Distinct epigenetic signatures elucidate enhancer-gene relationships that delineate CIMP and non-CIMP colorectal cancers

Allen Chong[1,5], Jing Xian Teo[2], Kenneth H.K. Ban[3,4]

[1]Department of Pathology, National University of Singapore, 119074 Singapore

[2]Cancer Science Institute, National University of Singapore, 117599 Singapore

[3]Department of Biochemistry, National University of Singapore, 117596 Singapore

[4]Institute of Molecular and Cell Biology, 138673 Singapore

[5]Present address: Shanxi Guoxin Caregeno Medical Laboratories, Taiyuan, Shanxi Province, China

Correspondence to: *Allen Chong,* ***e-mail:*** *chongallen@yahoo.com*
Kenneth H.K. Ban, ***e-mail:*** *kenneth_ban@nuhs.edu.sg*

Keywords: *DNA methylation, enhancers, epigenetics, bioinformatics, colon cancer*

Received: *December 10, 2015* ***Accepted:*** *March 14, 2016* ***Published:*** *March 30, 2016*

ABSTRACT

Epigenetic changes, like DNA methylation, affect gene expression and in colorectal cancer (CRC), a distinct phenotype called the CpG island methylator phenotype ("CIMP") has significantly higher levels of DNA methylation at so-called "Type C loci" within the genome. We postulate that enhancer-gene pairs are coordinately controlled through DNA methylation in order to regulate the expression of key genes/biomarkers for a particular phenotype.

Firstly, we found 24 experimentally-validated enhancers (VISTA enhancer browser) that contained statistically significant (FDR-adjusted q-value of <0.01) differentially methylated regions (DMRs) (1000bp) in a study of CIMP versus non-CIMP CRCs. Of these, the methylation of 2 enhancers, 1702 and 1944, were found to be very well correlated with the methylation of the genes *Wnt3A* and *IGDCC3*, respectively, in two separate and independent datasets.

We show for the first time that there are indeed distinct and dynamic changes in the methylation pattern of specific enhancer-gene pairs in CRCs. Such a coordinated epigenetic event could be indicative of an interaction between (1) enhancer 1702 and *Wnt3A* and (2) enhancer 1944 and *IGDCC3*. Moreover, our study shows that the methylation patterns of these 2 enhancer-gene pairs can potentially be used as biomarkers to delineate CIMP from non-CIMP CRCs.

INTRODUCTION

Toyota *et al.* [1] first coined the term "CIMP" and defined the CpG island methylator phenotype as a subset of colorectal cancers (CRCs) which showed DNA methylation in 7 cancer-specific differentially methylated loci (termed Type C for cancer-specific – the other 19 differentially methylated loci that they studied were found to be age-specific and termed Type A). In contrast to Type A methylation, Type C methylation is relatively infrequent in primary colorectal cancer, and is never observed in normal colon mucosa. Some studies have since tried to re-classify the CIMP phenotype into CIMP-high, CIMP-low and CIMP-negative based on different panels of markers [2, 3].

Holliday, Pugh and Riggs were first to suggest that DNA methylation of cytosines in the context of CpG dinucleotides may represent an epigenetic mark associated with gene silencing [4, 5]. CpG density is not evenly distributed within the genome, but rather, shows a bimodal distribution. Regions with an elevated CpG content, so-called CpG islands (CGIs), overlap with the transcriptional start sites (TSSs) of approximately 60–70% of all human genes [6, 7], whereas, regions of low CpG density are frequently located outside the TSS [8, 9]. DNA methylation at gene promoters is important for transcriptional regulation, with dense promoter hypermethylation around the TSS often being associated with gene repression [10, 11]. In contrast, the

hypomethylation of CpG sites has been associated with the overexpression of oncogenes within cancer cells [12].

However, in addition to gene promoters, distal regulatory regions such as enhancers, silencers and boundary elements are also often required to establish correct gene expression patterns in mammalian cells [13]. Transcription factors bind enhancers, which play key roles in the control of cell-type-specific gene expression [14–20]. A typical mammalian cell contains thousands of active enhancers, and it has been estimated that there may be ~1 million enhancers active in all human cells [21–23].

Enhancers play an important role in gene regulation and this is evident from mutations within enhancers that have the potential to generate a plethora of transcriptional alterations through both loss- and gain-of-function effects, leading to a gradient of phenotypic severities. A clear example of this is the dysregulation of *SHH* expression and limb malformations. *SHH* expression in a region of limb buds known as the zone of polarizing activity (ZPA) is necessary for limb patterning [24]. This expression pattern is governed by a long-range enhancer element about 1 Mb from *SHH*, known as the ZPA regulatory sequence (ZRS). Point mutations within ZRS have been linked to a congenital disease leading to extra digits known as preaxial polydactyly [25], whereas deletion of the entire ZRS in mice led to a truncation of limbs [26]. If point mutations within enhancers can cause a phenotypic change, then it is not surprising that an epigenetic change, such as DNA methylation, can also affect the activity of an enhancer and therefore cause a phenotypic change [27, 28].

Although it was originally thought that enhancers regulate only a single nearby promoter, many observations over the past 25 years point to a more complex interplay. Enhancers can control multiple neighboring genes [29–32], sometimes over hundreds of kb and often skipping one or more genes [33, 34]. With the advent of next-generation sequencing, there have been much interest in exploring the enhancer-promoter interactome across several cell types [35–38]. As it was previously shown that correlation between enhancer and promoter histone modification patterns can be used to infer their interactions [39], we therefore extended this approach and searched for well-correlated DNA methylation events between enhancers and genes that were unique to the CIMP/non-CIMP CRC phenotypes.

RESULTS AND DISCUSSION

Systematic mapping of differentially methylated regions (DMRs) and differentially methylated enhancers (DMEs) across 11 colorectal cancer cell lines

To explore the genome-wide DNA methylation state at single-base resolution across 11 CRC cells, we applied the method known as MethylCap-BS-seq, previously described by Brinkman *et al.* [40]. This involves the enrichment of methylated DNA by capturing with a monoclonal antibody against a methyl-DNA binding protein domain (MBD) followed by bisulfite conversion and deep sequencing to directly assess DNA methylation levels in captured chromatin fragments (see Methods). The 11 CRC cell lines include 6 CIMPs (Colo205, DLD1, HCT116, HT29, LIM2405, RKO) and 5 non-CIMPs (Caco-2, Colo320, LIM1215, SW403, SW480) [2, 41]. The sequence reads were mapped using Bismark [42] and DNA methylation profiles for the 11 cell lines were analyzed with the R package, methylKit [43].

In this study, we created 2 groups, namely, CIMP and non-CIMP, from the 11 cell lines and allowed methylKit to determine regions that are differentially methylated between the CIMP and non-CIMP groups through a logistic regression test [43]. Each differentially methylated region (DMR) identified by methylKit had a window size of 1000 bp. To ensure a very high level of stringency in the detection of the DMRs between the CIMP and non-CIMP, the limiting criteria for identifying these DMRs were that these regions would have a percent methylation difference of >50% with a q-value (i.e. FDR-adjusted) of <0.01. In total, 16,092 regions were discovered by methylKit to be differentially methylated between the CIMP and non-CIMP groups (Supplementary File 2).

We compared these 16,092 DMRs against 1729 experimentally validated human gene enhancers from VISTA Enhancer Browser [44] and found that 24 DMRs were found completely within 24 different gene enhancers.

