



DNA methylation of shelf, shore and open sea CpG positions distinguish high microsatellite instability from low or stable microsatellite status colon cancer stem cells

Rosa Visone^{‡,1,2}, Maria Giulia Bacalini^{‡,3}, Simone Di Franco^{‡,4}, Manuela Ferracin⁵, Maria Luisa Colorito⁴, Sara Pagotto^{1,2}, Noemi Laprovitera⁵, Danilo Licastro⁶, Mirco Di Marco^{1,2}, Emanuela Scavo⁴, Cristian Bassi⁷, Elena Saccenti⁷, Annalisa Nicotra⁴, Maria Grzes^{3,8}, Paolo Garagnani⁵, Vincenzo De Laurenzi^{1,2}, Nicola Valeri⁹, Renato Mariani-Costantini^{1,2}, Massimo Negrini⁷, Giorgio Stassi⁴ & Angelo Veronese^{*,2,10}

¹Department of Medical, Oral & Biotechnological Sciences, G. d'Annunzio University, Chieti-Pescara, Italy

²Center of Aging Science & Translational Medicine (CeSI-MeT), G. d'Annunzio University, Chieti, Italy

³IRCCS Istituto delle Scienze Neurologiche di Bologna, Bologna, Italia

⁴Cellular & Molecular Pathophysiology Laboratory, Department of Surgical, Oncological & Stomatological Sciences, University of Palermo, Palermo, Italy

⁵Department of Experimental, Diagnostic & Specialty Medicine (DIMES), University of Bologna, Bologna, Italy

⁶CBM S.c.r.l. Area Science Park, Trieste, Italy

⁷Department of Morphology, Surgery & Experimental Medicine, University of Ferrara, Ferrara, Italy

⁸Department of Molecular Biology, Institute of Genetics & Animal Breeding of the Polish Academy of Sciences, Jastrzebiec, Poland

⁹Division of Molecular Pathology, The Institute of Cancer Research, London, UK

¹⁰Department of Medicine & Aging Science, G. d'Annunzio University, Chieti-Pescara, Italy

*Author for correspondence: Tel.: +39 087 154 1498; a.veronese@unich.it

‡Authors contributed equally

Aim: To investigate the genome-wide methylation of genetically characterized colorectal cancer stem cell (CR-CSC) lines. **Materials & methods:** Eight CR-CSC lines were isolated from primary colorectal cancer (CRC) tissues, cultured and characterized for aneuploidy, mutational status of CRC-related genes and microsatellite instability (MSI). Genome-wide DNA methylation was assessed by MethylationEPIC microarray. **Results:** We describe a distinctive methylation pattern that is maintained following *in vivo* passages in immune-compromised mice. We identified an epigenetic CR-CSC signature associated with MSI. We noticed that the preponderance of the differentially methylated positions do not reside at CpG islands, but spread to shelf and open sea regions. **Conclusion:** Given that CRCs with MSI-high status have a lower metastatic potential, the identification of a MSI-related methylation signature could provide new insights and possible targets into metastatic CRC.

First draft submitted: 18 September 2018; Accepted for publication: 18 January 2019; Published online: 8 May 2019

Keywords: colon cancer stem cells • DNA methylation • MSI • MSS

Colorectal cancer (CRC) is the fourth leading cause of cancer-related deaths worldwide in males and the third in females [1]. Despite being one of the most preventable cancers, the overall 5-year survival rate has not improved in the past 15 years, amounting to 64.9% in 2013 compared with 64.8% in 2000 [2,3]. This multifactorial disease is thought to develop from intestinal crypt stem cells as a result of genetic, epigenetic and microenvironmental alterations [4–8]. Accumulating evidence supports the hypothesis that a small fraction of tumor initiating cells, called colorectal cancer stem cells (CR-CSCs) play a crucial role in CRC initiation, progression, dissemination and therapeutic resistance [6,9,10]. This subpopulation of cells exhibits biological characteristics of stem cells, such as self-renewing ability and a potential to differentiate into different types of progenitors [9,11,12]. CR-CSCs are characterized by alterations that lead to aberrant activation of signaling networks crucial for disease progression, including the constitutive activation of Wnt/ β -catenin and receptor tyrosine kinases (RTKs) pathways, and induction of the

epithelial–mesenchymal transition (EMT). All these pathways are normally activated in healthy stem cells at the base of the colonic crypts and are associated with the invasive/metastatic phenotype of CRC [13]. Among the markers that are specifically expressed by CR-CSCs there are CD44, CD166, CD133 antigens, and leucine-rich repeat-containing G-protein-coupled receptor 5 (LGR5), together with functional markers such as aldehyde dehydrogenase 1 family member A1 (ALDH1) and β -catenin (CTNNB1) [14,15].

In addition, CR-CSCs resist conventional chemo- and radio-therapy. This is attributed to several survival mechanisms, including: relatively quiescent nature, meaning that they have proliferative capacity but are not often cycling; greater ability to repair DNA and higher expression of the ATP-binding cassette (ABC) membrane transporters [16–19]. Therefore, even though CR-CSCs account for less than 1% of the tumor bulk, they are thought to represent the major source of recurrent disease after treatment, and to be capable of reconstituting the tumor at the original and/or distant sites, ultimately leading to therapeutic failure.

Several studies suggest that aberrant DNA methylation is one of the mechanisms driving tumor onset, development, progression and recurrence [20–24]. In CRC, hypermethylation is observed at CpG-rich promoter regions and results in transcriptional gene silencing [21]. On the other hand, hypomethylated regions are often localized in open sea areas of the genome and are linked with chromosomal instability (CIN), gene activation and loss of imprinting [25,26]. However, these data mostly derive from studies performed on primary tumors, characterized by intrinsic cellular heterogeneity [27], which dampens their biological relevance.

Here we report the epigenetic profiles of eight CR-CSC lines and compare these profiles to those publicly available for non neoplastic colorectal mucosa, adenomas, CRCs and matching metastases. Our results show that: seven out of eight CR-CSC lines display an intragenetic homogeneity; each cell lines disclose a distinctive epigenetic profile; the methylation status of 262 CpG can distinguish the microsatellite instability (MSI) status of the cell lines and that these CpGs are mostly located on shelf, shore and open sea genomic positions.

Materials & methods

Isolation & culture of CR-CSCs

CRC specimens were obtained from patients undergoing CRC resection, in accordance with the ethical standards of the institutional committee of the University of Palermo. CR-CSCs were isolated as previously described [28–30]. Briefly, tumor samples were mechanically and enzymatically digested with collagenase and hyaluronidase for 1 h at 37°C. Then the obtained digests have been cultured in ultralow adhesion flasks to select CSCs that do not undergo anoikis. The culture medium is based on serum-free advanced Dulbecco's Modified Eagle Medium: Nutrient Mixture F-12 (DMEM/F12), supplemented with EGF and bFGF. Cells are cultured in these conditions for several passages (3–4 months depending on the parental tumor histological grading and the number of CSCs in tumor specimens) and expanded by both enzymatic and mechanical dissociation. Then cells are characterized and eventually sorted for the expression of putative CSC markers. Finally, CR-CSCs are tested for their tumorigenic potential by subcutaneous injection in non-obese diabetic (NOD) severe combined immunodeficiency (SCID) gamma mice (NSG). Cells were maintained at 37°C in a 5% CO₂ humidified incubator. Short tandem repeat (STR) DNA profiling was carried out for the authentication of CR-CSC lines using a 24-locus STR kit (GlobalFiler™ STR kit, Thermo Fisher Scientific, MA, USA). STR analysis was performed on the ABIPRISM 3130 genetic analyzer (Thermo Fisher Scientific) following the manufacturer's instructions. DNA profiles of isolated CR-CSC lines were matched with their relative patient tumor tissues to ensure no contamination with standard cell lines and their identity.

MSI analysis

For MSI analysis, genomic DNA (gDNA) was extracted from CR-CSCs using the DNeasy Blood & Tissue kit (Qiagen, Hilden, DE) according to the manufacturer's instructions. Isolated gDNA was subjected to a multiplex PCR assay using the GeneQuality CC-MSI kit (AB ANALITICA, Padova, IT) following standard protocol and then analyzed by capillary electrophoresis on Genetic Analyzer ABIPRISM 3130 (Thermo Fisher Scientific). Detection of MSI status was carried out using a set of microsatellite markers including BAT25, BAT26, D2S123, D17S250, D5S346 (Bethesda Panel) and BAT40, NR21, NR24, D18S58, TGF β R2 (MSI-high, ≥ 4 markers unstable; MSI-low; 1–3 markers unstable).

Animal procedures

All the *in vivo* experiments have been performed in accordance with the Animal Care Committee Guidelines of the University of Palermo, using three mice per condition. The tumorigenic potential of CR-CSCs was assessed by harvesting 5×10^5 dissociated tumor cells, resuspending them into 100 μ l of 1:1 Matrigel (BD Matrigel Matrix Growth Factor Reduced, Becton, Dickinson and Company, NJ, USA) and subcutaneously injecting cells in NOD/SCID mice. Tumor growth was monitored and measured twice per week following cell injection using an electronic caliper.

Tumor volume was calculated using the formula: $(\text{largest measured diameter} \times [\text{smallest measured diameter}]^2 \times \pi/6)$ for up to 15 weeks. At the end points, when subcutaneous tumor reached 2 cm in diameter, or when mice showed the first signs of suffering (Italian Ministry of Health, authorization no. 154/2017-PR), animals were sacrificed accordingly to Directive 2010/63/EU Guidelines (D.lgs 26/2016). Xenograft tumors were collected and used for paraffin-embedded tissues, DNA/RNA/protein extraction and cell isolation.

Cell viability assay

The cell viability assay was assessed for up to 96 h, using the CellTiter 96[®] AQueous One Solution Cell Proliferation Assay kit (MTS) according to the manufacturer's instruction (Promega, WI, USA). To evaluate the CR-CSCs chemosensitivity to oxaliplatin, 3000 cells were plated into ultralow attachment 96-well plates (Corning, NY, USA) and exposed to vehicle or 50 μ M oxaliplatin (cat. 09512, Sigma Aldrich, MO, USA). The results were analyzed by using a multiwell plate reader (GDV programmable MPT Reader DV 990BV6).

Reverse transcription quantitative PCR

Total RNA from CR-CSCs was purified by using the RNeasy Mini Kit (Qiagen) and retro-transcribed with the iScript[™] gDNA Clear cDNA Synthesis Kit as recommended by the manufacturer (Bio-Rad, CA, USA). Relative expression level of selected genes was assessed by real-time PCR using the PrimePCR Custom Panel (Bio-Rad) according to manufacturer's instructions. Each sample was normalized to GAPDH and HPRT1 reference genes to obtain the Δ Ct $[(2^{-[\text{Ct}_{\text{gene}} - \text{Ct}_{\text{GAPDH}}]}) \text{ Ct}]$. Data, collected from QuantStudio 7 Flex Real-Time PCR System (Thermo Fisher Scientific) were analyzed with the PrimePCR Analysis software (Bio-Rad).

