T-ChIC: multi-omic detection of histone modifications and full-length transcriptomes in the same single cell

Peter Zeller [# 1,2], Marloes Blotenburg [# 1,2], Vivek Bhardwaj[1,2], Buys Anton de Barbanson[1,2], Fredrik Salmén[1,2], Alexander van Oudenaarden[1,2 *]

[1] Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences), Oncode Institute, Utrecht, The Netherlands.

[2] University Medical Center Utrecht, Utrecht, The Netherlands.

* Correspondence: a.vanoudenaarden@hubrecht.eu (A.v.O.)
# These authors contributed equally

Abstract

Epigenetic mechanisms, such as histone modifications, are key regulators of transcription and maintenance of cell identity. While our knowledge concerning cell type-specific histone modifications has constantly increased, we still know little about the interplay between epigenetics and transcription at the level of the individual cell. To gain an understanding of this process, we developed T-ChIC (Transcriptome + Chromatin ImmunoCleavage), a method allowing for the detection of full-length transcripts and histone modifications in the same single cell. We applied this technique to an *in vitro* model of gastrulation and monitored the coordinated dynamics of the transcriptome and active and repressive histone modifications as mouse embryonic stem cells differentiate into the three germ layers. Our analysis reveals a germ layer-dependent coupling between chromatin regulation and transcriptional states.

Introduction

During multicellular organism development, individual cells need to specify towards different germ layers and cell types in the correct space and time to form the complex mature body plan. In order for genes to become activated, the promoter region needs to be accessible for gene regulatory factors to bind. The Polycomb (PRC) and Trithorax (TrxG) group proteins deposit H3K27me3 and H3K4me3 respectively and play fundamental roles through their opposing function: while H3K4me3 marks active gene promoters and plays an essential role in PolII mediated transcription[1], H3K27me3 represses gene expression. On the onset of gastrulation, the embryo undergoes vast rearrangements and greatly increases in size and complexity, requiring the use of sensitive single-cell methods to disentangle cell type-specific gene regulation. Advances in single-cell RNA sequencing have provided a detailed understanding of cell type complexity and dynamics in developing systems, both *in vivo* and *in vitro*[2–4], and the first multi-omic atlases have started to elucidate the role of other epigenetic

layers including DNA methylation and chromatin accessibility during this process[5]. However, while the role of histone modifications has been studied in selected populations in bulk, their cell type-specific distributions remain unexplored at the single-cell level.

A variety of techniques that are based on antibody tethering of proteinA-Tn5 transposase[6–8] or proteinA-micrococcal nuclease (pA-MN), such as sort-assisted Chromatin ImmunoCleavage (sortChIC)[9–12] have been developed to study chromatin modifications at the single-cell level. However, to study the interplay between transcription and chromatin modifications and to tease apart cell type-specific profiles, co-acquisition of the cell's nascent transcriptome is necessary. For the development of multi-omics approaches recovering histone modification distribution and transcriptome from the same single cell pA-Tn5 methods that omit technically challenging DNA fragment processing steps have been used[13–16]. While some of these approaches allow the processing of large cell numbers by applying combinatorial barcoding[13,14] or 10x microfluidics[15], these techniques only capture the 3'-end of transcripts, which largely inhibits the identification of unspliced nascent RNA most closely reflecting ongoing transcriptional activity. The temporal disconnect between accumulated spliced RNA and active transcription can hinder a more detailed understanding of the interplay between chromatin regulation and active transcription.

Here, we developed T-ChIC (Transcriptome + Chromatin ImmunoCleavage), a novel multi-omic method to profile full-length transcriptome and histone modifications in the same single cell. This method combines VASA-seq for profiling of the full-length transcriptome with sortChIC as a read-out of chromatin modifications, which achieves the sensitivity of PaMN in addition to the high specificity demonstrated by PaTn5-based approaches. We demonstrate that T-ChIC provides a highly sensitive and specific read-out of genome-wide chromatin modification distributions in combination with the full transcriptome. We apply this technique to an *in vitro* model of gastrulation to study the epigenetic changes underlying lineage specification and provide cell type-specific characterisations of chromatin states during differentiation.

Results

T-ChIC profiles the full transcriptome and histone modification distribution in the same single cell with high sensitivity and specificity

In order to measure the transcriptome and histone modifications in the same single cell, we developed T-ChIC (Transcriptome + Chromatin Immunocleavage). To facilitate the profiling of histone modifications, an adaptation of the sortChIC method was used, which achieves high sensitivity and specificity, in particular when characterising repressive histone modifications[12]. For characterisation of the full-length transcriptome, we integrated the VASA-seq protocol[17] with sortChIC.

To detect both modalities, cells are fixed in ethanol, incubated with an antibody against the histone modification of interest, and protein A-MNase fusion is added that binds to the histone modification-bound antibody. G1-phase single cells are sorted into 384-well plates containing primers for transcript recovery, and MNase is activated through the addition of calcium, followed by MNase inactivation and cell lysis. This results in DNA double strand breaks surrounding the histone modification of interest. To profile the transcriptome, RNA is fragmented, repaired and poly-A-tailed to provide full-transcript coverage. Reverse transcription is performed using transcript primers containing a unique molecular identifier

(UMI) and transcript cell-specific barcode. This is followed by end repair of DNA fragments and ligation of adapters containing a ChIC cell barcode and UMI. Subsequently, plates are pooled, second strands are synthesised and products are amplified through in vitro transcription and PCR before sequencing (Figure 1A). Because the cell-specific barcodes are unique for the ChIC and transcript read-outs, they allow computational separation of both signals in each cell.

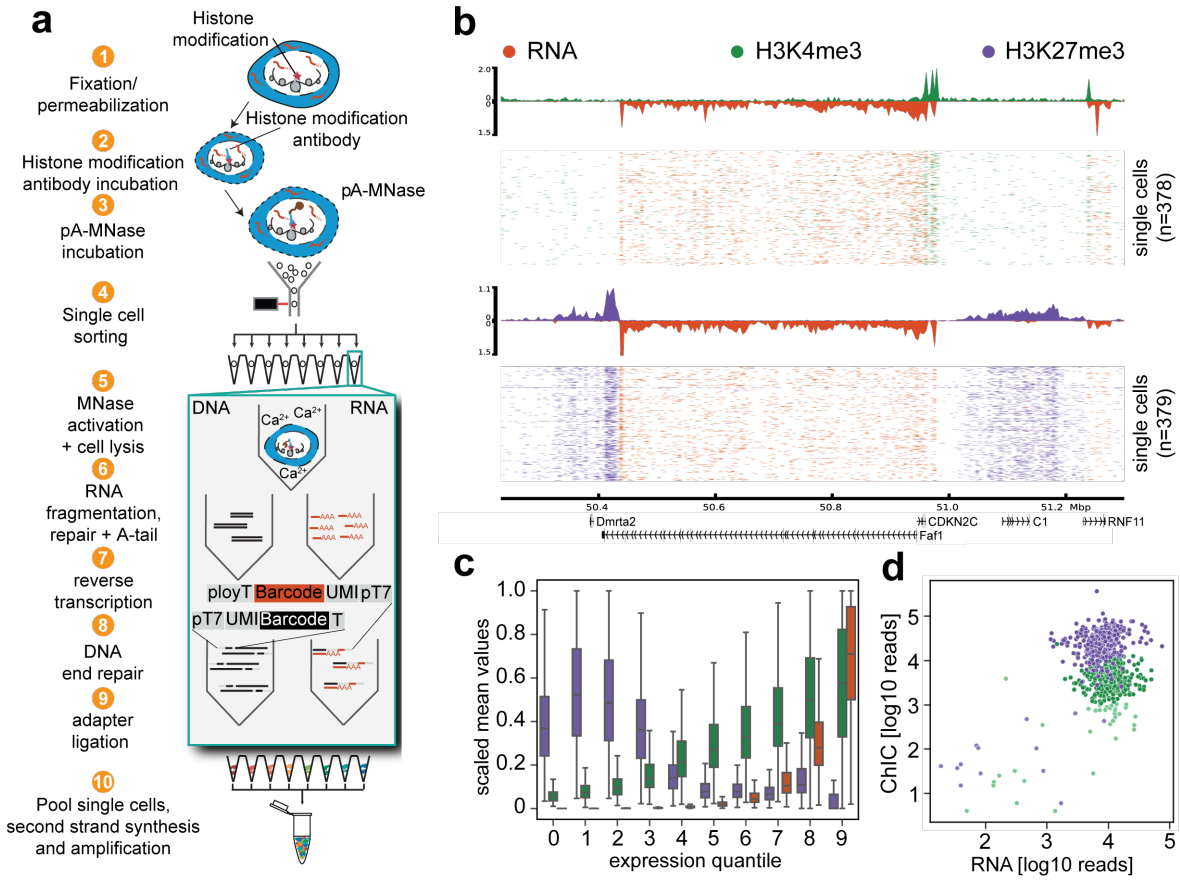


Figure 1 | T-ChIC allows the simultaneous recovery of the full length transcript and histone mark position from the same single cell. a) T-ChIC experimental workflow including cell fixation and Protein A-MN targeting in bulk (1-3), sorting of single cells into 384-well plates by FACS (4), followed by separate processing of RNA and DNA fragments inside the plate and introducing RNA- and DNA-specific cell barcodes (5-9). After pooling, fragments are linearly amplified by IVT before sequencing handles are introduced in a final indexing PCR (10). b) Single-cell coverage plots from K562 cells processed for transcriptome (orange) and H3K4me3 (top, green) or H3K27me3 (bottom, purple) over a 1 Mb region of Chromosome 2. Coverage plots above represent the average signals. c) Boxplots of single-cell read averages over genes split into quantiles of log10 expression. RNA (orange), H3K4me3 (green) and H3K27me3 (purple). Values are scaled per layer (RNA, H3K4me3 and H3K27me3), mean values are shown in supplementary fig1 b. d) Uniquely mapped transcripts (x-axis) and chromatin reads (y-axis) recovered for K562 cells treated for H3K4me3 (green) or H3K27me3 (purple). Pale dots represent QC-fail cells.

