Exploring tumor heterogeneity
Fessler, E.

Citation for published version (APA):
Fessler, E. (2016). Exploring tumor heterogeneity

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: http://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, Singel 425, 1012 WP Amsterdam, The Netherlands. You will be contacted as soon as possible.
Chapter 6

An epigenetic regulatory network associated with poor prognosis colorectal cancer is installed at the premalignant stage.


Manuscript in preparation.
Abstract
Colorectal cancer (CRC) has long been thought of as a molecularly well-characterized disease in which the progression from premalignant lesions to carcinomas follows a defined route. The tubular adenoma (TA) was recognized as the only precursor lesion able to spawn colorectal carcinomas. However, over the last decade reports have described the progression of sessile serrated adenomas (SSAs) to a malignant stage, thus complementing the classical path of CRC development with a serrated neoplasia pathway. SSAs have been suggested to be precursor lesions of the mesenchymal, poor prognosis subtype of colon cancer based on similarities in gene expression profiles. Yet, due to its short history and relatively low prevalence, this pathway lacks the same wealth of information that is available for the classical route. Here we set out to gain more insight into the molecular make-up of SSAs. We analyzed gene expression profiles of TA and SSA samples and demonstrate that unifying biological programs are active in SSAs and colon cancers of the mesenchymal subgroup. Employing a regulatory network approach, we have previously shown that microRNAs (miRs) of the miR-200 family tune the majority of genes differentially expressed in the mesenchymal CRC subtype. The members of the miR-200 family are lower expressed in mesenchymal tumors and thus allow the expression of, for instance, the zinc-finger E-box binding homeobox transcription factor ZEB1, which is a well-known inducer of the epithelial-mesenchymal transition (EMT) program. Inhibition of miR-200 expression is achieved by methylation of the miR-200 promoter regions, suggesting that this epigenetic mark plays a crucial role in determining the activity of the EMT program. Here we show that the molecular circuit which defines the mesenchymal CRC subtype, is already active in serrated precursor lesions, indicating it to be a feature installed early in tumorigenesis and maintained throughout this specific adenoma-carcinoma sequence.
Introduction

Colorectal cancer (CRC) is a broadly studied malignancy of which the developmental path, the so-called adenoma-carcinoma sequence, was described already 40 years ago and molecularly defined in 1990 (1, 2). However, despite its detailed characterization, CRC still poses a major clinical challenge representing the third most common cause of cancer-related deaths (3). Gene expression-based classification of CRC samples into four distinct consensus molecular subtypes (CMSs) led to the recognition of a poor prognosis, mesenchymal group of CRCs (4). In addition, this mesenchymal subgroup appears to respond poorly to chemo- as well as targeted therapy. Molecular markers commonly used in the clinic unfortunately do not allow the identification of this subtype, pointing to the need to gain more biological insight and generate diagnostic tools for these patients. Recently, we have reported that a microRNA (miR) regulatory network, involving the miR-200 family, regulates a large proportion of genes differentially expressed in the mesenchymal subgroup (CMS4). It has been shown that the miR-200 family members miR-200a, miR-200b, miR-200c, miR-429, and miR-141 play a role in orchestrating the epithelial-mesenchymal transition (EMT) (5-7), thus likely accounting for the mesenchymal appearance of the poor prognosis CMS4. The miR-200 family members are organized as polycistronic transcripts on chromosome 1 (miR-200ba429) and chromosome 12 (miR-200c141) (8) and their expression can be epigenetically silenced by promoter methylation (9-12). miR-200 promoter methylation levels are higher in tumors belonging to the mesenchymal CMS4 and present a tool to identify patients belonging to this subtype and displaying worse clinical outcome (13).

CRCs of the CMS4 group may in part develop following an alternative route of colorectal tumorigenesis, namely the serrated neoplasia pathway (14). In contrast to the tubular adenoma (TA), which has been recognized as a premalignant lesion leading to CRC via the classical adenoma-carcinoma pathway already in 1975 (1), alternative forms of premalignant, serrated, lesions were recognized and attributed malignant potential only in the last two decades. The term serrated lesions comprises multiple distinct forms of serrated adenomas (15), of which the sessile serrated adenoma (SSA) has been reported to be able to progress to a colorectal malignancy (16, 17). Gene expression-based profiling revealed that this specific type of benign polyp is closely related to the mesenchymal, poor prognosis subtype of colon cancer (14). To gain further insight into the relation of SSAs and CMS4 tumors, we herein analyzed gene expression data of SSA and TA samples with regard to biological programs defining the distinct CMSs. We show that pathways and signatures present in CMS4 tumors are also active in SSAs. Additionally, the miR-200 regulatory network defining the poor prognosis subtype is already installed at the benign state, as miR-200 family members are lowly expressed in SSA lesions due to promoter methylation. Our data thus indicate, that the molecular features of poor prognosis CRC –
expression of EMT-associated genes and miR-200 promoter methylation – are already active at a premalignant stage.

