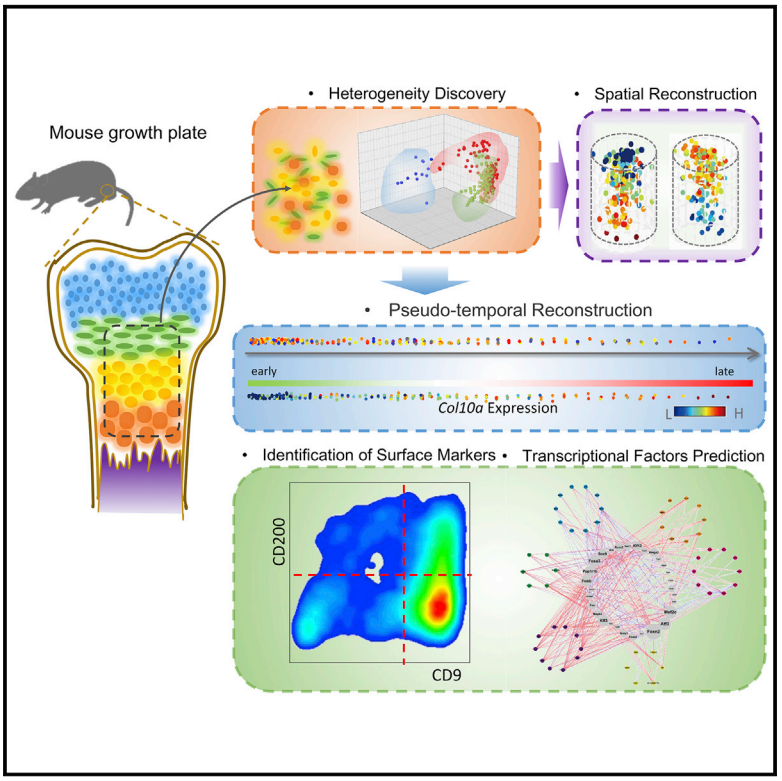
Article

Cell Reports

Systematic Reconstruction of Molecular Cascades Regulating GP Development Using Single-Cell RNA-Seq

Graphical Abstract



Authors

Junxiang Li, Haofei Luo, Rui Wang, ..., Yi Yao, Geng Tian, Qiong Wu

Correspondence

wuqiong@tsinghua.edu.cn

In Brief

Li et al. have developed an unsupervised clustering approach called Sinova to analyze single-cell RNA-seq data. By using this pipeline to analyze single-cell RNA-seq data from developing GPs in mice, they have generated a spatial and temporal map of the GP and identified molecular networks involved in GP development.

Highlights

- Unbiased clustering of single cells from GPs identifying stage transitions
- Temporal and spatial reconstruction of GP maturation
- Prediction of the TF portfolio directing GP maturation
- Recapitulation of TF portfolio using an in vitro screening system

Accession Numbers

GSE76157

Li et al., 2016, Cell Reports *15*, 1467–1480
May 17, 2016
http://dx.doi.org/10.1016/j.celrep.2016.04.043

Cell Reports
Article

Systematic Reconstruction of Molecular Cascades Regulating GP Development Using Single-Cell RNA-Seq

Junxiang Li,[1,2,5] Haofei Luo,[2,5] Rui Wang,[1,2] Jidong Lang,[3] Siyu Zhu,[2] Zhenming Zhang,[1,2] Jianhuo Fang,[3] Keke Qu,[1,2] Yuting Lin,[3] Haizhou Long,[1,2,4] Yi Yao,[1,2,4] Geng Tian,[3] and Qiong Wu[1,2,4,*]
[1]MOE Key Laboratory of Bioinformatics, Center for Synthetic and System Biology, Tsinghua University, Beijing, China
[2]School of Life Sciences, Tsinghua University, Beijing 100084, China
[3]School of Medicine, Tsinghua University, Beijing 10084, China
[4]Center for Synthetic & System Biology, Tsinghua University, Beijing 10084, China
[5]Co-first author
*Correspondence: wuqiong@tsinghua.edu.cn
http://dx.doi.org/10.1016/j.celrep.2016.04.043

SUMMARY

The growth plate (GP) comprising sequentially differentiated cell layers is a critical structure for bone elongation and regeneration. Although several key regulators in GP development have been identified using genetic perturbation, systematic understanding is still limited. Here, we used single-cell RNA-sequencing (RNA-seq) to determine the gene expression profiles of 217 single cells from GPs and developed a bioinformatics pipeline named Sinova to de novo reconstruct physiological GP development in both temporal and spatial high resolution. Our unsupervised model not only confirmed prior knowledge, but also enabled the systematic discovery of genes, potential signal pathways, and surface markers CD9/CD200 to precisely depict development. Sinova further identified the effective combination of transcriptional factors (TFs) that regulates GP maturation, and the result was validated using an in vitro EGFP-Col10a screening system. Our case systematically reconstructed molecular cascades in GP development through single-cell profiling, and the bioinformatics pipeline is applicable to other developmental processes.

INTRODUCTION

Mammalian body dimensions are primarily determined by the lengthening of skeletal structures; this lengthening is controlled by growth plate (GP) development, and the GP is composed of at least four hierarchical zones that are arranged in columns that are aligned parallel to the longitudinal axis of the bone: the resting zone (RZ), the proliferative zone (PZ), the prehypertrophic zone (PHZ), and the hypertrophic zone (HZ) (Yeung Tsang et al., 2014) (Figure 1A, left). RZ chondrocytes serve as a reservoir of cells that later form the GP. The proliferative zone is located just below the RZ and consists of proliferative chondrocytes; these cells rapidly proliferate and undergo stepwise terminal differentiation, forming the hypertrophic zone (Yeung Tsang et al., 2014). The progressive transformation of proliferative chondrocytes to hypertrophic cells followed by ossification is the basic process underlying both bone lengthening in adolescence and fracture repair after injury (Emons et al., 2011; Scotti et al., 2013; Yeung Tsang et al., 2014). Abnormal growth development may cause severe skeletal dysfunction (Kornak and Mundlos, 2003). Therefore, understanding the precise molecular mechanisms of GP development is critical for understanding both physiological and pathological phenomena.

Classical approaches to understand developmental regulatory mechanisms include either creating an in vitro differentiation model or characterizing model animals with genetic perturbation at a few time points throughout development. Although in vitro models that mimic GP development have been reported, these models require a long cell-culture interval, with increasing system instability, and only partially recapitulate the process (Mueller and Tuan, 2008; Yao and Wang, 2013). Therefore, studies of GP development primarily rely on mouse models using either genetic perturbations or fluorescent tracing (Leung et al., 2011; Yang et al., 2014b; Yeung Tsang et al., 2014). Despite recent insights into the indispensable signaling pathway regulators and specific biomarkers, a more general and systematic understanding of this process remains incomplete.

Single-cell transcriptome sequencing has recently increased in popularity because it enables the gene expression profiling of single cells (Navin, 2015; Trapnell, 2015). It is difficult to restore the temporal and spatial pattern of individual cells due to a loss of this information in major single-cell techniques, as well as to identify the systematic regulators during the stage transition of development (Achim et al., 2015; Durruthy-Durruthy et al., 2014; Shin et al., 2015; Treutlein et al., 2014; Waldhaus et al., 2015; Wen and Tang, 2014). Conceptually similar to other developmental scenarios, the GP is a structure that harbors a sequential differentiation process across the proliferative zone toward the hypertrophic zone (Li et al., 2015); we reasoned that each single cell from postnatal GPs individually represents a given differentiation stage, and we sequenced the transcriptome of 217

