

5 **Participants and response-to-medication assessment**

The study group comprises patients diagnosed with schizophrenia (SCZ) according to the Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition (DSM-IV) and attending the regular outpatient services at the Department of Psychiatry, Jawaharlal Institute of Postgraduate Medical Education and Research (JIPMER) in Puducherry, India. The patients were all Tamil speaking and recruited between 10 August 2014 and August 2016. The inclusion criteria were: diagnosed schizophrenia, age above 18, a total score at the Positive and Negative Syndrome Scale (PANSS) of at least 30 and newly prescribed risperidone. Patients who were prescribed antipsychotics other than risperidone, substance abusing patients and pregnant or lactating women were excluded from the study. Treatment compliance was confirmed by the patients' 15 family members or caregivers and patients who did not take risperidone for five consecutive days during the study period were excluded from the study. Duration of psychosis varied between one month and 12 years (median = 21 months, IQR (interquartile range) [6 months, 48 months], mean = 32 months). Whereas some participants had already been treated for SCZ at some point during their life, none of them received any antipsychotic medication at least four weeks before the baseline visit. At the baseline, the participants started 20 treatment with 4-8 mg of risperidone per day. All except one were comedicated with at least one additional medicament to alleviate side effects: clonazepam for sleep (0.5-1 mg/day as required before going to bed) and trihexyphenidyl for extrapyramidal symptoms (2-4 mg/day).

Treatment response was assessed using the total PANSS score. Briefly, symptom severity at the baseline and follow-up visit four weeks later was calculated as a sum of scores on a positive, negative and a general psychopathology scale. Patients with at least 20% reduction of PANSS score after four weeks of treatment 25 were categorized as good responders, while the rest were considered as bad responders [1].

Patients' characteristics summary, including the differences between the good and bad responders can be found in Table 1 in the article. The cohort comprises 12 Tamil women and 16 Tamil men, aged 18-60 (mean = 32.9, median = 32.5). The majority (N = 20, 71.5%) were non-smokers and never consumed alcohol (N = 23, 88%). We found no significant difference in age, weight, alcohol and tobacco use or prescribed 30 risperidone dose between the good and bad responders. Informed consent was obtained from each study participant and the patient's legally acceptable representative (LAR) or a family member. The Institutional Ethics Committee approved the study protocol before the commencement of the study (IEC Project No. JIP/IEC/4/2013/189).

35 **Sampling and generation of raw methylome data**

The blood of 28 recruited patients was collected at the baseline (V = 7 ml) and follow-up visit (V = 5ml) between 9-11 AM, resulting in a total of 56 samples. Blood cells were separated by centrifugation immediately after sample collection and were stored at -80°C until the DNA extraction. DNA was extracted using QIAamp Blood Midi Kit (Qiagen, Germany) according to the manufacturer's protocol. All DNA 40 samples were quantified in fluorescence (Quant-IT kits, ThermoFischer Scientific), in duplicate. DNA quality and integrity were assessed with TapeStation 4200 (Agilent technologies Inc., Santa Clara, CA, U.S.A.) and by PCR amplification for 10 % of the extracted samples and all had a DNA integrity number (DIN) higher than 7. Bisulfite conversion of 1µg of genomic DNA was done by using EpiTect Fast 96 Bisulfite Kit (Qiagen, Germany) and genome-wide methylation was screened using Infinium Methylation

45 EPIC BeadChip [2] on the CNRGH automated platform, following the manufacturer's protocol (Illumina Inc., USA).

Quality control of methylation data

50 Quality control, preprocessing, statistical analysis and data visualization were all done in R statistical framework [3], version 4.1.3).

Raw intensity data was imported into R using the *minfi* v. 1.38.0 package [4,5]. The first step was the exclusion of low quality probes, as they can negatively influence the normalization procedure [6], and of low-quality samples, which could affect statistical inference [7]. Briefly, we first identified low-intensity probes by calculating detection p-values (threshold = 0.05) based on the non-specific fluorescence and performed the control metrics check using the *ewastools* v. 1.7 package [7,8]. In addition, we ran quality checks with *ENmix* v. 1.28.8 [9] and *minfi* [4,5] packages. We then checked for mismatches between the predicted and actual sex and between the predicted genotypes for paired samples [7]. In addition, we checked if any of the samples fulfilled at least two of the following criteria: median of unmethylated (U) and methylated (M) log raw intensities < 10 (as defined by the *minfi* "plotQC" function), failed *ewastools* control metrics [7] or *ENmix* v. 1.28.8 [9] quality check (samples below 3x SD of bisulfite control intensities or samples with >5% low quality probes, based on the detection p-values and number of beads). All samples passed our quality control criteria. Regarding the probes, we removed previously identified 44,570 cross-reactive probes (binding to multiple sites in the genome [10,11]) and low-quality probes fulfilling any of the following criteria: 25% detection p-values above 0.05, the intensity above 3x IQR of average U+M intensities, probes with high (>3x IQR) SD over the beads or the probes with the mean intensities calculated from less than median 5 beads per sample [6]. Sex chromosome probes and probes with SNPs at the target site were not excluded at this step. The quality control reduced the number of probes from 865,859 to 821,287.

