Altered 5-hydroxymethylcytosine landscape in primary gastric adenocarcinoma

Hao Liu[1], Tianlei Xu[2], Ying Cheng[2], Michael Jin[2], Margaret Chang[2], Emily G. Allen[2], Peng Jin[3*], Xudong Wang[1*]

[1]Jilin University, China, [2]Emory University, United States, [3]Emory University School of Medicine, United States

Submitted to Journal:
Frontiers in Genetics

Specialty Section:
Epigenomics and Epigenetics

Article type:
Original Research Article

Manuscript ID:
435387

Received on:
06 Nov 2018

Frontiers website link:
www.frontiersin.org

Conflict of interest statement

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

Author contribution statement

PJ and XW conceived the projects. HL, TX, YC, MJ, and MC performed the experiments and data analyses. EA performed the data analyses. EA, PJ and XW wrote the manuscript.

Keywords

gastric cancer, 5-Hydroxymethylcytosine, TET, c-Myc, Enhancer

Abstract

Word count: 178

Multiple factors, including molecular, genetic, and epigenetic changes, have been associated with the formation and progression of gastric cancer. Cytosine methylation has long been recognized as a critical epigenetic mark in the mammalian genome. Recent research has implicated the involvement of 5-hydroxymethylcytosine (5hmC), a DNA base derived from 5-methylcytosine (5mC), via oxidation by ten-eleven translocation (TET) enzymes, in DNA methylation-related plasticity. Here we show that gastric tumors display significant loss of 5hmC. Using matched distant normal, peripheral and tumor primary tissues, we performed genome-wide profiling of 5hmC and identified differentially hydroxymethylated regions (DhMRs) specifically associated with gastric tumors. Gene Ontology (GO) analyses indicated that DhMRs (both loss-of-5hmC and gain-of-5hmC) were enriched among the genes involved in specific pathways. Interestingly the binding motif of hypoxia-inducible factor 1 (HIF1) is enriched among both peripheral and tumor DhMRs, while the Myc-binding motif is specifically enriched among only tumor DhMRs. Tumor progression analyses revealed a unique set of DhMRs that correlate with tumor progression. These data together suggest that the alteration of 5hmC could potentially contribute to the tumorigenesis of gastric tumors.

Funding statement

This work was supported in part by the National Institutes of Health [NS051630, NS091859, and NS097206 to P.J.].

Ethics statements

(Authors are required to state the ethical considerations of their study in the manuscript, including for cases where the study was exempt from ethical approval procedures)

Does the study presented in the manuscript involve human or animal subjects: No

Data availability statement

Generated Statement: The datasets generated for this study can be found in NCBI GEO, GSE106220.

Altered 5-hydroxymethylcytosine landscape in primary gastric adenocarcinoma

Hao Liu[1#], Tianlei Xu[2#], Ying Cheng[3#], Michael H Jin[3], Margaret Y Chang[3], Emily G Allen[3], Peng Jin[3]*, Xudong Wang[1]*

[1]Department of Gastroenterological Surgery, The Second Hospital, Jilin University, Changchun, Jilin Province, China; [2]Department of Mathematics and Computer Science, [3]Department of Human Genetics, School of Medicine, Emory University, Atlanta, GA, USA;

#H. Liu, T. Xu and Y. Cheng contributed equally to this work.

***Corresponding authors:**
Peng Jin, Ph.D.
Department of Human Genetics
Emory University School of Medicine
Atlanta, GA 30322
USA
Email: peng.jin@emory.edu

Xudong Wang, PhD
Department of Gastroenterological Surgery
The Second Hospital, Jilin University
Changchun, Jilin Province
China
Email: wangxudong1971@126.com

Abstract

Multiple factors, including molecular, genetic, and epigenetic changes, have been linked to gastric cancer formation and progression. Cytosine methylation (5mC) has been well studied and recognized as a critical epigenetic mark in the mammalian genome. The recent discovery of 5-hydroxymethylcytosine (5hmC), which generated by ten-eleven translocation (TET) enzymes, provide new perspective to understand DNA methylation-related plasticity. Here we show that gastric tumors display significant loss of 5hmC. Using matched distant normal, peripheral and tumor primary tissues, we performed genome-wide profiling of 5hmC and identified differentially hydroxymethylated regions (DhMRs) specifically associated with gastric tumors. Gene Ontology (GO) analyses indicated that DhMRs (both loss-of-5hmC and gain-of-5hmC) were enriched among the genes involved in specific pathways. Interestingly the binding motif of hypoxia-inducible factor 1 (HIF1) is enriched among both peripheral and tumor DhMRs, while the Myc-binding motif is specifically enriched among only tumor DhMRs. Tumor progression analyses revealed a unique set of DhMRs that correlate with tumor progression. These data together suggest that the alteration of 5hmC could potentially contribute to the tumorigenesis of gastric tumors.

Key words: Gastric cancer, 5-hydroxymethylcytosine, TET, c-Myc, Enhancer

Introduction

Gastric cancer (GC) is one of the most common cancers and the leading causes of global cancer mortality (Ferlay et al., 2015). Recently, a substantial number of studies have revealed that GC is a multistage pathological state (Tan and Yeoh, 2015). The formation and progression of GC can be triggered by various factors, including internal molecular/genetic changes and external environmental exposures (e.g.., viral/bacterial infections, smoking, and diet) (Ajani et al., 2017).

"Epigenetics" is a term that defines the heritable changes in gene expression, where the DNA sequence is not altered (Egger et al., 2004). In general, there are three well-studied epigenetic marks, including histone modification, DNA methylation and non-coding RNA-mediated silencing (Padmanabhan et al., 2017). Disrupting any of the above-mentioned epigenetic systems can result in the dysregulated gene expression and, thereafter, induce a number of multi-system disorders and neoplasia. For example, previous studies have identified several key genetic alterations that link to gastric malignancy, including mutations in chromatin modifier gene ARID1A (AT-rich interaction domain 1A) and amplifications in HER2 (human epidermal growth factor receptor 2), FGFR2 (fibroblast growth factor receptor 2) and MET (mesenchymal epithelial transition) (Ajani et al., 2017). Moreover, the recent histone modification study in primary GCs have revealed a pervasive reprogramming of the gastric super-enhancer landscape during tumorigenesis (Ooi et al., 2016).

In addition to histone modifications, cytosine methylation (5mC) can influence transcriptional states via modifying DNA-protein interactions (Jaenisch and Bird, 2003). The dysregulation at epigenetic level, including hypermethylation on promoter CpG island, is the commonly known molecular cause in human neoplasia. The hypermethylation on mismatch repair gene hMLH1 (human mutL homolog 1) promoter region is the main underlying mechanism for microsatellite instability in GC (Keller et al., 1996). Another example is the methylation level on p16 (cyclin-dependent kinase inhibitor 2A) gene. It has been reported that, although mutation of the p16 gene is infrequent, p16 hypermethylation is common in GC with a higher incidence in the intestinal type, suggesting a vital role of DNA methylation-mediated p16 gene inactivation in GC (Shim et al., 2000).

