Integration of a Computational Pipeline for Dynamic Inference of Gene Regulatory Networks in Single Cells

Kyung Dae Ko[1*], Stefania Dell'Orso[2], Aster H. Juan[1], and Vittorio Sartorelli[1,3*]

[1]Laboratory of Muscle Stem Cells and Gene Regulation

[2]Genome Technology Unit

National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS), NIH

Bethesda, MD 208292, USA

[3]Lead Contact

*Correspondence: kyungdae.ko@nih.gov; sartorev@mail.nih.gov

SUMMARY

Single-cell RNA-seq permits the characterization of the molecular expression states of individual cells. Several methods have been developed to spatially and temporally resolve individual cell populations. However, these methods are not always integrated and some of them are constrained by prior knowledge. Here, we present an integrated pipeline for inference of gene regulatory networks. The pipeline does not rely on prior knowledge, it improves inference accuracy by integrating signatures from different data dimensions and facilitates tracing variation of gene expression by visualizing gene-interacting patterns of co-expressed gene regulatory networks at distinct developmental stages.

INTRODUCTION

The sequential and dynamic establishment and dismantling of gene regulatory networks (GRNs) instruct uncommitted and progenitor cells to adopt or avoid branching lineage choices (Davidson, 2006). Bulk transcriptomes have provided considerable insights and fostered the discovery and characterization of GRNs (Trapnell et al., 2010; Wang et al., 2009). However, bulk transcriptomes provide population-based averaged measurements which blur cell heterogeneity and developmental dynamics of asynchronous cell populations. Single-cell transcriptome technologies (scRNA-seq) capture cell heterogeneity and thus are useful for the discovery of cell populations, identification of cell mutants, and quantification of subpopulations (Linnarsson and Teichmann, 2016). Leveraging on the ability of generating thousands of individual measurements, methods have been developed to spatially and temporally resolve cell populations. Clustering and dimensionality reduction algorithms such as PCA, tSNE, and diffusion maps permit the identification and enumeration of cell types among cell populations (Satija et al., 2015) (Butler et al., 2018). Temporal trajectories are generated by pseudotime ordering of single cells to identify unique transition paths among different cell states (Trapnell et al., 2014) (Qiu et al., 2017) or by predicting future states of gene expression based on measurements of unspliced and spliced transcripts (La Manno et al., 2018). Using spatial or temporal information from clustering and trajectories, boolean models (Woodhouse et al., 2018), co-expression analysis (Allen et al., 2012), and multivariate information theory-based algorithms (Chan et al., 2017) have been successfully employed to infer GRNs. However, their accuracy depends on the size of the network (Fiers et al., 2018) and methods of normalization (Crow et al., 2016). To alleviate these issues, SCENIC (Aibar et al., 2017) combines co-expression with DNA binding motif enrichment analysis and SINCERA (Guo et al., 2015) makes use of scRNA-seq specific cell-type gene signatures. Because their power to infer GRNs is knowledge-based, the use of these models is constrained by the availability of annotated datasets (Fiers et al., 2018).

Here, we present an integrated pipeline for GRNs inference which uses clustering, temporal, and biological signatures extracted directly from scRNA-seq datasets. To evaluate its predictive power, we apply it to datasets derived from differentiating human

pluripotent stem cells. The pipeline correctly identifies signaling pathways activated and dismantled at specific stages of cell differentiation and reveals the composition of gene hubs underlying discrete GRNs in pluripotent and committed human cells.

RESULTS

Pipeline Workflow

The integrated computational pipeline for single cell gene regulatory network (IMSGEN) (Figure 1A) starts with the identification of transcripts corresponding to candidate genes in top 100 gene transcripts with average log fold changes among all cell clusters. Gene signatures are then employed to identify signaling pathways and gene ontology (GO) enrichment within each cluster. The clusters are temporally ordered by cell re-clustering using principal component analysis for dimension reduction and minimum spanning tree (MST) methods for trajectory modeling to predict the temporal relations among the clusters. After temporal ordering of cell clusters, distance matrices of dynamic and cluster-specific gene-interacting patterns are employed to infer GRNs by corrected gene interacting (CGI) maps and force-directed graph (FDG) network model (Fruchterman, 1991). This approach permits the identification and visualization of GRN transitions occurring in distinct cell states independent of cell-type and annotation biases.

Clustering and Temporal Ordering of Human Pluripotent Cells undergoing Mesoderm Differentiation

We tested IMSGEN performance on experimental data generated from scRNA-seq of 498 human pluripotent cells undergoing mesoderm differentiation (Loh et al., 2016). scRNA-seq was performed on cells induced to differentiate and captured at specific developmental stages: 51 human pluripotent stem cells (embryonic stem cells, ESCs), cells differentiated into anterior and middle primitive streaks (59 APS and 22 MPS cells, respectively), paraxial mesoderm (67 PXM cells), lateral mesoderm (LatM 55 cells), somitomere (76 cells), early somites (36 cells), dermomyotome (67 cells), and sclerotome (65 cells) (Figure 1B). We aggregated scRNA-seq datasets without *a priori* knowledge of

the developmental stage of the individual cells and clustered gene expression data using PCA and graph-based clustering method (Butler et al., 2018). Cells segregated into 8 clusters, representing each developmental stage with the exception of APS and MPS populations which could not be individually and correctly identified (Figure S1A). To evaluate somitogenesis (left branch of Figure 1B), we temporally ordered cell clusters by independent component analysis (ICA) for dimension reduction and MST for trajectory model (Figure S1B). Signatures for the individual clusters were generated by differential gene expression of the top 100 expressed genes (Figure 1C, Table S1) which were subsequently employed to perform pathway and GO enrichment analyses for the different clusters (Figure 1D, Figure S1C, Table S2). As expected, pathways regulating pluripotency of stem cells were identified and, consistent with their role in mesoderm induction (Cheung et al., 2012) (Gertow et al., 2013) (Loh et al., 2016), WNT and TGFβ pathways were also captured by this analysis. This unbiased approach permitted the identification of metabolic (glycolysis-gluconeogenesis) and several other pathways not directly queried in (Loh et al., 2016) (Figure 1D and see below).

Identification of Signaling Pathways and Visualization of Gene Regulatory Networks

WNT and TGFβ pathways play key roles in mesoderm formation starting from human pluripotent stem cells (Cheung et al., 2012) (Gertow et al., 2013) (Loh et al., 2016). However, the composition, structure and temporal formation of WNT and TGFβ GRNs occurring at discrete differentiation states have not been elucidated. To infer GRNs in cells undergoing somitogenesis, we queried sub-datasets for WNT and TGFβ signaling pathways (Figure 1C,D and Table S1) and generated pairwise matrices to represent gene co-expression using normalized distance methods (Figure 2). In these matrices, gene proximity (red in Figure 2) indicates co-expression while gene distance (blue in Figure 2) indicates absence of co-expression, thus allowing to evaluate the presence of functionally connected and related modules (Stuart et al., 2003) (Nguyen and Lio, 2009). This analysis revealed formation of WNT GRN in ESC and APS (Figure 2A). In PXM and somitomere cells, the GRN's strength was reduced and further decreased in early somite and

sclerotome cells (Figure 2A). Similarly, gene interactions within a TGFβ GRN present in ESC and APS were diminished in PXM and somitomere cells and continued to decline in early somite and sclerotome cells (Figure 2B). These findings are consistent with an inhibitory role exerted by both WNT and TGFβ signaling pathways on early somitogenesis (Loh et al., 2016). Using the same approach, an embryonic morphogenesis GRN was revealed to be gradually established and to progressively increase its connectivity as cells progressed from pluripotent to more differentiated cell states (Figure 2C), reflecting morphogenesis that occurs during cell differentiation. Appropriate temporal expression of the somitomere-specific MESP2, HEYL, and HOPX, and somite-specific MEOX1 and PARAXIS (TCF15) and FOXC2 genes (Loh et al., 2016) confirmed that the matrices accurately represent the individual cell developmental stages (Figure S1D). Thus, using an unbiased approach, our computational pipeline correctly identified and ordered GRNs for pathways known to regulate human pluripotent cell differentiation. Converting similarity matrices to adjacency matrices, we visualized co-expression networks using FDG network model (Fruchterman, 1991). This way, we could identify the nodes (genes), edges (gene connectivity) and overall structures of the WNT, TGFβ, and embryonic morphogenesis GRNs at each different stages of somitogenesis (Figure S2A,B). Both WNT and TGFβ networks increased their connectivity during the transition from pluripotency (ESC) to APS. Genetic interactions were pruned and refined at later stages of somitogenesis (Figure 2A,B). In early somite cells, the transcription factors Smad2, TCF7L2 (TCF-4), the TCF-4 interacting corepressor CtBP1, and calcineurin (PPP3CA), known to be involved in mesoderm formation (Dunn et al., 2004) (Kardon et al., 2003) (Hogan et al., 2003), were found to establish a WNT subnetwork (Figure S2A). A TGFβ subnetwork revealed connectivity between the TGFβ-stimulated Rho-associated kinase ROCK1, important for somitogenesis (Wei et al., 2001) and the activin A receptor ACVR2B in early somite cells (Figure S2B). SMAD2 and the serine/threonine Protein Phosphatase 2 (PPP2R1A) were equally connected in both WNT and TGFβ subnetworks, confirming cross-talk of the two pathways (Attisano and Wrana, 2013). In contrast to the dismantling of the TGFβ and WNT GRNs, a GRN composed of genes related to embryogenic morphogenesis (Figure S2C) gradually increased connectivity acquiring

additional nodes and edges in cells undergoing differentiation. Thus, IMSGEN was able to identify and dissect the composition, structure and temporal formation of GRNs in human pluripotent cells undergoing mesoderm differentiation.

DISCUSSION

Inference of GRNs from scRNA-seq data provides important clues to understand gene expression dynamics in developing systems. The pipeline described here, IMSGEN, complements and integrates existing methods. The salient characteristics of the pipeline are its independence from annotation biases, improved accuracy of inference integration of transcriptional signatures from different data dimensions, and easy visualization of gene interacting patterns and co-expressed GRNs. The pipeline performed well when tested with published data and its use can be extended to analyze GRNs during cellular development in any cell type and organism.

ACKNOWLEGMENTS

We thank Dr. Hong-Wei Sun (Biodata Mining and Discovery Section, NIAMS) for critical reading of the manuscript. This work was supported by the Intramural Research Program of the National Institute of Arthritis, and Musculoskeletal and Skin diseases (NIAMS) at the National Institutes of Health (NIH grants AR041126 and AR041164).

AUTHOR CONTRIBUTIONS

Conceptualization, K.D.K and V.S.; Software Design, K.D.K; Writing K.D.K, V.S.; Biological expertise and advice, S.D.O, A.H.J, V.S.

DECLARATION OF INTERESTS

The authors declare no competing interests

WEB RESOURCES

GitHub, https://github.com/holyone09/lmsgen

Seurat, https://satijalab.org/seurat

Pathfind, https://cran.r-project.org/web/packages/pathfindR/index.html

Metascope, http://metascape.org/gp/index.html#/main/step1

Pheatmap, https://cran.r-project.org/web/packages/pheatmap/index.html

Igraph, https://igraph.org/redirect.html

REFERENCES

Aibar, S., Gonzalez-Blas, C.B., Moerman, T., Huynh-Thu, V.A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J.C., Geurts, P., Aerts, J., *et al.* (2017). SCENIC: single-cell regulatory network inference and clustering. Nature methods *14*, 1083-1086.

Allen, J.D., Xie, Y., Chen, M., Girard, L., and Xiao, G. (2012). Comparing statistical methods for constructing large scale gene networks. PLoS One *7*, e29348.

Attisano, L., and Wrana, J.L. (2013). Signal integration in TGF-beta, WNT, and Hippo pathways. F1000Prime Rep *5*, 17.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology *36*, 411-420.