Methylation EPIC bead Chip analysis

A total of 1 μ g of extracted DNA was bisulphite-converted with the EZ-96 DNA Methylation-Gold[™] Kit, used according to the manufacturer's protocol (Zymo Research, CA, USA). Next, the 850 K DNA methylation array by Infinium MethylationEPIC BeadChip (Illumina, CA, USA) was performed on 4 μ g of bisulphite-converted DNA, following the Illumina Infinium HD Methylation protocol. This array includes 853,307 cytosine positions of the human genome (CpG sites, non-CpG sites and random SNPs). The methylation score for each CpG was represented as a β -value according to the fluorescent intensity ratio representing any value between 0 (unmethylated) and 1 (completely methylated). Arrays were scanned by HiScan (Illumina). DNA methylation data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO) and are accessible through GEO Series accession number GSE123367 (www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE123367).

Statistical analyses

Raw signal intensities generated using the Infinium MethylationEPIC BeadChip were extracted from .*idat* files using the minfi Bioconductor package. The following filtering criteria were applied: samples with more than 5% of sites with a detection p-value greater than 0.05 were removed (0 samples removed); probes with a detection p-value greater than 0.05 in more than 1% of the samples were removed (5059 probes removed). For each probe, β values ($\text{Meth}/[\text{Meth} + \text{Unmeth}]$) were calculated. Exploratory analysis of DNA methylation profiles was performed using RnBeads platform (<https://rnbeads.org/index.html>) [31]. Statistical analyses were performed using R software. For differentially methylated probes (DMPs) identification, the ANOVA test was applied and probes having Benjamini–Hochberg (BH) p-value < 0.05 were selected. For differentially methylated regions (DMRs) identification, we applied the analytical pipeline described in Bacalini *et al.* [32]. Briefly, we used the Infinium EPIC annotation to group the CpG probes located on the same island, shore or shelf associated to a gene, and for each of region we evaluated the differences in DNA methylation between the groups under investigation using a multivariate analysis of variance (MANOVA). In the comparison between Gr1 CR-CSCs and Gr2 CR-CSCs, DMRs with a BH-corrected p-value < 0.05 were selected. In the comparison between CR-CSCs and CRC samples

from datasets GSE42752 and GSE53051, a more stringent selection criteria for DMRs was adopted (Bonferroni corrected p -value < 0.01 , at least two adjacent CpG sites with >0.2 concordant difference of methylation between CRCs and CR-CSCs), in order to minimize potentially confounding batch effects. Hierarchical clustering was performed using the `heatmap.2` function in R `gplots` package, using default settings (euclidean distance function, complete agglomeration method). K-means clustering was performed using the R function `kmeans`, while the adjusted rand index was calculated using the `randIndex` function in the `flexclust` R package.

Library preparation & sequencing

Library preparation of exonic regions from 20 genes (*ACVR1B*, *APC*, *BRAF*, *CSMD3*, *CTNNB1*, *EDNRB*, *FAM123*, *FBXW7*, *GNAS*, *GPC6*, *KIAA1804*, *KRAS*, *NRAS*, *PIK3CA*, *PTPRD*, *SMAD2*, *SMAD4*, *SOX9*, *TCF7L2*, *TP53*; Supplementary Table 1) was performed according to the Ion AmpliSeq Library Kit 2.0 protocol (Thermo Fisher Scientific), starting with 20 ng of gDNA. Two 16-cycle multiplex amplification reactions of the regions of interest were performed by using AmpliSeq custom oligos. An Ion Xpress Barcode Adapters Kit (Thermo Fisher Scientific) was used to add Ion Torrent-specific motifs to amplicons. For purification, an Agencourt AMPure XP reagent (BeckmanCoulter, CA, USA) was used. Final libraries were quantified by using the Bioanalyzer instrument with the High Sensitivity DNA Kit (Agilent, CA, USA), diluted and pooled together in equimolar amounts. 25 μ l of a 26 pM pool of all libraries was mixed with Ion Sphere Particles and clonally amplified in an emulsion PCR, performed in accordance with the Ion PGM™ Hi-Q™ View OT2 Kit protocol and using the Ion OneTouch instrument (Thermo Fisher Scientific). Enrichment-System and Dynabeads MyOne Streptavidin C1 magnetic beads (Thermo Fisher Scientific) were used to enrich template-positive Ion Sphere Particles. Enriched samples were loaded onto Ion 318 chip and sequenced by using the Ion Torrent PGM, following the Ion PGM™ Hi-Q™ View Sequencing Kit protocol.

Sequencing data analysis & variant identification

Sequencing data analysis was conducted by using the Torrent Suite software v. 5.4.0 (Thermo Fisher Scientific). Briefly, low-quality reads were removed, adapter sequences trimmed and alignment against a reference genome (hg19) performed by using the Torrent Mapping Alignment Program. The Variant Caller plugin was used to identify variations from the reference sequence. The annotation of the variants was performed with ANNOVAR [33]. Insertions and deletions belonging to homopolymeric regions were removed, because sequencing error rate is high in these regions. The mutation waterfall plot was generated using the Bioconductor package GenVisR [34].

Copy number variation (CNV)

Copy number variation (CNV) was performed by reverse transcription quantitative PCR (RT-qPCR) using both the TaqMan technology (STUB1 ID: Hs02061495_cn; COX19 ID: Hs02372251_cn; JAG2 ID: Hs03069093; HES5 ID: Hs02713003_cn; Thermo Fisher Scientific) and the Roche Assay Design Center software to identify primers and relative UPL probes (https://lifescience.roche.com/en_it/brands/universal-probe-library.html; KRAS F TGTATGGGCTGTGACATTGC - KRAS R CCACCTGTTCTTCCACCATC, UPL #68; TP53 F TGTTCTTGCAGTTAAGGGTTAGTTT - TP53 R TGAAGTGGGCCCTACCTA, UPL#05; DLEU2 F ACGTTGTGCAGAACTTGAGAC - DLEU2 R TCTAAGCAACCTGGATTTTACACA, UPL#68). Briefly, 10 ng of DNA was amplified using the Universal Mastermix (Roche, Basel, CH), specific primers, fluorescent probes and the Human TaqMan® Copy Number Reference Assays RNase P as internal standard reference (Thermo Fisher Scientific). Each sample was analyzed in triplicate on CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad). Each well was normalized to RNase P to obtain the Δ Ct ($2^{-[\text{FAM dye Ct} - \text{VIC RNaseP dye Ct}]}$). All samples were then normalized to a calibrator (three DNA samples from healthy donors) to determine $\Delta\Delta$ Ct. Samples were considered to be carrying deletion or amplification if loss/acquisition of genetic material at the corresponding chromosome region was $\geq 25\%$ compared with the average values of normal controls.

Results

Genetic characterization of CR-CSC lines

We sequenced the hotspot mutation regions of 20 CRC-related genes (Supplementary Table 1) in eight primary CR-CSC lines obtained as described before [9,29,30] (Table 1). We focused on the most relevant genes associated with CRC development, including *TP53*, *APC*, *KRAS* and *SMAD4* [35,36], which respectively showed 8, 13, 5 and 6 nucleotide variations (NVs) affecting the coding sequences in 6, 8, 5 and 4 cell lines. Of note is the missense SNP

Table 1. Clinical and molecular features of colorectal cancer patients.

| Pt# | Gender | Age at diagnosis | Site | Grading | TNM | MSI status [†] |
|-----|--------|------------------|-------------|---------|------|-------------------------|
| #5 | M | 68 | Right colon | G3 | IIIB | MSI-H |
| #8 | F | 57 | Cecum | G3 | na | MSI-L |
| #9 | F | 53 | Sigma | G2 | na | MSI-L |
| #11 | F | 77 | Right colon | G2 | IIIB | MSS |
| #12 | M | 76 | Right colon | G2 | IIIC | MSI-H |
| #13 | M | 63 | Right colon | G3 | IIIC | MSI-L |
| #14 | F | 47 | Right colon | G2 | IIA | MSI-H |
| #21 | F | 60 | Right colon | G2 | na | MSI-L |

[†]MSI status is referred to the patient respective CR-CSC line.
CR-CSC: Colorectal cancer stem cell; MSI: Microsatellite instability; MSS: Microsatellite stable.

rs459552 in *APC* whose minor allele ‘T’ (MAF, 0.2019/24418-ExAC, NCBI-dsNP Short Genetic Variations) was present in all the CR-CSC lines. However, the association of *APC rs459552* with cancer remains unclear [37,38]. The allele frequencies of the NVs in *TP53*, *APC*, *KRAS* and *SMAD4* were mostly around 50 or 100%, hinting at hetero- or homozygosity and clonal homogeneity. However, in four cell lines, the genes mutations showed peculiar allele frequencies of 33% or 61–70%, pointing to the presence of three allele sets (Figure 1A; Supplementary Table 2). To verify this, we focused on *KRAS*, which revealed three non-synonymous NVs with a frequency of approximately 67% in three cell lines. We examined all the *KRAS* NVs (synonymous and non-synonymous) in those cell lines. In one of them, CR-CSC#12, we found the *rs1137282* synonymous polymorphism at a frequency of 36.6%, in other words, close to 33%, the frequency expected for three allele copies (Supplementary Table 2). This suggests that, at least in CR-CSC#12, there are three *KRAS* alleles. We then considered the frequencies of all the NVs found in the investigated genes in the eight cell lines (Figure 1B). Altogether, CR-CSC#5, #12 and #14 showed NV frequencies mostly around 50% (Figure 1B), pointing to near-diploid status, while CR-CSC#8, #11, #13 and #9 displayed allele frequencies suggesting amplification to three (33 and 66%) or four allele copies (25 and 75%). Only CR-CSC#21 presented a more complicated pattern of allele frequencies, which could reflect genetic and clonal heterogeneity.

To verify if the distribution pattern of the NVs frequencies reflected aneuploidy, we analyzed the CNVs of seven chromosomal regions known to be amplified or deleted in CRC (Figure 1C). The results of NVs frequency analysis were confirmed by CNV analysis. In fact, compared with the controls (healthy donors, HDs), CR-CSC#5, #12, and #14 (Group 1; Gr1) showed a moderate level of aneuploidy; CR-CSC#8, #11, #13 and #19 (Group 2; Gr2) displayed several DNA copy number alterations and CR-CSC#21 (Group 3; Gr3), showed a pattern intermediate between Gr1 and Gr2.

These results indicate that seven out of eight CR-CSC lines are genetically homogeneous, while one, CR-CSC#21, could comprise multiple subclones, in agreement with the fact that its *TP53* mutation presents a frequency of 70%, with no variation in gene copy number.