Until recently, a novel cytosine modification 5-hydroxymethylcytosine (5hmC) was discovered. The Ten-eleven translocation (TET) family dioxygenases oxidize 5mC to generate 5hmC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). The subsequent studies further demonstrated that TET proteins can further oxidize 5hmC to 5-formylcytosine and 5-carboxylcytosine, which are recognized and excised by mammalian DNA glycosylase TDG and subsequently converted to cytosine through base excision repair (Cortazar et al., 2011; Cortellino et al., 2011; He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011; Pfaffeneder et al., 2011; Zhang et al., 2012), resulting in active DNA demethylation in mammals. TET-mediated epigenetic modification is critical in embryonic stem cell (ES) maintenance, normal myelopoiesis, myeloid leukemia, and gliomas, pointing to the biological importance of 5hmC

modification in the regulation of ES pluripotency, tissue homeostasis, and disease pathogenesis (Dawlaty et al., 2011; Doege et al., 2012; Gu et al., 2011; Ito et al., 2010; Ko et al., 2010; Koh et al., 2011; Lian et al., 2012a; Lian et al., 2012b; Moran-Crusio et al., 2011; Tan and Shi, 2012; Williams et al., 2011; Xu et al., 2011). In particular, 5hmC abundance is significantly reduced in many types of human cancer, such as melanoma (Lian et al., 2012b), prostate, breast and colon cancers (Haffner et al., 2011), hepatocellular carcinoma (Liu et al., 2013), and especially in central nervous system tumors (Jin et al., 2011). However, the genome-wide analyses of 5hmC in GC tissues have not been conducted.

In this study, using matched distant normal, peripheral and tumor primary tissues from six patients, we show the loss of 5hmC is specifically associated with gastric tumors. Genome-wide profiling of 5hmC identified differentially hydroxymethylated regions (DhMRs) specifically associated with GC. Gene Ontology (GO) analyses indicated that DhMRs (both loss-of-5hmC and gain-of-5hmC) were enriched among the genes involved in specific pathways. Interestingly, the binding motif of hypoxia-inducible factor 1 (HIF1) is enriched among both peripheral and tumor DhMRs while the Myc-binding motif is specifically enriched among tumor DhMRs. Furthermore, we performed tumor progression analyses and identified a unique set of DhMRs that correlate with tumor progression. Finally, we observed significant overlap between DhMRs and previously identified super-enhancers. These data together suggest that the alteration of 5hmC could potentially contribute to the tumorigenesis of gastric tumors.

Results

5-hydroxymethylcytosine is significantly reduced in gastric tumors

Previous studies have found 5-hydroxymethylcytosine (5hmC) abundance significantly reduced in several different types of human cancer. To explore the role of 5hmC in GC, we collected six sets of fresh gastric tissues, including gastric tumor (Tumor/T), adjacent gastric tissue to the tumor (Peripheral/P) and distant normal gastric tissue (Normal/N). We first examined 5hmC abundance by dot blotting using an anti-5hmC antibody. Pairwise comparisons revealed that gastric tumor tissue (T) showed a significant reduction in 5hmC compared to the normal gastric tissue (N) ($p<0.001$; Figure 1A), with an average of 0.35-fold 5hmC relative to normal tissue. No significant difference was observed between peripheral tissue (P) and normal tissue (N), suggesting that the reduction of 5hmC is a hallmark for GC.

Genome-wide 5hmC profiling in gastric tumors

To understand the alteration of 5hmC in gastric tumors, we investigated 5hmC genome-wide distribution in all three tissue types using a chemical labeling and affinity purification method coupled with high-throughput sequencing (Song et al., 2011). Firstly, global 5hmC levels were determined by counting and plotting normalized 5hmC mapped reads in each 10-kb binned human genome among the three sets of tissues. Bins with lower 5hmC reads in tumor or peripheral tissue compared to normal tissue are shown in blue, whereas bins with higher 5hmC reads in tumor or peripheral tissue compared to normal are shown in red. In the comparison of tumor tissue to normal tissue, we found an overall higher number of

bins in blue, indicating a general reduction of global 5hmC in tumor tissue(Figure 1B, blue versus red), which is consistent with our dot blot analyses. Intriguingly, there were more red bins than blue bins in the comparison of peripheral tissues to normal tissue (Figure 1C), suggesting that the 5hmC abundance could potentially increase in peripheral tissues; however, we did not detect a significant difference by dot blot.

Identification and characterization of differentially hydroxymethylated regions (DhMRs)

To identify the differentially hydroxymethlated CpG sites, edgeR (Version: edgeR_3.8.6) was used to normalize read counts, estimate dispersion and perform differential tests. Tumor, Peripheral and Normal tissues from the same individual were grouped together to adjust for any within-subject effects. Sites with p-values smaller than 0.01 were considered significantly different as DhMRs, and included in the downstream analyses. The genome-wide distribution of DhMRs are shown in Figure 2A. Overall, 17,318 5hmC-containing genomic regions display reduced 5hmC while 1,344 loci show increased 5hmC in tumor vs. normal tissues, consistent with the overall reduction of 5hmC abundance observed earlier. Interestingly, in peripheral tissues vs. normal tissues, more genomic loci display increased 5hmC signals (4,795) than reduced (1,876) (Table 1, Figure 2B).

To investigate the relationship of DhMRs between these two comparisons, we overlapped the regions with gain-of-5hmC or loss-of-5hmC from the two comparison groups. Interestingly,

1123 of the 1344 (83.5%) DhMRs from the T/N gain-of-5hmC DhMRs overlapped with the P/N gain of 5hmC DhMRs, suggesting similar 5hmC changes in both tumor and peripheral tissues. However, the loss-of-5hmC regions from these two groups were very distinct: only 6 common loss-of-5hmC regions were identified as overlapping (0.3% of P/N and 0.03% of T/N) (Figure 2B).

We next characterized the genome-wide distributions of DhMRs and evaluated their enrichment or depletion on featured genomic regions by comparing to the expected values. DhMRs were enriched on intragenic regions (e.g., exons, UTRs and promoters) (Figure 2C and 2D, log2 fold change from expected > 0) while depleted in intergenic regions. The results supported the previous observations suggesting that intragenic 5hmC may be associated with gene regulation (Song et al., 2011).

Aberrant 5hmC changes in gastric tumors and peripheral tissue influence the core biological pathways and potential binding affinity of selective transcription factors

To understand whether dynamic 5hmC changes have impact on the relative biological processes, we mapped the these GC-associated DhMRs to the hg19 human reference sequence and further subjected these genes to Gene Ontology (GO) pathway analysis (Huang da et al., 2009). Interestingly, several biological processes, such as phosphate metabolic process, protein amino acid phosphorylation, and intracellular signaling cascade, were affected by gain-of-5hmC in both peripheral and tumor tissues (Figure 3A and 3C). GO

analysis also revealed unique pathways associated with either peripheral or tumor tissues. For instance, the pathway related to Ras protein was uniquely featured in peripheral tissue, but not in tumor tissue. Among the loss-of-5hmC loci, the overlap between peripheral and tumor tissues was relatively small compared to those with gain-of-5hmC (e.g., Ras protein transduction) (Figure 3B and 3D). The loss-of-5hmC loci in peripheral tissues were enriched with genes involved in programmed cell death, while the loss-of-5hmC loci associated with tumor tissues were enriched with genes involved in cytoskeleton organization and intracellular signaling. These observations suggest the unique 5hmC alterations associated with peripheral and tumor tissues, reflecting different stages of tumorigenesis.

Given that many transcription factors are sensitive to DNA covalent modification dynamics, we employed motif search algorithms to comprehensively predict the potential transcription factor binding sites on the DhMRs that we identified. Interestingly, the binding motif of hypoxia-inducible factor 1 (HIF1), a critical transcription factor responsible for controlling the transcriptome in response to normoxia and hypoxia, were enriched in both peripheral and tumor DhMRs (Figure 4A and 4B). Also, specifically among tumor DhMRs, Myc (both c-Myc and n-Myc) binding sites were enriched. These results suggest that the altered 5hmC modifications could potentially alter the binding of both HIF1 and Myc during gastric tumorigenesis.

