(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) **World Intellectual Property Organization**
International Bureau

(43) **International Publication Date**
28 March 2019 (28.03.2019)

W 1P O | **PCT**

(10) **International Publication Number**
WO 2019/060450 Al

(51) **International Patent Classification:**
C12N 5/07 (2010.01) ***CI2N 15/85*** (2006.01)
C12N 15/63 (2006.01)

(21) **International Application Number:**
PCT/US20 18/05 1808

(22) **International Filing Date:**
19 September 2018 (19.09.2018)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**

62/560,674	19 September 2017 (19.09.2017)	US
62/561,047	20 September 2017 (20.09.2017)	US

(71) **Applicants: THE BROAD INSTITUTE, INC.** [US/US]; 415 Main Street, Cambridge, Massachusetts 02142 (US). **MASSACHUSETTS INSTITUTE OF TECHNOLOGY** [US/US]; 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **WHITEHEAD INSTITUTE FOR BIOMEDICAL RESEARCH** [US/US]; 455 Main Street, Cambridge, Massachusetts 02142 (US).

(72) **Inventor; and**
(71) **Applicant: RIGOLLET, Philippe** [US/US]; **c/o** 77 Massachusetts Avenue, Cambridge, Massachusetts 02139 (US).

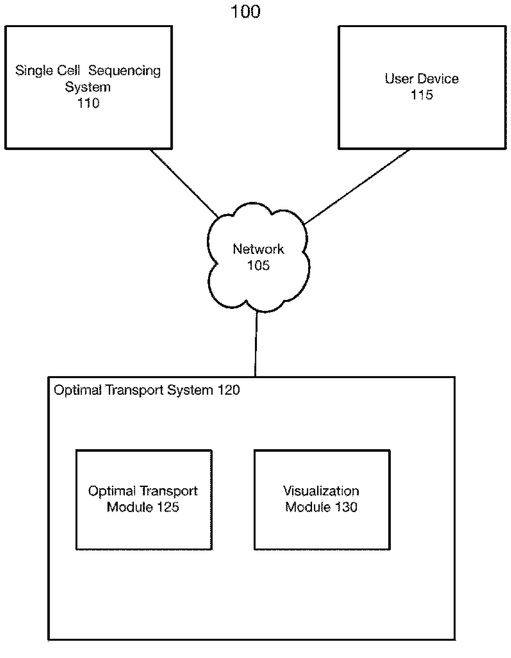
(72) **Inventors: SCHIEBINGER, Geoffrey;** c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **SHU, Jian;** c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **TABAKA, Marcin;** c/o 415 Main Street, Cambridge, Massachusetts 02 142 (US). **CLEARY, Brian;** c/o **77** Massachusetts Avenue, Cambridge, Massachusetts 02139 (US). **REGEV, Aviv;** c/o 415 Main Street, Cambridge, Massachusetts 02142 (US). **LANDER, Eric S.;** c/o 415 Main Street, Cambridge, Massachusetts 02142 (US).

(74) **Agent: XIN, Xiaoban** et al.; Johnson, Marcou & Isaacs, LLC, P.O. Box 691, Hoschton, Georgia 30548 (US).

(81) **Designated States** *(unless otherwise indicated, for every kind of national protection available):* AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP, KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

(84) **Designated States** *(unless otherwise indicated, for every kind of regional protection available):* ARIPO (BW, GH,

(54) **Title:** METHODS AND SYSTEMS FOR RECONSTRUCTION OF DEVELOPMENTAL LANDSCAPES BY OPTIMAL TRANSPORT ANALYSIS


100
Single Cell Sequencing System 110
User Device 115
Network 105
Optimal Transport System 120
Optimal Transport Module 125
Visualization Module 130

FTG. 1

(57) **Abstract:** Methods and compositions for producing induced pluripotent stem cell by introducing nucleic acids encoding one or more transcription factors including Obox6 into a target cell.

[Continued on next page]

GM, KE, LR, LS, MW, MZ, NA, RW, SD, SL, ST, SZ, TZ, UG, ZM, ZW), Eurasian (AM, AZ, BY, KG, KZ, RU, TJ, TM), European (AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, RS, SE, SI, SK, SM, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, KM, ML, MR, NE, SN, TD, TG).

Published:

— *with international search report (Art. 21(3))*

— *before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))*

METHODS AND SYSTEMS FOR RECONSTRUCTION OF DEVELOPMENTAL LANDSCAPES BY OPTIMAL TRANSPORT ANALYSIS

CROSS-REFERENCE TO RELATED APPLICATIONS

[0001] This application claims the benefit of U.S. Provisional Application Nos. 62/560,674, filed September 19, 2017 and 62/561,047, filed September 20, 2017. The entire contents of the above-identified applications are hereby fully incorporated herein by reference.

TECHNICAL FIELD

[0002] The subject matter disclosed herein is generally directed to methods and systems for analyzing the fates and origins of cells along developmental trajectories using optimal transport analysis of single-cell RNA-seq information over a given time course.

BACKGROUND

[0003] In the mid-20th century, Waddington introduced two images to describe cellular differentiation during development: first, trains moving along branching railroad tracks and, later, marbles following probabilistic trajectories as they roll through a developmental landscape of ridges and valleys (1, 2). These metaphors have powerfully shaped biological thinking in the ensuing decades. The recent advent of massively parallel single-cell RNA sequencing (scRNA-Seq) (3-7) now offers the prospect of empirically reconstructing and studying the actual "landscapes", "fates" and "trajectories" associated with complex processes of cellular differentiation and de-differentiation—such as organismal development, long-term physiological responses, and induced reprogramming—based on snapshots of expression profiles from heterogeneous cell populations undergoing dynamic transitions (6-1 1).

[0004] To understand such processes in detail, general approaches are needed to answer key questions. For any given system, we would like to know: What classes of cells are present at each stage? For the cells in each class, what was their origin at earlier stages, what are their potential fates at later stages, and what is the actual outcome of a given cell? To what extent are events along a path synchronous or asynchronous? What are the genetic regulatory programs that control each path? What are the intercellular interactions between classes of cells? Answering these questions would provide insights into the nature of developmental processes: How deterministic or stochastic is the process—that is: if, and how early, does it become determined that a particular cell or an entire cell class is destined to a specific fate? For a given origin and target fate, is there only a single path to the target, or are there multiple developmental paths? To what extent is the process cell-intrinsic, driven by intracellular mechanisms that do not require ongoing external inputs, or externally regulated, being affected by other contemporaneous cells? For artificial processes such as induced reprogramming, there are additional questions: What off-target cell classes arise? To what extent do cells activate normal developmental programs vs. unnatural hybrid programs? How can the efficiency of reprogramming be improved?

[0005] Experimental approaches to such questions have typically involved studying bulk populations or identifying subsets of cells based on activation of one or a few genes at a specific time (e.g., reporter genes or cell-surface markers) and tracing their subsequent fate. These experiments are severely limited, however, by the need to choose subsets of cells a priori and develop distinct reagents to study each subset. For example, studies of cellular reprogramming from fibroblasts to induced pluripotent cells (iPSCs) have largely relied on RNA- and chromatin-profiling studies of bulk cell populations, together with fate-tracing of cells based on a limited set of markers (e.g., Thyl and CD44 as markers of the fibroblast state, and ICAM1, Oct4, and Nanog as markers of partial reprogramming) (12-16).

[0006] Computational approaches based on single-cell gene expression profiles offer a complementary approach with broader molecular scope, because one can readily define classes of cells based on any expression profile at any stage. The remaining challenge is to reliably infer their trajectories across stages.

[0007] Several pioneering papers have introduced methods to infer cellular trajectories (9, 10, 17-29). Early studies recognized that cellular profiles from heterogeneous populations can provide information about the temporal order of asynchronous processes—enabling intermediate transitional cells to be ordered in "pseudotime" along "trajectories", based on their state of cell differentiation (18). Some approaches relied on k-nearest neighbor graphs (18) or binary trees (9). More recently, diffusion maps have been used to order cell state transitions. In this case, single-cell profiles are assigned to densely populated paths through diffusion map space (20, 21). Each such path is interpreted as a transition between cellular fates, with trajectories determined by curve fitting, and cells "pseudotemporally ordered" based on the diffusion distance to the endpoints of each path. Whereas initial efforts focused mostly on single paths, more recent work has grappled with challenges of branching, which is critical for understanding developmental decisions (10, 11, 21).

[0008] While these pioneering approaches have shed important light on various biological systems, many important challenges remain. First, because many methods were initially designed to extract information about stationary processes (such as the cell cycle or adult stem cell differentiation) in which all stages exist simultaneously, they neither directly model nor explicitly leverage the temporal information in a developmental time course *(29).* Second, a single cell can undergo multiple temporal processes at once. These processes can dramatically impact the performance of these models, with a notable example being the impact of cell proliferation and death *(29).* Third, many of the methods impose strong structural constraints on the model, such as one-dimensional trajectories and zero-dimensional branch points. This is of particular concern if development follows the flexible "marble" rather than the regimented "tracks" models, in Waddington's frameworks.

SUMMARY

[0009] In one aspect, the present disclosure includes a method of producing induced pluripotent stem cell comprising introducing a nucleic acid encoding Obox6 into a target cell to produce an induced pluripotent stem cell. In some embodiments, the methods further comprises introducing into the target cell at least one nucleic acid encoding a reprogramming factor selected from the group consisting of: Gdf9, Oct3/4, Sox2, Soxl, Sox3, Soxl5, Soxl7, Klf4, Klf2, c-Myc, N-Myc, L-Myc, Nanog, Lin28, Fbxl5, ERas, ECAT15-2, Tell, beta-catenin, Lin28b, Sall l, Sall4, Esrrb, Nr5a2, Tbx3, and Glisl. In some embodiments, the method further comprises introducing into the target cell at least one nucleic acid encoding a reprogramming factor selected from the group consisting of: Oct4, Klf4, Sox2 and Myc. In some embodiments, the nucleic acid encoding Obox6 is provided in a recombinant vector. In some embodiments, the vector is a lentivirus vector. In some embodiments, the nucleic acid encoding the reprogramming factor is provided in a recombinant vector. In some embodiments, the method further comprises a step of culturing the cells in reprogramming medium. In some embodiments, the method further comprises a step of culturing the cells in the presence of serum. In some embodiments, the method further comprises a step of culturing the cells in the absence of serum. In some embodiments, the induced pluripotent stem cell expresses at least one of a surface marker selected from the group consisting of: Oct4, SOX2, KLf4, c-MYC, LIN28, Nanog, Glisl , TRA-160/TRA-1-81/TRA-2-54, SSEA1, SSEA4, Sal4, and Esrbbl. In some embodiments, the target cell is a mammalian cell. In some embodiments, the target cell is a human cell or a murine cell. In some embodiments, the target cell is a mouse embryonic fibroblast. In some embodiments, the target cell is selected from the group consisting of: fibroblasts, B cells, T cells, dendritic cells, keratinocytes, adipose cells, epithelial cells, epidermal cells, chondrocytes, cumulus cells, neural cells, glial cells, astrocytes, cardiac cells, esophageal cells, muscle cells, melanocytes, hematopoietic cells, pancreatic cells, hepatocytes, macrophages, monocytes, mononuclear cells, and gastric cells, including gastric epithelial cells.

[0010] In another aspect, the present disclosure includes a method of producing an induced pluripotent stem cell comprising introducing at least one of Obox6, Spic, Zfp42, Sox2, Mybl2, Msc, Nanog, Hesxl and Esrrb into a target cell to produce an induced pluripotent stem cell.

[0011] In another aspect, the present disclosure includes a method of producing an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell to produce an induced pluripotent stem cell.

[0012] In another aspect, the present disclosure includes a method of increasing the efficiency of production of an induced pluripotent stem cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell.

[0013] In another aspect, the present disclosure includes a method of increasing the efficiency of production of an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell to produce an induced pluripotent stem cell.

[0014] In another aspect, the present disclosure includes an isolated induced pluripotential stem cell produced by the methods disclosed herein.

[0015] In another aspect, the present disclosure includes a method of treating a subject with a disease comprising administering to the subject a cell produced by differentiation of the induced pluripotent stem cell produced by the methods disclosed herein.

[0016] In another aspect, the present disclosure includes a composition for producing an induced pluripotent stem cell comprising Obox6 in combination with reprogramming medium.

[0017] In another aspect, the present disclosure includes a composition for producing an induced pluripotent stem cell comprising one or more of the factors identified in or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6 in combination with reprogramming medium.

[0018] In another aspect, the present disclosure includes use of Obox6 for production of an induced pluripotent stem cell.

[0019] In another aspect, the present disclosure includes use of a factor identified in or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6 for production of an induced pluripotent stem cell.

[0020] In another aspect, the present disclosure includes a method of increasing the efficiency of reprogramming a cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell.

[0021] In another aspect, the present disclosure includes a method of increasing the efficiency of reprogramming a cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5 and Table 6, into a target cell to produce an induced pluripotent stem cell.

[0022] In another aspect, the present disclosure includes a computer-implemented method for mapping developmental trajectories of cells, comprising: generating, using one or more computing devices, optimal transport maps for a set of cells from single cell sequencing data obtained over a defined time course; determining, using one or more computing devices, cell regulatory models, and optionally identifying local biomarker enrichment, based on at least the generated optimal transport maps; defining, using the one or more computing devices, gene modules; and generating, using the one or more computing devices, a visualization of a developmental landscape of the set of cells.

[0023] In some embodiments, determining cell regulatory models comprise sampling pairs of cells at a first time and a second time point according to transport probabilities. In some embodiments, the method further comprises using the expression levels of transcription factors at the earlier time point to predict non-transcription factor expression at the second time point. In some embodiments, identifying local biomarker enrichment comprises identifying transcription factors enriched in cells having a defined percentage of descendants in a target cell population. In some embodiments, the defined percentage is at least 50% of mass. In some embodiments, defining gene modules comprises partitioning genes based on correlated gene expression across cells and clusters. In some embodiments, partitioning comprises partitioning cells based on graph clustering. In some embodiments, graph clustering further comprises dimensionality reduction using diffusion maps. In some embodiments, the visualization of the developmental landscape comprises high-dimensional gene expression data in two dimensions. In some embodiments, the visualization is generated using force-directed layout embedding (FLE). In some embodiments, the visualization provides one or more cell types, cell ancestors, cell descendants, cell trajectories, gene modules, and cell clusters from the single cell sequencing data.

[0024] In another aspect, the present disclosure includes a computer program product, comprising: a non-transitory computer-executable storage device having computer-readable program instructions embodied thereon that when executed by a computer cause the computer to execute the methods disclosed herein.

[0025] In another aspect, the present disclosure includes a system comprising: a storage device; and a processor communicatively coupled to the storage device, wherein the processor executes application code instructions that are stored in the storage device and that cause the system to executed the methods disclosed herein.

[0026] In another aspect, the present disclosure includes a method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Gdf9 into a target cell to produce an induced pluripotent stem cell.

[0027] These and other aspects, objects, features, and advantages of the example embodiments will become apparent to those having ordinary skill in the art upon consideration of the following detailed description of illustrated example embodiments.

BRIEF DESCRIPTION OF THE DRAWINGS

[0028] An understanding of the features and advantages of the present invention will be obtained by reference to the following detailed description that sets forth illustrative embodiments, in which the principles of the invention may be utilized, and the accompanying drawings of which:

[0029] **FIG. 1** - is a block diagram depicting a system for mapping developmental trajectories of cells, in accordance with certain example embodiments

[0030] **FIG. 2** - is a block flow diagram depicting a method for mapping development trajectories of cells, in accordance with certain example embodiments.

[0031] **FIG. 3** - is a diagram showing data S_i from a generic branching developmental process. The x-axis represents the time and the y-axis represents expression.

[0032] **FIG. 4** - provides a schematic of a regulatory vector file which gives rise to a time-dependent probability distribution.

[0033] **FIGs. 5A-5G - (FIGs. 5A-5B)** Waddington's classical analogies of cells undergoing differentiation, initially (1936) illustrated by railroad cars on switching tracks **(FIG. 5A)** and later (1957) by marbles rolling in a landscape **(FIG. 5B),** with trajectories shaped by hills and valleys. **(FIGs. 5C-E)** Differentiation processes in which the ultimate fate of individual cells (filled dots) is **(C)** predetermined **(FIG. 5D)** not predetermined, or **(FIG. 5E)** progressively determined. Arrows indicate possible transitions, and color represents cell fate, with red and blue indicating distinct fates, light red and light blue indicating partially determined fates, and grey indicating undetermined fate. **(FIG. 5F)** Illustration of transported mass. A transport map, , describes how a point x at one stage (X) is redistributed across all points (denoted by "") at the subsequent stage (Y). **(FIG. 5G)** Transport maps computed from a time series of samples taken from a time-varying distribution. Between each pair of time points, a transport map redistributes the cells observed at time to match the distribution of cells observed at time.

[0034] **FIGs. 6A-6C - (FIG. 6A)** Representation of reprogramming procedure and time points of sample collection. (Top) Mouse embryos (E13.5) were dissected to obtain secondary MEFs (2° MEF), which were reprogrammed into iPSCs. In Phase-1 of reprogramming (light blue; days 0-8), doxycycline (Dox) was added to the media to induce ectopic expression of reprogramming factors *(Oct4, Kl/4, Sox2,* and *Myc).* In Phase-2 (days 9-16), Dox was withdrawn from the media, and cells were grown either in the presence of 2i (light red) or serum (light green). Samples were also collected from established iPSC lines reprogrammed from the same 2° MEFs, maintained in either 2i or serum conditions (far right in each time course). Individual dots along the time course indicate time points of scRNA-Seq collection, with two dots indicating biological replicates. **(FIG. 6B)** Number of scRNA-Seq profiles from each sample collection that passed quality control filters. **(FIG. 6C)** Bright field images of day 0 (Phase l-(Dox)) and day 16 cells during reprogramming in (Phase-2(2i)) and (Phase-2(serum)) culture conditions.

[0035] **FIGs. 7A-7F** - scRNA-Seq profiles of all 65,781 cells were embedded in two-dimensional space using FLE, and annotated with indicated features. **(FIG. 7A)** Unannotated layout of all cells. Each dot represents one cell. **(FIGs. 7B-7C)** Annotation by time point (color) and biological feature, with Phase-2 points from either **(FIG. 7B)** 2i condition or **(FIG. 7C)** serum condition. Phase-1 points appear in both **(FIG. 7B)** and **(FIG. 7C).** Individual cells are colored by day of collection, with grey points (BC, background color) representing Phase-2 cells from serum (in **FIG. 7B)** or 2i (in **FIG. 7C). (FIG. 7D)** Annotation by cell cluster. Cells were clustered on the basis of similarity in gene expression. Each cell is colored by cluster membership (with clusters numbered 1-33). **(FIGs. 7E-7F)** Annotation by gene signature **(FIG. 7E)** and individual gene expression levels **(FIG. 7F).** Individual cells are colored by gene signature scores (in **FIG. 7E)** or normalized expression levels (in **FIG. 7F;** , where E is the number of transcripts of a gene per 10,000 total transcripts).

[0036] **FIGs. 8A-8F - (FIG. 8A)** Schematic representation of the major cluster-to-cluster transitions (see Table 10 for details[BC17]). Individual arrows indicate transport from ancestral clusters to descendant clusters, with colors corresponding to the ancestral cluster. For each descendant cluster, arrows were drawn when at least 20% of the ancestral cells (at the previous time point) were contained within a given cluster (self-loops not shown). Arrow thickness indicates the proportion of ancestors arising from a given cluster. **(FIG. 8B)** Heatmap depiction of cluster descendants in 2i condition. In each row of the heatmap, color intensity indicates the number of descendant cells ("mass", normalized to a starting population of 100 cells) transported to each cluster at the subsequent time point (see Table 10 for details). Clusters with highly-proliferative cells *(e.g.,* cluster 4) transport more total mass than clusters with lowly-proliferative cells *(e.g.,* cluster 14). **((FIG. 8C)** Depiction of divergent day 8 descendant distributions for two clusters of cells at day 2 (cluster 4 (left) and cluster 6 (right). Color intensity indicates the distribution of descendants at day 8, with bright teal indicating high probability fates and gray indicating low probability fates. **(FIG. 8D)** Enrichment of the ancestral distributions of iPSCs, Valley of Stress, and alternative fates (neuron-like and placenta-like) in clusters of day 2 cells. The red horizontal dashed line indicates a null-enrichment, where a cluster contributes to the ancestral distribution in proportion to its size. Cluster 4 has a net positive enrichment because its descendants are highly proliferative, while cluster 6 has a net negative enrichment because its descendants are lowly proliferative. **(FIG. 8E)** and **(FIG. 8F)** Ancestral trajectories of indicated populations of cells at day 16 (iPSCs, placental, neural-like cells, *etc)* in serum **(FIG. 8E)** and 2i **(FIG. 8F).** Clusters used to define the indicated populations are shown in parentheses. Colors indicate time point. Sizes of points and intensity of colors indicate ancestral distribution probabilities by day (color bars, right; BC, background color, representing cells from the other culture condition).

[0037] **FIGs. 9A-9D - (FIG. 9A)** Classification of genes into 14 groups based on similar temporal expression profiles along the trajectory to successful reprogramming. Averaged gene expression profiles for each group, in 2i and serum conditions (left). Heatmap for genes within each group, with intensity of color indicating log2-fold change in expression relative to day 0 (middle). Representative genes and top terms from gene-set enrichment analysis for each group (right). **(FIG. 9B)** Comparison of FACS and *in* silico sorting experiments. Scatterplot shows reprogramming efficiencies determined by FACS sort and growth experiments (blue triangles) *(16)* and our computationally inferred trajectories (red squares). The specific cell surface markers used for the *in silico* and experimental methods are indicated. Reprogramming efficiencies for these categories (calculated both experimentally and *in silico)* are normalized to the percentage of EGFP+ colonies in CD44$^-$ ICAM1$^+$ Nanog$^+$ condition (details found in *Appendix 5).* **(FIG. 9C)** Schematic of regulatory model in which TF expression in ancestral cells is predictive of gene expression in descendant cells. **(FIG. 9D)** Onset of iPSC-associated TFs in 2i (left) and serum (right). (Top) Mean expression levels weighted by iPSC ancestral distribution probabilities (Y axis) of *Nanog, Obox6,* and *Sox2* at each day (X axis). (Bottom) Normalized expression of TF modules "A" and "B" from our regulatory model (as in **FIG. 9B)** that were associated with gene expression in iPSCs.

[0038] **FIGs. lOA-lOC - (FIGs. 10A-10B)** Bright field and fluorescence images of iPSC colonies generated by lentiviral overexpression of *Oct4, Kl/4, Sox2,* and *Myc* (OKSM) with either an empty control, *Zfp 42* or *Obox6* expression cassette, in either Phase-l(Dox)/Phase-2(2i)

(FIG. 10A) and Phase- l(Dox)/Phase-2(serum) **(FIG. 10B)** conditions (indicated). Cells were imaged at day 16 to measure Oct4-EGFP $^+$ cells. Bar plots representing average percentage of Oct4-EGFP $^+$ colonies in each condition on day 16 are included below the images. Shown are data from one of five independent experiments, with three biological replicates each. Error bars represent standard deviation for the three biological replicates. **(FIG. IOC)** Schematic of the overall reprogramming landscape highlighting: the progression of the successful reprogramming trajectory, alternative cell lineages, and specific transition states (Horn of Transformation). Also highlighted are transcription factors (orange) predicted to play a role in the induction and maintenance of indicated cellular states, and putative cell-cell interactions between contemporaneous cells in the reprogramming system.

[0039] **FIGs. 11A-11D** - Single-cell RNA-Seq quality metrics. **(FIG. 11**A) Correlation between number of genes and tran- scripts per cell (loglO transformed). Cells with fewer than 1000 genes detected were filtered out. The color gradient represents cell density. **(FIG. 11B)** Variation in single cell data depicted by correlation between transcript levels (loglO transformed average transcript counts) detected in biological replicates generated from day 10 samples in 2i conditions. Pearson correlation coefficient (r) is given. The color gradient represents cell density. **(FIG. 11C)** Biological variation in single cell data depicted by correlation between tran- script levels (loglO transformed average transcript counts) detected in iPSCs and MEFs. Pearson correlation coefficient (r) is given. The color gradient represents cell density. **(FIG. 11D)** Correlogram visualizing correlation between single cell gene expression profiles between various time points and their biological replicates. In this plot, the correlation coefficients (circles) are colored according to their values, ranging from 0.75 (blue) to 1 (red). The size of the circles represents the magnitude of the coefficient. The replicates within the timepoints are denoted with suffixes 1 and 2.

[0040] **FIGs. 12A-12C** - Comparison of various dimensionality reduction methods to visualize single cell RNA- Seq data. High-dimensional structure of single-cell expression data was embedded in low-dimensional space for visualization using **(FIG. 12A)** the Force-directed Layout Embedding algorithm (FLE) (directed graph approach) and the t-Distributed Stochastic Neighbor Embedding algorithm (t-SNE) with **(FIG. 12B)** principal components and **(FIG. 12C)** diffusion maps as input parameters.

[0041] **FIG. 13** - Visualization of gene modules across reprogramming time points. Expression profiles of all 65,781 cells studied were embedded in two-dimensional space, using force-directed layout embed- ding (FLE). The layouts were annotated by single-cell z-scores for 44 gene modules (details in *Table 1).* The color gradient represents the distribution of z-scores across all cells for a given gene module.

[0042] **FIGs. 14A-14B** - Characterization of cell clusters. **(FIG. 14A)** Heatmap representing the enrichment of cells from the indicated samples at various time points and culture conditions across 33 different clusters. The color gradient represents the range of cell fractions from 0-0.25. **(FIG. 14B)** Heatmap depicting the enrichment of correlated gene modules within specific cell clusters. The color gradient represents the average gene module scores at the indicated cell clusters. Specific cell clusters that show highly correlated gene module scores were numerically labeled as shown

[0043] **FIG. 15** - Visualization of individual gene expression levels.Normalized expression levels [log2(E+l)] for indicated genes were used to annotate force-directed layout embedding (FLE) graphs generated from the expression profiles of 65,781 cells. E represents the number of transcripts of a gene per 10,000 total transcripts

[0044] **FIGs. 16A-16E** - Distribution of gene signatures. **(FIG. 16A)** Distribution of proliferation scores for cells at day 0 (solid black). Proliferation scores were calculated from combined expression levels of Gl/S and G2/M cell cycle genes (see *Appendix 5).* Normal mixture modeling (dashed line) was used to classify the cells based on proliferation scores into non-cycling (red) and cycling (blue) cells (top). Visualization of the cycling and non-cycling of cells on FLE at day 0 (bottom). **(FIG. 16B)** Violin plots of single-cell scores for indicated gene signatures and Shisa8 expression levels in clusters 3, 4, 5, and 6. **(FIG. 16C)** Violin plots of single cell scores for indicated gene signatures in clusters 7, 8, and 18. **(FIG. 16D)** Bar plots of normalized expression levels [log2(E+l)] for indicated genes, where E is the number of transcripts of a gene per 10,000 total transcripts. **(FIG. 16E)** Single-cell scores for indicated gene signatures across all 33 cell clusters.

[0045] **FIGs. 17A-17C** - Heatmap depiction of origins and fates of cells inferred from optimal transport. Heatmap depiction of cluster descendants in **(FIG. 17A)** serum condition, and cluster ancestors in **(FIG. 17B)** 2i and **(FIG. 17C)** serum conditions. Each row of the heatmap in

(FIG. 17A) shows how the descendants of the cells in a particular cluster are distributed over all clusters. Color intensity indicates the number of descendant cells ("mass", normalized to a starting population of 100 cells) transported to each cluster at the next time point. Each column of the heatmaps in **(FIG. 17B, FIG. 17C)** shows how the ancestors of a particular cluster are distributed over all clusters. *Table 10* contains the specific numerical values.

[0046] **FIGs. 18A-18F** - Potential cell-cell interactions across the reprogramming time course. **(FIG. 18A)** Temporal pattern of the net potential for paracrine signaling between contemporaneous cells. Each dot represents the aggregated interaction score across all ligand-receptor pairs for a given combination of clusters (all 149 detected ligands). The aggregate interaction score is defined as a sum of individual interaction scores. **(FIG. 18B)** As in A, but genes specific to SASP signature are considered (20 detected ligands). **(FIG. 18C)** Heatmap representing the aggregate interaction scores on day 16 cells in 2i condition for ligands specific to SASP signature. Rows correspond to clusters of cells expressing ligands. Columns correspond to clusters of cells expressing cognate receptors. Only clusters containing more than 1% of cells from day 16 (2i) are shown. **(FIGs. 18D-18F)** Potential ligand-receptor pairs ranked by their standardized interaction scores calculated from the permuted data (see *Appendix 5* for details). Ligand-receptor pairs between **(FIG. 18D)** valley of stress cells (clusters 11-17) and iPSCs (clusters 28-33) on day 16 (2i), **(FIG. 18E)** valley of stress cells and preneural/neural-like cells (clusters 23, 26, and 27) on day 16 (serum), and **(FIG. 18F)** placental-like cells (clusters 24 and 25) and valley of stress cells on day 12 (2i)

[0047] **FIGs. 19A-19F -** Gene modules and associated transcription factors based on optimal transport. Using optimal transport trajectories, TF levels in cells at time t are used to predict the activity levels of gene modules in descendant cells at time t + 1. Gene modules are learned during model training to capture coherent expression programs. For five modules **(FIGs. 19A-19E),** bar plots depict the top 50 genes in the module (black), and the top 20 TFs each associated with positive (red) and negative (blue) module activity. **(FIGs. 19A- 19B)** Two modules that are active in cells with placental identity. **(FIG. 19C)** A module active in cells with neural identity. **(FIG. 19D-19E)** Two modules active in successfully reprogrammed cells. **(FIG. 19F)** Enrichment analysis of TFs in day 12 cells with high (>80%) vs. low (<20%) probability of successful reprogramming. Dot size and color represent percentage of day 12 cells expressing the indicated TF in high- or low-probability cells. Bar heights indicate the fold enrichment in high- vs. low-probability cells.

[0048] **FIGs. 20A-20C** - Effect of overexpression of *Obox6* and *Zpf42* on reprogramming efficiency. **(FIG. 20A)** Percentage of Oct4-EGFP+ cells at day 16 of reprogramming from secondary MEFs by lentiviral overexpression of *Oct4, Kl/4, Sox2,* and *Myc* (OKSM) combined with either *Zfp42*, *Obox6,* or an empty control, in either 2i or serum conditions. Oct4-EGFP+ cells were measured by flow cytometry. Plot includes the percentage of Oct4-EGFP+ cells in three biological replicates (for *Zfp42* and *Obox6* overexpression, or an empty control) from five independent experiments (Exp). **(FIG. 20B, FIG. 20C)** Number of Oct4-EGFP+ colonies at day 16 of reprogramming from primary MEFs by lentiviral overexpression of individual *Oct4, Kl/4, Sox2,* and *Myc* combined with either *Zfp42*, *Obox6,* or an empty control in **(FIG. 20B)** 2i and **(FIG. 20C)** serum conditions. Plot includes the number of Oct4-EGFP+ cells in three biological replicates (for *Zfp42* and *Obox6* overexpression, or an empty control) from two independent experiments (Exp).

[0049] **FIGs. 21A-21E** - X-chromosome reactivation. **(FIGs. 21A**-21C) Boxplots showing X/Autosome expression ratio (left panel) and Xist expression log2(E+l) across individual cells by clusters (right panel): **(FIG. 21A)** all cells, **(FIG. 21B)** phase-l(Dox) and phase-2(2i) cells, **(FIG. 21C)** phase-l(Dox) and phase-2(serum) cells. **(FIGs. 21D-21F)** - X/Autosome expression ratio and A6, A7 activation pattern changes along the successful trajectory determined by optimal transport: Relative gene expression changes of individual genes from A6 **(FIG. 21D)** and A7 **(FIG. 21E)** activation patterns (gray solid lines). Black and blue solid lines correspond to average relative expression of genes and average X/Autosome expression ratios, respectively. **(FIG. 21F)** Comparison between activation of A6 and A7 programs (average relative expression) with X/Autosome expression ratio. Distribution of X/Autosome expression ratios **(FIG. 21G)** and A7 scores **(FIG. 21H)** across all cells. Dotted lines represent threshold values used in classification of cells that reactivated X-chromosome (> 1.4) and upregulated A7 genes (> 0.25).

[0050] **FIGs. 22A-22C** - Single-cell expression levels were used to identify cells with aberrant expression in large chromosomal regions. **(FIG. 22A)** Whole chromosome aberrations were detected in 1% of all cells. Each dot represents one chromosome (X axis) in a single cell with significant aberrations (FDR 10%), with violin plots capturing the distributions of dots. The net expression of these chromosomes relative to the average expression across all cells (Y axis) is 1.7-fold higher (median, left panel) and 2.2-fold lower (right panel), indicating whole chromosome gain and loss, respectively. The median relative expression levels are slightly higher (lower) than the 1.5-fold (2-fold) increase (decrease) that would be expected from a true chromosomal gain (loss) because our statistics are conservative in calling significant events but allow for a long tail of high (low) expression. **(FIG. 22B)** Visualization of cells with significant subchromosomal aberrations (red) in FLE. **(FIG. 22C)** Bar plots depict the fraction of cells in each cluster with significant subchromosomal (25-200Mbp) aberrations (FDR 10%).

[0051] **FIGs. 23A-23F** - Modeling developmental processes with optimal transport. Waddington-OT: a probabilistic model for developmental processes. **(FIG. 23A)** A temporal progression of a time-varying distribution $\mathbf{P}_t$ (left) can be sampled to obtain finite empirical distributions of cells $\hat{\mathbf{P}}_{t_i}$ at various time points t_1, t_2, t_3 (right). Over short time scales, the unknown true coupling, $Y_{tl,}t_2$, is assumed to be close to the optimal transport coupling, π_{t_1,t_2}, which can be approximated by $\pi_{tl,t2}$ computed from the empirical distributions $\hat{\mathbf{P}}_{t_1}$and $\hat{\mathbf{P}}_{t_2}$. **(FIGs. 23B-23F)** Simulated data and analysis performed by Waddington-OT. **(FIG. 23B)** Single-cell profiles (individual dots) are embedded in two dimensions and colored by the time of collection. Optimal transport can be used to calculate the descendant trajectories **(FIG. 23C)** and ancestor trajectories **(FIG. 23D)** of any subpopulation of interest (cells highlighted in black; color indicates time). Ancestor distributions of distinct subpopulations can be compared to calculate their shared ancestry **(FIG. 23E)** (ancestors of each population shown in red and blue, shared ancestors in purple). **(FIG. 23F)** The expression of gene signatures (left; green, high expression; grey, low expression) can be predicted from the earlier expression of transcription factors (middle; black, high expression; grey, low expression) in a gene regulatory model by analyzing trends along ancestor trajectories. In the plot at right, at each time point, the height of the curve depicts the average expression in the ancestors of cells in the leftmost tip.

