

FIGURE LEGENDS

Figure S1. Schema of HELP-GT assay. HELP-tagging libraries were prepared by the process described previously (Suzuki et al. 2010) with some modifications as follows. Based on the assumption that most of the hydroxymethylated cytosines come from methylated cytosines, we added a step of β -GT prior to MspI or HpaII digestions (Song, 2010; Josse and Kornberg, 1962). The final library was sequenced by multiplexing in-house adapter primers using Illumina HiSeq 2000.

Figure S2. 5-hmC and 5-mC estimation by the HELP-GT assay are reproducible. Unsupervised clustering of two independent replicates shows high degree of concordance for both 5-hmC and 5-mC analysis (A, C). A high degree of concordance was also seen in gene specific localization of 5-hmC and 5-mC sites for replicates (B, D) as shown for the representative control sample.

Figure S3. 5-hmC correlates with gene expression at proximal and intragenic regions: 5-hmC and 5-mC loci were mapped relative to RefSeq transcripts expressed at different levels in pancreatic cancer (Panc Ca1) cells. RefSeq transcripts were divided into two bins based on gene expression level and 5-hmC or 5-mC genomic loci reads falling in 10-bp bins centered on transcription start sites or end sites. Proximal and intragenic enrichment of 5-hmC is seen in highly expressed genes (A, B). 5-mC levels are decreased around TSS and enriched in intragenic areas for highly expressed genes. (C, D).

Figure S4. Acquisition of 5-hmC at different promoters *MMP11* (A) *VAV2* (B) and *LATS2* (C) show increased 5-hmC at promoter regions as shown by brown marks. 5-mC marks are shown as downward blue lines. The scale is from 0 to 100. Top panel shows RNA-seq data demonstrating increased expression of *MMP11*, *VAV2* and *LATS2* in pancreatic cancer cells. *RBM38* (D) shows increased 5-hmC at promoter regions as shown by brown marks. 5-mC marks are shown as downward blue lines. The scale is from 0 to 100. Top panel shows RNA-seq data demonstrating increased expression of *RBM38* in pancreatic cancer cells.

Figure S5. 5hmC loci are distributed throughout the genome Circos plots of 5-hmC loci (orange bars) and gene expression (by RNA seq, green bars) show genome wide prevalence of 5-hmC loci in both control (A) and cancer cell lines Panc Ca1 (B) and Panc Ca2 (C). The length of the bars corresponds to amount of 5-hmC and magnitude of expression.

Table S1:

Genomic locations and primers for 5-hmC validations:

Chromosome	Position	Gene	Fwd primer	Rev primer
chr19	39658315	PAK4	CTTGGGACCAGCTGAGACAG	CCCAGAGCTAGGACCATCCA
chr21	38603328	TMPRSS3a	GGACAAGTGTCCCACGTCAT	ACCTGTTCTCCTGGTCCTCA
chr6	170594425	DLL1	TGGGGTCACACTCGTCAATC	AGGGGAGCTACACTTGCTCT
chr9	136748793	VAV2c	CCCAAGTCATCCCCAGCC	ATTGCACAGGTGACCGGAAG
chr7	923330	GET4d	CCCTATCTGGAAGGCACGTC	GGTGACTAAGCCAGGACAGG
chr22	24111338	MMP11	TCTGGGTGACTTTGCAGTGG	AGCAGTGATGAGGGACAGT
chr16	15028924	NIPIa	GACGGTGCAGATGTCCCATA	TGCTCAGGACAGGGATGAGA
chr16	88990095	CBFA2T3	AGCAGCTGGGTTGTGAAGG	GTTCCAAAGCCGACCTCCAT
chr9	136746964	VAV2a	TACTACTCCAGGCAGCGTGT	ACTTGCTGCGTGTCTGATCT
chr10	43614886	RET	TGTGTCCACCCCCTTACTCA	CAGCTTGGGTCTTCCAGGAG
chr9	136844169	VAV2e	TTTAGGGTCTCCTCCAGCCA	ACGAAAGCTCTTCCGAAGCC
chr9	136845219	VAV2f	AGTAGGGTGGAGAACAGGCT	GAGACAGAGAACCAGGTGCC
chr7	922846	GET4e	GTGACTCCTGGCAGAGCAG	ATTTACAAGGACCCCTCCG
chr19	5041526	KDM4B	AGCCAGGCTTTCTCTGATTTT	GTTTGTAAACACGCTGTGGGC
chr15	40396398	BMF	CAGCAGACTCAACCCTCCTC	TCAGTGCATTGCAGACCAGT