Chan, T.E., Stumpf, M.P.H., and Babtie, A.C. (2017). Gene Regulatory Network Inference from Single-Cell Data Using Multivariate Information Measures. Cell Syst *5*, 251-267 e253.

Cheung, C., Bernardo, A.S., Trotter, M.W., Pedersen, R.A., and Sinha, S. (2012). Generation of human vascular smooth muscle subtypes provides insight into embryological origin-dependent disease susceptibility. Nature biotechnology *30*, 165-173.

Crow, M., Paul, A., Ballouz, S., Huang, Z.J., and Gillis, J. (2016). Exploiting single-cell expression to characterize co-expression replicability. Genome biology *17*, 101.

Davidson, E.H. (2006). The Regulatory Genome: Gene Regulatory Networks in Development and Evolution (Academic Press).

Dunn, N.R., Vincent, S.D., Oxburgh, L., Robertson, E.J., and Bikoff, E.K. (2004). Combinatorial activities of Smad2 and Smad3 regulate mesoderm formation and patterning in the mouse embryo. Development *131*, 1717-1728.

Fiers, M., Minnoye, L., Aibar, S., Bravo Gonzalez-Blas, C., Kalender Atak, Z., and Aerts, S. (2018). Mapping gene regulatory networks from single-cell omics data. Brief Funct Genomics *17*, 246-254.

Fruchterman, T.M.J.R.E.M. (1991). Graph drawing by force-directed placement. J Software: Practice and Experience *21*, 1129-1164.

Gertow, K., Hirst, C.E., Yu, Q.C., Ng, E.S., Pereira, L.A., Davis, R.P., Stanley, E.G., and Elefanty, A.G. (2013). WNT3A promotes hematopoietic or mesenchymal differentiation from hESCs depending on the time of exposure. Stem Cell Reports *1*, 53-65.

Guo, M., Wang, H., Potter, S.S., Whitsett, J.A., and Xu, Y. (2015). SINCERA: A Pipeline for Single-Cell RNA-Seq Profiling Analysis. PLoS Comput Biol *11*, e1004575.
Hogan, P.G., Chen, L., Nardone, J., and Rao, A. (2003). Transcriptional regulation by calcium, calcineurin, and NFAT. Genes & development *17*, 2205-2232.
Kardon, G., Harfe, B.D., and Tabin, C.J. (2003). A Tcf4-positive mesodermal population provides a prepattern for vertebrate limb muscle patterning. Developmental cell *5*, 937-944.
La Manno, G., Soldatov, R., Zeisel, A., Braun, E., Hochgerner, H., Petukhov, V., Lidschreiber, K., Kastriti, M.E., Lonnerberg, P., Furlan, A.*, et al.* (2018). RNA velocity of single cells. Nature *560*, 494-498.
Linnarsson, S., and Teichmann, S.A. (2016). Single-cell genomics: coming of age. Genome biology *17*, 97.
Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M.*, et al.* (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451-467.
Nguyen, V.A., and Lio, P. (2009). Measuring similarity between gene expression profiles: a Bayesian approach. BMC Genomics *10 Suppl 3*, S14.
Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nature methods *14*, 979-982.
Satija, R., Farrell, J.A., Gennert, D., Schier, A.F., and Regev, A. (2015). Spatial reconstruction of single-cell gene expression data. Nature biotechnology *33*, 495-502.
Stuart, J.M., Segal, E., Koller, D., and Kim, S.K. (2003). A gene-coexpression network for global discovery of conserved genetic modules. Science *302*, 249-255.
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. Nature biotechnology *32*, 381-386.
Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., and Pachter, L. (2010). Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology *28*, 511-515.
Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. Nature reviews Genetics *10*, 57-63.
Wei, L., Roberts, W., Wang, L., Yamada, M., Zhang, S., Zhao, Z., Rivkees, S.A., Schwartz, R.J., and Imanaka-Yoshida, K. (2001). Rho kinases play an obligatory role in vertebrate embryonic organogenesis. Development *128*, 2953-2962.
Woodhouse, S., Piterman, N., Wintersteiger, C.M., Gottgens, B., and Fisher, J. (2018). SCNS: a graphical tool for reconstructing executable regulatory networks from single-cell genomic data. BMC Syst Biol *12*, 59.

FIGURE LEGENDS

Figure 1. (A) Pipeline's flowchart. (B) Scheme representing alternative mesoderm differentiation choices of human pluripotent stem cells. Left branch, somitogenesis; right branch, cardiogenesis (Loh et al., 2016). (C) Heatmap of top 100 most variably expressed genes in human pluripotent stem cells (ESC) and at the indicated stages of mesoderm differentiation. (D) Identification of pathways enriched in human pluripotent stem cells (ESC) and at the indicated stages of mesoderm differentiation.

Figure 2 (A-C) Distance matrices indicating gene connectivity (highest connectivity, red; lowest connectivity, blue) for the WNT (A), TGFβ (B), and Embryonic morphogenesis (C) genes in human pluripotent stem cells (ESC) and at the indicated stages of mesoderm differentiation.

Figure S1 (A) tSNE-based clustering of human pluripotent stem cells (ESC) and of cells at different stages of mesoderm differentiation. (B) Pseudotime ordering of ESC and cells at different stages of somitogenesis. (C) Pathway enrichment representation based on pathways identified in Figure 1C. The size of the symbols for differentially-expressed genes (DEGs) is proportional to gene number (10,20, or 30 genes, respectively) and p-values of the single pathways indicated in the -log10 red color scale m (lower p-values, brighter red color). (D) Heatmap of MESP2, HEYL, HOPX, MEOX1, FOXC2 and TCF15 (PARAXIS) transcripts at different stages of mesoderm differentiation.

Figure S2 (A-C) Force-directed networks of WNT (A), TGFβ (B), and Embryonic morphogenesis (C) genes in human pluripotent stem cells (ESC) and at the indicated stages of mesoderm differentiation. Red edges indicate positive and grey edges negative gene correlation.

Table S1. List of the top 100 differentially expressed genes in human pluripotent stem cells (ESC) and in different stages of mesoderm differentiation.

Table S2. Pathways enriched in human pluripotent stem cells (ESC) and in different stages of mesoderm differentiation.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Single cell RNA-seq data of human pluripotency	(Loh et al., 2016)	ftp://ftp.ncbi.nlm.nih.gov/geo/samples/GSM2257nnn/GSM2257302/
Software and Algorithms		
R package source code	This paper	https://github.com/holyone09/lmsgen
Seurat	(Butler et al., 2018)	https://satijalab.org/seurat/
Pathfind	Ulgen E. et al. 2018	https://cran.r-project.org/web/packages/pathfindR/index.html
Metascope	(Zhou et al., 2019)	http://metascape.org/gp/index.html#/main/step1
Pheatmap	Kolde R. et al. 2015	https://cran.r-project.org/web/packages/pheatmap/index.html
Igraph	Csardi G. et al. 2006	https://igraph.org/redirect.html

Contact for Reagent and Resource Sharing

Further information and requests for reagents should be directed to Lead Contact Vittorio Sartorelli (sartorev@mail.nih.gov).

Method Details

The pipeline (Figure 1A) consist of five modules, and each module is implemented by R programming language with its package. To evaluate the effectiveness of analyzing results from the pipeline, we applied it to scRNA-seq datasets of human pluripotent stem cells (Loh et al., 2016).

Extracting spatial, biological, and temporal signatures from datasets

The first step of the pipeline is to reduce the matrix of UMI counts or gene expressing values such as TPMs (Transcripts per millions) into PCA dimension. Based on Seurat R package (Butler et al., 2018), cells are clustered using graph-based clustering methods with PCA values and top-ranking genes are collected among clusters calculating average

log fold changes. Next, the results of clustering and top-ranking gene list are transferred into the modules to identify biological signatures and reconstruct the temporal orders of clustered groups separately. Using Pathfind R package (Ulgen E. et al. 2018), we predict signaling pathways with low p-value (<0.01) for entire cell populations and gather candidate genes related to the pathways involved in cell differentiation. In addition, genes are extracted from GO (Gene Ontology) enrichment analysis ($p < 0.01$) of Metascope (Zhou et al., 2019). Finally, lists of genes are preprocessed to visualize gene interaction. To reconstruct temporal orders of clusters, we reduce the dimension of the distance matrix to PCA after calculating euclidean distances among scRNAseq data. Then, temporal signatures are restored from PCA by Mclust and MST (Minimum Spanning Tree) algorithms (Xu et al., 2002). Finally, we reorganize spatial clusters following temporal signatures.

Visualization of gene-interacting patterns and GRNs

Using pre-processed gene lists, we generate the temporal submatrices of selected gene expression from an original expressing matrix. To visualize the patterns of gene interactions in the temporal submatrices, we generate the matrices of similarity calculating normalized distance methods with eq.(1), and we draw correlational heatmaps using the matrices of similarity using Pheatmap R package (Kolde R. et al. 2015).

$$\mathrm{Similarity}(x,y) = \sqrt{\sum_i^n (x_i - y_i)^2} \Big/ \mathrm{MAX}\left(\sqrt{\sum_i^n (x_i - y_i)^2}\right) \qquad (1)$$

Converting these similarity matrices to adjacency matrices for co-expression networks, we visualize co-expression networks in each state using Igraph R package (Csardi G. et al. 2006) and the Fruchterman-Reingold layout algorithm (Fruchterman, 1991). Genes with concordant expression levels are closely positioned forming multiple hubs within a given GRN.

Data and Software Availability

Source code and installation of the R package is available at https://github.com/holyone09/lmsgen under Open-source R package under 'GPL (version 2 or later)'.

REFERENCES

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nature biotechnology *36*, 411-420.

Csardi G., Nepusz T., The igraph software package for complex network research, InterJournal, Complex Systems 2006;1695.

Kolde R., pheatmap: Pretty Heatmaps. R package version 1.0.8. 2015

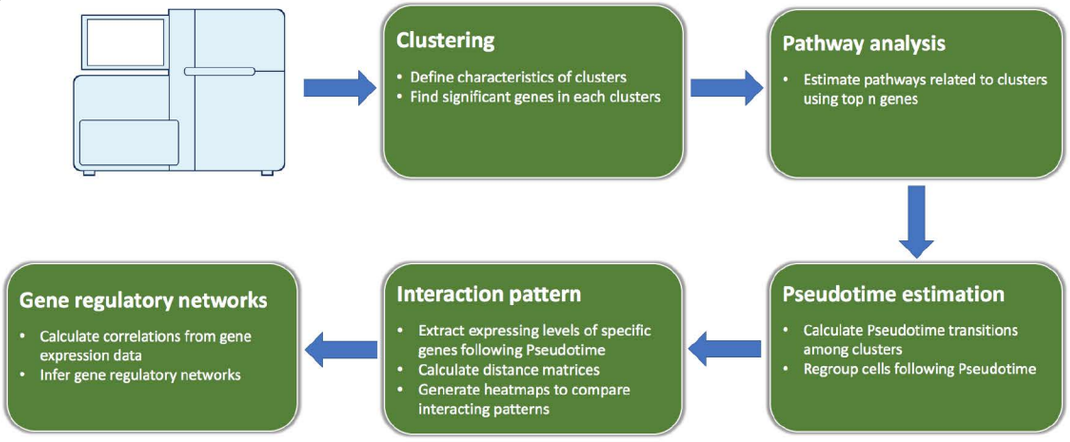
Fruchterman, T.M.J.R.E.M. (1991). Graph drawing by force-directed placement. J Software: Practice and Experience *21*, 1129-1164.

Loh, K.M., Chen, A., Koh, P.W., Deng, T.Z., Sinha, R., Tsai, J.M., Barkal, A.A., Shen, K.Y., Jain, R., Morganti, R.M.*, et al.* (2016). Mapping the Pairwise Choices Leading from Pluripotency to Human Bone, Heart, and Other Mesoderm Cell Types. Cell *166*, 451-467.

