nature research

Corresponding author(s): Magdalena Zernicka-Goetz

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	The basic command used for data preparation. Tophat -p 16 -g 1 -N 2 --microexon- search --no-novel-juncs -G refGene.gtf -o out/ genome R1.fastq R2.fastq Cufflinks -p 16 -o out/ -g refGene.gtf accepted_hits.bam

Data analysis

- Immunofluorescence and time-lapse images: analyzed with Fiji (version 1.0) or Imaris (Bitplane).
- Statistical analyses (excluding the deep-sequencing data): analyzed with GraphPad Prism (version 6) and Microsoft Excel
- Analysis of deep-sequecing data, analysed with;
R (version: 3.4.0)
Python (version:2.7.13)
Matlab (version: 2015a)
Tophat (version: 2.0.4)
Cufflinks (version: 2.0.4)
DAVID (Database for Annotation, Visualisation and Integrated Discovery, version:6.8)
MEGA (version:7.0)
Cluster 3.0

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

RNA–seq data that support the findings of this study have been deposited in the Gene Expression Omnibus (GEO) under accession code GSE110105. LCM sequencing data is available under accession number GSE65924. Source Data for RT–qPCR experiments (Fig. 4c, Supplementary Fig. 2i) and quantifications of the immunofluorescence data (Fig. 1f, Fig. 2a-b-c, Fig. 3e, Supplementary Fig. 2d-f-g, Supplementary Fig. 4c, Supplementary Fig. 5d, h) and the differentially expressed gene list (Fig. 7b) have been provided in Supplementary Table 3. 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

No statistical method was performed to predetermine sample size.

Sample size for each experiment described in the paper based on our previous experience and the work of other groups using embryos and stem cells as experimental model systems.

Data exclusions

Culture experiments: on principle, data were only excluded for failed experiments, reasons for which included suboptimal culture conditions.

ETX embryo samples for deep-sequencing analysis: those samples that did not pass the quality controls were excluded from the analysis The quality of the reads was evaluated using the FASTQC tool. Density distribution of gene expression for all samples was also plotted to assess sample consistency. Raw reads were mapped to mm10 version of the mouse genome using the Tophat2 v2.0.4 program. Samples with raw sequencing reads count > 10 million and mapping ratio > 50% were retained for further analysis. We calculated fragment per kilobase per million (FPKM) as expression level using Cufflinks v2.0.2 with default parameters. Genes with the FPKM value > 1.0 in at least one sample across all samples were retained for further analysis. Finally, the expression levels were transformed to logarithmic space by using log2 (FPKM +1). With these criteria, 3 gastrulating ETX embryos were retained for further analysis (GE-A, GE-B and GE-C) and since they were cut in half, each side was arbitrarily called side 1 or side 2.

Replication

In the study, all attempts at replication were successful.
Each result described in the paper is based on at least two independent biological replicates but very often an experiment is based on more

than two experiments. Figure legends indicate the number of independent experiments performed in each analysis.

Randomization

For experiments were our samples were exposed to chemical inhibitors, samples were randomly allocated to control and experimental groups.

Blinding

Investigators were not blinded.
This study compares ETX embryos to natural embryos. The experiments were descriptive in nature.

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

The following antibodies were used:
Oct 3/4 (mouse), C-10, Santa cruz sc-5279 1:200, Lot:E1017
Tbr2/Eomes (rabbit) abcam ab23345 1:400 Lot:GR305614-2
Brachyury/T (goat) Santa cruz sc-17745, C-19, 1:500 (discontinued)
Brachyury/T (goat) R&D systems AF2085 1:500, Lot:KQP0617031
Podocalyxin (rat) Clone 192703 R&D systems MAB1556 1:400, Lot:IPF0317051
GFP (rat) Clone GF090R Nacalai biochemicals 04404-84 ,1:2000, Lot:M8E4658
Tfap2c (rabbit) Santa cruz, H-77, sc-8977, 1:200 (discontinued)
Laminin (rabbit) Sigma L9393 1:400, Lot:046M4837V
Collagen IV (rabbit), abcam ab19808, 1:200, Lot:GR261208-14
E-cadherin (rat) Clone ECCD-2 Life Technologies (Thermofisher scientific) 13-1900, 1:400, Lot:74960345A
N-cadherin (mouse) BD Biosciences, 610920, 1:400, Lot:7124538
GM130 (mouse) Clone 35/GM130 BD Biosciences 610822 1:500
Anti-phospho-Histone H3 (Ser10), Millipore, 06-570, 1/500, Lot:29480
Gata6 (Goat) R&D Systems, AF1700, 1:200, Lot:KWT0316081
Gata4 (Goat) Santa cruz sc-1237, C20, 1:200, Lot:J2015
Foxa2 (Rabbit) Cell signalling technologies, D56D6, 8186 1:200, Lot:3
Sox17 R&D Systems, AF1924, 1:200, Lot:KGA091621
Otx2 (Goat) R&D Systems, AF1979, 1:200, Lot:KNO0615111 and Lot:KNO0616081
Lefty1 (goat) R&D Systems, AF746, 1:2000, Lot: CMM021805A
F-actin (Phalloidin 488) Life Technologies (Thermofisher scientific), A12379, 1:1000, Lot:1896875
Alexa 488 (Donkey anti-rat) Life Technologies (Thermofisher scientific), A21208, 1:500, Lot:1900239
Alexa 568 (Donkey anti-mouse) Life Technologies (Thermofisher scientific), A10037, 1:500, Lot:1917938
Alexa 647 (Donkey anti-rabbit) Life Technologies (Thermofisher scientific), A31573, 1:500, Lot:190516
Alexa 647 (Donkey anti-goat) Life Technologies (Thermofisher scientific), A21447, 1:500, Lot:1917928

