Research Article

Epigenomics

Distinct cytosine modification profiles define epithelial-to-mesenchymal cell-state transitions

Min Kyung Lee[1] [iD], Meredith S Brown[2], Owen M Wilkins[3] [iD], Diwakar R Pattabiraman[‡,2] [iD]
& Brock C Christensen*,[‡,1,2,4] [iD]
[1]Department of Epidemiology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA
[2]Department of Molecular & Systems Biology, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA
[3]Norris Cotton Cancer Center, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, USA
[4]Department of Community & Family Medicine, Geisel School of Medicine at Dartmouth, Lebanon, NH 03756, USA
*Author for correspondence: Brock.C.Christensen@dartmouth.edu
[‡]Equal contributions as senior authors

Background: Epithelial-to-mesenchymal transition (EMT) is an early step in the invasion-metastasis cascade, involving progression through intermediate cell states. Due to challenges with isolating intermediate cell states, genome-wide cytosine modifications that define transition are not completely understood. **Methods:** The authors measured multiple DNA cytosine modification marks and chromatin accessibility across clonal populations residing in specific EMT states. **Results:** Clones exhibiting more intermediate EMT phenotypes demonstrated increased 5-hydroxymethylcytosine and decreased 5-methylcytosine. Open chromatin regions containing increased 5-hydroxymethylcytosine CpG loci were enriched in EMT transcription factor motifs and were associated with Rho GTPases. **Conclusion:** The results indicate the importance of both distinct and shared epigenetic profiles associated with EMT processes that may be targeted to prevent EMT progression.

Epithelial-to-mesenchymal transition (EMT) is an early step in the invasion-metastasis cascade, involving progression through a number of cellular states. It is a process by which epithelial cells lose specific properties such as apical-basal polarity detach from the basement membrane to gain mesenchymal properties such as front-back polarity and motility [1]. Rather than being a binary conversion from an epithelial to a mesenchymal state, the EMT encompasses a step-wise progression to a mesenchymal cell state whereby the cells could display intermediate/hybrid phenotypes of both epithelial and mesenchymal cells [2,3]. As metastasis is responsible for the majority of deaths in cancer patients [4,5], it is critical to understand the molecular underpinnings of EMT.

Cells that reside in an intermediate state display more plasticity than the cells on either ends of the EMT spectrum [3,6–8]. In addition to increased plasticity, intermediate cells have been shown to harbor stem cell characteristics such as self-renewal and increased expression of pluripotent genes [9–11]. Although it is evident that there are intermediate phases when transitioning from epithelial to mesenchymal states [12–14], experimental isolation of these specific states has proven challenging. Consequently, the molecular and functional characteristics and of the intermediate states and their contribution to metastasis are poorly understood.

DNA methylation is a well-studied epigenetic mark, mostly known for its role in regulating gene expression. Methylation of cytosines (5-methylcytosine [5mC]) can occur in the context of CpG dinucleotides and the reaction is catalyzed by DNA methyltransferase enzymes (DNMTs). Ten eleven translocation (TET) enzymes can oxidize methylcytosine to form 5-hydroxymethylcytosine (5hmC), then 5-formylcytosine (5fC) and finally 5-carboxylcytosine (5caC) [15]. Oxidized cytosines can then be deaminated by activation-induced cytidine deaminase

(AID), then undergo thymine DNA glycosylase-mediated base excision repair to an unmethylated cytosine. While around 80% of mammalian CpG dinucleotides are estimated to be methylated [16,17], hydroxymethylation accounts for a relatively modest proportion of overall cytosine modification and varies greatly with tissue type [18,19]. Although 5hmC levels are low in relation to 5mC in human tissues, it is most highly enriched in brain and breast tissues, relative to other tissue types [20]. While a number of studies have shown the importance of DNA methylation in EMT, these studies used traditional bisulfite treatment to measure 5mC, which does not resolve 5hmC [21–26]. 5hmC can be estimated from comparing oxidized-bisulfite treatment to bisulfite-treated DNA [27], as traditional bisulfite treatment does not distinguish 5mC from 5hmC. In comparison with general repression of transcription from 5mC, 5hmC is positively associated with transcriptional activity and gene expression [28,29]. Whether the association is a consequence of passive dilution of 5mC via DNA demethylation or is due to functional actions of 5hmC is yet unclear and is likely context dependent. However, growing evidence suggests that 5hmC contributes directly to gene regulation in several specific contexts, aside from its role in DNA demethylation. At the chromatin level, 5hmC has been shown to increase DNA flexibility and mechanical stability, as well as nucleosome accessibility [30]. Transcription factors and their binding sites have been associated with being colocalized with TET and 5hmC [31–34], which provides a possible 5hmC mechanism of gene expression regulation through transcription factor recruitment [35].

Although decreased global 5hmC is consistently observed in cancer [36–39], few studies have measured cancer-associated 5hmC changes at nucleotide-resolution. 5hmC maintenance has been associated with protecting against CpG island hypermethylation, which commonly occurs in cancer [40–44]. Measures of breast tissue nucleotide-specific 5hmC revealed enrichment within breast-specific enhancers and transcriptionally active chromatin [45]. In estrogen receptor (ER)/progesterone receptor (PR)-negative breast cancer particularly, loss of 5hmC is associated with poor prognosis [39]. As DNA methylation alterations occur early in breast carcinogenesis and are related to prognosis [46,47], a better understanding of 5hmC in breast cancer and EMT is needed.

In concert with DNA methylation, chromatin accessibility regulates transcription and cell reprogramming [48]. Interactions with different nuclear macromolecules such as transcription factors and histone modifications shape the topology of chromatin [48]. Specific chromatin accessibility states have been implicated in regulating EMT. Putative enhancers, defined by promoter-distal H3K27ac and H3K4me1 histone modifications, have been shown to recruit key EMT transcription factors such as NF-κB and AP-1 in epithelial cells in comparison with TGF-β-treated mesenchymal cells [49–51]. In addition, motifs of key EMT transcription factors (AP-1, ETS) were enriched in accessible chromatin regions of TGF-β transformed mammary epithelial cells [52]. Although transcription factors influencing EMT and metastasis-associated chromatin accessibility have been identified [53–56], gaps in the knowledge of chromatin accessibility changes in non-TGF-β-induced EMT cells and cells in EMT intermediate/hybrid states still remain due to challenges in isolating cells in these states. Moreover, a better understanding of the relationship between cytosine modifications and chromatin conformation is needed.

This study provide a nucleotide-resolution genome-scale map of cytosine modifications and chromatin accessibility for phenotypes spanning the EMT spectrum. It addresses gaps in the understanding of epigenomic changes in the intermediate/hybrid states on the EMT spectrum. Using a novel model derived from ER/PR-negative breast cancer cells to study terminal and intermediate EMT states, the authors demonstrate substantial differences in the cytosine modification profiles of cells in intermediate EMT states, particularly increases in 5hmC enriched in key EMT transcription factor motifs. Further, the authors utilize novel, integrative multi-component epigenetic analysis to show cytosine modifications coordinate with chromatin accessibility, especially at promoters to regulate transcription.

Methods
Cell culture
Single cell clones, the isolation and characterization methods of which are detailed in Brown *et al.* [8], were used. To summarize, six single-cell clones were isolated from SUM149PT cells to represent different points of the EMT spectrum. Position on the EMT spectrum was determined by cell morphology, flow cytometry analysis of CD44 and CD104 markers and mRNA expressions of *ZEB1/2*. Graphic representation of each clone's position on the EMT spectrum can be found in Figure 1. Transwell assays to measure migration and invasion were conducted and reported for each clone in Brown *et al.* [8].
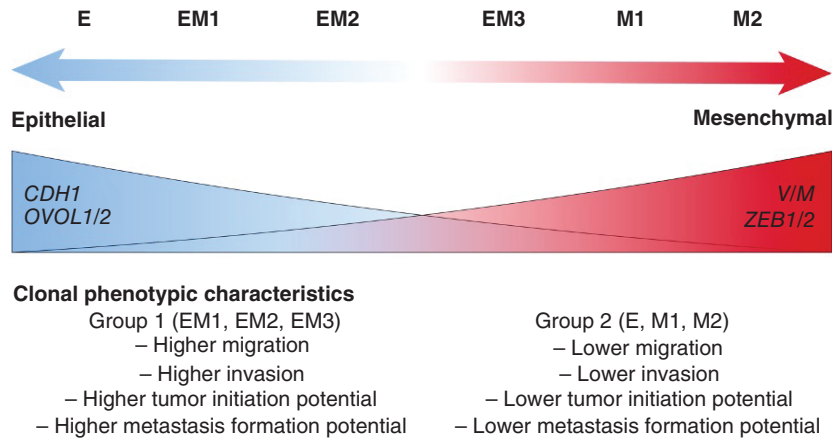
Figure 1.   Summary of characteristics of isolated single cell clones that reside in specific positions of the epithelial-to-mesenchymal transition spectrum. Specific data for each clone that are summarized in this figure are reported in Brown *et al.* [8]. Gene expression of epithelial markers (*CDH1* and *OVOL1/2*) is highest on the most epithelial-like clone and decreases sequentially as clones display more mesenchymal characteristics. Gene expression of mesenchymal markers (*VIM* and *ZEB1/2*) is lowest in the most epithelial-like clone and increases sequentially as clones display more mesenchymal characteristics.

