Odze RD, Kawasaki T, Brahmandam M, Kirkner GJ, Laird PW, et al. Correlation of pathologic features with CpG island methylator phenotype (CIMP) by quantitative DNA methylation analysis in colorectal carcinoma. *Am J Surg Pathol*. 2006;30:1175-83.
7. Samowitz WS, Curtin K, Ma KN, Schaffer D, Coleman LW, Leppert M, et al. Microsatellite instability in sporadic colon cancer is associated with an improved prognosis at the population level. *Cancer Epidemiol Biomarkers Prev*. 2001;10:917-23.
8. Herman JG, Umar A, Polyak K, Graff JR, Ahuja N, Issa JP, et al. Incidence and functional consequences of hMLH1 promoter hypermethylation in colorectal carcinoma. *Proc Natl Acad Sci U S A*. 1998;95:6870-5.
9. Menigatti M, Truninger K, Gebbers JO, Marbet U, Marra G, Schar P. Normal colorectal mucosa exhibits sex- and segment-specific susceptibility to DNA methylation at the hMLH1 and MGMT promoters. *Oncogene*. 2009;28:899-909.
10. Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. *J Mol Diagn*. 2007;9:305-14.
11. Sobin LH, Wittekind CH. UICC: TNM classification of malignant tumors. Wiley-Liss. 2002;6th ed.
12. Kim MS, Lee J, Sidransky D. DNA methylation markers in colorectal cancer. *Cancer Metastasis Rev*. 2010;29:181-206.

13. Weber M, Davies JJ, Wittig D, Oakeley EJ, Haase M, Lam WL, et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat Genet.* 2005;37:853-62.
14. Worthley DL, Whitehall VL, Buttenshaw RL, Irahara N, Greco SA, Ramsnes I, et al. DNA methylation within the normal colorectal mucosa is associated with pathway-specific predisposition to cancer. *Oncogene.* 2010;29:1653-62.
15. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, Laczko E, Kurowski MA, Bujnicki JM, Menigatti M, Luz J, Ranalli TV, Gomes V, Pastorelli A, Faggiani R, Anti M, Jiricny J, Clevers H, Marra G. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res* 2007;5:1263-75.

## Tables

**Table 1.** Clinical characteristics of the study cohorts

A) Colorectal cancer patients

<b>Patients</b>	<b>All</b>	<b>Female</b>	<b>Male</b>
Number	106 (100%)	39 (36.8%)	67 (63.2%)
Age (y)			
Median	69.5	66	71
Mean	68.3	64.6	70.4
Standard deviation	12.4	14.5	10.5
Normal distribution	Yes <sup>1</sup> /Yes <sup>2</sup> /No <sup>3</sup>	Yes <sup>1</sup> /Yes <sup>2</sup> /Yes <sup>3</sup>	No <sup>1</sup> /Yes <sup>2</sup> /Yes <sup>3</sup>
Tumor location			
Right	32 (30.2%)	10 (9.4%)	22 (20.8%)
Left	74 (69.8%)	29 (27.4%)	45 (42.5%)
Stage of disease (UICC/AJCC)			
I	22 (20.8%)	10 (9.4%)	12 (11.3%)
II	24 (22.6%)	8 (7.5%)	16 (15.1%)
III	32 (30.2%)	13 (12.3%)	19 (17.9%)
IV	28 (26.4%)	8 (7.5%)	20 (18.9%)

<sup>1</sup> Kolmogorov-Smirnov normality test; <sup>2</sup> D'Agostino & Pearson omnibus normality test; <sup>3</sup> Shapiro-Wilk normality test.

## B) Healthy individuals

<b>Subjects</b>	<b>All</b>	<b>Female</b>	<b>Male</b>	
Number	31	16	15	
Age (y)				
Median	62	63.5	60	
Mean	63.6	64.8	62.4	P = 0.423 <sup>1</sup>
Standard deviation	8.23	8.89	7.56	
Distribution	Normal	Normal	Normal	

Distribution was assessed with the Kolmogorov-Smirnov normality test, D'Agostino & Pearson omnibus normality test and Shapiro-Wilk normality test. <sup>1</sup> Unpaired t test.

**Table 2.** Methylation cut-off, cancer sensitivity and specificity

<b>Target</b>	<b>Cut-off PMA (%)</b>	<b>Cancer Sensitivity (%)</b>	<b>Cancer Specificity (%)</b>
<i>FOXF1</i>	4	66	99.1
<i>MGMT</i>	5	35.8	100
<i>CA4</i>	5	20.8	100
<i>NPY1R</i>	5	18.9	100
<i>IFITM1</i>	7	15.1	100
<i>hMLH1</i>	6	14.2	100
<i>GREM1</i>	6.5	13.2	100

Cut-off values (percentage of methylated alleles, PMA) were established for each gene promoter to discriminate cancer tissue from cancer-associated mucosa with  $\geq 99\%$  specificity. Tumors with a PMA value above this cut-off are considered hypermethylated. Cancer sensitivity represents the frequency of hypermethylated tumors.

## Figure legends

### Figure 1

Cancer-specific Methylation of the Investigated Gene Promoters. A, Cancer-specific methylation levels: Scatter plot showing the percentage of methylated alleles (PMA) for each gene in matched CRC, CAM and blood samples (n = 106). The gray bar indicates the  $\geq 99\%$  cancer-specific methylation threshold, ranging from 4% (*FOXF1*) to 7% (*IFITMI*). B, Frequency of hypermethylated CRCs. Tumors showing methylation levels above the  $\geq 99\%$  cancer-specific threshold are classified as hypermethylated in a particular target. Shown is the frequency of hypermethylated cancers according to location. \*: P < 0.05. C, CRC coverage by the seven marker panel. Methylation above the  $> 99\%$  cancer-specific threshold in either of the seven gene promoter panel discriminates 82% of all tumors and 94% of right-sided tumors from CAM.

### Figure 2

Clustering of Target Methylation in Colorectal Cancer. A, Distribution of Hypermethylated tumors according to Gender and Location. Hypermethylated tumors in a particular target are represented by a grey box. B, Heat Map of unsupervised hierarchical clustering. Methylation levels are represented by different colors with red indicating the highest and black the lowest values.

### Figure 3

Methylation clustering in cancer patients and healthy individuals. Concomitant methylation between the investigated targets is shown for normal mucosa from healthy individuals (A), cancer-associated mucosa from patients (B), and tumor tissue (C). Red lines indicate concomitant methylation as assessed by Fisher's exact test of contingency tables for all target pairs or repeated measures non-parametric and parametric (ANOVA) tests. Accordingly, thin black lines indicate absent concomitant methylation. The gray area illustrates the concordant methylation of the newly described CIMP targets *C44*, *NPY1R* and *GREM1* preserved in all three tissues.

### Figure 4

Methylation levels of newly defined and classical CIMP targets in cancer patients and healthy individuals. Shown are median methylation beta values of genome-wide promoter methylation analyses in CIMP-positive (CIMP+ PT) and CIMP-negative (CIMP- PT) cancers as well as normal colorectal tissue and one blood sample from healthy individuals (NM). P-values for the

difference in beta values between normal tissues from individuals aged 75 years and CIMP-positive cancers are indicated. NM Age 50/75: Normal mucosa from healthy individuals aged 50 and 75 years.

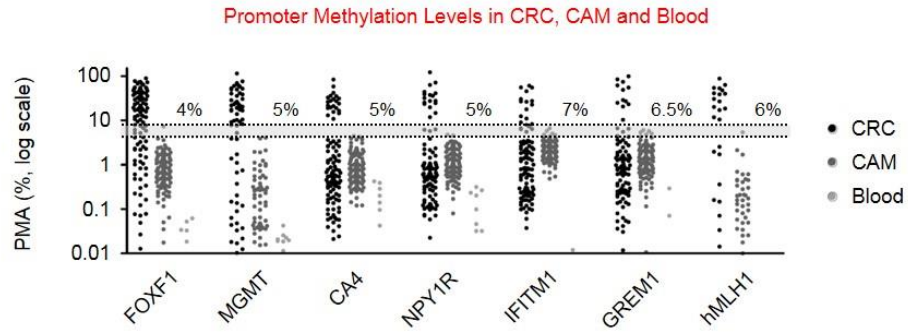
### **Figure 5**

Model for the pathogenesis of CIMP-positive colorectal cancers (serrated pathway). A differentiated epithelial cell population along the colorectum is constitutively low-level methylated in *FOXF1*, *CA4*, *NPY1R* and *GREM1* (*constitutive* methylation markers). Critical (epi)genetic events in these cells, e.g. activating *BRAF*<sup>V600E</sup> mutation and/or aberrant methylation of *hMLH1* (*oncogenic* type marker) contribute to malignant transformation. Clonal expansion of this cell population, occurring typically in the proximal colon, leads to cancer-specific methylation of various targets and the establishment of the CIMP phenotype. met: methylated; unmet: unmethylated.