Of the 24 enhancers identified by methylKit to contain a DMR, we wanted to prioritize and narrow down the field of 24 enhancers and only study those enhancers which we considered most important. How then do we decide which of the 24 enhancers were most important? The size of the enhancers were generally larger than the 1000 bp window that defined the DMRs. Therefore, we chose to calculate the exact methylation levels for the entire enhancer region for all 24 enhancers in all 11 CRC cell lines (Supplementary Table 1). Using the methylation values for each enhancer region in the 11 CRC cells, we performed a two-tailed t-test to ensure that there is a significant difference in methylation states of the 24 enhancers in the CIMP and non-CIMP cell lines (Supplementary Table 1). Out of the 24, only 3 enhancers were found to have a statistically significant difference ($p<0.05$) in methylation state between the CIMP and non-CIMP groups. These 3 human enhancer elements are **1702** [chr1:18,958,671-18,960,284 (GRCh37/hg19)], **285** (chr10:102,414,915-102,415,578) and **1944** (chr15:65,377,669-65,381,418). Thus, in this way, we chose 3 differentially methylated enhancers (DMEs) for further study. These 3 DMEs are hypermethylated in the CIMP cell lines and hypomethylated in the non-CIMP group (Supplementary Table 1).

DMRs show significant correlation between enhancer and other gene regulatory regions

We paired each enhancer with all intrachromosomal DMRs and performed a Pearson Correlation test. For the enhancer element 1702, there were 2 DMRs (chr1:228194001-228195000 and chr1:240118001-240119000) with methylation states that correlated strongly with the enhancer's methylation state (Pearson Correlation Coefficient, $-0.85 \geq r \geq 0.85$, FDR adjusted p-value ≤ 0.05). Of these 2 DMRs, the region, chr1:228194001-228195000 (which we will refer to as "DC1A" from here on), was found to overlap the promoter of the *Wnt3A* gene (Table 1). We define here the promoter as the region 1500 bp upstream of the gene's transcription start site (TSS). The correlation of methylation states between element 1702 and DC1A is a positive one (Correlation coeffiecient, $r = 0.853$, FDR-adjusted p-value = 0.02) and therefore, DC1A, like 1702, is hypermethylated in the CIMP cell lines and hypomethylated in the non-CIMP group (Supplementary Table 2). The Wnt/beta-catenin signaling pathway has previously been associated with tumor progression in CRC [45–47] and therefore, it is not surprising to find this association of enhancer 1702 with the gene of a signaling pathway that initiates the carcinogenic process in CRC.

The second DMR (chr1:240118001-240119000) (which we will refer to as "DC1B") also had a methylation state that correlated strongly with those of element 1702 ($r = -0.855$, FDR-adjusted p-value = 0.02) (Supplementary Table 2). However, DC1B is located in an intergenic region. Previous studies by the ENCODE Consortium have shown this to be a DNase hypersensitive area, an indication that this is a gene regulatory region since such regulatory regions tend to be DNase-sensitive [23, 48] (Figure 1A). Since the discovery of DNase hypersensitive sites 30 years ago, they have been used as markers of regulatory DNA regions [23]. In fact, Caco-2 is among the 9 cells that have shown this region to be DNase-sensitive and therefore suggesting that this is potentially a regulatory region in CRC cells. A closer look shows that DC1B is 39,251 bp downstream from the muscarinic 3 cholinergic receptor gene, *CHRM3* (Table 1). The human HT29 and H508 colon cancer cells have been shown to express *CHRM3* and this muscarinic receptor is believed to be responsible for CRC cell migration and invasion [49–51]. In a cell invasion model, acetylcholine-induced HT29 cell invasion could be blocked by atropine [49]. Interestingly, there is evidence to also suggest an interplay of Chrm3 and beta-catenin signaling which is responsible for intestinal mucosal differentiation and neoplasia [51].

The methylation state for enhancer element 1944 correlated well with only one intrachromosomal DMR (chr15: 65628001- 65629000; which we will refer to as "DC15A") ($r = -0.949$, FDR-adjusted p-value = 0.0019) (Supplementary Table 2). DC15A overlaps intron 3, exon 4 and intron 4 of the *IGDCC3* gene (Table 1). A detailed look at the methylation profile for the region covered by DC15A shows that most mapped reads are found between exon 4 and intron 4 (Supplementary Figure 1). Again, the ENCODE Consortium has identified *IGDCC3* exon 4 and intron 4 to contain DNase I hypersensitive sites (Figure 1B) [23, 48]. Moreover, this region was also found to contain binding sites for various transcription factors (namely, SIN3A, CTCF and RAD21) through ENCODE's transcription factor ChIP-seq. The relationship between methylation and gene expression is complex, with high levels of gene expression often associated with low promoter methylation [52] but elevated gene body methylation [53], and the causality relationships have not yet been determined. Therefore, this correlative relationship between element 1944 and DC15A may yet represent a means by which element 1944 exerts control of *IGDCC3* expression. *IGDCC3* (also known as *PUNC*) is most similar to the *Deleted in Colorectal Cancer* gene *DCC*, with 41, 42, and 47% identity in the second, third and fourth lg domains, respectively [54–56]. Interestingly, *IGDCC3* is also one of eight genes whose expression patterns strongly overlapped with known regions of high Wnt activity during embryonic development [57, 58].

For enhancer element 285, we could find no statistically significant correlations with any intrachromosomal DMRs.

Recurring enhancer-gene methylation signatures found in an independent set of human CRCs

We reanalyzed data from a recent MethylCap-seq of tissues from 24 human patients with primary sporadic CRC (GEO Series: GSE39068) [59]. Here, we aligned the reads with Bowtie 2 [60] and used MEDIPS [61] for analysis of the data. We obtained the methylation levels of the regions covered by enhancers 1702 and 1944, and DC1A, DC1B and DC15A from MEDIPS and performed a Pearson Correlation test to validate the correlations that we observed in our group of 11 CRC cells.

The methylation levels of enhancer 1702 and DC1A again showed good correlation across the 24 human CRC tissues ($r = 0.634$, p-value = 0.00088) (Supplementary Table 3). This finding definitely supports our observation in the 11 CRC cell lines and suggests this correlated methylation pattern within CRC cells is maintained because there is a functional relationship between enhancer 1702 and DC1A (within the *Wnt3A* promoter). A previous study showed that correlation between enhancer and promoter histone modification patterns can be used to infer their interactions [39]. As such, it is possible that other correlated epigenetic changes, such as DNA methylation, between enhancers and promoters may also be indicative of their interaction, especially when such patterns are well conserved within a particular cell type. Enhancer 1702 and DC1A are separated by a distance

Table 1: Summary of information of related enhancer-DMR pairs

Enhancer (VISTA)	Chromo-some	Enhancer's Chromosomal Coordinates	DMR	DMR's Chromosomal Coordinates	Overlapping or Closest Gene (with/to DMR)	Gene Accession	Gene's Chromosomal Coordinates	Strand (Gene)	Distance of Enhancer from Gene TSS (bp)
1702	1	18958671-18960284	DC1A	228194001-228195000	Homo sapiens wingless-type MMTV integration site family, member 3A (*WNT3A*)	NM_033131	228194722-228248972	+ve	209,236,051
1702	1	18958671-18960284	DC1B	240118001-240119000	Homo sapiens cholinergic receptor, muscarinic 3 (CHRM3)	NM_000740	239549876-240078750	+ve	568,125
1944	15	65377669-65381418	DC15A	65628001-65629000	Homo sapiens immunoglobulin superfamily, DCC subclass, member 3 (*IGDCC3*)	NM_004884	65619464-65670378	-ve	288,960

Enhancer-DMR pairs that have a high degree of correlation in their DNA methylation patterns are shown in this table. For the purpose of this study, we have annotated the differentially methylated regions (DMRs) as DC1A, DC1B and DC15A. Two DMRs, DC1A and DC15A, were found to overlap genes, *Wnt3A* and *IGDCC3*, respectively, while the third DMR, DC1B, was found to be in close proximity to the gene, CHRM3. DC1B is located downstream of CHRM3 and so, although it is 568,125 bp from CHRM3's transcription start site (TSS), it is, in reality, only 39,251 bp from the 3'-end of this gene.

