Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/4/9/eaau6190/DC1

Supplementary Materials for

A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons

Alexey Kozlenkov, Junhao Li, Pasha Apontes, Yasmin L. Hurd, William M. Byne, Eugene V. Koonin, Michael Wegner, Eran A. Mukamel*, Stella Dracheva*

*Corresponding author. Email: emukamel@ucsd.edu (E.A.M.); stella.dracheva@mssm.edu (S.D.)

Published 26 September 2018, *Sci. Adv.* **4**, eaau6190 (2018)
DOI: 10.1126/sciadv.aau6190

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/4/9/eaau6190/DC1)

Figure S1

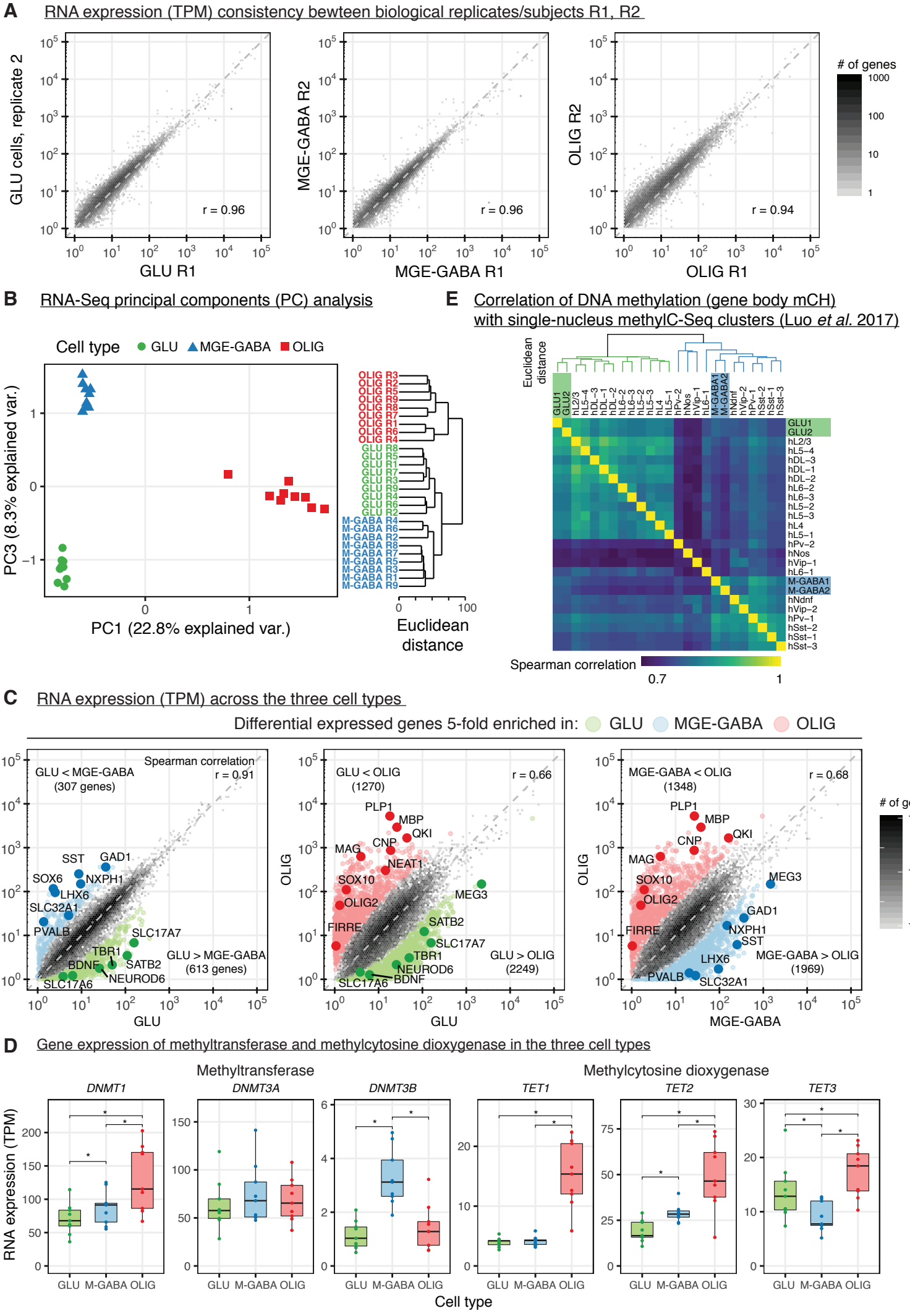
A
RNA expression (TPM) consistency bewteen biological replicates/subjects R1, R2
GLU cells, replicate 2
GLU R1
r = 0.96
MGE-GABA R2
MGE-GABA R1
r = 0.96
OLIG R2
OLIG R1
r = 0.94
# of genes
1000
100
10
1
B
RNA-Seq principal components (PC) analysis
Cell type
GLU
MGE-GABA
OLIG
PC3 (8.3% explained var.)
PC1 (22.8% explained var.)
Euclidean distance
E
Correlation of DNA methylation (gene body mCH) with single-nucleus methylC-Seq clusters (Luo et al. 2017)
Euclidean distance
Spearman correlation
0.7
1
C
RNA expression (TPM) across the three cell types
Differential expressed genes 5-fold enriched in:
GLU
MGE-GABA
OLIG
GLU < MGE-GABA (307 genes)
Spearman correlation r = 0.91
GLU > MGE-GABA (613 genes)
GLU < OLIG (1270)
r = 0.66
GLU > OLIG (2249)
MGE-GABA < OLIG (1348)
r = 0.68
MGE-GABA > OLIG (1969)
# of genes
100
10
1
D
Gene expression of methyltransferase and methylcytosine dioxygenase in the three cell types
Methyltransferase
Methylcytosine dioxygenase
DNMT1
DNMT3A
DNMT3B
TET1
TET2
TET3
RNA expression (TPM)
GLU M-GABA OLIG
Cell type