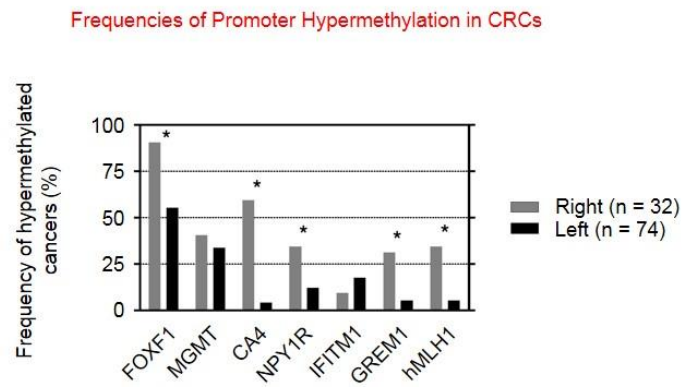


Figure 1

A



B



C

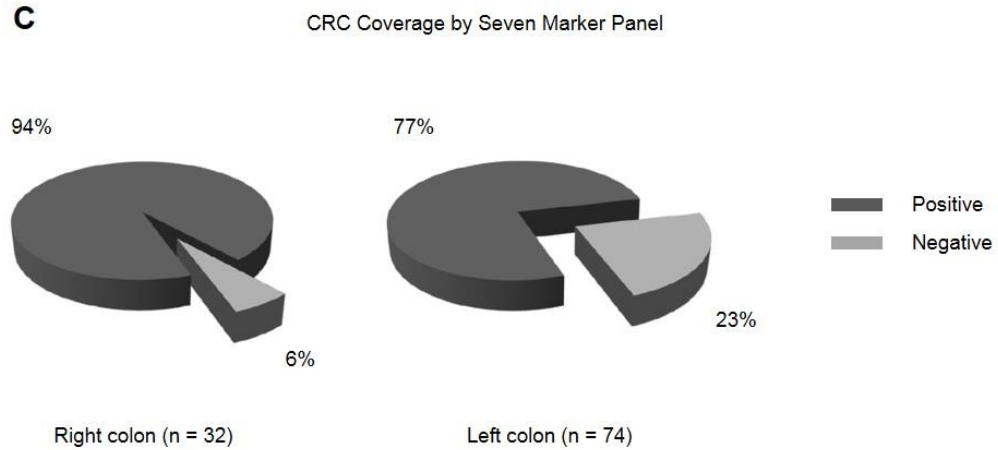


Figure 2

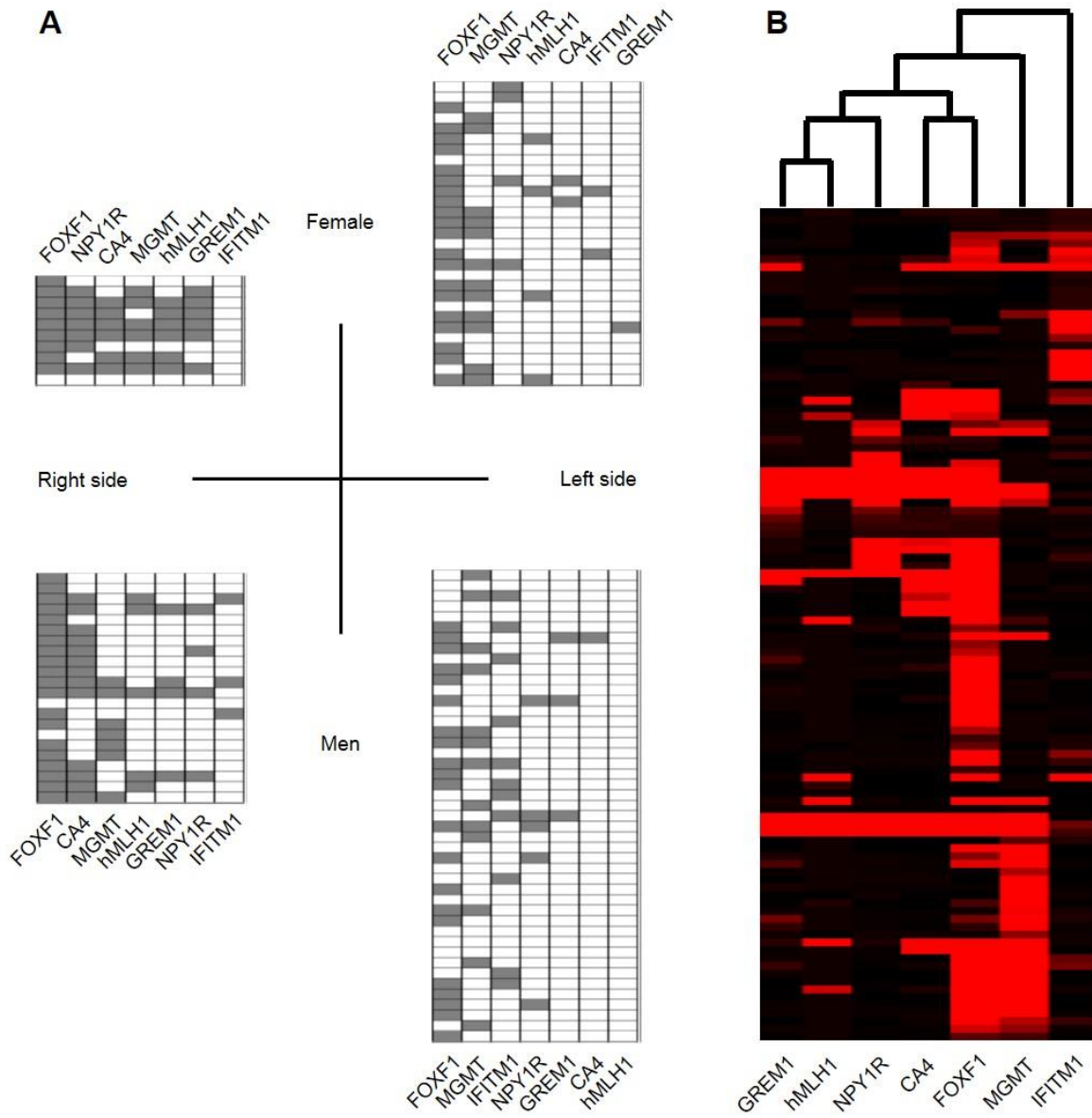


Figure 3

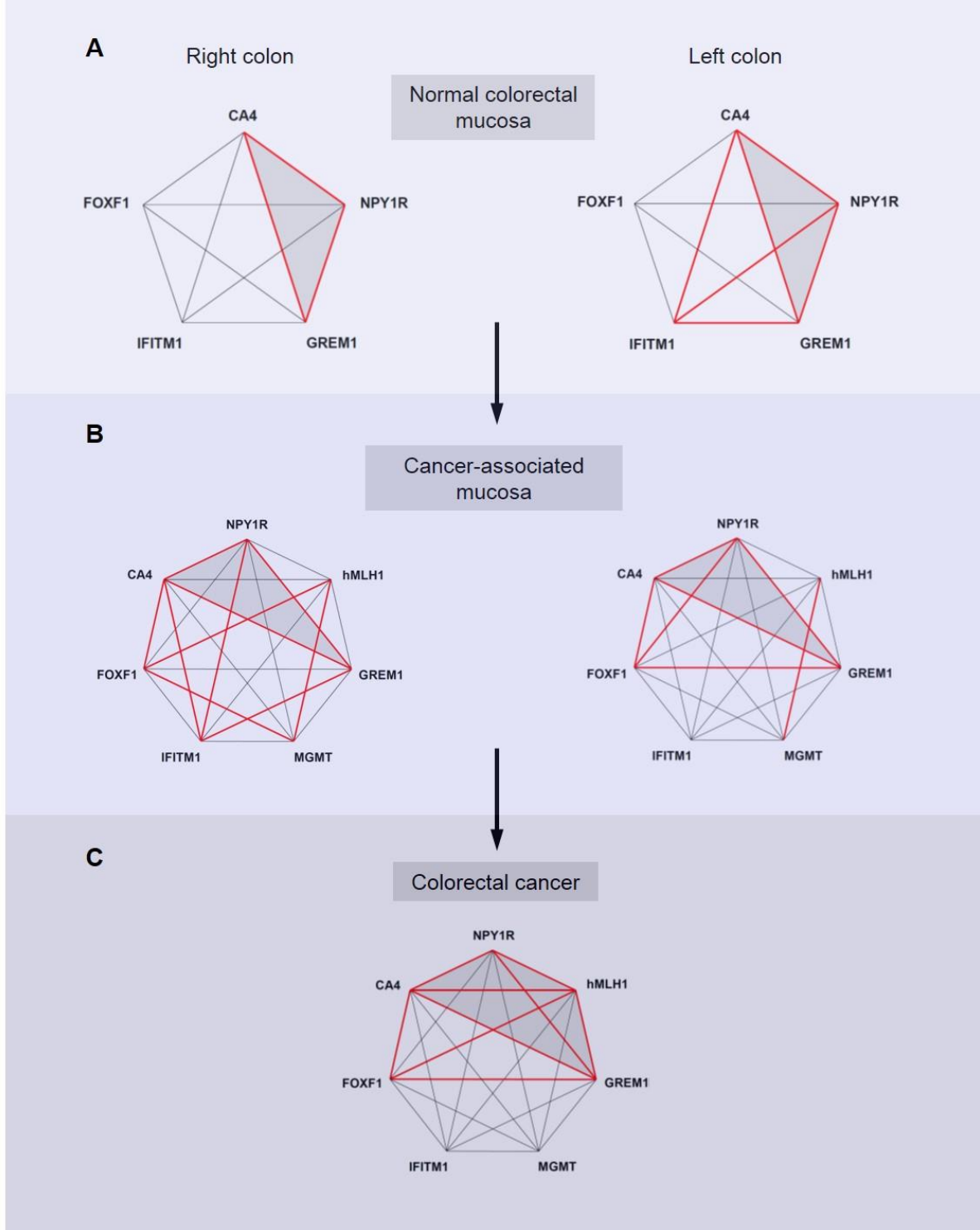


Figure 4

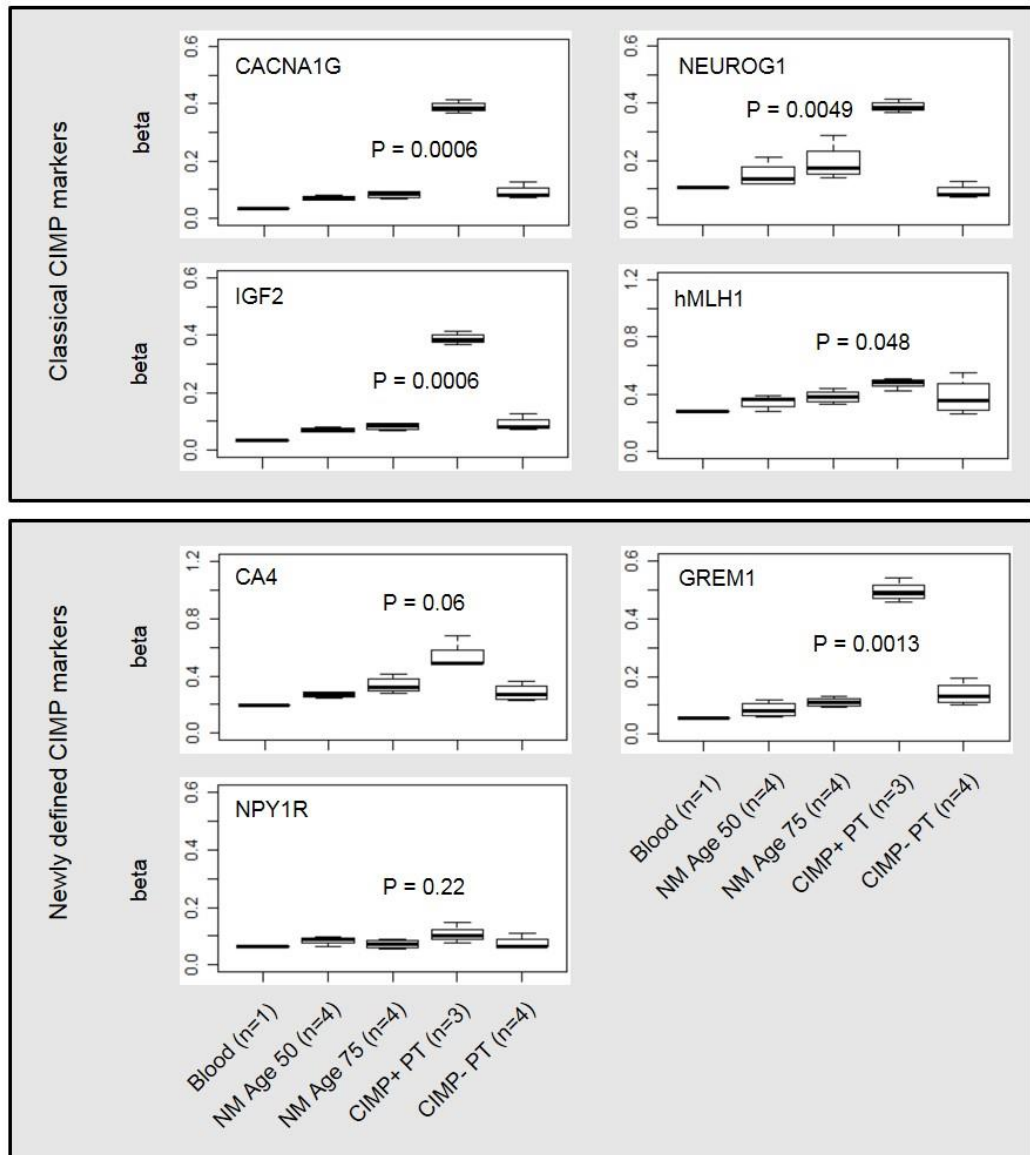
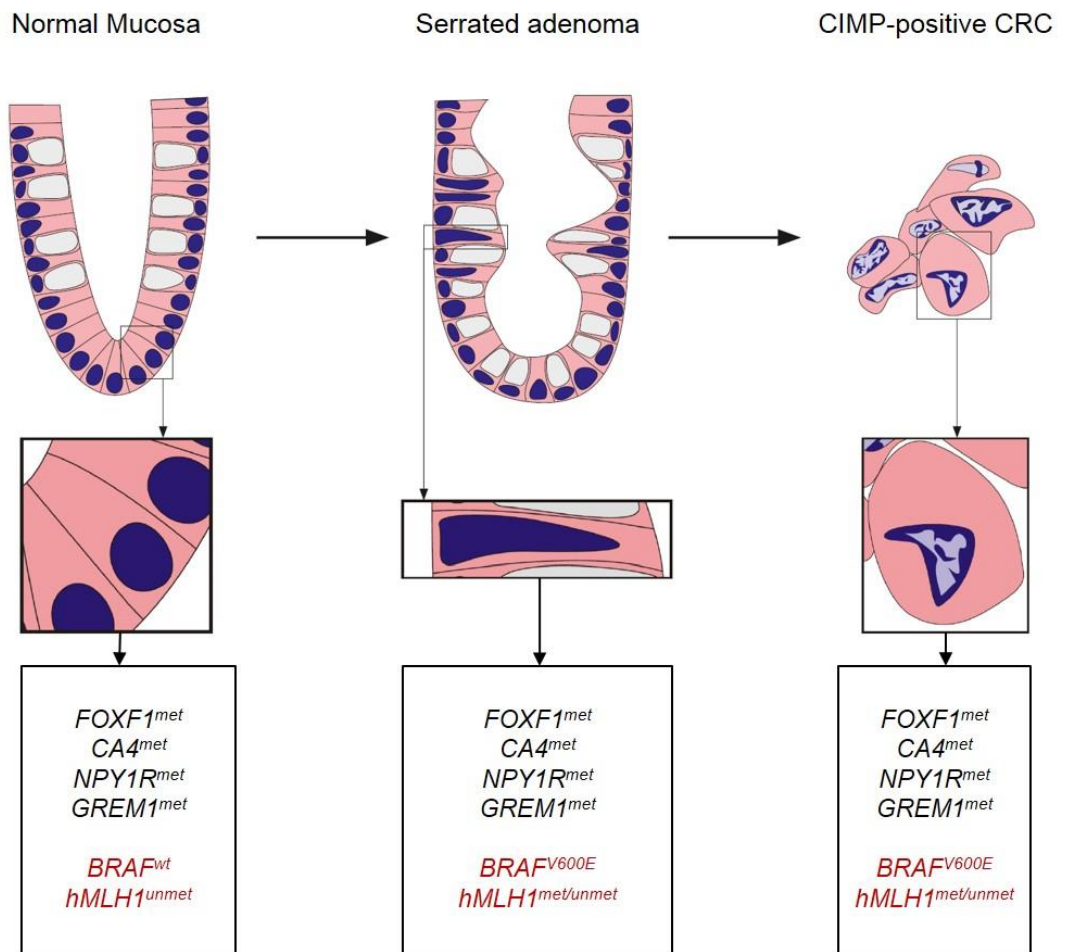


Figure 5



## Supplemental data

### Supplementary Materials and Methods

#### Reanalysis of the transcriptome of 32 colorectal adenomas and matched mucosa tissues

Transcriptome analysis of 32 colorectal adenomas and matched mucosa tissues was performed by Sabates-Bellver et al. (1) as previously described. Briefly, the polyp sample and three biopsies of normal mucosa from the same colon segment were immersed in RNAlater (Ambion). Total RNA was extracted (RNeasy Mini kit, Qiagen) from homogenized tissue samples (5-15 mg), and its integrity was verified by capillary gel electrophoresis (Bio Analyzer, Agilent Technologies). Complementary RNA (15 µg/sample), synthesized and labeled as previously described, was hybridized with the Affymetrix U133 Plus 2.0 array, which contains in situ synthesized oligonucleotides representing the entire human genome (54,675 probes). We imported the raw gene expression data generated by GeneChip Operating Software (Affymetrix) (GEO record GSE8671) into ROCHE Affymetrix Chip Experiment Analysis (RACE-A) Software (2) and normalized per chip (i.e., to the median of all values on a given array) and per gene (i.e., to the median expression level of the given gene across all samples).

#### Cell lines

CO115, SW48 (both microsatellite instable) and GP5d human colorectal cancer cell lines were grown according to a routine protocol. Briefly, they were incubated in Dulbecco's Modified Eagle's Medium (DMEM) containing 10% fetal bovine serum (FBS), 1% (V/V) Penicillin/Streptomycin and 2 mM L-Glutamine. The cells were kept at 37°C in an atmosphere containing 4% CO<sub>2</sub> and were harvested in an exponential growth phase.

#### Methylation analysis, locus-normalized quantitative methylation-specific PCR and bisulfite sequencing

Methylation of the CpG-rich promoter regions of *FOXF1*, *CA4*, *NPY1R*, *GREM1*, *hMLH1*, *MGMT* and *IFITM1* was analyzed by quantitative methylation specific PCR (qMSP). Strand-specific primers complementary to the *methylated* or *unmethylated* (*M*- and *U*-primers) promoter sequence were used. All qMSP primer sequences are listed in Supplementary Table S11. Additionally, we used methylation *insensitive* *S*-primers to standardize for the DNA input after sodium bisulfite treatment. The *S*-primer amplicon did *not* contain any CpG dinucleotides and was chosen in the close proximity to the interrogated promoter to *standardize* for its locus, thereby eliminating ploidy aberration bias. We therefore termed this assay *locus-normalized*

*qMSP (ln-qMSP)*. qMSP reactions were performed with the Rotor-Gene 3000 real time thermal cycler (Corbett Life Science, Sydney, Australia) and the QuantiTect SYBR Green Kit (Qiagen). Each reaction was performed in a volume of 15  $\mu$ l and contained 0.5  $\mu$ M of each primer and 1.5  $\mu$ l of bisulfite converted DNA. Real time PCR conditions were as followed: activation of polymerase at 95°C for 15 minutes followed by 50 cycles with denaturation at 94°C for 10 seconds; primer annealing at 57°C (*FOXF1*), 58°C (*hMLH1*, *MGMT*), 59°C (*GREM1*, *NPY1R*), 61°C (*CA4*) or 52°C (*IFITM1*) for 25 seconds and extension at 62°C for 20 seconds or at 60°C for 30 seconds (*IFITM1*). In order to ensure product specificity, a melting point analysis was performed after every run. The correct size of PCR products was verified by agarose gel electrophoresis.

For methylation quantification standard curves were generated by real time amplification of serial 1:10-dilutions (from 100 to 0.1%) of M.sssI methylase and sodium bisulfite treated blood DNA with M- and S-primers (Supplementary Fig. S5A). The quality of positive control DNA was proven by unsuccessful amplification with U-primers under optimized conditions. The tester DNA was amplified by two separate real-time PCR runs with the M- and the S-primer set in technical duplicates each. The *percentage of methylated alleles (PMA)* was calculated by the ratio between the values of the amplification with the M- and the S-primer set (for a calculation example see Supplementary Fig. S5B). The *percentage of unmethylated alleles (PUA)* was determined analogously using the U- and S-primer sets. Here, peripheral blood DNA served as a methylation negative control to establish the standard curves. Real time amplification profiles, melting and standard curves for all newly established targets are shown in Supplementary Fig. S5A, C-F.

Complementarity experiments using locus-normalized qMSP showed that the sum of the values obtained by amplifications of defined mixtures of methylated (99, 90, 10, and 1%) and unmethylated DNA with the M- and U-primers was close to 100% (from 90 to 105%) for the validation experiments indicating an accurate measurement of methylated alleles. In consequence, quantitative methylation results in 106 tumor samples did not yield values of more than 100% methylation. Values greater than 100% may result from cancer-associated phenomena like loss of heterozygosity (LOH) or chromosomal abnormalities (aneuploidies), and are frequently seen when using other techniques, which do not normalize for the locus under investigation.