70 Within-array normalization

To reduce the amount of technical variation within a single array, we performed a three-step (background, dye-bias and design-bias correction) normalization of the data. Specifically, we used the *noob* method (*minfi* v. 1.38.0), based on the out-of-band intensities of all probes in the array [12], for the background correction, followed by dye-bias correction using RELIC (REgression on Logarithm of Internal Control probes), which adjusts the green intensities based on their predicted values from linear regression with the corresponding red intensities as predictors [13]. Finally, we performed design-bias correction, accounting for the differences in the dynamic range of type I and type II probes, using the BMIQ (Beta Mixture Quantile dilation) method [14] implemented in the *wateRmelon* v. 1.36.0 package [15].

80 Estimation of blood cell type composition from the methylation data

Blood is a mixture of cell types with specific DNAm profiles. Proportions of these cells vary with age and physiological conditions [16]. Thus, leukocyte composition is an important source of DNAm variability, and failing to account for it can result in spurious conclusions due to confounding with other factors (for review see [17]). To estimate the relative abundances of six leukocyte types (neutrophils/granulocytes, CD4+ and CD8+ T-cells, B-cells, monocytes and NK cells), we used the method implemented in the R package *flow.sorted.blood.EPIC* v. 1.10.1 on the background-corrected data as advised by the authors [18]. As the blood cell types *per se* were not of primary interest in this study and six additional variables could negatively affect inference, we reduced the dimensionality of the leukocyte composition data by performing a principal component analysis (PCA), using the *vegan* v. 2.5-7 package [19] on the six cell types and extracting the axes that accounted for the majority of the observed variation. As the first principal component (PC1)

accounted for 84.6% of variation, we used it instead of composition data to present the cell type heterogeneity in our analyses. This axis strongly positively correlated with the proportion of neutrophils and negatively with other leukocyte types (Supplementary Figure 1).

95 ***Estimation of unwanted variation (batch effects)***

Although the methods, such as the empirical Bayes method developed by Johnson et al. [20] can remove unwanted variation directly from the methylation data, we decided not to use this approach, as it may result in false positives, especially so for unbalanced designs and small sample sizes [21,22]. We thus opted for the inclusion of batch effects as covariates in the statistical analysis. We first created a “control matrix”, as described in Fortin et al. [23], summarizing control and out-of-band probes, whose values should be independent of biological variation. Other factors describing technical variation included chip identity, position on the chip and sampling centre. To avoid overfitting due to a high number of independent variables in the models and to prevent false positives due to a tendency of multilevel factors to appear significant in the statistical analysis, we decided to reduce the dimensionality of the batch effect data by performing a factor analysis using the *FactoMineR* v. 2.4 package [24] with the control matrix, chip ID, position on the chip and sampling centre as input variables. We then selected the first two axes, describing 42% of the unwanted variation (Supplementary Figure 2), as covariates for the statistical analysis.

Exploratory multivariate analysis

In order to examine global blood methylation patterns, we used *FactoMineR* to perform a PCA on all 56 samples. Considering that using all of the CpGs would be computationally expensive and most of the probes were barely variable anyway, we first filtered out all the probes with SD of beta-values distribution < 0.1 , as methylation differences $\Delta\beta < 0.1$ are often due to technical rather than biological variation [6]. In addition, we removed the probes situated on the sex chromosomes as they would obscure autosomal DNAm variation. This filtering step reduced the size of the dataset to 39,239 autosomal probes. This initial PCA (Supplementary Figure 3) revealed two outlier individuals (P9 and P14) that were excluded from further analysis. Of note, these two individuals had nothing conspicuous in common (they differed by sex and response to treatment from one another, had distinct DNAm profiles according to the PCA and did not differ from other patients either by duration of psychosis or comedication). In both cases, the samples from the same individual grouped together, indicating that the lack of similarity to other samples was probably due to unknown sources of biological variation (e.g. genetics or physiological condition). All the reported results are based on the remaining 52 samples from 26 individuals (19 good and seven bad responders).