Tumor Progression Analyses

Since we have three distinct tissues: tumor, peripheral tissue and distant normal tissue, we performed a Tumor-Peripheral-Normal tissue progression test. This analysis aimed to identify the regions that displayed a gradual gain- or loss-of-5hmC along the progression of tissue changes. In this analysis, T, P and N tissues were coded as the three stages of a continuous progression and quantified as level 2, 1 and 0, respectively. At the differential calling step, the stages were used as the alternative column to group in the design matrix while adjusting for individual effects. Our progression analysis revealed a distinct set of DhMRs (74 loci with gain-of-5hmC and 3481 loci with loss-of-5hmC; Table 1). The changes of 5hmC within these regions are shown in Figure 5A as a heatmap, which correlates with the disease progression.

Genomic features associated with this unique set of DhMRs were similar to the DhMRs of both peripheral and tumor tissues (Figure 5B compared to Figure 2C and 2D). However, the function of DhMRs as characterized in functional enrichment of GO terms are distinct from the previous two groups of DhMRs (Figure 5C compared to Figure 3). In particular, among the loci with loss-of-5hmC in the progression analyses, genes involved in splicing and nucleoside regulators were uniquely enriched.

DhMRs overlap with the previously identified super-enhancers

Super-enhancers refer to enhancer clusters occurring in close proximity, are larger in size than typical enhancers, and exhibit higher transcription factor binding densities and are

strongly associated with key cell identity regulators. Super-enhancers are enriched in disease-associated genetic variants and are acquired by cancer cells at key oncogenes. By analyzing more than one hundred epigenomic profiles from primary GCs, normal gastric tissues and cell lines, earlier works have identified super-enhancers associated with GC tumorigenesis and revealed a genome-wide reprogramming of super-enhancer landscape during GC tumorigenesis (Ooi et al., 2016). To correlate the 5hmC dynamics with super-enhancers, we examined the change of 5hmC at the top 100 super-enhancers identified in the previous study and performed enrichment analysis. The analyses were performed by comparing the two ratios: (1) (DhMRs overlapping with super enhancers)/(all DhMRs); (2) (super enhancers)/(the whole genome). The fold change of the ratio (1) vs. (2) is shown in Figure 6 (in log2 scale). In addition, binomial tests were performed to evaluate the significance of enrichment using a one-sided null hypothesis and the probability of one nucleotide being inside the super enhancer, i.e. ratio (2), as the percentage of success. We observed significant enrichment of the identified super-enhancers among the DhMRs of both peripheral and tumor tissues.

Discussion

In the past decade, increasing evidence suggests that epigenetic regulation plays a vital role in the initiation and progression of human cancer, including gastric cancer (Jones and Baylin, 2002). Not only as a transient intermediate in DNA active demethylation process, 5hmC is reported to be enriched and stable in mammalian tissues and recognized by specific 5hmC-binding proteins (Spruijt et al., 2013; Szulwach et al., 2011). Also, it is known that 5hmC has biological functions in a number of developmental and neurological diseases and cancer (Cheng et al., 2015; Lian et al., 2012b; Szulwach et al., 2011). Thus, 5hmC could potentially contribute to the tumorigenesis of gastric tumors. In the present study, for the first time, we profiled the genome-wide 5hmC in matched distant normal, peripheral and tumor primary tissues from gastric cancer patients, and our results show a specific association between the loss of 5hmC in gastric tumor tissues. GO analyses indicated that DhMRs (both loss-of-5hmC and gain-of-5hmC) were enriched among the genes involved in a number of selective pathways. Interestingly, the binding motif of hypoxia-inducible factor 1 (HIF1) is enriched among both peripheral and tumor DhMRs, while the Myc-binding motif is specifically enriched among tumor DhMRs. The tumor progression analyses identified a unique set of DhMRs that correlate with gastric tumor progression. In addition, we observed significant overlap between DhMRs and previously identified super-enhancers. These data together suggest that the alteration of 5hmC could potentially contribute to the tumorigenesis and progression of gastric tumors.

As the first epigenetic mark that was linked to tumorigenesis (Feinberg and Vogelstein, 1983), DNA methylation provides a stable gene silencing mechanism, and it is commonly known that the hypermethylation in promoter regions causes the inactivation of certain tumor-suppressor genes (Kulis and Esteller, 2010). In addition, studies have found that hypomethylation could also result in the activation of genes that are important in cancer (Feinberg and Tycko, 2004). It is interesting that our tumor progression analyses suggest that the loss-of-5hmC is more significantly correlated with gastric tumor progression. The GO analysis revealed the loss-of-5hmC loci of peripheral and tumor tissues are distributed in genes involved in distinct pathways, with the exception of a few well-known pathways that are associated with the tumorigenesis of gastric tumors (e.g., Ras protein signal transduction). It has been reported that an oncogene *BRAF* (serine/threonine-protein kinase B-Raf), encoding a RAF protein in the downstream of RAS pathway, is somatically mutated in many human cancers, including gastric cancer (Davies et al., 2002). This suggests that aberration of the RAS pathway could contribute to the pathogenesis of stomach cancer. Therefore, the loss-of-5hmC may result in the dysregulation of the RAS genes and contribute to the gastric tumor progression.

It was once considered that only protein-coding DNA (about 2% of the hg19) contained valuable information, while the rest of genomic non-coding DNA was considered 'junk DNA'. However, more recent genome-wide studies have revealed understandingof the hiding information in the human genome (Consortium, 2012), where more than the expected

number of non-coding regions have been found to encode regulatory information. For example, it has been reported that the levels of a long non-coding RNA, H19, were substantially upregulated in the cells and tissues of GC when compared to normal controls, resulting in partial p53 inactivation (Yang et al., 2012). Previous global profiling of chromatin marks that are associated with distinct regulatory elements have revealed the alterations of hundreds of epigenomic promoter and predicted enhancers in GC (Muratani et al., 2014; Ooi et al., 2016; Qamra et al., 2017). In this study, we overlapped the top 100 super-enhancers identified in the previous studies and performed enrichment analyses. Many of the DhMRs of both peripheral and tumor tissues overlap with predicted super-enhancers, suggesting the dynamic change of 5hmC may influence the expression of non-coding regions and adjacent genes. It will be important to investigate the roles of 5hmC in our identified intergenic DhMRs in gastric cancer patients.

In summary, we profiled the genome-wide distribution of 5hmC in matched distant normal, peripheral and tumor primary tissues from gastric cancer patients. Our results showed a specific association between the loss of 5hmC and gastric tumor tissue. The DhMRs located within genes associated with a number of selective pathways and were enriched with motifs that are involved in gastric tumorigenesis. Also, we observed significant overlap between DhMRs and previously identified super-enhancers. These data together suggest that the alteration of 5hmC could potentially contribute to the tumorigenesis of gastric tumors.

Materials and Methods

Sample collection

The protocol of this study was approved by the Ethical Committee of The Second Hospital, Jilin University. Patients recruited for this study were provided with informed consent to participate and written informed consents were obtained from all the tissue donors of this study. Sets of fresh gastric tissues were collected from six patients, including gastric tumor (Tumor/T), adjacent gastric tissue to the tumor (Peripheral/P) and distant normal gastric tissue (Normal/N). Samples were stored at -80^0 C until DNA was extracted.