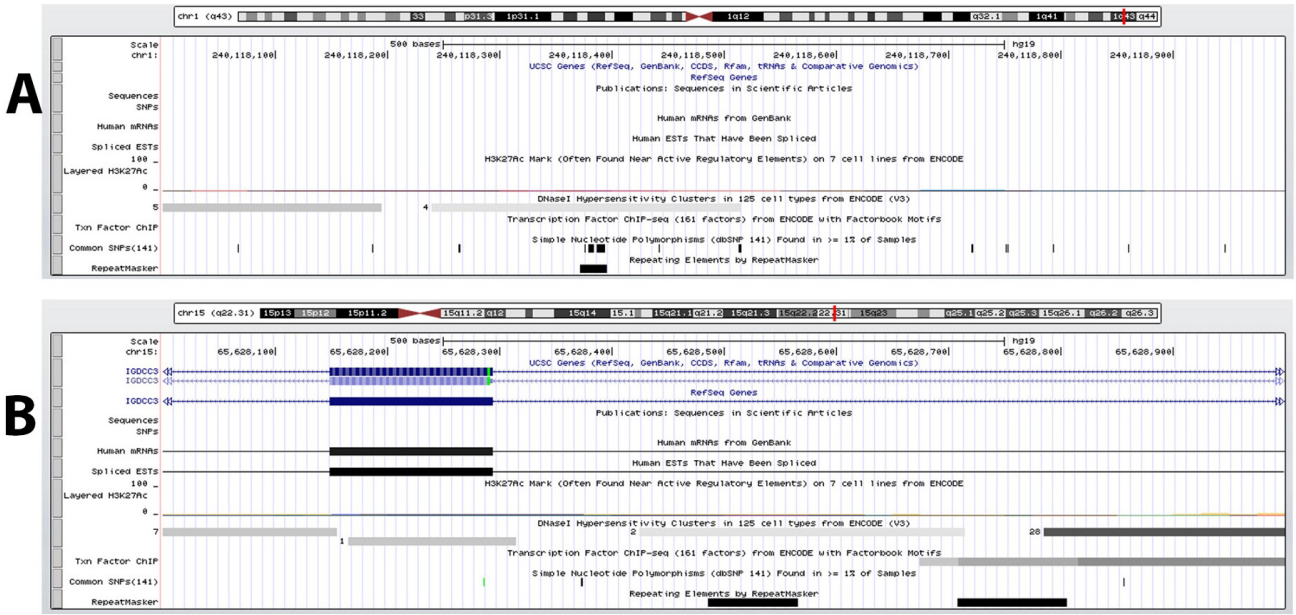

Figure 1: DNase hypersensitive sites identified within DMRs. A. DC1B (chr1:240118001-240119000), identified as a DMR between CIMP and non-CIMP cells and located in an intergenic region downstream of the *CHRM3* gene, is shown to possess DNase hypersensitive sites in 9 separate cell types by the ENCODE Consortium. **B.** DC15A (chr15: 65628001- 65629000) overlaps intron 3, exon 4 and intron 4 of the *IGDCC3* gene and has also be shown by ENCODE to be a highly DNase sensitive site by 38 cell types. Although both these regions are not located within a promoter region but these evidence indicate that DC1B and DC15A are regulatory DNA regions.

of 209 Mb. For a long time, the greatest known distance between an enhancer (ZRS) and a gene (*Shh*) was about 1Mb [25] but more recently, lineage-specific enhancers were discovered about 2Mb from the *Myc* promoter [62]. There is increasing evidence to show that distal regulatory elements can function over long distances and even on a different chromosome from their target genes [38, 63, 64]. A closer look at the data from a recent ChIA-PET assay [38] shows that of the 40,000 RNA polymerase II-bound interaction pairs discovered in this study, at least about 250 interacting pairs of loci were separated by a distance of greater than 100 Mb in both embryonic and neural stem cells (FDR ≤ 0.05). Thus, although uncommon, such long-range interactions spanning more than 100 Mb are indeed possible. A second plausible explanation for the well-correlated methylation patterns of enhancer 1702 and DC1A could also be that enhancer 1702 controls the expression of a gene or set of genes that have a functional physiological relationship with *Wnt3A* and thus, to ensure a coordinated expression of functionally related genes, the regulatory elements for these genes are coordinately regulated by methylation.

The methylation levels of enhancer 1944 and DC15A showed good-to-moderate correlation across the 24 human CRC tissues (r = 0.493, p-value = 0.014) (Supplementary Table 3). DC15A is located within the gene body of *IGDCC3* and enhancer 1944 is 238,046 bp downstream of *IGDCC3*. Enhancer 1944's proximity to the *IGDCC3* gene and its coordinated methylation with DC15A within *IGDCC3* leads us to believe that enhancer 1944 controls the transcriptional regulation of *IGDCC3*.

We found no correlation in the DNA methylation pattern between enhancer 1702 and DC1B across the 24 human CRC tissues (r=0.089, p-value = 0.679) (Supplementary Table 3). In short, of the 3 enhancer-gene pairs, we were able to validate all but one with a second independent dataset.

Conservative motifs within enhancers and DMRs suggest functional dependencies

We analyzed the sequences of enhancers 1702 and 1944 along with DMRs, DC1A and DC15A, (see Supplementary File 1 for FASTA sequences) for conserved motifs because we believe that these loci with well-correlated DNA methylation patterns across colon cancer phenotypes must play a functional role in the transcriptional regulation of genes responsible for these phenotypes. Studies have previously shown that functionally-related genes are regulated through a unique combination of conserved transcription factor binding sites (TFBSs) [65–67]. Using MEME (Multiple EM for Motif Elicitation) [68], a Web service available on MEME suite [69], two motifs (p-values ranging from 2.33×10^{-7} to 3.45×10^{-17}) were found consistently in all 4 sequences (Figure 2A). For each motif, we used TOMTOM [70] to search against the JASPAR CORE collection of vertebrate TFBS motifs [71] for similarities to known TFBSs. The results of TOMTOM offers an E-value - the expected number of times that the given query would be expected to match a target as well or better than the observed match in a randomized target database of the given size – and we chose a cutoff of E-value < 1, as previously suggested by Vandenbon *et al.* [67].

The first motif identified by MEME (which we will refer to as "M1") is only 12 bp in length (Figure 2A). A detailed bioinformatics analysis of M1 using TOMTOM showed it to be a good match to the SRY (p-value = 0.0025; E-value = 0.52) and FOXP2 (p-value = 0.0044; E-value = 0.91) binding sites (Figure 2B). There is evidence to indicate that there is some interplay between FOXP2 and Wnt signaling [72, 73]. Promoters of genes of the Wnt signaling pathway were among 100 most significantly-enriched by FOXP2 ChIP-chip experiments performed on the SH-SY5Y neuronal cell line [73]. Conversely lef1, a member of the Lef/Tcf family of transcription factors activated by Wnt signaling, has also been shown to regulate *FOXP2* during embryogenesis, and novel *FOXP2* enhancers were also found to be lef1-dependent [72]. There are also several studies showing that SRY can antagonize the Wnt/beta-catenin transcriptional activities and repress Wnt target genes during differentiation and development [74–76]. It may be possible that SRY or FOXP2 binds to this site under different cellular conditions to elicit either a repression or expression of Wnt signaling components and other functionally-related genes that contribute to the CIMP or non-CIMP phenotype in CRC.