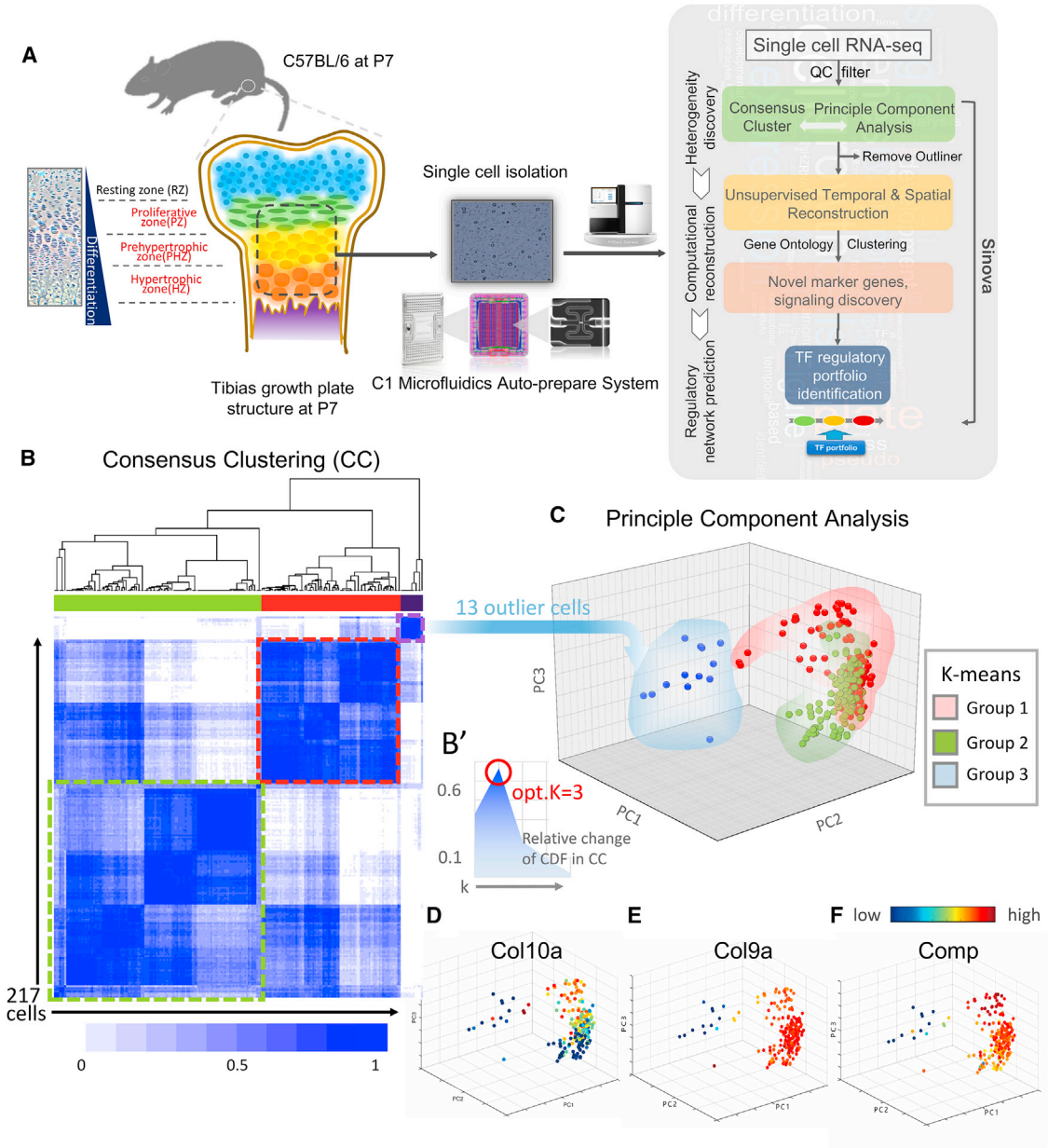


Figure 1. Single-Cell Profiling Reveals Cell Heterogeneity throughout the GP

The distal cartilage structure of the tibia at postnatal day 7 (P7) consisted of four zones that were sequentially differentiated: the RZ, the PZ, the PHZ, and the HZ. The PZ, PHZ, and HZ generally compose the GP structure.

(A) Schematic diagrams of the single-cell analysis of cell populations from the GP. The H&E staining graph and diagram illustrate four cell types comprising tibia distal cartilage free from invasion by circulating cell types in the C57BL/6 mouse at P7. The dashed-line box indicates the microdissected structure of the GP across the three zones that were used for the single-cell analysis. These microstructures were trypsinized to yield a single-cell suspension and trapped using the Fluidigm C1 auto prepare system for the downstream single-cell mRNA-seq experiment (for details, see Movie S1 protocols).

(B) CC of 217 single cells by transcriptome analysis revealed at least three populations.

(B′) Optimal cluster numbers were determined in the delta plot to reveal the relative changes in the cumulative density function (CDF) in the CC.

(C) In parallel, PCA indicated the relative similarity across the 217 cells in three-dimensional space. K-means clustering further grouped the cells into three clusters marked with different colors in the PCA plot. A dissociated cluster (blue), contained outlier cells that exactly matched a distal cluster (purple dash line box) in the CC, as indicated by the arrow (B and C).

(D–F) Matrix protein gene expression distributions of *Col10a* (D), *Col9a* (E), and *Comp* (F) in the 217 single cells. The gene expression levels are indicated by the color of the bar.

individual cells from GPs. We developed a pipeline, "Sinova" (Figure 1A, right), to integrate multiple unsupervised analyses and reconstruct GP development both temporally and spatially at the single-cell level. The reconstructed model not only confirmed prior knowledge, but also enabled the systematic discovery of genes with expression pattern across GPs. Notably, the de novo-identified surface-marker combination (CD9/CD200) predictively classified cells according to different cell stages in the GPs; this classification was effectively validated by fluorescence-activated cell sorting (FACS) sorting. By further combining gene clusters and cell-stage delineation analysis, Sinova de novo identified the effective transcriptional factor (TF) portfolios regulating GP maturation, and these were experimentally cross-validated by an EGFP-Col10a knockin reporter screening system. Our results indicate how functional genes and TFs are combined into a master regulatory network at the systems level during GP development and depict this process with more precision than ever. Our bioinformatic pipeline demonstrates the ability to resolve a complete developmental process by interpreting single-cell expression profiles in an unbiased manner and is readily transferable to other in vivo scenarios and is complementary to traditional animal genetic studies.

RESULTS

Isolation of GP Tissue from a Postnatal Day 7 Mouse and Single-Cell Profiling

We precisely isolated pure mouse postnatal GPs free from invasion by circulating cell types from the blood vessels on postnatal day 7 (P7) and completely removed the perichondrial structures (Figure S1). These microdissected structures contain the cell layers from the proliferative zone (PZ) to the hypertrophic zone (HZ) (Figure 1A, left). To further prevent the contamination by cells of hematopoietic origin (CD45 positive) during isolation, red blood cells were lysed, and CD45-negative cell populations were sorted from the trypsinized structures for downstream single-cell analysis. We next applied a Fluidigm C1 automatic microfluidic platform to randomly capture and analyze the transcriptomes of 217 single cells via single-cell mRNA-sequencing (mRNA-seq) (Figure 1A, middle). We sequenced an average of 3.8 million reads and mapped more than 9000 genes with confident expression in at least ten single-cell samples (for details on quality control, refer to Figure S2).

Cell Heterogeneity in the GP Revealed by Single-Cell Profiling

On P7, the GP consists of cells at the sequential stage of differentiation, from early proliferative chondrocytes to late-stage hypertrophic chondrocytes. To broadly characterize potential heterogeneity, we measured the similarity of the 217 single-cell samples across the expression profiles of more than 9,000 sequenced genes. Consensus clustering (CC) (Swift et al., 2004; Wahyudi et al., 2011) was used to account for the correlation between the broadly similar cells throughout development of the GP. At least three cell populations were revealed after CC, among which most of the cells belonged to two large groups, and 13 cells belonged to a smaller, less similar group (Figures 1B, 1B′, and S3).

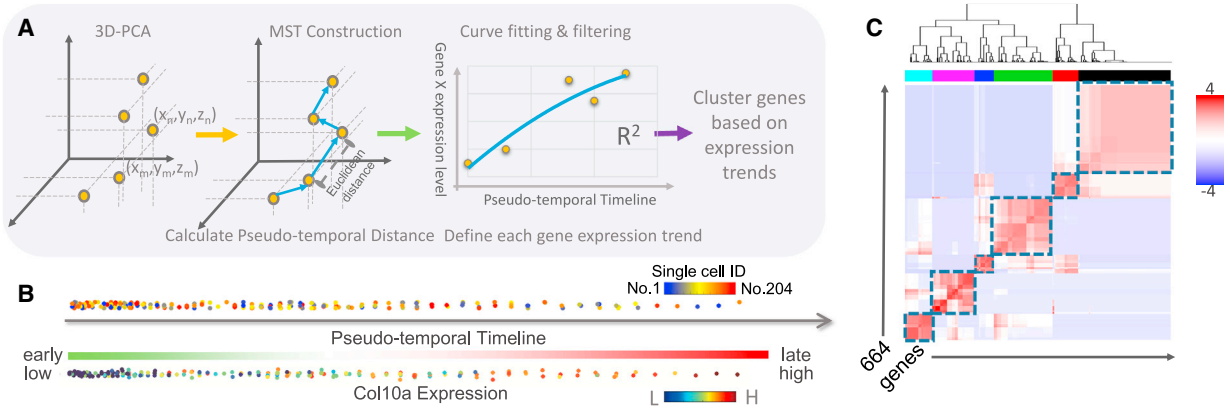
In parallel, we performed a dimensionality reduction-based principle component analysis (PCA) to reduce the high-dimension expression data across these single cells to three dimensions and capture innate single-cell characteristics. We then applied k-means clustering to group the cell population into three clusters in three-dimensional space (Figure 1C). Notably, the outlier cells grouped together (marked in blue) in the PCA, which dissociated from the bulk population and were consistent with 13 samples in the distal cluster in CC (Figures 1B and 1C). Two mixed clusters (marked in green and red) roughly corresponded to two large clusters in CC. The agreement of the results from two independent and unsupervised methods supported the rationale of the cell heterogeneity identification.

Matrix proteins are diversely expressed in cartilage tissues. For example, collagen IX (Col9a) and cartilage oligomeric matrix protein (Comp) are putatively expressed in cartilage tissues, whereas collagen X (Col10a) is specifically expressed in the hypertrophic zone of the GP cartilage (Myllyharju, 2014). We mapped the expression levels of these genes in the 217 cells projected in PCA. Both Col9a and Comp were putatively expressed 1in two large populations, whereas Col10a was differentially expressed, with significantly higher expression in the yellow cluster (Figures 1D–1F). However, none of these genes were markedly expressed in the small cluster of 13 cells. Therefore, we assumed that the 204 cells in the large clusters originated from the GP, whereas the 13 cells were unidentified outliers and were thus excluded from the downstream analysis (see Table S1 for the potential lineage identity prediction of the 13 cells). Collectively, we revealed diverse cell heterogeneity based on the transcriptome expression profiles and confidently identified broad cell populations from the GP.

Reconstruction of the Temporal Dynamics of GP Development via Pseudo-Temporal Ordering

Because GP development and many other developmental contexts are stepwise sequential processes, we developed a generalized bioinformatics pipeline, "Sinova," to systematically resolve the developmental process from single-cell mRNA-seq (mRNA-seq) data. Sequenced single cells lost their in vivo temporal and spatial information after isolation; therefore, we first attempted to re-order the 204 single cells according to their differentiation stage in the GP.

As data representing these single cells were projected to three-dimensional (3D) PCA space, the Euclidean distance in the 3D-PCA reflected the transcriptional similarities among single cells. Here, we assumed single cells with greater similarity were at mutually closer stages of the differentiation process and referred to the minimal spanning tree (MST)-based algorithm adapted from "Monocle" (Magwene et al., 2003; Trapnell et al., 2014) to re-order these single cells into a pseudo-temporal quasi timeline corresponding to the in vivo developmental process. In practice, Sinova constructed the MST based on the 3D-PCA space to computationally ligate similar cells in a stepwise manner and identified the longest path across the MST, which corresponds to the longest developmentally pseudo-temporal sequence of these single cells. Sinova calculated the Euclidean



D

Gene Clusters	GO enrichment	Representative genes
Cluster 1 — Constant upregulation	Reproductive System Development Bone Development Osteoblast Differentiation Skeletal System Development Ossification Cellular Development Pattern Specification Process	**Known participated genes:** *Mef2c,Ibsp,Smo,Npr2,Ihh,Smad3,Epha2,Gtg2, Foxa2,Bmpr2,Pcsk6,Cited2,Foxa2,Rora,Gpx4.* **Genes with most significant expression trend:** *Ptgis, Map1lc3b, Mef2c, Klf3, Rftn2, Cep85l, CD200, Frzb, Ccdc8c, Tbc1d1, Snai1, Lama4, Enpp2, Arf1, Foxa2, Atf, Lgr6, 3632451O06Rik, Npr2, Bckdk.*
Cluster 2 — Transient upregulation	Generation of Precursor Metabolites Endocrine System Development Regulation of Hormone Levels Ossification Bone Development	**Known participated genes:** *Bmp2, Klf10,Chrdl1,Gnas, Scarb1,Col10a1,Comp.* **Genes with most significant expression trend:** *Ugp2, Col10a1, Smpd3, Lmcd1, Adk, Sparc, Lgmn, Cpe, Slc13a5, Rnasek, Ntan1, Atp1b3, Shisa4, Comp, Cd24a, Eps8l2, Lxn, Tuba1a, Smim14, Myo5a*
Cluster 3 — Delayed-upregulation	Vasculature Development Insulin Receptor Signaling Pathway Response to Insulin Stimulus Response to Peptide Hormone Stimuli Enzyme Linked Receptor Signaling Biomineral Formation	**Known participated genes:** *Klf5, Akt1, Dlx3, Vhl, Id1/2/3,Mmp13, Spp1, Smad1 , Dmp1, Tnfsf11, Ubb, Srgn, Pik3r2/3, Sorbs1,Akt1.* **Genes with most significant expression trend:** *Plin2, Carhsp1, Dlx3, S1pr1, Id3, Epas1, 4933409K07Rik, Hes1, Flii, Fam13b, Tmsb4x, Rhod, Itpkb, Cst3, Pik3r2, Isg20, Lonrf3, Trpt1, Fn1, Kank1*
Cluster 4 — Tide wave	Cellular Homeostasis Cellular Chemical Homeostasis Cellular Ion Homeostasis Chemical Homeostasis Skeletal& Muscular Syst.Dev.	**Known participated genes:** *Scd2, Smad7, Foxa3, Scd1, Thbs1.* **Genes with most significant expression trend:** *Ssr1, Panx3, Ssr4, CD9, Serinc5, Chsy1, Sgms2, Scd1, Capn6, Stk38l, Nucb2, Deb1, Ppa1, Tmbim4, Sidt2, Slc2a1, Trps1, Chad, Sema3d, Scd2.*
Cluster 5 — Delayed-downregulation	Cartilage Development Sterol Biosynthetic Process Structural Molecule Activity Steroid Biosynthetic Process Skeletal System Development Tissue Development	**Known participated genes:** *Col9a1, Lect1, Sox9, Col11a1/2, Col11a1, Hmgcr, Insig1, Hmgcs1, Sc4mol, Col27a1, Spry2.* **Genes with most significant expression trend:** *Lmna, Sdc4, Spry2, Sox9, Arglu1, Foxn2, AI597468, Kdelr2, Rps25, Ywhae, Srsf2, Kdm6b, Actr2, Lrrc8a, Golt1b, Rbm26, Rrn3, Mrfap1, Hmga2, Insig1.*
Cluster 6 — Transient downregulation	RNA binding Nucleotide Binding Translation Initiation Factor Activity Translation Factor Activity Transcription Initiation	**Known participated genes:** (Null) **Genes with most significant expression trend:** *Ncl, Cycs, Srxn1, Ptges3, Nup98, Gas5, Mat2a, Hnrnpu, Gpn2, G3bp1, Csnk2a2, Rap2b, S100a10, Nufip2, Cct6a, Sox11, Hspa8, Rdh13, Eif4a1, Trabd2b*

(legend on next page)

distance as the pseudo-temporal interval. Based on this, the expression fluctuation of each gene was projected across the pseudo-temporal ordering. Sinova then fit each gene using a nonlinear function, and genes with defined pseudo-temporal dynamics were clustered (Figure 2A, also see Experimental Procedures for details).

Using the Sinova pipeline, we re-ordered the 204 single cells from the GP in an unsupervised manner across the pseudo-temporal ordering and determined the beginning and end of the timeline based on the expression levels of the known hypertrophic master marker Col10a (Figure 2B). Next, more than 600 genes that were dynamically regulated were identified with a high correlation threshold (see Experimental Procedures). Because each single cell was represented as a quasi time-series point, k-means clustering was used to computationally cluster these genes based on their high-resolution pseudo-time-series expression profiles. We applied 10,000 iterations of random k-means to optimally classify these genes into six clusters based on co-occurrence probability (Figure 2C, also refer to Experimental Procedures). Subsequently, genes with similar expression trends were clustered (k = 6) and visualized with a fitting curve after smoothing (Figure 2D, left). The top 30 genes of each cluster with highest significance were listed (Table S2), and the top five genes of each cluster were also re-validated by single-cell qPCR (Figure S5) to demonstrate the accuracy of the single-cell RNA-sequencing (RNA-seq) data.

To determine the accuracy of the unsupervised algorithm, we inspected the gene ontology (GO) profile enrichment among each cluster of genes (Figure 2D, right). Notably, hypertrophic differentiation-related GO terms were accumulated in the constantly upregulated and transiently upregulated clusters (clusters 1 and 2), including TFs that were previously identified in hypertrophic regulation, well-known matrix proteins, and Ihh and Bmp signaling, which is activated in the prehypertrophic and hypertrophic zones (Minina et al., 2001). Growth hormone phosphatidylinositol signaling and angiogenic genes were closely associated with the delayed upregulation cluster (cluster 3), supporting the assumption of late hypertrophy and later endochondral bone formation in this cluster (Gerber and Ferrara, 2000; Ulici et al., 2008). The expression of the delayed downregulation cluster (cluster 5), which enriched steroid hormone signaling and non-hypertrophic chondrocyte-associated genes, was maintained in the early stage and decreased in the late stage. Transcription and translational activity-associated GO terms were downregulated across the pseudo-timeline in the transient downregulation cluster (cluster 6), correlating with the cessation of the cell cycle in proliferative chondrocytes upon entering hypertrophy (Yang et al., 2014a; Yeung Tsang et al., 2014). Collectively, the reconstruction of pseudo-temporal development successfully recaptured most of the findings of previous genetic model studies without referring to prior knowledge.

Identification of Molecular Cascades Regulating GP through Proximo-Distal Axis Reconstruction

The morphology of the GP resembles a cylinder, and cells corresponding to stepwise developmental stages are arranged in unidirectional "waterfall" structures along the proximo-distal axis (PDA) (Li et al., 2015). Upon re-inspecting the 3D-PCA and pseudo-temporal ordering, which transcriptionally reflect the developmental relationships, we sought to re-map the single cells into a cylinder space and view the gene expression distribution along the longitudinal axis, resembling the PDA of the GP in vivo (Figure 3A, also refer to Experimental Procedures). The distance from the center of the cylinder does not actually reflect the physical location in the GP. To demonstrate the effectiveness of our model, we microdissected the GP structure into three layers corresponding to the distal (PZ), middle (PHZ), and proximal (HZ) positions and compared the expression levels of selected genes in each mechanically dissected layer sample with the expression signature of single cells in the spatial trisection of the pseudo-PDA model (Figure 3B).

Among the genes that shared a significant spatial expression pattern, we classified them into (1) matrix protein and regulators; (2) TFs and co-factors; (3) signaling and related regulators; and (4) metabolic regulators, and selectively presented and explained them below (Figures 3C–3F and S4 for additional genes and corresponding pseudo-PDA patterns).

(1) Matrix protein and regulators (Figure 3C): as expected, the expression of the well-recognized hypertrophic differentiation marker Col10a was barely detectable in the distal proliferative zone but began to increase across the PDA axis. The terminal hypertrophic markers Mmp13, Spp1, and Dmp1 were only expressed on the bottom of the proximal layer, supporting previous findings (Komori, 2010; Yeung Tsang et al., 2014). Collagen IX (Col9) (Zhang et al., 2003), which was previously regarded as a putative marker across the GP, exhibited slightly higher expression in the middle layers compared to the

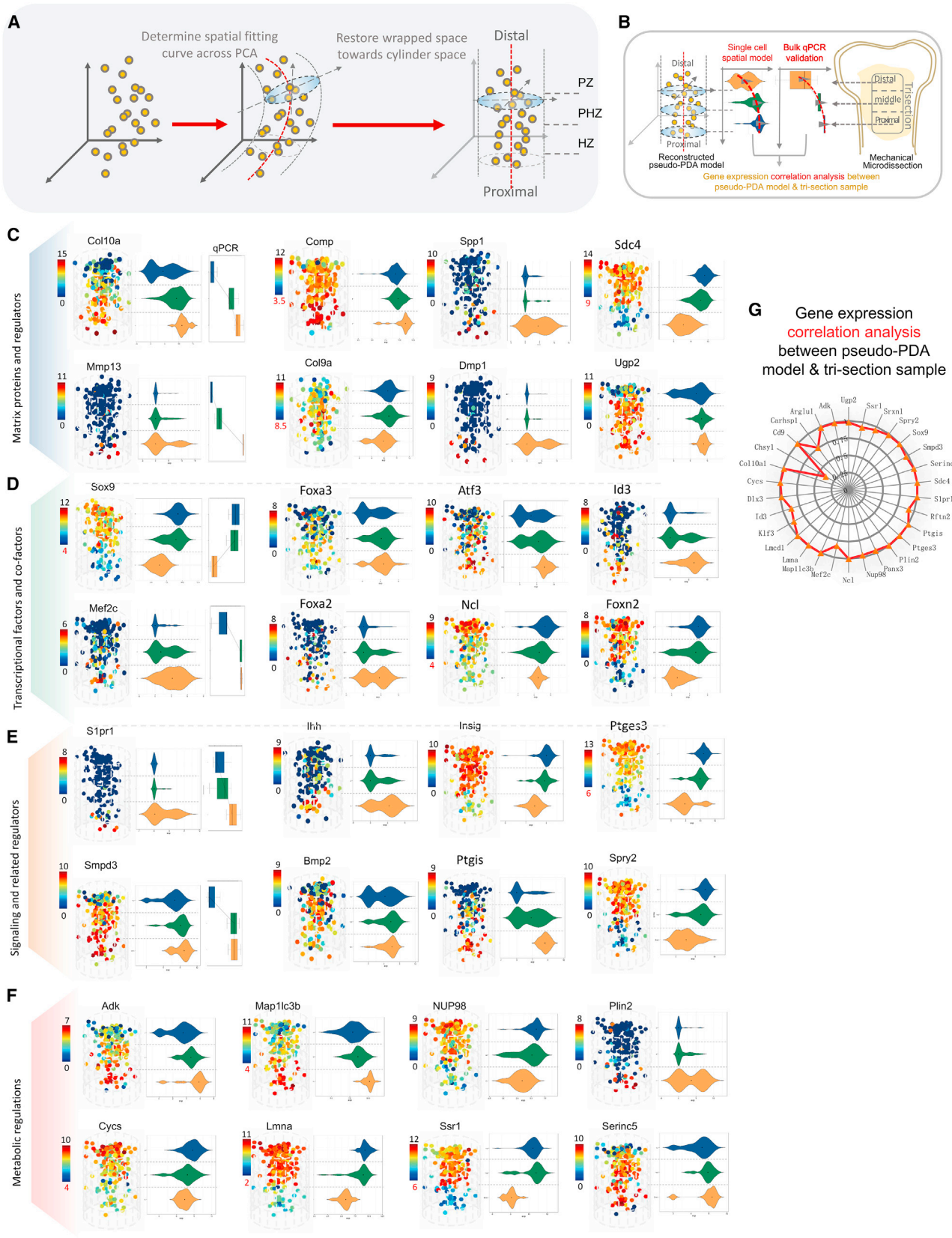
Figure 2. Reconstruction of the Temporal Dynamics of GP Development

(A) Workflow of the temporal dynamics reconstruction via pseudo-temporal ordering.

(B) Upper: cells were projected across the pseudo-temporal quasi-timeline based on the ordering and were color coded based on the original cell ID on the microfluidic single-cell trapping system; mixed ID sequences across the pseudo-temporal quasi-timeline indicate the unbiased single-cell trapping process. Lower: the expression levels of the hypertrophic marker Col10a were mapped across the single-cell points and used to determine the beginning and end of the timeline.

(C) Iterations of random k-means clustering identified six optimal clusters among the 664 genes with significant pseudo-temporal expression trends.

(D) Six gene clusters with similar expression trends and functional gene identification. Left: the genes were clustered in six groups based on relative expression with the indicated trends. All of the genes in each cluster were revealed in the contour map, and the clustering centers of each expression trend are shown in the red curves. Middle: the most relevant GO terms associated with the genes in each cluster are listed. Enrichment significance scores for each GO term are shown as histograms (dark blue). Right: previously reported or experimentally validated genes that participate in GP or cartilage development in each cluster are listed under "known participating genes"; genes with the most significant expression trends (top 20 sorting order based on multiple testing) in each cluster are listed under "genes with the most significant expression trends," among which genes that overlapped with the "known participating genes" are marked in red.



(legend on next page)

distal and proximal layers. Ugp2, which encodes UDP-glucose pyrophosphorylase, was highly expressed in the PHZ and HZ and has been proposed to regulate hyaluronan synthesis in hypertrophic cartilage (Magee et al., 2001). These examples demonstrate a more precise definition of spatial gene expression compared to the current understanding obtained using traditional methods.

(2) TFs and co-factors (Figure 3D): consistent with the results of previous studies, the expression of the early chondrogenesis master TF Sox9 was lower in the proximal HZ, whereas Mef2c was induced in several single cells in the PHZ and HZ, presumably to regulate hypertrophic differentiation (Arnold et al., 2007; Yeung Tsang et al., 2014). Interestingly, although Foxa2/3 were recently identified as a critical factor in hypertrophic chondrocyte maturation in genetic mouse models (Ionescu et al., 2012), our pseudo-temporal-pattern indicated that Foxa3 expression was induced early in the PZ-PHZ, whereas Foxa2 expression was restricted mainly to the PHZ and HZ. Notably, Ncl, which encodes nucleolin, was highly expressed in the PZ compared to the PHZ and HZ, consistent with the function of nucleolin in repressing the transactivation of Mmp13 (Figure 3C). We also determined that several TFs, including Atf3, Id3, and Foxn2 (Koivisto et al., 2014; Zhao et al., 2011), which participate primarily in hypertrophic cardiomyopathy, shared a clear spatial expression pseudo-temporal-pattern signature, indicating potential roles in directing hypertrophic chondrocyte differentiation.

(3) Signaling and related regulators (Figure 3E): we successfully recapitulated the spatial information of the important Indian hedgehog signaling molecule gene Ihh (Wongdee et al., 2013), which is secreted primarily by hypertrophic cells in the HZ but not by terminal hypertrophic chondrocytes. In contrast, Bmp2 signaling appears to be activated early, as in the PHZ. Interestingly, Smpd3, which is highly induced from PHZ to HZ, encodes a phosphodiesterase that is involved in ceramide and subsequent sphingosine-1-phosphate (S1p) biosynthesis (Van Brocklyn and Williams, 2012). The S1p signal may be subsequently received downstream by S1P receptor (S1pr1)-expressing cells among terminal hypertrophic chondrocytes, revealing a potential role in cell survival regulation via the sphingolipid biostat in the HZ. We also identified the high expression of Spry2, which is responsible for potential angiogenic inhibition (Wietecha et al., 2011), in the PZ and PHZ but not the HZ. These findings support a mechanism in which spatial blood vessel induction is restricted primarily to the HZ and is responsible for later blood vessel invasion and subsequent secondary ossification center formation.

(4) Metabolic regulators (Figure 3F): among genes with the most obvious shared pseudo-PDA signatures, we identified groups that were closely associated with various metabolic regulatory pathways but with no clear function in GP development. For instance, Nup98, Cycs, Lmna, and Ssr1 were predominantly expressed in the PZ/PHZ and reduced in the HZ, which is consistent with the significantly reduced metabolic activity in late-stage hypertrophic chondrocytes (Terpstra et al., 2003). Map1lc3b, a gene that encodes a protein with a predicted function in autophagy (Li et al., 2013), may be associated with death/survival regulation in the HZ. These genes merit further exploration in systematic studies of metabolic regulation during GP development.

Finally, we selectively compared expression signatures in the computationally reconstructed pseudo-PDA model with qPCR-tested relative expression level in mechanical trisection samples for 30 dynamically regulated genes across the pseudo-PDA axis (as explained in Figure 3B). The correlation analysis revealed identical trends in the selected genes between the reconstructed PDA patterns and the mechanical microdissected bulk sample (Figure 3G), validating the rationality of this PDA reconstruction.

Identification and Validation of Surface-Marker Combinations to Sort Cell Subpopulations from GPs

The separation and enrichment of cell subpopulation would facilitate the in vitro study system of cartilage development and skeletal disorders. Surface proteins are differentially expressed among different types of cells, which could be exploited to sort a specific cell type from a mixed population through FACS. Here, the surface markers CD200, CD24, and CD9 were dynamically expressed across the reconstructed pseudo-PDA pattern (Figures 4A–4C). Notably, CD9 was more highly expressed in the middle PHZ layer, whereas CD200 was predominantly expressed in the HZ layer. In contrast, CD24 was universally expressed across the GP. We next performed in situ immunostaining and flow cytometry to sort single-cell suspensions from the GP to selectively validate CD200 and CD24a. Consistent with the reconstructed model, in the fluorescence-activated cell sorting (FACS) analysis, more than 85% of the cells from the GP were CD24a-positive, whereas approximately 30% of cells were CD200 positive (Figures 4B and 4C, middle). In situ

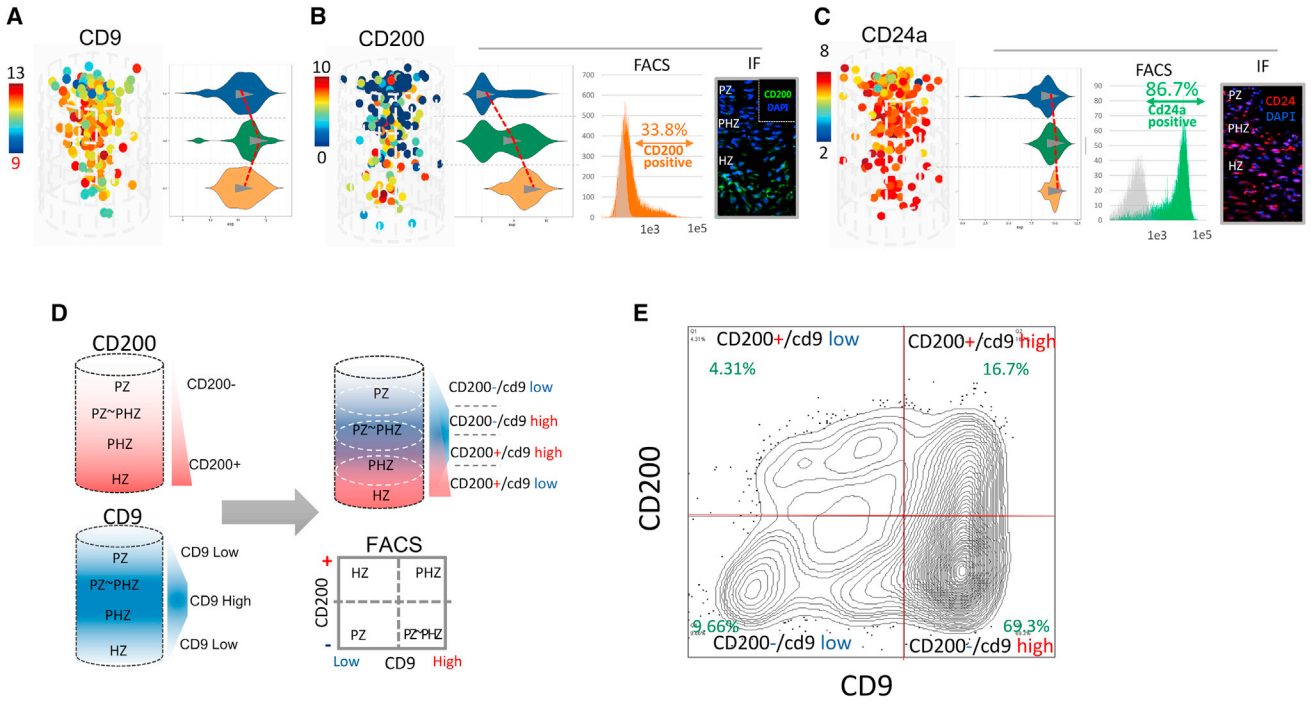
Figure 3. Proximo-Distal Axis Reconstruction and Identification of Regulators in the GP