To assess the difference in magnitude of medication-induced methylome changes between the good and bad responders, we calculated pairwise Mahalanobis (multivariate) distances between the paired samples, and used these as a dependent variable in a linear regression model (Mahalanobis distances \sim response). In order to identify CpGs significantly correlated (absolute value of the correlation coefficient ≥ 0.2) with the main axes (PC1 and PC2) representing the methylome variation, and to assess the relationship between the risperidone treatment response and other factors (baseline and follow-up PANSS scores, visit, sex, smoking, alcohol use, blood cell type composition, chip number, position on the chip, patient identity, unwanted variation) and these axes, we calculated the corresponding correlations using *dimdesc* function from the *FactoMineR* package[24].

Gene annotation and pathway analyses

In order to gain insight into the function of genes associated with the CpGs significantly correlated with either positive or negative sides of the two main principal components (PCs), we conducted a gene-set

enrichment analysis for each of the four groups of the CpGs separately (i.e. PC1-, PC1+, PC2-, PC2+, see Figure 1D) using the *gometh* method [25] implemented in the *missMethyl* v. 1.26.1 package [26] and a CpG universe reduced to the variable 39,239 probes to avoid sample source bias [27]. The method accounts for the number of CpGs annotated to a single gene as well as for the CpGs annotated to multiple genes. We tested for the enrichment of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways and gene ontology (GO) categories. Both databases contain a large number of categories/pathways to be tested, strongly reducing the likelihood for any of the p-values to remain significant at any of the commonly used significance cutoffs after the multiple testing correction in the small datasets. However, gene-set enrichment analysis is a multi-step procedure and some authors argue that a p-value adjustment does not correctly account for the family-wise error rate and advise to use raw p-values and term ranking to select potentially interesting pathways and processes instead [28]. Therefore, we decided to keep the terms/pathways with raw $p \leq 0.01$ and containing at least three differentially methylated genes in case of GO terms, or 5% of differentially methylated genes for KEGG pathways. In addition, we included only the KEGG pathways with at least 10% genes covered by the CpG universe and pathway size between 15-500 genes to reduce the effect of large pathways on the analysis [29]. In order to simplify the interpretation of the GO enrichment analysis results, we used the *GOxploreR* v. 1.2.4 package to prioritize GO terms [30] and *GOsemSim* v. 2.18.1 package [31,32] to group them based on their semantic similarity (0.7 similarity threshold, Relevance method [33]). We visualized the results using *GOxploreR* v. 1.2.4 [30], *rrvgo* v. 1.4.4 [34,35] and *pheatmap* v. 1.0.12 [36] packages. In addition, we performed a gene-set enrichment analysis of the variable 39,239 autosomal CpGs with all probes on the chip that passed our quality control step ($N = 821,287$) as the universe in order to identify functional categories that were over-represented in the variable portion of the blood methylome of our cohort.

Association between individual CpG methylation levels and risperidone response

In order to explore the relationship between the treatment response and the methylation values of the CpGs annotated to the genes associated with the enriched functional terms, we ran generalized linear mixed models with beta family (*glmmTMB* v. 1.1.2.3 package [37]) using visit, response (defined either as binary variable or %PANSS improvement) and visit x response interaction as variables of interest. We ran two models for each CpG-response type combination, i.e. with and without covariates in order to account for other likely sources of methylome variation. The covariates included: leukocyte composition, sex, smoking and two variables describing the unwanted variation (s. the procedure above). Patient identity was added as a random variable (model formula: CpG beta value \sim visit * response + smoking + sex + blood cell type composition PC1 + unwanted variation PC1 + unwanted variation PC2 + (1|patient)). We removed the models with convergence, multicollinearity, outlier, uniformity and dispersion issues (*DHARMA* v.0.4.6, [38]). We compared these models to the corresponding null models with a likelihood ratio test (LRT) to obtain a model significance and subsequently adjusted the p-values for multiple testing using the Benjamini-Hochberg procedure. However, due to the exploratory nature of our analysis, we included all the models with raw p-values ≤ 0.05 in the results. Still, we clearly indicate which models remained significant at 0.05 level following the B-H correction. In addition, we wanted to determine if the CpGs associated with the same gene behave in a similar way. To do this, we calculated the proportion of differentially methylated vs. total number of probes associated with the genes of interest on the chip. We also checked for the experimental evidence of correlation between the blood and brain methylation values of these CpGs using IMAGE-CpG tool [39]. Finally, we used the top twenty CpGs associated with the main effect of response or the CpGs with the absolute value of coefficient estimates from *glmmTMB* ≥ 0.18 for the response x visit interaction to try to predict the response to treatment using the stability variable selection method (package *mboost* v. 2.9-5 [40,41]). The reason we limited the variable selection procedure based on the response-specific CpGs to twenty is computational feasibility.