Given that metastatic capacity is associated with CIN [39–41] and with CD44v6 expression on CR-CSCs [30], we measured CD44v6 levels on the CR-CSC cell lines. As expected Gr2 and Gr3, that had higher rates of aneuploidy, also showed higher CD44v6 surface expression compared with Gr1 (Figure 1D). Overall, differences in CD44v6 expression were not significant, but higher CD44v6 expression correlated with increased levels of CRC stemness markers (Figure 1E). Indeed, critical genes involved in embryonic development, EMT and invasiveness, including *DKK1*, *FNI* and *TWIST*, showed higher expression in Gr2 compared with Gr1 (Figure 1E, left panel). Gr3 CR-CSC line (#21) show over-expression of genes related to cancer-stemness including TGFβs, WNT5A, ZEB1, MMP9, SERPINE1 and Vimentin (VIM), when compared with the other groups of CR-CSC lines (Figure 1E, central & left panel). Given that MSI is a key characteristic of CRC and is mutually exclusive with CIN, we analyzed the MSI status of the CR-CSCs. As expected, Gr1 displayed high microsatellite instability (MSI-H), whereas Gr2 and Gr3 showed low or stable microsatellite status (MSI-L, MSS; Table 1). This indicates that a defined MSI status can be present at the CR-CSC level.

DNA methylation profile of CR-CSC lines correlates with MSI status of CR-CSCs

We analyzed the genome-wide DNA methylation profile of the CR-CSC lines using the Illumina Infinium MethylationEPIC BeadChip. Unsupervised clustering based on all methylation values showed three main clusters, respectively comprising: CR-CSC#5, #12, #14; CR-CSC#8, #19, #11, #13 and CR-CSC#21 (Figure 2A), which matched the three groups previously identified through genetic profiling (Gr1, Gr2 and Gr3). When we plotted

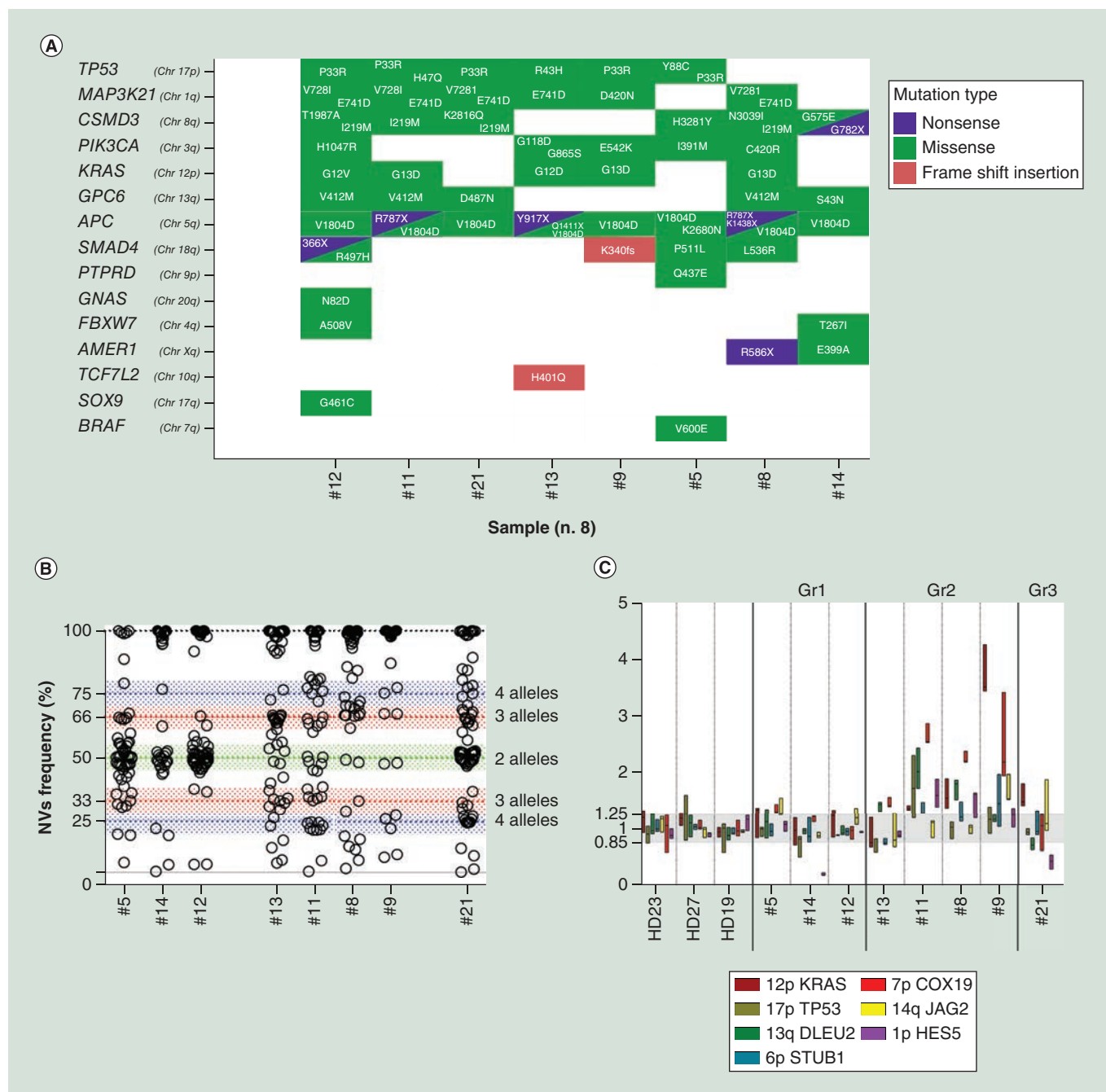


Figure 1. Genetic analysis of the colorectal cancer stem cell lines. (A) Mutational waterfall plot of eight colorectal cancer stem cell lines (only nucleotide variation affecting the coding sequence are represented). **(B)** Frequency analysis of all detected nucleotide variations. **(C)** Copy number values were determined by quantitative PCR; genes data were normalized to the endogenous references *RNASEP* with the $2^{-\Delta\Delta Ct}$ method. Data represent the mean \pm standard deviation of at least three technical replicates. **(D)** CD44v6 dosage by fluorescence activated cell sorter. **(E)** cancer stem cells-related gene expression analysis of Gr1-Gr2-Gr3 colorectal cancer stem cells. Fold variation >4 is shown in colour (Gr1 = yellow; Gr2 = blue; Gr3 = green).

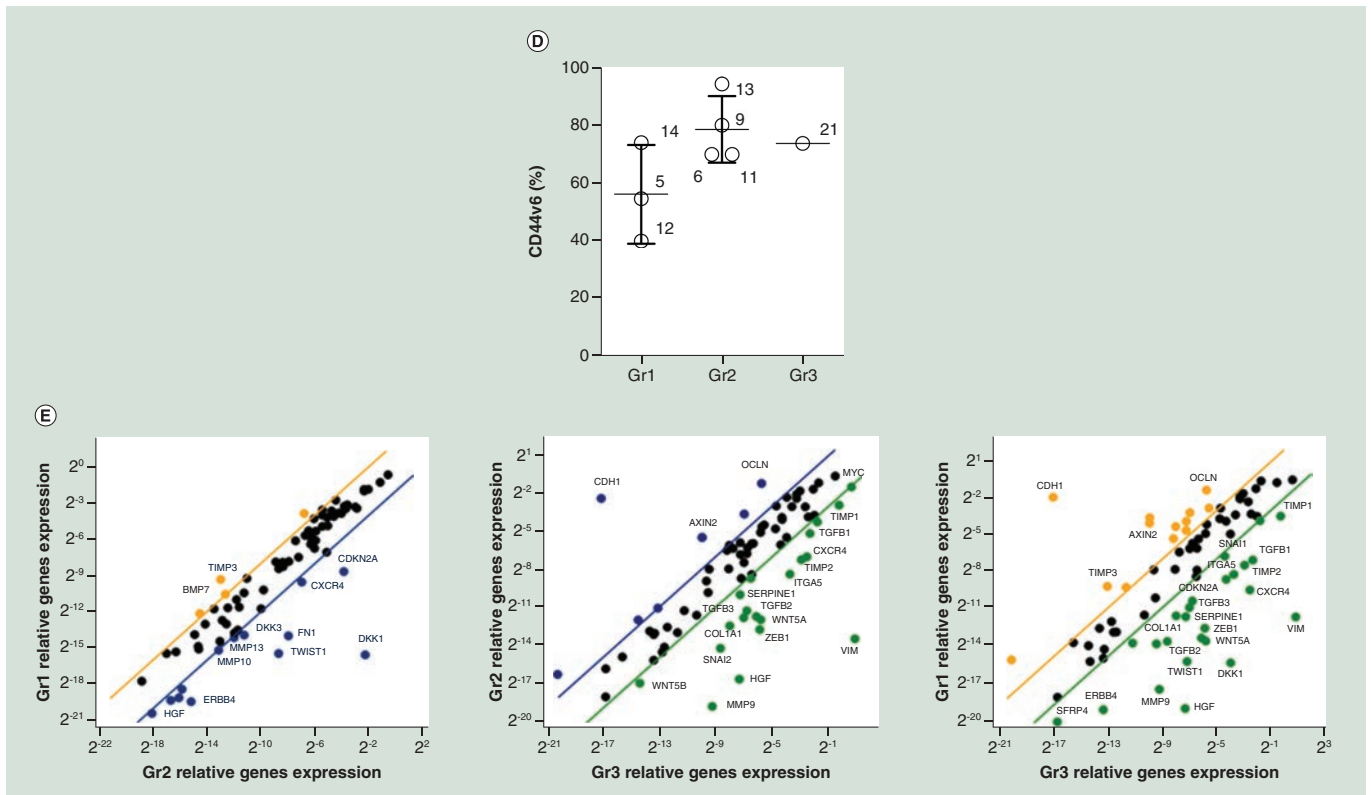


Figure 1. Genetic analysis of the colorectal cancer stem cell lines (cont.). (A) Mutational waterfall plot of eight colorectal cancer stem cell lines (only nucleotide variation affecting the coding sequence are represented). (B) Frequency analysis of all detected nucleotide variations. (C) Copy number values were determined by quantitative PCR; genes data were normalized to the endogenous references *RNASEP* with the $2^{-\Delta\text{ct}}$ method. Data represent the mean \pm standard deviation of at least three technical replicates. (D) CD44v6 dosage by fluorescence activated cell sorter. (E) cancer stem cells-related gene expression analysis of Gr1-Gr2-Gr3 colorectal cancer stem cells. Fold variation >4 is shown in colour (Gr1 = yellow; Gr2 = blue; Gr3 = green).

the density distribution of the methylation β values, we noticed that only the Gr1 CR-CSCs had a peak between 0.3 and 0.5 (Figure 2B). To rule out the possibility that this peculiar distribution was the result of a technical artifact, we performed the same analysis after normalization of raw data using the method proposed by Fortin *et al.* [42]. Gr1 samples maintained an evident peak of intermediate methylation values respect to non-normalized data (Supplementary Figure 1A). To gain insight on this pattern, we divided the probes according to genomic localization: CpG islands; shores, in other words, regions up to 2 kb from CpG island; shelves, in other words, regions from 2 to 4 kb from CpG island and open sea, in other words, the rest of the genome [43,44]. We found that the peak at intermediated DNA methylation values in Gr1 was due to CpG probes mapping in shores, shelves and open sea, but not in CpG islands (Figure 2C). Given that Gr1 displayed molecular features associated to favorable prognosis when found in tumor sample patients (MSI-H; Table 1), we investigated the differences in DNA methylation between Gr1 and Gr2 in order to identify genes whose aberrant epigenetic regulation could be linked to more aggressive tumor phenotype.

