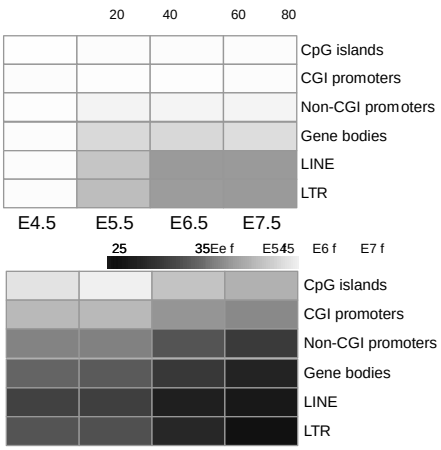
Article

Multi-omics profiling of mouse gastrulation at single-cell resolution

Ricard Argelaguet [1,17], Stephen J. Clark [2,17]*, Hisham Mohammed [2,17], L. Carine Stapel [2,17], Christel Krueger [2], Chantriolnt-Andreas Kapourani [3,4], Ivan Imaz-Rosshandler [5,6], Tim Lohoff [2,5], Yunlong Xiang [7,8], Courtney W. Hanna [2,9], Sebastien Smallwood [2], Ximena Ibarra-Soria [10], Florian Buettner [11], Guido Sanguinetti [3], Wei Xie [7,8], Felix Krueger [12], Berthold Göttgens [5,6], Peter J. Rugg-Gunn [2,5,6,9], Gavin Kelsey [2,9], Wendy Dean [13], Jennifer Nichols [5], Oliver Stegle [1,14,15]*, John C. Marioni [1,10,16]* & Wolf Reik [2,9,16]*

Formation of the three primary germ layers during gastrulation is an essential step in the establishment of the vertebrate body plan and is associated with major transcriptional changes [1–5]. Global epigenetic reprogramming accompanies these changes[6–8], but the role of the epigenome in regulating early cell-fate choice remains unresolved, and the coordination between di erent molecular layers is unclear. Here we describe a single-cell multi-omics map of chromatin accessibility, DNA methylation and RNA expression during the onset of gastrulation in mouse embryos. The initial exit from pluripotency coincides with the establishment of a global repressive epigenetic landscape, followed by the emergence of lineage-speci c epigenetic patterns during gastrulation. Notably, cells committed to mesoderm and



5,000 genes tested, we identified 125 genes the expression of which shows significant correlation with promoter DNA methylation and 52 with expression significantly correlated with chromatin accessibility (Fig. 1g, Extended Data Fig. 4, Supplementary Tables 1, 2). These loci largely comprise markers of early pluripotency and germ cells, such as Dppa4, Zfp42, Tex19.1 and Pou3f1 (Fig. 1g, Extended Data Fig. 4), which are repressed, coinciding with the global increase in methylation and decrease in accessibility. In addition, this analysis identified genes, including Trap1a and Zfp981, that may have unknown roles in development. Notably, of the genes that are upregulated after E4.5, only 39 and 9 show a significant correlation between RNA expression and promoter methylation or accessibility, respectively (Extended Data Fig. 4). This suggests that the upregulation of these genes is probably controlled by other regulatory elements.

Characterizing germ-layer epigenomes

To understand the relationships between all three molecular layers during germ-layer commitment we next applied multi-omics factor analysis (MOFA)[15] to cells collected at E7.5. MOFA performs unsupervised dimensionality reduction simultaneously across multiple data modalities, thereby capturing the global sources of cell-to-cell variability via a small number of inferred factors. Notably, the model makes use of multimodal measurements from the same cells, thereby detecting coordinated changes between the different data modalities.