The second motif (which we will call "M2") is 40 bp long (Figure 2A). An analysis with TOMTOM shows that this motif contains a putative binding site for Tcf3 (p-value = 0.0029; E-value = 0.59) (Figure 2C). Tcf3 is believed to play a role in Wnt-stimulated self-renewal of pluripotent mouse embryonic stem (ES) cells because it was found to bind the same genes as the core stem cell self-renewal circuit transcription factors, Oct4, Sox2 and Nanog [77]. It is generally accepted that the accumulation of stabilized beta-catenin in the presence of the Lef/Tcf family of transcription factors results in their translocation to the nucleus where they activate Wnt-responsive genes [78]. A recent study showed that genetic ablation of Tcf3 in ES cells replaced the requirement of exogenous *Wnt3A* for self-renewal of ES cells, which the authors suggest demonstrates that inhibition of Tcf3-repressor is the downstream effect of Wnt signaling [79]. Our analysis, however, suggest that Tcf3 may also be able to act upstream through its binding of the *Wnt3A* promoter and other related enhancers to affect the expression of *Wnt3A* and other physiologically-related genes. The role of *Wnt3A* in stimulating self-renewal (shown in pluripotent ES cells) could also therefore suggest its role in tumorigenesis in CRC.

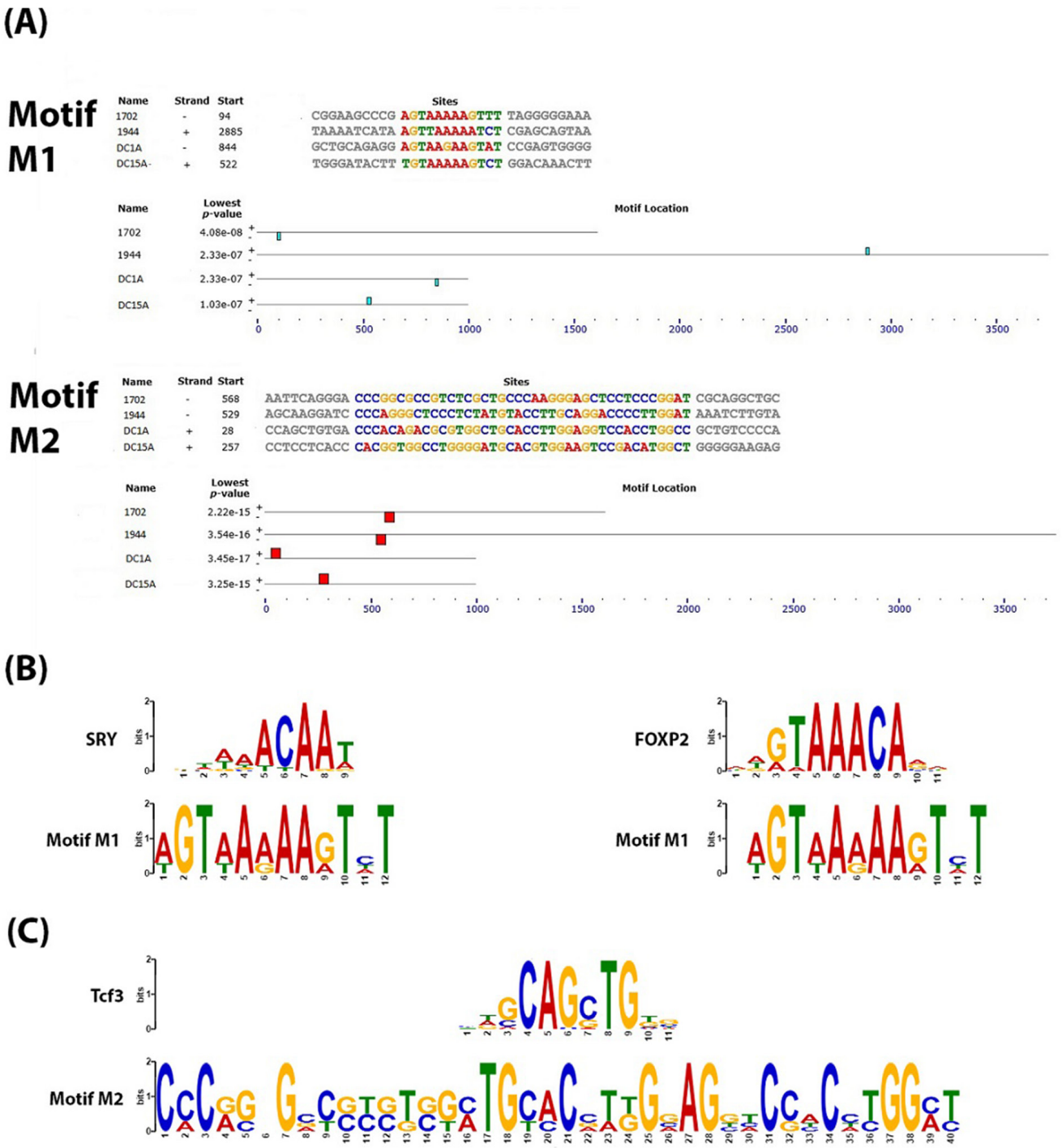

Figure 2: Conserved motifs identified within the DNA sequences of the enhancer-DMR pairs. A. Output from an MEME analysis showing the relative location of 2 conserved motifs which we called "M1" and "M2" within the DNA sequences of the enhancers and DMRs. **B.** Output of TOMTOM analysis of M1 showing similarity of this motif to known validated transcription factor binding sites for SRY and FOXP2 (E-value >1) **C.** Output of TOMTOM analysis of M2 showing this motif to contain the transcription factor binding site for Tcf3 (E-value >1).

CONCLUSION

We took a novel approach to analyze DNA methylation patterns in CRCs and our analysis revealed a strong correlation in DNA methylation patterns between specific enhancer-gene pairs that are distinct between phenotypes. To the best of our knowledge, this is the first study to show specific DNA methylation of enhancers and genes that are coordinately regulated in a cell-specific manner. We observed that both element 1702 and DC1A are more highly methylated in CIMP than in non-CIMP cells. DC1A lies within the promoter of *Wnt3A*. On the other hand, the DNA methylation pattern for enhancer 1944 and DC15A is inversely related: in CIMP cells, enhancer 1944 is highly methylated and DC15A is unmethylated while the opposite is true in non-CIMP cells. DC15A lies with the gene body of *IGDCC3*, spanning specifically, intron 3, exon 4 and intron 4. So, what does all this mean? The relationship between methylation and gene expression is complex, with high levels of gene expression often associated with low promoter methylation [52] but elevated gene body methylation [53]. Thus, we postulate from these DNA methylation patterns that perhaps *Wnt3A* and *IGDCC3* are lowly or not expressed in the CIMP phenotype. *IGDCC3* (also known as *PUNC*) is most similar to the *Deleted in Colorectal Cancer* gene *DCC*, with 41, 42, and 47% identity in the second, third and fourth Ig domains, respectively [54–56]. Interestingly, *IGDCC3* is also one of eight genes whose expression patterns strongly overlapped with known regions of high Wnt activity during embryonic development [57, 58]. Thus, it is perhaps not surprising if the expression of *IGDCC3* is closely regulated with *Wnt3A* in CRCs with low expression seen in CIMP phenotype and high expression observed in non-CIMP phenotypes.