First, we used the ANOVA test to compare Gr1 and Gr2, and identified 262 DMPs (BH corrected p-value < 0.05 ; Figure 3A and Supplementary Table 3). A hierarchical clustering heatmap confirmed that Gr1 and Gr2 were clearly separated (Figure 3B) and showed that a large fraction of DMPs had intermediate methylation values in Gr1 that resembled the previously observed pattern (Figure 2B–C). 83 of the 262 DMPs had DNA methylation values between 0.3 and 0.5 in all the Gr1 samples but in none of the Gr2 samples. 12 out of these 83 probes mapped in the CpG islands, 13 in shores, 4 in shelves and 54 in open sea regions.

Thereafter, we used an alternative analytical pipeline specifically designed to identify DMRs, in other words, regions where multiple adjacent CpG sites were differentially methylated among the compared groups [32]. Focusing on CpG islands, shores and shelves associated to genes, we identified 282 DMRs (BH corrected p-value < 0.05 ;

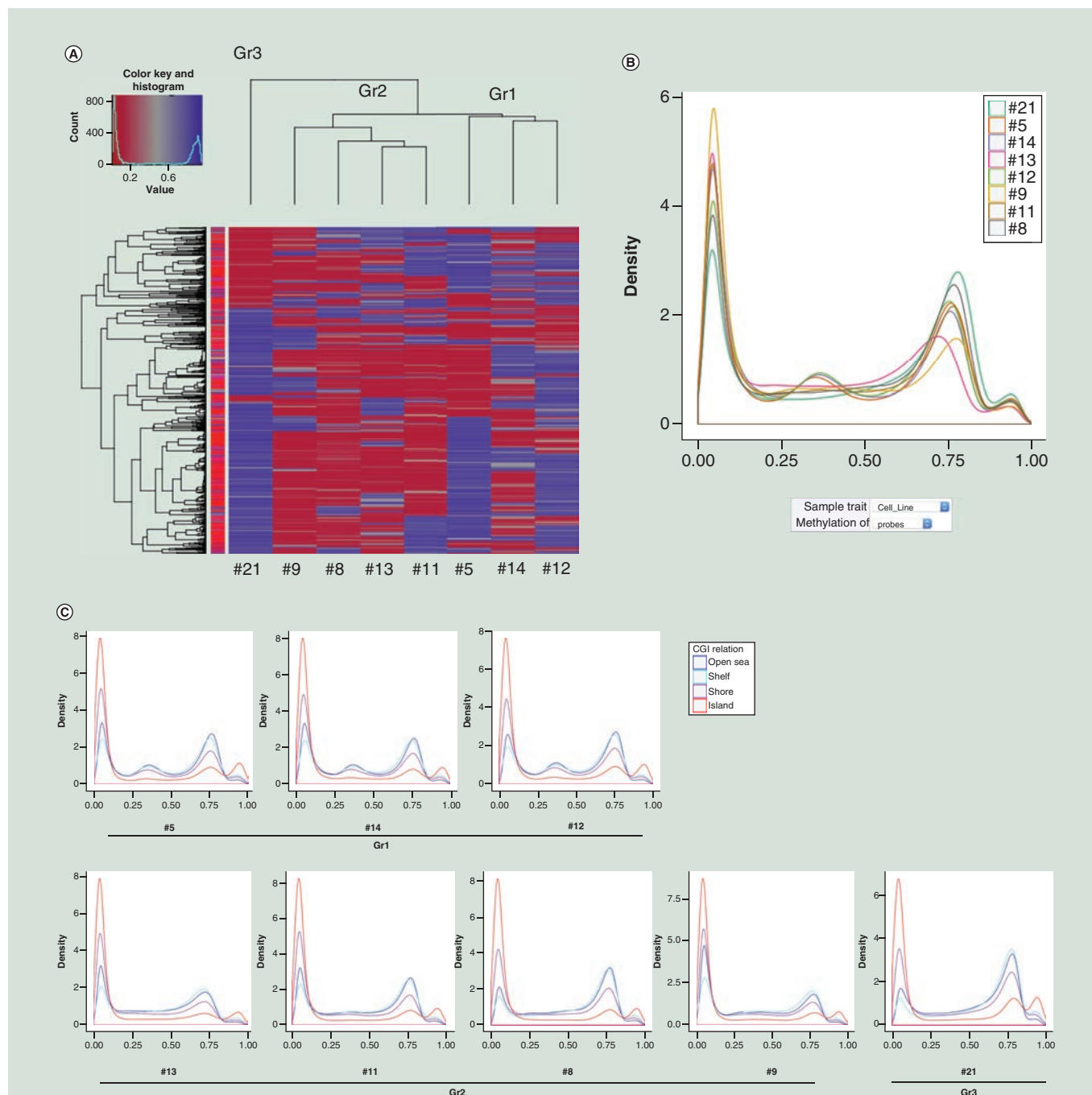


Figure 2. DNA methylation profile of colorectal cancer stem cell lines. (A) Hierarchical clustering of colorectal cancer stem cell samples based on all Infinium EPIC methylation values. The heatmap displays only the top 1000 CpG sites with the highest variance across all samples. **(B)** β value density estimation of the Infinium EPIC CpG sites in the eight colorectal cancer stem cell lines. **(C)** For each sample, β value density estimation of the Infinium EPIC CpG sites divided according to probe category. Not normalized data are plotted.

Supplementary Table 4). Among the genes, *MERTK*, *CPE*, *HEY2* and *FOXO3* stood out because of their known association with cancer aggressiveness and metastatic potential [45–49].

Genome-wide methylation analysis of CR-CSC xenografts

As demonstrated by unsupervised clustering, the CR-CSCs and the CSCs derived from the xenografts formed after their engraftment into NOD-SCID (CR-XenoCSCs) tended to maintain similar genome-wide DNA methylation

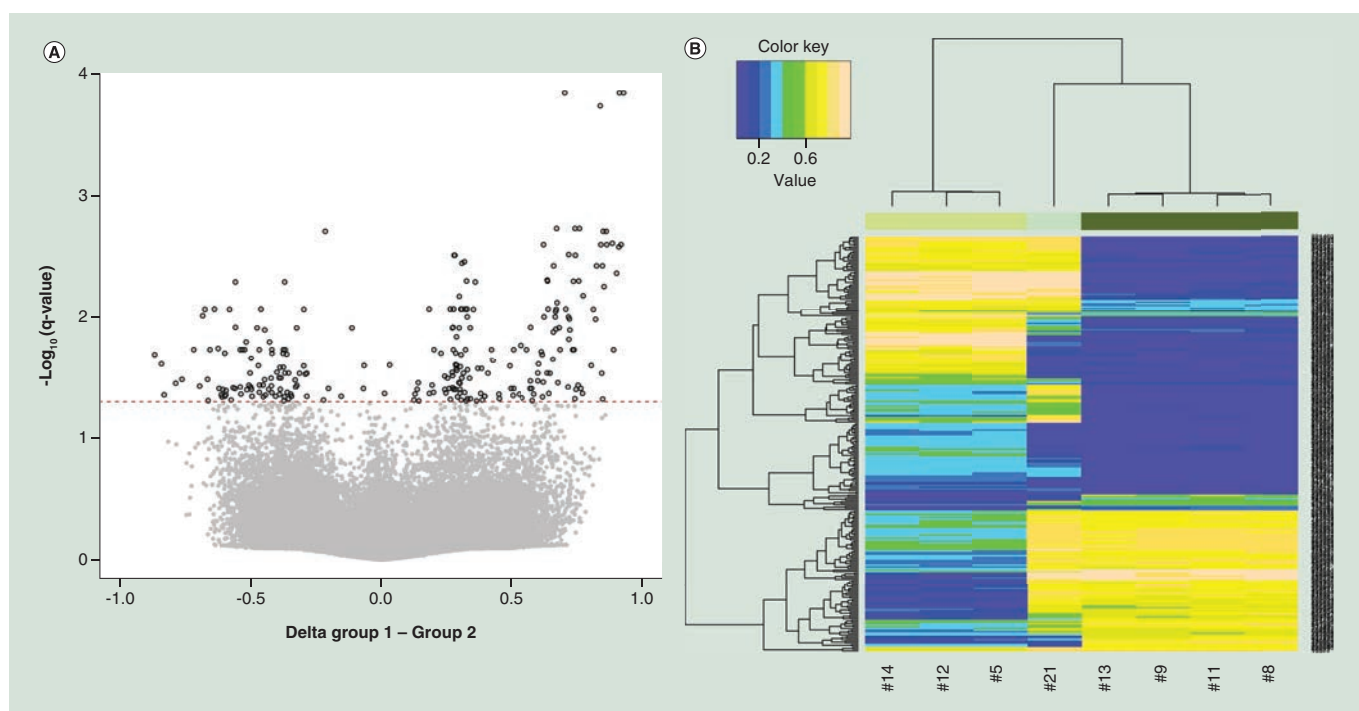


Figure 3. Gr1 versus Gr2 DNA methylation profile. (A) Volcano plot of differentially methylated probes between Gr1 and Gr2. The difference between mean DNA methylation values in Gr1 and Gr2 samples is plotted on the x axis, while the nominal Benjamini–Hochberg-corrected p-value (q-value) is reported on the y axis ($-1 \times \log_{10}$ scale). The dotted line corresponds to a Benjamini–Hochberg-corrected p-value of 0.05. (B) Hierarchical clustering of samples based on the methylation values of the 262 differentially methylated probes between Gr1 and Gr2.

profiles (Figure 4A). Accordingly, the peculiar distribution of β values of Gr1 and Gr2 samples was also maintained in CR-XenoCSCs (Supplementary Figure 1B). Paired t-test identified 3176 DMPs (BH corrected $p < 0.05$; Supplementary Table 5) between the CR-CSC and CR-XenoCSCs pairs, but intrapair methylation differences were low, mostly > -0.1 or < 0.1 (Figure 4B).