To validate our In-qMSP results we analyzed *FOXF1*, *CA4*, *NPY1R*, *GREM1* and *IFITM1* promoter methylation in two primary tumors and one cancer-associated mucosa sample (the latter only for *FOXF1*) by bisulfite sequencing (Supplementary Fig. S6, primer sequences are given in Supplementary Table S12). After bisulfite conversion of genomic DNA CpG-rich promoter regions were PCR amplified with True Start Taq Polymerase (Fermentas; Burlington, Canada). PCR conditions were as followed: Polymerase activation at 95°C for 2 minutes followed by 40 cycles of denaturation at 95°C for 20 seconds, primer annealing at 51°C (*GREM1* distal, 2.5 mM MgCl<sub>2</sub>, 1 μM each primer), 53°C (*GREM1* proximal, 1.5 mM, 0.5 μM), 57°C (*FOXF1* proximal, *NPY1R*, *IFITM1*, all 2.5 mM, 0.5 μM), 62°C (*CA4*, 2.5 mM, 0.5 μM) or 63°C (*FOXF1* distal, 2.5 mM, 0.5 μM) for 25 seconds and primer extension at 72°C for 50 seconds. PCR was completed by a final extension at 72°C for 15 minutes. To avoid an amplification bias towards more unmethylated sequences the primers each covered one CpG site at the very 5'-end and were designed complementary to the methylated sequence (3) (Supplementary Table S12). The amplification efficiencies were equalized by adjusting the PCR conditions in real time PCR reactions (standardized with the S-primer sets from the qMSP assay) visualized with EvaGreen (Biotium, USA).

The correct, unbiasedly amplified PCR products were purified by gel extraction (QIAEX Gel Extraction Kit, Qiagen) and subsequently incubated with true Start Taq Polymerase and 0.2 mM ATP to re-establish 3'-A overhangs usually introduced by Taq polymerases and lost to a certain extent during gel extraction. Purified PCR products were cloned into the pCR<sup>®</sup>4-TOPO<sup>®</sup> vector provided with the TOPO TA Cloning Kit for Sequencing (Invitrogen; Carlsbad, USA). Chemically competent E.coli cells (also provided in the Kit) were transformed and grown on Ampicillin (0.1 mg/ml) containing LB-agar plates. Inserts were checked by Colony PCR and plasmid DNA of positive clones was purified with the peqGOLD Plasmid Miniprep Kit I (Peqlab; Erlangen, Germany). Inserts were sequenced using the Big Dye Terminator Sequencing Kit (Applied Biosystems; Foster City, USA).

### **DNA methylation and mRNA expression in colorectal tissue**

Simultaneous purification of DNA and RNA from 10 pairs of matched CRC and CAM samples was performed with the AllPrep DNA/RNA/Protein Mini Kit<sup>®</sup> (Qiagen, Hilden, Germany). Up to 30 mg of tissue were disrupted in 600 μl Buffer RLT and 300 μl Zirconia beads using the FastPrep<sup>®</sup> cell disrupter (Qbiogene, Carlsbad, USA). The quantity of the purified DNA and RNA was determined spectrophotometrically. DNA methylation and RNA expression were analyzed by In-qMSP and quantitative RT-PCR, respectively (*see there*).



## Quantitative Reverse Transcriptase-PCR

Complementary DNA (cDNA) was synthesized from total RNA using the RevertAid<sup>®</sup> H Minus First Strand cDNA Synthesis Kit (Fermentas) and random hexamer primers (0.5  $\mu$ M). Synthesized cDNA was diluted 200 times and subjected to qRT-PCR reaction on the Rotor-Gene 3000 real time thermal cycler (Corbett Life Science) using the QuantiTect SYBR Green Kit (Qiagen). The following PCR cycling conditions were used: Hot Start Taq polymerase activation at 95°C for 15 minutes followed by 50 cycles of denaturation at 94°C for 10 seconds, primer annealing at 59°C (*CA4*, *IFITM1*, *FOXF1*, *ACTB*, *GAPDH*, *GREM1*) or 60°C (*NPY1R*) for 25 seconds and extension at 62°C for 20 seconds (primers are listed in Supplementary Table S13). CAM from patient 69, which showed expression for all target genes, served as a positive control and for the establishment of the standard curves. Tester cell line cDNA was analyzed in technical triplicates. Amplification signals of the target genes were normalized to the average signal of the housekeeping genes  $\beta$ -Actin (*ACTB*) and glyceraldehyde-3-phosphate dehydrogenase (*GAPDH*). Transcription levels were normalized to the expression of the housekeeper genes and to the target gene expression of CAM from patient 31.

## Reexpression experiments

To correlate the promoter methylation levels of the newly established target genes *CA4*, *NPY1R*, *FOXF1*, *IFITM1* and *GREM1* with their expression we analyzed mRNA levels in human CRC cell lines (CO115, SW48 and GP5d) before and after treatment with 5-Aza-cytidine (DNA methyltransferase inhibitor) and/or Trichostatin A (Histone deacetylase inhibitor). After 24 h of culture in Dulbecco's Modified Eagle Medium (Sigma; Hamburg, Germany) supplemented with 10% fetal calf serum, 2 mM L-Glutamine (Sigma), 60U/ml Penicillin (Sigma) and 10 $\mu$ g/ml Streptomycin (Sigma) the medium was replaced with medium containing 1  $\mu$ M 5'-Aza-C or 0.5  $\mu$ M TSA and incubated for five and two days, respectively. The medium was renewed every 24 h. For the double treatment, cells were initially treated with 5-Aza-cytidine followed by Trichostatin A treatment. The cells were harvested and the genomic DNA and total RNA were extracted with TRI REAGENT (Sigma). Promoter methylation and mRNA expression were analyzed by In-qMSP and quantitative RT-PCR, respectively (*see there*).

## Supplementary figure legends

### **Supplementary Figure 1. Methylation levels in tumor tissue (CRC), cancer-associated mucosa (CAM) and blood.**

Bars and whiskers indicate mean PMA (Percentage of methylated alleles) values with standard deviations. \*: The methylation level for hMLH1 in blood is lower than 0.001% and therefore is not depicted.

### **Supplementary Figure 2. DNA methylation and gene expression of FOXF1 in colorectal cancer and cancer-associated mucosa.**

DNA methylation and gene expression of *FOXF1* were measured in matched cancer tissue (CRC) and cancer-associated colorectal mucosa (CAM) from 10 patients. The mRNA expression of the investigated samples is indicated as multiples of the mRNA expression in CAM from patient 31 (mRNA expression of patient 31 = 1). The P-value for the inverse correlation of DNA methylation and gene expression is given. PMA: Percentage of methylated alleles.

### **Supplementary Figure 3. Promoter demethylation and gene reexpression in cell lines.**

Three cell lines (CO115, SW48 and GP-5d) were treated with either 5-Azacytidine (5AzaC), Trichostatin A (TSA) or both. GP-5d was not methylated in *CA4*, *NPY1R*, *IFITM1*, and *GREM1*. mRNA expression is relative to the cancer-associated mucosa (CAM) from patient 69 (see Supplementary Fig. S2).

### **Supplementary Figure 4. Relation of Hypermethylated Tumors to Clinical Characteristics.**

Frequencies of hypermethylated CRC in relation to gender (A) and tumor stage (B). \*:  $P < 0.05$ .

### **Supplementary Figure 5. Locus-normalized quantitative Methylation Specific PCR (In-qMSP) for CA4 (a), NPY1R (c), IFITM1 (d), FOXF1 (e) and GREM1 (f).**

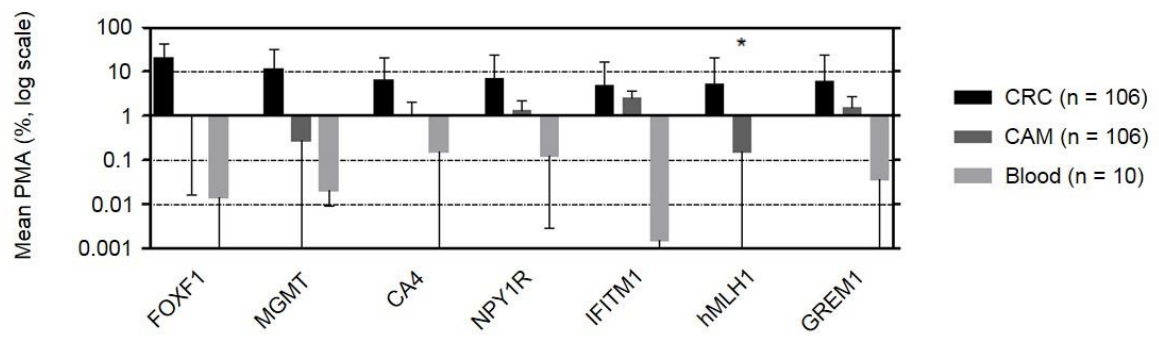
Amplification (I), melting (II) and standard curves (III) for the real time PCR assays with the M- (methylation specific) and S-primers (standardization, methylation *insensitive*) are shown (a, c-f). M. sssI methylase and sodium bisulfite treated blood DNA was used as positive control, which was serially diluted 1:10 up to 1:1'000. A negative control (sodium bisulfite treated peripheral blood DNA) and a no template control (NTC) were included in the assay. Melting curve analyses verified the specificity of the PCR reactions (absence of primer dimers and unspecific PCR products). b) The  $C_t$  (threshold cycle) values of the amplifications of a tester DNA with

the M- and S-primer sets were compared with the pre-established standard curves and the relative level of both amplicons (%) determined. The percentage of methylated alleles (PMA) in the tester DNA was then calculated by the ratio between the levels of the M- and the S-amplicons. qMSP assay data for *hMLH1* and *MGMT* was already shown in (4).

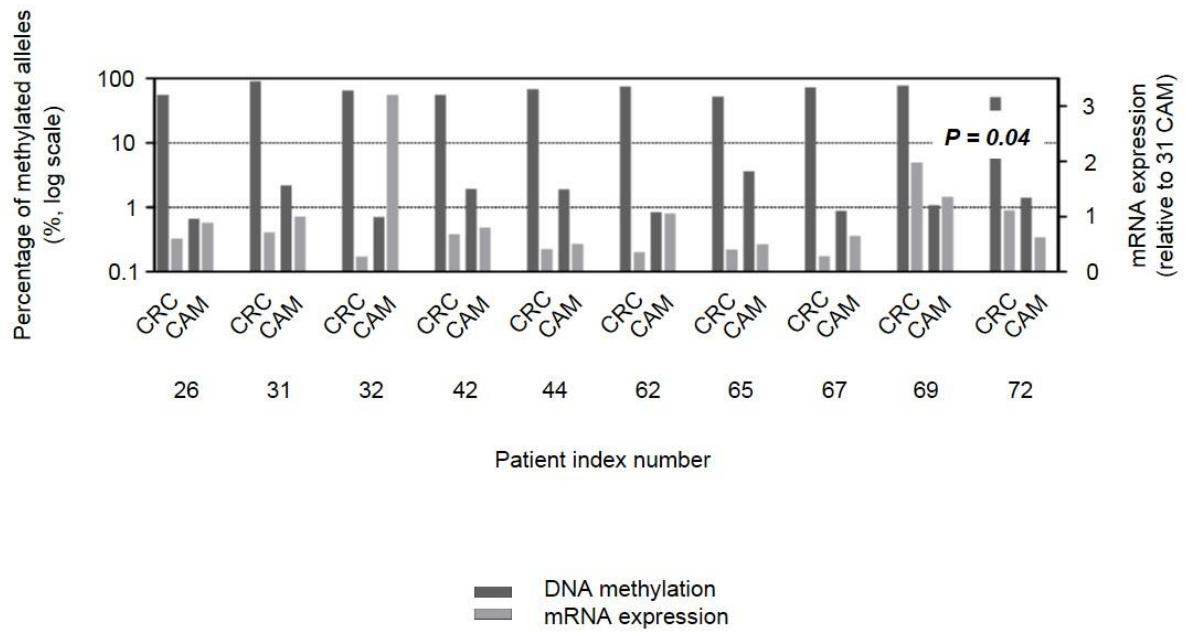
**Supplementary Figure 6. Bisulfite sequencing of the newly established targets.**

*FOXF1*, *CAA*, *NPY1R*, *IFITM1* and *GREM1* promoter regions were bisulfite sequenced in two cancer tissues (CRC) each. Additionally, *FOXF1* methylation was assessed in one cancer-associated colorectal mucosa (CAM). Distances from transcription start site (TSS) are indicated at the left border. Each circle represents a CpG site with an empty circle (○) symbolizing an unmethylated and a filled circle (●) a methylated state.  $\Gamma \rightarrow$  : Transcription start site;  $\dashrightarrow$  : In-qMSP primer location.

