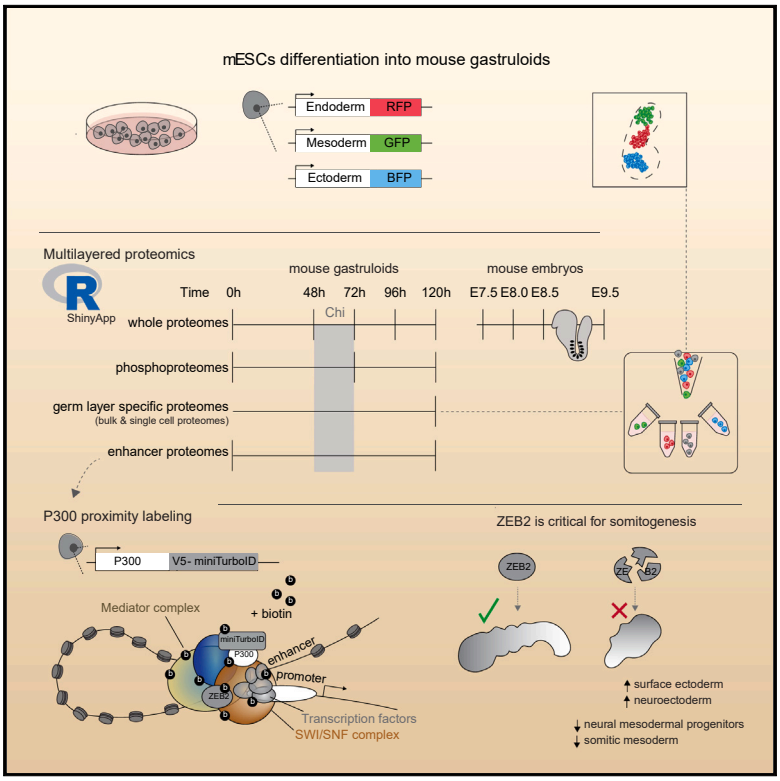
Cell Stem Cell

Resource

Deciphering lineage specification during early embryogenesis in mouse gastruloids using multilayered proteomics

Graphical abstract



Highlights

- Mouse gastruloid formation is associated with global rewiring of the (phospho) proteome
- The three germ layers exhibit distinct protein expression profiles
- P300 proximity labeling reveals global enhancer interactomes
- ZEB2 plays a key role in mouse and human somitogenesis

Authors

Suzan Stelloo, Maria Teresa Alejo-Vinogradova, Charlotte A.G.H. van Gelder, ..., Maarten A.F.M. Altelaar, Harmjan R. Vos, Michiel Vermeulen

Correspondence

s.stelloo@science.ru.nl (S.S.), michiel.vermeulen@science.ru.nl (M.V.)

In brief

Stelloo et al. applied mass spectrometry-based proteomics technologies on gastruloids and mouse embryos to characterize global protein expression dynamics during early mouse embryonic development. Using P300 proximity labeling, they identified gastruloid-specific enhancer-binding proteins. Subsequent investigation revealed a role for ZEB2, a gastruloid-specific transcription factor, during mouse and human somitogenesis.

Stelloo et al., 2024, Cell Stem Cell *31*, 1072–1090
July 5, 2024
https://doi.org/10.1016/j.stem.2024.04.017

CellPress

CellPress
OPEN ACCESS

Cell Stem Cell

Resource

Deciphering lineage specification during early embryogenesis in mouse gastruloids using multilayered proteomics

Suzan Stelloo,[1,*] Maria Teresa Alejo-Vinogradova,[1,9] Charlotte A.G.H. van Gelder,[2,9] Dick W. Zijlmans,[1] Marek J. van Oostrom,[3] Juan Manuel Valverde,[4,5] Lieke A. Lamers,[1] Teja Rus,[1] Paula Sobrevals Alcaraz,[2] Tilman Schäfers,[6] Cristina Furlan,[7] Pascal W.T.C. Jansen,[1] Marijke P.A. Baltissen,[1] Katharina F. Sonnen,[3] Boudewijn Burgering,[2] Maarten A.F.M. Altelaar,[4,5] Harmjan R. Vos,[2] and Michiel Vermeulen[1,8,10,*]

[1]Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Oncode Institute, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands
[2]Molecular Cancer Research, Center for Molecular Medicine, Oncode Institute, University Medical Center Utrecht, Utrecht University, 3584 CG Utrecht, the Netherlands
[3]Hubrecht Institute, KNAW (Royal Netherlands Academy of Arts and Sciences), University Medical Center Utrecht, 3584 CT Utrecht, the Netherlands
[4]Biomolecular Mass Spectrometry and Proteomics, Bijvoet Center for Biomolecular Research and Utrecht Institute for Pharmaceutical Sciences, Utrecht University, 3584 CA Utrecht, the Netherlands
[5]Netherlands Proteomics Center, 3584 CH Utrecht, the Netherlands
[6]Department of Molecular Developmental Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University Nijmegen, 6525 GA Nijmegen, the Netherlands
[7]Laboratory of Systems and Synthetic Biology, Wageningen University & Research, 6708 WE Wageningen, the Netherlands
[8]Division of Molecular Genetics, Netherlands Cancer Institute, 1066 CX Amsterdam, the Netherlands
[9]These authors contributed equally
[10]Lead contact
*Correspondence: s.stelloo@science.ru.nl (S.S.), michiel.vermeulen@science.ru.nl (M.V.)
https://doi.org/10.1016/j.stem.2024.04.017

SUMMARY

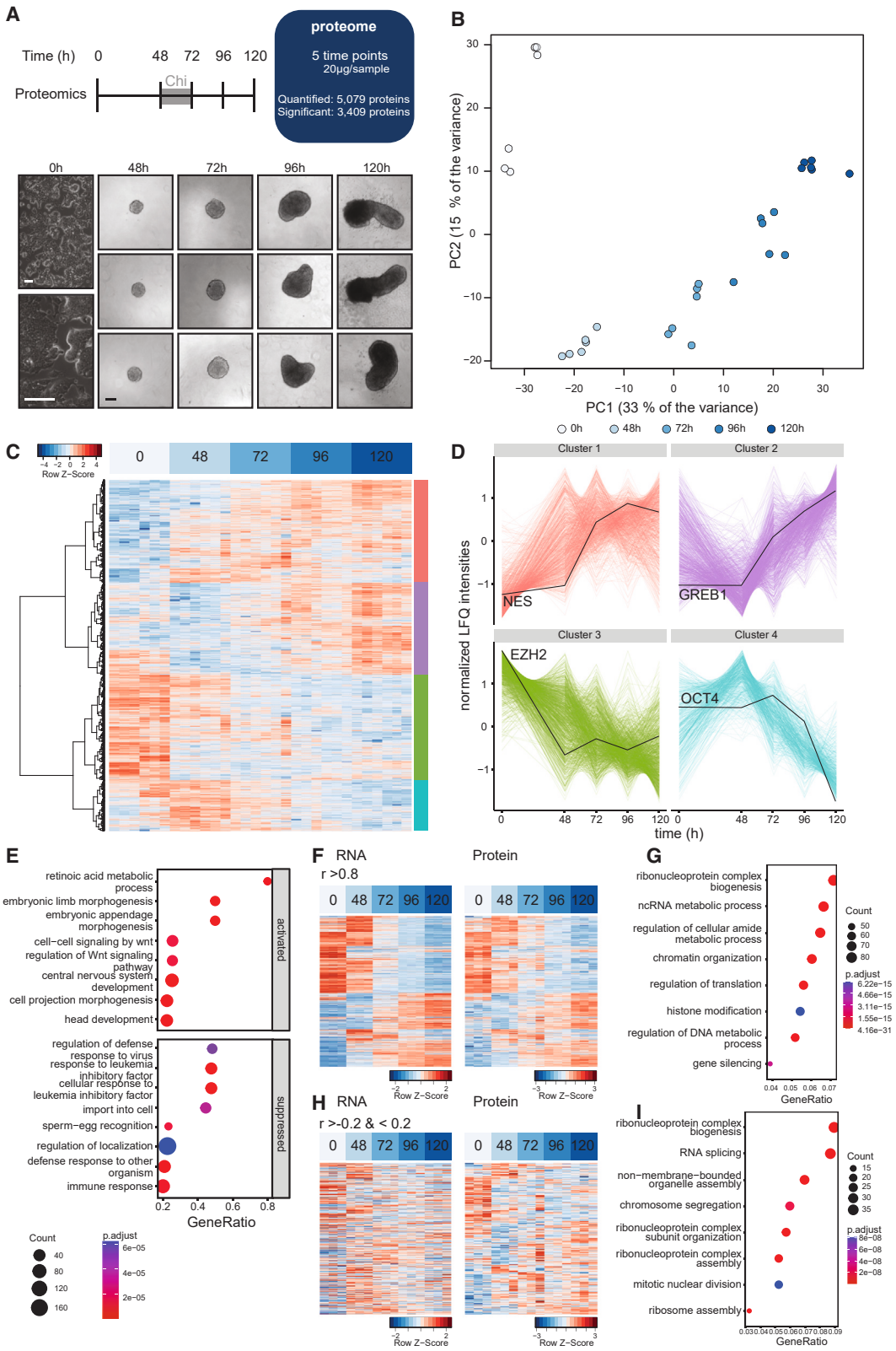
Gastrulation is a critical stage in embryonic development during which the germ layers are established. Advances in sequencing technologies led to the identification of gene regulatory programs that control the emergence of the germ layers and their derivatives. However, proteome-based studies of early mammalian development are scarce. To overcome this, we utilized gastruloids and a multilayered mass spectrometry-based proteomics approach to investigate the global dynamics of (phospho) protein expression during gastruloid differentiation. Our findings revealed many proteins with temporal expression and unique expression profiles for each germ layer, which we also validated using single-cell proteomics technology. Additionally, we profiled enhancer interaction landscapes using P300 proximity labeling, which revealed numerous gastruloid-specific transcription factors and chromatin remodelers. Subsequent degron-based perturbations combined with single-cell RNA sequencing (scRNA-seq) identified a critical role for ZEB2 in mouse and human somitogenesis. Overall, this study provides a rich resource for developmental and synthetic biology communities endeavoring to understand mammalian embryogenesis.

INTRODUCTION

In recent years, the significance of embryo-like structures generated from mouse embryonic stem cells (mESCs) has become increasingly apparent.[1–3] One of these models, gastruloids, recapitulate critical aspects of early embryogenesis, including the formation of the three germ layers. Thus far, the specification of different cell types within gastruloids has extensively been characterized using imaging, bulk and single-cell RNA sequencing (scRNA-seq), and epigenomics technologies,[4–9] while proteomics-based studies of early mammalian development are scarce.[10–13] Proteomics-based studies are important because the integration of transcriptome and proteome data has revealed that global protein and mRNA expression levels do not correlate well.[14–17] In addition, posttranslational modifications can significantly impact protein levels and function. However, to date, mass spectrometry-based proteomics analyses on post-implantation embryos or stem-cell-based models of embryo development (stembryos) have not been conducted.

In this study, we profiled the dynamic (phospho) proteome of mESCs during gastruloid differentiation and in equivalent mouse embryo stages. Furthermore, using a triple fluorescent germ

CellPress
OPEN ACCESS



(legend on next page)

layer reporter line, we sorted mesoderm, ectoderm, and endoderm from gastruloids and characterized their proteome. We further implemented the CellenOne platform, which allowed us to quantify single-cell proteomes in gastruloids to a depth of ~1,500 proteins per cell. To identify chromatin-associated proteins and transcription factors that may be involved in lineage specification, we made use of P300 proximity labeling in mESCs and gastruloids. Gastruloid-specific P300-proximal transcription factors were functionally investigated using a dTAG degron approach. Targeted degradation of ZEB2 during gastruloid formation revealed that a loss of ZEB2 impairs differentiation to neuromesodermal progenitor (NMP) cells, which concomitantly results in an increased proportion of ectoderm cells compared with wild-type cells. Collectively, integrative multi-omics and perturbation-based analyses provide insights into lineage-specific transcriptional programs during early mammalian embryogenesis. These omics datasets can be explored through a Rshiny web application: https://mouse-gastruloids-omics.shinyapps.io/gtlshiny/.

RESULTS

Time-resolved proteome of mESC differentiation toward gastruloids

To assess proteome dynamics during gastruloid formation, we conducted two biological experiments, collecting samples at five time points from undifferentiated mESCs to 120 h gastruloids, following a previously described protocol (Figure 1A).[18] Label-free quantification (LFQ) mass spectrometry identified 5,079 proteins across both experiments. An overlap of 87% and 95% proteins was observed across the two biological replicates (Figure S1A). The correlation between the samples was high (Pearson $r > 0.9$) (Figure S1B). Principal component analysis (PCA) separates the samples by time points, revealing distinct protein expression patterns during differentiation (Figure 1B). To investigate protein expression dynamics in time, we performed differential expression analysis using hierarchical clustering (Figures 1C and 1D). Differentially expressed proteins display a similar protein abundance distribution compared with all quantified proteins, and highly abundant proteins are enriched in translation and RNA processing (Figure S1C). Proteins in cluster 3 were mainly downregulated upon differentiation and include SOX2 and members of the PRC2 complex. Proteins in cluster 4, such as OCT4 and DNMT3A, showed the highest expression in 48 and 72 h gastruloids. Cluster 1 contains proteins (e.g., CTNNB1, NES, and T) that get upregulated during gastruloid differentiation, while proteins in cluster 2 were downregulated at 48 h followed by a strong induction at later time points. As expected, Gene Ontology (GO) enrichment analysis revealed suppression of "cellular response to leukemia inhibitory factor (LIF)" in 120 h gastruloids, while animal organ morphogenesis, Wnt signaling, and terms related to neuron development were enriched (Figure 1E).

To explore the relationship between global mRNA and protein expression dynamics, we generated RNA sequencing (RNA-seq) libraries of all time points. Comparison of our RNA-seq data to previously published gastruloid RNA-seq data[4] revealed highly concordant transcriptome profiles (Figures S1D and S1E). The expression of specific genes associated with distinct embryonic stages was evident (Figure S1F). A subset of differentially expressed proteins showed similar dynamics at the protein and transcript level (Spearman correlation of 0.6, Figure S1G), which is consistent with previous studies.[14,15] Genes that show a strong correlation between protein and mRNA expression dynamics (Spearman mRNA-protein correlation > 0.8, $n = 1,220$) are enriched for GO terms associated with metabolic and epigenetic processes (Figures 1F and 1G). By contrast, genes showing a low correlation ($r > -0.2$ and < 0.2) are enriched for biological processes associated with ribonucleoprotein complex and cell cycle ($n = 418$; Figures 1H and 1I; Table S1). The identification of genes that show a discordance between mRNA and protein expression levels during differentiation represents a valuable resource for studying the intricate regulation of mRNA and protein expression dynamics during early embryonic development.

Comparison of temporal protein expression profiles between mouse gastruloids and natural embryos

To date, proteomics-based profiling of mouse embryos has exclusively been performed using pre-implantation embryos. Given that mouse gastruloids mimic key aspects of post-implantation development, comparative proteomics-based profiling of gastruloids and post-implantation mouse embryos is imperative (Figure 2A). We selected embryonic day (E)7.5 to

Figure 1. Time-resolved proteome profiling of mESC differentiation toward 120 h gastruloids

(A) Schematic overview. Proteomes were profiled at 0, 48, 72, 96, and 120 h during gastruloid differentiation. Phase contrast images of representative mESC cultures and gastruloids at different time points. The lower image of mESCs (0 h) is a higher magnification of the top image. Scale bars, 100 μm. Summary of the number of proteins quantified and significant using ANOVA multiple sample test with Benjamini-Hochberg correction.

(B) PCA using the LFQ intensities of overlapping proteins (5,079). Different blue tints denote the different time points.

(C) Heatmap of significantly expressed proteins across time points. Hierarchical cluster assignment is denoted by the right-side color bar on the heatmap. Cluster 1, 994 proteins; cluster 2, 899 proteins; cluster 3, 1,021 proteins; and cluster 4, 495 proteins.

(D) Plot of differentially expressed proteins per cluster. Scaled mean expression for each protein is shown and one protein per cluster is highlighted in black.

(E) gseGO results showing the top 8 suppressed and activated biological processes in 120 h gastruloids. Dot size indicates the number of enriched genes in each term. Dot color represents the adjusted p value.

(F) Row-matched heatmaps showing the relative mRNA (left) and protein (right) expression of significantly changing proteins for which sample-wise mRNA-protein correlation was higher than 0.8.

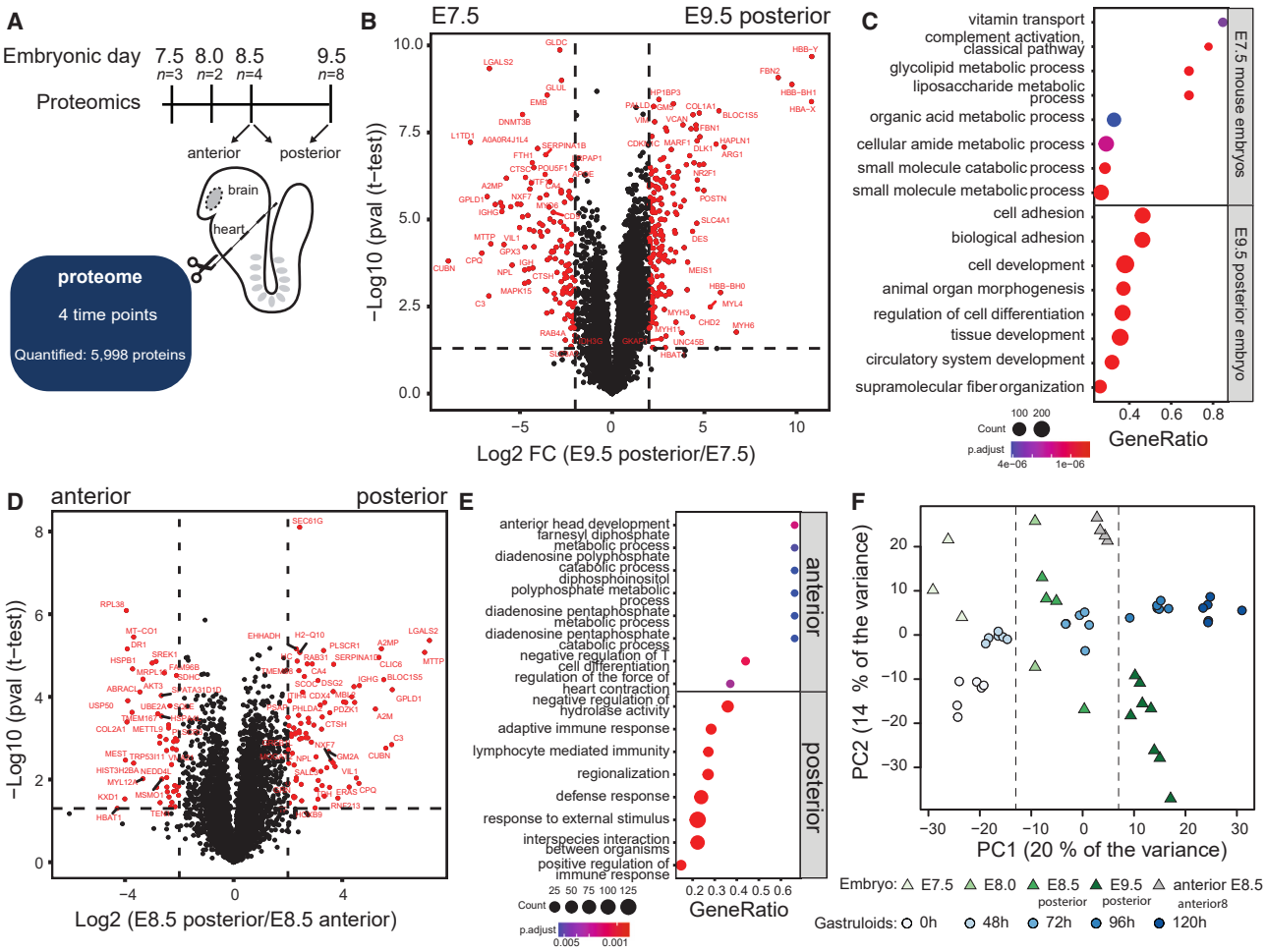
(G) GO terms enriched for transcripts and proteins exhibiting high sample-wise mRNA-protein correlation (shown in F).

(H) Row-matched heatmaps showing the relative mRNA (left) and protein (right) expression of significantly changing proteins for which sample-wise mRNA-protein correlation was lower than 0.2.

(I) GO terms enriched transcripts and proteins exhibiting low sample-wise mRNA-protein correlation (shown in H).

See also Figure S1 and Table S1.

CellPress
OPEN ACCESS



(legend on next page)

E9.5 embryos as previous transcriptome profiling revealed that gastruloid differentiation roughly corresponds to cell populations from E6.5 (24 h gastruloids), E7.5–8.0 (72 h gastruloids), and E8.5–9.5 (120 h gastruloids) embryos. Given that gastruloids exhibit developmental features and organization comparable to the posterior part of the embryo, mouse E8.5 and 9.5 embryos were dissected into anterior and posterior portions (lacking the head region) or only posterior portions, respectively. In total, we identified 5,998 proteins in mouse embryos. PCA of all quantified proteins revealed a time-dependent separation on component 1 accurately reflecting embryonic developmental time (Figure S2A). The observed differential expression of embryonic globin proteins and myosin heavy chains proteins, particularly expressed at E9.5, is indicative of processes associated with erythropoiesis and cardiac muscle development, respectively (Figures 2B and 2C). Mining public scRNA-seq data from E6.5 to 8.5 mouse embryos confirmed increased expression of embryonic globin genes, particularly at later time points during embryonic development (Figure S2B). In E7.5 embryos, we observed the expression of pluripotency-associated proteins such as POU5F1, DNMT3B, and UTF1. Principal component 2 captures the differences between anterior and posterior samples (Figure S2A). Differentially expressed proteins include known posterior (CDX4 and HOXB9) and anterior proteins (DDX10) (Figure 2D). Particularly significant is the anterior expression of RPL38, a protein involved in axial skeletal patterning by specifically controlling the translation of Hox mRNA (Figure 2E).[19]

We then compared the proteomic profiles of mouse embryos with gastruloids. In total, 4,736 proteins were detected in both mouse gastruloids and mouse embryo datasets. Proteins detected exclusively in mouse embryos are enriched for GO terms related to pattern specification processes, immune response, and heart development (Figures S2C and S2D). PCA illustrated separation based on developmental stage, E7.5 mouse embryo proteomes most closely resemble mESCs and 48 h gastruloid proteomes, E8.0–8.5 mouse embryos resemble 72 h gastruloids, and E9.5 mouse embryos align best with 96–120 h gastruloids (Figures 2F and S2E). Similar changes in the expression of specific proteins associated with pluripotency, neuronal development, and somitic differentiation were observed (Figures 2G and 2H). In conclusion, proteomic profiling reveals a good overlap in protein expression profiles between mouse gastruloids and mouse embryos.

Time-resolved phosphoproteome of mESC differentiation toward gastruloids

To identify dynamic signaling pathways during gastruloid formation, we profiled the phosphoproteome of undifferentiated mESCs, 72 h gastruloids, and 120 h gastruloids (Figure 3A). We identified 20,380 phosphorylation sites, of which 11,695 sites were identified in at least three replicates of one condition with a localization probability score > 0.75 (Figures 3B and 3C). Nearly 6,000 significantly regulated phosphorylation sites were identified on 2,366 unique proteins, the majority of which are phosphorylated in undifferentiated mESCs (Figure 3D). We next assessed the overlap between the proteome and phosphoproteome to account for those changes in phosphosite abundances due to changes in total protein abundance. ~30% of the phosphorylated proteins were not quantified in whole proteomes. Among the overlapping proteins, 1,151 proteins show differential expression and phosphorylation, while 412 proteins undergo phosphorylation changes, irrespective of changes in protein abundance (Figure S3A; Table S2). Analyzing signaling pathways in gastruloids revealed that shear stress signaling is repressed in 120 h gastruloids while signaling pathways related to RNA splicing and embryonic development are active (Figures S3B and S3C). Interestingly, differential splicing patterns are known to be associated with embryonic stem cell differentiation.[20,21] Next, we employed the kinase enrichment analysis (KEA2) tool to identify upstream kinases responsible for protein phosphorylations in mESCs and gastruloids. This revealed increased activity of CDK2, CDK1, GSK3B, P38-MAPK14, and DYRK2 in undifferentiated mESCs compared with gastruloids (Figure 3E). The decreased activity of GSK3B is expected given that gastruloids are cultured with CHIR99021, a GSK3B inhibitor, for 24 h (between time points 48 and 72 h).

We identified more phosphorylated transcription factors in 120 h gastruloids compared with 72 h gastruloids and undifferentiated mESCs (9.2% versus 4.1% and 3.0%, respectively; Figure 3F). Most of these belong to the C2H2-type zinc-finger transcription factor family, which is the largest transcription factor family in the mouse genome (Figures 3G and 3H). For example, phosphorylated residues for members of the Krüppel-like family (KLF), known pluripotency regulators,[23] were detected in undifferentiated mESCs. We also observed several proteins to be phosphorylated on multiple sites, including SALL1. Phosphorylation of SALL1 inhibits its interaction with NuRD and affects its

Figure 2. Proteomic profiling of mouse embryos and comparison to mouse gastruloids

(A) Schematic overview. Proteomes were profiled across four embryonic stages.

(B) Volcano plot showing the protein expression level changes between mouse embryos at E7.5 and 9.5. The $\log_2$ fold change of LFQ intensities was plotted against the adjusted $-\log_{10}$ (p value). Proteins with a $\log_2$ fold change > 2 and p value < 0.05 are colored in red.

(C) gseGO results showing the top 8 enriched biological processes in E7.5 or 9.5 mouse embryos. Dot size indicates the number of enriched genes in each term. Dot color represents the adjusted p value.

(D) Volcano plot showing the protein expression level changes between anterior and posterior samples from E8.5 mouse embryos. The $\log_2$ fold change of LFQ intensities was plotted against the adjusted $-\log_{10}$ (p value). Proteins with a $\log_2$ fold change > 2 and p value < 0.05 are colored in red.

(E) gseGO results showing the top 8 enriched biological processes in anterior or posterior halves of E8.5 mouse embryos. Dot size indicates the number of enriched genes in each term. Dot color represents the adjusted p value.

(F) PCA using the LFQ intensities of overlapping proteins (4,736 proteins). The different blue tints denote the different time points of gastruloids, while the green tints correspond to the different mouse embryo stages. Different shapes distinguish mouse embryos from gastruloids.

(G) Heatmap of scaled expression of proteins associated with embryonic development in mouse gastruloids.

(H) Heatmap of scaled expression of proteins associated with embryonic development in mouse embryos.

See also Figure S2.

CellPress
OPEN ACCESS

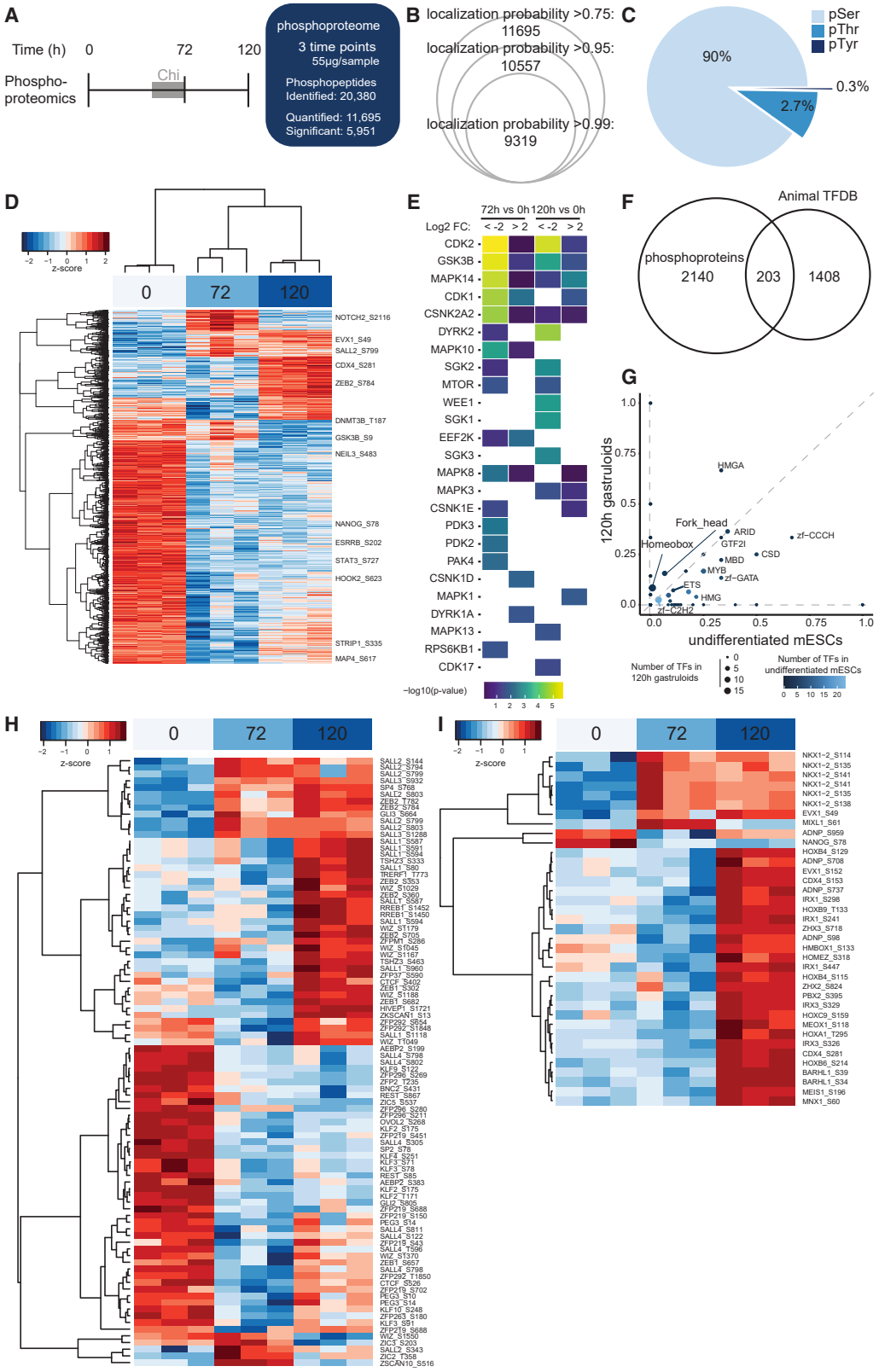


Figure 3. Time-resolved phosphoproteome profiling of mESC differentiation toward 120 h gastruloids

(A) Schematic overview. Phosphoproteomes were profiled at 0, 72, and 120 h during mESC differentiation toward gastruloids. Summary of the number of phosphopeptides identified, quantified, and significant as assessed using ANOVA multiple sample test with Benjamini-Hochberg correction.

(legend continued on next page)

transcriptional ability.[24] Additionally, we identified differential phosphorylation on CTCF (S402), which is known to reduce the affinity of CTCF for DNA.[25] In contrast to stem-cell-specific transcription factor phosphorylation, several members of the homeodomain-containing transcription factor family were found to be specifically phosphorylated in 120 h gastruloids (Figures 3G and 3I). Exceptions include NANOG, phosphorylated in mESCs and MIXL1, phosphorylated in 72 h gastruloids, which corresponds to their total protein expression dynamics (Figure S3D, NANOG only detected at 0 h but not quantified). The preferential phosphorylation of homeodomain-containing transcription factors in late-stage gastruloids is consistent with their established role as important regulators of cell fate switches during early mammalian embryogenesis.

The three germ layers exhibit distinct protein expression profiles

To identify germ layer-specific proteomes, we generated a reporter line in which the germ layers are distinguishable by different fluorophores (Figure 4A). We inserted TagBFP, encoding for a blue fluorescent protein, at the C terminus of the MT1 gene in a previously established mESC line expressing BRA-GFP and SOX17-mStrawberry (RFP)[26] (Figure S4A). In gastruloids, the BRA-GFP gene is expressed in mesodermal cells (mesodermal precursor cells and presomitic mesoderm [PSM]), SOX17-RFP in endoderm, and MT1-BFP in primordial-germ-cell-like or extra-embryonic ectoderm (here referred to as ectoderm for simplicity) (Figure S4B). Triple-negative cells mainly correspond to differentiated mesoderm and neuroectoderm. Analysis of 120 h gastruloids revealed ~4% RFP^+ cells, ~6% BFP^+ cells, and >20% GFP^+ cells (Figure 4B), and subsequent RNA expression analysis of the fluorescence-activated cell sorting (FACS)-purified populations confirmed expression of germ layer-specific marker genes (Figure S4C). Using proteomics, we identified 4,795 proteins of which 2,708 were differentially expressed between the four cell populations (Figure 4C, ANOVA test, p value < 0.05). Proteomic profiling of FACS-sorted cell populations from the dual reporter cell line revealed that the introduction of the third reporter had no discernible effects on the proteomes of the RFP^+ and GFP^+ cell populations (Figures S4D and S4E). PCA of the 1,000 most variable-expressed proteins revealed distinct clustering (Figure S4F). The triple-negative population and the BRA-GFP+ cells cluster close to each other, which is consistent with the fact that the majority of cell types in gastruloids correspond to mesodermal subtypes (clusters 1–7, Figure S4B). To further explore the cell-type-specific proteomes, we plotted the differentially expressed proteins, which revealed differential expression of known markers of endoderm (CLDN6 and FOXA2), ectoderm (POU5F1 and UTF1), and mesoderm (MEST and ALDH1A2) (Figure 3C). Next, we explored protein expression of the differentially expressed genes for clusters 6/7 (mesoderm), 10 (endoderm), and 12 (ectoderm) as previously identified with scRNA-seq of 120 h gastruloids.[5] The majority of the corresponding proteins are upregulated in the expected cell populations (Figures S4G–S4I). Additionally, we explored the biological processes and signaling pathways associated with the four isolated cell populations (Figures S4J–S4M). As expected, mesodermal and retinoic acid signaling are enriched in BRA-GFP^+ cells. Gene sets related to cell junctions and cell motility are enriched in SOX17-RFP^+ cells and terms related to neurodevelopmental processes in the triple-negative population. Furthermore, we observed the enrichment of metabolic processes, response to LIF, and histone methyltransferase complex in MT1-BFP^+ cells. Next, we further assessed which members of other protein complexes are co-regulated. Besides increased protein abundances for members of the PRC2 complex, minichromosome maintenance complex, prefoldin complex, T complex protein 1, and eIF3 complex were observed in MT-BFP+ cells (Figure 4D). For SOX17-RFP^+ cells, we observed an increased abundance for the members of the mitochondrial membrane ATP synthase complex, vacuolar ATPase complex, and actin-related protein 2/3 complex (Figure 4E).

SCP captures germ layer-specific cellular heterogeneity

We next applied single-cell proteomics (SCP) on SOX17-RFP+, BRA-GFP+, MT1-BFP+ cells, and undifferentiated mESCs (Data S1). In total, 560 single cells passed quality control, resulting in 2,259 identified proteins with at least two unique peptides, with an average of 1,541 proteins per single cell (Data S1). PCA of 675 differentially expressed proteins (ANOVA, q < 0.01) revealed clear separation of the different cell populations on the first component (Figure 5A), and this separation is independent of the tandem mass tag (TMT) label (Figure S5A). However, the clustering does reflect the TMT multiplexing to some extent, possibly due to the non-random distribution of the cell types on the proteoCHIPs (Figures S5B and S5C). Among the top 15 proteins contributing to principal component 1 are proteins involved in RNA processing (e.g., NSRP1, GTPBP1, and UTP6) and transcriptional regulation, including CHD8 and CARM1 (Table S3). Expression of selected proteins is displayed on PCA projections (Figure 5B). The predominant expression of ANXA5 in SOX17-RFP^+ cells is concordant with scRNA-seq

(B) Number of phosphosites detected according to the localization probability score, which refers to the likelihood that a particular phosphorylation site on a peptide is correctly assigned to a specific amino acid residue.
(C) Distribution of phosphosites identified among serine, threonine, and tyrosine residues.
(D) Hierarchical clustering of significantly changing phosphosites (ANOVA, Benjamini-Hochberg correction, FDR 0.05).
(E) Predicted upstream kinases based on significantly changing phosphosites (t test 72 h gastruloids or 120 h gastruloids versus 0 h, $\log_2$ FC >2 and p value < 0.05).
(F) Venn diagram illustrating the overlap between significantly detected phosphoproteins and 1,611 transcription factors from the AnimalTFDB v4.0 database.[22]
(G) Scatterplot showing the fraction of transcription factors identified as significantly phosphorylated in 120 h gastruloids and mESCs. Dot size depicts the number of phosphorylated transcription factors in each transcription factor family for 120 h gastruloids. Dot color depicts the number of phosphorylated transcription factors in each transcription factor family for undifferentiated mESCs.
(H) Heatmap of significantly detected phosphosites on transcription factors belonging to the C2H2 zinc-finger (C2H2-ZF) family of transcription factors.
(I) Heatmap of significantly detected phosphosites on transcription factors belonging to the homeodomain transcription factor family.
See also Figure S3 and Table S2.

CellPress
OPEN ACCESS

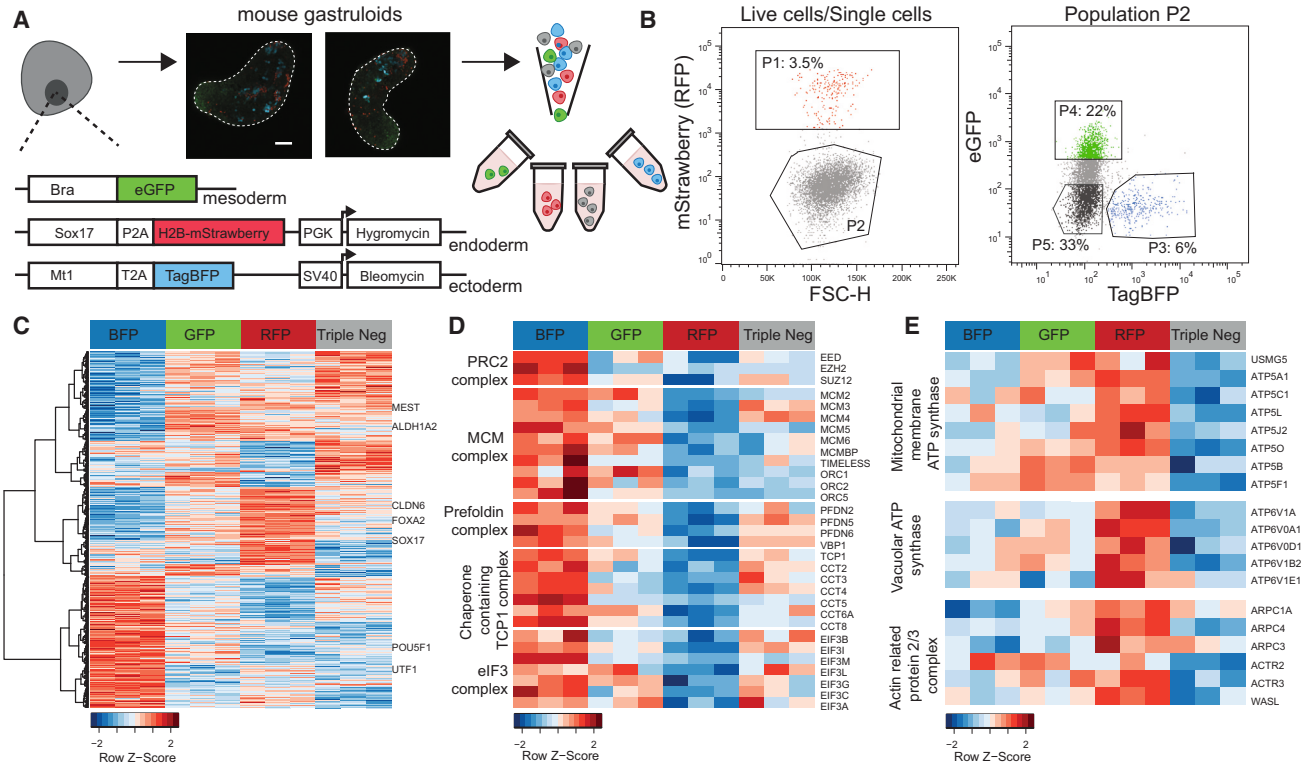


Figure 4. The three germ layers exhibit distinct protein expression profiles

(A) Schematic representation of the fluorescent reporter mESC line. T2A-TagBFP followed by a BGH poly(A) signal, and a bleomycin resistance cassette was inserted at the endogenous *Mt1* gene locus in the dual BRA-GFP, SOX17-RFP reporter mESC line.[26] Upon differentiation of mESCs to gastruloids, the three reporter proteins are expressed. Scale bars, 100 μm.

(B) Flow cytometry gating strategy to isolate the fluorescent reporter cells and a triple-negative cell population. Single cells were gated after doublet discrimination to further gate on mStrawberry^{+} (P1) and mStrawberry^{-} cells (P2). The mStrawberry negative cells are further gated into BFP^{+} cells (P3), GFP^{+} cells (P4), and a triple-negative cell population (P5).

(C) Heatmap of 2,708 significantly expressed proteins (ANOVA multiple-sample test with Benjamini-Hochberg correction) in the triple-negative (gray), BRA-GFP^{+} (green), SOX17-RFP^{+} (red), and MT1-TagBFP^{+} (blue) cell populations.

(D) Heatmap selecting significantly expressed subunits belonging to the same protein complex, which show higher expression in MT1-TagBFP^{+} cells.

(E) Heatmap selecting significantly expressed subunits belonging to the same protein complex, which show higher expression in SOX17-RFP^{+} cells.

See also Figure S4.

data, which showed high ANXA5 expression in SOX17^{+} endothelial cells.[5] Furthermore, SOX17 protein expression was detected in 89 out of 144 SOX17-RFP^{+} cells. mESCs express higher levels of ESRRB, NDUFV1, and NEDD8, and mesodermal cells express higher levels of TGM1. After successfully identifying unique protein expression profiles in sorted cell populations, we conducted an unbiased SCP analysis utilizing unsorted dissociated gastruloids (Data S1). We identified 2,088 proteins with an average of 1,424 proteins per single cell (Data S1). Integration of both experiments depended on the use of an identical carrier proteome (equal ratio of mESCs, RFP^{+}, GFP^{+}, BFP^{+}, and unsorted cells). Within a uniform manifold approximation and projection (UMAP) projection, a clear separation of undifferentiated mESCs and germ layer cell clusters can be observed (Figures 5C and S5D). To assess the robustness of the clustering, we randomly shuffled the input matrix and performed UMAP analysis. The lack of structure in the randomized data suggests that the identified clustering in the original data is biologically meaningful (Figure S5E). Importantly, just a few gastruloid-derived single cells cluster with mESCs, while the majority of the gastruloid-derived cells cluster with BRA-GFP^{+} cells, indicating that these are likely cells of mesodermal origin (Figure 5C). This corresponds to the predominant presence of mesodermal cell subtypes in gastruloids.[5] We conducted another SCP experiment exclusively with BRA-GFP^{+} cells to profile more GFP^{+} cells. To boost cell-type-specific protein identification, we used a carrier proteome containing only BRA-GFP^{+} cells.[27] In total, 363 single cells passed quality control, resulting in 1,822 identified proteins with at least two unique peptides, with an average of 1,605 proteins per single cell (Figure S5F). We identified three subcellular populations within the BRA-GFP^{+} cell population (Figures 5D and S5G). Aggregated protein expression of gene markers derived from annotated cell types of gastruloid scRNA-seq data[8] distinguishes PSM, somite, and NMP cellular subtypes (Figure 5D). Comparing SCP data with proteins previously identified as highly expressed in BRA-GFP^{+} cells (Figure 4C, bulk proteomics data), we observed an overlap of 40 proteins (Table S4), 13 of which are overlapping with RNA-based mesodermal markers. Although a considerable number of proteins detected in BRA-GFP^{+} cell proteomes overlap with published scRNA-seq data, we have

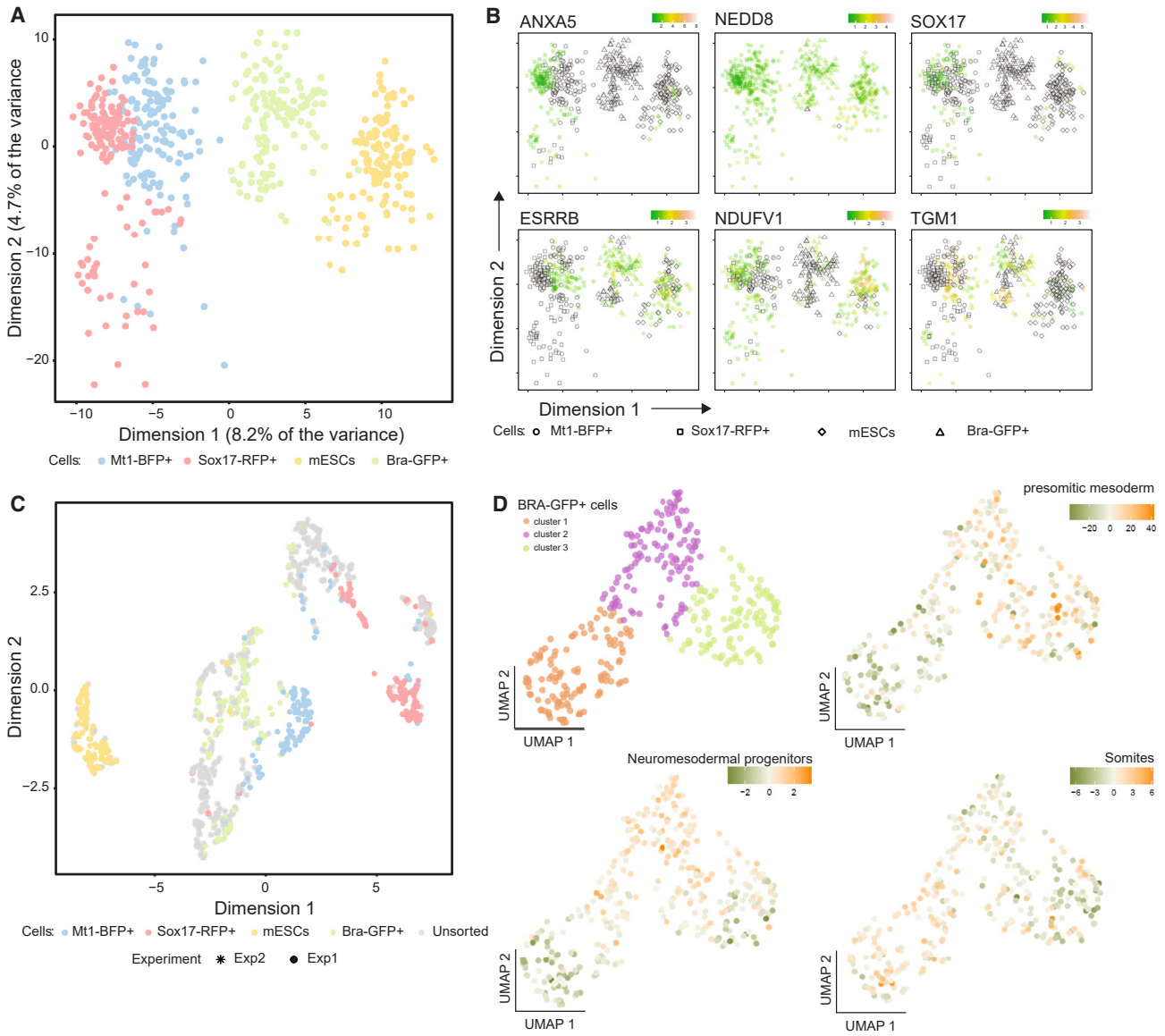


Figure 5. SCP captures germ layer-specific cellular heterogeneity

(A) PCA plot showing the distribution of cells based on the expression of 675 proteins. Each dot represents a single cell, and the color indicates the cell type as in (C).

(B) PCA embedding colored by the expression of ANXA5, NEDD8, SOX17, ESRRB, NDUFV1, and TGM1. Color indicates normalized reporter intensity.

(C) UMAP projection of integrated datasets of the sorted and unsorted SCP experiments.

(D) UMAP projection of BRA-GFP$^+$ cells based on the expression of 1,822 proteins. Upper left: each cell colored according to K-means clustering. Other three UMAPs show the *Z* score normalized and aggregated $\log_2$ abundances of proteins of which the corresponding RNA is identified as a marker for neuromesodermal progenitors (TRHAP3, SUMO1, CBX5, CYCS, HMGN2, and NDUFA12), pre-somitic mesoderm (PAFAH1B3, ALDH1A2, NACA, and RAB8B), and somite (YBX3, RFC5, HMGA2, ANXA2, and DSP) in scRNA-seq.[8]

See also Figure S5, Data S1, and Tables S3 and S4.

also identified numerous novel protein markers specific to BRA-GFP$^+$ cell subtypes (Figure S5H). Another notable observation is the evident enrichment of keratins within one of the clusters (cluster 1, Figure S5I). Keratins are often excluded in scRNA-seq analysis and labeled as contaminants in proteomics experiments. However, the detection of unique peptides specific to mouse keratins confirms their presence and rules out contamination (Figure S5J). This suggests a possible functional significance of keratins within a specific subset of mesodermal cells, potentially facilitating their adaptation to the growth and morphogenesis changes in gastruloids.