To assess the data quality generated, we performed T-ChIC with either H3K27me3 or H3K4me3 antibodies in K562 cells and investigated the specificity and sensitivity of the resulting signal. We observed specific enrichment of transcript reads over gene bodies, with H3K4me3 covering promoter regions of transcribed genes and H3K27me3 covering untranscribed regions (Figure 1B, S1A). To quantify the co-occurrence of H3K4me3 and H3K27me3 signal with transcription, we grouped genes into quantiles according to their

expression and determined the average coverage of both histone modifications per cell. While highly expressed genes show co-occurrence with H3K4me3 signal, lowly or non-expressed genes show a higher H3K27me3 coverage, which corresponds to the known active and repressive nature of these modifications (Figure 1C, S1B). Transcript recovered with T-ChIC correlated well with VASA-seq, and chromatin modification distributions profiled with T-ChIC correspond to results obtained with sortChIC (Figure S1C). Additionally, the recovered number of transcript reads was independent of the number of ChIC reads and the histone modification recovered, excluding significant signal interference between the two modalities (Figure 1D). Finally, we compared our multi-omic method with other recently published protocols profiling both transcriptome and chromatin modifications in single cells. Both in terms of RNA and ChIC signal, our protocol recovers the highest number of fragments (Figure S1D). In addition, although all methods showed similar enrichment of H3K4me3 on active genes, T-ChIC was the only method recovering the expected anti-correlation between the repressive H3K27me3 and gene transcription without the need for normalisation (Figure S1E).

Application of T-ChIC to gastruloids reveals epigenetic regulation during cell differentiation

Next, we applied T-ChIC to gastruloids. These are assemblies of mouse embryonic stem cells (mESCs) that over the course of several days grow out into polarised structures with an anteroposterior axis and cell types derived from all three germ layers[18–20]. We chose this system because of its scalability and faithful simulation of various processes occurring during *in vivo* embryonic development, including differentiation of cells into a complex mixture of cell types. To allow for extended differentiation, we embedded the gastruloids in matrigel at 96 h after aggregation (AA) as reported before[21,22], and modified the protocol to be extended to 7 days AA resulting in gastruloids showing reproducible morphologies. Using this improved protocol, we generated gastruloids and sampled across a time course from 72 until 168 h AA with 24-hour intervals (Figure 2A). We also included mESCs at the time point of aggregation.

First, we confirmed that the sampling time point or histone mark profiled did not affect the transcript data quality. Between the two target histone marks, we recover similar numbers of total counts as well as genes detected, and observe minor differences in the percentage of mitochondrial reads per cell (Figure S2A). The filtered transcriptome data consists of 38,117 single cells, with 18,279 H3K4me3 co-profiled cells and 19,838 H3K27me3 co-profiled cells intermixed in UMAP-space (Figure S2B). We observed a clear differentiation axis from early to late time points, with three major trajectories and a pluripotency cluster consisting of early time points (Figure 2B). Next, we performed detailed cell typing based on expression of known marker genes (Figures 2C, S2C-D). We identified the expected variety of cell types derived from the three germ layers, including the previously reported cardiac mesoderm, somites, endoderm, PGCs and neural progenitors[18,19,21,22]. Notably, especially in the later time points we recovered some previously unrecorded cell types such as erythrocytes and motor neurons, including several neural tube subtypes. We noticed a gradual differentiation from pluripotent cell types towards the three lineages, followed by specification into more discrete cell types (Figures 2D-E). For subsequent analyses, we grouped the cells into the three lineages or classified cells as pluripotent. (Figure 2F, Figure S2E). For all identified lineages, expected time-point distributions and similar ratios of H3K4me3- and H3K27me3-profiled cells are recovered (Figure S2E-F).

To further explore these differentiation trajectories, we calculated transcriptome-based pseudotime and displayed the gene expression patterns of pluripotency and differentiation markers of the three lineages (Figures 2F-G). As expected, we observe expression of

pluripotency markers early in pseudotime, followed by gene expression of symmetry breaking factors Brachyury (*T*) and *Cdx2*, and finally sequential activation of lineage specifying genes including *Neurod4* (ectoderm), *Sox17* (endoderm), and *Meox1* (mesoderm).

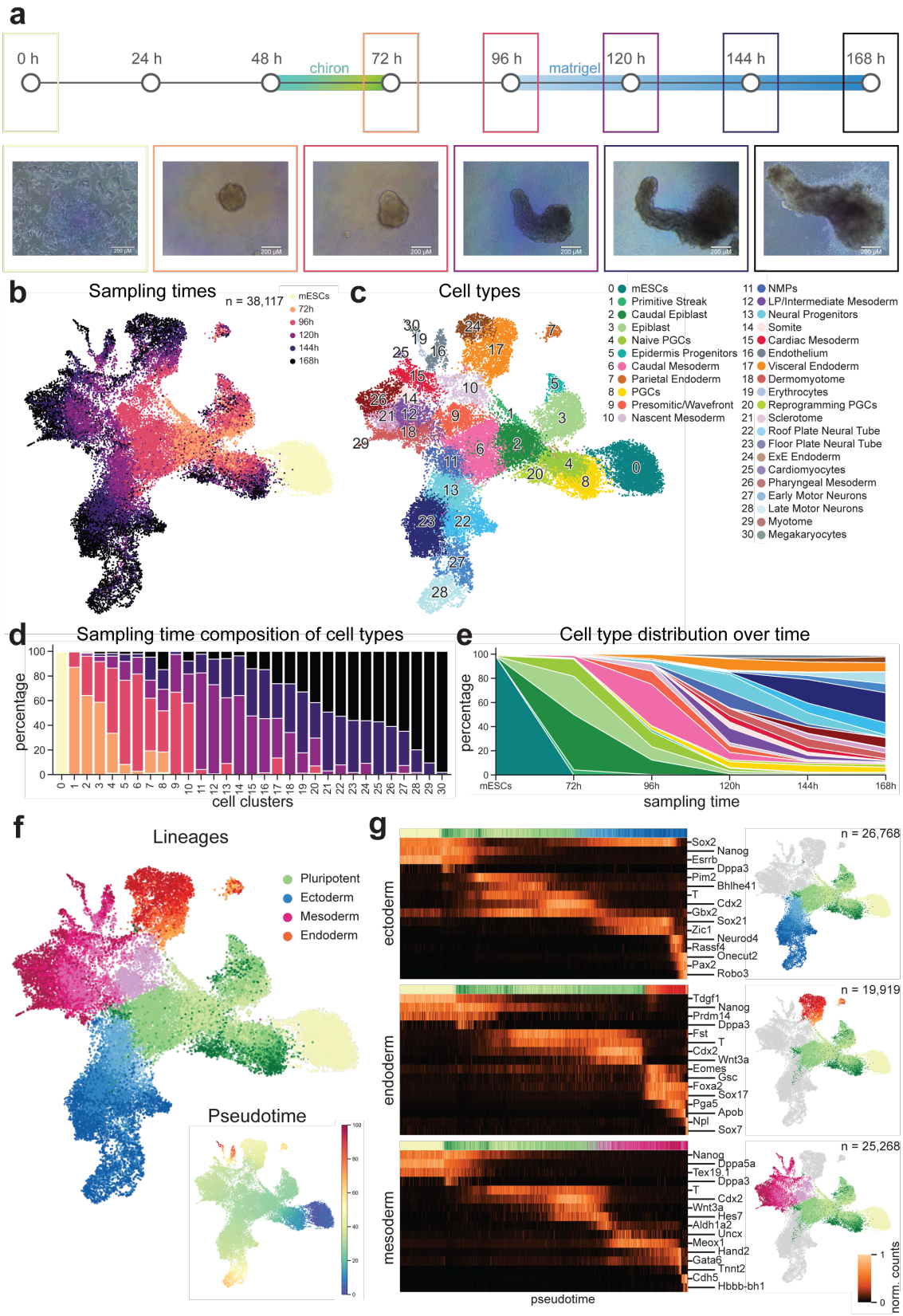


Figure 2 | Time course of gastruloid development recovers lineage commitment to the three germ layers. a) Overview of sampling points during the gastruloid time course and the representative gastruloid morphologies. b) Transcriptome UMAP of the gastruloid dataset (n = 38,117) coloured by sampling time point of cells profiled. c) Transcriptome UMAP coloured by identified cell types. Pluripotent cell types are shaded in green, endodermal cell types are shaded in orange, mesodermal cell types are shaded in red/purple, and ectodermal cell types are shaded in blue. d) Sampling time point distribution per cell type. Numbers on the x-axis correspond to numbered cell types in c. e) Cell type distribution across time points sampled. Colours represent cell types in c. f) Transcriptome UMAP coloured by lineage. Insert: transcriptome UMAP coloured by pseudo time. g) Gene expression patterns across pseudotime for marker genes of ectoderm (top, n = 26,768), endoderm (middle, n = 19,919), and mesoderm (bottom, n = 25, 268). All trajectories include pluripotent cells.

T-ChIC recovers epigenetic gene regulation during gastruloid development

Next, we investigated how chromatin modifications H3K4me3 and H3K27me3 change during gastruloid development. For both modifications, we aggregated counts across the whole genome in 50 kb windows and selected cells based on minimal unique molecules and a local cell-neighbourhood-specific genomic enrichment, resulting in 17,916 and 18,956 high quality cells for H3K4me3 and H3K27me3 respectively (Figure S3A-C). We then performed Latent Semantic Analysis (LSA) followed by UMAP embedding with colours representing the transcriptome-based cell classification, and compared the obtained results (Figure 3A). We recover similar cluster complexity for both H3K4me3 and H3K27me3, separating the three germ layers and pluripotency clusters. In general, we recover comparable amounts of H3K4me3-related fragments over the experimental time course, with a small increase in mESCs. H3K27me3 signal on the other hand decreases during differentiation (Figure S3A).

For further analysis of genic chromatin changes relative to their transcription we selected counting windows based on the distributions of both marks surrounding transcribed and silent genes (Figure S3D). For H3K4me3, we counted the number of fragments in ±5 kb surrounding the transcription start sites (TSS), and for H3K27me3, we counted inside the first 10 kb into the gene body. Visualising gene-specific counts across the three modalities shows a co-occurrence of H3K4me3 with gene expression and anti-correlation with H3K27me3 in single cells (Figure 3B). However, while known lineage-specific genes such as *Sox17*, *Prtm8*, and *Hand2* show a strong H3K27me3 enrichment in cells not expressing the gene, the pluripotency gene *Nanog* does not show H3K27me3-mediated silencing in differentiated cell types. At the same time, H3K4me3 enrichment is not limited to cells expressing the gene but can be also found in less differentiated cells of the same trajectory.

One of the most prominent examples of epigenetic gene regulation in development is the collinear derepression of the Hox gene cluster, which was also described to occur in gastruloids in a bulk time course experiment[19]. The relative enrichment of RNA and histone mark distributions along the HoxA cluster in single cells 7 days after aggregation show celltype-related derepression of the Hox cluster (Figure 3C, S3E). Pluripotent and endoderm cells are characterised by the absence of transcript and H3K4me3 along with H3K27me3 coverage over the full length of the HoxA cluster. Mesoderm and ectoderm cells show a stepwise withdrawal of H3K27me3 across the domain, accompanied by the activation of the underlying Hox genes. Subsequently, we followed the 5 top differentially expressed genes per cell type along the three differentiation trajectories (Figure 3D). This showed that most genes that are expressed during late time points are repressed by H3K27me3 in pluripotent cells. In contrast, genes expressed in pluripotent cell types accumulated H3K27me3 in a germ layer-specific manner, with endoderm cells silencing most pluripotency genes, mesoderm cells silencing a subset, and ectoderm cells accumulating repressive H3K27me3 on only a small subset of genes. This observation is consistent with the abundance of the three modalities per

cell type, with large variation in H3K27me3, but only moderate differences for the transcriptome and H3K4me3 signal (Figure S3F). H3K27me3 abundance is highest in pluripotency and endoderm cell types, while mesoderm displays intermediate levels and ectoderm contains the lowest abundance of this histone modification. To conclude, morphological and transcriptional changes characterised in the gastruloid system are underlined by epigenetic alterations. Generally, promoters of transcribed genes are covered by H3K4me3, and genes covered by H3K27me3 stay unexpressed. In addition, we do observe lineage- and differentiation-specific differences in the usage of both histone modifications, with H3K27me3 strongly regulating lineage-specific genes in pluripotent and endoderm cell types.

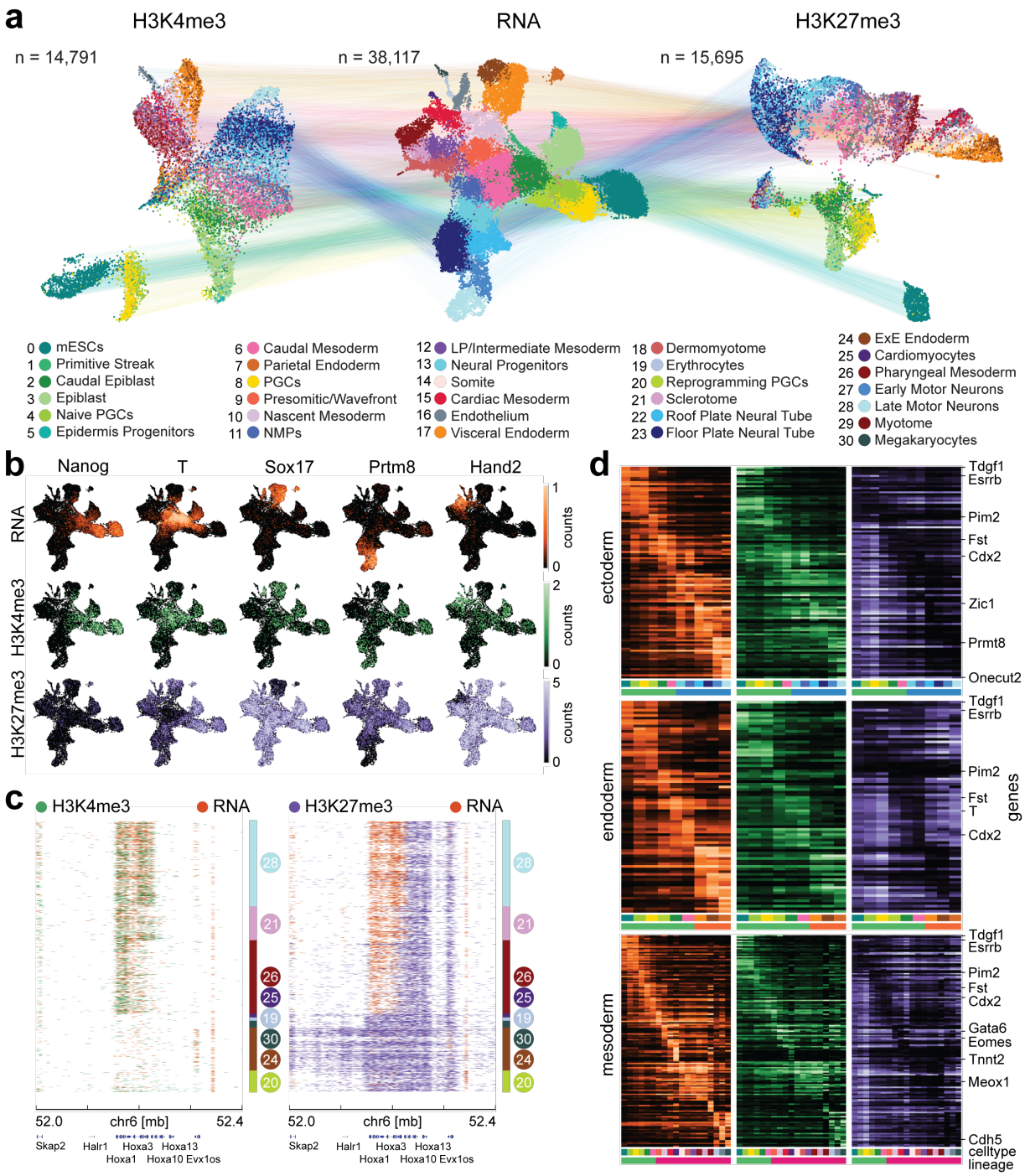


Figure 3 | T-ChIC recovers epigenetic changes underlying transcriptional programs during gastrulation. a) Three UMAPs generated based on each of the 3 modalities. Gene expression counts are used for RNA (middle), and 50 kb full genome counts were used for H3K4me3 (left) and H3K27me3 (right). Dots are coloured based on transcript-based cell types identification (Figure 2c). b) Transcript-based UMAPs coloured by gene expression counts for RNA, promoter counts for H3K4me3 and gene counts for H3K27me3 of selected marker genes. c) Single-cell coverage tracks of the HoxA cluster on chromosome 6. Stacked colour bar (right) indicates selected cell clusters from a. Left: RNA (orange) and H3K4me3 (green), right: RNA (orange) and H3K27me3 (purple) are shown. d) Heatmaps of mean log10 RNA counts (orange), mean H3K4me3 promoter counts (green) and mean H3K27me3 gene counts (purple) are shown over top 5 differentially expressed genes per cell type along the three germ layer trajectories.

Discussion

In this study, we introduce T-ChIC, a novel single-cell method that allows the recovery of full-length transcriptome and histone modification distribution with high sensitivity and accuracy. We demonstrate that our technique is particularly accurate in the recovery of repressive chromatin modifications, where it outperforms Tn5-based technologies. Repressive histone modifications are especially important to characterise, as they form a boundary for transcriptional gene activation and thereby have the potential to restrict gene expression and lineage decisions.

We apply T-ChIC to an extended gastruloid system and profile H3K4me3 and H3K27me3 distributions over a time course of seven days. We demonstrate that differential gene expression is accompanied by corresponding changes in epigenetic modifications. This includes the Hox gene cluster, which is known to display collinear expression in gastruloids [19,23]. Here, we show that H3K27me3 and H3K4me3 profiles divide the *Hox* cluster into an active and an inactive compartment, which correspond to collinear activation observed in vivo. As expected, promoters of transcribed genes are covered by H3K4me3, while genes covered by H3K27me3 stay unexpressed. This occurs in a germ layer-dependent manner, with ectoderm cells repressing fewer pluripotent genes by H3K27me3 compared to mesoderm and endoderm. We thereby provide a time- and cell type-resolved atlas of epigenetic changes during gastruloid development, which can serve as a first step in understanding the epigenetic control of transcription upon differentiation.

Methods

Cell culture

K562 cells (ATCC® CCL-243™) were grown in RPMI 1640 Medium GlutaMAX™, supplemented with 10 % FCS, Pen-Strep and non-essential amino acids. After harvesting, cells were washed 3 times with room temperature PBS before continuing with the cell fixation.

129S1/SvImJ / C57BL/6 mESCs[24] were cultured in a humidified incubator (5 % CO2, 37 °C) on 0.1 % gelatin-coated cell culture dishes in ESLIF medium (GMEM (Gibco) containing 10 % fetal bovine serum (FBS, Gibco, #16141-079, lot 2215500RP), 1 % Sodium Pyruvate, 1 % non-essential amino acids, 1 % GlutaMAX, 1 % penicillin-streptomycin, 0.2 % β-mercaptoethanol, and 1000 units/mL mouse leukemia inhibitory factor (mLIF, ESGRO)). Cells were split every second day by washing with PBS0, dissociating with TrypLE for ~5 min, centrifuging at 300 g for 3 min, and replating the pellet 1:5 on gelatinised plates. 2i medium

was prepared as follows: 48.1 % DMEM/F12 (Gibco) and 48.1 % Neurobasal (Gibco) containing 0.5 % N-2 supplement (Gibco, #7502048), 1 % B-27 supplement (Gibco, #17504044), 1 % GlutaMAX, 1.1 % penicillin-streptomycin, 0.2 % β-mercaptoethanol, 1000 units/mL mouse leukemia inhibitory factor (mLIF, ESGRO), 0.03 % Chiron (CHIR99021, Tocris #4423) and 0.01 % PD032509 (Sigma, #PZ0162). Before gastruloid formation, mESCs were cultured for 2 days in 2i medium followed by 2 days in ESLIF medium and aggregation. All cell lines were routinely tested and confirmed to be free of mycoplasma.

Gastruloid formation

Gastruloids were generated as described before[18,21] with the following adaptations: N2B27 medium (Takara, #Y40002) was passed through a 0.22 µm filter before use and 600 cells were used for aggregation instead of 300. For all time points from 96 h AA onwards, gastruloids were embedded in 10 % matrigel (Corning, #356231, lot 0006004) in N2B27 medium as follows: 6-well plates were coated with 1 mL 10 % matrigel in N2B27 and left to solidify at 37 °C for >1 h. Gastruloids were harvested at 96 h AA from 96-well plates, pooled, and placed on ice as shortly as possible. All medium was removed from gastruloid pools and replaced with 10 % matrigel in N2B27, and a maximum of 32 individual gastruloids in matrigel were plated per 6-well and distributed evenly. At 120 h AA and 144h AA, gastruloids were either harvested or medium was refreshed by removing and adding 1 mL N2B27.

Dissociation and fixation of mESCs and gastruloids

mESCs were dissociated as described above, resuspended in PBS0 supplemented with 5 % FBS, and filtered through a 35 µm filter (Falcon, 352235).

Non-matrigel-embedded gastruloids were pooled in protein low-bind tubes (#0030122216), washed with PBS0, and incubated with 1 mL trypsin (Gibco) for ~5 min at 37 °C. Gastruloids were mechanically broken up into a single-cell suspension by pipetting with a P200 pipette, resuspended in N2B27 medium, and washed twice with PBS0. Embedded gastruloids were removed from matrigel as follows: N2B27 medium was removed from 6-well plates, replaced with 1 mL Corning Cell Recovery Solution (#354253), and incubated on ice for ~15 min. Gastruloids were flushed from matrigel solution using a P1000 pipette and collected in protein low-bind tubes, followed by dissociation as described above.

From here, all steps were performed at 4 °C. Cells were washed once in cold PBS0, resuspended at 1 million cells per 300 µL PBS0, pre-cooled ethanol was added while vortexing to a final concentration of 70 %, and cells were mildly fixed at -20 °C for 1 h. After, cells were washed once in antibody incubation buffer (47.5 mL H2O RNAse free, 1 mL 1 M HEPES pH 7.5 (Invitrogen), 1.5 mL 5 M NaCl, 3.6 mL pure spermidine solution (Sigma Aldrich), 0.05 % Tween-20, EDTA free protease inhibitor, 1 U/µL RNasin® Plus Ribonuclease Inhibitor, 2 mM EDTA; buffer identical to T-ChIC wash buffer + 2 mM EDTA) at 4 °C, resuspended in antibody incubation buffer supplemented with 10 % DMSO, and aliquoted at 0.5 million cells per tube for long-term storage at -80 °C.

T-ChIC protocol

PA-MN targeting

(1) After thawing, cells were pelleted at 500 g for 4 min and washed once in T-ChIC wash buffer (47.5 mL H2O RNAse free, 1 mL 1 M HEPES pH 7.5 (Invitrogen), 1.5 mL 5 M NaCl, 3.6 mL pure spermidine solution (Sigma Aldrich), 0.05 % Tween-20, EDTA free protease inhibitor, 0.2 U/µL RNasin® Plus Ribonuclease Inhibitor; buffer identical to basic T-ChIC buffer + protease and RNAse inhibitor). (2) Cells were resuspended in 200 µL antibody incubation

buffer per 0.5 million cells and were aliquoted into 0.5 mL protein low binding tubes containing the primary histone mark antibody (1:200 dilution for H3K4me3 and 1:100 for H3K27me3) diluted in 200 µL antibody incubation buffer. (3) Cells were incubated overnight at 4 °C on a roller, (4) followed by washing once with 500 µL T-ChIC wash buffer. (5) Afterwards, cells were resuspended in 500 µL T-ChIC wash buffer containing a total of 1 U/µL RNAsin plus, PaMN (3 ng/mL) and Hoechst 34580 (5 mg/mL) and (6) incubated for 1 h at 4 °C on a roller.
(7) Finally, cells were washed an additional 2 times with 500 µL T-ChIC wash buffer before passing through a 70 mm cell strainer (Corning, 431751) for sorting.

FACS sorting
(8) For all experiments, cells were gated for G1 cell cycle stage based on Hoechst staining and forward scatter and trigger pulse width to further remove cell doublets on an Influx FACS machine. Cells were sorted into 384-well plates containing 5 µL sterile-filtered mineral oil (Sigma Aldrich) and 50 nl RNA recovery primer (28 ng/ml, supplementary table 1) per well using index sorting, which records FACS information for every sorted well. To further increase sorting specificity of single cells, we used the following sort settings: objective: single, number of drops = 1, extra coincidence = complete empty (no signal in the previous and next drop), and phase mask = center 10/16 (cell is in the middle of the sorted drop).
For both marks we sorted 3 plates of K562 cells, 6 plates with mESCs, 6 plates with 72 h AA gastruloids and 12 plates for all later gastruloid timepoints. For H3K4me3, clone G.532.8 (Invitrogen, MA5-11199) was used at 1:400 dilution and for H3K27me3, clone C36B11 (NEB, 9733S) was used at 1:200 dilution. Data was collected using BD FACS software (version 1.2.0.124).

Protein A-MN activation
Volumes for the following steps of the protocol were distributed using a Nanodrop II system (Innovadyme) and plates were spun for 2 min at 4 °C and 2000 g after each reagent addition. (9) 100 nL of Basic T-ChIC buffer (47.5 mL Nuclease-free water, 1 mL 1 M HEPES pH 7.5 (Invitrogen), 1.5 mL 5 M NaCl, 3.6 mL pure spermidine solution (Sigma Aldrich), 0.05% Tween20) containing 3 mM $CaCl_2$, was added per well to induce Protein A-MN-mediated chromatin digestion. (10) For digestion, plates were incubated for 30 min in a PCR machine set at 4 °C. (11) The reaction was stopped by adding 100 nL of the stop solution (40 mM EGTA to chelate Ca^{2+} and stop MN (Thermo, 15425795), 1.5% NP40, 2 mg/mL temperature labile proteinase K (NEB, P8111S), and 12.5 mM MgCl2 for RNA fragmentation). (12) Plates were incubated in a PCR machine for 20 min at 4 °C, followed by the release of chromatin and digestion of PaMN by proteinase K at 37 °C for 2 h and heat inactivation at 55 °C for 20 min. Afterwards, plates can be stored at -80 °C until further processing.

Library preparation
(13) After thawing, RNA was fragmented at 85 °C for exactly 2 min before cooling down on ice. (14) RNA fragments were repaired and poly-A-tailed through the addition of 150 nL RNA repair mix per well and incubation at 37 °C for 1 h. Per plate, the RNA repair mix contains 18.75 µL Nuclease-free water, 12 µL 0.1 M DTT, 0.9 µL 1 M MgCl2, 18 µL 1 M KCl, 16.8 µL 1 M Tris with pH 8, 4.5 µL 50% PEG8000, 0.45 µL 20 mg/mL BSA, 5.25 µL 0.1 mM ATP, 6 µL RNase-OUT, 1.05 µL polyA Polymerase, and 7.20 µL T4 PNK.

(15) Repaired RNA fragments were reverse transcribed by adding 100 nL of RT mix per well and incubation at 50 °C for 1 h. Per plate, the RT mix contains 38.7 µL Nuclease-free

water, 3 µL 0.1 M DTT, 4.5 µL 1 M KCl, 3 µL 50% PEG8000, 0.3 µL 20 mg/mL BSA, 4.5 µL 10 mM dNTPs, and 6 µL SuperScript III.

(16) DNA fragments were blunt-ended by adding 150 nL DNA repair mix per well and incubation for 30 min at 37 °C, followed by 20 min at 75 °C for enzyme inactivation. Per plate, the DNA repair mix contains 54.6 µL Nuclease-free water, 1.2 µL 5 M NaCl, 10.5 µL 1 M Tris pH 7.5, 1.5 µL 1 M MgCl2, 4.5 µL 0.1 M DTT, 4.5 µL 50% PEG8000, 0.45 µL 20 mg/mL BSA, 3.9 µL 100 mM ATP, 1.95 µL 10 mM dNTPs, 3 µL TL Exo1, 1.95 µL Klenow large fragment, and 1.95 µL T4 PNK.

(17) Blunt fragments were subsequently A-tailed by adding 150 nL A-tailing mix per well and incubation for 15 min at 72 °C. Per plate, the A-tailing mix contains 76.7 µL Nuclease-free water, 6 µL 1 M KCl, 4.5 µL 50% PEG8000, 0.45 µL 20 mg/mL BSA, 1.2 µL 100 mM dATP, and 1.2 µL Taq360.

Next, fragments were ligated to forked adapters containing T-overhangs (see supplementary table 2 for sequences). (18) For adapter ligation, 50 nL of 5 mM adapters in 50 mM Tris pH 7 was added to each well with the Mosquito HTS (ttp labtech). (19) After centrifugation, 150 nL of adapter ligation mix was added and (20) plates were incubated for 20 min at 4 °C, followed by 16 h at 16 °C for ligation and 10 min at 65 °C for ligase inactivation. Per plate, the adapter ligation mix contains 17.1 µL Nuclease-free water, 2.1 µL 1 M MgCl2, 2.7 µL 1 M Tris with pH 7.5, 40.5 µL 0.1 M DTT, 4.5 µL 50% PEG8000, 0.45 µL 20 mg/mL BSA, 2.1 µL 100 mM ATP, and 21 µL 400,000 units/mL T4 DNA Ligase.

(21) Before pooling, 1 mL of Nuclease-free water was added to each well to minimise material loss. (22) Ligation products were pooled by centrifugation into oil-coated VBLOK200 Reservoirs (ClickBio) at 500 g for 2 min and (23) the liquid phase was transferred into 1.5 mL Eppendorf tubes and (24) was purified by centrifugation at 13000 g for 1 min and transferred into a fresh tube. Steps 23 and 24 were performed twice. (25) DNA fragments were purified using Ampure XP beads (Beckman Coulter - prediluted 1 in 8 in bead binding buffer – 1 M NaCl, 20 % PEG8000, 20 mM Tris pH 8, 1 mM EDTA) at a bead-to-sample ratio of 0.8. (26) After 15 min incubation at RT, beads were washed twice with 1 mL 80% ethanol, resuspending the beads during the first wash and (27) resuspension in 17 µL Nuclease-free water. After 2 min elution, the resuspended beads were transferred into a fresh 0.5 mL tube.
(28) Second-strand synthesis was performed per plate using the NEBNext® Ultra™ II Non-Directional RNA Second Strand Synthesis Module. For this, 2 µL 10x reaction buffer was mixed with the sample on ice, followed by addition of 1 µL enzyme, gentle mixing, and incubation at 16 °C for 2.5 h and 65 °C for 20 min.

(29) A second cleanup was performed after second strand synthesis by adding 26 µL of undiluted Ampure XP beads. The beads were resuspended in 8 µL Nuclease-free water. (30) After the cleanup, the DNA was linearly amplified by in vitro transcription with 12 µL of the MEGAscript™ T7 Transcription Kit (Fisher Scientific, AMB13345) for 12 h at 37 °C. (31) The produced RNA was purified using RNA Clean XP beads (Beckman Coulter) with a 1.3 beads-to sample ratio (26 µL beads), followed by sample resuspension in 12 µL Nuclease-free water.

(32) At this stage, 6 µL aRNA can be stored at -80 °C, while continuing with the remaining 6 µL. Next, rRNA was depleted using 2 µL Hybridisation Buffer (1000 mM NaCl in 500 mM Tris pH 7.5) and 2 µL 25 µM rRNA depletion oligos (supplementary table 3). (33) Oligos were hybridised to their rRNA targets by incubation at 95 °C for 2 min, followed by a gradual temperature decrease of 0.1 °C/s down to 45 °C. Once at 45°C, 10 µL RNAse mix (2 µL thermostable RNAseH (Epicentre), 8 µL RNAse buffer: 250 mM NaCl and 50 mM MgCl2 in 125 mM Tris pH 7.5) at 45 °C was added and samples were incubated at 45 °C for another

30 min. (34) DNA template and unused depletion probes were removed through addition of 4 µL RNase-free DNAse (Promega) with 5 µL 10 mM CaCl2 in 30 µL Nuclease-free water and incubated at 37 °C for 30 min. (35) RNA was purified using 80 µL RNA Clean XP beads per sample and elution in 6 µL water.

(36) 5 µL supernatant was transferred to a new 0.5 mL Eppendorf tube for RNA adapter ligation. For this, 1 µL 20 µM RA3 (/5rApp/TGGAATTCTCGGGTGCCAAGG/3SpC3/) was added, mixed and incubated for 2 min at 70 °C, followed by cooling on ice. On ice, 4 µL RNA ligation mix was added, mixed by pipetting and incubation for 1 h at 25 °C (Lid temperature should not exceed 25 °C). Per sample, RNA ligation mix contains 1 µL T4 RNA ligase 2, truncated (NEB), 1 µL T4 RNA ligase reaction buffer (NEB), 1 µL RNAseOUT (Invitrogen), and 1 µL Nuclease-free water. (37) RNA was primed for reverse transcription by adding 1 µL 10 mM dNTPs and 2 µL 20 mM reverse transcription primer - RTP (GCCTTGGCACCCGAGAATTCCA) and (38) hybridization through incubation at 65 °C for 5 min and direct cool-down on ice. (39) Reverse transcription was performed by addition of 4 µL first strand buffer (Invitrogen), 1 µL 0.1 M DTT, 1 µL RNaseOUT (Invitrogen), 1 µL SuperscriptIII (Invitrogen) and 1 µL Nuclease-free water and (40) incubating the mixture at 50 °C for 1 h followed by 15 min at 70 °C.

(41) Single-stranded DNA was purified using 1 µL RNAseA (Thermo Fisher) and incubation for 30 min at 37 °C. (42) DNA was cleaned up with 22 µL Ampure XP beads, and resuspension in 20 µL Nuclease-free water. 10 µL can be stored at -80 °C.
(43) The library PCR to add the Illumina small RNA barcodes and handles was performed by adding 25 µL NEBNext Ultra II Q5 Master Mix (NEB, M0492L), 11 µL Nuclease-free water and 2 µL RP1 and RPIx primers each (10 mM, for details see Illumina smallRNA indexing primer sequences) to the 10 µL single-stranded DNA product. (44) PCR incubation: activation for 30 sec at 98 °C, 8 cycles 10 sec at 98 °C, 30 sec at 60 °C, 30 sec at 72 °C, final amplification 10 min at 72 °C (45) PCR products were cleaned by 2 consecutive DNA bead clean-ups with a 0.8X bead-to-sample ratio, eluting in 30 µL Nuclease free water after the first round and (46) in 11 µL for the final product. Abundance was determined using a dsDNA hs QUBIT, and after adjusting concentrations to ~2 ng/ µL quality was assessed using a DNA hs bioanalyzer.

Pa-MN production
The Pa-MN fusion protein was produced following the methods section in [25].

Sequencing
Final DNA libraries were sequenced paired-end 100 bp, on either a Novaseq or NextSeq2000.

Transcriptome analysis
Fastq files were processed into .loom files using the T-ChIC Snakemake (see https://github.com/marloes3105/tchic/tree/main/workflows/1.snakemake-workflow/, v1.0). The pipeline works as follows: reads were demultiplexed for CelSeq2 barcodes using SingleCellMultiOmics' (SCMO) demux.py (version 0.1.30), polyA tails were trimmed with SCMO trim_vasa.py and adapters were trimmed with cutadapt (version 4.1), followed by mapping of the reads with STAR (version 2.5.3a) to the 129/B6 SNP-masked GRCm38 mouse genome (Ensembl 97). Reads were tagged and deduplicated using SCMO bamtagmultiome.py, and velocyto (version 0.17.17) was run to generate output .loom files. Downstream analysis was performed with Scanpy (version 1.8.1). Cells with less than 1.000 or more than 50.000 reads, and less than 100 genes detected were filtered out. Cells with more than 0.6 percent mitochondrial reads were also excluded. Mitochondrial reads, Malat1,

SnoRNAs, and ribosomal reads were excluded from further analysis. Counts were normalised to 1.000 transcripts per cell and logarithmized. The number of counts, percentage of mitochondrial reads, and ChIC mark profiled for the same cell were regressed out and each gene was scaled to unit variance with a maximum of 10.

Principal component analysis was performed and used to filter out cells with low ChIC quality (see cell filtering). Neighbours were calculated with sc.pp.neighbors with settings n_neighbors = 5, n_pcs = 50, metric = manhattan and the UMAP was generated with sc.tl.umap with settings min_dist = 0.4, spread = 0.6. Pseudotime was calculated using cellrank's (version 1.1.0) pseudotime embedding in combination with scvelo (version 0.2.4). In short, terminal states were calculated based on lineage + day clusters ('lin_day') using cr.tl.terminal_states with setting weight_connectivities 0.2, and initial states were calculated with cr.tl.initial_states. Lineages were computed with cr.tl.lineages. Finally, pseudotime was calculated with scv.tl.recover_latent_time with settings root_key = 'initial_states_probs' and end_key = 'terminal_states_probs'.

ChIC analysis

Fastq files were processed into count tables using the T-ChIC Snakemake (see https://github.com/marloes3105/tchic/tree/main/workflows/1.snakemake-workflow/, v1.0). The pipeline works as follows: reads were demultiplexed for T-ChIC barcodes with a hamming distance of 0 using SingleCellMultiOmics' (SCMO) demux.py (version 0.1.30), followed by adapter trimming using cutadapt (version 4.1). Reads were mapped paired-end with bwa mem (version 0.7.16a) to the 129/B6 SNP-masked GRCm38 mouse genome (Ensembl 97) with the following parameters: -M -I 1000. Reads were tagged with SCMO bamtagmultiome.py using a blacklist for specific genomic regions[26] and duplicates and reads with an insert size above 1000, mapping quality below 50, or non-TA ligation motif were removed.

To exclude low quality cells, transcript-based PCs were used to determine neighbourhoods of comparable cells. The local chromatin enrichment of each cell was compared to the neighbourhood average. To determine the neighbourhood average, pseudo-bulks of total count normalised cells were created and the highest covered bins containing 80% of the reads were determined. Next, the read fraction of the cell of interest in those neighbourhood-enriched bins was determined and cells with an enrichment of 2 std below average were marked as low quality. By definition, the enrichment distribution is around 80%. The script that was used is cellSelection_PCA.py with parameters: -neighbors *number of neighbors to use* -cutoff *minimal number of reads* -chic_csv *genome wide ChIC count table/merged_count_table_50000.csv* -trans_csv *PCs exported from scanpy* -o *folder name for output*. For all datasets, 200 neighbours were used, and minimum reads per cell were set at 300 for H3K4me3 and 900 for H3K27me3.

Comparison with other protocols

The chromatin and transcriptome count tables for published protocols were downloaded from their GEO repositories (Xiong et. al. 2021[13]: GSE158435; Zhu et. al. 2021[14]: GSE152020 ; Rang et. al. 2022[27]: GSE184036). To make the data biologically comparable, we only selected ES cells (Xiong et. al.[13]: 6993 cells; Zhu et. al.[14]: 13999 cells; Rang et. al.[27]: 2892 cells; this study: 3441 cells). For the chromatin fraction, Rang *et al.* data contained counts aggregated in 10kb genome-wide bins. We therefore aggregated the chromatin counts from all other studies (including ours) on the same bins. For the transcriptome, we used the gene-level UMI counts from all studies. To compare the variance in chromatin counts based on ranked gene expression, we ordered genes by their mean expression across all cells in their respective

study and divided them into 10 fractions. We then overlapped the gene promoters (TSS +-1kb) with the chromatin bins, summed up the counts in overlapping bins per gene, and plotted the average, log-scale counts per gene across cells in the respective study.

Code and data availability

All data generated for this project, code used and annotated notebooks for performing the analysis and generating figures can be found on github (https://github.com/marloes3105/tchic, release v1.0).

Acknowledgements

The authors thank the Hubrecht Sorting Facility (HSP) and Utrecht Sequencing Facility (USEQ) for cell sorting and sequencing. The 129S1/SvImJ / C57BL/6 mESC line was a gift from Matyas Flemr and Marc Bühler. Computational unmixing of VASA and ChIC barcodes was developed by Barbanson BioTech. This work was supported by a European Research Council Advanced grant (ERC-AdG 742225-IntScOmics) and a Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP award (NWO-CW 714.016.001). P.Z. was funded by HFSP (LT000209/2018-L) and Marie Skłodowska-Curie Actions (798573). V.B. was funded by EMBL LTF (ALTF 1197–2019). The Oncode Institute is partially funded by the KWF Dutch Cancer Society.

Contributions

P.Z., M.B., and A.v.O. designed the project; P.Z. and M.B. developed techniques with the help of F.S.; P.Z and M.B. performed experiments; M.B. and B.A.d.B. wrote the T-ChIC demultiplexing and pre-processing pipeline. P.Z., M.B, and V.B. analysed the data. P.Z., M.B., V.B. and A.v.O, wrote the manuscript.

References

1. Wang, H. *et al.* H3K4me3 regulates RNA polymerase II promoter-proximal pause-release. *Nature* **615**, 339–348 (2023).
2. Ibarra-Soria, X. *et al.* Defining murine organogenesis at single-cell resolution reveals a role for the leukotriene pathway in regulating blood progenitor formation. *Nat. Cell Biol.* **20**, 127–134 (2018).
3. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).
4. Lohoff, T. *et al.* Integration of spatial and single-cell transcriptomic data elucidates mouse organogenesis. *Nat. Biotechnol.* **40**, 74–85 (2022).
5. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).
6. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat. Commun.* **10**, 1–10 (2019).
7. Wu, S. J. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat. Biotechnol.* **39**, 819–824 (2021).

8. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat. Biotechnol.* **39**, 825–835 (2021).

9. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.e11 (2019).

10. Ku, W. L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat. Methods* **16**, 323–325 (2019).

11. Ku, W. L., Pan, L., Cao, Y., Gao, W. & Zhao, K. Profiling single-cell histone modifications using indexing chromatin immunocleavage sequencing. *Genome Res.* **31**, 1831–1842 (2021).

12. Zeller, P. *et al.* Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. *Nat. Genet.* **55**, 333–345 (2023).

13. Xiong, H., Luo, Y., Wang, Q., Yu, X. & He, A. Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. *Nat. Methods* **18**, 652–660 (2021).

14. Zhu, C. *et al.* Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat. Methods* **18**, 283–292 (2021).

15. Xie, Y. *et al.* Droplet-based single-cell joint profiling of histone modifications and transcriptomes. *Nat. Struct. Mol. Biol.* **30**, 1428–1433 (2023).

16. Sun, Z. *et al.* Joint single-cell multiomic analysis in Wnt3a induced asymmetric stem cell division. *Nat. Commun.* **12**, (2021).