[0052] **FIGs. 24A-24H** - A single cell RNA-Seq time course of iPSC reprogramming. **(FIG. 24A)** Representation of reprogramming procedure and time points of sample collection. (Top) Mouse embryos (E13.5) were dissected to obtain secondary MEFs (2° MEF), which were reprogrammed into iPSCs. In Phase-1 of reprogramming (light blue; days 0-8), doxycycline

(Dox) was added to the media to induce ectopic expression of reprogramming factors *(Oct4, Kl/4, Sox2,* and *Myc).* In Phase-2 (days 9-18), Dox was withdrawn from the media, and cells were grown either in the presence of 2i (light red) or serum (light green). Samples were also collected from established iPSC lines reprogrammed from the same 2° MEFs, maintained in either 2i or serum conditions (far right in each time course). Individual dots indicate time points of scRNA-Seq collection. **(FIGs. 24B-24E)** scRNA-Seq profiles of all 251,203 cells (individual dots) were embedded in two-dimensional space using FLE, and annotated with indicated features. **(FIG. 24B)** Unannotated layout of all cells, with the density of cells in each region indicated by intensity. **(FIG. 24C)** Cells colored by time point, with Phase-2 points from either 2i condition (left) or serum condition (right). Phase-1 points appear in both subplots. Grey points represent Phase-2 cells from the other condition. **(FIG. 24D)** In different regions of the FLE, cells have distinct expression patterns of six major gene signatures (average expression z-score of genes in a signature indicated by red color bar). Gene signature activity and trajectory analysis were used to define the major cell sets **(FIG. 24E)** and to establish the overall flow through the landscape **(FIG. 24F)** (schematic representation). **(FIG. 24G)** The relative abundance (y-axis) of each cell set (colored lines) is plotted over time (x-axis) in 2i (top) and serum (bottom). **(FIG. 24H)** Validation via geodesic interpolation in serum condition. Data at withheld timepoints (x-axis) are interpolated using data at the neighboring timepoints. Interpolation is done using a null estimator of independent coupling (blue) and the optimal transport coupling (red), with the distance between interpolated and withheld data indicated on the y-axis. The distance between two batches of withheld data at the same point is shown in green. Shaded regions indicate standard deviations over independent samples of the coupling map.

[0053] **FIGs. 25A-25H** - In initial stages of reprogramming, cells progress toward stromal or MET fates. **(FIG. 25A)** Cells in the stromal region have higher expression of gene signatures (red color bar, average z-score) and individual genes (red color bar, log(TPM+l)) that are associated with stromal activity and senescence. Ancestors of day 18 stromal cells are visualized on the FLE **(FIG. 25B)** (colored by day, intensity indicates probability), and expression trends along this ancestor trajectory **(FIG. 25C)** are depicted for gene signatures (left) and individual transcription factors (TFs; right). The ancestors of day 8 MET cells **(FIG. 25D)** have a distinct trajectory and gene signature trends **(FIG. 25E),** and show differential expression of several TFs

(FIG. 25F) (dashed line, average TPM in stromal ancestors; solid line, average TPM in MET ancestors). **(FIG. 25G, FIG. 25H)** The MET and stromal fates are gradually specified from day 0 through 8. Color bar in **(FIG. 25G)** indicates log-likelihood of obtaining stromal vs. MET fate. **(FIG. 25H)** The extent to which the stromal ancestor distribution has diverged (y-axis) from all other fates at each point in time (x-axis). The divergence is quantified as ½ times the total variation distance between the ancestor distributions.

[0054] **FIGs. 26A-26F** - iPSCs emerge from cells in the MET Region. **(FIG. 26A)** Ancestors of day 18 iPSCs in 2i (left) and serum (right) are visualized on the FLE (colored by day, intensity indicates probability). Cells in the iPSC region express pluripotency marker genes **(FIG. 26B)** (red color bar, log(TPM+l)) and diverge from alternative fates also arising from the MET region (neural, epithelial, and trophoblast) from days 8-12 **(FIG. 26C)** (divergence between pairs of lineages indicated by individual lines; green line, divergence between iPSC and all others). **(FIG. 26D)** Expression trends along the ancestor trajectory in serum are depicted for gene signatures (left) and individual transcription factors (right). **(FIG. 26E)** A signature of X reactivation (left; red color bar, average z-score) and *Xist* expression (right; log(TPM + 1)) visualized on the FLE. **(FIG. 26F)** Trends in X-inactivation, X-reactivation and pluripotency along the iPSC trajectory in 2i. The values on the axis refer to average expression across early (black) and late (red) pluripotency activation genes, *Xist* average expression (log(TPM+l), orange) and X/Autosome expression ratio (blue) along the iPSC trajectory.

[0055] **FIGs. 27A-27G** - Extra-embryonic and neural-like cells emerge during reprogramming. Subpopulations of trophoblast- **(FIGs. 27A-27C)** and neural-like **(FIGs. 27D-27G)** cells are found in the late stages of reprogramming. Ancestors of day 18 trophoblasts are visualized on the FLE **(FIG. 27A)** (colored by day, intensity indicates probability), and expression trends along the ancestor trajectory in serum **(FIG. 27B)** are depicted for gene signatures (left) and individual transcription factors (right). **(FIG. 27C)** Cells in the trophoblast cell set were re-embedded by FLE, and scored for signatures of trophoblast progenitors (TP), spiral artery trophoblast giant cells (SpA-TGC), and spongiotrophoblasts (SpTB). Colors indicate significant expression of TP, SpA-TGC, and SpTB signatures (4ogl0(FDR q-value)), or expression of labyrinthine trophoblast marker gene *Gcml* (red color bar, log(TPM + 1)). Ancestors of day 18 cells in the neural region are visualized on the FLE **(FIG. 27D)**

(colored by day, intensity indicates probability), and expression trends along the ancestor trajectory in serum **(FIG. 27E)** are depicted for gene signatures (left) and individual transcription factors (right). **(FIG. 27F)** Cells with radial glial (RG) and differentiated subtype signatures begin to appear around day 12 (x-axis, time; y-axis, relative abundance in serum). **(FIG. 27G)** All cells in the neural region we re-embedded by FLE, and scored for significant expression of differentiated signatures (OPC, astrocyte, cortical neurons; color, -loglO(FDR q-value)), or annotated by expression of markers of inhibitory and excitatory neurons (red color bars, log(TPM + 1)). OPC, oligodendrocyte precursor cells.

[0056] **FIGs. 28A-28K** - Paracrine signaling and genomic aberrations. **(FIG. 28A)** Schematic of the paracrine signaling interaction scores. High potential interaction occurs between two groups of contemporaneous cells in which one group secretes a ligand and a second group expresses a cognate receptor. **(FIG. 28B)** Temporal pattern of the net potential for paracrine signaling between contemporaneous cells in serum condition. Each dot represents the aggregated interaction score across all ligand-receptor pairs for a given combination of clusters (Figure S5A, all 180 detected ligands). The aggregate interaction score is defined as a sum of individual interaction scores. **(FIGs. 28C-E)** Potential ligand-receptor pairs between ancestors of stromal cells and iPSCs **(FIG. 28C),** neural-like cells **(FIG. 28D),** and trophoblasts **(FIG. 28E),** ranked by their standardized interaction scores calculated from the permuted data (see STAR Methods for details). **(FIGs. 28F-H)** Individual cells on the FLE colored by the expression level (log(TPM+l)) of ligands (upper row) and receptors (lower row) for top interacting pairs between stromal cells and iPSCs **(FIG. 28F),** neural-like cells **(FIG. 28G),** and trophoblasts **(FIG. 28H).** **(FIGs. 28I-28K)** Evidence for genomic aberrations was found at the level of whole chromosomes **(I)** and sub-chromosomal regions spanning 25 housekeeping genes **(FIGs. 28J, 28K).** **(FIG. 281)** Average expression of housekeeping genes on chromosomes (numbered on x-axis) in single cells (dots with violin plots) with evidence of genomic amplification (left panel) or loss (right panel), relative to all cells without evidence of aberrations (y-axis, relative expression). **(FIG. 28J)** Individual cells on the FLE are colored by statistical significance (-logl0(q-value), colorbar) of evidence for sub-chromosomal aberrations. **(FIG. 28K)** Average expression of genes on chromosome 15 in trophoblast-like cells with evidence of a recurrent subchromosomal amplification (FDR 10%, region indicated by red lines), relative to trophoblast-like cells without evidence of amplification in this region (y-axis, relative expression).

[0057] **FIGs. 29A-29D** - *Obox6* enhances reprogramming. **(FIG. 29A)** For cells (individual dots) at each timepoint (x-axis), the log-likelihood ratio of obtaining iPSCs fate vs non iPSCs fate in 2i is depicted on the y-axis. Cells expressing *Obox6* are highlighted in red. **(FIG. 29B)** Bright field and fluorescence images of iPSC colonies generated by lentiviral overexpression of *Oct4, Klf4, Sox2,* and *Myc* (OKSM) with either an empty control, *Zfp42* or *Obox6* expression cassette, in Phase-1(Dox)/Phase-2(2i). **(FIG. 29C)** Bar plots representing average percentage of Oct4-EGFP$^+$ colonies in 2i on day 16. Data shown is one of five independent experiments, with three biological replicates each. Error bars represent standard deviation for the three biological replicates. **(FIG. 29D)** Schematic of the overall reprogramming landscape in serum highlighting: the progression of the successful reprogramming trajectory (represented in black), alternative cell lineages and subtypes within these lineages (Stromal in blue, trophoblast-like in red, neural in green and epithelial in orange), and specific transition states (MET in purple). Also highlighted are transcription factors predicted to play a role in the transition to indicated cellular states (as indicated by the specific color), and putative cell-cell interactions between contemporaneous cells in the reprogramming system. i and e Neurons refers to inhibitory and excitatory neurons respectively.

[0058] **FIGs. 30A-30G** - Related to **FIGs. 24A-24H:** Validation, stability, and comparison to pilot study. **(FIGs. 30A-30C)** Unbalanced transport can be used to tune growth rates. **(FIG. 30A)** When the unbalanced regularization parameter is large (=16), growth constraints are imposed strictly, and the input growth (x-axis; determined by gene signatures — see STAR Methods) is well-correlated to the output growth (y-axis; implicit growth rate determined from the transport map). **(FIG. 30B)** When the unbalanced parameter is small (=1), the growth constraints are only loosely imposed, allowing implicit growth rates to adjust and better fit the data. **(FIG. 30C)** The correlation of output vs input growth as a function of . **(FIG. 30D)** Validation by geodesic interpolation for 2i conditions. As in **FIG. 24H** (which shows serum), the red curve shows the performance of interpolating held-out time points with optimal transport. The green curve shows the batch-to-batch Wasserstein distance for the held-out time points, which is a measure of the baseline noise level. The blue curve shows the performance of a null model (interpolating according to the independent coupling, including growth). **(FIGs. 30E-30F)** Comparison to pilot dataset. **(FIG. 30E)** Trends in signature scores along ancestor trajectories to iPSC, Stromal, Neural, and Trophoblast cell sets. Trends for the pilot dataset are shown with open circles and trends for the large dataset are shown with solid lines. **(FIG. 30F)** Shared ancestry results for pilot dataset (solid lines) and for the larger dataset (dashed lines). **(FIG. 30G)** Bright field images of day 2 (Phase l-(Dox)), day 4 (Phase l-(dox)) and day 18 cells during reprogramming in (Phase-2(2i)) and (Phase-2(serum)) culture conditions. BF (bright field). GFP (Oct4-GFP).

[0059] **FIGs. 31A-31F** - Related to **FIGs. 25A-25H** Divergence of Stromal and MET fates during the initial stages of reprogramming. **(FIGs. 31A-31B)** Cells from the stromal region were re-embedded by FLE, and scored for signatures of long-term cultured MEFs (left) or stromal cells in the embryonic mesenchyme (right) found in the Mouse Cell Atlas **(FIG. 31A),** or from signatures derived from genes co-expressed (see STAR-Methods) with *Cxcll2*, *Ifltml,* or *Matn4* in the stromal cell set **(FIG. 31B)** (red color bars, average z-score of expression). **(FIG. 31C)** Ectopic OKSM expression levels are predictive of MET fate. The y-axis shows correlation between OKSM expression and the log-likelihood of obtaining MET fate. Color (red vs blue) distinguishes the two batches at each time point (x-axis). **(FIG. 31D)** *Fut9+* and *Shisa8+* expression patterns visualized in a fate-divergence layout. Each dot represents a single cell, colored by expression of either *Fut9* (left) or *Shisa8* (right). The x-axis shows time of collection and the y-axis shows the log-likelihood ratio of obtaining MET vs Stromal fate, as predicted by optimal transport. **(FIG. 31E)** The Stromal region is a terminal destination as evidenced by (1) the large flow of cells into the region around day 9 (green spike, first and second panels) and (2) essentially zero flow out of the region (blue curves, first and second panels). By contrast, the MET region is a transient state as evidenced by the blue curves in the right two panels showing significant transitions out of MET. **(FIG. 31F)** Day 0 MEFs (DO; black dots) we re-embedded together with cells from the stromal set (red dots) in a TSNE plot.

[0060] **FIGs. 32A-32C** - Related to **FIGs. 26A-26F:** iPSCs. **(FIG. 32A)** Cells with significant expression of 2 cell (2C), 4 cell (4C), 8 cell (8C), 16 cell (16C) and 32cell (32C) signatures at an FDR of 10% on iPSC-specific FLE. **(FIG. 32B)** Overlap between different early embryonic stages. The horizontal bars show the number of cells identified as 2C, 4C, 8C, 16C, or

32C. The vertical bars indicate the number of cells in each possible combination of these cell sets (e.g. 2C and 4C). **(FIG. 32C)** Heatmap showing trends in expression of 1479 variable genes (STAR-Methods) along the ancestor trajectory to iPSCs. Color indicates fold-change in expression relative to day 0 (white). Each row shows the mean expression trend for a single gene, where the mean is computed with respect to the ancestor distribution. Genes are clustered into groups with similar trends. Terms on the right indicate significant gene set enrichment (GSEA, all adjusted p-values < 0.01) in one of several databases (M, MSigDB; BP, GO biological process; W, WikiPathways; C, chromosome; CC, GO cellular component).

[0061] **FIGs. 33A-33E** - Related to **FIGs. 27A-27G:** Trophoblast and Neural subtypes. **(FIG. 33A)** Expression of individual marker genes (red color bars, log(TPM +1); see also Table S2) for each subtype on the trophoblast FLE (as in Figure 5C). TP, trophoblast progenitors; SpA-TGC, spiral artery trophoblast giant cells; SpTB, spongiotrophoblasts; LaTB, labyrinthine trophoblasts. **(FIG. 33B)** Cells with a gene signature of extra-embryonic endoderm (XEN) arise in a single batch on day 15.5 (red color bar, average z-score). **(FIGs. 33C-33E)** Cells in the neural region were re-embedded by tSNE and annotated with various features. **(FIG. 33C)** Marker gene expression (red color bar, log(TPM + 1)) of neural subtypes on the neural tSNE. **(FIG. 33D)** Cells with significant expression (black dots) of indicated signatures from the Allen Mouse Brain Atlas on the neural tSNE at an FDR of 10%. OPC refers to oligodendrocyte precursor cells. **(FIG. 33E)** Cells in the neural region present from days 12.5-14.5 (left) or days 17-18 (right).

[0062] **FIGs. 34A-34E** - Related to **FIGs. 28A-28K:** Temporal patterns of paracrine signaling. **(FIG. 34A)** Cell clusters determined by Louvain-Jaccard community detection algorithm. **(FIG. 34B)** Temporal pattern of the net potential for paracrine signaling between contemporaneous cells in 2i condition. Each dot represents the aggregated interaction score across all ligand-receptor pairs for a given combination of clusters from **(FIG. 34A)** (see STAR Methods for details). **(FIGs. 34C-34E)** Changes in the standardized interaction scores for top ligand-receptor pairs between ancestors of stromal cells and ancestors of iPSCs **(FIG. 34C),** neural-like cells **(FIG. 34D),** and trophoblast cells **(FIG. 34E).**

[0063] **FIGs. 35A-35B** - Related to **FIGs. 29A-29D:** Comparison with alternate methods. **(FIG. 35A)** Monocle2 computes a graph upon which each cell is embedded. The graph, which consists of 5 segments, is visualized in the upper-left pane. The 5 segments are visualized on our FLE in the 5 remaining panels of **(FIG. 35A).** Segment 1 (green) consists of day 0 cells together with day 18 Stromal cells. Segments 2 and 3 consist of cells from day 2 - 8 that supposedly arise from Segment 1 cells. Segment 3 gives rise to Segments 4 (purple) and 5 (red). Segment 4 contains the cells we identify as on the MET region and Segment 5 contains the iPSCs, Trophoblasts, and Neural populations, which Monocle2 infers come directly from the non-proliferative cells in segment 3. **(FIG. 35B)** URD computes a graph representing random walks from a collection of tips to a root. This graph, which consists of **7** segments, is visualized in the upper-left pane. The **7** segments are visualized on our FLE in the remaining panels of **(FIG. 35B).** Segment 1 (magenta) contains the day 0 MEF cells. The first bifurcation occurs on day 0.5, where segment 2 (consisting of day 0.5 cells) splits off from segment 3 (consisting of day 12-18 Stromal cells). Segment 2 splits to give rise to Segment 4 (consisting of day 2 cells) and Segment 5 consisting of day 12-18 Trophoblasts and Epithelial cells. Segment 4 splits on day 3 to give rise to Segment 6 (consisting of a diverse population including day 3 cells and day 14-18 iPSCs) and Segment **7** (consisting of a diverse population including day 3 cells and day 12-18 Neural-like cells).

[0064] **FIGs. 36A-36F** - Related to **FIGs. 29A-29D:** *Obox6* + *Obox6* graphs. **(FIGs. 36A-36C)** Identical to **FIGs. 29A-29C** except here we show results for serum conditions. **(FIG. 36D)** Percentage of Oct4-EGFP+ cells at day 16 of reprogramming from secondary MEFs by lentiviral overexpression of *Oct4, Kl/4, Sox2,* and *Myc* (OKSM) combined with either *Zfp42, Obox6,* or an empty control, in either 2i or serum conditions. Oct4-EGFP+ cells were measured by flow cytometry. Plot includes the percentage of Oct4-EGFP+ cells in three biological replicates (for *Zfp42* and *Obox6* overexpression, or an empty control) from five independent experiments (Exp). **(FIG. 36E, FIG. 36F)** Number of Oct4-EGFP+ colonies at day 16 of reprogramming from primary MEFs by lentiviral overexpression of individual *Oct4, Kl/4, Sox2,* and *Myc* combined with either *Zfp42*, *Obox6,* or an empty control in **(FIG. 36E)** 2i and **(FIG. 36F)** serum conditions. Plot includes the number of Oct4-EGFP+ cells in three biological replicates (for *2fp42* and *Obox6* overexpression, or an empty control) from two independent experiments (Exp).

[0065] **FIG. 37** - Effects of GDF9 on reprogramming efficiency.

[0066] **FIG. 38** shows adding GDF9 to the medium resulted in more iPSCs.

DETAILED DESCRIPTION OF THE EXAMPLE EMBODIMENTS

General Definitions

[0067] Unless defined otherwise, technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this disclosure pertains. Definitions of common terms and techniques in molecular biology may be found in Molecular Cloning: A Laboratory Manual, 2nd edition (1989) (Sambrook, Fritsch, and Maniatis); Molecular Cloning: A Laboratory Manual, 4th edition (2012) (Green and Sambrook); Current Protocols in Molecular Biology (1987) (F.M. Ausubel et al. eds.); the series Methods in Enzymology (Academic Press, Inc.): PCR 2: A Practical Approach (1995) (M.J. MacPherson, B.D. Hames, and G.R. Taylor eds.): Antibodies, A Laboraotry Manual (1988) (Harlow and Lane, eds.): Antibodies A Laboraotry Manual, 2nd edition 2013 (E.A. Greenfield ed.); Animal Cell Culture (1987) (R.I. Freshney, ed.); Benjamin Lewin, Genes IX, published by Jones and Bartlet, 2008 (ISBN 0763752223); Kendrew *et al.* (eds.), The Encyclopedia of Molecular Biology, published by Blackwell Science Ltd., 1994 (ISBN 0632021829); Robert A. Meyers (ed.), Molecular Biology and Biotechnology: a Comprehensive Desk Reference, published by VCH Publishers, Inc., 1995 (ISBN 9780471 185710); Singleton *et al,* Dictionary of Microbiology and Molecular Biology 2nd ed., J. Wiley & Sons (New York, N.Y. 1994), March, Advanced Organic Chemistry Reactions, Mechanisms and Structure 4th ed., John Wiley & Sons (New York, N.Y. 1992); and Marten H. Hofker and Jan van Deursen, Transgenic Mouse Methods and Protocols, 2nd edition (201 1) .

[0068] As used herein, the singular forms "a", "an", and "the" include both singular and plural referents unless the context clearly dictates otherwise.

[0069] The term "optional" or "optionally" means that the subsequent described event, circumstance or substituent may or may not occur, and that the description includes instances where the event or circumstance occurs and instances where it does not.

[0070] The recitation of numerical ranges by endpoints includes all numbers and fractions subsumed within the respective ranges, as well as the recited endpoints.

[0071] The terms "about" or "approximately" as used herein when referring to a measurable value such as a parameter, an amount, a temporal duration, and the like, are meant to encompass variations of and from the specified value, such as variations of +/-10% or less, *+1-5%* or less, +/-1% or less, and +/-0.1% or less of and from the specified value, insofar such variations are appropriate to perform in the disclosed invention. It is to be understood that the value to which the modifier "about" or "approximately" refers is itself also specifically, and preferably, disclosed.

[0072] Reference throughout this specification to "one embodiment", "an embodiment," "an example embodiment," means that a particular feature, structure or characteristic described in connection with the embodiment is included in at least one embodiment of the present invention. Thus, appearances of the phrases "in one embodiment," "in an embodiment," or "an example embodiment" in various places throughout this specification are not necessarily all referring to the same embodiment, but may. Furthermore, the particular features, structures or characteristics may be combined in any suitable manner, as would be apparent to a person skilled in the art from this disclosure, in one or more embodiments. Furthermore, while some embodiments described herein include some but not other features included in other embodiments, combinations of features of different embodiments are meant to be within the scope of the invention. For example, in the appended claims, any of the claimed embodiments can be used in any combination.

[0073] All publications, published patent documents, and patent applications cited herein are hereby incorporated by reference to the same extent as though each individual publication, published patent document, or patent application was specifically and individually indicated as being incorporated by reference.

Overview

[0074] Embodiments disclosed herein provide methods and systems intended to reflect Waddington's image of marbles rolling within a development landscape. It captures the notion that cells at any position in the landscape have a *distribution* of both probable origins and probable fates. It seeks to reconstruct both the landscape and probabilistic trajectories from scRNA-seq data at various points along a time course. Specifically, it uses time-course data to infer how the probability distribution of cells in gene-expression space evolves over time, by using the mathematical approach of Optimal Transport (OT). The utility of this method is demonstrated in the context of reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs). However, the same method may be applied to other cell development and biological context where an understanding of cell orgins, trajectories, and fates is needed. For ease of reference, the methods disclosed herein and in their various embodiments may be referred to collectively as "Waddington-OT." As demonstrated herein, Waddington-OT readily rediscovers known biological features of reprogramming, including that successfully reprogrammed cells exhibit an early loss of fibroblast identity, maintain high levels of proliferation, and undergo a mesenchymal-to-epithelial transition before adopting an iPSC-like state *(12).* In addition, by exploiting single-cell resolution and the new model, it also extends these results by (1) identifying alternative cell fates, including senescence, apoptosis, neural identity, and placental identity; (2) quantifying the portion of cells in each state at each time point; (3) inferring the probable origin(s) and fate(s) of each cell and cell class at each time point; (4) identifying early molecular markers associated with eventual fates; and (5) using trajectory information to identify transcription factors (TFs) associated with the activation of different expression programs. In particular, TFs that are putative regulators of neural identity, placental identity, and pluripotency during reprogramming, and we experimentally demonstrate that one such TF, *Obox6,* enhances reprogramming efficiency are provided. Together, the data provide a high-resolution resource for studying the roadmap of reprogramming, and the methods provide a general approach for studying cellular differentiation in natural or induced settings.

[0075] Prior to describing implementation of the methods in detail, the following overview and definitions utilized in execution of the method are defined.

[0076] scRNA-seq may be obtained from cells using standard techniques known in the art. A collection of mRNA levels for a single cell is called an *expression profile* and is often represented mathematically by a vector in *gene expression space.* This is a vector space that has a dimension corresponding to each gene, with the value of the ith coordinate of an expression profile vector representing the number of copies of mRNA for the ith gene. Note that real cells only occupy an integer lattice in gene expression space (because the number of copies of mRNA is an integer), but it is assumed herein that cells can move continuously through a real-valued G dimensional vector space.

[0077] As an individual cell changes the genes it expresses over time, it moves in gene expression space and describes a trajectory. As a population of cells develops and grows, a distribution on gene expression space evolves over time. When a single cell from such a population is measured with single cell RNA sequencing, a noisy estimate of the number of molecules of mRNA for each gene is obtained. The measured expression profile of this single cell is represented as a sample from a probability distribution on gene expression space. This sampling captures both (a) the randomness in the single cell RNA sequencing measurement process (due to sub-sampling reads, technical issues, etc.) and (b) the random selection of a cell from a population. This probability distribution is treated as nonparametric in the sense that it is not specified by any finite list of parameters.

[0078] A precise mathematical notion for a *developmental process* as a generalization of a stochastic process is provided below. A goal of the methods disclosed herein is to infer the ancestors and descendants of subpopulations evolving according to an unknown developmental process. While not bound by a particular theory, this may be possible over short time scales because it is reasonable to assume that cells don't change too much and therefore it can be inferred which cells go where.

[0079] In certain example embodiments, the following definitions to define a precise notion of the developmental trajectory of an individual cell and its descendants are used. It is a continuous path in gene expression that bifurcates with every cell division.

Formally, consider a cell x(o) $\in \mathbb{R}^G$. *Let* k(t) ≥ 0 *specify the number of descendants at time* t, *where* k(0) = 1. *A single cell developmental trajectory is a continuous function*

$$x : [0, T) \rightarrow \underbrace{\underline{\mathbb{R}^G \text{ x } \mathbb{R}^G \text{ x } \ldots \text{ x } \mathbb{R}^G}}_{k(t)\ \textbf{times}}.$$

This means that x(t) is a k(t)-tuple of cells, each represented by a vector $\mathbb{R}^G$*:*

$$x(t) = (x_1(t), \ldots, x_{k(t)}(t)) \quad .$$

Cells xi(t), ..., xk(t)(t) as the descendants ofx(o).

[0080] $\mathbb{R}^G$ and R^G are used interchangeably.

[0081] Note that the temporal dynamics of an individual cell cannot be directly measured because scRNA-Seq is a destructive measurement process: scRNA-Seq lyses cells so it is only possible to measure the expression profile of a cell at a single point in time. As a result, it is not possible to directly measure the descendants of that cell, and it is (usually) not possible to directly measure which cells share a common ancestor with ordinary scRNA-Seq. Therefore the full trajectory of a specific cell is unobservable. However, one can learn something about the probable trajectories of individual cells by measuring snapshots from an evolving population.

[0082] Published methods typically represent the aggregate trajectory of a population of cells with a graph. While this recapitulates the branching path traveled by the descendants of an individual cell, it may over-simplify the stochastic nature of developmental processes. Individual cells have the potential to travel through different paths, but in reality any given cell travels one and only one such path. The methods disclosed herein help to describe this potential, which might not be a represented by a graph as a union of one dimensional paths.

[0083] Instead, a developmental process is defined to be a time-varying distribution on gene expression space. The word distribution is used to refer to an object that assigns mass to regions of $\mathbb{R}^G$. Note that a distinction is made between distribution and *probability* distribution, which necessarily has total mass 1. Distributions are formally defined as generalized functions (such as the delta function δ_χ) that act on test functions. A used herein a "distribution" is the same as a measure. One simple example of a distribution of cells is that a set of cells $x_1, \ldots, x_n$ can be represented by the distribution

$$\mathbb{P} = \sum_{i=1}^{n} \delta_{x_i}.$$

Similarly, a set of single cell trajectories may be represented *xj(t)*, $\ldots, x_n(t)$ with a distribution over trajectories. A developmental process $\mathbb{P}_*$ is a time-varying distribution on gene expression space. A developmental process generalizes the definition of *stochastic process.* A developmental process with total mass 1 for all time is a (continuous time) stochastic process, *i.e.* an ordered set of random variables with a particular dependence structure. Recall that a stochastic process is determined by its temporal dependence structure, *i.e.* the coupling between random variables at different time points. The coupling of a pair of random variables refers to the structure of their joint distribution. The notion of coupling for developmental processes is the same as for stochastic processes, except with general distributions replacing probability distributions.

[0084] A coupling of a pair of distributions P, Q on R^G is a distribution π on $R^G \times R^G$ with the property that π has P and Q as its two marginals. A coupling is also called a transport map.

[0085] As a distribution on the product space R^G x R^G, a transport map π assigns a number π(A, B) to any pair of sets A,B ⊂ R^G ·

$$\pi(A, B) = \int_{x \in A} \int_{y \in B} \pi(x, y) dx dy.$$

When π is the coupling of a developmental process, this number π(A, B) represents the mass transported from A to B by the developmental process. This is the amount of mass coming from A and going to B. When a particular destination is note specified, the quantity π(A, ·) specifies the full distribution of mass coming from A. This action may be referred to as*pushing* A through the transport map π. More generally, we can also push a *distribution* μ forward through the transport map π via integration

$$\mu \mapsto \int \pi(x, \cdot) d\mu(x).$$

The reverse operation is referred to as pulling a set B back through π. The resulting distribution π(·, B) encodes the mass ending up at B. Distributions μ can also be pulled back through π in a similar way:

$$\mu \mapsto \int \pi(\cdot, y) d\mu(y).$$

This may also be referred as *back-propagating* the distribution μ (and to pushing μ forward as *forward propagation).*

[0086] Recall that a stochastic process is Markov if the future is independent of the past, given the present. Equivalently, it is fully specified by its couplings between pairs of time points. A general stochastic process can be specified by further higher order couplings. Markov developmental processes, which are defined in the same way:

[0087] *A Markov developmental process* P_t *is a time-varying distribution on* R^G *that is completely specified by couplings between pairs of time points.* It is an interesting question to what extent developmental processes are Markov. On gene expression space, they are likely not Markov because, for example, the history of gene expression can influence chromatin modifications, which may not themselves be reflected in the observed expression profile but could still influence the subsequent evolution of the process. However, it is possible that developmental processes could be considered Markov on some augmented space.

[0088] A definition of descendants and ancestors of subgroups of cells evolving according to a Markov developmental process is now provided. The earlier definition of descendants is extended as follows: *Consider a set cf cells* S ⊂: R^G, *which live at time* x_1 *are part cf a population cf cells evolving according to a Markov developmental process* P^. *Let* π *denote the transport map for V^from time \\ to time x^- The descendants cf* S *at time* x_2 *are obtained by pushing* S *through the transport map* π. Note that if a developmental process is not Markov, then the descendants of S are not well defined. The descendants would depend on the cells that gave rise to S, which we refer to as the *ancestors* of S.

[0089] Definition 6 (ancestors in a Markov developmental process). *Consider a set cf cells* S ⊂ R^G, *which live at time* t_2 *and are part cf a population cf cells evolving according to a Markov developmental process* P^. *Let* π *denote the transport map for V^from time* x_2 *to time X\. The ancestors ofS at time* t_1 *are obtained by pushing* S *through the transport map* π.

Empirical developmental processes

[0090] In certain aspects, a goal of the embodiments disclosed herein is to track the evolution of a developmental process from a scRNA-Seq time course. Suppose we are given input data consisting of a sequence of sets of single cell expression profiles, collected at T different time slices of development. Mathematically, this time series of expression profiles is a sequence of sets $S_1,...,S_T$ ⊂: R^G collected at times ï$_1$,...,ï$_T$ ∈ R.

[0091] Developmental time series. *A developmental time series is a sequence cf samples from a developmental process* P_t *on* R^G. *This is a sequence cf sets* $S_1, \dots, S_N$ ⊂: R^G. *Each* S_i *is a set cf expression profiles in* R^G *drawn i.i.d from the probability distribution obtained by normalizing the distribution* P_{t_i} *tohavetotalmassX.* From this input data, we form an empirical version of the developmental process. Specifically, at each time point tj we form the empirical probability distribution supported on the data x ∈ S_i is formed. This is summarized inin the following definition:

[0092] Empirical developmental process. *An empirical developmental process* $\hat{P}_t$ *is a time vary-ing distribution constructedfrom a developmental time course* $S_1, \dots, S_N$ *:*

$$\hat{\mathbf{P}}_{t_i} = \frac{1}{|S_i|} \sum_{x \in S_i} \delta_x.$$

he empirical developmental process is undefined for $t \notin \{t_1, \ldots, t_N\}$.

[0093] Our goal is to recover information about a true, unknown developmental process P_t from the empirical developmental process $\hat{P}_t$. The measurement process of single cell RNA-Seq destroys the coupling, and the observed empirical developmental process does not come with an informative coupling between successive time points. Over short time scales, it is reasonable to assume that cells do not change too much and therefore inferences regarding which cells go where and estimate the coupling.

[0094] This may be done with *optimal transport:* the transport map π that minimizes the total work required for redistributing $\hat{P}_{tj}$ to $\hat{P}_{ti+1}$. is selected. One motivation for minimizing this objective, is a deep relationship between optimal transport and dynamical systems that provides a direct connection to Waddington's landscape: the optimal transport problem can formulated as a *least-action advection* of one distribution into another according to an unknown velocity field (see Theorem 1 in Section 6 below). At a high level, differentiation follows a velocity field on gene expression space, and the potential inducing this velocity field is in direct correspondence with Waddington's landscape^.

Optimal transport for scRNA-Seq time series

[0095] A process for how to compute probabilistic flows from a time series of single cell gene expression profiles by using optimal transport (SI) is provided. The embodiments disclosed herein show how to compute an optimal *coupling* of adjacent time points by solving a convex optimization problem.

[0096] Optimal transport defines a metric between probability distributions; it measures the total distance that mass must be transported to transform one distribution into another. For two measures P and Q on R^G, a *transport plan* is a measure on the product space $R^G \times R^G$ that has marginals P and Q. In probability theory, this is also called a *coupling.* Intuitively, a transport plan π can be interpreted as follows: if one picks a point mass at position x, then $\pi(\chi, \cdot)$ gives the distribution over points where x might end up.

[0097] If c(x, y) denotes the cost[2] of transporting a unit mass from x to y, then the expected cost under a transport plan π is given by

$$\iint c(x,y)\pi(x,y)dxdy.$$

The optimal transport plan minimizes the expected cost subject to marginal constraints:

$$\begin{aligned} \underset{\pi}{\text{minimize}} \quad & jj \; c(x\,,y\,)\pi(x\,,y)dxdy \\ \text{subject to} \quad & \int \pi(x\,,\text{-})dx \;=\; \mathbb{Q} \\ & \int \pi(\cdot,y)dy = \mathbb{P}. \end{aligned}$$

[0098] Note that this is a linear program in the variable π because the objective and constraints are both linear in π. Note that the optimal objective value defines the *transport distance* between P and Q (it is also called the Earthmover's distance or Wasserstein distance). Unlike most other ways to compare distributions (such as KL-divergence or total variation), optimal transport takes the geometry of the underlying space into account. For example, the KL-Divergence is infinite for any two distributions with disjoint support, but the transport distance between two unit masses depends on their separation.

