

Supplementary Notes

Evolutionary conservation of DNA methylation signatures in the human brain

We found that two chimpanzee cortical samples had distinctive epigenetic signatures compared with human-derived samples (Supplementary Fig. 2A). To examine whether this was attributable to sequence difference in the probe sequence between human and chimpanzee, we compared all probe sequences (about 50 mer in length per probe) between the two species. Of the 1505 probes, sequences of 867 probes showed complete sequence conservation between the two species. From these, we next removed genes on the X chromosomes to exclude the effect of X inactivation, due to the inclusion of one female chimpanzee sample. Clustering analysis using the resultant 833 probes revealed that chimpanzee cortex samples were closely clustered together with human bulk cortex samples (Fig. 3C). Therefore, the epigenetic signature in the prefrontal cortex appeared to be highly conserved between two species, despite striking differences in key environmental factors, such as diet, in addition to the genetic differences. We found that 39 CpG sites showed a large difference in DNA methylation levels (average beta value difference >0.20) between human and chimpanzee cortex samples (see Supplementary Fig. 2B and Supplementary Table 3). They included transcription factors like HOX genes (HOXA9, HOXB13, and HOXB2) and epigenetic regulators such as MBD2 and

HDAC5, suggesting that such genes might play a role in generating differences in transcriptome. To assess the effect of local sequence context around the measured CpG site, we also compared sequence identity outside the probe region (upstream and downstream 500 bp from the measured CpG site) between the two species. We did not observe a significant difference in the sequence identity between the 39 CpG sites showing large methylation differences between two species ($98.77\% \pm 0.62\%$) and the remaining 794 probes ($98.91\% \pm 0.61\%$, $P = 0.082$). In addition, sequence identity around the probe region and methylation differences were not correlated ($N = 833$, $R = -0.024$, $P = 0.492$). Therefore, we concluded that the methylation differences in the 39 CpG sites were not simply caused by the effect of the local sequence context of the probe region; rather, they were associated with species-specific epigenetic changes.

Affymetrix promoter tiling array data analysis and extraction of MRs

We obtained about 2000 positive regions per sample in the initial MAT analysis (Supplementary Fig. 4AB). To consider the effect of the use of the specific restriction enzymes and number of CpGs included in the resultant fragment on data analysis, we introduced the following filtering procedure (Supplementary Fig. 4C). We first performed *in silico* restriction enzyme digestion analysis around every identified

positive region. We found that GC content (GC%) and the number of CpGs included in the digested fragment around the positive regions showed a bimodal distribution (Supplementary Fig. 4D). We considered that most of the regions of low GC% fragments should be artifacts because signals of these fragments could be affected by the nonspecific binding of MBD2-beads or weak affinity of MBD2-beads to nondensely methylated CpG-containing DNA fragments. Thus, we filtered out the region whose corresponding fragments had a low number of CpGs (<5) for further analysis in the current study. After filtering, the peaks with low GC% were successfully suppressed and the remaining regions contained DNA fragments with high GC% (≥ 60).

We identified about 1500 positive regions per sample. Consistent with the global hypomethylation in neuronal nuclei, we observed slightly fewer MRs in neuronal nuclei ($N = 1470$ on average) compared to nonneuronal nuclei ($N = 1747$) (Supplementary Fig. 4B).

Effect of methylation signal on gene expression in the brain

We previously performed gene expression analysis in the human prefrontal cortices using the Affymetrix HU133A array in 34 control subjects (Iwamoto et al. 2005).

Among the probe sets on the array, we found that 322 probe sets were associated with

the common MR near (within 10 kb upstream or downstream from the gene), or within the gene (intragenic). Qualitative examination of these MR-associated expression probes showed that expression of 109 probe sets (33.9%) was not detected in any of the 34 subjects (Supplementary Fig. 6). We found significant enrichment of unexpressed or absent-called probe sets in the MR-associated genes compared to all probe sets on the array (Fisher exact test, $P = 2.6 \times 10^{-5}$). This relationship was dependent on the location of the MR with respect to the gene and on the status of the GC content. Enrichment of the unexpressed probe sets was found in the probe sets whose MR was located in the upstream ($P = 1.9 \times 10^{-3}$). By contrast, no significant enrichment was found in the probe sets whose MR was located in the intergenic region (data not shown). We also found that upstream MRs that were included in the CpG island were strongly associated with silencing gene expression ($P = 1.9 \times 10^{-5}$).

