Genome-wide analysis of focal DNA hypermethylation in *IDH*-mutant AML samples

Elisabeth R. Wilson[1], Nichole Helton[1], Sharon E. Heath[1], Robert S. Fulton[3], Christopher A. Miller[1,3], Jacqueline E. Payton[2], John S. Welch[1], Matthew J. Walter[1], Peter Westervelt[1], John F. DiPersio[1], Daniel C. Link[1], Timothy J. Ley[1], and David H. Spencer[1,2,3]

[1] Department of Medicine, Division of Oncology, Section of Stem Cell Biology,

[2] Department of Pathology and Immunology,

[3] McDonnell Genome Institute,

Washington University, St. Louis, MO, USA

Running title: Methylation phenotype of *IDH*-mutant AML

The authors have no conflicts of interest relating to the work described in this manuscript.

* Corresponding author

David H. Spencer, MD, PhD

Assistant Professor

Division of Oncology

Department of Medicine

Washington University School of Medicine

Campus Box 8007

660 South Euclid Avenue

St. Louis, MO. 63110

Abstract

Altered DNA methylation is a common feature of acute myeloid leukemia (AML) and is thought to play a significant role in disease pathogenesis. Gain of function mutations in *IDH1* or *IDH2* result in widespread but highly focal regions of hypermethylation across the genome that occurs due to the production of 2-hydroxyglutarate that inhibits TET-mediated demethylation. We used whole-genome bisulfite sequencing to identify canonical regions of DNA hypermethylation that are associated specifically with *IDH1* and *IDH2* mutations in primary AML samples. Consistent with previous reports, *IDH* mutant *(IDH*^{mut}) AMLs were the most hypermethylated among all mutationally-defined AML categories analyzed. We observed notable differences in the degree of hypermethylation associated with *IDH* mutation type, with *IDH1*^{mut} AMLs having more profound hypermethylation at specific regions than *IDH2*^{mut} samples. AMLs with biallelic inactivating mutations in *TET2* displayed more modest DNA methylation changes compared to normal hematopoietic stem/progenitor cells, but methylation in these samples was increased in the *IDH*^{mut}-specific regions, providing further support that these mutations act on the same TET-mediated demethylation pathway. Focal hypermethylated regions in *IDH*^{mut} AML samples tended to occur in regions with low steady state methylation levels in normal stem/progenitor cells, which implies that both DNA methylation and demethylation pathways are active at these loci. Indeed, analysis of AML samples containing mutations in both *IDH1* or *IDH2* and *DNMT3A*^{R882} were less hypermethylated, providing evidence that focal *IDH*^{mut}-associated hypermethylation is mediated by DNMT3A. *IDH*^{mut}-specific regions of hypermethylation were largely distinct from CpG island hypermethylation, and showed a significant enrichment for putative enhancers. Analysis of three-dimensional genome interactions from primary hematopoietic cells showed that differentially methylated enhancers formed direct interactions with highly expressed genes, including *MYC* and *ETV6*. Taken together, these results suggest that focal hypermethylation in *IDH*-mutant AML cells occurs by disrupting the balance of DNA methylation and demethylation, which is highly active in genomic regions involved in gene regulation.

Introduction

DNA methylation changes in acute myeloid leukemia (AML) occur by perturbing the balance between processes that generate 5-methylcytosine (5mC) and those that remove it[1,2]. In both normal and malignant hematopoietic cells, *de novo* DNA methylation is catalyzed primarily by the DNA methyltransferase DNMT3A[3,4], which can methylate unmethylated DNA substrates. Demethylation occurs both passively during DNA synthesis in the absence of DNMT1-mediated propagation of hemi-methylated DNA, and actively via the hydroxylation of 5mC by the TET family of iron-dependent, α-KG-dependent hydroxylases and subsequent removal of the modified cytosine residues. Disruption of these opposing forces in AML can result in either increased or decreased DNA methylation compared to normal hematopoietic cells. However, the regions of altered methylation are highly focal and quite canonical among patient with similar mutations, and the regions affected by altered DNA methylation comprise less than 1% of the CpGs in the genome. We recently showed that increased methylation at CpG islands is present in nearly all AML subtypes, likely mediated by DNMT3A, and is therefore a common epigenetic change associated with leukemic transformation[5]. In addition, specific DNA methylation patterns correlate with recurrent AML mutations that directly influence DNA methylation pathways, including the *DNMT3A* ^{R882} mutation that impairs DNA methylation activity, which results in a focal, canonical hypomethylation phenotype[5].

Mutations in *IDH1* and *IDH2* are also associated with altered DNA methylation patterns[6,7] that are thought to occur by disrupting the active pathway of DNA demethylation. *IDH1* and *IDH2* encode metabolic enzymes that are normally not involved in DNA methylation, but mutant *IDH* alleles produce 2-hydroxyglutarate (2HG)[8], a neometabolite that inhibits the TET family of enzymes[9] and therefore reduces active demethylation. Indeed, analysis of DNA methylation in primary AML samples using array-based technologies and reduced-representation bisulfite sequencing (RBBS) has shown that DNA methylation is increased in specific genomic regions in samples with *IDH* mutations[6,10]. While the direct effects of these DNA methylation changes on gene regulation have been challenging to identify, the contribution of these mutations to leukemogenesis has been established in mouse models. Expression of either *IDH1*^{R132H} or

IDH2^{R140Q} blocks normal hematopoietic differentiation and promotes myeloproliferation [11–13], with full transformation into AML occurring in the presence of additional cooperating mutations[13]. These studies demonstrate the functional contribution of mutations in *IDH1* and *IDH2* to AML development, and suggest that this may occur by disrupting the normal balance between DNA methylation and demethylation in AML initiating cells.

Although previous studies using targeted DNA methylation approaches have established the general effects of *IDH1* and *IDH2* mutations on DNA[6,7,10,14], a comprehensive, genome-wide analysis of the methylation patterns in primary AML samples with these mutations has not yet been described. It is not yet clear whether differences exist between the methylation phenotypes associated with mutations in *IDH1* vs. *IDH2* (hereafter referred to as *IDH*^{mut}), or whether *IDH*^{mut}-associated methylation changes occur at distinct genomic loci from DNMT3A-mediated CpG island hypermethylation, that is common to nearly all AML samples. In addition, although *IDH* mutations are thought to cause hypermethylation via inhibition of TET enzymes, the similarity of the methylation phenotypes of these mutations not yet clear. Here, we perform a genome-wide analysis of the DNA methylation phenotype of primary AML samples with recurrent mutations in *IDH1* or *IDH2,* using whole-genome bisulfite sequencing (WGBS). We incorporate analysis of WGBS data from normal hematopoietic cells and additional primary AML samples with a range of mutational profiles to define the methylation phenotypes specifically associated with *IDH* mutations, and to determine the extent to which these patterns are present in AML samples with *TET2* mutations. We also integrate these data with data on other epigenetic modifications and three-dimensional genome architecture from primary AML samples to characterize the functional genomic elements that are affected when the balance of DNA methylation and demethylation pathways is disrupted in AML cells.

Materials and Methods

Patient samples

Primary AML samples and normal hematopoietic cells were obtained from presentation AML and normal bone marrow aspirates, respectively, following informed consent using protocol (201011766) approved by the Human Research Protection Office at Washington University as described previously ((10,11); Table S1). All experiments with primary AML samples used total bone marrow cells after estimating the leukemic purity.

Whole genome bisulfite and oxidative bisulfite sequencing and data analysis

Whole-genome bisulfite sequencing data for 38 of the samples used herein was described previously[5]. Data for XX additional samples was generated as described using 50ng of input DNA with the Swift Accel-NGS Methyl-Seq library preparation kit. Oxidative bisulfite sequencing libraries were prepared following treatment of 200ng of input DNA with the TrueMethyl oxBS module (Cambridge Epigenetix) prior to bisulfite conversion and subsequent Swift library construction. All WGBS and OxBS libraries generated for this study were sequenced on either HiSeqX or NovaSeq 6000 instruments to obtain 2x150 paired-end reads to the coverage levels shown in Table S1. Data were aligned to the hg38/GRCh38 human reference and processed into methylated and unmethylated read counts using biscuit with default parameters. DMRs were identified between AML groups using methylated and unmethylated read counts from all CpGs in each sample using DSS[15] and subsequent filtering to retain regions with >10 CpGs and a difference in mean methylation between the groups of 0.2. 5hmC values for the indicated samples were obtained by subtracting the methylation ratio from OxBS data from WGBS data at all CpGs with a coverage of 10. Statistical analysis of differential methylation in DMRs was performed on the sum of the raw methylated/unmethylated counts using methylKit. Mutation-specific DMRs were identified using a hierarchical clustering approach. This was performed on mean methylation values at each DMR for mutationally-defined AML groups in R via the Euclidean distances calculated via the matrix computation function, 'dist', with default parameters. Clustering topology was determined computationally to identify DMRs where a single outlier branch represented a specified AML group mutational subclass under study.