As input to the model we used RNA-sequencing (RNA-seq) data across protein-coding genes and DNA methylation and chromatin accessibility data across putative regulatory elements. This includes promoters and germ-layer-specific chromatin immunoprecipitation with DNA sequencing (ChIP–seq) peaks for distal H3K27ac (enhancers) and H3K4me3 (transcription start sites) [16] (Extended Data Fig. 5). MOFA identified six factors, with the top two (sorted by variance explained) capturing the emergence of the three germ layers (Fig. 2a, b). Notably, MOFA links the variation at the gene-expression level to concerted methylation and accessibility changes at lineage-specific enhancer marks (Fig. 2c). By contrast, epigenetic changes at promoters or at H3K4me3-marked regions show much weaker associations with germ-layer formation (Fig. 2a, Extended Data Fig. 6, Supplementary Tables 3, 4). This supports other studies that have identified distal enhancers as lineage-driving regulatory regions [8,17–19]. Inspection of gene–enhancer associations identified enhancers linked to key germ-layer markers including Lefty2 and Mesp2 (mesoderm), Foxa2 and Bmp2 (endoderm), and Bcl11a and Sp8 (ectoderm) (Fig. 2c, Extended Data Fig. 7). Notably, ectoderm-specific enhancers display fewer associations than their mesoderm and endoderm counterparts, a finding that is explored further below.

The four remaining factors correspond to additional transcriptional and epigenetic signatures related to anterior–posterior axial patterning (factor 3), notochord formation (factor 4), mesoderm patterning (factor 5) and cell cycle (factor 6) (Extended Data Fig. 8).

Finally, we sought to identify transcription factors that could drive or respond to epigenetic changes in germ-layer commitment. Integrating differential-expression information with motif enrichment at differentially accessible loci revealed that lineage-specific enhancers were

endoderm enhancers. In both cases, changes in methylation and accessibility co-occur, suggesting tight co-regulation of the two epigenetic layers.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-019-1825-8 .

1. Peng, G. et al. Spatial transcriptome for the molecular annotation of lineage fates and cell identity in mid-gastrula mouse embryo. Dev. Cell 36, 681…697 (2016).
2. Mohammed, H. et al. Single-cell landscape of transcriptional heterogeneity and cell fate decisions during mouse early gastrulation. Cell Rep. 20 , 1215…1228 (2017).
3. Wen, J. et al. Single-cell analysis reveals lineage segregation in early post-implantation mouse embryos. J. Biol. Chem. 292, 9840…9854 (2017).
4. Pijuan-Sala, B. et al. A single-cell molecular map of mouse gastrulation and early organogenesis. Nature 566 , 490…495 (2019).
5. Chan, M. M. et al. Molecular recording of mammalian embryogenesis. Nature 570 , 77…82 (2019).
6. Auclair, G., Guibert, S., Bender, A. & Weber, M. Ontogeny of CpG island methylation and speci icity of DNMT3 methyltransferases during embryonic development in the mouse. Genome Biol. 15, 545 (2014).
7. Lee, H. J., Hore, T. A. & Reik, W. Reprogramming the methylome: erasing memory and creating diversity. Cell Stem Cell 14, 710…719 (2014).
8. Zhang, Y. et al. Dynamic epigenomic landscapes during early lineage speci ication in mouse embryos. Nat. Genet . 50 , 96…105 (2018).
9. Macaulay, I. C. et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat. Methods 12, 519…522 (2015).
10. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. Nat. Biotechnol . 33 , 285…289 (2015).
11. Angermueller, C. et al. Parallel single-cell sequencing links transcriptional and epigenetic heterogeneity. Nat. Methods 13, 229…232 (2016).
12. Clark, S. J. et al. scNMT-seq enables joint pro iling of chromatin accessibility DNA methylation and transcription in single cells. Nat. Commun . 9, 781 (2018).
13. Cao, J. et al. Joint pro iling of chromatin accessibility and gene expression in thousands of single cells. Science 361, 1380…1385 (2018).
14. Smith, Z. D. et al. A unique regulatory phase of DNA methylation in the early mammalian embryo. Nature 484 , 339…344 (2012).

Methods

No statistical methods were used to predetermine sample size. The

Samples were similar overall regarding total mapped read numbers, distribution of reads and ChIP enrichment.

To best represent the underlying ChIP–seq signal, different methods to define enriched genomic regions were used for H3K4me3 and H3K27ac marks. For H3K4me3, a SeqMonk implementation of MACS[42] with the local rescoring step omitted was used ($P < 10^{-15}$, fragment size 300 bp), and enriched regions closer than 100 bp were merged. Peaks were called separately for each lineage. For H3K27ac, reads were quantitated per 500-bp tiles correcting per million total reads and excluding duplicate reads. Smoothing subtraction quantification was used to identify local maxima (value >1), and peaks closer than 500 bp apart were merged. Lineage-specific peak annotations exclude peaks that are also present in one of the other lineages, and only peaks present in al15.2 (ot.7 (os c, ent in)]p02 Tc 5) 0 Td676both reldi63 eaces]TJ 0 ic

pair of factors across the ten trials. All inferred factors were consistently found in all model instances.