DNA methylation & hydroxymethylation

DNA conversion & methylation/hydroxymethylation profiling

DNA from each clone of similar passage numbers was extracted using DNeasy Blood and Tissue kit (Catalog ID 69504; Qiagen, Hilden, Germany). DNA was quantified with Qubit 3.0 Fluorometer (Life Technologies, CA, USA). ~2 μg of DNA underwent oxidative-bisulfite conversion to measure both 5mC and 5hmC using the TrueMethyl OxBS Module (Catalog ID 0414-32; Nugen, CA, USA). Epigenome-wide DNA methylation profiling was performed using the Infinium MethylationEPIC Bead Chips (Illumina, Inc., CA, USA) at the Norris Cotton Cancer Center Genomics Shared Resource Core.

Quality control & processing

Raw intensity files produced from the MethylationEPIC Bead Chips were preprocessed using the *minfi* R/Bioconductor analysis pipeline (v1.34.0) annotation file version *ilm10b4.hg19* [57,58]. Six hundred ninety-five technical probes and 33,360 SNP-associated probes were excluded. Quality control was performed using *ENmix* R package. A total of 301,580 probes that failed to meet a detection p-value of 0.00005 in >30% of the samples and 5% of the CpGs were excluded. The high number of CpGs that failed to pass quality control may have been due to oxidation further damaging the DNA on top of the bisulfite treatment and signal distributions being distorted from the oxidation measurement as the quality control measures were developed for bisulfite-converted DNA. After these exclusions, 545,515 CpGs remained for analysis. The filtered data were then normalized using *preprocessFunnorm* in *minfi* to remove unwanted technical variation.

Annotations of CpGs such as genomic context or relation to CpG island were provided in the Illumina EPIC B4 manifest and UCSC hg19 reference genome files. "Promoter," "Intergenic," "Intron" and "Exon" genomic contexts were defined by finding overlapping genomic regions of the CpGs and each context using the UCSC hg19 reference genome annotation. "DNase hypersensitive site" context was defined by having a record in the "DNase_Hypersensitive_NAME" in the annotation. "Gene body" transcriptional context was defined by having a "Body" in the *UCSC_RefGene_Group*. Likewise, "3′ UTR" and "5′ UTR" regions were defined by having "UTR3" and "UTR5," respectively, in the *UCSC_RefGene_Group*. Relation to CpG island was defined by "Relation_to_UCSC_CpG_Island" in the Illumina EPIC annotation file. If no record of relation to the CpG island was indicated, the CpG was considered to be in the "Open Sea" region. For analysis testing enrichment of CpGs measured on the Illumina EPIC array to ATAC regions, the GRCh38 annotation file from Zhou *et al.* was used [59].

CpGs annotated to open chromatin regions were defined by their overlap with open chromatin regions from assay for transposase-accessible chromatin with sequencing (ATAC-seq) data. CpGs were determined to be in enhancers if they were located in distal intergenic regions (within 10–15 kbps upstream and downstream of the gene) of the

ATAC-seq consensus peaks. CpGs were determined to be in open promoters if they were located in promoters of the ATAC-seq consensus peaks.

5hmC estimation

5hmC beta values were estimated using the *fitOxBS* function in the *OxyBS* package [60]. Instead of naive subtraction of signals from oxidative-bisulfite-treated probes from bisulfite-only-treated probes, the *OxyBS* package uses maximum likelihood estimation of the signal intensities from the oxidative-bisulfite-treated and bisulfite-treated DNA from the Illumina EPIC array to determine the parameters for unmethylated, hydroxymethylated and methylated CpGs.

Analysis

Principal component analyses were performed using 5hmC and 5mC beta values using the *princomp* function in R. Differential methylation and hydroxymethylated analyses were conducted using *limma* (v3.44.3) and *qvalue* (v2.20.0) R packages in R (v4.0.2) [61,62]. Differentially methylated and hydroxymethylated CpGs were identified by fitting into a linear regression model, testing for differences in beta values CpG-by-CpG in the groups of clones based position on the EMT spectrum (distal vs intermediate). Linear regression models were fit by using *lmFIt* and *eBayes* functions. E, EM1, M2 and P were considered as distal clones. The intermediate group comprised EM2, EM3 and M1 clones. The differentially methylated CpGs were deemed to be significant at the q-value threshold of 0.01.

Differentially hydroxymethylated and methylated CpGs were compared with the 545,515 CpGs used in analyses to test for enrichment at specific genomic contexts using Fisher's exact test. Functional significance of these CpGs was assessed using the Genomic Regions Enrichment of Annotations Tool (GREAT) [63].

Assay for transposase-accessible chromatin with sequencing

ATAC-seq & preprocessing

ATAC-seq for two replicates per clone was performed as described in Buenrostro *et al.* [64]. Similar passage number (+/- 1 passage) of the clones to the DNA methylation and hydroxymethylation measurements were used. The same processing methods and detailed descriptions can be found in Brown *et al.* [8].

Briefly, ATAC-seq data were then processed using the publicly available ENCODE ATAC-seq pipeline (www.encodeproject.org/pipelines/ENCPL792NWO/). Illumina adapter and transposase sequences were trimmed using *Cutadapt* [65] (v1.9.1) with parameters "–minimum-length 5 -e 0.1." Trimmed reads were aligned to hg38 human genome using *Bowtie2* [66] (v2.2.6) in "–local" mode with parameters "-X 2000 -k 2." Duplicate reads were identified and filtered from final alignments using *MarkDuplicates* (Picard Tools [67]). To account for insertion of adapter sequences by the transposase, alignments were converted to tagAlign files and shifted +4 bp and -5 bp on the + and – strands, respectively. *MACS2* [68] (v2.1.1) *callpeak* command with parameters "–shift -75 –extsize 150 –nomodel –keep-dup all –call-summits -p 1.0E-10" were used to call peaks. The peaks were filtered against the ENCODE hg38 blacklist. The Irreproducible Discovery Rate (IDR) method was used to identify a set of reproducible peaks across biological replicates using an IDR threshold of 0.05.

ATAC-seq analysis

Principal component analyses were performed using variance stabilizing transformed ATAC-seq counts using the *princomp* function in R. Low-level regions were filtered out using *filterByExpr* using *edgeR* (v3.30.3) [69]. Open chromatin regions containing dhmCpGs were annotated using *TxDb.Hsapiens.UCSC.hg38.knownGene* R annotation file package and the *annotatePeak* function in *ChIPseeker* (v1.24.0) [70,71]. Enriched biological pathways associated with the differentially accessible regions were identified using the *ReactomePA* (v1.32.0) [72].

The authors tested for overrepresentation of transcription factor (TF) binding site motifs of dhmCpGs containing consensus ATAC peaks compared with all ATAC peaks. The authors scanned these peaks for TF motif occurrences using R package *motifmatchr* [73]. Position frequency matrices for human TF motifs used as input to *motifmatchr* were downloaded using R packages *JASPAR2020* [74] and *TFBSTools* [75]. Overrepresented TF motifs in each peak set were identified through hypergeometric testing using the *phyper* R function, with all peaks identified in that clone used as the background set. TF motifs with a false discovery rate (FDR)-adjusted hypergeometric p-value < 0.05 were deemed to be overrepresented.

RNA sequencing

RNA extraction & preprocessing

RNA was collected using the Qiagen RNeasy plus kit (Catalog ID: 74034; Qiagen, Hilden, Germany) and quantified using a NanoDrop (Thermo Fisher Scientific – ND-2000-US-CAN). The same processing methods and detailed descriptions can be found in Brown *et al.* [8].

To summarize, raw single-end RNA sequencing (RNA-seq) data were trimmed of polyA sequences and low-quality bases using *Cutadapt* (v2.4) [65]. Reads were aligned to human genome hg38 using *STAR* (v 2.7.2b) [76] with parameters "–outSAMattributes NH HI AS NM MD –outFilterMultimapNmax 10 –outFilterMismatchNmax 999 –outFilterMismatchNoverReadLmax 0.04 –alignIntronMin 20 –alignIntronMax 1000000 –alignMatesGapMax 1000000 –alignSJoverhangMin 8 –alignSJDBoverhangMin 1." Quality of alignments was assessed using *CollectRNASeqMetrics* (Picard Tools) [67] and duplicate reads were identified (but retained) with *MarkDuplicates* (Picard Tools). Gene-level abundance estimates were generated using *RSEM* (v1.3.2) [77] using the rsem-calculate-expression command with the parameters "–strandedness reverse –fragment-length-mean 313 –fragment-length-sd 91."

Results

The authors utilized a previously derived model of six single cell clones from SUM149PT, a heterogeneous ER-/PR-inflammatory breast cancer line, that represent cell states present along the EMT spectrum. The EMT state of each clone was determined by cell morphology, flow cytometry for CD44 and CD104 markers and immunofluorescence staining for vimentin/e-cadherin, as well gene expression of canonical EMT markers (*SNAI1*, *ZEB1*, *CDH1*, *VIM* and others), detailed in previous work [8]. More epithelial-like clones had low CD44 and high CD104 expression, while more mesenchymal-like clones had high CD44 and low CD104 expression. Intermediate clones had high CD44 and high CD104 expression. *VIM* and *ZEB1/2* increased in expression along with progressive position on the EMT spectrum, while *CDH1* and *OVOL1/2* decreased in expression (Figure 1, gene expression data in Brown *et al.*) [8]. These clones were ranked as epithelial (E), three distinct intermediates (EM1, EM2 and EM3) and two unique mesenchymal-like clones (M1 and M2) and were compared here with the parental cell line (P). Phenotypically, the intermediate clones (EM1, EM2, EM3) displayed higher migratory and invasive behavior, and higher tumor initiation and metastasis formation potential, compared with the clones on either edges of the EMT spectrum (E, M1, M2) (Figure 1, specific data for each clone reported in Brown *et al.*) [8].