Results

Unifying biological programs are active in sessile serrated adenomas and colon cancers of the mesenchymal subtype. Sessile serrated adenomas (SSAs) present distinct colorectal cancer (CRC) precursor lesions. Molecularly and morphologically they can be differentiated from classical tubular adenomas (TAs) by their serrated morphology (Supplementary Figure S1A) and frequently by the presence of a $BRAF^{V600E}$ mutation and DNA hypermethylation (CpG island methylator phenotype positive; CIMP+) (Supplementary Figure S1B) (18-21). Gene expression data point to an overlap between mesenchymal colon cancers and SSAs (14), suggesting that a proportion of CMS4 cancers might originate from this distinct precursor lesion. Indeed, genes highly expressed in mesenchymal CMS4 vs CMS1-3 tumor samples were enriched in SSAs compared with TA samples (Figure 1A). Also on protein level SSAs related to CMS4 tumors, as the caudal-type homeobox transcription factor 2 (CDX2), whose expression is frequently lost in the mesenchymal colon cancer subtype (14), was downregulated in the epithelium of SSAs compared with TAs (Figure 1B). To gain more insight into biological similarities, we studied the expression of genes associated with specific biological processes that have previously been analyzed in CMS1-4 cancers (4). In particular, the expression patterns of migration and matrix degradation signatures, which were highly enriched in CMS4 samples (Supplementary Figure S2A and B), were examined. This revealed that SSAs, in comparison with TA samples, displayed elevated expression of both signatures (Figure 1C and Supplementary Figure S2B). More importantly, the same subset of genes within each signature that was highly expressed in CMS4 samples was also highly expressed in SSAs (Supplementary Figure S2C and D), highlighting the similarity between SSA precursor lesions and CMS4 carcinomas. One of the dominating biological programs in CMS4 tumors is the epithelial-mesenchymal transition (EMT) (4, 14), illustrated by the enrichment of a range of EMT-associated genes in this subgroup compared with CMS1-3 samples (Supplementary Figure S3A and B) (22, 23). Applying EMT signatures to gene expression data from TAs and SSAs revealed the higher expression of EMT-associated genes also in SSA polyps (Supplementary Figure S3A and B), and again a strong overlap between the genes highly expressed in CMS4 tumors and SSA lesions could be detected (Figure 1D). Taken together, these results indicate that SSAs are potential precursor lesions of CMS4 carcinomas as their gene expression profiles exhibit strong overlap in specific biological processes.
An epigenetic regulatory network associated with poor prognosis is installed at the premalignant stage.

A miR-200 regulatory network is active in precursor lesions. We have previously shown that a microRNA (miR) regulatory network is a determinant of subtype-specific gene expression and that the dominant cluster in this network was generated by the members of the miR-200 family. We and others have demonstrated that miR-200a, miR-200b, miR-200c, miR-429, and miR-141 regulate genes associated with EMT (5-7), and furthermore, our data show that the miR-200 family tunes the majority of genes differentially expressed between CMS4 and CMS1-3 tumors (13). To study the role of the miR regulatory network at the premalignant stage, we projected gene expression data of the adenoma set (GSE45270 (14)) onto the miR network of tumor samples from the TCGA set (24) (Figure 2). This approach revealed that most of the genes