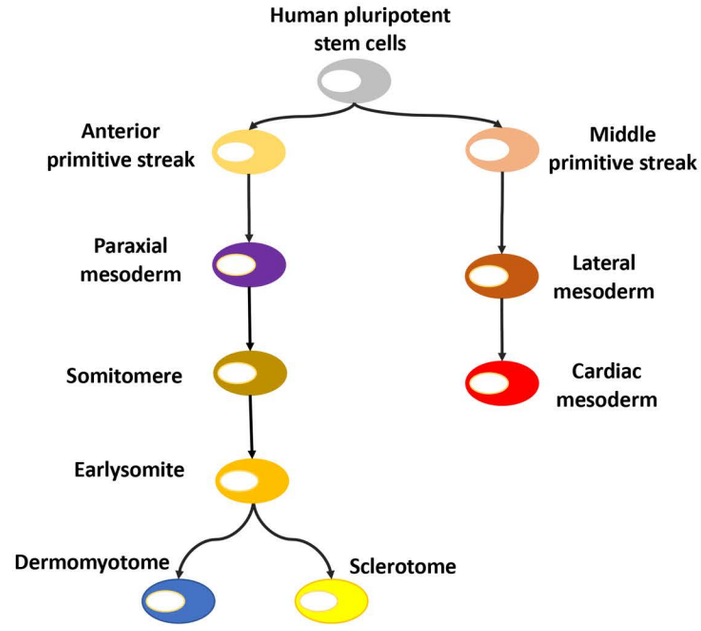
Xu, Y., Olman, V., and Xu, D. (2002). Clustering gene expression data using a graph-theoretic approach: an application of minimum spanning trees. Bioinformatics *18*, 536-545.

Zhou, Y., Zhou, B., Pache, L., Chang, M., Khodabakhshi, A.H., Tanaseichuk, O., Benner, C., and Chanda, S.K. (2019). Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nature communications *10*, 1523.

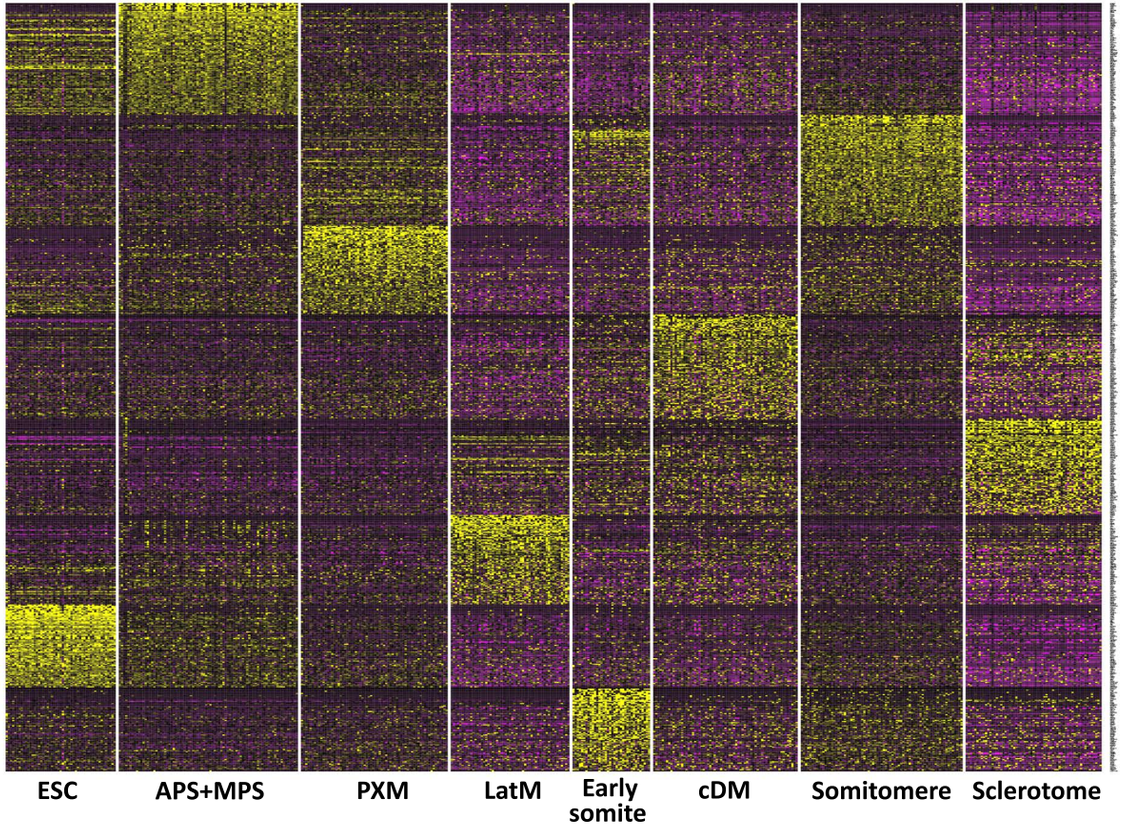
A.



B.

