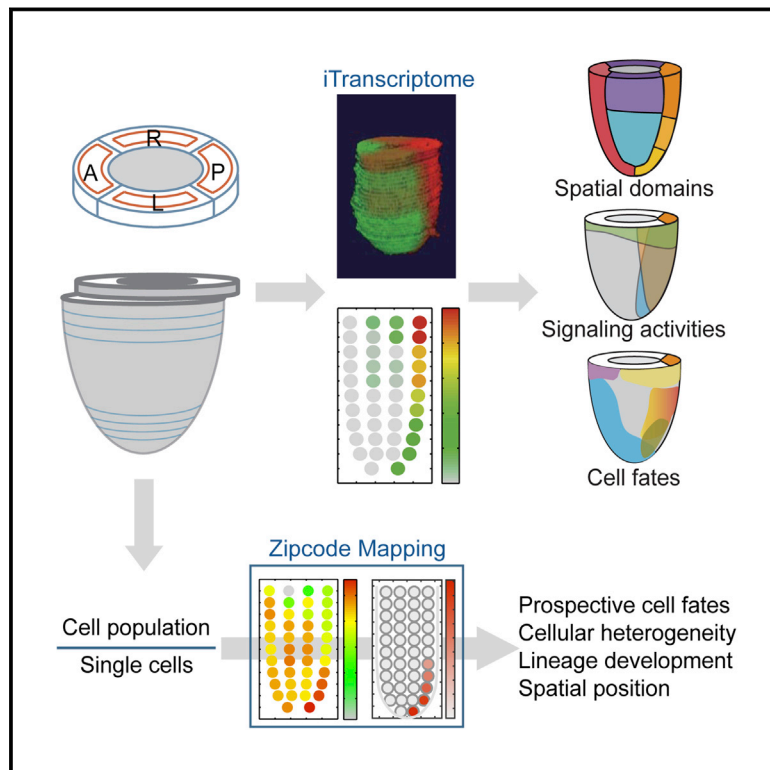


Developmental Cell

Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo

Graphical Abstract



Authors

Guangdun Peng, Shengbao Suo, Jun Chen, ..., Patrick P.L. Tam, Jing-Dong J. Han, Naihe Jing

Correspondence

jdhan@picb.ac.cn (J.-D.J.H.), njing@sibcb.ac.cn (N.J.)

In Brief

Peng et al. apply high-resolution RNA-seq to mid-gastrulation mouse embryos to collate a spatial transcriptome resource. 3D quantitative data rendition enables spatial gene expression pattern visualization in a web-based database and identifies zip code marker genes for mapping single epiblast cell position in the embryo by gene expression profile concordance.

Highlights

- Spatial transcriptome collated from RNA-seq data of the epiblast of mouse gastrula
- Rendition of expression pattern of 20,000 genes as digital whole-mount and corn plot
- Domains of transcription and signaling activity correlate with epiblast cell fates
- Mapping cell identity and home address in the epiblast by zip code marker genes

Accession Numbers

GSE65924



Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo

Guangdun Peng,^{1,4} Shengbao Suo,^{2,4} Jun Chen,^{1,4} Weiyang Chen,² Chang Liu,¹ Fang Yu,¹ Ran Wang,¹ Shirui Chen,¹ Na Sun,² Guizhong Cui,¹ Lu Song,¹ Patrick P.L. Tam,³ Jing-Dong J. Han,^{2,*} and Naihe Jing^{1,*}

¹State Key Laboratory of Cell Biology, Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

²Key Laboratory of Computational Biology, CAS Center for Excellence in Molecular Cell Science, Collaborative Innovation Center for Genetics and Developmental Biology, Chinese Academy of Sciences-Max Planck Partner Institute for Computational Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences, Shanghai 200031, China

³Embryology Unit, Children's Medical Research Institute and School of Medical Sciences, Sydney Medical School, The University of Sydney, Westmead, NSW 2145, Australia

⁴Co-first author

*Correspondence: jdhan@picb.ac.cn (J.-D.J.H.), njing@sibcb.ac.cn (N.J.)

<http://dx.doi.org/10.1016/j.devcel.2016.02.020>

SUMMARY

Gastrulation of the mouse embryo entails progressive restriction of lineage potency and the organization of the lineage progenitors into a body plan. Here we performed a high-resolution RNA sequencing analysis on single mid-gastrulation mouse embryos to collate a spatial transcriptome that correlated with the regionalization of cell fates in the embryo. 3D rendition of the quantitative data enabled the visualization of the spatial pattern of all expressing genes in the epiblast in a digital whole-mount in situ format. The dataset also identified genes that (1) are co-expressed in a specific cell population, (2) display similar global pattern of expression, (3) have lineage markers, (4) mark domains of transcriptional and signaling activity associated with cell fates, and (5) can be used as zip codes for mapping the position of single cells isolated from the mid-gastrula stage embryo and the embryo-derived stem cells to the equivalent epiblast cells for delineating their prospective cell fates.

INTRODUCTION

Gastrulation is a critical milestone of early embryogenesis in mammals when the primary germ layers are formed and the multipotent embryonic cells are allocated to the progenitors of tissue lineages within the germ layers. Morphogenesis of the germ layers during gastrulation entails a complex mechanism that regulates the proliferation, movement, and patterning of cell populations, and the choreography of switches in genetic and signaling activity that may drive lineage specification and tissue modeling in the embryo (Arnold and Robertson, 2009; Kojima et al., 2014b; Pfister et al., 2007; Solnica-Krezel and Sepich, 2012; Tam and Behringer, 1997; Tam and Loebel, 2007). Fate

mapping of the mouse embryo has shown that there is distinctive regionalization of cell fates in the germ layers before gastrulation is completed (Beddington, 1981, 1982; Tam and Behringer, 1997), and lineage tracing studies showed that concurrent with the formation of primary germ layer, progressive restriction of differentiation potency of embryonic cells takes place, culminating in the generation of lineage-restricted progenitors for specific tissue types (Beddington, 1981, 1982; Cajal et al., 2012; Lawson et al., 1991; Li et al., 2013). However, the molecular controls in time and space that underpin the exit of cells from the multipotent state, the specification of lineage-restricted progenitors and the regionalization of cell fates pertaining to the establishment of the body plan are not fully known.

A holistic knowledge of the activity of the genome specifically at gastrulation is essential for gleaning a better understanding of the molecular mechanism for lineage specification and embryonic patterning. Developmental changes in transcriptional activity during gastrulation have been studied by genome-wide microarray analysis of whole embryo or the epiblast (Kojima et al., 2014a; Mitiku and Baker, 2007). These studies have revealed that genes associated with regulation of pluripotency, germ layer formation, transcriptional regulation, cell metabolism, and transport and ion homeostasis are highly enriched during gastrulation, and the progression to organogenesis is accompanied by the activation of tissue-specific genes and morphogenetic drivers. Of special significance is the finding of a major overhaul of the transcriptome at the transition from mid- to late gastrulation (Kojima et al., 2014b), which reflects the downregulation of the pluripotency genes and concurrent acquisition of the ectoderm propensity and restriction of the mesendoderm fate. In the context of derivation of pluripotent stem cells, the dismantling of the pluripotency gene network marks the most advanced developmental stage at which epiblast stem cells could be isolated from the epiblast (Kojima et al., 2014a; Osorno et al., 2012). Taking into consideration the regionalization of cell fates in the germ layers, the mid-gastrulation stage is therefore the most advanced stage that cells at the transition between pluripotency and lineage-restricted state co-exist in the epiblast. However, neither whole embryo nor whole-epiblast transcriptome is

adequate to provide the molecular annotation of the body plan in the gastrulating embryo. Hence, it is imperative to undertake an analysis of the transcriptome at the resolution of discrete cell populations at anatomically defined sites in the epiblast.

In this study, we documented the spatial transcriptome of the epiblast of the mid-gastrulation (E7.0 late mid-streak stage) mouse embryo. We performed low-input RNA sequencing (RNA-seq) analysis on individual populations, of about 20 epiblast cells each, that were sampled by laser capture microdissection at known positional addresses in single embryos. This has enabled the collation of a genome-wide profile of the expressed genes in each location-defined cell population, which can be rendered into a spatial display of the transcriptional architecture for the whole epiblast. From the three mid-gastrulation embryos, the spatially defined and quantified expression data of over 20,000 genes are collated into a web-based database. This *in silico* spatial transcriptome (*iTranscriptome*) enables the visualization of the expression pattern of specific genes by 3D rendering and query for the cohort of genes displaying specific syn-expression patterns. In addition, the *iTranscriptome* of the mid-gastrulation embryo has a unique attribute of zip code mapping that enables pinpointing the position, at sufficient precision, of single cells isolated from the epiblast and identifying the equivalent cell population of embryo-derived stem cells such as the epiblast stem cells.

RESULTS

Construction of the Spatial Transcriptome

For profiling the gene expression pattern of cell populations in different regions of the epiblast, samples were harvested by laser capture microdissection (LCM) from non-overlapping sites in late mid-streak stage (E7.0) mouse embryo (Downs and Davies, 1993). To minimize the noise in the transcriptome created by between-embryo variance, the analysis was performed on single embryos in triplicate (E1, E2, and E3). Each embryo was cryo-sectioned serially, and alternate sections in the distal-proximal series were used for sample collection (S) and reference templates (R) for digital 3D reconstruction, respectively (Figures

1A, S1A, and S1B). A population of approximately 20 cells were captured from four quadrants of each sample section: anterior (A), posterior (P, which contains the primitive streak) and left lateral (L) and right lateral (R) (Figures 1A and S1C), except for the most distal section (S1) where only cells in the anterior and posterior positions were harvested. Altogether, 42 anatomical regions from 11 sections per embryo were sampled. Given that the late mid-streak epiblast contains about 2,000 cells (Power and Tam, 1993; Snow, 1977; Tam and Beddington, 1992), the capture of 42 × ~20 cells represented over 80% of the 1,000 cells in the sample sections (i.e., 40% of total epiblast population). To enhance the efficiency of RNA-seq with low RNA input, a single-cell RNA-seq technique was adapted and optimized for analyzing the small-cell-number LCM samples.

RNA-seq was performed at a depth of over 20 million reads per library to achieve sufficient sequencing depth saturation (Figure S1D). After removing samples with low sequencing quality, data analyses were carried out on the three embryos. For embryo E1, which provided the reference transcriptome for data mining, the dataset of a missing sample was re-created computationally (Figure S1E). Mapping of the raw reads onto mouse mm10 genome revealed that transcripts of between 13,500 and 14,500 unique genes (fragment per kilobase per million (FPKM) >1.0 in at least two samples for each embryo) were detected in individual embryos (altogether about 20,000 unique genes from the three embryos), and the samples in each quadrant have similar expression density distribution (Figure S1F).

3D Visualization of the Spatial Transcriptome as Digital WISH

Whole-mount *in situ* hybridization (WISH) has been the most practiced means to study the spatial pattern of gene expression in gastrulation-stage mouse embryos, but it is inherently a low-throughput gene-by-gene approach. Our RNA-seq data that integrated gene expression level (by transcript counts) with the precise position of the cell population made it feasible to reconstruct a quasi-quantitative 3D expression map for every transcript in a single epiblast in WISH format. An algorithm was

Figure 1. Spatial RNA-Seq Analysis of the Mid-gastrulation Embryo

(A) Experimental strategy: cells were captured by laser capture microdissection from four quadrants: anterior (A), posterior (P), and lateral (left/right, L/R), of each sample section (S1–S11) of the epiblast of late mid-streak stage (E7.0) C57BL/6 embryos, and analyzed by RNA-seq. The reference sections (R1–R11) were used as templates for 3D reconstruction of the embryo for data visualization.

(B) The presentation of the spatial pattern of gene expression in the corn plot. Each dot in the plot represents the cell sample at the specific positional address, and the color indicates the level of gene expression computed from the transcript counts in the RNA-seq dataset.

(C) Whole-mount *in situ* hybridization (WISH) result of *T* gene expression (upper paired panels): The *in situ* image was digitally scanned in 11 slices and expression values for anterior (A), left (L), and posterior (P) regions for each slice were quantified by the average color intensity and plotted in the corn plot format (values were scaled to [0, 1], see Supplemental Experimental Procedures). As scanning was performed from the left side of the embryo, only the expression values of the left quadrants were shown. Digital (d)-WISH (lower paired panels): display of the RNA-seq data for the *T* gene on the reconstructed template of the embryo and the corn plot (FPKM values were scaled to [0, 1]).