**Figure 1.** Gene expression profiles of sessile serrated adenomas display high similarity to colon cancers of the mesenchymal subtype. (A) Genes highly expressed in CMS4 vs CMS1-3 tumor samples are enriched in SSA compared with TA samples revealed by GSEA. (B) Immunohistochemical staining of the epithelial subtype marker CDX2 on TA and SSA samples. (C) GSEA using a matrix remodeling signature (GO) applied to a gene expression data set comprising SSAs and TAs. (D) Genes present in an experimentally-derived EMT signature (Taube et al. 2010) are highly expressed in SSA vs TA samples and the same expression pattern can be observed in CMS4 compared to CMS1-3 tumor samples of the AMC-AJCCII-90 data set.
present in this network were highly expressed in SSA compared with TA polyps, thus following
the same trend that was observed for CMS4 vs CMS1-3 cancers, confirming the similarity
between SSAs and CMS4 tumors. Zooming in on the genes present in an experimentally-derived
EMT signature (22) revealed that most gene products were inversely correlated with the
respective miR. This supports the idea that members of the miR-200 family have a negative
impact on EMT-associated genes in precursor lesions (Figure 2). Moreover, these results indicate

Figure 2. Genes regulated by miRs lowly expressed in CMS4 tumors are highly expressed in SSA polyps. Differential gene expression data of 6 SSAs and 7 TAs is projected on a miR network derived from tumor samples of the TCGA data set. Displayed is the differential miR expression between CMS1-3 and CMS4 tumors of the TCGA data set (green: lowly expressed in CMS4). Genes predicted to be regulated by the miRs are shown based on their differential expression between SSA and TA samples (blue: lowly expressed in SSA, orange: highly expressed in SSA). The connection between miRs and genes is depicted in red (induction) or blue (repression) based on the influence of the miR on gene expression. EMT-associated genes from the Taube EMT signature are highlighted as rectangles.
An epigenetic regulatory network associated with poor prognosis is installed at the premalignant stage. That SSA lesions harbor the same gene regulatory mechanisms as CMS4 tumors, highlighting their potential affiliation to the same adenoma-carcinoma sequence.

**Epigenetic silencing of miR-200 expression is already installed in precursor lesions.** To further investigate the role of the miR-200 family members at the premalignant stage, we generated a set of 20 TAs and 25 SSAs. Quantitative real-time PCR revealed that all members of the miR-200 family were significantly lower expressed in SSAs compared to TAs (Figure 3A). The miR-200 family members are located on two chromosomes as polycistronic transcripts (miR-200ba429 on chromosome 1 and miR-200c141 on chromosome 12) (8). Their expression can be regulated on multiple levels, for example by the zinc-finger E-box binding homeobox transcription factor ZEB1, which represses their expression (5-7), but also by epigenetic silencing (9-12). We have previously shown that the methylation of the miR-200 promoter regions can serve as a tool to identify CMS4 tumors and that this epigenetic mark is instrumental in controlling miR-200 expression. Since DNA methylation represents a metastable event that can be inherited by daughter cells, we speculated that these methylation marks are already installed in the precursor

![Figure 3.](https://example.com/figure3.png)

**Figure 3.** Differential expression of miR-200 family members in SSAs and TAs is regulated by promoter methylation. (A) Detection of miR-200 family members by quantitative real-time PCR in 20 TA and 25 SSA samples (Boxes extend from the 25th to the 75th percentile, the lines indicate the median. The whiskers are drawn to the 5th and the 95th percentile; outliers are depicted as dots. *P < 0.05, **P < 0.01, ***P < 0.001; P-values are based on two-tailed Student's t-tests). (B) Methylation of the miR-200ba429 and miR-200c141 promoter regions in adenoma samples determined by pyrosequencing (n = 20 for TA and n = 16 for SSA, P < 0.001; P-values are based on two-tailed Student's t-tests. Box plots represent the same parameters as described in (A)). (C) The methylation of the miR-200ba429 and miR-200c141 loci is highly correlated. (D) The methylation of the miR-200ba429 locus correlates with the expression of miR-200b.
lesions and maintained throughout the adenoma-carcinoma sequence. Indeed, both promoter regions were significantly higher methylated in SSAs compared with TAs (Figure 3B). Furthermore, the methylation of both regions was highly correlated (Figure 3C). The expression of miR-200b inversely correlated with the methylation of its respective locus (Figure 3D), indicating the methylation mark to be decisive for miR-200 expression. Premalignant SSAs are therefore already endowed with the molecular circuitry that relieves the inhibitory constraint on the EMT program. Our data thus demonstrate that a program that was thought to be responsible for aggressiveness of late stage disease can already be installed at a premalignant stage of CRC.