The downstream characterization of the model output included several analyses. (1) Variance decomposition: quantification of the fraction of variance explained (R^2) by each factor in each view, using a coefficient of determination[15]. (2) Visualization of weights/loadings: the model learns a weight for every feature in each factor, which can be interpreted as a measure of feature importance. Features with large weights (in absolute value) are highly correlated with the factor values. (3) Visualization of factors: each MOFA factor captures a different dimension of cellular heterogeneity. All together, they define a latent space that maximizes the variance explained in the data (under some important sparsity assumptions[15]). The cells can be visualized in the latent space by plotting scatter plots of combinations of factors. (4) Gene set enrichment analysis: when inspecting the weights for a given factor, multiple features can be combined into a gene set-based annotation. For a given gene set G, we evaluate its significance via a parametric *t*-test (two-sided), whereby we compare the mean of the weights of the foreground set (features that belong to the set G) with the mean of the weights in the background set (features that do not belong to the set G). Resulting *P* values are adjusted for multiple testing using the Benjamini–Hochberg procedure from which significant pathways are called (FDR <10%).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Raw sequencing data together with processed files (RNA counts, CpG methylation reports, GpC accessibility reports) are available in the Gene Expression Omnibus under accession number GSE121708 Processed data can be downloaded from ftp://ftp.ebi.ac.uk/pub/databases/scnmt_gastrulation.

Code availability

All code used for analysis is available at https://github.com/rargelaguet/scnmt_gastrulation.

31. Hu, [illegible] Tet and TDG mediate DNA [illegible]methylation essential for mesenchymal-to-epit[illegible]sition in somatic cell rep[illegible]mming. Cell Stem Cell 14, 512–522 (2014).
32. Mac[illegible] et al. Separation and [illegible]el sequencing of the genomes and tran[illegible]f single cells using G[illegible]q. Nat. Protoc. 11, 2081–2103 (2016).
33. Pice[illegible] et a[illegible]length RNA-seq f[illegible]ngle cells using [illegible]rotoc. [illegible] 171[illegible] (2014).
34. Cla[illegible]J. et al. [illegible]ne-wide base-[illegible]ution mappi[illegible]NA methylation [illegible]gle cell[illegible]g single-c[illegible]ul ite sequen[illegible]scBS-seq). Nat. Protoc. 12, 53[illegible]7 (2017).
35. Kim[illegible]angmead, [illegible]alzberg, S. [illegible]SAT: a fast spliced aligner with low [illegible]ory requ[illegible]ents. Nat. Me[illegible] 12, 357[illegible] (2015).
36. Liac[illegible]Smyth, G. K. & S[illegible]featur[illegible]nts: an ef i[illegible]neral purpose p[illegible]n for ass[illegible]g sequence reads [illegible]omic[illegible]ures. Bioin[illegible]tics 30, 923–9[illegible]14).
37. Yate[illegible]et al. Ensembl 2016[illegible]eic[illegible]s Res. 4[illegible], D710–D716 (20[illegible]
38. Lun[illegible] L., McCarthy, D. J. & [illegible]. A step-b[illegible]work low for l[illegible]nalysis of s[illegible]cell RNA-seq data with [illegible]tor. F1000[illegible]5, 212[illegible]16).
39. Kru[illegible] [illegible] 1a . HISw ma[illegible]gle(-lquencin[illegible]6 Td [5[illegible]atics

Extended Data Fig. 1 | scNMT-seq quality controls. a, b, Number of observed cytosines in CpG (red; a) or GpC (blue; b) contexts respectivel

Extended Data Fig. 2 | See next page for caption.

Extended Data Fig. 3 | See next page for caption.