Identifying enhancer-associated proteins reveals gastruloid-enriched transcription factors

Next, to facilitate the identification of transcription factors that may be important to drive cell-type specification in gastruloids,

CellPress
OPEN ACCESS

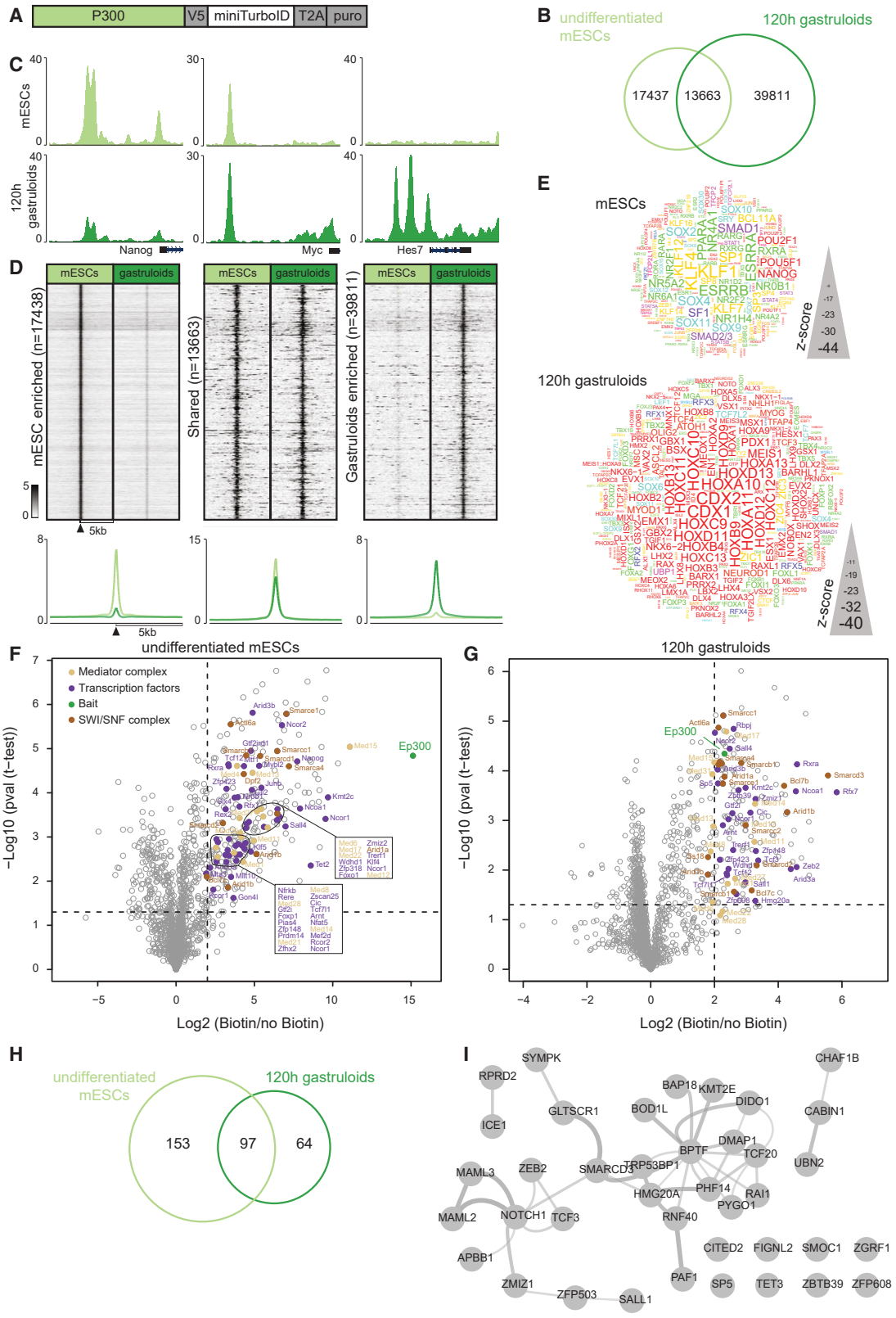


Figure 6. Enhancer proteomic profiling identifies transcriptional regulators in mESCs and gastruloids

(A) CRISPR-Cas9-mediated endogenous gene tagging of the C terminus of *Ep300* gene with V5-miniTurboID followed by a cleaving peptide (T2A) and puromycin coding sequence.

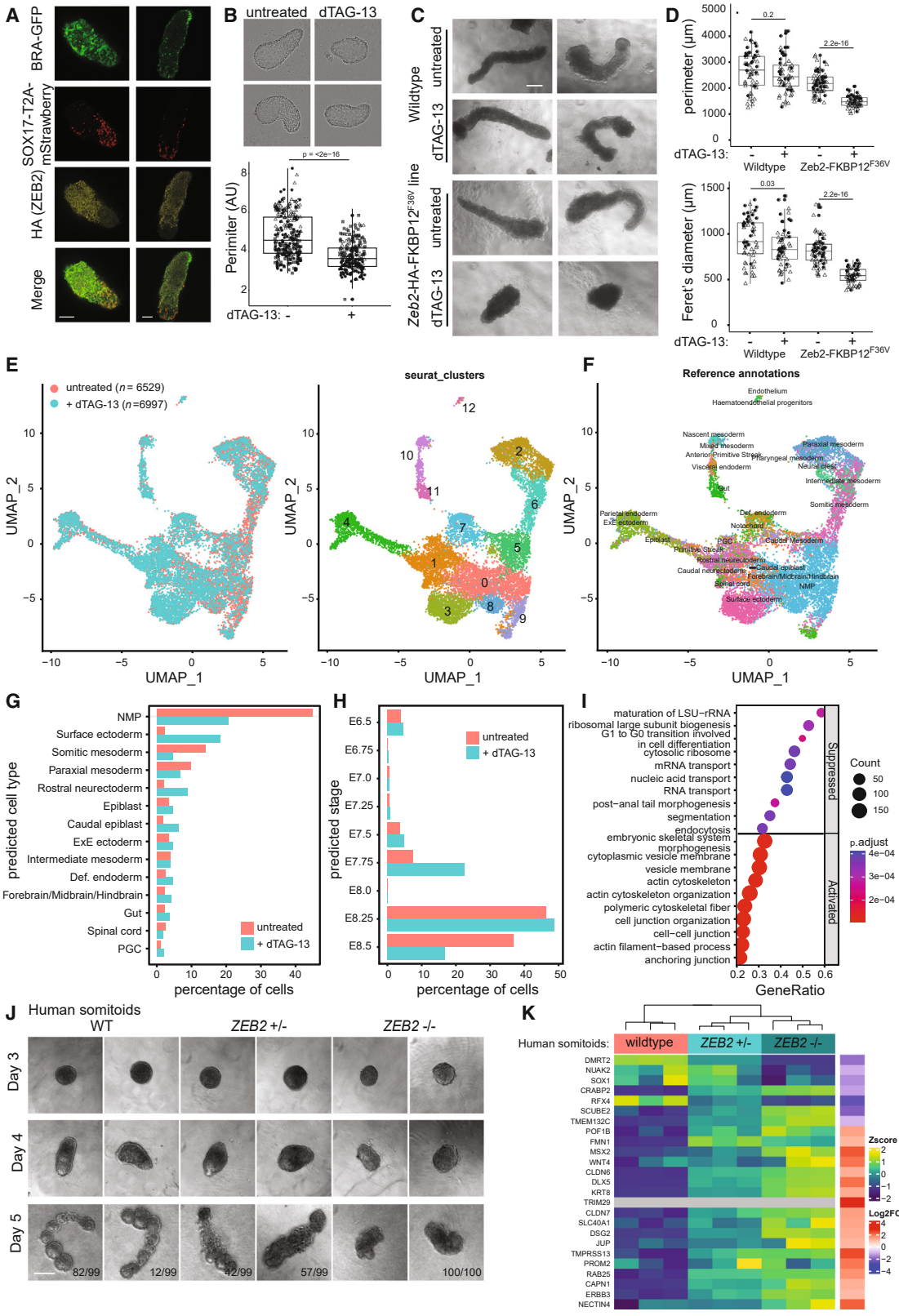
(legend continued on next page)

we aimed to identify global enhancer interactomes in mESCs and gastruloids. P300 acetylates H3K27 at active enhancers. We therefore reasoned that P300 proximity biotinylation would facilitate the identification of enhancer-interacting proteins, including stage-specific transcription factors, in mESCs and gastruloids. We endogenously tagged the C terminus of P300 with a V5 tag and the biotin ligase miniTurboID (Figure 6A), successfully creating a homozygous knockin cell line, which was further characterized by western blotting and immunofluorescence (Figures S6A–S6D). We mapped active enhancers in undifferentiated mESCs and gastruloids by performing V5 chromatin immunoprecipitation sequencing (ChIP-seq) experiments. Reassuringly, the genomic binding sites of tagged P300 in mESCs correlate with published P300 ChIP-seq datasets (Figure S6E). In mESCs and gastruloids, we detected 31,100 and 53,474 P300 binding sites, respectively (Figure 6B). As expected, P300 binding occurs mainly at introns and distal intergenic regions in both conditions (Figure S6F). Next, the P300 regions were classified into three region sets: (1) enriched in mESCs, (2) shared, and (3) enriched in gastruloids (Figure 6C). Shared peaks had a higher average ChIP-seq signal than peaks enriched in gastruloids or mESCs only (Figure 6D). As expected, motif analysis revealed pluripotency transcription factor motifs (NANOG, SOX2, and POU5F1) at mESC-enriched P300 sites (Figure 6E). By contrast, gastruloid-enriched P300 sites show an overrepresentation of homeodomain motifs. Other motifs enriched at P300 binding sites in gastruloids include ZIC1/3/4, TCF7L2, SOX6, and Forkhead box proteins. Next, we employed ANANSE (ANalysis Algorithm for Networks Specified by Enhancers) to determine the influence of each transcription factor on gastruloid differentiation (Figures S6G and S6H).[28] ANANSE integrates genome accessibility from P300 binding sites, motif enrichment at these sites, and the differential expression of the proximal target genes and transcription factor. ANANSE predicted well-known transcription factors important in stem cell maintenance (Figure S6G). Moreover, HOX transcription factors were identified as essential for gastruloid formation (Figure S6H). Additionally, atrial markers like HEY1, NR2F2, and NR2F1 and less well-studied factors such as RFX4 were predicted to be important for gastruloid differentiation.

The presence of a transcription factor motif does not necessarily indicate protein binding. Additionally, proteins that do not rely on sequence-specific binding will be missed. P300 proximity labeling coupled with quantitative mass spectrometry provides an opportunity to map the protein composition at enhancers in their native context in an unbiased manner. In total, 250 proteins were identified as P300 proximal proteins in mESCs (Figure 6F). Expectedly, P300-miniTurboID was the most highly enriched protein in biotin-treated mESCs (Figure 6F). We also observed enrichment of Mediator, SWI/SNF complex members, and numerous transcription factors, including SALL4, NANOG, TCF7L1, and KLF4. For one of the interactors, PAXIP1, we performed a reciprocal proximity labeling experiment using mESCs expressing endogenous PAXIP1-miniTurboID (Figures S6I–S6K). Proximity labeling in gastruloids yielded 161 P300 proximal proteins (Figure 6G). P300 protein itself was not as highly enriched in gastruloids as compared with mESCs, which is likely due to self-biotinylation caused by endogenous biotin in the Ndiff227 medium used to culture gastruloids. Without exogenous biotin stimulation, more P300 peptides were detected in gastruloids (80 peptides) compared with untreated mESCs (<10 peptides). We identified 64 gastruloid-specific, P300 proximal proteins, 39 of which were only detected proximal to P300 in gastruloids (no peptides in mESCs) (Figures 6H and 6I). The majority of these gastruloid-specific P300 proximal proteins showed equal or higher expression in gastruloids compared with mESCs (Figure S6L). We identified two gastruloid-specific P300 proximal SWI/SNF complex subunit proteins, SMARCD3 and GLTSCR1. SMARCD3 is a subunit of the neural progenitors-specific chromatin remodeling complex (npBAF) and neuron-specific chromatin remodeler complex (nBAF).[29,30] GLTSCR1 has been identified as a subunit of the subcomplex GBAF that binds to promoter regions marked with H3K4me3 in mESCs.[31,32] Although GLTSCR1 was only identified and quantified as P300 proximal interactor in gastruloids, we hypothesize that GBAF may redistribute from promoters to enhancers during differentiation, as GBAF has been shown to bind enhancers in other cell types.[33–36] Other gastruloid-specific P300 proximal proteins include all members of the putative chromatin-associated complex known as PRTH or PHF14 complex, which is associated with neurodevelopmental disorders.[37–39] Notably, just a few homeodomain transcription factors were enriched in gastruloids. Specifically, ZEB2 and NANOG exhibited a $\log_2$ FC > 2 and p value < 0.05, while HOXB9, CDX2, EVX1, and ADNP showed a $\log_2$ FC > 1. However, based on the ANANSE results, we expected more HOX proteins proximal to P300 in gastruloids

(B) Venn diagram of overlap between P300 binding sites in undifferentiated mESCs and gastruloids. Peaks identified in both replicates were used.
(C) Snapshots of P300 chromatin binding at three example loci in mESCs (light green) and 120 h gastruloids (dark green). The genomic coordinates are chr6:122700035–122709083 (left), chr15:61977635–61986007 (middle), and chr11:69115928–69125218 (right). y axes indicate the ChIP-seq signal in fragments per kilobase per million (FPKM) reads mapped.
(D) Heatmap visualizing P300 ChIP-seq signal (FPKM) in mESC and 120 h gastruloids. Data are centered at P300 peaks, depicting a 5-kb window around the peak. Binding events are subdivided into mESC-enriched sites (17,437 sites), shared mESC- and gastruloid-binding sites (shared, 13,663 sites), and gastruloid-enriched sites (39,811 sites). Average signal (FPKM) of P300 for all sites (bottom).
(E) Motif enrichment at mESC and gastruloid-enriched P300 binding sites. Font size represents the Z score (the smaller the Z score, the more enriched, the larger the font size), and colors represent transcription factor families.
(F) Volcano plot of proteins identified in P300 proximity labeling experiments in undifferentiated mESCs. Enrichment of P300 and proximity interactors is shown as $\log_2$ fold enrichment of LFQ intensities of biotin treatment over LFQ intensity of untreated (x axis) plotted against the $-\log_{10} p$ value. Selected proteins are colored and labeled with gene symbols.
(G) Volcano plot of proteins identified in P300 proximity labeling experiments in gastruloids.
(H) Overlap in significantly enriched P300 proximal proteins ($\log_2$ FC > 2 and p value < 0.05) between mESCs and gastruloids.
(I) STRING protein-protein interaction network of 39 proteins, which are only detected proximal to P300 in 120 h gastruloids.
See also Figure S6 and Data S1.

CellPress
OPEN ACCESS



(legend on next page)

(Figure S6H). To assess the cause of this discrepancy, we assessed HOX protein expression in gastruloids. DNA affinity purifications with HOX-motif-containing oligos in gastruloid extracts efficiently pulled down homeodomain transcription factors, including MEIS2, HOXB9, HOXA5, CDX2, MEOX1, and three PBX proteins (Data S1). The PBX proteins are known to serve as heterodimeric binding partners for HOX proteins. To further confirm the expression of HOX proteins in gastruloids, we performed immunofluorescence for HOXC8 and HOXB4. Both approaches confirm the expression of HOX proteins in gastruloids (Figure S6M). Finally, we observed the enrichment of a variety of transcription regulators involved in notch signaling (NOTCH1 and MAML2/3), Wnt signaling (PYGO1 and SP5), and less well-characterized proteins (ZFP608, ZBTB39, and ICE1) as proximal to P300 in gastruloids. Further experiments are required to investigate a putative role for these proteins in gastrulation.

Perturbing transcription factor expression during gastruloid formation

To examine whether perturbation of P300 proximal transcription factors impacts gastruloid formation, we endogenously tagged three candidate genes (*Rxrα*, *Sp5*, and *Zeb2*) with a hemagglutinin (HA) tag and FKBP12^{F36V} degron tag in mESCs (Data S1). SP5 and ZEB2 were selected since these were only identified as P300 proximal transcription factors in gastruloids and are known regulators of Wnt signaling and cellular differentiation, respectively.[40,41] RXRα was selected given a described role for retinoic acid signaling in early lineage specification.[42,43] Following validation of the knockin lines (Data S1), we assessed the genome-wide binding of the tagged transcription factors (Data S1). Genome browser snapshots at three representative loci show overlap of the transcription factors with P300 binding (Data S1). Heatmap-based visualization revealed substantial overlap between RXRα and P300 in mESCs and gastruloids, respectively (Data S1). RXRα binding sites mainly map to distal intergenic regions and introns in mESCs or gastruloids (Data S1). For ZEB2, the majority of binding sites in gastruloids are also located at distal intergenic regions and introns. By contrast, the majority of SP5 binding sites in gastruloids map to promoters, whereas ~35% localize at enhancers. The motifs for each transcription factor are enriched underneath the ChIP-seq peaks (Data S1). Next, to assess the importance of RXRα, SP5, and ZEB2 during early embryonic development, we depleted these proteins at different time points during gastruloid formation. RNA-seq-based analyses revealed that perturbation of RXRα and SP5 during gastruloid formation does not impact global gene expression patterns and is not associated with an apparent phenotype (Data S1). This was somewhat surprising, given the recently described role for retinoic acid signaling in gastruloid differentiation,[44] which we were able to reproduce (Data S1). We therefore hypothesize that all-trans retinoic acid (ATRA) signaling is likely facilitated by other members of the RXR family or through non-canonical ATRA activities.[45] By contrast, ZEB2 perturbation was associated with a clear phenotype, which is described in detail below.

Disruption of ZEB2 impairs mouse and human somitogenesis

Next, we investigated ZEB2's role in gastruloid formation. ZEB2 protein expression was undetectable in undifferentiated mESCs and is upregulated during gastruloid differentiation from 96 h onwards. Staining for ZEB2 protein in 120 h gastruloids revealed no overlap with SOX17-RFP^{+} cells (Figure 7A), which is consistent with the absence of *Zeb2* mRNA expression in the *Sox17*$^{+}$ scRNA clusters in both gastruloids and mouse embryos (Figures S7A and S7B). To determine the effect of ZEB2 depletion on gastruloid formation, ZEB2-HA-FKBP12^{F36V} expressing mESC were treated with dTAG-13 at 72 h up to 120 h during (somitic) gastruloid differentiation (Figures 7B and 7C). Degradation of ZEB2 during "regular" gastruloid differentiation resulted in more ovoid-shaped gastruloids than elongated gastruloids as compared with untreated gastruloids (Figure 7B). ZEB2 depletion in gastruloids embedded in matrigel (somitic gastruloids) failed to elongate, whereas dTAG-13 treatment did not perturb cell viability and differentiation in wild-type somitic gastruloids (Figures 7C and 7D; Video S1). To further investigate this phenotype, we performed bulk and scRNA-seq. For scRNA-seq, we obtained 6,529 untreated cells and 6,997 cells from ZEB2-depleted gastruloids after quality control. Data integration and

Figure 7. Perturbation of Zeb2 expression in gastruloids

(A) Immunofluorescence staining for HA in gastruloids generated from *Zeb2*-HA-FKBP12^{F36V} expressing mESCs. Scale bars, 100 μm.

(B) Phase contrast images and perimeter quantification of regular 120 h gastruloids generated from *Zeb2*-HA-FKBP12^{F36V} expressing mESCs. Gastruloids were treated with DMSO (−) or with dTAG-13 (+) at time points 72 and 96 h. Images were acquired with an IncuCyte device (4× objective). Reported *p* values are determined with two-tailed t test and point shapes indicate biological replicates.

(C) Phase contrast images of somitic gastruloids generated from wild-type mESCs and *Zeb2*-HA-FKBP12^{F36V} expressing mESCs. Gastruloids were treated with DMSO (−) or with dTAG-13 (+) at time points 72 and 96 h. Scale bars, 200 μm.

(D) Quantification of the perimeter and Feret's diameter of somitic gastruloids treated without dTAG-13 (−) and with dTAG-13 (+) at 72 and 96 h. Shapes indicate biological replicates. Reported *p* values are determined with two-tailed t test.

(E) UMAP colored by cells from untreated or dTAG-13-treated somitic gastruloids generated from *Zeb2*-HA-FKBP12^{F36V} expressing mESCs (left) or Seurat cluster identity (right).

(F) UMAP embedding overlay showing annotation of cell types based on label transfer from the reference mouse gastrulation atlas.[46]

(G) Cell types were annotated by label transfer from publicly available mouse gastrulation scRNA-seq atlas.[46]

(H) Embryonic stage annotated by label transfer from publicly available mouse gastrulation scRNA-seq atlas.[46]

(I) fGSEA results showing the top 10 suppressed and activated biological processes in ZEB2-depleted somitic gastruloids. Dot size indicates the number of enriched genes in each term. Dot color represents the adjusted *p* value.

(J) Phase contrast images of human somitoid development using wild-type human iPSCs (WT), homozygous (ZEB2−/−), and heterozygous (ZEB2+/−) *ZEB2* KO iPSCs. The occurrence of the different morphologies is given.

(K) Heatmap of scaled gene expression in human somitoids. The selected subset of genes comprises those identified as differentially expressed in mouse somitic gastruloids. The $\log_2$ fold changes for these genes in mouse somitic gastruloids are shown on the right.

See also Figure S7, Data S1, and Video S1.

UMAP projection revealed 13 clusters with different proportions of untreated and ZEB2-depleted cells (dTAG-13 treated) among the clusters (Figure 7E; Data S1). Clusters 1 and 3 are enriched for dTAG-13 treated cells, whereas four clusters (clusters 5, 6, 8, and 9) are enriched for untreated cells. We used the reference atlas from mouse embryos to predict the cell types within this dataset.[46] The majority of the cells are classified as NMPs (Figures 7F and 7G). Expression of well-known markers for ectoderm (or primordial germ cell) (*Pou5f1*, *Utf1*, and *Mt1*), endoderm (*Sox17*, *Foxa2*, and *Epcam*), cardiac (*Gata6* and *Kdr*), mesoderm (*T/Bra*, *Sp5*, and *Tbx6*), somites (*Dll1* and *Hes7*), and surface ectoderm (*Krt8* and *Krt18*) is concordant with the mapping to the *in vivo* cell types (Figure S7C). Having assigned predicted cell types, we observed that the cellular composition of ZEB2-depleted gastruloids shifts from more NMPs and somitic mesoderm-rich gastruloids to enrichment in surface ectoderm cells and rostral neuroectoderm cells (Figure 7G). Further, looking at the assignment of embryonic stage: more ZEB2-depleted cells are assigned to embryonic stage E7.75 (22.48% ZEB2-depleted cells versus 7.6% untreated cells), while fewer ZEB2-depleted cells are assigned to E8.5 (16.78% ZEB2-depleted cells versus 36.87% untreated cells) (Figure 7H). Next, we performed differential gene expression analysis between untreated and dTAG-13-treated cells. The upregulated genes in ZEB2-depleted gastruloids are involved in cell-cell junctions and actin cytoskeleton organization (Figure 7I). GO terms linked to suppressed pathways are involved in post-anal tail morphogenesis and segmentation. The differentially expressed genes identified in bulk RNA-seq were validated in scRNA-seq (Figures S7D and S7E). Promoters of these differentially expressed genes are occupied by P300 and ZEB2 (Figure S7F). To construct the lineage differentiation trajectory of gastruloids, we used Monocle3.[47] The trajectory begins with cells classified as epiblast cells and Exe ectoderm (cluster 10), followed by a path that leads to ectodermal lineages and then NMPs. Additionally, the trajectory splits into three branches: one toward surface ectoderm; the second branch encompasses endodermal lineages, such as definitive endoderm, gut, visceral endoderm, endothelium, and hematopoietic progenitors; and the third branch consists of mesodermal lineages, such as somitic mesoderm, intermediate mesoderm, and paraxial mesoderm (Figure S7G). In conclusion, loss of ZEB2 pushes toward an expansion of surface ectoderm and neuroectoderm cells concomitant with a decrease in somitic mesoderm.