17. Salmen, F. *et al.* High-throughput total RNA sequencing in single cells using VASA-seq. *Nat. Biotechnol.* **40**, 1780–1793 (2022).

18. Van Den Brink, S. C. *et al.* Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. *Dev.* **141**, 4231–4242 (2014).

19. Beccari, L. *et al.* Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).

20. Turner, D. A., Rué, P., Mackenzie, J. P., Davies, E. & Martinez Arias, A. Brachyury cooperates with Wnt/β-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. *BMC Biol.* **12**, 8–11 (2014).

21. van den Brink, S. C. *et al.* Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. *Nature* **582**, 405–409 (2020).

22. Veenvliet, J. V. *et al.* Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. *Science (80-.).* **370**, (2020).

23. Rekaik, H. *et al.* Sequential and directional insulation by conserved CTCF sites underlies the Hox timer in stembryos. *Nat. Genet.* **55**, (2023).

24. Flemr, M. & Bühler, M. Single-step generation of conditional knockout mouse embryonic stem cells. *Cell Rep.* **12**, 709–716 (2015).

25. Schmid, M., Rè Se Durussel, T. & Laemmli, U. K. Technique ChIC and ChEC: Genomic Mapping of Chromatin Proteins ble, and significant amounts are lost into the pellet dur-ing centrifugation. While ChIP is highly successful when applied to solu. *Mol. Cell* **16**, 147–157 (2004).

26. Amemiya, H. M., Kundaje, A. & Boyle, A. P. The ENCODE Blacklist: Identification of Problematic Regions of the Genome. *Sci. Rep.* **9**, 1–5 (2019).

27. Rang, F. J. *et al.* Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol. Cell* **82**, 1956-1970.e14 (2022).

Supplemental figures

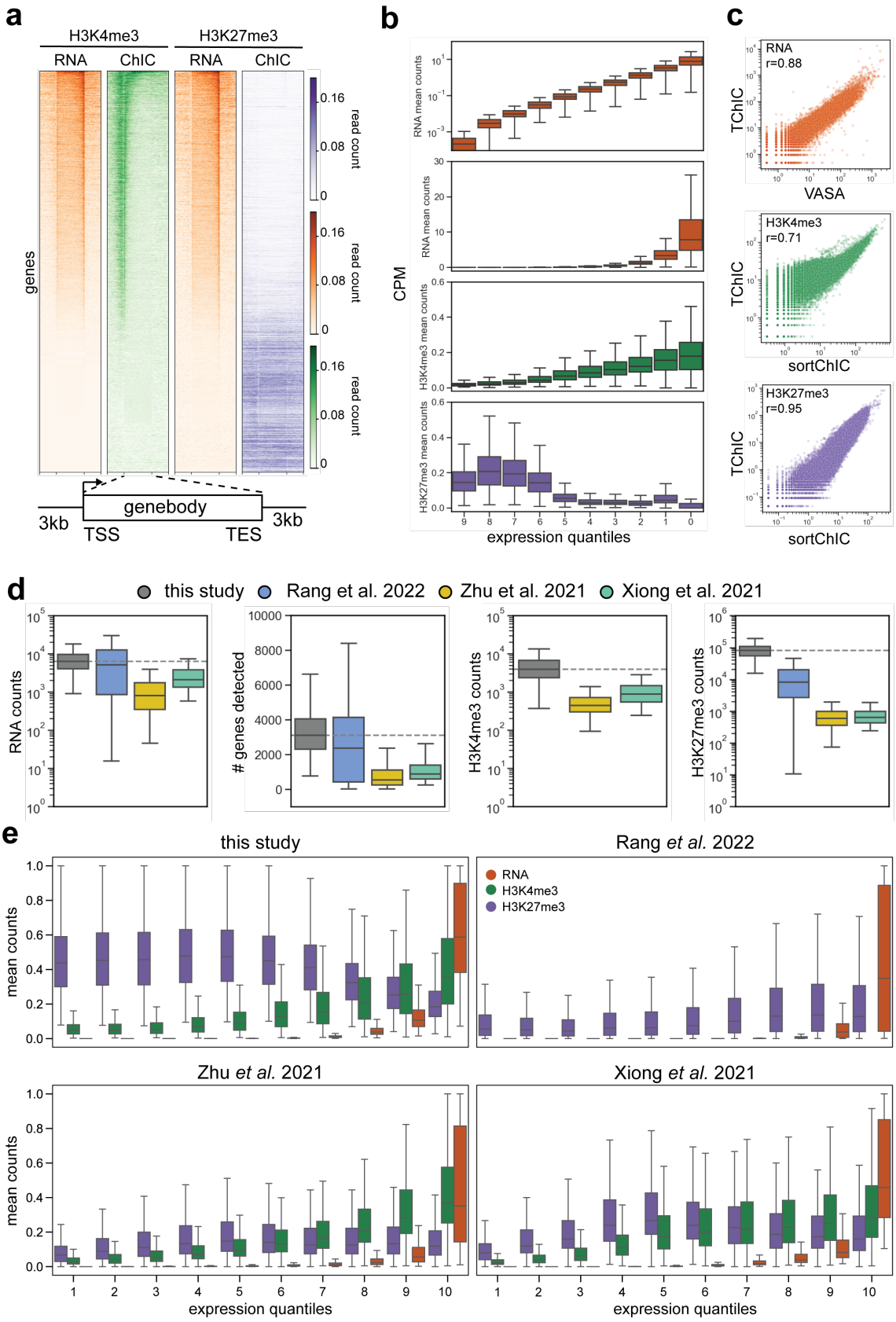


Figure S1 | T-ChIC recovers biologically expected Chromatin distributions and outcompetes alternative methods in sensitivity and selectivity. a) Heatmap of read distribution in a window of 3 kb around protein-coding gene bodies. Pseudo-bulk results from one plate of K562 cells per histone modification are shown. b) Boxplots of single-cell read averages over genes split into quantiles of log10 expression. From top to bottom, the following measurements are shown: RNA log10 counts (orange), RNA counts (orange), H3K4me3 (green) and H3K27me3 (purple). c) Scatter plots showing Pearson correlation of T-ChIC to VASA-seq and published ChIC results for K562 cells. shown are RNA (top, orange), H3K4me3 (middle, green) and H3K27me3 (bottom, purple). d-e) Comparison of RNA and chromatin sensitivity (d) and selectivity (e) with published single-cell methods that allow simultaneous measurement of transcription and histone modification distribution. Rang et al. is based on DamID, which uses transgenic cell lines expressing Dam fusion proteins in combination with life cell sorting into 384 well plates. Zhu et al. and Xiong et al. are based on cut&tag (Pa-Tn5) and the library prep is performed by combinatorial barcoding. d) From left to right, boxplots of uniquely mapped transcript reads, number of genes with at least one read, uniquely mapped H3K4me3 and H3K27me3 reads are shown. e) Boxplots of single cell read averages over genes split into quantiles of log10 expression are shown for RNA (orange), H3K4me3 (green) and H3K27me3 (purple). Values are scaled per modality and one plot is shown per study.

