nature neuroscience

Corresponding Author: Pamela Sklar

Manuscript Number: # NN-RS56376-T

Manuscript Type: Article

Main Figures: 7

Supplementary Figures: 12

Supplementary Tables: 4

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
+ -												

Representative figures

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

   If so, what figure(s)?

Yes, Supplementary Figure 10 shows immunohistochemistry/ staining.

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 #)?

These images were downloaded from Allen Human Brain Atlas.

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.

We sequenced a cohort of SCZ and control subjects that is an order of magnitude larger than early RNA sequencing studies. The sample size is 258 from subjects with schizophrenia, 279 control subjects, and 55 affective disorder subjects whose samples were used for eQTL estimation. We detected 16,423 genes by RNA-seq at sufficient quantity per subject for statistical analyses (based on Ensembl models). The sample size is the largest available.

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

   Where (section, paragraph #)?

Statistical tests are justified and described here:

Figure 1: MatrixEQTL Regression, FDR<0.05; Kolmogorov-Smirnov (KS) test, p<0.05 (caption, enrichment)
Figure 2: Sherlock, Bayesian model and using Frequentist equivalent inference (Sherlock p_corrected ≤ 0.05, figure); MatrixEQTL Regression & logistic regression p-values, in figure.
Figure 3 : Student's t-test, p<0.05 (figure)
Figure 4 : Student's t-test, p<0.05 (figure)
Figure 5 : Limma regression, testing case status in Best Average Model (FDR<0.050); Pearson correlation (text)
Figure 6: WGCNA, no test; Connectivity and enrichment, permutation tests (Bonferroni correction)
Figure 7: Power calculations use standard statistical theory.
Table 1: No test
Supplementary Fig. 1: Logistic regression, t-test of slope parameter (predict SCZ).
Supplementary Fig. 2. No test.
Supplementary Fig. 3: Pearson correlation, no test.
Supplementary Fig. 4. Limma, BIC model selection for best average model; hierarchical Clustering, reduce dimension of batch covariate; model fit, no test; Surrogate Variable Analysis, no test.
Supplementary Fig. 5: Sherlock adjusted for multiple testing.
Supplementary Fig. 6: Spearman correlation, p<0.05, t-test
Supplementary Fig. 7: No test.
Supplementary Fig. 8: Limma-based Regression, FDR < 0.05
Supplementary Fig. 9: Pearson correlation, p<0.05
Supplementary Figure 10. In situ hybridization, no test
Supplementary Fig. 11: Isoforms, Limma regression, FDR<0.050
Supplementary Fig. 12: CiberSort and CellMix deconvolution (mixture model, no test of significance)

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

Statistical tests are defined in Online methods, Supplementary Text, and Figure legends.

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 #)?

Tests were chosen to match the data and assumptions of the model. See Online methods and Supplementary Text.

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 #)?

Estimated variances are typically displayed in the figures and were roughly the same for all contrasts.

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

Almost all tests performed were two-sided. We performed one-sided tests when the direction of deviation was established a priori and these were noted as such.

e. Are there adjustments for multiple comparisons?

Yes, tests in the paper were adjusted for multiple testing, either using Bonferroni or FDR.

3. To promote transparency, *Nature Neuroscience* has stopped allowing bar graphs to report statistics in the papers it publishes. If you have bar graphs in your paper, please make sure to switch them to dot-plots (with central and dispersion statistics displayed) or to box-and-whisker plots to show data distributions.

We gave presented the distributions of the data by violin plots, when appropriate.

4. Are criteria for excluding data points reported?

Was this criterion established prior to data collection?

Where is this described (section, paragraph #)?

All quality control (QC) procedures and their criteria are reported. In a genomics paper of this type, QC criteria are typically based on the properties of the data, rather than set a priori. QC procedures are described in Online Methods and Supplementary Text.

5. 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 #)?

To control for batch effects, multiple randomization steps were introduced and DNA and RNA isolation, library preparation and RNA sequencing were performed at one site, Mount Sinai.

6. 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 #)?

This is not applicable to our study.

7. For experiments in live vertebrates, is a statement of compliance with ethical guidelines/regulations included?

Where (section, paragraph #)?

Yes, all zebrafish assays were performed utilizing the wild-type ZDR strain in accordance with standard zebrafish husbandry practices at Duke University (Online Methods).

8. Is the species of the animals used reported?

   Where (section, paragraph #)?

   zebrafish

9. Is the strain of the animals (including background strains of KO/ transgenic animals used) reported?

   Where (section, paragraph #)?

   NA

10. Is the sex of the animals/subjects used reported?

    Where (section, paragraph #)?

    NA

11. Is the age of the animals/subjects reported?

    Where (section, paragraph #)?

    Yes, (Online Methods)

12. For animals housed in a vivarium, is the light/dark cycle reported?

    Where (section, paragraph #)?

    NA

13. For animals housed in a vivarium, is the housing group (i.e. number of animals per cage) reported?

    Where (section, paragraph #)?

    NA

14. For behavioral experiments, is the time of day reported (e.g. light or dark cycle)?

    Where (section, paragraph #)?

    NA

15. Is the previous history of the animals/subjects (e.g. prior drug administration, surgery, behavioral testing) reported?