Recently, a human model of somitogenesis has been developed,[48] which provides an excellent opportunity to validate findings from the mouse somitogenesis model to human. We generated homozygous and heterozygous *ZEB2* knockout human induced pluripotent stem cell (iPSC) lines (Data S1). When cultured in matrigel for 24 h, wild-type human iPSCs formed structures containing multiple somite-like structures (Figure 7J). By contrast, the *ZEB2* knockout iPSCs failed to elongate (Figure 7J; Data S1). The somitoids with heterozygous *ZEB2* loss displayed an intermediate phenotype, characterized by elongation but without segmented structures. We next assessed whether differentially expressed genes identified in ZEB2-depleted mouse somitic gastruloids exhibit differential expression in human somitoids. Indeed, initial validation by quantitative reverse-transcription PCR (RT-qPCR) confirmed that four genes (*CLDN7*, *SCUBE2*, *KRT8*, and *DLX5*) that were found to be upregulated upon ZEB2 depletion in mouse somitic gastruloids were also upregulated in *ZEB2* knockout human somitoids, whereas the downregulated gene *DMRT2* showed concordant downregulation (Data S1). Somitoids with heterozygous *ZEB2* loss display intermediate effects on gene expression of the same genes. RNA-seq analysis further confirmed the concordance between mouse somitic gastruloids and human somitoids, particularly for those differentially expressed genes identified in mouse somitic gastruloids (Figure 7K; Data S1). However, a higher number of differentially expressed genes was detected in wild-type human somitoids versus *ZEB2* knockout conditions (n = 1,124, adjusted p value < 0.05 and absolute $\log_2$ fold change > 2) (Data S1). We hypothesize that these differences can be explained by the fact that we made use of a transient protein degradation strategy in mouse gastruloids, whereas a full *ZEB2* knockout was used in human somitoids. Nevertheless, the upregulated genes in human *ZEB2* knockout somitoids showed significant enrichment for surface ectoderm marker genes, while downregulated genes are enriched for markers of paraxial mesoderm (Data S1), which agrees with our findings in mouse somitic gastruloids (Figure 7G). In conclusion, ZEB2 is essential for the formation of segmented structures during mouse and human somitogenesis.

DISCUSSION

Here, we exploited the scalability of mouse gastruloids to profile their dynamic (phospho) proteome in a time-resolved manner. In addition to many well-known marker proteins, numerous less well-characterized proteins are dynamically expressed during gastruloid formation, providing a list of target proteins for future experiments aimed at elucidating their putative role in gastruloid formation. Additionally, investigating dynamic phosphorylation sites on these proteins can shed further light on their biological functions. We also profiled the proteomes of sorted germ layer cells. Here, the most distinct difference was observed between SOX17-RFP$^+$ (endoderm) and MT1-BFP$^+$ (ectoderm) cell populations. However, discrimination between BRA-GFP$^+$ (mesoderm) cells and triple-negative cells was less clear, likely due to the predominance of mesodermal subtypes among the triple-negative cells. Of note, utilizing fluorescent reporter lines requires time-consuming CRISPR-based genetic manipulation and the selection of appropriate reporter genes. In some cases, the specificity of the chosen reporter gene may be insufficient to discriminate between different cell types. For example, SOX17 is expressed in both endoderm and endothelial cells, and these cells have a distinct RNA expression profile.[5] Bulk proteomics analysis of SOX17$^+$ cells thus results in a mixed proteome of both cell types. The recent emergence of SCP technology offers a promising solution to overcome these issues. With advancements in the technology, the average number of detected proteins per cell has increased from 767 proteins in the first study to 1,500–2,000 proteins per cell using the CellenOne platform workflow.[49,50] Here, this approach enabled the quantification of >1,400 proteins in single cells, facilitating the distinction of germ layer cell populations and mESCs resembling bulk proteomics. Notably, SOX17$^+$ cells exhibit two distinct clusters on the PCA component 2, possibly reflecting endoderm and

endothelial cells (Figures 5A and 5C). In the coming years, further improvements in throughput and proteome depth will enhance the ability to discriminate cell clusters and annotate cell types without prior knowledge. A further interesting direction to explore would be to conduct parallel profiling of the transcriptome and proteome in individual cells to better understand gene expression regulation in complex developmental biological systems.[51] Additionally, live-cell microscopy combined with fluorescence-based methods would be a valuable and complementary approach for validating protein dynamics at single-cell resolution.

Most gene expression studies apply sequence-based methods to infer gene regulatory networks, which are biased toward transcription factors with annotated DNA binding motifs. In addition, transcriptional repressors are often not considered in gene regulatory network algorithms.[28,52] Tagging *Ep300* with miniTurboID has proven here to be effective in characterizing enhancer-bound proteins, including transcription factors and chromatin-modifying enzymes. Follow-up experiments revealed that ZEB2 loss during gastruloid formation results in elongation defects. Loss of ZEB2 in somitic gastruloids results in an expansion of surface ectoderm and neuroectoderm concomitant with a decrease in somitic mesoderm. Previous studies have shown that the surface ectoderm and somites have a complex interplay during embryonic development, where the former provides important signals (direct cell-cell contacts, Wnt signals) to regulate the elongation of somitic mesoderm and somite epithelialization.[53–55] Precise control in fate specification to neuroectoderm and surface ectoderm seems to be crucial for proper somitogenesis. Consistently, *Zeb2* knockout mice display multiple developmental defects, including abnormalities in somite formation, myogenic differentiation, neural plate, and neural crest cells.[56–58] Additionally, we were able to reproduce the defective somite phenotype in human somitoids. These findings suggest a conserved role for ZEB2 in regulating the development of these crucial cell types across species. Interestingly, human somitoids generated from heterozygous *ZEB2* knockout cells display an intermediate phenotype. Notably, in humans, haploinsufficiency of *ZEB2* has been linked to Mowat-Wilson syndrome, a neurological disorder.[59] Altogether, these discoveries highlight the importance of ZEB2 in early embryogenesis. To investigate the importance of other gastruloid-specific, p300 proximal proteins, a genetic screen targeting these genes could uncover additional regulators of early embryonic development and/or somitogenesis.

Limitations of the study

Although we provide a thorough characterization of the proteome in gastruloids, this quantified proteome is not comprehensive. Despite the application of various affinity enrichment methods, key transcription factors, such as classical HOX proteins, were sparsely identified in our experiments. This was unexpected given the strong enrichment of HOX motifs in P300-bound regions and given the prominent RNA expression patterns of *HOX* genes in gastruloids. Low-abundant proteins are notoriously difficult to detect, but the development of new technologies, such as data-independent acquisition mass spectrometry (DIA-MS) analysis and algorithms based on deep learning,[60,61] hold great promise for their improved detection in the future. Finally, although gastruloids are very useful for studying key aspects of early mammalian embryogenesis, particularly due to their scalability and susceptibility of mESCs to genome engineering technologies, these embryo-like structures do not (yet) fully recapitulate the complexity and cellular diversity of natural embryos. Systematic comparisons, such as the proteome comparisons presented in this study, are therefore essential to investigate similarities and discrepancies between these embryo-like structures and natural embryos. Furthermore, these insights could lead to advancements in generating more complex mouse three-dimensional (3D) gastruloid models.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Cell lines
  - Generation of cell lines
  - Mouse embryos
- METHOD DETAILS
  - Mouse gastruloids and somites
  - Mouse embryoid bodies (EBs)
  - Human somitoids
  - Mouse embryo isolation
  - (Phospho)proteomics
  - Streptavidin pulldowns
  - DNA pulldown
  - LC-MS/MS and downstream analysis
  - Single cell proteomics
  - HA pulldown followed by western blotting
  - ChIP-seq
  - RNA-seq
  - RT-qPCR
  - Ananse analysis
  - Single-cell RNA-seq
  - Immunofluorescence
  - Perimeter quantification
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.stem.2024.04.017.

ACKNOWLEDGMENTS

We would like to thank Rob Woestenenk, Paul Ruijs, Laura Wingens, Tom van Oorschot, Nick C.M. Pelzers, Susan Zwakenberg, and Marjolein Vliem for their technical support with cell sorting. We thank Marieke Willemse and Gert-Jan Bakker from the Radboud Technology Center for Microscopy for training and assistance with confocal microscopy and Okolab, respectively. We thank Maarten van der Sande and Siebren Frölich for seq2science support and Sybren Rinzema for sequencing data management. We would also like to thank Susanne van den Brink and Sandra de Vries for their valuable advice on gastruloid culturing and Stéfane Nedelec for advice on HOX proteins. Additionally, we thank all members of the Vermeulen lab for their input and suggestions. We thank the laboratory of Iftach Nachman (Tel Aviv University) for sharing the dual BRA/SOX17 reporter cell line. This work was supported by a VENI grant from the Netherlands Organisation for Scientific Research (NWO, VI.Veni.212.076); Pluripotent Stem cells for Inherited Diseases and Embryonic Research

(PSIDER) grant from ZonMw (10250042110004); EPIC-XS (project number 823839), funded by the Horizon 2020 programme of the European Union; and the NWO-funded Netherlands Proteomics Centre through the National Road Map for Large-scale Infrastructures program X-Omics (project 184.034.019). J.M.V. is supported by scholarships from the Ministry of Science and Technology of Costa Rica (MICITT) and the University of Costa Rica (UCR). The Vermeulen and Burgering labs are part of the Oncode Institute, which is partly funded by the Dutch Cancer Society (KWF). The research reported in this publication was supported by Oncode Accelerator, a Dutch National Growth Fund project under grant number NGFOP2201.

AUTHOR CONTRIBUTIONS

Conceptualization, M.V. and S.S.; methodology, S.S., K.F.S., H.R.V., and C.A.G.H.v.G.; formal analysis, S.S., M.T.A.-V., C.A.G.H.v.G., P.S.A., and J.M.V.; investigation, S.S., M.T.A.-V., L.A.L., D.W.Z., M.J.v.O., J.M.V., C.A.G.H.v.G., P.S.A., T.R., and M.P.A.B.; data curation, S.S., P.S.A., and H.R.V.; resources, P.W.T.C.J. and C.F.; writing—original draft, S.S., M.T.A.-V., and M.V.; writing—review and editing, all authors; visualization, S.S., M.T.A.-V., C.A.G.H.v.G., P.S.A., and T.S.; supervision, S.S., M.A.F.M.A., K.F.S., B.B., H.R.V., and M.V.; funding acquisition, S.S., M.V., M.A.F.M.A., J.M.V., and B.B.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 24, 2023
Revised: January 10, 2024
Accepted: April 19, 2024
Published: May 15, 2024

REFERENCES

1. Bolondi, A., Kretzmer, H., and Meissner, A. (2022). Single-cell technologies: a new lens into epigenetic regulation in development. Curr. Opin. Genet. Dev. *76*, 101947. https://doi.org/10.1016/j.gde.2022.101947.
2. van den Brink, S.C., and van Oudenaarden, A. (2021). 3D gastruloids: a novel frontier in stem cell-based in vitro modeling of mammalian gastrulation. Trends Cell Biol. *31*, 747–759. https://doi.org/10.1016/j.tcb.2021.06.007.
3. El Azhar, Y., and Sonnen, K.F. (2021). Development in a Dish-In Vitro Models of Mammalian Embryonic Development. Front. Cell Dev. Biol. *9*, 655993. https://doi.org/10.3389/fcell.2021.655993.
4. Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature *562*, 272–276. https://doi.org/10.1038/s41586-018-0578-0.
5. van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A., and van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature *582*, 405–409. https://doi.org/10.1038/s41586-020-2024-3.
6. Braccioli, L., van den Brand, T., Saiz, N.A., Fountas, C., Celie, P.H.N., Kazokaitė-Adomaitienė, J., and de Wit, E. (2022). Identifying cross-lineage dependencies of cell-type specific regulators in gastruloids. Preprint at bioRxiv. https://doi.org/10.1101/2022.11.01.514697.
7. Rosen, L.U., Stapel, L.C., Argelaguet, R., Barker, C.G., Yang, A., Reik, W., and Marioni, J.C. (2022). Inter-gastruloid heterogeneity revealed by single cell transcriptomics time course: implications for organoid based perturbation studies. Preprint at bioRxiv. https://doi.org/10.1101/2022.09.27.509783.
8. Suppinger, S., Zinner, M., Aizarani, N., Lukonin, I., Ortiz, R., Azzi, C., Stadler, M.B., Vianello, S., Palla, G., Kohler, H., et al. (2023). Multimodal characterization of murine gastruloid development. Cell Stem Cell *30*, 867–884.e11. https://doi.org/10.1016/j.stem.2023.04.018.
9. Merle, M., Friedman, L., Chureau, C., Shoushtarizadeh, A., and Gregor, T. (2024). Precise and scalable self-organization in mammalian pseudo-embryos. Nat. Struct. Mol. Biol. 1–7. https://doi.org/10.1038/s41594-024-01251-4.
10. Gao, Y., Liu, X., Tang, B., Li, C., Kou, Z., Li, L., Liu, W., Wu, Y., Kou, X., Li, J., et al. (2017). Protein Expression Landscape of Mouse Embryos during Pre-implantation Development. Cell Rep. *21*, 3957–3969. https://doi.org/10.1016/j.celrep.2017.11.111.
11. Wang, S., Kou, Z., Jing, Z., Zhang, Y., Guo, X., Dong, M., Wilmut, I., and Gao, S. (2010). Proteome of mouse oocytes at different developmental stages. Proc. Natl. Acad. Sci. USA *107*, 17639–17644. https://doi.org/10.1073/pnas.1013185107.
12. Zhang, P., Ni, X., Guo, Y., Guo, X., Wang, Y., Zhou, Z., Huo, R., and Sha, J. (2009). Proteomic-based identification of maternal proteins in mature mouse oocytes. BMC Genomics *10*, 348. https://doi.org/10.1186/1471-2164-10-348.
13. Dang, Y., Zhu, L., Yuan, P., Liu, Q., Guo, Q., Chen, X., Gao, S., Liu, X., Ji, S., Yuan, Y., et al. (2023). Functional profiling of stage-specific proteome and translational transition across human pre-implantation embryo development at a single-cell resolution. Cell Discov. *9*, 10. https://doi.org/10.1038/s41421-022-00491-2.
14. Jarnuczak, A.F., Najgebauer, H., Barzine, M., Kundu, D.J., Ghavidel, F., Perez-Riverol, Y., Papatheodorou, I., Brazma, A., and Vizcaíno, J.A. (2021). An integrated landscape of protein expression in human cancer. Sci. Data *8*, 115. https://doi.org/10.1038/s41597-021-00890-2.
15. Lindeboom, R.G., van Voorthuijsen, L., Oost, K.C., Rodríguez-Colman, M.J., Luna-Velez, M.V., Furlan, C., Baraille, F., Jansen, P.W., Ribeiro, A., Burgering, B.M., et al. (2018). Integrative multi-omics analysis of intestinal organoid differentiation. Mol. Syst. Biol. *14*, e8227. https://doi.org/10.15252/msb.20188227.
16. Cheng, Z., Teo, G., Krueger, S., Rock, T.M., Koh, H.W.L., Choi, H., and Vogel, C. (2016). Differential dynamics of the mammalian mRNA and protein expression response to misfolding stress. Mol. Syst. Biol. *12*, 855. https://doi.org/10.15252/msb.20156423.
17. Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. Nature *473*, 337–342. https://doi.org/10.1038/nature10098.
18. Baillie-Johnson, P., van den Brink, S.C., Balayo, T., Turner, D.A., and Martinez Arias, A. (2015). Generation of Aggregates of Mouse Embryonic Stem Cells that Show Symmetry Breaking, Polarization and Emergent Collective Behaviour In Vitro. J. Vis. Exp. *105*, 53252. https://doi.org/10.3791/53252.
19. Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Ishijima, J., Shiroishi, T., and Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell *145*, 383–397. https://doi.org/10.1016/j.cell.2011.03.028.
20. Revil, T., Gaffney, D., Dias, C., Majewski, J., and Jerome-Majewska, L.A. (2010). Alternative splicing is frequent during early embryonic development in mouse. BMC Genomics *11*, 399. https://doi.org/10.1186/1471-2164-11-399.
21. Lu, X., Zhao, Z.A., Wang, X., Zhang, X., Zhai, Y., Deng, W., Yi, Z., and Li, L. (2018). Whole-transcriptome splicing profiling of E7.5 mouse primary germ layers reveals frequent alternative promoter usage during mouse early embryogenesis. Biol. Open *7*, bio032508. https://doi.org/10.1242/bio.032508.
22. Shen, W.K., Chen, S.Y., Gan, Z.Q., Zhang, Y.Z., Yue, T., Chen, M.M., Xue, Y., Hu, H., and Guo, A.Y. (2023). AnimalTFDB 4.0: a comprehensive animal transcription factor database updated with variation and expression annotations. Nucleic Acids Res. *51*, D39–D45. https://doi.org/10.1093/nar/gkac907.
23. Bialkowska, A.B., Yang, V.W., and Mallipattu, S.K. (2017). Kruppel-like factors in mammalian stem cells and development. Development *144*, 737–754. https://doi.org/10.1242/dev.145441.

24. Lauberth, S.M., Bilyeu, A.C., Firulli, B.A., Kroll, K.L., and Rauchman, M. (2007). A phosphomimetic mutation in the Sall1 repression motif disrupts recruitment of the nucleosome remodeling and deacetylase complex and repression of Gbx2. J. Biol. Chem. *282*, 34858–34868. https://doi.org/10.1074/jbc.M703702200.

25. Luo, H., Yu, Q., Liu, Y., Tang, M., Liang, M., Zhang, D., Xiao, T.S., Wu, L., Tan, M., Ruan, Y., et al. (2020). LATS kinase-mediated CTCF phosphorylation and selective loss of genomic binding. Sci. Adv. *6*, eaaw4651. https://doi.org/10.1126/sciadv.aaw4651.

26. Pour, M., Kumar, A.S., Farag, N., Bolondi, A., Kretzmer, H., Walther, M., Wittler, L., Meissner, A., and Nachman, I. (2022). Emergence and patterning dynamics of mouse-definitive endoderm. iScience *25*, 103556. https://doi.org/10.1016/j.isci.2021.103556.

27. Dwivedi, P., and Rose, C.M. (2022). Understanding the effect of carrier proteomes in single cell proteomic studies - key lessons. Expert Rev. Proteomics *19*, 5–15. https://doi.org/10.1080/14789450.2022.2036126.

28. Xu, Q., Georgiou, G., Frölich, S., van der Sande, M., Veenstra, G.J.C., Zhou, H., and van Heeringen, S.J. (2021). ANANSE: an enhancer network-based computational approach for predicting key transcription factors in cell fate determination. Nucleic Acids Res. *49*, 7966–7985. https://doi.org/10.1093/nar/gkab598.

29. Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron *55*, 201–215. https://doi.org/10.1016/j.neuron.2007.06.019.

30. Staahl, B.T., and Crabtree, G.R. (2013). Creating a neural specific chromatin landscape by npBAF and nBAF complexes. Curr. Opin. Neurobiol. *23*, 903–913. https://doi.org/10.1016/j.conb.2013.09.003.

31. Alpsoy, A., and Dykhuizen, E.C. (2018). Glioma tumor suppressor candidate region gene 1 (GLTSCR1) and its paralog GLTSCR1-like form SWI/SNF chromatin remodeling subcomplexes. J. Biol. Chem. *293*, 3892–3903. https://doi.org/10.1074/jbc.RA117.001065.

32. Gatchalian, J., Malik, S., Ho, J., Lee, D.S., Kelso, T.W.R., Shokhirev, M.N., Dixon, J.R., and Hargreaves, D.C. (2018). A non-canonical BRD9-containing BAF chromatin remodeling complex regulates naive pluripotency in mouse embryonic stem cells. Nat. Commun. *9*, 5139. https://doi.org/10.1038/s41467-018-07528-9.

33. Jefimov, K., Alcaraz, N., Kloet, S.L., Värv, S., Sakya, S.A., Vaagenso, C.D., Vermeulen, M., Aasland, R., and Andersson, a.R. (2018). The GBAF chromatin remodeling complex binds H3K27ac and mediates enhancer transcription. Preprint at bioRxiv. https://doi.org/10.1101/445148.

34. Michel, B.C., D'Avino, A.R., Cassel, S.H., Mashtalir, N., McKenzie, Z.M., McBride, M.J., Valencia, A.M., Zhou, Q., Bocker, M., Soares, L.M.M., et al. (2018). A non-canonical SWI/SNF complex is a synthetic lethal target in cancers driven by BAF complex perturbation. Nat. Cell Biol. *20*, 1410–1420. https://doi.org/10.1038/s41556-018-0221-1.

35. Brien, G.L., Remillard, D., Shi, J., Hemming, M.L., Chabon, J., Wynne, K., Dillon, E.T., Cagney, G., Van Mierlo, G., Baltissen, M.P., et al. (2018). Targeted degradation of BRD9 reverses oncogenic gene expression in synovial sarcoma. eLife *7*, e41305. https://doi.org/10.7554/eLife.41305.

36. Loo, C.S., Gatchalian, J., Liang, Y., Leblanc, M., Xie, M., Ho, J., Venkatraghavan, B., Hargreaves, D.C., and Zheng, Y. (2020). A Genome-wide CRISPR Screen Reveals a Role for the Non-canonical Nucleosome-Remodeling BAF Complex in Foxp3 Expression and Regulatory T Cell Function. Immunity *53*, 143–157.e8. https://doi.org/10.1016/j.immuni.2020.06.011.

37. Eberl, H.C., Spruijt, C.G., Kelstrup, C.D., Vermeulen, M., and Mann, M. (2013). A map of general and specialized chromatin readers in mouse tissues generated by label-free interaction proteomics. Mol. Cell *49*, 368–378. https://doi.org/10.1016/j.molcel.2012.10.026.

38. Vetrini, F., McKee, S., Rosenfeld, J.A., Suri, M., Lewis, A.M., Nugent, K.M., Roeder, E., Littlejohn, R.O., Holder, S., Zhu, W., et al. (2019). De novo and inherited TCF20 pathogenic variants are associated with intellectual disability, dysmorphic features, hypotonia, and neurological impairments with similarities to Smith-Magenis syndrome. Genome Med *11*, 12. https://doi.org/10.1186/s13073-019-0623-0.

39. Slager, R.E., Newton, T.L., Vlangos, C.N., Finucane, B., and Elsea, S.H. (2003). Mutations in RAI1 associated with Smith-Magenis syndrome. Nat. Genet. *33*, 466–468. https://doi.org/10.1038/ng1126.

40. Stryjewska, A., Dries, R., Pieters, T., Verstappen, G., Conidi, A., Coddens, K., Francis, A., Umans, L., van IJcken, W.F.J., Berx, G., et al. (2017). Zeb2 Regulates Cell Fate at the Exit from Epiblast State in Mouse Embryonic Stem Cells. Stem Cells *35*, 611–625. https://doi.org/10.1002/stem.2521.

41. Huggins, I.J., Bos, T., Gaylord, O., Jessen, C., Lonquich, B., Puranen, A., Richter, J., Rossdam, C., Brafman, D., Gaasterland, T., and Willert, K. (2017). The WNT target SP5 negatively regulates WNT transcriptional programs in human pluripotent stem cells. Nat. Commun. *8*, 1034. https://doi.org/10.1038/s41467-017-01203-1.

42. Iturbide, A., Ruiz Tejada Segura, M.L., Noll, C., Schorpp, K., Rothenaigner, I., Ruiz-Morales, E.R., Lubatti, G., Agami, A., Hadian, K., Scialdone, A., and Torres-Padilla, M.E. (2021). Retinoic acid signaling is critical during the totipotency window in early mammalian development. Nat. Struct. Mol. Biol. *28*, 521–532. https://doi.org/10.1038/s41594-021-00590-w.

43. Mark, M., Ghyselinck, N.B., and Chambon, P. (2009). Function of retinoic acid receptors during embryonic development. Nucl. Recept. Signal. *7*, e002. https://doi.org/10.1621/nrs.07002.

44. Mantziou, V., Baillie-Benson, P., Jaklin, M., Kustermann, S., Arias, A.M., and Moris, N. (2021). In vitro teratogenicity testing using a 3D, embryo-like gastruloid system. Reprod. Toxicol. *105*, 72–90. https://doi.org/10.1016/j.reprotox.2021.08.003.

45. Nhieu, J., Lin, Y.L., and Wei, L.N. (2020). Noncanonical retinoic acid signaling. Methods Enzymol. *637*, 261–281. https://doi.org/10.1016/bs.mie.2020.02.012.

46. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., et al. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490–495. https://doi.org/10.1038/s41586-019-0933-9.

47. Cao, J., Spielmann, M., Qiu, X., Huang, X., Ibrahim, D.M., Hill, A.J., Zhang, F., Mundlos, S., Christiansen, L., Steemers, F.J., et al. (2019). The single-cell transcriptional landscape of mammalian organogenesis. Nature *566*, 496–502. https://doi.org/10.1038/s41586-019-0969-x.

48. Sanaki-Matsumiya, M., Matsuda, M., Gritti, N., Nakaki, F., Sharpe, J., Trivedi, V., and Ebisuya, M. (2022). Periodic formation of epithelial somites from human pluripotent stem cells. Nat. Commun. *13*, 2325. https://doi.org/10.1038/s41467-022-29967-1.