ChIP-seq for histone modifications

ChIP-seq was performed using ChIPmentation[16] with the following antibodies: H3K27me3 (9733S), and H3K27ac (8173S) from Cell Signaling Technology, and H3K4me1 (ab1012) from Abcam. Sequencing was performed on a NovaSeq 6000 (Illumina, San Diego, CA) to obtain ~50 million 150 bp paired-end reads. Data were analyzed via adapter trimming with trimgalore and alignment to GRCh38 using bwa mem[17]. Normalized coverage for visualization and analysis used the deeptools "bamCoverage" tool[18], and peaks were called with MACS2[19]. Statistical comparisons with DESeq2[20] used raw fragment counts at peak summits, and visualizations were prepared with Gviz[21].

TF binding enrichment and motif analysis

Transcription factor binding sites for enrichment analysis were obtained from the ENCODE Regulation tracks on the UCSC genome browser. Overlaps were computed with bedtools[22]. Motif analysis was performed with Homer using the findMotifsGenome.pl tool[23] with custom background sequence selection (otherwise default parameters) using *IDH*^{mut}-specific DMRs vs. hypermethylated DMRs identified in at least 2 AML groups.

RNA-seq analysis

RNA-seq data were obtained from the AML TCGA study[14]. TPM values were obtained using kallisto[24] and gene counts were generated using featureCounts[25].

Hi-C data analysis

Hi-C data were obtained from previous studies of 3D genome interactions in primary AML samples[26] and normal hematopoietic stem/progenitors[27]. All libraries were generated using MboI digestion prior to proximity ligation and data were analyzed using the juicer pipeline[28]. Loops were identified with HICCUPs and were subsequently analyzed using bedtools[22] to identify overlap with genes and putative enhancers. Visualizations used the GenomicInteractions and Gviz R packages[21].

Results

Primary AML samples with IDH1 or IDH2 mutations are hypermethylated at specific regions with low methylation in normal hematopoietic cells

Whole genome bisulfite sequencing (WBGS) was performed using 15 primary bone marrow aspirate samples from AML patients with canonical *IDH* mutations, including seven each with *IDH1*$^{R132C/G}$ or *IDH2*^{R140Q}, and one with an *IDH2*^{R172K} allele (referred to hereafter as *IDH*^{mut}). These data were analyzed along with WGBS data from 36 other primary AML samples representing eight mutational categories, including five with biallelic loss-of-function mutations in *TET2*, and primary CD34+ hematopoietic stem/progenitor cells (HSPCs) from six normal adult bone marrow donors[5]. All AML samples were previously sequenced using whole genome and/or whole exome sequencing[14,29], which defined the mutational landscape in these samples and confirmed that the mutations affecting DNA methylation were present in the dominant leukemic clone in the sample (Figure 1A). Importantly, the *IDH*^{mut} AML samples did not contain mutations in *DNMT3A* and *TET2,* to minimize the effects of other mutations on DNA methylation patterns. WGBS produced a mean of 12x coverage (range: 7-26) for at least 28 million CpGs in the human reference sequence across all samples. Genome-wide methylation patterns displayed the expected associations with CpG density, with high overall methylation across the genome and low methylation at CpG islands (Figures 1B and 1C). *IDH*^{mut} AML samples had genome-wide methylation levels similar to normal HSPCs and AMLs with other mutations (Figure 1B), but were slightly more methylated at CpG islands compared with normal HSPCs and AMLs with *DNMT3A*^{R882} mutations which have diminished CpG island hypermethylation (*IDH*^{mut} vs. HSPC P=4 x 10-6, *IDH*^{mut} vs. *DNMT3A*^{R882} P=4 x 10-6; see Figure 1C). CpG island methylation levels were similar between *IDH*^{mut} AML samples and AMLs with wild type *IDH1* and *IDH2* genes (mean CGI methylation of *IDH*^{mut} cases = 0.353, mean of all other groups = 0.325; P= 0.1; see Figure 1C), indicating that *IDH* mutations do not result in an exaggerated CpG island hypermethylation phenotype.

Differential methylation analysis was performed using established methods[15] for each of the mutation-defined AML groups vs. normal HSPCs to characterize mutation-specific methylation changes. This identified a mean of 4,689 (1,114-8,386) Differentially Methylated Regions (DMRs) across the eight categories (minimum methylation difference of 0.2, FDR<0.05); of these DMRs, between 6% and 97% were hypermethylated in the AML samples vs. HSPCs (see Figure 1D). As expected, AML samples with *DNMT3A*^{R882} had the most hypomethylated DMRs compared to HSPCs (97% of 5,747 DMRs), whereas *IDH*^{mut} AML samples had the most hypermethylated DMRs, with 99% of the 6,960 identified regions having a methylation level at least 0.2 higher than HSPCs. *TET2*-mutant samples were also hypermethylated, but at fewer loci, with 74% of the 2,991 identified DMRs having higher methylation in the AML samples. The fewest DMRs were identified in samples with either *IDH1* or *IDH2* mutations and *DNMT3A*^{R882}, which is consistent with previous studies suggesting that AML samples with both mutations have abrogated methylation phenotypes, compared to individually mutated samples[10]. Analysis of the DMRs identified in *IDH*^{mut} samples demonstrated that 29% were associated with promoters and 43% occurred in CpG islands, which were similar to the frequencies observed for commonly hypermethylated DMRs in other AML subtypes. Interestingly, although *IDH* mutations are thought to lead to hypermethylation by inhibiting active demethylation, most *IDH*^{mut}-specific DMRs occur in regions with low methylation in normal hematopoietic cells. For example, 72% of the *IDH*^{mut} DMRs had a mean methylation < 0.3 in both HSPCs and more mature myeloid cell populations. This suggests that increased methylation in these regions in *IDH-mutant* AML samples results from either aberrant methylation, or impaired demethylation caused by the *IDH* mutations.

IDH^{mut}*-specific methylation changes are distinct from AML-associated CpG island hypermethylation and are influenced by IDH mutation type*

We next sought to identify the methylation changes that are uniquely associated with *IDH1* and/or *IDH2* mutations to determine whether their methylation phenotypes are similar to each other and distinct from CpG island hypermethylation (Figure 2A). To this end, we performed

hierarchical clustering of the mean methylation values for each group of *IDH*^{mut} and *IDH*^{wt} AML samples at the DMRs identified above and used the clustering topology to identify regions where the single outlier branch represented the *IDH*^{mut} AML samples (Figure 2B, see also Methods). We then clustered AMLs with either *IDH1* or *IDH2* mutations with all other AMLs to identify DMRs with *IDH*^{mut}-specific methylation (samples with either *TET2* and/or *DNMT3A*^{R882} mutations were excluded given that they may share methylation changes with *IDH*^{mut} samples and lack CpG island hypermethylation[5,6], respectively). This approach identified 3,928 *IDH1*^{mut}-specific and 1,821 *IDH2*^{mut}-specific DMRs, of which 90% and 79% were hypermethylated with respect to normal HSPCs, respectively (see Figure 2C-E). Consistent with the analysis above, most DMRs displayed low methylation in normal cells, with 55% of *IDH1*^{mut}-specific and 71% of *IDH2*^{mut}-specific the loci having a methylation level <0.3 in HSPCs and mature myeloid cells (Figure 2C-E). There was extensive overlap between the DMRs specific to each mutation (94%, 5,403 of 5,749 total DMRs) and AML samples with either mutation were hypermethylated at both DMR sets (Figure 2E). However, hierarchical clustering demonstrated considerable variability in methylation levels between the *IDH1*^{mut} and *IDH2*^{mut} samples, with clear stratification of the samples by mutation type (Figure 2F). This was most striking for the *IDH2*^{mut} samples, which included three AMLs with markedly lower methylation levels at the *IDH2*^{mut}-specific DMRs (Figure 2F-G) that was not explained by mutation abundance, sample purity, or cellular maturation. Samples with *IDH2* mutations also had lower methylation levels at the combined set of *IDH1*^{mut}-specific and *IDH2*^{mut}-specific DMR loci (*IDH1*^{mut}=0.70, *IDH2*^{mut}= 0.54; p-value= 0.04), although they remained hypermethylated relative to HSPCs (Figure 2G). This difference was not related to the abundance of the mutant *IDH* allele (all samples had VAFs > 30%) and did not correlate with other recurrent mutations, including *NPM1c* (4 in *IDH1*^{mut} and 3 in *IDH2*^{mut} samples). This demonstrates that *IDH1* and *IDH2* mutations generally affect methylation at the same loci, but the level of hypermethylation depends on mutation type, and can be influenced by mutation-extrinsic factors.