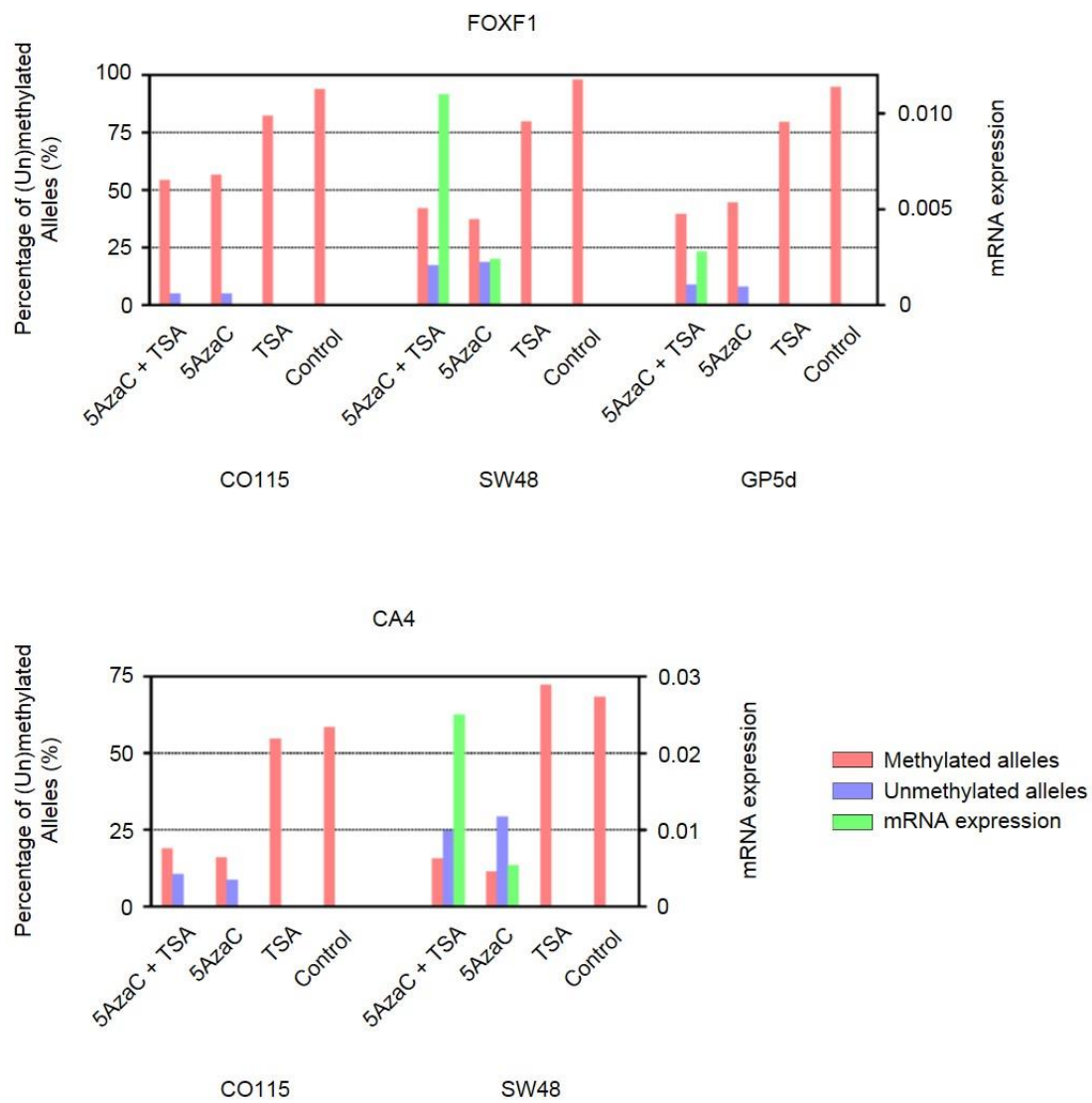
Supplementary Figure 1

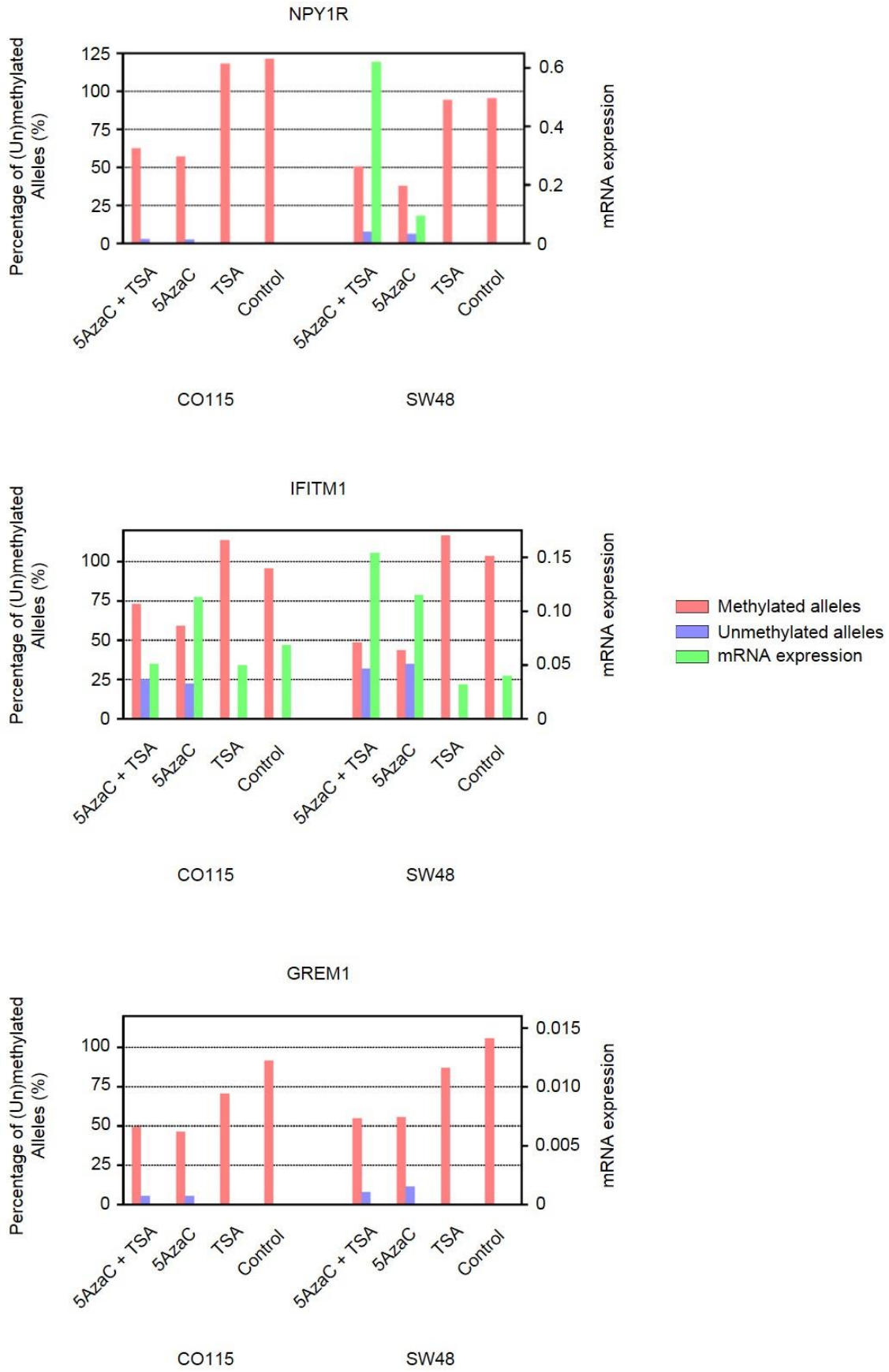


Supplementary Figure 2



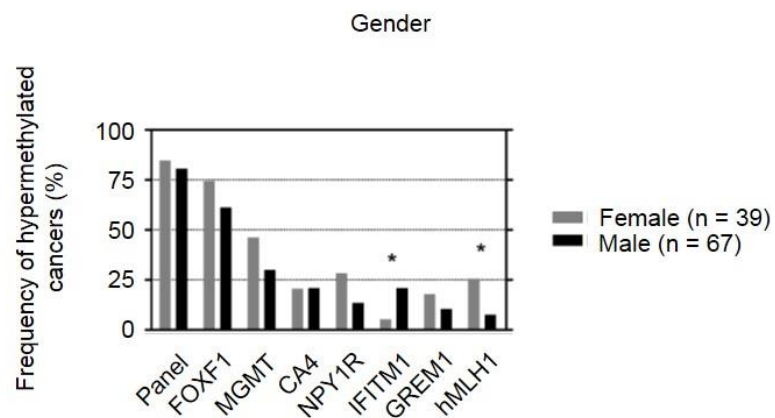
Supplementary Figure 3



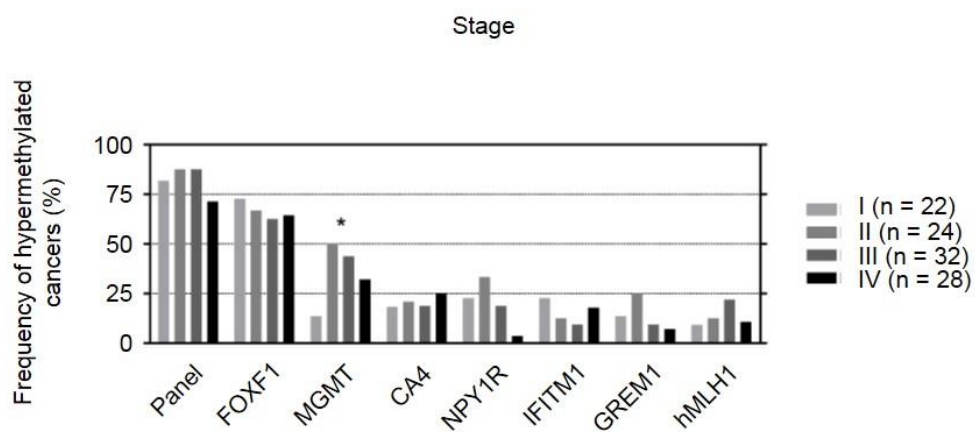


# Supplementary Figure 4

**A**

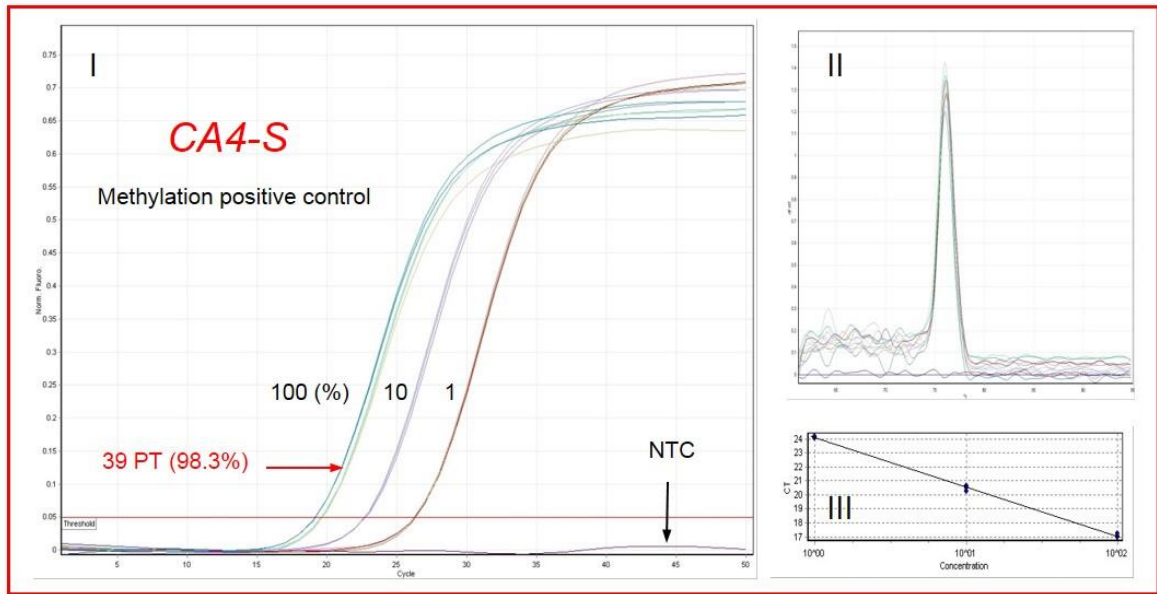
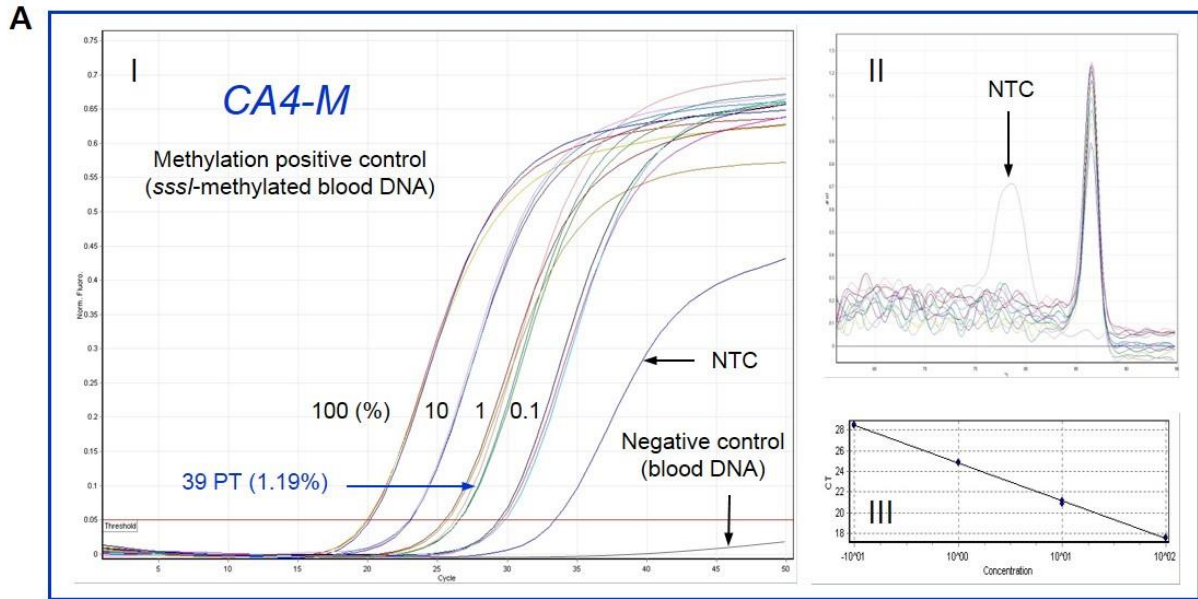


**B**





Supplementary Figure 5



**B**

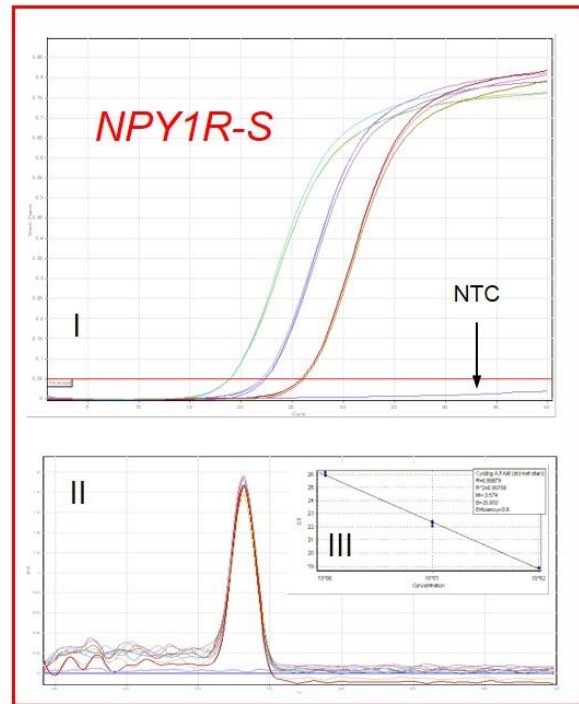
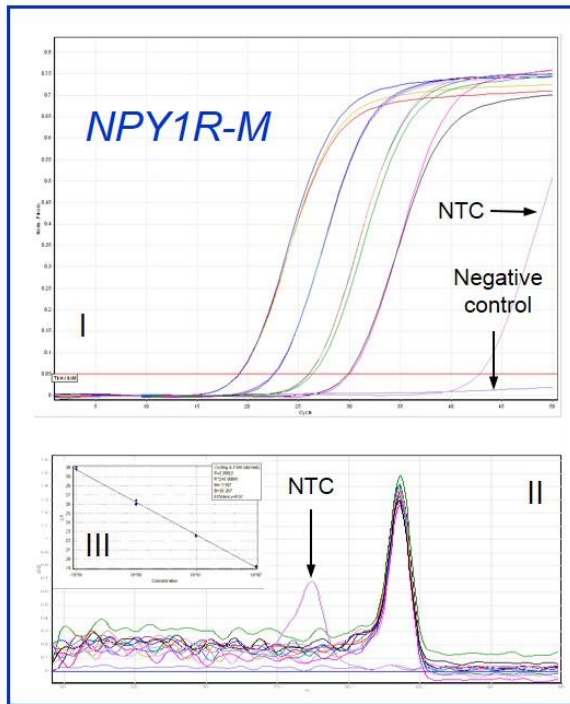
Calculation of the percentage of methylated alleles (PMA) in tumor DNA (39 PT):

$$\text{M-amplicon level \%} / \text{S-amplicon level \%} \times 100 \%$$

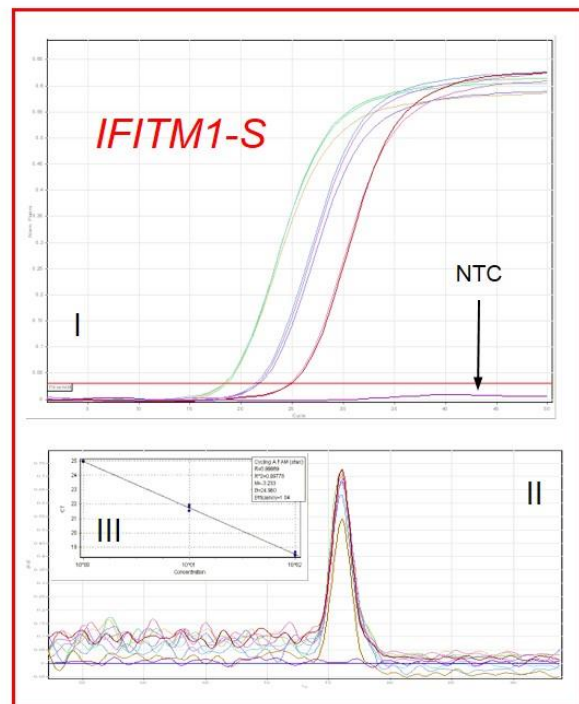
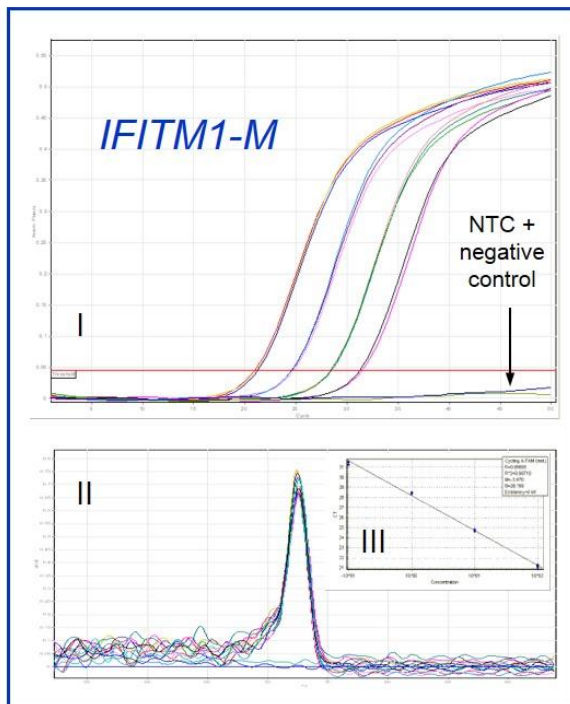


$$1.19\% / 98.3\% \times 100\% = 1.21\%$$

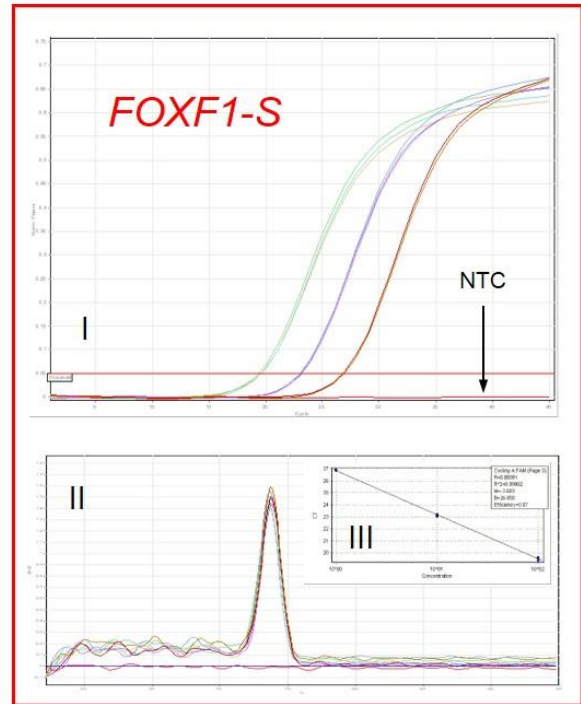
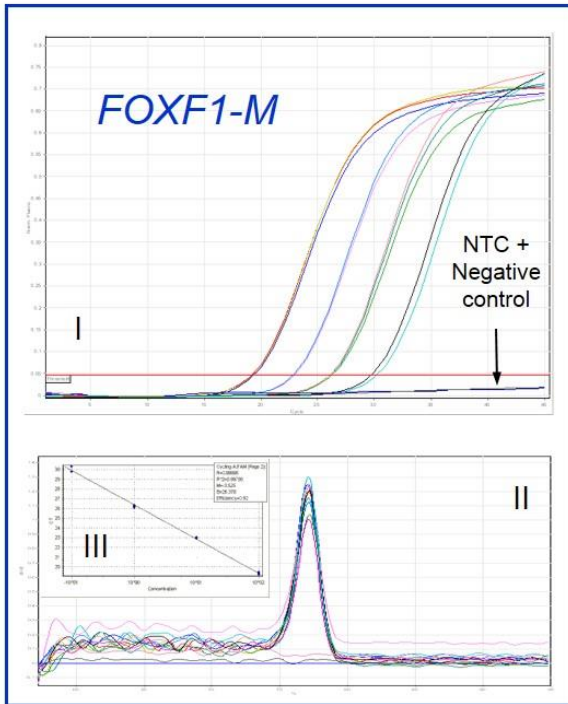
C



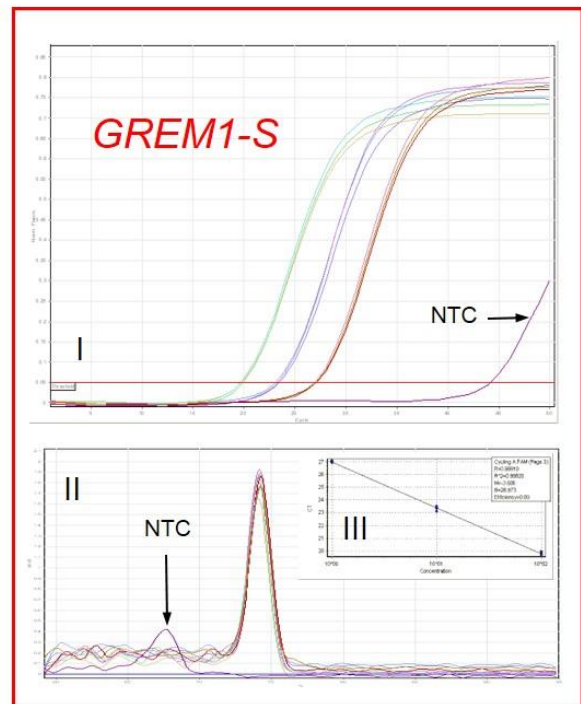
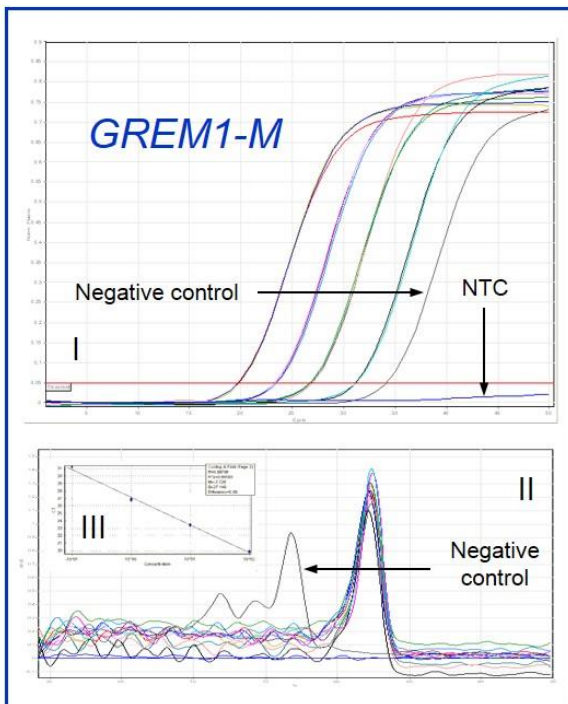
D



