nature neuroscience

Corresponding Author: Christopher A. Walsh

Manuscript Number: NN-A56531D

Manuscript Type: Article

Main Figures: 7

Supplementary Figures: 1

Supplementary Tables: 15

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 #
+ -	2a	hypergeometric test	Results para 13	1993	observation counted from exome data	Results para 13	-	-	p = 0.024, 0.041 and 0.32 respectively	Results para 13	-	-
+ -	2b	Wilcoxon rank sum test	Results para 14	1048	odds ratios calculated using a previously published method	Results para 14	-	-	In Fig. 2B	Results para 14	-	-
+ -	3	Wilcoxon rank sum test	Results para 15	16	odds ratios calculated using a previously published method	Results para 15	-	-	p = 5.4×10-3 for top hit	Results para 15	-	-
+ -	4c	unpaired t-test	Results para 24	3	observation from qPCR results	Results para 24	all replicate points are shown in red dots	Fig. 4 legend	p = 0.015 and 0.0031	Fig. 4c	-	-
+ -	1c	Pearson's correlation	Results para 5	N=49 for CloneSeq, N=46 for Pyroseq, N=42 for MiSeq	validated mutations using 3 technologies - exact variants found in Supp Table 3	Results para 5	R2 = 0.85 for CloneSeq and MiSeq, R2 = 0.63 for CloneSeq and Pyroseq	Results para 5	-	-	-	-

Representative figures

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

   If so, what figure(s)?

   Fig. 4c - qPCR results

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 #)?

   Yes, methods paras 18-20.

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 obtained the largest possible set of ASD exomes, published and unpublished, in order to detect the largest number of de novo and mosaic mutations.
   Mentioned in Results para 1.
   Given that we had >5900 exomes, and that the rate of mosaics from Group C is 7.5%, this allowed us to detect a substantial number of mosaics to perform analyses on the properties of these mosaics.

2. Are statistical tests justified as appropriate for every figure? Where (section, paragraph #)?	Yes. Listed in detail above.
a. If there is a section summarizing the statistical methods in the methods, is the statistical test for each experiment clearly defined?	Yes. Listed above and in the methods section.
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 #)?	Yes, log2 transformed expression values from qPCR were used for Fig. 4c.
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 #)?	We used non-parametric tests, such as Fisher's Exact Test and Wilcoxon rank sum test, whenever possible to compare the groups. Described in Methods para 20.
d. Are tests specified as one- or two-sided?	Yes.
e. Are there adjustments for multiple comparisons?	Yes.
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.	Done.
4. Are criteria for excluding data points reported? Was this criterion established prior to data collection? Where is this described (section, paragraph #)?	Yes. We described the filters we used to exclude samples with a high rate of de novo and mosaic mutations. The criteria was based on analyzing the data, since we expect the rates of these mutations to follow a Poisson distribution and removed outlier samples. Described in Methods para 2.
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 #)?	No randomization was used.
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 #)?	No blinding was done.
7. For experiments in live vertebrates, is a statement of compliance with ethical guidelines/regulations included? Where (section, paragraph #)?	Not applicable.

8. Is the species of the animals used reported?

   Where (section, paragraph #)?

Not applicable.

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

   Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

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

       Where (section, paragraph #)?

Not applicable.

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

    Where (section, paragraph #)?

Not applicable.

    a. How were the criteria for exclusion defined?

       Where is this described (section, paragraph #)?

Not applicable.

    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 #)?

Not applicable.

Reagents

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

   Not applicable.

   a. Is antibody catalog number given?

      Where does this appear (section, paragraph #)?

      Not applicable.

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

      Where does this appear (section, paragraph #)?

      Not applicable.

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 #)?

      No.

   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.

      Not applicable.

   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 #)?

      - Source of the cell-lines: N2A cells from ATCC
      - Cell lines were authenticated using mouse-specific qPCR primers
      - Cell lines were negative for mycoplasma contamination.

      Described in Methods para 19.

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 #)?

We have listed a data availability statement at the end of the Methods section, but will check that the latest datasets are uploaded, and will update the accession codes prior to publication.

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 our scripts (written in R or Perl) are available upon request, and have been uploaded to the website in the text. We have also described the methods in detail under the Methods section.

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.

Done.

Human subjects

1. Which IRB approved the protocol?

   Where is this stated (section, paragraph #)?

   Research performed on samples and data of human origin was conducted according to protocols approved by the institutional review boards of Boston Children's Hospital and Beth Israel Deaconess Medical Center. Methods para 1.

2. Is demographic information on all subjects provided?

   Where (section, paragraph #)?

   Information about the number of subjects are shown in Supplementary Table 1, and the detailed IDs for each cohort is shown in Supplementary Table 2. More information about the subjects can be found in previous publications (Iossifov I et al., 2014 and De Rubeis S et al., 2014).

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

   Where (section, paragraph #)?

   Yes, sex has been defined for each individual in Supplementary Table 2. More information about the subjects can be found in previous publications (Iossifov I et al., 2014 and De Rubeis S et al., 2014).

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

   Where (section, paragraph #)?

   Yes. Methods para 2.

5. How well were the groups matched?

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

   We compared the probands to their unaffected siblings.

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

   Where (section, paragraph #)?

   Not applicable as the identity of the individuals are anonymous to the researchers.

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

   Where (section, paragraph #)?

   Not applicable.

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?

   Not applicable.

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

      Where (section, paragraph #)?

      Not applicable.

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

   Where (section, paragraph #)?

   Not applicable.

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

   Not applicable.

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.	Not applicable.
5. Is the task design clearly described? Where (section, paragraph #)?	Not applicable.
6. How was behavioral performance measured?	Not applicable.
7. Is an ANOVA or factorial design being used?	Not applicable.
8. For data acquisition, is a whole brain scan used? If not, state area of acquisition.	Not applicable.
a. How was this region determined?	Not applicable.
9. Is the field strength (in Tesla) of the MRI system stated?	Not applicable.
a. Is the pulse sequence type (gradient/spin echo, EPI/spiral) stated?	Not applicable.
b. Are the field-of-view, matrix size, slice thickness, and TE/TR/ flip angle clearly stated?	Not applicable.
10. Are the software and specific parameters (model/functions, smoothing kernel size if applicable, etc.) used for data processing and pre-processing clearly stated?	Not applicable.
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 #)?	Not applicable.
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 #)?	Not applicable.
13. How were anatomical locations determined, e.g., via an automated labeling algorithm (AAL), standardized coordinate database (Talairach daemon), probabilistic atlases, etc.?	Not applicable.
14. Were any additional regressors (behavioral covariates, motion etc) used?	Not applicable.
15. Is the contrast construction clearly defined?	Not applicable.
16. Is a mixed/random effects or fixed inference used?	Not applicable.

a. If fixed effects inference used, is this justified?	Not applicable.
17. Were repeated measures used (multiple measurements per subject)?	Not applicable.
a. If so, are the method to account for within subject correlation and the assumptions made about variance clearly stated?	Not applicable.
18. If the threshold used for inference and visualization in figures varies, is this clearly stated?	Not applicable.
19. Are statistical inferences corrected for multiple comparisons?	Not applicable.
a. If not, is this labeled as uncorrected?	Not applicable.
20. Are the results based on an ROI (region of interest) analysis?	Not applicable.
a. If so, is the rationale clearly described?	Not applicable.
b. How were the ROI's defined (functional vs anatomical localization)?	Not applicable.
21. Is there correction for multiple comparisons within each voxel?	Not applicable.
22. For cluster-wise significance, is the cluster-defining threshold and the corrected significance level defined?	Not applicable.

Additional comments

Additional Comments	