To determine whether *IDH*^{mut}-specific hypermethylation is similar to, or different from, CpG island hypermethylation common in other AML subtypes, we analyzed the CpG content and

genomic features of the *IDH*^{mut}-specific DMRs. Interestingly, both *IDH1*^{mut}-specific and *IDH2*^{mut}-specific DMRs were markedly different in their CpG density and amount of overlap with genomic annotations, compared with regions with AML-associated CpG island hypermethylation. For example, the combined set of *IDH*^{mut}-specific DMRs displayed significantly lower CpG density (mean CpG density of 0.89 vs. 1.26; P<0.0001) and less overlap with annotated CpG islands (23% vs. 54%) compared to 4,573 commonly hypermethylated regions identified in at least 2 other AML mutation categories (Figure 2H-I). Promoter regions were also underrepresented among the *IDH*^{mut}-specific DMRs (21% vs 31%; see Figure 2H), consistent with depletion of promoter-associated CpG islands among *IDH*^{mut}-specific DMRs. These data suggest that *IDH*-associated hypermethylation tends to occur in regions with different genomic features compared to those affected by AML-associated CpG island hypermethylation.

Hypermethylation in TET2^{mut} *AMLs overlaps with IDH*^{mut}*-specific hypermethylation but does not phenocopy the extent of methylation changes*

We next determined whether AML samples with biallelic loss-of-function mutations in *TET2* shared similar genome-wide patterns of hypermethylation to *IDH*^{mut} AMLs. The initial comparison of the *TET2*^{mut} AMLs vs. normal HSPCs performed above resulted fewer DMRs and a lower proportion of hypermethylated regions compared to *IDH*^{mut} samples (2,512 vs. 6,523 DMRs, and 78% vs 99% hypermethylated regions, respectively), consistent with previous reports that inactivation of *TET2* has a less profound effect on DNA methylation[6,10]. Identification of *TET2*mut-specific DMRs using the clustering approach described above (with *IDH*^{mut} and *DNMT3A*^{R882} AMLs excluded from the analysis) produced only 51 *TET2*^{mut}-specific DMRs, confirming that these AMLs do not display a strong hypermethylation phenotype (Figure 3A). Although many of these DMRs were hypermethylated relative to HSPCs (31 of 51), the frequency of hypermethylated DMRs was significantly less than that observed in either *IDH1*^{mut} or *IDH2*^{mut} AML samples (60% vs 91% and 73% respectively). *TET2*^{mut}-specific DMRs were not enriched for either CpG islands or promoters, compared to the set of commonly

hypermethylated DMRs (17% of *TET2* DMRs vs. 54% generic DMRs; 11% of *TET2* DMRs vs. 31% generic DMRs), suggesting these regions are unlikely to reflect CpG island hypermethylation.

To further investigate the interaction between *IDH* mutations and *TET2*-mediated demethylation, we compared the *TET2*^{mut}-specific and *IDH*^{mut}-specific DMRs and performed oxidative bisulfite sequencing[30] to measure 5-hydroxymethylation in samples with and without biallelic inactivating *TET2* mutations. Our analysis showed that 82% (42 of 51) of the *TET2*^{mut}-specific DMRs overlapped an *IDH*^{mut}-specific hypermethylated region; the reciprocal overlap was only 2%, owing to the much larger number of *IDH*^{mut}-specific DMRs (Figure 3B). Direct analysis of methylation in *TET2*^{mut} AMLs at the combined set of 4008 *IDH*^{mut}-specific DMRs also showed that it was not significantly increased compared to HSPCs (0.34 vs 0.28; p-value=0.12; Figure 3C-D). We next analyzed levels of 5-hydroxymethylation at hypermethylated regions using paired oxidative and standard whole genome bisulfite sequencing (oxWGBS and WGBS) of primary AML samples with (N=3) and without (N=1) *TET2* mutations. Sequencing generated at least 10x coverage for the paired datasets, and subtraction of oxWGBS from WGBS data, demonstrated very low levels of 5hmC across the genomes of these samples. However, regions with *IDH*^{mut}-specific hypermethylation had significantly higher levels of 5hmC (calculated by subtracting the percent of modified cytosines in these regions in oxWGBS from WGBS data) than in constitutively methylated heterochromatin and regions with CpG island hypermethylation (Figure 3E). 5hmC levels were diminished in all three of these regions in *TET2*^{mut} AML samples, indicating that 5hmC levels in these regions are dependent on active TET2 enzyme. Together, these data provide additional support that hypermethylation in *IDH*^{mut} AML cells occurs via inhibition of TET2-mediated demethylation, and suggests that the specific regions where TET2 is active become hypermethylated in *IDH*^{mut} AMLs.

DNA hypermethylation in IDH^{mut} *AML cells requires DNMT3A*

To assess whether *de novo* DNA methylation by DNMT3A contributes to *IDH*^{mut}-associated hypermethylation, we analyzed methylation levels at *IDH*^{mut}-specific DMRs in seven AML

samples with co-occurring *IDH1* (N=5) or *IDH2* (N=2) and *DNMT3A*^{R882} mutations. We specifically selected samples with *DNMT3A*^{R882} mutations for this analysis (rather than other heterozygous missense or nonsense alleles) since R882 mutations have a dominant negative phenotype, causing a more severe hypomethylation phenotype in AML compared to other mutations[4]. Interestingly, although the *DNMT3A*^{R882}/*IDH*^{mut} double mutant AMLs were still hypermethylated at *IDH*^{mut}-specific DMRs, the degree of hypermethylation was significantly less, with 63% of these regions having significantly lower DNA methylation levels than samples with *IDH* mutations alone (Figure 4A-C). This difference was observed in double mutant samples with either *IDH1* or *IDH2* mutations. Similar findings were observed in an extended set of 6 *DNMT3A*^{R882}/*IDH*^{mut} AML samples using methylation array data from the TCGA AML study[14]. To further characterize the extent to which this interaction occurs at regions with DNMT3A-mediated methylation, we analyzed DNA methylation levels in *DNMT3A*^{R882}/*IDH*^{mut} AML samples at DMRs that are hypomethylated in AMLs with *DNMT3A*^{R882} compared to normal HSPCs, and therefore have DNA methylation that is DNMT3A-dependent[5]. Surprisingly, these regions remained methylated in the *DNMT3A*^{R882}/*IDH*^{mut} double mutant samples, with 92% of the regions having significantly higher methylation than in AMLs with *DNMT3A*^{R882} alone. In fact, 77% of the regions had methylation levels that were not significantly different from normal HSPCs (Figure 4D-F). This observation is consistent with previous studies finding that *DNMT3A* and *IDH* mutations have antagonistic effects on DNA methylation. In addition, this analysis further implies that DNMT3A-mediated methylation and TET-mediated demethylation occur at the same places in the genome. Moreover, it shows that DNMT3A is required for a fully penetrant *IDH*^{mut}-associated hypermethylation phenotype, and that loss of TET2-mediated demethylation leads to near complete abrogation the *DNMT3A*^{R882} hypomethylation phenotype.