[0099] When the measures P and Q are supported on finite subsets of R^G, the transport plan is a matrix whose entries give transport probabilities and the linear program above is finite dimensional. In this context, *empirical distributions* are formed from the sets of samples $S_1, \ldots,$ S_T :

$$\hat{\mathbb{P}}_{t_i} = \frac{1}{|S_i|}\sum_{x\in S_i}\delta_x,$$

were δ_χ denotes the Dirac delta function centered at $x \in R^G$. These empirical distributions $\hat{P}_{ti}$ are definitely supported, and so it is possible solve the linear program[l]with $P=\hat{P}_{tj}$ and $Q=\hat{P}_{ti+1}$·

[00100] However, the classical formulation [1] does not allow cells to grow (or die) during transportation (because it was designed to move piles of dirt and conserve mass). When the classical formulation is applied to a time series with two distinct subpopulations proliferating at

different rates[3], the transport map will artificially transport mass between the subpopulations to account for the relative proliferation. Therefore, we modify the classical formulation of optimal transport in equation [1] is modified to allow cells to grow at different rates.

[00101] Is it assumed that a cell's measured expression profile x determines its growth rate g(x). This is reasonable because many genes are involved in cell proliferation (e.g. cell cycle genes). It is further assumed g(x) is a known function (based on knowledge of gene expression) representing the exponential increase in mass per unit time, but also note that the growth rate can be allowed to be miss-specified by leveraging techniques from *unbalanced transport* (S2). In practice, g(x) is defined in terms of the expression levels of genes involved in cell proliferation.

[00102] **Derivation of transport with growth:** For any cell x ∈ Sj-j, let r(x, y) be the fraction of x that transitions towards y. Then the amount of probability mass from x that ends up at y (after proliferation) is

$$r(x,\ y)g(x)^{\ \Delta t},$$

where $A_t = t_i+_1$ - t_i. The total amount of mass that comes from x can be written two ways:

$$\sum_{y \in S_{i+1}} r(x,\ y)g(x)^{\ \Delta t} \approx g(x)^{\ \Delta t} d\hat{\mathbb{P}}_{t_i}(x)\ .$$

This gives us a first constraint. Similarly, there is also the constraint that *the total mass observed at y is equal to the sum cf masses comingfrom each* x *and ending up at y*. In symbols,

$$d\hat{\mathbb{P}}_{t_{i+1}}(y) \sum_{x \in S_i} g(x)^{\Delta_t} \sim \sum_{x \in S_i} r\,(x\,, V)9(x)^{\Delta_t} \qquad \text{for each } y \in S_{i+1}\ .$$

The factor $_x \in$ gj g(x)$^{\Delta t}$ on the left hand side accounts for the overall proliferation of all the cells from S_i. Note that this factor is required so that the constraints are consistent: when one sums up both sides of the first constraint over x, this must equal the result of summing up both sides of the second constraint over y. Finally, for convenience these constraints are rewritten in terms of the optimization variable

$$\pi(x,\ y) = r(x,\ y)g(x)^{\ \Delta t}.$$

Therefore, to compute the transport map between the empirical distributions of expression profiles observed at time t_i and t_i+_1, the following linear program is set up:

$$\begin{aligned} &\underset{\pi}{\textbf{minimize}} && \sum_{x\in S_i}\sum_{y\in S_{i+1}} c(x,y)\pi(x,y) \\ &\textbf{subject to} && \sum_{x\in S_i} \pi\mathbf{r}(x,y)\ \textbf{sa}\ d\hat{\mathbb{P}}_{t_{i+1}}\mathbf{(y)} \sum_{x\in S_i} g\mathbf{(}x)^{\Delta_t} \\ & && \sum_{y\in S_{i+1}} \pi(x,y) \approx d\hat{\mathbb{P}}_{t_i}(x)g(x)^{\Delta_t} \end{aligned}$$

[00103] Regularization and algorithmic considerations: Fast algorithms have been recently developed to solve an entropically regularized version of the transport linear program (S3). Entropic regularization means adding the entropy H(π) = E_π log π to the objective function, which penalizes deterministic transport plans (a purely deterministic transport plan would have only one nonzero entry in each row). Entropic regularization speeds up the computations because it makes the optimization problem strongly convex, and gradient ascent on the dual can be realized by successive diagonal matrix scalings (S3). These are very fast operations. This scaling algorithm has also been extended to work in the setting of *unbalanced transport,* where equality constraints are relaxed to bounds on KL-divergence (S2). This allows the growth rate function g(x) to be misspecified to some extent.

[00104] Both entropic regularization and unbalanced transport may be used. To compute the transport map between the empirical distributions of expression profiles observed at time t_i and t_i+_1, the embodiments disclosed herein solve the following optimization problem:

$$\begin{aligned} &\underset{\pi}{\textbf{minimize}} && \sum_{x\in S_i}\sum_{y\in S_{i+1}} c(x,y)\pi(x,y) - \epsilon\mathrm{H}\{\pi) \\ &\textbf{subject to} && \mathrm{KL}\left[\sum_{x\in S_i}\pi(x,y)\middle\| d\hat{\mathbb{P}}_{t_{i+1}}(y)\sum_{x\in S_i} g(x)^{\Delta_t}\right] \le \frac{1}{\lambda_1} \\ & && \mathrm{KL}\left[\sum_{y\in S_{i+1}}\pi(x,y)\middle\| d\hat{\mathbb{P}}_{t_i}(x)g(x)^{\Delta_t}\right] \le \frac{1}{\lambda_2} \end{aligned}$$

where ε, λ_1 and λ_2 are regularization parameters. This is a convex optimization problem in the matrix variable π ∈ $R^{Ni\times Ni+1}$, $_w$h$_{ere}$ N_I = $|\mathbf{g}_I|$ j$_s$ the $_{num}$ ber $_0$f cells sequenced at time tj. It takes about 5 seconds to solve this unbalanced transport problem using the scaling algorithm of Chizat et al. 2016 (S2) on a standard laptop with Nj ≈ 5000. Note that the densities (on the discrete set Sj) of the empirical distributions specified in equation [2] are simply $d\hat{P}_t(x) = {}^1$.

However, in principle one could use nonuniform empirical distributions (e.g. i Ni if one wanted to include information about cell quality).

[00105] To summarize: given a sequence of expression profiles S_i, . . . , S_T , the optimization problem [5] for each successive pair of time points S_i, S_i+_1 is solved. This gives us a sequence of transport maps as illustrated in **FIG. 3.**

[00106] To make this more precise, consider a single cell y ∈ Sj. The column π(·, y) of the transport map π from tj_i to t_i describes the contributions to y of the cells in S_i-_1**.** This is the origin of y at the time point tj_i . Similarly, the row r(y, ·) of the transition map from tj to t_i+_1 describes the probabilities y would transition to cells in S_i+_1. These are the fates of y, i.e. the descendants of y.

[00107] The origin of y further back in time may be computed via matrix multiplication: the contributions to y of cells in S_i-_2 are given by a column of the matrix

$$\tilde{K}[i\text{-}2,i] = \pi_{[i-2,i-1]}\pi_{[i-1,i]}.$$

[00108] This matrix $\pi\tilde{}_{[i-2,}$i] represents the inferred transport from time point tj_2 to t_i, and note it with a tilde to distinguish it from the maps computed directly from adjacent time points. Note that, in principle, the transport between any non-consecutive pairs of time points S_i, Sj, may be directly computed but it is not anticipated that the principle of optimal transport to be as reliable over long time gaps.

[00109] Finally, note that expression profiles can be interpolated between pairs of time points by averaging a cell's expression profile at time t_i with its fated expression profiles at time t_i+_1**.**

Transport maps encode regulatory information

[00110] Transport maps can encode regulatory information, and provided herein are methods on how to set up a regression to fit a regulatory function to our sequence of transport maps. It is assumed that a cell's trajectory is cell-autonomous and, in fact, depends only on its own internal gene expression. We know this is wrong as it ignores paracrine signaling between cells, and we return to discuss models that include cell-cell communication at the end of this section. However, this assumption is powerful because it exposes the time-dependence of the stochastic process P_t as arising from pushing an initial measure through a differential equation:

$$\dot{x} = /(*)\cdot$$

[00111] Here f is a vector field that prescribes the flow of a particle x (see fig. 3 for a cartoon illustration of a distribution flowing according to a vector field). Our biological motivation for estimating such a function f is that it encodes information about the regulatory networks that create the equations of motion in gene-expression space.

[00112] We propose to set up a regression to learn a regulatory function f that models the fate of a cell at time t_i+_1 as a function of its expression profile at time ¾. For motivation that the transport maps might contain information about the underlying regulatory dynamics, we appeal to a classical theorem establishing a dynamical formulation of optimal transport.

[00113] Theorem 1 (Benamou and Brenier, 2001). *The optimal objective value cf the transport problem* [1] *is equal to the optimal objective value cf the following optimization problem:*

$$\begin{array}{ll} \underset{\rho,v}{\text{minimize}} & \int_0^1 \int_{\mathbb{R}^G} \|v(t,x)\|^2 \rho(t,x) dt dx \\ \text{subject to} & \rho(0,\cdot) = \mathbb{P}, \quad \rho(1,\cdot) = \mathbb{Q} \qquad \cdot \\ & \nabla \cdot (\rho v) = \frac{\partial \rho}{\partial t} \end{array}$$

[00114] In this theorem, v is a vector-valued velocity field that advects4 the distribution p from P to Q, and the objective value to be minimized is the kinetic energy of the flow (mass **x** squared velocity). Intuitively, the theorem shows that a transport map π can be seen as a point-to-point summary of a least-action continuous time flow, according to an unknown velocity field. While the optimization problem [8] can be reformulated as a convex optimization problem, and modified to allow for variable growth rates, it is inherently infinite dimensional and therefore difficult to solve numerically.

[00115] We therefore propose a tractable approach to learn a static regulatory function f from our sequence of transport maps. Our approach involves sampling pairs of points using the couplings from optimal transport, and solving a regression to learn a regulatory function that predicts the fate of a cell at time t_i+_1 as a function of its expression profile at time t_i:

[00116] **Regulatory network regression:** For each pair of time points ¾,¾+!, we consider the pair of random variables X_t ,X_t jointly distributed according to $r_{[t\ ,t\ ]}$, (which we obtained from

the i i+1 i i+1 transport map $\pi_{[t_i, t_{i+1}}$] by removing the effect of proliferation as in equation [3]). We set up the following optimization problem over regulatory functions f:

$$\min_{f \in \mathcal{F}} \quad \mathbb{E}_r \left\| \frac{X_{t_i} - X_{t_{i+1}}}{\Delta_t} - f(X_{t_i}) \right\|^2 .$$

Here F specifies a parametric function class to optimize over.

[00117] Cell non-autonomous processes: We conclude our treatment of gene regulatory networks by discussing an approach to cell-cell communication. Note that the gradient flow [8] only makes sense for cell autonomous processes. Otherwise, the rate of change in expression $\dot{x}$ is not just a function of a cell's own expression vector x(t), but also of other expression vectors from other cells. We can accommodate cell non-autonomous processes by allowing f to also depend on the full distribution P_t

$$\frac{dx}{dt} = f(x, \mathbb{P}_t).$$

4. Extensions to continuous time.

[00118] In this section we discuss how our method could be improved by going beyond pairs of time points to track the continuous evolution of P_t. We begin by pointing out a peculiar behavior of our method: whenever we have a time point with few sampled cells, our method is forced through an information bottleneck. As an extreme example - suppose we had a time point with only one cell. Everything would transition through that single cell, which is absurd! In this extreme case, we would be better off ignoring the time point. We therefore propose a smoothed approach that shares information between time slices and gracefully improves as data is added.

[00119] Our continuous-time formulation is based on locally-weighted averaging, an elementary interpolation technique. Recall that given noisy function evaluations yj ≈ f(xj), one can interpolate f by averaging the y_i for all x_i close to a point of interest x:

$$f(x) \approx \sum_i \alpha_i f(x_i),$$

where oq are weights that give more influence to nearby points

[00120] In our setup, we seek to interpolate a distribution-valued function P_t from the collections of i.i.d. samples S_i, . . . , S_T . We can interpolate a distribution-valued function by computing the *barycenter* (or centroid) of nearby time points with respect to the optimal transport metric. The *transport barycenter* of

$$\underset{\mathbb{Q}}{\textbf{minimize}} \quad \Sigma_{i=l}^{T} \ \alpha_i W^2(\mathbb{P}_i, \mathbb{Q}),$$

where W (P, Q) denotes the transport distance (or Wasserstein distance) between P and Q. The transport distance is defined by the optimal value of the transport problem [1]. The weights α_i can be chosen to interpolate about time point t by setting, for example,

$$\underset{\mathbb{Q}}{\textbf{minimize}} \quad \sum_{i=l}^{T} \alpha i G^2 (\hat{\text{¥}t\text{-}}, \mathbb{Q}),$$

where G(P, Q) denotes our modified transport distance from equation [5]. To solve this optimization problem, we can fix the support of Q to the samples observed at all time points $U^T_{i=\backslash} S_i$. Then we can applythescalingalgorithmforunbalancedbarycentersduetoChizatetal. (S2).

[00121] However, fixing the support of the barycenter ahead of time may not be completely satisfactory, and this motivates further research in the computation of transport barycenters: can we design an algorithm to solve for the barycenter Q without fixing the support in advance? Is there a dynamic formulation for barycenters analogous to the Brenier Benamou formula of Theorem 1, and can we leverage it to better learn gene regulatory networks?

[00122] Finally, we conclude this section with the observation that this continuous-time approach could pro-vide a principled approach to sequential experimental design. We can identify optimal time points for further data collection by examining the loss function (fit of barycenter) across time, and adding data where the fit is poor. Moreover, we could also use this continuous time approach to test the principle of optimal transport by withholding some time points and testing the quality of the interpolation against the held-out truth.

Example System Architectures

[00123] **FIG. 1** is a block diagram depicting a system for mapping developmental trajectories of cells using single cell sequencing data, in accordance with certain example embodiments. As depicted in **FIG. 1**, the system 100 includes network devices 110, 115, and 120, that are configured to communicate with one another via one or more networks 105. In some embodiments, a user associated with the user device 115, may have to install an application and/or make a feature selection to obtain the benefits of the techniques described herein.

[00124] Each network 105 includes a wired or wireless telecommunication means by which network devices (including devices 110, 135 and 140) can exchange data. For example, each network 105 can include a local area network ("LAN"), a wide area network ("WAN"), an intranet, an Internet, a mobile telephone network, or any combination thereof. Throughout the discussion of example embodiments, it should be understood that the terms "data" and "information" are used interchangeably herein to refer to text, images, audio, video, or any other form of information that can exist in a computer-based environment.

[00125] Each network device 110, 135 and 140 includes a device having a communication module capable of transmitting and receiving data over the network 105. For example, each network device 110, 135 and 140 can include a server, desktop computer, laptop computer, tablet computer, a television with one or more processors embedded therein and / or coupled thereto, smart phone, handheld computer, personal digital assistant ("PDA"), or any other wired or wireless, processor-driven device. In the example embodiment depicted in **FIG. 1,** the network devices (including systems 110, 115 and 120) are operated by end-users or consumers, merchant operators (not depicted), and feedback system operators (not depicted), respectively.

[00126] A user can use the application 112, such as a web browser application or a stand-alone application, to view, download, upload, or otherwise access documents or web pages via a distributed network 105. The network 105 includes a wired or wireless telecommunication system or device by which network devices (including devices 110, 115 and 120) can exchange data. For example, the network 105 can include a local area network ("LAN"), a wide area network ("WAN"), an intranet, an Internet, storage area network (SAN), personal area network (PAN), a metropolitan area network (MAN), a wireless local area network (WLAN), a virtual private network (VPN), a cellular or other mobile communication network, Bluetooth, NFC, or any combination thereof or any other appropriate architecture or system that facilitates the communication of signals, data, and/or messages. Throughout the discussion of example embodiments, it should be understood that the terms "data" and "information" are used interchangeably herein to refer to text, images, audio, video, or any other form of information that can exist in a computer based environment.

[00127] The communication application 112 can interact with web servers or other computing devices connected to the network 105, including the single cell sequencing system 110 and optimal transport system 120.

[00128] It will be appreciated that the network connections shown are example and other means of establishing a communications link between the computers and devices can be used. Moreover, those having ordinary skill in the art having the benefit of the present disclosure will appreciate that the single cell sequencing system 110, user device 115, and optimal transport system 120 illustrated in **FIG. 1** can have any of several other suitable computer system configurations. For example a user device 115 embodied as a mobile phone or handheld computer may not include all the components described above

Example Processes

[00129] The example methods illustrated in **FIG. 2** are described hereinafter with respect to the components of the example operating environment **100.** The example methods of **FIG. 2** may also be performed with other systems and in other environments

[00130] **FIG. 2** is a block flow diagram depicting a method **200** to determine developmental trajectories of cells, in accordance with certain example embodiments.

[00131] Method **200** begins at block **205,** where the optimal transport module **125** performs optimal transport analysis on single cell RNA-seq data (scRNA-seq) from a time course, by calculating optimal transport maps and using them to find ancestors, descendants and trajectories for any set of cells. Given a subpopulation of cells, the sequence of ancestors coming before it and descendants coming after it are referred to as its developmental trajectory. Further example of how development trajectories may be computed in block **205** is described in Example 1 below. Briefly, transport maps are calculated, as described above, between consecutive time points, with cells allowed to grow according to a gene-expression signature of cell proliferation. From these transport maps, the forward and backword transport possibilities can be calculated between any two classes of cells at any time points. For example, a successfully reprogrammed cell at day 16 and use back-propagation to infer the distribution over their precursors at day 12. This can then be further propagated back to day 11, and so one to obtain the ancestor distributions at all previous time points. From this trend in gene expression over time may be plotted. See **FIGs. 9A-9D.**

[00132] In certain example embodiments, an expression matrix may be computed by the optimal transport module 125 from the scRNA-Seq data. Sequence reads may be aligned to obtain a matrix U of UMI counts, with a row for each gene and column for each cell. To reduce variation due to fluctuations in the total number of transcripts per cell, we divide the UMI vector for each cell by the total number of transcripts in that cell. Thus we define the expression matrix E in terms of the UMI matrix U via:

$$E_{ij} = \frac{U_{ij}}{\sum_{i=1}^{G} U_{ij}} \times 10^4.$$

[00133] Two variance-stabilizing transforms of the expression matrix E may be used for further analysis. In particular

1. $\tilde{E}$ to be the log-normalized expression matrix. The entries of $\tilde{E}$ are obtained via

$$\tilde{E}_{ij} = \mathbf{log}\,(E\, i_j + 1)\ .$$

2. E~to be the truncated expression matrix. The entries of $E^{-}$are obtained by capping the entries of E at the 99.5% quantile.

[00134] At block **210,** the optimal transport module 125 determines cell regulatory models based on the optimal transport maps. In certain example embodiments, the optimal transport module 125 determines cell regulatory models based at least in part on the optimal transport maps. In certain example embodiments, the optimal transport module 125 may further identify local biomarker enrichment based at least in part on the optimal transport maps. An example implementation is described in further detail in Example 1 below. Transcription factors (TFs) that appear to play important roles along trajectories to key desitnations are identified by two approaches. The first approach involves constructing a global regulatory model. Pairs of cells at consecutive time points are sampled according to their transport probabilities; expression levels of Tfs in the cell at time *t* are used to predict expression levels of all non-TFs in the paired cell at time $t + 1$., under the assumption that the regulatory rules are constant across cells and time points. TFs may be excluded from the predicted set to avoid cases of spurious self-regulation). The second approach involves enrichment analysis. TFs are identified based on enrichment in cells at an earlier time point with a high probability (e.g. >80%) of transitioning to a given state vs. those with a low probability (e.g. <20%).

[00135] At block 215, the optimal transport module 125 may further define gene modules. In certain example embodiments, this step is optional. Cells may be clustered based on their gene-expression profiles, after performing two rounds of dimensionality reduction to increase statistical power in subsequent analyses. For the reprogramming data disclosed herein, the analysis partitioned 16,339 detected genes into 44 gene modules, which were then analyzed for enrichment of gene sets (signatures) related to specific pathways, cells types, and conditions. **(FIG. 13, Table 1).** Based on the expression profiles in each cell, signature scores were calculated (defined by curated gene sets) for relevant features including MEF identity, pluripotency, proliferation, apoptosis, senescence, X-reactivation, neural identity, placental identity and genomic copy-number variation.

Table 1

Clusters	Gene Modules	ID (Term)	q-Value	Database
1	GM4	GO:0036211 (protein modification process)	7.0 10-3	BP
	GM10	GO:001604 (cellular component organization)		BP
		GO:0036211 (protein modification process)		BP
		GO:0006325 (chromain organization)		BP
		GO:0016570 (histone modification)		BP
2	GM5	GO:0007049 (cell cycle)	9.6 10-123	BP
		GO:0000278 (mitotic cell cycle)	6.7 10-110	BP
		GO:0006260 (DNA replication)	6.7 10-55	BP
3	GM33	IPR001400 (Somatotropin)	9.0 10-06	1
		GO:0005179 (hormone activity)	3.3 10-09	MF
		R-MMU-1170546 (Prolactin receptor signaling)	7.0 10-15	R
		R-MMU-982772 (Growth hormone receptor signaling)	1.1 10-13	R
	GM40	GO:0045664 (regulation of neuron differentiation)		BP
4	GM8	GO:0030855 (epithelial cell differentiation)	2.6 10-11	BP
		GO:0060429 (epithelium development)	1.5 10-07	BP
		mmu04530 (Tight junction)	2.7 10-08	K
	GM14	GO:0001890 (placenta development)	2.5 10-5	BP
	GM42	GO:0016126 (sterol biosynthetic process)	4.8 10-38	BP
		Hallmark cholesterol homeostasis	8.0 10-29	M
5	GM2	GO:0009653 (anatomical structure morphogenesis)	5.8 10-29	BP
		GO:0050793 (regulation of developmental process)	1.6 10-25	BO

		GO:0031012 (extracellular matrix)	1.6 10-17	CC
	GM6	Lee Bmp2 Targets up	2.3 10-16	M
	GM7	GO:0034976 (response to endoplasmic reticulum stress)	3.8 10-16	BP
	GM9	GO:0072331 (signal transduction by p53 class mediator)	6.5 10-06	BP
		mmu04115 (p53 signaling pathway)	2.9 10-10	K
		HALLMARK_P53_PATHWAY	2.1 10-26	M
	GM23	GO:0043568 (positive regulation of insulin-like growth factor receptor signaling pathway)	1.0 10-4	BP
		GO:0005520 (insulin-like growth factor binding)	3.1 10-5	MF
	GM27	GO:0031012 (extracellular matrix)	2.9 10-3	CC
	GM32	GO:0006749 (glutathione metabolic process)	1.5 10-3	BP
		MOUSE_PWY-4061 (glutathione-mediated detoxification)	1.7 10-2	BI
	GM34	GO:0035456 (response to interferon-beta)	2.5 10-13	BP
		GO:0006952 (defense response)	8.0 10-11	BP
	GM35	GO:0006952 (defense response)	6.6 10-08	BP
		GO:0006958 (complement activation, classical pathway)	1.7 10-5	BP
	GM37	GO:0034097 (response to cytokine)	5.0 10-11	BP
		mmu04668 (TNF signaling pathway)	4.8 10-11	K
	GM43	Hallmark Tgf beta signaling	2.0 10-3	M
	GM44	GO:0009952 (ranterior/posterior pattern specification)	2.9 10 15	BP
		GO:0001501 (skeletal system development)	1.2 10-12	BP
6	GM13	Pasini Suzl2 Targets up	3.0 10-20	M
		WP1763 PluriNetWork	3.6 10-06	W
	GM18	Mikkelsen Pluripotent State up	2.2 10-3	M
	GM25	mouse chrX |X	1.1 10-3	C
7	GM22	GO:0007399 (nervous system development)	4.64 10-5	BP
		GO:0097458 (neuron part)	2.4 10-5	CC

[00136] In certain example embodiments, dimensionality reduction may be used to increase robustness. As a first step towards dimensionality reduction, genes that do not show significant variation are removed. The resulting variable-gene expression matrix may be denoted E_{var}.

[0100] A second round of dimensionality reduction may comprise non-linear mapping such as Laplacian embedding, or diffusion component embedding. While principal component analysis (PCA) is a traditional approach to reduce dimensionality, it is only typically appropriate for preserving linear structures. To accommodate nonlinear shapes in high-dimensional gene expression space, diffusion components which are a generalization of principal components were used.

[0101] The diffusion components defined in terms of a similarity function k : RG **x** RG → [0, ∞). For a pair (x, y) of G-dimensional gene-expression profiles, the similarity function — or *kernel* function — k(x, y) measures the similarity between x and y. We use the Gaussian kernel function

$$k(x,y) = e^{-\frac{\|\tilde{x}-\tilde{y}\|^2}{4\sigma^2}}.$$

Where *x* and *y* are log-transformed expression profiles (*i.e.* columns of $\tilde{E}'$)

[0102] The diffusion components are defined as the top eigenvectors of a certain matrix constructed by evaluating the kernel function for all pairs of expression profiles *xi*, ..., *XN*. Specifically, the kernel matrix *K* is formed with entries

$$Ki_j = k(x_i, xj),$$

and then the Laplacian matrix *L* is formed by multiplying *K* on the left and the right by $D^{-1/2}$, where *D* is a diagonal matrix with entries

$$D_{ii} = \sum_{j=1}^{N} k(x_i, x_j).$$

The Laplacian matrix *L* is given by

$$L = D^{-\frac{1}{2}} K D^{-\frac{1}{2}}.$$

The diffusion components are the eigenvectors vi, . . . , **VN** of L, sorted by eigenvalue. We embed the data in d dimensional diffusion component space by selecting the top d diffusion components vl, . . . , vd, and sending data point xi to the vector obtained by selecting the ith entry of vl, . . . , v20. The diffusion component embedding of an expression profile x may be denoted by Φd(x). The top 20 diffusion components were enriched for gene signatures related to biological processes, and therefore were elected to use the top 20 diffusion components to represent data (see below for details).

[00137] At block 215, the visualization module 130 generates a visualization of a developmental landscape of the set of cells. To visualize the developmental landscape, the dimensionality of the data is reduced with diffusion components (such as those described above), and then the data is embedded in two dimension with force-directed graph visualization. While alternative visualization methods, such as t-distributed Stochastic Neighbor Embedding (t-SNE), are well suited for identifying clusters, they do not preserve global structures by including repulsive forces between dissimilar points. In particular, these repulsive forces seem to do a good job of splaying out the spikes present in the diffusion map embedding. **FIGs. 7A-7F.**

[0103] The invention is further described in the following examples, which do not limit the scope of the invention described in the claims.

Methods for Inducing Pluripotent Stems cell

[0104] The invention provides for a method of producing an induced pluripotent stem cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell. In one embodiment, a nucleic acid encoding Obox6 is introduced into a target cell. The method may include a step of introducing into the target cell at least one nucleic acid encoding a reprogramming factor selected from the group consisting of: Oct3/4, Sox2, Soxl, Sox3, Soxl5, Soxl7, Klf4, Klf2, c-Myc, N-Myc, L-Myc, Nanog, Lin28, Fbxl5, ERas, ECAT15-2, Tell, beta-catenin, Lin28b, Sail 1, Sall4, Esrrb, Nr5a2, Tbx3, and Glisl, or selected from the group consisting of: Oct4, Klf4, Sox2 and Myc.

[0105] In one embodiment, the nucleic acid encoding Obox6 is provided in a recombinant vector, for example, a lentivirus vector. In another embodiment, the nucleic acid encoding the reprogramming factor is provided in a recombinant vector. The nucleic acid may be incorporated into the genome of the cell. The nucleic may not be incorporated into the genome of the cell.

[0106] The method may include a step of culturing the cells in reprogramming medium as defined herein. The method may also include a step of culturing the cells in the presence of serum or the absence of serum, for example, after a culturing step in reprogramming medium.

[0107] The induced pluripotent stem cell produced according to the methods of the invention can express at least one of a surface marker selected from the group consisting of: Oct4, SOX2, KLf4, c-MYC, LIN28, Nanog, Glisl , TRA-160/TRA-1-81/TRA-2-54, SSEA1, SSEA4, Sal4 and Esrbb 1.

[0108] The method can be performed with a target cell that is a mammalian cell, including but not limited to a human, murine, porcine or canine cell. The target cell can be a primary or secondary mouse embryonic fibroblast (MEF).The target cell can be any one of the following: fibroblasts, B cells, T cells, dendritic cells, keratinocytes, adipose cells, epithelial cells, epidermal cells, chondrocytes, cumulus cells, neural cells, glial cells, astrocytes, cardiac cells, esophageal cells, muscle cells, melanocytes, hematopoietic cells, pancreatic cells, hepatocytes, macrophages, monocytes, mononuclear cells, and gastric cells, including gastric epithelial cells.

[0109] The target cell can be embryonic, or adult somatic cells, differentiated cells, cells with an intact nuclear membrane, non-dividing cells, quiescent cells, terminally differentiated primary cells, and the like.

[0110] The invention also provides for a method of producing an induced pluripotent stem cell comprising introducing at least one of Obox6, Spic, Zfp42, Sox2, Mybl2, Msc, Nanog, Hesxl and Esrrb into a target cell to produce an induced pluripotent stem cell. In one embodiment, a nucleic acid encoding Obox6, Spic, Zfp42, Sox2, Mybl2, Msc, Nanog, Hesxl or Esrrb is introduced into a target cell.

[0111] The invention also provides a method of producing an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5 or Table 6 into a target cell to produce an induced pluripotent stem cell. . In one embodiment, a nucleic acid encoding a transcription factor identified in Table 2, Table 3, Table 4, Table 5 or Table 6 is introduced into a target cell.

Table 2

Genes detected in less than 1% of cells in clusters 1-27
Rhox2a
Myolf
Xlr3c
Stra8
Smtnl l
Tspo2
Aurkc
Dazl
Rhoxl
Crxos
Rbakdn
Smclb
Tuba3a
Sycp3
Apobec2
Obox6
Patl2
Platr3

Gpx6
1700013H 16Rik
Lncencl
Tell
Spic
Hsf2bp
Fkbp6
Arl l4epl
Pacsinl
Fam l83b
Dpys
Fmrlnb
Gm9732
Dppa4
Fam25c
Dppa2
Lrrc34
Trpm l
Khdc3
Col9a2
Magebl6
Hesxl
Myl7
Ly6g6e
Gm9
Gm l3580
Aard
Zfp42
Gm7325

Table 3

TF	frequency in high / frequency in low	frequency in high	frequency in low
Spic	15.63	38.5%	2.4%
Zfp42	17.41	33.4%	1.9%
Obox6	61.90	9.3%	0.1%
Sox2	11.68	33.5%	2.9%
Mybl2	22.55	17.2%	0.7%
Msc	20.37	16.9%	0.8%
Nanog	6.08	51.3%	8.4%

Hesxl	8.68	35.5%	4.1%
Esrrb	17.00	16.4%	1.0%

Bold: Intersection between global regulatory network and enrichment analysis

Table 4

Late pluripotency markers unique to successful trajectory

Genes detected in less than 1% of cells in clusters 1-27
Rhox2a
Myolf
Xlr3c
Stra8
Smtnll
Tspo2
Aurkc
Dazl
Rhoxl
Crxos
Rbakdn
Smclb
Tuba3a
Sycp3
Apobec2
Obox6
Patl2
Platr3
Gpx6
1700013H16Rik
Lncencl
Tell
Spic
Hsf2bp
Fkbp6
ArlHepl
Pacsinl
Faml83b
Dpys
Fmrlnb

Gm9732
Dppa4
Fam25c
Dppa2
Lrrc34
Trpml
Khdc3
Col9a2
Mageb16
Hesxl
Myl7
Ly6g6e
Gm9
Gml3580
Aard
Zfp42
Gm7325

Table 5

TF	**frequency in high / frequency in low**	**frequency in high**	**frequency in low**
Spic	15.63	38.5%	2.4%
Zfp42	17.41	33.4%	1.9%
Obox6	61.90	9.3%	0.1%
Sox2	11.68	33.5%	2.9%
Mybl2	22.55	17.2%	0.7%
Msc	20.37	16.9%	0.8%
Nanog	6.08	51.3%	8.4%
Hesxl	8.68	35.5%	4.1%
Esrrb	17.00	16.4%	1.0%

Bold: Intersection between global regulatory network and enrichment analysis

Table 6

Candidate Transcription Factors

Gene	Description	Reference
Spic	Spi-C transcription factor (Spi-l/PU.l related)	Roderick TH, Chromosomal inversions in studies of mammalian mutagenesis. Genetics. 1979 May;92(l Pt 1 Suppl):sl21-6
Zfp42	zinc finger protein 42	Hosier BA, et al., Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. Mol Cell Biol. 1989 Dec;9(12):5623-9
Obox6	oocyte specific homeobox 6	Ko MS, et al., Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development. 2000 Apr; 127(8): 1737-49
Sox2	SRY (sex determining region Y)-box 2	Lyon MF, et al., Dose-response curves for radiation-induced gene mutations in mouse oocytes and their interpretation. Mutat Res. 1979 Nov;63(l): 161-73
Mybl2	myeloblastosis oncogene-like 2	Lam EW, et al., Characterization and cell cycle-regulated expression of mouse B-myb. Oncogene. 1992 Sep;7(9): 1885-90
Msc	musculin	Robb L, et al., musculin: a murine basic helix-loop-helix transcription factor gene expressed in embryonic skeletal muscle. Mech Dev. 1998 Aug;76(l-2): 197-201
Nanog	Nanog homeobox	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Hesxl	homeobox gene expressed in ES cells	Thomas PQ, et al., F£ES-1, a novel homeobox gene expressed by murine embryonic stem cells, identifies a new class of homeobox genes. Nucleic Acids Res. 1992 Nov 11;20(21):5840
Esrrb	estrogen related receptor, beta	Pettersson K, et al., Expression of a novel member of estrogen response element-binding nuclear receptors is restricted to the early stages of chorion formation during mouse embryogenesis. Mech Dev. 1996 Feb;54(2):21 1-23
Rhox2a	reproductive homeobox 2A	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Myolf	myosin IF	Hasson T, et al., Mapping of

		unconventional myosins in mouse and human. Genomics. 1996 Sep 15;36(3):431-9
Xlr3c	X-linked lymphocyte-regulated 3C	Bergsagel PL, et al., Sequence and expression of murine cDNAs encoding Xlr3a and Xlr3b, defining a new X-linked lymphocyte-regulated Xlr gene subfamily. Gene. 1994 Dec 15;150(2):345-50
Stra8	stimulated by retinoic acid gene 8	Bouillet P, et al., Efficient cloning of cDNAs of retinoic acid-responsive genes in P19 embryonal carcinoma cells and characterization of a novel mouse gene, Stral (mouse LERK-2/Eplg2). Dev Biol. 1995 Aug;170(2):420-33
Smtnll	smoothelin-like 1	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Tspo2	translocator protein 2	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Aurkc	aurora kinase C	Tseng TC, et al., Protein kinase profile of sperm and eggs: cloning and characterization of two novel testis-specific protein kinases (AIE1, AIE2) related to yeast and fly chromosome segregation regulators. DNA Cell Biol. 1998 Oct; 17(10): 823-33
Dazl	deleted in azoospermia-like	Kasahara M, et al., Genetic mapping of a male germ cell-expressed gene Tpx-2 to mouse chromosome 17. Immunogenetics. 1991;34(2):132-5
Rhoxl	reproductive homeobox 1	Maclean JA 2nd, et al., Rhox: a new homeobox gene cluster. Cell. 2005 Feb 1 l;120(3):369-82
Crxos	cone-rod homeobox, opposite strand	Ko MS, et al., Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development. 2000 Apr; 127(8): 1737-49
Rbakdn	RB-associated KRAB zinc finger downstream neighbor (non-protein coding)	MGD Nomenclature Committee, 2/14/1995;
Smclb	structural maintenance of chromosomes IB	Biswas U, et al., Distinct Roles of Meiosis-Specific Cohesin Complexes in Mammalian Spermatogenesis. PLoS

