**Supplementary Information**

**Single value decomposition (SVD) analysis**

The batch effect of plate, well, sentrix id and sentrix position, and potential confounding effects of age and gender on the methylation data were checked by single value decomposition (SVD) analysis(Teschendorff et al., 2013). First, SVD deconvolution was performed on beta matrix, the components explained most variance in original data set would be extracted. Then the correlation significance between the components and each phenotype was checked. Figure S1 displayed the results of SVD analysis. Group and sentrix factors showed significant contribution to the variation of the methylation data. Those technical factors that showed significant contribution to the variation of the methylation data were corrected using the Combat function in the sva package(Leek, Johnson, Parker, Jaffe, & Storey, 2012).

**MRI Acquisition and Image Processing**

MRI examinations were performed on 3T SIEMENS TrioTim scanner equipped with a 32-chanel head coil located at the West China Hospital, Sichuan University. High resolution T1 weighted images and diffusion-weighted images were acquired.

PANDA pipeline tool (http://www.nitrc. org/projects/panda) and FMRIB Software Library (FSL, <http://www.fmrib.ox.ac.uk/fsl)> were used to preprocess the imaging data. Routine DTI preprocessing was performed, including head motion and eddy current correction, brain extraction, and tensor model fitting. Then, whole-brain deterministic fiber tractography was performed to generate three-dimensional streamlines characterizing neural fiber tract connectivity. Following tractography, we performed waypoint region of interest-based tract segmentation, and probability map-based fiber refinement using the 20-tract Johns Hopkins University white matter template (http://neuro.debian.net/pkgs/fsl-jhu-dti-whitematter-atlas.html). Tracts included the left and right anterior thalamic radiation, corticospinal tract, cingulum-cingulate, cingulum-hippocampus pathway, inferior fronto-occipital fasciculus, inferior longitudinal fasciculus, superior longitudinal fasciculus, uncinated fasciculus and arcuate fasciculus, and the forceps major of the splenium and the forceps minor of the genu of the corpus callosum. After tract identification, the diffusion measurement along the tract core, defined as the tract profile, was extracted from each fiber tract. Tracts were smoothed using a 10-point moving average filter to reduce local variation caused by imaging noise. One-way ANOVA was used to compare average fractional anisotropy values among the three subgroups followed by pairwise post hoc tests in tracts with significant group differences. Age and gender were included as covariates.

Cortical modeling and volumetric segmentation of structural MRI data were performed using the FreeSurfer package (version 5.3.0, http://surfer.nmr.mgh.harvard.edu/). Briefly, the procedure included automated registration to the Talairach space, normalization of intensity, segmentation of the subcortical white matter and gray matter volumetric structures, tessellation of the gray matter and white matter boundaries, and automated topology correction and surface deformation following intensity gradients to optimally place the gray/white and gray/CSF borders defined at the location with the greatest shift in signal intensity. Then, surface inflation and registration to a spherical atlas were performed. Cortical thickness was estimated as the average shortest distance from the gray and white matter boundary to the gray matter and cerebrospinal fluid boundary.The results for each subject were visually inspected to ensure accuracy of registration, skull stripping, segmentation, and cortical surface reconstruction. Maps of cortical thickness were smoothed using a circularly symmetric Gaussian kernel across the surface with a full width at half maximum (FWHM) of 15 mm. Comparisons of among subgroups were assessed using a vertex-by-vertex general linear model. The model included cortical thickness as a dependent factor and diagnosis as an independent factor, and also included age and gender as nuisance variables. The results were explored at threshold of uncorrected vertex-level p<0.05.

Besides, subcortical volumes were extracted for each participant. One-way ANOVA was used to compare subcortical volumes among the three subgroups followed by pairwise post hoc tests in tracts with significant group differences. Age and gender were included as covariates.

**Neuroanatomical features of subtypes**

There is no significant group difference in fractional anisotropy across the examined fiber tracts or in subcortical volumes. Patients in FES-P2 had relatively more prevalent cortical thickness increases, predominantly localized in cingulate, parietal and temporal cortical regions (Figure S2).