IDH^{mut}*-specific hypermethylated DMRs are enriched for enhancers*

Given the underrepresentation of CpG islands in *IDH*^{mut}-specific DMRs, we next asked whether these regions were associated with other functional regulatory elements. We first annotated

IDH^{mut}-specific DMRs using publicly available annotations of chromatin states in HSPCs via the ChromHMM model[31]. This identified potential enhancers as the most common elements in *IDH*^{mut} hypermethylated regions, with 44% of *IDH*^{mut} DMRs overlapping regions with the "enhancer state", which is a 2-fold enrichment over regions commonly hypermethylated in AML (Figure 5A). Similar analysis performed on DMRs detected in other AML subtypes (*MLL-ELL*, *RUNX1-RUNX1T1*, *CBFB-MYH11*, *NPM1c*) showed a different distribution of chromatin states, with an underrepresentation of the enhancer state. We further defined this association using ChIP-seq to detect the enhancer-associated H3K27ac and H3K4me1 modifications using 16 primary AML samples, including 14 that were *IDH*^{wt} and 2 that were *IDH*^{mut}. Analysis of these data defined 44,762 and 6,917 consensus peaks for H3K27ac and H3K4me1, respectively. H3K27me3 ChIP-seq data from 9 AML patients was also included to define 16,417 regions that were used in combination with the active marks to assign regions to active, poised, or weak enhancer states. Analysis of the *IDH*^{mut} DMRs showed that 47% overlapped with putative enhancers, with far fewer poised and weak regions (only 3% and 1% of DMRs, respectively) (Figures 5B-C). This pattern was markedly different from that observed at commonly hypermethylated regions in AML, which showed less enrichment for active enhancers (47% vs. 13% of DMRs; 3.6-fold enrichment) and greater signal for the repressive H3K27me3 mark (Figure 5D). We also analyzed *IDH*^{mut}-specific DMRs for transcription factor (TF) binding motifs, which resulted in several motifs characteristic of hematopoietic-associated TFs that are highly expressed in AML cells, including *SPI1* and other ETS family factors, *RUNX1*, and *MYC* (Figure 5E), providing further support that *IDH*^{mut}-specific hypermethylation occurs at regions with potential regulatory activity. However, quantitative analysis of H3K27ac signal in these regions in AML samples with and without *IDH* mutations did not identify changes in H3K27ac signal (p-value= 0.24, Figure 5F), suggesting that methylation status alone does not modify the enhancer state.

IDH^{mut}*-specific enhancer regions form direct interactions with highly expressed genes in AML cells*

Given the striking enrichment of *IDH*^{mut}-DMRs for putative enhancers, we next asked whether these sequences could be involved in controlling expression of genes relevant for AML pathogenesis. To this end, we analyzed three-dimensional (3D) genome interactions generated using in situ HiC from both normal HSPCs[27] and 3 primary AML samples[26] (all of which were wild type for *DNMT3A*, *IDH1*, *IDH2*, and *TET2*). Standard loop calling strategies were used to identify high-confidence chromatin contacts in individual samples, and a union of all detected loops was used to identify genes that form direct contacts with regions of *IDH*^{mut}-specific DNA hypermethylation. This analysis demonstrated that 25% (1047/4008) of all *IDH*^{mut}-specific DMRs overlapped a loop anchor, and 32% (322/1021) of DMRs in putative enhancers showed such an overlap (Figure 6A). We next determined the set of genes whose promoters occurred in a loop anchor opposite that of an *IDH*^{mut}-specific enhancer DMR (eDMR), to identify potential targets for these candidate enhancers (Figure 6B). This analysis resulted in a total of 750 target genes, including many that are highly expressed in AML cells (Figure 6C). For example, 67% of these genes ranked in the top 50th percentile of gene expression using RNA-seq data from 113 TCGA AML samples. This hypothesis was further supported when we looked at a second set of genes opposite loop anchors harboring generically hypermethylated enhancer regions, where a more modest 51% of generically hypermethylated enhancer targets fell in the top 50th percentile of gene expression (Figure 6C). Further analysis of 3D genome interactions containing *IDH*^{mut}-specific DMRs identified known and novel enhancers of genes important in hematopoiesis and AML. This included a distal enhancer of *MYC* with an established role in hematopoietic development[32–34] (Figure 6D), and previously unreported putative enhancers that form interactions with *SRSF3* (Figure 6E), *DOT1L,* and *ETV6.* Although we did not observe significant changes in expression of these genes between *IDH*^{mut} and *IDH*^{wt} AMLs, they tended to have higher expression in AML samples compared to CD34+ HSPCs, consistent with the observed trend towards increased expression of *IDH*^{mut}-eDMR targets in AML samples (Figure 6D-E).

Discussion

Recurrent gain-of-function *IDH1* and *IDH2* mutations in AML produce the neometabolite 2-hydroxyglutarate, which inhibits enzymes involved in the oxidation and removal of methylated

cytosines, which leads to DNA hypermethylation at several thousand specific regions of the genome. In this study, we used whole genome bisulfite sequencing of primary AML samples to demonstrate that *IDH*-mutant AMLs have genome-wide levels of methylation that are similar to AMLs with other mutations, but possess thousands of regions that are uniquely hypermethylated compared to both normal HSPCs and other AML samples. These *IDH*^{mut}-specific hypermethylated regions had lower CpG density and fewer CpG islands than loci with hypermethylation in other AML types, suggesting that the genomic features that direct *IDH*^{mut}-associated hypermethylation are distinct from the very common CpG island hypermethylation in AML. *IDH2*-mutant AML samples had less pronounced hypermethylation than those with *IDH1* mutations, but both were associated with increased methylation at a highly overlapping set of loci, suggesting that they act on the same pathway—but to a different degree. Analysis of WGBS data from AMLs with biallelic inactivating mutations in *TET2* showed that these samples had a far less dramatic methylation phenotype, but were more methylated than normal HSPCs at many of the same loci. Oxidative bisulfite sequencing also demonstrated detectable levels of 5hmC in AML cells in *IDH*^{mut}-specific hypermethylated regions that was absent in AML samples with inactivating *TET2* mutations, providing additional support for convergence of *IDH1*, *IDH2*, and *TET2* mutations on the TET-mediated DNA demethylation pathway. Regions with *IDH*^{mut}-specific hypermethylation were enriched for regions with enhancer-associated histone modifications and formed direct interactions with genes that are highly expressed in AML cells. This included distal regulatory sequences that control the expression of the important AML genes *MYC* and *DOT1L,* which are relevant for AML pathogenesis.

This study provides several new insights into the dynamics of *de novo* DNA methylation and demethylation in hematopoietic cells, and the mechanisms and potential consequences of DNA hypermethylation in AMLs with *IDH* mutations. First, *IDH*^{mut}-associated hypermethylation occurs at regions with relatively low levels of DNA methylation in normal hematopoietic cells. Because mutant *IDH* alleles impair demethylation, this implies that both *de novo* DNA methylation and TET-mediated demethylation must be highly active in these regions, despite their low steady state methylation levels. This is supported by the observation that AML

samples with co-occurring *IDH* and *DNMT3A*^{R882} mutations have attenuated hypermethylation in these regions. Active remodeling of DNA methylation at certain loci has been reported in other studies, which have found that *de novo* DNA methylation and demethylation are in a dynamic equilibrium[1,2]. The presence of *IDH* mutations in AML disrupts this balance, and results in increased methylation at very specific sites in the genome.

Focal competition of methylation and demethylation activity has been previously described at enhancers, supporting our observation that hypermethylated DMRs in *IDH*^{mut} AML are enriched for enhancers. While the mechanisms involved in recruiting the DNMT3A and TET enzymes to particular functional elements remain unclear[35], this enrichment suggests that these pathways are most active at loci undergoing active epigenetic regulation. Studies in pluripotent cells have shown that TET-mediated demethylation has the highest activity at poised enhancers, which suggests that high methylation turnover may facilitate regulation in response to stimuli, including cues to undergo differentiation. Similar forces may be at play in hematopoietic stem/progenitor cells. Our analysis of 3D genome interactions involving *IDH*^{mut}-specific hypermethylated enhancers showed that these regions contact genes that are highly expressed in AML samples, including genes with higher expression in AML compared with HPSCs, providing additional support for this hypothesis.

Taken together, our results suggest that *IDH*^{mut}-associated focal hypermethylation is mediated by DNMT3A, and potential enhancer elements are often the regions that are modified. Further studies will be necessary to interrogate individual loci to determine whether altered enhancer methylation influences the expression of genes that are relevant for AML pathogenesis.