		Genet. 2016 Oct;12(10):el006389
Tuba3a	tubulin, alpha 3A	Villasante A, et al., Six mouse alpha-tubulin mRNAs encode five distinct isotypes: testis-specific expression of two sister genes. Mol Cell Biol. 1986 Jul;6(7):2409-19
Sycp3	synaptonemal complex protein 3	Roderick TH, Chromosomal inversions in studies of mammalian mutagenesis. Genetics. 1979 May;92(l Pt 1 Suppl):sl21-6
Apobec2	apolipoprotein B mRNA editing enzyme, catalytic polypeptide 2	Hirano K, et al., Targeted disruption of the mouse apobec-1 gene abolishes apolipoprotein B mRNA editing and eliminates apolipoprotein B48. J Biol Chem. 1996 Apr 26;271(17):9887-90
Obox6	oocyte specific homeobox 6	Ko MS, et al., Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development. 2000 Apr; 127(8):1737-49
Patl2	protein associated with topoisomerase II homolog 2	Marnef A, et al., Distinct functions of maternal and somatic Patl protein paralogs. RNA. 2010 Nov;16(11):2094-107
Platr3	pluripotency associated transcript 3	Leo D, et al., Transgenic mouse models for ADHD. Cell Tissue Res. 2013 May 17
Gpx6	glutathione peroxidase 6	Roderick TH, Producing and detecting paracentric chromosomal inversions in mice. Mutat Res. 1971 Jan;l l(l):59-69
1700013H16Rik	RIKEN cDNA 1700013H16 gene	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Lncencl	long non-coding RNA, embryonic stem cells expressed 1	Lai KM, et al., Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs. PLoS One. 2015;10(4):e0125522
Tell	T cell lymphoma breakpoint 1	Narducci MG, et al., The murine Tell oncogene: embryonic and lymphoid cell expression. Oncogene. 1997 Aug 18;15(8):919-26
Spic	Spi-C transcription factor (Spi-l/PU.l related)	Roderick TH, Chromosomal inversions in studies of mammalian mutagenesis. Genetics. 1979 May;92(l Pt 1 Suppl):sl21-6
Hsf2bp	heat shock transcription factor	Kawai J, et al., Functional annotation of a

	2 binding protein	full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Fkbp6	FK506 binding protein 6	Coss MC, et al., Molecular cloning, DNA sequence analysis, and biochemical characterization of a novel 65-kDa FK506-binding protein (FKBP65). J Biol Chem. 1995 Dec 8;270(49):29336-41
Arll4epl	ADP-ribosylation factor-like 14 effector protein-like	Zambrowicz BP, et al., Wnkl kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. Proc Natl Acad Sci U S A. 2003 Nov 25;100(24):14109-14
Pacsinl	protein kinase C and casein kinase substrate in neurons 1	Plomann M, et al., PACSIN, a brain protein that is upregulated upon differentiation into neuronal cells. Eur J Biochem. 1998 Aug 15;256(1):201-1 1
Faml83b	family with sequence similarity 183, member B	Roderick TH, Chromosomal inversions in studies of mammalian mutagenesis. Genetics. 1979 May;92(l Pt 1 Suppl):sl21-6
Dpys	dihydropyrimidinase	Skarnes WC, et al., A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 201 1 Jun 16;474(7351):337-42
Fmrlnb	fragile X mental retardation 1 neighbor	Skarnes WC, et al., A conditional knockout resource for the genome-wide study of mouse gene function. Nature. 201 1 Jun 16;474(7351):337-42
Gm9732	predicted gene 9732	Roderick TH, Using inversions to detect and study recessive lethals and detrimentals in mice, in Utilization of Mammalian Specific Locus Studies in Hazard Evaluation and Estimation of Genetic Risk. 1983:135-67.
Dppa4	developmental pluripotency associated 4	Ko MS, et al., Large-scale cDNA analysis reveals phased gene expression patterns during preimplantation mouse development. Development. 2000 Apr; 127(8): 173 7-49
Fam25c	family with sequence similarity 25, member C	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Dppa2	developmental pluripotency associated 2	Ko MS, et al., Large-scale cDNA analysis reveals phased gene expression patterns

		during preimplantation mouse development. Development. 2000 Apr; 127(8): 1737-49
Lrrc34	leucine rich repeat containing 34	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Trpm1	transient receptor potential cation channel, subfamily M, member 1	Dickinson ME, et al., High-throughput discovery of novel developmental phenotypes. Nature. 2016 Sep 14;537(7621):508-514
Khdc3	KH domain containing 3, subcortical maternal complex member	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Col9a2	collagen, type IX, alpha 2	Dickinson ME, et al., High-throughput discovery of novel developmental phenotypes. Nature. 2016 Sep 14;537(7621):508-514
Mageb16	melanoma antigen family B, 16	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Hesx1	homeobox gene expressed in ES cells	Thomas PQ, et al., HES-1, a novel homeobox gene expressed by murine embryonic stem cells, identifies a new class of homeobox genes. Nucleic Acids Res. 1992 Nov 11;20(21):5840
Myl7	myosin, light polypeptide 7, regulatory	Lowey S, et al., Light chains from fast and slow muscle myosins. Nature. 1971 Nov 12;234(5324):81-5
Ly6g6e	lymphocyte antigen 6 complex, locus G6E	Kawai J, et al., Functional annotation of a full-length mouse cDNA collection. Nature. 2001 Feb 8;409(6821):685-90
Gm9	predicted gene 9	The FANTOM Consortium and RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group), The Transcriptional Landscape of the Mammalian Genome. Science. 2005;309(5740): 1559-1563
Gm13580	predicted gene 13580	Zambrowicz BP, et al., Wnkl kinase deficiency lowers blood pressure in mice: a gene-trap screen to identify potential targets for therapeutic intervention. Proc Natl Acad Sci U S A. 2003 Nov 25;100(24):14109-14
Aard	alanine and arginine rich domain containing protein	Roderick TH, et al., Nineteen paracentric chromosomal inversions in mice. Genetics.

		1974 Jan;76(l): 109-17
Zfp42	zinc finger protein 42	Hosier BA, et al., Expression of REX-1, a gene containing zinc finger motifs, is rapidly reduced by retinoic acid in F9 teratocarcinoma cells. Mol Cell Biol. 1989 Dec;9(12):5623-9
Gm7325	myomixer, myoblast fusion factor	Hansen J, et al., A large-scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. Proc Natl Acad Sci U S A. 2003 Aug 19;100(17):9918-22

[0112] The invention also provides a method of increasing the efficiency of production of an induced pluripotent stem cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell.

[0113] The invention also provides a method of increasing the efficiency of production of an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5 or Table 6 into a target cell to produce an induced pluripotent stem cell.

[0114] The invention also provides a method of increasing the efficiency of reprogramming of a cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell.

[0115] The invention also provides a method of increasing the efficiency of reprogramming a cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5 or Table 6 into a target cell to produce an induced pluripotent stem cell.

[0116]

[0117] The invention also provides for an isolated induced pluripotent stem cell produced by the methods of the invention.

[0118] The invention also provides a method of treating a subject with a disease comprising administering to the subject a cell produced by differentiation of the induced pluripotent stem cell produced by the methods of the invention.

[0119] The invention also provides for a composition for producing an induced pluripotent stem cell comprising Obox6 or any of the factors identified in Table 2, Table 3, Table 4, Table 5 or Table 6 in combination with reprogramming media.

[0120] The invention also provides for use of Obox6 or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5 or Table 6 for production of an induced pluripotent stem cell.

Definitions

[0121] As used herein, "pluripotent" as it refers to a "pluripotent stem cell" means a cell with the developmental potential, under different conditions, to differentiate to cell types characteristic of all three germ cell layers, i.e., endoderm (e.g., gut tissue), mesoderm (including blood, muscle, and vessels), and ectoderm (such as skin and nerve). Pluripotent cell as used herein, includes a cell that can form a teratoma which includes tissues or cells of all three embryonic germ layers, or that resemble normal derivatives of all three embryonic germ layers (i.e., ectoderm, mesoderm, and endoderm). A pluripotent cell of the invention also means a cell that can form an embryoid body (EB) and express markers for all three germ layers including but not limited to the following: endoderm markers-AFP, FOXA2, GATA4; mesoderm markers-CD34, CDH2 (N-cadherin), COL2A1, GATA2, HAND1, PECAMI, RUNX1, RUNX2; and Ectoderm markers-ALDHlAl, COL1A1, NCAM1, PAX6, TUBB3 (Tuj l).

[0122] A pluripotent cell of the invention also means a human cell that expresses at least one of the following markers: SSEA3, SSEA4, Tra-1-81, Tra-1-60, Rexl, Oct4, Nanog, Sox2 as detected using methods known in the art. A pluripotent stem cell of the invention includes a cell that stains positive with alkaline phosphatase or Hoechst Stain.

[0123] In some embodiments, a pluripotent cell is termed an "undifferentiated cell." Accordingly, the terms "pluripotency" or a "pluripotent state" as used herein refer to the developmental potential of a cell that provides the ability of the cell to differentiate into all three embryonic germ layers (endoderm, mesoderm and ectoderm). Those of skill in the art are aware of the embryonic germ layer or lineage that gives rise to a given cell type. A cell in a pluripotent state typically has the potential to divide in vitro for a long period of time, e.g., greater than one year or more than 30 passages.

[0124] As used herein, the term "induced pluripotent stem cells (iPSCs or "iPS cells)" refers to cells having similar properties to those of ES cells. In particular, an "iPSC" or "iPS cell" as used herein, includes an undifferentiated cell which is reprogrammed from somatic cells and have pluripotency and proliferation potency. However, this term is not to be construed as limiting in any sense, and should be construed to have its broadest meaning. As used herein, the term "pluripotent stem cell", as it refers to the cell produced by the claimed methods is synonymous with the term "iPS".

[0125] Obox6 and any of the other factors described herein can be used to generate induced pluripotent stem cells from differentiated adult somatic cells. In the preparation of induced pluripotent stem cells by using the factors of the present invention, types of cells to be reprogrammed are not particularly limited, and any kind of cells may be used. For example, matured somatic cells may be used, as well as somatic cells of an embryonic period. Other examples of cells capable of being generated into iPS cells and/or encompassed by the present invention include mammalian cells such as fibroblasts, mouse embryonic fibroblasts, B cells, T cells, dendritic cells, keratinocytes, adipose cells, epithelial cells, epidermal cells, chondrocytes, cumulus cells, neural cells, glial cells, astrocytes, cardiac cells, esophageal cells, muscle cells, melanocytes, hematopoietic cells, pancreatic cells, hepatocytes, macrophages, monocytes, mononuclear cells, and gastric cells, including gastric epithelial cells. The cells can be embryonic, or adult somatic cells, differentiated cells, cells with an intact nuclear membrane, non-dividing cells, quiescent cells, terminally differentiated primary cells, and the like. The pluripotent or multipotent cells of the present invention possess the ability to differentiate into cells that have characteristic attributes and specialized functions, such as hair follicle cells, blood cells, heart cells, eye cells, skin cells, placental cells, pancreatic cells, or nerve cells. In particular, pluripotent cells of the invention can differentiate into multiple cell types including but not limited to: cells derived from the endoderm, mesoderm or ectoderm, including but not limited to cardiac cells, neural cells (for example, astrocytes and oligodendrocytes), hepatic cells (for example, pancreatic islet cells), osteogentic, muscle cells, epithelial cells, chondrocytes, adipocytes, placental cells, dendritic cells and, haematopoietic and retinal pigment epithelial (RPE) cells.

[0126] Induced pluripotent stem cells may express any number of pluripotent cell markers, including: alkaline phosphatase (AP); ABCG2; stage specific embryonic antigen-1 (SSEA-1); SSEA-3; SSEA-4; TRA-1-60; TRA-1-81; Tra-2-49/6E; ERas/ECAT5, E-cadherin; III-tubulin; -smooth muscle actin (-SMA); fibroblast growth factor 4 (Fgf4), Cripto, Daxl; zinc finger protein 296 (Zfp296); N-acetyltransferase-1 (Natl); (ES cell associated transcript 1 (ECAT1);

ESG1/DPPA5/ECAT2; ECAT3; ECAT6; ECAT7; ECAT8; ECAT9; ECAT10; ECAT15-1; ECAT15-2; Fthll7; Sall4; undifferentiated embryonic cell transcription factor (Utfl); Rexl; p53; G3PDH; telomerase, including TERT; silent X chromosome genes; Dnmt3a; Dnmt3b; TRIM28; F-box containing protein 15 (Fbxl5); Nanog/ECAT4; Oct3/4; Sox2; Klf4; c-Myc; Esrrb; TDGF1; GABRB3; Zfp42, FoxD3; GDF3; CYP25A1; developmental pluripotency-associated 2 (DPPA2); T-cell lymphoma breakpoint 1 (Tell); DPPA3/Stella; DPPA4; other general markers for pluripotency, etc. Other markers can include Dnmt3L; Soxl5; Stat3; Grb2; SV40 Large T Antigen; HPV16 E6; HPV16 E7, -catenin, and Bmil. Such cells can also be characterized by the down-regulation of markers characteristic of the differentiated cell from which the iPS cell is induced. For example, iPS cells derived from fibroblasts may be characterized by down-regulation of the fibroblast cell marker Thyl and/or up-regulation of SSEA-1. It is understood that the present invention is not limited to those markers listed herein, and encompasses markers such as cell surface markers, antigens, and other gene products including ESTs, RNA (including microRNAs and antisense RNA), DNA (including genes and cDNAs), and portions thereof.

[0127] As used herein, "increases the efficiency" as it refers to the production of induced pluripotent stem cells, means an increase in the number of induced pluripotent stem cells that are produced, for example in the presence of Obox6 or one or more of the factors identified in Table 2, 3, 4, 5 or 6, as compared to the number of cells produced in the absence of Obox6 or one or more of the factors identified in Table 2, 3, 4, 5 or 6 under identical conditions. An increase in the number of induced pluripotent cells means an increase of at least 5%, for example, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%), or 100%) or more. An increase also means at least 5-fold more, for example, 5-fold, -fold, 20-fold, 30-fold, 40-fold, 50-fold, 60-fold, 70-fold, 80-fold, 90-fold, 100-fold, 500-fold, 1000-fold or more. Increases the efficiency also means decreasing the time required to produce an induced pluripotent stem cell, for example in the presence of Obox6 or one or more of the factors identified in Table 6, 7, 8, 9 or 10, as compared to the number of cells produced in the absence of Obox6 or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5 and Table 6. In the presence of Obox6 or any one of the factors identified in Table 2, Table 3, Table 4, Table 5 and Table 6, an iPSC can be formed between 5 and 30 days, between 5 and 20 days, between

10 and 20 days, for example 10 days, 11 days, 12 days, 13 days, 14 days, 15 days, 16 days, 17 days, 18 days, 19 days or 20 days after the addition of Obox6 or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5 and Table 6or following induction of expression of Obox6 or or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5 and Table 6.

[0128] Candidate transcriptional regulators to augment reprogramming efficiency include but are not limited to the transcription regulators presented in Tables 2, 3, 4, 5 and 6.

EXPERIMENTAL METHODS

1. Derivation of MEFs

[0129] Mouse embryonic fibroblasts (MEFs) were derived from E13.5 embryos with a mixed B6;129 background. The cell line used in this study was homozygous for ROSA26-M2rtTA, homozygous for a polycistronic cassette carrying *Pou5fl*, *Kl/4*, *Sox2*, and *Myc* at the *Collal* locus (18), and homozygous for an EGFP reporter under the control of the *Pou5fl* promoter. Briefly, MEFs were isolated from E13.5 embryos resulting from timed-matings by removing the head, limbs, and internal organs under a dissecting microscope. The remaining tissue was finely minced using scalpels and dissociated by incubation at 37°C for 10 minutes in trypsin-EDTA (Thermo Fisher Scientific). Dissociated cells were then plated in MEF medium containing DMEM (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (GE Healthcare Life Sciences), non-essential amino acids (Thermo Fisher Scientific), and GlutaMAX (Thermo Fisher Scientific). MEFs were cultured at 37°C and 4% $C0_2$ and passaged until confluent. All procedures, including maintenance of animals, were performed according to a mouse protocol (2006N000104) approved by the MGH Subcommittee on Research Animal Care.

2. Reprogramming assay

[0130] For the reprogramming assay, 20,000 low passage MEFs (no greater than 3-4 passages from isolation) were seeded in a 6-well plate. These cells were cultured at 37°C and 5% CO2 in reprogramming medium containing KnockOut DMEM (GIBCO), 10% knockout serum replacement (KSR, GIBCO), 10% fetal bovine serum (FBS, GIBCO), 1% GlutaMAX (Invitrogen), 1% nonessential amino acids (NEAA, Invitrogen), 0.055 mM 2-mercaptoethanol (Sigma), 1%) penicillin-streptomycin (Invitrogen) and 1,000 U/ml leukemia inhibitory factor

(LIF, Millipore). Day 0 medium was supplemented with 2 g/mL doxycycline Phase- l(Dox) to induce the polycistronic OKSM expression cassette. Medium was refreshed every other day. At day 8, doxycycline was withdrawn, and cells were transferred to either serum-free 2i medium containing 3 μM CHIR99021, 1 μM PD0325901, and LIF (Phase-2(2i)) (25) or maintained in reprogramming medium (Phase-2(serum)). Fresh medium was added every other day until the final time point on day 16. Oct4-EGFP positive iPSC colonies should start to appear on day 10, indicative of successful reprogramming of the endogenous *Oct4* locus.

3. Sample collection

[0131] A total of 66,000 cells were collected from twelve time points over a period of 16 days in two different culture conditions. Single or duplicate samples were collected at day 0 (before and after Dox addition), 2, 4, 6, and 8 in Phase-l(Dox); day 9, 10, 11, 12, 16 in Phase-2(2i); and day 10, 12, 16 in Phase-2(serum). Cells were also collected from established iPSCs cell lines reprogrammed from the same MEFs, maintained either in Phase-2(2i) conditions or in Phase-2(serum) medium. For all time points, selected wells were trypsinized for 5 mins followed by inactivation of trypsin by addition of MEF medium. Cells were subsequently spun down and washed with IX PBS supplemented with 0.1% bovine serum albumin. The cells were then passed through a 40 micron filter to remove cell debris and large clumps. Cell count was determined using Neubauer chamber hemocytometer to a final concentration of 1000 cells/ 1.

4. Single-cell RNA sequencing

[0132] Single-cell RNA-Seq libraries were generated from each time point using the 10X Genomics Chromium Controller Instrument (10X Genomics, Pleasanton, CA) and Chromium™ Single Cell 3' Reagent Kits vl (PN-120230, PN-120231, PN-120232) according to manufacturer's instructions. Reverse transcription and sample indexing were performed using the CI000 Touch Thermal cycler with 96-Deep Well Reaction Module. Briefly, the suspended cells were loaded on a Chromium controller Single-Cell Instrument to first generate single-cell Gel Bead-In-Emulsions (GEMs). After breaking the GEMs, the barcoded cDNA was then purified and amplified. The amplified barcoded cDNA was fragmented, Atailed and ligated with adaptors. Finally, PCR amplification was performed to enable sample indexing and enrichment of the 3' RNA-Seq libraries. The final libraries were quantified using Thermo Fisher Qubit dsDNA HS Assay kit (Q32851) and the fragment size distribution of the libraries were determined using the Agilent 2100 BioAnalyzer High Sensitivity DNA kit (5067-4626). Pooled libraries were then sequenced using Illumina Sequencing By Synthesis (SBS) chemistry.

5. Lentivirus Vector Construction and Particle Production

[0133] To test whether transcription factors (TFs) improve late-stage reprogramming efficiency, lentiviral constructs for the top candidates *Zfp42,* and *Obox6* were generated. cDNA for these factors were ordered from Origene (Zfp42-MG203929, and Obox6-MR2 15428) were cloned into the FUW Tet-On vector (Addgene, Plasmid #20323) using the Gibson Assembly (NEB, E261 1S). Briefly, the cDNA for each TF was amplified and cloned into the backbone generated by removing Oct4 from the FUW-Teto-Oct4 vector. All vectors were verified by Sanger sequencing analysis. For lentivirus production, FIEK293T cells were plated at a density of 2.6 x 10^6 cells/well in a 10cm dish. The cells were transfected with the lentiviral packaging vector and a TF-expressing vector at 70-80% growth confluency using the Fugene FID reagent (Promega E231 1) according to the manufacturer's protocols. At 48 hours after transfection, the viral supernatant was collected, filtered and stored at -80°C for future use.

6. Reprogramming efficiency of secondary MEFS together with individual TFs

[0134] We sought to determine the ability of the candidate TFs to augment reprogramming efficiency in secondary MEFs; the use of secondary MEFs for reprogramming overcomes limitations associated with random lentiviral integration events at variable genomic locations. Briefly, secondary MEFs were plated at a concentration of 20,000 cells per well of a 6-well plate. Cells were infected with virus containing *2fp42, Obox6,* or an empty vector and maintained in reprogramming medium as described above. At day 8 after induction, cells were switched to either Phase-2(2i) or Phase-2(serum). On day 16, reprogramming efficiency was quantified by measuring the levels of the EGFP reporter driven by the endogenous *Oct4* promoter. FACS analyses was performed using the Beckman Coulter CytoFLEX S, and the percentage of Oct4-EGFP+ cells was determined. Triplicates were used to determine average and standard deviation **(FIG. 10B).**

7. Reprogramming efficiency of primary MEFS with individual TFs and OKSM

[0135] In addition to demonstrating the ability of a TF to increase reprogramming efficiency in secondary MEFs, the performance of the TFs were independently tested in primary MEFs. To this end, lentiviral particles were generated from four distinct FUW-Teto vectors, containing

Oct4, Sox2, Kl/4, and *Myc,.* MEFs from the background strain *B6.Cg-Gt(ROSA)26Sortml(rtTA*M2)Jae/J x* B6;129S4-Pou5fltm2Jae/J were infected with these lentiviral particles, together with a lentivirus expressing tetracycline-inducible *Zfp42, Obox6* or no insert. Infected cells were then induced with 2μg/mL doxycycline in ESC reprogramming medium (day 0). At day 8 after induction, cells were switched to either Phase-2(2i) or Phase-2(serum). On day 16, the number of Oct4-EGFP+ colonies were counted using a fluorescence microscope. Triplicates for each condition used to determine average values and standard deviation.

EXAMPLES

Example 1

[0136] Computing trajectories with optimal transport

[0137] As noted above, for any pair of time points we compute a transport plan that minimizes the expected cost of redistributing mass, subject to constraints involving a proliferation score (see Appendix 1 for a precise statement of the optimization problem). To compute these transport matrices, we need to specify a cost function, a proliferation function, and numerical values for the regularization parameters.

[0138] Cost functions: We tried several different cost functions based on squared Euclidean distance in different input spaces. Specifically, for cells with expression profiles x and y, given by two columns of the expression matrix E, we specify a cost function c(x, y)

Expression space $$c_1(x, y) = /\!/ \chi^{-} - y^{-} /\!/ ^2$$

100 dimensional diffusion component space $$¾\ (\chi, y) = /\!/ \Lambda \phi\ _{100}(x) - \Lambda \phi\ _{1}Qo(y) /\!/ ^2$$

20 dimensional diffusion component space $$c_3(x, y) = /\!/ \Lambda \phi\ _{20}(\chi) - {}^{A}\Phi_{20}(y) /\!/ ^2$$

[0139] The bar above x^{-}, $y^{\sim}$ denotes that we apply the truncation transform from section 2, and Φ_d is the Laplacian embedding from section 3. Note that Φ_d has the log transform $x \rightarrow \tilde{x}$ built-in. In the equations above, A is a diagonal matrix containing the eigenvalues of the Laplacian matrix, raised to the power 8. Hence c2 and c3 are both truncated versions of the diffusion distance D4(x, y) from (S5).

[0140] The cost function c3 was used to report the numerical values in the main text, and we computed separate transport maps for 2i and serum. Note that all the cost functions cl, c2, c3 give largely similar results.

[0141] Proliferation function: We estimate the relative growth rate for every cell using the proliferation signature displayed in **FIG.** 7D in the main text. To transform the proliferation score into an estimate of the growth rate (in doublings per day), we first observed that the proliferation score is bimodally distributed over the dataset. We transformed the proliferation score so that the two modes were mapped to a growth ratio of 2.5 per day (this means that over 1 day, a cell in the more proliferative group is expected to produce 2.5 times as many offspring as a cell in the non-proliferative group). However, note that we allow for some laxity in the prescribed growth rate (see supplemental figure on input vs implied proliferation).

[0142] Regularization parameters: We employed the following strategy to select the regularization pa- rameters λ and ε. The entropy parameter ε controls the entropy of the transport map. An extremely large entropy parameter will give a maximally entropic transport map, and an extremely small entropy parameter will give a nearly deterministic transport map (but could also lead to numerical instability in the algorithm). We adjusted the entropy parameter until each cell transitions to between 10 and 50 percent of cells in the next time point, as measured by the Shannon diversity of the rows of the transport map.

[0143] The regularization parameter λ controls the fidelity of the constraints: as λ gets larger, the constraints become more stringent. We selected λ so that the marginals of the transport map are 95% correlated with the prescribed proliferation score.

[0144] Implementation: The scaling algorithm for unbalanced transport (S2) was implemented to compute optimal transport maps. This algorithm performs gradient ascent steps on the dual optimization problem. Because of the entropic regularization, these gradient ascent steps can be performed via diagonal matrix scalings. We implemented versions of the solver in both R and Python.

[0145] Experiments: Computational experiments were performed to evaluate the stability of our results to choice of cost function, regularization parameters, and subsampling the dataset.

[0146] The cluster-to-cluster origin were compared and fate tables for the different cost functions listed above, and consistent results were found. Moreover, the transport probabilities described above are all robust to choice of cost function.

[0147] A bootstrap analysis was performed on a batch of 100 subsamples consisting of 50% of the data from each time point. The variance in the cluster-to-cluster origin and fate tables is extremely small (see Table 7).

Table 7

MEF.identity	Pluripotency	Gl.S	G2.M	Cell.cycle	ER.stress	Epithelial.identity	ECM.rearrangement	Apoptosis	SASP	Neural.identity	Placental.identity	X.reactivation
Gm5571	Rhox5	Cdca7	Cbx5	Mcm4	Nck2	Cdhl	Sulfl	Ercc5	116	Vtn	4933433pl4rik	Gm21950
Rbfox2	Tdgfl	Mcm4	Aurkb	Smc4	Ankzf1	Tgml	Coll9al	Serpinb5	117	Ednrb	Esxl	Gm21364
Btbdl9	Utfl	Mcm2	Ckslb	Gtse1	Dnajb2	Cldn3	Col3al	Inhbb	Ilia	Sox21	Afapl	Gml4346
Actnl	Mkrnl	Rfc2	Cks2	Ttk	Rhbddl	Cldn4	Col5a2	SteaP3	Illb	Zeb2	Zfyve21	Gml4345
Gatad2a	Dppa5a	Ung	Hn1	Rangap1	Bcl2	Cldn7	Fnl	Btg2	1113	Hes5	Erv3	Gml4351
Med6	Uppl	Mcm6	Hmgb2	Ccnb2	Ubxn4	Cldnll	Ihh	Phlda3	1115	Fabp7	Atgl2	Gm3701
Mex3a	Chchd10	Rrm1	Anp32e	Cenpa	Yodl	Ocln	Col4a4	Tnni1	Cxcl15	Soxl	Lasll	Gm3706
Ccdc80	Klf2	Slbp	Lbr	Cenpe	Ppplrl5b	Epcam	Col4a3	Rgsl6	Cxcl1	Neurodl	Rbpl	Gml4347
Mex3c	Trapla	Pena	Tmpo	Cdca8	Faml29a	Crb3	Serpinb5	Ier5	Cxcl2	Pax3	Prl2bl	Gml0921
Sdpr	Mylpf	Atad2	Top2a	CkaP2	Edem3	Krt8	Fmod	Slcl9a2	Cxcl3	Pax6	Prl3dl	Gml0922
Pcdhb2	1700013H16Rik	Tipin	Tacc3	Rad51	Atf6	Krtl9	Elf3	Adck3	Ccl8	Cdh2	Rnf2	Gm3750
Triml6	AA467197	Mcm5	Tubb4b	Pena	Ufcl	Pkp3	Lamcl	Ephxl	Cell3	Sox9	Set	Gm3763
Obsll	Dhxl6	Uhrf1	Ncapd2	Ube2c	Atf3	Dsp	Tnr	Ptpn14	Ccl3	Sox2	Mrgprg	Mycs
Ephal	Mt2	Rpa2	Rangap1	Lbr	Manlbl	Pkpl	Dpt	Atf3	Ccl20	Id2	Aa763515	Gml4374
Stxlb	Ube2a	Dtl	Cdk1	CenPf	Toria		Ddr2	Notehi	Cell6	Hoxbl	Tfpi	Nudtll

Staul	Khdc3	Priml	Smc4	Birc5	Hspa5		Olfml2b	Rxra	Ccl26	Msxl	Etosl	AU022751
Serpinel	Pycard	Fen1	Kif2Ob	Dtl	Dab2ip		Tgfb2	Ralgds	Csf2	Msil	Slc5a6	NudtlO
Aa881470	Hsp90aal	Hells	Cdca8	Dscc1	Nfe2l2		Itga8	Akl	Csf3	Msi2	1600025ml7rik	Bmpl5
Coll2a1	Prrcl	Gmnn	CkaP2	Cbx5	Dnajc10		Adamtsl2	Stom	Ifng	Atohl	Gm9	Shroom4
2010300fl7rik	Hatl	Pold3	Ndc80	Usp1	Psmc3		Col5al	Ddb2	Mif	Rbfox3	Creb3l2	Dgkk
CcdclO2a	Calcoco2	NasP	Digap5	Hmmr	Creb311		Pomtl	Cd82	Areg	Map2	Bbx	Ccnb3
Nradd	Impa2	Chaflb	Hjurp	Wdr76	Thbs1		Eng	Ilia	Ereg	Tubb3	Prl3cl	Akap4
Pard6g	Saa3	Gins2	CkaP5	Ung	Eif2ak4		Lmxlb	Pcna	Nrg1		Mta3	Clcn5
Ntn4	Ooep	Pola1	Bub1	Hnl	Chac1		Gsn	Bmp2	Egf		Prl2al	Usp27x
5730471hl9rik	Bnip3	Msh2	Ckap2l	Cks2	Pdia3		Olfml2a	Trib3	Fgf2		Gm9112	Ppplr3f
Sepnl	Mtl	Casp8aP2	Ect2	Kif2Ob	Bcl2l11		Creb3ll	Procr	Hgf		Afapll2	Ppplr3fos
Pegl2	Asns	Cdc6	Kifl1	Cdk1	Ddrgkl		Hsdl7bl2	BlcaP	Fgf7		Erlin2	Foxp3
Dpysl3	Aldoa	Ubr7	Birc5	Slbp	Tmx4		Wtl	Ada	Vegfa		Pard3	Ccdc22
1110012d08rik	Tdh	Ccne2	Cdca2	Aurkb	Trib3		Greml	Fgfl3	Ang		Aifll	Cacnalf
Aktl	Gjb3	Wdr76	Nuf2	Kifl1	H13		Spintl	Irak1	Kitl		Dmrtcla	Syp
Zfp286	Rbpms2	Tyms	Cdca3	Ckslb	Edem2		Cst3	Tspy12	Cxcl12		4932442l08rik	Gml4703
Ubap2l	Prpsl	Cdc45	Nusapl	Blm	Cebpb		Fkbpla	Satl	Pigf		Gjb2	Prickle3
Samd4	Fam25c	Clspn	Ttk	Msh2	Ptpn1		Mmp9	Zmat3	IgfbP2		Gjb5	Plp2
Phc2	Eif2s2	Rrm2	Aurka	Gas213	Vapb		Sulf2	Hspa4l	IgfbP3		Slco5al	Magix
Mcam	Cenpm	Dscc1	Mki67	Tyms	Srpx		Atp7a	Slc7all	Igfbp4		Wdr61	Gpkow
Pla2g4c	Nanog	Rad51	Fam64a	HjurP	Aifm1		Noxl	Tm4sfl	Igfbp6		Kitl	Wdr45
Fzd7	Ndufa412	Usp1	Ccnb2	Hells	Ubqln2		Col4a6	Rap2b	IgfbP7		9430027b09rik	RP23-109E24.10

Pappa	Syce2	Exo1	Tpx2	Priml	Mbtps2		Prdx4	Fbxw7	MmPi		Tfrc	Praf2
Ptk7	Gml3251	Blm	Hjurp	Uhrf1	Uspl3		Gpm6b	S100a4	MmP3		Slc6a2	Ccdcl20
Nuakl	Taf7	Rad51aPi	Anln	Ndc80	Ufml		Egfl6	S10Oal0	MmplO		Wdr45	Tfe3
Ill7rd	Nudt4	Mlflip	Kif2c	Mcm6	Serp1		Postn	TxniP	Mmpl2		Zxda	Gripap1
Ptk2	Cox5a	E2f8	Cenpe	Rrm1	Creb314		Rxfpl	Nhlh2	Mmpl3		Prdx4	Kcndl
Ehd2	Sod2	Brip1	Gtsel	Mlflip	Tmem67		Sfrp2	Dnttip2	Mmpl4		Faml22b	Otud5
Lats2	SlOOa13		Kif23	Top2a	Ufll		Hapln2	Clca2	TimP2		Zxdb	Pim2
Hspg2	Fkbp6		Cdc20	Hmgb2	Ube2jl		Ctss	WwPi	Serpine1		Zxdc	Slc35a2
4930456gl4rik	Rhox9		Ube2c	Cene2	Vcp		Adamtsl4	Klf4	Serpinb2		Pip5kla	Pqbpl
4930429b21rik	Gdf3		CenPf	G2e3	Creb3		St7l	Ikbkap	Plat		Placl	Timml7b
Rps20	2700094K13Rik		Cenpa	Tmp0	Sec6lb		Colllal	Cdkn2a	Plau		Igf2as	Gml0491
Vgll3	Fmrlnb		Hmmr	Nusapl	Erp44		Npnt	Cdkn2b	Ctsb		Usp9x	Gml0490
Prrl5	Hmgn2		Ctcf	Ncapd2	AI314180		Cyr61	Jun	leaml		Psg28	Pcskln
Fbxl7	Ubald2		Psrcl	Mcm2	Jun		B4galtl	Slc35dl	leam3		Bmp8b	Eras
Maged2	Lactb2		Cdc25c	Kif2c	Casp9		Reck	Plk3	Tnfrsfllb		Fnl	Hdac6
Galntl4	Folrl		Nek2	Cdca2	Fbxo6		Tgfbrl	Rnfl9b	Tnfrsfla		Psg23	Gatal
Pdgfc	Gm7325		Gas213	NasP	Fbxo2		Col27al	Sfn	Tnfrsflb		Bmp8a	Glod5
Tmtc4	Agtrap		G2e3	Gmnn	Ube4b		P3hl	Fuca1	TnfrsflOb		Psg21	Gml4820
Tmtc3	Sppl			Cdc6	Ube2j2		Hspg2	Epha2	Fas		Dusp9	Suv39h1
Lpar4	Hells			Pold3	Psmc2		Vwal	Wrap73	Plaur		H19	Was
Pcdhl9	Dppa4			Ckap2l	Tmubl		Dnajb6	Mxd4	Il6st		Tmem37	Wdrl3
Eda2r	Gabarapl2			Fam64a	Tmeml2		Emilinl	Rchy1	Egfr		Mmpl5	Rbm3

					9							
Pcdhl8	Rhox6			Ubr7	Wfsl		Mpvl7	Iscu	Fnl		FamlOlb	Rbm3os
Gprl76	Rhoxl			Fen1	Ube2k		Apbb2	TriaPi			Phfl6	Tbcld25
LoclOO503471	Cdc5l			Bub1	Tbl2		Pdgfra	Prkabl			4930422n03rik	Ebp
Mical2	Texl9.1			Brip1	Get4		Ambn	Trafdl			Ada	Porcn
Dzipll	Trim28			Atad2	Bhlhal5		Dmpl	Pom121			Mmpla	Ftsjl
Hoxc6	Atp5gl			Psrc1	Creb312		Ibsp	Pdgfa			Gprl26	Slc38a5
Hoxc5	Sox2			Rrm2	Pdia4		Tfipll	Gadd45a			Arf2	SsxblO
Mettl4-psl	Jam2			Tipin	Eif2ak3		Eln	Vamp8			Tinagll	Ssxb9
Sec63	Fkbp3			Casp8aP2	Rnfl03		Plod3	Retsat			Mfi2	Ssxbl
Ikbip	Cox7b			Tubb4b	Aupl		Colla2	Tprkb			Rpn2	Ssxb2
Tsc22d2	Ash2l			Kif23	Itprl		Ndnf	Tgfa			Abhd2	Gml4459
2310076g05rik	Dut			Exo1	Edeml		Vhl	Mxd1			Hrctl	Ssxb6
Anxa6	Dtymk			Rfc2	Bbc3		Mfap5	Sec6lal			Adm	Ssxb3
Nfatc4	Gpx4			Pola1	Psmc4		Ercc2	Xpc			Abhd6	Ssxb8
Fnl	Eif4ebPi			Mki67	Bax		Bcl3	Ccnd2			Slc7al	Ssx9
Wnt9a	Morel			Tpx2	Ppplrl5a		Tgfbl	H2afj			Tead4	Ssxb5
Sorcs2	Fabp3			Aurka	Vimp		Mia	Ldhb			Mbnl3	Gm6592
Tmeffl	Zfp428			Anln	Rnfl21		Spint2	LrmP			Gprl	Gm5751
C79491	Aqp3			Chafl b	Anks4b		Aplpl	Tm7sf3			2900057el5rik	B630019K06Rik
Crlfl	Grhpr			HjurP	Ern2		Hpn	Tgfb1			Ldocl	Fthll7b
2610034e01rik	Higdla			Tacc3	Atp2al		Klk4	Sertad3			Adaml9	Fthll7c
Gjd4	Rpp25			Mc	Brsk2		Acan	Ceb			Rybp	Fthll7

				m5				pa				d
Ccngl	Rbpms			Anp32e	Ins2		Serpinhl	Klk8			Col4al	Fthll7e
Gprl24	Mmp3			DlgaP5	Ccnd1		Apbbl	Bax			Fndc3c1	Fthll7f
Fibin	Apobec3			Ect2	Map3k5		Ilk	Ppplrl5a			Col4a2	4930402K13Rik
8030476ll9rik	Spc24			Nuf2	Nrbf2		Ric8	Rpll8			4930502el8rik	Lancl3
Ddr2	Xlr3a			Cdc45	Derl3		Muc5ac	Aen			Pkn2	Gml4862
Arf4	Recll4			CkaP5	Ube2g2		Ctgf	Rrp8			Rlim	Xk
Ptprs	Mtf2			Ctcf	Tmem259		Nr2el	Ccp110			160001SilOrik	1700012L04Rik
Sprr2k	Snrpn			Clspn	Creb313		Nepn	Nuprl			Afp	Gml4501
Adm	Gml3580			Cdca7	Hsp9Obi		P4hal	Ptpre			Tmeml40	Cybb
A830029e22rik	Gmnn			Cdca3	Apaf1		Spock2	Hras			Fstl3	Gm5132
9230114kl4rik	Chmp4c			Rpa2	Ifng		Adamtsl4	Eps812			Ing4	Dynlt3
Extl3	Hsf2bp			Gins2	Os9		Mmpll	Ctsd			Taf7l	Hypm
Mecom	Polr2e			E2f8	Ddit3		Coll8al	Cd81			Sultlel	4930557A04Rik
Qsoxl	Blvrb			Cdc25c	Erlin2		Myf5	Perp			Olrl	Sytl5
Teadl	Ldhb			Nek2	Ppp2cb		Col4al	Rps12			2610019f03rik	Srpx
Snx7	Apocl			Cdc20	Ubxn8		Csgalnact1	Tpd5211			Fll	Rpgr
Cdkl4	Syngrl			Rad51aPi	Casp3		Comp	Sesn1			Fbxw8	Otc
Cdkn2a	Bexl				Pik3r2		Gfod2	Foxo3			Sema4c	Tspan7
Cdkn2b	Nr2c2aP				Amfr		Has3	Ddit4			Ctnnbip1	Gml0489
Ccnyll					Herpudl		Atxnll	Zfp365			Tfpi2	Midlip1
Tubb2a-ps2					Aars		Crispld2	Prmt2			ZbtblO	Gml4493

Aen					Selk		Foxfl	Mknk2			Mitf	Gm14483
Farpl					**Eroll**		Foxc2	Draml			Gpr50	Gm14474
4930402h24rik					Psmc6		Agt	Apaf1			Hic2	Gm14477
Sh3rf3					Trim13		Exoc8	Btgl			Tpbpb	Gm14476
Adaml9					Dnajc3		Eroll	Mdm2			Slc9a6	Gm14484
Ddbl					Casp4		Lgals3	Ddit3			Prl7dl	Gm14479
Cttn					Casp12		Ripk3	Gls2			Tpbpa	Gm14482
9230112e08rik					ScamP5		Loxl2	Dgka			Slco2al	Gm14478
Dbnl					Pml		Lcpl	Cdkn2aiP			Pkp2	Gm14475
Fyttdl					Parp16		Mmpl3	Hmoxl			9630050el6rik	Gm4906
Lrrcl5					Nckl		Mmp20	Rrad			Pvrl2	Bcor
Fkbpl0					Uba5		Col5a3	Cdh13			Zfp568	Gm14635
Trubl					Uspl9		Smarca4	Osginl			Vtcnl	Atp6ap2
Zdhhc20					Stt3b		Aplp2	Cgrrfl			Il6ra	1810030O07Rik
Stonl					Rnfl85		Mpzl3	Abhd4			Foxo4	Med14
Hoxdl3					Xbpl		Thsd4	Kifl3b			Hsp90b1	Usp9x
Nudt6					Erlec1		Anxa2	Rbl			Prl7cl	2010308F09Rik
Hoxdl2					Stc2		Myole	Nudtl5			Prl6al	Ddx3x
Prss23					Trp53		Nphp3	Tsc22dl			Cdh5	Nyx
9430030nl7rik					Aloxl5		Dagl	Casp1			Fgd6	Cask
Arntl2					Derl2		Lamb2	Stl4			Cysltr2	Gpr34
Sh3rfl					Trim25		Kif9	Ei24			Rhox6	Gpr82
Mrc2					Cdk5rap3		Sh3pxd2b	Vwa5a			Cdh3	Gm5382

Mdhl					Cede47		Adamts2	Zbtb16			Spp2	Gml4505
Rictor					Psmc5		Wnt3a	Rps271			Ziml	Drrl
Map4k5					Ernl		Mfap4	Mapkapk3			Flnb	Cyptl
Plcll					Nploc4		Serpinf2	Ip6k2			Rbbp7	Maoa
Septll					P4hb		Vtn	Tcn2			Map3k7	Maob
Ryk					Txndc5		Nfl	Lif			Rhox9	Ndp
Tgfb3					Faf2		Collal	Upp1			Whscll1	Efhc2
Ube2i					Ubqlnl		Ramp2	Ceng1			Slc38al	Fundcl
Tgfb2					Atgl0		Gfap	CyfiP2			1600012pl7rik	Dusp21
Zfp319					Thbs4		Sox9	Gnb211			Adra2b	Kdm6a
GmlO399					Col4a3bp		Erollb	Hint1			Pgf	4930578C19Rik
Fbxol7					Pik3r1		Nidi	Gm2a			1200009i06rik	Gm26652
Wnt5a					Pdia6		Foxf2	Hist3h2a			Mfsd7c	BC049702
Criml					Dnajb9		Foxcl	Alox8			Esam	Chst7
Midi					Tmxl		Ripkl	Trp53			Gprl07	Slc9a7
Displ					JkamP		Tfap2a	Taxibp3			Au015791	Rp2
Ubox5					Selll		Ecm2	Traf4			Arhgap8	Jade3
St7l					Psmc1		B4galt7	Cdk5rl			Ankrdl7	Rgn
Col5a2					Atxn3		Tgfbi	PpmI d			Cul7	Ndufbl1
Axl					Derll		Pxdn	Rad51c			2310067p03rik	RbmlO
Col5al					Rnfl39		Smocl	Tob1			Irs3	Ubal
Zyx					Foxred2		Ltbp2	Krtl7			Prl5al	Cdkl6
Ror2					Pla2g6		Flrt2	Hexim l			Fntb	Uspll
Wdfy3					Atf4		Fbln5	Fdxr			Tceanc	Araf
Amotl					Ep30		Egflam	Itgb			Lepr	Synl

2					0			4				
Yapl					Tmbim6		Tnfrsfllb	Sphk1			Tnfrsf9	Timp1
Phldb2					Txndell		Col 14a 1	Rhbdf2			Papola	Cfp
6330562c20rik					Sdf2l1		Has2	BaiaP2			Srd5al	Elk1
Ctnnd1					Ufdll		Ptk2	Dcxr			Clqtnfl	Uxt
Rock2					Eif2b5		Sex	Histlhlc			Slc38a4	Zfp182
Maspl					Nrros		Fblnl	Ninj1			Angpt4	Spaca5
Pvtl					Pdia5		Adamts20	Nol8			Ctla2a	Zfp300
Tnc					Gsk3b		Col2al	F2r			9930012kllrik	Ssxa1
Fbln2					Park2		Myhll	Ankra2			Mical3	Gm21876
Hdlbp					Stub1		Ccdc80	Plk2			Apoa4	4930453H23Rik
AtplOa					Pdia2		Abi3bp	Sdcl			Cul4b	Gm6938
Loxll					Crebrf		App	Gpx2			3632454l22rik	Gm26593
Loxl2					Bakl		Seracl	Zfp3611			Psg-psl	Agtr2
Fbln5					Rnf5		Pig	Fos			Lcor	Slc6a14
Ctgf					Atf6b		Smoc2	Ccnk			Tnfrsf22	Gm28269
Efnb2					Bag6		Hasl	Jag2			Tnfrsf23	Gm28268
Rxra					Flotl		Noxol	Ndrgl			Sosl	Klhl13
Ccnd2					Eif2ak2		Collla2	Pmml			Dlx3	Wdr44
Gpc2					PmaiPi		Tnxb	Plxnb2			Ippk	Gm4907
Ntf3					Tmx3		Tnf	Vdr			Htr2b	Gm4985
Kif5b					Syvn1		2300002M23Rik	CsrnP2			Duspl6	Gm27192
Slit2					Erlin1		Flotl	Acvrl b			Cdc73	Gm5934
Tpml							Hsp90ab1	Spl			1700025g04rik	Gm4297
Gpc4							Washl	Abat			Prl4al	Gm593

												5
Flnb							Vit	Socs1			Zfp655	Gm5169
4930555bllrik							Cyplbl	Abcc5			Slcl3a4	Gml993
Fine							Fshr	Trp63			Ceacam14	E330010L02Rik
C76332							Mkx	Fam162a			Ceacam15	Gm5168
Capn2							Lox	App			Trapla	Gm2012
Phlda3							Hpse2	Rab40c			Ceacam12	Gm2030
Map3k7							Kazaldl	Bak1			Gml6515	Six
MyhlO							Nfkb2	Def6			Ceacam13	Gml4525
D18ertd653e								Cdknla			4930447f24rik	Gm6121
Stox2								Tap1			Gzmd	Gml0230
Igf2r								Ier3			Foxj2	Gm2101
D15ertd621e								Polh			Fbxll9	GmlOO58
Arid5b								Ccnd3			Gzmc	Gm2117
TnfrsflOb								Hbegf			Gzmf	Gm4836
26100lle03rik								Hdac3			Gzme	GmlOl47
Ckap4								Rad9a			Gzmg	Gm2165
Efna2								Ctsf			Patl2	GmlOO96
Picalm								Slc3a2			3830417al3rik	Gm2200
CdhlO								Fas			Tspanl4	Gm26818
Ddahl											Handl	Gm3669
Uba3											AtxnlO	Gml0488
0610038b21rik											Mgat4a	E330016L19Rik

Gemin7											Unc50	Gm14632
Ubal											Il2rb	Gm7437
Fbnl											Ceacam11	Gm14974
Lhx9											Plekhgl	Gm10487
Eif4g2											Prl3bl	Gm21447
Vcl											Folrl	Spin2f
Bcl2l2											A83008OdOlrik	Gm2784
Cd276											Blzfl	Gm2777
Lrrc58											Zfp667	Gm21883
Wwc2											Fltl	Spin2e
Lpp											Usp27x	Gm21608
Aril											Hdac4	Gm21637
Ltbpl											Itgb3	Gm21645
Ltbp2											Sri	Gm2799
Wispl											Sema3f	Gmcl1l
Igflr											Prl3al	Gm5926
Rhobtb3											Bahdl	Gm21951
Faml98b											Sin3b	Gm21657
Cnn2											Gm2a	Gm21789
Glipr2											Serpinb9g	Gm2825
Sydel											Bend4	Spin2-ps6
Hhat											Bend5	Gm2863
Zmat3											Serpinb9b	Gm2854
Caldl											Serpinb9c	Gm2913
Pmepa1											Serpinb9d	Gm2927
E130112l23rik											Plekhhl	Gm2933

Bag2											221001lc24rik	Gm2964
Zfp583											Cd320	Gm21870
Pibfl											Ccnjl	Gm21681
Pmaip1											Entpd2	Spin2g
A130022jl5rik											Illr2	Gm21699
Bcl9l											Sfmbt2	Gm14552
Cpa6											170001lm02rik	Gm10486
D13ertd787e											Plekha7	Gm2309
Pabpc4 1											Sfrp5	Gm14553
Zfhx3											Ppplr3f	Gm14819
Itga5											Obsll	Dock11
Txnrdl											Slc23a3	Il13ra1
Htrlb											Tmem87b	Zcchc12
Hmga2											Epasl	Lonrf3
Sept2											Ccdc68	Gm6268
Lambl											Kdelr2	Gm14569
Zfp518b											Pramef12	Pgrmc1
Parva											Lrp8	Akap17b
Gulpl											Pard6b	Slc25a43
Shank1											PeglO	Slc25a5
Bmpl											N4bp2	Gm14549
Aktlsl											Pla2g4e	2310010G23Rik
Itga9											Fam78b	C330007P06Rik
Abccl											Arrdc3	Ube2a
Eda											Pla2g4d	Nkrf
B4galt											Rassf8	Gm150

2												08
Nidi											Au015836	Sept6
Ncaml											Csnkle	Sowahd
Shc2											Stagl	Rpl39
Uba6											Vnnl	Upf3b
Tradd											Tchhll	Nkap
Rtell											Plala	Akapl4
Bicd2											Slc45a4	Ndufal
Adamtsl2											Tex264	Rnfll3al
Hs2stl											Pcdhl2	Gm9
DlOertd610e											Ctr9	Rhoxl
Cyr61											Ccrlll	Rhox2a
Gtf3cl											Htatsfl	Rhox3a
Lbh											9030409gllrik	Rhox4a
Krt33b											Tspan9	Rhox3a2
Gm6607											Rassf6	Rhox4a2
D3wsu167e											4631402f24rik	Rhox2b
Zc3h7b											A2m	Rhox4b
7630403g23rik											Rimklb	Rhox2c
Tnpo2											LoclOO504569	Rhox3c
Cepl70											Apob	Rhox4c
Pdlim5											Tmeml50a	Rhox2d
Pdlim7											9130404d08rik	Rhox4d
Cad											Prl8a6	Rhox2e
Unc5b											Cts6	Rhox3e
2410018ll3rik											Prl8a8	Rhox4e
LoclOO216343											Prl8a9	Rhox2f
Glrx3											Cts3	Rhox3f

Kctd5											Krtl8	Rhox4f
Loc269472											Nrnll	Rhox3g
Myolc											Sfil	Rhox2g
4930562cl5rik											Tlr5	Rhox4g
Till											Rhou	Rhox3h
Sema3a											Arhgef6	Rhox2h
Itgbl											Tmeml85b	Rhox5
Nxn											Tram2	Rhox6
Tmem41b											Citedl	Rhox7a
Sec23a											Cited2	Rhox8
Gm22											Zfand2a	Rhox7b
Itgb5											Krt25	Rhox9
Dysf											Klk4	Btgl-psl
Thbsl											Tnfrsfll b	Btgl-ps2
Bc022687											2010204kl3rik	RhoxlO
Dnm3os											Torlaip2	Rhoxll
Rnd3											Fmrlnb	Rhoxl2
Pik3c2a											Ctsr	Rhoxl3
2810008m24rik											Ctsq	Zbtb33
Spred3											Prl8a2	Tmem255a
Senp5											Ctsm	Atplb4
Arll3b											Prl8al	Lamp2
Polr2e											Ctsj	Gm7598
Itgav											Mpzll	Cul4b
Igf2bp3											Stra6	Mctsl
											Bcap31	Clgaltlcl
											Cregl	Gml4565
											Tcfap2c	603049

												8E09Rik
											Prl7bl	Cyptl5
											Ghrh	Cyptl4
											4930486l24rik	Gria3
											Neurog2	Thoc2
											5430425jl2rik	Xiap
											Prl7al	Stag2
											Prl7a2	Gm43337
											Mirll99	Sh2dla
											TbcldlOa	Tenml
											Ralbpl	Gm362
											Pdgfra	Dcafl212
											Morc4	Dcafl211
											Rarres2	Prr32
											Arid3a	4930515L19Rik
											Lifr	Actrtl
											Shisa3	Gm29242
											Uevld	Smarca1
											Scnnlb	Ocrl
											Dnajbl2	Apln
											Brwd3	XpnpeP2
											Hhipll	Sash3
											Fbln7	Zdhhc9
											Maspl	Utpl4a
											Nrk	9530027J09Rik
											Pvr	Bcorll
											Atp2cl	Elf4
											Amot	Aifml
											1600014k23rik	Rab33a

											Tbrgl	Zfp280c
											Slitl	Slc25a14
											A730090h04rik	Gprll9
											4931406pl6rik	Rbmx2
											Opn3	Gm595
											Pdia4	Enox2
											B930054o08	Gml4696
											1700031f05rik	Gml4697
											Inhba	Arhgap36
											Inhbb	Olfrl320
											Helz	Olfrl321
											Sele	Igsfl
											Pdia6	Olfrl322
											Pdia5	Olfrl323
											Creb3	Olfrl324
											Efnal	Stk26
											Dlg5	Frmd7
											Procr	Rap2c
											Fgfrl	Mbnl3
											Gnb4	Hs6st2
											2310030g06rik	Usp26
											Gcml	1700080O16Rik
											Psgl8	Gpc4
											Goltlb	Gpc3
											Psgl9	Gml4582
											Psgl6	A630012P03Rik
											Slc2al	Ccdcl60
											Psgl7	Phf6

											Htra3	Hprt
											Klhll3	Gm28730
											Ets2	Placl
											Nppc	Faml22b
											Tgml	Faml22c
											Tmeml08	Mospd1
											Usp53	Etd
											Mark3	Gml4597
											Cbx8	Cxxlc
											Hspa5	Cxxla
											Spats2	Cxxlb
											Limk2	4930502E18Rik
											Mkl2	1700013H16Rik
											Shroom4	Zfp36l3
											Shroom1	Xlr
											Pou2f3	Gml6405
											Acvr2b	Gml6430
											Rbms2	Slxll
											Atg4b	3830403N18Rik
											Pappa2	Gm773
											Rbm25	1600025M17Rik
											Gm4793	Zfp449
											Nidi	Gm2155
											Uba6	Smiml012a
											Lamcl	Gm2174
											Slc40al	Ddx26b

											Hapln3	Gml0477
											Faml76a	Gm648
											Pdliml	Mmgtl
											Ube2q2	Slc9a6
											Au018091	Fhll
											Bdkrb2	Mtap7d3
											E130203bl4rik	Adgrg4
											SlOOg	Brs3
											4933402el3rik	Htatsfl
											Dapk2	Vglll
											Gmll985	Gml4718
											Fndc3b	Cd40lg
											Twsgl	Arhgef6
											Aldhla3	Rbmx
											Lnx2	Gm364
											Taf7	GprlOl
											Ai844869	Zic3
											Clecl2b	4930550L24Rik
											Prkcsh	Fgfl3
											Lama5	F9
											Tchh	Mcf2
											Lamal	Atpllc
											Rps6ka6	Gm7073
											Vhl	Gml4661
											Eps8l2	Sox3
											Polg	Gml4662
												Gml4664
												Cdrl
												Ldocl
												4933402E13Rik

												4931400O07Rik
												1700019B21Rik
												Gm6760
												3830417A13Rik
												Slitrk4
												Ctag2
												4930447F04Rik
												Slitrk2
												1700036O09Rik
												Gmll40
												Gml4692
												4933436l01Rik
												Fmrlos
												Fmrl
												Fmrlnb
												Gml4698
												Gm6812
												Gml4705
												Aff2
												170011lN16Rik
												1700020N15Rik
												Ids
												1110012L19Rik
												4930567H17Ri

												k
												BC023829
												Mamld1
												Mtml
												Mtmrl
												Cd99l2
												Gml6189
												Hmgb3
												Gpr50
												Vma21
												Gmll41
												Prrg3
												Fatel
												Cnga2
												Magea4
												Gabre
												Magea10
												Gabra3
												Gabrq
												Cetn2
												Nsdhl
												Gml4684
												Zfpl85
												Pnma5
												Pnma3
												Xlr4a
												Xlr3a
												Xlr5a
												Gml4685
												DXBay18
												Xlr5b
												Spin2d
												Xlr3b
												Xlr4b
												F8a

												Xlr4c
												Xlr3c
												Xlr5c
												RP23-95K12.13
												Zfp275
												Gml8336
												Gm26726
												Zfp92
												Trex2
												Haus7
												Bgn
												Atp2b3
												Dusp9
												Pnck
												Slc6a8
												Bcap31
												Abcdl
												Plxnb3
												Srpk3
												Idh3g
												Ssr4
												Pdzd4
												Llcam
												Arhgap4
												Avpr2
												NaalO
												Renbp
												Hcfcl
												Iraki
												Mecp2
												Opnlmw
												Tex28
												Tktll
												Flna
												Emd
												RpllO
												Dnasel

												11
												Taz
												Atp6ap1
												Gdil
												Fam50a
												Plxna3
												Lage3
												Ubl4a
												SlclOa3
												Fam3a
												Ikbkg
												G6pdx
												Gm6880
												Olfrl326-psl
												Olfrl325
												Gm5640
												Gm6890
												Gm5936
												Gab3
												Dkcl
												Mppl
												Smim9
												F8
												Fundc2
												Cmc4
												Mtcpl
												Brcc3
												Vbpl
												Gml5384
												Rab39b
												Gml5063
												Pls3
												Gml47

												15
												Gml4707
												Gml4717
												Cldn34b3
												Cldn34b4
												Cldn34d
												Tbllx
												Prkx
												Gml4742
												Pbsn
												Gml4744
												5430402E10Rik
												Obpla
												Gm5938
												Obplb
												Gml4743
												493048OEllRik
												Prrgl
												Fam47c
												Gm7173
												Mageb16
												Gm26775
												Tmem47
												4930595M18Rik
												Dmd
												Tsga8
												Fthll7a
												Tab3

												Gk
												Gml4764
												Gml4762
												5430427019Rik
												Samt3
												NrObl
												Mageb4
												Illrapl1
												Gm27000
												Pet2
												4932429P05Rik
												4930415L06Rik
												Gm44
												Gml4773
												Mageb2
												Gm5072
												Gm8914
												1700084M14Rik
												Gml4781
												Mageb5
												Mageb1
												Mageb18
												Gm5941
												1700003E24Rik
												BC061

												195
												Arx
												Polal
												Pcytlb
												Pdk3
												AU015836
												Gml4798
												Zfx
												Eif2s3x
												Klhll5
												Fam90alb
												Apoo
												Gml4827
												Maged1
												Gspt2
												Zxdb
												RP23-9K14.6
												Gm26617
												Spin4
												Arhgef9
												Amerl
												Asbl2
												Zc4h2
												Zc3hl2b
												170001ODOIRik
												Lasll
												Msn
												F630028O10Rik
												Vsig4
												Hsf3
												Heph
												Gprl65

												Pgrl5l
												Eda2r
												Ar
												Ophnl
												Yipf6
												Stard8
												Efnbl
												Gml4812
												Gml4809
												Gml4808
												Pjal
												Tmem28
												Eda
												Awat2
												Otud6a
												Igbpl
												Dgat2l6
												Awatl
												P2ry4
												Arr3
												Pdzdll
												Kif4
												Gdpd2
												Gml4902
												Dlg3
												Texll
												Slc7a3
												Snxl2
												Foxo4
												Gm614
												Gm20489
												H2rg
												Medl2
												Nlgn3
												Gjbl
												Zmym3

												Nono
												ItgblbP2
												Tafl
												Ogt
												Cxcr3
												Gm4779
												8030474K03Rik
												Nhsl2
												Rgag4
												Pin4
												Ercc6l
												Rps4x
												Citedl
												Hdac8
												Phkal
												Gm9112
												Dmrtclb
												Dmrtclcl
												Dmrtclc2
												170003lF05Rik
												Dmrtcla
												1700011M02Rik
												Napll2
												Cdx4
												Chicl
												Gm26952
												Tsx
												Gm26992
												Tsix
												Xist
												Jpx

												Ftx
												Zcchcl3
												Slcl6a2
												Rlim
												C77370
												Abcb7
												Uprt
												Zdhhcl5
												170012lL16Rik
												Magee2
												Pbdcl
												Magee1
												5330434G04Rik
												Cypt2
												Fgfl6
												Atrx
												Magtl
												Cox7b
												Atp7a
												Tlrl3
												Pgkl
												Taf9b
												Fnd3c2
												Fndc3c1
												Cysltrl
												Gm5127
												Zcchc5
												Lpar4
												P2rylO
												A630033H20Rik
												Gprl74

												Itm2a
												Tbx22
												2610002M06Rik
												Fam46d
												Gm732
												Gm379
												Brwd3
												Hmgn5
												Sh3bgr1
												Gm6377
												RP23-240M8.2
												Pou3f4
												Cylcl
												GmlOl12
												Rps6ka6
												Hdx
												RP23-466J17.3
												Texl6
												4933403O08Rik
												Apool
												Satll
												2010106E10Rik
												Zfp711
												Poflb
												Gml4936
												Chm
												Dach2
												Klhl4
												Ube2dnil

												Ube2dnl2
												4930555B12Rik
												Cpxcrl
												H2afb2
												Gml4920
												Gm28579
												Tgif2lx2
												Tgif2lx1
												Gml4929
												Pabpc5
												Pcdhllx
												H2afb3
												Napll3
												Gml7521
												Cldn34cl
												Astx6
												Srsx
												Gml7577
												Gml4951
												Astx2
												Gml7412
												Cldn34c2
												Gml4950
												Gml7467
												Cldn34c3
												Astx5
												Vmn2r121
												Astxla
												Gml75

												84
												Astx4a
												Gml7469
												Astx4b
												Astxlb
												Gml7361
												Gm21616
												Astx4c
												Gml7693
												Astxlc
												Gml7522
												Astx4d
												Gml7267
												Astx3
												493241lN23Rik
												Gm382
												492151lC20Rik
												Cldn34c4
												4930558G05Rik
												Diaph2
												Pcdhl9
												Gm26851
												Tnmd
												Tspan6
												Srpx2
												Sytl4
												Cstf2
												Noxl
												Xkrx
												Arll3a
												Trmt2b

												Tmem35
												Cenpi
												Drp2
												Taf7l
												Timm8al
												Btk
												Rpl36a
												Gla
												Hnrnph2
												Armcx4
												Armcx1
												Armcx6
												Armcx3
												Armcx2
												Nxf2
												Zmatl
												Gml5023
												Tceal6
												Pramel3
												Gm5128
												Gm7903
												AV320801
												Nxf7
												Prame
												Tcpllx2
												Tmsbl5a
												Armcx5
												Gprasp1
												Bhlhb9
												Gprasp

												2
												Arxes2
												Arxesl
												Bex2
												Nxf3
												Bex4
												Tceal8
												Tceal5
												Bexl
												Tceal7
												Wbp5
												Ngfrap1
												Kir3dl2
												Kir3dll
												Tceal3
												Tceall
												Morf4l2
												Glra4
												Plpl
												Rab9b
												H2bfm
												Tmsbl5 1
												Tmsbl5b2
												Tmsbl5bl
												Slc25a53
												Zcchcl8
												Faml99x
												Esxl
												Illrapl2
												Texl3a
												Nrk
												Serpina7
												4930513O06Rik

												4933428M09Rik
												Mumll1
												Trapla
												D330045A20Rik
												Rnfl28
												Tbcld8b
												Gml5013
												Ripplyl
												Cldn2
												Morc4
												Rbm41
												Nup62cl
												Pihlh3b
												Gml5046
												Frmpd3
												Prpsl
												Tsc22d3
												Mid2
												Eif2c5
												Texl3
												Vsigl
												Psmdl0
												Atg4a
												Col4a6
												Col4a5
												Irs4
												Gml5295
												Gml5294
												Gml5298
												Gucy2f

												Nxt2
												Kcnell
												Acsl4
												Tmem164
												Ammecrl
												Rgagl
												Chrdll
												Pak3
												Capn6
												Dcx
												A730046J19Rik
												Algl3
												Trpc5
												Trpc5os
												Zcchcl6
												Lhfpll
												Amot
												Htr2c
												Ill3ra2
												Lrch2
												Gml5128
												Gml5080
												Gml5107
												Gml5114
												Gm8334
												Gml5127
												Luzp4
												Gml5099
												Ott
												Gml5092
												Gml5093

												Gml5100
												Gml5085
												Gml5086
												Gml0439
												Gml5097
												Gml5091
												Gml5104
												Tmem29
												Apex2
												Alas2
												Pfkfbl
												Tro
												Maged2
												Gm27191
												Gnl3l
												Fgdl
												Tsr2
												Gml5138
												Wnk3
												A230072E10Rik
												Faml2Oc
												Phf8
												Huwel
												Hsdl7blO
												Ribcl
												Smcla
												Iqsec2
												Kdm5c
												Kantr
												Tspyl2
												Gprl73

												Cldn34a
												Shroom2
												Gprl43
												Usp51
												Mageh1
												Foxr2
												Rragb
												Klf8
												Ubqln2
												Cypt3
												Kctdl2b
												RP23-106P7.5
												2210013021Rik
												Spin2c
												Samtl
												4921511M17Rik
												GmlOO57
												Gml5140
												4930524N10Rik
												Samt4
												Samt2
												Cldn34bl
												Magea6
												Magea3
												Magea8
												Magea2
												Magea5
												Magea

												1
												Cldn34b2
												Satl
												Acot9
												Prdx4
												Ptchdl
												Gml5156
												Gml5155
												Phex
												Sms
												Mbtps2
												Yy2
												Smpx
												Gml5169
												Klhl34
												Cnksr2
												Rps6ka3
												Eiflax
												Map7d2
												A830080D01Rik
												Sh3kbp1
												Map3k15
												Pdhal
												Adgrg2
												Gml5241
												Phka2
												Gml5243
												Ppefl
												Rsl
												Cdkl5
												Gja6
												Scml2

												Gml5262
												Rai2
												Scmll
												Gml5205
												Nhs
												Gml5202
												Reps2
												Rbbp7
												Txlng
												Syapl
												Ctps2
												SlOOg
												Grpr
												Rnfl38rtl
												Apls2
												Zrsr2
												Car5b
												Siahlb
												Tmem27
												Ace2
												Bmx
												Pir
												Figf
												Piga
												Asbll
												Asb9
												Mospd2
												Fancb
												Gml7604
												Glra2
												Gemin8
												Gpm6b
												Ofdl
												Trappc2
												Rab9

												Tceanc
												Egfl6
												Gml5226
												Gml720
												Gml5230
												Gm8817
												Gml5232
												Gml5228
												Tmsb4x
												Tlr8
												Tlr7
												Prps2
												Gml5239
												Frmpd4
												Msl3
												Arhgap6
												Gml5261
												Amelx
												Hccs
												Gml5245
												Midi
												493340OAllRik
												Gml5726
												Gml5247
												Gm21887
												Asmt

[0148] As an additional validation, we modified an existing trajectory finding technique, Wishbone(SlO) — based on shortest paths in k-NN graphs—to include information about time and proliferation. This gives trajectories whose overall shape agrees with the transports displayed in **FIG.** 8A.

Learning gene regulatory networks

[0149] How to set up an optimization problem to solve for a regulatory function that fits the transport maps is described above.

[0150] In order to make this concrete, a function class F was specified over which to optimize. Consider a rectified-linear function class defined in terms of a specific generalized logistic function

$$\ell(x;k,b,y_0,x_0) = \frac{\quad}{y_0} \frac{k y \circ}{+ \{k - yo)e^{-b(x-x_0)}},$$

where k, b, yO, xO ∈ R are parameters of the generalized logistic function l(x). A function class F is defined consisting of functions f : RG → RG of the form

$$f(x) \;\; = U\pounds\{WTx),$$

where **1**is applied entry-wise to the vector WZx ∈ R^M to obtain a vector that we multiply against U ∈ RGxM . Here T ∈ RGTF ×G denotes a projection operator that selects only the coordinates of x that are transcription factors, and GTF is the number of transcription factors.