(D) Comparison of WISH (images, upper panels) and d-WISH (corn plot data, lower panels) for *Pou3f1*, *Sox2*, *Noto*, *Six3* (our data), *Otx2*, *Lefty2*, *Eomes*, *Wnt3a*, and *Tdgf1* (from EMAGE in eMouseAtlas). d-WISH reveals *Noto* expression in the distal posterior epiblast that is masked by the overlying endoderm when examined by WISH. *Six3* expression in the anterior proximal region was detected by d-WISH but not detected by WISH. Inset shows the reads coverage of *Six3* in sample 11A (data displayed by the genome browsing application IGV).

(E) The linear correlation analysis of d-WISH (RNA-seq) and *in situ* hybridization (image scan) data of genes shown in (C) and (D). Expression values are scaled and the Pearson correlation coefficient values are indicated for each plot.

(F) The functionalities of the *iTranscriptome* web portal: (1) search of expression pattern by gene, (2) search of co-expressing genes, (3) search of genes by expression pattern, and (4) zip code mapping to equivalent cell population or positional address by transcriptome comparison. The output examples were the *T* gene, top ten genes co-expressed with *T*, genes displaying a predefined expression pattern, and the mapping of cells from an ESC-derived EpiSC line (ESD-EpiSC).

See also Figure S1.

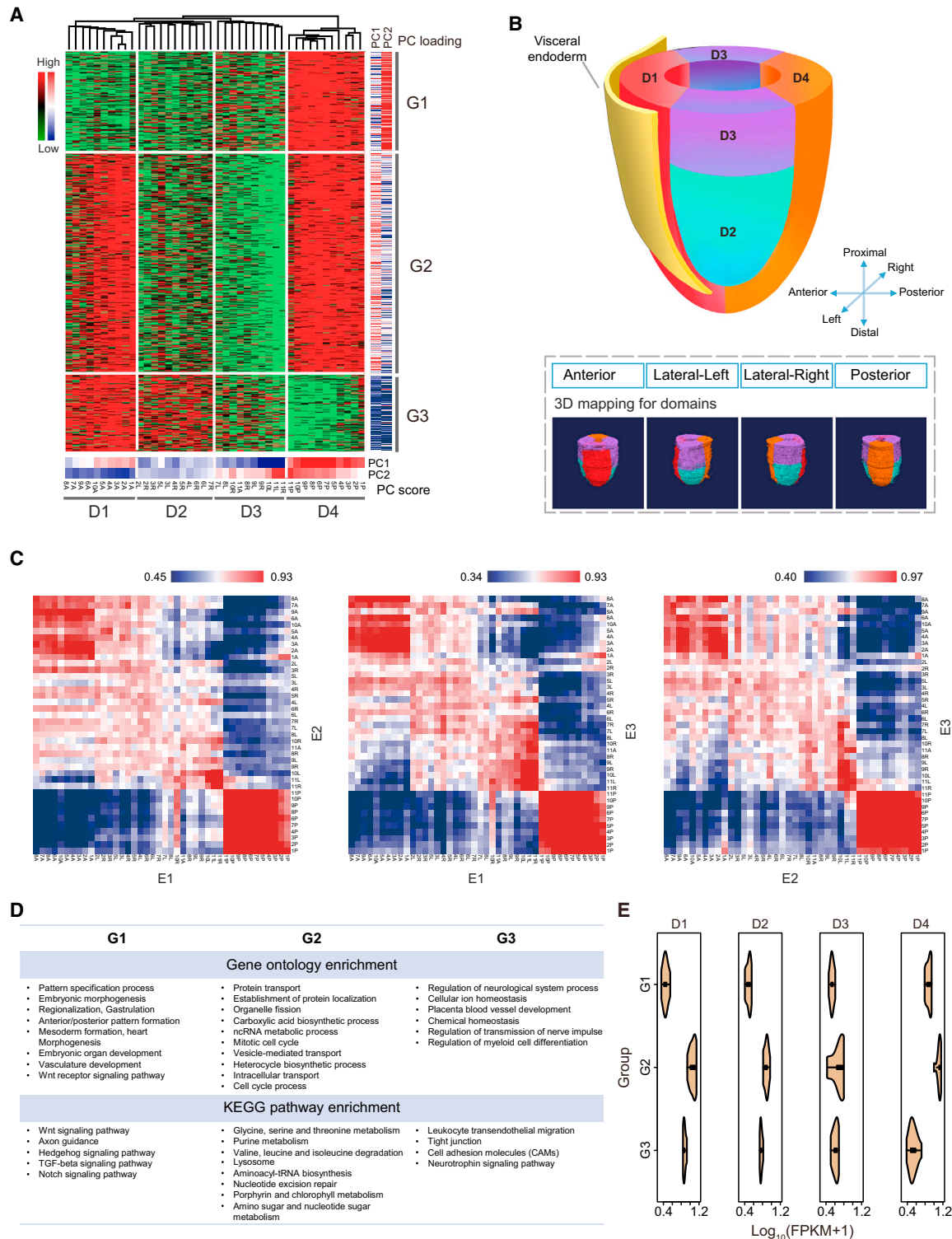


Figure 2. Spatial Transcriptome Reveals Transcription Domains of the Mid-gastrulation Embryo

(A) Clustering of differentially expressed genes (DEGs) of the expression domains in the reference embryo (embryo E1). Hierarchical clustering and BIC-SKmeans separated the laser capture microdissection samples and DEGs into four expression domains (D1–D4) and three gene groups (G1–G3), respectively. A re-sampling method was applied to identify gene groups significantly associated with the two major PCs, and the FDRs (one-sided Mann-Whitney test followed by BH correction) for G1, G2, and G3 are $4.1\text{e}-259$, 0 , $3.8\text{e}-196$, respectively. Right column: contribution of each gene (rows) to the first two PCs (PC loading). Bottom row: projection scores of the cell samples on the first two principal components (PC score).

(legend continued on next page)

devised for displaying the RNA-seq data by mapping to an anatomical template of the embryo. Virtually, the cylindrical embryo was transformed into a planar (2D) configuration. The position of each cell sample was represented by the grid node in a corn plot (so named for its resemblance to a corn cob), which is defined by the coordinate position in the cross-sectional plane (A, P, R, L) and the distal-proximal axis (1–11). Samples in the lateral quadrants (R, L) are displayed as parallel sets of nodes in the middle of the plot and the gene expression level is presented in a color-coded format (Figure 1B).

To validate that the corn plot is a faithful emulation of the spatial pattern of gene expression, we compared the digital display with conventional WISH, using the *Brachury* (*T*) gene as a worked example. For the WISH, the stained embryo was scanned in 11 slices and the staining intensity was digitally rendered as color gradient data (signal values) displayed in a corn plot (Figure 1C). A high degree of concordance was found between the ISH pattern in whole mount and corn plot (*T* in situ) with the digital gene expression data (digital (d)-WISH) displayed in the 3D reconstruction and the corn plot (*T* RNA-seq), respectively (Figure 1C). The veracity of the d-WISH (our data) against conventional WISH (our data and from EMAGE: <http://www.emouseatlas.org/emage/>) was further validated for examples of genes that were expressed in the anterior epiblast (*Pou3f1*, *Otx2*, and *Sox2*), and the posterior epiblast (*Lefty2*, *Eomes*, *TdGF1*, and *Wnt3*) (Figure 1D). The RNA-seq FPKM values were found to correlate significantly with the WISH signal values (Figure 1E).

The d-WISH analysis has a potentially higher sensitivity than WISH for detecting genes that are expressed weakly or in a confined region (e.g., *Six3* and *Noto*; Figure 1D). *Six3* expression in the anterior proximal epiblast was not detected by WISH but d-WISH has revealed reads of *Six3* transcripts by RNA-seq in region 11A of the epiblast (Figure 1D and inset). The spatial transcriptome also revealed the epiblast-specific component of the expression pattern: WISH of *Noto* did not distinguish between the expression in distal visceral endoderm and distal epiblast, while d-WISH revealed a discrete expression domain of *Noto* in the 1P–4P epiblast (Figure 1D). In these two cases, the difference in the sensitivity between d-WISH and WISH underpinned the poor correlation of FPKM and staining intensity value. In essence, d-WISH offers unprecedented resolution and data content: the expression results can be viewed in a spatial context, the data give a finer measurement of gene activity, and this genome-wide spatial transcriptome have documented both spatial and quantitative expression data of transcripts of over 20,000 unique genes in the epiblast.

To avail of the spatial transcriptome data as a resource to the scientific community, a web portal, *iTranscriptome* (for in silico transcriptome) at <http://www.itranscriptome.org>, has been es-

tablished to provide open access of the spatial transcriptome data (Figure 1F). The *iTranscriptome* offers data search functionalities in (1) pattern search by gene: querying and displaying the expression pattern of genes of interest in either a corn plot or a digitally reconstructed d-WISH format; (2) gene search by gene: searching for genes that share a similar expression pattern with the queried gene. For each queried gene, the guilt-by-association (GBA) method (Walker et al., 1999) is used to identify the co-expressed genes, and then returns the genes with a p value smaller than the input cutoff; and (3) search for a set of genes displaying a predefined (20 in the database) or customized (user-defined) expression pattern. Besides these functionalities, the *iTranscriptome* also enables matching queried cells to their in vivo counterparts in the epiblast (by zip code mapping; see details later).

Spatial Domains of Gene Expression

To identify the transcriptome domain in the mid-gastrulation embryo, genes that were expressed (FPKM > 1.0) in at least two samples in each embryo and with a variance in transcript level > 0.05 (transcript levels are measured as $\log_{10}(\text{FPKM}+1)$) across all samples were retrieved for further analysis. Unsupervised hierarchical clustering of cell samples using differentially expressed genes (DEGs; Figure S2A and Table S1) identified four spatial expression domains (D1, D2, D3, and D4) that display distinct gene expression patterns (Figure 2A). Principal component analysis (PCA) showed that PC1 and PC2 were sufficient for separating the samples into the same four domains (Figure 2A, bottom, heatmap for PC1 and PC2 scores).

Among the spatial expression domains, D1 encompassed all the anterior samples, whereas the posterior region samples were clustered into D4 aligning orderly by their distal-proximal positions (Figure 2A). D2 consisted of samples from the distal lateral region of the epiblast and D3 contains those from the proximal lateral region. The anterior (D1) and distal-lateral regions (D2) showed closer relations than with other domains, suggesting a developmental coherence of these two domains. This classification indicated that the epiblast at this stage can be partitioned into four distinctive domains of gene activity, each displaying remarkable spatial coherence of samples (Figure 2B).

The concordance (measured by Spearman rank correlation coefficients [RCC]) of the DEGs across all samples between the embryos was high (average RCC \pm SD, 0.74 ± 0.13 , Figure 2C). The significantly higher intra-domain correlation (average RCC \pm SD, 0.75 ± 0.04) compared with inter-domain correlations (average RCC \pm SD, 0.66 ± 0.04 , two-sided t test, $p = 3.9 \times 10^{-3}$) indicates that the spatial transcriptome is reproducible and the delineation of spatial expression domains is consistent among the mid-gastrulation embryos.

(B) Four expression domains in the embryo (upper panel) and the 3D presentation of the domains, visualized by DEG clustering in the four quadrants of the epiblast (lower panel).

(C) Heatmaps of Spearman rank correlation coefficient (RCC) analysis showing pairwise comparison of inter-domain DEG expression profiles across all samples of the three embryos (E1, E2, and E3). The RCC of each sample between any two embryos across all inter-domain DEGs was calculated. The samples in each embryo were sorted according to the results of hierarchical clustering in the reference embryo (E1).

(D) The GO and KEGG enrichment analysis for the three DEG gene expression groups (G1–G3, $p < 0.01$).

(E) Violin plot showing the mean expression values of each of three gene groups across the four sample domains. Each violin plot shows the frequency distribution of the mean transcript level (\log_{10} transformed FPKM) of DEGs per sample.

See also Figure S2 and Table S1.

Distinctive Functional Gene Ontology in the Spatial Expression Domains

Next, BIC-SKmeans (Zhang et al., 2013) was used to search for the minimal numbers of gene clusters based on DEGs that can account for the variations of the transcriptome across the 42 samples (Figure S2B). Three distinct gene groups (referred to as G1, G2, and G3; Figure 2A) were identified based on similarity in their expression patterns across different samples. G1 was highly expressed in the posterior epiblast. G2 was highly expressed in both the anterior and posterior epiblast. G3 was significantly upregulated in the anterior epiblast and also highly expressed in the distal lateral regions, further suggesting a developmental relationship between the cells in the anterior and distal lateral regions. Mapping the median expression level of each gene group in different samples onto the 3D embryo template (Figure S2C) revealed that the transcriptional relationships between samples recapitulate the spatial topography of the embryo. In addition, by a re-sampling method (Chung and Storey, 2015), we found these three groups significantly associated with the two major PCs ($p < 1.0e-195$; Figure 2A, right column).

We next examined the Gene Ontology (GO) terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways enriched in each group (Huang et al., 2009) (Figure 2D). G1, which consisted of genes highly expressed in the posterior region of the embryo, showed a high enrichment for embryonic morphogenesis and pattern specification under the regulation of WNT signaling, corroborating known functions for posterior development. On the contrary, we found relatively fewer function terms enriched for the G3, except for “regulation of neurological system process” and “tight junction” functions, consistent with epithelial morphogenesis and the neuroectoderm differentiation of cells in this domain. G2 genes, which were highly expressed in both anterior and posterior epiblasts, were significantly enriched for cell-cycle and metabolism functions. The average expression level of G2 genes was the highest in all domains, compared with other gene groups (Figure 2E), indicating that the epiblast is at a phase of rapid cell proliferation (Lawson et al., 1991; Snow, 1977). Interestingly, we also observed a high enrichment for epigenetic modifications in G2 expressed anterior and posterior domains (Figure S2D). Epithelial-mesenchymal transition (EMT) is an important morphogenetic process during gastrulation (Thiery et al., 2009). The EMT-regulated genes were strongly enriched in the posterior domain (D4) (Figure S2E), consistent with the heightened EMT activity in the primitive streak. These EMT genes were expressed at a lower level in the anterior domain (D1) (Figure S2E), where cells maintain an epithelial architecture. Whereas there was significant difference in the transcriptome between the anterior (D1) and posterior (D4) domains, analysis of the lateral domains showed that between 26 and 34 genes in each sample embryo displayed contralaterally different expression patterns (Figure S2F). To validate the DEG data on left-right asymmetrical expression of the genes identified in spatial transcriptome of embryos E1–E3, WISH was performed on six differentially expressed genes, each on more than three embryos (Figure S2G). Overall, there was no consistent left-right asymmetry in expression (Figure S2H). It is therefore likely that the left-right asymmetry has yet to be consistently established in the late mid-streak stage mouse embryo (Raya and Izpisua Belmonte, 2006).

The spatial transcriptome is a unique resource for identifying region-specific marker genes for epiblast cells of the mid-gastrulation embryo. Since PC1 and PC2 have accounted for the majority (~60%) of variances, and selection of genes by PCA loading could preserve more effective vectors on each variable (Wang et al., 2015), the top 40 genes with highest or lowest PCA loadings in the first two PCs (Table S2, 93.7% of them overlapped with DEGs, Fisher’s exact test, $p = 8.4e-58$) were re-analyzed by unsupervised k-means ($k=4$) clustering for all 42 samples. From this analysis, four module domains (MD1–MD4) and four module gene families (MG1–MG4) were identified (Figures 3A and S3A). The module domains MD1–MD4 were highly consistent with the four domains identified by DEGs (D1–D4). MG1, MG2, and MG3 showed similar expression patterns as G1, G2, and G3, respectively (Figure 2A). MG4 that was highly expressed in MD3 characterized the proximal lateral domain. The gene expression density distribution of each gene module in each domain (Figure S3B) was similar to that of DEGs (Figure 2E). Like G2, the MG2 that was common for anterior and posterior domains displayed the highest expression level. Therefore, we identified a smaller set of “region-specific” genes (Table S2), which are sufficient to characterize each expression domain (MD1–MD4) in the epiblast.

PCA analysis of these PCA loading genes (Table S2) showed that they constituted four sample groups (Figure 3B) and four distinct gene groups (Figure 3C). On the basis of co-expression with known markers (anterior epiblast, *Sox2*; posterior epiblast, *Mixl1*, *T*, and *Mesp1*), two gene groups were identified on the PC1 axis for the anterior and posterior regions of the epiblast (Figure 3C), and contained within each group were domain-specific signature genes. Among them, *Pou3f1*, a POU III subfamily transcription factor, which belongs to the anterior gene groups, was recently found to be associated with the neuroectoderm lineage (Zhu et al., 2014). Several newly identified region-specific genes were validated by WISH and the results agreed well with the domain classification: *Cbx7*, *Cldn7*, *Uchl1*, *Sall2*, *Utf1*, *Slc7a3*, and *Zfp462* were identified to be exclusively expressed in the anterior region and *Ccnd2*, *Sall3*, and *Sp5* were mainly expressed in the posterior region (Figures 3D and S3C). Therefore the spatial transcriptome enables identification of spatially restricted transcripts that could be putative lineage markers.

On the PC2 axis, two gene groups were differentially expressed in anterior/posterior (D1/D4) and proximal lateral domains (D3) (Figure 3C). Gene set enrichment analysis (GSEA) analysis based on KEGG pathways (false discovery rate [FDR] < 0.25) showed that the PC1 axis was enriched for the development-related pathways, such as the WNT and TGF- β signaling pathways. The PC2 axis was enriched for the cell-growth and metabolism-related pathway terms, such as spliceosome, purine metabolism, and cell cycle (Figure S3D), representing different functional attributes of the epiblast cells. For the four region-specific marker gene set, we further verified the expression pattern by microfluidic multiplex qPCR using the E1 LCM samples. Similar four spatial domains were identified by PCA analysis of the marker gene expression profiles (Figure S3E).

Taken together, unsupervised hierarchical clustering and PCA analysis have defined four spatial expression domains in the

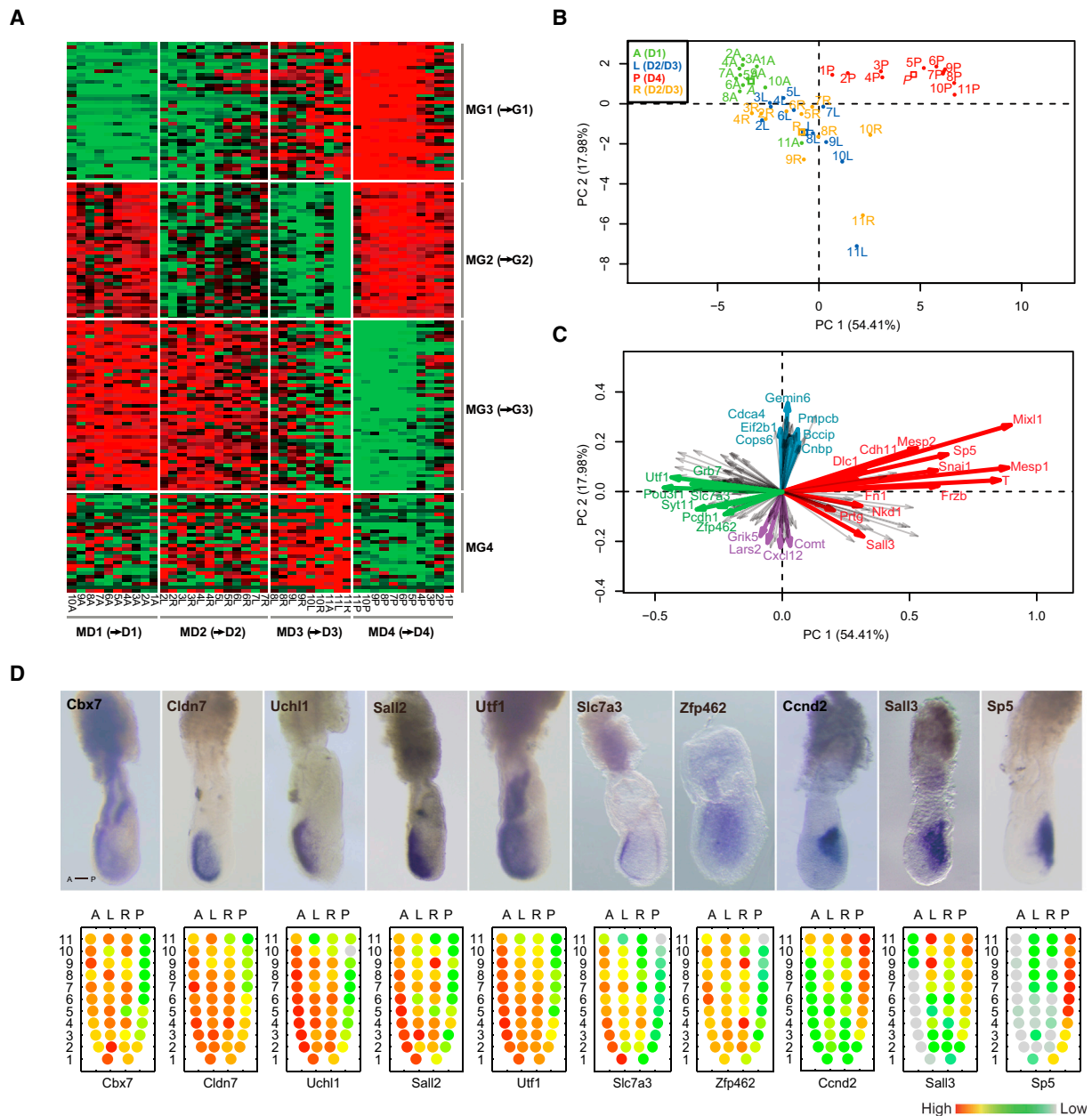


Figure 3. Spatial Transcriptome Uncovers Gene Markers for Different Expression Domains

(A) The heatmap of clustering analysis of marker genes. The samples are clustered by K-means ($k = 4$), and the marker genes are clustered by BIC-SKmeans. (B) Principal component analysis (PCA) of all cell samples and their projections on PC1 and PC2. PC1 discriminates between the anterior and the posterior samples, and PC2 between the proximal lateral samples and the others.

(C) PC projections showing the contribution of each gene to the first two PCs. The arrow tips indicate the correlation coefficient of the respective genes with each principal component. *Mixl1*, *Mesp1*, *T* represent the posterior enrichment. *Pou3f1* and *Utf1* indicate an anterior enrichment. *Cxcl12* and *Lars2* denote the correlation with the proximal lateral population. Only genes with cosine correlation >0.8 or >0.55 for the purple genes are shown.

(D) WISH validation (upper panels) of newly identified region-specific genes against the d-WISH data presented in corn plots (lower panels). Anterior, *Cbx7*, *Cldn7*, *Uchl1*, *Sall2*, *Utf1*, *Slc7a3*, and *Zfp462*; posterior, *Ccnd2*, *Sall3*, and *Sp5*.

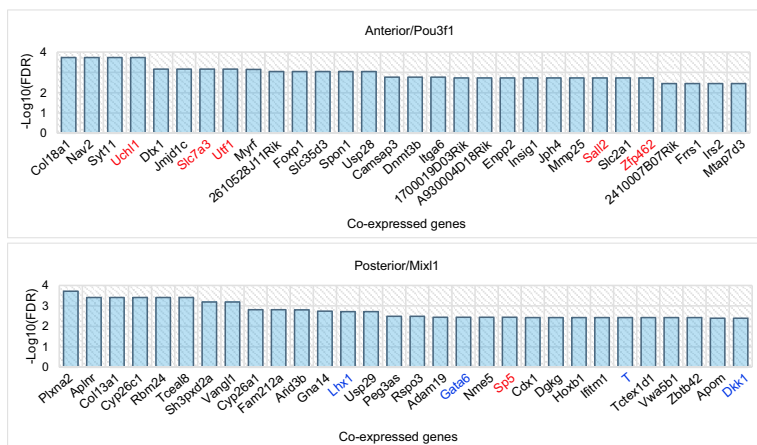
See also Figure S3 and Table S2.

epiblast of the mid-gastrulation embryo, which may be correlated with the regional variations of the molecular activity associated with a full spectrum of functional attributes such as cell proliferation, lineage differentiation, epithelial-mesenchyme transition and epigenetic modification.