In vitro & *in vivo* characterization of CR-CSC lines

We characterized the eight CR-CSCs for their proliferative and tumorigenic capacity through *in vitro* and *in vivo* experiments. The proliferative potential was assessed *in vitro* in presence or absence of oxaliplatin, to establish whether chromosomal or MSI status could correlate with proliferation and chemoresistance. Despite the low number of cell lines tested, our results show a less proliferative phenotype in the Gr1/MSI-H subset compared with the Gr2/MSS. Interestingly, CR-CSC #21 (Gr3) showed very low growth rate coupled with resistance to oxaliplatin (Supplementary Figure 2A–C). The little differences observed *in vitro* disappeared *in vivo* when the CR-CSCs were injected subcutaneously into NOD/SCID mice. No significant differences in tumor growth or survival rate of the xenografted mice were identified (Supplementary Figure 2D).

Comparison between the DNA methylation signatures of the CR-CSC lines & of patient-derived normal & neoplastic colorectal tissues

Two Infinium450k datasets were downloaded from the Gene Expression Omnibus – GEO repository. The first dataset (GSE42752) [50] included 22 primary CRCs, their matched healthy mucosa samples and 19 cancer-unrelated colon tissues. The second dataset (GSE53051) [25] included 18 normal colon tissues, 10 colon adenomas, 9 primary CRCs and 16 CRC lung or liver metastases. Only common probes (intersection between datasets: 450626 probes) were considered to compare the β value distributions in the Infinium450k and EPIC datasets. The DNA methylation landscape of the CRCs and of the CR-CSCs was largely different, with the CSCs generally hypomethylated compared with the CRCs (Figure 5A). We cannot exclude *a priori* that the observed differences are at least in part the result of a batch effect, as the CR-CSCs and the CRCs datasets have been generated by

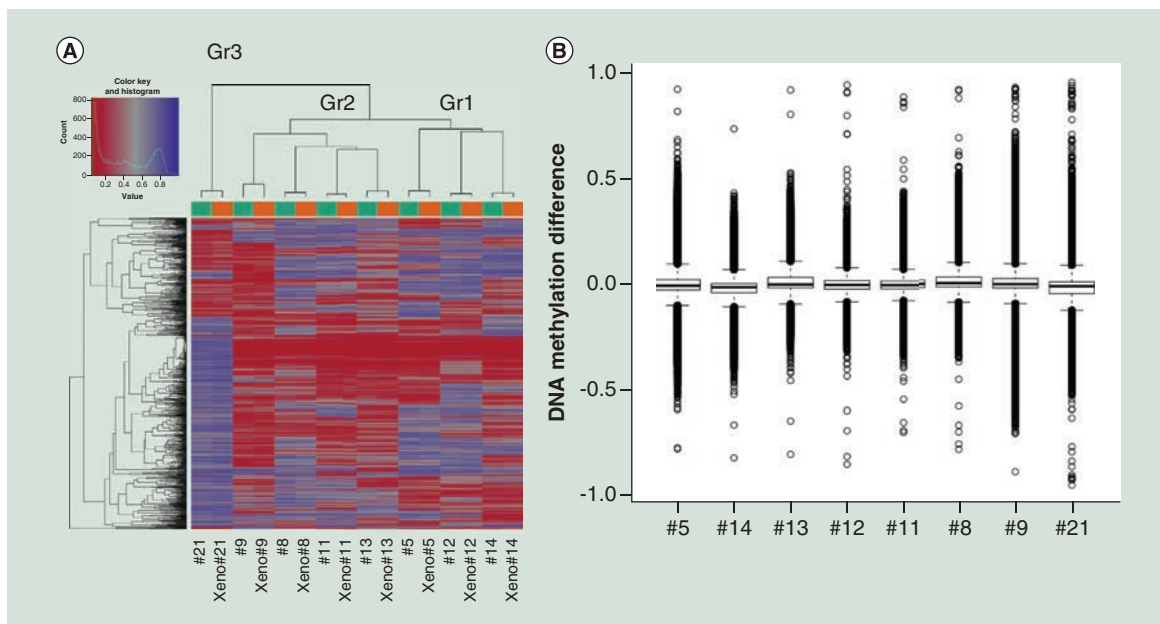


Figure 4. Colorectal cancer stem cell lines versus colorectal-Xeno cancer stem cell lines DNA methylation profile. (A) Hierarchical clustering of colorectal cancer stem cell. and colorectal-Xeno cancer stem cell samples based on all Infinium EPIC methylation values. The heatmap displays only the top 1000 CpG sites with the highest variance across all samples. **(B)** Boxplots of DNA methylation differences between colorectal cancer stem cell and colorectal-Xeno cancer stem cell pairs.

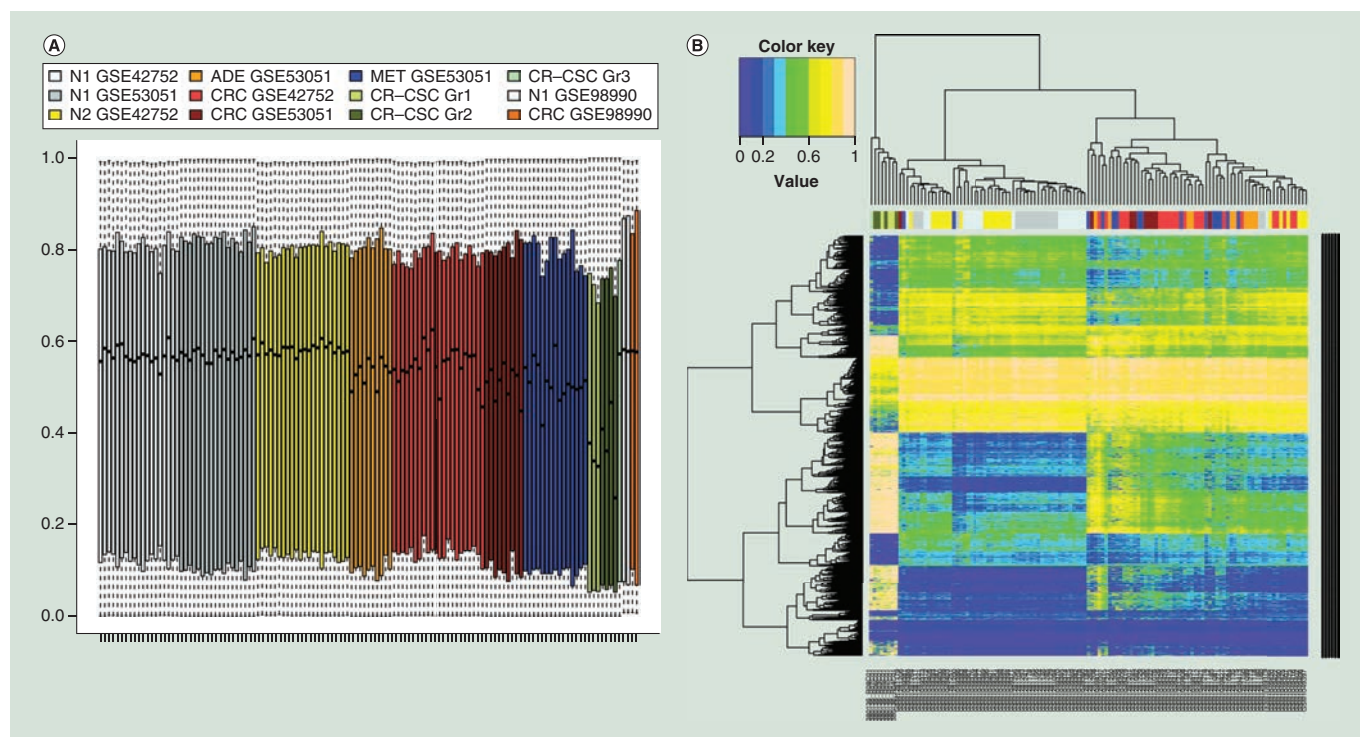


Figure 5. Comparison between DNA methylation profiles of colorectal cancer stem cell lines and primary tumors. (A) Boxplots of DNA methylation values in colorectal cancer stem cell lines and in cancer-unrelated colon tissue (N1), colorectal cancer healthy-matched mucosa (N2) ADE, (primary) CRC and MET CRCs samples included in GSE42752, GSE53051 and GSE98990 datasets. **(B)** Hierarchical clustering of the samples described in A (CR-CSC lines, GSE42752 and GSE53051 datasets), based on DNA methylation values of the 2967 differentially methylated regions between CR-CSCs and CRC samples. ADE: Adenoma; CRC: Colorectal cancer; CR-CSC: Colorectal cancer stem cell; MET: Metastatic.

different laboratories using different versions of the same microarray platform. However, two considerations suggest that the distinct patterns observed in CR-CSCs and CRCs have a true biological basis. First of all, a small dataset (GSE98990) containing two normal samples and two CRC samples, processed using the EPIC microarray, had a distribution of β values similar to that of the samples in GSE42752 and GSE53051 datasets and completely distinct from the CR-CSCs (Figure 5A). Secondly, hypomethylation of CR-CSCs was not generalized, as it was specific for shores, shelves and open sea CpGs, while a trend toward hypermethylation was observed for the distribution of the β values in CpG islands (Supplementary Figure 3).

To gain further insights into the DNA methylation differences between CRCs and CR-CSCs, we compared our dataset of CR-CSCs with CRCs samples from datasets GSE42752 and GSE53051. To minimize the potentially confounding batch effect discussed above and to maximize the identification of biologically relevant signals, we focused on the DMRs mapping in regions associated to genes [32]. Using stringent criteria (Bonferroni corrected p-value < 0.01, at least two adjacent CpG sites with >0.2 concordant difference of methylation between CRCs and CR-CSCs) we selected 2967 DMRs mapping in 2688 genes (Supplementary Table 6). Compared with CRCs, we found that in CR-CSCs DNA methylation increased at CpG islands (hypermethylation in 78% of the DMRs) but decreased in shores and shelves (hypomethylation in 84 and 97% of the DMRs, respectively).

We used the DMRs associated to genes to cluster the samples of normal mucosa, adenoma, primary CRC and metastases included in the GSE42752 and in GSE53051 datasets. Cluster analysis using the most significant CpG site within each DMR suggested that the epigenetic profile of the selected DMRs tended to accompany CRC progression (Figure 5B). Figure 6 shows examples of hypermethylation of gene promoters in CR-CSCs, intermediate methylation in CRCs and hypomethylation in non-neoplastic mucosa.

17 DMRs mapped in microRNA genes (Supplementary Table 6), including *miR-124-3* and *miR-129-2*, both found hypermethylated in CR-CSCs in this study and already described as hypermethylated in CRCs compared with normal tissue [51,52]. Moreover, we found 31 DMRs mapping in imprinted *loci* (Supplementary Table 6), including the *IGF2/H19* locus and the *KCNQ1* gene both at 11p15, associated in several studies with CRC [53–55].

We plotted the CpG methylation status of the genes associated to cell stemness [6]. We found these genes, with few exceptions, having complete promoter demethylation in CR-CSCs (Supplementary Figure 4), which could attest promoter accessibility of these genes to sustain stemness and/or the reprogramming of differentiated CRC cells toward the CR-CSC phenotype [56–58].