Validation of Affymetrix tiling array results by Nimblegen array

For the purpose of validation, we performed DNA methylation analysis using a different tiling array platform, the Nimblegen promoter tiling array ($N = 1$ for NeuN+ and NeuN- samples), which covers about 800 bp upstream and 200 bp downstream from the first exons or CpG island. Of the 660 common MRs detected by the Affymetrix array, 259

MRs are covered by the Nimblegen promoter tiling array (with ≥ 4 probes/MR). These MRs were used for cross-platform comparison. Of these, 235 (90.7%) MRs were detected to be methylated in both NeuN⁺ and NeuN⁻ samples in the Nimblegen array (Supplementary Fig. 5), a very high rate of validation for two distinct microarray platforms. In addition, among the remaining 24 inconsistent regions, 14 MRs were detected to be methylated in either NeuN⁺ or NeuN⁻ samples in the Nimblegen array. Despite the fundamental differences between the two arrays, such as probes, labeling method, and data-processing procedures, this high reproducibility suggests that these factors have a minimal effect on the current results.

Ontology analysis of the genes associated with MRs

Among the genes associated with common MRs, we found that three gene families (29 *PCDH*, 5 *RBMY*, and 10 *OR*) were included. Because gene families showed a significant influence on the ontology analysis, we chose one gene for each of three gene families as representatives, to scrutinize other features of methylated genes. Through the PANTHER analysis, we found that a large number of the genes have functions in protein metabolism/modification, signal transduction, or cell structure. Signal transduction-related genes included subcategories such as calcium-mediated signaling,

cell adhesion-mediated signaling, voltage-gated potassium channel, cation transport, and stress response. In relation to this, we found that several potassium channel genes (*KCNMB4*, *KCNH7*, *KCNJ12*, and *KCNJ5*) were methylated in both nuclei. The genes included in the other large category of protein metabolism/modification consisted of protein glycosylation, carbohydrate metabolism, and proteolysis. Among these genes, we found that glycosylation-related categories frequently appeared, including *A4GALT*, *B3GALT6*, *FUT6*, *FUT3*, *TMTC1*, *B3GNTL1*, and *B4GALNT4*.

We found that several categories are specifically overrepresented in genes associated with neuronal or nonneuronal MRs (Fig. 5B). Several gene families were specifically enriched in neuronal nuclei compared to nonneuronal nuclei. They included 7 *KIR*, 5 *DAZ*, and 10 *OR* genes. The *OR* genes were identified as neuronal nuclei-specific methylation, and they appear to be due to the consistent methylation of different regions of the promoter of olfactory receptors in neuronal nuclei compared to nonneuronal nuclei.

Ontology analysis in nonneuronal nuclei revealed clear overrepresentation of several functional categories (Fig. 5B). Most of the overrepresented categories were related to signal transduction and transcription. Signal transduction-related genes were mainly found in the intracellular signaling cascade or the receptor-mediated signaling pathway

and included protein kinases, phosphorylases, and G-protein modulators. Notably, transcription factors (TFs) mainly consisted of zinc finger proteins. These results possibly reflect the presence of wide-ranging differences in the transcriptome and cellular functions between neurons and nonneurons. Other prominent features included overrepresentation of cell structure and motility-related categories such as cytoskeletal protein, extracellular matrix, and intermediate filament, possibly reflecting the morphologic divergence of neuronal and nonneuronal cells. Cell cycle and cell adhesion-related genes are also extensively methylated in nonneuronal cells.

TF-binding sites in the MR

In the common MRs, we found that a number of TF-binding sites were overrepresented (Supplementary Fig. 11). Based on the similarity of binding sites, some of them can be grouped into the larger matrix family named CREB, EBOX, or HIFF family. The CREB family includes TFs such as ATF, XBP1, and CREB. They are involved in a wide range of biologic functions such as signal transduction, unfolded protein stress, cell cycle, and nervous system development. The EBOX family includes ubiquitously expressed TFs. Members of the EBOX family have functions in signal transduction, cell cycle, and apoptosis. The HIFF family includes TFs such as hypoxia inducible factors (HIF1A and

HIF3A), aryl hydrocarbon receptor nuclear translocators (ARNT, ARNT2, ARNTL, and ARNTL2), CLOCK, and neuronal PAS domain proteins (NPAS1 and NPAS2). In addition to the role of response to hypoxia and oxidative stress by HIFs, members of this family have roles in several important brain functions such as circadian rhythm and neuronal cell differentiation and development. The others are involved in non-brain-specific functions such as immune response and glycolysis. It should be noted that our results of TF-binding site analysis would be affected by the use of restriction enzymes to some extent. The TF-binding sites of the AT-rich motifs, which are recognized by the enzymes used, are significantly underrepresented in each analysis (data not shown).

Reference

Iwamoto, K., M. Bundo, and T. Kato. 2005. Altered expression of mitochondria-related genes in postmortem brains of patients with bipolar disorder or schizophrenia, as revealed by large-scale DNA microarray analysis. *Hum Mol Genet* **14**: 241-253.