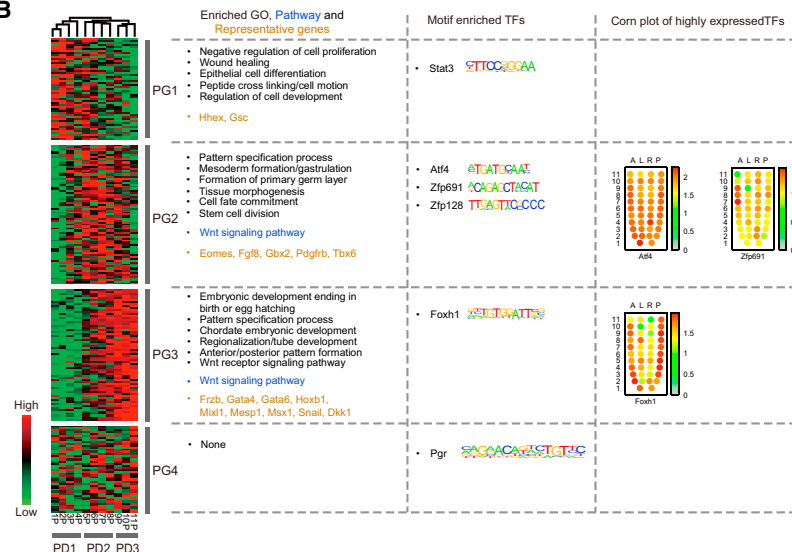
Subdomains of Transcriptional Activity in the Anterior and Posterior Epiblast

Next, we examined if the spatial transcriptome may reveal the molecular activity that is associated with the regionalization of cell fates in the anterior and posterior epiblast. To do this,

A



B



C

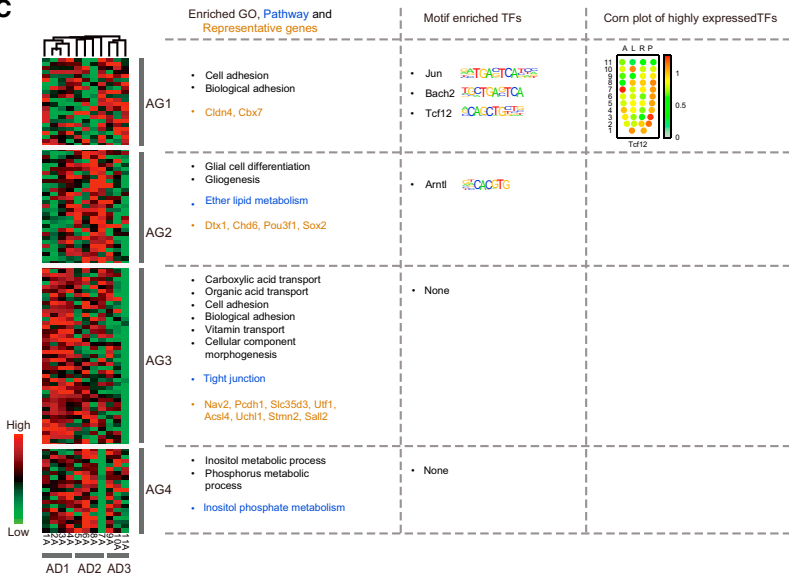


Figure 4. Subdomains of the Anterior and Posterior Epiblast

(A) Top 30 co-expressed genes in the anterior (anchored by *Pou3f1*) and posterior (by *Mixl1*) regions identified by guilt-by-association analysis (GBA). These co-expressed genes are sorted by the BH multiple testing corrected p values. Known domain-specific markers are highlighted in blue and WISH-validated newly identified markers in red.

(B and C) Subdomains of the posterior transcription domain (B) and the anterior transcription domain (C) delineated by clustering analysis using the combined set of genes co-expressed with the seed marker and genes of highest PCA loading. The subdomains of the samples are classified by hierarchical clustering and the gene groups are identified by BIC-SKmeans. The enriched GO/KEGG terms (one-sided Fisher's exact test, $p < 0.01$), representative subdomain genes, motif-enriched transcription factors (TFs) (known, $p < 0.01$; de novo, $p < 1.0e-12$) and corn plots of highly expressed TFs are shown for each subgroup.

See also Figure S4 and Tables S3 and S4.

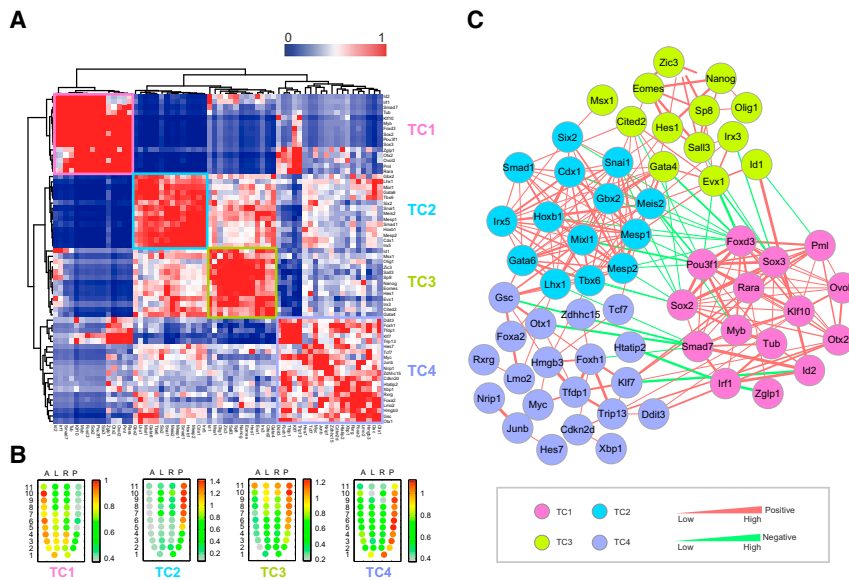


Figure 5. Transcription Factors Regulatory Network

(A) Hierarchical clustering by the connection specific index (CSI, CSI > 0.7) reveals four development-related transcriptional gene cliques (TC1–4).

(B) The corn plot shows the average expression values of TFs in each transcription factor clique (TC1–4).

(C) The co-expression network of development-related TFs based on the CSI values (CSI > 0.7). The node with different colors represents the TF cluster and the edges are the CSI value of two nodes. Red lines indicate the positive correlation and green lines indicate negative correlation. Edge weights are proportional to the CSI values of two correlated nodes.

See also Figures S5, S6, and Table S5.

Pou3f1 in the anterior region, and *Mixl1* in the posterior region, which showed the highest square cosine correlation to both PC1 and PC2 simultaneously, were selected as the seed marker gene (see [Experimental Procedures](#)). We expanded the gene set by including genes that were co-expressed with the seed marker gene through GBA analysis ([Walker et al., 1999](#)) (Figure 4A and Table S3). Functional enrichment analysis showed that the genes co-expressing with *Pou3f1* and *Mixl1* were related to neuron development and pattern specification, respectively, and the average expression pattern of the co-expressed genes was also highly consistent with the seed genes (Figures S4A–S4D).

Unsupervised hierarchical clustering of posterior samples based on the combined posterior gene set partitioned the posterior region into three subdomains: 1P to 4P as distal posterior subdomain (PD1), 5P to 8P as the intermediate posterior domain (PD2), and 9P to 11P as the proximal posterior domain (PD3) (Figure 4B). BIC-SKmeans analysis clustered the combined posterior marker gene set into four subgroups: posterior gene PG1–3, corresponding to distal, intermediate and proximal expressed gene groups respectively, while PG4 showed no proximal-distal bias (Figures 4B and S4E). Some known genes of mesodermal progenitors were represented in this marker set, such as *Hhex*, a PG1 gene, which is related to cardiac mesoderm formation ([Liu et al., 2014](#)), was specifically expressed in the distal subdomain (PD1); *Tbx6*, a PG2 gene, related to the specification of paraxial mesoderm ([Chapman et al., 1996](#)), was exclusively expressed in the intermediate domain (PD2), and *Gata4*, a PG3 gene associated with lateral mesoderm differentiation ([Rojas et al., 2005](#)) was strongly expressed in the proximal posterior domain (PD3). The expression pattern of these PG genes aligned well with the regionalization of mesodermal fates of cells in the primitive streak ([Tam and Loebel, 2007](#)). The different PG groups also displayed enriched biological functions, pathways, and the upstream transcriptional factors (TFs) (Figure 4B), which are associated with mesendoderm specification under the influence of WNT signaling. Consistent with this inference, enforced expression of *Foxh1*, a core upstream transcription factors for the PG3 proximal subdomain gene groups,

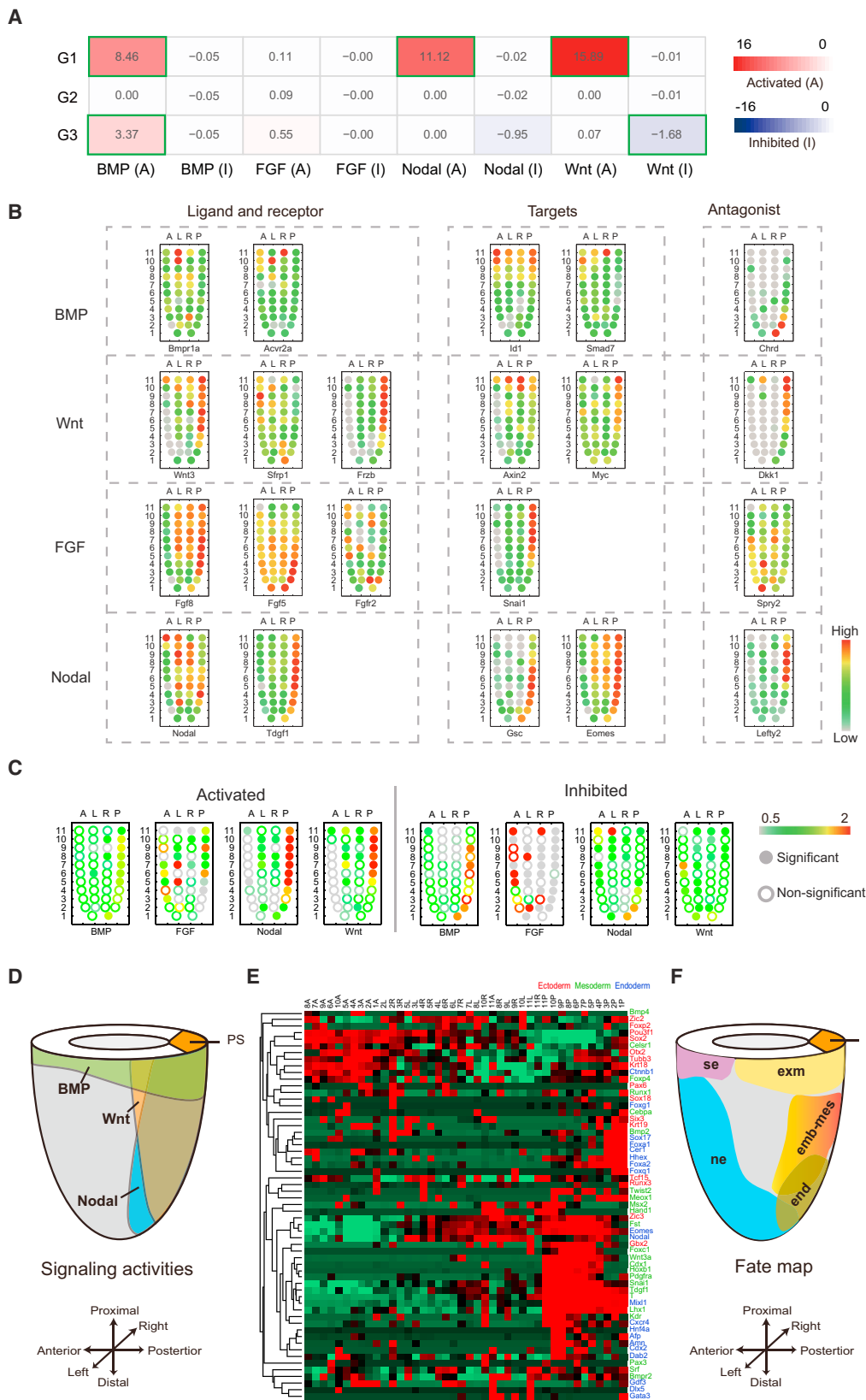
in embryonic stem cells (ESCs) led to the upregulation of PG3 genes, such as *Dkk1*, *Mesp1*, in a WNT-signaling-dependent manner (Figure S4F). This example corroborated with the previous finding that *Foxh1* functions as a potential regulator for mesendoderm progenitors ([Hoodless et al., 2001](#); [Yamamoto et al., 2001](#)).

To identify gene signatures that correlated with the three distinct subdomains in the posterior region, a domain specificity score was defined and used (see [Supplemental Experimental Procedures](#)). The expression patterns of the top five genes for PD1–3 were found to be spatially matched to each subdomain (Figure S4G), further confirming that the data allow detection of a wide range of spatial expression patterns at the genome-wide level. The full dataset (Table S4) provides a useful resource for further research in cell-fate determinants.

The findings of the aforementioned analysis have led to the identification of subregions of the apparently homogeneous transcription domain in the posterior epiblast, thus we sought to find out if there are also subregions in the anterior epiblast. Unsupervised clustering has identified three proximal-distal (AD1–3) subdomains (Figures 4C and S4E). However, the difference in expression profiles was not as clear-cut as that of the posterior domain, and some genes showed uniform expression in the anterior samples (e.g., AG1 and AG3). The biological functions, pathways, and TFs for each of these subdomains were also less well defined. These results suggest that the lineage fates of cells in the anterior epiblast are not as well segregated as those in the posterior epiblast at mid-gastrulation.

Regionalization of Transcription Factor Regulatory Network Activity

To characterize the transcriptional regulation network in the epiblast, the Connection Specificity Index (CSI) ([Bass et al., 2013](#)) of all the development-related TFs in DEGs (TFs with GO terms related to development; Table S5) was calculated and subjected to unsupervised clustering analysis. The TFs formed four distinct module TF cliques (TC1–TC4, Figure 5A). Interestingly, these four modules were roughly divided into two mutually exclusive expression categories based on the average expression level of genes in each TF clique. TC2, TC3, and TC4 showed



(legend on next page)

a posterior-specific expression pattern, whereas TC1 genes were exclusively expressed in the anterior region (Figure 5B). This mutually exclusive expression pattern suggests that these two categories of TFs may have divergent function in the epiblast.

In order to reveal the underlying connections between these TF groups, we further constructed a TF co-expression network on the basis of their expression correlation CSI (CSI > 0.7). Consistent with the clustering result, the development-related TFs formed four groups in the network recapitulating TC1–4 (Figure 5C). TFs within each group showed positive interactions with each other, indicating that TFs of each group may form a combinatorial regulatory circuitry. Two TF groups that showed strong negative correlation in expression represented the anterior and posterior domains, respectively (TC1 and TC2). Since negative correlation often occurs between genes regulating switch-like behavior or alternative cellular states (Xue et al., 2007), the negative correlation between TC1 and TC2 may suggest that these TFs contribute to the divergence of cell identity in the anterior and posterior epiblast. Indeed, many hub TFs (e.g. *Mixl1*, *Snai1*, *Gsc*, *Sox2*) in the network (Figure 5C) are associated with the determination of anterior versus posterior fate (Arnold and Robertson, 2009; Robb and Tam, 2004). Many of the TFs in combinations (e.g. *Pou3f1*, *Sox2* and *Sox3*, *Snai1* and *Smad1*) also shared development-related targets (Figures S5A–S5D).

Similar analysis on all differentially expressed TF showed the presence of similar anterior and posterior TF cliques (AITC) (Figures S6A–S6C). The GO enrichment analysis showed that a clique enriched for terms associated with metabolism was less represented in the development-related TF groups (AITC2; Figures S6D and S6E), suggesting this clique of TFs and its enriched function “metabolism and cell cycle” may have insignificant impact on the divergence of cell fates of the anterior and posterior epiblast.

Regionalization of Signaling Pathway Activity

In order to chart the activity of signaling pathways in the epiblast, we examined the enrichment of signaling target genes of the WNT, BMP, FGF, and Nodal pathways (using microarray data from published perturbation experiments). The results of the enrichment analysis revealed that target genes of active WNT, BMP, and Nodal signaling pathways were significantly enriched

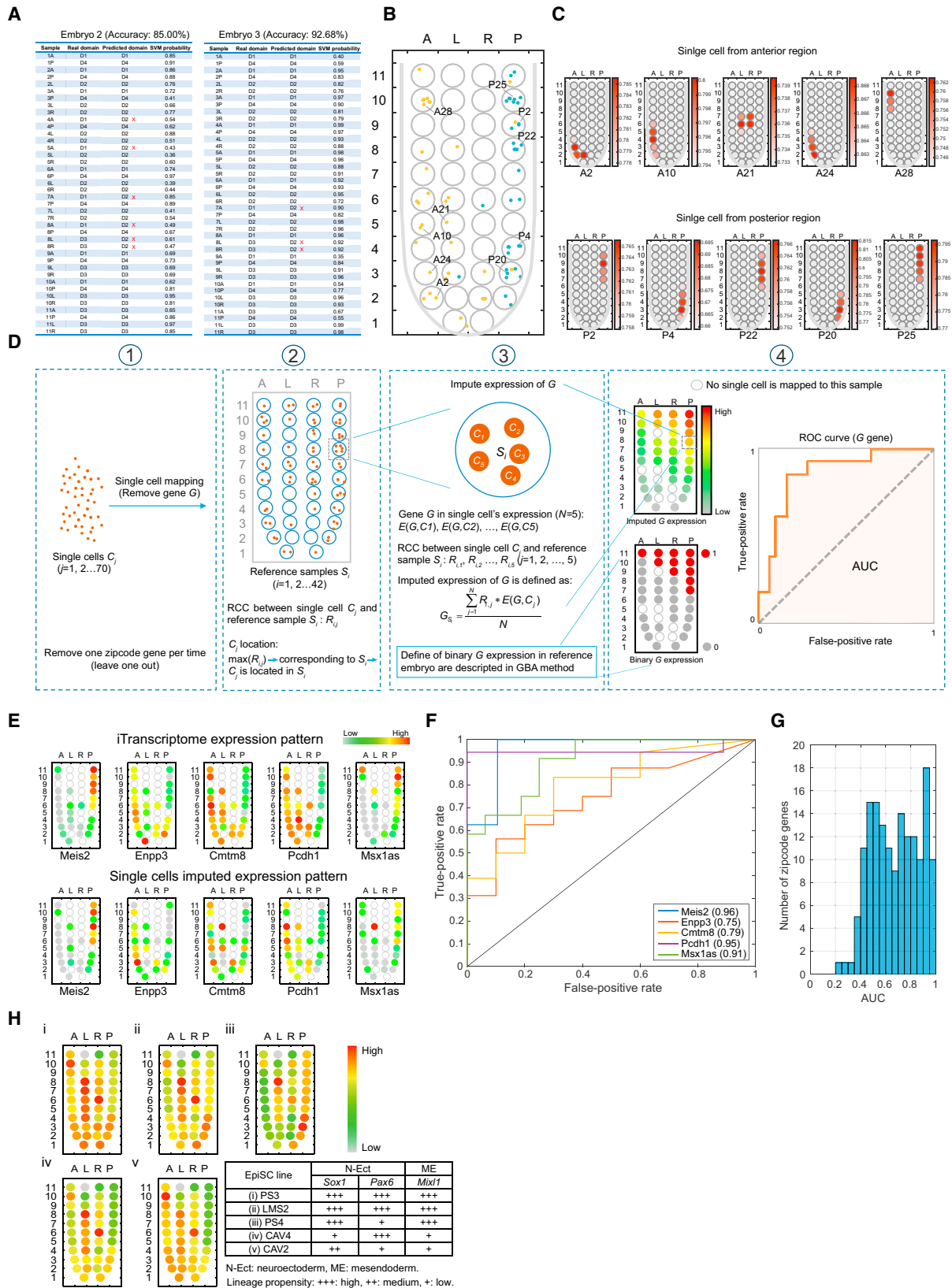
in the posterior gene group (G1) (Figure 6A, one-sided Fisher's exact test with FDR < 0.01), whereas FGF signaling targets showed no obvious spatial enrichment. Key signaling pathway factors, including ligands, targets, and antagonists all showed enhanced expression in the posterior domain (Figure 6B). In contrast, the anterior epiblast was largely devoid of FGF/Wnt/Nodal activity, while expression of BMP responsive genes was detected in the proximal anterior and lateral epiblast (Figure 6B). GSEA analysis revealed that cells in the posterior epiblast displayed activated Nodal/WNT activity and inhibitory BMP activity, whereas the anterior epiblast displayed both inhibitory and activated FGF activity (Figure 6C, FDR < 0.25). Of interest is that the regionalization of signaling activity in the epiblast revealed by the spatial transcriptome (Figure 6D) is related to the prospective lineage fate of the epiblast cells signified by the region-specific expression of germ layer markers (Figure 6E) and results of previous fate mapping studies (Figure 6F). For example, the BMP active domain is populated by progenitors of surface ectoderm and extraembryonic mesoderm, the Nodal/WNT active domain contains the mesendoderm progenitors, and the anterior epiblast with least active signaling activity is fated for neuroectoderm. The correlation of the gene expression pattern with the prospective lineage fates of cells at a specific region of the epiblast provides the basis for mapping exogenous cells to their equivalent cell types in the mid-gastrulation epiblast by zip code mapping.

Zip Code Mapping of Single Cells and Embryo-Derived Stem Cells

Analysis of the spatial transcriptome of the transcription domains has defined a small set of domain-specific signature genes (Figures 3A–3C, Table S2) that would enable mapping the position of cells in the epiblast on the basis of the concordance of the gene expression profile. To examine whether this gene set in the D1–D4 domains was consistent across different embryos, a machine learning method was used to analyze the cell samples of embryos E2 and E3 against the reference embryo E1. Using the expression values of these genes as the input to train a support vector machine (SVM), and then applying the trained model, the domain label for each sample in E2 and E3 can be predicted at 85.0% and 92.7% accuracy, respectively (Figure 7A), suggesting that this gene set could be used to map the position of individual epiblast cells sourced independently from other

Figure 6. Signaling Regulatory Network

- (A) The enrichment for target genes of development-related signaling pathways in the inter-domain DEG gene groups (G1–G3, tested by one-sided Fisher's exact test followed by BH correction). Red and blue indicate activated enrichment and inhibitory enrichment, respectively. The value in each cell is the $-\log_{10}(\text{FDR})$ and the green border indicates significant enrichments (FDR < 0.01).
- (B) Corn plots of the expression pattern of key component genes of development-related signaling pathway in each sample. The pathway components include ligands and receptors, downstream targets, and antagonists.
- (C) Corn plots showing the GSEA analysis of the target genes related to the activated and inhibitory modes of signaling pathways in all samples. The fold changes (divided by mean expression value of each gene across all samples) of DEGs were used as the rank list for GSEA analysis. A hollow dot with a red or green border indicates enrichment FDR < 0.25 by GSEA.
- (D) The territory of active BMP, Wnt, and Nodal signaling in the epiblast revealed by the spatial transcriptome.
- (E) The expression pattern of molecular markers of the precursor cells of the three germ layers across the domains encompassing epiblasts of the anterior (A), left (L), right (R), and posterior (P) regions. Ectoderm genes (red) were expressed predominantly in the anterior and lateral domains, while mesoderm (green) and endoderm (blue) genes were preferentially expressed in the posterior domain.
- (F) Regionalization of prospective cell fates revealed by fate mapping studies and correlated with molecular annotation by germ layer markers. emb-mes, embryonic mesoderm; end, endoderm; exm, extraembryonic mesoderm; ne, neuroectoderm; ps, primitive streak; se, surface ectoderm. Crossed arrows mark the three embryonic axes.



(legend on next page)

embryos of the late to mid-streak stage (see later). This finding raised the possibility that these signature genes (designated as the zip code genes; [Table S2](#)) may also be used for mapping samples of cells of other sources to the equivalent epiblast cell population. To this end, RCCs were computed between each queried sample and all 42 LCM samples based on the expression values of the zip code genes, resulting in 42 RCC values for each sample. The RCC data were displayed in a corn plot (zip code mapping, Corn plotter; [Figure 1F](#)), with the higher RCC indicating a strong probability of an epiblast position. The protocol of zip code mapping is available in the online users' manual on the *iTranscriptome* portal (also see [Supplemental Experimental Procedures](#)).

To demonstrate the utility of zip code mapping, 70 single cells isolated from the anterior and posterior fragments (each including part of the lateral domain) of the epiblast of one mid-gastrulation embryo (stage-matched with the embryos used for generating the spatial transcriptome) were subjected to single-cell RNA-seq. Zip code mapping showed that most of the cells had an RCC value above 0.7 across all these 42 samples ([Figures S7A and S7B](#)). A position with the maximum RCC (single best fit) was retained as the location of this particular single cell. The results showed that 66 of 70 single cells can be mapped back to the original half-epiblast ([Figure 7B](#)). To evaluate whether the single best fit of mapping for a single cell is reliable, we calculated a p value for each single cell through the permutation of zip code genes for 10,000 times ([Figure S7C](#)). We found that all the mapping results were significant ($p < 0.05$). With best-RCC-based single-cell mapping, some single cells could be mapped to a number of adjoining locations. To resolve this ambiguity, a smoothing procedure was used instead to render the single cell's location as a diffused domain that is defined by the top smoothed RCC values (diffused domain of the ten examples are colored in the corn plot; [Figure 7C](#)). Good agreement was found between the maximum RCC ([Figure 7B](#), the same cells in [Figure 7C](#) were numbered) and the smoothed RCC results ([Figure 7C](#)) in pinpointing the single cell's position.

To further evaluate the accuracy of the zip code mapping, we attempted to infer a spatial expression pattern for each zip code gene. First, the mapped embryo position of each of the 70 single

cells were recorded and the weighted average expression of the particular zip code gene in each node of the corn plot was calculated based on single-cell RNA-seq values in this position ([Satija et al., 2015](#)). In this way, a cross-validation was performed by taking out the zip code gene one at a time from the mapping exercise, zip code mapping was then reiterated to re-infer the location of the single cells. Finally, the imputed expression pattern for the held-out gene was compared with its spatial expression pattern as documented in the *iTranscriptome* ([Figure 7D](#)). This reiterative analysis revealed good concordance between the *iTranscriptome* expression pattern and the imputed pattern for the held-out genes ([Figure 7E](#)). When individual 42 sample regions can be correctly classified into the expressed (1) or not expressed (0) state (binarization defined by GBA analysis) as gold standards, the average area under the curve (AUC) for all zip code genes approached 0.7 ([Figures 7F and 7G](#)), indicating that the zip code mapping of the single cells can be achieved with high confidence. However, while most zip code genes performed well (i.e., $AUC > 0.6$), a subset of genes performed less satisfactorily. The reduced fidelity of imputed expression to the cell position might be affected by insufficient coverage of the single cells isolated from a particular location (e.g., only one single cell has been mapped to region 9A) and the range of innate cellular heterogeneity.

We further applied zip code mapping to identify the epiblast cells that are developmentally similar to the epiblast stem cells (EpiSCs). In agreement with our previous finding that EpiSCs are congruent with the anterior primitive streak cells of the mid-gastrulation embryo ([Kaufman-Francis et al., 2014; Kojima et al., 2014a](#)), cells of ten embryo-derived EpiSC lines and one ES cell-derived EpiSC line (ESD-EpiSC) ([Zhang et al., 2010](#)) were mapped to epiblast cell populations that encompassed those in the distal posterior domains (1P–4P) ([Figures 7H, i–iii and 1F](#)). Interestingly, EpiSCs with enhanced mesendoderm propensity displayed the transcriptome characteristics more of the posterior epiblast (the presumptive mesoderm and endoderm progenitors in the anterior primitive streak, [Figure 7H iii](#); cf. [Figure 6F](#)), while EpiSCs with enhanced ectoderm propensity were more similar to the anterior and distal lateral epiblasts that are fated for the ectoderm progenitor ([Figure 7H iv, v](#), cf. [Figure 6F](#)).

Figure 7. Zip Code Mapping of Single Epiblast Cells and Stem Cells

- (A) The SVM prediction of domain label for each sample in embryos E2 and E3 using the scaled expression values of marker genes from reference embryo E1 as the input features. The accuracy is calculated by comparing the predicted and real domain labels. X, imperfect match.
- (B) Zip code mapping of single cells isolated from the anterior and posterior fragments of the epiblast of mid-gastrulation embryo. The positions of the single cells in the epiblast are predicted based on the single-best fit maximum RCC value. Yellow dots, cells from the anterior epiblast fragment; blue dots, cell from the posterior epiblast fragment. Cells that were also mapped to diffused domains ([Figure 7C](#)) by the smoothing procedure are numbered.
- (C) The location of single cells (numbered in [Figure 7B](#), five single cells each from anterior and posterior regions) defined as diffused domains based on the top smoothed RCC values. Nodes of the six top values are shown for each single cells in the corn plots.
- (D) The workflow of assessment of single-cell zip code mapping accuracy (detailed method in the [Supplemental Experimental Procedures](#)).
- (E) The mapping results of the *iTranscriptome* expression pattern (derived from spatial RNA-seq) and the imputed expression pattern of a zip code gene (see [Supplemental Experimental Procedures](#) for details of the analysis). The results of five randomly selected held-out zip code genes are presented. The reference sample regions that the single cells were not mapped to are shown as hollow spots.
- (F) Receiver operating characteristics results showing the accuracy of the imputed expression levels versus the binarized expression values from the spatial RNA-seq data for the five randomly selected zip code genes in (C). AUC values are in brackets.
- (G) Histogram summarizing AUC values of all zip code genes.
- (H) Zip code mapping of EpiSCs to five patterns of equivalent in vivo cell populations in the epiblast: (i) posterior-distal, lateral, and anterior epiblast, (ii) posterior-distal and distal lateral epiblast, (iii) posterior epiblast and lateral-intermediate epiblast, (iv) lateral epiblast and distal posterior epiblast, and (v) lateral epiblast and anterior epiblast. The table shows examples of EpiSC lines with different propensity of neuroectoderm and mesendoderm differentiation revealed by the expediency of activation of the lineage markers (data from GEO: GSE46227).

See also [Figure S7](#).

Zip code mapping by spatial transcriptome can therefore be applied for characterizing the spectrum of prospective cell fates of embryo-derived stem cells such as the epiblast stem cells.

DISCUSSION

Post-implantation mouse development, especially during gastrulation, is a complex process involving rapid cell proliferation, extensive tissue re-modeling and significant changes in transcriptional architecture. It is technically challenging to study the molecular activities underlying the mouse gastrulation due to the limited amount of tissues that could be procured from an individual embryo. Moreover, the rapid changes in genetic activity require a genome-wide and unbiased gene expression profiling of embryos within a narrow window of development. In this study, we have profiled the transcriptome by low-input RNA-seq analysis of groups of about 20 cells in the epiblast of single mid-gastrulation (late mid-streak) stage embryos that were systematically sampled by laser capture microdissection and with all the spatial information retained. On the basis that the genome-wide transcriptome data are integrated with the positional address of the cell samples from a single embryo, we termed this approach single-embryo spatial RNA-seq. Analysis of the single-embryo spatial transcriptome enabled the characterization of expression domains in the epiblast (that correspond to the regionalization of cell fates), the identification of region-specific marker genes for each domain, the inference on the interaction network for TFs and signaling pathways, and the generation of an in-depth quantitative in silico expression map of over 20,000 transcribed genes in the epiblast of mid-gastrulation embryos in a 3D topographic context.

Fate mapping studies of the gastrulating embryo have demonstrated that the progenitors of the major germ layer tissues reside in different regions in the epiblast (Tam and Behringer, 1997; Tam et al., 2006). However, fate maps reveal little about the molecular activities underlying cell fate specification. Although the current study only focuses on one developmental stage, we identified gene expression domains and subdomains that represent the different transcriptional states in the epiblast, providing a high-fidelity molecular annotation for the fate map of germ layer progenitors. Many of the signature genes are known to be functionally relevant to germ layer specification. However, a large number of genes with uncharacterized function could be important new regulators and await further investigation. In this light, the mouse spatial RNA-seq dataset is an invaluable resource for elucidating the functional genomics of lineage specification and differentiation.

The advent of high-resolution transcriptome analysis, particularly of single cells, opens a venue to investigate in depth the molecular mechanism of embryogenesis, tissue patterning, lineage allocation, and differentiation (Klein et al., 2015; Macosko et al., 2015; Morrison et al., 2015; Tang et al., 2010; Xue et al., 2013; Yan et al., 2013). For single-cell analysis of tissues, organs, or embryos where topographical location of the cells has a significant bearing on the study of regionalized cellular heterogeneity, it is imperative to know the positional address of the cell of interest. While barcoding or unique gene expression profiling may identify the genetic materials sourced from individual cells (Wilson et al., 2015), an effective systematic coding of the cells by

position is not yet feasible. Spatial mapping of cell positions have been reported recently; however, these studies either revealed spatial resolution primarily in a 2D context with a much lower number of genes detected (Zechel et al., 2014), or required the assembly of sets of known position markers to build a reference map of positional addresses (Achim et al., 2015; Satija et al., 2015), and thus may have limitations in wider application when a detailed expression pattern of gene markers is not available for precise positioning of the single cells in the tissue or embryo. An unbiased spatial transcriptome of the zebrafish embryo has been generated by the tomo-seq technique (Junker et al., 2014). As it is based on the tomographic reconstruction of the RNA-seq data from multiple embryos, the dataset might be subject to between-embryo heterogeneity even if the samples were nominally of the same developmental stages. Analysis of our spatial transcriptome of the mouse epiblast through the unbiased interrogation of the gene expression profile of small cell populations from precisely defined anatomical sites has allowed us to define a set of position-related genes. This gene set could be used as the zip code for mapping the position of a specific cell population of unknown address to the best-matched transcriptional domain in the epiblast. As a validation of this approach, 94% of the single cells were mapped to the correct halves of the embryo and each cell to a predicted position with a high degree of precision (Figure 7B). The unique attributes of the positional dataset are (1) it is a stage-specific reference for the epiblast of embryos at a defined developmental stage, and (2) the dataset is generated from the region-specific quantitative profile of the genome-wide transcriptome of cell populations at known locations. The latter attribute distinguishes our analytical approach (based on d-WISH) from that of two recently published studies (Achim et al., 2015; Satija et al., 2015), where the spatial information is derived from the binary scores of gene expression (based on conventional ISH) at each grid position. It is envisaged that with further training of the dataset using transcriptome data of additional single cells harvested from cell populations at defined positions, the zip code functionality could be enhanced to allow precise mapping the position of every single cell to the mid-gastrulation epiblast. Work is in progress for the acquisition of new single-cell data to be used in machine learning for zip code mapping, and the generation of a spatial transcriptome for the full range of stages of gastrulation. Our spatial transcriptome dataset can also be used for pinpointing the best-fit cell populations that match the embryo-derived stem cells. We demonstrated the utility of zip code mapping to delineate the heterogeneous lineage propensity (the spectrum of cell fates displayed upon differentiation) of EpiSCs derived from the epiblast and the ESCs. The mapping functionality of the spatial transcriptome is a useful adjunct to the elucidation of the cell potency and lineage property of multipotent stem cells.

The transcriptome profiling of the current study provide a snapshot of the molecular activity at one stage of gastrulation. Extension of this study to cover the whole period of gastrulation would allow a longitudinal study of lineage hierarchy and diversification of cell populations based on the history of expression of lineage-specific marker genes, gleaned from the expression profile or by lineage-gene reporter expression, in selected cell lineages throughout gastrulation. Taken together, the spatial RNA-seq data of the *iTranscriptome*

provide an unprecedented resource of information on the lineage status and differentiation trajectories of epiblast cells and transcriptional and signaling activity that are vital for lineage specification and differentiation, and such knowledge gleaned from the spatial transcriptome has significant implications for understanding the mechanism of embryogenesis and the biological attributes of stem cells.

EXPERIMENTAL PROCEDURES

Embryo Laser Capture Microdissection and RNA Isolation

Late mid-streak (~E7.0) embryos (Downs and Davies, 1993) of C57BL/6J mice were embedded in OCT compound and cryosectioned serially at 15 μ m. Alternate sections were mounted on polyethylene terephthalate-coated slides and on regular glass slides. Frozen sections were allowed to quickly thaw at room temperature and then dehydrated in ice-cold 100% ethanol for 30 s. Fixation was performed in 75% ethanol for 30 s, then the slides were stained for 30 s with 1% cresyl violet acetate solution (Sigma-Aldrich, prepared in 75% ethanol), dehydrated in a series of 75%, 95%, 100% ethanol (30 s for each step), and finally subjected to LCM on an MMI Cellcut Plus system (MMI, Zurich, Switzerland). Sections mounted on plain slides were stained with 1% cresyl violet acetate solution for histology and imaged for the construction of the 3D embryo template for data presentation.

Approximately 20 cells from four positions in each section were harvested by LCM. Cell samples were lysed in 50 μ l of 4 M guanidine isothiocyanate solution (GuSCN; Invitrogen, 15577-018) at 42°C for 10 min. The volume of the lysate was adjusted to 200 μ l by nuclease-free water, and was further concentrated by ethanol precipitation in the presence of 1/10 volume of acetate sodium (pH 5.7, 3 M; Ambion) and 2 μ l of carrier glycogen (20 mg/ml; Roche). Total RNA pellets were dissolved in lysis solution and used as a template for low-cell number RNA-seq.