Dot blot assay

Genomic DNA was extracted by using phenol-chloroform and then quantified by Nanodrop 2000 (Thermo, USA). After being denatured in 1M NaOH for 2 hours, 100-300ng genomic DNA was spotted onto a charged nylon-based membrane at room temperature for 30 minutes and then baked at 75°C for 1 hour. The membrane was blocked with 5% non-fat milk for 1 hour, and washed in PBS three times. The blot was then incubated with rabbit polyclonal anti-5hmC antibody (Active Motif; 1:10000) at 4°C overnight. On the following day, the blot was washed in PBS three times and incubated with peroxidase-conjugated anti-rabbit IgG secondary antibody for 1 hour. The signal was visualized by using ECL-Plus system (Amersham Pharmacia Biotech) and quantified.

5hmC-specific chemical labeling, affinity purification, and sequencing

Genomic DNA was used for 5hmC genome-wide profiling. 5hmC enrichment was performed using a previously described procedure with an improved selective chemical labeling method (28). DNA libraries were generated following the Illumina protocol for "Preparing Samples for ChIP Sequencing of DNA" (Part# 111257047 Rev. A) using 25-50 ng of input genomic DNA or 5hmC-captured DNA to initiate the protocol.

Bioinformatic data analyses

Processing of sequencing data was performed as previously described (55). Briefly, FASTQ sequence files from biological replicates were concatenated and aligned to the *Homo sapiens* reference genome (GRCh37/hg19) using Bowtie 0.12.6, keeping only unique non-duplicate genomic matches with no more than 2 mismatches within the first 25 bp. Hydroxymethlation levels on each CpG site were evaluated by read counts. To determine differentially hydroxymethylated CpG sites/regions, we used edgeR (Version: edgeR_3.8.6) to normalize read counts, estimate dispersion and perform differential tests. P-values smaller than 0.01 are considered significant to call DhMRs.

Motif detection within DhMRs was performed with Homer. DhMRs were resized to 50bp windows fixed at the center. Homer (v 4.8.3) was used to scan DhMRs for motifs using "findMotifsGenome.pl" with standard background (random genomic sequences sampled according to GC content of peak sequences) and parameters (mask repeats/lower case sequence, using 200bp fragment size for motif finding).

Gene Ontology analyses were performed as previously described using the DAVID Bioinformatics Resources 6.7 Functional Annotation Tool. Gene sets were identified by joining subsets of DhMRs with RefSeq tables obtained from the UCSC genome browser tables.

Statistical analysis

GraphPad Prism software was used for the statistical analyses. Either a t-test or one-way ANOVA was performed to evaluate differences between the groups. All results were expressed in mean ± SEM. P-values<0.05 was considered statistically significant different.

Acknowledgements

This work was supported in part by the National Institutes of Health [NS051630, NS091859, and NS097206 to P.J.].

Conflict of Interest Statement: None declared.

Reference:

Ajani, J.A., Lee, J., Sano, T., Janjigian, Y.Y., Fan, D., and Song, S. (2017). Gastric adenocarcinoma. Nat Rev Dis Primers *3*, 17036.

Cheng, Y., Bernstein, A., Chen, D., and Jin, P. (2015). 5-Hydroxymethylcytosine: A new player in brain disorders? Experimental neurology *268*, 3-9.

Consortium, E.P. (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57-74.

Cortazar, D., Kunz, C., Selfridge, J., Lettieri, T., Saito, Y., MacDougall, E., Wirz, A., Schuermann, D., Jacobs, A.L., Siegrist, F.*, et al.* (2011). Embryonic lethal phenotype reveals a function of TDG in maintaining epigenetic stability. Nature *470*, 419-423.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D.*, et al.* (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W.*, et al.* (2002). Mutations of the BRAF gene in human cancer. Nature *417*, 949-954.

Dawlaty, M.M., Ganz, K., Powell, B.E., Hu, Y.C., Markoulaki, S., Cheng, A.W., Gao, Q., Kim, J., Choi, S.W., Page, D.C.*, et al.* (2011). Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell *9*, 166-175.

Doege, C.A., Inoue, K., Yamashita, T., Rhee, D.B., Travis, S., Fujita, R., Guarnieri, P., Bhagat, G., Vanti, W.B., Shih, A.*, et al.* (2012). Early-stage epigenetic modification during somatic cell reprogramming by Parp1 and Tet2. Nature *488*, 652-655.

Egger, G., Liang, G., Aparicio, A., and Jones, P.A. (2004). Epigenetics in human disease and prospects for epigenetic therapy. Nature *429*, 457-463.

Feinberg, A.P., and Tycko, B. (2004). The history of cancer epigenetics. Nature reviews Cancer *4*, 143-153.

Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature *301*, 89-92.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., and Bray, F. (2015). Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer *136*, E359-386.

Gu, T.P., Guo, F., Yang, H., Wu, H.P., Xu, G.F., Liu, W., Xie, Z.G., Shi, L., He, X., Jin, S.G.*, et al.* (2011). The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. Nature *477*, 606-610.

Haffner, M.C., Chaux, A., Meeker, A.K., Esopi, D.M., Gerber, J., Pellakuru, L.G., Toubaji, A., Argani, P., Iacobuzio-Donahue, C., Nelson, W.G.*, et al.* (2011). Global 5-hydroxymethylcytosine content is significantly reduced in tissue stem/progenitor cell compartments and in human cancers. Oncotarget *2*, 627-637.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L.*, et al.* (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA.

Science *333*, 1303-1307.
Huang da, W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nature protocols *4*, 44-57.
Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. (2010). Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. Nature *466*, 1129-1133.
Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. Science *333*, 1300-1303.
Jaenisch, R., and Bird, A. (2003). Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. Nat Genet *33 Suppl*, 245-254.
Jin, S.G., Jiang, Y., Qiu, R., Rauch, T.A., Wang, Y., Schackert, G., Krex, D., Lu, Q., and Pfeifer, G.P. (2011). 5-Hydroxymethylcytosine is strongly depleted in human cancers but its levels do not correlate with IDH1 mutations. Cancer research *71*, 7360-7365.
Jones, P.A., and Baylin, S.B. (2002). The fundamental role of epigenetic events in cancer. Nature reviews Genetics *3*, 415-428.
Keller, G., Grimm, V., Vogelsang, H., Bischoff, P., Mueller, J., Siewert, J.R., and Hofler, H. (1996). Analysis for microsatellite instability and mutations of the DNA mismatch repair gene hMLH1 in familial gastric cancer. Int J Cancer *68*, 571-576.
Ko, M., Huang, Y., Jankowska, A.M., Pape, U.J., Tahiliani, M., Bandukwala, H.S., An, J., Lamperti, E.D., Koh, K.P., Ganetzky, R., *et al.* (2010). Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature *468*, 839-843.
Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G., *et al.* (2011). Tet1 and tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. Cell Stem Cell *8*, 200-213.
Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. Science *324*, 929-930.
Kulis, M., and Esteller, M. (2010). DNA methylation and cancer. Adv Genet *70*, 27-56.
Lian, Christine G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., *et al.* (2012a). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell *150*, 1135-1146.
Lian, C.G., Xu, Y., Ceol, C., Wu, F., Larson, A., Dresser, K., Xu, W., Tan, L., Hu, Y., Zhan, Q., *et al.* (2012b). Loss of 5-hydroxymethylcytosine is an epigenetic hallmark of melanoma. Cell *150*, 1135-1146.
Liu, C., Liu, L., Chen, X., Shen, J., Shan, J., Xu, Y., Yang, Z., Wu, L., Xia, F., Bie, P., *et al.* (2013). Decrease of 5-hydroxymethylcytosine is associated with progression of hepatocellular carcinoma through downregulation of TET1. PloS one *8*, e62828.
Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. J Biol Chem *286*, 35334-35338.