Discussion
As we demonstrate herein, EMT and the miR-200 regulatory network, both dominating characteristics in the mesenchymal group of CRCs, are already installed in SSAs, precursor lesions that are associated with CMS4 tumors. Thus, even though the EMT process is thought to occur late in tumor progression and to mark the time point when a localized tumor is able to spread to distant organs (25), we show that already at a premalignant stage in CRC development this program can be active. This points to the fact that early in CRC formation SSA lesions possess features allowing them to progress into aggressive and metastatic types of CRC, highlighting the need for early identification and thorough surveillance. Indeed, in a mouse model of pancreatic cancer, activation of EMT and subsequent dissemination of cells was observed before full malignant conversion (26). In a follow up study Rhim et al. detected pancreatic epithelial cells in the circulation of patients with premalignant lesions (27).

It remains unresolved what proportion of SSAs develops into CMS4 tumors and what fraction gives rise to other subtypes. In particular, based on their genetic make-up (frequently \(BRAF^{V600E}\)-mutated and CIMP+) (18-21), SSA lesions could also be potential precursor lesions for MSI+ CMS1 cancers (28, 29). Indeed, based on gene expression, SSAs are predicted to progress to either CMS1 or CMS4 tumors (30). In line with this notion is the observation that some genes highly expressed in CMS4 and SSA samples of the signatures studied herein also displayed higher expression in CMS1 compared to CMS2+3 tumors (Figure 1D and Supplementary Figure S2C and D). Even though being lower in the average of tumor samples, some CMS1 tumors display high levels of miR-200 promoter methylation. Whereas this epigenetic mark is responsible for tuning CMS4-specific gene expression, it could be a bystander in CMS1 tumors with overall higher DNA methylation levels. While the reason for increased miR-200 promoter methylation in SSAs remains to be examined, we recently described that high activity of the TGF\(\beta\) signaling pathway could direct SSA lesions to develop to the mesenchymal CRC subtype (30). It is important to note that not all CMS4 cancers display \(BRAF^{V600E}\) mutations, while this
An epigenetic regulatory network associated with poor prognosis is installed at the premalignant stage.

An oncogenic hit was detected in the majority of SSAs studied herein (24/25, 96%; Supplementary Figure S1B), suggesting the existence of other precursor lesions developing to CMS4 malignancies. It is important to unravel specific adenoma-carcinoma sequences in depth in the future to be able to select the premalignant lesions developing to the poor prognosis CMS4. It will be crucial to define the signaling molecules and pathways driving distinct developmental routes. This warrants further studies comprising larger gene expression data sets of premalignant lesions of CRC. Our data demonstrate that still benign SSAs possess the molecular makeup of CMS4 tumors and we speculate that the formation of a distinct microenvironment provides cues for malignant conversion to this specific CRC subtype. Hence, our data stress the need to identify such tumors already at an early stage, as they might acquire a highly malignant potential early in tumorigenesis.

**Materials and methods**

**Sample collection.** Human tissues were obtained in accordance with the legislation in the Netherlands. Collection of adenomas was approved by the Medical Ethical Committee (MEC 09/146 and MEC 05/071; Academic Medical Center (AMC), Amsterdam) and tissue was collected following written informed consent of patients. Tubular adenomas (TAs) were obtained from familial adenomatous polyposis (FAP) patients. Sessile serrated adenomas (SSAs) were collected from serrated polyposis syndrome (SPS) patients.

5x20 μm frozen tissue sections were cut using a cryotome and stored at -80°C until further use. RNA was extracted using the miRNAeasy micro kit (Qiagen) or the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen) from 20 TA and 19 SSA polyps. In the detection of miR-200 family expression we included 6 SSA polyps used for gene expression analysis (GSE45270) (14).

Matching gDNA was extracted for all 20 TAs and for 16 SSAs using the High Pure PCR Template Preparation Kit (Roche) or the AllPrep DNA/RNA/miRNA Universal Kit (Qiagen). BRAF mutation status was determined by quantitative real-time PCR in a 10 μl total reaction volume consisting of 20 ng gDNA or cDNA, 5 μl SYBR Green (Roche), 1 μl forward and reverse primers (10 μM), and 3 μl H2O. Each sample was measured in triplicate for each primer pair on a Roche Light Cycler 480 II in accordance with the manufacturer’s instructions (annealing temperature of 60°C). gDNA or cDNA of a BRAF\textsuperscript{wildtype} and a BRAF\textsuperscript{V600E}-mutated cell line was included in each run as control. The resulting C\textsubscript{p}-values of the mutant primer pair was subtracted from the wild-type one. Samples were scored as BRAF\textsuperscript{V600E}-mutant if this difference was < 5. Primer sequences: BRAF\textsuperscript{wildtype}-forward: 5'-AGG TGA TTT TGG TCT AGC TAC AGT, BRAF\textsuperscript{V600E}-forward: 5'-AGG TGA TTT TGG TCT AGC TAC AGA, BRAF\textsuperscript{reverse}: 5'-TAG TAA CTC AGC AGC ATC TCA GGG C.