E

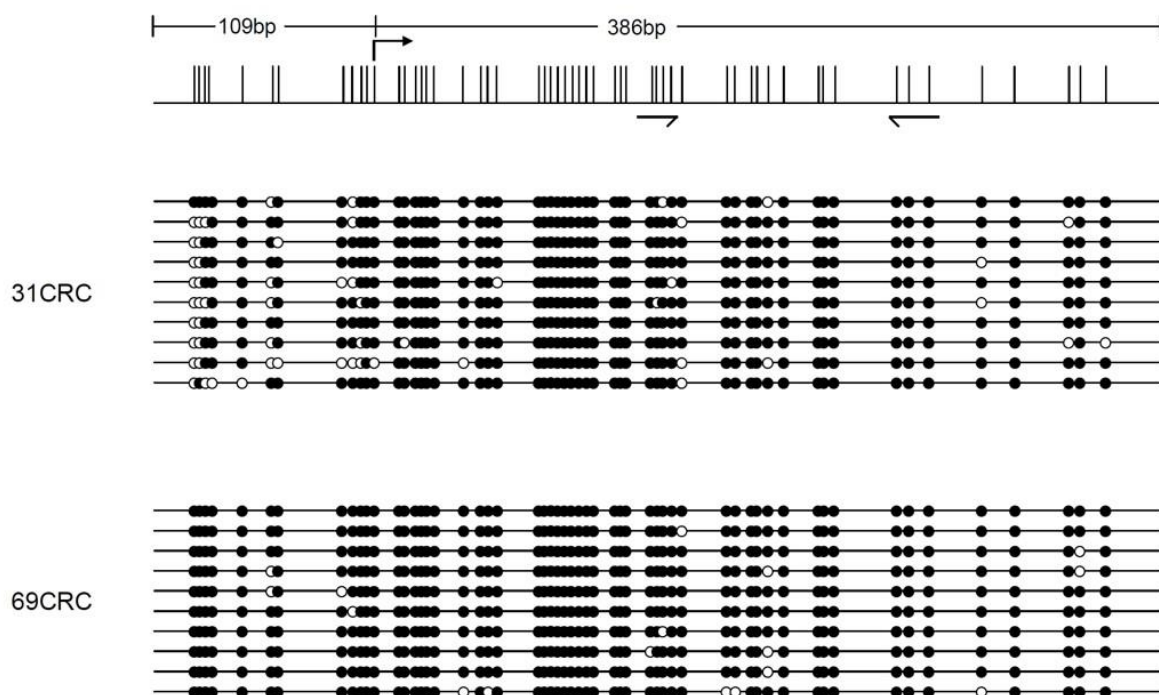


F

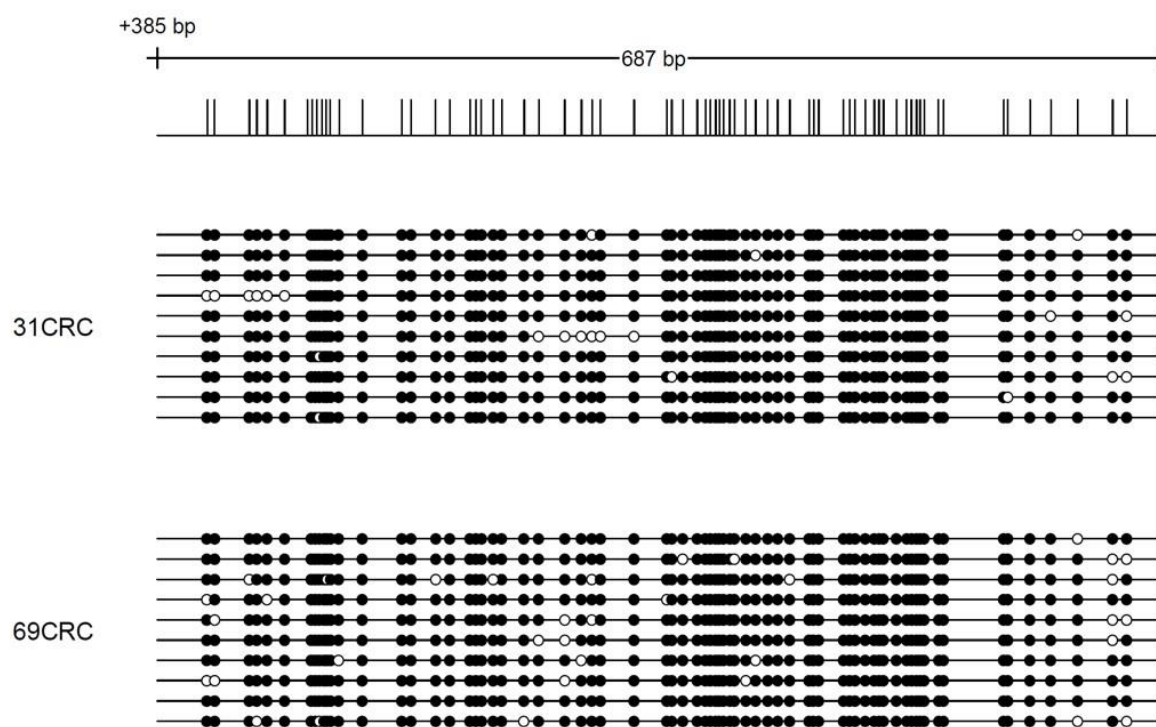


Supplementary Figure 6

FOXF1 proximal

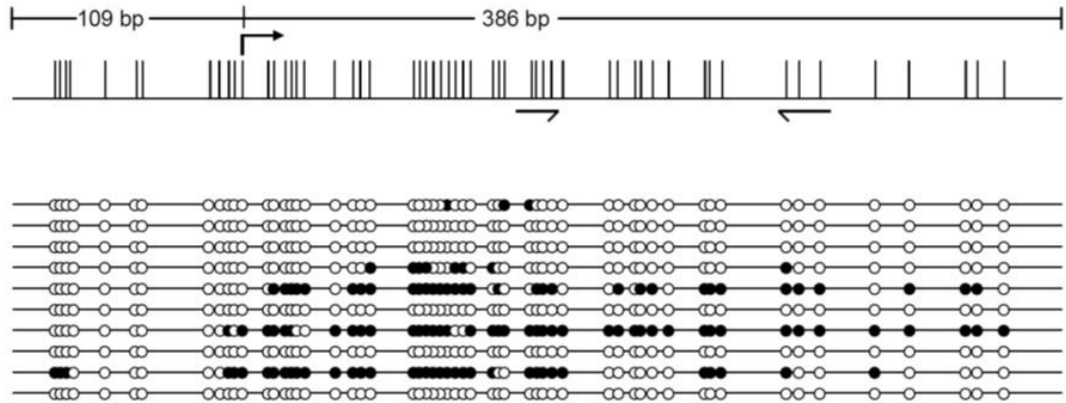


FOXF1 distal



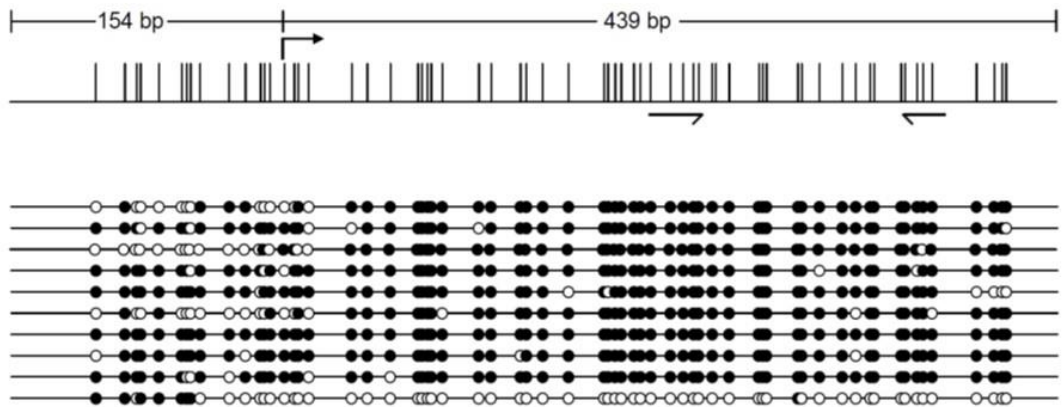


FOXF1 proximal



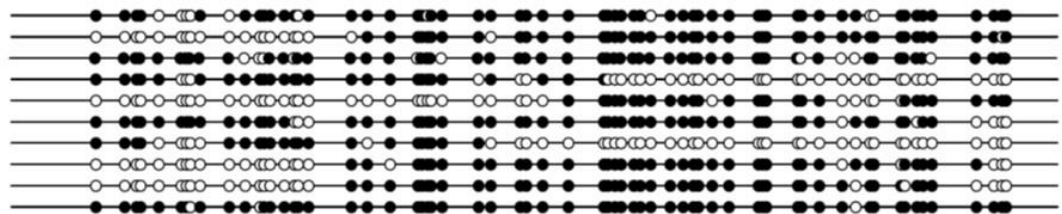
31CAM

CA4

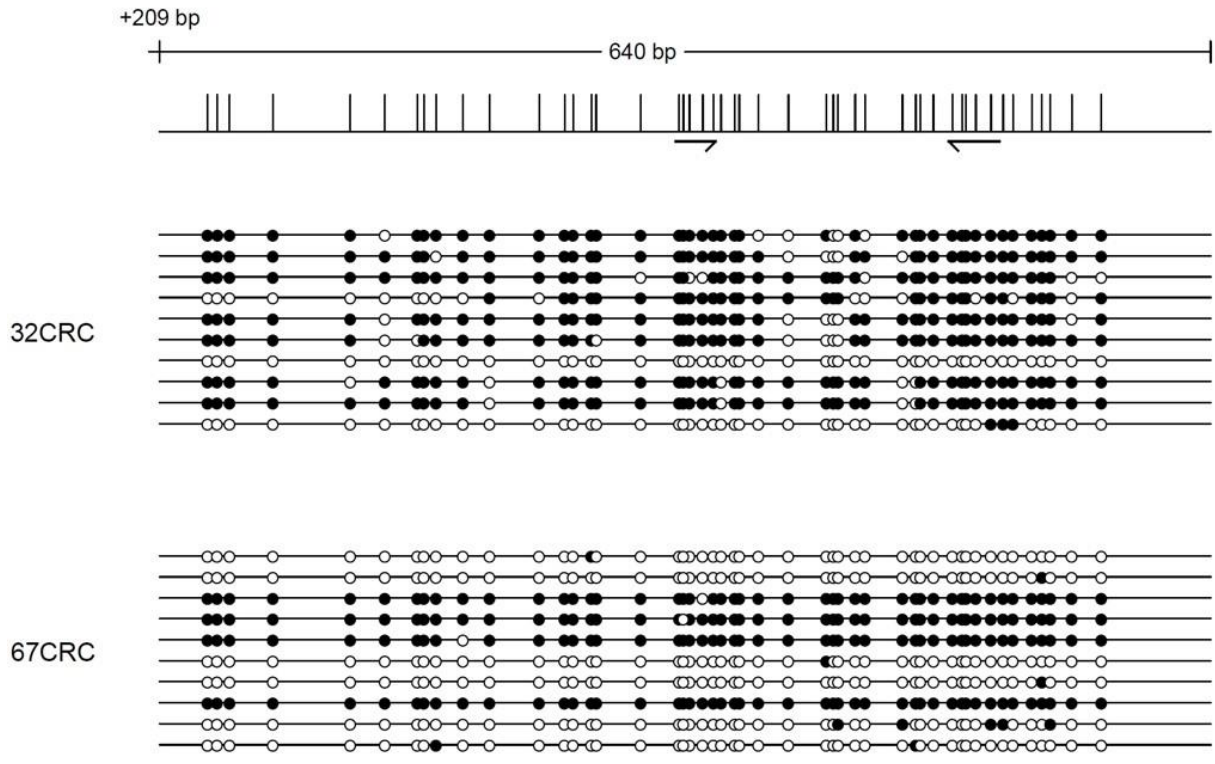


31CRC

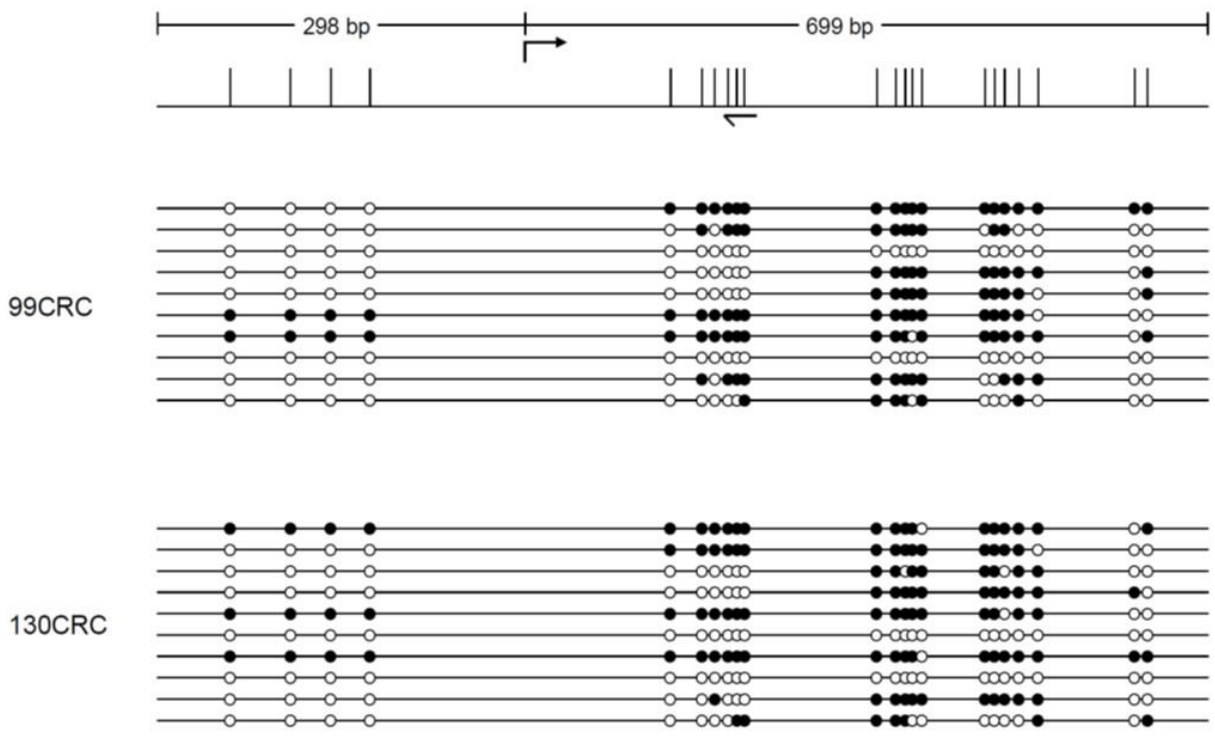
99CRC



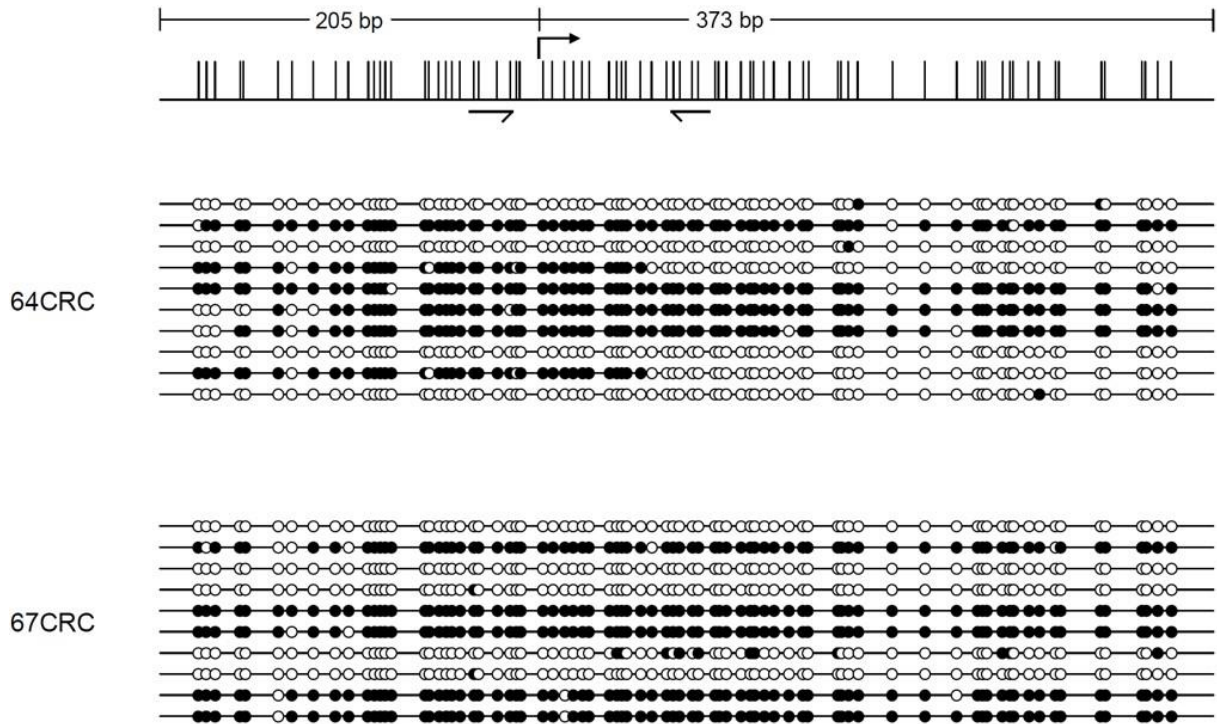
# NPY1R



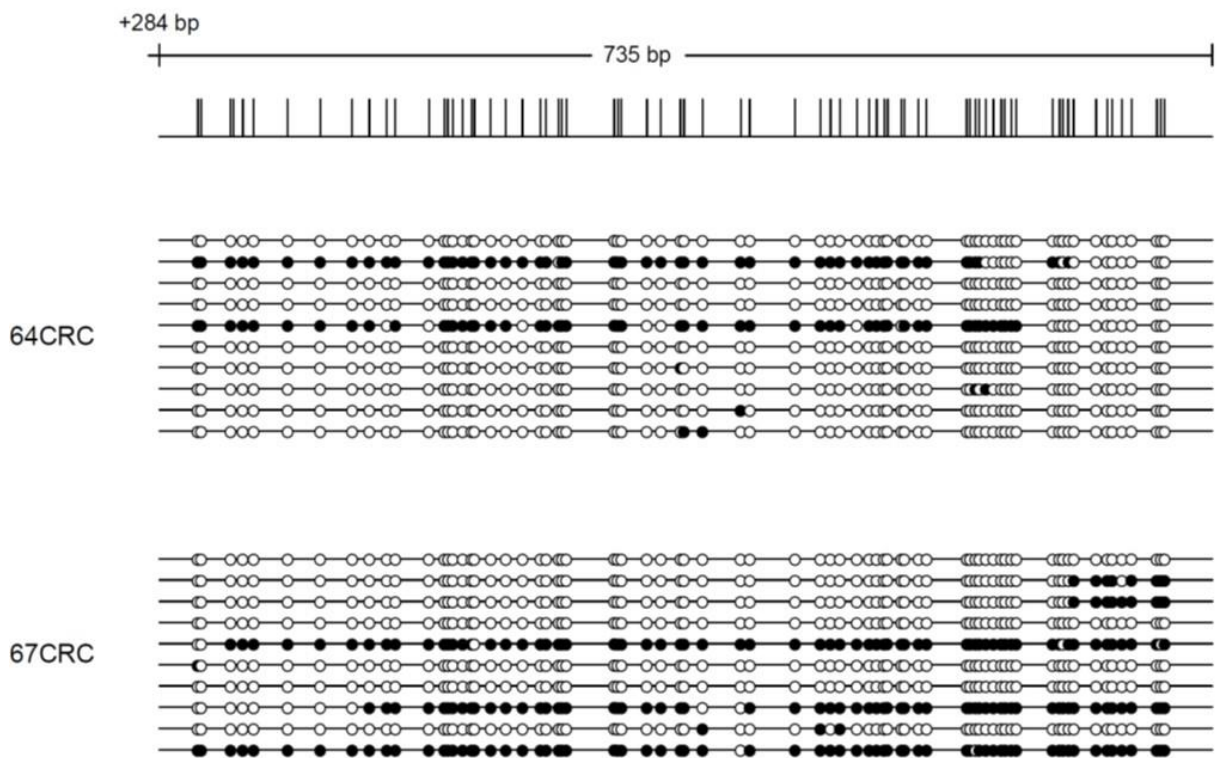
# IFITM1



### GREM1 proximal



### GREM1 distal



## *Supplementary Tables*

**Supplementary table 1.** Mean methylation levels in matched cancers, cancer-associated mucosa and blood

<b>Target</b>	<b>Mean PMA (%)</b>		
	<b>CRC</b>	<b>CAM</b>	<b>Blood</b>
<i>FOXF1</i>	21.09 ± 22.63	0.99 ± 0.97	0.01 ± 0.02
<i>MGMT</i>	11.51 ± 20.83	0.28 ± 0.66	0.02 ± 0.01
<i>CA4</i>	6.52 ± 13.55	1.09 ± 0.93	0.16 ± 0.16
<i>NPY1R</i>	6.94 ± 17.54	1.28 ± 0.96	0.12 ± 0.12
<i>IFITM1</i>	5.00 ± 11.58	2.49 ± 1.18	0.002 ± 0.004
<i>hMLH1</i>	5.38 ± 15.30	0.16 ± 0.59	0.0003 ± 0.0006
<i>GREM1</i>	6.21 ± 17.29	1.56 ± 1.27	0.04 ± 0.09

Values are means with standard deviations. CRC: Colorectal cancer; CAM: Cancer associated mucosa; PMA: Percentage of methylated alleles.



