nature research

Corresponding author(s): Shankar Srinivas, Antonio Scialdone

Last updated by author(s): Sep 1, 2021

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 for data collection
Data analysis	We used existing published sequence analysis packages, as detailed in the methods. Including biomaRt R, velocyto v0.17.17, RaceID package v0.1.5, pypairs' v3.1.1, 'transIndel' v0.1, DESeq2 v3.11, Seurat v3.0 and Salmon v0.17.

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

The raw data from our study can be downloaded from ArrayExpress under accession code: E-MTAB-9388. The processed data may be downloaded from www.human-gastrula.net. Datasets used as references include; Mouse gastrula data: E-MTAB-6967; Pre-implantation embryo data: E-MTAB-3929; ENSEMBL database (GRCh38.p13); MSigDB database; Signaling Database CellPhoneDB.

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	The sample size of this study was dictated by the availability of human CS7 embryos. Embryos at this stage of human development are extremely rare thereby significantly impacting sample size.
Data exclusions	Disaggregated cells from the embryo were excluded on the basis of a Live/Dead stain (described in Methods), to select for live cells. Post sequencing quality control (detailed in manuscript) was used to exclude poor quality sequenced cells.
Replication	Due to the rarity and difficulty in collecting human embryos at this stage of development replication was not possible.
Randomization	Given this study characterises a single human embryo randomisation was not necessary.
Blinding	Given this study characterises a single human embryo blinding was not required.

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	Anti-Ecadherin antibody (3195, Cell Signaling Technology, 1:200) and anti-Rabbit IgG Alexa Fluoro 568 antibody (A10042, Invitrogen, 1:1000) were used in this study
Validation	Anti-Ecadherin antibody has been validated by western blot analysis and cited 1686 times (https://www.cellsignal.co.uk/products/primary-antibodies/e-cadherin-24e10-rabbit-mab/3195).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	Human ESCs (H9/WA09 line; WiCell), Vallier Laboratory
Authentication	Validated using fingerprinting
Mycoplasma contamination	Confirmed negative
Commonly misidentified lines (See ICLAC register)	No commonly misidentified cell lines were used in the study

Animals and other organisms

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

Laboratory animals	Wild-type Mus musculus embryos were obtained using C57BL/6 males crossed with 8-16 week old CD1 females.
Wild animals	This study did not use any wild animals.
Field-collected samples	This study did not use any field-collected samples.
Ethics oversight	All animal experiments complied with the UK Animals (Scientific Procedures) Act 1986, approved by the local Biological Services Ethical Review Process and were performed under UK Home Office project licenses PPL 30/3420 and PCB8EF1B4.

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	A single CS7 human embryo.
Recruitment	The sample was collected by the Human Developmental Biology Resource (HDBR - https://www.hdbr.org/general-information). The material was collected after appropriate informed written consent from the donor, by medical termination.
Ethics oversight	HDBR has approval from the UK National Research Ethics Service (London Fulham Research Ethics Committee (18/LO/0822) and the Newcastle and North Tyneside NHS Health Authority Joint Ethics Committee (08/H0906/21+5)) to function as a Research Tissue Bank for registered projects. The HDBR is monitored by The Human Tissue Authority (HTA) for compliance with the Human Tissue Act (HTA; 2004). This work was done as part of project #200295 registered with the HDBR.

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