nature research

Corresponding author(s): Joseph G. Gleeson, Jonathan Sebat

Last updated by author(s): Nov 15, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used to collect the data in this study.
Data analysis	Commercial and available software used directly to analyze and visualize data: QuantaSoft (1.7.4.0917) and QuantaSoft Analysis Pro (1.0.596), GraphPad Prism 5, R (3.4.1 and ggbio package), Python (3.6.5 and matplotlib, seaborn, pysam, pandas, scipy modules), CNView, Strelka 2 (v2.9.2), MuTect 2 (v2.1), MosaicHunter (v1.0), HipSTR (v0.6), Pysim, Picard's MarkDuplicates (v1.83), BWA (v0.7.8), Genome Analysis ToolKit (GATK version 3.5-0-g36282e4), SnpEff (v4.2), SnpSift (v4.2), Triodenovo (v0.06), BWA mem (version 0.7.15-r1140), sambamba (version 0.6.6), samtools (v1.9), and bedtools (v2.25.0). Variant calling was done using standard pipelines and programs as described in the methods section of this paper.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated during the current study will be made available in public data bases and are available upon request from J.G.G. on reasonable request. Additionally, summary tables of all data are included as supplementary information files.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size for deep sequencing was based on previous mosaic studies on blood and the expected number of de novo mutations per trio. Based on this, we estimated that we require at least 10 trios to obtain a minimum of 500 DNMs for interrogation (current n=14 trios with n=912 dSNVs), to obtain >10 mosaic variants for analysis. Sequencing depth was determined by theoretical considerations (binomial model) and simulated data. For the analysis of pathogenic DNMs, based on previous sperm mosaicism studies on epileptic disorders, we estimated that around 10% of pathogenic variants may be mosaic. Thus, we expected that we need at least 15 DNM families to observe a mosaic variant with ~80% chance (n=20 pathogenic DNMs).
Data exclusions	No full data sets were excluded in this study. Individual variants were filtered based on previously established best practices for variant calling and mosaicism detection.
Replication	This is a descriptive study of fixed, unique cohort. Replication was not attempted.
Randomization	This is a descriptive study. No randomization was performed.
Blinding	No groups were allocated by the scientists.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	Human participants were derived from two autism cohorts (O.D. and J.S.). The study included parents, affected children and unaffected children of all sexes and ages.
Recruitment	The sole inclusion criterion was a medical diagnosis of autism spectrum disorder with or without epilepsy and a molecular diagnosis by next generation sequencing. The patient cohort was solely used for genetic analysis and the ascertainment of appropriate samples was performed as stated in our IRB protocol. In short, fathers of individuals with autism were previously enrolled through independent studies and subjected to whole genome or exome sequencing. All fathers with a pathogenic DNM who agreed to be recontacted for additional studies were informed of this study. As the existence of sperm mosaicism is independent of any possible selection bias (e.g. religious conflicts, history of vasectomy), willingness to participate and provide a semen sample is expected to be independent from measured outcomes.
Ethics oversight	IRB at UC, San Diego

Note that full information on the approval of the study protocol must also be provided in the manuscript.