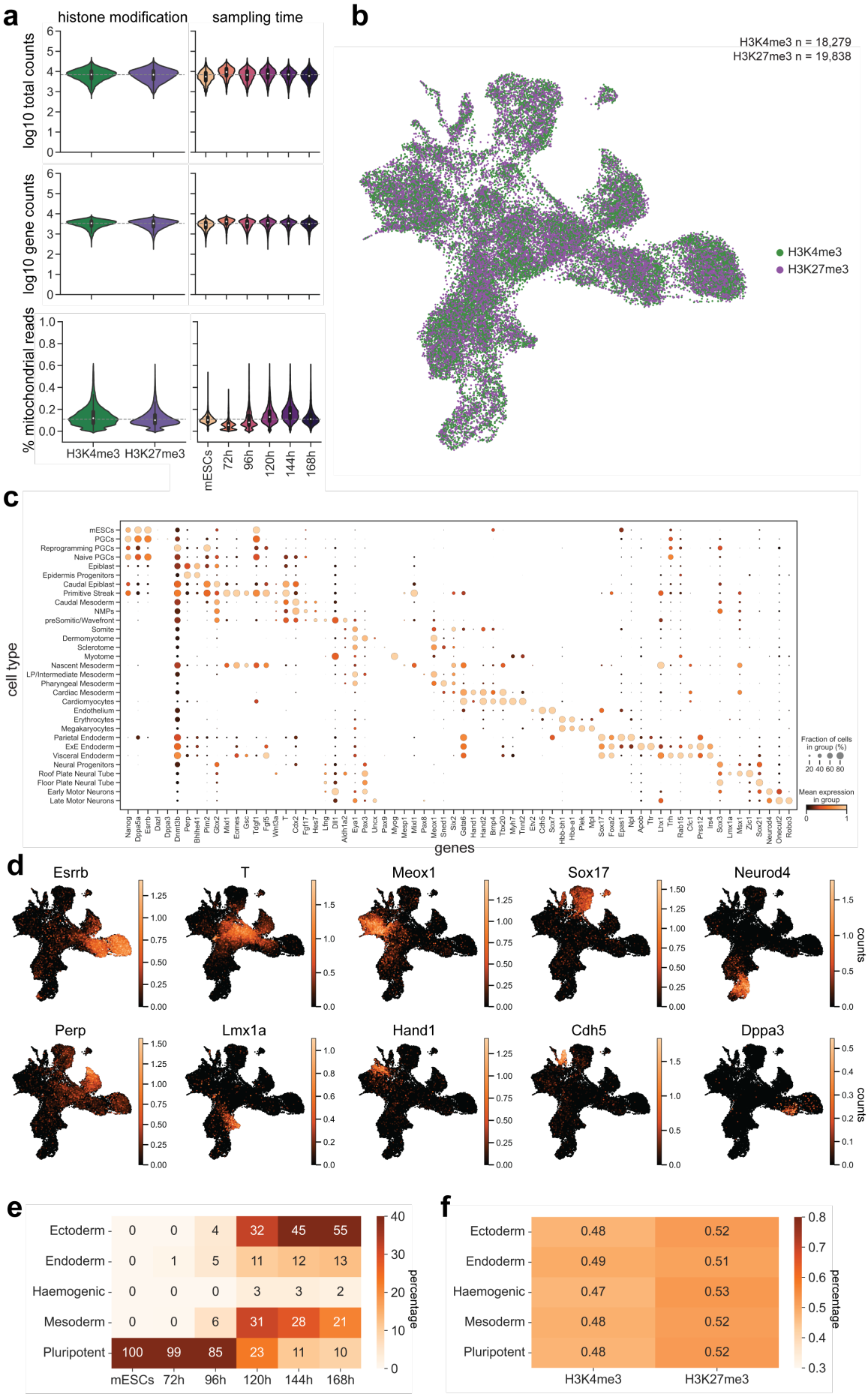
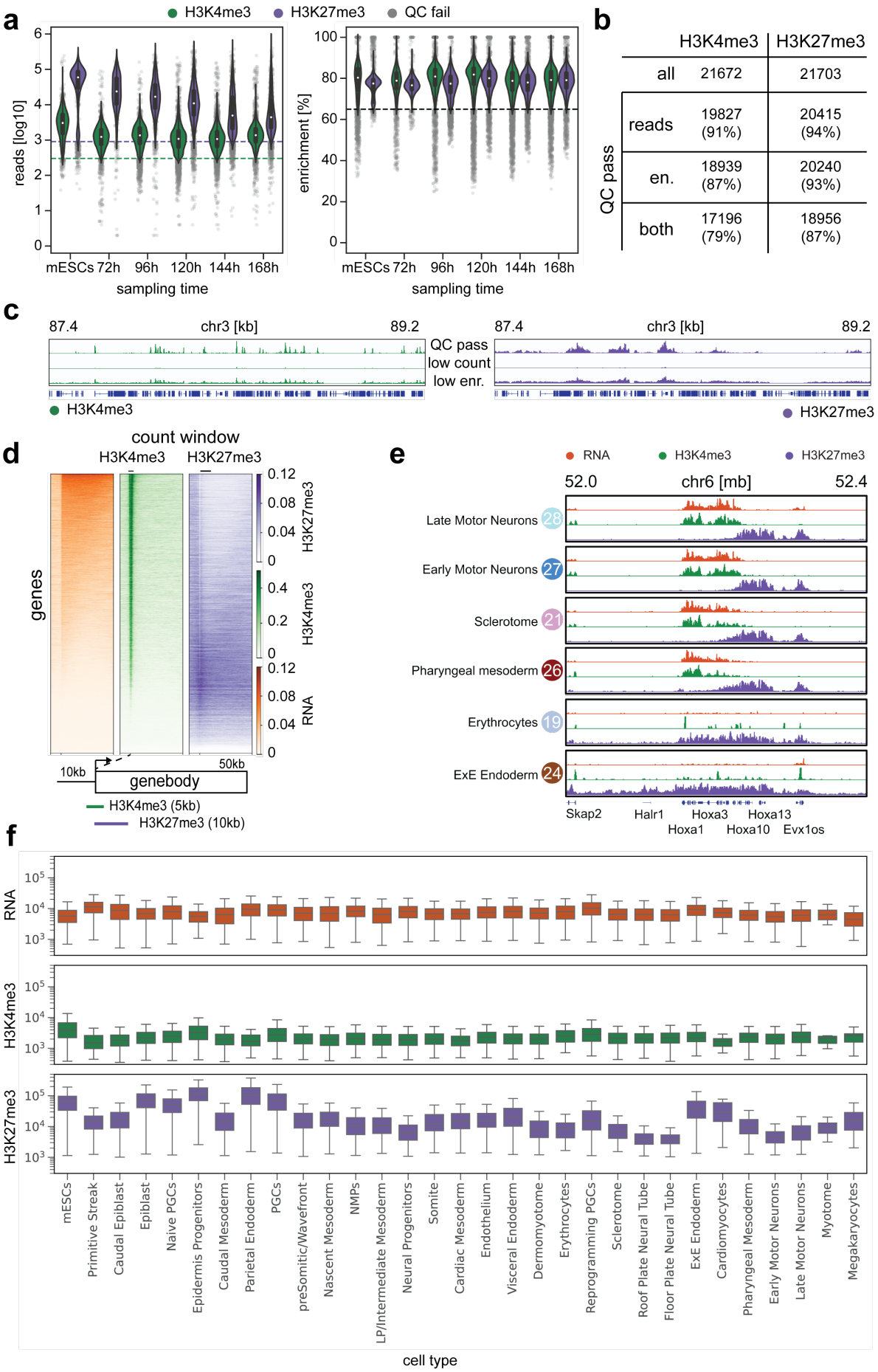

a
histone modification
sampling time
log10 total counts
log10 gene counts
% mitochondrial reads
H3K4me3
H3K27me3
mESCs
72h
96h
120h
144h
168h
b
H3K4me3 n = 18,279
H3K27me3 n = 19,838
H3K4me3
H3K27me3
c
cell type
genes
Fraction of cells in group (%)
Mean expression in group
d
Esrrb
T
Meox1
Sox17
Neurod4
Perp
Lmx1a
Hand1
Cdh5
Dppa3
counts
e
Ectoderm
Endoderm
Haemogenic
Mesoderm
Pluripotent
percentage
f
Ectoderm
Endoderm
Haemogenic
Mesoderm
Pluripotent
H3K4me3
H3K27me3
percentage

Figure S2 | Gastruloid transcript QC. a) Transcriptome QC split by histone mark profiled simultaneously (left) or sampling time point (right). The following QC metrics are shown: log10 values of total counts per cell (top), log10 values of total genes recovered per cell (middle), and percentage of mitochondrial reads per cell (bottom). b) Transcriptome UMAP coloured by the histone modifications profiled together with RNA: H3K4me3 (green, n = 18,279) and H3K27me3 (purple, n = 19,838). c) Dot plot showing expression of marker genes used for cell type annotation in Figure 2C. Dot colour represents the mean expression of marker genes per specified cluster, and dot size represents the fraction of cells within the group that show expression of the gene. d) Transcriptome UMAPs showing expression levels and patterns of selected marker genes: *Esrrb*, pluripotency; *T*, symmetry breaking and elongation factor; *Meox1*, early mesoderm; Sox17, endoderm; *Neurod4*, ectoderm; *Perp*, epiblast; *Lmx1a*, neural tube; *Hand1*, heart progenitors; *Cdh5*, endothelium; *Dppa3*, PGCs. e) Percent of lineages recovered for each sampling time point. f) Percent of lineages recovered for each chromatin modification profiled.



		H3K4me3	H3K27me3
	all	21672	21703
QC pass	reads	19827 (91%)	20415 (94%)
	en.	18939 (87%)	20240 (93%)
	both	17196 (79%)	18956 (87%)

Figure S3 | Gastruloid Chromatin QC. a) Chromatin QC split by sampling time point and histone modification profiled. Left: log10 uniquely mapped reads per cell, right: fraction of reads assigned to genomic bins enriched in local neighbourhoods. b) Number and percentage of cells passing QC filtering. c) Coverage plots showing the signal obtained from 1000 randomly selected cells passing QC, or failing due to low reads or enrichment. d) Heatmap of read distribution 10 kb upstream and 50 kb into the gene body of gastruloid pseudo-bulk data. RNA (orange, left), H3K4me3 (green, middle) and H3K27me3 (purple, right) are shown. Lines on top indicate counting windows for gene-based analysis. e) Pseudo-bulk representation per cluster of single-cell coverage plots shown in 3c. A 400 kb window on chromosome 6 covering the Hoxa cluster is shown. f) Log10 uniquely mapped reads per cell are shown, split per cell type for RNA (orange, top), H3K4me3 (green, middle) and H3K27me3 (purple, bottom).