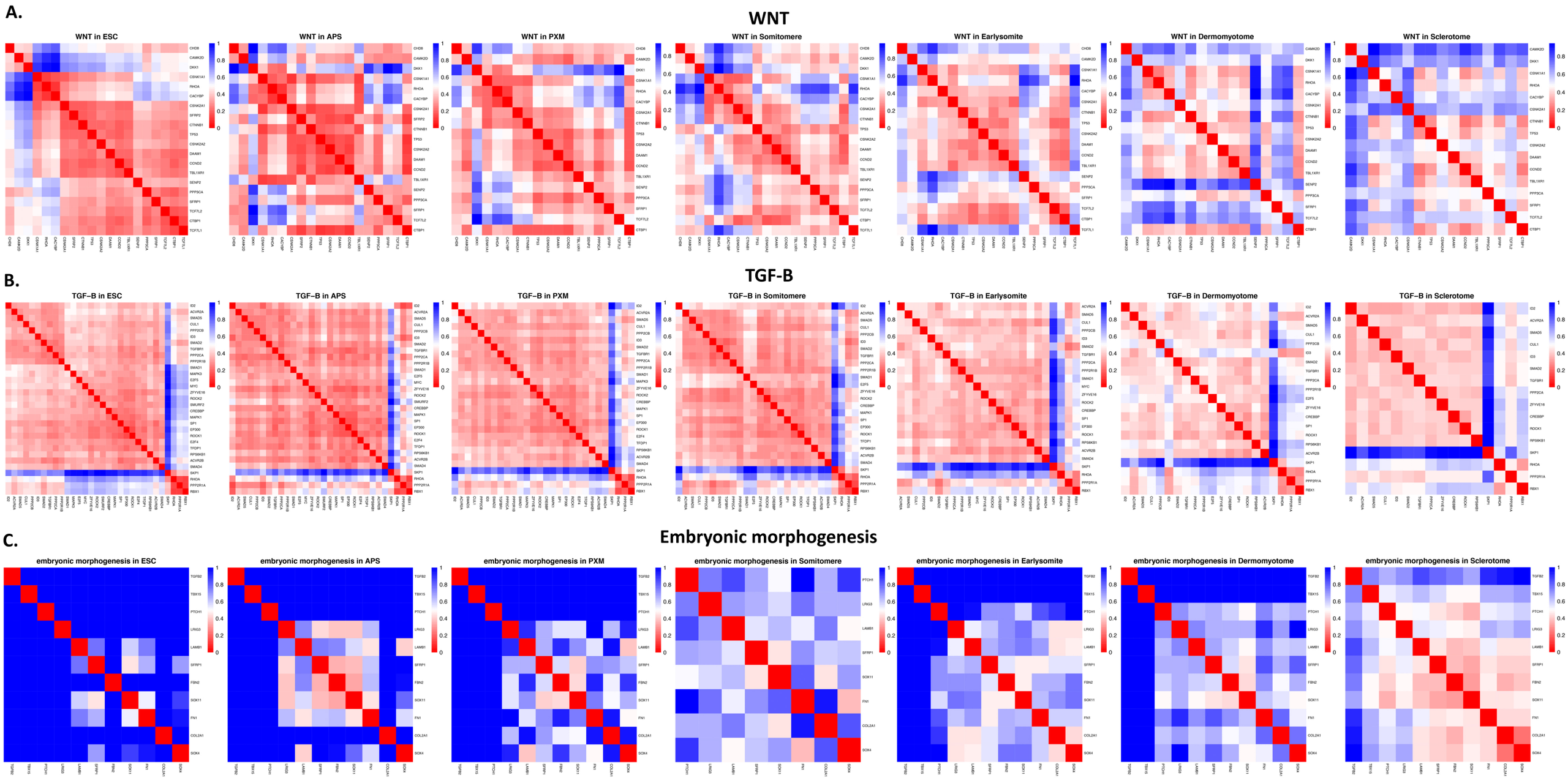


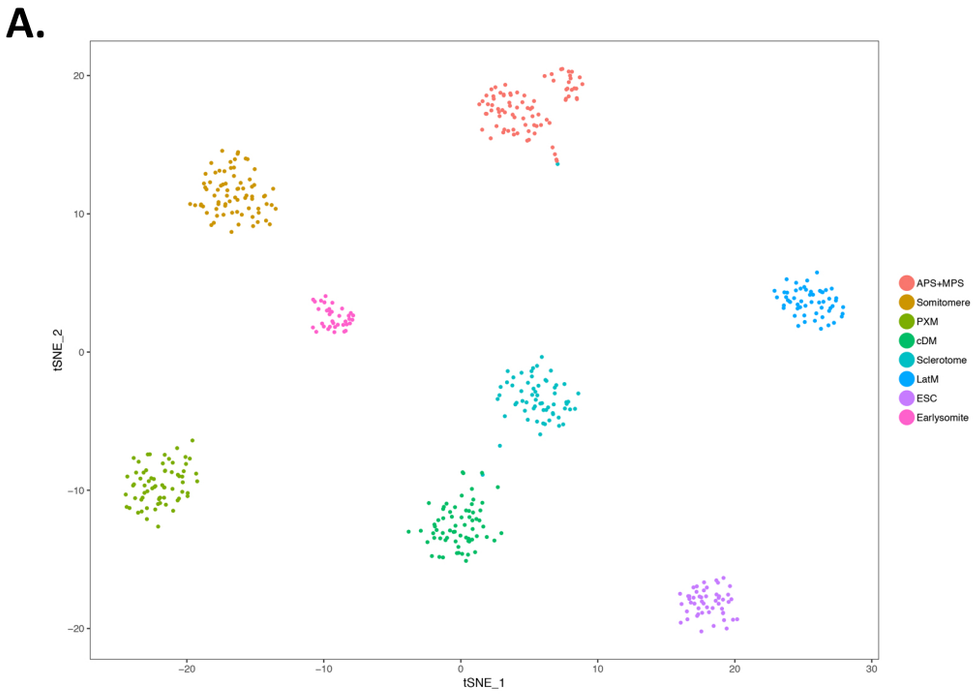
C.