Table S2:

Genomic distribution of HpaII/MspI sites

Genomic Position	Number of sites
Total	2,326,002
CpG islands	288,413
CpG shores	255,321
Promoter	313,138
Gene Body	950,168

Table S3:

MspI, HpaII and β GT-MspI counts of Panc Ca1 samples for the loci that were validated to have 5hmC marks.

Gene Name	HpaII counts	GT-MspI counts	MspI counts
VAV2	1	1	3
CBFA2T3	4	3	14
MMP11	1	1	4
NIPI	3.285714286	3.142857143	8.714285714
PAK4	4	5	14
DLL1	1	5	14
RET	3	2	12
TMPRSS3	1	3	6
GET4	0.333333333	2.333333333	8.666666667

Figure S1

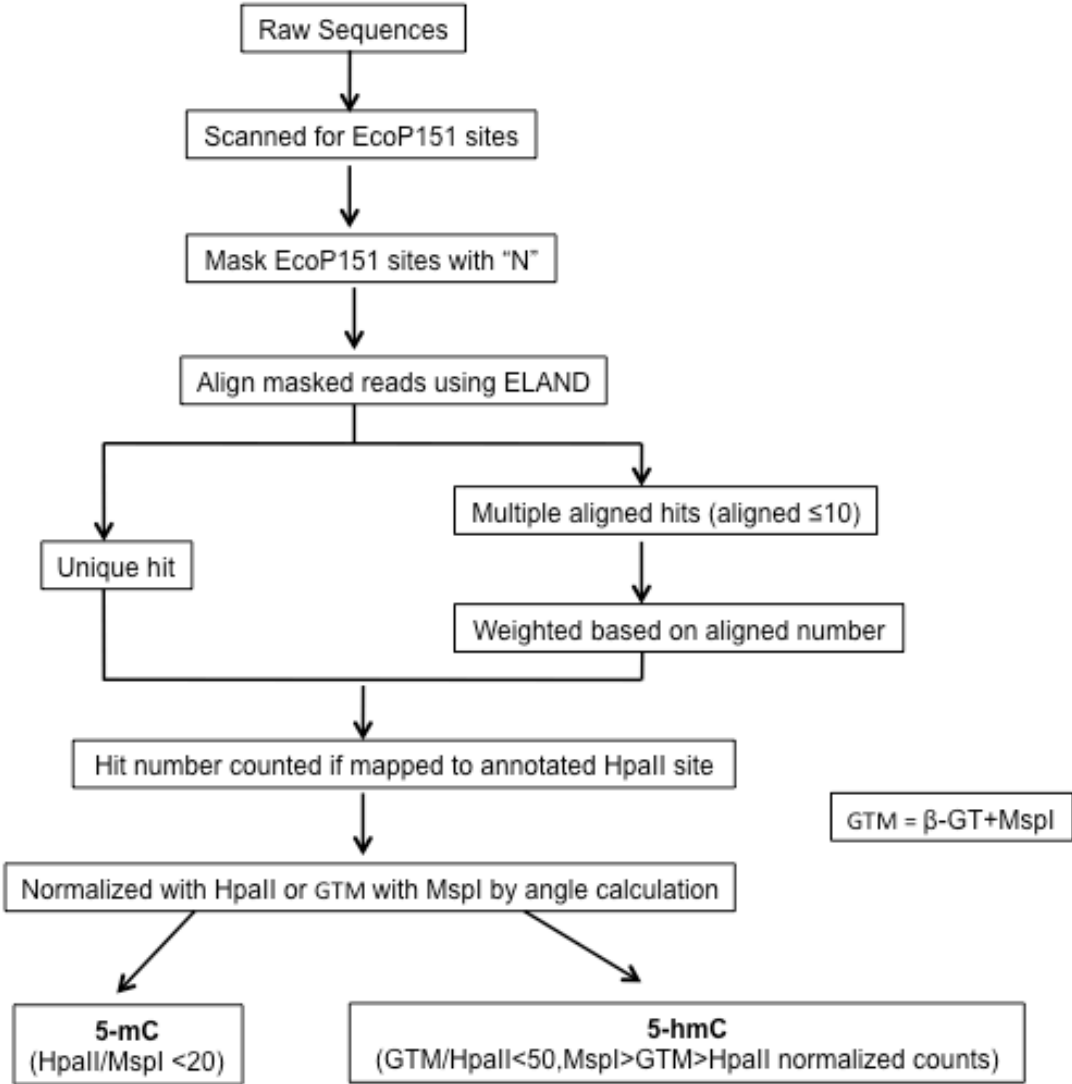


Figure S2

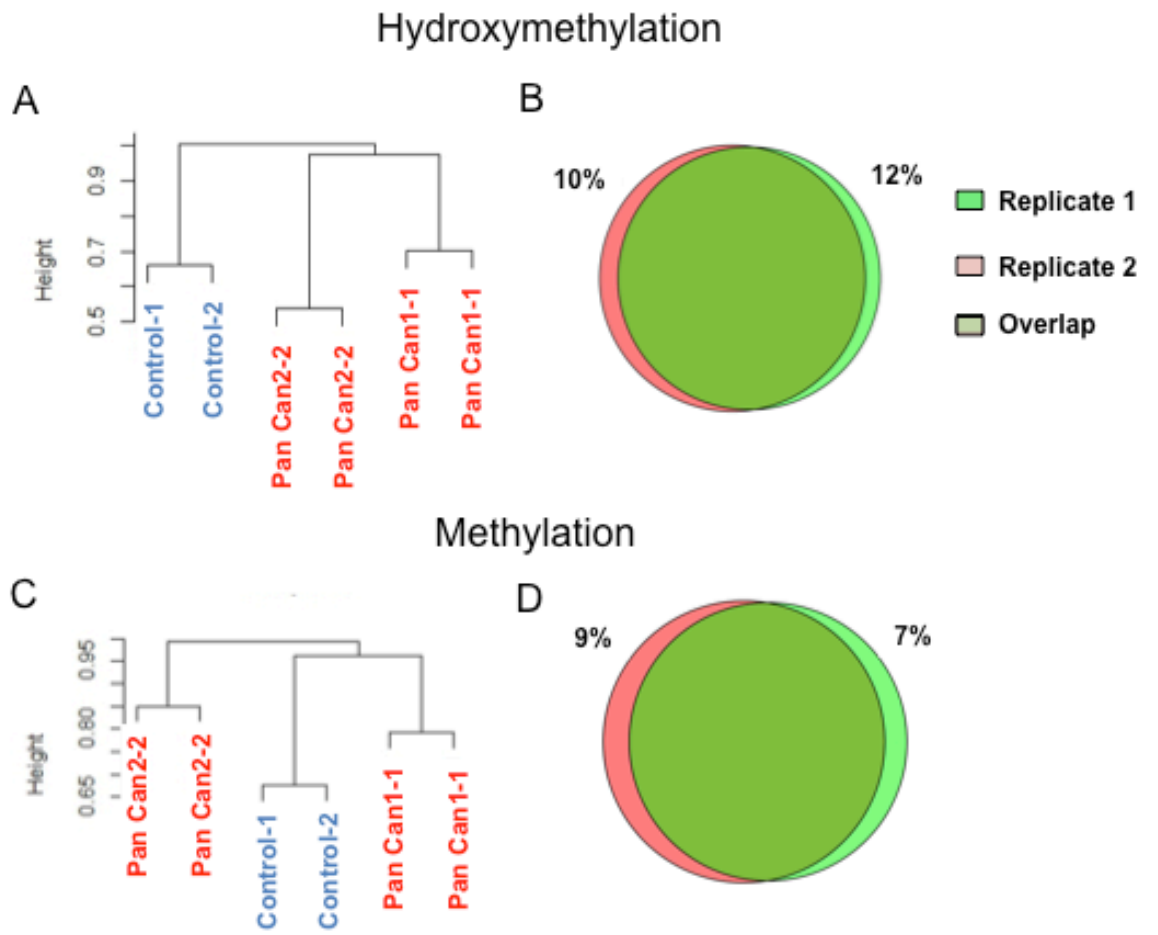


Figure S3

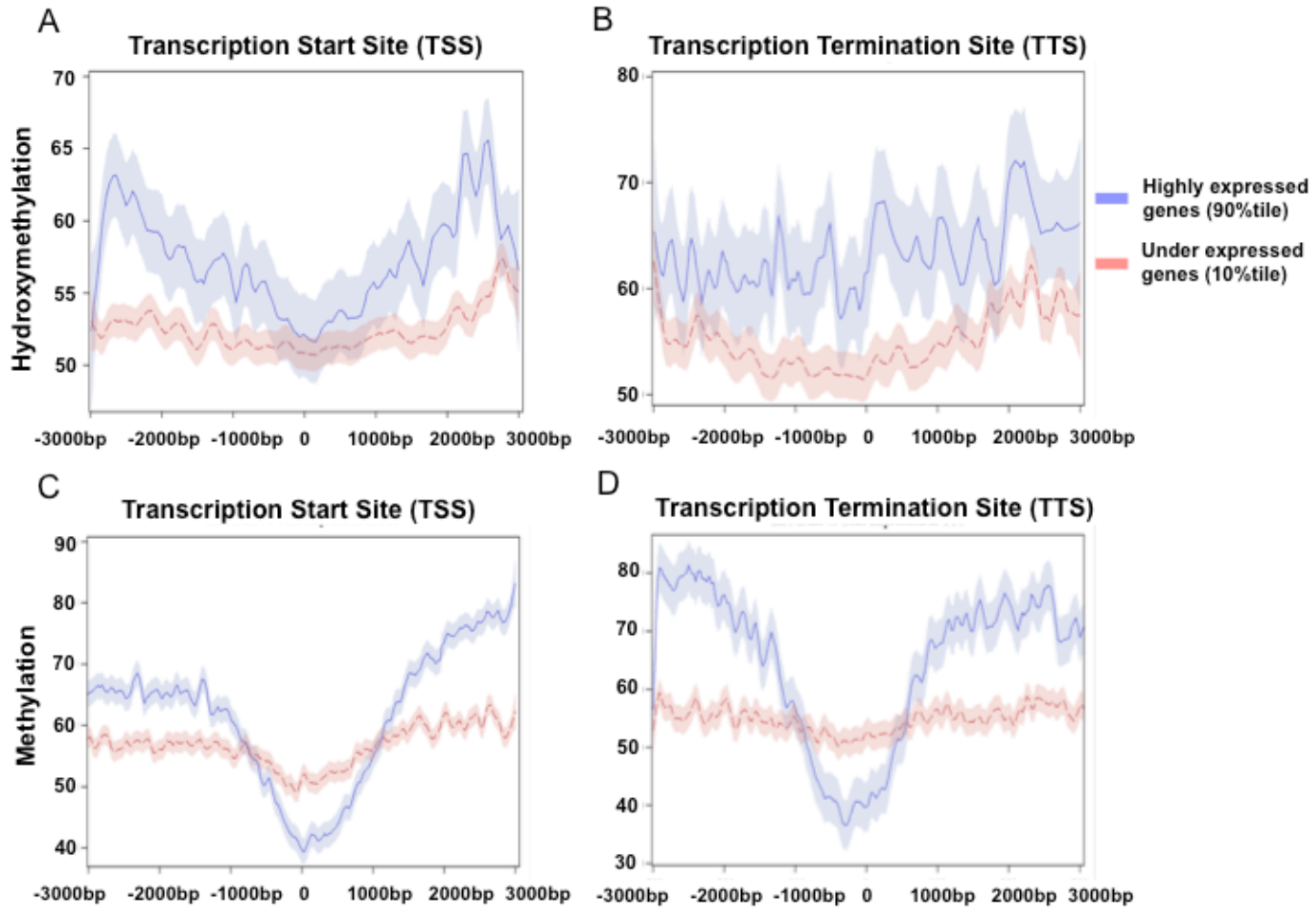