Extended Data Fig. 3 | Global methylation and chromatin accessibility dynamics. a, b, Distribution of DNA methylation (a) and chromatin accessibility levels (b) per stage and genomic context. When aggregating over genomic features, CpG methylation and GpC accessibility levels (%) are computed assuming a binomial model, with the number of trials being the total number of observed CpG (or GpC) sites and the number of successes being the number of methylated CpG (or GpC) sites (Methods). Notably, this implies that DNA methylation and chromatin accessibility are quantified as a percentage and are not binarized into low or high states. As this figure shows, the distribution of DNA methylation and chromatin accessibility across loci (after aggregating measurements across all cells per stage) is largely continuous and does not show bimodality. Hence, a binary approach similar to that sometimes used for differentiated cell types would not provide a good representation of the data. c, d, Box plots showing the distribution of genome-wide CpG methylation levels (c) or GpC accessibility levels (d) per stage and lineage. Each dot represents a single cell. Box plots show median levels and the first and third quartile, whiskers show 1.5× the interquartile range. At a significance threshold of 0.01 (t-test, two-sided), the global DNA methylation levels differ between embryonic and extra-embryonic lineages, but the global chromatin accessibility levels do not. e , f, Dimensionality reduction of DNA methylation (e) and chromatin accessibility (f) data. To perform dimensionality reduction while handling the large amount of missing values, we used a Bayesian factor analysis model (Methods). Scatter plots of the first two latent factors (sorted by variance explained) for models trained with cells from the indicated stages are shown. From E4.5 to E6.5, cells are coloured by embryolmnd extra-mbrls dif rlaasfyoa

Extended Data Fig. 5 | See next page for caption.

Extended Data Fig. 5 | Characterization of lineage-specific H3K27ac and H3K4me3 ChIP–seq data. a, Percentage of peaks overlapping promoters (±500 bp of TSS of annotated mRNAs (Ensembl v.87); lighter colour) and not overlapping promoters (distal peaks, darker colour). H3K27ac peaks tend to be distal from the promoters, marking putative enhancer elements [53]. H3K4me3 peaks tend to overlap promoter regions, marking TSS [54]. b, Venn diagrams showing overlap of peaks for each lineage, for distal H3K27ac (left) and H3K4me3 (right). This shows that H3K27ac peaks tend to be lineage-specific, whereas H3K4me3 peaks tend to be shared between lineages. c, Illustrative example of the ChIP–seq profile for the ectoderm marker Cxcl12 . The top tracks show wiggle plots of ChIP–seq read density (normalized by total read count) for lineage-specific H3K27ac and H3K4me3. The coding sequence is shown in black. The bottom tracks show the lineage-specific peak calls (Methods). H3K27ac peaks are split into distal (putative enhancers) and proximal to the promoter. d, Left, bar plot of the fraction of E7.5 lineage-specific enhancers (n = 691 for ectoderm, 618 for endoderm and 340 for mesoderm) that are uniquely marked by H3K27ac in either E10.5 midbrain, E12.5 gut or E10.5 heart. Right, heat map displaying H3K27ac levels at individual lineage-specific enhancers (n = 2,039 for ectoderm, 1,124 for endoderm and 631 for mesoderm) in more differentiated tissues. E7.5 enhancers are predominantly marked in their differentiated-tissue counterparts (midbrain for ectoderm, gut for endoderm and heart for mesoderm).

Extended Data Fig. 6 | Differential DNA methylation and chromatin accessibility analysis at E7.5 for different genomic contexts. a, Bar plots showing the fraction (left) or the total number (right) of differentially methylated (red) or accessible (blue) loci (FDR <10%, y axis) per genomic context (x axis). Each subplot corresponds to the comparison of one cell type (group A) against cells comprising the other cell types present at E7.5 (group B). In the graphs on the right, positive values indicate an increase in DNA methylation or chromatin accessibility in group A, whereas negative values indicate a decrease in DNA methylation or chromatin accessibility. Differential analysis of DNA methylation and chromatin accessibility was performed independently for each genomic element using a two-sided Fisher's exact test of equal proportions (Methods). b , Scatter plots showing differential DNA methylation (x axis) versus chromatin accessibility (y axis) analysis at promoters. Ectoderm versus non-ectoderm cells (left), endoderm versus non-endoderm cells (middle) and mesoderm versus non-mesoderm cells (right) are shown. Each dot corresponds to a gene (n = 2,038). Labelled black dots highlight genes with lineage-specific RNA expression that show significant differential methylation or accessibility in their promoters (FDR <10%).