**DMPs detection accounting for age and gender effect**

DMPs were also detected accounting for the age and gender effect. The statistical significance of methylation value difference between groups were calculated via a linear regression model, with age and gender added as covariates. R package limma was used for the calculation. P value was adjusted with Benjamini-Hochberg correction. DMPs that met all of the following criteria were kept and the rest were filtered out: 1) average expression value above 0.2; 2) adjusted p value below 0.01; 3) no SNP on the probe site; and 4) a minimum difference of 0.05 in beta value between groups.

Generally, DMPs detection analysis revealed that P2 was more divergent from healthy controls than P1 in terms of blood methylation levels in both FES and LTS cohorts. There were 17,241 DMPs between FES-P2 and controls and only 53 DMPs between FES-P1 and controls, with the between group differences of beta values above 0.05. Over-representation analysis of genes associated with FES-P2 DMPs via GO BP indicated an enrichment of genes involved in neutrophil activity (Figure S3A). There were 59,383 DMPs were found between LTS-P2 and controls and only 2 DMPs between LTS-P1 and controls, with the between group differences of beta values above 0.05. Over-representation analysis of genes associated with LTS-P2 DMPs via GO BP indicated an enrichment in sites related to biological process involved in neuron development (Figure S3B).

DMPs were also examined with between group difference of beta value above 0.1. There were1264 DMPs between FES-P2 and controls and only 2 DMPs between FES-P1 and controls. Examination of those DMPs with greater group differences in FES-P2 via GO BP indicated enrichment of genes involved in neutrophil activity (Figure S3C). There were 15380 DMPs between LTS-P2 and controls and zero DMPs between FES-P1 and controls. Examination of those DMPs with greater group differences in LTS-P2 via GO BP also indicated enrichment of genes involved in neutrophil activity (Figure S3D).

Shared DMPs were examined on methylation sites that met the following criteria in both cohorts: 1) average expression value above 0.2; 2) adjusted p value below 0.01; and 3) a minimum difference of 0.05 in beta value between groups. There were 15,968 shared DMPs in FES-P2 and LTS-P2 ((Figure S4A). Enrichment analysis of genes associated with shared DMPs via GO BP indicated an enrichment related to biological process involved in immune cell activity, especially neutrophil activity (Figure S4B).

Difference in methylation profile between P2 subtype in FES and LTS were also explored. There were 51380 DMPs between FES-P2 and FES-P2 patients with the between group differences of beta values above 0.05. Over-representation analysis of genes associated with DMPs via GO BP indicated an enrichment of genes top-ranking involved in neuron development and axonogenesis (Figure S4C).