Enhancer 1702 and *Wnt3A* are separated by a distance of more than 200 Mb while enhancer 1944 is 238 Kb downstream of *IGDCC3*. As these specific enhancer-gene pairs are coordinately regulated through DNA methylation, it would be natural to ask if these enhancers are directly responsible for regulating these genes. The record for the greatest distance between an

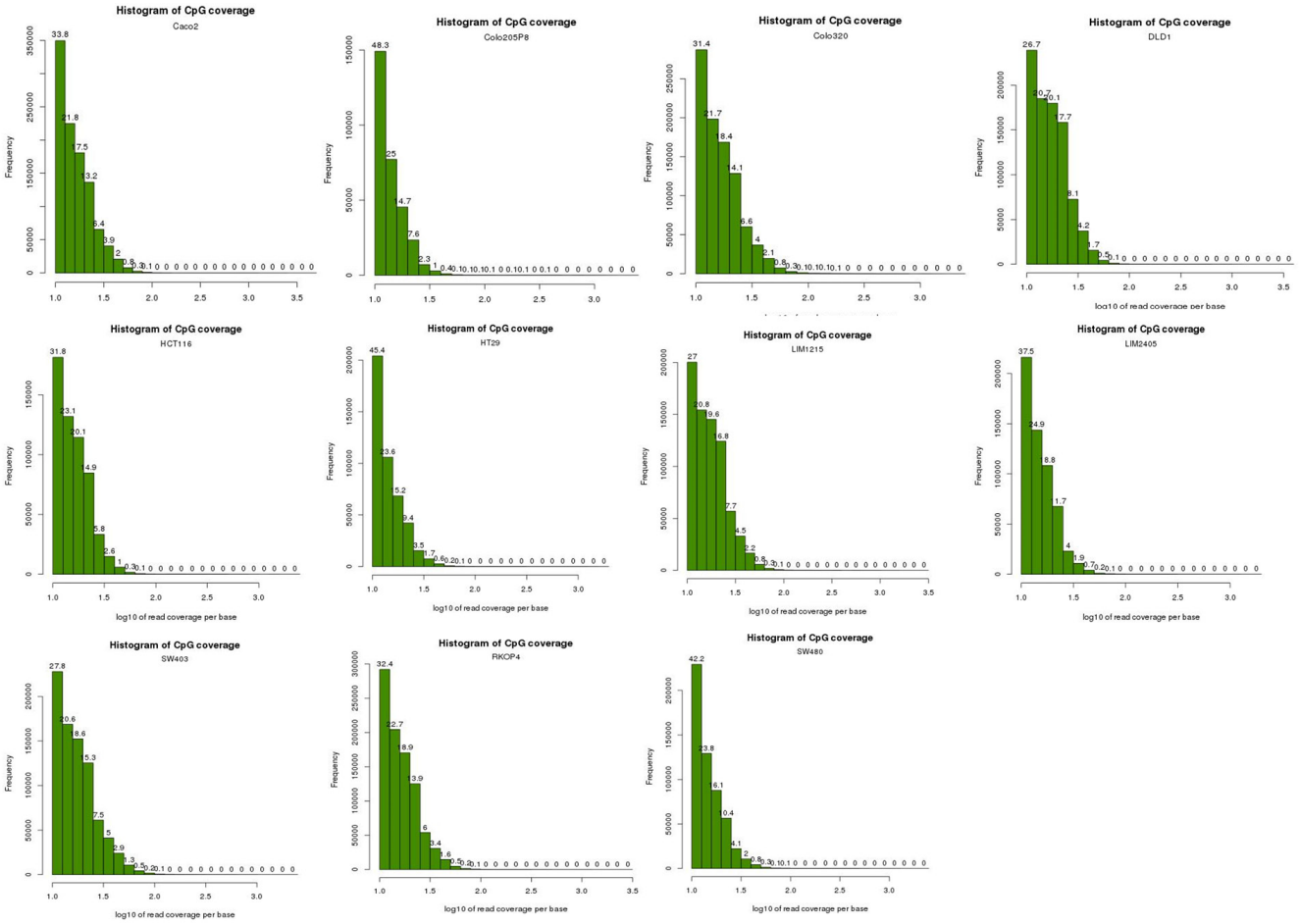

Figure 3: Histogram plots of read coverage per base for the 11 cell lines. These histogram plots show that the NGS data for the 11 cell lines do not suffer from PCR amplification bias. Experiments that suffer from PCR duplication bias will have a secondary peak on the right-hand side of the histogram. As no secondary peak is observed in our data for all 11 cell lines, this indicates that our data do not suffer from PCR duplication bias.

enhancer and a gene was held by the ZRS enhancer and *Shh* gene (about1Mb) [25] until recently, when lineage-specific enhancers were discovered about 2Mb from the *Myc* promoter [62]. There is increasing evidence to show that distal regulatory elements can function over long distances and even on a different chromosome from their target genes [38, 63, 64]. A close look at the data from a recent ChIA-PET assay [38] shows that of the 40,000 RNA polymerase II-bound interaction pairs discovered, at least about 250 interacting pairs of loci were separated by a distance of greater than 100 Mb in both embryonic and neural stem cells (FDR $\leq$ 0.05). Thus, although uncommon, such long-range interactions spanning more than 100 Mb are indeed possible. A second plausible explanation for the well-correlated methylation patterns could be that enhancer 1702 and 1944 control the expression of a gene or set of genes that have a functional relationship with *Wnt3A* and *IGDCC3*, respectively. Thus, to ensure a coordinated expression of functionally related genes, the regulatory elements for these genes are coordinately regulated by methylation.

Normally, enhancer regions are predicted by DNAse I hypersensitivity. However, it generally requires a considerable effort to identify just what genes are regulated by these enhancers. If the former argument is true and enhancer 1702 does regulate the expression of *Wnt3A*, then this would be a very exciting find not only because it would be the longest known distance between an interacting enhancer and gene but it also means that by studying the correlation of epigenetic changes between enhancers and genes, one would be able to elucidate the interaction of enhancer-gene pairs.

Gene expression is commonly used as a biomarker to identify various clinical phenotypes. Current classification of CIMP/non-CIMP is based on either the CIMP panel suggested by Issa [80] or by Weisenberger et al. [81] and both panels have selected a subset of genes whose methylation marks are supposed to represent the CIMP phenotype. Through this study, we observe that enhancer-gene pair methylation can also be an effective method for delineating phenotypes and therefore, can be used as a clinical biomarker. The various panels of methylation markers introduced to delineate CIMP and non-CIMP [2, 3, 80, 81] don't always agree on what is CIMP and what is non-CIMP since the methylation markers that were selected for each panel are slightly different. Furthermore, there is also the problem of "grey areas" when not all the markers on a panel are found to be methylated. Not surprisingly, using a combination of two features (the methylation of the enhancer and of the gene) may prove to be a much better predictor of an outcome than using just one feature. We show that by looking at just the methylation level of 1702 and *Wnt3A* and/or 1944 and *IGDCC3*, these methylation signatures can potentially be used to delineate CIMP from non-CIMP CRCs.

Finally, mutations in *APC* and other genes of the Wnt pathway are characteristics of the chromosomal instability (CIN) phenotype of CRC [82]. Here, our results suggest that the Wnt pathway is also responsible in defining the CIMP/non-CIMP phenotype through epigenetic events involving *Wnt3A*. Up until now, involvement of the Wnt pathway in CRC has only been suggested for the CIN phenotype.

MATERIALS AND METHODS

Cell Line DNA samples

DNA samples from LIM1215 and LIM2405 colorectal cancer cell lines were provided by John Mariadason. The remaining DNA samples were extracted from cell lines obtained from American Type Culture Collection (Manassas, VA) and cultured under recommended conditions. Genomic DNA was extracted using the DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions.