[0151] The following optimization over matrices U ≡ RGxM and W ≡ RMxGTF

$$\begin{aligned} \min_{U,W} \quad & \mathbb{E}_r \left\| \frac{X_{t_i} - X_{t_{i+1}}}{\Delta_t} - U\ell(WTX_{t_i}) \right\|^2 + \eta_1 \|U\|_1 + \eta_2 \|W\|_1, + \eta_3 \|W\|_2^2 \\ \text{s.t.} \quad & U \geq 0. \end{aligned}$$

where (Xti , Xti+i) is a pair of random variables distributed according to the normalized transport map *r* and // U // i denotes the sparsity-promoting l_1 norm of U, viewed as a vector (that is, the sum of the absolute value of the entries of U). Each rank one component (row of U or column of W) gives us a group of genes controlled by a set of transcription factors. The regularization parameters ηι and η_2 control the sparsity level (i.e. number of genes in these groups).

[0152] **Implementation:** A stochastic gradient descent algorithm was designed to solve [10]. Over a sequence of epochs, the algorithm samples batches of points (Xti , Xti+i) from the transport maps, computes the gradient of the loss, and updates the optimization variables U and

W. The batch sizes are determined by the Shannon diversity of the transport maps: for each pair of consecutive time points, the Shannon diversity S was computed of the transport map, then randomly sample max(S x 10-5, 10) pairs of points to add to the batch. We run for a total of 10, 000 epochs.

[0153] This algorithm was implemented in Python.

7. Clustering cells

[0154] Cells were clustered using the Louvain-Jaccard community detection algorithm (SI 9-S21) in 20 dimensional diffusion component space. This algorithm maximizes the Louvain modularity — a value between - 1 and 1 that measures the density of links inside communities compared to links between communities.

[0155] As a first step, the 20-nearest neighbor graph in 20 dimensional diffusion component space (computed on cells from both 2i and serum) were computed. The edges are weighted in this graph by the Jaccard similarity coefficient. The resulting graph was partitioned into clusters using the Louvain community detection algorithm (SI 9) implemented in the function multilevel. community from the R pack- age IGRAPH (1.0.1) (S22). The default parameters for automatically selecting the number of clusters gave us 33 clusters, displayed in **FIG. 7D.**

8. Gene correlation modules reveal biological signatures

[0156] In this section technique for identifying modules of correlated genes are described, with the goal of revealing coherent biological processes.

[0157] The procedure consists of two steps. In the first step, the Graphical Lasso (S23) was used to compute a regularized estimate of the covariance matrix for the 66,000 expression profiles. The Graphical Lasso fits a covariance matrix to the data, regularized so that the inverse of the covariance matrix is *sparse* (i.e. has only a few non-zeros). The motivation for selecting a sparse inverse covariance is based on the fact that if a collection of observations have a multivariate Gaussian distribution with mean μ and covariance Σ, then the zero pattern of Σ- 1 completely specifies the conditional independence structure of the observations:

$\Sigma_{ij}^{-1} = 0 \iff$ **variables i and j are conditionally independent given the other variables.**

Let $\Theta = \Sigma^{-1}$ and let S denote the empirical covariance for our expression profiles

[0158] The Graphical Lasso maximizes the Gaussian log likelihood:

$$\underset{\Theta}{\text{maximize}} \quad \log\det\Theta - \mathrm{tr}(S\Theta) - \rho\|\Theta\|_1 .$$

Here $\|\Theta\|_1$ is a regularization term that promotes sparse solutions. The optimal Θ is a (regularized) maximum-likelihood estimate of the inverse covariance matrix Σ^{-1} for a Gaussian ensemble.

[0159] Gene modules were identifed as tightly knit communities in the network specified by Θ (see below). Based on these gene modules, we then identified gene signatures related to specific pathways, cell types, and conditions. We did this by functional enrichment analysis (see below). The gene modules are displayed in **FIG. 13.**

[0160] Computing gene modules: The glasso package was used (S23) to solve the graphical lasso optimization problem. The regularization parameter p was tuned to achieve a desirable sparsity level for Θ. In particular, we select a value of p that gave around 10, 000 total genes (i.e. 10, 000 non-zero rows and columns of Θ).

[0161] Viewing Θ as an adjacency matrix defining a network of genes, we partitioned the network using with the Infomap community detection algorithm (S24) from the R package IGRAPH (vl.1.0) (S22), retaining modules that contain more than 10 genes. This yields 44 gene modules, each consisting of a set of genes. The modules are visualized in **FIG. 13.**

[0162] **Functional enrichments:** Functional enrichment analysis was performed on the gene sets defined by the modules using the findGO.pl program from the HOMER suite (Hypergeometric Optimization of Motif Enrichment, version: 4.9.1) (S12) with Benjamini and Hochberg correction for multiple hypothesis testing (retaining terms at adjusted p-value < 0.05). All genes that passed quality-control filters were used as a background set.

[0163] This yielded a set of biological signatures related to each module.

[0164] Computing scores from gene sets Given a set of genes (coming from a gene module or biological signature), cells were scored based on their gene expression. In particular, for a given cell the z-score for each gene in the set was determined. The z-scores were then truncated at 5 or -5, and define the signature of the cell to be the mean z-score over all genes in the gene set. The scores for the gene modules are visualized in **FIG. 13** and the scores for the biological signatures are visualized in **FIGs. 7A-7F.**

Example 2 Reprogramming to iPSCs as a test case for analysis of developmental landscapes.

[0165] WADDINGTON-OT was used to analyze the reprogramming of fibroblasts to iPSCs (39-42).

[0166] Studies have applied scRNA-Seq, but they have involved only several dozen cells or several dozen genes (13, 43). Studies have proposed that reprogramming involves two "transcriptional waves," with gain of proliferation and loss of fibroblast identity followed by transient activation of developmental regulators and gradual activation of embryonic stem cell (ESC) genes (12). Some studies (16, 44, 45), have noted strong upregulation of lineage-specific genes from unrelated lineages (e.g., related to neurons), but it has been unclear whether this largely reflects disorganized gene activation by TFs or coherent differentiation of specific (off-target) cell types (45).

[0167] scRNA-seq profiles of 65,781 cells were collected across a 16-day time course of iPSC induction, under two conditions **(FIGs. 6A,6B).** An efficient "secondary" reprogramming system was used (46), as described hereinbelow.

[0168] Mouse embryonic fibroblasts (MEFs) were obtained from a single female embryo homozygous for ROSA26-M2rtTA, which constitutively expresses a reverse transactivator controlled by doxycycline (Dox), a Dox-inducible polycistronic cassette carrying Pou5fl (Oct4), Klf4, *Sox2,* and *Myc (OKSM),* and an EGFP reporter incorporated into the endogenous *Oct4* locus (Oct4-IRES-EGFP). MEFs were plated in serum-containing induction medium, with Dox added on day 0 to induce the OKSM cassette (Phase-l(Dox)). Following Dox withdrawal at day 8, cells were transferred to either serum-free N2B27 2i medium (Phase-2(2i)) or maintained in serum (Phase-2(serum)). Oct4 EGFP+ cells emerged on day 10 as a reporter for "successful" reprogramming to endogenous Oct4 expression **(FIG. 6C).** Single or duplicate samples were collected at the various time points **(FIG. 6A),** single cell suspensions were generated and scRNA-Seq **(Table 8, FIGs. 11A-11D)** was performed. Samples were also collected from established iPSC lines reprogrammed from the same MEFs, maintained in either 2i or serum conditions. Overall, 68,339 cells were programed to an average depth of 38,462 reads per cell **(Table 8).** After discarding cells with less than 1,000 genes detected, a total of 65,781 cells were retained, with a median of 2,398 genes and 7,387 unique transcripts per cell.

Table 8

Sample (Day)	Phase	Number of Cells	Number of cells (filtered)	Number of reads	Mean Reads per Cells	Median Genes per Cell	Median UMI Counts per Cell	cDNA PCR Duplication %
DO	Dox	4241	4060	111,286,101	26240	2446	6495	50.5
D2-1	Dox	2909	2890	143,713,479	49403	2867	8401	55.6
D2-2	Dox	2758	2729	109,907,870	39850	2521	6271	70.2
D4-1	Dox	2889	2882	126,824,856	43899	2447	7349	57.3
D4-2	Dox	3976	3962	99,109,221	24926	2386	7446	34.1
D6-1	Dox	3676	3198	132,565,146	36062	1453	3147	84
D6-2	Dox	3534	3168	99,748,307	28225	1533	3567	76.5
D8-1	Dox	2177	2142	98,462,446	45228	2332	8216	65.7
D8-2	Dox	3677	2625	95,807,550	26055	1486	3862	62.6
D9-1	2i	2445	2441	122,451,561	50082	2843	11799	51.8
D9-2	2i	2183	2174	125,014,976	57267	2734	11183	57
DlO-1	2i	2878	2878	129,837,247	45113	2625	9570	58.1
D10-2	2i	2620	2619	126,364,110	48230	2647	9930	59.5
Dll	2i	1532	1529	119,736,956	78157	2892	10744	65.9
D12-1	2i	5144	5139	158,679,538	30847	2269	6299	41
D12-2	2i	2156	2155	112,512,277	52185	2651	8633	54.8
D16	2i	4621	4500	117,242,910	25371	2203	7761	39.5
iPSCs	2i	2917	2916	139,441,360	47803	3172	12775	38.2
D10	serum	2094	2088	115,832,953	55316	2717	9733	58.4
D12	serum	2913	2895	96,402,567	33093	2711	8819	44.2
D16	serum	3875	3703	119,329,130	30794	1953	4984	53.6
iPSCs	serum	3124	3088	128,207,617	41039	2637	9689	46.1
	Total	68339	65781					
			Average depth per cell:	38,462				

Example 3 The reprogramming landscape reveals relationships among biological features.

[0169] WADDINGTON-OT was used to generate a transport map across the cells in the time course described in the previous example. Based on similarity of expression profiles, the 16,339 detected genes were partitioned into 44 gene modules and the 65,781 cells into 33 cell clusters. Some of the clusters contained cells from more than one time point, reflecting asynchrony in the reprogramming process. The landscape of reprogramming was explored by identifying cell subsets of interest (e.g., successfully reprogrammed cells at day 16, or each of the cell clusters), studying the trajectories to and from these subsets (e.g., characterizing the pattern of gene expression in ancestors at day 8 of successfully reprogrammed target cells at day 16), and considering contemporaneous interactions between them. The analyses were visualized in a two-dimensional embedding using FLE (Fig. 7A), annotated in various ways. FLE reflects better global structures in the data presented herein than other modes of visualization (Figs. 12A-12C). These annotations include time points and growth conditions (Figs. 7B,7C), gene modules (Figs. 13, 14A-14B, Table 1), cell clusters (Fig. 7D, Fig. 14A-14D, Table 9), expression of gene signatures (curated gene sets associated with specific cell types, pathways, and responses, such as MEF identity, proliferation, pluripotency, and apoptosis; Fig. 7E, Table 7), expression of individual genes (Fig. 7F, Fig. 15), and ancestor and descendant distributions (Figs. 8A-8F). Extensive sensitivity analysis showed that key biological results for the reprogramming data were largely robust to the details of the formulation. Finally, the WADDINGTON-OT landscape was compared to the landscapes produced by various graph-based methods. The results show the following. Cell trajectories start at the lower right corner at day 0, proceed leftward to day 2 and then upward towards two regions identified as the Valley of Stress and the Horn of Transformation (Fig. 7B, Fig. 8A). The Valley is characterized by signatures of cellular stress, senescence, and, in some regions, apoptosis (Fig. 7E); it appears to be a terminal destination. By contrast, the Horn is characterized by increased proliferation, loss of fibroblast identity, a mesenchymal-to-epithelial transition (Fig. 7E), and early appearance of certain pluripotency markers (e.g., Nanog and Zfp42, Fig. 7F), which are predictive features of successful reprogramming (47). Some of the cells in the Horn proceed toward pre-iPSCs by day 12 and iPSCs by day 16, while others encounter alternative fates of placental-like development and neurogenesis (in serum, but not 2i condition; Figs. 7B, 7C). A more detailed account of the landscape is in the following examples.

Table 9

	Phase-l(Dox)					**Phase-2 (2i)**						**Phase-2 (serum)**			
Cluster	**DO**	**D2**	**D4**	**D6**	**D8**	**D9**	**D10**	**Dll**	**D12**	**D16**	**iPSCs**	**D10**	**D12**	**D16**	**iPSCs**
1	**97.4**	**0.1**	**0.0**	**0.0**	**0.1**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.1**	**0.4**	**0.1**	**0.9**

2	2.0	0.3	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.1
3	0.1	22.0	0.9	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	0.0	31.7	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
5	0.2	33.5	0.1	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.0
6	0.0	12.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
7	0.0	0.1	60.7	5.8	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
8	0.0	0.0	23.9	8.3	2.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
9	0.0	0.0	0.9	16.5	16.8	1.2	0.0	0.0	0.0	0.0	0.0	1.0	0.0	0.0	0.0
10	0.0	0.0	0.0	2.4	15.1	19.3	0.5	0.3	0.0	0.0	0.0	21.8	0.0	0.1	0.0
11	0.0	0.0	0.0	0.2	1.3	22.6	14.1	7.1	1.5	0.1	0.0	14.4	2.9	0.7	0.1
12	0.2	0.0	0.0	0.0	0.0	3.2	16.0	11.4	9.7	1.1	0.6	3.0	13.9	2.6	0.2
13	0.1	0.0	0.0	0.0	0.4	9.1	11.5	8.6	3.4	0.2	0.0	18.1	16.8	1.8	0.1
14	0.0	0.0	0.0	0.0	0.0	0.2	2.9	4.8	12.3	1.4	1.5	0.0	2.5	0.6	0.0
15	0.0	0.0	0.0	0.0	0.0	0.1	1.2	5.6	11.6	6.2	5.3	0.0	0.2	0.6	0.0
16	0.0	0.0	0.0	0.0	0.0	0.7	5.9	14.2	16.0	2.5	0.0	0.3	1.0	1.5	0.0
17	0.0	0.0	0.0	0.0	0.0	0.6	10.5	11.9	6.7	0.2	0.0	0.0	0.9	0.2	0.0
18	0.0	0.1	12.5	15.9	1.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
19	0.0	0.0	0.0	10.6	27.5	11.6	0.0	0.1	0.0	0.0	0.0	5.6	0.0	0.0	0.0
20	0.0	0.0	0.6	31.7	20.0	4.3	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.0	0.0
21	0.0	0.0	0.0	8.5	15.5	24.9	0.1	0.1	0.1	0.0	0.0	32.5	0.2	0.6	0.1
22	0.0	0.0	0.0	0.0	0.0	1.6	25.8	10.1	0.5	0.1	0.0	1.2	1.0	0.3	0.1
23	0.0	0.0	0.0	0.0	0.0	0.1	0.3	0.1	0.5	0.1	0.0	0.7	29.2	16.5	1.7
24	0.0	0.0	0.0	0.0	0.0	0.3	8.6	11.6	6.3	1.6	0.1	0.2	16.8	7.7	0.1
25	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.3	7.3	0.4	0.0	0.0	0.0	0.1	0.0
26	0.0	0.0	0.0	0.0	0.0	0.1	0.6	1.0	0.3	0.1	0.0	0.0	0.8	30.7	0.0
27	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.6	0.1	0.0	0.0	0.0	3.0	0.0
28	0.0	0.0	0.0	0.0	0.0	0.0	1.8	12.7	23.0	2.3	0.7	0.6	12.7	0.6	0.0
29	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	31.6	0.0	0.0	0.0	1.1	0.0
30	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	33.4	0.1	0.0	0.1	0.4	0.0

31	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**15.4**	**1.6**	**0.0**	**0.1**	**23.3**	**1.1**
32	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.1**	**6.6**	**95.5**
33	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.1**	**3.1**	**90.2**	**0.0**	**0.0**	**0.8**	**0.1**

Example 4

[0170] Predictive markers of reprogramming success are detectable by day 2.

[0171] The vast majority (>98%) of cells at day 0 fall into a single cluster characterized by a strong signature of MEF identity, with clear bimodality in the proliferation signature (Fig. 16A). By day 2 after Dox treatment, cells show high levels of expression of the OKSM cassette and have begun to diverge in their responses (clusters 3, 4, 5, 6, Fig. 7D). Overall, they score highly for expression signatures of proliferation, MEF identity, and endoplasmic reticulum (ER) stress (reflecting high secretion in mesenchymal cells) (Fig. 7E).

[0172] However, the cells exhibit considerable heterogeneity, seen most clearly by comparing the cells in clusters 4 and 6, which vary in their expression signatures and in their fates (Figs. 8A, 8B and Figs. 17A-17C). While cells in both clusters are highly proliferative, cells in cluster 4 have begun to lose MEF identity, show lower ER stress, and have higher OKSM-cassette expression, while cells in cluster 6 have the opposite properties (FIGs. 7D, 7E and Fig. 16B). The cells in the two clusters show clear differences in their enrichment in the ancestral distribution of iPSCs (Fig. 8D). The majority (54%) of the day 2 ancestors of iPSCs lie in cluster 4, while only a small fraction (3%) lie in cluster 6. Clusters 4 and 6 also show clear differences in their descendants (Figs. 8A, 8C and Fig. 17A): the descendants of cells in cluster 6 are strongly biased toward the Valley of Stress (e.g., 81% of Cluster 6 cell descendants are in clusters 8-11 by day 8 vs. 18% for cluster 4), while cluster 4 is strongly biased toward the Horn of Transformation (e.g., 81% in clusters 19-21vs. 12% for cluster 6).

[0173] The strongest difference in gene expression between clusters 4 and 6 was seen for Shisa8 (detected in 67% vs. 3% of cells in clusters 4 and 6, respectively) (Fig. 7F, fig. 16B) and Shisa8+ cells are enriched among the day 2 ancestors of iPSCs (Fig. 16B). Notably, Shisa8 is strongly associated with the entire trajectory toward successful reprogramming (Fig. 7F): it is expressed in the Horn, pre-iPSCs, and iPSCs, but not in the Valley or in the alternative fates of neurogenesis and placental development. The expression pattern of Shisa8 is similar to, but stronger than, that of Fut9 (Fig. 15), a known early marker of successful reprogramming that synthesizes the surface glyco-antigen SSEA-1 (12). Shisa8 is a little-studied mammalian specific member of the Shisa gene family in vertebrates, which encodes single-transmembrane proteins that play roles in development and are thought to serve as adaptor proteins (48). The analysis suggests that Shisa8 may serve as a useful early predictive marker of eventual reprogramming success and may play a functional role in the process.

Example 5 Cells in the valley of stress induce a Senescence Associated Secretion Phenotype (SASP).

[0174] By day 4, cells display a bimodal distribution of properties that is strongly correlated with their eventual descendants: cells in cluster 8 (low proliferation, high MEF identity, Fig. 7D, E and Fig. 16C) have 95% of their descendants in the Valley (Figs. 8A, 8B and Fig. 17A), while cells in cluster 18 (high proliferation, low MEF identity, Figs. 7D, 7E and Fig. 16C) have 94% of their descendants in the Horn (Figs. 8A, 8B and Fig. 17A and Table 10). Cells in cluster 7 show intermediate properties and have roughly equal probabilities of each fate (Fig. 8A, 8B and Fig. 17A).

Table 10

Cluster	To 1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
From 1		0.001	0.920	0.980	0.978	0.987	0.001	0.001	0.000	0.000	0.000	0.001	0.008	0.001	0.002	0.003
2		0.790	0.000	0.003	0.003	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3		0.000	0.012	0.005	0.000	0.000	0.206	0.166	0.012	0.002	0.002	0.000	0.000	0.000	0.000	0.000
4		0.007	0.058	0.002	0.000	0.000	0.265	0.044	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5		0.106	0.008	0.003	0.006	0.003	0.293	0.298	0.004	0.000	0.000	0.001	0.000	0.000	0.000	0.000
6		0.000	0.000	0.000	0.007	0.010	0.100	0.074	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.000
7		0.000	0.001	0.000	0.000	0.000	0.131	0.169	0.383	0.143	0.040	0.000	0.005	0.000	0.000	0.000
8		0.000	0.000	0.000	0.000	0.000	0.003	0.240	0.171	0.126	0.018	0.000	0.005	0.000	0.000	0.000
9		0.002	0.000	0.000	0.000	0.000	0.000	0.006	0.163	0.197	0.062	0.031	0.168	0.021	0.001	0.046
10		0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	0.3	0.0	0.0

		05	00	00	00	00	00	00	11	63	88	83	93	77	25	37
11		0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.001	0.031	0.216	0.081	0.211	0.085	0.065
12		0.012	0.000	0.004	0.000	0.000	0.000	0.000	0.000	0.000	0.020	0.127	0.032	0.166	0.269	0.152
13		0.012	0.001	0.003	0.000	0.000	0.000	0.000	0.001	0.000	0.013	0.112	0.236	0.085	0.514	0.578
14		0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.017	0.002	0.028	0.037	0.017
15		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.001	0.006	0.005
16		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.005	0.003	0.025	0.026
17		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.003	0.003	0.003	0.026	0.027
18		0.000	0.000	0.000	0.000	0.000	0.002	0.003	0.201	0.079	0.013	0.003	0.001	0.000	0.000	0.000
19		0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.029	0.120	0.357	0.123	0.272	0.036	0.001	0.032
20		0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.018	0.172	0.270	0.047	0.052	0.001	0.000	0.002
21		0.010	0.000	0.000	0.004	0.000	0.000	0.000	0.001	0.094	0.075	0.021	0.036	0.035	0.001	0.005
22		0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.004	0.001	0.006	0.003	0.002
23		0.027	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.005	0.004	0.001	0.021	0.004	0.003
24		0.010	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.002	0.001	0.005	0.003	0.002
25		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
26		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
27		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
28		0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
29		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
30		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
31		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
32		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
33		0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Table 10 (Cont'd)

Cluster	**To 17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**	**30**	**31**	**32**	**33**
From 1	0.003	0.003	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.004	0.006	0.000	0.006	0.002	0.001	0.006	0.001
2	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
3	0.000	0.051	0.001	0.004	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
4	0.000	0.276	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
5	0.000	0.009	0.000	0.001	0.000	0.000	0.001	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
6	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
7	0.000	0.578	0.183	0.340	0.044	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
8	0.000	0.008	0.008	0.001	0.005	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
9	0.026	0.004	0.047	0.003	0.073	0.011	0.001	0.005	0.000	0.000	0.000	0.001	0.000	0.001	0.000	0.000	0.000
10	0.058	0.000	0.033	0.001	0.069	0.080	0.065	0.026	0.015	0.001	0.001	0.009	0.001	0.003	0.000	0.001	0.000
11	0.111	0.000	0.003	0.001	0.006	0.005	0.000	0.000	0.000	0.007	0.012	0.001	0.012	0.004	0.003	0.012	0.001
12	0.084	0.000	0.000	0.000	0.000	0.014	0.000	0.000	0.000	0.025	0.046	0.002	0.043	0.015	0.009	0.041	0.004
13	0.650	0.000	0.001	0.000	0.001	0.015	0.000	0.000	0.000	0.037	0.066	0.003	0.057	0.020	0.011	0.055	0.005
14	0.006	0.000	0.000	0.000	0.000	0.003	0.000	0.000	0.000	0.006	0.010	0.000	0.010	0.004	0.002	0.010	0.001
15	0.002	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
16	0.020	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.002	0.000	0.002	0.001	0.000	0.002	0.000
17	0.015	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.001	0.002	0.000	0.001	0.000	0.000	0.001	0.000
18	0.000	0.064	0.264	0.227	0.116	0.007	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
19	0.014	0.003	0.143	0.057	0.107	0.104	0.050	0.073	0.017	0.001	0.000	0.045	0.003	0.013	0.000	0.002	0.000
20	0.001	0.006	0.304	0.309	0.336	0.276	0.011	0.005	0.000	0.001	0.000	0.002	0.000	0.001	0.000	0.000	0.000
21	0.006	0.000	0.014	0.052	0.235	0.387	0.339	0.260	0.083	0.032	0.013	0.744	0.021	0.082	0.006	0.017	0.003
22	0.001	0.000	0.000	0.000	0.000	0.008	0.014	0.001	0.001	0.008	0.007	0.000	0.009	0.003	0.002	0.008	0.001
23	0.001	0.000	0.000	0.000	0.005	0.076	0.498	0.008	0.089	0.663	0.396	0.005	0.243	0.076	0.047	0.223	0.021
24	0.001	0.000	0.000	0.000	0.001	0.010	0.020	0.622	0.793	0.145	0.201	0.011	0.197	0.111	0.095	0.183	0.067

25	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
26	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.061	0.228	0.000	0.000	0.000	0.000	0.000	0.000
27	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.005	0.000	0.000	0.000	0.000	0.000	0.000
28	0.000	0.000	0.000	0.000	0.000	0.001	0.000	0.000	0.000	0.006	0.004	0.174	0.364	0.640	0.804	0.406	0.885
29	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.002	0.002	0.002	0.002	0.001
30	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.004	0.003	0.003	0.004	0.002
31	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.009	0.008	0.007	0.010	0.004
32	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.001	0.001	0.000	0.015	0.010	0.008	0.016	0.005
33	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000

[0175] Along the trajectory from cluster 8 to the Valley (days 10-16; Fisg. 8A, 8B and 8E,F), cells show a strong decrease in cell proliferation (Fig. 7E), accompanied by increased expression of various cell-cycle inhibitors, such as Cdkn2a, which encodes pi6, an inhibitor of the Cdk4/6 kinase and halts Gl/S transition (Fig. 7F), Cdknla (p21), and Cdkn2b (pl5) (Fig. 16D), which peaks in the Valley. The cells show increased expression of D-type cyclin gene Ccnd2 (Figs. 15, 16D) associated with growth arrest (49). A subset of the cells in the Valley (29%; clusters 12 and 14) showed high activity for a gene module that is correlated with a p53 pro-apoptotic signature, compared to all other cells inside the Valley (p-value< 10-16, average difference 0.17, Mest) and outside the Valley (p-value< 10-16, average difference 0.32, Mest) (Fig. 7E, fig. 16E).

[0176] Cells in the Valley also show activation of signatures of extracellular-matrix (ECM) rearrangement and secretory functions (Fig. 7E, Fig. 16E). Because these properties are consistent with a senescence associated secretory phenotype (SASP), a SASP signature involving 60 genes (50) was used. Cells with this signature appear on day 10 and continue through day 16, consistent with previous reports concerning the timing of onset of stress-induced senescence (50) (Fig. 7E, Fig. 16E).

[0177] SASP, which has key roles in wound healing and development that are relevant for reprogramming biology, includes the expression of various soluble factors (including 116), chemokines (including 118), inflammatory factors (including Ifng), and growth factors (including Vegf) that can promote proliferation and inhibit differentiation of epithelial cells (50). Recent reports have suggested that secretion of 116 and other soluble factors by senescent cells can enhance reprogramming (51). Although detectable levels of 116 mRNA were present in only a small fraction of cells both in 2i and serum (0.2%) at days 12 and 16 (0.34% in all cells), the overall SASP signature was evident in 72% of cells in the Valley (vs. 11% elsewhere, primarily in day 0 MEFs). This suggests that the senescent cells in the Valley are likely to have paracrine effects on cells that successfully emerge from the Horn.

Example 6 Other cells at day 4 are strongly biased toward the Horn of Transformation.

[0178] For the remaining cells at day 4, the forward trajectory is characterized by high proliferation and loss of MEF identity (Figs. 7B, 7E), and the descendants are strongly biased toward the Horn at day 8 (Figs. 8A, 8B and Fig. 17A and Table 10). The Horn is distinguished as a point of transformation, where cells that have lost their mesenchymal identity are beginning their transitions to an epithelial fate. As discussed below, a minority of cells in the Horn have begun to express activators of a pluripotency expression program.

[0179] Following Dox withdrawal and media replacement on day 8, the cells in the Horn adopt one of four alternative outcomes by day 12 (senescence, neuronal program, placental program, and pre-iPSCs). Roughly half appear to become senescent, migrating through clusters 19 and 10 to the Valley (Fig. 8A). The fate of the remaining cells is strongly influenced by the culture medium. In serum conditions, the proportion of these cells that transition to neuronal, placental and pre-iPSC states is 62%, 13% and 26%, respectively. By contrast, the proportions in 2i condition are 3%, 37% and 59% (Table 10). These results are consistent with the presence in the 2i medium of two small-molecule inhibitors to inhibit differentiation, including one reported to inhibit neuronal differentiation (52).

Example 7

[0180] Neuronal-like and placental-like cells arise during reprogramming.

[0181] Two unusual cell populations were analyzed: placental-like cells (clusters 24 and 25, Figs. 7B, 7D and Figs. 8A, 8B, 8E, 8F) at day 12 and neural-like cells (clusters 26 and 27, Figs. 7B, 7D and Figs. 8A, 8B, 8E, 8F) at day 16. The first group was characterized by high activity of two gene modules enriched in signatures for "epithelial cell differentiation," "placenta development," and "reproductive structure development," while the second group showed high activity of signature for "neuron differentiation," "axon development," and "regulation of nervous system development" (Table 1, and Figs. 7B, 8C, 8E).

[0182] Both populations showed a substantial decrease in proliferation (Fig. 7E, fig. 16E). To explore if a common mechanism was responsible for this change, 98 cell-cycle related genes (53) were examined to identify those that were differentially upregulated in the placenta and neural clusters compared to all other clusters. The most distinctive characteristic was the high expression of Cdknlc, which encodes a cell-cycle inhibitor (p57) that promotes Gl arrest (Fig. 7F) and is required for maintenance of some adult stem cells (54). Other features are also shared between these two alternative lineages and adult stem cells-including the expression of Lgr5, a marker of adult epithelial stem cells in certain tissues (55) (Fig. 15).

[0183] The neural-like cells reside in a large "spike" observed at day 16 in serum but not 2i conditions (16% vs. 0.1% of cells), presumably due to differentiation inhibitors in the latter conditions. Cells near the base of the spike (cluster 26, Fig. 7D and Figs. 8E, 8F) expressed neural stem-cell markers (including Pax6 and Sox2, Fig. 7E, fig. 15), while cells further out along the spike (cluster 27, Fig. 7D) expressed markers of neuronal differentiation (including Neurog2 and Map2, fig. 15). The cells thus appear to span multiple stages of neurogenesis along the length of the spike (Fig. 7E).

[0184] Analysis of the developmental landscape suggests a potential mechanism for triggering neural differentiation. The ancestors of neural-like cells are largely found in cluster 23 on day 12 (Figs. 8A, 8F and fig. 17C and Table 10). At least 19% of cells in cluster 23 express Cntfr, an 116-family receptor that plays a critical role in neuronal differentiation and survival (56) (Fig. 7F); the true proportion is likely to be higher because the gene has low expression. Contemporaneously, senescent cells in the Valley at day 12 express activating ligands (Crlfl and Clcfl) of Cntfr (fig. 15). Thus, neural differentiation may be triggered by paracrine signals from senescent cells to Cntfr-expressing cells.

[0185] The placental-like cells express high levels of certain imprinted genes on chromosome 7 (Cdknlc, Igf2, Peg3, H19 and Ascl2; Fig. 7F, Fig. 15), as well as TFs (Cdx2 and Soxl7) associated with placental development (57, 58) (Fig. 15). They also show elevated levels of an ER stress signature (Fig. 3E), consistent with the secretory nature of placental cells and observations of placental cells in vivo (59). Analysis was performed to address whether the placental-like cells resembled recently described extraembryonic endodermal (XEN) cells from an iPSC reprogramming study (44). It was found that they do not share the distinctive XEN signature of the cells disclosed in that analysis. The proportion of cells in the placental-like population decreased substantially from day 12 to day 16 in 2i conditions, although the optimal-transport analysis could not confidently infer whether the decrease is due to cells dying, being overtaken by faster-growing cells, or transitioning to other fates (fig. 14A).

[0186] The following two tables provide a list of candidate reprogramming factors.

Example 8

Trajectory to successful reprogramming reveals a continuous program of gene activation.

[0187] We next studied the trajectory leading to reprogramming (Figs. 8D, 8E), which passes through pre-iPSCs (cluster 28; Figs. 8A, 8B) at day 12 en route to iPSC-like cells at day 16. The iPSC-like cells in serum conditions (which reside in cluster 31) closely resemble fully reprogrammed cells grown in serum (cluster 32). By contrast, the iPSC-like cells under 2i conditions are spread across three clusters (cluster 29-31). While the cells in cluster 31 resemble fully reprogrammed cells grown in 2i (cluster 33), those in cluster 29 show distinct properties suggestive of partial differentiation. In particular, cluster 29 shows lower proliferation, lower Nanog expression, and increased expression of genes related to differentiation (Figs. 7D, 7F).

[0188] In contrast to initial descriptions of reprogramming as involving two "waves" of gene expression, the trajectory of successful reprogramming reveals a more complex regulatory program of gene activity (Fig. 9A). By grouping genes according to their temporal patterns of activation in cells on the OT-defined trajectory to successful reprogramming, a rich collection of markers for particular stages can be obtained **(Fig. 9A).** In particular, 47 genes that appear late in successfully reprogrammed cells (for example, *Obox6, Spic, Dppa4)* were identified. These genes may provide useful markers to enrich fully reprogrammed iPSCs **(Table 2).**

Example 9

Paracrine signaling from the Valley may influence late stages of reprogramming.

[0189] The simultaneous presence of multiple cell types raises the possibility of paracrine signaling, with secreted factors from one cell type binding to receptors on another cell type. One such potential interaction above, is SASP+ cells in the Valley secreting Crlfl, Clcfl and neural-like cells on days 12 and 16 expressing the cognate receptor Cntfr.

[0190] To systematically identify potential opportunities for paracrine signaling, we defined an interaction score, IA,B,X,Y,ï, as the product of (1) the fraction of cells in cluster A expressing ligand X and (2) the fraction of cells in cluster B expressing the cognate receptor Y, at time t. Using a curated list of 149 expressed ligands and their associated receptors, we studied potential interactions between all pairs of clusters for each ligand-receptor pair, as well as the aggregate signal across all pairs and across those pairs related to the SASP signature. The potential for paracrine signaling varied sharply across the time course, as well as across cell types. Potential interactions are initially high, as cells with MEF identity retain their secretory functions; drop dramatically by day 6 (Fig. 18A), after cells have lost their MEF identity (Fig. 7B, 7C, 7E); rise steadily from day 8 to day 11, as secretory cells in the Valley emerge; and then drop again from days 12 to 16, as the abundance of cells in the Valley decreases (Fig. 18A). The same pattern is seen when considering only the 20 ligands in the SASP signature (Fig. 18B).

[0191] Notably, potential interactions are observed between cells in the Valley and each of iPSC, neural-like and placental-like cells. At day 16, cells in the Valley (clusters 15 and 16) express SASP ligands, while iPSCs (clusters 29-33) express receptors for these ligands (Fig. 18C), with the highest frequency seen for the chemokine Cxcll2 and receptor Dpp4 (Fig. 18D). As noted above, at days 12 and 16, the ligands Crlfl and Clcfl cells are expressed in the Valley while their receptor Cntfr is expressed in the neural spike (Fig. 7E, Fig. 18E). The interaction between Cntfr and Crlfl is ranked as the top interaction among all ligand-receptor pairs (Fig. 18E).