Mouse Embryo Single-Cell RNA-Seq

E7.0 C57BL/6E embryos were harvested and incubated in 0.5% Trypsin and 2.5% pancreatin at 4°C. The epiblast dissected free from the extraembryonic tissues, the mesoderm, and endoderm, was bisected into the anterior and the posterior fragments. These fragments were dissociated into single cells by digestion with 50 μ l of Accutase (STEMCELL Technologies) at 37°C for 15 min. The single-cell suspension was diluted in PBS-BSA solution and individual cells were picked into PCR tubes by micropipette for modified Smart2-seq (Picelli et al., 2014).

All animal experiments were performed in compliance with the guidelines of the Animal Ethical Committee of the Institute of Biochemistry and Cell Biology, Shanghai Institutes for Biological Sciences, Chinese Academy of Sciences.

Identification of Inter-domain Differentially Expressed Genes

To find the DEGs between different domains, we first used the unsupervised hierarchical clustering method to classify all samples into four domains (based on a distinctly separated dendrogram; Figure S2A) by using all genes expressed (FPKM > 1.0) in more than two samples and with a variance in transcript level ($\log_{10}(\text{FPKM} + 1)$) across all samples greater than 0.05. Next, to identify inter-domain differentially expressed genes, the genes in each domain were compared with other three domains using the t test in MATLAB with 10,000 permutations (Huber et al., 2002). Genes were defined as DEGs if they exhibited a p value ≤ 0.01 and a mean fold change of >2.0 or <0.5. 1887 DEGs were thus identified.

Clustering Analysis for Inter-domain DEGs

We used our recently adaptive clustering algorithm based on the Bayesian Information Criterion (BIC) to determine the optimal number of gene clusters, and coupled this with deep clustering using our developed unsupervised super k-means algorithm (called BIC-SKmeans) (Zhang et al., 2013). Hierarchical clustering with correlation distance metric was used for sample clustering. Before clustering, we used the Z score to normalize gene expression levels across all samples.

Mouse Embryo 3D Model Reconstruction

The mouse embryo 3D model is visualized by Vaa3D (Peng et al., 2010). For 3D visualization of the gene expression pattern, four regions were first colored in each 2D image of the sample slice, performed by a MATLAB-based custom-written program, and the slices were assembled in a 3D format for visualization by Vaa3D. For 3D visualization of the four transcription domains, each region was colored in the 2D image of the sample slice before the input of the colored slice images into Vaa3D.

Identifying Additional Marker Genes for Anterior and Posterior Regions Based on Seed Genes

To uncover previously unknown anterior and posterior specific marker genes besides the marker genes that are identified by PCA loading (using the FactoMineR package in R), a “seed gene” for each of these two domains was defined as the gene having the highest square cosine (\cos^2) correlation to both PC1 and PC2 simultaneously, which is carried out using the FactoMineR package (the plot.PCA function with parameter “select” as \cos^2). GBA (Walker et al., 1999) analysis was used to find the co-expressed genes with the seed genes. To do this, all gene expression values were first Z score normalized across all samples. Then, a binary expression matrix was built by defining a gene as expressed (defined as 1) in a given sample if the normalized expression value was larger than or equal to 0, and as not expressed (defined as 0) if it was smaller than 0. Finally, the p values for observing a given co-expression of two genes were calculated using hypergeometric distribution, followed by Benjamini Hochberg (BH) multiple testing correction. Only co-expressed genes with $\text{FDR} \leq 0.05$ were identified as additional marker genes for both anterior and posterior regions.

Comparison between RNA-Seq Data and In Situ Data

To calculate the correlation of RNA-seq and in situ data for a particular gene, the expression values were scale to [0, 1] using the formula $(E_i - \min(E)) / (\max(E) - \min(E))$, ($i = 1, \dots, 42$), where E is the expression values of this gene in all 42 samples, and then Pearson correlation coefficients were calculated.

Zip Code Mapping for Single Cells

To test the utility of zip code gene set for mapping the position of single cells isolated separately from the anterior and posterior epiblast fragments, the Spearman RCCs between the expression values of the zip code genes of each single cell and all 42 reference samples of the reference embryo were calculated, resulting in 42 RCC values for each testing cell. The single cell was then mapped to the sample region of the reference epiblast that has the maximum RCC value.

To evaluate whether the single best fit of mapping for a single cell is reliable, we calculated a p value for each single cell through the permutation of zip code genes. Briefly, the expression of zip code genes in each sample in reference embryos were randomly permuted 10,000 times, and for each permutation, all single cells were mapped to a reference embryo by the above procedure. Then, for each single cell, 10,000 locations were predicted in the reference embryo and compared with the original location (without permutation). Finally, a p value, which measures the ratio of correctly predicted locations (without permutation) in all permutations, was calculated to evaluate the significance of mapping for each single cell.

For the RCC-based single-cell mapping, the location is determined based on the maximum RCC value for each single cell. There were, however, cases where a single cell could match to a few adjoined locations. A smoothing procedure was applied to fit this single cell's location as a diffused domain (see the method in the next section).

Inner Working of the Zip Code Mapping Utility

1. Top 40 genes with highest or lowest PCA loadings in the first two PCs were selected from genes that were expressed (FPKM > 1.0) in at least two samples and with a variance in transcript level ($\log_{10}(\text{FPKM} + 1) > 0.05$) across all samples as the zip code genes (158 zip code genes were identified).
2. Calculate the Spearman RCCs between 42 reference samples and the user's query transcriptome data based on the expression of these zip

code genes. The RCC values for each query sample were visualized in a corn plot. The higher RCC indicates a strong probability of a matching position.

3. A smoothing procedure was applied to fit the query sample's location as a diffused domain. In the smooth processing, 42 RCC values between all samples of reference embryo and the query sample were computed. Each RCC value was mapped on the corn plot by calculating the average value of this RCC and the RCCs of cell samples of adjoining regions in the same spatial domain. The position of the single cell was displayed as a diffused domain, encompassing regions with top smoothed RCC values, in the corn plot (e.g., Figure 7C).

Zip Code Mapping for EpiSCs

The gene expression datasets of ten mouse EpiSC lines retrieved from the GEO database (GEO: GSE46227) and the data of one EpiSC line generated by in vitro conversion of ESCs (ESD-EpiSC) were used for zip code mapping. Spearman RCCs were computed between each query sample and all 42 samples of the reference embryo based on the expression values of the zip code genes. The RCC values for each EpiSC sample were collated in a corn plot.

Web Service

The django package in python was used to construct the web site. The SQLite database was used to store the original gene expression data, the GBA values of any two genes in E1, and the mouse gene annotation data (downloaded from NCBI). The interfaces enable the user to search for the expression pattern of queried genes, and for genes that display a similar expression profile to a query gene or pattern. In addition, transcriptome data (from microarray or RNA-seq analysis) can be submitted to the web portal for zip code mapping onto the spatial transcriptome for delineating cell identity.

ACCESSION NUMBERS

The RNA-seq data reported in this paper were deposited in NCBI Gene Expression Omnibus under accession number GEO: GSE65924.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, seven figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.devcel.2016.02.020>.

AUTHOR CONTRIBUTIONS

N.J., P.P.L.T., and J.D.J.H. conceived the study. N.J. and J.D.J.H. supervised the project. N.J., P.P.L.T., J.D.J.H., and G.P. designed the experiments. G.P., J.C., F.Y., C.L., S.C., G.C., and L.S. performed experiments. S.S., R.W., and N.S. analyzed the sequencing data. W.C. wrote algorithms for 3D image construction and in situ image analysis. G.P., P.P.L.T., J.D.J.H., S.S., and N.J. wrote the paper with the help of all other authors.

ACKNOWLEDGEMENTS

We thank the Uli Schwarz public laboratory platform in PICB, SIBS, CAS for technical support, Fuchou Tang for help with the single-cell sequencing technology and insightful ideas, Janet Rossant for invaluable suggestions, Pingyu Liu and Qingqing Zhu for assistance and providing materials, Yun Qian for the care of animals, and all the laboratory members for discussions and helpful suggestions. This work was supported by grants from the Strategic Priority Research Program of the Chinese Academy of Sciences (XDA01010201 to N.J., XDA01010303 to J.D.J.H.), National Key Basic Research and Development Program of China (2014CB964804, 2015CB964500, 2015CB964803), and National Natural Science Foundation of China (91219303, 31430058, 31401261, 91329302, 31210103916, and 91519330). P.P.L.T. is a Senior Principal Research Fellow of the National Health and Medical Research Council of Australia (grants 1003100 and 1110751).

Received: September 25, 2015

Revised: February 10, 2016

Accepted: February 22, 2016

Published: March 21, 2016

REFERENCES

- Achim, K., Pettit, J.B., Saraiva, L.R., Gavriouchkina, D., Larsson, T., Arendt, D., and Marioni, J.C. (2015). High-throughput spatial mapping of single-cell RNA-seq data to tissue of origin. *Nat. Biotechnol.* **33**, 503–509.
- Arnold, S.J., and Robertson, E.J. (2009). Making a commitment: cell lineage allocation and axis patterning in the early mouse embryo. *Nat. Rev. Mol. Cell Biol.* **10**, 91–103.
- Bass, J.I.F., Diallo, A., Nelson, J., Soto, J.M., Myers, C.L., and Walhout, A.J.M. (2013). Using networks to measure similarity between genes: association index selection. *Nat. Methods* **10**, 1169–1176.
- Beddington, R.S. (1981). An autoradiographic analysis of the potency of embryonic ectoderm in the 8th day postimplantation mouse embryo. *J. Embryol. Exp. Morphol.* **64**, 87–104.
- Beddington, R.S. (1982). An autoradiographic analysis of tissue potency in different regions of the embryonic ectoderm during gastrulation in the mouse. *J. Embryol. Exp. Morphol.* **69**, 265–285.
- Cajal, M., Lawson, K.A., Hill, B., Moreau, A., Rao, J., Ross, A., Collignon, J., and Camus, A. (2012). Clonal and molecular analysis of the prospective anterior neural boundary in the mouse embryo. *Development* **139**, 423–436.
- Chapman, D.L., Agulnik, I., Hancock, S., Silver, L.M., and Papaioannou, V.E. (1996). Tbx6, a mouse T-box gene implicated in paraxial mesoderm formation at gastrulation. *Dev. Biol.* **180**, 534–542.
- Chung, N.C., and Storey, J.D. (2015). Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554.
- Downs, K.M., and Davies, T. (1993). Staging of gastrulating mouse embryos by morphological landmarks in the dissecting microscope. *Development* **118**, 1255–1266.
- Hoodless, P.A., Pye, M., Chazaud, C., Labbe, E., Attisano, L., Rossant, J., and Wrana, J.L. (2001). FoxH1 (Fast) functions to specify the anterior primitive streak in the mouse. *Genes Dev.* **15**, 1257–1271.
- Huang, D.W., Sherman, B.T., and Lempicki, R.A. (2009). Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57.
- Huber, W., von Heydebreck, A., Sultmann, H., Poustka, A., and Vingron, M. (2002). Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** (Suppl 1), S96–S104.
- Junker, J.P., Noel, E.S., Guryev, V., Peterson, K.A., Shah, G., Huisken, J., McMahon, A.P., Berezikov, E., Bakkers, J., and van Oudenaarden, A. (2014). Genome-wide RNA tomography in the zebrafish embryo. *Cell* **159**, 662–675.
- Kaufman-Francis, K., Goh, H.N., Kojima, Y., Studdert, J.B., Jones, V., Power, M.D., Wilkie, E., Teber, E., Loebel, D.A., and Tam, P.P. (2014). Differential response of epiblast stem cells to Nodal and Activin signalling: a paradigm of early endoderm development in the embryo. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **369**, <http://dx.doi.org/10.1098/rstb.2013.0550>.
- Klein, A.M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187–1201.
- Kojima, Y., Kaufman-Francis, K., Studdert, J.B., Steiner, K.A., Power, M.D., Loebel, D.A., Jones, V., Hor, A., de Alencastro, G., Logan, G.J., et al. (2014a). The transcriptional and functional properties of mouse epiblast stem cells resemble the anterior primitive streak. *Cell Stem Cell* **14**, 107–120.
- Kojima, Y., Tam, O.H., and Tam, P.P. (2014b). Timing of developmental events in the early mouse embryo. *Semin. Cell Dev. Biol.* **34C**, 65–75.
- Lawson, K.A., Meneses, J.J., and Pedersen, R.A. (1991). Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* **113**, 891–911.