Fig. S1. RNA-seq validates the identities of FANS-sorted populations. (A) High consistency of gene expression across biological replicates/subjects. Transcripts Per Million (TPM) values of genes in two donors (H372, R1 and H276, R2) are shown for the three cell types. r, Spearman correlation coefficient. **(B)** Principal components analysis using $\log_{10}$TPM of all genes shows clear separation of samples from the three cell types. The dendrogram to the right was made using hierarchical clustering with complete linkage and Euclidean distance of the Spearman correlation coefficients of $\log_{10}$TPM across all samples; n=9 samples per each cell type. **(C)** Pairwise comparison of gene expression between the three cell types. TPM values are shown in the scatter plots, and differentially expressed genes called by edgeR (FDR<0.05) with at least 5-fold enrichment are highlighted. Examples of marker genes in each cell type are labelled. r, Spearman correlation coefficient. **(D)** Gene expression of DNA methyltransferases (DNMT1, DNMT3A and DNMT3B) and methylcytosine dioxygenases (TET1, TET2 and TET3) for each cell type. Dots overlaying the boxplot show expression levels in each individual sample, and significant differential expression (reported by edgeR using raw read counts, FDR<0.05) is marked with asterisks. **(E)** Heatmap showing Spearman correlation coefficients across Glu and MGE-GABA samples in the current dataset (shaded in green and blue) and the cell sub-clusters identified in single-nucleus methylC-Seq *(Luo et al. 2017)*. Correlations were calculated using gene body non-CpG methylation levels (mCH). Glu (GLU), glutamatergic neurons; MGE-GABA (or M-GABA), medial ganglionic eminence derived GABAergic interneurons; OLIG, oligodendrocytes.