Finally, we analyzed the GSE42752 and GSE53051 datasets to evaluate the DNA methylation values of 262 DMPs that distinguish Gr1 and Gr2 CR-CSCs (Supplementary Table 3). Of these, 135 were included in both datasets. Unsupervised clustering using these 135 probes showed that the Gr1 CR-CSCs clustered separately from the other samples (Figure 7). Furthermore, the GEO samples tended to cluster according to CRC progression from normal mucosa to CRC. We performed the same analyses for the 83 DMPs that had methylation values between 0.3 and 0.5 in Gr1 but not in Gr2, 38 of which were included in GSE42752 and GSE53051 datasets (Supplementary File 1). Clustering of GEO samples according to CRC progression tended to be recapitulated also using this small set of probes (Supplementary Figure 5; K-means clustering with $k = 4$, adjusted rand index: 0.38).

Discussion

We analyzed the global methylation status of eight primary CR-CSC lines obtained from both primary CRCs and their respective xenograft-derived tumors. Genetic investigation revealed a high grade of homogeneity in each of the seven out of eight CR-CSC lines, thus permitting a remarkable degree of accuracy in the epigenetic analysis. Only one out of eight lines presented a genetic profile that could reflect the presence of subclones. The homogeneity of the CR-CSCs was also epigenetically maintained after rounds of *in vitro* and *in vivo* passages, highlighting the experimental and technical sustainability to the propagation and study of the CR-CSC lines. These cells presented a typical CRC cells mutational landscape involving *TP53*, *KRAS*, *SMAD4* and *APC* genes [35]. However, MSI status differed in these cells, a finding that could be relevant for therapeutic targeting. In fact, high immunogenicity, induced by neo-antigen expression of neo-antigen through mismatch repair deficiency [59] could be relevant for the eradication of the CR-CSC fraction of CR-CSCs within MSI-H CRCs.

By using three different techniques, we were able to discriminate the MSI-H CR-CSCs (Group 1) from those with MSS/MSI-L features and CIN phenotype (Group 2). These two groups were distinguishable based on: frequencies of genetic variations in 20 CRC associated genes investigated by next generation sequencing (NGS), CNV analysis of seven chromosomal regions frequently deleted or amplified in CRC or methylation status of 262

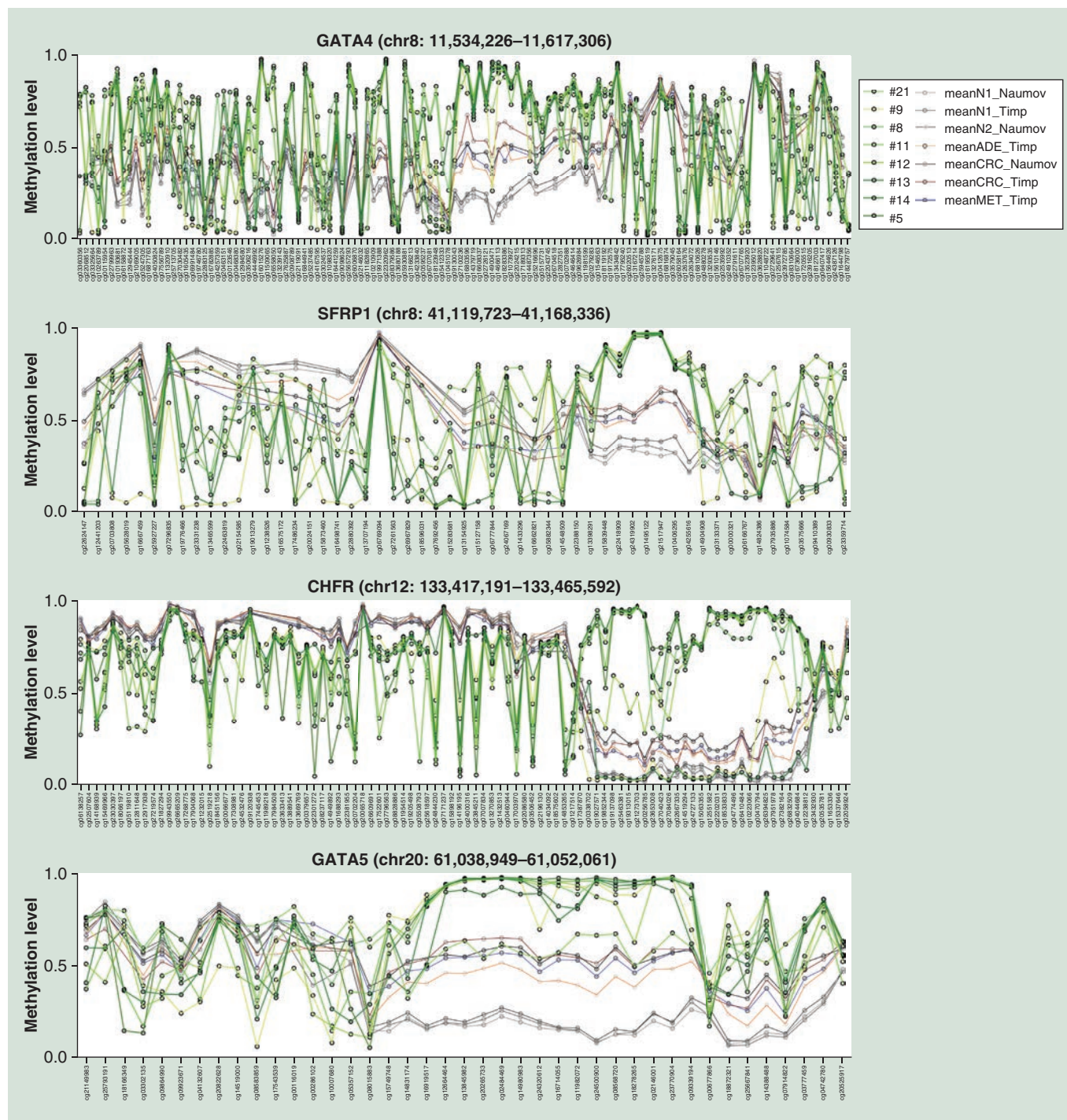


Figure 6. Methylation levels of a selection of differentially methylated regions between colorectal cancer stem cells and colorectal cancers.

CpG dinucleotides. Moreover, we saw a weak increase in surface level of the metastatic marker CD44v6 [30] in Group 2 relative to Group 1.

The MSI-H CRCs rarely metastasize [60]. In this respect MSI-H status is mutually exclusive with CIN, which has been associated with metastatic potential [39]. Here, we identified a CR-CSC DNA methylation signature associated to CIN. However, *in vivo* experiments did not confirm greater aggressiveness of the CIN relative to the MSI CR-CSCs, since no differences in tumorigenic potential were identified between the two groups. This

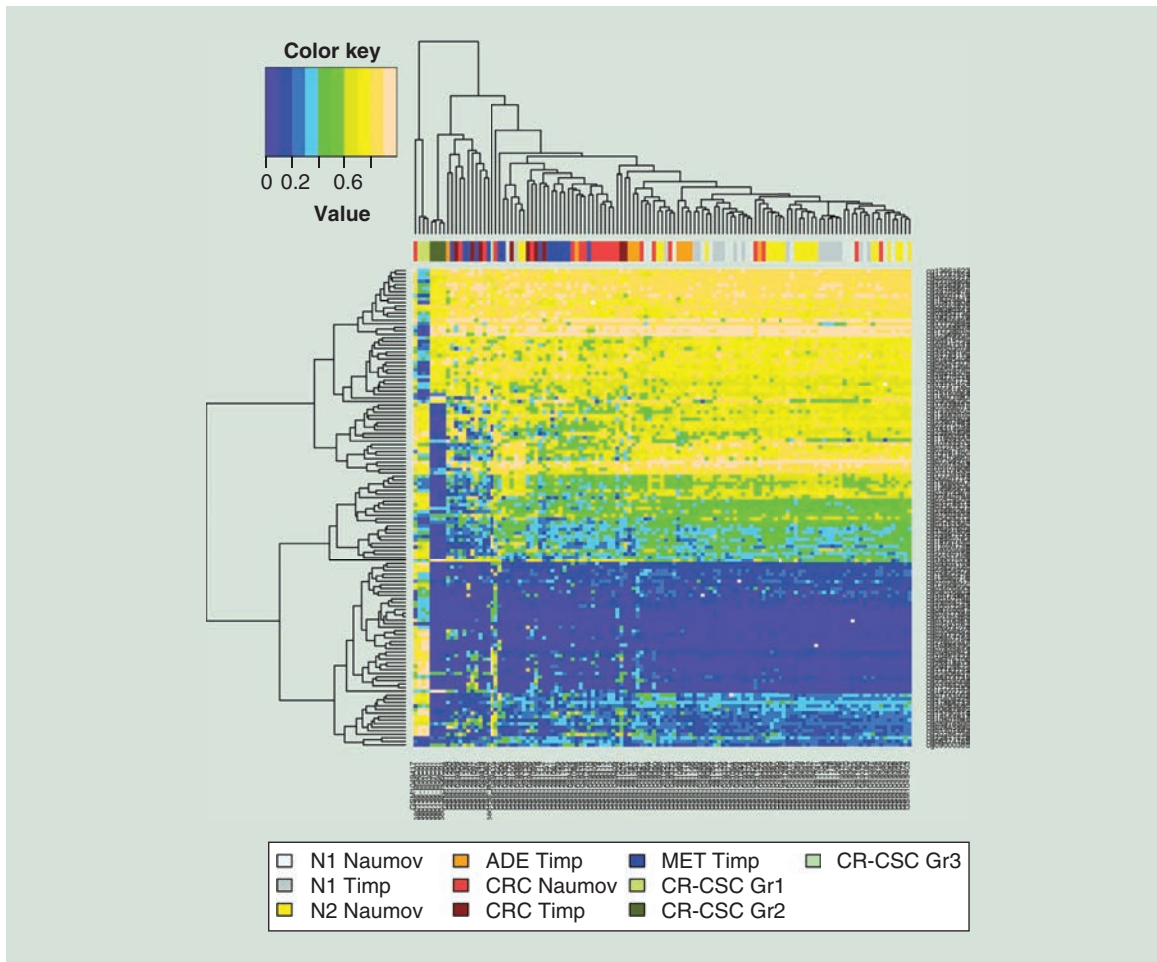


Figure 7. Clustering of GSE42752 and GSE53051 samples on the basis of differentially methylated probes between Gr1 and Gr2 colorectal cancer stem cell lines.

Hierarchical clustering of CR-CSC lines, normal mucosa, adenoma, primary and metastatic CRCs samples included in GSE42752 and in GSE53051 datasets, based on DNA methylation values of the differentially methylated probes that distinguish Gr1 and Gr2 CR-CSC lines and that are included in the Infinium 450 k design (135 probes).