- Li, L., Liu, C., Biechele, S., Zhu, Q., Song, L., Lanner, F., Jing, N., and Rossant, J. (2013). Location of transient ectodermal progenitor potential in mouse development. *Development* 140, 4533–4543.
- Liu, Y., Kaneda, R., Leja, T.W., Subkhankulova, T., Tolmachov, O., Minchiotti, G., Schwartz, R.J., Barahona, M., and Schneider, M.D. (2014). Hhex and Cer1 mediate the Sox17 pathway for cardiac mesoderm formation in embryonic stem cells. *Stem Cells* 32, 1515–1526.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M., et al. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* 161, 1202–1214.
- Mitiku, N., and Baker, J.C. (2007). Genomic analysis of gastrulation and organogenesis in the mouse. *Dev. Cell* 13, 897–907.
- Morrison, J.A., Box, A.C., McKinney, M.C., McLennan, R., and Kulesa, P.M. (2015). Quantitative single cell gene expression profiling in the avian embryo. *Dev. Dyn.* 244, 774–784.
- Osorno, R., Tsakiridis, A., Wong, F., Cambray, N., Economou, C., Wilkie, R., Blin, G., Scotting, P.J., Chambers, I., and Wilson, V. (2012). The developmental dismantling of pluripotency is reversed by ectopic Oct4 expression. *Development* 139, 2288–2298.
- Peng, H., Ruan, Z., Long, F., Simpson, J.H., and Myers, E.W. (2010). V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat. Biotechnol.* 28, 348–353.
- Pfister, S., Steiner, K.A., and Tam, P.P. (2007). Gene expression pattern and progression of embryogenesis in the immediate post-implantation period of mouse development. *Gene Expr. Patterns* 7, 558–573.
- Picelli, S., Faridani, O.R., Bjorklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nat. Protoc.* 9, 171–181.
- Power, M.A., and Tam, P.P. (1993). Onset of gastrulation, morphogenesis and somitogenesis in mouse embryos displaying compensatory growth. *Anat. Embryol. (Berl)* 187, 493–504.
- Raya, A., and Izpisua Belmonte, J.C. (2006). Left-right asymmetry in the vertebrate embryo: from early information to higher-level integration. *Nat. Rev. Genet.* 7, 283–293.
- Robb, L., and Tam, P.P. (2004). Gastrula organiser and embryonic patterning in the mouse. *Semin. Cell Dev. Biol.* 15, 543–554.
- Rojas, A., De Val, S., Heidt, A.B., Xu, S.M., Bristow, J., and Black, B.L. (2005). Gata4 expression in lateral mesoderm is downstream of BMP4 and is activated directly by Forkhead and GATA transcription factors through a distal enhancer element. *Development* 132, 3405–3417.
- Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* 33, 495–502.
- Snow, M. (1977). Gastrulation in the mouse: growth and regionalization of the epiblast. *J. Embryol. Exp. Morphol.* 42, 293–303.
- Solnica-Krezel, L., and Sepich, D.S. (2012). Gastrulation: making and shaping germ layers. *Annu. Rev. Cell Dev. Biol.* 28, 687–717.
- Tam, P.P., and Beddington, R.S. (1992). Establishment and organization of germ layers in the gastrulating mouse embryo. *Ciba Found. Symp.* 165, 27–41, [discussion: 42–9].
- Tam, P.P., and Behringer, R.R. (1997). Mouse gastrulation: the formation of a mammalian body plan. *Mech. Dev.* 68, 3–25.
- Tam, P.P., and Loebel, D.A. (2007). Gene function in mouse embryogenesis: get set for gastrulation. *Nat. Rev. Genet.* 8, 368–381.
- Tam, P.P., Loebel, D.A., and Tanaka, S.S. (2006). Building the mouse gastrula: signals, asymmetry and lineages. *Curr. Opin. Genet. Dev.* 16, 419–425.
- Tang, F., Barbacioru, C., Bao, S., Lee, C., Nordman, E., Wang, X., Lao, K., and Surani, M.A. (2010). Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell RNA-Seq analysis. *Cell Stem Cell* 6, 468–478.
- Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. (2009). Epithelial-mesenchymal transitions in development and disease. *Cell* 139, 871–890.
- Walker, M.G., Volkmut, W., Sprinzak, E., Hodgson, D., and Klingler, T. (1999). Prediction of gene function by genome-scale expression analysis: prostate cancer-associated genes. *Genome Res.* 9, 1198–1203.
- Wang, B., Yan, X.F., and Jiang, Q.C. (2015). Loading-based principal component selection for PCA integrated with support vector data description. *Ind. Eng. Chem. Res.* 54, 1615–1627.
- Wilson, N.K., Kent, D.G., Buettner, F., Shehata, M., Macaulay, I.C., Calero-Nieto, F.J., Sanchez Castillo, M., Oedekoven, C.A., Diamanti, E., Schulte, R., et al. (2015). Combined single-cell functional and gene expression analysis resolves heterogeneity within stem cell populations. *Cell Stem Cell* 16, 712–724.
- Xue, H., Xian, B., Dong, D., Xia, K., Zhu, S., Zhang, Z., Hou, L., Zhang, Q., Zhang, Y., and Han, J.D. (2007). A modular network model of aging. *Mol. Syst. Biol.* 3, 147.
- Xue, Z., Huang, K., Cai, C., Cai, L., Jiang, C.Y., Feng, Y., Liu, Z., Zeng, Q., Cheng, L., Sun, Y.E., et al. (2013). Genetic programs in human and mouse early embryos revealed by single-cell RNA sequencing. *Nature* 500, 593–597.
- Yamamoto, M., Meno, C., Sakai, Y., Shiratori, H., Mochida, K., Ikawa, Y., Saijoh, Y., and Hamada, H. (2001). The transcription factor FoxH1 (FAST) mediates Nodal signaling during anterior-posterior patterning and node formation in the mouse. *Genes Dev.* 15, 1242–1256.
- Yan, L., Yang, M., Guo, H., Yang, L., Wu, J., Li, R., Liu, P., Lian, Y., Zheng, X., Yan, J., et al. (2013). Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1131–1139.
- Zechel, S., Zajac, P., Lonnerberg, P., Ibanez, C.F., and Linnarsson, S. (2014). Topographical transcriptome mapping of the mouse medial ganglionic eminence by spatially resolved RNA-seq. *Genome Biol.* 15, 486.
- Zhang, K., Li, L., Huang, C., Shen, C., Tan, F., Xia, C., Liu, P., Rossant, J., and Jing, N. (2010). Distinct functions of BMP4 during different stages of mouse ES cell neural commitment. *Development* 137, 2095–2105.
- Zhang, W., Liu, Y., Sun, N., Wang, D., Boyd-Kirkup, J., Dou, X., and Han, J.D. (2013). Integrating genomic, epigenomic, and transcriptomic features reveals modular signatures underlying poor prognosis in ovarian cancer. *Cell Rep.* 4, 542–553.
- Zhu, Q., Song, L., Peng, G., Sun, N., Chen, J., Zhang, T., Sheng, N., Tang, W., Qian, C., Qiao, Y., et al. (2014). The transcription factor Pou3f1 promotes neural fate commitment via activation of neural lineage genes and inhibition of external signaling pathways. *Elife* 3, e02224.

Update

Developmental Cell

Volume 55, Issue 6, 21 December 2020, Page 802–804

DOI: <https://doi.org/10.1016/j.devcel.2020.11.018>

Correction

Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo

Guangdun Peng, Shengbao Suo, Jun Chen, Weiyang Chen, Chang Liu, Fang Yu, Ran Wang, Shirui Chen, Na Sun, Guizhong Cui, Lu Song, Patrick P.L. Tam, Jing-Dong J. Han,^{*} and Naihe Jing^{*}

^{*}Correspondence: jdhan@picb.ac.cn (J.-D.J.H.), njing@sibcb.ac.cn (N.J.)
<https://doi.org/10.1016/j.devcel.2020.11.018>

(Developmental Cell 36, 681–697; March 21, 2016)

During preparation of Figure 1A, the left and right sides all of tissue images were labeled in reverse and were noted in this way in the figure and figure legend in the originally published version of this article. The originally published paper reported that there was no left-right asymmetry at E7.0 stage, so the L-R inversion did not change our results, and the overall conclusion is unaffected. However, the authors have noticed this error and are correcting their paper. The corrected Figure 1 and Figure 1A legend are below. The authors apologize for any confusion generated by the original figure.

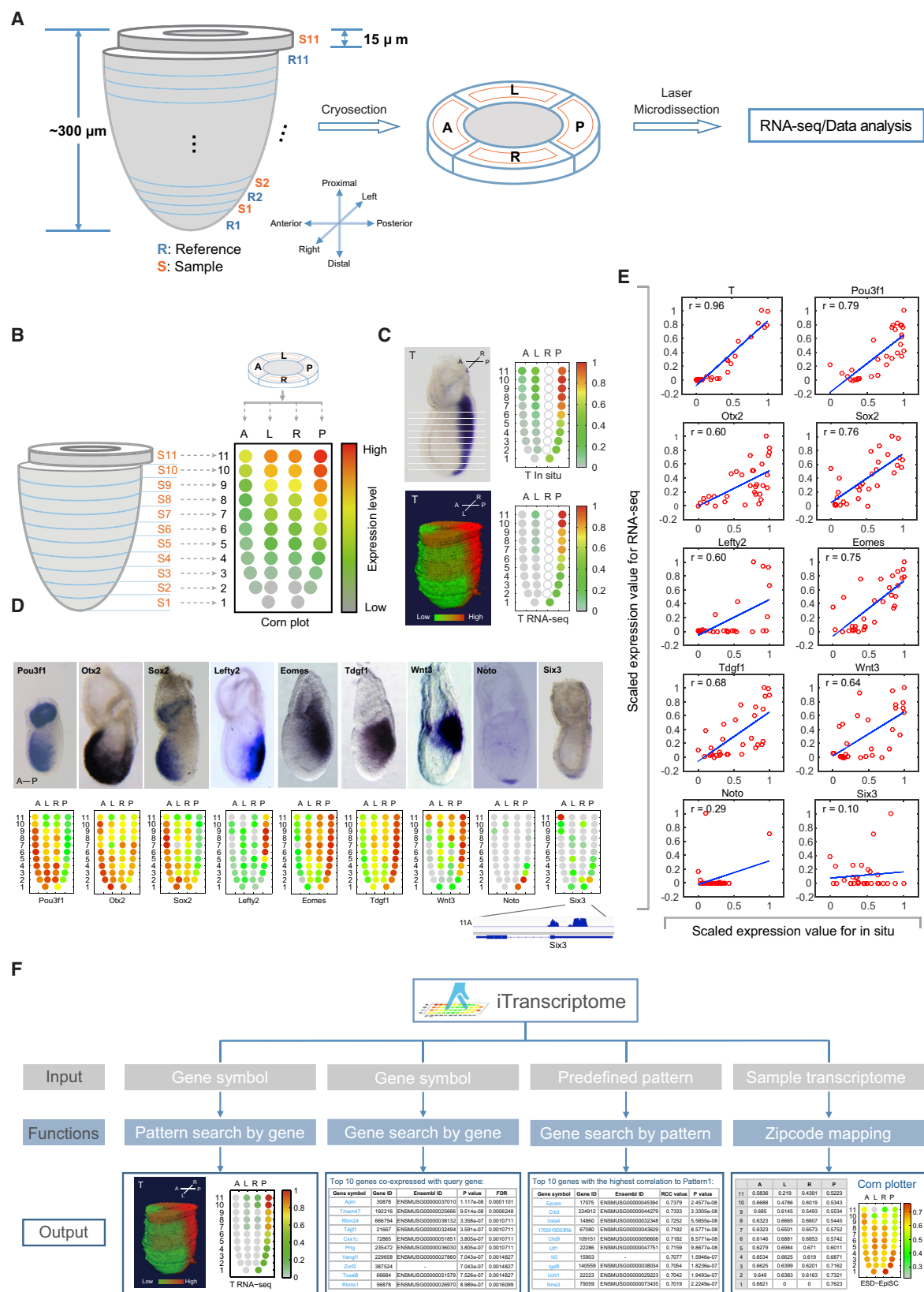


Figure 1A. Spatial RNA-Seq Analysis of the Mid-gastrulation Embryo

(A) Experimental strategy: cells were captured by laser capture microdissection from four quadrants: anterior (A), posterior (P), and lateral (left/right, L/R), of each sample section (S1–S11) of the epiblast of late mid-streak stage (E7.0) C57BL/6 embryos, and analyzed by RNA-seq. The reference sections (R1–R11) were used as templates for 3D reconstruction of the embryo for data visualization. The left and right quadrants were mirrored positions from the real section, so L and R should be considered in the respective right and left side viewed from the primitive streak.