This information is also included in Supplementary Table 1.

Validation

The subcellular localization of all the proteins analyzed in this study has been previously reported in mouse embryo/stem cell studies (shown as immunoflourescence labelling).

Oct4: it specifically stained the epiblast at all stages tested, as expected (Science 356, doi:10.1126/science.aal1810).
Eomes: it correctly stained Exe and embryonic VE as reported and as expected (Science 356, doi:10.1126/science.aal1810).
Bry: it correctly stained mesoderm at 6.5 and later as reported and expected (Dev Biol 288, 363-371).
Podocalyxin: correclty stained polarized lumen in the embryo as reported elsewhere (Nature 552, 239-243).
Tfap2c: it correctly stained the ExE lineage in postimplantation embryos and PGC after gastrulation as reported and expected (Science 356, doi:10.1126/science.aal1810).
Laminin: it correctly stained the basement membrane between visceral endoderm and Exe or epiblast, as reported elsewhere and as expected (Dev Dyn 241, 270-283)

Collagen: same as above (Dev Dyn 241, 270-283)
E-cadherin: it correctly stained the basolateral side of cells in the embryo as reported and as expected (Science 356, doi:10.1126/science.aal1810)
Ncadherin: it correctly displayed upregulation on the side of the epiblast undergoing epithelial to mesenchymal transition as reported and as expected (eLife 2018;7:e32839)
Gm130: it correctly displayed the punctate Golgi morphology (Nature 552, 239-243).
Anti phospho histone 3 it correctly labelled mitotic cells, as expected (Science, 360(6384):99-102 2018)
Gata6: it correctly labelled the visceral endoderm in postimplantation embryos (eLife 2018;7:e32839)
Gata4: same as above (Science 356, doi:10.1126/science.aal1810)
Foxa2: it was correctly expressed in the visceral endoderm and in the gastrulation stage mesoderm (Development 136, 1029-1038)
Sox17: it was correctly expressed in the visceral endoderm (eLife 2018;7:e32839)
Otx2: it was correctly expressed in the epiblast and embryonic visceral endoderm of post-implantation embryos (eLife 2018;7:e32839)
Lefty1: it was correctly expressed specifically on the anterior side of the embryonic visceral endoderm of developing postimplantation embryos starting at E5.5 and beyond (Dev. Biol. 402 (2015) 175-191).
F-actin: it correctly stained the cell membrane and was apically enriched, as expected and reported it correctly stain (Development, 138(2011) 3011-3020).

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	The following cell lines were used: Mouse embryonic stem cells: E14 WT mESCs (kindly provided by Prof. Austin Smith, Stem Cell Institute, Cambridge, UK), Nodal HBE-YFP ESCs and T:GFP ESCs kindly provided by Dr. Alfonso Martinez Arias, Cambridge, UK) Mouse trophoblast stem cells: wild-type TSCs (a gift from Jenny Nichols, University of Cambridge, UK) Mouse extraembryonic endoderm cells: wild-type XENs (a gift from Ellen Na, University of Charité, Germany) and EGFP XENs (a gift from Dr. Peter Rugg-Gunn Babraham Institute, Cambridge).
Authentication	Cells were maintained in conditions to preserve stem cell character and prevent differentiation. Plates were inspected for morphological evidence of differentiation (altered colony morphology in ESC cultures or presence of trophoblast giant cells in TSC cultures..etc) and plates with differentiated cells were discarded. Furthermore, cell identities were confirmed routinely by immunoflourescence marker expressions.
Mycoplasma contamination	Cell lines were routinely tested for mycoplasma contamination by PCR and confirmed that they were negative for mycoplasma contamination.
Commonly misidentified lines (See ICLAC register)	The cells we used are not part of this database

Animals and other organisms

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

Laboratory animals	6-week old mice (Mus musculus) were used to obtain mouse embryos for this study. The following strains and genetically-modified models were used: F1 (C57Bl6xCBA), MF1, CD1, and T:GFP (both males and females).
Wild animals	no wild animals were used in the study.
Field-collected samples	no field-collected samples were used in the study.