Two microgram of gDNA were subjected to bisulfite conversion using the EpiTect Bisulfite Kit (Qiagen) according to the manufacturer’s instructions. TaqMan-based methylation specific quantitative real-time PCR (MethyLight) was used to determine the CIMP-status. 8 CIMP-specific markers (CACNA1G, IGF2, NEUROG1, RUNX3, SOCS1, MLH1, CRABP1, CDKN2A) and a methylation-independent normalization control (ALU) were included in the analysis. Primer/probe sequences and protocol were described previously (14). Markers with a percent of methylated reference (PMR) value > 10 were considered positive for methylation; samples were defined as CIMP high (≥ 6 out of 8 CIMP markers have a PMR > 10), low (1-5 out of 8 CIMP markers have a PMR > 10) or negative (0 markers have a PMR > 10).

Gene set enrichment analysis. Gene set enrichment analysis (GSEA) was performed for two EMT gene sets (only upregulated genes of the Taube EMT signature were included), matrix remodeling (GO term; REACTOME; REACT_118779), and migration (MSigDB; http://software.broadinstitute.org/gsea/msigdb/cards/WU_CELL_MIGRATION.html) (22, 23, 31, 32). GSEA were performed using the default settings of the Broad Institute’s web application tool (http://www.broadinstitute.org/gsea/index.jsp). P-values indicating the significance of enrichment were estimated by 1,000 permutations.

Regulatory network inference. We employed a network-based approach to study the regulatory relationships between miRs and potential target genes as described in (13). Gene expression data are log\textsubscript{2}-transformed RPKM profiles (n = 270) in the TCGA Illumina GA data set, while the miR data are log-transformed RPM (reads per
million miR mapped, \( n = 255 \) obtained from (24). Together, 200 patient samples have both miR and gene expression data. We focused on 30 miRs downregulated (fold change < 0.71, FDR < 0.001) in CMS4 vs CMS1-3, and 1,437 genes differentially expressed between CMS4 and CMS1-3 (\(|\log_{2}\text{FC}| > 0.85, \text{FDR} < 0.001\). The miR expression data and gene expression data were standardized independently and merged for network inference. The bioconductor RTN package was employed to infer the regulatory network.

**Adenoma microarray data analysis.** Microarray data for 6 TAs and 7 SSAs (GSE45270) (14) were normalized and summarized using frozen robust multiarray analysis (fRMA) (33). Gene expression profiles were then collapsed from probe sets to unique symbols based on the criterion that the probe set of highest overall expression is selected. Differential expression analysis was performed using the limma package (34). Log2 fold changes in gene expression between SSAs and TAs were projected onto the regulatory network (Figure 2). The median-centered expression levels of signature genes for EMT, matrix remodeling, and migration were used to generate the heatmaps in Figure 2D, Supplementary Figure S2C and D, respectively.

**Quantitative real-time PCR.** RNA was extracted as described above. One microgram of total RNA was reverse transcribed to cDNA using the HiFlex buffer of the miScript II RT Kit (Qiagen). Quantitative real-time PCR was performed using the miScript SYBR Green PCR Kit (Qiagen) together with the following miScript Primer assays: Hs_RNU6-2_11, Hs_SNORA74A_11, Hs_miR-200b_3, Hs_miR-200a_1, Hs_miR-429_1, Hs_miR-200c_1, and Hs_miR-141_1 according to the manufacturer’s instructions. Data shown was normalized to the expression of SNORA74A, normalization to RNU6-2 yielded comparable results.