Moran-Crusio, K., Reavie, L., Shih, A., Abdel-Wahab, O., Ndiaye-Lobry, D., Lobry, C., Figueroa, M.E., Vasanthakumar, A., Patel, J., Zhao, X.*, et al.* (2011). Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell *20*, 11-24.
Muratani, M., Deng, N., Ooi, W.F., Lin, S.J., Xing, M., Xu, C., Qamra, A., Tay, S.T., Malik, S., Wu, J.*, et al.* (2014). Nanoscale chromatin profiling of gastric adenocarcinoma reveals cancer-associated cryptic promoters and somatically acquired regulatory elements. Nature communications *5*, 4361.
Ooi, W.F., Xing, M., Xu, C., Yao, X., Ramlee, M.K., Lim, M.C., Cao, F., Lim, K., Babu, D., Poon, L.F.*, et al.* (2016). Epigenomic profiling of primary gastric adenocarcinoma reveals super-enhancer heterogeneity. Nat Commun *7*, 12983.
Padmanabhan, N., Ushijima, T., and Tan, P. (2017). How to stomach an epigenetic insult: the gastric cancer epigenome. Nat Rev Gastroenterol Hepatol *14*, 467-478.
Pfaffeneder, T., Hackner, B., Truss, M., Munzel, M., Muller, M., Deiml, C.A., Hagemeier, C., and Carell, T. (2011). The discovery of 5-formylcytosine in embryonic stem cell DNA. Angew Chem, Int Ed *50*, 7008-7012.
Qamra, A., Xing, M., Padmanabhan, N., Kwok, J.J.T., Zhang, S., Xu, C., Leong, Y.S., Lee Lim, A.P., Tang, Q., Ooi, W.F.*, et al.* (2017). Epigenomic Promoter Alterations Amplify Gene Isoform and Immunogenic Diversity in Gastric Adenocarcinoma. Cancer Discov *7*, 630-651.
Shim, Y.H., Kang, G.H., and Ro, J.Y. (2000). Correlation of p16 hypermethylation with p16 protein loss in sporadic gastric carcinomas. Lab Invest *80*, 689-695.
Song, C.X., Szulwach, K.E., Fu, Y., Dai, Q., Yi, C., Li, X., Li, Y., Chen, C.H., Zhang, W., Jian, X.*, et al.* (2011). Selective chemical labeling reveals the genome-wide distribution of 5-hydroxymethylcytosine. Nat Biotechnol *29*, 68-72.
Spruijt, C.G., Gnerlich, F., Smits, A.H., Pfaffeneder, T., Jansen, P.W., Bauer, C., Munzel, M., Wagner, M., Muller, M., Khan, F.*, et al.* (2013). Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. Cell *152*, 1146-1159.
Szulwach, K.E., Li, X., Li, Y., Song, C.X., Wu, H., Dai, Q., Irier, H., Upadhyay, A.K., Gearing, M., Levey, A.I.*, et al.* (2011). 5-hmC-mediated epigenetic dynamics during postnatal neurodevelopment and aging. Nat Neurosci *14*, 1607-1616.
Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L.*, et al.* (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science *324*, 930-935.
Tan, L., and Shi, Y.G. (2012). Tet family proteins and 5-hydroxymethylcytosine in development and disease. Development *139*, 1895-1902.
Tan, P., and Yeoh, K.G. (2015). Genetics and Molecular Pathogenesis of Gastric Adenocarcinoma. Gastroenterology *149*, 1153-1162 e1153.
Williams, K., Christensen, J., Pedersen, M.T., Johansen, J.V., Cloos, P.A., Rappsilber, J., and Helin, K. (2011). TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature *473*, 343-348.
Xu, Y., Wu, F., Tan, L., Kong, L., Xiong, L., Deng, J., Barbera, A.J., Zheng, L., Zhang, H., Huang, S.*, et al.* (2011). Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1

hydroxylase in mouse embryonic stem cells. Mol Cell *42*, 451-464.
Yang, F., Bi, J., Xue, X., Zheng, L., Zhi, K., Hua, J., and Fang, G. (2012). Up-regulated long non-coding RNA H19 contributes to proliferation of gastric cancer cells. The FEBS journal *279*, 3159-3165.
Zhang, L., Lu, X., Lu, J., Liang, H., Dai, Q., Xu, G.L., Luo, C., Jiang, H., and He, C. (2012). Thymine DNA glycosylase specifically recognizes 5-carboxylcytosine-modified DNA. Nat Chem Biol *8*, 328-330.

Table 1. Number of differentially hydroxymethylated regions (DhMRs)

	Gain of 5hmC	Loss of 5hmC
P/N	4795	1876
T/N	1344	17318

	Gain of 5hmC	Loss of 5hmC
TPN progression	74	3481

Figure legends

Figure 1. 5-hydroxymethylcytosine is significantly reduced in gastric tumor

A. The 5hmC specific dot blot revealed a significant decrease of 5hmC in gastric tumors. Genomic DNA was collected from fresh gastric tissues of 6 patients, including gastric tumor (Tumor/T), adjacent gastric tissue to the tumor (Peripheral/P) and distant normal gastric tissue (Normal/N). The 5hmC specific dot blot was performed, and the 5hmC signal densities were quantified by ImageJ software. Error bars indicate mean ± SEM; One-way ANOVA with Tukey's *post hoc* test; $P < 0.001$.

B-C. Global change of 5hmC levels in gastric tumors. Genome-wide normalized 5hmC mapped reads were counted for each 10kb binned human genome comparing the gastric tumor (Tumor/T) to distant normal gastric tissue (Normal/N) (B) and adjacent gastric tissue to the tumor (Peripheral/P) compared to the distant normal gastric tissue (Normal/N) (C). Bins with lower 5hmC reads in tumor or peripheral compared to normal are shown in blue, and bins with higher 5hmC reads in tumor or peripheral compared to normal are shown in red. Compared to normal tissue, there was a general reduction of global 5hmC in tumor tissue (B) and an increase of global 5hmC in peripheral tissue (C). Bins containing more than ten 5hmC reads are shown.

Figure 2. Identification and characterization of differentially hydroxymethylated regions (DhMRs)

A. Circular map view of DhMRs on chromosomes. Each human chromosome is shown in the

black-and-white outer track. Read ratios between P and N (P/N) were calculated on these loci and are shown on the inner track, and 5hmC read ratios between T and N (T/N) are shown on the outer track. DhMRs with increased 5hmC signals compared to normal tissues were defined as Gain-of-5hmC (red), while those with decreased 5hmC signals were defined as Loss-of-5hmC (blue).

B. Gain-of-5hmC and loss-of-5hmC signals from the P/N and T/N groups were overlapped to determine the relationships of DhMRs between these two comparisons.

C-D. Identified Gain- and Loss-of-5hmC of T/N (C) and P/N (D) signals were annotated to genomic regions.

Figure 3. Gene ontology (GO) analysis with gain- and loss-of-5hmC

Genes with DhMRs on their gene body were defined as DhMR-associated genes. The Database for Annotation, Visualization and Integrated Discovery (DAVID) analysis was used for the DhMRs-associated genes in P/N (A and B) and T/N (C and D) groups. Significance of these processes is indicated by the rank of -log10 P value.

Figure 4. Motif search analysis of DhMRs

Motif search using DhMRs identified in P/N (A) and T/N (B) groups. Results for both comparisons suggested a common potential HIF1 binding alteration to these 5hmC dynamic regions. Motif search was performed by HOMER (Hypergeometric Optimization of Motif EnRichment).

