nature portfolio

Corresponding author(s): Joseph G. Gleeson

Last updated by author(s): Dec 28, 2022

Reporting Summary

Nature Portfolio 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 Portfolio 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	WES, AmpliSeq TASeq were performed on Illumina instruments using their proprietary platform
Data analysis	Codes for data analysis pipelines as well as codes to generate the figures are freely available on GitHub at https://github.com/shishenyxx/MCD_mosaic. Specifically we used the following software for data analysis: Primer3 with a Python v3.7.3 wrapper; BWA v3.7.16a; Sambamba v0.6.7; SAMtools v1.7 for WES, v1.9 for Ampliseq and TASeq; Picard v2.12.1 for WES, v2.18.27 for AmpliSeq and TASeq; Strelka2 v2.9.2; GATK v3.7.0; MuTect2 from GATK v3.8.1 for WES, v4.0.4 for AmpliSeq; MosaicHunter v1.0; DeepMosaic (https://github.com/Virginiaxu/DeepMosaic); BEDTools v2.27.1; R v4.0.1; r-gplots v3.1.1; ggplot2 v3.3.5; r-seurat v4.0.5; matplotlib v3.5.0; demuxlet v1.0; maftools v2.6.05; Cell Ranger v6.0.2; r-wgcna v1.69; Cytoscape v3.9.0; ClueGO v2.5.8; GraphPad Prism v9.4.0; Mutalisk (http://mutalisk.org/); STRING v11 (https://string-db.org/)

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 Portfolio 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 description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

WES and AmpliSeq data are deployed on NIMH Data Archive under study number 1484 “Comprehensive multi-omic profiling of somatic mutations in malformations

of cortical development" and SRA under accession number PRJNA821916: "Comprehensive multi-omic profiling of somatic mutations in malformations of cortical development". The BSMN neurotypical brain data are available at NIMH Data Archive (NDA study 644, 792 and 919, https://nda.nih.gov/study.html?tab=result&id=644, https://nda.nih.gov/study.html?tab=result&id=792 and https://nda.nih.gov/study.html?tab=result&id=919) and SRA: PRJNA736951. The raw and processed snRNAseq dataset was deposited in Gene Expression Omnibus (GEO) under accession number GSE218022.
gnomAD: https://gnomad.broadinstitute.org/
COSMIC: https://cancer.sanger.ac.uk/cosmic/
STRING: https://string-db.org/
Single-cell RNA-seq data from developing cortex (Nowakowski et al., 2017): https://cells.ucsc.edu/?ds=cortex-dev
GRCh37 genome accession number: PRJNA31257

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	No sample size calculation was performed. We collected resected tissues from as many cases as possible within a 10 year period of recruitment. The sample size of 283 patients is substantially larger than the largest previous cohort study to date, enabling discovery of many novel recurrent and non-recurrent somatically mutated genes within and beyond the mTOR pathway.
Data exclusions	A union of different pipelines was selected to achieve maximum sensitivity. The 1181 mosaic candidates from the combined lists were refined using the following criteria: (1) the variant had no less than three reads for the alternative allele; (2) the variant detected from the patient brain tissue had a 95% binomial confidence interval of VAF between 0.5% (0.3% in phase 1) and 50%; (3) the variant exhibited a gnomAD allele frequency lower than 0.0025 (not applied in phase 1) . Variants that existed in the 1000 Genome Project (phase 3) were also excluded from the analysis. Variants from both exome data sources were tested, a combination of tissue-specific mosaic variants and tissue-shared mosaic variants were collected, and the credible interval of VAFs was calculated using a Bayesian-based method described previously60. To filter for candidate MCD disease-causing variants called from the BSMN common pipeline or DeepMosaic in phase 2, and variants called in phase 3, we further filtered out synonymous variants in coding regions, variants with CADD Phred score <20, and candidates that fell out of coding regions and were not predicted to affect splicing, annotated by ANNOVAR and BEDtools (v.2.27.1); the annotation scripts are provided on GitHub (https://github.com/shishenyxx/MCD_mosaic). We excluded the mosaic candidate from the list for TASeq validation experiment if (1) the primer pair poorly amplified the target region; (2) there was not enough amount of DNA for TASeq. A total of 554 mosaic candidates were subjected for TASeq validation.
Replication	We identified several recurrently mutated genes found in at least two different patient resected brain samples. We report that there is a low likelihood that this would occur by chance. Several genes were mutated in only a single patient, which will require replication in future cohorts. This was described in the main manuscript and supplementary note. All the experiments relevant to mouse modeling were biologically replicated at least three times in Fig. 3 and Extended Data Fig. 6b, and at least two times in for Extended Data Fig.6a. For the single-nuclei RNAseq, two controls and 5 MCD cases were used for the replication of converging transcriptomic signatures.
Randomization	We did not perform randomization because this is descriptive study showing the genetic landscape of MCDs.
Blinding	All quantification data of staining results using mouse models were generated in a double blinded manner. Blinding was not performed to the other data because data processing was performed by unbiased, automatic bioinformatic pipelines.

