nature neuroscience

Corresponding Author:	Jonathan Mill	# Main Figures:	4
Manuscript Number:	NN-BC52734-T	# Supplementary Figures:	13
Manuscript Type:	Article	# Supplementary Tables:	17
		# Supplementary Videos:	0

Reporting Checklist for Nature Neuroscience

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. For more information, please read Reporting Life Sciences Research.

Please note that in the event of publication, it is mandatory that authors include all relevant methodological and statistical information in the manuscript.

Statistics reporting, by figure

- Please specify the following information for each panel reporting quantitative data, and where each item is reported (section, e.g. Results, & paragraph number).
- Each figure legend should ideally contain an exact sample size (n) for each experimental group/condition, where n is an exact number and not a range, a clear definition of how n is defined (for example x cells from x slices from x animals from x litters, collected over x days), a description of the statistical test used, the results of the tests, any descriptive statistics and clearly defined error bars if applicable.
- For any experiments using custom statistics, please indicate the test used and stats obtained for each experiment.
- Each figure legend should include a statement of how many times the experiment shown was replicated in the lab; the details of sample collection should be sufficiently clear so that the replicability of the experiment is obvious to the reader.
- For experiments reported in the text but not in the figures, please use the paragraph number instead of the figure number.

Note: Mean and standard deviation are not appropriate on small samples, and plotting independent data points is usually more informative. When technical replicates are reported, error and significance measures reflect the experimental variability and not the variability of the biological process; it is misleading not to state this clearly.

		TEST USED		n			DESCRIPTIVE STATS (AVERAGE, VARIANCE)		P VALUE		DEGREES OF FREEDOM & F/t/z/R/ETC VALUE	
	FIGURE NUMBER	WHICH TEST?	SECTION & PARAGRAPH #	EXACT VALUE	DEFINED?	SECTION & PARAGRAPH #	REPORTED?	SECTION & PARAGRAPH #	EXACT VALUE	SECTION & PARAGRAPH #	VALUE	SECTION & PARAGRAPH #
example	1a	one-way ANOVA	Fig. legend	9, 9, 10, 15	mice from at least 3 litters/group	Methods para 8	error bars are mean +/- SEM	Fig. legend	p = 0.044	Fig. legend	F(3, 36) = 2.97	Fig. legend
example	results, para 6	unpaired t-test	Results para 6	15	slices from 10 mice	Results para 6	error bars are mean +/- SEM	Results para 6	p = 0.0006	Results para 6	t(28) = 2.808	Results para 6

		TEST USED		n			DESCRIPTIVE STATS (AVERAGE, VARIANCE)		P VALUE		DEGREES OF FREEDOM & F/t/z/R/ETC VALUE	
	FIGURE NUMBER	WHICH TEST?	SECTION & PARAGRAPH #	EXACT VALUE	DEFINED?	SECTION & PARAGRAPH #	REPORTED?	SECTION & PARAGRAPH #	EXACT VALUE	SECTION & PARAGRAPH #	VALUE	SECTION & PARAGRAPH #
+ -	1	Linear regression	Online Methods, Methylation QTL (mQTL) analyses	Fetal (n = 166)	Number of post-mortem brain samples for each tissue-type	Main text paragraph 3/ Supplementary Table 1	Effect size (% DNA methylation change per allele)	Supplementary Table 3	N/A	N/A	N/A	N/A
+ -	3	N/A	N/A	N/A	N/A	N/A		fig legend		Supplementary Table		
+ -	2a	Linear regression	Online Methods, Methylation QTL (mQTL) analyses	Fetal (n = 166) PFC (n = 75) STR (n = 80) CER (n = 77)	Number of post-mortem brain samples for each tissue-type	Supplementary Table 1	Effect size (% DNA methylation change per allele)	Supplementary Table 3	NA	Supplementary Table 3	NA	NA
+ -	2b,c	Multilevel linear regression model	Online Methods, (Heterogeneity model)	Fetal (n = 166) PFC (n = 75) STR (n = 80) CER (n = 77)	Number of post-mortem brain samples for each tissue-type	Supplementary Table 1	Effect size (% DNA methylation change per allele)	Heterogeneity Supplementary table	b) P = 7.23x10-40 c) P = 8.39x10-36	Heterogeneity Supplementary table	b) 2 df; chisq statistic = 180.25. c) 2df; chisq statistic = 161.53.	
+ -	4	Multilevel linear regression model; Linear regression	Online Methods, (Methylation QTL (mQTL) analyses; Heterogeneity model)	Fetal (n = 166) PFC (n = 75) STR (n = 80) CER (n = 77)	Number of post-mortem brain samples for each tissue-type	Supplementary Table	Effect size (% DNA methylation change per allele)	SCZ variants supp table	NA	SCZ variants supp table	NA	NA

Representative figures

1. Are any representative images shown (including Western blots and immunohistochemistry/staining) in the paper?

   If so, what figure(s)?

N/A

2. For each representative image, is there a clear statement of how many times this experiment was successfully repeated and a discussion of any limitations in repeatability?

   If so, where is this reported (section, paragraph #)?

N/A

Statistics and general methods

1. Is there a justification of the sample size?

   If so, how was it justified?

   Where (section, paragraph #)?

   Even if no sample size calculation was performed, authors should report why the sample size is adequate to measure their effect size.

The fetal brain "discovery" cohort represents a unique sample of fetal brain tissues spanning 56 to 166 days post-conception - we used all available samples at the time, representing to our knowledge the largest set of human fetal brain tissues to be genomically profiled. Despite a conservative Bonferroni corrected significance threshold we obtained > 16,000 significant mQTLs in our discovery sample; the vast majority of these replicated in our smaller replication samples, indicating that the discovery sample was of sufficient size for the analysis.

2. Are statistical tests justified as appropriate for every figure?

   Where (section, paragraph #)?

Statistical tests are described in Figure Legends, and the text referring to them.

   a. If there is a section summarizing the statistical methods in the methods, is the statistical test for each experiment clearly defined?

The Online Methods paragraphs: Methylation QTL (mQTL) analyses, detail the primary statistical analysis and Online Methods paragraphs, Enrichment of regulatory regions, Enrichment of GWAS variants, Heterogenity model, and Colocalisation analysis detail additional analyses.

   b. Do the data meet the assumptions of the specific statistical test you chose (e.g. normality for a parametric test)?

      Where is this described (section, paragraph #)?

NA

   c. Is there any estimate of variance within each group of data?

      Is the variance similar between groups that are being statistically compared?

      Where is this described (section, paragraph #)?

NA

   d. Are tests specified as one- or two-sided?

two-sided: mQTL analyses; heterogeneity model; enrichments of regulatory features.
one-sided: Enrichment of GWAS variants

   e. Are there adjustments for multiple comparisons?

Bonferroni adjusted significance threshold is used throughout

3. Are criteria for excluding data points reported?

   Was this criterion established prior to data collection?

   Where is this described (section, paragraph #)?

Yes: Online Methods Genome-wide quantification of DNA methylation; Genome-wide SNP genotyping; Imputation.
Data points were only excluded for quality control reasons, or because the location of probes could cause a technical artefact (e.g. probes overlying SNPs)

