nature portfolio

Corresponding author(s): Kristen Brennand

Last updated by author(s): 07-24-2026

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	For cue-specific MPRAs, glutamatergic NGN2-neurons were generated from hiPSCs and harvested at Day 24 (n=2 mid-PRS control lines 2607 (XY) and 3182(XX) with 2-3 replicates per exposure per donor). For RNA-seq and ATAC-seq data, glutamatergic NGN2-neurons were generated from hiPSCs and harvested at Day 24 (n=2 mid-PRS control lines 2607(XY) and 3182(XX) with 3-4 replicates per exposure per donor). For ATAC-seq, cells from 4 wells of a 12-well-plate were combined into a single 15mL conical tube for each donor/condition/replicate, cells were resuspended in 310uL of DMEM (Thermofisher #10566-016)-10% FBS, filtered (37um), frozen in DMEM-10% FBS-10% DMSO and submitted to the Yale Genomics Core for library preparation and sequencing. RNA/DNA for MPRA was collected from ~6million cells pooled across 3 well per donor/replicate/condition. Cells were washed three times and harvested using AllPrep DNA/RNA mini kit (Qiagen). For RNA-sequencing 250-500k cells per replicate/condition/donor were collected and RNA harvested using RNA isolation kits (Qiagen) and submitted to the Mount Sinai Genomics Core for sample QC and sequencing.
Data analysis	RNA Sequencing libraries were prepared using the Kapa Total RNA library prep kit from 4 replicates per donor (2 biological replicates) per cue/ vehicle treatment. Paired-end sequencing reads (100bp) were generated on a NovaSeq platform. Raw reads were aligned to hg38 using STAR aligner100 (v2.5.2a) and gene-level expression were quantified by featureCounts101 (v1.6.3) based on Ensemble GRCh38 annotation model. Genes with over 10 counts per million (CPM) in at least four samples were retained. After filtering, the raw read counts were normalized by the voom function in limma and differential expression was computed by the moderated t-test implemented in limma. Differential gene expression analysis was performed between each cue and paired vehicle. Bayes shrinkage (limma::eBayes) estimated modified t- and p- values and identified differentially expressed genes (DEGs) based on an FDR <= 0.05 (limma::TopTable). ATAC sequencing library prep and sequencing were performed by the Yale Sequencing Core. The adaptor sequence for pair-end sequencing was removed using trim_galore and sequencing quality measure by FastQC and MutliQC. Data was aligned with Bowtie against hg38 reference genome including rare SNVs. Mitochondrial reads were removed, sam files sorted and indexed, and converted to compressed BAM files with

samtools and ATAC peaks called using Genrich. Differential peak activity analysis was performed with edgeR and csaw.

Processing of MPRA sequencing data. Barcode-CRE association was performed as previously described using the association utility of MPRAflow v2.3.5 (run as:). We demultiplexed the indexed DNA and RNA libraries and generated fastq files with bcl2fastq v2.20 and used the count utility of MPRAflow 2.3.5We filtered out barcode-CRS pairs with low DNA counts (<15) and those with RNA counts, but no DNA counts, leaving 3,668 CRSs with a minimum requirement of 10 barcodes each. quantification. We assessed the activity of CRS compared to the average activity of all CRS with mpralm() (mpra v.1.16.0) followed by eBayes() abd topTable() (limma v3.50.3) after correcting for donor and replicate (~ 1 + donor + replicate). Sequences that had positive logFC (>0) and a met a statistical significance threshold at FDR < 0.1 were labeled as "active CRS" while those with negative logFC and FDR<0.1 were labeled as "repressed CRS". Only "active CRS" were used in downstream analysis of differential variant effects.

variant-specific differential analysis: We applied mpralm() (mpra v.1.16.0) followed by topTable() (limma v3.50.3) to detect allelic effects in MPRA active CRS (MPRA sequence pairs with one of sequence active in any condition). We set a statistical significance threshold at FDR < 0.1 to define emVars.

cue-by-variant interaction analysis: Cue-by-variant interaction effects were assessed with ~ cue:variant + donor + rep. Only MPRA CRS with significant allelic differences (MPRA-emVars) in at least one condition were used to test for interaction effects. We set a statistical significance threshold at FDR < 0.1 to define interaction emVars.

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

DATA AVAILABILITY

All source donor hiPSCs have been deposited at the Rutgers University Cell and DNA Repository (study 160; http://www.nimhstemcells.org/).

The raw and processed high-throughput sequencing data generated in this study have been deposited on the Gene Expression Omnibus (GEO) database under accession code GSE341428 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE341428). The secondary data and summary statistics generated in this study are provided in the Supplementary Information/Source Data file and are available through Synapse (syn75962204; https://www.synapse.org/Synapse:syn75962204/wiki/). Hi-C data used for annotation in this study are previously published and available through www.synapse.org/#!Synapse:syn12979101 (registration required; Data Download—Study "iPSC-HiC" and through the PsychENCODE Knowledge Portal (https://psychencode.synapse.org/). The PsychENCODE Knowledge Portal is a platform for accessing data, analyses, and tools generated through grants funded by the National Institute of Mental Health (NIMH) PsychENCODE program. Data are available for general research use according to the following requirements for data access and data attribution: (https://psychencode.synapse.org/DataAccess). Track annotations are publicly available from BrainScope (https://brainscope.gersteinlab.org/), and IGV (https://igv.org/app/). GWAS summary statistics are publicly available from the Psychiatric Genomics Consortium (https://pgc.unc.edu/for-researchers/download-results/). Common Mind Consortium eQTL summary statistics can be accessed through a data cces request on the NIMH Data Archive (NDA; https://nda.nih.gov/). GWAS annotations were downloaded from the GWAS catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads).

CODE AVAILABILITY

The full analysis pipeline (including code and processed data objects) used for analysis of RNA-seq, ATAC-seq, and MPRA data evaluation is publicly available through Synapse (syn75962204; https://www.synapse.org/Synapse:syn75962204).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Two human-induced pluripotent stem cell-lines were used in this study - one XX and one XY karyotipic sex. However, this study is not powered to make any claims related to sex differences.
Reporting on race, ethnicity, or other socially relevant groupings	N/A
Population characteristics	hiPSCs (human-induvded pluripotent stem cells) were derived from an XX and XY donor, representing neurotypical controls with no history of psychiatric disorders and who are of European ancestry.
Recruitment	NA. Recruitment of donors was not a part of this study. All lines were selected from a previously reported case/control hiPSC cohort of childhood onset SZ (COS) (Hoffman et al, Nat Comm 2017)
Ethics oversight	Yale University and Mount Sinai Institutional Review Board waived ethical approval for this work. Ethical approval was not required because the hiPSC lines, lacking association with any identifying information and widely accessible from a public repository, are thus not considered to be human subjects research. Post-mortem data are similarly lacking identifiable information and are not considered human subjects research.