The authors first measured genome-scale cytosine-specific DNA methylation (5mC) and hydroxymethylation (5hmC) levels, using the Illumina EPIC methylation array. As expected [19], a relatively small subset of measured CpGs were hydroxymethylated, with average 5hmC beta values much smaller than that of 5mC across all clones (Figure 2A & B). Average 5hmC beta values and 5mC beta values were negatively correlated, at marginal significance (R = -0.72; p = 0.071), with increased global 5hmC and decreased global 5mC abundance in intermediate clones (EM2, EM3, M1; Figure 2B & C).

To identify which distal clone (E or M2) each clone along the EMT spectrum were similar to, the authors compared each 5hmC profile of EM1, EM2, EM3 and M1 with the 5hmC and 5mC profile of clones on the extreme ends of the EMT spectrum (E and M2). 5hmC profiles of EM1, EM2 and EM3 had more 5hmC profiles similar to the 5hmC profile of E, in which the number of CpGs with little to no change was higher in E compared with in M2 (Figure 3A). The 5hmC profile of M2 had a very similar number of CpGs with little to no change in comparison with 5hmC profiles of E and M2. The 5hmC profiles of EM1, EM2, EM3 and M1 were all more similar to the 5mC profile of E than to the 5mC profile of M2 (Figure 3B). The results suggest that EM1, EM2, EM3 and M1 clones likely were derived from the most epithelial clone and provide models of states on the EMT.

Genome-wide DNA cytosine modification profiles in EMT clones

To determine associations between EMT phenotypes (migratory and invasive behavior) of clones and DNA cytosine modifications, first, the authors analyzed correlations between global 5mC and 5hmC beta values with average migration and invasion levels that had previously been determined in Brown *et al.* [8]. There were no statistically significant correlations between global DNA cytosine modification levels and migration and invasion levels (Supplementary Figure 1A–D).

In addition to correlations between global levels of DNA cytosine modifications, the authors conducted an epigenome-wide association study to identify specific CpGs that are associated with high migration and invasive properties. Migration and invasion assays from Brown *et al.* indicated that clones (EM1, EM2, EM3, P) with greater than the median migration and invasion levels were determined to have high migratory and invasive properties
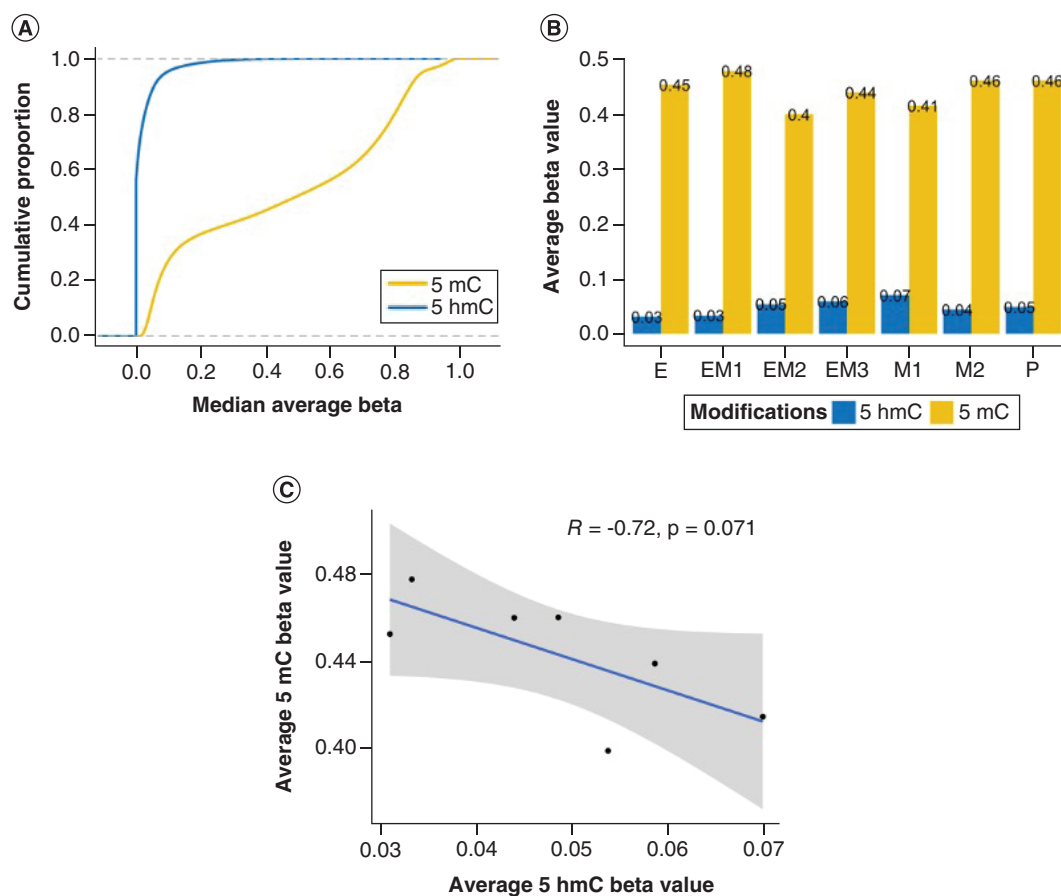
Figure 2.   5-hydroxymethylcytosine (5hmC) and 5-methylcytosine (5mC) levels in the epithelial-to-mesenchymal transition clonal cell line model. (A) Cumulative density of median 5hmC and 5mC beta values. **(B)** Average 5hmC and 5mC beta values per clone. **(C)** Pearson correlation of 5hmC beta values and 5mC beta values.

(Supplementary Figure 2A) [8]. While it was surprising that the EM1, EM2 and EM3 clones were more migratory and invasive than the mesenchymal clones, previously established traits of mesenchymal cells did not discern between mesenchymal and intermediate states when determining migratory and invasive behavior. It is possible that because these cell states were not distinguished, the migratory and invasive behavior of the intermediate clones influenced the notion that mesenchymal cells were more likely to be migratory [78]. Only one differentially hydroxymethylated CpG was determined to be associated with high migratory and invasive cellular phenotypes under the FDR <0.1 significance level (Supplementary Figure 2B). There were no differentially methylated CpGs associated with high migratory and invasive cellular phenotypes under the FDR <0.1 significance level (Supplementary Figure 2C).

To compare genome-scale similarity of DNA methylation profiles among all clones, the authors compared the 5hmC and 5mC beta values using principal component analysis (PCA). PCA results indicated that 5hmC and 5mC beta values clustered into two distinct groups: one group of E, EM1, M2 and another group with EM2, EM3, M1 (Figure 4A & B). In downstream analyses for this study, EM2, EM3 and M1 were defined as intermediate clones and E, EM1, M2 and P were defined as distal clones. These two groups were slightly different from groupings identified by the clones' cellular phenotypes summarized in Figure 1 and in the original development of the model. Furthermore, the groups identified by genome-scale 5mC and 5hmC beta values were different than PCA clustering from chromatin accessibility profiles from ATAC-seq (Supplementary Figure 3A) and gene expression profiles from RNA-seq (Supplementary Figure 3B). Non-negative matrix factorization hierarchical clustering with 5mC, 5hmC and chromatin accessibility profiles revealed similar clustering results from RNA-seq and ATAC-seq (Supplementary Figure 3C). Following the PCA results, distinct grouping of clones into intermediate and distal was supported by unsupervised hierarchical clustering of the top 5% most variable CpGs (27,276 CpGs), which were chosen based on the distribution of variances across CpGs (Supplementary Figure 4A & B). Unsupervised clustering
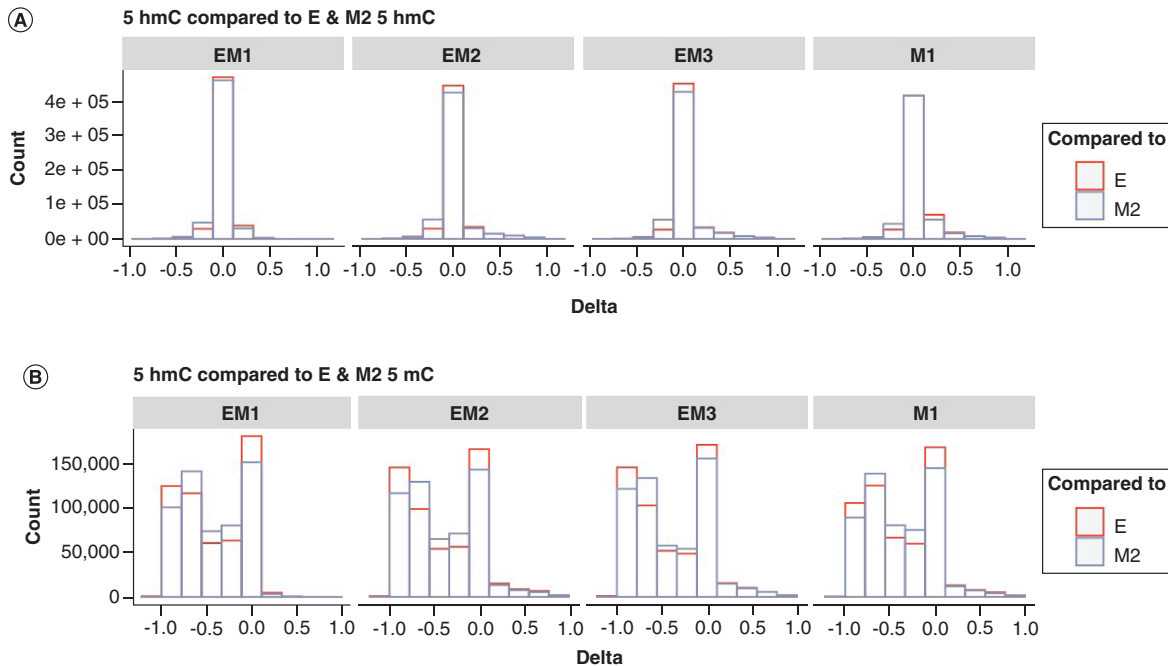
Figure 3.   Clones in between the most extremes of the epithelial-to-mesenchymal transition spectrum are more similar to the most epithelial clone. (A) Delta change in 5-hydroxymethylcytosine (5hmC) in EM1, EM2, EM3 and M1 compared with 5hmC in E and M2. Comparison with E indicated with red boxes. Comparison with M2 indicated with blue boxes. **(B)** Delta change in 5hmC in EM1, EM2, EM3 and M1 compared with 5-methylcytosine (5mC) in E and M2. Comparison with E indicated with red boxes. Comparison with M2 indicated with blue boxes.

identified highly distinct intermediate and distal clone clusters (Figure 4C & D) and highlighted the greater relative abundance of 5hmC in intermediate clones compared with distal clones (Figure 2B) at the CpG-specific level.