**miR-200 promoter methylation analysis.** gDNA was extracted and bisulfite converted as described above. Methylation detection was carried out using the PyroMark PCR system (Qiagen). Primers were designed using the PyroMark Assay Design Software 2.0 (Qiagen) and span 110 basepairs including 6 CpGs in the miR-200ba429 promoter region and 140 basepairs including 11 CpGs in the miR-200c141 promoter region. PCR reactions and pyrosequencing were performed following the manufacturer’s instructions. Depicted is the average methylation of all 6 or 11 CpGs taken together. Primer sequences: miR-200ba429-forward: 5’-GGT TTG AAT TGA TTT TTT GTG TTA GG; miR-200ba429-reverse: 5’-CCT CAA CCA AAA TCA AAC CTC A. miR-200c141-forward: 5’-ATT GTA GAG GGG GGA TGA G; miR-200c141-reverse: 5’-CCA AAT TAC AAT CCA AAC CC. miR-200c141-sequencing primer: 5’-GAT GAG GGT GGG TAA.

**Immunohistochemistry.** Four-micrometer sections of formalin-fixed paraffin-embedded tissues were used for immunohistochemistry. After rehydration, antigen retrieval was performed using a 10 mM sodium citrate buffer at pH 6.0 (Vector Laboratories) for 20 min at 98°C. Endogenous peroxidase was blocked with 3% H2O2 (VWR) in PBS for 15 min at RT followed by incubation in ultraV block (Immunologic) for 10 min at RT. Sections were incubated in primary antibody dilution at 4°C over night (anti-CDX2 (1:100, BioGenex, clone CDX2-88, MU392A)). Subsequently, post-antibody blocking (Immunologic) was added for 20 min at RT followed by Poly-HRP-Anti-mouse/rabbit/rat IgG (Immunologic) for 30 min at RT. Washing steps were performed using PBS. Sections were incubated in Bright DAB solution (Immunologic), rinsed in dH2O, and counterstained using hematoxylin (Sigma). After dehydration, slides were mounted using Pertex (HistoLab). Images were taken with a Leica TCS-SP2.

**Statistical analysis.** The correlation of miR-200 methylation and miR-200 expression was performed using linear regression and the \( P \)-values were based on Pearson correlation. For all other comparisons, unpaired two-tailed Student’s t-tests were used. A \( P \)-value of < 0.05 was considered significant.

**Acknowledgements**

This work was supported by an AMC Graduate School PhD Scholarship (EF), KWF grants (UvA2009-4416 and UvA2012-5735) KWF/Alpe d’Huzes program CONNECTION, MLDS focus project FP13-07 and NWO gravitation (JPM), KWF grants (UVA2011-4969 and UVA2014-7245), Worldwide Cancer Research (14-1164), the Maag Lever Darm Stichting (MLDS-CDG 14-03), and the European Research Council (EG-StG 638193) (LV), CityU Start-up Grant for New Faculty (7200455) and an early career scheme grant from the Research Grants Council of Hong Kong SAR (9048033) (XW).
An epigenetic regulatory network associated with poor prognosis is installed at the premalignant stage.

References

7. U. Burk et al., A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. *EMBO reports* **9**, 582-589 (2008).
Supplementary figures

**Supplementary Figure S1.** Morphologically and molecularly SSAs and TAs are distinct entities. (A) Representative histology of SSA and TA polyps, revealing the serrated morphology of SSA lesions. (B) *BRAF* mutation (white: *BRAF* wildtype, black: *BRAF*V600E) and CIMP status (white: CIMP negative, grey: CIMP low, black: CIMP high) in the SSA and TA samples used in this study (asterisks indicate data not available (n.a.)).
An epigenetic regulatory network associated with poor prognosis is installed at the premalignant stage.

Supplementary Figure S2. Unifying biological programs are active in CMS4 colon cancer and SSA samples. (A) Genes present in a matrix remodeling signature (GO) are highly enriched in CMS4 vs CMS1-3 tumor samples of the AMC-AJCCII-90 data set. (B) GSEA of a migration signature (MSigDB) in CMS4 vs CMS1-3 tumor samples of the AMC-AJCCII-90 data set (left) and in SSAs vs TAs (right). (C and D) Analyzing the expression of genes in (C) a matrix remodeling (GO) and (D) a migration (MSigDB) signature in the AMC-AJCCII-90 and the adenoma data set reveals that for each signature the same set of genes is highly expressed in CMS4 tumors and SSA lesions.
Supplementary Figure S3. Genes present in EMT signatures are highly enriched in CMS4 vs CMS1-3 colon cancer and in TA vs SSA samples. (A and B) GSEA of two EMT signatures ((A) Taube et al. 2010, (B) Gröger et al. 2012) in CMS4 vs CMS1-3 tumor samples of the AMC-AJCCII-90 data set (left) and in SSAs compared with TAs (right).