49. Ctortecka, C., Hartlmayr, D., Seth, A., Mendjan, S., Tourniaire, G., and Mechtler, K. (2022). An automated workflow for multiplexed single-cell proteomics sample preparation at unprecedented sensitivity. Preprint at bioRxiv. https://doi.org/10.1101/2021.04.14.439828.

50. Budnik, B., Levy, E., Harmange, G., and Slavov, N. (2018). SCoPE-MS: mass spectrometry of single mammalian cells quantifies proteome heterogeneity during cell differentiation. Genome Biol. *19*, 161. https://doi.org/10.1186/s13059-018-1547-5.

51. Fulcher, J.M., Markillie, L.M., Mitchell, H.D., Williams, S.M., Engbrecht, K.M., Moore, R.J., Cantlon-Bruce, J., Bagnoli, J.W., Seth, A., Paša-Tolić, L., and Zhu, Y. (2022). Parallel measurement of transcriptomes and proteomes from same single cells using nanodroplet splitting. Preprint at bioRxiv. https://doi.org/10.1101/2022.05.17.492137.

52. Aibar, S., González-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., et al. (2017). SCENIC: single-cell regulatory network inference and clustering. Nat. Methods *14*, 1083–1086. https://doi.org/10.1038/nmeth.4463.

53. Correia, K.M., and Conlon, R.A. (2000). Surface ectoderm is necessary for the morphogenesis of somites. Mech. Dev. *91*, 19–30. https://doi.org/10.1016/s0925-4773(99)00260-9.

54. Rifes, P., Carvalho, L., Lopes, C., Andrade, R.P., Rodrigues, G., Palmeirim, I., and Thorsteinsdóttir, S. (2007). Redefining the role of ectoderm in

somitogenesis: a player in the formation of the fibronectin matrix of presomitic mesoderm. Development *134*, 3155–3165. https://doi.org/10.1242/dev.003665.

55. Capdevila, J., Tabin, C., and Johnson, R.L. (1998). Control of dorsoventral somite patterning by Wnt-1 and beta-catenin. Dev. Biol. *193*, 182–194. https://doi.org/10.1006/dbio.1997.8806.

56. Maruhashi, M., Van De Putte, T., Huylebroeck, D., Kondoh, H., and Higashi, Y. (2005). Involvement of SIP1 in positioning of somite boundaries in the mouse embryo. Dev. Dyn. *234*, 332–338. https://doi.org/10.1002/dvdy.20546.

57. Van de Putte, T., Maruhashi, M., Francis, A., Nelles, L., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2003). Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. Am. J. Hum. Genet. *72*, 465–470. https://doi.org/10.1086/346092.

58. Di Filippo, E.S., Costamagna, D., Giacomazzi, G., Cortés-Calabuig, Á., Stryjewska, A., Huylebroeck, D., Fulle, S., and Sampaolesi, M. (2020). Zeb2 Regulates Myogenic Differentiation in Pluripotent Stem Cells. Int. J. Mol. Sci. *21*, 2525. https://doi.org/10.3390/ijms21072525.

59. Zweier, C., Albrecht, B., Mitulla, B., Behrens, R., Beese, M., Gillessen-Kaesbach, G., Rott, H.D., and Rauch, A. (2002). "Mowat-Wilson" syndrome with and without Hirschsprung disease is a distinct, recognizable multiple congenital anomalies-mental retardation syndrome caused by mutations in the zinc finger homeo box 1B gene. Am. J. Med. Genet. *108*, 177–181. https://doi.org/10.1002/ajmg.10226.

60. Doerr, A. (2015). DIA mass spectrometry. Nat. Methods *12*, 35. https://doi.org/10.1038/nmeth.3234.

61. Tiwary, S., Levy, R., Gutenbrunner, P., Salinas Soto, F., Palaniappan, K.K., Deming, L., Berndl, M., Brant, A., Cimermancic, P., and Cox, J. (2019). High-quality MS/MS spectrum prediction for data-dependent and data-independent acquisition data analysis. Nat. Methods *16*, 519–525. https://doi.org/10.1038/s41592-019-0427-6.

62. Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc Natl Acad Sci U S A *107*, 21931–21936. https://doi.org/10.1073/pnas.1016071107.

63. Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. Cell *168*, 442–459. https://doi.org/10.1016/j.cell.2016.12.016.

64. Lee, B.K., Jang, Y.J., Kim, M., LeBlanc, L., Rhee, C., Lee, J., Beck, S., Shen, W., and Kim, J. (2019). Super-enhancer-guided mapping of regulatory networks controlling mouse trophoblast stem cells. Nat Commun *10*, 4749. https://doi.org/10.1038/s41467-019-12720-6.

65. Suz, S. (2024). western blots of miniTurboID- or degron-tagged mESCs. Mendeley Data. vV1. https://doi.org/10.17632/2ypp9yfwfr.1.

66. Hansen, M., Varga, E., Wüst, T., Brouwer, N., Beauchemin, H., Mellink, C., van der Kevie-Kersemaekers, A.M., Möröy, T., van der Reijden, B., von Lindern, M., and van den Akker, E. (2017). Generation and characterization of human iPSC line MML-6838-Cl2 from mobilized peripheral blood derived megakaryoblasts. Stem Cell Res. *18*, 26–28. https://doi.org/10.1016/j.scr.2016.12.004.

67. Schmid-Burgk, J.L., Höning, K., Ebert, T.S., and Hornung, V. (2016). CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. Nat. Commun. *7*, 12338. https://doi.org/10.1038/ncomms12338.

68. Birkhoff, J.C., Korporaal, A.L., Brouwer, R.W.W., Nowosad, K., Milazzo, C., Mouratidou, L., van den Hout, M.C.G.N., van IJcken, W.F.J., Huylebroeck, D., and Conidi, A. (2023). Zeb2 DNA-Binding Sites in Neuroprogenitor Cells Reveal Autoregulation and Affirm Neurodevelopmental Defects, Including in Mowat-Wilson Syndrome. Genes (Basel) *14*, 629. https://doi.org/10.3390/genes14030629.

69. Grand, R.S., Burger, L., Gräwe, C., Michael, A.K., Isbel, L., Hess, D., Hoerner, L., Iesmantavicius, V., Durdu, S., Pregnolato, M., et al. (2021). BANP opens chromatin and activates CpG-island-regulated genes. Nature *596*, 133–137. https://doi.org/10.1038/s41586-021-03689-8.

70. Guidi, C.D. (2021). Generation of induced pluripotent stem cells (iPSCs) lines deficient for genes associated with neurodevelopmental diseases using CRISPR/Cas9 technology. https://www.diva-portal.org/smash/get/diva2:1564336/FULLTEXT01.pdf.

71. Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K., and Kuhn, R. (2015). Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. Nat Biotechnol *33*, 543–548. https://doi.org/10.1038/nbt.3198.

72. Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. Nat. Methods *6*, 359–362. https://doi.org/10.1038/nmeth.1322.

73. Rappsilber, J., Mann, M., and Ishihama, Y. (2007). Protocol for micro-purification, enrichment, pre-fractionation and storage of peptides for proteomics using StageTips. Nat. Protoc. *2*, 1896–1906. https://doi.org/10.1038/nprot.2007.261.

74. Post, H., Penning, R., Fitzpatrick, M.A., Garrigues, L.B., Wu, W., MacGillavry, H.D., Hoogenraad, C.C., Heck, A.J., and Altelaar, A.F. (2017). Robust, Sensitive, and Automated Phosphopeptide Enrichment Optimized for Low Sample Amounts Applied to Primary Hippocampal Neurons. J Proteome Res. *16*, 728–737. https://doi.org/10.1021/acs.jproteome.6b00753.

75. Santos-Barriopedro, I., van Mierlo, G., and Vermeulen, M. (2021). Off-the-shelf proximity biotinylation for interaction proteomics. Nat. Commun. *12*, 5015. https://doi.org/10.1038/s41467-021-25338-4.

76. Mao, Y., Chen, P., Ke, M., Chen, X., Ji, S., Chen, W., and Tian, R. (2021). Fully Integrated and Multiplexed Sample Preparation Technology for Sensitive Interactome Profiling. Anal. Chem. *93*, 3026–3034. https://doi.org/10.1021/acs.analchem.0c05076.

77. Lachmann, A., and Ma'ayan, A. (2009). KEA: kinase enrichment analysis. Bioinformatics *25*, 684–686. https://doi.org/10.1093/bioinformatics/btp026.

78. Yu, G., Wang, L.G., Han, Y., and He, Q.Y. (2012). clusterProfiler: an R package for comparing biological themes among gene clusters. Omics *16*, 284–287. https://doi.org/10.1089/omi.2011.0118.

79. vanderAa, C., and Gatto, L. (2023). The Current State of Single-Cell Proteomics Data Analysis. Curr. Protoc. *3*, e658. https://doi.org/10.1002/cpz1.658.

80. Singh, A.A., Schuurman, K., Nevedomskaya, E., Stelloo, S., Linder, S., Droog, M., Kim, Y., Sanders, J., van der Poel, H., Bergman, A.M., et al. (2019). Optimized ChIP-seq method facilitates transcription factor profiling in human tumors. Life Sci. Alliance *2*, e201800115. https://doi.org/10.26508/lsa.201800115.

81. van der Sande, M., Frölich, S., Schäfers, T., Smits, J.G.A., Snabel, R.R., Rinzema, S., and van Heeringen, S.J. (2023). Seq2science: an end-to-end workflow for functional genomics analysis. PeerJ *11*, e16380. https://doi.org/10.7717/peerj.16380.

82. Lerdrup, M., and Hansen, K. (2020). User-Friendly and Interactive Analysis of ChIP-Seq Data Using EaSeq. Methods Mol. Biol. *2117*, 35–63. https://doi.org/10.1007/978-1-0716-0301-7_2.

83. Kuilman, T., Velds, A., Kemper, K., Ranzani, M., Bombardelli, L., Hoogstraat, M., Nevedomskaya, E., Xu, G., de Ruiter, J., Lolkema, M.P., et al. (2015). CopywriteR: DNA copy number detection from off-target sequence data. Genome Biol. *16*, 49. https://doi.org/10.1186/s13059-015-0617-1.

84. Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

85. Zheng, G.X.Y., Terry, J.M., Belgrader, P., Ryvkin, P., Bent, Z.W., Wilson, R., Ziraldo, S.B., Wheeler, T.D., McDermott, G.P., Zhu, J., et al. (2017).

Massively parallel digital transcriptional profiling of single cells. Nat. Commun. *8*, 14049. https://doi.org/10.1038/ncomms14049.

86. Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W.M., 3rd, Zheng, S., Butler, A., Lee, M.J., Wilk, A.J., Darby, C., Zager, M., et al. (2021). Integrated analysis of multimodal single-cell data. Cell *184*, 3573–3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.

87. Ouyang, J.F., Kamaraj, U.S., Cao, E.Y., and Rackham, O.J.L. (2021). ShinyCell: simple and sharable visualization of single-cell gene expression data. Bioinformatics *37*, 3374–3376. https://doi.org/10.1093/bioinformatics/btab209.

88. Vianello, S.D.G., Mehmet, Rossi, G., and Lutolf, M. (2020). Protocol to immunostain Gastruloids (LSCB, EPFL). protocols.io. https://doi.org/10.17504/protocols.io.7tzhnp6.

89. Gritti, N., Lim, J.L., Anlaş, K., Pandya, M., Aalderink, G., Martínez-Ara, G., and Trivedi, V. (2021). MOrgAna: accessible quantitative analysis of organoids with machine learning. Development *148*, dev199611. https://doi.org/10.1242/dev.199611.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
anti-HA clone 3f10	Roche	11867423001; RRID:AB_390918
anti-rat Ig/HRP	Agilent Dako	P0450; RRID:AB_2630354
anti-V5	Invitrogen	R960CUS; RRID:AB_2792973
anti-p300	Santa Cruz	sc-585; RRID:AB_2231120
Streptavidin-HRP	Invitrogen	S911
anti-HA	Invitrogen	71-5500; RRID:AB_2533988
anti-HOXC8	Sigma	HPA028911; RRID:AB_10602236
anti-HOXB4	DSHB	I12; RRID:AB_2119288
FluoTag®-X2 anti-TagFP AlexaFluor647	Nanotag Biotechnologies	N0502-AF647; RRID:AB_3075936
Goat anti-Rat Alexa Fluor 647	Invitrogen	A-21247; RRID:AB_141778
Goat anit-Rabbit IgG Alexa Fluor 647	Invitrogen	21245; RRID:AB_141775
IRDye 680RD Donkey anti-Mouse IgG	LI-COR Biosciences	926-68072; RRID:AB_10953628
IRDye 800CW Goat anti-Rabbit IgG	LI-COR Biosciences	926-32211; RRID:AB_621843
Chemicals, peptides, and recombinant proteins		
Leukemia inhibitory factor (LIF)	produced in-house	N/A
Gelatin from bovine skin	Sigma	G9391
CHIR99021	Axon Medchem	1386
Matrigel	Corning	356231
dTAG-13	Tocris Biosciences	6605
ROCK-inhibitor	Sigma	Y0503
All-trans-Retinoic acid	Sigma	R2625
SB431542	Stem cell Technologies	72232
DMH1	Stem cell Technologies	73632
bFGF	PeproTech	AF-100-18B
StemFit Basic04 Complete Type	Ajinomoto group	N/A
Vitronectin XF	Stem Cell Technologies	100-0763
NDiff227	TaKaRa	Y40002
Dodecyl Maltoside (DDM)	Sigma	D5172
Biotin	Life technologies	B20656
Streptavidin Sepharose High-Performance beads	Cytiva	15511301
anti-HA agarose beads	Sigma	A2095
Ethidium bromide	Sigma	E1510
Disuccinimidyl glutarate	Thermo Fisher Scientific	20593
Complete protease inhibitors	Roche	45-4693132001
Iodoacetamide	Sigma	1149
TMT18	Thermo Fisher Scientific	A52045
Hydroxylamine	Sigma	467804
Trypsin	Promega	V5113
Critical commercial assays		
ProteoCHIP	Cellenion	CPS-1216-3
Nucleofector Lonza Kit P3	Lonza	V4XP-3024
KAPA RNA HyperPrep Kit	KAPA Biosystems	08105952001
KAPA RiboErase Kit	KAPA Biosystems	07962274001
KAPA ChIP HyperPrep Kit	KAPA Biosystems	07962363001

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Chromium Single Cell 3' Reagent kit (v3.1 Chemistry)	10x genomics	1000128
Deposited data		
Time course proteomes, germ layer proteomes, phosphoproteome and proximity labeling	This study	PXD041309
Single cell proteomics	This study	PXD041328, PXD048347
RNA-seq, ChIP-seq and scRNA-seq	This study	GSE229029
Time course RNA-seq of mouse gastruloids	Beccari et al.[4]	GSE106227
P300 ChIP-seq from mESCs	Creyghton et al.[62] Chronis et al.[63] Lee et al.[64]	GSM594600, GSM594601, GSM2417169, GSM3019270
scRNA-seq mouse gastruloids and embryos	Suppinger et al.[8] Pijuan-Sala et al.[46] Van den Brink et al.[5]	GSE229513, GSE123187
Original western blot images	Mendeley data	Suz[65] (https://doi.org/10.17632/2ypp9yfwfr.1)
Experimental models: Cell lines		
dual Bra-GFP/Sox17-RFP mESC reporter	Pour et al.[26]	N/A
MML-6838-Cl2 human iPSCs	Hansen et al.[66]	N/A
Oligonucleotides		
knock-in gRNA for *Mt1*: ACAGCACGTGCACTTGTCCG	Schmid-Burgk et al.[67]	N/A
knock-in gRNA for *Ep300*: TGTCTAGTGTACTCTGTGAG	N/A	N/A
knock-in gRNA for *Paxip1:* CCATCAGTTAAATTTATAT	N/A	N/A
knock-in gRNA for *Zeb2*: GGAAACCAAATCAGACCACG	Birkhoff et al.[68]	N/A
knock-in gRNA for *Sp5*: CGCGGGACCTATGAGCGCAC	N/A	N/A
knock-in gRNA for *Rxrα*: GGCACCACATCAAGCCACCT	N/A	N/A
bacterial gRNA: GTGTTGTGGACTGCGGCGGTCGG	Grand et al.[69]	N/A
knockout gRNA for *ZEB2*: CATTGGCCTCTGGCGTGCCA and TTGTAGCCCCGGTCGCAGTA	Guidi[70]	N/A
PCR primers	Table S5	N/A
Recombinant DNA		
pU6-(BbsI)_CBh-Cas9-T2A-mCherry	Chu et al.[71]	addgene #64324
pCAS9-mCherry-Frame +2	Schmid-Burgk et al.[67]	addgene #66941

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be provided by the lead contact, Michiel Vermeulen (michiel.vermeulen@science.ru.nl).

Materials availability

Plasmids and cell lines generated in this study are available upon request.

Data and code availability

- ChIP-seq, RNA-seq, scRNA-seq have been deposited to Gene Expression Omnibus (GEO) database: GSE229029. Single cell proteomics data have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier: PXD041328 and PXD048347 (BRA-GFP+ cells). All other mass spectrometry data have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the dataset identifier PXD041309. Original western blot images have been deposited at Mendeley and are publicly available as of the date of publication. The DOI is listed in the key resources table (https://doi.org/10.17632/2ypp9yfwfr.1).
- The code to analyze single cell proteomics data is available at: https://github.com/PSobrevalsAlcaraz/SCP_Stelloo.et.al.2023. Interactive visualization of processed datasets is available at https://mouse-gastruloids-omics.shinyapps.io/gtlshiny/.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Cell lines

The dual BRA-GFP (Brachyury), SOX17-RFP reporter mouse ES cell line[26] was cultured on 0.15% (w/v) gelatin-coated dishes in Dulbeccco's modified Eagle's medium (DMEM, Gibco, 31966) supplemented with 15% fetal bovine serum (Biowest, S1810), 2mM GlutaMAX (Gibco, 35050038), sodium pyruvate (Gibco, 11360039), non-essential amino acids (Gibco, 11140035), 50U/mL penicillin-streptomycin (Gibco, 15140122), 100μM β-mercaptoethanol (Sigma, M3148) and leukemia inhibitory factor (LIF; produced in-house). MML-6838-Cl2 human induced pluripotent stem cells[66] were cultured on vitronectin coated plates in StemFit Basic04CT medium (Ajinomoto) supplemented with 50U/mL penicillin-streptomycin (Gibco). All lines were routinely tested for mycoplasma contamination and copy number profiles were assessed in the derived cell lines of the triple reporter cell line (Data S1).

Generation of cell lines

For the generation of the triple reporter cell line, CRISPaint method[67] was used to endogenously tag the ectoderm marker gene *Mt1* with tagBFP. The CRISPaint universal donor plasmid (addgene #80970) was modified to in-frame insert T2A-tagBFP followed by a BGH poly(A) signal and a bleomycin resistance cassette. The guide RNA near the stopcodon of *Mt1* (5′-ACAGCACGTG CACTTGTCCG-3′) was cloned into pU6-(BbsI)_CBh-Cas9-T2A-mCherry vector (addgene #64324). The dual reporter cell line was transfected with the guide RNA plasmid, donor plasmid and frame selector 2 (addgene #66941) using Lipofectamine 3000 according to the manufacturer's protocol (Invitrogen). After 10 μg/ml zeocin (Invivogen, ant-zn-1) selection, clones were validated with Sanger sequencing.

For endogenous tagging of *Ep300*, a homology donor plasmid was generated containing a left homology arm of 700bp, a linker sequence, miniTurboID, V5, T2A-puromycin and a right homology arm of 300bp. The guide RNA used for *Ep300* tagging is 5′-TGTCTAGTGTACTCTGTGAG-3′. The guide RNA 5′-CCATCAGTTAAATTTATAT-3′ was used for tagging *Paxip1* with miniTurboID. For the generation of the degron cell lines, we used homology donor plasmids with a 3x HA sequence and $FKBP12^{F36V}$ flanked by 150bp homology arms. The target guide RNAs used are 5′-GGAAACCAAATCAGACCACG-3′ for *Zeb2*,[68] 5′-CGCGGGACCTAT GAGCGCAC-3′ for *Sp5* and 5′-GGCACCACATCAAGCCACCT-3′ for *Rxrα*. Cells were transfected with the donor plasmid, target guide RNA, and a bacterial guide RNA (5′-GTGTTGTGGACTGCGGCGGTCGG-3′)[69] to linearize the donor plasmid. Two days after transfection, $mCherry^{+}$ cells were single cell-sorted into 96 well plates, expanded and genotyped. Primers used for verification of edited loci in single clones are listed in Table S5.

To generate a *ZEB2* knockout in MML-6838-Cl2 iPSCs, we applied the same strategy as described previously.[70] In brief, MML-6838-Cl2 iPSCs were transfected with two guides (5′-CATTGGCCTCTGGCGTGCCA-3′ and 5′-TTGTAGCCCCGGTCGCAGTA-3′ cloned in addgene plasmid #48139) using an Amaxa nucleofector II device with program DS-137 and nucleofector Lonza Kit P3. Medium was supplemented with 10μM Y-27632 ROCK-inhibitor (Sigma, Y0503) for 24 h. At 48 h post transfection, the cells were treated with 0.25μg/mL puromycin (Invivogen, ant-pr-1) for 24h. Single cell derived clones were screened with primers listed in Table S5.

Mouse embryos

All animals were housed and bred according to institutional guidelines, and procedures were performed in compliance with Standards for Care and Use of Laboratory Animals with approval from the Hubrecht Institute ethical review board.

METHOD DETAILS

Mouse gastruloids and somites

Gastruloids were generated as previously described.[5,18] Briefly, 300 single cells were FACS sorted with a BD FACSMelody™ cell sorter into U-bottomed 96 wells (Greiner Bio-one; 650185) containing 40μL NDiff227 (Takara, Y40002). After 48 h, the aggregates were treated with 3μM CHIR99021 (Axon 1386) and medium was refreshed every day with pre-equilibrated NDiff227. For gastruloids embedded in matrigel, four '96 h gastruloids' were transferred with a P1000 pipette with the tip cut-off to 1.5mL Eppendorf tubes and resuspended in 10% matrigel (Corning, 356231) and transferred to a 48-well Greiner Bio-one; 677180). For time lapse imaging, plates were placed in a temperature and CO_2 chamber (37°C and 5% CO_2) and brightfield images were acquired with 10 min intervals for ~24 h using an Okolab 2.0. Brightfield images were acquired with a Zeiss Axiophot microscope (10x objective) with an Axiocam camera connected to a computer running Zen software. For degradation of the $FKBP12^{F36V}$ degron tagged transcription factors, cell lines were treated with 100nM dTAG-13 (Tocris Biosciences, 6605) or DMSO as control. For experiments with all-trans-Retinoic acid (ATRA, Sigma, R2625), ATRA was added to Ndiff227 medium at 0 h, 48 h, 72 h and 96 h during gastruloid differentiation at the indicated concentrations.

Mouse embryoid bodies (EBs)

mESCs were cultured in medium without LIF in bacterial dishes at a density of 1 million cells/dish. Medium was changed every other day and cells were harvested at day 6. Pelleted cells were lysed in 5 volumes (cell pellet size) of RIPA buffer (150mM NaCl, 50mM Tris pH 8.0, 1mM EDTA, 10% glycerol, 1% NP40, 1mM DTT and complete protease inhibitors (CPI, Roche)) and further processed as described for HA pulldown below.

Human somitoids

Human somitoids were generated as previously described.[48] iPSCs were sorted with a BD FACSMelody™ cell sorter into U-bottomed 96 wells (Greiner Bio-one; 650970) (350cells/well) containing 50ul somitoid induction medium (NDiff227 (Takara) supplemented with 10μM SB431542 (Stem cell Technologies, 72232), 10μM CHIR99021 (Axon 1386), 2μM DMH1 (Stem cell technologies, 73632), and 20ng/ml bFGF (PeproTech, AF-100-18B) with 10μM Y-27632 ROCK-inhibitor (Sigma, Y0503). After sorting, the plates were centrifuged for 2 min at 150g. Twenty four hours after sorting, 150μl somitoid medium was added and medium was refreshed with NDiff227 at 48 h and 72 h. At 96 h, medium was replaced with NDiff227 medium containing 10% matrigel (Corning, 356231). Brightfield images were acquired with a Zeiss Axiophot microscope (10x objective) with Axiocam camera connected to a computer running Zen software.