Methyl-binding domain-bisulfite DNA sequencing (MethylCap-BS-seq)

Methylated DNA was captured using the MethylCap kit (Diagenode, Liege, Belgium) according to manufacturer's instructions. Briefly, 2 ug of genomic DNA was randomly sheared by sonication using a Bioruptor UCD-200 (Diagenode) to generate DNA fragments peaking around 300bp. DNA was incubated with H6-GST-MBD (Methyl-DNA Binding Domain) protein and rotated at 4oC for 2 hours before adding magnetic beads. Following an additional one hour incubation, the beads were isolated and subjected to serial washing before the captured methylated DNA was eluted from the beads with low-salt, medium-salt and high-salt buffers in the kit. The medium salt and high salt eluate was combined for each sample and used for bisulfite sequencing. To construct a DNA library for bisulfite sequencing, the Illumina ChIP seq sample preparation kit (Illumina, San Diego, CA) with modification was used. Briefly, methylation enriched DNA was treated as per manufacturer's instruction before ligation with methylated adaptor oligonucleotides (Illumina) instead of the regular unmethylated adaptor oligonucleotides in the kit. DNA was then bisulfite treated using the EZ DNA methylation-gold kit (Zymo Research, Irvine, CA) before being amplified using Illumina PE 1.0 and PE 2.0 primers for 14 cycles. Constructed libraries were quantified using the Illumina library quantification qPCR protocol and sequenced on Genome Analyzer IIx (Illumina) at 36bp.

Bisulfite-treated DNA methylation sequence analysis

Sequencing reads were filtered from the adapter sequences using the wrapper script, Trim Galore! [83] which automates quality and adapter trimming through the use of Cutadapt [84] and FastQC [85]. We performed

bisulfite-treated read alignment to hg19 genome using the Bismark Bisulfite Mapper (v0.13.0) alignment software with default parameters, with only uniquely mapped reads kept for DNA methylation calling [42]. The reads from the 11 cell lines were divided into two groups of 6 CIMPs (Colo205, DLD1, HCT116, HT29, LIM2405, RKO) and 5 non-CIMPs (Caco-2, Colo320, LIM1215, SW403, SW480) and input to methylKit [43]. Sequence reads were not de-duplicated to remove potential PCR duplicates because a histogram plot of read coverage per base for each of the 11 cells (Figure 3) indicate that there is no PCR duplication bias present - experiments suffering from PCR duplication bias will have a secondary peak towards the right hand side of the histogram. Furthermore, Bainbridge et al. [86] had previously demonstrated in their study that de-duplicating reads that share the same coordinates could result in a loss of as much as 20% of actual reads that do not represent PCR duplicates for single-ended reads. In addition, a study reported by Illumina (unpublished) suggests that PCR duplicates may represent as little as 1% of all reads. In this study, scientists at lllumina prepared paired-end libraries using Illumina MID-tagged adapters, but instead of a finite set of known MID sequences, these adapters were constructed with random bases where the barcode would be. Thus, for each cluster they had three data points to compare: reads 1 and 2 and their respective alignment positions on the reference genome, plus the random 6bp sequence in the MID position. A read would need to match all three of these to be called a PCR duplicate. When they added these random tags, they found that the number of identified duplicates dropped from 8% to ~1%. Thus, based on our analysis and these studies, we chose not to de-duplicate the reads in our data. methylKit models the methylation per CpG site within a logistic regression between the CIMP and non-CIMP groups. A sliding linear model (SLIM) method is used to determine q-values from *p*-values to correct for multiple hypothesis testing [87] and thus, enabled us to determine the statistically significant DMRs (1000bp windows) between these 2 groups of cells. Methylation values for CpG dinucleotides are calculated by dividing number of methylated Cs by total coverage on that base. The methylation value calculated in this way is sometimes referred to as the beta-value (β-value). Methylation levels for DMRs or enhancers were calculated by taking the mean methylation of CpG dinucleotides within these regions. Intrachromosomal enhancer-to-DMR Pearson Correlation and other statistical calculations were performed in Microsoft Excel. We judged the strength of the correlation based on the following scale: r from +0.5 to +1.0 and from -0.5 to -1.0 constitutes a good to excellent correlation, while r from ±0.49 to 0 was considered moderate to nil. *p*-values were FDR-adjusted with the p.adjust package in R.

MethylCap-seq DNA methylation analysis

We downloaded the MethylCap-seq data for 24 human patients with primary sporadic CRC (GEO Series: GSE39068) [59] from Gene Expression Omnibus (GEO). The short reads from this study was aligned with Bowtie 2 [60] against the hg19 genome and we used MEDIPS [61] to obtain the methylation levels (measured by reads per kilobase per million mapped reads, rpkm) of the desire regions. Bowtie 2 / MEDIPS was used in this case because these are not bisulfite-treated sequences and it would be inappropriate to use the Bismark / methylKit method. Again, Pearson Correlation calculations were performed in Microsoft Excel.

Motif analysis

The DNA sequence for enhancers 1702 and 1944 along with DMRs, DC1A and DC15A, were entered into MEME (Multiple EM for Motif Elicitation) [68], a Web service available on MEME suite [69]. MEME can be used to discover motifs in a group of related nucleotide or peptide sequences. A MEME motif is a sequence pattern that occurs repeatedly in one or more sequences in the input group. MEME can be used to discover novel patterns because it bases its discoveries only on the input sequences and not on any prior knowledge (such as databases of known motifs). In this case, only motifs that were present in all 4 DNA sequences were taken for further analysis in TOMTOM [70]. TOMTOM was used to search against the JASPAR CORE collection of vertebrate TFBS motifs [71] for similarities of the motifs identified by MEME, M1 and M2, to known TFBSs.

ACKNOWLEDGMENTS

We wish to thank the lab of Dr. Richie Soong for sharing the DNA methylation data of the 11 CRC cell lines and especially, to Dr. Mengchu Wu for her work in generating this NGS data.

COMPETING INTERESTS

The authors declare that they have no competing interests.

Author contribution

A.C conceived the project, designed the experiment, analyzed the data and drafted the manuscript. J.X.T. performed the experiment. K.H.K.B. provided analysis, technical input and feedback on the draft manuscript.