Figure S2

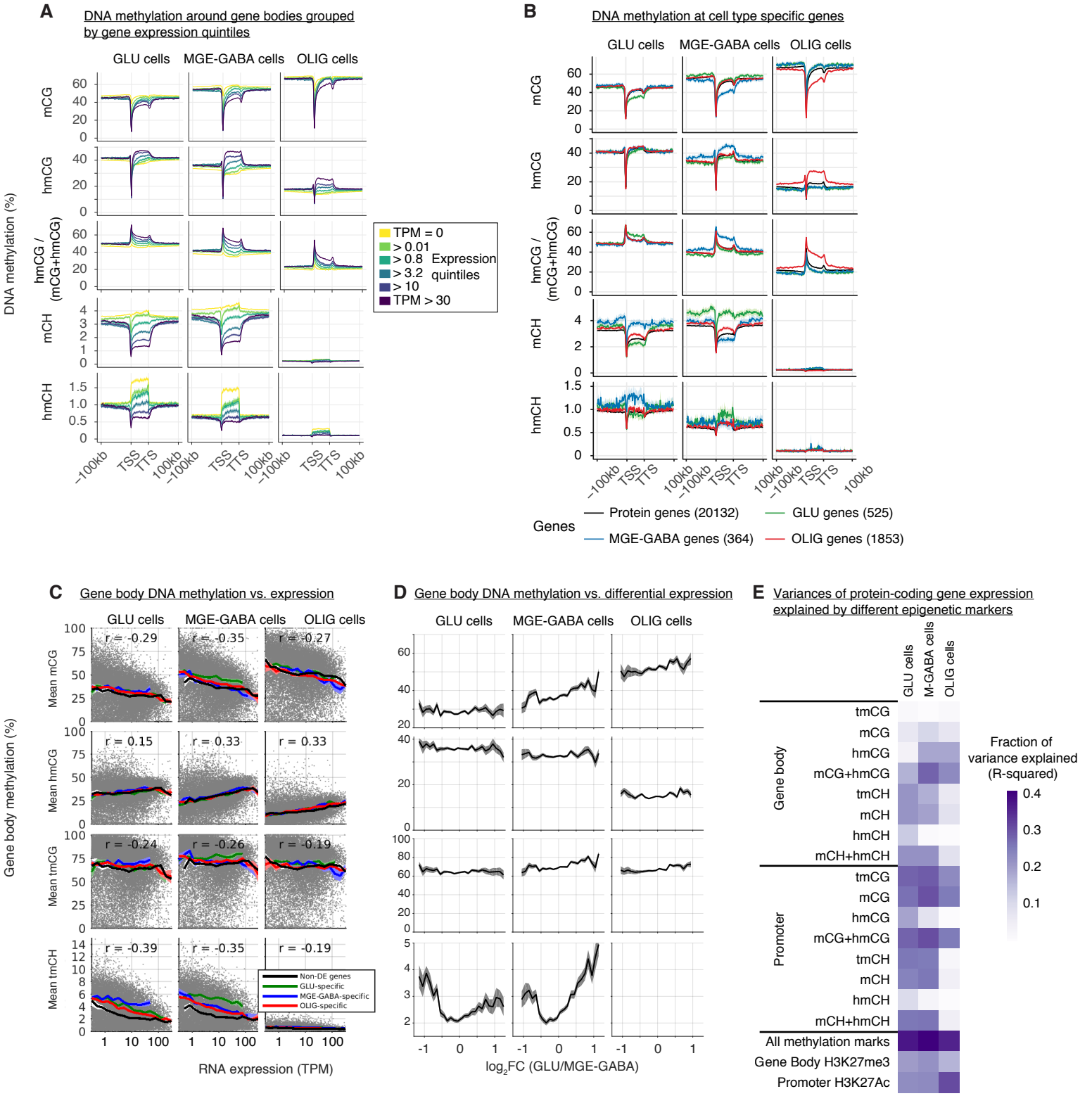
A
DNA methylation around gene bodies grouped by gene expression quintiles
GLU cells
MGE-GABA cells
OLIG cells
DNA methylation (%)
mCG
hmCG
hmCG / (mCG+hmCG)
mCH
hmCH
TPM = 0
> 0.01
> 0.8
> 3.2
> 10
TPM > 30
Expression quintiles
−100kb
TSS
TTS
100kb
B
DNA methylation at cell type specific genes
Genes
Protein genes (20132)
GLU genes (525)
MGE-GABA genes (364)
OLIG genes (1853)
C
Gene body DNA methylation vs. expression
Gene body methylation (%)
Mean mCG
Mean hmCG
Mean tmCG
Mean tmCH
r = -0.29
r = -0.35
r = -0.27
r = 0.15
r = 0.33
r = 0.33
r = -0.24
r = -0.26
r = -0.19
r = -0.39
r = -0.35
r = -0.19
Non-DE genes
GLU-specific
MGE-GABA-specific
OLIG-specific
RNA expression (TPM)
D
Gene body DNA methylation vs. differential expression
log2FC (GLU/MGE-GABA)
E
Variances of protein-coding gene expression explained by different epigenetic markers
GLU cells
M-GABA cells
OLIG cells
Gene body
Promoter
tmCG
mCG
hmCG
mCG+hmCG
tmCH
mCH
hmCH
mCH+hmCH
All methylation marks
Gene Body H3K27me3
Promoter H3K27Ac
Fraction of variance explained (R-squared)
0.4
0.3
0.2
0.1

Fig. S2. Cell type–specific gene body DNA methylation patterns in neurons and OLIG associate with expression and differential expression. (A, B) DNA methylation profiles around gene bodies grouped by gene expression quintiles in each cell type (A) and cell type-specific gene sets (B). Gene bodies of different genes were binned into 30 bins of equal length, normalized by length of each gene. Flanking regions were extended up to 100 kb from both ends of the genes and binned into bins of 2 kb. Y-axis shows the average methylation levels of these bins in each gene set. In (B), Glu-specific genes and MGE-GABA-specific genes were defined as differential expressed genes (FDR<0.05) in the Glu vs. MGE-GABA comparison with at least 2-fold differences; OLIG-specific genes were defined as differential expressed genes (FDR<0.05) that have 2-fold higher expression in OLIG in both OLIG vs. Glu and OLIG vs. MGE-GABA comparisons. TSS, transcription start site; TTS, transcription termination site. **(C, D)** Gene body methylation levels as a function of gene expression in each cell type (C) and a function of differential expression fold changes (FC) between Glu and MGE-GABA cells (D). In (C) the relationship between gene expression (RNA-Seq) and DNA methylation is shown for every gene (gray points); cell type-specific genes were defined as in (B), and non-DE genes were defined as genes that were not differentially expressed (FDR>0.05) with FC<1.1 in the Glu vs. MGE-GABA comparison. Solid lines show the mean methylation levels and the shaded areas denote standard errors of the means. r, Spearman correlation ($p < 10^{-10}$). tmCH levels in OLIG cells were too low and hence not shown in (D). **(E)** Epigenetic markers were used in linear regression to model gene expression ($\log_{10}$(TPM)) in each of the three cell types. Fractions of variance explained (R^2) from the models that used different features are shown in the heatmap.