The authors next used a candidate gene approach to investigate if EMT-associated 5hmC and 5mC loci distinguished intermediate from distal clones. They performed unsupervised clustering on beta values of 439 CpGs annotated to epithelial genes (*CDH1, CLDN1, EPCAM, ITGAB4, KRT8* and *OCLN)*, mesenchymal genes (*CDH2, FN1, ITGB1, MMP19, MMP2* and *VIM*) and EMT-related TFs (*SNAI1, SNAI2, TWIST1, ZEB1* and *ZEB2*). Intermediate clones clustered separately from distal clones for both 5mC- and 5hmC-associated genes, and a subset of CpGs annotated predominantly to epithelial genes (*OCLN, CDH1, KRT8* and *EPCAM*) had high 5hmC among intermediate clones in cluster #4 (Figure 5A & B), many of which tracked to promoter regions (Figure 5C). Together, it suggests a potential role of 5hmC in regulating epithelial genes during the EMT process.

To determine if overall 5hmC and 5mC abundance was related to expression of cytosine-modifying enzymes (DNMTs and TETs), the authors leveraged RNA-seq to test the correlation of average methylation and gene expression levels. Only *TET1* gene expression was significantly positively correlated with global average 5hmC beta values ($R = 0.86$; $p = 0.024$), and none were correlated with 5mC (Supplementary Figure 5A & B). 5hmC and 5mC beta values of DNMT and TET CpGs with unsupervised clustering did not identify extensive variation in cytosine states at cytosine modification enzyme genes (Supplementary Figure 5C). However, a small subset of CpGs (n CpGs = 18 of 241 total), located within TETs (*TET1*: 33%; *TET2*: 28%; *TET3*: 17%), exhibited higher 5hmC in intermediate clones (Supplementary Figure 5C).

Together, these findings suggest there are variable patterns of genome-wide 5hmC and 5mC based on clonal EMT status, not with clonal phenotypes.

Differential methylation & hydroxymethylation in intermediate clones

Next, the authors conducted an epigenome-wide association study (EWAS) comparing cytosine modifications at the nucleotide level to identify differential cytosine modifications between intermediate and distal clones. Overall, they identified 17,862 significantly differentially hydroxymethylated CpGs (dhmCpGs) (FDR < 0.01), between distal and intermediate clones, almost all of which had increased in 5hmC in the intermediate clones (Figure 6A
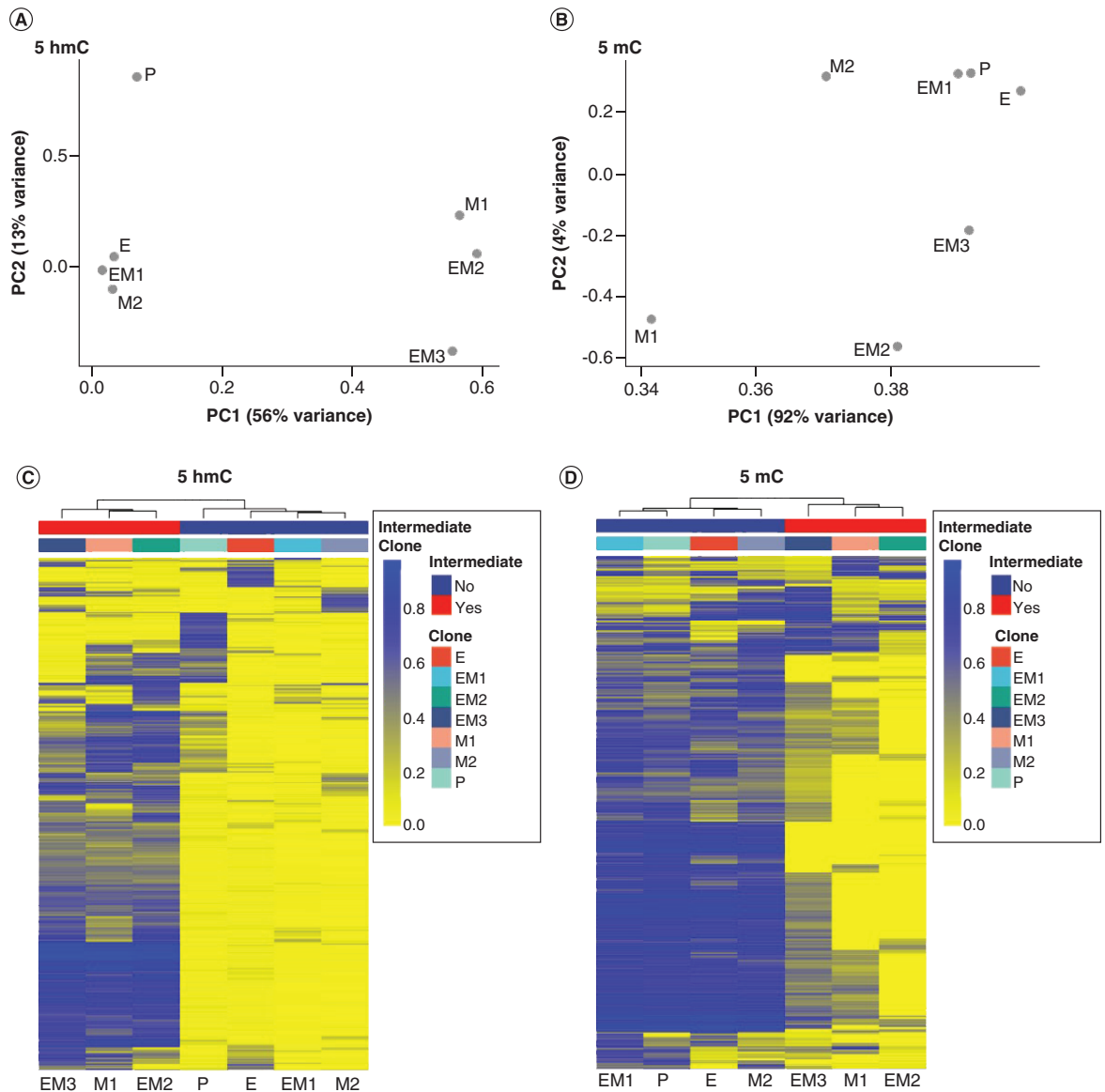
Figure 4.   Distal and intermediate clones have distinct methylation and hydroxymethylation profiles. Results from principal component analysis of **(A)** 5hmC and **(B)** 5mC beta values. Heatmap of unsupervised clustering of the top 5% (27,276 CpGs) most variable **(C)** 5hmC and **(D)** 5mC CpGs. Color scale ranges from yellow (low beta value) to blue (high beta value). Horizontal tracking bars indicate clones and position on the EMT spectrum.