4. Define the method of randomization used to assign subjects (or samples) to the experimental groups and to collect and process data. If no randomization was used, state so. Where does this appear (section, paragraph #)?	Samples were randomized by gender for array processing. Online Methods Genome-wide quantification of DNA methylation.
5. Is a statement of the extent to which investigator knew the group allocation during the experiment and in assessing outcome included? If no blinding was done, state so. Where (section, paragraph #)?	N/A
6. For experiments in live vertebrates, is a statement of compliance with ethical guidelines/regulations included? Where (section, paragraph #)?	N/A
7. Is the species of the animals used reported? Where (section, paragraph #)?	N/A
8. Is the strain of the animals (including background strains of KO/ transgenic animals used) reported? Where (section, paragraph #)?	N/A
9. Is the sex of the animals/subjects used reported? Where (section, paragraph #)?	Yes; Supplementary Table 1.
10. Is the age of the animals/subjects reported? Where (section, paragraph #)?	Yes; Supplementary Table 1.
11. For animals housed in a vivarium, is the light/dark cycle reported? Where (section, paragraph #)?	N/A
12. For animals housed in a vivarium, is the housing group (i.e. number of animals per cage) reported? Where (section, paragraph #)?	N/A
13. For behavioral experiments, is the time of day reported (e.g. light or dark cycle)? Where (section, paragraph #)?	N/A
14. Is the previous history of the animals/subjects (e.g. prior drug administration, surgery, behavioral testing) reported? Where (section, paragraph #)?	N/A

a. If multiple behavioral tests were conducted in the same group of animals, is this reported?

Where (section, paragraph #)?

N/A

15. If any animals/subjects were excluded from analysis, is this reported?

Where (section, paragraph #)?

Yes: Online Methods. Methylation QTL (mQTL) analyses.

a. How were the criteria for exclusion defined?

Where is this described (section, paragraph #)?

Samples were only excluded if they did not pass our stringent quality control metrics: Online Methods: Methylation QTL (mQTL) analyses.

b. Specify reasons for any discrepancy between the number of animals at the beginning and end of the study.

Where is this described (section, paragraph #)?

Failed methylation array QC, failed SNP array QC. Online Methods: Genome-wide quantification of DNA methylation; Genome-wide SNP genotyping; Methylation QTL (mQTL) analyses.

Reagents

1. Have antibodies been validated for use in the system under study (assay and species)?

NA

a. Is antibody catalog number given?

Where does this appear (section, paragraph #)?

NA

b. Where were the validation data reported (citation, supplementary information, Antibodypedia)?

Where does this appear (section, paragraph #)?

NA

2. Cell line identity

a. Are any cell lines used in this paper listed in the database of commonly misidentified cell lines maintained by ICLAC and NCBI Biosample?

Where (section, paragraph #)?

N/A

b. If yes, include in the Methods section a scientific justification of their use--indicate here in which section and paragraph the justification can be found.

N/A

c. For each cell line, include in the Methods section a statement that specifies:

- the source of the cell lines
- have the cell lines been authenticated? If so, by which method?
- have the cell lines been tested for mycoplasma contamination?

Where (section, paragraph #)?

N/A

Data deposition

Data deposition in a public repository is mandatory for:

a. Protein, DNA and RNA sequences
b. Macromolecular structures
c. Crystallographic data for small molecules
d. Microarray data

Deposition is strongly recommended for many other datasets for which structured public repositories exist; more details on our data policy are available here. We encourage the provision of other source data in supplementary information or in unstructured repositories such as Figshare and Dryad.

We encourage publication of Data Descriptors (see Scientific Data) to maximize data reuse.

1. Are accession codes for deposit dates provided? Where (section, paragraph #)?	The methylation data are publically available through GEO and can be found under accession numbers: GSE58885, GSE61431, GSE61380. The genotype data is deposited in dbGAP. Online Methods, Genome-wide quantification of DNA methylation.

Computer code/software

Any custom algorithm/software that is central to the methods must be supplied by the authors in a usable and readable form for readers at the time of publication. However, referees may ask for this information at any time during the review process.

1. Identify all custom software or scripts that were required to conduct the study and where in the procedures each was used.	No custom software was used, scripts are available as a supplementary item.
2. If computer code was used to generate results that are central to the paper's conclusions, include a statement in the Methods section under "**Code availability**" to indicate whether and how the code can be accessed. Include version information as necessary and any restrictions on availability.	All software used in this analysis has been referenced. Analysis scripts are available for download from http://epigenetics.essex.ac.uk/mQTL/ and this has been stated in the manuscript.