Figure S3

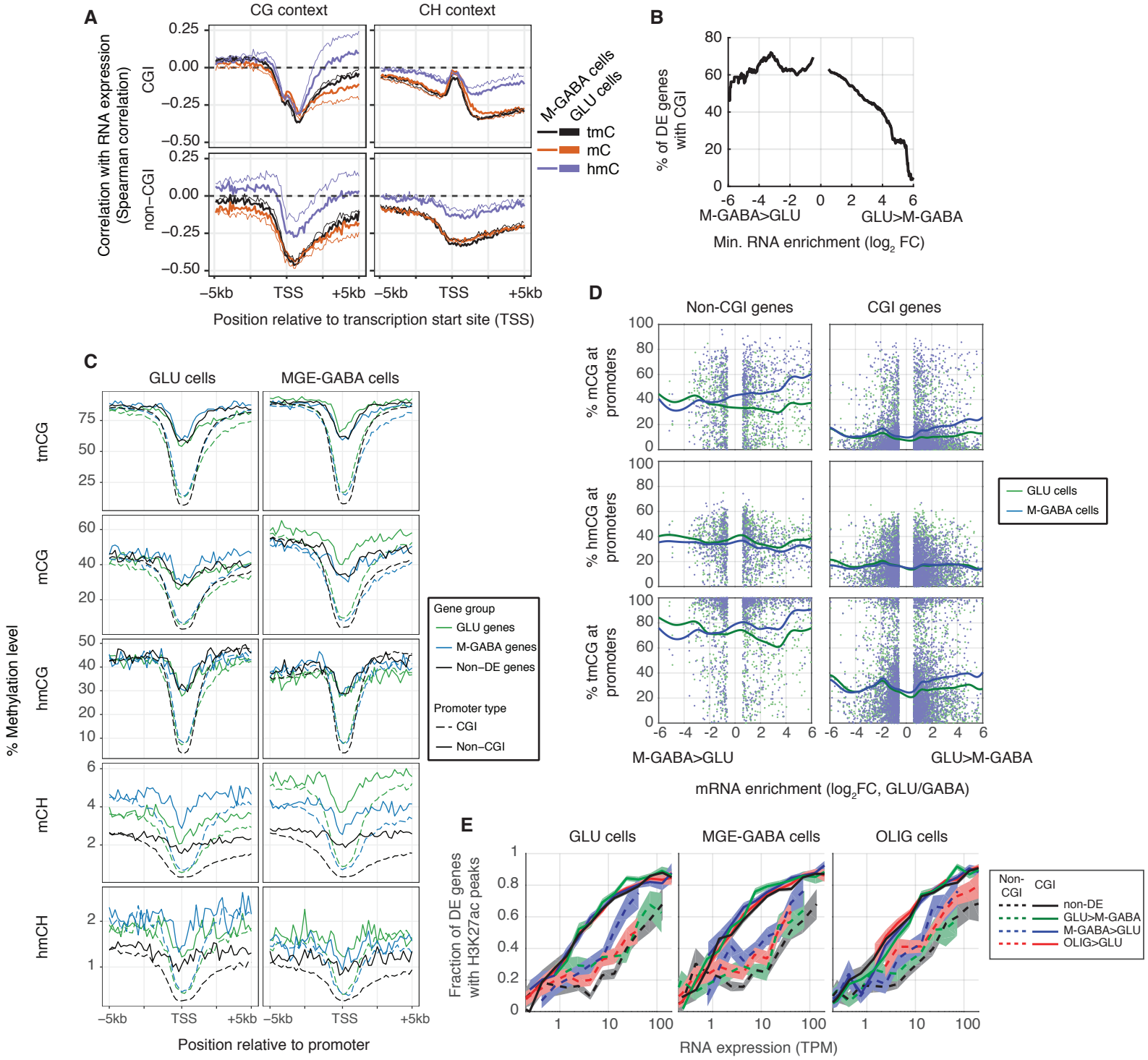
A
CG context
CH context
Correlation with RNA expression (Spearman correlation)
CGI
non−CGI
M-GABA cells
GLU cells
tmC
mC
hmC
−5kb
TSS
+5kb
Position relative to transcription start site (TSS)
B
% of DE genes with CGI
M-GABA>GLU
GLU>M-GABA
Min. RNA enrichment (log2 FC)
C
GLU cells
MGE-GABA cells
% Methylation level
tmCG
mCG
hmCG
mCH
hmCH
Gene group
GLU genes
M-GABA genes
Non−DE genes
Promoter type
CGI
Non−CGI
Position relative to promoter
D
Non-CGI genes
CGI genes
% mCG at promoters
% hmCG at promoters
% tmCG at promoters
GLU cells
M-GABA cells
M-GABA>GLU
GLU>M-GABA
mRNA enrichment (log2FC, GLU/GABA)
E
GLU cells
MGE-GABA cells
OLIG cells
Fraction of DE genes with H3K27ac peaks
RNA expression (TPM)
Non-CGI
CGI
non-DE
GLU>M-GABA
M-GABA>GLU
OLIG>GLU