& Supplementary Table 1), including EMT-associated genes such as *SNAI1* and *TWIST1*. There were 7903 significantly differentially methylated CpGs (dmCpGs) (FDR <0.01), most of which had decreased in 5mC in intermediate clones (Figure 6B & Supplementary Table 2), including EMT-associated cell type markers *CDH1* and *MMP19*. For further downstream analyses, dhmCpGs were subset for only CpGs increasing in 5hmC. dmCpGs were subset for only CpGs decreasing in 5mC. Among CpGs with increased 5hmC and decreased 5mC, only 33 CpGs overlapped (Figure 6C). Expanding to the gene level, 1365 genes had both dhmCpGs and dmCpGs among intermediate clones (Figure 6D). Genomic contexts with enrichment of dhmCpGs were generally depleted among dmCpGs (Figure 6E & Supplementary Table 3). While dhmCpGs were enriched in regulatory regions (open chromatin regions, enhancers, 5'UTR, promoters, TSS1500, TSS200) and in the first exon, dmCpGs were enriched within exons and introns, suggesting different cytosine modifications act on different genomic regions in regulating the EMT process. The results suggest that while some differential cytosine modification mark may act
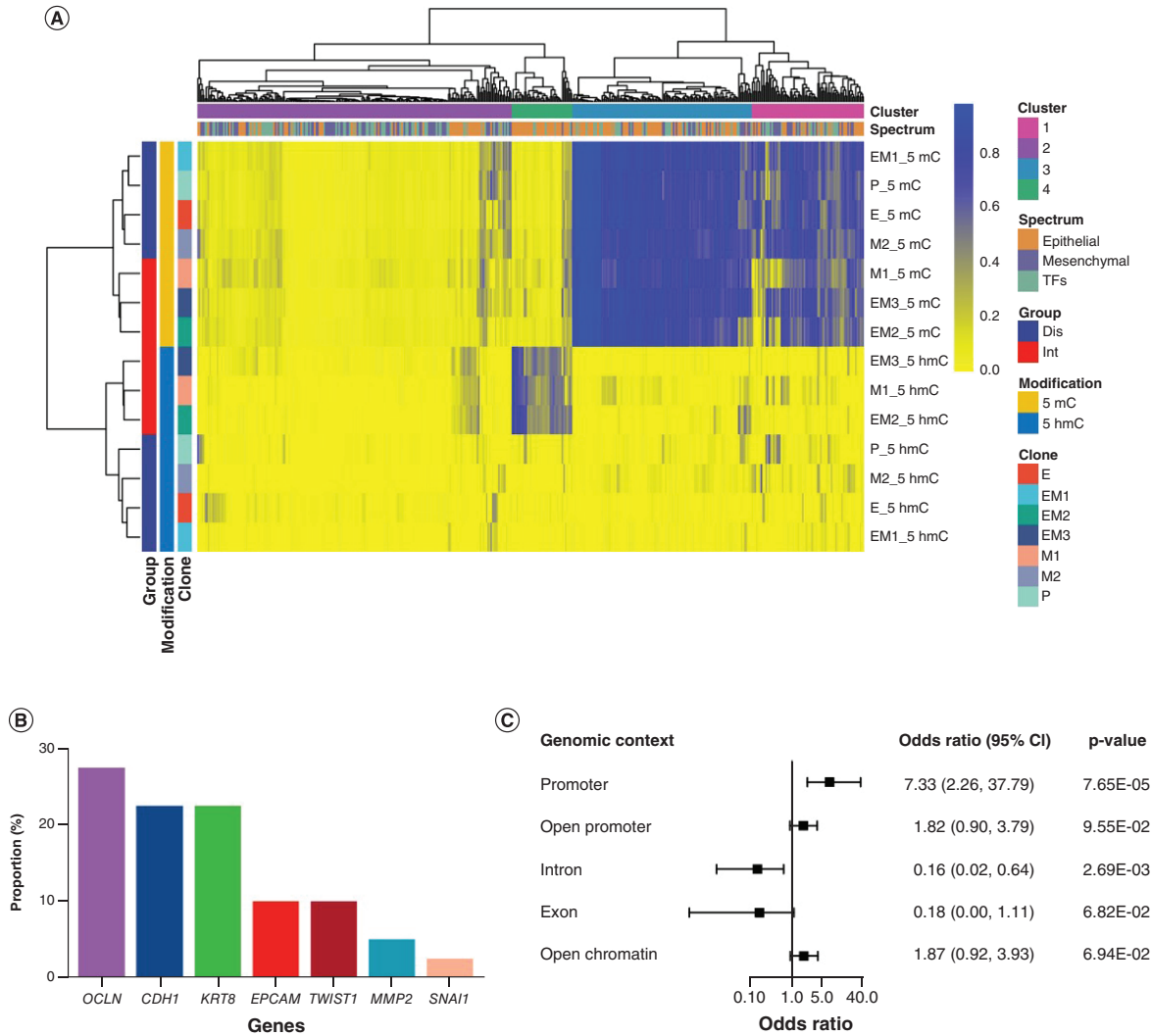
Figure 5.   Intermediate clones have higher hydroxymethylation among epithelial genes. (A) Heatmap of unsupervised clustering of 5hmC and 5mC in a set of 231 CpGs within epithelial genes (*CDH1, CLDN1, EPCAM, ITGB4, KRT8,* and *OLCN*), mesenchymal genes (*CDH2, FN1, ITGB1, MMP19, MMP2,* and *VIM*), and transcription factors (*SNAI1, SNAI2, TWIST1, ZEB1,* and *ZEB2*). Vertical tracking bars indicate DNA modification, clones, and position on the EMT spectrum. Horizontal tracking bars indicate EMT marker group (epithelial genes, mesenchymal genes, transcription factors) and hierarchical clustering group from when height = 1.6. **(B)** Proportions of the genes annotated to the 40 CpGs in cluster #4 of the hierarchical clustering from the heatmap. **(C)** Enrichment of genomic contexts of the CpGs in hierarchical cluster 4. 40 CpGs in cluster 4 were compared to all 231 CpGs in EMT-related genes using Fisher's test.

on the same gene, generally the two DNA cytosine modification marks act on different regions of the genome to coordinate EMT processes.

GREAT analyses revealed that dhmCpGs were associated with fatty acid-related molecular functions (MFs), such as peroxisomal fatty-acyl-CoA transporter activity (fold enrichment [FE]: 19.00) and long-chain fatty acid transporter activity (FE: 7.46), as well as RNA polymerase II TF-related molecular functions such as RNA polymerase II TF sequence-specific DNA binding (FE: 1.18) and RNA polymerase II regulatory region DNA binding (FE: 1.17) (Supplementary Figure 6A). Similarly, dmCpGs were associated with RNA polymerase II-related molecular functions such as RNA polymerase II transcription coactivator binding (FE: 7.62) and cofactor binding (FE: 6.95) (Supplementary Figure 6B). Additionally, dmCpGs were associated with metal ion transmembrane activity (FE: 1.44). Collectively, these results support the role of differential cytosine modifications in RNA polymerase II-related regulation of transcription to influence the intermediate EMT phenotype.
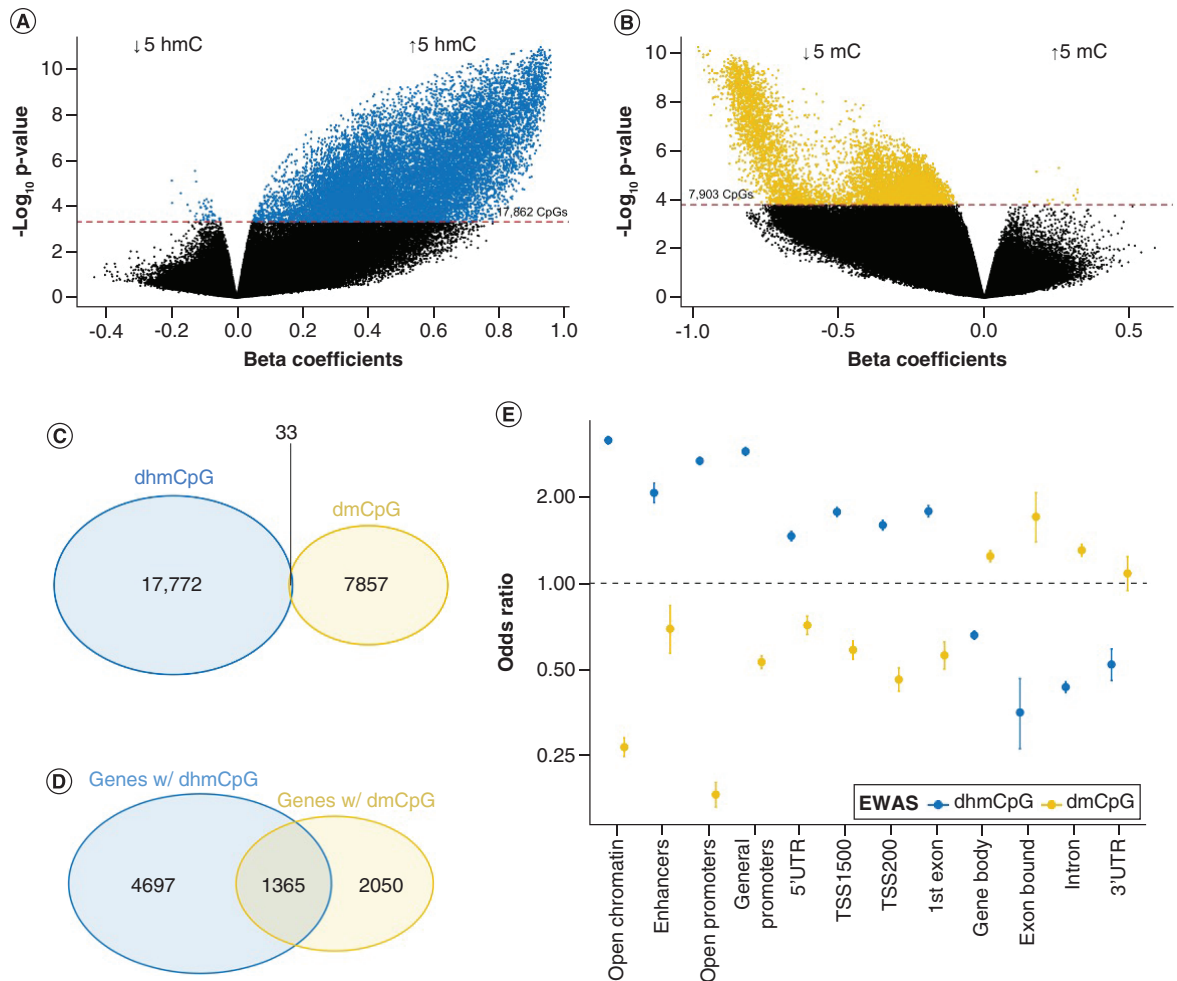
Figure 6.   Differential 5-hydroxymethylcytosine (5hmC) CpGs in intermediate clones compared to distal clones are distinct from the differential 5-methylcytosine (5mC) CpGs. Volcano plots indicating **(A)** 17,862 significantly differentially hydroxymethylated CpGs (dhmCpGs) and **(B)** 7903 significantly differentially methylated CpGs (dmCpGs) under false discovery rate (FDR) q-value of 0.01, in intermediate clones in comparison with distal clones. Red dashed lines indicate the -log10 (p-value) at FDR q-value of 0.01. Venn diagrams comparing **(C)** dhmCpGs versus dmCpGs and **(D)** genes annotated to dhmCpGs versus genes annotated to dmCpGs. dhmCpGs were subset for only CpGs increasing in 5hmC. dmCpGs were subset for only CpGs decreasing in 5mC. **(E)** Enrichment of dhmCpGs and dmCpGs at different genomic contexts. Odds ratios calculated by Fisher's exact test. dhmCpG enrichment indicated in blue. dmCpG enrichment indicated in yellow.