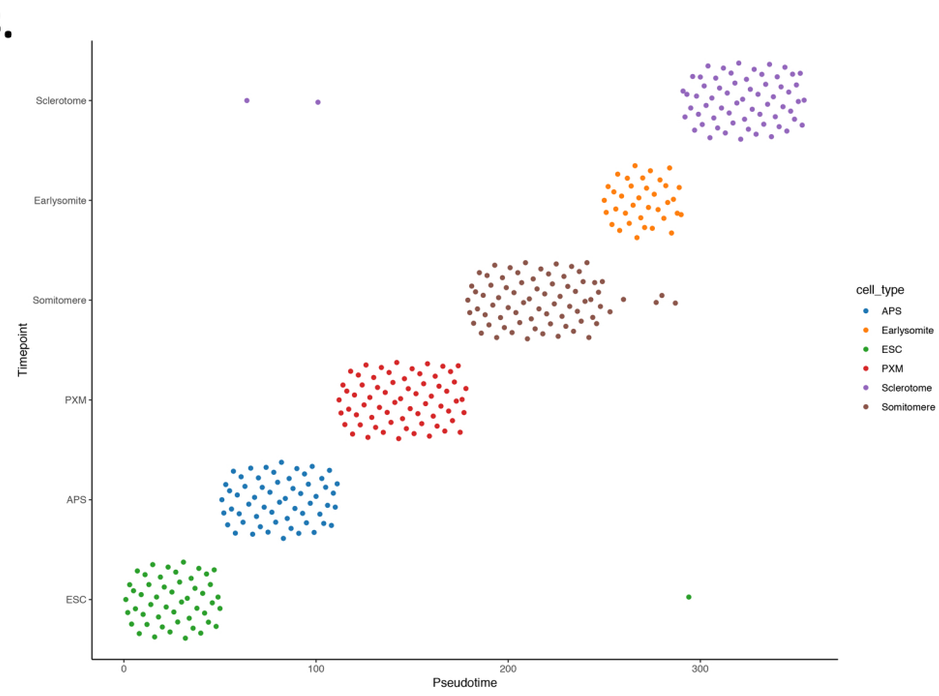


D.

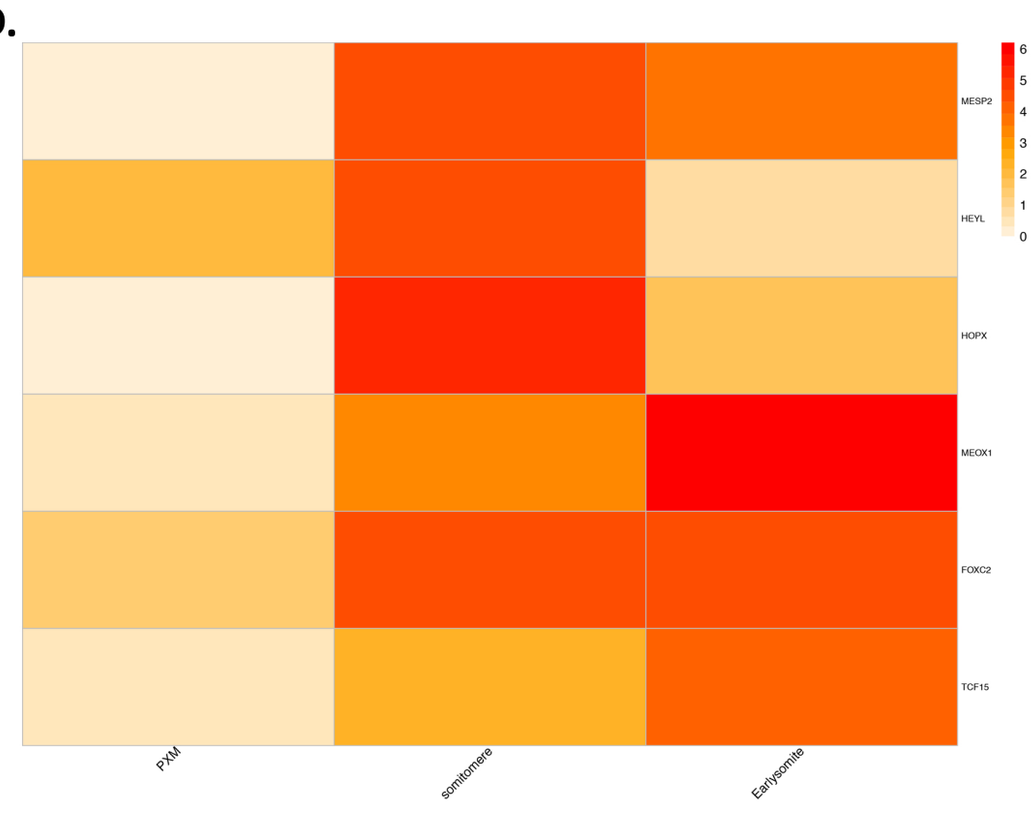
ID	Pathway	Fold_Enrichment	occurrence	lowest_p	highest_p	Upregulated gene	State related to upregulated gene
hsa03010	Ribosome	30.54470	10	4.3e-10	3.3e-05	RPS27L, RPL11, RPL8, RPL18, RPL34, RPL37A, RPL22L1, RPS5, RPS3, MRPS6	ESC, APS+MPS, Earlysomite, somitomere, cDM
hsa00010	Glycolysis - Gluconeogenesis	19.77838	10	4.5e-08	9.6e-07	ALDOA, ALDOC, ALDH3A2, LDHA, PKM, ENO1, GAPDH, TPI1, GPI, PGK1	ESC, APS+MPS, somitomere
hsa04110	Cell cycle	65.43666	10	4.7e-08	5.7e-06	ORC6, MAD2L2, YWHAE, YWHAZ, TP53, CDKN1C, TGFB2, CCNB2, CDK4, CCND2, TFDP2	ESC,LatM, somitomere, cDM, Sclerotome
hsa05130	Pathogenic Escherichia coli infection	112.93210	10	1.5e-07	1.5e-07	TUBA1A, TUBA1C, YWHAZ, KRT18, EZR, RHOA, TUBB2A, TUBB2B	APS+MPS, PXM, somitomere
hsa04550	Signaling pathways regulating pluripotency of stem cells	60.98333	10	1.9e-07	1.9e-07	WNT5A, WNT8A, PIK3R3, FGFR2, ID1, ID2, ID3, ID4, HESX1, SKIL, MEIS1, HAND1, ISL1, NODAL	ESC, APS+MPS,LatM, Earlysomite, somitomere, cDM
hsa04145	Phagosome	53.80882	10	3.2e-07	3.2e-07	TUBA1A, TUBA1C, TUBB2A, TUBB2B, ATP6V1F, LAMP2	APS+MPS,PXM, somitomere, cDM
hsa04540	Gap junction	79.19913	10	1.7e-06	1.7e-06	GJA1, TUBB2A, TUBB2B, TUBA1A, TUBA1C, PDGFRB, PDGFA	APS+MPS,PXM, Earlysomite, somitomere, Sclerotome
hsa04714	Thermogenesis	42.54651	10	1.9e-06	1.2e-04	ACTL6A, ZNF516, COX7A2, COX7C, COX1, COX2, COX3, CYTB, ND4, ND4L, UQCRB, UQCRC1, UQCR11, UQCRQ, NDUFAF4, NDUFAF7, ATP6, ATP8	ESC,APS+MPS,PXM,LatM, Earlysomite, somitomere, Sclerotome
hsa05418	Fluid shear stress and atherosclerosis	37.24173	10	4.2e-06	6.4e-05	PIK3R3, NQO1, MGST1, MGST2, GSTO1, TXN, SQSTM1, MMP2, PDGFA, RHOA, TP53, CAV1, CALM1, CALM2, HSP90AB1, SDC4	ESC,APS+MPS,PXM,LatM, Earlysomite, somitomere,cDM, Sclerotome
hsa03050	Proteasome	85.09302	10	8.1e-06	5.7e-03	PSMA5, PSMA3, PSMB5	APS+MPS
hsa00190	Oxidative phosphorylation	64.87589	10	8.1e-06	6.5e-04	ND4L, ND4, COX7C, COX7A2, CYTB, UQCRB, UQCRC1, UQCR11, UQCRQ, COX1, COX2, COX3, ATP6, ATP8, ATP6V1F, PPA1	ESC,APS+MPS,PXM,LatM, Earlysomite, somitomere,cDM, Sclerotome
hsa03040	Spliceosome	52.38368	10	8.4e-06	5.0e-02	SNRPE, SF3B2, SF3B6, SRSF6, SRSF7	APS+MPS,PXM, cDM, Sclerotome
hsa04310	Wnt signaling pathway	49.18011	10	1.3e-05	2.1e-05	PPP3CA, WNT5A, RHOA, DAAM1, PRICKLE2, CCND2, TCF7L2, LEF1, CSNK1A1, TBL1XR1, TP53, CSNK2A2, DKK1, DKK4, CER1, SFRP1, SFRP2, WNT8A, BAMBI, SERPINF1	ESC,APS+MPS,PXM,LatM, Earlysomite, somitomere,cDM, Sclerotome
hsa04350	TGF-beta signaling pathway	66.83105	10	1.5e-04	1.5e-04	NODAL, ID1, ID2, ID3, ID4, FST, RHOA, TGFB2, DCN, AMHR2, NOG, BAMBI, TGIF1	APS+MPS,PXM,LatM, somitomere, Sclerotome