Figure 5. Tumor progression analyses

A. Tumor-Peripheral-Normal tissue progression (T-P-N progression) test was used to identify the regions that displayed a graduate gain- or loss- of 5hmC along the progression of tissue changes. Briefly, T, P and N tissues were coded as the three stages of a continuous progression and quantified as level 2, 1 and 0, respectively. At the differential calling step, the stages were used as the alternative column to group in the design matrix with individual effect adjusted. Our progression analysis revealed a distinct set of DhMRs (74 Gain-of-5hmC and 3481Loss-of-5hmC DhMRs). The changes of 5hmC within these regions were illustrated by a heatmap, which correlate with the disease progression.

B. Identified Gain- and Loss-of-5hmC of T-P-N progression DhMRs were annotated to various genomic regions.

C. Gene ontology (GO) analysis with Gain- and Loss-of-5hmC of T-P-N progression DhMRs reveal distinct functions that differ from those DhMRs in P/N and T/N. In particular, those loss-of-5hmC associated genes that are involved in splicing and nucleoside regulators were uniquely enriched.

Figure 6. Overlaps between DhMRs and super-enhancers

Enrichment analysis was performed by examining the 5hmC change at the top 100 super-enhancers identified in the previous study. The analyses were performed by comparing the

two ratios: (1) the ratio between DhMRs overlapping with super enhancers, and the whole range of DhMRs; (2) the ratio between the super enhancers and the whole genome. The fold change of the ratio (1) vs. (2) is highest in P/N, followed by T/N and least in TPN-progression. A binomial test was performed to evaluate the significance of enrichment using a one-sided null hypothesis and the probability of one nucleotide being inside the super enhancer, i.e. ratio (2), as the percentage of success.

Figure 1.

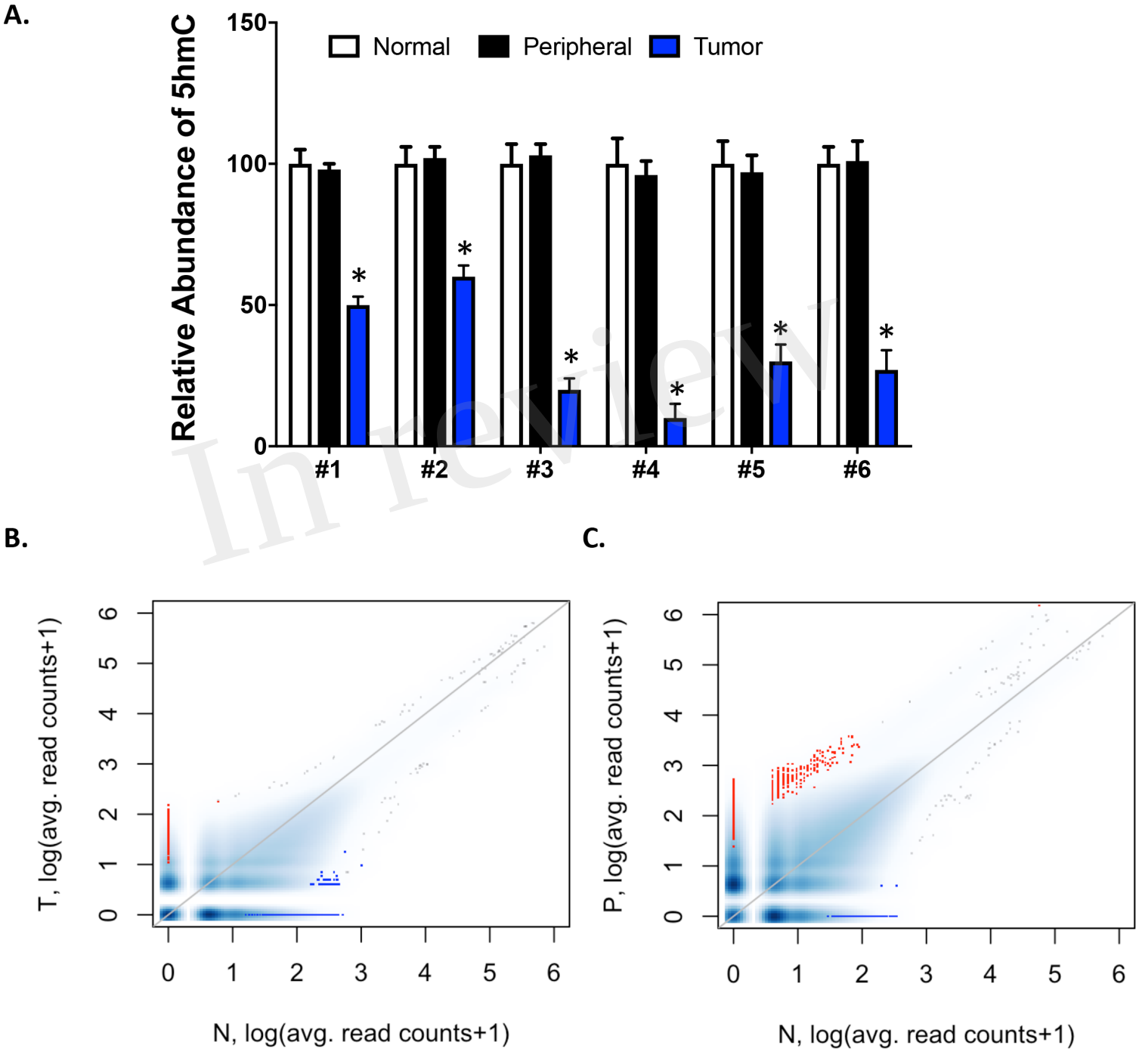


Figure 2.