**Cell-type decomposition analysis accounting for age and gender effect.**

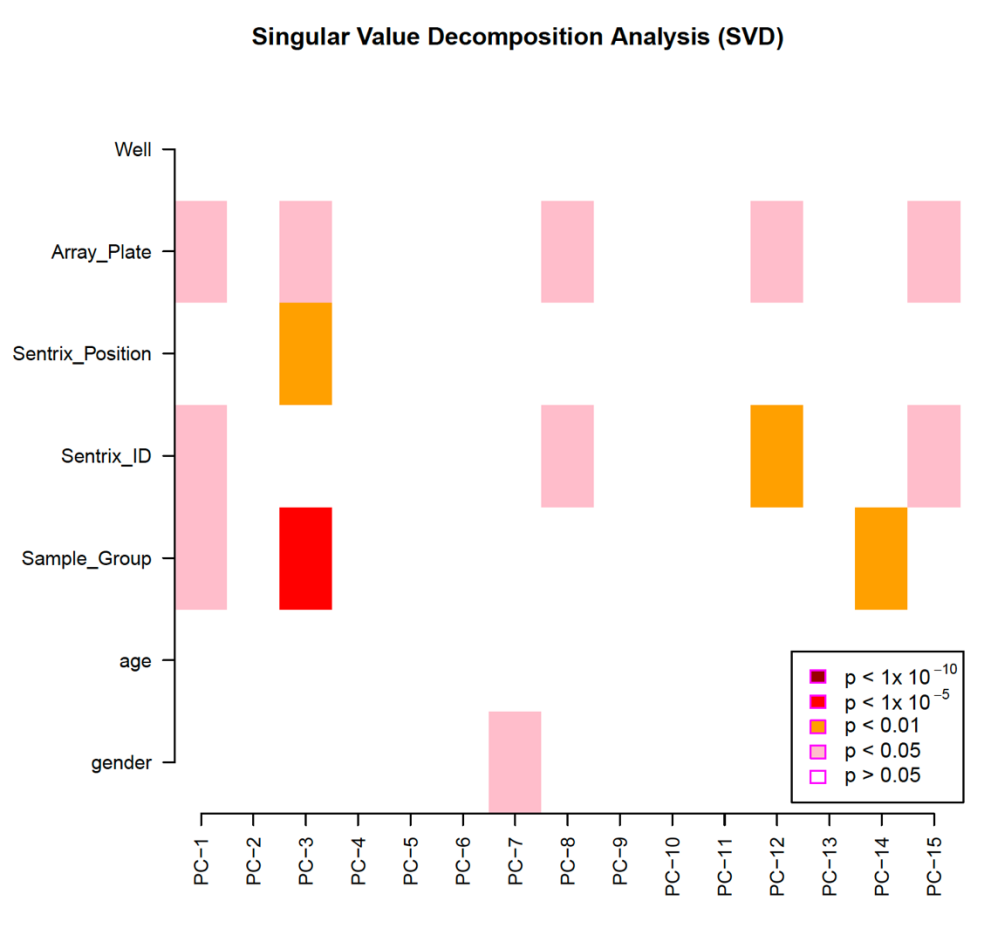
A reference-based method was used via GLINT(Rahmani et al., 2017) on the DNA methylation profile to estimate blood cell-type proportion in each sample. The statistical significance of cell-type proportion difference between groups were calculated via a linear regression model, with age and gender added as covariates. R package limma was used for the calculation. Cell-type decomposition analysis identified a significantly higher proportion of neutrophils and significantly lower proportion of lymphocytes, including CD4+ T-cells, CD8+ T-cells and natural killer (NK) cells in both P2 patients in LTS and FES cohort, compared to their matched healthy controls and P1 subtype patients (Figure S5).

**Reference**

Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E., & Storey, J. D. (2012). The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics (Oxford, England), 28*(6), 882-883. doi:10.1093/bioinformatics/bts034

Rahmani, E., Yedidim, R., Shenhav, L., Schweiger, R., Weissbrod, O., Zaitlen, N., & Halperin, E. (2017). GLINT: a user-friendly toolset for the analysis of high-throughput DNA-methylation array data. *Bioinformatics (Oxford, England), 33*(12), 1870-1872. doi:10.1093/bioinformatics/btx059

Teschendorff, A. E., Marabita, F., Lechner, M., Bartlett, T., Tegner, J., Gomez-Cabrero, D., & Beck, S. (2013). A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics (Oxford, England), 29*(2), 189-196. doi:10.1093/bioinformatics/bts680