Mouse embryo isolation

Mice were housed in a standard condition with 12:12 h light:dark cycle in temperature-controlled rooms with food and water ad libitum. All mouse embryos were isolated after timed mating of F1 WT strain males and females. 12.00 on the day of an observed vaginal plug was considered as E0.5. During isolation extraembryonic tissues and membranes were removed with tweezers. For embryo collection at E8.5 and E9.5 surgical cuts were made posterior to the heart and only tissues posterior to the cut were collected, while at E7.5 whole embryonic tissue was isolated. For E7.5 embryos, a total of 47 embryos were collected from 6 pregnant mice and distributed into three replicates. The three replicates comprised 16, 11, and 20 embryos. For E8.0, two replicates were conducted, with 19 and 11 tails/trunks analyzed in each, derived from three pregnant mice. Sixteen E8.5 embryos (posterior material), derived from two pregnant mice, were distributed evenly across four replicates. All anterior material of E8.5 embryos were pooled together. For E9.5 stage, individual embryos were harvested from one pregnant mouse. Upon collection in 1.5ml Eppendorf tubes, all liquid was removed with a 30G needle and samples were snap frozen in liquid nitrogen for further processing.

(Phospho)proteomics

To harvest gastruloids for (phospho)proteomics experiments, 96-well plates were flipped upside down in 3D printed collection plates and centrifuged for 3 min at 300g. The gastruloids were washed twice with ice-cold PBS, snapfrozen with liquid nitrogen and stored at -80°C. Concurrently with harvesting gastruloids for one of the time course experiments, we collected gastruloids for RNA-seq. For germ layer specific proteomes, the collected gastruloids were dissociated by pipetting up and down with pre-warmed trypsin-EDTA. Trypsin was inactivated with DMEM supplemented with 15% FBS. Subsequently, cells were washed twice with ice-cold PBS and resuspended in PBS. Sorting was performed on a BD FACSAria™ instrument and 100,000 cells per replicate were collected in protein LoBind tubes (Thermo scientific, 90410) containing 200μL PBS. Sorted cells were centrifuged at 4°C for 10 min at 300×*g* and supernatant was removed with a 30-gauge needle with syringe.

Frozen pellets were lysed with SDS lysis buffer (4% SDS, 1mM DTT, 100mM Tris pH 7.5) and incubated for 3 min at 95°C. Samples were sonicated for five cycles (30 s on/30 s off, Bioruptor Pico). Filter Aided Sample-Preparation (FASP)[72] was performed on a total of 20μg of protein per sample for whole proteomes or 50μg of protein per sample for phosphoproteomes. Cell lysates were mixed with Urea buffer (8M urea, 0.1M Tris pH 8.5, 5mM DTT) and loaded onto Centrifugal Filters (Microcon, MRCF0R030). Filters were washed twice with Urea buffer, incubated with 0.05M iodoacetamide in Urea buffer for 10 min, and washed three times more with Urea buffer followed by three washes with 0.05M ammonium bicarbonate (Fluka Analytical, 09830). Samples were digested overnight with trypsin (0.1μg/ml, 1:100) at 37°C. For whole proteomes, peptide solutions were acidified with trifluoroacetic acid (TFA) and desalted using C18 Stagetips.[73] For phosphoproteomics, phosphorylated peptides were enriched using Fe(III)-NTA 5μL cartridges on the AssayMap BRAVO platform (Agilent Technologies).[74]

Streptavidin pulldowns

Cells were treated with 50μM biotin (B20656, Life technologies) for 1 h at 37°C. Cells were harvested by scraping and gastruloids by centrifugation. Pelleted cells were washed twice with PBS and resuspended in five volumes of RIPA buffer (150mM NaCl, 50mM Tris pH 8.0, 1mM EDTA, 10% glycerol, 1% NP40, 1mM DTT and CPI). Lysates were rotated for 1 h at 4°C and centrifuged for 5 min at 21,000g at 4°C. The supernatant was collected and protein concentration was measured using BCA protein assay (Thermo FIsher Scientific). Approximately 500μg of protein per reaction was incubated with 15μl prewashed Streptavidin Sepharose High-Performance beads (15511301, Cytiva) and 2μl of ethidium bromide (Sigma, E1510) for 2 h in a rotation wheel at 4°C. The beads were washed twice with RIPA buffer, twice with PBS + 1% NP40 and twice with PBS. Beads were resuspended in Laemmli buffer (120mM Tris, 20% Glycerol, 4% SDS, 100mM DTT and CPI) and boiled for 10 min at 95°C for western blot. Antibodies used are streptavidin-HRP (Invitrogen, S911), P300 (Santa Cruz Biotechnology, sc-585), V5 (Invitrogen, R960), IRDye secondary antibodies (LI-COR Biosciences, 926-68072 and 926-32211). As described previously, for mass spectrometry, beads were resuspended in 50μl elution buffer (2M Urea, 100mM Tris pH 8.0, 10mM DTT) and incubated for 20 min in a thermo shaker at 1,250rpm at room temperature.[75] Iodoacetamide (50mM) was added to the beads and further incubated for 10 min in a thermo shaker at 1,250rpm in the dark. Then, 0.25μg trypsin was added, followed by incubation in a thermo shaker at 1,250rpm for 2 h at room temperature. The supernatant was collected to new tubes and further digested overnight at room temperature with an additional 0.1μg of trypsin. Peptide solutions were acidified with TFA and desalted using C18 Stagetips. For visualization of P300 proximity labeling data in the ShinyApp, we used the code provided on https://github.com/FredHutch/interactiveVolcano.git.

DNA pulldown

DNA oligonucleotides containing the HOX motif or a mutated HOX motif with the forward strand containing a 5'- biotin moiety were ordered from Integrated DNA Technologies (Table S5). The forward oligo (20μM) and 30μM of the complementary reverse oligo were mixed in annealing buffer (10mM HEPES pH 8.0, 50mM NaCl and 1mM EDTA). Oligos were annealed by heating to 95°C for 10 min before cooling to room temperature. Subsequently, 10pmol of annealed oligo were incubated with prewashed 1-3μl streptavidin sepharose beads for 30 min in a rotation wheel at 4°C. Two C18 discs were inserted into a 200μL pipette tip, which was then placed onto a homemade adapter. The following steps were performed as previously described for one-tip pulldowns.[76] Steps: (1) conditioning the C18 with 60μl of MeOH at 1000g for 1 min; (2) blocking the C18 with 60μl of 2% (w/w) SDS at 1500g for 1 min; (3) loading the DNA oligos coupled onto the streptavidin beads at 1500g for 1 min; (4) twice 50μg protein lysate loading at 100g for 30 min or longer when necessary; (5) washing with 60μl of NP40 buffer (1% NP40, 50 mM Tris pH 7.4, 150 mM NaCl) at 1500g for 5 min for four times; (6) activation of C18 with 60μl of Buffer B (80% (v/v) acetonitrile and 0.1% (v/v) formic acid) at 1500g for 1 min; (7) reduction by loading 20μL of 10mM TCEP and 50mM triethylammonium bicarbonate buffer (TEAB, Sigma #T7408) for 15 min at room temperature (RT) and then discarding the solution at 1000g for 30s; (8) digestion and alkylation by loading 2μl of digestion buffer containing 0.25 μg/μl of trypsin (Promega #V5113), 20mM chloroacetamide, and 50mM TEAB for 1 h at 37°C; (9) washing with 60μl buffer A (0.1% formic acid) at 1500g for 1 min; (10) dimethyl labeling twice with 150μl of labeling reagent (16.2μl 37% CH2O (light) or 30.0 μL 20% CD2O (medium) plus 6mg sodium cyanoborohydride in 3mL of labeling buffer (10mM NaH2PO4, 35mM Na2HPO4) at 1500g for 10 min; (11) washing with 60μL buffer A (0.1% formic acid) at 1500g for 1 min. Labeled samples were eluted with Buffer B while combining the respective light and medium labeled pairs into the same tube.

LC-MS/MS and downstream analysis

Phosphoproteomics samples were analyzed using an Ultimate 3000 uHPLC system coupled to an Orbitrap Exploris 480 (Thermo Fisher Scientific). Peptides were separated using a nanoflow rate of 300 nL/min on an analytical column (ID of 75μm and 50cm length; packed in-house with 2.7μm Poroshell EC-C18 particles (Agilent)). We used a two-system buffer consisting of solvent A (0.1% Formic Acid in water) and B (0.1% Formic Acid in 80% ACN). A 98 min gradient from 9% to 36% of solvent B, followed by 5 min wash with 99% solvent B and a 10 min column equilibration with 9% solvent B was used. MS1 scans were acquired from 375 to 1600 m/z at 60,000 resolution (at 200 m/z). RF lens (%) was set to 40 and the AGC target was set to 'standard' with the maximum injection time mode set to 'auto'. A minimum intensity threshold of 50,000 was used to trigger an MS2 scan. Precursors were selected for fragmentation with an isolation window of 1.4 m/z. AGC target and injection window were also set as 'standard' and 'auto' for MS2 scans. Precursors were fragmented with an HCD collision energy of 28%. MS2 scans were acquired from 120 m/z with a 30,000 resolution at 200 m/z. Precursors were added to the dynamic exclusion list for 16 seconds after being fragmented once. Raw MS spectra from phosphoproteomics samples were searched using MaxQuant software (version 2.0.1.0) against a mouse UniProt database (fasta file downloaded 201708). The default settings were used, with the following exceptions: methionine oxidation, protein N-term acetylation and phosphorylation of serine, threonine and tyrosine were set as variable modifications. Cysteine carbamidomethylation was set as a fixed modification. 'Match between runs' was enabled with the default parameters, and fractions were set so that matching was only done between replicates. Phosphosite data was analyzed using Perseus software. Potential contaminants, reverse sequences (decoys), and phosphorylation sites with <0.75 localization probability score were filtered out. Only phosphosites quantified in at least three replicates from the same condition were kept for further analysis, and missing values were replaced from a normal distribution using the default settings. Intensities were $\log_2$ transformed and subjected to ANOVA multiple sample test with Benjamini-Hochberg correction. Significantly changing phosphosites were z-scored for visualization. Significantly downregulated and upregulated sites (based on t-test results with a $\log_2$ fold change cutoff of > 2 and a student t-test p-value of <0.05) were submitted to KEA2.[77] A list of transcription factors was obtained from the AnimalTFDB v4.0 database.[22]

Whole proteomes, proximity labeling experiments and DNA pulldowns were analyzed using an online Easy-nLC 1000 (Thermo Fisher Scientific) coupled to an Orbitrap Exploris 480 (Thermo Fisher Scientific). MS1 spectra were acquired at 120,000 resolution with a scan range from 350 to 1300 m/z, normalized AGC target of 300% and maximum injection time of 20ms. The top 20 most intense ions with a charge state 2-6 from each MS1 scan were selected for fragmentation by HCD. MS2 resolution was set at 15,000 with a normalized AGC target of 75%. Raw MS spectra were analyzed using MaxQuant software (version 1.6.0.1) with standard settings. For dimethyl labelled samples, the respective built in N-terminal and lysine modification for dimethyl labeling was specified under ''labels''. Data was searched against the mouse UniProt database (fasta file downloaded 201706) using the integrated search engine. Potential contaminants, reverse sequences, 'only identified by site' and proteins identified by only one peptide were excluded from the analysis using Perseus software. Proteins quantified in all triplicates of at least one sample group were considered for downstream analysis. Next, missing LFQ values were imputed for statistical analysis using 'replace missing values from normal distribution' function with default settings. Data obtained from the two biological time course experiments, each comprising three technical replicates per timepoint, were merged and subjected to the ''removeBatchEffect'' function in limma Rpackage before statistical analysis. Statistical analysis was performed using ANOVA multiple sample test with Benjamini-Hochberg correction for multiple hypothesis testing and Student's t-test for proximity labeling experiments. Gene ontology analysis was performed using the gseGO and enrichGO functions of R package clusterProfiler.[78]

Single cell proteomics

50nl of master mix (0.2% DDM (Cayman Chemical Company), 100mM TEAB (Sigma-Aldrich), 10ng/μl LysC&Trypsin (Promega)) was dispensed in each nanowell of a proteoCHIP (Cellenion, France) using the cellenONE (Cellenion, France), at 22°C and at 85% humidity.[49] Gastruloids were resuspended in degassed PBS (Sigma) and single cells with a diameter between 13 and 30μm and maximum elongation of 2.0 were dispensed in each well. A subset of gastruloids was first sorted for fluorescence expressing germ layer cells using FACS (Aria III), followed by further processing for single cell proteomics on the cellenONE. Morphological parameters were recorded during cell dispensing. Another 50nl master mix was added and the temperature of the heating block was elevated to 50°C for 2 h to allow protein digestion. For TMTpro 18 labeling, 100mM TMT (all channels except for 126 and 127C) in anhydrous acetonitrile was deposited in each well and incubated at room temperature for 1 h. The carrier was created by mixing endoderm, ectoderm, mesoderm, unsorted cells, and mESCs at equal ratios, and lysed following the same protocol as the single cell samples. The carrier sample was labeled with TMT126 and a 20-cell equivalent was printed with the cellenONE to each single cell sample before mass spectrometry analysis. The 127C channel was left empty to avoid potential contamination from the 126 carrier channel. Quenching was performed by dispensing 50 nL of 0.5% hydroxylamine (Sigma) and incubating for 15 min. Samples were acidified and diluted with 150nL of 2% formic acid, and pooled via centrifugation at 1,500rpm for 2 min to the proteoCHIP funnel. ProteoCHIP funnels were incubated at 4°C to freeze the hexadecane oil layer and cleaned up. Eluted peptide mixes were vacuum-dried *in* and resuspended in buffer A (0.1% formic acid) for mass spectrometry analysis.

Samples were separated on a 20-cm pico-tip column (50μm ID, New Objective) packed *in house* with C-18 material (1.9μm aquapur gold, dr. Maisch) using a two-step 140 min gradient (5% to 25% ACN/0.2% FA in 85 min, and to 45% in 30 min) using an easy-nLC 1200 system (Thermo Fisher Scientific). Peptides were electro-sprayed directly into an Orbitrap Eclipse Tribrid Mass Spectrometer (Thermo Fisher Scientific). The column temperature was maintained at 45°C using a column oven (Sonation). Spray voltage was set to 2.1kV, funnel RF level at 60, and the transfer capillary temperature at 275°C. The FAIMS device was set at standard resolution and a carrier gas flow of 3.8, and a constant CV of -50. The MS was operated in DDA mode, and full scans were acquired in the Orbitrap with a resolution of 120,000 and a scan range from 375-1200 m/z, with an AGC target of 3e6 and an automatically determined maximum injection time. Up to 10 most intense precursor ions were selected for HCD fragmentation at a normalized collision energy of 32%, after reaching the AGC target of 1e5 or maximum injection time of 118 ms. MS/MS was acquired at a resolution of 60,000, with an exclusion duration of 120s and with a fixed first mass of 110 m/z.

RAW data files were processed using Thermo Proteome Discoverer (version 2.4) and Sequest HT search engine allowing for variable methionine oxidation, protein N-terminal acetylation, and protein N-terminal methionine loss. TMTpro of peptide N-termini and lysine residues were set as static modifications. The protein database consisted of the Uniprot TrEMBL protein database of Mus Musculus. Enzyme specificity was set for trypsin, with a maximum of two allowed missed cleavages. The precursor mass tolerance was set at 10ppm, and fragment mass tolerance was set to 0.02Da. TMTpro 18 reporter ion quantification was performed using the Reporter Ions Quantifier node, with a co-isolation threshold of 50% and Minimal Channel Occupancy of 0. Results were filtered using a 1% FDR cut-off at the protein and peptide level.

The downstream analysis pipeline was adapted from SCP R Package.[79] Proteins were screened to ensure they contained at least two unique peptides. The peptide spectrum matches files were then extracted from Thermo Proteome Discoverer. Quality control was performed by selecting peptide spectrum matches fitting the expected single cell to carrier ratio and a false discovery rate (FDR) of 0.01. The filtered peptide spectrum matches were normalized to the sum of the single cells and combined at the peptide level using their median. To explore the peptide data, the median relative intensity and median coefficient of variation per cell were computed. The peptide data was normalized using median and was log transformed, and the data was aggregated to the protein level using their median. The normalized protein data was subjected to a filtering process based on missingness, where proteins found present in at least 70% of the cells of one cell type were extracted for subsequent analysis. The remaining proteins were imputed using random draws from a manually defined distribution relative to the original data distribution. To address batch effects, *ComBat* function from the *sva* package (version 3.38.0) was used with proteoCHIPs serving as batch covariates and accounting for the different known cell types in the model matrix. Imputation was performed separately for each experiment to account for the stochasticity of the DDA mode. ANOVA analysis was then conducted on the imputed data, correcting for multiple testing with the Benjamini-Hochberg method and selecting proteins with an adjusted p-value of less than 0.01. The remaining proteins were used for the downstream analysis, including PCA and UMAP. To test robustness of clustering, the data was randomly shuffled and applied to the same downstream analysis.

HA pulldown followed by western blotting

Approximately 500μg of protein per reaction was incubated with 15μl prewashed anti-HA agarose beads (Sigma, A2095) and 2μl of ethidium bromide (Sigma, 10mg/ml) for 90 min in a rotation wheel at 4°C. The beads were washed twice with RIPA buffer, twice with PBS + 1% NP40 and twice with PBS. Beads were resuspended in Laemmli buffer (120mM Tris, 20% Glycerol, 4% SDS, 100mM DTT, CPI and bromophenol blue) and boiled for 10 min at 95°C after which proteins were separated on an SDS–PAGE gel. Proteins were transferred from the gel to a nitrocellulose membrane using wet transfer. Nitrocellulose membranes were blocked with 5% milk in 0.1% Tween-PBS for 30 min followed by 1 h primary antibody (anti-HA clone 3f10, Roche, 11867423001) and 1h secondary antibody (anti-rat Ig/HRP, Agilent Dako, P0450). The experiment was performed once.

ChIP-seq

Chromatin immunoprecipitations were performed as described previously.[80] In brief, trypsinized gastruloids were crosslinked in solution A with 2mM disuccinimidyl glutarate (Thermo Fisher Scientific #20593) for 25 min, followed by 1% formaldehyde methanol-free (Thermo Fisher Scientific, 28906) for 20 min. Chromatin extracts were sonicated for 2-4 cycles of 30 sec on, 30 sec off using a Diagenode Bioruptor Pico. For each ChIP, 4μg of V5 Tag antibody (Thermo Fisher Scientific, R960CUS) or 4μg of HA antibody (a mix of 2μg from Roche 11867423001 and 2μg from Invitrogen 71–5500) was pre-conjugated to 40ul of Protein A/G magnetic beads (Invitrogen, 10009D and 10008D). Immunoprecipitated DNA was processed for library preparation using the KAPA HyperPrep Kit (KAPA Biosystems, 07962363001), barcoded with NEXTflex DNA barcodes (IDT technologies) and paired-end sequenced using an Illumina NextSeq500. Sequence reads were processed using the seq2science pipeline (v0.7.1).[81] Briefly, paired-end reads were trimmed with fastp (v0.20.1, default settings) and aligned with bwa-mem2 (v2.2.1, options '-M') to the Mus musculus genome assembly GRCm38.p6. Aligned reads were filtered based on mapping quality (MAPQ $\geq$ 30) and filtering out duplicate reads and reads in the ENCODE blacklisted regions. Peaks were called with MACS2 (v2.2.7) with default settings in BAMPE mode. The peaks called in both biological replicates were used for analysis. Genome browser snapshots were generated with Easeq (v1.111),[82] motif analysis was performed using the SeqPos motif tool with default settings and genomic region enrichment analysis was performed with CEAS (http://cistrome.org/ap/). Publicly available ChIP-seq data used in this study is available from GEO (GSE24164, GSE90893 and GSE110950) and re-analyzed with the seq2science pipeline. CopywriteR Rpackage (v1.0.2) was used to obtain copy number profiles from input and ChIP samples.[83]

RNA-seq

RNA was extracted using Quick-RNA MicroPrep kit (Zymo Research, R1051) according to the manufacturer's instruction with DNAseI treatment included. Libraries were generated from 40-500ng RNA with the KAPA RNA HyperPrep Kit with RiboErase (KAPA Biosystems, 08105952001 and 07962274001). Fragmentation and priming were performed at 94°C for 6 min. NEXTflex DNA barcodes were used for adaptor ligation. Libraries were amplified with 6-10 PCR cycles followed by a two steps library amplification cleanup using a 0.8x bead-based cleanup and then an 1x bead cleanup. Library size was determined using the High Sensitivity DNA bioanalyzer kit on a Bioanalyzer 2100 system (Agilent) and concentration was measured using the dsDNA High Sensitivity Assay (Denovix). Libraries were paired-end sequenced (42bp or 59bp) on an Illumina NextSeq500 or NextSeq2000. Sequence reads were processed using the seq2science pipeline (v0.7.1).[81] In short, paired-end reads were trimmed with fastp (v0.20.1, default settings) and aligned with STAR (v2.7.6a, default settings) to the Mus musculus genome assembly GRCm38.p6. Aligned reads were filtered based on a minimum mapping quality of 255 and duplicate reads and reads in the ENCODE blacklisted regions were filtered out. Sample sequencing strandedness was inferred using RSeQC (v4.0.0) and number of reads per gene were measured with HTSeq count (v0.12.4). Deseq2 (v1.34.0) was used for downstream analysis.[84] Genes with low counts were filtered out and the lfcShrink function was used to obtain differentially expressed genes. Publicly available RNA-seq data used in this study is available from GEO (GSE106227) and analyzed with the seq2science pipeline.

RT-qPCR

Extracted RNA was used for cDNA synthesis using iScript cDNA Synthesis Kit (Bio-Rad, 1708891). Real-time PCR analysis was performed using iQ SYBR Green Supermix (Bio-Rad, 1708886) and run on a CFX96 Real-Time system (Bio-Rad). The data were normalized to housekeeping gene TBP. Primer sequences are listed in Table S5.

Ananse analysis

For gene regulatory network analysis for gastruloids and mECS the computational approach ANANSE was used (v0.4.0).[28] Briefly, the ananse binding command line was used to predict specific transcription factor binding using P300 ChIP-seq data from mESCs and gastruloids. To build the gene regulatory networks we used 1) the output from ananse binding, 2) RNA-seq data processed to obtain gene-level transcripts per million (TPM) values from gastruloids and mESCs (3) GRCm38 genome assembly. To obtain the TPM files the reads from the RNA seq data were quantified using Kallisto (v0.48.0) and the index file was build using the Ensemble mouse transcriptome GRCm38. The influence score was calculated using the gene regulatory networks from gastruloids and mESC, to generate a differential regulatory network. The number of top edges used for this differential network was 500 (-i 500.000).

Single-cell RNA-seq

Gastruloids (~32) were harvested from matrigel with icecold PBS and spun down at 300g for 5 min at 4°C. Gastruloids were trypsinized for 5 min at 37°C and resuspended in DMEM supplemented with 15% FBS. Cells were washed with cold PBS and resuspended in PBS containing 7-AAD (Invitrogen, 00-6993-50) and 0.04% BSA. Live cells were sorted at 4°C with a BD FACSMelody™ cell sorter in DNA LoBind tubes containing PBS with 0.04% BSA and centrifuged at 300g for 5 min. Per sample, 10,000 cells were processed using the 10x Genomics Chromium Controller and the Chromium Single Cell 3' Reagent kit (v3.1 Chemistry) following the standard manufacturer's protocol. The resulting libraries were sequenced on an Illumina Nextseq500 (1x 28bp + 1x56bp + 1x8bp index). Raw sequencing data was processed using Cell Ranger (v6.1.2, 10x Genomics)[85] and further processed with Seurat (v4.1.0).[86] Seurat objects were created with the parameters min.cell=3 and min.features=200 set. In addition, cells with at least 2000 but no more than 8000 genes detected and less than 10% mitochondrial reads were kept. To integrate the two conditions, scTransform data, using the SelectIntegrationFeatures(), PrepSCTIntegration(), FindIntegrationAnchors(), and IntegrateData() functions with default

options were run. PCA was then run on the top 3000 variable genes and the data was then clustered (resolution = 0.5 and dimensions=30). Differentially expressed genes were identified with Seurat's PrepSCTFindMarkers. Cell type and embryo stage annotation was performed using SeuratFindTransferAnchors and TransferData functions using the mouse gastrulation atlas as a reference (R package MouseGastrulationData).[46] Monocle3 was used to order single cells along pseudotime.[47] For visualization of the scRNA data in the ShinyApp, we used the ShinyCell R package.[87]

Immunofluorescence

Immunofluorescence was performed as previously described.[88] For gastruloids grown in matrigel, the medium was removed and the gastruloids were harvested in icecold PBS and washed twice with PBS. Gastruloids were fixed in 4% formaldehyde in PBS while rocking for 2 h at 4°C. Samples were incubated for 1 h in PBS containing 10% FBS and 0.2% Triton-X100 followed by 24 h incubation with primary antibody and then secondary Alexa Fluor 647 antibody (Invitrogen, A-21247 or A-21245) at 4°C for 24 h. Preconjugated Alexa Fluor 647 anti-TagFP (Nanotag Biotechnologies, N0502-AF647) was used to stain TagBFP protein. The following primary antibodies were used: anti-HA clone 3f10, anti- HOXC8 (Sigma, HPA028911) and anti-HOXB4 (DSHB antibodies, I12). For validation of the P300-miniTurboID expressing cell line, mESCs were grown on gelatin-coated coverslips for a maximum of 48h. Cells were fixed with 4% formaldehyde for 10 min, washed twice with PBS, permeabilized with 0.3% Triton-X100 in PBS for 10 min and blocked with 0.5% BSA in PBS for 30 min. Cells were incubated with anti-V5 (Invitrogen, R960) for 1 h followed by PBS washes and 1 h incubation with Avidin-FITC (Invitrogen, A821) and secondary Alexa Fluor 647 antibody. Excess antibodies were removed with PBS washes. Coverslips were mounted with Fluoromount G mounting medium with DAPI (Invitrogen, 00-4959-52). Images were acquired using a Zeiss LSM880 with ZEN software.