[0192] At day 12, many placental-like cells express the ligand Igf2 while cells in the Valley express receptors *Igflr and Igf2r* **(Fig. 18F).**

Example 10

X-chromosome reactivation follows activation of early and late pluripotency genes.

[0193] The reversal of X-chromosome inactivation in female cells is known to occur in the late stages of reprogramming and is an example of chromosome-wide chromatin remodeling. A recent study (60) reported that X-reactivation follows the activation of various pluripotency genes, based on immunofluorescence and RNA FISH in single cells. To assess X-reactivation, from scRNA-Seq data, each cell was characterized with respect to signatures of X-inactivation (Xist expression), X-reactivation (proportion of transcripts derived from X-linked genes, normalized to cells at day 0), and early and late pluripotency genes. Along the trajectory to successful reprogramming (but not elsewhere, **Fig. 7E),** cells at day 12 show strong downregulation of *Xist* but do not yet display X-reactivation. X-reactivation is complete at day 16, with the signature having risen from 1.0 to -1.6, consistent with the expected increase in X-chromosome expression (61). Analysis of the trajectory confirms that activation of both early and late pluripotency genes precedes *Xist* downregulation and X-reactivation.

Example 11

Some cell populations are enriched for aberrant genomic events.

[0194] Anaylsis was done to identify other coherent increases or decreases in gene expression across large genomic regions, which might indicate the presence of copy-number variations (CNVs) in specific cells. Particularly, analysis done to identify whole chromosome aberrations, demonstrated that 0.9% of cells showed significant up- or down-regulation across an entire chromosome; the expression-level changes were largely consistent with gain or loss of a single chromosome.

[0195] Next, evidence of large subchromosomal events was identified by analyzing regions spanning 25 consecutive housekeeping genes (median size -25 Mb). Significant events were found in -0.8% of cells. The frequency was highest (2.8%) in cluster 14, consisting of cells in the Valley of Stress enriched for a DNA damage-induced apoptosis signature. The frequency was 2-to-3-fold lower in other cells in the Valley (enriched for senescence but not apoptosis), in cells en route to the Valley (clusters 8 and 11), and in fibroblast-like cells at days 0 and 2. Notably, it was much lower (6-fold) in cells on the trajectory to successful reprogramming (Figs. 22B, 22C). Direct experimental evidence would be needed to confirm these events, and to clarify if the aberrations were preexisting in the MEF population, or if they accumulated during the course of reprogramming. 9

Example 12

Inferred trajectories agree with experimental results from cell sorting.

[0196] To test the accuracy of the probabilistic trajectories calculated for each cell based on optimal transport, results based on the trajectories were compared to experimental data from a recent study of reprogramming of secondary MEFs (16). In that study, cells were flow-sorted at day 10, based on the cell-surface markers CD44 and ICAM1 and a Nanog-EGFP reporter gene, and each sorted population was grown for several days thereafter to monitor reprogramming success. Gene expression profiles were obtained from each population at day 10 and CD44-ICAMl+Nanog+ population at day 15, together with mature iPSCs and ESCs. Reprogramming efficiency was lowest for CD44+ICAM-Nanog- cells, intermediate for CD44-ICAMl+Nanog- and CD44-ICAMl-Nanog+ cells, and highest for CD44-ICAMl+Nanog+ cells.

[0197] The flow-sorting-and-growth protocol was emulated in silico, by partitioning cells based on transcript levels of the same three genes at day 10 and predicting the fates of each population at day 16 based on the inferred trajectory of each cell in the optimal transport model. The computational predictions showed good agreement with these earlier experimental results (Fig. 5B), with respect to both reprogramming efficiency and changes in gene-expression profiles. In particular, the *in silico* results showed 93% correlation with results from the earlier study concerning relative reprogramming efficiencies for six categories of sorted cells (p value= 0.0023) **(Fig. 9B*).*** Notably, the computationally inferred trajectory of double positive cells rapidly transitioned toward iPSCs and continued in this direction through the end of the time course **(Fig. 9B).** Only one category (CD44-ICAM+Nanog-) differed significantly.

[00138] Differences may reflect the fact that experimental protocols were not identical *(e.g.,* the earlier study (16) maintains continuous expression of OSKM and supplements the medium with an ALK-inhibitor and vitamin C).

Example 13

Inferring transcriptional regulators that control the reprogramming landscape.

[0198] The optimal transport map provides an opportunity to infer regulatory models, based on association between TF expression in ancestors and gene expression patterns in descendants. TFs were identified by two approaches **(Fig. 9C):** (i) a global regulatory model, to identify modules of TFs and target genes and (ii) enrichment analysis, to identify TFs in cells having many vs.few descendants in a target cell population of interest. Gene regulation along the trajectories to placental-like and neural-like cells was examined **(Fig. 19).** For placental-like cells, the analysis pointed to 22 TFs **(Figs. 19A, 19B** and **Table 3).** Of the four most enriched *(Pparg, Cebpa, Gcml,* and *Gata2),* all have been reported to play roles in placenta development

(62). For example, *Gcml* was detected in 42% of cells at day 10 with a high proportion (>80%) of descendants in the placental-like fate but only 0.7% of those cells with a low proportion (<20%) (57-fold enrichment). For neural-like cells, the analysis pointed to 10 TFs *(Pax3, Msxl, Msx3, Sox3, Soxll, Tall, Enl, Foxa2, Gbx2,* and *Foxbl).* All have been implicated in various aspects of neural development **(fig.** ***19C)*** *(62-70).*

[0199] Additional analysis focused on identifying TFs that play roles along the trajectory to successful reprogramming (Fig. 9D and fig. 19D, 19E). The global regulatory model generated two regulatory modules, A and B, with 61 TFs in module A, 16 in module B, and 11 in both (Figs. 19D, 19E).

[0200] Module A involves target genes active across clusters 29-31, while Module B involves target genes that are more active in cluster 31, which contains more fully reprogrammed cells. The TFs in these modules are progressively activated across the trajectory of successful reprogramming. For Module B, the TFs are active in 13% of cells in the Horn on day 8, while target-gene activity is evident (at >80% of the levels observed in iPSCs) in 1.3%, 10%, and 21% of their descendant cells in days 10, 11, and 12 in 2i conditions; the pattern in serum conditions is similar, although with lower overall frequency (11%) of cells by day 12). The onset of TFs and target genes in Module A lags by 1-2 days (Fig. 9D).

[0201] To identify TFs likely to play a key role in the final stages of reprogramming, we used enrichment analysis to identify TFs enriched in cells at day 12 with a high vs. low proportion (>80% vs. <20%) of successfully reprogrammed descendants and then focused on the intersection of this set with the 66 TFs from the global regulatory analysis above. The analysis pointed to 9 TFs associated with a high probability of success in the late stages of reprogramming (Fig. 19F). Of these, five (Sox2, Nanog, Hesxl, Esrrb, Zfp42) have establishedroles in regulation of pluripotency *(71-73),* while the remaining four *(Obox6, Spic, Mybl2,* and *Msc)* have not previously been implicated. Among these novel factors, *Obox6* stands out as having the greatest enrichment in high- vs. low-probability cells (68-fold, 9.3% vs -0.14%) **(fig. 19F).**

Example 14

Forced expression of *Obox6* enhances reprogramming.

[0202] Obox6 was identified by the regulatory analysis described herein as strongly correlating to reprogramming success. Obox6 (oocyte-specific homeobox 6) is a homeobox gene of unknown function that is preferentially expressed in the oocyte, zygote, early embryos and embryonic stem cells (74).

[0203] To test whether Obox6 also plays an active role in the process of reprogramming, experiments were performed to address whether expressing Obox6 along with OKSM during days 0-8 can boost reprogramming efficiency. Secondary MEFs were infected with a Dox-inducible lentivirus carrying either Obox6, the known pluripotency factor Zfp42 (73), or no insert as a negative control. Both Obox6 and Zpf42 increased reprogramming efficiency of secondary MEFs by ~2-fold in 2i and even more so in serum. The results were confirmed in multiple independent experiments (Figs. 10A and 10B, and fig. 20). Assays in primary MEFs showed similar increases in reprogramming efficiency (fig. 20). These results demonstrate the importance of Obox6 in the context of cellular reprogramming.

[0204] Figs. 1OA-IOC demonstrate the effect of overexpression of Obox6 and Zpf42 on reprogramming efficiency in secondary MEFs. Figs. 10 A and 10B show bright field and fluorescence images of iPSC colonies generated by lentiviral overexpression of *Oct4, Kl/4, Sox2,* and *Myc* (OKSM) with either an empty control, *2fp42* or *Obox6* expression cassette, in either Phase-l(Dox)/Phase-2(2i)(A) and Phase- l(Dox)/Phase-2(serum) (B) conditions (indicated). Cells were imaged at day 16 to measure Oct4-EGFP+ cells. Bar plots representing average percentage of Oct4-EGFP+ colonies in each condition on day 16 are included below the images. Shown are data from one of five independent experiments, with three biological replicates each. Error bars represent standard deviation for the three biological replicates. Figure 6C is a schematic of the overall reprogramming landscape highlighting: the progression of the successful reprogramming trajectory, alternative cell lineages, and specific transition states (Horn of Transformation). Also highlighted are transcription factors (orange) predicted to play a role in the induction and maintenance of indicated cellular states, and putative cell-cell interactions between contemporaneous cells in the reprogramming system.

Example 15

Definition of gene signatures

[0205] From gene set enrichment analysis of 44 gene modules (Table 1, Figs. 12A-12C), significant enrichments for terms that shed light on the reprogramming landscape were found. Analysis was done to investigate whether similar expression patterns from well-defined gene signatures could be identified. To investigate this, a list of gene sets from various databases of gene signatures was curated (see Table 11, a list of genes for each gene signature is shown in Table 2). A pluripotency gene signature was determined.

[0206] Differential gene expression analysis was performed between two groups of cells: mature iPSCs and cells along the time course DO to D16, and the top 100 genes with increased expression in mature iPSCs were identified. A proliferation gene signature was obtained by combining genes expressed at Gl/S and G2/M phases. For epithelial and neural gene signatures, canonical markers of epithelial and neuronal cell lineage markers, respectively were collected.

Table 11.: List of gene signatures used in this work. List of genes for each gene signature are shown in Table 2.

Gene Signature	Source
MEF identity	Mouse Gene Atlas (S29, S30)
Pluripotency	this work, iPSCs vs. DO to D16 cells
Proliferation	G1/S and G2/M genes, (S31)
ER stress	GO:0034976, Biological Process Ontology
Epithelial identity	(S32–S35)
ECM rearrangement	GO:0030198, Biological Process Ontology
Apoptosis	Hallmark P53 Pathway, MSigDB
Senescence	Table 1 in (S36)
Neural identity	(S37-S43)
Placental identify	Mouse Gene Atlas, (S29, S30)
X reactivation	chromosome X

Computing descendant distributions for clusters of cells

[0207] The descendant distributions for the 33 clusters of cells, some of which span multiple days were computed. To put each cluster on equal footing, 100 cells in each cluster were initialized. These 100 cells were distributed proportionally over the days represented in the cluster.

For each day d and cluster i, let n_d^i denote the number of day d cells in cluster i. We denote the total number of cells in cluster %by $N^i = \Sigma_d n_d^i$. With this notation, we initialize 100 x $\frac{n_d^i}{N^i}$ cells in cluster i on day d and compute the descendant distribution of these cells at the next time point. We denote this descendant distribution by $D_d^{\cdot}$. We then compute the mass of this descendant distribution residing in each cluster j by summing up the mass $D_d^{\cdot}$ assigns to each cell in cluster j. Finally, to obtain the i,j entry of the cluster - cluster transition table, wc sum over d.

This give the total mass transferred from from cluster i to cluster j, per 100 cells initialized in cluster i. We compute this separately for 2i and serum.

Extraembryonic gene signatures

[0208] Previous reports have shown that extraembryonic endoderm stem cells (XEN) were induced in the reprogramming process in parallel of reprograming to iPSCs (S48). To determine if XEN cells were induced in the reprogramming system described herein, the XEN gene signature from in vivo XEN cells, trophoblast and placental gene signatures was analyzed **(Table 12).** While a small fraction of cells (180 cells) displays a high XEN score at day 16 (under serum condition), a larger fraction of cells in clusters 24 and 25 displays high trophoblast and placental signature scores . This indicates that the alternative placental-like cell lineage does not share the distinctive XEN signature as previously reported.

Gene .Signature	Genes	Reference
XEN	Dab2 Fst Pdgfra Pthlr Gata6 Foxql Fxyd3 Tet3 Sox 17 Foxa2 Lama l Lamb l Gata4 Krt8	(S49)
Trophoblast	Ascl2 Bmp4 Bmp8b Cdx2 Elf5 Eomes Esrrb Els2 Fgfr2 Grn Tgf2 .Tade 1 Lipg Pcsfc6 Ptpra Smad3 Snai 1 Tead4 Tfap2c Vavl Yap l Gata3 Krt7 Krtl 8	(S50)
Placenta	Table A 1	

Table 12.: List of XEN, trophoblast and placenta gene signatures

Example 16

Identifying markers for reprogramming success

[0209] To gain further insights into the mechanisms of reprogramming success, categories of genes that changed their expression in characteristic patterns (Figs. 5A-5G) along the successful trajectory determined by optimal transport were characterized. Genes that exhibited significant changes along the trajectory (2,872 genes) were clustered using k-means clustering and the number of clusters was determined by the gap statistic (S44). 14 distinct expression patterns among cells that would end up succesfully reprogrammed (Table 10) were identified. Genes were divided into two obvious patterns, upregulated (Al to A10) and downregulated (Al l to A14). After dox induction, a large number of genes that were mainly involved with MEF identify were downregulated. Instead of "two waves" indicated by a previous report (S45), continuous activation patterns after dox induction were observed. In early stage of reprogramming, they were involved with metabolic changes and were targets of Myc (Al to A3). In late stage (A6 and A7) they were associated with activation of pluripotency networks. Two categories of pluripotency-associated genes were identifed. Genes in category A6 gradually upregulated after dox withdrawal, such as Nanog, Sox2, Dppa3 (early pluripotency-associated genes). Genes in category A7 upregulated after genes in A6, such as Obox6, Dppa4 (late pluripotency-associated genes).

[0210] Genes that were upregulated preferentially in cells that were successfully reprogrammed from A6 and A7 were identifed. The fraction of cells in clusters 28 to 33 vs. all other clusters were calculated. By setting a threshold of 1%, genes that were expressed in less than 1% of cells in all other clusters were ranked. 47 genes that were preferentially expressed in the late stage of reprogramming on successful trajectory and were mostly absent from other cells (Table 10) were identified.

Example 17

Cell-Cell Interactions

[0211] To characterize potential cell-cell interactions between contemporaneous cells during reprogramming, a list of ligands and receptors found in the GO database were collected. The set of ligands (415 genes) is a union of three gene sets from the following GO terms: 1) cytokine activity (GO:0005125), 2) growth factor activity (GO:0008083), and 3) hormone activity (GO:0005179). The set of receptors (2335 genes) is defined by the GO term receptor activity (GO:0004872). Next, a curated database of mouse protein-protein interactions (S46) was used to identify 580 potential ligand-receptor pairs. Two aspects of potential cell-cell interactions in the data were the focus of the analysis: 1) determining global trends in the expression of all potential contemporaneous ligand-receptor pairs across the reprogramming time course and 2) ranking individual ligand-receptor pairs at a specific day and condition. First, an interaction score

Iλ,Bχγ ,i as the product of (1) the fraction of cells *{FAX ,t)* in cluster A expressing ligand X at time *t* and (2) the fraction of cells *(FB. Y,t)* in cluster B expressing the cognate receptor Y at time t was defined. Aggregate interaction score *IA,B,t* was defined as a sum of the individual interaction scores across all pairs:

$$I_{A,B,t} = \sum_{All\ X\ ¾pairs} I_{A,B,X,Y,t} = \sum_{Ail\text{-}X\text{-}Y\ pairs} F_{A,X,t}F_{B,Y,t}$$

[0212] The aggregate interaction scores for all combinations of cell clusters in **figs. 18A-B** were depicted. Second, individual ligand-receptor pairs at a given day and condition between cell subsets of interest were examined. Values of the interaction scores IA,B,X,Y,ï are high for ubiquitously expressed ligands and receptors at a given day and may be nonspecific to a pair of cell subsets of interest. Thus, permutations were used to generate an empirical null distribution of interaction scores between two random groups of cells. In each of the 10,000 permutations, two groups R1 and R2 of 100 cells each from time t were selected and the interaction score between the ligand in group R1 and the receptor in group R2 was calculated. Each ligand-receptor interaction score was standardized by taking the distance between the interaction score *lA,B,x,Yt* and the mean interaction score in units of standard deviations from the permuted data *((IA,B,X, Y,t — mean(lRi,R2 ,x,Y,t))/sd(lRi,R2,xj,t))*. Examples of standardized interaction scores ranked by their values are depicted in **Figs. 18D-F.**

Example 18

X-chromosome reactivation

[0213] Analysis was performed to identify X-chromosome reactivation from our scRNA-seq dataset. The set of all detected genes (16,339) was split to X-chromosomal and autosomal genes. Then the mean X/autosome expression ratio for each cell (normalized by the average X/autosome expression ratio at day 0 cells) as a measurement of X-chromosome reactivation was calculated.

[0214] The mean X/Autosome expression ratio reached mean value of 1.6 in late stage of reprogramming indicating X-chromosome reactivation . Interestingly, cells in cluster 32 (mature iPSCs in serum) had their X-chromosome inactivated but no Xist expression, which might be due to partial differentiation of iPSCs in serum condition or that the established female iPSCs lost one of their X chromosomes, which happens frequently in serum cultured female ESCs or iPSCs but less often in 2i cultured female ESCs/iPSCs (S47). This was specific to mature iPSCs in serum as day-16 cells in serum exhibited similar X-chromosome reactivation to day 16 cells in 2i

[0215] Downregulation of Xist expression (cluster 28, day 12 cells) preceded X-chromosome reactivation (clusters 29,30,31, and 33; day 16, mature iPSCs) (Figs. 21A-21C). The upregulation of early and late pluripotency genes (activation pattern A6 and A7, respectively) preceded X-chromosome reactivation (Figs. 21D-21F).

[0216] The fraction of cells that activated late pluripotency genes A7 and reactivated the X-chromosome were analyzed. The X/Autosome expression ratio and A7 gene signature score show bimodal distribution across all cells (fig. 21G and fig. 21H, respectively). We classified cells to those that had reactivated their X-chromosome if the X/Autosome expression ratio > 1.4 and those that induced A7 genes if the A7 average z-score > 0.25 **(figs. 21G, 21H).** Using the above thresholds the fraction of cells in clusters 28-33 that reactivated their X-chromosome and activated the A7 program **(Table 13) were calculated.** Around a 10-fold difference is observed in the percentage of cells that upregulated A7 genes and reactivated X chromosome in clusters 28 and 32.

Cluster	28	29	30	31	32	33
X/A	7.6	79.3	84.2	89.1	7.2	81.9
A7	72,9	98.9	99.7	99.1	93.3	99.1

Table 13. Percentage of cells in clusters 28-33 that exhibited X-chromosome reactivation and induction of A7 genes.

Example 19

Identifying large chromosomal aberrations

[0217] Methodology. Two types of analysis were performed to detect aberrant expression in large chromosomal regions. First, analysis was performed to identify cells with significant up- or down-regulation at the level of entire chromosomes. Second, analysis was performed to identify cells with significant subchromosomal aberrations spanning windows of 25 consecutive broadly-expressed genes. Empirical p-values and false discovery rates (FDRs) for both analyses were computed by randomly permuting the arrangement of genes in the genome, as described below.

[0218] Permutations for both types of analysis are done as follows. In each of 100,000 permutations the labels of genes in the entire dataset were randomly shuffled, while preserving the genomic positions of genes (with each position having a new label each time) and the expression levels in each cell (so that each cell has the same expression values, but with new labels). Either whole chromosome or subchromosomal aberration scores for each cell were calculated. To identify whole-chromosome aberrations scores in each cell, the sum of expression levels in 25Mbp sliding windows along each chromosome, with each window sliding IMbp so that it overlaps the previous window by 24Mbp was calculated. For each window in each cell, the Z-score of the net expression, relative to the same window in all other cells was calculated. The fraction of windows on each chromosome with an absolute value Z-score > 2 was counted. This fraction serves as the whole-chromosome aberration score for each chromosome in each cell. To assign a p-value to the whole-chromosome score for cellj chromosomej, the empirical probability that the score for cellj chromosomej in the randomly permuted data was at least as large as the score in the original data was calculated.

[0219] Subchromosomal aberration scores were computed as follows. The 20% of genes with the most uniform expression across the entire dataset were identified. This is done by calculating the Shannon Diversity (eentropy(gene)) for each gene, and taking the 20% of genes with the largest values. Using these genes, the sum of expression in sliding windows of 25 consecutive genes, with each window sliding by one gene and overlapping the previous window (on the same chromosome) by 24 genes was calculated. In each window, the Z-score relative to all cells at day 0 was calculated. The net subchromosomal aberration score for a cell is calculated as the 12-norm of the Z-scores across all windows. To assign a p-value to the subchromosomal aberration score for celli, the empirical probability that the score for celli in the randomly permuted data was at least as large as the score in the original data was calculated.

[0220] For subchromosomal aberration scores chromosomal aberrations (vs. locally coordinated programs of gene expression) were enriched for by excluding recurrent events. Recurrent events were identified by clustering cells based on their aberration profiles (net expression levels across all windows). Clustering was completed by calculating the SVD of all aberration profiles, and performing KMeans clustering on the the top 10 singular vectors (with k=100). For each cluster, we quantified cluster compactness and separation using the silhouette score. Cells that were in compact, well-separated clusters (with a silhouette score > 0.08) were removed from consideration for subchromosomal aberrations.

[0221] For both types of scores, p-values were used to calculate false discovery rates (FDRs). To identify cells with aberrations at an FDR of q, the largest p-value, $\hat{p}$ was identified, such that pN/sum(p< p), where N represents the total number of p-values for a score and sum(p< $\hat{p}$) represents the number of p-values less than $\hat{p}$.

[0222] Since recurrent aberrations are expected in this setting (due to clonal expansion) cells based on clustering recurrent patterns were not removed. Applied to these data, this method detected aberrations in 35% of malignant cells (classified in the original study as containing significant copy number variation) and 0% of non-malignant cells (FDR 5%). This demonstrates the specificity and conservative nature of the approach.

[0223] Results. The results of this analysis are displayed in Figs. 22A-22C. In analysis designed to look for whole chromosome aberrations, it was found that 0.9% of cells showed significant up- or downregulation across an entire chromosome; the expression-level changes were largely consistent with gain or loss of a single chromosome (AHA). Next, analysis performed to look for evidence of large subchromosomal events, found significant events in 0.8%o of cells. The frequency was highest (2.8%>) in cluster 14, consisting of cells in the Valley of Stress enriched for a DNA damage-induced apoptosis signature. The frequency was 2-to-3-fold lower in other cells in the Valley (enriched for senescence but not apoptosis), in cells en route to the Valley (clusters 8 and 11), and in fibroblast-like cells at days 0 and 2. Notably, it was much lower (6-fold) in cells on the trajectory to successful reprogramming **(Figs. 22B, 22C).** Direct experimental evidence would be needed to confirm these events, and to clarify if the aberrations were preexisting in the MEF population, or if they accumulated during the course of reprogramming.

Example 20

[0224] Forced expression of transcriptional regulators enhances reprogramming.

[0225] To test whether any of the transcriptional regulators provided in Tables 2, 3 and 4, for example, Obox6, Spic, Zfp42, Sox2, Mybl2, Msc, Nanog, Hesxl and Esrrb, play an active role in the process of reprogramming, experiments are performed to address whether expressing these transcription regulators along with OKSM during days 0-8 can boost reprogramming efficiency. Secondary MEFs or primary MEFS are infected with a Dox-inducible lentivirus carrying any one of the transcription regulators provided in Tables 2, 3 and 4, the known pluripotency factor Zfp42 (73), or no insert as a negative control. Reprogramming efficiency is assessed in 2i or in serum. Multiple independent experiments are performed. An increase in reprogramming efficiency by a transcriptional regulator identifies the regulator as important in the context of cellular reprogramming.

[0226] Reprogramming efficiency is assessed by analyzing bright field and fluorescence images of iPSC colonies generated by lentiviral overexpression of Oct4, Klf4, Sox2, and Myc (OKSM) with either an empty control, Zfp42 or an expression cassette for any one of the transcription regulators provided in Tables 2, 3 and 4, in either Phase-l(Dox)/Phase-2(2i)(A) and Phase- l(Dox)/Phase-2(serum). Cells are imaged at day 16 to measure Oct4-EGFP+ cells. Bar plots representing average percentage of Oct4-EGFP+ colonies in each condition on day 16 are generated. Error bars represent standard deviation for biological replicates.

Example 20

[0227] Reconstruction of developmental landscapes by optimal-transport analysis of single-cell gene expression across time sheds light on reprogramming

[0228] Here, we introduced Waddington-OT, a new approach for studying developmental time courses to infer ancestor-descendant fates and model the regulatory programs that underlie them. We applied Waddington-OT to reconstruct the landscape of reprogramming from 315,000 scRNA-seq profiles, collected mostly at half-day intervals across 18 days. We revealed a wider range of developmental programs than previously recognized. Cells gradually adopted either a terminal stromal state or a mesenchymal-to-epithelial transition state. The latter gave rise to populations related to pluripotent, extra-embryonic, and neural cells, with each harboring multiple finer subpopulations. We predicted transcription factors controlling various fates, of which we showed that Obox6 enhanced reprogramming efficiency. We also found rich potential for paracrine signaling. Our approach shedded new light on the process and outcome of reprogramming and provided a framework applicable to diverse temporal processes in biology.

[0229] In the mid-20th century, Waddington introduced two metaphors that shaped biological thinking about cellular differentiation during development: first, trains moving along branching railroad tracks and, later, marbles following probabilistic trajectories as they roll through a developmental landscape of ridges and valleys (Waddington, 1936, 1957). Empirically reconstructing and studying the actual landscapes, fates and trajectories associated with cellular differentiation and de-differentiation — such as in organismal development, long-term physiological responses, and induced reprogramming — requires general approaches to answer questions such as: What classes of cells are present at each stage? What was their origin at earlier stages? What are their likely fates at later stages? What genetic regulatory programs control their dynamics? To what extent are events synchronous vs. asynchronous? To what extent are they stochastic vs. deterministic? Is there only a single path to a given fate, or are there multiple developmental paths?

[0230] Traditional approaches based on bulk analysis of cell populations were not well suited to addressing these questions, because they did not provide general solutions to two challenges: discovering the cell classes in a population and tracing the development of each class. Progress had historically relied on ad hoc approaches for each question asked (e.g., sorting and following the development of a particular cell class by using an antibody to a class-specific cell-surface protein or a reporter construct).

[0231] The first challenge has recently been largely solved by the advent of single-cell RNA-Seq (scRNA-Seq) (Klein et al., 2015; Kumar et al., 2014; Macosko et al., 2015; Ramskold et al., 2012; Shalek et al., 2013; Tanay and Regev, 2017; Tang et al., 2009; Wagner et al., 2016), which allowed cell classes to be discovered based on their expression profiles. The second challenge remained a work-in-progress. ScRNA-seq now offered the prospect of empirically reconstructing developmental trajectories based on snapshots of expression profiles from heterogeneous cell populations undergoing dynamic transitions (Bendall et al., 2014; Marco et al., 2014; Setty et al., 2016; Tanay and Regev, 2017; Trapnell et al., 2014; Wagner et al., 2016). But, to trace the trajectories of cell classes, one may connect the discrete 'snapshots' produced by scRNA-Seq into continuous 'movies.' At least at present, one may not be able to follow expression profiles of the same cell and its direct descendants across time because current methods may destroy cells to profile their state. While various approaches have been developed to record information about cell lineage, they currently provide only very limited information about a cell's state at all earlier time points (Daniel T. Montoro et al., 2018; Kester and van Oudenaarden, 2018; McKenna et al., 2016).

[0232] Comprehensive studies of cell trajectories thus relied heavily on computational reconstruction of paths in gene-expression space. Pioneering work introduced various methods to infer trajectories (Bendall et al., 2014; Cannoodt et al., 2016; Haghverdi et al., 2015; Matsumoto and Kiryu, 2016; Qiu et al., 2017; Rashid et al., 2017; Rostom et al., 2017; Setty et al., 2016; Street et al., 2017; Trapnell et al., 2014; Weinreb et al., 2017; Welch et al., 2016; Zwiessele and Lawrence, 2016). Profiles of heterogeneous populations can provide information about the temporal order of asynchronous processes—enabling cells to be ordered in pseudotime along trajectories, based on their state of differentiation (Bendall et al., 2014). Some approaches used k-nearest neighbor graphs (Bendall et al., 2014) or binary trees (Trapnell et al., 2014) to connect cells into paths. More recently, diffusion maps have been used to order cell-state transitions, by assigning cells to densely populated paths in diffusion-component space (Haghverdi et al., 2015; Haghverdi et al., 2016). Each such path was interpreted as a transition between cellular fates, with trajectories determined by curve fitting and cells pseudotemporally ordered based on the diffusion distance to the endpoints of each path. Recent work has grappled with incorporating branching paths, which were critical for understanding developmental decisions, and have been applied to analyze whole-organism development in zebrafish, frog, and planaria (Briggs et al., 2018; Farrell et al., 2018; Fincher et al., 2018; Plass et al., 2018; Wagner et al., 2018).

[0233] While these approaches have shed important light on various biological systems, many important challenges remain. First, most methods neither directly modeled nor explicitly leveraged the temporal information in a developmental time course (Weinreb et al., 2017) because they were designed to extract information about stationary processes (such as adult stem cell differentiation or the cell cycle) in which all stages existed simultaneously across a single population of cells. However, with the rapidly decreasing cost of scRNA-Seq, time-courses may soon be commonplace. Second, many methods model trajectoried in the language of graph theory which imposesed strong structural constraints on the model, such as one-dimensional trajectories ("edges") and zero-dimensional branch points ("nodes"). Yet, some biological systems may show a gradual divergence of fates that were not captured well by these models

(Briggs et al., 2018; Farrell et al., 2018; Wagner et al., 2018). Third, few methods were able to account for cellular growth and death during development. One method capable of modeling nonuniform cellular growth rates was Population Balance Analysis (Weinreb et al., 2017). However, this method assumed the population of cells is in equilibrium, and therefore it was not suited for analyzing dynamical systems where the distribution of cells changed over time.

[0234] One case in point was the challenge of understanding cellular reprogramming—such as converting fibroblasts to induced pluripotent stem cells (iPSCs) or trans-differentiating one mature cell type into another. These non-natural processes involved the transient overexpression of a set of transcription factors (TFs) designed to push a cell out of its current state and toward a new fate, even in the absence of the usual developmental context. Reprogramming had great therapeutic potential, but it still tends to be slow, inefficient, and asynchronous (Takahashi and Yamanaka, 2016). Single-cell analysis of trajectories during reprogramming could shed light on questions such as: What is the full range of cell classes that arise during reprogramming? What are the developmental paths that lead to reprogramming and to any alternative fates? Which cell intrinsic factors and cell-cell interactions drive progress along these paths? To what extent do cells activate normal developmental programs vs. unnatural hybrid programs? Can the programs that are activated provide information about the normal developmental landscape? Can the information gleaned be used to improve the efficiency of reprogramming toward a desired destination?

[0235] In particular, reprogramming of fibroblasts to induced pluripotent stem cells (iPSCs), as pioneered by Yamanaka (Hou et al., 2013; Shu et al., 2013; Takahashi and Yamanaka, 2006; Yu et al., 2007), has been largely characterized to date by a combination of fate-tracing of cells based on a handful of markers (e.g., Thyl and CD44 as markers of the fibroblast state, and ICAM1, Oct4, and Nanog as markers of successful reprogramming), together with RNA- and chromatin-profiling studies of bulk cell populations (Buganim et al., 2012; Hussein et al., 2014; O'Malley et al., 2013; Polo et al., 2012; Tonge et al., 2014). With limited cellular resolution, the profiling studies have provided only coarse-grained analyses, such as describing two "transcriptional waves," with gain of proliferation and loss of fibroblast identity followed by transient activation of developmental regulators and gradual activation of embryonic stem cell (ESC) genes (Polo et al., 2012). Some studies (Mikkelsen et al., 2008; O'Malley et al., 2013;

Parenti et al., 2016), including from our own group (Mikkelsen et al., 2008), have noted strong upregulation of several lineage-specific genes from unrelated lineages (e.g., neurons), but it has been unclear whether this reflects coherent differentiation of specific cell types or disorganized gene expression (Kim et al., 2015; Mikkelsen et al., 2008). Most studies that used single-cell methods to study genetic reprogramming have involved few genes or few cells (Buganim et al., 2012, Kim et al., 2015). Recently, a study (Zhao et al., 2018) profiled -36,000 cells during chemical reprogramming, but focused only on a single bifurcation separating successful and failed trajectories.

[0236] Here, we described a framework, implemented in a method called Waddington-OT, that aimed to capture the notion that cells at any time were drawn from a probability distribution in gene-expression space and cells at any time and position within the landscape had a distribution of both probable origins and probable fates **(FIGs. 23A-23F).** It then used scRNA-seq data collected across a time-course to infer how these probability distributions evolved over time, by using the mathematical approach of Optimal Transport (OT). We applied and tested this framework in the context of scRNA-seq data we profiled from more than 315,000 cells, sampled across a dense time course over 18 days under two different reprogramming conditions. We found that reprogramming unleashed a much wider range of developmental programs and subprograms than previously recognized, resulting in multiple large distinct populations of cells related to pluripotent, extraembryonic, neural, and stromal cells, with evidence for large-scale genomic amplifications and deletions in trophoblast-like and stromal-like cells. Within each population, there were subsets with distinct programs associated with specific cell types in vivo, including programs associated with 2-, 4-, 8-, 16-, and 32-cell stage embryos; with several distinct types of trophoblasts and primitive endoderm; with astrocytes, oligodendrocytes, and neurons; and with a wider range of stromal cells than MEFs. Trajectory analysis with Waddington-OT showed that differentiation among these classes occurred gradually, including an early gradual transition to either stroma-like cells or a mesenchymal-to-epithelial transition state, with the latter state serving as the ancestor population of both eventual iPSC-like cells and extraembryonic and neural. These differentiation fates were predicted by various sets of TFs, including well studied factors and others not previously implicated. We tested one TF found by our analysis to be associated with pluripotency and showed that it enhanced reprogramming efficiency. Finally, we also found evidence for potential paracrine interactions between the stromal cells and other cell types, which may be important cell extrinsic forces in reprogramming, and for genomic aberrations in certain cells types, with different features in stromal cells and trophoblasts.