References

1. Charlton, J. *et al.* TETs compete with DNMT3 activity in pluripotent cells at thousands of methylated somatic enhancers. *Nat Genet* **52**, 819–827 (2020).

2. Ginno, P. A. *et al.* A genome-scale map of DNA methylation turnover identifies site-specific dependencies of DNMT and TET activity. *Nat Commun* **11**, 2680 (2020).

3. Challen, G. A. *et al.* Dnmt3a is essential for hematopoietic stem cell differentiation. *Nat Genet* **44**, 23 (2012).

4. Russler-Germain, D. A. *et al.* The R882H DNMT3A Mutation Associated with AML Dominantly Inhibits Wild-Type DNMT3A by Blocking Its Ability to Form Active Tetramers. *Cancer Cell* **25**, 442–454 (2014).

5. Spencer, D. H. *et al.* CpG Island Hypermethylation Mediated by DNMT3A Is a Consequence of AML Progression. *Cell* **168**, 801-816.e13 (2017).

6. Figueroa, M. E. *et al.* Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* **18**, 553–567 (2010).

7. Akalin, A. *et al.* Base-Pair Resolution DNA Methylation Sequencing Reveals Profoundly Divergent Epigenetic Landscapes in Acute Myeloid Leukemia. *Plos Genet* **8**, e1002781 (2012).

8. Losman, J.-A. *et al.* (*R*)-2-Hydroxyglutarate Is Sufficient to Promote Leukemogenesis and Its Effects Are Reversible. *Science* **339**, 1621–1625 (2013).

9. Xu, W. *et al.* Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of α-Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* **19**, 17–30 (2011).

10. Glass, J. L. *et al.* Epigenetic Identity in AML Depends on Disruption of Nonpromoter Regulatory Elements and Is Affected by Antagonistic Effects of Mutations in Epigenetic Modifiers. *Cancer Discov* **7**, 868–883 (2017).

11. Moran-Crusio, K. *et al.* Tet2 Loss Leads to Increased Hematopoietic Stem Cell Self-Renewal and Myeloid Transformation. *Cancer Cell* **20**, 11–24 (2011).

12. Sasaki, M. *et al.* IDH1(R132H) mutation increases murine haematopoietic progenitors and alters epigenetics. *Nature* **488**, 656 (2012).

13. Kats, L. M. *et al.* Proto-Oncogenic Role of Mutant IDH2 in Leukemia Initiation and Maintenance. *Cell Stem Cell* **14**, 329–341 (2014).

14. Network, C. G. A. R., TJ, L. & C, M. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. *New Engl J Medicine* **368**, 2059–2074 (2013).

15. Feng, H., Conneely, K. N. & Wu, H. A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Res* **42**, e69–e69 (2014).

16. Schmidl, C., Rendeiro, A. F., Sheffield, N. C. & Bock, C. ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* **12**, 963–965 (2015).

17. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *undefined* (2013).

18. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).

19. Zhang, Y. *et al.* Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**, R137 (2008).

20. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).

21. Hahne, F. & Ivanek, R. Methods in Molecular Biology. *Methods Mol Biology Clifton N J* **1418**, 335–351 (2016).

22. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

23. Heinz, S. *et al.* Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* **38**, 576–589 (2010).

24. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* **34**, 525–527 (2016).

25. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

26. Ghasemi, R., Struthers, H., Wilson, E. R. & Spencer, D. H. Contribution of CTCF binding to transcriptional activity at the HOXA locus in NPM1-mutant AML cells. *Leukemia* 1–13 (2020) doi:10.1038/s41375-020-0856-3.

27. Zhang, X. *et al.* Large DNA Methylation Nadirs Anchor Chromatin Loops Maintaining Hematopoietic Stem Cell Identity. *Mol Cell* **78**, 506-521.e6 (2020).

28. Durand, N. C. *et al.* Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Syst* **3**, 95–98 (2016).

29. Klco, J. M. *et al.* Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. *Jama* **314**, 811–822 (2015).

30. Kernaleguen, M. *et al.* Epigenome Editing, Methods and Protocols. *Methods Mol Biology* **1767**, 311–349 (2018).

31. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**, 2478–2492 (2017).

32. Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Gene Dev* **27**, 2648–2662 (2013).

33. Pulikkan, J. A. *et al.* CBFβ-SMMHC Inhibition Triggers Apoptosis by Disrupting MYC Chromatin Dynamics in Acute Myeloid Leukemia. *Cell* **174**, 172-186.e21 (2018).

34. Bahr, C. *et al.* A Myc enhancer cluster regulates normal and leukaemic haematopoietic stem cell hierarchies. *Nature* **553**, 515–520 (2018).

35. Weinberg, D. N. *et al.* The histone mark H3K36me2 recruits DNMT3A and shapes the intergenic DNA methylation landscape. *Nature* **573**, 281–286 (2019).

Figure 1. Genome-wide DNA methylation summary of primary samples representing seven AML mutational subtypes and normal CD34 positive HSPCs. A) Summary of mutation status in 51 primary AML patients with WGBS data. B) Average methylation levels across ~28,000,000 CpGs in normal bone marrow (NLBM) and AML subtypes harboring somatic mutations or gene fusions effecting epigenetic modifiers (Normal HSPC, n=6; IDH, n=15; TET2, n=5; DNMT3A R882, n=6; DNMT3A R882/IDH1/2, n=7; NPM1, n=4; NK, n=4; CBFB MYH11, n=3; MLLELL, n=3; RUNX1 RUNX1T1, n=3). C) Average CpG island methylation levels in NLBM and AML subtypes. D) Number of differentially methylated regions identified for each AML subtype compared with normal HSPCs. Blue bars represent regions of hypomethylation with respect to normal HSPCs while red bars represent regions of hypermethylation compared to normal HSPCs. Average number of CpGs per DMR (top panel) and DMR length (bottom panel) are shown for each AML subtype. E) Average CpG density across DMRs identified in each AML subtype.

Figure 2. Characterization of *IDH* mutation specific DMRs. A) Distribution of shared hypermethylated DMRs amongst AML subtypes where DMRs are binned based on number of AML subtypes sharing the same DMR region. B) Example of hierarchical clustering tree topology identifying an *IDH* specific DMR. Aggregate DMR methylation for individual AML subtypes are shown in bottom panel with *IDH* samples indicated in red and NLBM indicated in blue. C) Average DMR methylation across *IDH*^{mut}-specific DMRs in *IDH* mutant samples versus average methylation of all other AMLs. D) Aggregate methylation levels across all *IDH1/2*^{mut}-specific DMRs in *IDH*^{mut} and *IDH*^{wt} AML subtypes. E) Methylation locus heatmap showing average subgroup methylation values for all *IDH*^{mut}-specific DMRs, where each column is centered over the DMR with the window extending 5kb up- and down-stream the DMR center point. F) Mean DMR methylation across all *IDH1/2*^{mut}-specific DMRs (rows) in 15 individual *IDH*^{mut} cases (columns). G) Aggregate DMR methylation across 3,928 *IDH1*^{mut} specific DMRs and 1,821 *IDH2*^{mut} specific DMRs respectively. H) Fraction of functional genomic elements overlapping generically hypermethylated DMRs, *IDH1*^{mut} specific DMRs, and *IDH2*^{mut} specific DMRs. I) Distribution of CpG densities across generically hypermethylated regions in primary AML and *IDH1*^{mut}- and *IDH2*^{mut}-specific DMRs.

Figure 3. Hypermethylation in *TET2*^{mut} AMLs is subtle and overlaps a subset of *IDH*^{mut}-specific DMRs. A) Average DMR methylation across 2512 in *TET2*^{mut} DMRs called vs. C34+ HSPCs (black dots) and 52 in *TET2*^{mut}-specific DMRs (red dots) in *TET2* mutant samples versus average methylation in all other AMLs. B) Intersection of in *TET2*^{mut}-specific and *IDH1/2*^{mut}-specific DMRs. C) Aggregate methylation over *IDH*^{mut}-specific DMRs in *IDH*^{mut} and in *TET2*^{mut} AML and CD34+ HSPCs. D) Methylation locus heatmap of average subgroup methylation values for all *IDH*^{mut}-specific DMRs (rows), where each column is centered over the DMR with the window extending 5kb up- and down-stream the DMR center point. E) Average 5hmc (WGBS minus oxWGBS) levels in in *TET2*^{mut} and in *TET2*^{wt} AML samples at 4650 ChromHMM heterochromatic regions, 4586 generically hypermethylated regions, and 4008 *IDH*^{mut}-specific hypermethylated DMRs.