Fig. S3. Association of epigenetic modifications at promoters with gene expression in different brain cell types. (A) Spearman correlation between methylation levels in 100 bp bins flanking TSS and expression of protein-coding genes. mCG is most strongly correlated with expression in the region ~1 kb downstream of the TSS for CGI and non-CGI genes. CGI genes were defined as genes that contain at least one CpG island in the TSS ±2 kb regions of any of their transcripts. Non-CGI genes were defined as genes that do not have any CpG island within the TSS+/-10kb regions of all of their transcripts. TSS, transcription start site. **(B)** Compared with MGE-GABA-enriched genes, Glu-enriched genes were depleted for CGI promoters. The minimum $\log_2$FC was the threshold used to select Glu-enriched and MGE-GABA-enriched genes, and the proportion of genes that have CpG islands within the TSS+/-2kb region among the selected genes is shown. **(C)** Methylation profiles around TSS. Mean methylation levels in 200 bp bins flanking TSS are shown. Gene groups were defined as in Fig. S2C. Promoter types were defined as in (A). MGE-GABA cells have increased mCG at Glu-enriched genes, but the converse is not found. **(D)** Promoter methylation levels as a function of gene differential expression fold-change. Methylation levels of differentially expressed genes (FDR<0.05) in Glu and MGE-GABA cells are shown in green and blue dots, respectively, and the curves are cubic smoothing splines generated from the dots. Promoter regions were defined as TSS+/-2kb. **(E)** Gene expression vs. the presence of H3K27ac peak in the promoter region. In all three cell types, gene expression correlates with H3K27ac at both non-CGI and CGI promoters. Compared to CGI genes, a significantly lower fraction of non-CGI genes contained H3K27ac peaks in their promoters. Gene groups were defined as in Fig.S2C. Promoter types were defined as in (A).

Figure S4

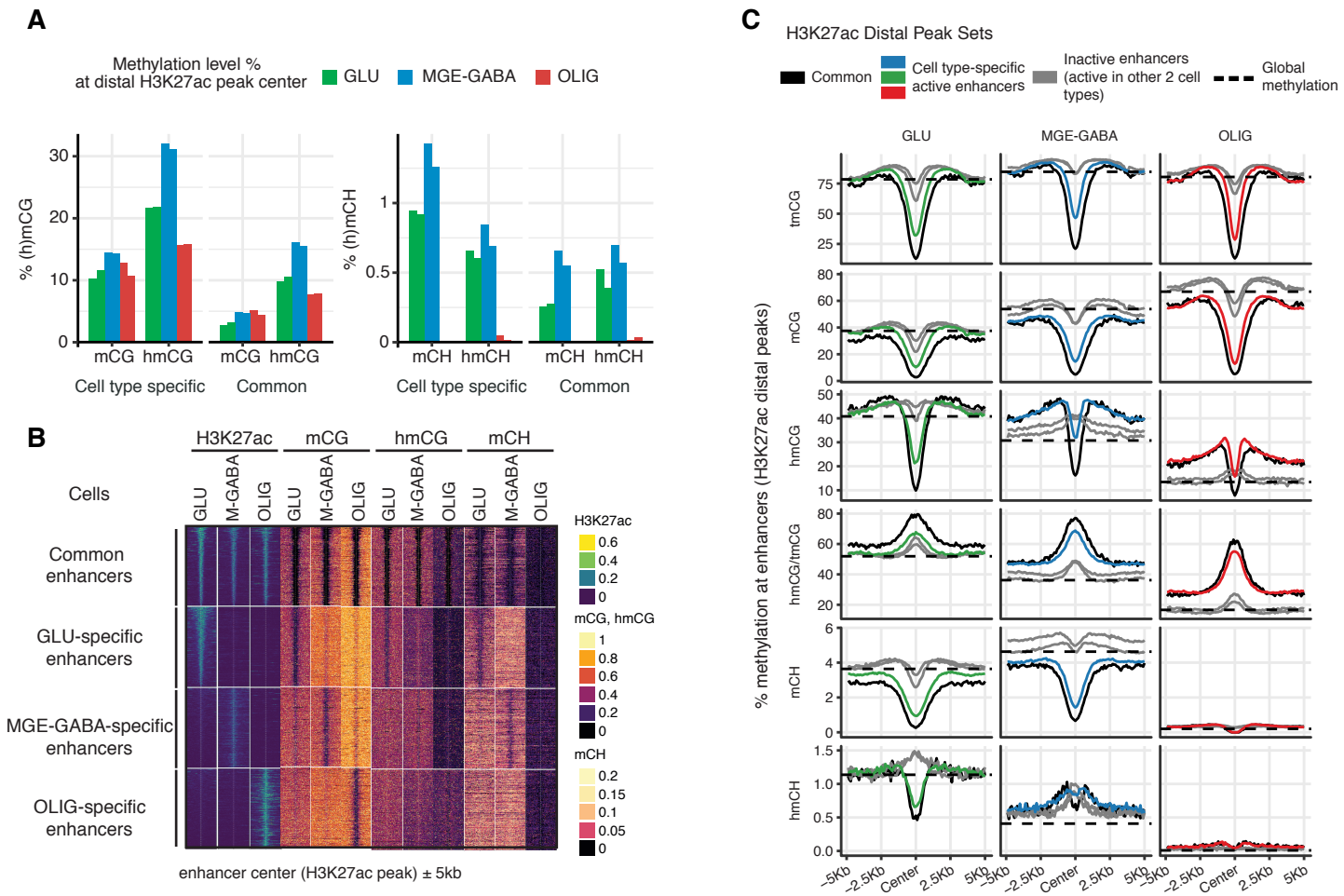

Fig. S4. DNA methylation profiles at cell type–specific distal gene regulatory elements. (A) DNA methylation levels in the three cell types at the center of cell type-specific and common enhancers (distal H3K27ac peaks). MGE-GABA cells are enriched for hmCG at cell type-specific peaks. **(B)** Heatmap of H3K27ac enrichment and methylation levels at +/-5kb regions flanking the cell type-specific and common enhancers. **(C)** Mean profiles of DNA methylation marks in 100 bp bins flanking distal H3K27ac peak centers. Genome-wide methylation levels (as in Fig.1B) are shown in horizontal dashed lines.