Figure S4

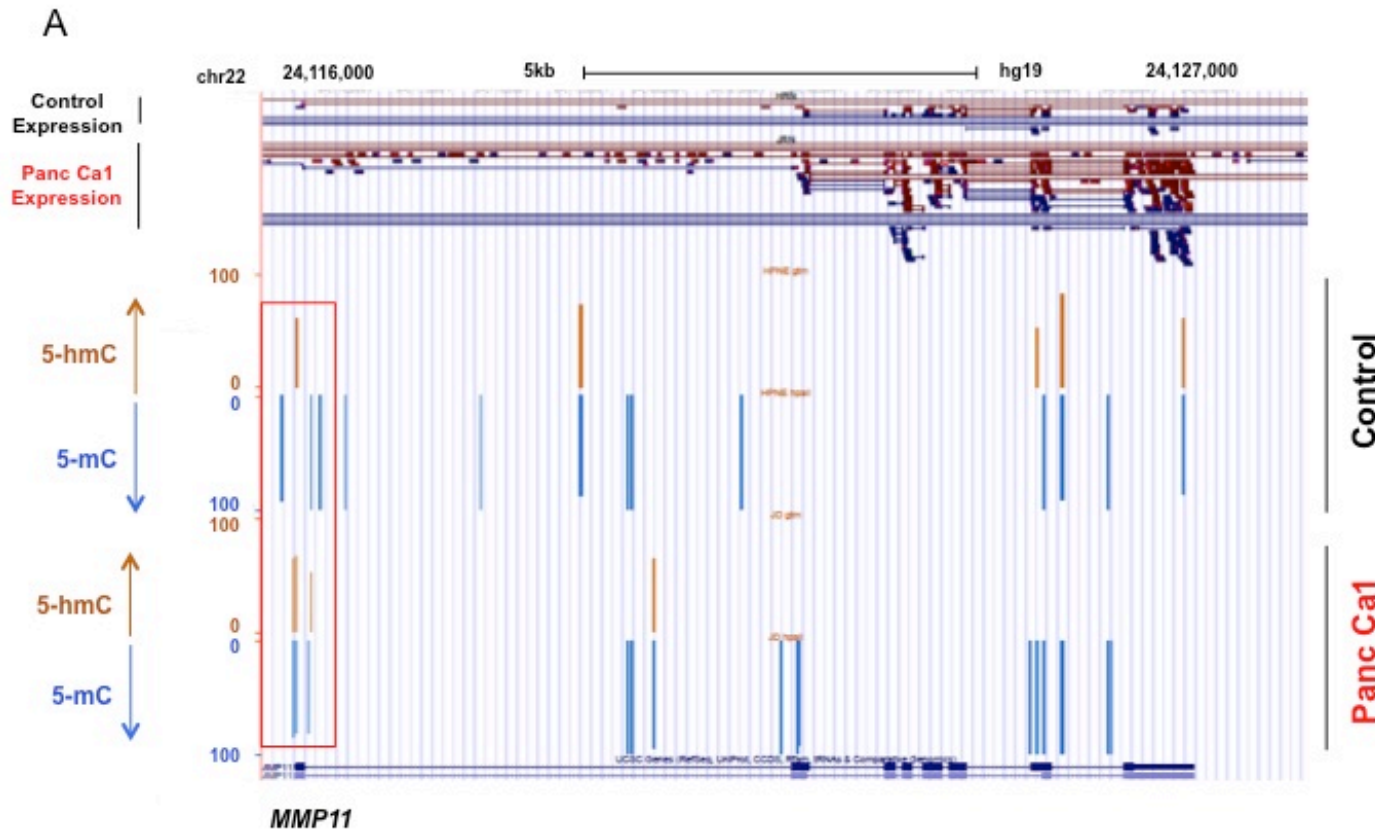


Figure S4

B

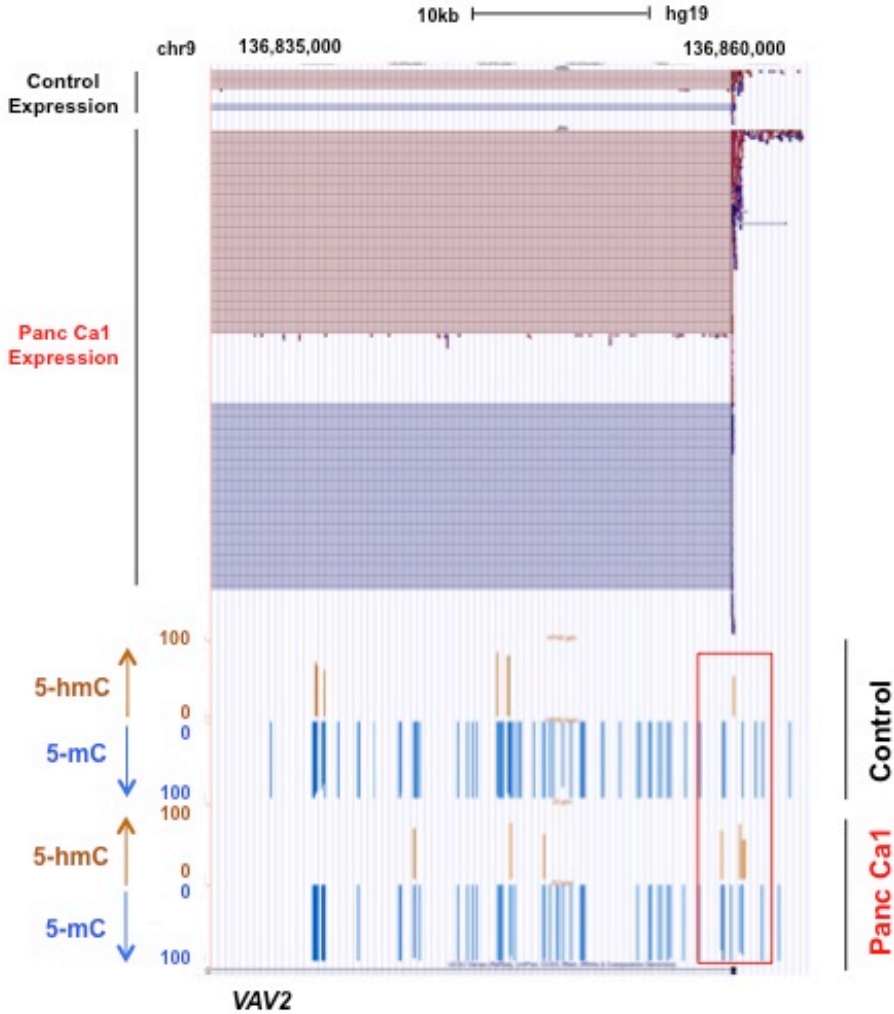


Figure S4