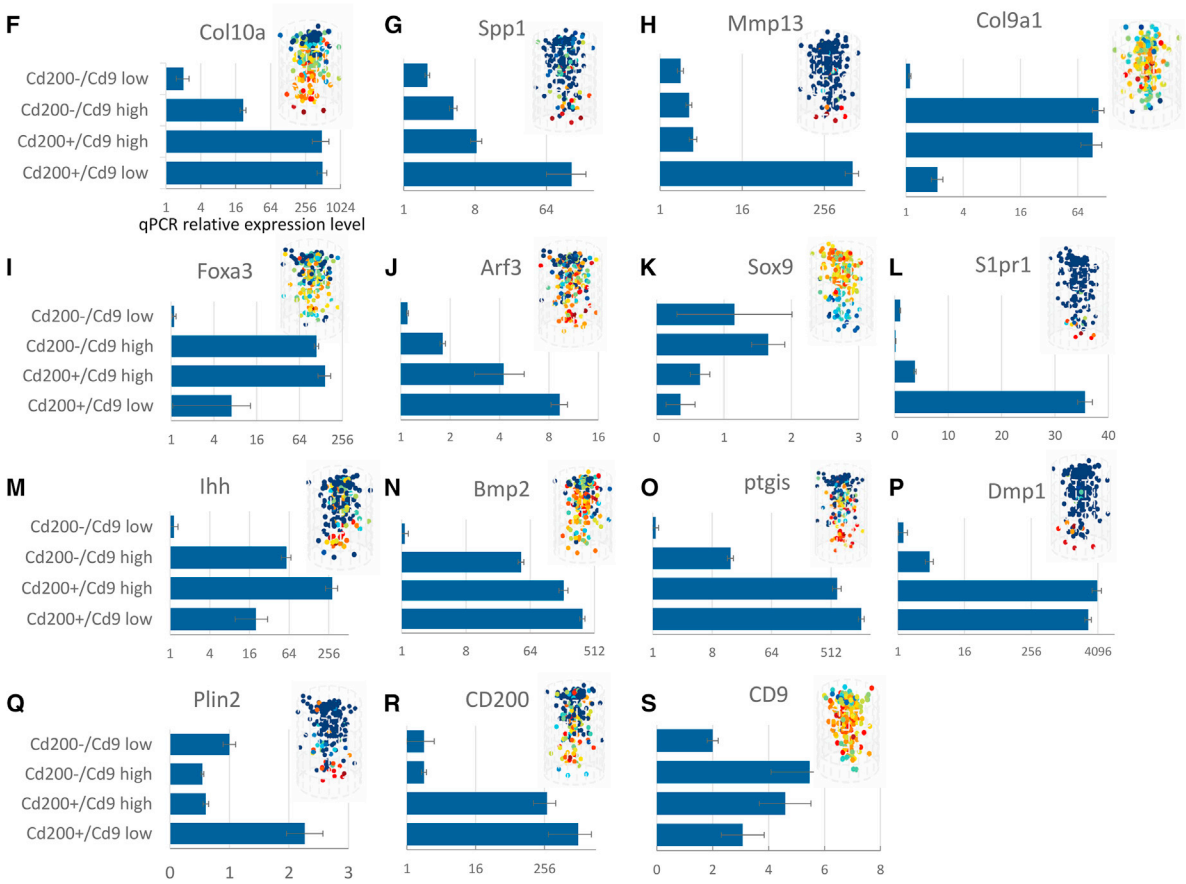
(A) Workflow of the PDA reconstruction of the GP based on single-cell RNA-seq (see Experimental Procedures for details).

(B) Diagram of correlation analysis compared the fitting trends of the selected genes between the mechanically dissected three-layer bulk samples and the spatial trisection of the reconstructed pseudo-PDA pattern.

(C–F) Spatial distribution of the selected genes in the pseudo-PDA pattern of the GP. Heatmaps were used to reveal the RPKM (reads per kilo-bases per million reads) of the gene in each single cell at log scale. The colored bars are indicated on the left side of each plot. The violin plot on the left reveals each gene's expression level distribution in the trisection of the pseudo-PDA pattern. qPCR of gene expression from the mechanically dissected bulk sample is displayed where applicable. Selected genes with dynamic expression distributions were classified as matrix proteins and regulators (C), TFs and co-factors (D), signaling and related regulators (E), and metabolic regulators (F).

(G) Correlation analysis of 30 genes shown by radar map indicated identical trends between the reconstructed pseudo-PDA model and the mechanical microdissected bulk sample. PCC, Pearson correlation coefficient.





(legend on next page)

immunofluorescence (IF) staining of the GP sections also validated the spatial distribution of CD200 and CD24a (Figures 4B and 4C, right), which were well consistent with the reconstructed pseudo-PDA pattern, as well as a previous report (Belluoccio et al., 2010).

We next sought to ask whether these "in silico"-discovered spatial patterns of surface markers from our model could be practically used to sort functional subpopulations in GPs. Theoretically, four CD200/CD9 signatures, namely, CD200-/CD9 low, $CD200^{-}$/CD9 high, $CD200^{+}$/CD9 high, and $CD200^{+}$/CD9 low, can approximately distinguish among the PZ, PZ-PHZ, PHZ, and HZ cells, respectively, according to our computationally reconstructed PDA model (Figure 4D). We analyzed the expression of the marker genes in the four subpopulations sorted by FACS (Figure 4E). Notably, the expression levels of three marker genes and the relative expression levels of CD9 and CD200 in the four FACS-sorted populations correlated highly with the pseudo-PDA patterns (Figures 4F–4S, compared to each pseudo-PDA pattern in the left corner). These cross-validated results further verified our model, and the single-cell spatial reconstruction enabled the successful identification of surface-marker combinations to effectively and precisely sort single cells at different biological stages across the GP.

Reconstruction of the TF Regulatory Network during GP Development

TFs and co-factors are critical participants in the regulation of GP development. Because of the difficulties in establishing in vitro differentiation systems that precisely reflect real physiological conditions and the low-throughput and tedious methods that are required to establish mouse genetic models, our understanding of the precise transcriptional regulation networks remains incomplete. To systematically identify the TF portfolios (combinations) that contribute to GP development, we screened genes with transcriptional activity among genes that were dynamically regulated within a pseudo-temporal timeline. A total of 36 TFs were identified from this dataset. Among these TFs, several previously validated regulators from different experimental contexts were successfully recapitulated (Table S3). To further predict the functionally important TF portfolio in GP development, we reasoned that a TF with more significance would share high co-expression correlations with a larger number of genes that were dynamically regulated during development. The binary expression patterns (Figure 5A) of genes from single-cell profiling suggested the presence of discrete "on" and "off" states among individual cells, and the distribution of the "on" part was regarded as a continuous fraction of the expression profile. Consequently, to comprehensively evaluate the association of each given gene pair, such as Foxa3/Col10a (Figure 5B), both characters within a gene pair should be considered as co-existing binary and continuous expression patterns. We thus methodologically calculated the connectivity score by integrating both the odds ratio (OR) (for binary relationship) and the Spearman rank correlation (for continuous relationships) (Pina et al., 2015), which comprehensively reflected the co-expression patterns among the TFs and genes across the pseudo-temporal timeline (Figure 5C; Supplemental Experimental Procedures).

To reduce network complexity, we first globally and comprehensively interrogated the co-expression patterns among 36 TFs and 60 of the most significantly regulated genes (top ten genes each from six clusters) (Figure 5D). The network revealed a higher positive connectivity between the TFs and the transiently upregulated cluster (cluster 2). In contrast, two downregulated clusters (cluster 5 and cluster 6) shared a significantly higher proportion of negative associations with the TFs, revealing global regulation by repression (Figure 5D).

The global connectivity score for each TF indicated the regulatory potential for the genes across the entire developmental process. By evaluating both the score and the TF pseudo-temporal expression pattern, we statistically predicted the participation of both globally upregulated positive and downregulated negative TF regulators during GP development (Figure 5E). Notably, several TFs that were previously shown genetically to be critical for GP development, including Mef2c, Atf3, Foxa2/3, and Sox9, were successfully re-captured as having a high regulatory potential. We also systematically predicted several previously unknown TFs, such as the Klf family and Fos family, with relatively significant connectivity scores throughout development. These TFs represent subjects for future genetic model analyses (see Figure 5E and Table S4 for compete list).