Potential roles of 5hmC in regulating EMT

As increased hydroxymethylation and decreased methylation are traditionally associated with increased expression, the authors wanted to determine whether the dhmCpGs and dmCpGs were acting in regions of open chromatin as identified by ATAC-seq. Out of 42,510 open chromatin regions containing a CpG that was measured on the Illumina EPIC array, 12.03% of the open chromatin regions contained dhmCpGs, in contrast to 1.59% of the open chromatin regions containing dmCpGs (Figure 7A). Interestingly, the only pathways significantly associated with the open chromatin regions containing dhmCpGs were related to the Rho family of GTPase, which have been extensively shown to function as cellular switches in coordinating cell polarity and migration by regulating the cytoskeleton (Figure 7B) [79]. Expression of the majority of genes in the Rho GTPase cycle pathway is high in EM1, EM2 and EM3 clones (Figure 7C).

To identify additional molecular processes dhmCpGs in open chromatin regions may regulate, the authors conducted TF motif enrichment analysis. Motif enrichment analysis found 571 TFs significantly associated with open chromatin regions with dhmCpGs in intermediate clones compared with only four TFs in distal clones under
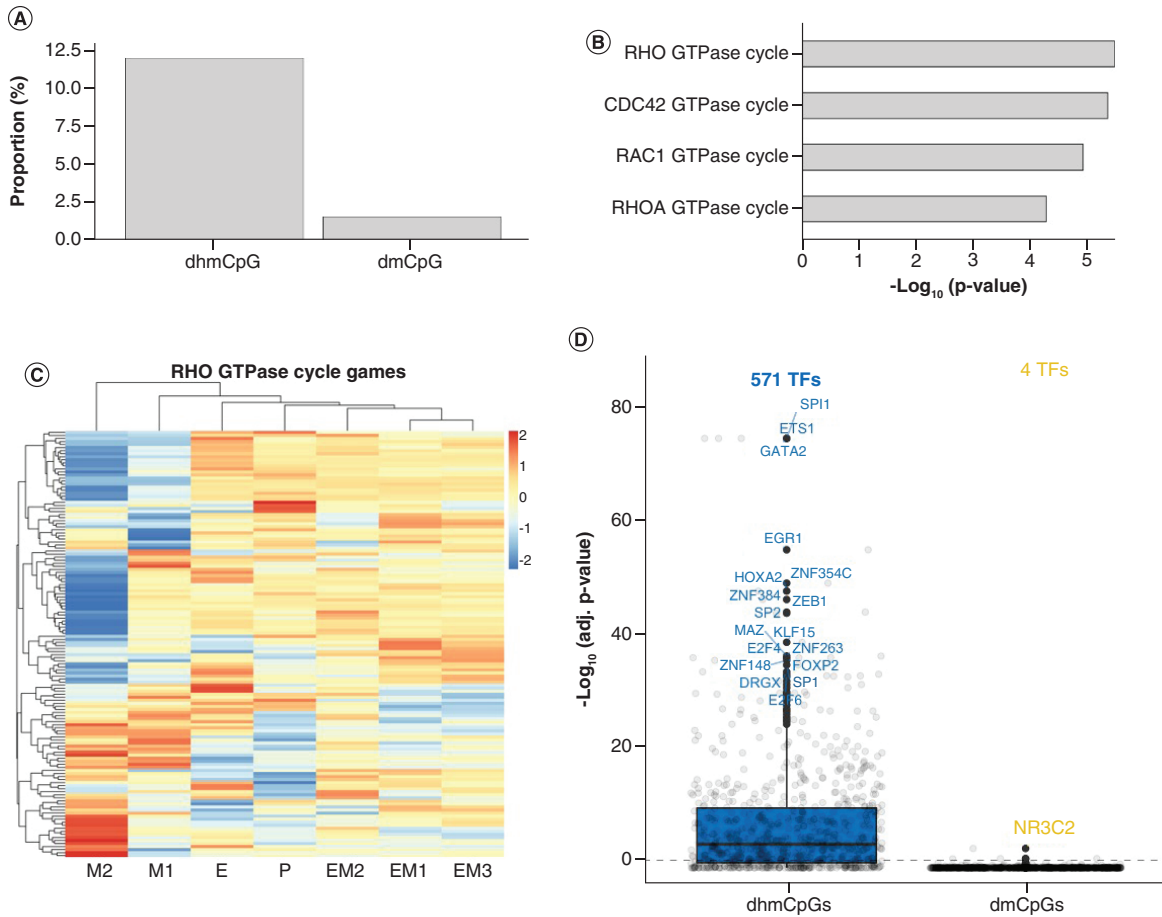
Figure 7. Differentially hydroxymethylated CpGs (dhmCpGs) in open chromatin regions are associated with the Rho GTPase family and epithelial-to-mesenchymal transition-specific transcription factor motifs. (A) Proportion of open chromatin regions with dhmCpGs and differentially methylated CpGs (dmCpGs) in open chromatin regions containing CpGs analyzed from the Illumina Methylation EPIC array. **(B)** Reactome pathways associated with open chromatin regions containing dhmCpGs. **(C)** Gene expression z-scores of genes in the Rho GTPase reactome pathway for each clone. Red indicates high expression. Blue indicates low expression. **(D)** Transcription factor motifs associated with open chromatin regions containing dhmCpGs and dmCpGs.

the FDR <0.05 threshold (Figure 7D & Supplementary Table 4). In the intermediate clones, motifs for key EMT TFs (ZEB1 and SNAI2) were enriched among open chromatin regions with dhmCpG, implicating 5hmC in EMT process-associated gene regulation. In addition, motifs for GRHL2, a suggested EMT pioneer TF that has been shown to be associated with epigenetic remodeling, also were enriched in consensus open chromatin regions with dhmCpGs, but not in consensus open chromatin regions with dmCpGs (Supplementary Table 4) [80,81]. While not known specifically to play roles in EMT, other TF motifs, particularly motifs of GATA2 and SPI1, were also found to be in open chromatin regions with dhmCpGs. Together, these results suggest that an increase in 5hmC may play a regulatory role in the EMT process by acting in Rho GTPase-associated genes and acting on binding sites of EMT-associated TFs.

Discussion

Widely used standard bisulfite conversion used to study DNA methylation is unable to distinguish between 5mC and 5hmC. Using a tandem oxidative-bisulfite treatment approach, the authors measured both cytosine modifications to understand their unique distribution across distal and intermediate EMT states. The majority of previous studies measuring 5hmC have been limited to global 5hmC levels in tissues of heterogeneous cell types, including tumors, where extremely low levels of 5hmC were observed [38,82,83]. Here, identifying differences in cell-state-specific, nucleotide-specific 5hmC is a strength of this approach. The intermediate clones in the EMT model

system suggests that genome-wide patterns of hydroxymethylation are associated with specific EMT phenotypes, suggesting a potential role of 5hmC in mediating EMT-related processes. Moreover, through a multi-component approach of epigenome profiling, the authors show that EMT phenotypes are underscored by substantial epigenetic differences.

Previous work establishing this model system has demonstrated that the intermediate clones represent a population of tumor cells with high migratory and invasive properties. The authors identify that open chromatin regions with dhmCpGs are particularly associated with the Rho family of GTPases, the family of GTPases that regulates cell polarity and migration by coordinating the cytoskeleton [79]. Rho GTPases have been well documented to play a role in EMT in tumors [84]. While Rho GTPases have been implicated in tumor progression, mutations in Rho proteins are not common and do not favor the initiation or progression of tumors, which has called for the study of other mechanisms in deregulation Rho proteins [85]. The present study suggests that increasing 5hmC may be implicated in the EMT, which in turn may contribute to deregulation of Rho proteins. In addition, the authors show that dhmCpGs are associated with motifs of key EMT TFs, which may indicate that the recruitment of various TFs by 5hmC may be a potential mechanism regulating the intermediate clones' high migratory and invasive potential. These results suggest that targeting increases in 5hmC in intermediate cells may impede the maintenance of this state and/or force lineage commitment, effects that could lead to altered metastatic propensity.

Prior literature has already indicated that DNA methylation states change during TGF-β-induced EMT [86]. Similarly, this study's natural (non-TGF-β-induced) EMT model suggests that DNA cytosine modifications exhibit altered genome-wide patterns during the EMT process. These results indicate that these altered patterns may regulate the existence of cells in various EMT states, thereby enabling tumor heterogeneity. Alterations in cytosine modifications and chromatin accessibility toward a less repressive state suggest that the multi-level epigenome is essential in regulating the dynamics of EMT.

Finally, this study highlights the importance of multi-component measures of epigenetic states. Utilizing ATAC-seq in combination with 5mC and 5hmC methylation array profiles allowed for the identification of the significance of the Rho GTPases that was not evident in only DNA cytosine modification analyses. Moreover, combined datasets allowed for the identification of the potential role of 5hmC regulating EMT-related TFs. However, the array-based approach may not have revealed CpG loci in relevant accessible chromatin, a limitation that may be overcome with a whole genome bisulfite, oxidative-bisulfite sequencing approach. It highlights the complex epigenetic landscape that is required in the EMT process.