****

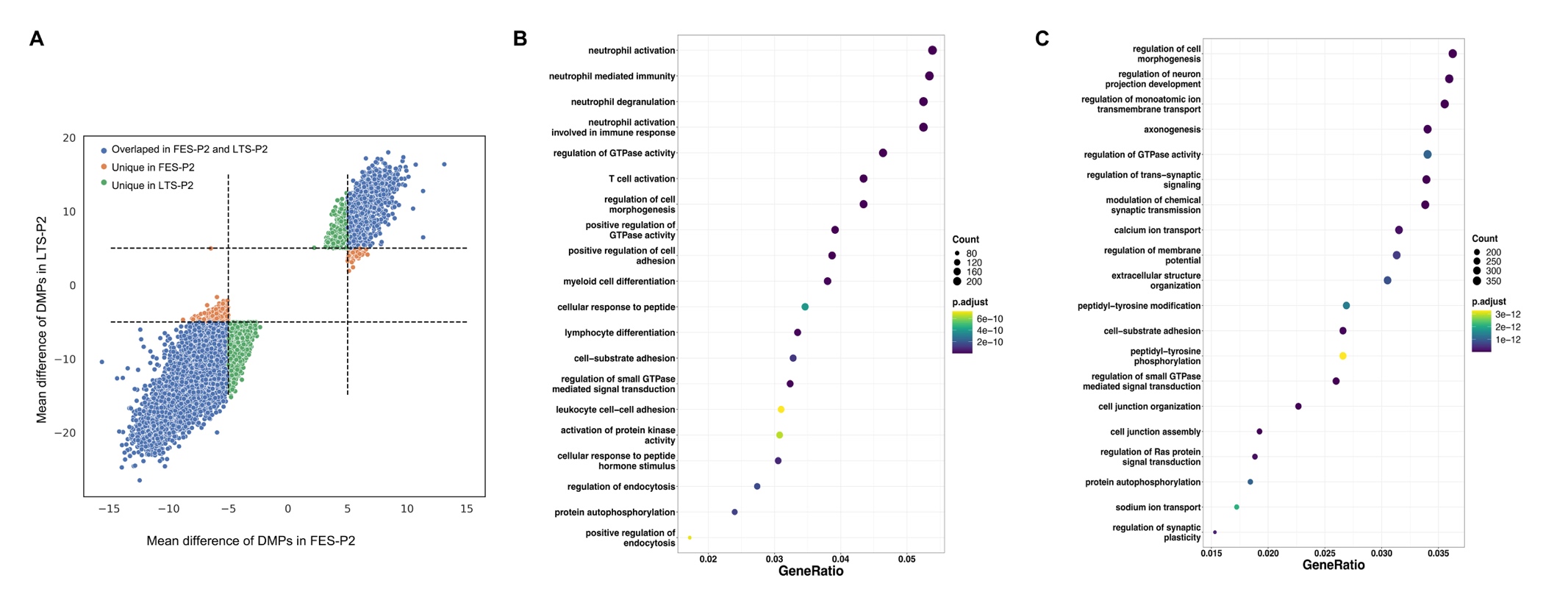
**Figure S1.** Single value decomposition (SVD) analysis on methylation data. The horizontal line shows the major component explained the variances in the data. Color in the boxes indicates different levels of significance. The darker the color is, the more significant the component are correlated with the phenotype.



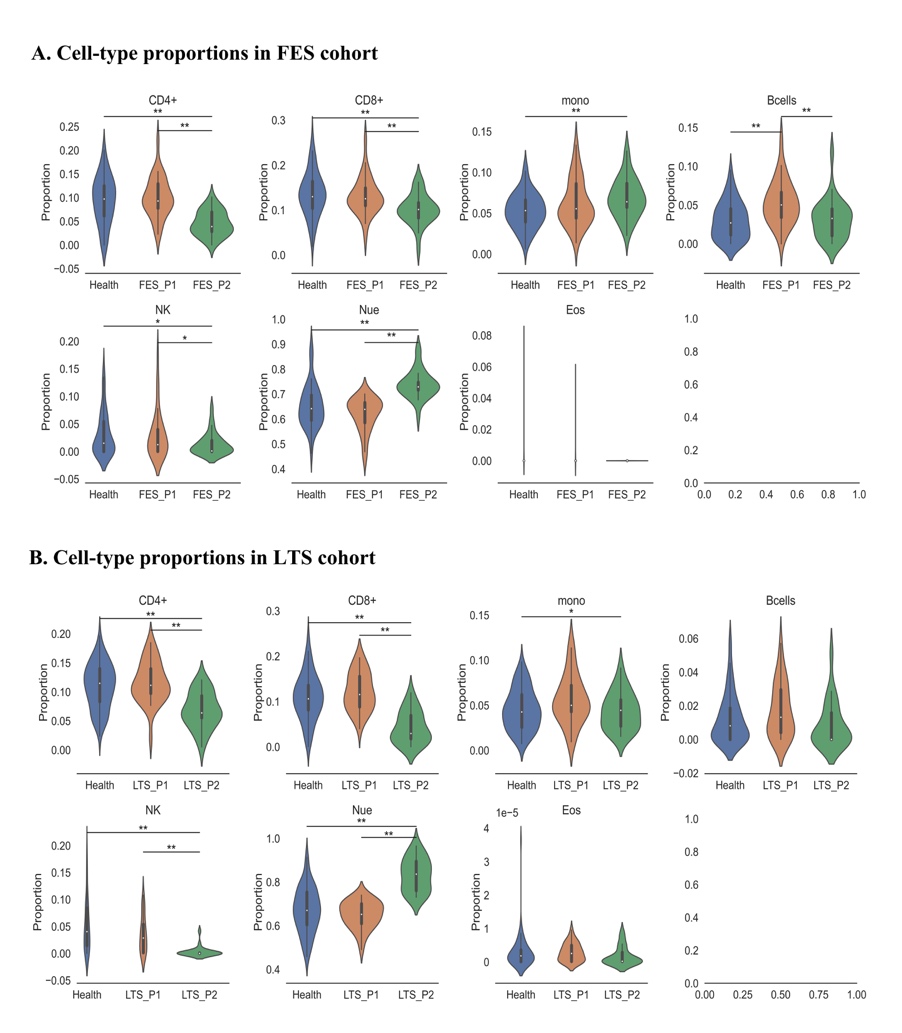
**Figure S2**. Group difference in cortical thickness among patient subgroups and healthy controls. The results were explored at threshold of uncorrected vertex-level p<0.05. Blue indicate negative statistics (thinner cortex), and red indicate positive statistics (thicker cortex).