The single-cell expression data provided a high-resolution developmental profile that permitted the TF regulatory network to be recaptured across sub-divided developmental stages. We next divided the pseudo-temporal quasi timeline into three sections by unsupervised k-means clustering (corresponding to the predefined distal, middle, and proximal stages) and sought to precisely infer each TF contribution in two transition intervals by our algorithm (Figure 5F). In this end, Sinova characterized the integrative co-expression patterns among each TF and the remaining dynamically regulated genes among single cells attributed to each of the transition intervals (Supplemental Experimental Procedures). Notably, although TF portfolios for two transitions highly overlapped in global analysis (Figure 5E), there

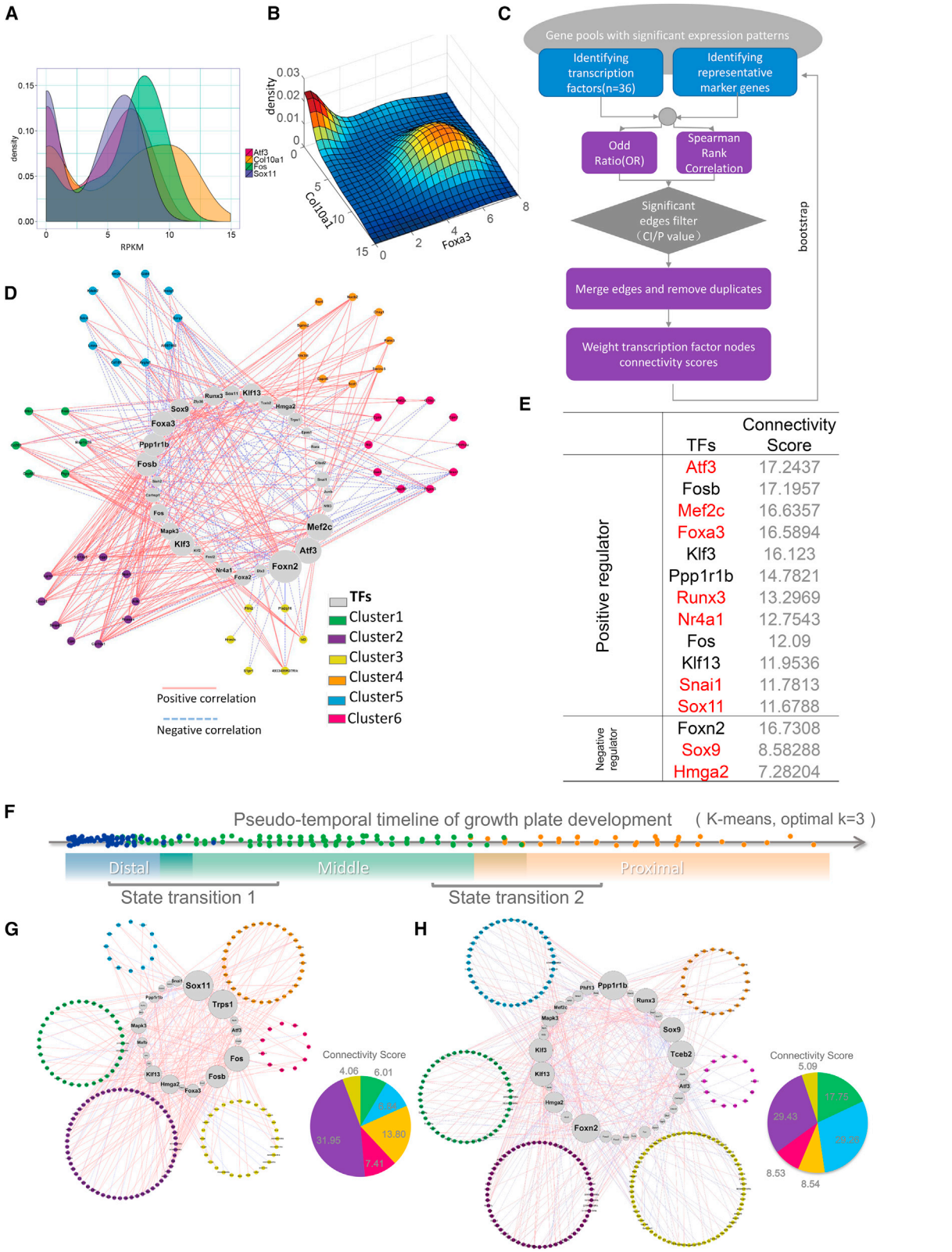
Figure 4. De Novo Identification and Validation of Surface-Marker Combinations to Sort Subpopulations of Cells from the GP

(A–C) Spatial distribution of the surface markers CD9, CD200, and CD24a in the pseudo-temporal pattern of the GP. Heatmaps were used to reveal the RPKM of the gene in each single cell at log scale. The colored bars are indicated on the left side of each plot. The violin plot on the left reveals each gene's expression level distribution in the trisection of the pseudo-pattern. The expression distributions of the surface markers CD200 (B) and CD24a (C) were experimentally validated via immuno-fluorescence-activated cell sorting (FACS) and in situ IF staining of GP sections from mouse tibias at P7.

(D) Schematic diagram illustrating the CD200/CD9 FACS sorting strategy to classify the four spatial regions, PZ ($CD200^{-}$/CD9 low), PZ-PHZ ($CD200^{-}$/CD9 high), PHZ ($CD200^{+}$/CD9 high), and HZ ($CD200^{+}$/CD9 low).

(E) FACS analysis of CD200/CD24 surface protein expression in single-cell suspensions from the GP. Four gating regions are indicated.

(F–S) The relative gene expression across the four CD200/CD9 sorting regions was analyzed by qPCR. The computationally reconstructed spatial gene expression signatures of each gene tested are shown adjacent to the bar charts.



(legend on next page)

was obviously a different TF contribution priority between the first- and second-state transition process, suggesting the precise and sequential roles of these TFs at the sub-stage level (Figures 5G, 5H, and S6 [box for TF portfolio 1/2]; Table S4 for a full list of TFs and respective connectivity scores). In the first-state transition, the revealed associations between TFs and genes in the transient-upregulation cluster (cluster 2) were established for a majority, as indicated by the sum of connectivity scores (Figure 5G, and inserted pie chart). However, during the second-state transition stage, the TFs were more evenly associated with each cluster, and there was a significantly increased association with the delayed downregulation cluster (cluster 5) (Figure 5H). Additionally, while positive regulatory associations predominated across the first-state transition, more negative regulatory associations were found in the terminal differentiation of the GP, increasing the complexity of the gene regulatory network (GRN).

Finally, we interrogated the GO enrichment for potentially regulated genes that were significantly associated with TFs in the two-state transition processes. For the first-state transition, GO terms were associated with ion exchange, signaling molecules, and granule transport or secretion; in contrast, extracellular matrix secretion, ossification, and metabolic activity were enriched in the second-state transition (Table S4). Based on these de novo findings, we systematically reconstructed a GP development model from an unsupervised single cell analysis. In this model (Figure S6), the groups of TFs listed in portfolio 1 potentially activated the ion exchange and expression of signaling molecules, including Bmp2 and Ihh, during the transition from distal PZ to PHZ. The secretion of these molecules, caused by enhanced Golgi activity, regulates the ossification process downstream in the HZ. The second wave of regulation from PHZ toward terminal HZ was driven by either the enhanced expression of TFs, including Prr1rqb, RunX3, and Mef2c, or the repression of TFs, such as Foxn2 and Sox9. The TFs in this portfolio reduce the metabolic activity of cells and induce the secretion of lipid derivatives among the cells in the HZ. Additionally, these TFs strongly activate ECM production, corresponding to the terminal ossification process.

Identification and Validation of Effective TFs Portfolio that Facilitate GP Maturation

Because it is difficult to recapitulate GP hypertrophic maturation in in vitro systems, current approaches mainly rely on tissue-engineered explants in combination with chemical pre-induction. Although several TFs and co-factors have been revealed by independent genetic perturbation experiments, effective TF portfolios that direct GP maturation have not been fully revealed. Among the 36 TFs that were identified from a single-cell analysis, 33/36 TFs were upregulated during the pseudo-temporal timeline. We next sought to determine whether these positive TFs with high connectivity scores predicted by the Sinova algorithm could be effective contributors to GP maturation. Because Col10a is a standard marker for GP maturation, as corroborated by previous studies, as well as by our single cell reconstructed pattern, we thus generated an EGFP reporter knockin within the Col10a gene coding sequence in the mouse chondrocyte progenitor cell line ATDC5 (Figure 6A). This engineered cell line, named ATDC5 EGFP-Col10a, did not express Col10a, as indicated by EGFP florescence at the basal level. We next generated a lentiviral expression vector pool comprising 33 TFs to infect ATDC5 EGFP-Col10a, along with puromycin resistance screening (Figure 6B). Notably, 28.3% of the total cells were EGFP positive in groups that were infected with vectors from the expression pool compared to the basal control, indicating the functionality of 33 positive TFs identified from the single-cell dataset. Subsequently, cell populations with both EGFP-high and EGFP-low profiles were sorted, and their relative TF accumulation was compared (Figure 6C). Rationally, one would expect that TFs with higher functional priority toward GP maturation would share higher accumulation in the EGFP-high population; thus, we screened out 12 TFs with three times higher accumulation in the EGFP-high population. A total of eight of 12 TFs overlapped with the corresponding top 12 predicted TFs with the highest connectivity scores, further illustrating the reliability of the Sinova TF regulatory prediction algorithm (Figure 6D).

To further demonstrate the effectiveness of the TF portfolio, we transfected a vector pool comprising eight overlapped TFs and the eight TF expression vectors individually into the freshly isolated mouse primary chondrocyte. Interestingly, the transfection of the pool of eight TFs significantly elevated the Col10a gene expression, even with a transfection efficiency of ~30%, but eight TF single transfections with equal dose only moderately regulated Col10a expression, highlighting the collective effects of the identified TFs (Figure 6E). Taken together, these results demonstrate that a portfolio of eight TFs, as identified using a combination of bioinformatics prediction and pooled screening, can effectively facilitate GP maturation.

Figure 5. Reconstruction of the TF Regulatory Network during GP Development

(A and B) Representative binary expression pattern for Atf3, Col10a1, Fos, and Sox11 (A), and an example of the density distribution between Foxa3 and Col10a1 reflecting the both on/off and continuous nature of gene-gene association (B).

(C) Workflow of the TF regulatory network prediction algorithm adapted to single-cell developmental reconstruction.

(D) Global TF regulatory network across the GP developmental process, inferred by the OR and Spearman rank correlation. Gray nodes, TFs; colored nodes represent the respective gene expression clusters identified in Figure 2, as indicated in the legend. Solid red lines, positive associations; dashed blue lines, negative associations. The node size of the TFs is proportional to the relative connectivity in each network.

(E) Representative predicted upregulated (positive) and downregulated (negative) TFs are listed based on the connectivity score ranks. Previously genetically validated critical regulators during GP development are marked in red. Refer to Table S2 for the full list.

(F) De novo identification of three sub-stages through k-means clustering across the pseudo-temporal quasi timeline of GP development. Distal stage, blue; middle stage, green; proximal stage, orange. Two-state transition processes are indicated.