Extended Data Fig. 7 | Illustrative examples of putative epigenetic regulation in enhancer elements during germ-layer commitment. a–c, Box and violin plots showing the distribution of RNA expression (log

Extended Data Fig. 8 | See next page for caption.

Extended Data Fig. 8 | Characterization of MOFA factors. a, Factor 1 as mesoderm commitment factor. Left, RNA-expression loadings for factor 1. Genes with large positive loadings increase expression in the positive factor values (mesoderm cells). Middle, scatter plot of factor 1 (x axis) and factor 2 (y axis) values. Each dot corresponds to a single cell, coloured by the average methylation levels of the top 100 enhancers with highest loading. Right, as the middle panel, except cells are coloured by the average accessibility levels. b , Factor 2 as the endoderm commitment factor. Left, RNA-expression loadings for factor 2. Genes with large positive loadings increase expression in the positive factor values (endoderm cells). Middle, scatter plot of factor 1 (x axis) and factor 2 (y axis) values. Each dot corresponds to a single cell, coloured by the average methylation levels (%) of the top 100 enhancers with highest loading. Right, as the middle panel, but cells are coloured by the average accessibility levels. c, Characterization of MOFA factor 3 as anteroposterior axial patterning and mesoderm maturation. Left, bee swarm plot of factor 3 values, grouped and coloured by cell type. The mesoderm cells are subclassified into nascent and mature mesoderm (Extended Data Fig. 2). Right, gene set enrichment analysis of the gene loadings of factor 3. The top most significant pathways from MSigDB C2 [55] (Methods) are shown. d , Characterization of MOFA Factor 6 as cell cycle. Left, bee swarm plot of factor 6 values, grouped by cell type and coloctw57 (d by c)-16.3 ()-16.3 (g)-9.9 (t)6.1 (r)-22.1 (m (09 (

Extended Data Fig. 10 | E7.5 ectoderm enhancers contain a mixture of pluripotency and neural signatures with different epigenetic dynamics. a, Scatter plot showing H3K27ac levels for individual ectoderm enhancers (n = 2,039) quantified in serum-grown ES cells (pluripotency enhancers, x axis) versus E10.5 midbrain (neuroectoderm enhancers, y axis). H3K27ac levels in the two lineages are negatively correlated (Pearson's R = −0.44), indicating that most enhancers are either marked in ES cells or in the brain. The top 250 enhancers that show the strongest differential H3K27ac levels between midbrain and ES cells (blue for midbrain-specific enhancers and grey for ES cell-specific enhancers) are highlighted. b , Density plots of H3K27ac levels in ES cells versus E10.5 midbrain. H3K27ac levels are negatively correlated at E7.5 ectoderm enhancers, but not in E7.5 endoderm (n = 1,124) or mesoderm enhancers (n = 631).c, Profiles of DNA methylation (red) and chromatin accessibility (blue) along the epiblast–ectoderm trajectory. Panels show different genomic contexts: E7.5 ectoderm enhancers that are specifically marked by H3K27ac in the midbrain (middle) or ES cells (bottom) (highlighted populations in a). Running averages of 50-bp windows around the centre of the ChIP–seq peaks (2 kb upstream and downstream) are shown. Solid lines display the mean across cells (within a given lineage) and shading displays the s.d. Dashed horizontal lines represent genome-wide background levels for DNA methylation (red) and chromatin accessn

Extended Data Fig. 11 | See next page for caption.