**Supplementary table 2.** Association of CRC hypermethylation with gender.

	Female (n = 39)	Male (n = 67)	
<b>FOXF1</b>			
Hypermethylated ( $\geq 4\%$ )	74.4% (29)	61.2% (41)	<i>P=0.205</i>
Not hypermethylated ( $<4\%$ )	25.6% (10)	38.8% (26)	
<b>MGMT</b>			
Hypermethylated ( $\geq 5\%$ )	46.2% (18)	29.9% (20)	<i>P=0.0988</i>
Not hypermethylated ( $<5\%$ )	53.8% (21)	70.1% (47)	
<b>CA4</b>			
Hypermethylated ( $\geq 5\%$ )	20.5% (8)	20.9% (14)	<i>P=1.000</i>
Not hypermethylated ( $<5\%$ )	79.5% (31)	79.1% (53)	
<b>NPY1R</b>			
Hypermethylated ( $\geq 5\%$ )	28.2% (11)	13.4% (9)	<i>P=0.0744</i>
Not hypermethylated ( $<5\%$ )	71.8% (28)	86.6% (58)	
<b>IFITM1</b>			
Hypermethylated ( $\geq 7\%$ )	5.2% (2)	20.9% (14)	<i>P=0.0461</i>
Not hypermethylated ( $<7\%$ )	94.8% (37)	79.1% (53)	
<b>GREM1</b>			
Hypermethylated ( $\geq 6.5\%$ )	17.9% (7)	10.4% (7)	<i>P=0.3729</i>
Not hypermethylated ( $<6.5\%$ )	82.1% (32)	89.6% (60)	
<b>hMLH1</b>			
Hypermethylated ( $\geq 6\%$ )	25.6% (10)	7.5% (5)	<i>P=0.0182</i>
Not hypermethylated ( $<6\%$ )	74.4% (29)	92.5% (62)	
<b>Panel</b>			
Hypermethylated in $\geq 1$ marker	84.6% (33)	80.6% (54)	<i>P=0.79</i>
No hypermethylated marker	15.4% (6)	19.4% (13)	
<b>Age</b>			
Median (y)	66	71	<i>P=0.0181<sup>1</sup></i>
Mean (y)	64.59	70.43	
Standard deviation	14.46	10.46	
Standard error of the mean	2.316	1.277	
Normally distributed	Yes <sup>4</sup> /Yes <sup>5</sup> /Yes <sup>6</sup>	No <sup>4</sup> /Yes <sup>5</sup> /Yes <sup>6</sup>	
<b>Location of tumor</b>			
Right side	25.6% (10)	32.8% (22)	<i>P=0.5138<sup>2</sup></i>
Left side	74.4% (29)	67.2% (45)	

<b>Stage</b>			
I	10	12	
II	8	16	
III	13	19	
IV	8	20	<i>P=0.6043</i> <sup>3</sup>

---

Statistic test applied is two-sided Fisher's exact test where not differently indicated. <sup>1</sup> two sided unpaired t-test; <sup>2</sup> two-sided Fisher's exact test; <sup>3</sup> Chi square test, <sup>4</sup> Kolmogorov-Smirnov normality test; <sup>5</sup> D'Agostino & Pearson omnibus normality test; <sup>6</sup> Shapiro-Wilk normality test.

Values in brackets are absolute numbers.

**Supplementary table 3.** Association of CRC hypermethylation with tumor location

	<b>Right side (n = 32)</b>	<b>Left side (n = 74)</b>	
<b>FOXF1</b>			
Hypermethylated ( $\geq 4\%$ )	90.6% (29)	55.4% (41)	<b><i>P=0.0003</i></b>
Not hypermethylated ( $< 4\%$ )	9.4% (3)	44.6% (33)	
<b>MGMT</b>			
Hypermethylated ( $\geq 5\%$ )	40.6% (13)	33.8% (25)	<i>P=0.5157</i>
Not hypermethylated ( $< 5\%$ )	59.4% (19)	66.2% (49)	
<b>CA4</b>			
Hypermethylated ( $\geq 5\%$ )	59.4% (19)	4.1% (3)	<b><i>P&lt;0.0001</i></b>
Not hypermethylated ( $< 5\%$ )	40.6% (13)	95.9% (71)	
<b>NPY1R</b>			
Hypermethylated ( $\geq 5\%$ )	34.4% (11)	12.2% (9)	<b><i>P=0.0133</i></b>
Not hypermethylated ( $< 5\%$ )	65.6% (21)	87.8% (65)	
<b>IFITM1</b>			
Hypermethylated ( $\geq 7\%$ )	9.4% (3)	17.6% (13)	<i>P=0.3812</i>
Not hypermethylated ( $< 7\%$ )	90.6% (29)	82.4% (61)	
<b>GREM1</b>			
Hypermethylated ( $\geq 6.5\%$ )	31.3% (10)	5.4% (4)	<b><i>P=0.0008</i></b>
Not hypermethylated ( $< 6.5\%$ )	68.7% (22)	94.6% (70)	
<b>hMLH1</b>			
Hypermethylated ( $\geq 6\%$ )	34.4% (11)	5.4% (4)	<b><i>P=0.0003</i></b>
Not hypermethylated ( $< 6\%$ )	65.6% (21)	94.6% (70)	
<b>Panel</b>			
Hypermethylated in $\geq 1$ marker	93.8% (30)	77% (57)	<i>P=0.053</i>
No hypermethylated marker	6.2% (2)	23% (17)	
<b>Age</b>			
Median (y)	75	67	<i>P=0.1562<sup>1</sup></i>
Mean (y)	70.88	67.16	
Normally distributed	Yes <sup>2, 3</sup>	Yes <sup>2, 3</sup>	
Standard deviation	14.21	11.37	
<b>Gender</b>			
Female	10	29	<i>P=0.5138</i>
Male	22	45	

<b>Stage</b>		
I	6	16
II	8	16
III	9	23
IV	9	19

---

*P=0.9568<sup>4</sup>*

Right sided tumors derived from the cecum to the splenic flexure and left sided from the splenic flexure to the rectum. Statistic test applied is two-sided Fisher's exact test where not differently indicated. <sup>1</sup> two sided unpaired t-test; <sup>2</sup> Kolmogorov-Smirnov normality test; <sup>3</sup> D'Agostino & Pearson omnibus normality test, <sup>4</sup> Chi square test. Values in brackets are absolute numbers.

Supplementary table 4. Association of CRC hypermethylation with stage of disease.

	I (n=22)	II (n=24)	III (n=32)	IV (n=28)	
<b>FOXF1</b>					
Hypermethylated ( $\geq 4\%$ )	72.7% (16)	66.7% (16)	62.5% (20)	64.3% (18)	$P = 0.88$
Not hypermethylated ( $< 4\%$ )	27.3% (6)	33.3% (8)	37.5% (12)	35.7% (10)	
<b>MGMT</b>					
Hypermethylated ( $\geq 5\%$ )	13.6% (3)	50% (12)	43.8% (14)	32.1% (9)	$P = 0.049$
Not hypermethylated ( $< 5\%$ )	86.4% (19)	50% (12)	56.2% (18)	67.9% (19)	
<b>CA4</b>					
Hypermethylated ( $\geq 5\%$ )	18.2% (4)	20.8% (5)	18.8% (6)	25% (7)	$P = 0.92$
Not hypermethylated ( $< 5\%$ )	81.8% (18)	79.2% (19)	81.2% (26)	75% (21)	
<b>NPY1R</b>					
Hypermethylated ( $\geq 5\%$ )	22.7% (5)	33.3% (8)	18.8% (6)	3.6% (1)	$P = 0.051$
Not hypermethylated ( $< 5\%$ )	77.3% (17)	66.7% (16)	81.2% (26)	96.4% (27)	
<b>IFITM1</b>					
Hypermethylated ( $\geq 7\%$ )	22.7% (5)	12.5% (3)	9.4% (3)	17.9% (5)	$P = 0.55$
Not hypermethylated ( $< 7\%$ )	77.3% (17)	87.5% (21)	90.6% (29)	82.1% (23)	
<b>GREM1</b>					
Hypermethylated ( $\geq 6.5\%$ )	13.6% (3)	25% (6)	9.4% (3)	7.1% (2)	$P = 0.24$
Not hypermethylated ( $< 6.5\%$ )	86.4% (19)	75% (18)	90.6% (29)	92.9% (26)	
<b>hMLH1</b>					
Hypermethylated ( $\geq 6\%$ )	9.1% (2)	12.5% (3)	21.9% (7)	10.7% (3)	$P = 0.50$
Not hypermethylated ( $< 6\%$ )	90.9% (20)	87.5% (21)	78.1% (25)	89.3% (25)	

<b>Panel</b>	81.8% (18) 18.2% (4)	87.5% (21) 12.5% (3)	87.5% (28) 12.5% (4)	71.4% (20) 28.6% (8)	<i>P</i> = 0.35
Hypermethylated in ≥1 marker					
No hypermethylated marker					
<b>Age</b>					
Median (y)	72	76.5	64.5	66	<i>P</i> = 0.008 <sup>1</sup>
Mean (y)	71.09	74 *	63.69 *	66.43	<i>P</i> < 0.01 <sup>2</sup>
Normally distributed	Yes <sup>3, 4, 5</sup>	Yes <sup>3, 4, 5</sup>	Yes <sup>3, 4, 5</sup>	Yes <sup>3, 4, 5</sup>	
Standard deviation	11.18	9.514	13.99	11.39	
Standard error (SEM)	2.384	1.942	2.473	2.153	
<b>Gender</b>					
Female	10	8	13	8	
Male	12	16	19	20	<i>P</i> = 0.60
<b>Colonic location</b>					
Right side	6	8	9	9	
Left side	16	16	23	19	<i>P</i> = 0.96

Statistic test applied is Chi square test where not differently indicated. <sup>1</sup> 1-way ANOVA; <sup>2</sup> Tukey's Multiple Comparison Test (No other significant correlation between neither of two pairs of stage groups could be found); <sup>3</sup> Kolmogorov-Smirnov normality test, <sup>4</sup> D'Agostino & Pearson omnibus normality test. <sup>5</sup> Shapiro-Wilk normality test. Values in brackets are absolute numbers.

**Supplementary table 5.** Concomitance of target hypermethylation in colorectal cancer.

<b>Target</b>							
<b>FOXF1</b>	<b>FOXF1</b>						
<b>MGMT</b>	0.134	<b>MGMT</b>					
<b>CA4</b>	< <b>0.0001</b>	1.000	<b>CA4</b>				
<b>NPY1R</b>	0.066	0.796	<b>0.001</b>	<b>NPY1R</b>			
<b>IFITM1</b>	0.399	0.161	0.515	0.296	<b>IFITM1</b>		
<b>hMLH1</b>	<b>0.002</b>	0.152	< <b>0.0001</b>	<b>0.001</b>	1.000	<b>hMLH1</b>	
<b>GREM1</b>	<b>0.031</b>	0.132	< <b>0.0001</b>	< <b>0.0001</b>	1.000	< <b>0.0001</b>	<b>GREM1</b>

Contingency tables for each target pair were calculated using Fisher's exact test.

**Supplementary table 6.** Clinical data on samples used for methylome analysis

<b>Sample type</b>	<b>Sample ID</b>	<b>Age (years)</b>	<b>Gender</b>	<b>Location</b>
Blood (Healthy subject)	1	30	F	na
Normal mucosa (Healthy aged 50)	1	50	F	R
	2		F	
	3		M	
	4		M	
Normal mucosa (Healthy aged 75)	1	75	F	R
	2		F	
	3		M	
	4		M	
CIMP+ Cancer (BRAF V600E)	1	78	F	R
	2	82	F	
	3	73	M	
CIMP- Cancer (BRAF wt)	1	63	F	R
	2	70		
	3	84		
	4	85		

F: Female; M: Male; R: Right-sided; Left-sided; wt: wildtype; na: not applicable.



Supplementary table 7. Methylation of cancer-associated mucosa dissected by gender and colonic location.

Target	FOXF1		MGMT		CA4		NPY1R		IFITM1		hMLHI		GREM1	
Colonic side	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
<b>Methylated samples (%)</b>														
<b>All (n=106)*</b>	100 (32)	97.3 (72)	46.9 (15)	59.5 (44)	100 (32)	98.6 (73)	100 (32)	98.6 (73)	100 (32)	100 (74)	56.3 (18)	28.4 (21)	96.9 (31)	95.9 (71)
<b>Female (n=39)<sup>1</sup></b>	100 (10)	96.6 (28)	70 (7)	51.7 (15)	100 (10)	100 (29)	100 (10)	100 (29)	100 (10)	100 (29)	90 (9)	20.7 (6)	100 (10)	89.7 (26)
<b>Male (n=67)<sup>2</sup></b>	100 (22)	97.8 (44)	36.4 (8)	64.4 (29)	100 (22)	97.8 (44)	100 (22)	97.8 (44)	100 (22)	100 (45)	40.9 (9)	33.3 (15)	95.5 (21)	100 (45)
<b>Methylated alleles</b>														
<b>All (n=106)</b>														
Mean	0.96	1.00	0.31	0.26	1.53	0.90	1.64	1.13	2.06	2.68	0.32	0.09	2.15	1.30
Standard deviation	0.83	1.03	0.79	0.60	1.10	0.79	1.16	0.81	1.04	1.19	0.99	0.28	1.54	1.05
Median	0.68	0.74	0	0.04	1.04	0.67	1.25	0.87	2.03	2.55	0.02	0	1.95	1.03
Normal Distribution <sup>3</sup>	No	No	No	No	No	No	No	No	Yes	Yes	No	No	No	No
<b>Female (n=39)</b>														
Mean	0.71	0.80	0.82	0.29	1.32	0.91	1.61	1.18	1.93	2.57	0.92	0.07	2.21	1.01
Standard deviation	0.64	0.61	1.28	0.82	0.94	0.86	1.12	0.92	1.22	1.12	1.66	0.17	1.64	0.79
Median	0.50	0.61	0.25	0.02	0.87	0.75	1.14	0.88	1.89	2.38	0.32	0	1.95	0.83
Normal Distribution <sup>3</sup>	No	No	No	No	No	No	Yes	No	Yes	Yes	No	No	No	No
<b>Men (n=67)<sup>1</sup></b>														
Mean	1.07	1.13	0.08	0.25	1.63	0.89	1.65	1.10	2.12	2.75	0.05	0.10	2.13	1.49
Standard deviation	0.90	1.22	0.18	0.41	1.17	0.75	1.20	0.75	0.97	1.24	0.08	0.34	1.54	1.15
Median	0.74	0.94	0.0	0.04	1.23	0.59	1.28	0.86	2.03	2.65	0	0	1.96	1.11
Normal Distribution <sup>3</sup>	No	No	No	No	Yes	No	No	No	Yes	Yes	No	No	Yes	No

Values in brackets are absolute numbers; \* Right sided cancers: n=32, left sided cancers: n=74; <sup>1</sup> Female: n=39 (Right side: n=10; left side: n=29); <sup>2</sup> Male: n=67 (Right side: n=22; Left side: n=45); Age (y): Women: Right side: 67.8±17.8 (mean±SD), 74 (median), left side: 63.5±18.3, 64; Men: Right side: 72.3±12.5, 75, left side: 69.5±9.4, 69; Age is normally distributed in all 4 groups (Kolmogorov-Smirnov normality test) and there is no significant difference of means (One way ANOVA followed by Tukey's multiple comparison test); <sup>3</sup> Kolmogorov-Smirnov normality test.

Supplementary table 8. Concomitance of target methylation in cancer-associated mucosa (CAM).

Target	Right colon				
<b>FOXF1</b>	<b>FOXF1</b>				
<b>MGMT</b>	<b>&gt;0.05</b>	<b>MGMT</b>			
<b>CA4</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>	<b>CA4</b>		
<b>NPY1R</b>	<b>&lt;0.05</b>	<b>&lt;0.001</b>	<b>&gt;0.05</b>	<b>NPY1R</b>	
<b>IFITM1</b>	<b>&lt;0.01</b>	<b>&lt;0.001</b>	<b>&gt;0.05</b>	<b>&gt;0.05</b>	<b>IFITM1</b>
<b>hMLH1</b>	<b>&gt;0.05</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>hMLH1</b>
<b>GREM1</b>	<b>&lt;0.01</b>	<b>&lt;0.001</b>	<b>&gt;0.05</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>
					<b>GREM1</b>

Target	Left colon				
<b>FOXF1</b>	<b>FOXF1</b>				
<b>MGMT</b>	<b>&lt;0.001</b>	<b>MGMT</b>			
<b>CA4</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>	<b>CA4</b>		
<b>NPY1R</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>	<b>&gt;0.05</b>	<b>NPY1R</b>	
<b>IFITM1</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>IFITM1</b>
<b>hMLH1</b>	<b>&lt;0.001</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>	<b>&lt;0.001</b>	<b>hMLH1</b>
<b>GREM1</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>	<b>&gt;0.05</b>	<b>&gt;0.05</b>	<b>&lt;0.001</b>
					<b>GREM1</b>

P-values of different medians for each target pair were calculated using a repeated measures Friedman non-parametric test (absence of normal distribution was confirmed by Kolmogorov-Smirnov normality test), followed by a Dunn's multiple comparison test. Non-significant differences of medians are written in bold letters.