(G and H) TF regulatory networks among state transition 1 and state transition 2 are shown. The detailed formatting and legend are consistent with (D). The sum of connectivity scores in each cluster is presented in the pie chart within each state transition process.

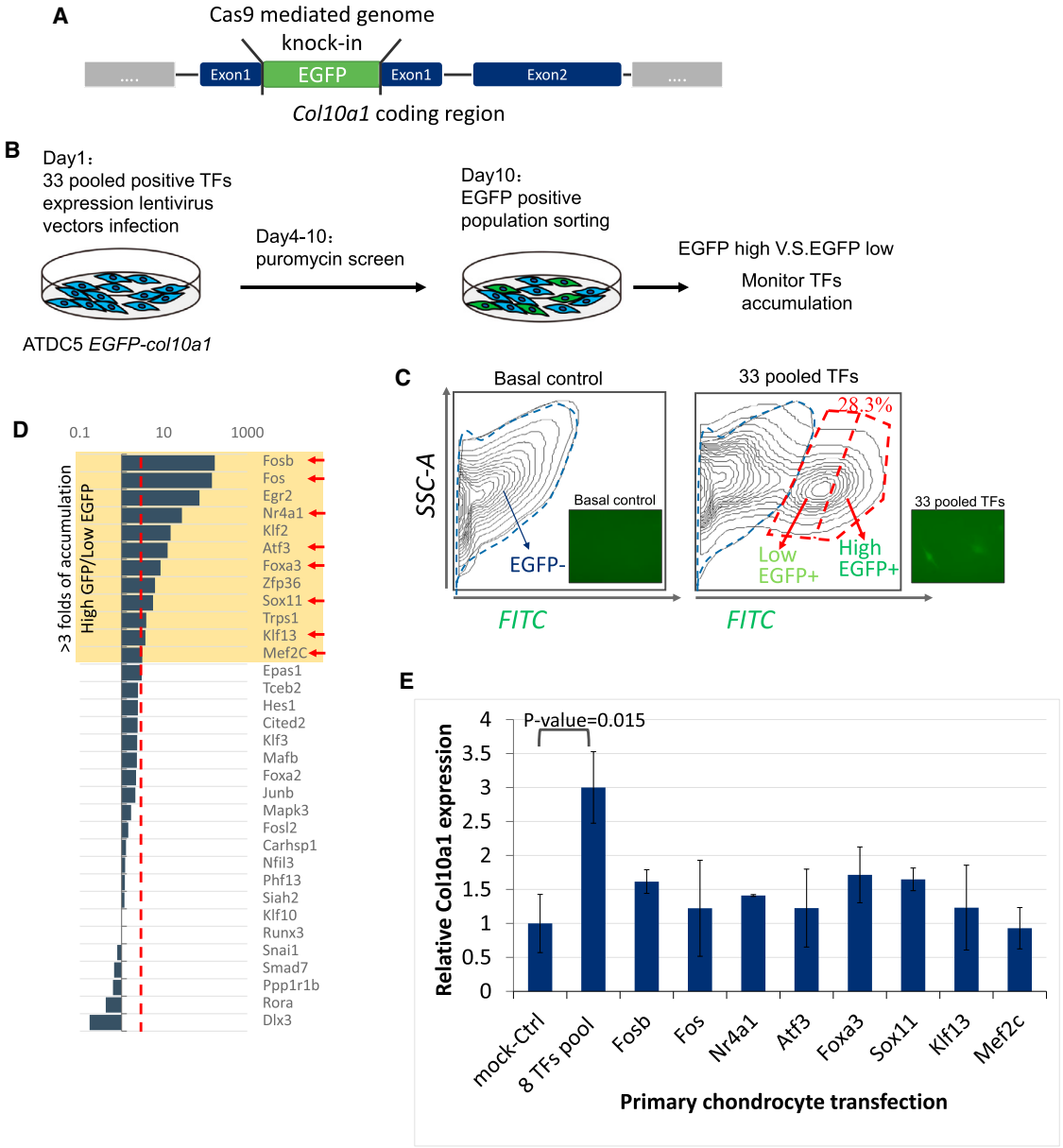


Figure 6. Identification of Effective TFs Portfolio Direct GP Maturation

(A) Diagram of EGFP knockin at the first exon of Col10a coding region in ATDC5 chondrocyte progenitor cell line.

(B) Diagram of TFs pool screening strategy based on lentivirus hybrid infection into the ATDC5 EGFP-Col10a cell line.

(C) Flow cytometry analysis revealed EGFP subpopulation activation after 33 pooled TFs infection. Activated gating in contrast to the basal control is indicated in the red dashed line and was further divided into low-EGFP$^+$ and high-EGFP$^+$ subpopulations in sorting process. Representative florescent views of culture ATDC5 EGFP-Col10a are shown on the right side in each experimental condition.

(D) TF accumulation was detected by qPCR between high GFP$^+$ and low GFP$^+$ subpopulations. A total of 12 TFs with a greater than 3-fold enrichment were highlighted with a yellow box, and the TFs that overlapped with the top 12 predicted TFs with the highest connectivity score are indicated with red narrows.

(E) qPCR of Col10a marker activation upon indicated TF/TFs pool transfection in primary chondrocyte cells. The total transfection dose was equal among each experimental group. p value and error bars were indicated.

DISCUSSION

Recent developments in single-cell gene expression profiling permit alterations in gene expression to be tracked among single cells directly isolated from tissues, thus enabling a systematic elucidation of the regulatory machinery under in vivo conditions. Because the GP is a well-arranged structure with unidirectional differentiation and a cylindrical shape, we assumed that cells

that were isolated from the GP structure would be at different differentiation stages across the GP under physiological conditions (Li et al., 2015).

Here, we developed a pipeline called "Sinova," which uses an unbiased bioinformatics approach to determine the transcriptional heterogeneity among 217 single cells. The computational model allowed us to reorder single cells throughout development based on their transcriptome similarities in an unsupervised manner. We then spatially reconstructed the GP, and a modeled 3D pseudo-PDA pattern enabled the discovery and analysis of gene makers across the GP. For example, we clearly in silico recaptured the spatial pattern of one of the important known TFs contributing to hypertrophy, Mef2C, which is well correlated with a rigorous and elegant animal genetic study (Arnold et al., 2007). Notably, both temporally and spatially resolved "high-resolution" maps of the GP not only enabled a systematic analysis of the "in situ" regulation mechanism but also de novo identified surface-marker combinations that enabled the effective sorting of cell subpopulations at four stages in GPs. To identify effective TF combinations that direct a given development, Sinova further applied a hybrid association algorithm that took the advantage of single-cell expression profiles to estimate the regulatory potential of each TF along a pseudo-temporal timeline that represents a model of the GP. Our estimation results highly overlapped with the experimental screening outcome, which conceptually justified the rationality of the algorithm.

Sinova shared at least several key differences from the existing single-cell analysis packages. First, Sinova used two unsupervised statistic methods to identify a reliable cell population in continuous manner and remove outliers. Second, unlike "Monocle," which requires multiple prior information (Trapnell et al., 2014), Sinova requires only the gene expression matrix as input, which is more general for in vivo development processes. Third, Sinova could further exploit the gene correlation relationship to de novo identify the potential TFs that were biologically meaningful to the development. On the other hand, Sinova could partially reconstruct the tissue based on the similarity among single cells at the global transcriptome level, while "Seurat" reconstructed the tissue based on the in situ landmark gene hybridization database through a machine-learning concept (Satija et al., 2015).

In conclusion, we developed the Sinova pipeline using a combination of computational models to successfully reconstruct GP development and potential spatial distributions based on the transcriptomes of 217 single cells, and we systematically identified functional genes and proposed regulatory networks that have not been previously reported in traditional experimental systems. Our model enhances the systematic understanding of the complexity of GP development. In addition, our bioinformatic pipeline is also applicable to other tissues and organ development processes in diverse contexts.

EXPERIMENTAL PROCEDURES

Mice and GP Isolation

C57BL/6 wild-type mice at postnatal day 7 were used for GP isolation under the approval of the Institutional Animals Care and Use Committee of Tsinghua University. For detailed procedures, refer to the Supplemental Experimental Procedures and Movie S1.

Sinova Pipeline and Algorithms

Pipeline analysis started from the expression matrix, with each column representing a single cell and each raw representing a gene. The analysis procedures were based on three sequential modules: (1) unsupervised cell population clustering; (2) pseudo-temporal ordering and PDA reconstruction; and (3) TF regulatory network prediction and analysis.

Statistical Analyses

For qPCR results, the data were expressed as the mean, and the SD was expressed as an error bar. Statistical comparisons were made between multiple groups based on Student's t test, and p values were calculated with Microsoft Excel. Advanced bioinformatic statistical analyses are described in Supplemental Experimental Procedures.

Detailed additional methods are described in Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the single cell sequencing raw data reported in this paper is GEO: GSE76157.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, six figures, four tables, and one movie and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.04.043.

A video abstract is available at http://dx.doi.org/10.1016/j.celrep.2016.04.043#mmc8.

AUTHOR CONTRIBUTIONS

J. Li, Q.W., and G.T. initiated this project. J. Li, S.Z., J.F., and Y.L. performed the single-cell RNA-seq experiments. J. Li, H.L., and J. Lang performed the data analysis. J. Li, R.W., Z.Z., and K.Q. performed the animal experiments. J. Li, S.Z., and R.W. performed all of the other experiments. J. Li, H.L., and Q.W. prepared the manuscript.

ACKNOWLEDGMENTS

We thank Professor Weidong Tian, Hao Han, Zhi Lu, and Jianzhong Xi for comments on the manuscript. We appreciate Professor Xiao Yang and Dr. Guan Yang for their assistance in the experimental design. We appreciate Hao Ding and Le Xu for their assistance with the MST construction. We thank Dr. Ivan Hajnal for a critical reading and correction of the manuscript. We thank Bionova (Beijing) Biotech Co., Ltd. for the assistance with the bioinformatics analysis and software packaging. We thank Suzhou GenoArray Co., Ltd. for the assistance on the TFs screening experiments. This work was supported by the Natural Sciences Foundation of China (grant no. 31470933, 31170940), the National High Technology Research and Development (863 Program) (grant no. 2012AA020503, 2013AA020301, and 2012AA02A702), the 973 Basic Research Fund (grant no. 2012CB725200), and the Tsinghua University Initiative Scientific Research Program (no. 20131089199).