Data and code availability

185 All data including R objects and figures that are not included as a supplementary material are available on Figshare. Raw methylation intensity data are deposited in the ArrayExpress repository under the accession number E-MTAB-11921. All scripts necessary to reproduce the presented analysis are available upon request from the authors.

190 References

1. Haas M, Unis AS, Armenteros J, Copenhaver MD, Quiroz JA, Kushner SF. A 6-week, randomized, double-blind, placebo-controlled study of the efficacy and safety of risperidone in adolescents with schizophrenia. *J. Child Adolesc. Psychopharmacol.* 19(6), 611–621 (2009).
2. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. *Epigenomics.* 8(3), 389–399 (2016).
3. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. R Foundation for Statistical Computing, Vienna, Austria. Available from: <http://www.R-project.org/>.
4. Aryee MJ, Jaffe AE, Corrada-Bravo H, *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics.* 30(10), 1363–1369 (2014).
5. Fortin J-P, Triche TJ, Hansen KD. Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. *Bioinformatics.* 33(4), 558–560 (2017).
6. Dedeurwaerder S, Defrance M, Bizet M, Calonne E, Bontempi G, Fuks F. A comprehensive overview of Infinium HumanMethylation450 data processing. *Brief. Bioinform.* 15(6), 929–941 (2014).
7. Heiss JA, Just AC. Identifying mislabeled and contaminated DNA methylation microarray data: an extended quality control toolset with examples from GEO. *Clin. Epigenetics.* 10(1), 73 (2018).
8. Murat K, Grüning B, Poterlowicz PW, Westgate G, Tobin DJ, Poterlowicz K. Ewastools: Infinium Human Methylation BeadChip pipeline for population epigenetics integrated into Galaxy. *GigaScience* [Internet]. 9(5) (2020). Available from: <https://academic.oup.com/gigascience/article/9/5/giaa049/5836679>.
9. Xu Z, Niu L, Taylor JA. The ENmix DNA methylation analysis pipeline for Illumina BeadChip and comparisons with seven other preprocessing pipelines. *Clin. Epigenetics.* 13(1), 216 (2021).
10. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics Data.* 9, 22–24 (2016).
11. Pidsley R, Zotenko E, Peters TJ, *et al.* Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. *Genome Biol.* 17(1), 208 (2016).
12. Triche TJ, Weisenberger DJ, Van Den Berg D, Laird PW, Siegmund KD. Low-level processing of Illumina Infinium DNA Methylation BeadArrays. *Nucleic Acids Res.* 41(7), e90–e90 (2013).