ADE: Adenoma; CRC: Colorectal cancer; CR-CSC: Colorectal cancer stem cell; MET: Metastatic.

result could reflect the fact that increased immunogenic potential [59,61–63], a hallmark of MSI-H CRCs, cannot be evaluated in *in vivo* models based on immune-deficient mice. Furthermore, it is worth to note that the most significant DMR between Gr1 and Gr2 fits with the DNA regulative elements of the *MERTK* gene, methylated in the MSI-H and demethylated in the MSI-L/MSS CR-CSCs, respectively. *MERTK* is a tyrosine kinase receptor that possesses several roles in cellular biology, including immune suppression through PDL-1 expression [64]. Thus, our data suggest a possible role of epigenetics in immune escape, independently of neo-antigen expression. Another significant difference between MSI-H and MSS CRCs is the lack of benefit of adjuvant 5-FU-based chemotherapy in stage II colon MSI-H cancer patients [65]. We did not find DMRs in genes associated with 5-FU resistance [66,67], but we registered *miR-193a* hypermethylation in Gr1 *versus* Gr2 (Supplementary Table 4), accordingly with its methylation and involvement in 5-FU resistance in hepatocellular carcinoma cells [68]. However, this does not exclude that our DMR regions, which are mostly shore, shelves and open sea, could be involved in tumor drug resistance.

One of the most relevant epigenetic differences between the MSI-H (Gr1) and MSS/MSI-L (Gr2) subsets is that most of the significant DMPs are localized in shelf, shore and open sea CpG positions. This is evident in the light of the Gr1 specific-peak around 0.4 of β values, composed mostly of shelf, shore and open sea CpGs. Due to the homogeneity of our cell lines, we assume that this value identifies CpGs that are monoallelically methylated,

although we cannot exclude a specific CpGs hemimethylated pattern in CR-CSCs [69]. In the view of mono-allelic methylation, since Irizarry *et al.* show that most of the cancer- and tissue-specific CpG methylations occurs at shore positions and not at CpG islands of gene promoter [43], we acknowledged the possibility that the MSS/MSI-L (Gr2) could have lost its tissue-specific epigenetic footprint when compared with the MSI group (Gr1) and/or gained other methylation hallmarks.

Other studies correlated epigenetics to aneuploidy, but whether CIN could be caused [70,71] from or be the cause [26,72] of aberrant DNA methylation is still unresolved. Our specific epigenetic signatures separating CR-CSCs with or without CIN could provide insights into this unsolved issue by indicating new genomic regulative elements for the chromosomal stability.

Next, we compared the methylation profiles of the CR-CSCs with those of non-neoplastic mucosa, adenoma, primary and metastatic CRC and found a progression in methylation alterations from normal mucosa to CR-CSCs. The overall hypomethylation of the CR-CSCs does not reflect the methylation status of the different genomic compartments. In the CR-CSCs, CGIs were essentially hypermethylated, whereas shelves and shores were hypomethylated. However, a more specific analysis identified demethylation of several gene promoters controlling cell stemness. This suggests that demethylation of these genes is not necessary for cellular differentiation but relevant/required for possible reprogramming of differentiated cells to a stem status. Therefore, as already stated for intestinal stem cell differentiation, gene regulation by DNA methylation might play a role in the differentiation steps of colorectal tumorigenesis [22,73]. However, given the molecular complexity of gDNA methylation, it is unclear how this could occur. For example, an unexpected molecular mechanism was recently described by Roulois *et al.*, who demonstrated that the demethylating agent 5-aza-2-deoxycytidine targets CRC-initiating cells by inducing viral mimicry through activation of endogenous double-strand RNAs [74].

Conclusion

In conclusion, our findings provide further characterization of CR-CSCs and could help in the identification of novel targets to be used for the eradication of resistant stem-like cancer cells in preclinical and clinical settings. Moreover, the genetic and epigenetic homogeneity of the CR-CSCs derived from both primary and xenograft tumors suggests that these models could be used to test the effectiveness of therapies against the CSC compartment *in vitro* and *in vivo*.

Future perspective

CR-CSCs are implicated in tumor development, metastasis and therapy resistance; however, little is known about their epigenetic features. The results presented in this manuscript pave the way for further researches that will functionally characterize the involvement of CR-CSC DNA methylation in the regulation of cellular stemness behavior and of metastatic potential, taking into account the differences that we have described between CR-CSCs with or without MSI. It will be challenging to understand if these epigenetics alterations are affecting, consequence or both of the CR-CSC phenotype, with particular regard to those CpGs that do not reside within CpG islands. Finally, future research should address the prognostic potential of the methylation profiles of CR-CSC lines, especially when purified from primary tumors of early-stages CRC patients.

Supplementary data

To view the supplementary data that accompany this paper please visit the journal website at: www.futuremedicine.com/doi/full/10.2217/epi-2018-0158

Acknowledgment

The authors thank T Terranova for editing our work.

Author contributions

A Veronese and R Visone designed the study. G Stassi provided CR-CSCs from CRC patients. MG Bacalini performed all the methylation bioinformatic analysis with contribution from R Visone, M Ferracin and A Veronese. S Di Franco did *in vivo* experiments and cellular characterization with contribution from A Nicotra and ML Colorito. E Saccenti, C Bassi, E Scavo and S Pagotto performed genetic characterization with contributions from N Laprovitera, S Di Franco and M Grzes. D Licastro did the genome-wide methylation assay. A Veronese, R Visone, M Ferracin, MG Bacalini, P Garagnani, S Di Franco, G Stassi, N Valeri, R Mariani-Costantini,

V De Laurenzi and M Negrini analyzed, interpreted and discussed the results. A Veronese, R Visone, MG Bacalini and S Di Franco wrote the manuscript. All the authors critically reviewed and revised the paper.

Financial & competing interests disclosure

This work was financed by Italian Ministry of Health funds (Ricerca Finalizzata GR-2011-02350699) to AV, by grants from Associazione Italiana per la Ricerca sul Cancro (AIRC) to GS (IG 16746 and 5 × 1000 9979), and partially, by a generous private donation from Anna Bin in memory of Sergio Bin. SDF is a fellowship recipient of Umberto Veronesi foundation. The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed. The authors declare no conflict of interests.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

CRC specimens were obtained from patients undergoing CRC resection, in accordance with the ethical standards of the institutional committee of the University of Palermo. Animal work was undertaken in accordance with Italian Ministry of Health authorization no. 154/2017-PR, animals were sacrificed accordingly to Directive 2010/63/EU guidelines (D.lgs 26/2016).

Open access

This work is licensed under the Creative Commons Attribution 4.0 License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

Summary points

- Accumulating evidence support the central role of cancer stem cells in colorectal cancer (CRC) initiation and progression.
- The identification of a reliable signature characterizing colorectal cancer stem cells (CR-CSCs) with different tumorigenic/metastatic capacities is crucial to design innovative therapeutic regimens.
- We genetically and epigenetically characterized eight CR-CSC lines isolated from primary CRC tissues.
- The analysis of nucleotide and copy number variations, microsatellite instability (MSI) and expression levels of genes related to cancer stemness concordantly distinguished the CR-CSCs lines in three groups.
- Unsupervised clustering of DNA methylation profiles reproduced the classification of the CR-CSC lines in three groups.
- CR-CSCs with high MSI exhibited a higher proportion of probes occurring in shores, shelves and open sea regions displaying intermediate DNA methylation levels.
- These distinctive DNA methylation profiles were maintained after *in vivo* passages in immuno-deficient mouse models.
- We defined an epigenetic signature distinguishing CR-CSCs with high and low MSI.
- The identification of these specific methylation signatures of CR-CSCs could have important implications in the clinical setting for the management of CRC patients.

References

Papers of special note have been highlighted as: ● of interest; ●● of considerable interest

1. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. *CA Cancer J. Clin.* 65(2), 87–108 (2015).
2. SEER Cancer Stat Facts: Colon and Rectum Cancer. National Cancer Institute. (2017) <https://seer.cancer.gov/statfacts/html/colorect.html>
3. Siegel R, Desantis C, Jemal A. Colorectal cancer statistics, 2014. *CA Cancer J. Clin.* 64(2), 104–117 (2014).
4. Fiori ME, Villanova L, De Maria R. Cancer stem cells: at the forefront of personalized medicine and immunotherapy. *Curr. Opin. Pharmacol.* 35, 1–11 (2017).
5. Barker N, Ridgway RA, Van Es JH *et al.* Crypt stem cells as the cells-of-origin of intestinal cancer. *Nature* 457(7229), 608–611 (2009).
6. Zeuner A, Todaro M, Stassi G, De Maria R. Colorectal cancer stem cells: from the crypt to the clinic. *Cell Stem Cell* 15(6), 692–705 (2014).
7. Sangiorgi E, Capecchi MR. Bmi1 is expressed *in vivo* in intestinal stem cells. *Nat. Genet.* 40(7), 915–920 (2008).
8. Barker N, Van Es JH, Kuipers J *et al.* Identification of stem cells in small intestine and colon by marker gene Lgr5. *Nature* 449(7165), 1003–1007 (2007).

9. Ricci-Vitiani L, Lombardi DG, Pilozzi E *et al.* Identification and expansion of human colon-cancer-initiating cells. *Nature* 445(7123), 111–115 (2007).
10. Jordan CT, Guzman ML, Noble M. Cancer stem cells. *N. Engl. J. Med.* 355(12), 1253–1261 (2006).
11. Meacham CE, Morrison SJ. Tumour heterogeneity and cancer cell plasticity. *Nature* 501(7467), 328–337 (2013).
12. Easwaran H, Tsai HC, Baylin SB. Cancer epigenetics: tumor heterogeneity, plasticity of stem-like states, and drug resistance. *Mol. Cell* 54(5), 716–727 (2014).
13. Gehart H, Clevers H. Tales from the crypt: new insights into intestinal stem cells. *Nat. Rev. Gastroenterol. Hepatol.* 16(1) 19–34 (2019)
- **A review article describing the new insight in the control of homeostasis and regeneration of the intestinal epithelium.**
14. Douville J, Beaulieu R, Balicki D. ALDH1 as a functional marker of cancer stem and progenitor cells. *Stem Cells Dev.* 18(1), 17–25 (2009).
15. Medema JP. Cancer stem cells: the challenges ahead. *Nat Cell Biol.* 15(4), 338–344 (2013).
16. Soteriou D, Fuchs Y. A matter of life and death: stem cell survival in tissue regeneration and tumour formation. *Nat. Rev. Cancer* 18 187–201 (2018).
17. Botchkina G. Colon cancer stem cells – from basic to clinical application. *Cancer Lett.* 338(1), 127–140 (2013).
18. Huang EH, Wicha MS. Colon cancer stem cells: implications for prevention and therapy. *Trends Mol. Med.* 14(11), 503–509 (2008).
19. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell* 14(3), 275–291 (2014).
20. Esteller M, Herman JG. Cancer as an epigenetic disease: DNA methylation and chromatin alterations in human tumours. *J. Pathol.* 196(1), 1–7 (2002).
21. Song L, Li Y. The role of stem cell DNA methylation in colorectal carcinogenesis. *Stem Cell Rev.* 12(5), 573–583 (2016).
22. Kaaij LT, Van De Wetering M, Fang F *et al.* DNA methylation dynamics during intestinal stem cell differentiation reveals enhancers driving gene expression in the villus. *Genome Biol.* 14(5), R50 (2013).
23. Toh TB, Lim JJ, Chow EK. Epigenetics in cancer stem cells. *Mol. Cancer* 16(1), 29 (2017).
24. De Sousa EMF, Colak S, Buikhuisen J *et al.* Methylation of cancer-stem-cell-associated Wnt target genes predicts poor prognosis in colorectal cancer patients. *Cell Stem Cell* 9(5), 476–485 (2011).
25. Timp W, Bravo HC, McDonald OG *et al.* Large hypomethylated blocks as a universal defining epigenetic alteration in human solid tumors. *Genome Med.* 6(8), 61 (2014).
26. Rodriguez J, Frigola J, Vendrell E *et al.* Chromosomal instability correlates with genome-wide DNA demethylation in human primary colorectal cancers. *Cancer Res.* 66(17), 8462–9468 (2006).
27. Prasetyanti PR, Medema JP. Intra-tumor heterogeneity from a cancer stem cell perspective. *Mol. Cancer* 16(1), 41 (2017).
28. Lombardo Y, Scopelliti A, Cammareri P *et al.* Bone morphogenetic protein 4 induces differentiation of colorectal cancer stem cells and increases their response to chemotherapy in mice. *Gastroenterology* 140(1), 297–309 (2011).
29. Todaro M, Alea MP, Di Stefano AB *et al.* Colon cancer stem cells dictate tumor growth and resist cell death by production of interleukin-4. *Cell Stem Cell* 1(4), 389–402 (2007).
30. Todaro M, Gaggianesi M, Catalano V *et al.* CD44v6 is a marker of constitutive and reprogrammed cancer stem cells driving colon cancer metastasis. *Cell Stem Cell* 14(3), 342–356 (2014).
31. Assenov Y, Muller F, Lutsik P, Walter J, Lengauer T, Bock C. Comprehensive analysis of DNA methylation data with RnBeads. *Nat. Methods* 11(11), 1138–1140 (2014).
32. Bacalini MG, Boattini A, Gentilini D *et al.* A meta-analysis on age-associated changes in blood DNA methylation: results from an original analysis pipeline for Infinium 450 k data. *Aging* 7(2), 97–109 (2015).
33. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* 10(10), 1556–1566 (2015).
34. Skidmore ZL, Wagner AH, Lesurf R *et al.* GenVisR: genomic visualizations in R. *Bioinformatics* 32(19), 3012–3014 (2016).
35. Fumagalli A, Drost J, Suijkerbuijk SJ *et al.* Genetic dissection of colorectal cancer progression by orthotopic transplantation of engineered cancer organoids. *Proc. Natl Acad. Sci. USA* 114(12), E2357–E2364 (2017).
36. Drost J, Van Jaarsveld RH, Ponsioen B *et al.* Sequential cancer mutations in cultured human intestinal stem cells. *Nature* 521(7550), 43–47 (2015).
37. Picelli S, Zajac P, Zhou XL *et al.* Common variants in human CRC genes as low-risk alleles. *Eur. J. Cancer* 46(6), 1041–1048 (2010).
38. Campo C, Kohler A, Figlioli G *et al.* Inherited variants in genes somatically mutated in thyroid cancer. *PLoS ONE* 12(4), e0174995 (2017).
39. Bakhroum SF, Ngo B, Laughney AM *et al.* Chromosomal instability drives metastasis through a cytosolic DNA response. *Nature* 553(7689), 467–472 (2018).

- **A research article describing how chromosomal instability drive metastatic potential through the chronic activation of the cytosolic DNA sensing pathway.**
- 40. Jamal-Hanjani M, Wilson GA, Mcgranahan N *et al.* Tracking the evolution of non-small-cell lung cancer. *N. Engl. J. Med.* 376(22), 2109–2121 (2017).
- 41. Turajlic S, Swanton C. Metastasis as an evolutionary process. *Science* 352(6282), 169–175 (2016).
- 42. Fortin JP, Labbe A, Lemire M *et al.* Functional normalization of 450 k methylation array data improves replication in large cancer studies. *Genome Biol.* 15(12), 503 (2014).
- 43. Irizarry RA, Ladd-Acosta C, Wen B *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat. Genet.* 41(2), 178–186 (2009).
- **A research article showing that most methylation alterations in colon cancer occur in specific-tissue CpG island shores.**
- 44. Sandoval J, Heyn H, Moran S *et al.* Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. *Epigenetics* 6(6), 692–702 (2011).
- 45. Cummings CT, Deryckere D, Earp HS, Graham DK. Molecular pathways: MERTK signaling in cancer. *Clin. Cancer Res.* 19(19), 5275–5280 (2013).
- 46. Liang XH, Li LL, Wu GG *et al.* Upregulation of CPE promotes cell proliferation and tumorigenicity in colorectal cancer. *BMC Cancer* 13, 412 (2013).
- 47. Lee TK, Murthy SR, Cawley NX *et al.* An N-terminal truncated carboxypeptidase E splice isoform induces tumor growth and is a biomarker for predicting future metastasis in human cancers. *J. Clin. Invest.* 121(3), 880–892 (2011).
- 48. Vinson KE, George DC, Fender AW, Bertrand FE, Sigounas G. The Notch pathway in colorectal cancer. *Int. J. Cancer* 138(8), 1835–1842 (2016).
- 49. Bullock MD, Bruce A, Sreekumar R *et al.* FOXO3 expression during colorectal cancer progression: biomarker potential reflects a tumour suppressor role. *Br. J. Cancer* 109(2), 387–394 (2013).
- 50. Naumov VA, Generozov EV, Zaharjevskaya NB *et al.* Genome-scale analysis of DNA methylation in colorectal cancer using Infinium HumanMethylation450 BeadChips. *Epigenetics* 8(9), 921–934 (2013).
- 51. Lujambio A, Ropero S, Ballestar E *et al.* Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer Res.* 67(4), 1424–1429 (2007).
- 52. Bandres E, Agirre X, Bitarte N *et al.* Epigenetic regulation of microRNA expression in colorectal cancer. *Int. J. Cancer* 125(11), 2737–2743 (2009).
- 53. Veronese A, Lupini L, Consiglio J *et al.* Oncogenic role of miR-483-3p at the IGF2/483 locus. *Cancer Res.* 70(8), 3140–3149 (2010).
- 54. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487(7407), 330–337 (2012).
- 55. Cui H, Cruz-Correa M, Giardiello FM *et al.* Loss of IGF2 imprinting: a potential marker of colorectal cancer risk. *Science* 299(5613), 1753–1755 (2003).
- 56. Chaffer CL, Brueckmann I, Scheel C *et al.* Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. *Proc. Natl Acad. Sci. USA* 108(19), 7950–7955 (2011).
- 57. Bjerkvig R, Tysnes BB, Aboody KS, Najbauer J, Terzis AJ. Opinion: the origin of the cancer stem cell: current controversies and new insights. *Nat. Rev. Cancer* 5(11), 899–904 (2005).
- 58. Plaks V, Kong N, Werb Z. The cancer stem cell niche: how essential is the niche in regulating stemness of tumor cells? *Cell Stem Cell* 16(3), 225–238 (2015).
- 59. Germano G, Lamba S, Rospo G *et al.* Inactivation of DNA repair triggers neoantigen generation and impairs tumour growth. *Nature* 552(7683), 116–120 (2017).
- **A research article demonstrating how mismatch repair deficiency drives the generation of neoantigens with subsequent increase of cancer cell immunogenicity.**
- 60. Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet* 383(9927), 1490–1502 (2014).
- 61. Chang K, Taggart MW, Reyes-Uribe L *et al.* Immune profiling of premalignant lesions in patients with lynch syndrome. *JAMA Oncol.* 4(8), 1085–1092 (2018) (Epub ahead of print).
- 62. Phillips SM, Banerjee A, Feakins R, Li SR, Bustin SA, Dorudi S. Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic. *Br. J. Surg.* 91(4), 469–475 (2004).
- 63. Le Gouvello S, Bastuji-Garin S, Aloulou N *et al.* High prevalence of Foxp3 and IL17 in MMR-proficient colorectal carcinomas. *Gut* 57(6), 772–779 (2008).
- 64. Nguyen KQ, Tsou WI, Calarese DA *et al.* Overexpression of MERTK receptor tyrosine kinase in epithelial cancer cells drives efferocytosis in a gain-of-function capacity. *J. Biol. Chem.* 289(37), 25737–25749 (2014).

65. Kawakami H, Zaanan A, Sinicrope FA. Microsatellite instability testing and its role in the management of colorectal cancer. *Curr. Treat Options Oncol.* 16(7), 30 (2015).
66. Zhang N, Yin Y, Xu SJ, Chen WS. 5-Fluorouracil: mechanisms of resistance and reversal strategies. *Molecules* 13(8), 1551–1569 (2008).
67. Kunicka T, Prochazka P, Krus I *et al.* Molecular profile of 5-fluorouracil pathway genes in colorectal carcinoma. *BMC Cancer* 16(1), 795 (2016).
68. Ma K, He Y, Zhang H *et al.* DNA methylation-regulated miR-193a-3p dictates resistance of hepatocellular carcinoma to 5-fluorouracil via repression of SRSF2 expression. *J. Biol. Chem.* 287(8), 5639–5649 (2012).
69. Shao C, Lacey M, Dubeau L, Ehrlich M. Hemimethylation footprints of DNA demethylation in cancer. *Epigenetics* 4(3), 165–175 (2009).
70. Chango A, Abdennebi-Najar L, Tessier F *et al.* Quantitative methylation-sensitive arbitrarily primed PCR method to determine differential genomic DNA methylation in Down Syndrome. *Biochem. Biophys. Res. Commun.* 349(2), 492–496 (2006).
71. Davidsson J, Veerla S, Johansson B. Constitutional trisomy 8 mosaicism as a model for epigenetic studies of aneuploidy. *Epigenetics Chromatin* 6(1), 18 (2013).
72. Herrera LA, Prada D, Andonegui MA, Duenas-Gonzalez A. The epigenetic origin of aneuploidy. *Curr. Genomics.* 9(1), 43–50 (2008).
73. Sheaffer KL, Kim R, Aoki R *et al.* DNA methylation is required for the control of stem cell differentiation in the small intestine. *Genes Dev.* 28(6), 652–664 (2014).
74. Roulois D, Loo Yau H, Singhanian R *et al.* DNA-demethylating agents target colorectal cancer cells by inducing viral mimicry by endogenous transcripts. *Cell* 162(5), 961–973 (2015).