Received: December 15, 2015
Revised: February 5, 2016
Accepted: April 6, 2016
Published: May 5, 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, M.A., Kim, Y., Czubryt, M.P., Phan, D., McAnally, J., Qi, X., Shelton, J.M., Richardson, J.A., Bassel-Duby, R., and Olson, E.N. (2007). MEF2C

transcription factor controls chondrocyte hypertrophy and bone development. Dev. Cell *12*, 377–389.

Belluoccio, D., Etich, J., Rosenbaum, S., Frie, C., Grskovic, I., Stermann, J., Ehlen, H., Vogel, S., Zaucke, F., von der Mark, K., et al. (2010). Sorting of GP chondrocytes allows the isolation and characterization of cells of a defined differentiation status. J. Bone Miner. Res. *25*, 1267–1281.

Durruthy-Durruthy, R., Gottlieb, A., Hartman, B.H., Waldhaus, J., Laske, R.D., Altman, R., and Heller, S. (2014). Reconstruction of the mouse otocyst and early neuroblast lineage at single-cell resolution. Cell *157*, 964–978.

Emons, J., Chagin, A.S., Sävendahl, L., Karperien, M., and Wit, J.M. (2011). Mechanisms of GP maturation and epiphyseal fusion. Horm. Res. Paediatr. *75*, 383–391.

Gerber, H.P., and Ferrara, N. (2000). Angiogenesis and bone growth. Trends Cardiovasc. Med. *10*, 223–228.

Ionescu, A., Kozhemyakina, E., Nicolae, C., Kaestner, K.H., Olsen, B.R., and Lassar, A.B. (2012). FoxA family members are crucial regulators of the hypertrophic chondrocyte differentiation program. Dev. Cell *22*, 927–939.

Koivisto, E., Jurado Acosta, A., Moilanen, A.M., Tokola, H., Aro, J., Pennanen, H., Säkkinen, H., Kaikkonen, L., Ruskoaho, H., and Rysä, J. (2014). Characterization of the regulatory mechanisms of activating transcription factor 3 by hypertrophic stimuli in rat cardiomyocytes. PLoS ONE *9*, e105168.

Komori, T. (2010). Regulation of bone development and extracellular matrix protein genes by RUNX2. Cell Tissue Res. *339*, 189–195.

Kornak, U., and Mundlos, S. (2003). Genetic disorders of the skeleton: a developmental approach. Am. J. Hum. Genet. *73*, 447–474.

Leung, V.Y., Gao, B., Leung, K.K., Melhado, I.G., Wynn, S.L., Au, T.Y., Dung, N.W., Lau, J.Y., Mak, A.C., Chan, D., and Cheah, K.S. (2011). SOX9 governs differentiation stage-specific gene expression in GP chondrocytes via direct concomitant transactivation and repression. PLoS Genet. *7*, e1002356.

Li, T., Su, L., Zhong, N., Hao, X., Zhong, D., Singhal, S., and Liu, X. (2013). Salinomycin induces cell death with autophagy through activation of endoplasmic reticulum stress in human cancer cells. Autophagy *9*, 1057–1068.

Li, Y., Trivedi, V., Truong, T.V., Koos, D.S., Lansford, R., Chuong, C.M., Warburton, D., Moats, R.A., and Fraser, S.E. (2015). Dynamic imaging of the GP cartilage reveals multiple contributors to skeletal morphogenesis. Nat. Commun. *6*, 6798.

Magee, C., Nurminskaya, M., and Linsenmayer, T.F. (2001). UDP-glucose pyrophosphorylase: up-regulation in hypertrophic cartilage and role in hyaluronan synthesis. Biochem. J. *360*, 667–674.

Magwene, P.M., Lizardi, P., and Kim, J. (2003). Reconstructing the temporal ordering of biological samples using microarray data. Bioinformatics *19*, 842–850.

Minina, E., Wenzel, H.M., Kreschel, C., Karp, S., Gaffield, W., McMahon, A.P., and Vortkamp, A. (2001). BMP and Ihh/PTHrP signaling interact to coordinate chondrocyte proliferation and differentiation. Development *128*, 4523–4534.

Mueller, M.B., and Tuan, R.S. (2008). Functional characterization of hypertrophy in chondrogenesis of human mesenchymal stem cells. Arthritis Rheum. *58*, 1377–1388.

Myllyharju, J. (2014). Extracellular matrix and developing GP. Curr. Osteoporos. Rep. *12*, 439–445.

Navin, N.E. (2015). The first five years of single-cell cancer genomics and beyond. Genome Res. *25*, 1499–1507.

Pina, C., Teles, J., Fugazza, C., May, G., Wang, D., Guo, Y., Soneji, S., Brown, J., Edén, P., Ohlsson, M., et al. (2015). Single-Cell Network Analysis Identifies DDIT3 as a Nodal Lineage Regulator in Hematopoiesis. Cell Rep. *11*, 1503–1510.

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.

Scotti, C., Piccinini, E., Takizawa, H., Todorov, A., Bourgine, P., Papadimitropoulos, A., Barbero, A., Manz, M.G., and Martin, I. (2013). Engineering of a functional bone organ through endochondral ossification. Proc. Natl. Acad. Sci. USA *110*, 3997–4002.

Shin, J., Berg, D.A., Zhu, Y., Shin, J.Y., Song, J., Bonaguidi, M.A., Enikolopov, G., Nauen, D.W., Christian, K.M., Ming, G.L., and Song, H. (2015). Single-Cell RNA-Seq with Waterfall Reveals Molecular Cascades underlying Adult Neurogenesis. Cell Stem Cell *17*, 360–372.

Swift, S., Tucker, A., Vinciotti, V., Martin, N., Orengo, C., Liu, X., and Kellam, P. (2004). Consensus clustering and functional interpretation of gene-expression data. Genome Biol. *5*, R94.

Terpstra, L., Prud'homme, J., Arabian, A., Takeda, S., Karsenty, G., Dedhar, S., and St-Arnaud, R. (2003). Reduced chondrocyte proliferation and chondrodysplasia in mice lacking the integrin-linked kinase in chondrocytes. J. Cell Biol. *162*, 139–148.

Trapnell, C. (2015). Defining cell types and states with single-cell genomics. Genome Res. *25*, 1491–1498.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N.J., Livak, K.J., Mikkelsen, T.S., and Rinn, J.L. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat. Biotechnol. *32*, 381–386.

Treutlein, B., Brownfield, D.G., Wu, A.R., Neff, N.F., Mantalas, G.L., Espinoza, F.H., Desai, T.J., Krasnow, M.A., and Quake, S.R. (2014). Reconstructing lineage hierarchies of the distal lung epithelium using single-cell RNA-seq. Nature *509*, 371–375.

Ulici, V., Hoenselaar, K.D., Gillespie, J.R., and Beier, F. (2008). The PI3K pathway regulates endochondral bone growth through control of hypertrophic chondrocyte differentiation. BMC Dev. Biol. *8*, 40.

Van Brocklyn, J.R., and Williams, J.B. (2012). The control of the balance between ceramide and sphingosine-1-phosphate by sphingosine kinase: oxidative stress and the seesaw of cell survival and death. Comp. Biochem. Physiol. B Biochem. Mol. Biol. *163*, 26–36.

Wahyudi, G., Wasito, I., Melia, T., and Budi, I. (2011). Robust consensus clustering for identification of expressed genes linked to malignancy of human colorectal carcinoma. Bioinformation *6*, 279–282.

Waldhaus, J., Durruthy-Durruthy, R., and Heller, S. (2015). Quantitative High-Resolution Cellular Map of the Organ of Corti. Cell Rep. *11*, 1385–1399.

Wen, L., and Tang, F. (2014). Reconstructing complex tissues from single-cell analyses. Cell *157*, 771–773.

Wietecha, M.S., Chen, L., Ranzer, M.J., Anderson, K., Ying, C., Patel, T.B., and DiPietro, L.A. (2011). Sprouty2 downregulates angiogenesis during mouse skin wound healing. Am. J. Physiol. Heart Circ. Physiol. *300*, H459–H467.

Wongdee, K., Thonapan, N., Saengamnart, W., Krishnamra, N., and Charoenphandhu, N. (2013). Bromocriptine modulates the expression of PTHrP receptor, Indian hedgehog, and Runx2 proteins in the GP of lactating rats. Mol. Cell. Biochem. *381*, 191–199.

Yang, G., Zhu, L., Hou, N., Lan, Y., Wu, X.M., Zhou, B., Teng, Y., and Yang, X. (2014a). Osteogenic fate of hypertrophic chondrocytes. Cell Res. *24*, 1266–1269.

Yang, L., Tsang, K.Y., Tang, H.C., Chan, D., and Cheah, K.S. (2014b). Hypertrophic chondrocytes can become osteoblasts and osteocytes in endochondral bone formation. Proc. Natl. Acad. Sci. USA *111*, 12097–12102.

Yao, Y., and Wang, Y. (2013). ATDC5: an excellent in vitro model cell line for skeletal development. J. Cell. Biochem. *114*, 1223–1229.

Yeung Tsang, K., Wa Tsang, S., Chan, D., and Cheah, K.S. (2014). The chondrocytic journey in endochondral bone growth and skeletal dysplasia. Birth Defects Res. C Embryo Today *102*, 52–73.

Zhang, P., Jimenez, S.A., and Stokes, D.G. (2003). Regulation of human COL9A1 gene expression. Activation of the proximal promoter region by SOX9. J. Biol. Chem. *278*, 117–123.

Zhao, Q., Beck, A.J., Vitale, J.M., Schneider, J.S., Gao, S., Chang, C., Elson, G., Leibovich, S.J., Park, J.Y., Tian, B., et al. (2011). Developmental ablation of Id1 and Id3 genes in the vasculature leads to postnatal cardiac phenotypes. Dev. Biol. *349*, 53–64.