REFERENCES

1. Toyota M, Ahuja N, Ohe-Toyota M, Herman JG, Baylin SB, Issa JP. CpG island methylator phenotype in colorectal cancer. Proceedings of the National Academy of Sciences of the United States of America. 1999; 96:8681-8686.
2. Ahmed D, Eide PW, Eilertsen IA, Danielsen SA, Eknaes M, Hektoen M, Lind GE, Lothe RA. Epigenetic and genetic features of 24 colon cancer cell lines. Oncogenesis. 2013; 2:e71.
3. Ogino S, Kawasaki T, Kirkner GJ, Kraft P, Loda M, Fuchs CS. Evaluation of markers for CpG island methylator phenotype (CIMP) in colorectal cancer by a large population-based sample. The Journal of molecular diagnostics. 2007; 9:305-314.
4. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975; 187:226-232.
5. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenetics and cell genetics. 1975; 14:9-25.
6. Illingworth RS, Bird AP. CpG islands—'a rough guide'. FEBS letters. 2009; 583:1713-1720.
7. Saxonov S, Berg P, Brutlag DL. A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. Proceedings of the National Academy of Sciences of the United States of America. 2006; 103:1412-1417.
8. Bird A, Taggart M, Frommer M, Miller OJ, Macleod D. A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. Cell. 1985; 40:91-99.
9. Tykocinski ML, Max EE. CG dinucleotide clusters in MHC genes and in 5' demethylated genes. Nucleic acids research. 1984; 12:4385-4396.
10. Feinberg AP. Phenotypic plasticity and the epigenetics of human disease. Nature. 2007; 447:433-440.
11. Portela A, Esteller M. Epigenetic modifications and human disease. Nature biotechnology. 2010; 28:1057-1068.
12. Jones PA, Laird PW. Cancer epigenetics comes of age. Nature genetics. 1999; 21:163-167.
13. West AG, Fraser P. Remote control of gene transcription. Human molecular genetics. 2005; 14 Spec No 1:R101-111.
14. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. Cell. 2011; 144:327-339.
15. Calo E, Wysocka J. Modification of enhancer chromatin: what, how, and why? Molecular cell. 2013; 49:825-837.
16. Levine M, Tjian R. Transcription regulation and animal diversity. Nature. 2003; 424:147-151.
17. Ong CT, Corces VG. Enhancer function: new insights into the regulation of tissue-specific gene expression. Nature reviews Genetics. 2011; 12:283-293.
18. Panne D. The enhanceosome. Current opinion in structural biology. 2008; 18:236-242.
19. Spitz F, Furlong EE. Transcription factors: from enhancer binding to developmental control. Nature reviews Genetics. 2012; 13:613-626.
20. Xie W, Ren B. Developmental biology. Enhancing pluripotency and lineage specification. Science. 2013; 341:245-247.
21. Bernstein BEB, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., Epstein CBF, S., Harrow, J., Kaul, R. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57-74.
22. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, Barrera LO, Van Calcar S, Qu C, Ching KA, Wang W, Weng Z, Green RD, Crawford GE, Ren B. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature genetics. 2007; 39:311-318.
23. Thurman RE, Rynes E, Humbert R, Vierstra J, Maurano MT, Haugen E, Sheffield NC, Stergachis AB, Wang H, Vernot B, Garg K, John S, Sandstrom R, Bates D, Boatman L, Canfield TK, et al. The accessible chromatin landscape of the human genome. Nature. 2012; 489:75-82.
24. Kleinjan DA, van Heyningen V. Long-range control of gene expression: emerging mechanisms and disruption in disease. American journal of human genetics. 2005; 76:8-32.
25. Lettice LA, Heaney SJ, Purdie LA, Li L, de Beer P, Oostra BA, Goode D, Elgar G, Hill RE, de Graaff E. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. Human molecular genetics. 2003; 12:1725-1735.
26. Sagai T, Hosoya M, Mizushina Y, Tamura M, Shiroishi T. Elimination of a long-range cis-regulatory module causes complete loss of limb-specific Shh expression and truncation of the mouse limb. Development. 2005; 132:797-803.
27. Kennedy A, Schmidt EM, Cribbs AP, Penn H, Amjadi P, Syed K, Read JE, Green P, Gregory B, Brennan FM. A novel upstream enhancer of FOXP3, sensitive to methylation-induced silencing, exhibits dysregulated methylation in rheumatoid arthritis Treg cells. European journal of immunology. 2014; 44:2968-2978.
28. Zhang B, Xing X, Li J, Lowdon RF, Zhou Y, Lin N, Zhang B, Sundaram V, Chiappinelli KB, Hagemann IS, Mutch DG, Goodfellow PJ, Wang T. Comparative DNA methylome analysis of endometrial carcinoma reveals complex and distinct deregulation of cancer promoters and enhancers. BMC genomics. 2014; 15:868.
29. Cajiao I, Zhang A, Yoo EJ, Cooke NE, Liebhaber SA. Bystander gene activation by a locus control region. The EMBO journal. 2004; 23:3854-3863.
30. Gomez-Skarmeta JL, Rodriguez I, Martinez C, Culi J, Ferres-Marco D, Beamonte D, Modolell J. Cis-regulation of achaete and scute: shared enhancer-like elements drive their coexpression in proneural clusters of the imaginal discs. Genes & development. 1995; 9:1869-1882.

31. Nickol JM, Felsenfeld G. Bidirectional control of the chicken beta- and epsilon-globin genes by a shared enhancer. Proceedings of the National Academy of Sciences of the United States of America. 1988; 85:2548-2552.
32. Spitz F, Gonzalez F, Duboule D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. Cell. 2003; 113:405-417.
33. Lower KM, Hughes JR, De Gobbi M, Henderson S, Viprakasit V, Fisher C, Goriely A, Ayyub H, Sloane-Stanley J, Vernimmen D, Langford C, Garrick D, Gibbons RJ, Higgs DR. Adventitious changes in long-range gene expression caused by polymorphic structural variation and promoter competition. Proceedings of the National Academy of Sciences of the United States of America. 2009; 106:21771-21776.
34. Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, et al. Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proceedings of the National Academy of Sciences of the United States of America. 2002; 99:7548-7553.
35. He B, Chen C, Teng L, Tan K. Global view of enhancer-promoter interactome in human cells. Proceedings of the National Academy of Sciences of the United States of America. 2014; 111:E2191-2199.
36. Sanyal A, Lajoie BR, Jain G, Dekker J. The long-range interaction landscape of gene promoters. Nature. 2012; 489:109-113.
37. van Arensbergen J, van Steensel B, Bussemaker HJ. In search of the determinants of enhancer-promoter interaction specificity. Trends in cell biology. 2014; 24:695-702.
38. Zhang Y, Wong CH, Birnbaum RY, Li G, Favaro R, Ngan CY, Lim J, Tai E, Poh HM, Wong E, Mulawadi FH, Sung WK, Nicolis S, Ahituv N, Ruan Y, Wei CL. Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature. 2013; 504:306-310.
39. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, Zhang X, Wang L, Issner R, Coyne M, Ku M, Durham T, Kellis M, Bernstein BE. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature. 2011; 473:43-49.
40. Brinkman AB, Gu H, Bartels SJ, Zhang Y, Matarese F, Simmer F, Marks H, Bock C, Gnirke A, Meissner A, Stunnenberg HG. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. Genome research. 2012; 22:1128-1138.
41. Mouradov D, Sloggett C, Jorissen RN, Love CG, Li S, Burgess AW, Arango D, Strausberg RL, Buchanan D, Wormald S, O'Connor L, Wilding JL, Bicknell D, Tomlinson IP, Bodmer WF, Mariadason JM, et al. Colorectal cancer cell lines are representative models of the main molecular subtypes of primary cancer. Cancer research. 2014; 74:3238-3247.
42. Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011; 27:1571-1572.
43. Akalin A, Kormaksson M, Li S, Garrett-Bakelman FE, Figueroa ME, Melnick A, Mason CE. methylKit: a comprehensive R package for the analysis of genome-wide DNA methylation profiles. Genome biology. 2012; 13:R87.
44. Visel A, Minovitsky S, Dubchak I, Pennacchio LA. VISTA Enhancer Browser—a database of tissue-specific human enhancers. Nucleic acids research. 2007; 35:D88-92.
45. Najdi R, Holcombe RF, Waterman ML. Wnt signaling and colon carcinogenesis: beyond APC. Journal of carcinogenesis. 2011; 10:5.
46. Kang YJ, Park HJ, Chung HJ, Min HY, Park EJ, Lee MA, Shin Y, Lee SK. Wnt/beta-catenin signaling mediates the antitumor activity of magnolol in colorectal cancer cells. Molecular pharmacology. 2012; 82:168-177.
47. Fan K, Li N, Qi J, Yin P, Zhao C, Wang L, Li Z, Zha X. Wnt/beta-catenin signaling induces the transcription of cystathionine-gamma-lyase, a stimulator of tumor in colon cancer. Cellular signalling. 2014; 26:2801-2808.
48. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012; 489:57-74.
49. Belo A, Cheng K, Chahdi A, Shant J, Xie G, Khurana S, Raufman JP. Muscarinic receptor agonists stimulate human colon cancer cell migration and invasion. American journal of physiology Gastrointestinal and liver physiology. 2011; 300:G749-760.
50. Peng Z, Heath J, Drachenberg C, Raufman JP, Xie G. Cholinergic muscarinic receptor activation augments murine intestinal epithelial cell proliferation and tumorigenesis. BMC cancer. 2013; 13:204.
51. Raufman JP, Shant J, Xie G, Cheng K, Gao XM, Shiu B, Shah N, Drachenberg CB, Heath J, Wess J, Khurana S. Muscarinic receptor subtype-3 gene ablation and scopolamine butylbromide treatment attenuate small intestinal neoplasia in Apcmin/+ mice. Carcinogenesis. 2011; 32:1396-1402.
52. Kass SU, Landsberger N, Wolffe AP. DNA methylation directs a time-dependent repression of transcription initiation. Current biology. 1997; 7:157-165.
53. Jones PA. The DNA methylation paradox. Trends in genetics: TIG. 1999; 15:34-37.
54. Salbaum JM. Punc, a novel mouse gene of the immunoglobulin superfamily, is expressed predominantly in the developing nervous system. Mechanisms of development. 1998; 71:201-204.
55. Salbaum JM. Genomic structure and chromosomal localization of the mouse gene Punc. Mammalian genome. 1999; 10:107-111.
56. Fearon ER, Cho KR, Nigro JM, Kern SE, Simons JW, Ruppert JM, Hamilton SR, Preisinger AC, Thomas G, Kinzler KW, et al. Identification of a chromosome 18q

gene that is altered in colorectal cancers. Science. 1990; 247:49-56.