Human subjects

1. Which IRB approved the protocol? Where is this stated (section, paragraph #)?	Fetal samples: Ethical approval for the HDBR was granted by the Royal Free Hospital research ethics committee under reference 08/H0712/34 and Human Tissue Authority (HTA) material storage license 12220 Adult samples: NHS reference number 08/MRE09/38; the HTA license number for the LBBND brain bank is 12293 This is stated in Online Methods.
2. Is demographic information on all subjects provided? Where (section, paragraph #)?	Supplementary Table 1.
3. Is the number of human subjects, their age and sex clearly defined? Where (section, paragraph #)?	Supplementary Table 1.

4. Are the inclusion and exclusion criteria (if any) clearly specified?

   Where (section, paragraph #)?

   Samples were only excluded if they failed quality control. This is fully descfibed in the Online Methods.

5. How well were the groups matched?

   Where is this information described (section, paragraph #)?

   N/A

6. Is a statement included confirming that informed consent was obtained from all subjects?

   Where (section, paragraph #)?

   Online Methods, Human brain samples

7. For publication of patient photos, is a statement included confirming that consent to publish was obtained?

   Where (section, paragraph #)?

   N/A

fMRI studies

For papers reporting functional imaging (fMRI) results please ensure that these minimal reporting guidelines are met and that all this information is clearly provided in the methods:

1. Were any subjects scanned but then rejected for the analysis after the data was collected?

   N/A

   a. If yes, is the number rejected and reasons for rejection described?

      Where (section, paragraph #)?

      N/A

2. Is the number of blocks, trials or experimental units per session and/or subjects specified?

   Where (section, paragraph #)?

   N/A

3. Is the length of each trial and interval between trials specified?

   N/A

4. Is a blocked, event-related, or mixed design being used? If applicable, please specify the block length or how the event-related or mixed design was optimized.

   N/A

5. Is the task design clearly described?

   Where (section, paragraph #)?

   N/A

6. How was behavioral performance measured?

   N/A

7. Is an ANOVA or factorial design being used?

   N/A

8. For data acquisition, is a whole brain scan used?

   If not, state area of acquisition.

   N/A

a. How was this region determined?	N/A
9. Is the field strength (in Tesla) of the MRI system stated?	N/A
a. Is the pulse sequence type (gradient/spin echo, EPI/spiral) stated?	N/A
b. Are the field-of-view, matrix size, slice thickness, and TE/TR/ flip angle clearly stated?	N/A
10. Are the software and specific parameters (model/functions, smoothing kernel size if applicable, etc.) used for data processing and pre-processing clearly stated?	N/A
11. Is the coordinate space for the anatomical/functional imaging data clearly defined as subject/native space or standardized stereotaxic space, e.g., original Talairach, MNI305, ICBM152, etc? Where (section, paragraph #)?	N/A
12. If there was data normalization/standardization to a specific space template, are the type of transformation (linear vs. nonlinear) used and image types being transformed clearly described? Where (section, paragraph #)?	N/A
13. How were anatomical locations determined, e.g., via an automated labeling algorithm (AAL), standardized coordinate database (Talairach daemon), probabilistic atlases, etc.?	N/A
14. Were any additional regressors (behavioral covariates, motion etc) used?	N/A
15. Is the contrast construction clearly defined?	N/A
16. Is a mixed/random effects or fixed inference used?	N/A
a. If fixed effects inference used, is this justified?	N/A
17. Were repeated measures used (multiple measurements per subject)?	N/A
a. If so, are the method to account for within subject correlation and the assumptions made about variance clearly stated?	N/A
18. If the threshold used for inference and visualization in figures varies, is this clearly stated?	N/A
19. Are statistical inferences corrected for multiple comparisons?	N/A
a. If not, is this labeled as uncorrected?	N/A

20. Are the results based on an ROI (region of interest) analysis? N/A

    a. If so, is the rationale clearly described? N/A

    b. How were the ROI's defined (functional vs anatomical localization)? N/A

21. Is there correction for multiple comparisons within each voxel? N/A

22. For cluster-wise significance, is the cluster-defining threshold and the corrected significance level defined? N/A

Additional comments

Additional Comments