A.
WNT
WNT in ESC
WNT in APS
WNT in PXM
WNT in Somitomere
WNT in Earlysomite
WNT in Dermomyotome
WNT in Sclerotome
B.
TGF-B
TGF−B in ESC
TGF−B in APS
TGF−B in PXM
TGF−B in Somitomere
TGF−B in Earlysomite
TGF−B in Dermomyotome
TGF−B in Sclerotome
C.
Embryonic morphogenesis
embryonic morphogenesis in ESC
embryonic morphogenesis in APS
embryonic morphogenesis in PXM
embryonic morphogenesis in Somitomere
embryonic morphogenesis in Earlysomite
embryonic morphogenesis in Dermomyotome
embryonic morphogenesis in Sclerotome


A.
tSNE_2
tSNE_1
APS+MPS
Somitomere
PXM
cDM
Sclerotome
LatM
ESC
Earlysomite


B.
Timepoint
Sclerotome
Earlysomite
Somitomere
PXM
APS
ESC
Pseudotime
cell_type
APS
Earlysomite
ESC
PXM
Sclerotome
Somitomere


C.
Ribosome
Cell cycle
Pathogenic Escherichia coli infection
Signaling pathways regulating pluripotency of stem cells
Phagosome
Glycolysis – Gluconeogenesis
Gap junction
Thermogenesis
Fluid shear stress and atherosclerosis
Spliceosome
Wnt signaling pathway
Oxidative phosphorylation
Parkinson's disease
Cellular senescence
Non–alcoholic fatty liver disease (NAFLD)
Non–small cell lung cancer
Alzheimer's disease
Glioma
Melanoma
Viral carcinogenesis
Pancreatic cancer
Chronic myeloid leukemia
Huntington's disease
Proteoglycans in cancer
Cardiac muscle contraction
Bacterial invasion of epithelial cells
ECM–receptor interaction
TGF–beta signaling pathway
Small cell lung cancer
HTLV–I infection
Colorectal cancer
Central carbon metabolism in cancer
Pentose phosphate pathway
Neurotrophin signaling pathway
Human papillomavirus infection
Measles
Prostate cancer
Hepatitis B
Apoptosis – multiple species
Breast cancer
Ferroptosis
Bladder cancer
HIF–1 signaling pathway
Renin secretion
Hepatocellular carcinoma
PI3K–Akt signaling pathway
Proteasome
Adrenergic signaling in cardiomyocytes
Amyotrophic lateral sclerosis (ALS)
Apoptosis
Kaposi's sarcoma–associated herpesvirus infection
Oocyte meiosis
Pathways in cancer
Mitophagy – animal
Protein processing in endoplasmic reticulum
Focal adhesion
Epstein–Barr virus infection
p53 signaling pathway
Shigellosis
Cysteine and methionine metabolism
Estrogen signaling pathway
Axon guidance
Regulation of actin cytoskeleton
Herpes simplex infection
Tight junction
Adherens junction
Longevity regulating pathway
Gastric acid secretion
Protein digestion and absorption
Acute myeloid leukemia
Rap1 signaling pathway
Salivary secretion
Aldosterone synthesis and secretion
Dopaminergic synapse
Leukocyte transendothelial migration
Hypertrophic cardiomyopathy (HCM)
Sphingolipid signaling pathway
Vascular smooth muscle contraction
Dilated cardiomyopathy (DCM)
Ribosome biogenesis in eukaryotes
Oxytocin signaling pathway
Fold Enrichment
−log10(lowest−p)
# of DEGs


D.
MESP2
HEYL
HOPX
MEOX1
FOXC2
TCF15
PXM
somitomere
Earlysomite