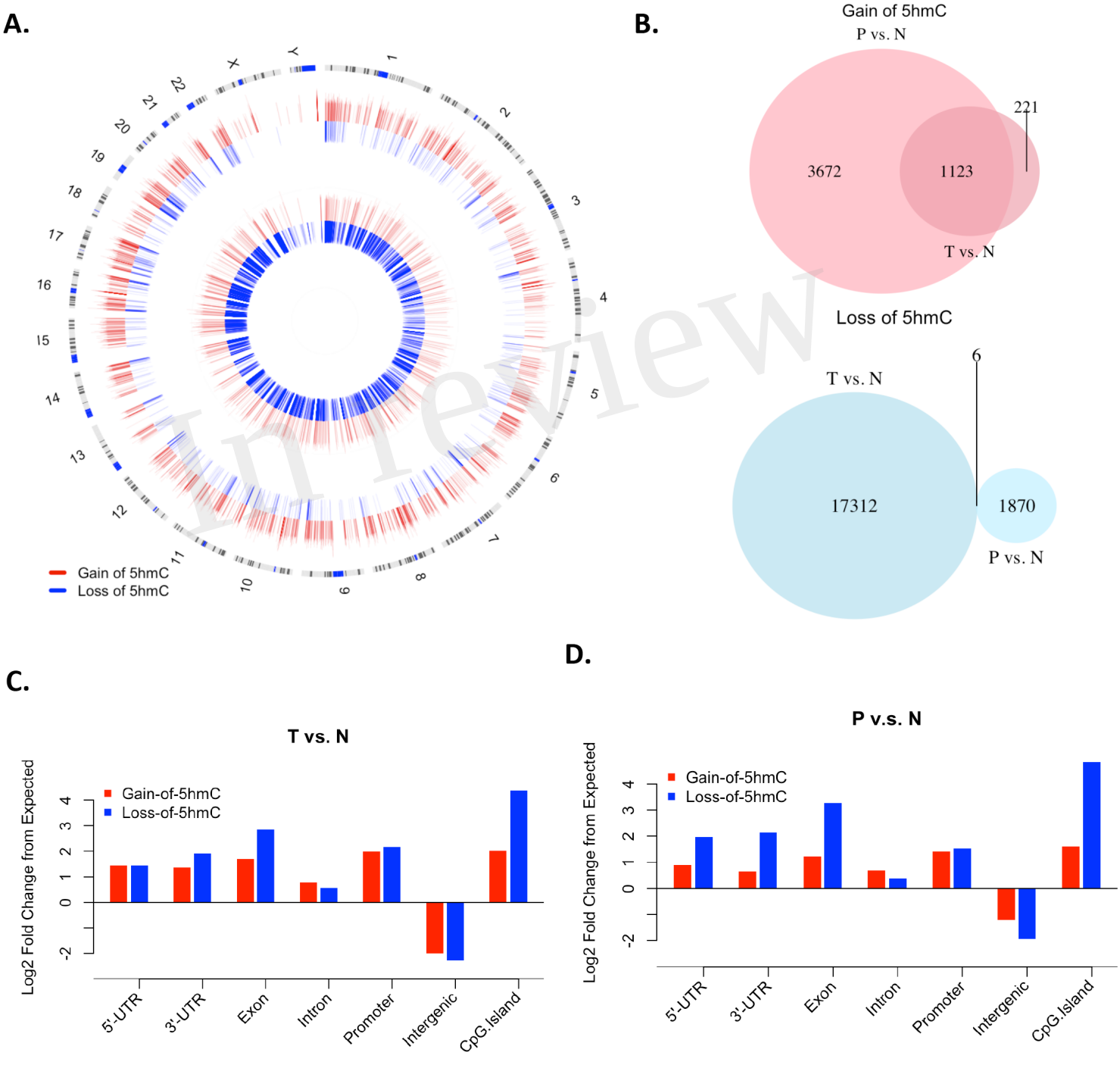
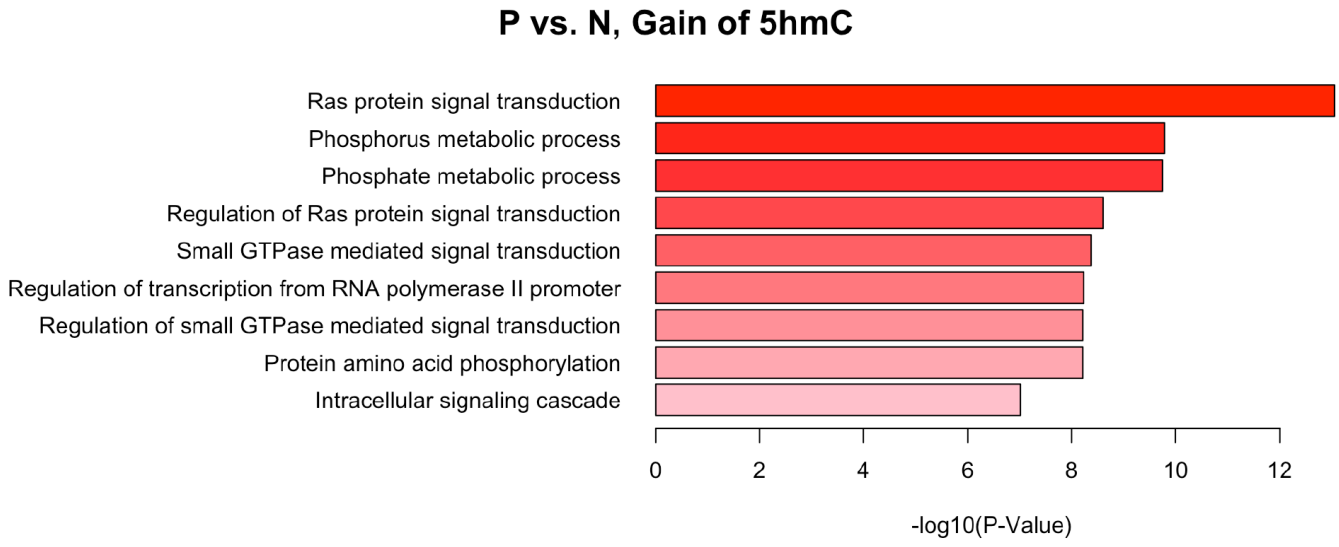


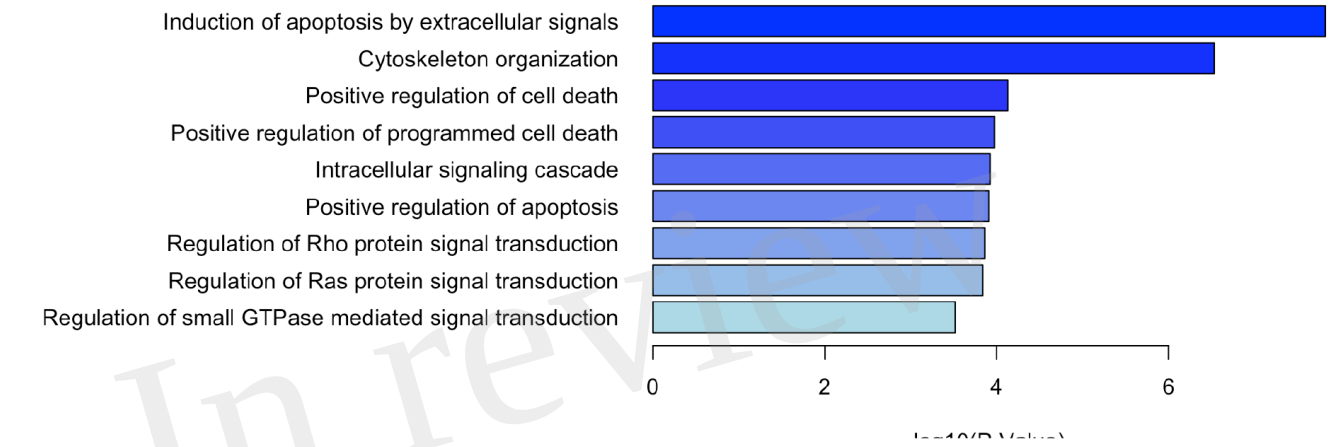
Figure 3.

Figure 3.TIF

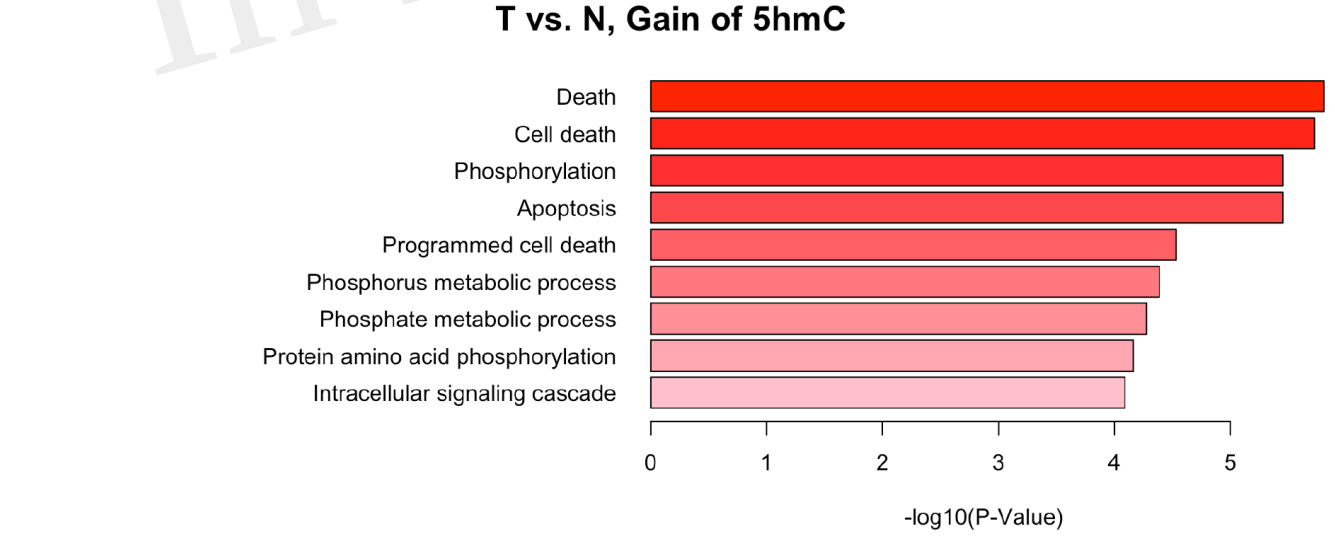
A.