Perimeter quantification

Perimeter was quantified with the software MOrgAna (version 0.1.1).[89] For this a training set ranging from 10 to 50 images with their respective masks was used to train the model with default settings, downscaling of 0.25, edge size of 2, pixel extraction of 0.5 and extraction bias of 0.5. Masks were inspected and those that showed an incorrect mask were manually mapped. The resulting data was exported and analyzed with R.

QUANTIFICATION AND STATISTICAL ANALYSIS

The statistical details of experiments can be found in the figure legends, figures, results, and method details.

Cell Stem Cell, Volume *31*

Supplemental Information

Deciphering lineage specification during early embryogenesis in mouse gastruloids using multilayered proteomics

Suzan Stelloo, Maria Teresa Alejo-Vinogradova, Charlotte A.G.H. van Gelder, Dick W. Zijlmans, Marek J. van Oostrom, Juan Manuel Valverde, Lieke A. Lamers, Teja Rus, Paula Sobrevals Alcaraz, Tilman Schäfers, Cristina Furlan, Pascal W.T.C. Jansen, Marijke P.A. Baltissen, Katharina F. Sonnen, Boudewijn Burgering, Maarten A.F.M. Altelaar, Harmjan R. Vos, and Michiel Vermeulen

Figure S1

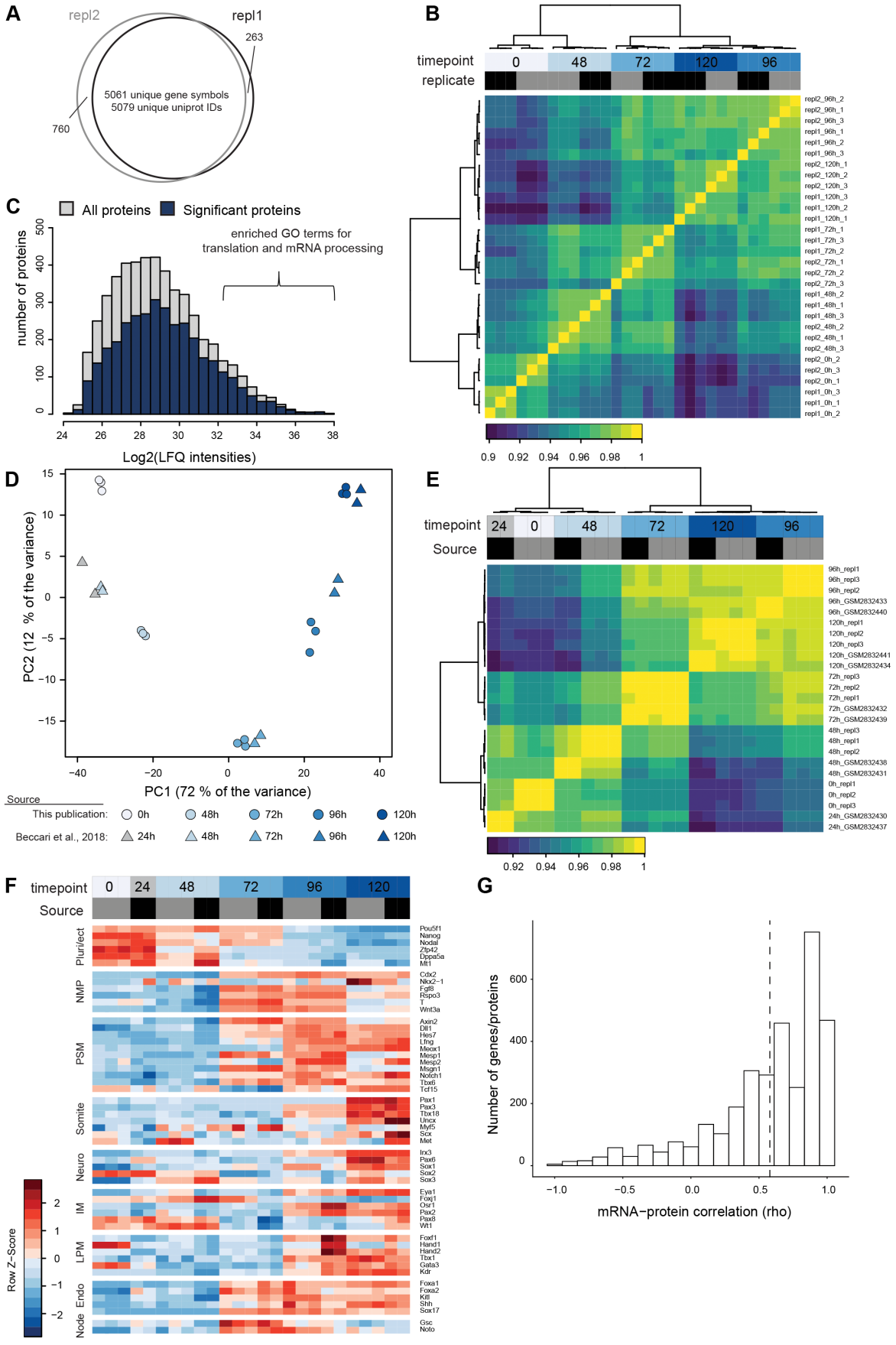

A
repl2
repl1
263
760
5061 unique gene symbols
5079 unique uniprot IDs
B
timepoint
replicate
0
48
72
120
96
0.9
0.92
0.94
0.96
0.98
1
C
All proteins
Significant proteins
enriched GO terms for
translation and mRNA processing
number of proteins
Log2(LFQ intensities)
D
PC2 (12 % of the variance)
PC1 (72 % of the variance)
Source
This publication:
Beccari et al., 2018:
0h
24h
48h
72h
96h
120h
E
timepoint
Source
24
0
48
72
120
96
F
timepoint
Source
Pluri/ect
NMP
PSM
Somite
Neuro
IM
LPM
Endo
Node
Row Z-Score
G
Number of genes/proteins
mRNA−protein correlation (rho)

Figure S1 Proteomic and transcriptomic profiling during gastruloid differentiation, related to Figure 1

(A) Overlap of quantified proteins (proteins quantified in all triplicates of at least one timepoint) from two biological experiments.

(B) Pearson's correlation of log2 normalized LFQ intensities between samples. The samples from the different time points and replicates are color coded and indicated on top of the heatmap.

(C) Histogram of the distribution of log2 normalized protein intensities. All quantified proteins are indicated in grey and significantly expressed proteins in blue.

(D) Principal component analysis after batch correction using the top 1,000 most variable expressed genes. The different blue tints denote the different time points, while the different shapes represent the two RNA-seq time course datasets either from this publication or from GSE106225 [s1].

(E) Heatmap shows the Pearson correlation coefficient of gene expression for all pairwise combinations of samples in the two datasets. The column side color bar for 'Source' labels the different datasets: in black samples from GSE106225[s1] and in grey samples from this publication.

(F) Heatmap of scaled expression of selected genes associated with embryonic development. Pluri, pluripotency; ect, ectoderm; NMP, neuromesodermal progenitors; PSM, presomitic mesoderm; LPM, lateral plate mesoderm; IM, intermediate mesoderm; Endo, endoderm. The column side color bar for 'Source' labels the different datasets: in black samples from Beccari et al.,2018[s1] and in grey samples from this publication.

(G) Histogram showing the distribution of gene-wise mRNA-protein correlations computed as Spearman's Rho (x-axis). The dashed line indicates the median correlation.

Figure S2

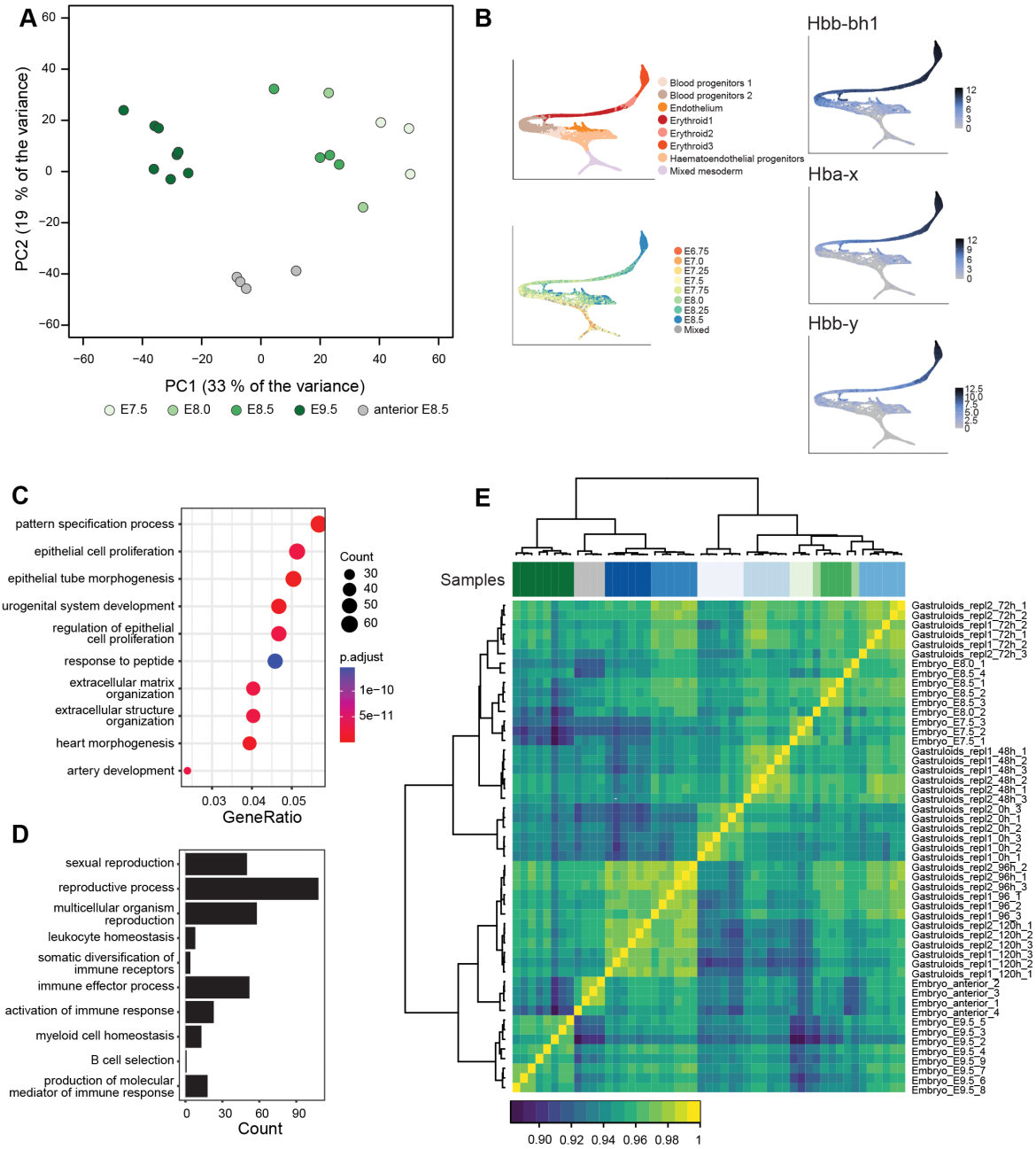

A
PC2 (19 % of the variance)
PC1 (33 % of the variance)
E7.5
E8.0
E8.5
E9.5
anterior E8.5
B
Blood progenitors 1
Blood progenitors 2
Endothelium
Erythroid1
Erythroid2
Erythroid3
Haematoendothelial progenitors
Mixed mesoderm
E6.75
E7.0
E7.25
E7.5
E7.75
E8.0
E8.25
E8.5
Mixed
Hbb-bh1
Hba-x
Hbb-y
C
pattern specification process
epithelial cell proliferation
epithelial tube morphogenesis
urogenital system development
regulation of epithelial cell proliferation
response to peptide
extracellular matrix organization
extracellular structure organization
heart morphogenesis
artery development
GeneRatio
Count
p.adjust
D
sexual reproduction
reproductive process
multicellular organism reproduction
leukocyte homeostasis
somatic diversification of immune receptors
immune effector process
activation of immune response
myeloid cell homeostasis
B cell selection
production of molecular mediator of immune response
Count
E
Samples

Figure S2 Label-free protein quantification in mouse embryos, related to Figure 2

(A) Principal component analysis using all quantified proteins (proteins quantified in all samples of at least one condition).

(B) Cell clustering based on scRNA-seq data of the blood cell lineages (erythroid, haemato-endothelial, blood progenitor, endothelial and mixed mesoderm groups (n=15,875 cells)[S2]. Top left: cells are coloured by cell type. Bottom left: cells are coloured by embryonic stage. Right: Expression of embryonic globin genes in each cell.

(C) GO terms associations with proteins detected exclusively in mouse embryos and not gastruloids as obtained through the enrichGO function using clusterProfiler.

(D) GO terms associations with proteins detected exclusively in mouse embryos and not gastruloids as obtained through the groupGO function using clusterProfiler.

(E) Heatmap shows the Pearson correlation coefficient of protein expression for all pairwise combinations of samples. In the column color side bar, mouse embryo samples are shown in green, while gastruloid samples are denoted in blue, with the color intensity indicating the developmental time from light to dark.

Figure S3

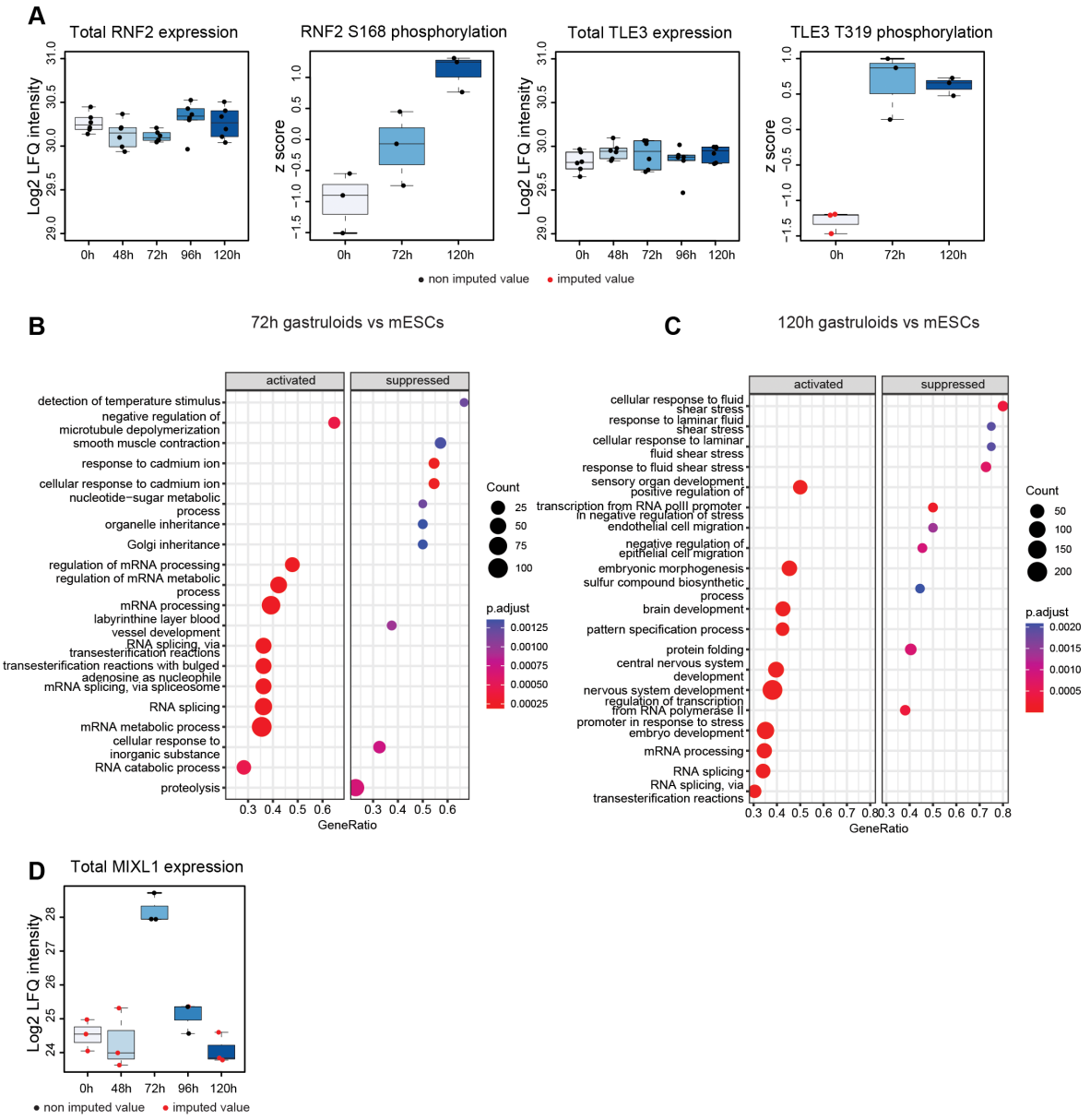


Figure S3 Phosphoproteomic analysis in gastruloids, related to Figure 3

(A) Total protein expression (log2 LFQ intensity) and phosphoprotein expression (z-score) for RNF2 and TLE3. Each dot represent a sample and red dots indicate imputed values.

(B) gseGO results showing the top 10 suppressed and activated biological processes in 72h gastruloids. A ranked list of all phosphorylated proteins was used as input, in which duplicates were removed keeping the entries with the largest absolute fold change. The size of each dot shows the number of enriched genes in each term. The color of each dot represents the adjusted p-value.

(C) gseGO results showing the top 10 suppressed and activated biological processes in 120h gastruloids.

(D) Total protein expression (log2 LFQ intensity) for MIXL1. Each dot represent a sample and red dots indicate imputed values.

Figure S4

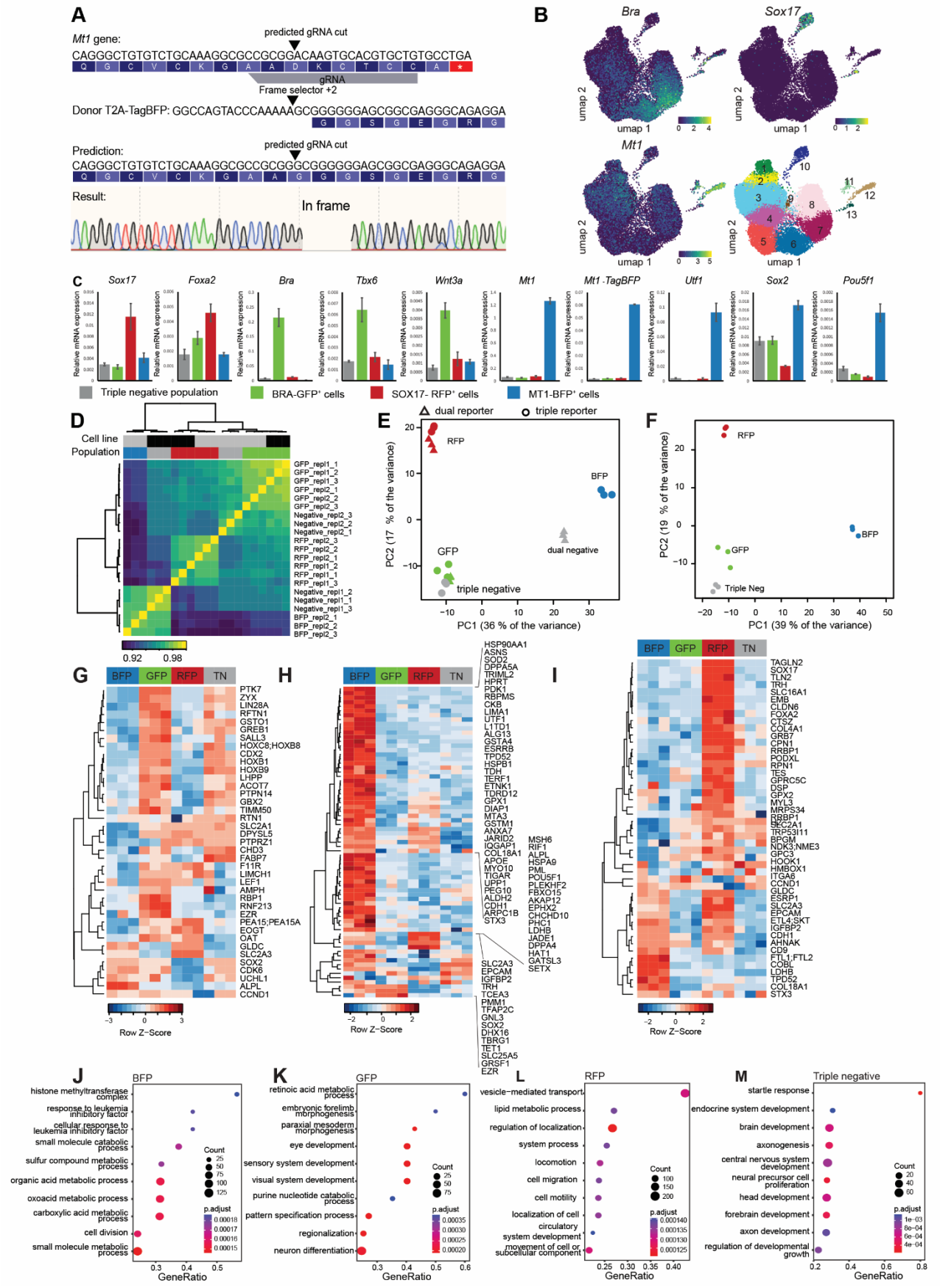

A
Mt1 gene:
predicted gRNA cut
CAGGGCTGTGTCTGCAAAGGCGCCGCGGACAAGTGCACGTGCTGTGCCTGA
gRNA
Frame selector +2
Donor T2A-TagBFP: GGCCAGTACCCAAAAAGCGGGGGGAGCGGCGAGGGCAGAGGA
Prediction:
CAGGGCTGTGTCTGCAAAGGCGCCGCGGGCGGGGGGAGCGGCGAGGGCAGAGGA
Result:
In frame
B
Bra
Sox17
Mt1
umap 1
umap 2
C
Sox17
Foxa2
Bra
Tbx6
Wnt3a
Mt1
Mt1 -TagBFP
Utf1
Sox2
Pou5f1
Relative mRNA expression
Triple negative population
BRA-GFP+ cells
SOX17- RFP+ cells
MT1-BFP+ cells
D
Cell line
Population
E
dual reporter
triple reporter
RFP
BFP
GFP
dual negative
triple negative
PC1 (36 % of the variance)
PC2 (17 % of the variance)
F
PC1 (39 % of the variance)
PC2 (19 % of the variance)
Triple Neg
G
H
I
BFP
GFP
RFP
TN
Row Z-Score
J
BFP
K
GFP
L
RFP
M
Triple negative
Count
p.adjust
GeneRatio

Figure S4 Tagging MT1 with TagBFP to generate a germ layer reporter mESC line, related to Figure 4

(A) Schematic of endogenous *Mt1* gene tagging with T2A-TagBFP in the dual BRA/SOX17 mESC reporter. The top panel shows the gRNA target sequence and location. The middle panel shows the predicted in-frame tagging with frame selector 2. The bottom panel shows the Sanger sequencing chromatogram of genomic fusions of *Mt1* with T2A-TagBFP.

(B) UMAPs showing the expression of *Bra*, *Sox17* and *Mt1* in publicly available scRNA-seq dataset from 120h gastruloids[3]. Cluster annotation: 1) Cardiac, 2) Paraxial mesoderm, 3) Differentiated somite, 4) Somite, 5) Differentiation front, 6) PSM, 7) neuromesodermal progenitors, 8) Spinal cord, 9) Mesenchyme, 10) Endothelium, 11) Allantois, 12) PGC-like/EXE ectoderm and 13) Endoderm.