Conclusion

Our study addresses current gaps in the understanding of the roles of specific cytosine modifications (5mC and 5hmC) in EMT and their associations with other epigenetic changes. Clones exhibiting intermediate EMT phenotypes had distinct, more open epigenetic states with increased 5hmC, decreased 5mC and more accessible chromatin compared with clones exhibiting more distal EMT phenotypes. Open chromatin regions containing CpG loci with increased 5hmC enriched in motifs of key EMT TFs, ZEB1 and SNAI2, indicate the likelihood of multi-component epigenetic regulation during EMT. Epigenetic profiles at the cytosine and chromatin levels associated with EMT processes that contribute to gene regulation may be targeted to prevent the progression of EMT.

Future perspective

The roles of cell-state-specific epigenomic changes, specifically in multiple DNA cytosine modification marks, in regulating EMT are only just beginning to be identified. Utilizing multiple genome-wide epigenomic assays will improve the understanding of how different parts of the epigenome interact to regulate EMT, which may yield new therapeutic targets to prevent EMT. With novel epigenetic targets, therapeutic strategies to prevent cancer progression into metastasis may be developed for clinical use.

Summary points

- Epithelial-to-mesenchymal transition (EMT) is an early step in the invasion-metastasis cascade, involving progression through a number of cellular states.
- Due to challenges with isolating intermediate cell states, genome-wide cytosine modification mechanisms that define transition are not completely understood.
- This study provides a nucleotide-resolution genome-scale map of cytosine modifications and chromatin accessibility for phenotypes spanning the EMT spectrum to address gaps in understanding epigenomic changes in the intermediate/hybrid states during EMT.
- The study utilized a previously derived model of six single cell clones from SUM149PT, a heterogeneous estrogen receptor/progesterone receptor-negative inflammatory breast cancer line, that represent phenotypes across the EMT spectrum.
- Variable patterns of genome-wide 5-hydroxymethylcytosine (5hmC) and 5-methylcytosine (5mC) exist based on clonal EMT status.
- A total of 17,862 significantly differentially hydroxymethylated CpGs, almost all of which were increased in 5hmC, and 7903 significantly differentially methylated CpGs, most of which were decreased in 5mC, were identified in intermediate clones.
- Some differentially hydroxymethylated CpGs and differentially methylated CpGs were in EMT-associated genes such as *SNAI1* and *TWIST1* and in epithelial or mesenchymal cell markers such as *CDH1* and *MMP19*.
- Open chromatin regions containing increased 5hmC CpGs were associated with the Rho family of GTPases, proteins that have been extensively shown to function as cellular switches in coordinating cell polarity and migration by regulating the cytoskeleton.
- In the intermediate clones, motifs for key EMT transcription factors (ZEB1 and SNAI2) were enriched among open chromatin regions with increased 5hmC CpGs, implicating 5hmC in EMT process-associated gene regulation.
- The results suggest 5hmC may play a regulatory role in the EMT process.

Author contributions

MK Lee and MS Brown carried out the experiments. MS Brown and DR Pattabiraman conceived the original epithelial-to-mesenchymal transition experimental design. MK Lee conducted analyses with help from OM Wilkins and BC Christensen. BC Christensen supervised the project. All authors discussed the results and contributed to the final version of the manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review board approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations.

Data sharing statement

The datasets used and/or analyzed during the current study are available in the GEO repository, under accession numbers GSE181036 and GSE172613.

Open access

References

1. Nieto MA, Huang RY, Jackson RA, Thiery JP. EMT: 2016. *Cell* 166(1), 21–45 (2016).

2. Jolly MK, Boareto M, Huang B *et al.* Implications of the hybrid epithelial/mesenchymal phenotype in metastasis. *Front. Oncol.* 5(JUN), 155 (2015).

3. Pastushenko I, Brisebarre A, Sifrim A *et al.* Identification of the tumour transition states occurring during EMT. *Nature* 556(7702), 463–468 (2018).

4. Chaffer CL, Weinberg RA. A perspective on cancer cell metastasis. *Science (80-.)* 331(6024), 1559–1564 (2011).

5. Hanna D, Rogers MS, Straume O. Are 90% of deaths from cancer caused by metastases? *Cancer Med.* 8, 5574–5576 (2019).

6. Nieto MA. Epithelial plasticity: a common theme in embryonic and cancer cells. *Science (80-.)* 342(6159), 1234850 (2013).

7. Jolly MK, Somarelli JA, Sheth M *et al.* Hybrid epithelial/mesenchymal phenotypes promote metastasis and therapy resistance across carcinomas. *Pharmacol. Ther.* 194, 161–184 (2019).

8. Brown MS, Abdollahi B, Wilkins OM *et al.* Dynamic plasticity within the EMT spectrum, rather than static mesenchymal traits, drives tumor heterogeneity and metastatic progression of breast cancers. *bioRxiv* doi:2021.03.17.434993 (2021).

9. Grosse-Wilde A, Fouquier D'hérouël A, McIntosh E *et al.* Stemness of the hybrid epithelial/mesenchymal state in breast cancer and its association with poor survival. *PLoS One* 10(5), e0126522 (2015).

10. Bierie B, Pierce SE, Kroger C *et al.* Integrin-β4 identifies cancer stem cell-enriched populations of partially mesenchymal carcinoma cells. *Proc. Natl Acad. Sci. USA* 114(12), e2337–e2346 (2017).

11. Ognjenovic NB, Bagheri M, Mohamed GA *et al.* Limiting self-renewal of the basal compartment by PKA activation induces differentiation and alters the evolution of mammary tumors. *Dev. Cell.* 55(5), 544–557.e6 (2020).

12. Goetz H, Melendez-Alvarez JR, Chen L, Tian XJ. A plausible accelerating function of intermediate states in cancer metastasis. *PLoS Comput. Biol.* 16(3), e1007682 (2020).

13. Hong T, Watanabe K, Ha Ta C, Villarreal-Ponce A, Nie Q, Dai X. An Ovol2-Zeb1 mutual inhibitory circuit governs bidirectional and multi-step transition between epithelial and mesenchymal states. *PLoS Comput. Biol.* 11(11), 1004569 (2015).

14. Huang RYJ, Wong MK, Tan TZ *et al.* An EMT spectrum defines an anoikis-resistant and spheroidogenic intermediate mesenchymal state that is sensitive to e-cadherin restoration by a src-kinase inhibitor, saracatinib (AZD0530). *Cell Death Dis.* 4(11), e915 (2013).

15. Tahiliani M, Koh KP, Shen Y *et al.* Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science (80-.)* 324(5929), 930–935 (2009).

16. Janitz K, Janitz M. Assessing epigenetic information. In: *Handbook of Epigenetics: The New Molecular and Medical Genetics* Tollefsbol T (Ed.). Elsevier/Academic Press, Amsterdam, The Netherlands, 173–181 (2011).

17. Cortellino S, Xu J, Sannai M *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* 146(1), 67–79 (2011).

18. Tost J. DNA methylation: an introduction to the biology and the disease-associated changes of a promising biomarker. *Mol. Biotechnol.* 44(1), 71–81 (2010).

19. Cui XL, Nie J, Ku J *et al.* A human tissue map of 5-hydroxymethylcytosines exhibits tissue specificity through gene and enhancer modulation. *Nat. Commun.* 11(1), 6161 (2020).

20. Nestor CE, Ottaviano R, Reddington J *et al.* Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* 22(3), 467–477 (2012).

21. Pistore C, Giannoni E, Colangelo T *et al.* DNA methylation variations are required for epithelial-to-mesenchymal transition induced by cancer-associated fibroblasts in prostate cancer cells. *Oncogene* 36(40), 5551–5566 (2017).

22. Tahara T, Shibata T, Okubo M *et al.* DNA methylation status of epithelial-mesenchymal transition (EMT) – related genes is associated with severe clinical phenotypes in ulcerative colitis (UC). *PLoS One* 9(10), e107947 (2014).

23. Carmona FJ, Davalos V, Vidal E *et al.* A comprehensive DNA methylation profile of epithelial-to-mesenchymal transition. *Cancer Res.* 74(19), 5608–5619 (2014).

24. Dumont N, Wilson MB, Crawford YG, Reynolds PA, Sigaroudinia M, Tlsty TD. Sustained induction of epithelial to mesenchymal transition activates DNA methylation of genes silenced in basal-like breast cancers. *Proc. Natl Acad. Sci. USA* 105(39), 14867–14872 (2008).

25. Rajić J, Dinić S, Usković A *et al.* DNA methylation of miR-200 clusters promotes epithelial to mesenchymal transition in human conjunctival epithelial cells. *Exp. Eye Res.* 197, 108047 (2020).

26. Choi SK, Pandiyan K, Eun JW *et al.* Epigenetic landscape change analysis during human EMT sheds light on a key EMT mediator TRIM29. *Oncotarget* 8(58), 98322–98335 (2017).

27. Booth MJ, Ost TWB, Beraldi D *et al.* Oxidative bisulfite sequencing of 5-methylcytosine and 5-hydroxymethylcytosine. *Nat. Protoc.* 8(10), 1841–1851 (2013).

28. Kumar S, Chinnusamy V, Mohapatra T. Epigenetics of modified DNA bases: 5-methylcytosine and beyond. *Front. Genet.* 9, 640 (2018).

29. Pastor WA, Aravind L, Rao A. TETonic shift: biological roles of TET proteins in DNA demethylation and transcription. *Nat. Rev. Mol. Cell Biol.* 14(6), 341–356 (2013).

30. Ngo TTM, Yoo J, Dai Q *et al.* Effects of cytosine modifications on DNA flexibility and nucleosome mechanical stability. *Nat. Commun.* 7, 10813 (2016).

31. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. *Genome Biol.* 12(6), R54 (2011).

32. Tekpli X, Urbanucci A, Hashim A *et al.* Changes of 5-hydroxymethylcytosine distribution during myeloid and lymphoid differentiation of CD34+ cells. *Epigenetics Chromatin* 9(1), 21 (2016).