B.



C.



D.

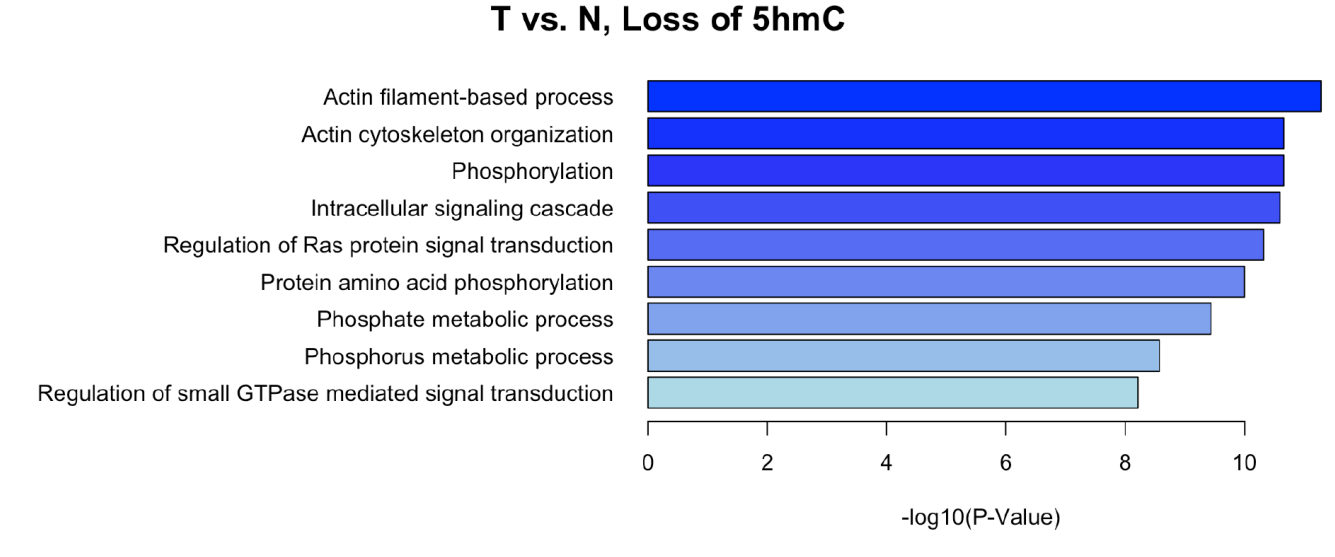


Figure 4.

A.

Transcription Factor	Motif
HIF-1b	
USF1	

B.

Transcription Factor	Motif
c-Myc	
HIF-1b	
n-Myc	
CLOCK	
MAX	
bHLHE40	
USF1	
BMAL1	
ZFX	
ZNF711	

Figure 5.

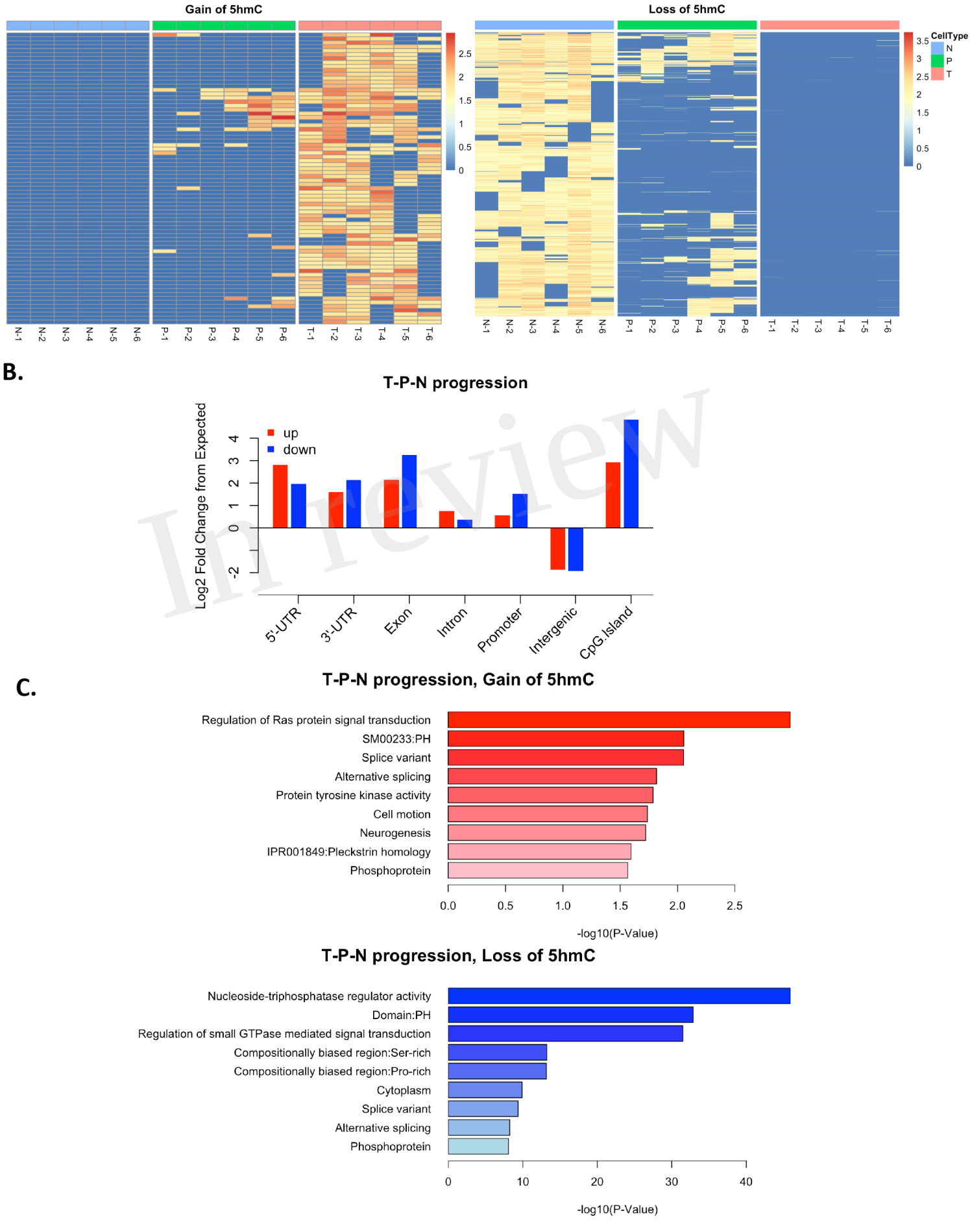


Figure 6.TIF

Figure 6.