Figure S5

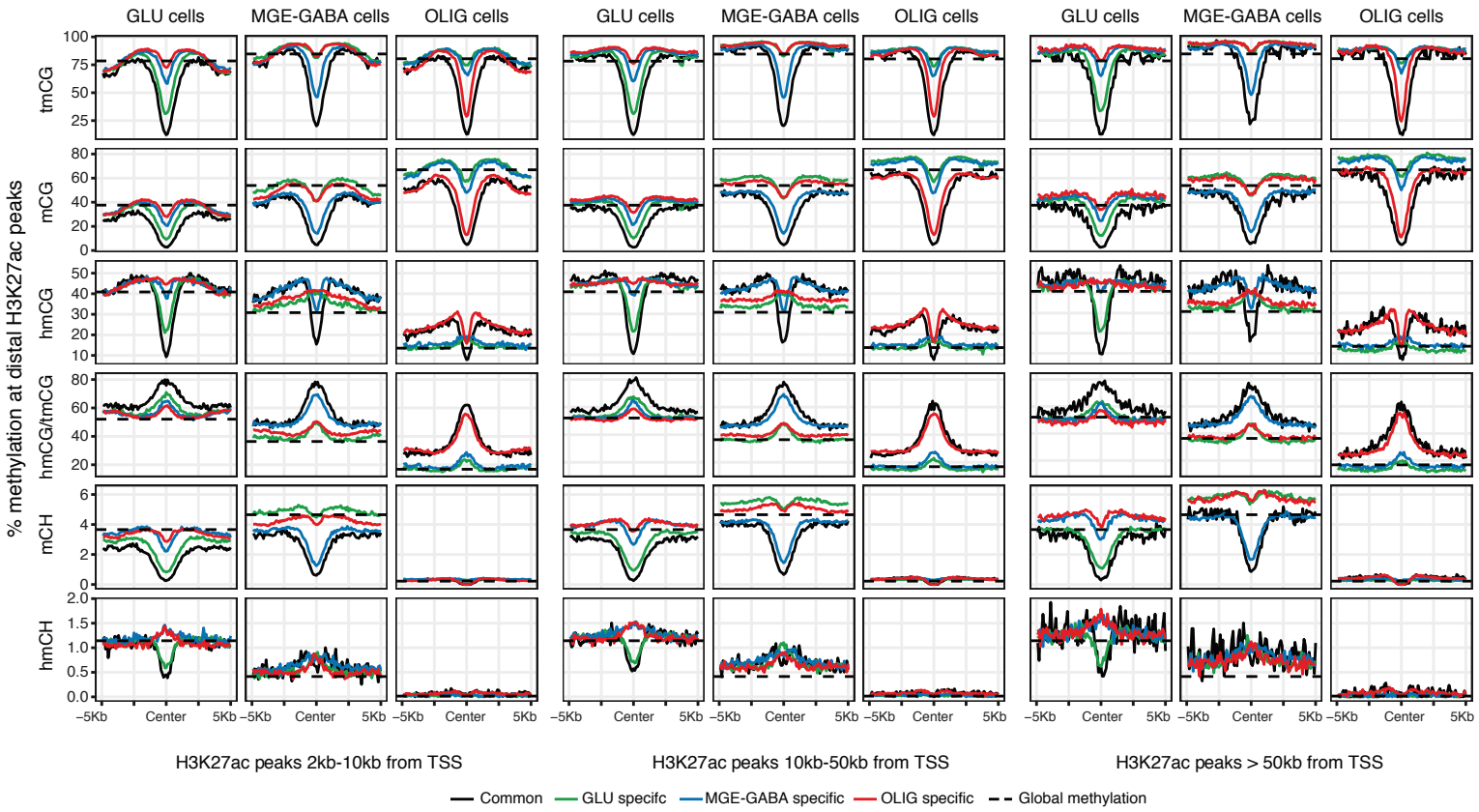

Fig. S5. DNA methylation profiles at cell type–specific distal H3K27ac peaks situated at different distances from TSS. Mean profile of DNA methylation marks at common and cell type-specific distal H3K27ac peaks situated at 2kb-10kb (left panel), 10kb-50kb (central panel) and >50kb (right panel) distance from TSS. Genome-wide methylation levels (as in Fig.1B) are shown in horizontal dashed lines. TSS, transcription start site.

Figure S6

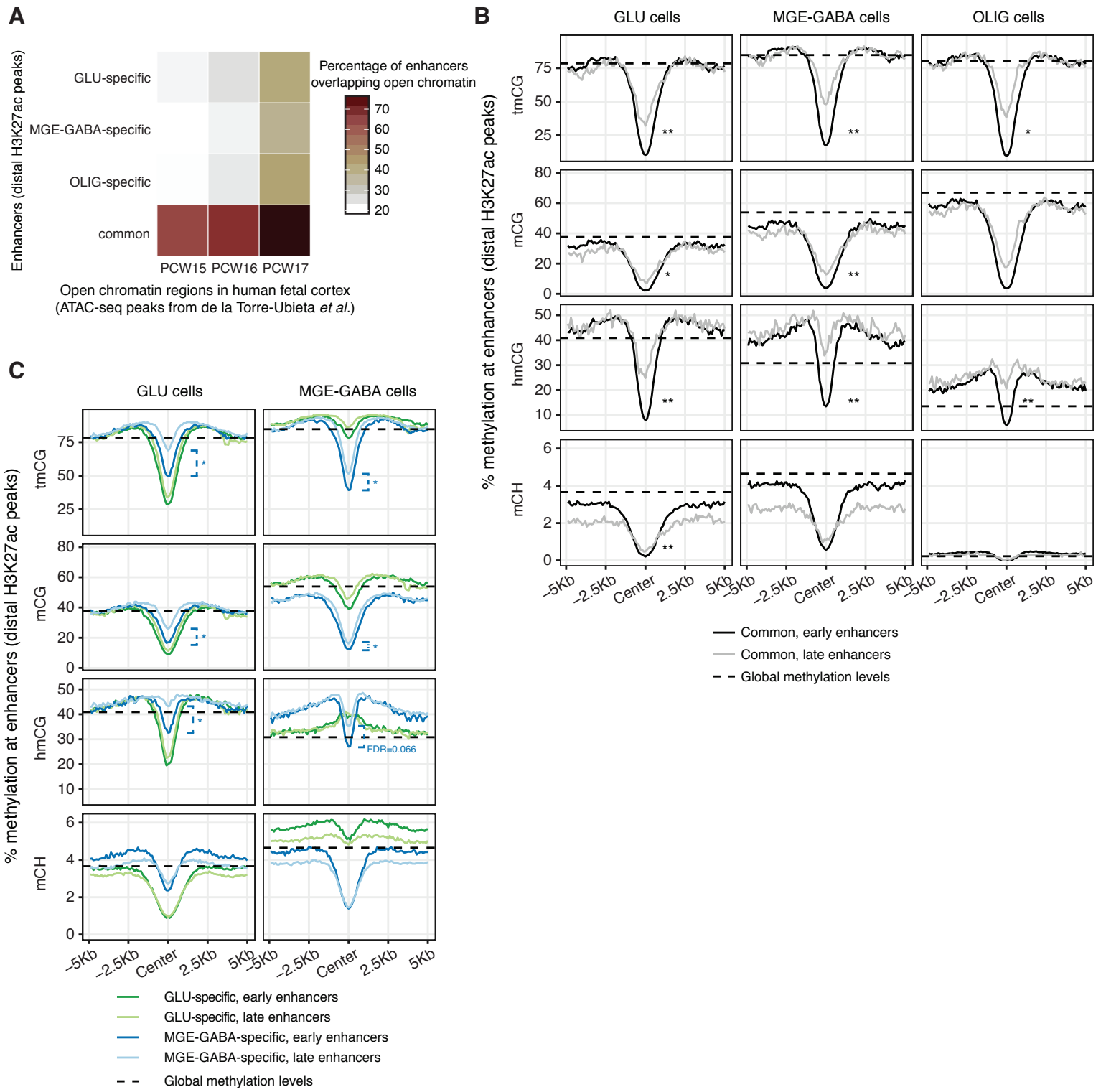
A
Enhancers (distal H3K27ac peaks)
GLU-specific
MGE-GABA-specific
OLIG-specific
common
PCW15 PCW16 PCW17
Percentage of enhancers overlapping open chromatin
70
60
50
40
30
20
Open chromatin regions in human fetal cortex
(ATAC-seq peaks from de la Torre-Ubieta et al.)
B
GLU cells
MGE-GABA cells
OLIG cells
% methylation at enhancers (distal H3K27ac peaks)
tmCG
mCG
hmCG
mCH
–5Kb
–2.5Kb
Center
2.5Kb
5Kb
Common, early enhancers
Common, late enhancers
Global methylation levels
C
GLU cells
MGE-GABA cells
% methylation at enhancers (distal H3K27ac peaks)
tmCG
mCG
hmCG
mCH
FDR=0.066
–5Kb
–2.5Kb
Center
2.5Kb
5Kb
GLU-specific, early enhancers
GLU-specific, late enhancers
MGE-GABA-specific, early enhancers
MGE-GABA-specific, late enhancers
Global methylation levels