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

Antibodies

Antibodies used

<Primary antibodies>
Serial number, Epitope, Dilution ratio, Host, Clonality, Clone number, Supplier, Catalog number
1. Phospho-S6, 1:800, Rabbit, Monoclonal, D68F8, Cell Signaling, 5364S
2. NeuN, 1:100, Mouse, Monoclonal, A60, Sigma-Aldrich, MAB377X
3. GFP, 1:500, Chicken, Polyclonal, NA, Aves Labs, GFP-1020
<Secondary antibodies>
Serial number, Name, Dilution ratio, Host, Clonality, Supplier, Catalog number
1. Alexa Fluor Goat 488 chicken IgY (H+L), 1:1000, Goat, Polyclonal, Invitrogen, A-11039
2. Alexa Fluor 594 donkey anti-rabbit IgG (H+L), 1:1000, Donkey, Polyclonal, Invitrogen, R37119

Validation

1. Phospho-S6 Ribosomal Protein (Ser240/244) (D68F8) XP® Rabbit mAb detects endogenous levels of ribosomal protein S6 only when phosphorylated at Ser240 and Ser244 (Supplier website). Specificity was demonstrated in dopaminergic-neuron-specific Rictor OK mice (PMID:35881440, RRID:AB_10694233).
2. Anti-NeuN AlexaFluor 488 conjugated antibody (MAB377X) reacts with most neuronal cell types throughout the nervous system of mice including cerebellum, cerebral cortex, hippocampus, thalamus, spinal cord and neurons in the peripheral nervous system including dorsal root ganglia, sympathetic chain ganglia and enteric ganglia. The immunohistochemical staining is primarily in the nucleus of the neurons with lighter staining in the cytoplasm (supplier website). Specificity was demonstrated by neuron-specific histone modification signatures in flow cytometry-based sorted NeuN+ compared to NeuN- nuclei in human (PMID:32111906, RRID:AB_2149209).
3. The GFP antibody detects GFP protein expressed in various organisms. Specificity was demonstrated by showing the immunostaining result that GFP antibody only stains GFP-expressing virus injected regions and cell types in mice (PMID:31948733, RRID:AB_10000240).
4. Alexa Fluor Goat 488 chicken IgY (H+L) (RRID:AB_2534096)
5. Alexa Fluor 594 donkey anti-rabbit IgG (H+L) (RRID:AB_2556547).

Animals and other organisms

Policy information about studies involving animals; ARRIVE guidelines recommended for reporting animal research

Laboratory animals

Pregnant Crl:CD1(ICR) mice (embryonic day 14) for mouse modeling were purchased from Charles River Laboratory. Temperature and humidity were kept as ~18-23°C and 40-60% in the vivarium, respectively.
Experimental age for immunofluorescence staining was embryonic day 18 or postnatal day 21.

Wild animals

No wild animals were used in this study

Field-collected samples

No field-collected samples were used in this study

Ethics oversight

All work with mice was performed in accordance with UCSD IACUC protocol S15113.

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

Human research participants

Policy information about studies involving human research participants

Population characteristics

The characteristics of solved patients including gender, genotypic information, details of diagnosis, surgical outcomes were ascertained and summarized in Supplementary table 4.

Recruitment

We recruited a cohort of 283 individuals from the 'FCD Neurogenetics Consortium' (see the member list in the manuscript). These individuals were diagnosed with FCD type I, II, III, HME, or TSC and underwent surgical resection to treat drug-resistant epilepsy between 2013 and 2021. Any cases that underwent surgical resection due to environmental factors, for example, stroke, or acute trauma, were excluded.

Ethics oversight

The study protocol was approved by the UC San Diego IRB (#140028). Informed consent was obtained from all participants or their legal guardians at the time of enrollment.

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

Magnetic resonance imaging

Experimental design

Design type	Standard clinical MRI sequences
Design specifications	NA
Behavioral performance measures	NA

Acquisition

Imaging type(s)	Standard clinical
Field strength	0.5-1.5T
Sequence & imaging parameters	Standard clinical T1 and T2
Area of acquisition	NA

Diffusion MRI ☐ Used ☒ Not used

Preprocessing

Preprocessing software	Standard clinical
Normalization	NA
Normalization template	NA
Noise and artifact removal	NA
Volume censoring	NA

Statistical modeling & inference

Model type and settings	NA
Effect(s) tested	NA

Specify type of analysis: ☐ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	NA
Correction	NA

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis