nature research

Corresponding author(s): Melissa H Little

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	Bulk RNAseq counts were obtained by mapping using the STAR aligner (V2.4.0h1) and summarising reads over genes using featureCounts (v1.4.6). Counts for single cell RNAseq were obtained by running Cell Ranger (v1.3.1). qRT–PCR analyses were performed with GoTaq qPCR Master Mix (Promega) by an ABI PRISM 7500 96 real-time PCR machine.
Data analysis	The R programming language (R3.4.0) was used to perform the vast majority of the statistical analysis. Within R, the following Bioconductor and CRAN packages were used to perform specific analyses: edgeR v3.17.10, limma v3.31.22, RUVSeq v1.10.0, Mfuzz v2.36.0, org.Hs.eg.db v3.4.1, lme4 v1.1-13, scatterplot3d v0.3-40, NMF v0.20.6, RColorBrewer v1.1-2, cellrangerRkit v1.1.0 and Seurat v2.0.0. The ToppGene suite (https://toppgene.cchmc.org/) was used to further explore gene set enrichment analysis. Image processing was performed using Imaris v8.2 and FIJI v2.0.0. Scripts with the statistical analysis of all RNAseq and single cell RNAseq data are available from the Oshlack github page (https://github.com/Oshlack/OrganoidVarAnalysis). All figures can be reproduced using these scripts.

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 upon request. 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 and analysed during the current study are available in the NCBI Gene Expression Omnibus44, under accession numbers GSE88942, GSE70101, GSE89044, GSE99468, GSE99469, GSE99582, GSE107305, GSE107230, GSE108291, GSE88942 and GSE114802.
Summarised count data for the bulk RNAseq data is available from the Oshlack github page (https://github.com/Oshlack/OrganoidVarAnalysis).

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

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

Sample size	Samples sizes were chosen to allow enough degrees of freedom to estimate variance components in the random effects model and at least two to three replicates at each time point of differentiation. In total 50 whole organoids and 12 epithelial replicates from pooled organoids were sequenced using RNAseq. For single cell sequencing, 8323 single cells from 4 whole organoids were sequenced. The study design is extensively described in the manuscript.
Data exclusions	A single sample from Batch 4 was excluded after library preparation as the library size for this sample was less than a million, with the median library size 24 million for all other CRL1502 clone 32 day 18 samples.
Replication	The objective of this entire manuscript is to examine reproducibility and transferability of a protocol. As such the experiments were repeated multiple times.
Randomization	Not relevant.
Blinding	Blinding was not relevant to the study. All analysis was performed in an unbiased fashion.

Reporting for specific materials, systems and methods

Materials & experimental systems

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

Methods

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

Unique biological materials

Policy information about availability of materials

Obtaining unique materials	The CRL1502 Clone 32 iPSC cell line was derived from an ATCC fibroblast cell line. This is subject to ATCC MTA requirements.

Antibodies

Antibodies used	Mouse anti-CDH1 (E-Cadherin), BD Biosciences, 610181, Clone 36/E-Cadherin, 1:300 Goat anti-GATA3, R&D Systems, AF2605, Polyclonal, 1:300 Sheep anti-NPHS1, R&D Systems, AF4269, Polyclonal, 1:300 Rabbit anti-JAGGED1, Abcam, ab7771, Polyclonal, 1:300
Validation	All antibodies react with human antigen, as described by the manufacturer. All antibodies have been previously validated for immunofluorescence in human iPSC derived organoids (Takasato et al, 2015). CDH1 and GATA3 antibodies have been validated for immunofluorescence in human fetal kidney (Lindstrom et al, 2018). NPHS1 antibody has been validated for immunofluorescence in human adult kidney (Sharma et al, 2016 PMID: 27849017).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	CRL1502 Clone 32 – Wild type iPSC reprogrammed from human skin fibroblasts as detailed in Briggs et al. Integration-free induced pluripotent stem cells model genetic and neural developmental features of down syndrome etiology. Stem Cells. 2013;31(3):467-78 RG_0019.0149.C6 – Wild type human iPSC line reprogrammed from human skin fibroblasts by iPSC Reprogramming Facility, MCRI, Melbourne. "Patient" (RG_0118.0143.UN47) – Patient derived iPSC line reprogrammed from patient fibroblasts by Dr Sara Howden, Kidney Development, Disease and Regeneration Group, MCRI, Melbourne. "Gene Corrected" (RG_0118.0143.GC2) – Gene corrected patient derived iPSC line reprogrammed from patient fibroblasts by Dr Sara Howden, Kidney Development, Disease and Regeneration Group, MCRI, Melbourne.
Authentication	The authentification of the CRL1502 clone 32 cell line has been described previously (Briggs et al, Stem Cells, 2013; Takasato et al, 2015). The RG_0019.0149.C6 cell line was authenticated for pluripotency based on immunocytochemistry for the pluripotency markers OCT4, SOX2, NANOG, SSEA-4, TRA-1-81 and TRA-1-60. Karyotype was also checked with no evidence of anomaly. Two iPSC lines derived from the IFT140(-/-) patient (patient and gene-edited) were authenticated for pluripotency by flow cytometry for pluripotency markers CD9, EpCAM and SSEA4. IFT140 genotype and gene correction were confirmed by Sanger sequencing. Molecular karyotype was normal for both lines.
Mycoplasma contamination	All lines tested negative for Mycoplasma contamination by PCR.
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	A single human iPSC line was derived from a healthy consenting adult female. Ethics approval for the derivation of human induced pluripotent stem cells line RG_0019.0149.C6 was approved as a subproposal of HREC/14/QRBW/34 by The Human Research Ethics Committee of the Royal Brisbane and Women's and Hospital. Two iPSC lines were derived from an 11 year old girl with Mainzer Saldino syndrome secondary to IFT140 mutations. This research was conducted with approval from the human research ethics committees of the Lady Cilento Children's Hospital (HREC/15/QRCH/126), the University of Queensland (MREC Approval 2014000453) and the Royal Brisbane and Women's Hospital (HREC/14/QRBW/34), including research governance approval at all sites. Written informed consent was obtained from the girl's family.
Recruitment	The 11 year old girl was recruited from a renal genetics clinic under the same HREC approvals listed above.