nature research

Corresponding author(s): Eric D. Siggia, Ali H. Brivanlou

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	Microscopy data was collected on the Zeiss Inverted LSM 780 laser scanning confocal microscope and on the Yokagawa/Olympus CellVoyager spinning disk microscope, equipped with Zeiss Zen 2.3 and CV1000 software, respectively. BD FACSDiva II was used for FACS sorting.
Data analysis	Data was analyzed using a combination of codes: Fiji (ver2.0.0), Ilastik (1.2.0) and in-house watershed cell segmentation software (available upon request) for cell segmentation, and Python 3 for data handling and statistical analyses.

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

Numerical source data for all figures have been provided as Supplementary Table 2. All other data supporting the findings of this study are available from the corresponding author on reasonable request.

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	The population fractions (depicted in pie charts throughout the article) for a given experimental condition or cell line were done by imaging a number of epithelial colonies without bias then counting the populations with a unique marker combination. In addition, separate experiments were run to collect more data at higher resolution for quantification, by only imaging the colonies with a given marker combination, and these were not included in population statistics in the pie charts. The n, reported in each figure caption and throughout the text, related to microscopy imaging, represents the number of individual 3D colonies that were imaged for that combination of immunofluorescence markers or, in case of live-cell imaging, for that particular transgenic cell line. The main result was repeated at least 7 times independently over the course of at least 18 months using immunofluorescence staining and in at least 3 independent experiments with live-cell imaging. The n, relating to filter experiments, represent the number of individually run experiments on separate filter-grown tissues.
Data exclusions	There was no exclusion of numerical data based on outliers per se. As described in the text, the patterned 3D colonies were categorized based on the combination of expressing markers and each category separately analyzed, in both IF and live-cell imaging. After establishing the symmetry-breaking phenotype using BRA/SOX2 dichotomy, further derivative markers for better characterization of the phenomenon (e.g., N-CAD, SNAIL, COLL IV, etc.) were carried out by only imaging the symmetry-broken structures based on criteria explained in the text. In live-cell imaging, few examples were discarded either due to very poor fluorescence signal that prevented categorization (usually if the colony is suspended far above the dish surface or moves up), due to cell death, presence of too many loose cells unrelated to the colony, or adhesion to the dish surface. Any nonspecific cell clusters such as those that failed to form an epithelium as well as large agglomerates that occur in 3D cultures in general (organoids or otherwise) were not imaged. In live-cell imaging, as described in the text, the data analysis was carried out until the onset of EMT as the separation of cells from the colony obfuscates the symmetry-breaking metrics.
Replication	The self-organization of hESCs into a 3D human epiblast model and the BMP4-concentration-dependent patterning was confirmed in multiple experiments as described. The key result was replicated in 7 independent experiments probing the BRA/SOX2 asymmetry over the course of at least 20 months and in 3 independent experiments using live-cell imaging. Symmetry breaking was further observed with many other experiments probing other signatures of the primitive streak (as described in the text). The main result in 3D was further replicated in an equivalent 2D system, as described in the text.
Randomization	There were no experimental groups that could be randomized.
Blinding	There were no animal or human subjects in the experiments, and the experimenters had no control over the outcome of cell biology experiments, therefore no blinding was done.

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

Antibodies

Antibodies used

We list the antibodies and dilutions for the following antigens: BRACHYURY (R&D Systems AF-2085; 1:300), SOX2 (Cell Signaling 3579; 1:200), OCT4 (BD Biosciences 611203; 1:200), NANOG (R&D Systems AF-1997, 1:150), SNAIL (R&D Systems AF-3639; 1:150), COLLAGEN IV (Abcam Ab-6586; 1:200), EZRIN (Sigma-Aldrich E8897; 1:200), GATA3 (ThermoFisher, MA1-028; 1:100), GATA3 (Abcam, ab124288, 1:200), GATA6 (Cell Signaling 5851S; 1:200), E-CADHERIN (Cell Signalling 3195; 1:200), N-CADHERIN (BioLegend 350802; 1:200), ZO-1 (Invitrogen 617300; 1:200), SOX17 (Abcam, ab84990, 1:200), pSMAD1 (Cell Signaling, 9516S, 1:200), donkey anti-goat Alexa Fluor 488 (Thermo-Fisher A-11055, 1:500), donkey anti-rabbit Alexa Fluor 555 (Thermo-Fisher A-31572, 1:500), donkey anti-mouse Alexa Fluor 647 (Thermo-Fisher A-31571, 1:500).

Validation

All antibodies used have been tested and validated by the supplier and also by other researchers in the field. The dilution of each antibody was used based on a thorough prior investigation both in our labs and in other labs.

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)

The RUES2 and RUES1 cell lines were previously created in our lab and are listed in the NIH Human Embryonic Stem Cell Registry. They were derived under the approval from the Tri-Institutional Stem Cell Initiative Embryonic Stem Cell Research Oversight (Tri-SCI ESCRO) Committee, an independent committee charged with oversight of research with human pluripotent stem cells and embryos to ensure conformance with University policies, and guidelines from the U.S. National Academy of Sciences (NAS)

Authentication

The cell lines and their derivatives were thoroughly authenticated by sequencing, karyotyping, and by verifying that they remain pluripotent in extended culture and that they can give rise to all three germ layers under controlled differentiation conditions.

Mycoplasma contamination

The lines tested negative for mycoplasma contamination.

Commonly misidentified lines
(See ICLAC register)

No commonly misidentified cell lines were used.

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☒ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation

hESCs epithelia differentiated on filters were dissociated with Accutase and resuspended in MEF-CM supplemented with Rock Inhibitor Y27632 . The single-cell suspension was run through a cell strainer to remove eventual cell clumps.

Instrument

BD FACS Aria II

Software

BD FACSDiva

Cell population abundance

Live-to-dead gating was determined based on forward scattering. On average, 70% of the cells were gated as live.

Gating strategy

To determine SOX2-high and SOX2-low populations, gating of the SOX2-high population was determined in the undifferentiated sample. This led, in the differentiated sample, to 67% of cells in the SOX2-high gate and 25.5% of the cells in the SOX2-low gate. Please, see Figure S7C.

☒ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.