Figure 4. *DNMT3A*^{R882}/*IDH* double mutant AMLs exhibit an intermediate methylation phenotype between *IDH1*^{mut} hypermethylation and R882 mutant driven hypomethylation. A) Methylation locus heatmap of average subtype methylation across *IDH1*^{mut} DMRs (rows) in IDH mutant, *DNMT3A*^{R882}/IDH double mutant, and *DNMT3A*^{R882} AMLs, and CD34+ HSPCs. B) Distribution of *IDH*^{mut}-specific DMR methylation levels by AML subtype. C) Example *IDH*^{mut}-specific DMR locus within ETV6 gene demonstrating intermediate methylation phenotype of double mutant samples with respect to *IDH*^{mut} and *DNMT3A*^{R882} mutant AML. D) Methylation locus heatmap of average subtype methylation across *DNMT3A*^{R882} DMRs called vs. CD34+ HSPCs in IDH mutant, *DNMT3A*^{R882}/*IDH* double mutant, and *DNMT3A*^{R882} AMLs, and CD34+ HSPCs. E) Distribution of *DNMT3A*^{R882} DMR methylation levels by AML subtype. F) Example *DNMT3A*^{R882} DMR locus within the MLLT1 gene demonstrating hypomethylation phenotype of *DNMT3A*^{R882} mutant samples with respect to *IDH*^{mut} and *DNMT3A*^{R882}/*IDH* double mutant AML.

Figure 5. *IDH*^{mut}-specific hypermethylation is enriched for putative enhancer elements. A) Distribution of CD34+ HSPC ChromHMM chromatin states represented in *IDH*^{mut}-specific DMRs. Enrichment of chromatin states within *IDH*^{mut}-specific DMRs is given with respect to the frequency of states overlapping generically hypermethylated DMRs. B) Enhancer based

annotation of generically hypermethylated regions, *IDH1*^{mut}, and *IDH2*^{mut} DMRs, where DMRs intersecting an H3K27ac peak alone or in combination with H3K4me1 constitute active enhancers, an H3K27ac peak in combination with H3K27me3 constitutes poised enhancers, and H3K4me1 alone constitutes a weak enhancer profile. C) Examples of intragenic and genic enhancer regions exhibiting *IDH1*^{mut}, *IDH2*^{mut}, or *IDH1/2*^{mut} hypermethylation compared with CD34+ HSPCs and other AML subtypes. D) Heatmap of enhancer histone modifications and heterochromatin modifications over *IDH*^{mut}-specific DMRs (left) and generic hypermethylation (right) in CD34+ HSPCs (n=4/7/7 H3K27ac/H3K3me1/H3K27me3 samples respectively), *IDH*^{mut} AML (n=3), and *IDH*^{wt} AML samples (n=9/10/24 H3K27ac/H3K3me1/H3K27me3 respectively). E) Differential active enhancer signal (H3K27ac) for all AML-associated putative enhancers (black dots) compared to putative enhancers intersecting an *IDH1/2*^{mut}-specific DMR (red dots). F) HOMER motif enrichment analysis of *IDH1/2*^{mut}-specific DMRs with respect to a background set of generically hypermethylated regions. G) Enrichment analysis of TF binding events for 445 TFs within *IDH1/2*^{mut}-specific DMRs.

Figure 6. *IDH*^{mut}-specific enhancer DMRs interact with highly expressed genes. A) Locus heatmap of mean subtype methylation across *IDH*^{mut}-specific DMRs including annotated overlap with gene promoters (green), putative active enhancers (purple), and FitHiC loop anchors (blue). B) Schematic of eDMR target gene identification based on intersection of FitHiC loop-anchor regions, *IDH*^{mut}-specific eDMRs, and gene body coordinates. C) Scatter plot comparison of average normalized gene expression of all expressed genes in *IDH*^{mut} AML samples and CD34+ HSPCs (black dots) and eDMR target genes (red dots). D) Scatter plot comparison of average normalized gene expression of all expressed genes in *IDH*^{mut} AML samples and CD34+ HSPCs (black dots) and generic hypermethylated enhancer targets (red dots). E) Example *IDH*^{mut} -eDMR locus displaying robust interactions with the MYC promoter. A zoomed in view of the locus demonstrates focal enhancer hypermethylation in IDH1mut (purple) and IDH2mut (green) samples compared with CD34+ HSPCs (blue). Normalized MYC expression is shown for 2 CD34 samples, 7 and 15 IDH1 and ID2mut samples, and 91 *IDH*^{wt} samples. F) Example *IDH*^{mut} -eDMR locus displaying robust interactions with the SRSF3

promoter. A zoomed locus view demonstrates focal enhancer hypermethylation in *IDH1*^{mut} (purple) and *IDH2*^{mut} (green) samples compared with CD34+ HSPCs (blue). Normalized SRSF2 expression is shown for CD34 samples, *IDH1*^{mut} and *IDH2*^{mut} samples, and *IDH*^{wt} samples (see E for sample sizes).

Figure 1.

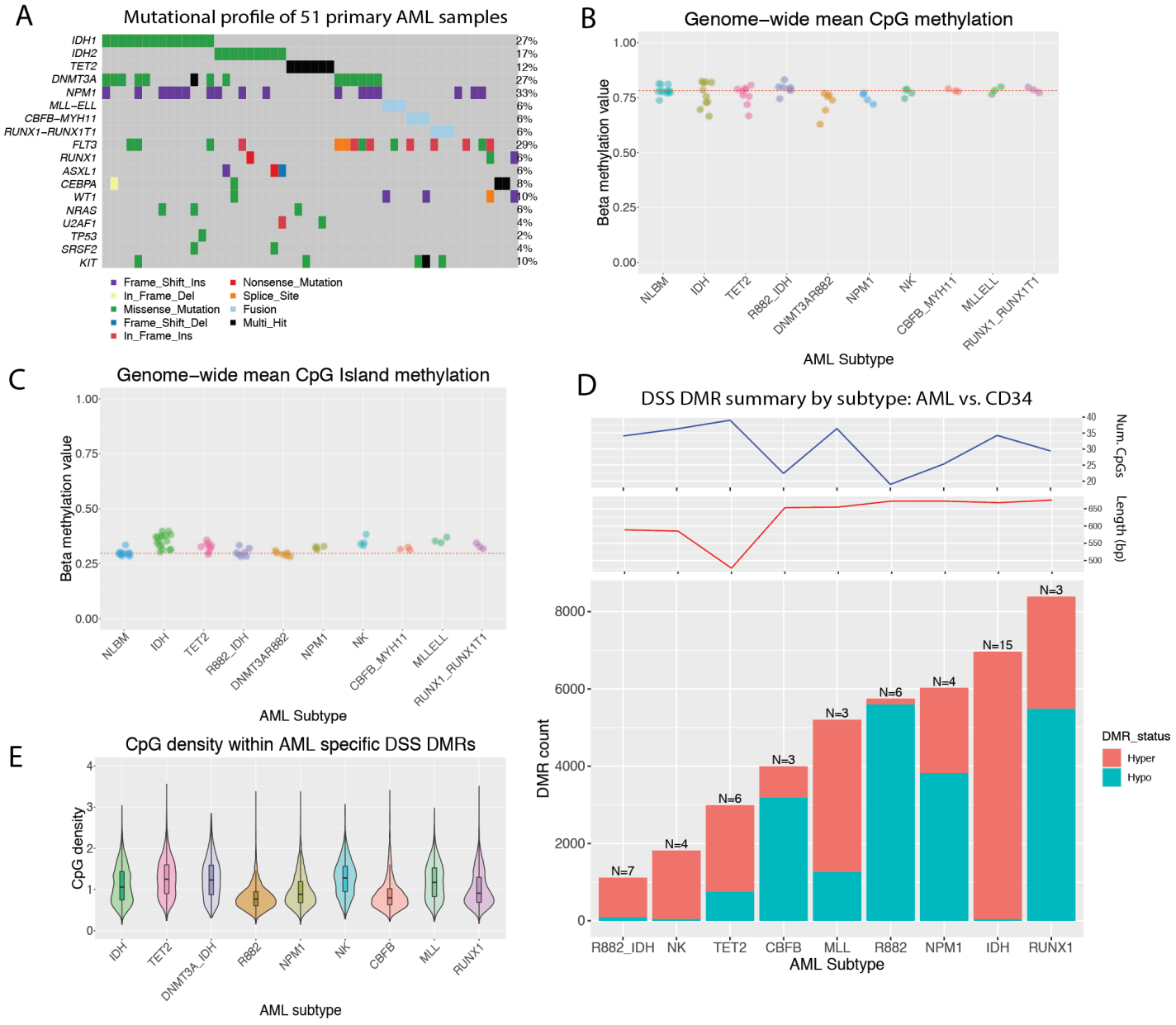

Figure 2.

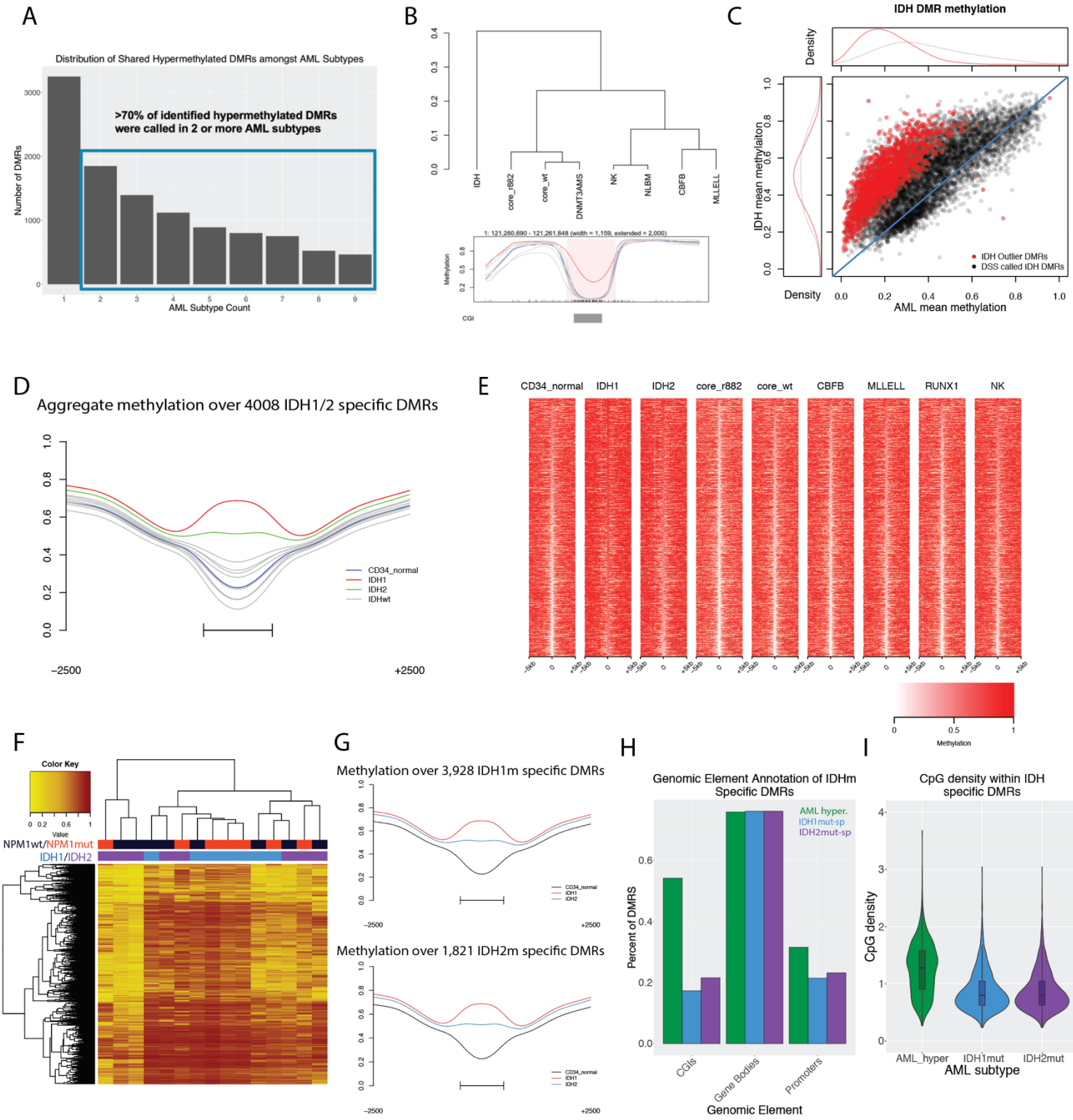

Figure 3.

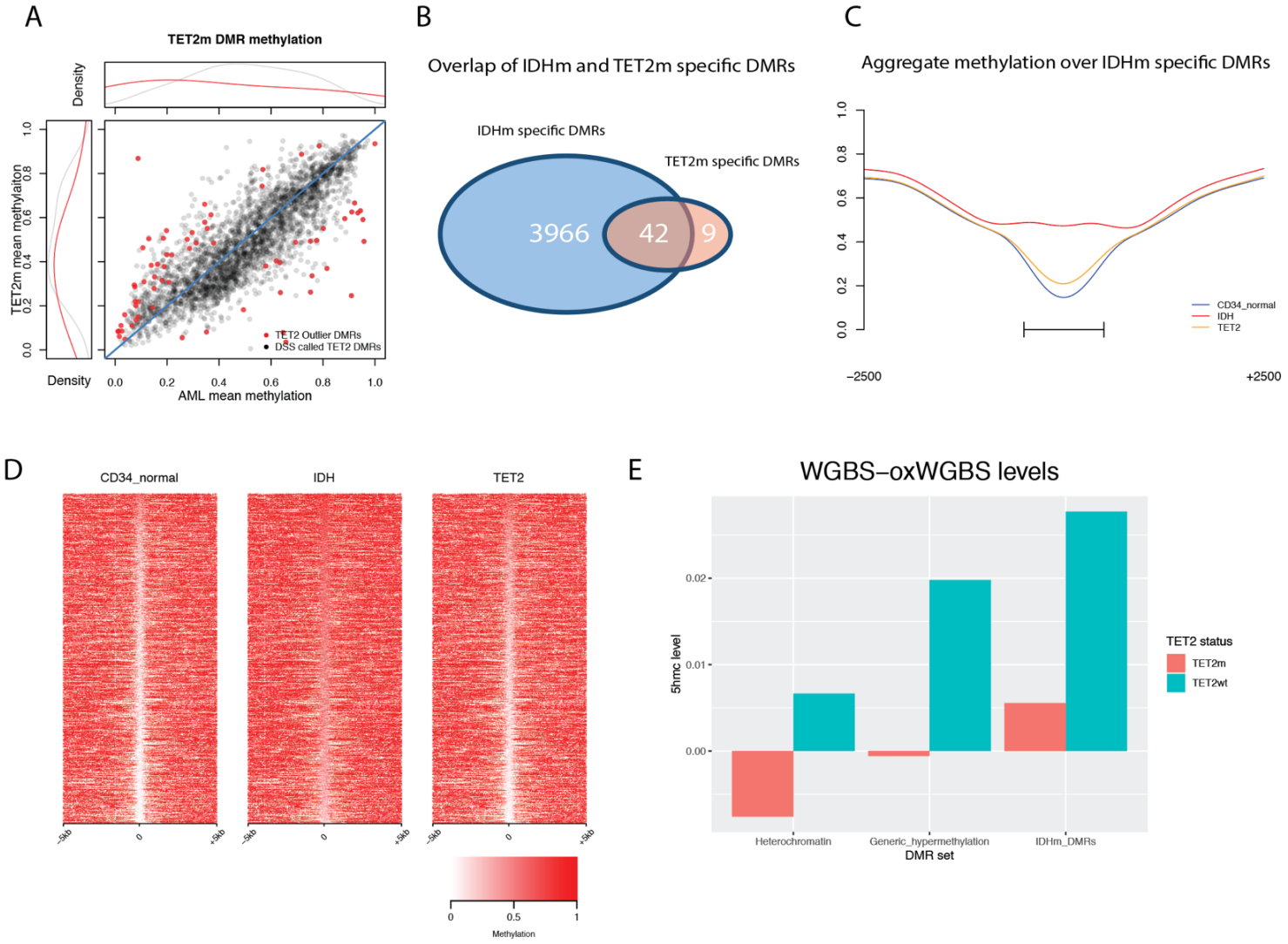

Figure 4.