**Figure S3**. Over-representation analysis of DMPs accounting for the age and gender effect. (A) Over-representation analysis of DMPs between FES-P2 and controls (with beta difference>0.05) associated genes in terms of GO biological process. (B) Over-representation analysis of DMPs between LTS-P2 and controls (with beta difference>0.05) associated genes in terms of GO biological process. (C) Over-representation analysis of DMPs between FES-P2 and controls (with beta difference>0.1) associated genes in terms of GO biological process. (D) Over-representation analysis of DMPs between LTS-P2 and controls (with beta difference>0.1) associated genes in terms of GO biological process. Only top 20 over-represented terms were displayed. NMF:non-negative matrix factorization; FES: patients with first-episode schizophrenia; DMPs: differentially methylated probes.



**Figure S**4. (A) Shared DMPs of subtype P2 in FES and LTS cohorts accounting for age and gender effect. (B) Enrichment analysis of genes contained shared DMPs between P2 subtype in FES and LTS via Gene Ontology Biological Process. (C) Enrichment analysis of genes contained DMPs discovered between P2 subtype in FES and LTS via Gene Ontology Biological Process. Only top 20 over-represented terms were displayed. FES: patients with first-episode schizophrenia; LTS: patients with long-term ill schizophrenia; DMPs: differentially methylated probes.

****

**Figure S5.** Cell-type decomposition analysis accounting for the age and gender effect. (A) Comparisons of cell-type proportions derived from DNA methylation data in FES cohort. (B) Comparisons of cell-type proportions derived from DNA methylation data in LTS cohort. Pairwise comparison between control and the two patient subtypes were applied. \*p < 0.05, \*\*p < 0.01. FES: patients with first-episode schizophrenia; LTS: patients with long-term ill schizophrenia.

**Table S1**. Results of meta-analysis of differences in cell-type proportion estimates between P2 subtype and controls across two cohorts.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Fixed effects model** | | | **Random effects model** | | | **Heterogeneity p-Value** |
| Mean difference | SE | p-Value | Mean difference | SE | p-Value |
| **CD4+ T-cells** | -0.060 | 0.008 | <0.001 | -0.060 | 0.008 | <0.001 | 0.378 |
| **CD8+ T-cells** | -0.054 | 0.010 | 0.000 | -0.056 | 0.020 | 0.006 | 0.038 |
| **Monocytes** | 0.004 | 0.005 | 0.408 | 0.002 | 0.016 | 0.903 | 0.001 |
| **B cells** | -0.007 | 0.004 | 0.091 | -0.007 | 0.004 | 0.091 | 0.734 |
| **NK cells** | -0.036 | 0.008 | 0.000 | -0.039 | 0.015 | 0.010 | 0.074 |
| **Eosinophils** | <0.001 | <0.001 | 0.077 | <0.001 | <0.001 | 0.077 | 0.688 |
| **Neutrophils** | 0.146 | 0.018 | <0.001 | 0.160 | 0.060 | 0.008 | 0.001 |