Supplementary table 9. Methylation of normal colorectal mucosa dissected by gender and colonic location.

Target	<i>FOXF1</i>		<i>CA4</i>		<i>NPY1R</i>		<i>IFITM1</i>		<i>GREM1</i>	
Colonic side	Right	Left	Right	Left	Right	Left	Right	Left	Right	Left
<b>Methylated sam- ples<sup>2</sup> (%)</b>										
<b>All (n=31)</b>	90.3 (28)	96.8 (30)	100 (31)	96.8 (30)	100 (31)	100 (31)	100 (31)	100 (31)	93.5 (29)	100 (31)
<b>Female (n=16)</b>	87.5 (14)	93.8 (15)	100 (16)	93.8 (15)	100 (16)	100 (16)	100 (16)	100 (16)	100 (16)	100 (16)
<b>Male (n=15)</b>	93.3 (14)	100 (15)	100 (15)	100 (15)	100 (15)	100 (15)	100 (15)	100 (15)	86.7 (13)	100 (15)
<b>Methylated alleles</b>										
<b>All (n=31)</b>										
Mean	0.71	0.98	2.52	2.15	2.33	2.10	1.66	2.20	2.23	2.04
Standard deviation	0.41	0.54	1.04	1.06	1.19	1.02	0.73	0.91	1.69	1.09
Median	0.78	0.98	2.41	2.15	2.16	1.98	1.58	2.08	1.91	1.76
Normal Distribution <sup>1</sup>	Yes	Yes	Yes	Yes	Yes/Yes/	Yes/No/N	Yes/No/N	Yes/No/N	Yes/No/N	Yes/No/Y
<b>Female (n=16)</b>										
Mean	0.68	1.10	2.80	2.25	2.28	2.21	1.66	2.06	2.41	2.27
Standard deviation	0.41	0.58	1.15	1.14	1.15	1.10	0.47	0.89	1.72	1.13
Median	0.74	1.06	2.65	2.17	2.23	1.96	1.72	1.89	2.02	1.98
Normal Distribution <sup>1</sup>	Yes	Yes	Yes	Yes	Yes	Yes/No/N	Yes	Yes/No/N	Yes	Yes/No/Y
<b>Male (n=15)</b>										
Mean	0.74	0.85	2.21	2.05	2.38	1.97	1.66	2.35	2.04	1.80
Standard deviation	0.43	0.48	0.83	1.00	1.26	0.95	0.95	0.94	1.70	1.02
Median	0.84	0.70	2.16	1.72	2.14	1.98	1.33	2.12	1.70	1.63
Normal Distribution <sup>1</sup>	Yes	Yes	Yes	Yes	Yes	Yes	Yes/Yes/	Yes/No/N	Yes	Yes

Values in brackets are absolute numbers. <sup>1</sup>Normality test used are Kolmogorov-Smirnov normality test, D'Agostino & Pearson omnibus normality test and Shapiro-Wilk normality test (in this order); <sup>2</sup> PMA  $\geq 0.01$ .

**Supplementary table 10.** Concomitance of target methylation in normal colorectal mucosa (NM)

Target	Right colon				Left colon			
<b>FOXF1</b>	<b>FOXF1</b>				<b>FOXF1</b>			
<b>CA4</b>	<0.001	<b>CA4</b>			<0.001	<b>CA4</b>		
<b>NPY1R</b>	<0.001	> <b>0.05</b>	<b>NPY1R</b>		<0.001	> <b>0.05</b>	<b>NPY1R</b>	
<b>IFITM1</b>	<0.001	<0.001	<0.01	<b>IFITM1</b>	<0.001	> <b>0.05</b>	> <b>0.05</b>	<b>IFITM1</b>
<b>GREM1</b>	<0.001	> <b>0.05</b>	> <b>0.05</b>	<b>GREM1</b>	<0.001	> <b>0.05</b>	> <b>0.05</b>	<b>GREM1</b>

P-values of different means for each target pair were calculated using repeated measures ANOVA (normal distribution was confirmed by Kolmogorov-Smirnov normality test), followed by a Tukey's multiple comparison test. Non-significant differences of means are written in bold letters.

Supplementary table 11. Ln-qMSP primer sequences.

Gene	Specification	Forward	Reverse	Location of amplicon
<i>FOXF1</i>	M	ggt tat gga ttt cgc gtc gtt c	aaa tet cgc tca acg tca aac g	+114 to +266
	U	gfg gfg tgg tta tgg att ttg t	aac taa taa atc tca ctc aac atc aaa ca	+111 to +273
	S	gta ttt tta tga aga gtt tgg gga gga g	aaa act cct aac cca aac cac ac	-6'450 to -6'303
<i>MGMT</i>	M	agg cgt cgt tgc gtt tgt atc g <sup>s</sup>	cgc cga acc taa aac aat cta cg <sup>s</sup>	-317 to -188
	U	ggg gga agt tgg gaa ggt gtt g <sup>s</sup>	caa cac ata ccc aat aca aca aca cca <sup>s</sup>	-331 to -166
	S	gfg ttt gga gga aaa ggg gtt	cta ccc ctc act aaa caa cca	-2'873 to -2'744
<i>C44</i>	M	ttt cgg gtt cgc tgt cgc tc	ctc gtc tgc cgc aac cac g	+221 to +374
	U	tgt tgg ggt ttt ggg ttg tgt gtg t	act ccc aac tca tct cac caa acc aca	+213 to +382
	S	ggg tat tgg ttg aag ata aag ggg aat g	ttc tcc cac tct tac aat aca ccc tac c	-53'316 to -53'166
<i>NPY1R</i>	M	cgc cgt cgt tta cgc tgc tac	cgc gac tgc att aca acc gaa aaa cg	+581 to +726
	U	tgg gtt tgg tgt gtt gtt tag tgt t	aaa aca ccc aat act aca aaa cac aac tca	+573 to +747
	S	gga taa ttt tta ggt ttt ggt ttg agg gag	tta cat ata ccc aac cct caa tct tta cc	-6'294 to -6'158
<i>IFITM1</i>	M	tta tat ttt att ggt ttt tgg tta att tat taa ttt ata aat ag	att ttc tac gta aaa cga aaa acc g	+100 to +270
	U	*	tft aaa aaa tat aat tft cta cat aaa aca aaa aac ca	+100 to +283
	S	tat agg gag tag gtt ttt ggt ggt ttt g	tcc cct act cta ctc taa aat ccc c	+13'599 to +13'508
<i>GREM1</i>	M	cgt cgg tat tta aac ggg aga cg	caa aac cgc cga aac tgc acg	-134 to -3
	U	ggt ttt ggg tgt tgt gtt ttg ttg	cac aac caa aac cac caa aac tca	-155 to +3
	S	gfg tga ata tgg tgg ttt gfg gta tg	cac aac taa ata aca aac atc cac tcc tac	-9'512 to -9'372
<i>hMLH1</i>	M	acg tag acg ttt tat tag ggt cgc <sup>s</sup>	tcc gaa aaa cga taa aac cct ata c <sup>s</sup>	-656 to -502
	U	gat gta gat gtt tta tta ggg ttg tgt <sup>s</sup>	tat cac cac ctc atc ata act acc ca <sup>s</sup>	-657 to -533
	S	agt ttg gag tgg tag gtt ttt aga gg	cat ccc caa att ctt act cct tct ttc	+982 to +1'134



Sequences are indicated in 5' -3' -direction; M: specific for methylated sequences; U: specific for unmethylated sequences; S: standardization primers; location of amplicon is indicated relative to transcription start site (TSS); \* same sequence as forward M-primer of *IFITM1*; § Primer sequence described in (4).

**Supplementary table 12.** Sequences of primers used for bisulfite sequencing

<b>Gene</b>	<b>Specification</b>	<b>Forward</b>	<b>Reverse</b>	<b>Location of amplicon</b>
<i>FOXF1</i>	Proximal	cgg tag agg tgt agt tag ttt	gcc caa aac cct taa ata act taa taa	-109 to +386
	Distal	cga gtg ttt tat taa gtt att taa ggg ttt tgg g	gac ctc ccc act cac ctt a	+351 to +1'038
<i>CA4</i>		gga ttt tag atg agg gat agg	tac taa cat ttc aaa cta aaa cac c	-154 to +439
<i>NPY1R</i>		cga aat tgg ttt tgg ggg att aa	ggt cca tca tta ctt ttc taa aaa aaa aaa cta aaa	+209 to +849
<i>IFITM1</i>		tta tat ttt att ggt ttt tgg tta att tat taa ttt ata aat ag	act aaa caa att ccc caa aac cac	+100 to +700
<i>GREM1</i>	Proximal	cgg ggg gag tag ata aag a	ggt tcc cta caa acc caa aac	-205 to +373
	Distal	cgg ggg tga att gtg aag	gaa act aaa acc tct aaa aca caa ttt aaa aaa t	+284 to +1'020

Proximal or distal indicates the location of the amplicon relative to the transcription start site.

**Supplementary table 13.** qRT-PCR primer sequences

<b>Gene</b>	<b>Forward</b>	<b>Reverse</b>	<b>Location of amplicon</b>
<i>FOXF1</i>	gta tct gca cca gaa cag cca c	gag aag aca aac tcc ttt cgg tca cac	953 to 1'058
<i>C44</i>	tga gca cta cga tgg cag aga g	tgt ctg ttc ctt gtc gta gta cag c	629 to 835
<i>NPY1R</i>	tct tat acc act ctc ctc tfg gfg c	cat gtt tct cct ttt tag gcg tat ata tat ctt g	893 to 998
<i>IFITM1</i>	ttc ata gca ttc gcc tac tcc gtg	aga ctg tca cag agc cga ata cc	339 to 517
<i>GREM1</i>	agc gcc acg cgt cga aag	gtg ctg ggc ctt gtc tgg c	109 to 282
<i>ACTB</i>	agc ctc gcc ttt gcc ga <sup>§</sup>	ctg gtg cct ggg gcg <sup>§</sup>	22 to 196
<i>GAPDH</i>	atc ttc ttt tgc gtc gcc ag	aat ccg tfg act ccg acc ttc	59 to 134

Sequences are indicated in 5'-3'-direction. Location of amplicon is relative to transcription start site; <sup>§</sup> Sequences according to (5).

## Supplementary References

1. Sabates-Bellver J, Van der Flier LG, de Palo M, Cattaneo E, Maake C, Rehrauer H, et al. Transcriptome profile of human colorectal adenomas. *Mol Cancer Res.* 2007;5:1263-75.
2. Certa U, Wilhelm-Seiler M, Foser S, Broger C, Neeb M. Expression modes of interferon-alpha inducible genes in sensitive and resistant human melanoma cells stimulated with regular and pegylated interferon-alpha. *Gene.* 2003;315:79-86.
3. Wojdacz TK, Hansen LL, Dobrovic A. A new approach to primer design for the control of PCR bias in methylation studies. *BMC Res Notes.* 2008;1:54.
4. Menigatti M, Truninger K, Gebbers JO, Marbet U, Marra G, Schar P. Normal colorectal mucosa exhibits sex- and segment-specific susceptibility to DNA methylation at the hMLH1 and MGMT promoters. *Oncogene.* 2009;28:899-909.
5. Kreuzer KA, Lass U, Landt O, Nitsche A, Laser J, Ellerbrok H, et al. Highly sensitive and specific fluorescence reverse transcription-PCR assay for the pseudogene-free detection of beta-actin transcripts as quantitative reference. *Clin Chem.* 1999;45:297-300.

Date of birth: 29th September 1979

**Education**

---

- 10/2009 – 01/2015 **PhD thesis in genetics** at the University of Basel, Department Biomedicine, Group Molecular Genetics, Prof. Primo Schär
- Focus: Epigenetics, Genetics, Carcinogenesis, Cancer progression, Cell biology
- *Elucidating a role of the DNA dioxygenase TET1 in the tumorigenesis of colorectal cancers displaying a CpG island methylator phenotype*
  - *Characterizing genetic and epigenetic changes during secondary tumor formation*
  - *Supervision of students during design, realisation and analysis of molecular genetic experiments in a practical course.*
- Grade: magna cum laude
- 03/2009 – 08/2009 **Research assistant** at the University of Basel, Department Biomedicine, Group Molecular Genetics, Prof. Primo Schär
- *Analyzing aberrant promoter DNA methylation in colorectal cancer*
- 06/2007 – 02/2009 **Master thesis in molecular biology** at the University of Basel, Department Biomedicine, Group Molecular Genetics, Prof. Primo Schär
- Identification and characterization of DNA methylation biomarkers in colorectal cancer
- Grade: 5.0 (good)
- 10/2003 – 03/2007 **Bachelor of Science in Biology** with major in molecular biology, University of Basel
- Grade: 4.8 (good)
- 03/2003 – 09/2003 **Assistant in a microbiological research laboratory**, Ciba Spezialitätenchemie Grenzach GmbH, Grenzach
- 04/2001 – 03/2003 **Training as chemical laboratory assistant**, Ciba Spezialitätenchemie Grenzach GmbH, Grenzach
- Grade: 1.5 (very good)
- 09/1997 – 07/2000 **Higher school certificate**, Ernährungswissenschaftliches Gymnasium, Lörrach
- Grade: 2.0 (good)

## Languages

---

<b>German</b>	Native speaker
<b>English</b>	proficient, oral and written
<b>French</b>	basic knowledge, oral and written

## IT skills

---

<b>Office applications</b>	Proficient in Word, Excel, PowerPoint
<b>Statistics</b>	Proficient in Prism GraphPad
<b>Bioinformatics</b>	Genome Browser, Ingenuity Pathway Analysis (IPA)

## Publications

---

- “Modulation of Age- and cancer-Associated DNA Methylation change in the Healthy colon by Aspirin and lifestyle”, Faiza Noreen, Martin Röösl, Pawel Gaj, Jakub Pietrzak, **Stefan Weis**, Patric Urfer, Jaroslaw Regula, Primo Schär, Kaspar Truninger, *J Natl Cancer Inst*, 106(7), published online June, 2014
- “Loss of TET1 explains a CpG island methylator phenotype and altered cell plasticity in a subset of colorectal cancers”, **Weis, S.**, Muraro, M.G., Noreen, F., Mele, V., Wirz, A., Truninger, K., Schär, P., manuscript in preparation
- “The 8p21.3 encoded SHOCA-2 acts as a tumor suppressor in colorectal cancer via repression of STAT3 activation”, Sébastien Loeffler, Michal Kovac, **Stefan Weis**, Salvatore Piscuoglio, Saulius Zuklys, Marcel Keller, Katrin Hafen, Daniel Hess, Kaspar Truninger, Inti Zlobec, Luigi Terracciano, Karl Heinimann, Primo Schär and Georg A. Holländer, manuscript in preparation

**References** are available on demand