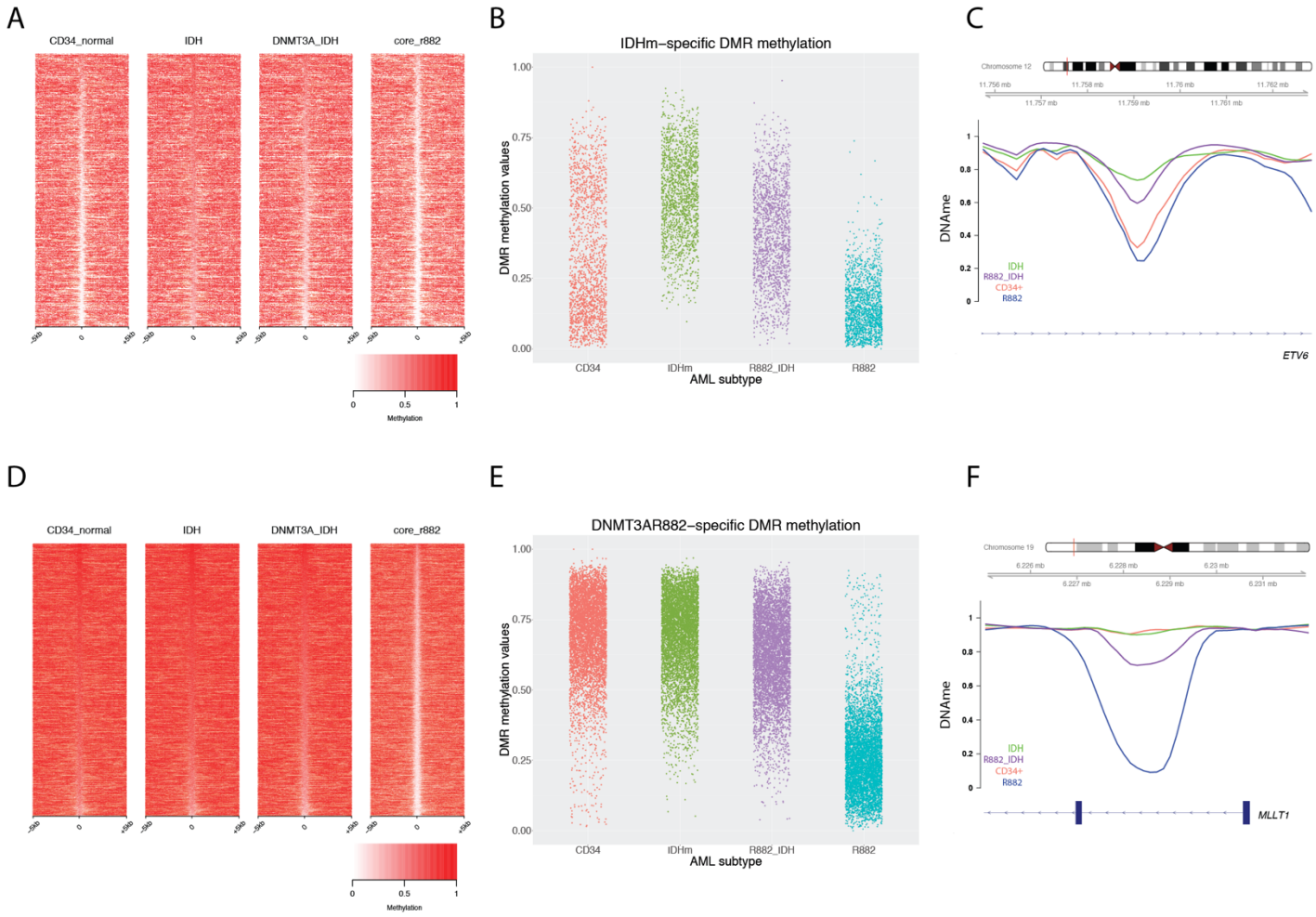

Figure 5.

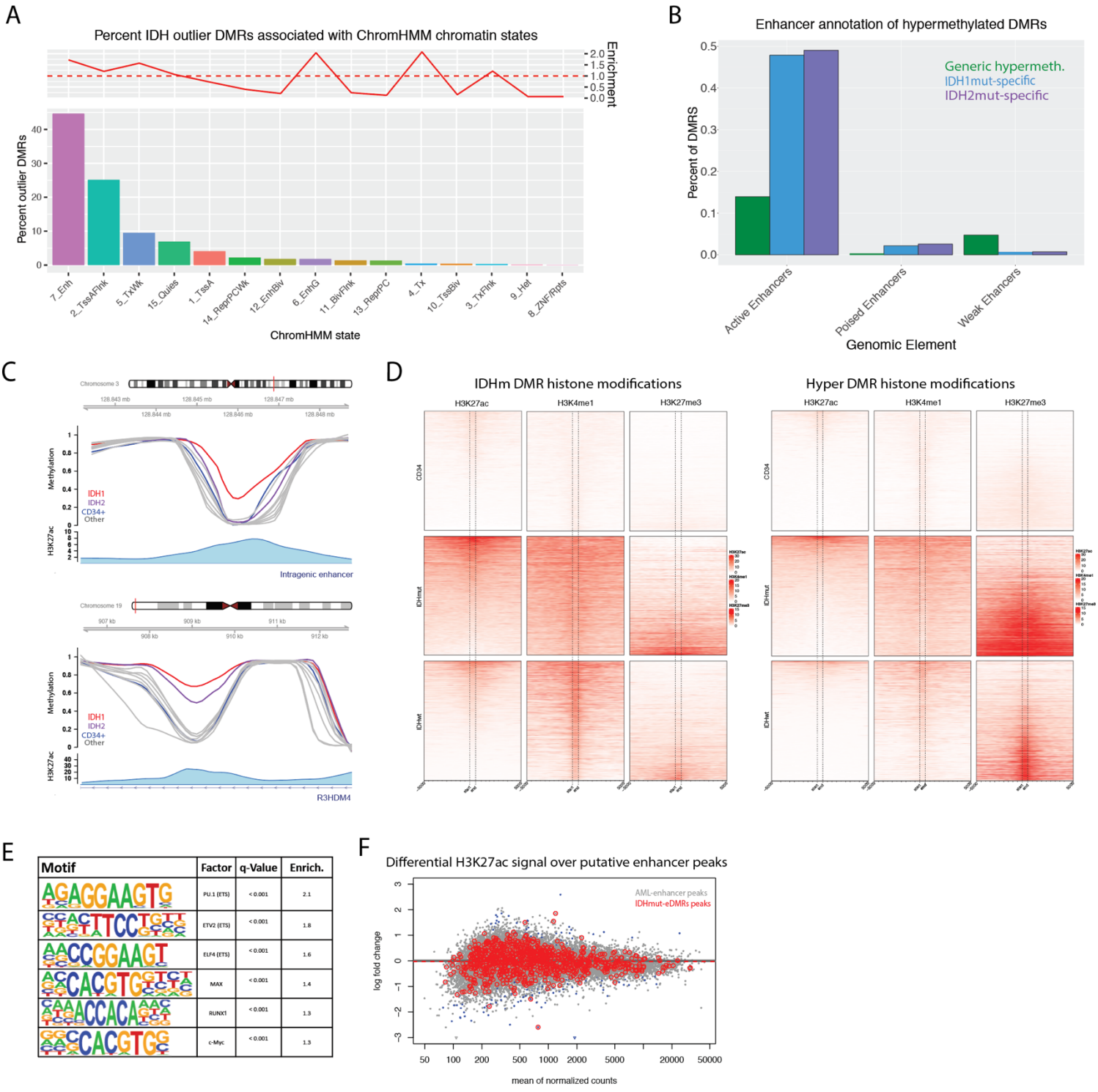

Figure 6.