Extended Data Fig. 11 | Silencing of ectoderm enhancers precedes activation of mesoderm and endoderm enhancers. a, Reconstructed mesoderm (top) and endoderm (bottom) commitment trajectories using a diffusion pseudotime method applied to the RNA-expression data (Methods). Scatter plots of the first two diffusion components are shown, with cells coloured according to their lineage assignment (n = 1,154 for endoderm and n = 1,511 for mesoderm). For both cases, ranks along the first diffusion component are selected to order cells according to their differentiation state. b , DNA methylation (red) and chromatin accessibility (blue) dynamics of lineage-defining enhancers along the mesoderm (top) and endoderm (bottom) trajectories. Each dot denotes a single cell (n = 387 for endoderm and n = 474 for mesoderm) and black curves represent non-parametric locally estimated scatterplot smoothing regression estimates. In addition, for each scenario we fit a piecewise linear regression model for epiblast, primitive streak and mesoderm or endoderm cells (vertical lines indicate the discretized lineage transitions). For each model fit, the slope (r) and its significance level are displayed in the top (for nonsignificant, 0.01<*P < 0.1 and **P < 0.01). c, Density plots showing differential DNA methylation (x axis) and chromatin accessibility (y axis) at lineage-defining enhancers calculated for each of the lineage transitions.

Extended Data Fig. 12 | See next page for caption.

Extended Data Fig. 12 | Embryoid bodies recapitulate the transcriptional, methylation and accessibility dynamics of the embryo. a, Embryoid bodies show high transcriptional similarity to gastrulation-stage embryos. Top left, UMAP projection of RNA expression for the embryoid body dataset (n = 775). Cells are coloured by lineage assignment and shaped by genotype (WT or Tet TKO). Bottom left, UMAP projection of stages E6.5 to E8.5 of the atlas dataset (no extra-embryonic cells) with the nearest neighbours that were used to assign cell type labels to the scNMT-seq embryoid body dataset coloured in red (WT) or blue (Tet TKO). Middle, UMAP projection of embryoid body cells coloured by the relative RNA expression of marker genes. Right, scatter plot of the differential gene expression ($\log_2$ normalized counts) between different assigned lineages for embryoid bodies (x axis) versus embryos (y axis). Each dot represents one gene. Pearson correlation coefficient with corresponding P value (two-sided) are displayed. Lines show the linear regression fit. The top-four genes with the largest differential expression are highlighted in red. b, Global DNA methylation and chromatin accessibility levels during embryoid body differentiation. Top, box plots showing the distribution of genome-wide CpG methylation (left) or GpC accessibility levels (right) per time point and lineage (compare with Extended Data Fig. 3). Each dot represents a single cell (only wild-type cells are used). Box plots show median levels and the first and third quartile, whiskers show 1.5× the interquartile range. Bottom, heat map of DNA methylation (left) or chromatin accessibility (right) levels per time point and genomic context (compare with Fig. 1e, f). c, Ectoderm enhancers are more methylated in Tet TKO compared with wild-type epiblast cells in vivo. Bar plots show the mean (bulk) DNA methylation levels for ectoderm (left), endoderm (middle) and mesoderm (right) enhancers in E6.5 epiblast cells [25]. For each genotype, two replicates are shown. d, Profiles of DNA methylation (red) and chromatin accessibility (blue) at lineage-defining enhancers quantified over different lineages across embryoid body differentiation (only wild-type cells). Running averages in 50-bp windows around the centre of the ChIP–seq peaks (2 kb upstream and downstream) are shown. Solid lines display the mean across cells and shading displays the corresponding s.d. Dashed horizontal lines represent genome-wide background levels for methylation (red) and accessibility (blue).

Stephen J Clark
Wolf Reik
John C Marioni
Corresponding author(s): Oliver Stegle

Last updated by author(s): 18 October 2019

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 Authors & Referees 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 Kvo Ç }uu}v š •š• •Z}µo • Œ] •}o oÇ Ç v u V • Œ] u}Œ }u‰o Æ š Zv]‹µ •]v šZ D šZ
☒	☐	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. šŒ) with confidence intervals, effect sizes, degrees of freedom and W value noted ']À W À oµ • • Æ š À oµ • ÁZ v À Œ •µ]š o X
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

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	Sample sizes were determined in order to obtain at least 50 cells for each germ layer.
Data exclusions	Regions of coverage outliers and extreme strand bias excluded as these were assumed to be alignment artefacts.
Replication	For each developmental stage, we collected cells from at least 3 individual embryos and results were consistent across embryos.
Randomization	This is not relevant since we did not use different experimental groups or conditions in our study.
Blinding	This is not relevant since we did not use different experimental groups or conditions in our study.

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