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

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	Across all experiments a total of 2 hiPSC donors of mid-SCZ PRS with no family history of psychiatric disorders; one of XX and XY karyotypic sex were used with 2-4 technical replicates per donor/per condition.
Data exclusions	CRS with fewer than 10 barcodes across replicates are excluded from the MPRA analysis.
Replication	For MPRA a library of ~9,000 synthesized sequences representing ~4,500 SNPs (biallelic) and cloned into a lentiMPRA vector. MiSeq of the library and filtering after demultiplexing determined each sequence had a minumun of 10 barcodes and a median of 25 barcodes and per sequence and an average of 50 barcodes per sequence. We performed a power analysis to determine size limitations of our MPRA; with 10000 SNPs, 50 barcodes per SNP, activity standard deviation of 1 (typical range = 0.3-2), and 3 biological replicates, the power of a t-test to detect differential variant shifts of 0.75 or greater (≥ 5 times as much mRNA per input DNA) at a Bonferroni corrected alpha=0.05 level is 100%. For each replicate in each of the two donors (biological replicates) 3 wells of a 6well plate were combined (~6 million cells) to represent a single replicate for a each cue or vehicle treatment. For phenotypic assays, 20 wells were imaged per condition across a minimum of 2 independent cell lines, with 9 images acquired per well for synaptic puncta analysis. For MEA, from 6-12 biological replicates were analyzed per condition. RNA-sequencing analyses were performed with 4 technical replicates per donor (2 donors) per cue/vehicle treatment.
Randomization	cDNA and RNA from each donor/replicate/cue was barcoded and randomly pooled into two custom libraries and sequenced on a single on a NextSeq 2x50 on S2 flow cell (3.3-4.1 B reads/cell). Libraries were demultiplexed after sequencing. For ATAC-sequencing cells from 4 wells of a 12-well-plate were combined into a single pool for each condition.
Blinding	Following stem cell reprogramming, the clinical origin of each sample was blinded to those performing neuronal differentiations, RNA purification, and RNA sequencing in the validation of the cell lines. For the experiment, CRS identity and proportion of barcodes is inherently unknown - technical replicates were pooled and randomly divided across 2 lanes.

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
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

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

Antibodies

Antibodies used	Antibody Species Vendor Catalog # Dilution anti-MAP2 chicken Invitrogen, Abcam PA1-10005, ab5392 1:1000 anti-SYNAPSIN1 mous Synaptic Systems 106-011 1:500 Alexa 568 anti-Mouse Ms Life technologies A10037 1:500 Alexa 647 anti-Chicken Ck Life technologies A10042 1:500
Validation	All antibodies were previously validated (in Brennand et al 2011, Brennand et al 2015, and Ho et al 2017) for immunocytochemistry in human cells. Validation and references for Synapsin1 antibody provided at https://sysy.com/product/106011 Validation and references for MAP2 antibody provided at https://www.abcam.com/map2-antibody-ab5392.html

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	Cultured human fibroblast-reprogrammed hiPSCs Validated control hiPSCs were selected from a previously reported case/control hiPSC cohort of childhood onset SZ (COS) (Hoffman et al, Nat Comm 2017). The following controls were used for this study (hiPSC NPCs NSB2607(XY), NSB3128(XX). Commercial cell lines: HEK293T cells for virus generation: Verma Lab (https://jvi.asm.org/content/73/1/576)
Authentication	Samples were confirmed to key karyotypically normal using the Illumina Core Exome Genotyping Chip (Illumina, 20030770) and cnvPartition 3.2.0 (Illumina, Genome Studio). No cell lines displayed karyotypic abnormalities (no reported CNVs ≧2.5 MB in size). All reported CNVs are shown in each certificate of analysis. hiPSC lines were confirmed to be viable post-thaw, achieving a minimum of 50% confluency within 10 days). Sample identity testing was performed using the SNPTrace assay, confirming correct sample association between parental fibroblast and hiPSC line. Gene expression analysis was using a custom Nanostring panel27 to confirm expression of pluripotency markers such as POU5F1, NANOG, and SOX2, and lack of expression of early differentiation markers such as AFP (Mesoderm), SOX17 (Endoderm), and NR2F2 (Ectoderm). A scorecard panel was used to confirm propensity to differentiate. All hiPSC lines used in this study passed the above QC and have a certificate of analysis.
Mycoplasma contamination	As part of the hiPSC validation process, all samples were tested for the absence of Mycoplasma (Lonza, LT07-710) and confirmed to be sterile (Hardy Diagnostics, K82).
Commonly misidentified lines (See ICLAC register)	n/a

Plants

Seed stocks	*Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*
Novel plant genotypes	*Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*
Authentication	*Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*