13. Xu Z, Langie SAS, De Boever P, Taylor JA, Niu L. RELIC: a novel dye-bias correction method for Illumina Methylation BeadChip. *BMC Genomics*. 18(1), 4 (2017).
14. Teschendorff AE, Marabita F, Lechner M, *et al.* A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics*. 29(2), 189–196 (2013).
15. Pidsley R, Y Wong CC, Volta M, Lunnon K, Mill J, Schalkwyk LC. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics*. 14(1), 293 (2013).
16. Koestler DC, Christensen BC, Karagas MR, *et al.* Blood-based profiles of DNA methylation predict the underlying distribution of cell types. *Epigenetics*. 8(8), 816–826 (2013).
17. Jaffe AE, Irizarry RA. Accounting for cellular heterogeneity is critical in epigenome-wide association studies. *Genome Biol*. 15(2), R31 (2014).
18. Salas L, Koestler DC. FlowSorted.Blood.EPIC: Illumina EPIC data on immunomagnetic sorted peripheral adult blood cells. [Internet]. (2020). Available from: <https://github.com/immunomethylomics/FlowSorted.Blood.EPIC>.
19. Oksanen J, Blanchet FG, Friendly M, *et al.* vegan: Community Ecology Package [Internet]. . Available from: <https://CRAN.R-project.org/package=vegan>.
20. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 8(1), 118–127 (2007).
21. Price EM, Robinson WP. Adjusting for Batch Effects in DNA Methylation Microarray Data, a Lesson Learned. *Front. Genet.* [Internet]. 9 (2018). Available from: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00083/full>.
22. Zindler T, Frieling H, Neyazi A, Bleich S, Friedel E. Simulating ComBat: how batch correction can lead to the systematic introduction of false positive results in DNA methylation microarray studies. *BMC Bioinformatics*. 21(1), 271 (2020).
23. Fortin J-P, Labbe A, Lemire M, *et al.* Functional normalization of 450k methylation array data improves replication in large cancer studies. *Genome Biol*. 15(11), 503 (2014).
24. Lê S, Josse J, Husson F. FactoMineR: An R Package for Multivariate Analysis. *J. Stat. Softw.* 25(1), 1–18 (2008).
25. Maksimovic J, Oshlack A, Phipson B. Gene set enrichment analysis for genome-wide DNA methylation data. *Genome Biol*. 22(1), 173 (2021).
26. Phipson B, Maksimovic J, Oshlack A. missMethyl: an R package for analyzing data from Illumina's HumanMethylation450 platform. *Bioinformatics*. 32(2), 286–288 (2016).
27. Simillion C, Liechti R, Lischer HEL, Ioannidis V, Bruggmann R. Avoiding the pitfalls of gene set enrichment analysis with SetRank. *BMC Bioinformatics* [Internet]. 18(1) (2017). Available from: <http://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-017-1571-6>.

28. Alexa A, Rahnenführer J. Gene set enrichment analysis with topGO [Internet]. Available. Available from: <https://bioconductor.riken.jp/packages/3.2/bioc/vignettes/topGO/inst/doc/topGO.pdf>.
29. Karp PD, Midford PE, Caspi R, Khodursky A. Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics. *BMC Genomics*. 22(1), 191 (2021).
30. Manjang K, Tripathi S, Yli-Harja O, Dehmer M, Emmert-Streib F. Graph-based exploitation of gene ontology using GOxploreR for scrutinizing biological significance. *Sci. Rep.* 10(1), 16672 (2020).
31. Yu G. Gene Ontology Semantic Similarity Analysis Using GOSemSim [Internet]. In: *Stem Cell Transcriptional Networks: Methods and Protocols*. Kidder BL (Ed.), Springer US, New York, NY, 207–215 (2020) [cited 2021 Nov 3]. Available from: https://doi.org/10.1007/978-1-0716-0301-7_11.
32. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*. 26(7), 976–978 (2010).
33. Schlicker A, Domingues FS, Rahnenführer J, Lengauer T. A new measure for functional similarity of gene products based on Gene Ontology. *BMC Bioinformatics*. 7(1), 302 (2006).
34. Sayols S. rrvgo: Reduce + Visualize GO [Internet]. (2021). Available from: <https://bioconductor.org/packages/rrvgo/>.
35. Supek F, Bošnjak M, Škunca N, Šmuc T. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE*. 6(7), e21800 (2011).
36. Kolde R. pheatmap: Pretty Heatmaps [Internet]. (2019). Available from: <https://CRAN.R-project.org/package=pheatmap>.
37. Brooks ME, Kristensen K, Benthem KJ van, *et al.* glmmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *R J.* 9(2), 378–400 (2017).
38. Hartig F, Lohse L. DHARMA: Residual Diagnostics for Hierarchical (Multi-Level / Mixed) Regression Models [Internet]. (2022). Available from: <https://CRAN.R-project.org/package=DHARMA>.
39. Braun P, Han S, Nagahama Y, *et al.* 28 - IMAGE-CpG: development of a web-based search tool for genome-wide dna methylation correlation between live human brain and peripheral tissues within individuals. *Eur. Neuropsychopharmacol.* 29, S796 (2019).
40. Bühlmann P, Hothorn T. Boosting Algorithms: Regularization, Prediction and Model Fitting. *Stat. Sci.* 22(4), 477–505 (2007).
41. Hofner B, Boccuto L, Göker M. Controlling false discoveries in high-dimensional situations: boosting with stability selection. *BMC Bioinformatics*. 16(1), 144 (2015).