C

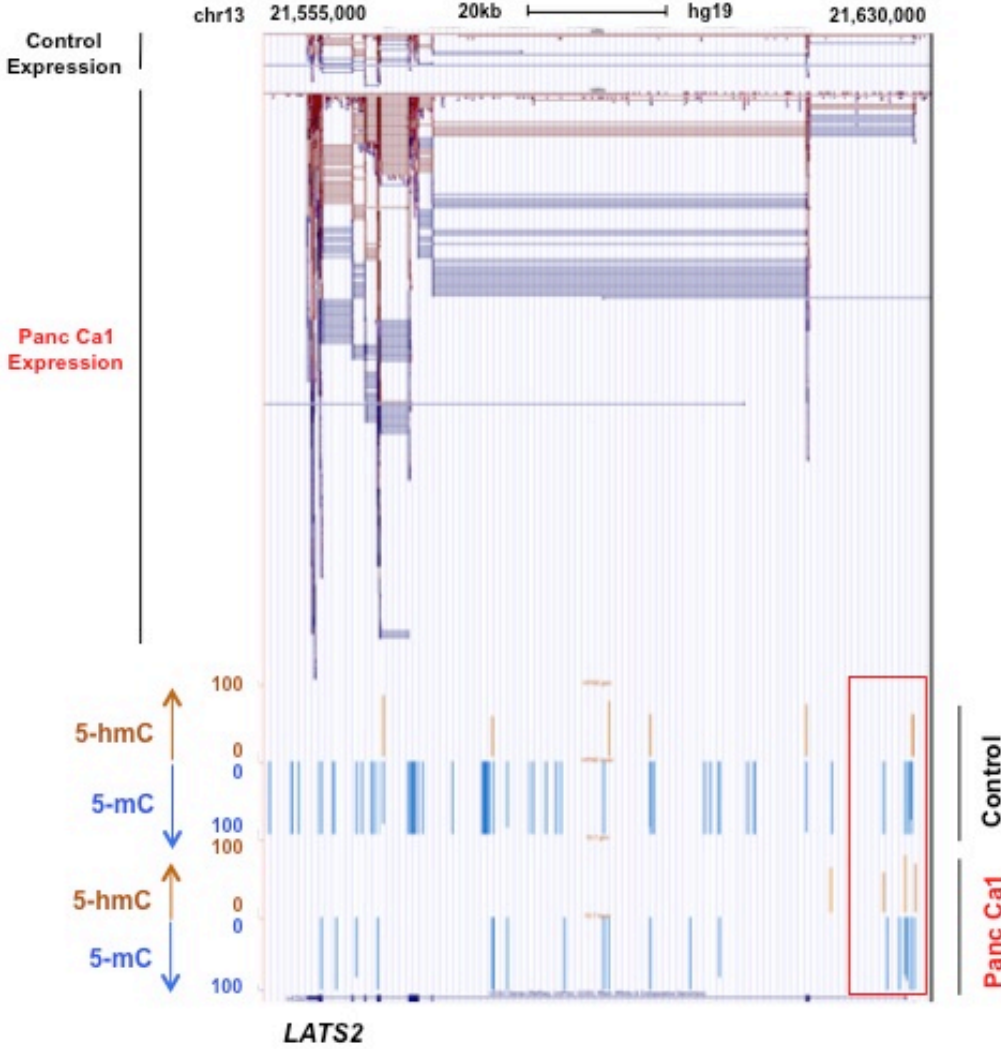


Figure S5

A

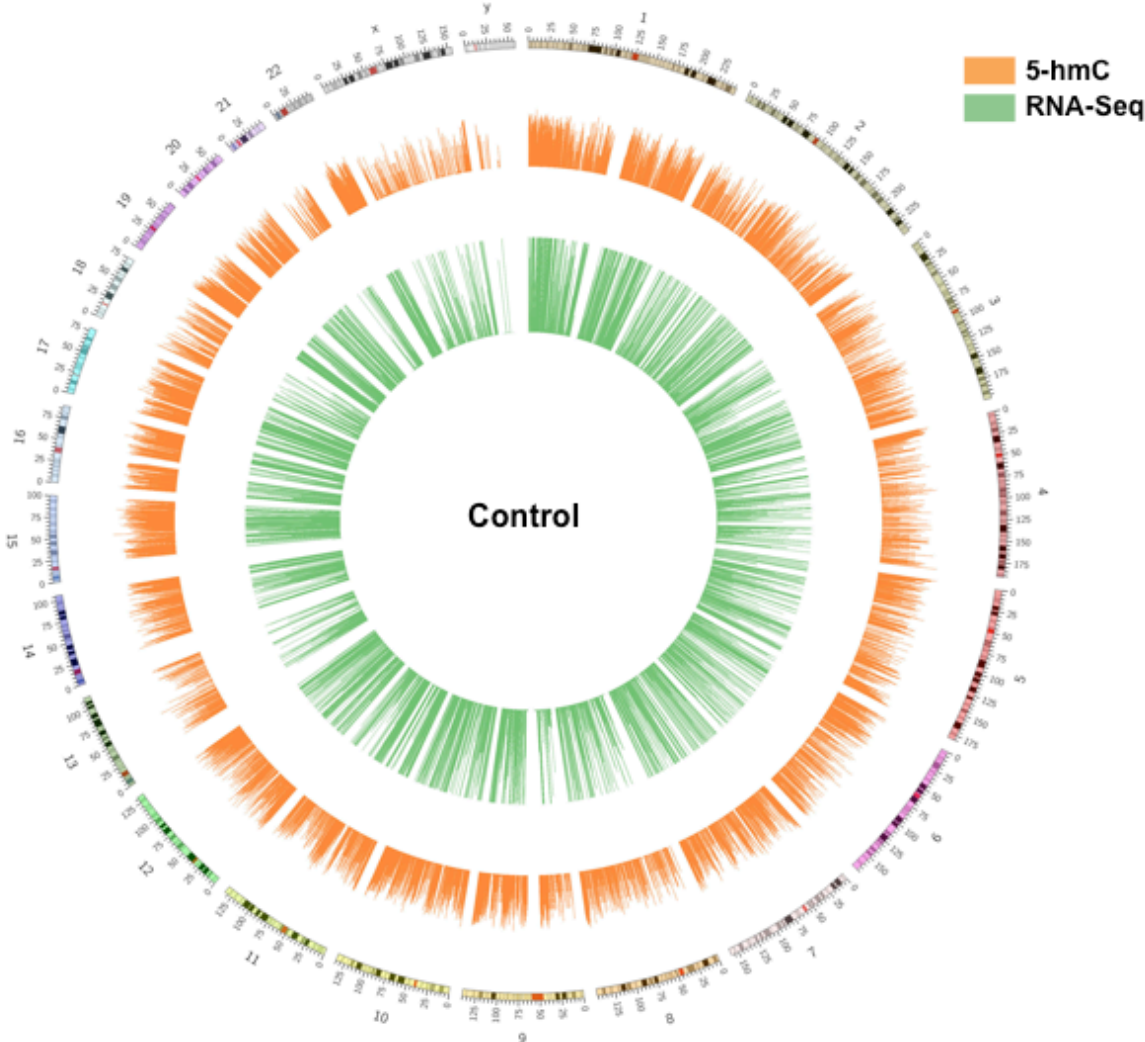


Figure S5

B



Figure S5

C