57. Lickert H, Cox B, Wehrle C, Taketo MM, Kemler R, Rossant J. Dissecting Wnt/beta-catenin signaling during gastrulation using RNA interference in mouse embryos. Development. 2005; 132:2599-2609.
58. Mohamed OA, Clarke HJ, Dufort D. Beta-catenin signaling marks the prospective site of primitive streak formation in the mouse embryo. Developmental dynamics. 2004; 231:416-424.
59. Simmer F, Brinkman AB, Assenov Y, Matarese F, Kaan A, Sabatino L, Villanueva A, Huertas D, Esteller M, Lengauer T, Bock C, Colantuoni V, Altucci L, Stunnenberg HG. Comparative genome-wide DNA methylation analysis of colorectal tumor and matched normal tissues. Epigenetics. 2012; 7:1355-1367.
60. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012; 9:357-359.
61. Lienhard M, Grimm C, Morkel M, Herwig R, Chavez L. MEDIPS: genome-wide differential coverage analysis of sequencing data derived from DNA enrichment experiments. Bioinformatics. 2014; 30:284-286.
62. Shi J, Whyte WA, Zepeda-Mendoza CJ, Milazzo JP, Shen C, Roe JS, Minder JL, Mercan F, Wang E, Eckersley-Maslin MA, Campbell AE, Kawaoka S, Shareef S, Zhu Z, Kendall J, Muhar M, et al. Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. Genes & development. 2013; 27:2648-2662.
63. Li G, Ruan X, Auerbach RK, Sandhu KS, Zheng M, Wang P, Poh HM, Goh Y, Lim J, Zhang J, Sim HS, Peh SQ, Mulawadi FH, Ong CT, Orlov YL, Hong S, et al. Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell. 2012; 148:84-98.
64. Spilianakis CG, Lalioti MD, Town T, Lee GR, Flavell RA. Interchromosomal associations between alternatively expressed loci. Nature. 2005; 435:637-645.
65. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, Akalin A, Schmeier S, Kanamori-Katayama M, Bertin N, Carninci P, Daub CO, Forrest AR, Gough J, Grimmond S, Han JH, et al. An atlas of combinatorial transcriptional regulation in mouse and man. Cell. 2010; 140:744-752.
66. Chong A, Zhang Z, Choi KP, Choudhary V, Djamgoz MB, Zhang G, Bajic VB. Promoter profiling and coexpression data analysis identifies 24 novel genes that are coregulated with AMPA receptor genes, GRIAs. Genomics. 2007; 89:378-384.
67. Vandenbon A, Kumagai Y, Akira S, Standley DM. A novel unbiased measure for motif co-occurrence predicts combinatorial regulation of transcription. BMC genomics. 2012; 13 Suppl 7:S11.
68. Bailey TL, Williams N, Misleh C, Li WW. MEME: discovering and analyzing DNA and protein sequence motifs. Nucleic acids research. 2006; 34:W369-373.
69. Bailey TL, Boden M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. Nucleic acids research. 2009; 37:W202-208.
70. Gupta S, Stamatoyannopoulos JA, Bailey TL, Noble WS. Quantifying similarity between motifs. Genome biology. 2007; 8:R24.
71. Mathelier A, Zhao X, Zhang AW, Parcy F, Worsley-Hunt R, Arenillas DJ, Buchman S, Chen CY, Chou A, Ienasescu H, Lim J, Shyr C, Tan G, Zhou M, Lenhard B, Sandelin A, et al. JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic acids research. 2014; 42:D142-147.
72. Bonkowsky JL, Wang X, Fujimoto E, Lee JE, Chien CB, Dorsky RI. Domain-specific regulation of foxP2 CNS expression by lef1. BMC developmental biology. 2008; 8:103.
73. Vernes SC, Spiteri E, Nicod J, Groszer M, Taylor JM, Davies KE, Geschwind DH, Fisher SE. High-throughput analysis of promoter occupancy reveals direct neural targets of FOXP2, a gene mutated in speech and language disorders. American journal of human genetics. 2007; 81:1232-1250.
74. Li Y, Zheng M, Lau YF. The sex-determining factors SRY and SOX9 regulate similar target genes and promote testis cord formation during testicular differentiation. Cell reports. 2014; 8:723-733.
75. Lau YF, Li Y. The human and mouse sex-determining SRY genes repress the Rspol/beta-catenin signaling. Journal of genetics and genomics. 2009; 36:193-202.
76. Bernard P, Sim H, Knower K, Vilain E, Harley V. Human SRY inhibits beta-catenin-mediated transcription. The international journal of biochemistry & cell biology. 2008; 40:2889-2900.
77. Marson A, Levine SS, Cole MF, Frampton GM, Brambrink T, Johnstone S, Guenther MG, Johnston WK, Wernig M, Newman J, Calabrese JM, Dennis LM, Volkert TL, Gupta S, Love J, Hannett N, et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell. 2008; 134:521-533.
78. Lien WH, Fuchs E. Wnt some lose some: transcriptional governance of stem cells by Wnt/beta-catenin signaling. Genes & development. 2014; 28:1517-1532.
79. Yi F, Pereira L, Hoffman JA, Shy BR, Yuen CM, Liu DR, Merrill BJ. Opposing effects of Tcf3 and Tcf1 control Wnt stimulation of embryonic stem cell self-renewal. Nature cell biology. 2011; 13:762-770.
80. Issa JP. CpG island methylator phenotype in cancer. Nature reviews Cancer. 2004; 4:988-993.

81. Weisenberger DJ, Siegmund KD, Campan M, Young J, Long TI, Faasse MA, Kang GH, Widschwendter M, Weener D, Buchanan D, Koh H, Simms L, Barker M, Leggett B, Levine J, Kim M, et al. CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. Nature genetics. 2006; 38:787-793.
82. Pino MS, Chung DC. The chromosomal instability pathway in colon cancer. Gastroenterology. 2010; 138:2059-2072.
83. Krueger F. Trim Galore! [http://wwwbioinformaticsbabrahamacuk/projects/trim_galore/]. 2012.
84. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnetjournal. 2011; 17:10-12.
85. Andrews S. FastQC. [http://wwwbioinformaticsbbsrcacuk/projects/fastqc/]. 2010.
86. Bainbridge MN, Wang M, Burgess DL, Kovar C, Rodesch MJ, D'Ascenzo M, Kitzman J, Wu YQ, Newsham I, Richmond TA, Jeddeloh JA, Muzny D, Albert TJ, Gibbs RA. Whole exome capture in solution with 3 Gbp of data. Genome biology. 2010; 11:R62.
87. Wang HQ, Tuominen LK, Tsai CJ. SLIM: a sliding linear model for estimating the proportion of true null hypotheses in datasets with dependence structures. Bioinformatics. 2011; 27:225-231.