nature research

Corresponding author(s): Peter J. Park, Christopher A. Walsh

Last updated by author(s): May 8, 2020

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 our Editorial Policies 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 data in this study.
Data analysis	We aligned FASTQ files using BWA-MEM version 0.7.8. We used MuTect2 with panel of normals from GATK version 3.5 to call variants. We identified somatic variants using a custom method called MosaicForecast, which is published separately. Our analysis also used RepeatMasker version 4.0.7, Meerkat version 0.189, ANNOVAR version 2017-07-17, LOFTEE version 1.0.3, CADD version 1.4, SIFT 4G, LRT (original version), MutationTaster2, MutationAssesor release 3, FATHMM version 2.3, Provean version 1.1.3, MetaSVM version 1.0, MetaLR (original version), M-CAP version 1.4, MutPred2, Eigen version 1.1, VGAM version 1.1-1, deconstructSigs version 1.8, SAMtools-1.10, and AsymTools (original version).

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 and 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

Whole-genome sequencing data is available from the National Institute of Mental Health Data Archive (DOI: 10.15154/1503337).

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	An attempt was made at procuring all donated autism brains in the world for which fresh-frozen prefrontal cortex was available. In total, 61 ASD brains were obtained. Due to the high cost of ultra-deep sequencing and importance of focusing resources on autism cases, 15 matched controls were selected and sequenced. These sample sizes are small compared to studies on peripheral autism DNA, but the sample sizes are very large in comparison to prior WGS studies of human brain.
Data exclusions	Two autism samples was excluded from analysis due to likely sequencing contamination. Extremely high rates of mutation calling and extremely low validation success rate were consistent with contamination at the level of sequencing. These criteria were not established prior to data analysis, but the inadequacy and contamination of these samples was abundantly obvious.
Replication	Deep re-sequencing of a large set of putative mosaic mutations was attempted, and confirmed a very high validation success rate. Protein damaging effects were also determined via multiple published softwares. These methods ensured replication of results.
Randomization	This is not relevant to our study. ADI-R scores were obtained whenever possible in order to verify that ASD cases did indeed have autism, and communication with brain banks confirmed that controls were neurologically normal.
Blinding	Blinding was not relevant to this study. In order to obtain samples from brain banks, it was necessary to specify and know disease status, and this information was known to all researchers throughout the study.

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 and archaeology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Neuro-2a cells from ATCC CCL-131
Authentication	Cell lines were not authenticated
Mycoplasma contamination	Regular mycoplasma testing was conducted during culture of cells used in this study and was negative.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.