(C) Relative mRNA levels of endoderm marker genes (*Sox17*, *Foxa2*), mesoderm marker genes (*T*, *Tbx6*, *Wnt3a*) and ectoderm marker genes (*Mt1*, *Utf1*, *Sox2*, *Pou5f1*). Fusion specific primers (forward primer annealing to Mt1 and the reverse primer to TagBFP) were used to detect Mt1-TagBFP mRNA. Data are represented as mean ± standard error.

(D) Heatmap shows the Pearson correlation coefficient of protein expression (batch corrected) for all pairwise combinations of samples. The column side color bar for 'Cell line' labels the two experiments: in black samples from the dual reporter cell line and in grey the samples from the triple reporter cell line.

(E) Principal component analysis after batch correction using the top 1,000 most variable expressed proteins.

(F) Principal component analysis using the LFQ intensities of the top 1,000 most variable proteins obtained from the triple negative (grey), BRA-GFP$^+$ (green), SOX17-RFP$^+$ (red) and MT1-TagBFP$^+$ (blue) cell populations.

(G) Heatmap of proteins encoding marker genes characteristic of cluster 6 presomitic mesoderm (PSM) and 7 neuromesodermal progenitors (NMPs). Of all quantified proteins, 39 proteins out of 135 marker genes were detected.

(H) Heatmap of proteins encoding marker genes characteristic of cluster 12 PGC-like/Exe ectoderm as determined by differential expression analysis between scRNA-seq clusters (ref). Of all quantified proteins, 71 proteins out of 96 marker genes were detected.

(I) Heatmap of proteins encoding marker genes characteristic of cluster 10 endoderm. Of all quantified proteins, 47 proteins out of 91 marker genes were detected.

(J) Dot plot showing the enriched GO terms for the differentially expressed proteins between MT1-BFP$^+$ cells (ectoderm) and all other cells. The size of each dot shows the number of enriched proteins in each term. The color of each dot represents the adjusted p-value.

(K) Dot plot showing the enriched GO terms for the differentially expressed proteins between BRA-GFP$^+$ cells (mesoderm) and all other cells. The size of each dot shows the number of enriched proteins in each term. The color of each dot represents the adjusted p-value.

(L) Dot plot showing the enriched GO terms for the differentially expressed proteins between SOX17-RFP$^+$ cells (endoderm) and all other cells. The size of each dot shows the number of enriched proteins in each term. The color of each dot represents the adjusted p-value.

(M) Dot plot showing the enriched GO terms for the differentially expressed proteins between triple negative cells and all other cells. The size of each dot shows the number of enriched proteins in each term. The color of each dot represents the adjusted p-value.

Figure S5

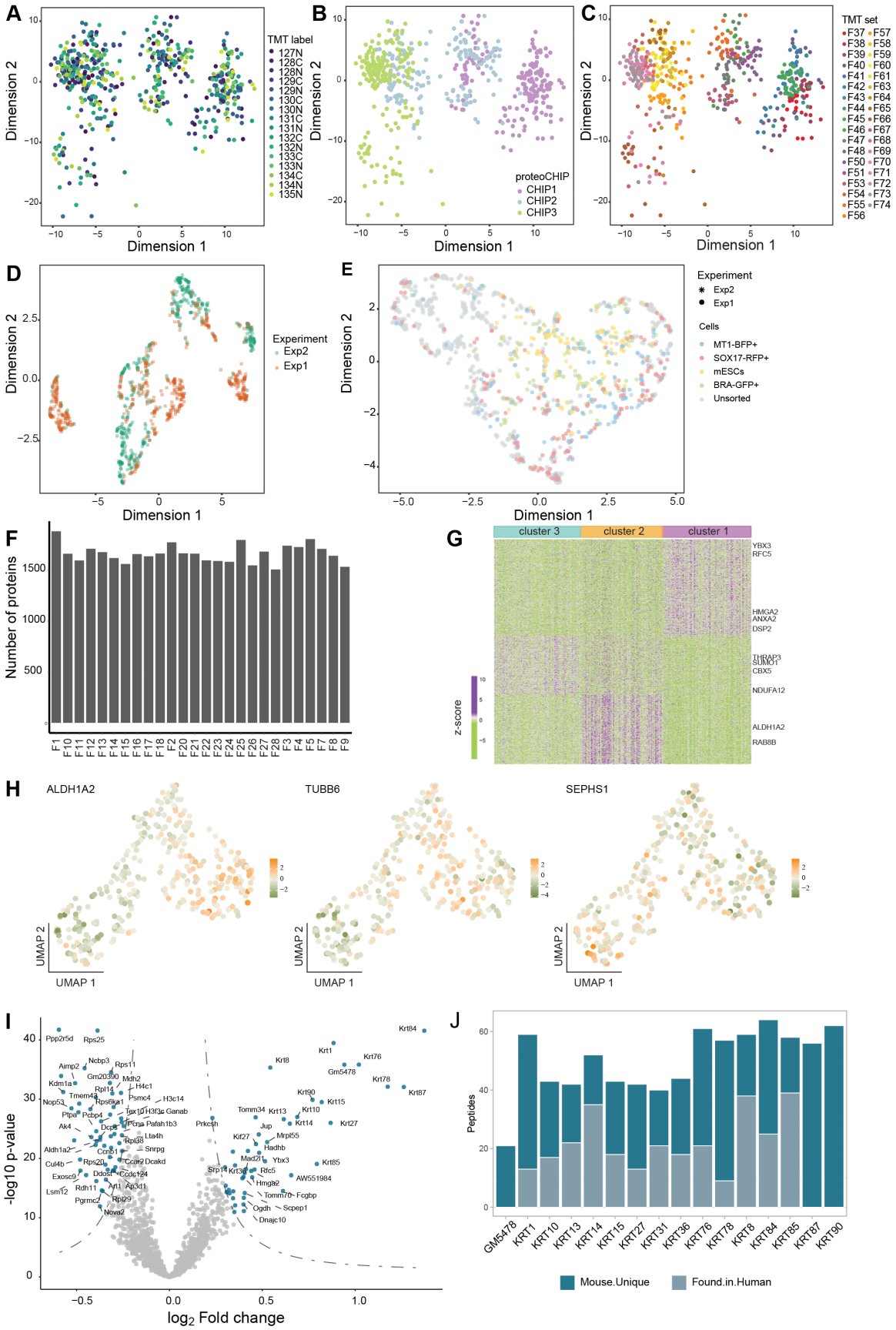

A
B
C
D
E
F
G
H
I
J
Dimension 1
Dimension 2
TMT label
127N
128C
128N
129C
129N
130C
130N
131C
131N
132C
132N
133C
133N
134C
134N
135N
proteoCHIP
CHIP1
CHIP2
CHIP3
TMT set
F37 F57
F38 F58
F39 F59
F40 F60
F41 F61
F42 F63
F43 F64
F44 F65
F45 F66
F46 F67
F47 F68
F48 F69
F50 F70
F51 F71
F53 F72
F54 F73
F55 F74
F56
Experiment
Exp2
Exp1
Cells
MT1-BFP+
SOX17-RFP+
mESCs
BRA-GFP+
Unsorted
Number of proteins
cluster 3
cluster 2
cluster 1
z-score
YBX3
RFC5
HMGA2
ANXA2
DSP2
THRAP3
SUMO1
CBX5
NDUFA12
ALDH1A2
RAB8B
ALDH1A2
TUBB6
SEPHS1
UMAP 1
UMAP 2
-log10 p-value
log2 Fold change
Peptides
Mouse.Unique
Found.in.Human
GM5478
KRT1
KRT10
KRT13
KRT14
KRT15
KRT27
KRT31
KRT36
KRT76
KRT78
KRT8
KRT84
KRT85
KRT87
KRT90

Figure S5 Quality control assessment of single cell proteomics experiments, related to Figure 5

(A) PCA plot showing the distribution of cells based on the expression of 675 proteins. Each dot represents a single cell, and the color indicates the TMT label.

(B) PCA plot showing the distribution of cells based on the expression of 675 proteins. Each dot represents a single cell, and the color indicates the proteoCHIP in which each sample was processed.

(C) PCA plot showing the distribution of cells based on the expression of 675 proteins. Each dot represents a single cell, and the color indicates the different TMT runs (35 multiplexed samples each containing 16 single cells).

(D) UMAP plot colored by the two single cell proteomics experiments. Single cell proteomics using the sorted cells is Exp1 and single cell proteomics using the unsorted cells is Exp2.

(E) UMAP plot based on randomizing the input matrix.

(F) Number of proteins per TMT18 set for the BRA-GFP$^+$ single cell experiment.

(G) Heatmap displaying significantly differentially expressed proteins (ANOVA, $q<0.01$) between the three defined BRA-GFP$^+$ clusters.

(H) UMAPs showing the z-score normalized log2 abundances of ALDH1A2, TUBB6 and SEPHS1.

(I) Volcano plot shows the log2 normalized protein expression between cluster 1 and the other two clusters.

(J) The number of unique peptides for mouse and human keratins.

Figure S6

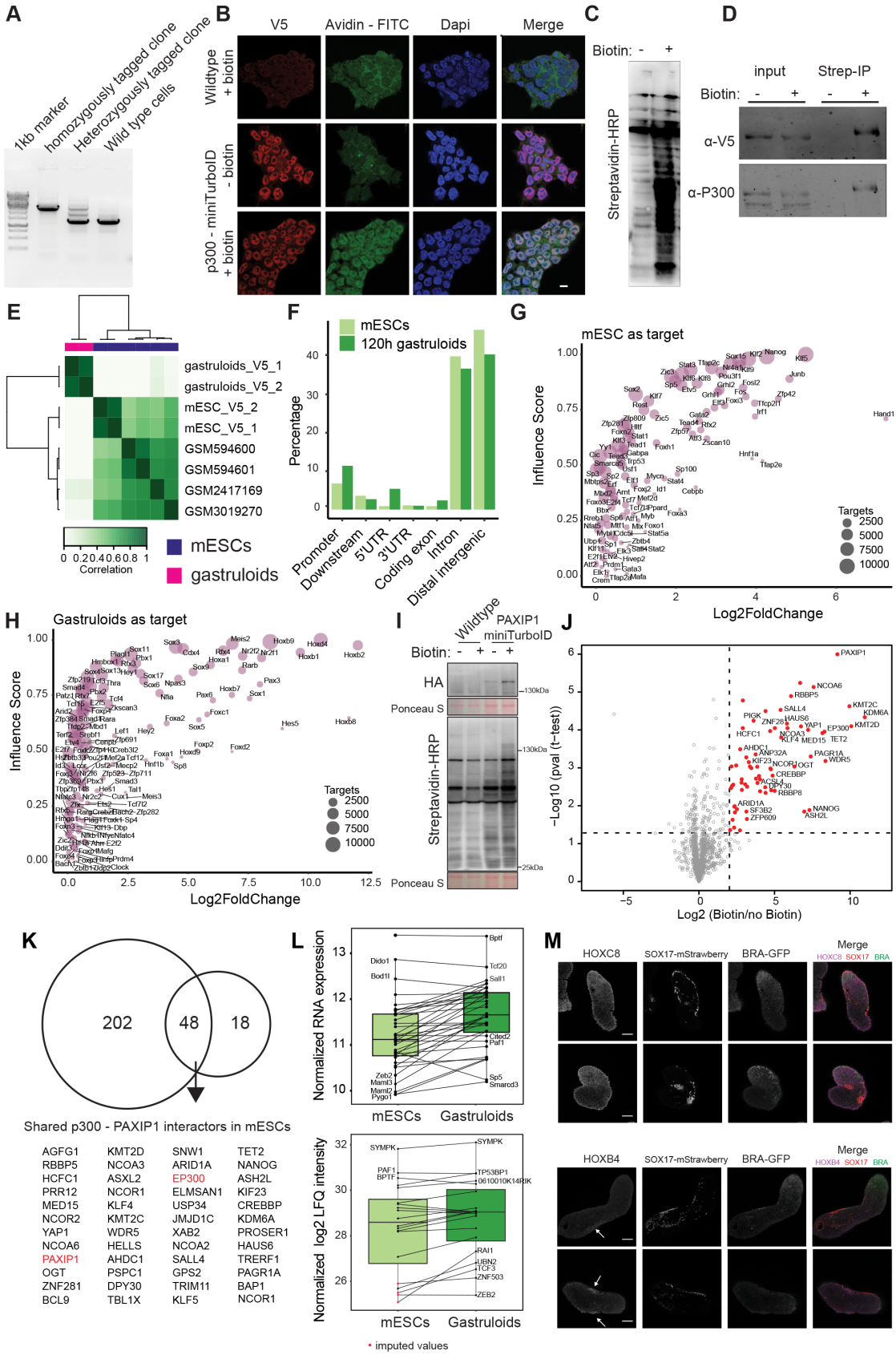

A
1kb marker
homozygously tagged clone
Heterozygously tagged clone
Wild type cells
B
V5
Avidin - FITC
Dapi
Merge
Wildtype + biotin
p300 - miniTurboID - biotin
p300 - miniTurboID + biotin
C
Biotin: - +
Streptavidin-HRP
D
input
Strep-IP
Biotin: - + - +
α-V5
α-P300
E
gastruloids_V5_1
gastruloids_V5_2
mESC_V5_2
mESC_V5_1
GSM594600
GSM594601
GSM2417169
GSM3019270
0 0.2 0.4 0.6 0.8 1
Correlation
mESCs
gastruloids
F
mESCs
120h gastruloids
Percentage
Promoter
Downstream
5'UTR
3'UTR
Coding exon
Intron
Distal intergenic
G
mESC as target
Influence Score
Log2FoldChange
Targets
2500
5000
7500
10000
H
Gastruloids as target
Influence Score
Log2FoldChange
Targets
2500
5000
7500
10000
I
Wildtype
PAXIP1 miniTurboID
Biotin: - + - +
HA
Ponceau S
Streptavidin-HRP
130kDa
25kDa
J
−Log10 (pval (t−test))
Log2 (Biotin/no Biotin)
PAXIP1
K
202
48
18
Shared p300 - PAXIP1 interactors in mESCs
AGFG1 KMT2D SNW1 TET2
RBBP5 NCOA3 ARID1A NANOG
HCFC1 ASXL2 EP300 ASH2L
PRR12 NCOR1 ELMSAN1 KIF23
MED15 KLF4 USP34 CREBBP
NCOR2 KMT2C JMJD1C KDM6A
YAP1 WDR5 XAB2 PROSER1
NCOA6 HELLS NCOA2 HAUS6
PAXIP1 AHDC1 SALL4 TRERF1
OGT PSPC1 GPS2 PAGR1A
ZNF281 DPY30 TRIM11 BAP1
BCL9 TBL1X KLF5 NCOR1
L
Normalized RNA expression
mESCs
Gastruloids
Normalized log2 LFQ intensity
mESCs
Gastruloids
• imputed values
M
HOXC8
SOX17-mStrawberry
BRA-GFP
Merge
HOXC8 SOX17 BRA
HOXB4
SOX17-mStrawberry
BRA-GFP
Merge
HOXB4 SOX17 BRA

Figure S6 P300-V5-miniTurboID mESCs to identify enhancers and enhancer binding proteins, related to Figure 6

(A) PCR to test the integration of V5-miniTurboID-T2A-puro at the *Ep300* genomic locus. The forward and reverse primer bind outside of homology arms resulting in a 2.8 kb fragment in case *Ep300* is tagged. Untagged results in a band of 1.3kb and heterozygously tagged clones will show both bands.

(B) Immunofluorescence staining for V5 and biotinylated proteins in control mESCs stimulated with biotin and P300-v5-miniturboID expressing cells stimulated without and with biotin. Scale bar = 10µm.

(C) Western blot of protein extracts from P300-miniturboID expressing cell line without (-) and with (+) treatment of biotin. Blots were incubated with HRP-Streptavidin.

(D) Streptavidin pulldown using whole cell extracts from P300-miniTurboID expressing mESCs incubated without or with biotin. The western blots were probed with anti-V5 and anti-p300 antibodies.

(E) Correlation heatmap based on peak occupancy. The clustering of the samples represents correlations between individual ChIP-seq samples on the basis of all called peaks.

(F) Genomic distribution of P300 binding across genomic features.

(G) The scatter plot shows the influence score for each transcription factor, which indicates to what extent the variation in gene expression between mESCs and gastruloids can be attributed to a transcription factor. The higher the influence score, the more important the transcription factor is predicted to be for mESCs. On the x-axis, the log2 fold change of RNA expression is plotted between mESCs and gastruloids.

(H) The scatter plot shows the influence score for each transcription factor, which indicates to what extent the variation in gene expression between mESCs and gastruloids can be attributed to a transcription factor. The higher the influence score, the more important the transcription factor is predicted to be for gastruloids. On the x-axis, the log2 fold change of RNA expression is plotted between gastruloids and mESCs.

(I) Western blot of protein extracts from PAXIP1-miniturboID expressing cell line without (-) and with (+) treatment of biotin. Blots were incubated with Streptavidin-HRP and anti-HA. Ponceau S stain serves as loading control.

(J) Volcano plot of proteins identified in PAXIP1 proximity labeling experiments in undifferentiated mESCs. Enrichment of PAXIP1 and proximity interactors is shown as fold enrichment of LFQ intensity of biotin treatment over LFQ intensity of untreated (x-axis) plotted against the −log10 transformed p-value. Significant hits are colored and labeled.

(K) Overlap of significant (log fold change >2 & p-value < 0.05) p300 proximity interactors and PAXIP1 proximity interactors.

(L) RNA and protein expression of gastruloid-enriched P300 proximity hits in undifferentiated mESCs and 120h gastruloids.

(M) Immunofluorescence staining for HOXC8 (top) and HOXB4 (bottom) in 120h gastruloids. Scale bar = 100µm.

Figure S7

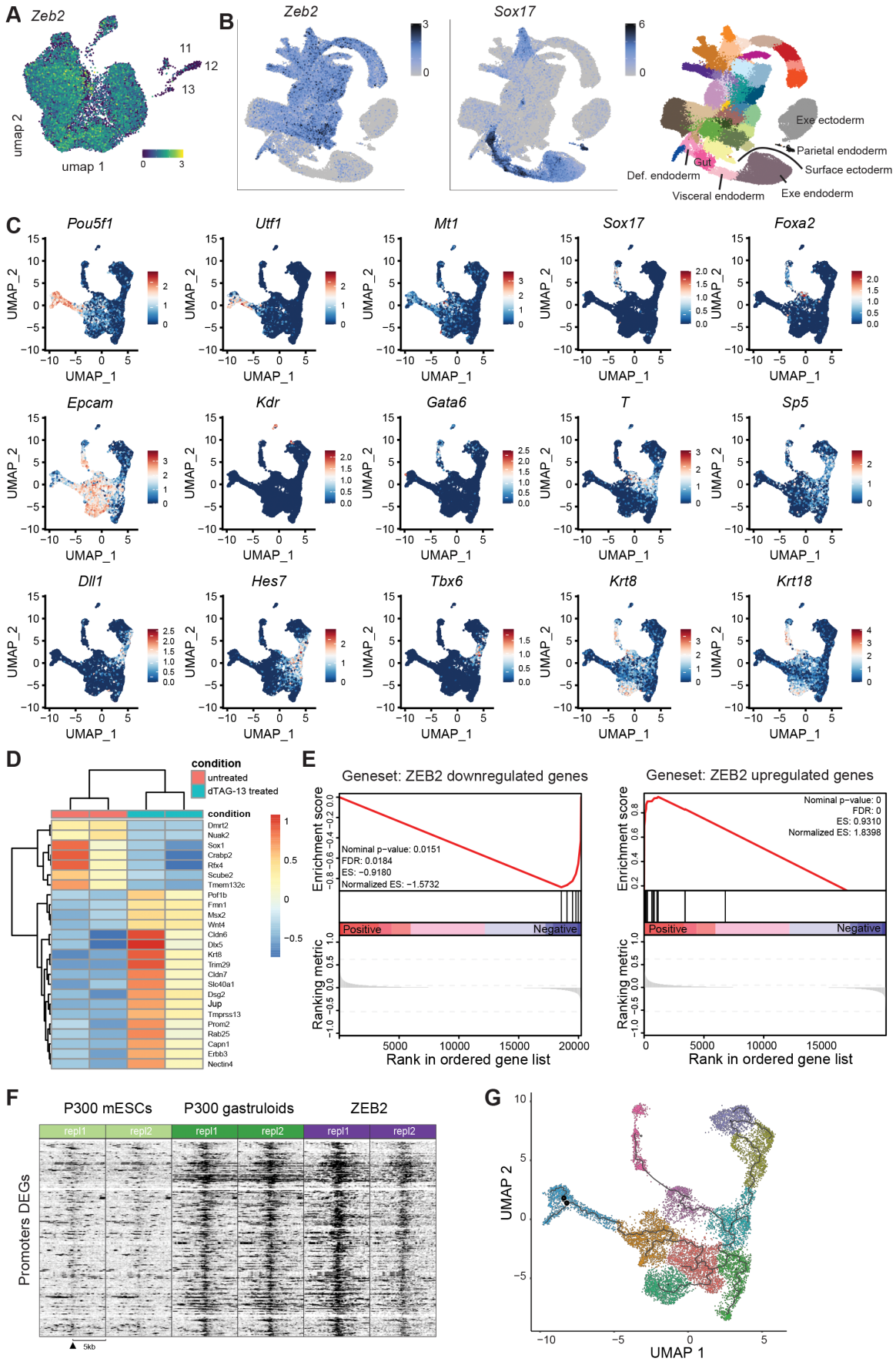

A
Zeb2
umap 2
umap 1
11
12
13
B
Zeb2
Sox17
Exe ectoderm
Parietal endoderm
Surface ectoderm
Exe endoderm
Visceral endoderm
Def. endoderm
Gut
C
Pou5f1
Utf1
Mt1
Sox17
Foxa2
Epcam
Kdr
Gata6
T
Sp5
Dll1
Hes7
Tbx6
Krt8
Krt18
UMAP_1
UMAP_2
D
condition
untreated
dTAG-13 treated
Dmrt2
Nuak2
Sox1
Crabp2
Rfx4
Scube2
Tmem132c
Pof1b
Fmn1
Msx2
Wnt4
Cldn6
Dlx5
Krt8
Trim29
Cldn7
Slc40a1
Dsg2
Jup
Tmprss13
Prom2
Rab25
Capn1
Erbb3
Nectin4
E
Geneset: ZEB2 downregulated genes
Nominal p-value: 0.0151
FDR: 0.0184
ES: −0.9180
Normalized ES: −1.5732
Positive
Negative
Enrichment score
Ranking metric
Rank in ordered gene list
Geneset: ZEB2 upregulated genes
Nominal p-value: 0
FDR: 0
ES: 0.9310
Normalized ES: 1.8398
F
P300 mESCs
P300 gastruloids
ZEB2
repl1
repl2
Promoters DEGs
5kb
G
UMAP 2
UMAP 1

Figure S7 single cell transcriptional changes upon ZEB2 loss in mouse gastruloids, related to Figure 7

(A) UMAP projection of publicly available scRNA-seq dataset of 120h gastruloids colored according to the expression of *Zeb2*[S3].

(B) UMAP projection of scRNA-seq dataset of mouse gastrulation atlas colored according to the expression of *Zeb2* and *Sox17*[S2].

(C) Expression of germ layer marker genes projected on the UMAP of scRNA-seq from somitic gastruloids.

(D) Heatmap showing the expression (z-score) of differentially expressed genes (adjusted p-value < 0.05 & absolute log2 foldchange > 1) between untreated and dTAG13 treated somitic gastruloids generated from Zeb2-HA-FKBP12^{F36V} expressing mESCs in bulk RNA-seq data.

(E) Gene set enrichment analysis of Zeb2 target genes in scRNA-seq data using custom genesets with the significantly differential expressed genes identified in bulk RNA-seq.

(F) Heatmap visualizing ChIP-seq signal (FPKM) for P300 in mESCs and P300 and ZEB2 in 120h gastruloids. Data are centered at promoters of differentially expressed genes, depicting a 5-kb window around the peak.

(G) Trajectories inferred by monocle3.

Supplemental References

S1. Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature *562*, 272-276. 10.1038/s41586-018-0578-0.

S2. Pijuan-Sala, B., Griffiths, J.A., Guibentif, C., Hiscock, T.W., Jawaid, W., Calero-Nieto, F.J., Mulas, C., Ibarra-Soria, X., Tyser, R.C.V., Ho, D.L.L., Reik, W., Srinivas, S., Simons, B.D., Nichols, J., Marioni, J.C., and Gottgens, B. (2019). A single-cell molecular map of mouse gastrulation and early organogenesis. Nature *566*, 490-495. 10.1038/s41586-019-0933-9.

S3. van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivie, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A., and van Oudenaarden, A. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature *582*, 405-409. 10.1038/s41586-020-2024-3.