Fig. S6. Comparison of adult enhancers (H3K27ac peaks) with regions of open chromatin (ATAC-seq) in the developing human fetal brain. Distal H3K27ac peaks in Glu and MGE-GABA neurons obtained in this study were overlapped with regions marked as open chromatin in the human fetal brain (post-conception weeks 15-17) based on ATAC-seq analysis from *(de la Torre-Ubieta et al. 2018*). **(A)** Overlap between distal H3K27ac peaks and open chromatin regions in the human fetal cortex. **(B, C)** Mean profile of DNA methylation marks at common (B) and cell type-specific (C) distal H3K27ac peaks. Genome-wide methylation levels (as in Fig.1B) are shown in horizontal dashed lines. PCW, post-conception week. Differences of methylation levels in the center of the regions were tested between early and late enhancers in each panel, and significances are indicated with asterisks (t-test, * FDR < 0.05; ** FDR < 0.01). There were no significant differences between early and late Glu-specific enhancers.

Figure S7

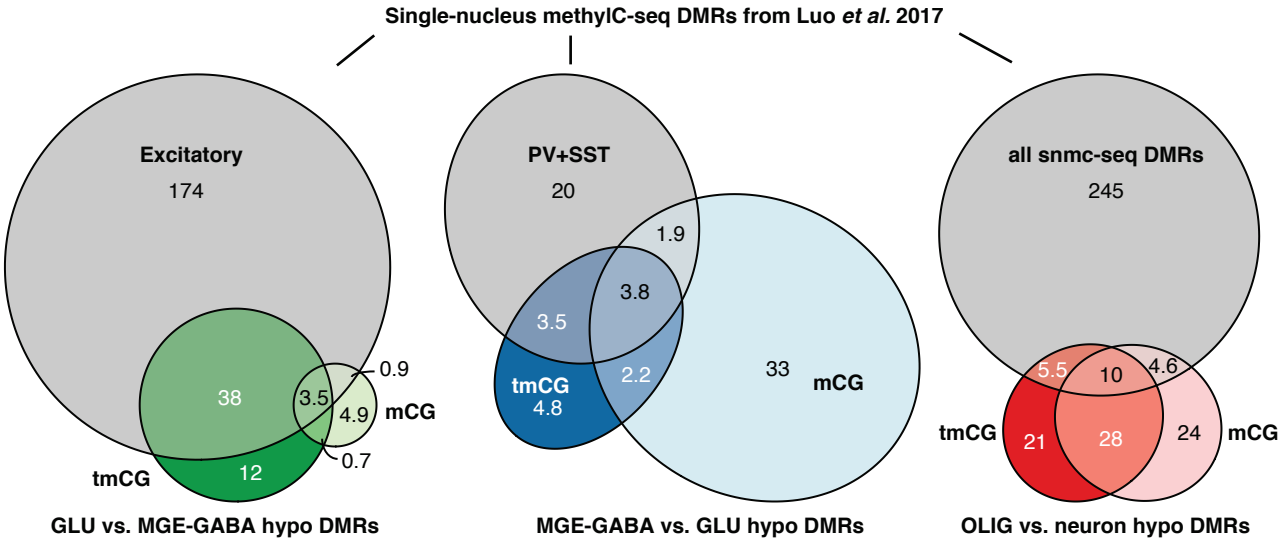

Fig. S7. Comparison of cell type–specific DMRs with DMRs from single-nucleus methylC-seq data. Venn diagrams show overlaps of cell-type specific DMRs identified in the current study (Fig. 4A; colored ovals) with DMRs identified in the published single-nucleus methylC-seq data *(Luo et al. 2017)* (grey ovals). Numbers of DMRs are shown in thousands.