    Where (section, paragraph #)?

    NA

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

       Where (section, paragraph #)?

       NA

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

    Where (section, paragraph #)?

    NA

    a. How were the criteria for exclusion defined?

       Where is this described (section, paragraph #)?

    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 #)?

       NA

Reagents

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

   Yes

   a. Is antibody catalog number given?

      Where does this appear (section, paragraph #)?

      Yes. See Online Methods.

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

      Where does this appear (section, paragraph #)?

2. Cell line identity

   NA

   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 #)?

   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.

   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 #)?

      Control fibroblast biopsies were obtained as part of a longitudinal study Dr. Judith Rapoport (NIMH). hiPSCs were derived at ISMMS as described previously 1, 2 using Cytotune Sendai virus (ThermoFisher Scientific). Control NPCs were differentiated at ISMMS with dual-SMAD inhibition 3 to improve yield, detailed methods as described 4; NPC validation as shown 1, 2. All hiPSC and NPCs in the laboratory are tested monthly using MycoAlert (Lonza) to ensure they remain mycoplasma free.

      See Online Methods

      References

      1. Lee, I.S., et al. Characterization of molecular and cellular phenotypes associated with a heterozygous CNTNAP2 deletion using patient-derived hiPSC neural cells. NPJ Schizophrenia 1, 15019 (2015).
      2. Topol, A., et al. Dysregulation of miRNA-9 in a Subset of Schizophrenia Patient-Derived Neural Progenitor Cells. Cell reports 15, 1024-1036 (2016).
      3. Chambers, S.M., et al. Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol 27, 275-280 (2009).
      4. Topol, A., Tran, N.N. & Brennand, K.J. A guide to generating and using hiPSC derived NPCs for the study of neurological diseases. Journal of visualized experiments : JoVE, e52495 (2015).

Data availability

Provide a Data availability statement in the Methods section under "Data availability", which should include, where applicable:

- Accession codes for deposited data
- Other unique identifiers (such as DOIs and hyperlinks for any other datasets)
- At a minimum, a statement confirming that all relevant data are available from the authors
- Formal citations of datasets that are assigned DOIs
- A statement regarding data available in the manuscript as source data
- A statement regarding data available with restrictions

See our data availability and data citations policy page for more information.

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.

Where is the Data Availability statement provided (section, paragraph #)?

All results are available through the CommonMind Knowledge Portal (www.synapse.org/CMC).

All data are available at https://www.nimhgenetics.org/available_data/commonmind/.

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.

All scripts used are available on CommonMind Knowledge Portal. https://bitbucket.org/commonmind/commonmind/src/9be020b462525cb1c3eca2c005543761d33ed26e/scripts/phaseI

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.

Key scripts used to analyze these data are available on CommonMind Knowledge Portal.

Human subjects

1. Which IRB approved the protocol?

   Where is this stated (section, paragraph #)?

   Mount Sinai, University of Pittsburgh and University of Pennsylvania, as appropriate. See Online Methods.

2. Is demographic information on all subjects provided?

   Where (section, paragraph #)?

   Raw data are provided, see Data Availability. Summary demographic data are available in Supplementary Table 1.

3. Is the number of human subjects, their age and sex clearly defined?

   Where (section, paragraph #)?

   Yes, raw data are provided, see Data Availability. Summary data are available in Supplementary Table 1.

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

   Where (section, paragraph #)?

   Described under QC in Supplementary Information.

5. How well were the groups matched?

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

   Groups were not matched, aside from University of Pittsburgh samples.

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

   Where (section, paragraph #)?

   See Online Methods.

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

   Where (section, paragraph #)?

   NA

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?

   NA

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

      Where (section, paragraph #)?

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

   Where (section, paragraph #)?

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

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.

5. Is the task design clearly described?

   Where (section, paragraph #)?

6. How was behavioral performance measured?

7. Is an ANOVA or factorial design being used?

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

   If not, state area of acquisition.

   a. How was this region determined?

9. Is the field strength (in Tesla) of the MRI system stated?

   a. Is the pulse sequence type (gradient/spin echo, EPI/spiral) stated?

   b. Are the field-of-view, matrix size, slice thickness, and TE/TR/flip angle clearly stated?

10. Are the software and specific parameters (model/functions, smoothing kernel size if applicable, etc.) used for data processing and pre-processing clearly stated?

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 #)?

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 #)?

13. How were anatomical locations determined, e.g., via an automated labeling algorithm (AAL), standardized coordinate database (Talairach daemon), probabilistic atlases, etc.?

14. Were any additional regressors (behavioral covariates, motion etc) used?

15. Is the contrast construction clearly defined?

16. Is a mixed/random effects or fixed inference used?

   a. If fixed effects inference used, is this justified?

17. Were repeated measures used (multiple measurements per subject)?

a. If so, are the method to account for within subject correlation and the assumptions made about variance clearly stated?

18. If the threshold used for inference and visualization in figures varies, is this clearly stated?

19. Are statistical inferences corrected for multiple comparisons?

a. If not, is this labeled as uncorrected?

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

a. If so, is the rationale clearly described?

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

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

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

Additional comments

Additional Comments