33. Costa Y, Ding J, Theunissen TW *et al.* NANOG-dependent function of TET1 and TET2 in establishment of pluripotency. *Nature* 495(7441), 370–374 (2013).

34. Xiong J, Zhang Z, Chen J *et al.* Cooperative action between SALL4A and TET proteins in stepwise oxidation of 5-methylcytosine. *Mol. Cell.* 64(5), 913–925 (2016).

35. He B, Zhang C, Zhang X *et al.* Tissue-specific 5-hydroxymethylcytosine landscape of the human genome. *Nat. Commun.* 12(1), 1–12 (2021).

36. Jin SG, Jiang Y, Qiu R *et al.* 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. *Cancer Res.* 71(24), 7360–7365 (2011).

37. Shenoy N, Bhagat TD, Cheville J *et al.* Ascorbic acid–induced TET activation mitigates adverse hydroxymethylcytosine loss in renal cell carcinoma. *J. Clin. Invest.* 129(4), 1612–1625 (2019).

38. Yang H, Liu Y, Bai F *et al.* Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. *Oncogene* 32(5), 663–669 (2013).

39. Tsai KW, Li GC, Chen CH *et al.* Reduction of global 5-hydroxymethylcytosine is a poor prognostic factor in breast cancer patients, especially for an ER/PR-negative subtype. *Breast Cancer Res. Treat.* 153(1), 219–234 (2015).

40. Skvortsova K, Masle-Farquhar E, Luu PL *et al.* DNA hypermethylation encroachment at CpG island borders in cancer is predisposed by H3K4 monomethylation patterns. *Cancer Cell* 35(2), 297–314.e8 (2019).

41. Arechederra M, Daian F, Yim A *et al.* Hypermethylation of gene body CpG islands predicts high dosage of functional oncogenes in liver cancer. *Nat. Commun.* 9(1), 1–16 (2018).

42. Gonzalez-Zulueta M, Bender CM, Yang AS *et al.* Methylation of the 5′ CpG island of the p16/CDKN2 tumor suppressor gene in normal and transformed human tissues correlates with gene silencing. *Cancer Res.* 55(20), 4531–5 (1995).

43. Herman JG, Latif F, Weng Y *et al.* Silencing of the VHL tumor-suppressor gene by DNA methylation in renal carcinoma. *Proc. Natl Acad. Sci. USA* 91(21), 9700–9704 (1994).

44. Jin C, Lu Y, Jelinek J *et al.* TET1 is a maintenance DNA demethylase that prevents methylation spreading in differentiated cells. *Nucleic Acids Res.* 42(11), 6956–6971 (2014).

45. Wilkins OM, Johnson KC, Houseman EA, King JE, Marsit CJ, Christensen BC. Genome-wide characterization of cytosine-specific 5-hydroxymethylation in normal breast tissue. *Epigenetics* 15(4), 398–418 (2020).

46. Fleischer T, Frigessi A, Johnson KC *et al.* Genome-wide DNA methylation profiles in progression to *in situ* and invasive carcinoma of the breast with impact on gene transcription and prognosis. *Genome Biol.* 15(8), 435 (2014).

47. Johnson KC, Koestler DC, Fleischer T *et al.* DNA methylation in ductal carcinoma *in situ* related with future development of invasive breast cancer. *Clin. Epigenetics* 7(1), 75 (2015).

48. Klemm SL, Shipony Z, Greenleaf WJ. Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* 20(4), 207–220 (2019).

49. Cieślik M, Hoang SA, Baranova N *et al.* Epigenetic coordination of signaling pathways during the epithelial-mesenchymal transition. *Epigenetics Chromatin* 6(1), 28 (2013).

50. Bakiri L, Macho-Maschler S, Custic I *et al.* Fra-1/AP-1 induces EMT in mammary epithelial cells by modulating Zeb1/2 and TGFβ expression. *Cell Death Differ.* 22(2), 336–350 (2015).

51. Hardy K, Wu F, Tu W *et al.* Identification of chromatin accessibility domains in human breast cancer stem cells. *Nucleus* 7(1), 50–67 (2016).

52. Arase M, Tamura Y, Kawasaki N *et al.* Dynamics of chromatin accessibility during TGF-β-induced EMT of Ras-transformed mammary gland epithelial cells. *Sci. Reports* 7(1), 1–14 (2017).

53. Denny SK, Yang D, Chuang CH *et al.* Nfib promotes metastasis through a widespread increase in chromatin accessibility. *Cell* 166(2), 328–342 (2016).

54. Xin L, Zhao R, Lei J *et al.* SND1 acts upstream of SLUG to regulate the epithelial–mesenchymal transition (EMT) in SKOV3 cells. *FASEB J.* 33(3), 3795–3806 (2019).

55. Frey WD, Chaudhry A, Slepicka PF *et al.* BPTF maintains chromatin accessibility and the self-renewal capacity of mammary gland stem cells. *Stem Cell Reports* 9(1), 23–31 (2017).

56. Stowers RS, Shcherbina A, Israeli J *et al.* Matrix stiffness induces a tumorigenic phenotype in mammary epithelium through changes in chromatin accessibility. *Nat. Biomed. Eng.* 3(12), 1009–1019 (2019).

57. Aryee MJ, Jaffe AE, Corrada-Bravo H *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* 30(10), 1363–1369 (2014).

58. Hansen K. IlluminaHumanMethylationEPICanno.ilm10b4.hg19: annotation for Illumina's EPIC methylation arrays. *R Packag. version 0.6.0.* (2017). https://bitbucket.com/kasperdanielhansen/Illumina_EPIC

59. Zhou W, Laird PW, Shen H. Comprehensive characterization, annotation and innovative use of Infinium DNA methylation BeadChip probes. *Nucleic Acids Res.* 45(4), e22 (2017).

60. Houseman EA, Johnson KC, Christensen BC. OxyBS: estimation of 5-methylcytosine and 5-hydroxymethylcytosine from tandem-treated oxidative bisulfite and bisulfite DNA. *Bioinformatics* 32(16), 2505–7 (2016).

61. Ritchie ME, Phipson B, Wu D *et al. limma* powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43(7), e47 (2015).

62. Storey JD, Bass AJ, Dabney A, Robinson D, Warnes G. qvalue: q-value estimation for false discovery rate control. (2019). http://github.com/jdstorey/qvalue

63. McLean CY, Bristor D, Hiller M *et al.* GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* 28(5), 495–501 (2010).

64. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 2015, 21.29.1–21.29.9 (2015).

65. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* 17(1), 10 (2011).

66. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* 9(4), 357–359 (2012).

67. Picard Tools – by Broad Institute. Github (2018). http://broadinstitute.github.io/picard/

68. Zhang Y, Liu T, Meyer CA *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9(9), R137 (2008).

69. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1), 139–140 (2009).

70. Maintainer B, Team B. TxDb.Hsapiens.UCSC.hg38.knownGene: annotation package for TxDb object(s). *R Packag. version 3.4.6.* (2019).

71. Yu G, Wang LG, He QY. ChIPseeker: an R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics* 31(14), 2382–2383 (2015).

72. Yu G, He QY. ReactomePA: an R/Bioconductor package for reactome pathway analysis and visualization. *Mol. Biosyst.* 12(2), 477–479 (2016).

73. Schep A. motifmatchr: fast motif matching in R. *R package version 1.14.0* (2020).

74. Fornes O, Castro-Mondragon JA, Khan A *et al.* JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* 48(D1), D87–D92 (2020).

75. Tan G, Lenhard B. TFBSTools: an R/Bioconductor package for transcription factor binding site analysis. *Bioinformatics* 32(10), 1555–1556 (2016).

76. Dobin A, Davis CA, Schlesinger F *et al.* Sequence analysis STAR: ultrafast universal RNA-seq aligner. 29(1), 15–21 (2013).

77. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-seq data with or without a reference genome. *BMC Bioinformatics* 12(1), 323 (2011).

78. Yang J, Mani SA, Donaher JL *et al.* Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* 117(7), 927–939 (2004).

79. Clayton NS, Ridley AJ. Targeting Rho GTPase signaling networks in cancer. *Front. Cell Dev. Biol.* 8, 222 (2020).

80. Chen AF, Liu AJ, Krishnakumar R, Freimer JW, Deveale B, Blelloch R. GRHL2-dependent enhancer switching maintains a pluripotent stem cell transcriptional subnetwork after exit from naive pluripotency. *Cell Stem Cell* 23, 226–238 (2018).

81. Chung VY, Tan TZ, Ye J *et al.* The role of GRHL2 and epigenetic remodeling in epithelial–mesenchymal plasticity in ovarian cancer cells. *Commun. Biol.* 2(1), 1–15 (2019).

82. Nestor CE, Ottaviano R, Reddington J *et al.* Tissue type is a major modifier of the 5-hydroxymethylcytosine content of human genes. *Genome Res.* 22(3), 467–477 (2012).

83. Lian CG, Xu Y, Ceol C *et al.* Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. *Cell* 150(6), 1135–1146 (2012).

84. Ungefroren H, Witte D, Lehnert H. The role of small GTPases of the Rho/Rac family in TGF-β-induced EMT and cell motility in cancer. *Dev. Dyn.* 247(3), 451–461 (2018).

85. Ellenbroek SIJ, Collard JG. Rho GTPases: functions and association with cancer. *Clin. Exp. Metastasis* 24(8), 657–672 (2007).

86. Gong F, Guo Y, Niu Y *et al.* Epigenetic silencing of TET2 and TET3 induces an EMT-like process in melanoma. *Oncotarget* 8(1), 315–328 (2017).