[0237] Results

[0238] Reconstruction of probabilistic trajectories by Optimal Transport

[0239] **A** goal of the study was to learn the relationship between ancestor cells at one time point and descendant cells at another time point: given that a cell has a specific expression profile at one time point, where will its descendants likely be at a later time point and where are its likely ancestors at an earlier time point? To this end, we modeled a differentiating population of cells as a time-varying probability distribution (i.e., stochastic process) on a high-dimensional gene expression space. By sampling this probability distribution Pt at various time points t, we aimed to infer how the differentiation process it modeled evolves over time **(FIG. 23A).** By sampling a large number of cells at a given time point, we approximated the distribution at that time point. However, this alone did not tell us the ancestor or descendant relationships between cells at different time points: Because different cells were sampled at different time points, we lost this temporal coupling of the stochastic process Pt that specified the joint distribution of expression between pairs of time points. In the absence of any constraint on cellular transitions (e.g., if cells may "jump" about gene-expression space arbitrarily rapidly), we could not infer the temporal coupling. But if we assumed that, over sufficiently short time periods, cells could only move relatively short distance, we could infer the temporal coupling by using the classical mathematical technique of optimal transport **(FIG. 23A, Methods).**

[0240] Optimal transport was originally developed by Monge in 1781 to redistribute earth for the purpose of building fortifications with minimal work (Villani, 2008). In the 1940s, Kantorovich generalized it to identify an optimal coupling of probability distributions via linear programming (Kantorovitch, 1958). This classical linear program minimized the total squared distance that earth travels, subject to conservation of mass constraints. Recent work, which added entropic regularization, dramatically accelerated the numerical computation of large-scale optimal transport problems (Chizat et al., 2017; Cuturi, 2013).

[0241] However, matching cells to their descendants differed in one important aspect: unlike earth or particles, cells can proliferate. We therefore modified the classical conservation of mass constraints to accommodate cell growth and death. In particular, we allowed the mass of cells to grow as cells proliferate and shrink as cells die (STAR Methods). By leveraging techniques from unbalanced transport (Chizat et al., 2017), we automatically learned cellular growth and death rates, initializing with prior estimates from signatures of cellular proliferation and apoptosis (STAR Methods).

[0242] Using optimal transport, we calculated couplings between consecutive time points and then inferred couplings over longer time-intervals by composing the transport maps between every pair of consecutive intermediate time points. We noted that the optimal-transport calculation (i) implicitly assumed that a cell's fate depended on its current position but not on its previous history (i.e., the stochastic process is Markov) and (ii) captured only the time-varying components of the distribution, rather than processes at dynamic equilibrium. We returned to these points in the Discussion.

[0243] We defined trajectories in terms of "descendant distributions" and "ancestor distributions" as follows. For any set C of cells at time ti , its "descendant distribution" at a later time ti+1 referred to the mass distribution over all cells at time ti+1 obtained by transporting C according to the transport maps (FIG. 23C). Branching events, for example, were revealed by the (potentially gradual) emergence of bimodality in the descendant distribution (FIG. 23C). Conversely, its "ancestor distribution" at an earlier time ti-1 was defined as a mass distribution over all cells at time ti-1, obtained by transporting C in the opposite direction (that is, as though one "rewinds" time) (FIG. 23D). Shared ancestry between two cell sets at ti was revealed by convergence of the ancestor distributions (FIG. 23E). The "trajectory from C" referred to the sequence of descendant distributions at each subsequent time point, and the trajectory to C similarly referred to the sequence of ancestor distributions (FIGs. 23C, 23D). For convenience below, we sometimes referred simply to the 'ancestors, 'descendants', and 'trajectories' of cells. These terms referred to probability distributions over a set of observed cells that served as proxies for the actual ancestors or descendants. In summary, we used the inferred coupling to calculate a distribution over representative ancestors and descendants at any other time. We then determined the expression of any gene or gene signature along a trajectory by computing the mean expression level weighted by the distribution over cells at each time point.

[0244] To identify TFs that regulated the trajectory, we inferred regulatory models by sampling cells from the joint distribution given by the couplings. We developed two approaches: one used 'local' enrichment analysis, identifying TFs that were enriched in cells having many vs. few descendants in the target cell population; a second built a global regulatory model, composed of modules of TFs and modules of target genes, to predict expression levels of target gene signatures (FIG. 23F, left) at later time points from expression levels of TFs at earlier time points (FIG. 23F, middle, right).

[0245] We implemented our approach in a method, Waddington-OT, for exploratory analysis of developmental landscapes and trajectories, including a public software package (STAR Methods). The method included: (1) Performing optimal-transport analyses on scRNA-seq data from a time course, by calculating optimal-transport maps and using them to find ancestors, descendants and trajectories; (2) Inferring regulatory models that drive the temporal dynamics by sampling pairs of cells from the joint distribution specified by the OT couplings; (3) Visualizing the developmental landscape in two dimensions, by using Force-Directed Layout Embedding (FLE) to visualize the graph of nearest neighbor relationships in diffusion component space (Jacomy et al., 2014; Weinreb et al., 2016; Zunder et al., 2015), and (4) annotating the landscape by cell types, ancestors, descendants, trajectories, gene expression patterns, and other features.

[0246] A dense experimental scRNA-Seq time course of iPS reprogramming

[0247] To study the trajectories of reprogramming, we generated iPSCs via a secondary reprogramming system (FIG. 24A), which is more efficient than derivation of iPSCs by primary infection (Stadtfeld et al., 2010). We obtained mouse embryonic fibroblasts (MEFs) from a single female embryo homozygous for ROSA26-M2rtTA, which constitutively expresses a reverse transactivator controlled by doxycycline (Dox), a Dox-inducible polycistronic cassette carrying Pou5fl (Oct4), Klf4, Sox2, and Myc (OKSM), and an EGFP reporter incorporated into the endogenous Oct4 locus (Oct4-IRES-EGFP). We plated MEFs in serum-containing induction medium, with Dox added on day 0 to induce the OKSM cassette (Phase-1(Dox)). Following Dox withdrawal at day 8, we transferred cells to either serum-free N2B27 2i medium (Phase-2(2i)) or maintained the cells in serum (Phase-2(serum)). Oct4-EGFP+ cells emerged on day 10 as a reporter for successful reprogramming to endogenous Oct4 expression (FIGs. 24A, 30G).

[0248] We performed two dense time-course experiments. In the first we collected -65,000 scRNA-seq profiles at 10 time points across 16 days, with samples taken every 48 hours. In the second we profiled -250,00 0 cells collected at 39 time points across 18 days, with samples taken every 12 hours (and every 6 hours between days 8 and 9) (FIG. 24A, Methods, Table 14). The density allows us to ensure that the model is fit on a smoothly progressing process, as well as to use some time points as test data for predictions (below). We also collected samples from established iPSC lines reprogrammed from the same MEFs, maintained in either 2i or serum conditions. The two experiments were consistent (STAR Methods). We focused on the second experiment, where we profiled 259,155 cells to an average depth of 46,523 reads per cell (Table 14). After discarding cells with less than 2,000 transcripts detected, we retained a total of 251,203 cells, with a median of 2,565 genes and 9,132 unique transcripts detected per cell.

Table 14 - Summary of single cell sequencing statistics and sample information.

Sample Name	Estimated Number of Cells	Mean Reads per Cell	Median Genes per Cell	Number of Reads	Valid Barcodes	Reads Mapped Confidently to Transcriptome	Reads Mapped Confidently to Exonic Regions	Reads Mapped Confidently to Intronic Regions	Reads Mapped Confidently to Intergenic Regions	Reads Mapped Antisense to Gene	Sequencing Saturation	Q30 Bases in Barcode	Q30 Bases in RNA Read	Q30 Bases in Sample Index	Q30 Bases in UMI	Fraction Reads in Cells	Total Genes Detected	Median UMI Counts per Cell
D0_Dox_C1	3495	17263	2308	60336236	98	62.7	66.1	10.8	5.4	4.4	17.4	97.9	90.9	95.8	97.7	92.2	16467	7421
D0_Dox_C2	1125	41979	3559	47227004	98	64.2	67.6	10.5	4.9	4.3	30.8	97.9	90.6	96.3	97.7	92.4	15884	15756
D0.5_Dox_C1	1220	65642	4258	80083266	97.9	63.4	66.9	11.3	5	4.4	38.7	97.9	90.6	95.8	97.7	95.5	16658	22429

12851	6263	8318	27357	48498	11247	2275	5041	7728	8215	18216
16911	15028	16161	17182	15562	17003	14980	15423	16143	16144	17099
90.3	89	94	91.8	78.5	92.5	87.1	92.7	95.6	94.4	93.5
97.5	97.4	97.5	97.7	97.6	97.6	97.6	97.8	97.5	97.5	97.5
96.2	95.8	96	95.5	96.1	96.1	95.9	96.3	95.6	95.9	96.1
87.8	85.3	88.2	91.3	89	90.5	88.8	90.6	87.4	87.6	87.1
97.8	97.7	97.8	97.9	97.9	97.9	97.8	98	97.8	97.8	97.7
22.5	12.8	13.5	33.3	64.6	18.9	10.2	13	14.7	15.8	26.1
4.6	6.6	5.2	4	4.7	3.5	4.4	3.9	4.2	4.4	4.2
5.2	2.9	7.4	9.2	3.1	9.8	3.3	5.8	4.4	3.3	3.8
10.2	9.7	11.4	12.6	8.9	12.4	7.9	10.7	9.4	9.5	9.1
65.7	73.6	55.8	50.2	74.9	47.6	75.6	60.4	69	73.7	71.9
61.9	67.8	51.8	47.4	71.1	45.3	71.9	57.5	65.4	69.8	68.1
98.3	98.1	98.1	97.9	98.3	97.9	98.2	98.4	98.3	98.2	98.2
72036482	17538332	49231019	1.7E+08	80424447	1.64E+08	23593131	37988832	59391343	58972209	1.3E+08
3230	2366	2776	4926	6159	3154	1007	1838	2296	2314	3630
32317	12500	21111	103491	253704	37710	4443	11931	15914	16055	41424
2229	1403	2332	1639	317	4360	5310	3184	3732	3673	3148
D0.5_Dox_C2	**D1_Dox_C1**	**D1_Dox_C2**	**D1.5_Dox_C1**	**D1.5_Dox_C2**	**D2_Dox_C1**	**D2_Dox_C2**	**D2.5_Dox_C1**	**D2.5_Dox_C2**	**D3_Dox_C1**	**D3_Dox_C2**

6138	3562	11428	16183	20437	20725	20293	28005	16917	12974	19034
15929	14788	16574	17265	17466	17681	17882	17837	17425	16996	18190
96.3	96.6	97	97.6	95.9	96.2	96.3	94.9	96	96	95.1
97.6	97.6	97.6	97.6	97.6	97.3	97.3	97	97.3	97.3	97.8
95.7	96.3	96.1	95.9	96	95.3	95.7	95.2	95.3	95.9	96
89.3	89.7	89.6	89.7	87.9	83.1	82.9	84.1	83.4	84.3	92
97.9	97.9	97.9	97.9	97.8	97.6	97.6	97.5	97.6	97.6	98
15.3	12.1	22.5	28.9	38.2	31.5	34.4	42.1	37.5	27.4	56.6
4.6	4.6	4.5	4.5	4.7	5.5	5.5	5.1	5.4	5	3.7
3.3	3	3	2.6	2.8	2.7	2.7	2.5	2.8	2.7	3.1
9	9	9	8.4	8.9	7.6	7.3	7.4	8	7.5	10
74.5	76.3	76.1	77.8	75.8	74.6	74.7	76.5	74.5	75.8	75.5
70.7	72.4	72.3	74	71.8	69.6	69.7	71.7	69.7	71.4	73
98.3	98.3	98.4	98.1	98.3	98.3	98.4	98.3	98.4	98.4	98.4
55079302	21741409	94013331	1.69E+08	1.88E+08	1.78E+08	2.01E+08	2.5E+08	1.48E+08	92501384	4.19E+08
1782	1284	2532	3078	3490	3460	3308	3986	3032	2586	3223
11906	6320	23014	34713	52881	49701	49996	77855	44353	28798	75461
4626	3440	4085	4877	3551	3576	4018	3209	3338	3212	5554
D3.5_Dox_C1	**D3.5_Dox_C2**	**D4_Dox_C1**	**D4_Dox_C2**	**D4.5_Dox_C1**	**D4.5_Dox_C2**	**D5_Dox_C1**	**D5_Dox_C2**	**D5.5_Dox_C1**	**D5.5_Dox_C2**	**D6_Dox_C1**

39404	32776	25293	27686	25478	19859	11274	6435	4995	6758	5702
18938	16277	17548	18209	18024	17416	16519	15616	15285	15657	15714
95.6	96.7	96.2	94.8	95.5	94.3	92.7	90.6	91.7	93.1	92.6
97.9	97.8	97.8	97.8	97.8	97.8	97.8	97.6	97.6	97.6	97.6
96.4	96.4	96	96.2	96	96	95.7	95.8	96.1	96	96
93.2	92.6	92.1	92.1	92.2	92	92.3	90.9	90.4	90.3	90.3
98.1	98	98	98	98	98	98	97.9	97.9	97.9	97.9
85.2	81.8	54.1	65.5	47.9	51.1	26.3	23.2	20.7	21.2	19.1
4	4.5	3.9	4.1	4	3.9	3.8	3.9	3.9	4.4	4.5
3.5	2.8	2.5	3.2	3	3.1	3.7	5.7	4.2	3.3	4.1
9.7	11.6	9.1	11.2	10.7	11.1	10	10.4	9.1	9.2	9.3
73.7	73.3	77.1	73.1	73.9	73.7	72.3	64.3	71.4	75.2	71.4
71.2	70.2	74.4	70.2	71.1	70.9	69.8	61.3	68.2	71.5	67.8
98.5	98.4	98.4	98.3	98.3	98.4	98.4	98.2	98.4	98.3	98.3
1.35E+09	1.55E+08	2.21E+08	4.31E+08	2.72E+08	1.78E+08	65541812	33456383	24003361	28066499	27516277
4897	4717	4114	4327	4154	3667	2494	1644	1374	1692	1587
471033	290563	85899	137190	80817	68735	26535	17805	11221	15122	12979
2868	535	2576	3138	3369	2591	2470	1879	2139	1856	2120
D6_Dox_C2	**D6.5_Dox_C1**	**D6.5_Dox_C2**	**D7_Dox_C1**	**D7_Dox_C2**	**D7.5_Dox_C1**	**D7.5_Dox_C2**	**D8_Dox_C1**	**D8_Dox_C2**	**D8.25_2i_C1**	**D8.25_2i_C2**

7892	6359	19378	14092	14336	12381	4785	5962	5629	10133	15871
15808	15972	16274	16219	16335	16274	15033	15231	15445	16266	16091
90.7	88.9	92.6	92.8	90.7	91.9	92.2	92.2	89.6	87.1	85.3
97.7	97.7	97.8	97.8	97.8	97.8	97.7	97.7	97.6	97.7	97.8
95.6	96.1	96.3	96.2	95.7	96	96.1	95.7	95.8	96.1	96.4
91.4	90.7	93.5	93.5	93.4	93.6	91.7	90.5	90.1	90.6	93.7
97.9	97.9	98	98	98	98	98	97.9	97.9	97.9	98
25.9	25.2	50.1	36.2	39.6	35.8	17.6	19.1	18.8	26.3	52.1
3.8	3.6	3.8	3.9	4	3.9	3.7	3.9	3.9	3.7	3.9
5.4	4.5	2.4	2.4	2.3	2.4	3.2	2.9	2.9	3.1	2.3
10.7	8.9	7.2	7	7.8	7.7	9	8.8	9.4	9.5	7.3
65	70.7	79.6	79.8	78.9	78.6	75.3	76.8	76	75	79.5
62.2	67.9	76.5	76.6	75.6	75.6	72.5	73.5	72.7	71.9	76.4
98.2	98.4	98.2	98	98	97.9	98.4	98.4	98.3	98.3	97.8
34670761	38854100	71646422	57753221	66514572	60937426	17654865	20225030	20630020	40237550	64328422
1901	1601	3119	2534	2653	2451	1333	1552	1529	2275	2817
22382	16332	60410	35193	40214	31754	9830	12257	12766	26367	59016
1549	2379	1186	1641	1654	1919	1796	1650	1616	1526	1090
D8.25_serum_C1	**D8.25_serum_C2**	**D8.5_2i_C1**	**D8.5_2i_C2**	**D8.5_serum_C1**	**D8.5_serum_C2**	**D8.75_2i_C1**	**D8.75_2i_C2**	**D8.75_serum_C1**	**D8.75_serum_C2**	**D9_2i_C1**

13794	6160	8071	9665	13737	8356	8383	5660	9422	7906	3321
15694	15502	15526	15662	15572	15936	15754	15323	15798	16178	14888
94.5	95	95.2	90.5	89.9	87.2	86.6	91.3	92.5	83.5	85.1
97.8	97.8	97.9	97.6	97.7	97.7	97.6	97.8	97.8	97.8	97.8
96.2	96.2	96	95.9	96.3	96.1	96.2	95.9	95.9	96	95.6
93.7	93.5	93.6	90.4	90.7	90.8	90.3	92.5	92.3	92.2	91.9
98	98	98	97.9	97.9	97.9	97.9	98	98	98	98
42.9	52.1	64.2	40.4	49.8	39.1	41.1	24.7	33.7	31.1	15.8
3.6	3	3.1	3.3	3.5	3.5	3.2	3.5	3.5	3.6	3.4
2.2	1.8	2	3.3	4	3.9	2.9	5.9	5	4.1	3.3
7	4.4	5.2	9.7	9.6	10.9	8.7	12	11.8	12.7	11.9
80.3	85.3	83.8	75.9	72.9	71.4	78.1	63.8	67.1	69.3	73.6
77.5	83.2	81.7	73.1	70	68.6	75.3	61.3	64.7	66.7	71.1
98.1	98.5	98.5	98.3	98.2	98.2	98.3	98.1	98.2	98.1	98.3
34630027	33750278	40057020	29703571	31593148	31931324	29811637	17333643	27704152	33583765	8955917
2753	1977	2317	2185	2732	2056	1892	1645	2358	2068	1210
36684	18322	32382	29973	52831	27622	26127	16523	30277	26013	7939
944	1842	1237	991	598	1156	1141	1049	915	1291	1128
D9_2i_C2	**D9_serum_C1**	**D9_serum_C2**	**D9.5_2i_C1**	**D9.5_2i_C2**	**D9.5_serum_C1**	**D9.5_serum_C2**	**D10_2i_C1**	**D10_2i_C2**	**D10_serum_C1**	**D10_serum_C2**

11465	9225	8158	6896	8173	8421	4054	4176	11511	14816	15611
16115	15697	15951	15650	15758	15560	15335	15379	16398	16538	17172
92.4	91.8	72.5	78.8	79.2	89.8	86	80.8	88.4	90.7	85.8
97.7	97.7	97.8	97.8	97.8	97.8	97.8	97.7	97.8	97.7	97.8
95.5	95.7	96	96.1	96.2	95.7	96.1	95.7	95.5	96.3	96.2
91.8	91.9	91.7	92.2	92	92.6	91.5	90.4	92	91.9	91.6
98	98	98	98	98	98	98	97.9	98	97.9	98
30.1	25.8	29	30.8	29.4	27.2	19.4	25.6	40.9	49	60.1
3.7	3.7	3.8	3.5	3.7	3.8	3.5	3.6	3.7	3.6	3.5
3.6	3.6	2.8	2.7	2.4	3	2.3	2	2.7	2.7	2.5
13	11.9	11.8	11	9.2	10.5	10.7	8.5	10.7	10.5	11.6
71.4	71.5	74.7	76	78.3	74.4	78.3	81.5	76.6	76.7	76.7
68.5	68.8	72	73.6	75.6	71.9	75.7	78.8	73.9	74.1	74.1
98.1	98.1	98.2	98.2	98.2	98.2	98.3	98.4	98.3	98.3	98.2
24523951	17574924	26189701	22243909	18033999	13379426	12888357	12788655	27834347	35823619	90774725
2717	2369	2313	2171	2171	2308	1585	1692	2783	3298	3586
31973	25324	27167	21765	23981	22188	9160	10612	38658	54360	77058
767	694	964	1022	752	603	1407	1205	720	659	1178
D10.5_2i_C1	**D10.5_2i_C2**	**D10.5_serum_C1**	**D10.5_serum_C2**	**D11_2i_C1**	**D11_2i_C2**	**D11_serum_C1**	**D11_serum_C2**	**D11.5_2i_C1**	**D11.5_2i_C2**	**D11.5_serum_C1**

5562	10044	12519	8119	7210	10070	15004	10108	21756	12776	11522
15665	16604	16529	16471	16513	16343	16879	16850	18479	16853	16820
86.2	86.2	85	84.8	85.4	84.3	86	84.7	81.5	66.3	49.1
97.8	97.8	97.8	97.7	97.7	97.7	97.7	97.7	97.7	97.7	97.7
95.6	96.2	96	96	96.2	96.1	96.1	96	96.1	96.1	95.8
91.9	92	91.4	91	90.6	91	91.2	90.8	90.8	90.8	90.8
98	98	98	98	97.9	97.9	97.9	97.9	97.9	98	98
23.6	51.4	55.3	35.4	29.9	37.9	47.7	35	67.1	56.4	72.9
3.5	4.1	3.8	3.6	3.6	4.1	4	3.7	3.8	4.3	4.3
2.4	2.8	2.7	2.4	2.3	2.9	2.9	2.3	2.4	3.1	2.8
10.9	8.6	7.8	9.4	9.3	8.5	8.4	8.5	8.8	8.8	8.3
77.4	77.1	78.7	78.7	79.2	76.8	76.8	79.7	79.2	75.5	76.8
74.9	74.3	76	76.1	76.4	73.7	73.8	76.8	76.3	72.1	73.4
98.2	98.5	98.5	98.4	98.4	98.4	98.4	98.3	98.3	98.3	98.3
15149367	34932625	36075300	27804384	27170840	22372820	36585438	30987716	1.66E+08	49269432	1.01E+08
1903	2523	2880	2468	2358	2560	3214	2816	4369	2938	2866
14238	42704	58092	25116	20552	32471	54768	29456	138451	75220	156892
1064	818	621	1107	1322	689	668	1052	1201	655	643
D11.5_serum_C2	**D12_2i_C1**	**D12_2i_C2**	**D12_serum_C1**	**D12_serum_C2**	**D12.5_2i_C1**	**D12.5_2i_C2**	**D12.5_serum_C1**	**D12.5_serum_C2**	**D13_2i_C1**	**D13_2i_C2**

12190	15494	5599	5146	5287	5360	15207	20543	10816	14705	12798
17377	18070	16769	15987	16853	16725	18525	18764	18461	18884	18532
77.6	85.4	74.6	75.3	77.2	71.1	91.6	93.6	87.9	87.1	81.5
97.8	97.8	97.8	97.7	97.8	97.8	97.8	97.7	97.7	97.7	97.8
96.3	96.1	96.3	95.7	95.9	95.9	96.3	96.2	96	96.1	96
92.1	92.2	92.5	90.8	90.9	91	91.8	91.7	91.4	91.5	92
98	98	98	97.9	98	97.9	98	97.9	97.9	98	98
73.7	67.1	69.4	52.4	70.2	68.1	37	42.1	39.5	50.7	36.7
4	4	5.7	5.3	5.6	5.5	4.9	4.8	4.1	3.9	5.6
2.4	2.5	4.9	5.4	3.4	3.5	3.1	3.1	2.5	2.6	3.4
9.6	10.3	16	15.7	14.9	14.6	7.6	7.4	11.2	10.4	9.3
78.1	77	65.4	63.3	70.3	70.6	77.5	77.6	77.2	77.6	74.4
75	73.8	60.7	59	65.9	66.3	73.3	73.5	73.7	74.3	69.7
98.3	98.3	97.5	97.5	97.3	97.3	98.3	98.4	98.3	98.3	98.3
97956936	1.09E+08	49186630	22110011	54618691	46544722	74206890	1.05E+08	70077873	97873582	79779089
3179	3646	1996	1853	2056	2126	3022	3577	2897	3539	2744
99956	93789	46666	26735	43074	42121	39097	54136	34487	56705	39164
980	1166	1054	827	1268	1105	1898	1938	2032	1726	2037
D13_serum_C1	**D13_serum_C2**	**D13.5_2i_C1**	**D13.5_2i_C2**	**D13.5_serum_C1**	**D13.5_serum_C2**	**D14_2i_C1**	**D14_2i_C2**	**D14_serum_C1**	**D14_serum_C2**	**D14.5_2i_C1**

15068	8409	14650	5664	7023	11915	5252	8467	9841	15905	13986
18770	18018	18580	18159	17960	18739	18103	18490	18358	19807	19970
89.7	78.9	79.2	85.3	92.1	66.9	63.9	94.4	94.3	76.9	82.2
97.8	97.8	97.8	97.8	97.7	97.8	97.8	97.7	97.7	97.8	97.8
95.6	96.1	96.4	96.2	95.7	95.7	96	96	96.3	95.9	96
92	91.6	91.9	91.6	91.5	91.5	91.6	91.6	92.1	92	91.9
98	98	98	98	97.9	98	98	97.9	97.9	98	98
33.7	42	59.7	61.6	38.4	39.9	46	21.3	23	66.5	54.1
5.3	4.9	4.1	7.9	6	3.9	5.1	4.5	4.3	4.3	4.4
3.3	2.7	2.4	5	4.8	2.3	2.9	3.4	3.4	2.9	3
8.7	12	10	18	14.1	10	13.5	7.7	7.6	12.6	12.5
75.4	75.8	78.9	63.1	67.5	79	74	76.8	76.8	74.5	74.2
71	71.6	75.6	56.2	62.2	75.7	69.5	72.7	73	70.8	70.4
98.3	98.2	98.4	97.4	97.9	98.3	97.8	98.2	98.3	98.2	98.2
78954514	45618882	1.05E+08	82113379	46137688	86460491	55835189	70721479	66435427	2.48E+08	1.65E+08
3074	2505	3705	1935	2111	3162	2007	1964	2143	3685	3347
37795	33892	76526	32100	20244	48958	25885	16535	19528	107956	64367
2089	1346	1377	2558	2279	1766	2157	4277	3402	2295	2556
D14.5_2i_C2	**D14.5_serum_C1**	**D14.5_serum_C2**	**D15_2i_C1**	**D15_2i_C2**	**D15_serum_C1**	**D15_serum_C2**	**D15.5_2i_C1**	**D15.5_2i_C2**	**D15.5_serum_C1**	**D15.5_serum_C2**

5076	9135	6791	8342	8471	5373	13361	6278	12668	10936	15523
17665	17761	18278	18336	18679	18374	19896	18796	18877	18501	19538
92.2	94.5	57	78.1	89.2	88.7	76.4	65.7	89.9	90.5	88.1
97.8	97.7	97.8	97.7	97.8	97.7	97.8	97.8	97.8	97.8	97.8
96.3	96.2	95.6	96.1	96.3	96.2	96.1	96.3	96.2	96.2	96.1
91.9	91.8	91.5	91.3	91.8	91.6	91.5	91.7	91.9	91.7	91.8
98	97.9	97.9	97.9	97.9	97.9	98	98	98	98	98
38.5	25.7	30.4	36.6	22.6	15.9	47.3	28.2	29.8	23.6	49.4
3.7	3.7	4	4.1	4.2	4.2	3.9	3.9	3.9	3.8	3.9
3.6	3.4	2.5	2.6	3.3	3.3	2.8	2.7	3.4	3.3	3
8.1	7.8	8.7	10.4	7.5	7.6	11.7	10.1	7.8	7.5	11.5
76.2	76.8	78.3	77.3	77.6	77.4	75.9	77.1	77.2	77.9	75.1
72.9	73.4	75	73.7	74	73.9	72.6	73.9	73.9	74.7	71.9
98.4	98.4	98.1	98.2	98.3	98.3	98.2	98.2	98.3	98.4	98.3
52290532	53190608	48558555	48904299	55829324	41911584	1.35E+08	52474229	67119554	46535861	96863752
1343	1921	2182	2467	2124	1618	3393	2119	2807	2539	3583
13315	18996	27763	28886	17424	10237	57651	22716	28918	22044	62052
3927	2800	1749	1693	3204	4094	2350	2310	2321	2111	1561
D16_2i_C1	**D16_2i_C2**	**D16_serum_C1**	**D16_serum_C2**	**D16.5_2i_C1**	**D16.5_2i_C2**	**D16.5_serum_C1**	**D16.5_serum_C2**	**D17_2i_C1**	**D17_2i_C2**	**D17_serum_C1**

12979	14477	10753	12806	9998	18060	17916	9840	9029	10626	20527
19729	18309	18452	19556	19155	18821	18566	19294	19023	17918	18049
86.3	92.1	92.2	85.1	87.9	90.9	90.6	80	77.3	96.4	96.2
97.8	97.8	97.8	97.8	97.8	97.8	97.8	97.8	97.8	97.7	97.7
96.2	96.3	95.9	96.3	96	96.2	96.3	96	96.4	96.1	96.1
91.5	92.1	91.8	91.4	91.8	92.6	92.5	92.3	92	91.3	90.9
98	98	98	97.9	98	98	98	98	98	98	97.9
42	40.2	28.2	44.1	36.5	58.2	54.8	62.7	48.1	20.2	28.8
4	3.8	4	4	3.8	3.9	3.7	4.1	3.9	5.1	5.3
3	3.2	3.1	2.9	2.7	3.5	3.4	3	2.8	3.8	3.8
11.5	6.9	6.9	10.3	10.1	6.3	6	10.4	9.3	9.7	9.5
75	78.6	78.7	76.4	77.7	77.5	78.2	75.5	76.8	71.6	71.7
71.6	75.4	75.4	73.1	74.6	74.3	75	72.1	73.6	67.7	67.6
98.3	98.5	98.4	98.4	98.4	98.4	98.4	98.3	98.3	98.2	98.3
96965300	59918421	54120470	86540688	62361742	1.39E+08	1.04E+08	1.18E+08	77222647	74406713	87759016
3300	2900	2474	3292	2849	2774	2761	2472	2322	2524	3649
45803	36580	22428	44221	29527	69937	63038	62257	40600	21467	46879
2117	1638	2413	1957	2112	1989	1648	1898	1902	3466	1872
D17_serum_C2	**D17.5_2i_C1**	**D17.5_2i_C2**	**D17.5_serum_C1**	**D17.5_serum_C2**	**D18_2i_C1**	**D18_2i_C2**	**D18_serum_C1**	**D18_serum_C2**	**DiPSC_2i_C1**	**DiPSC_2i_C2**

DiPSC_serum_C1	DiPSC_serum_C2
5247	4340
18112	21502
2241	2535
95034273	93322919
98.2	98.2
65.9	67.5
70.1	71.4
10.3	9.8
4.4	4
5.1	4.9
23.2	23.3
97.9	97.9
90.1	90.9
95.9	96.1
97.7	97.7
93.2	90.8
19202	19098
7777	9449

[0249] A model of the developmental landscape

[0250] We visualized the developmental landscape of the 251,203 cells in a two-dimensional FLE (FIG. 24B) and annotated it according to sampling time (FIG. 24C), expression scores of gene signatures, and expression of individual genes (FIG. 24D, Table 15).

Table 15 - List of genes comprising gene signatures.

MEF identity							
Gm5571	Il17rd	Gjd4	Prss23	Atp10a	Eif4g2	Gulp1	Sema3a
Rbfox2	Ptk2	Ccng1	9430030n17rik	Loxl1	Vcl	Shank1	Itgb1
Btbd19	Ehd2	Gpr124	Arntl2	Loxl2	Bcl2l2	Bmp1	Nxn
Actn1	Lats2	Fibin	Sh3rf1	Fbln5	Cd276	Akt1s1	Tmem41b
Gatad2a	Hspg2	8030476l19rik	Mrc2	Ctgf	Lrrc58	Itga9	Sec23a
Med6	4930456g14rik	Ddr2	Mdh1	Efnb2	Wwc2	Abcc1	Gm22
Mex3a	4930429b21rik	Arf4	Rictor	Rxra	Lpp	Eda	Itgb5
Ccdc80	Rps20	Ptprs	Map4k5	Ccnd2	Arl1	B4galt2	Dysf
Mex3c	Vgll3	Sprr2k	Plcl1	Gpc2	Ltbp1	Nid1	Thbs1
Sdpr	Prr15	Adm	11-Sep	Ntf3	Ltbp2	Ncam1	Bc022687
Pcdhb2	Fbxl7	A830029e22rik	Ryk	Kif5b	Wisp1	Shc2	Dnm3os
Trim16	Maged2	9230114k14rik	Tgfb3	Slit2	Igf1r	Uba6	Rnd3
Obsl1	Galntl4	Extl3	Ube2i	Tpm1	Rhobtb3	Tradd	Pik3c2a
Epha1	Pdgfc	Mecom	Tgfb2	Gpc4	Fam198b	Rtel1	2810008m24rik
Stx1b	Tmtc4	Qsox1	Zfp319	Flnb	Cnn2	Bicd2	Spred3
Stau1			Gm10399	4930555b11rik	Glipr2	Adamts12	Senp5

Thbsl Bc022687

Nidi Ncaml

Ltbpl Ltb 2

Gpc2 Ntf3

Plcll

Sprr2k

Rps20 Vgll3

Mex3c Sd r

Serpinel Aa881470 Coll2al 2010300f 17rik Cede 102a Nradd Pard6g Ntn4 5730471h 19rik Sepnl Peg 12 Dpysl3 1110012d 08rik Aktl Zfp286 Ubap21 Samd4 Phc2 Mcam Pla2g4c Fzd7 Pappa Ptk7 Nuakl	Tmtc3 Lpar4 Pcdhl9 Eda2r Pcdhl8 Gprl76 Loc 10050 3471 Mical2 Dzipll Hoxc6 Hoxc5 Mettl4-psl Sec63 Ikbip Tsc22d2 2310076g 05rik Anxa6 Nfatc4 Fnl Wnt9a Sorcs2 Tmeffl C79491 Crlfl 2610034e Olrik	Teadl Snx7 Cdkl4 Cdkn2a Cdkn2b Ccnyll Tubb2a-ps2 Aen Farpl 4930402h 24rik Sh3rf3 Adaml9 Ddbl Cttn 92301 12e 08rik Dbnl Fyttdl Lrrcl5 FkbplO Trubl Zdhhc20 Stonl Hoxdl3 Nudt6 Hoxdl2	Fbxol7 Wnt5a Criml Midi Displ Ubox5 St71 Col5a2 Axl Col5al Zyx Ror2 Wdfy3 Amotl2 Yapl Phldb2 6330562c 20rik Ctnndl Rock2 Maspl Pvtl Tnc Fbln2 Hdlbp	Fine C76332 Capn2 Phlda3 Map3k7 MyhlO D18ertd6 53e Stox2 Igf2r D15ertd6 21e Arid5b TnfrsflOb 261001 1e 03rik Ckap4 Efna2 Picalm CdhlO Ddahl Uba3 0610038b 21rik Gemin7 Ubal Fbnl Lhx9	Sydel Hhat Zmat3 Caldl Pmepal E1301 121 23rik Bag2 Zfp583 Pibfl Pmaipl A130022j 15rik Bcl91 Cpa6 D13ertd7 87e Pabpc41 Zfhx3 Itga5 Txnrdl Htrlb Hmga2 2-Sep Lambl Zfp5 18b Parva	Hs2stl D10ertd6 lOe Cyr61 Gtf3cl Lbh Krt33b Gm6607 D3wsul6 7e Zc3h7b 7630403g 23rik Tnpo2 Cepl70 Pdlim5 Pdlim7 Cad Unc5b 24100181 13rik Loc 10021 6343 Glrx3 Kctd5 Loc26947 2 Myolc 4930562c 15rik Till	Aril 3b Polr2e Itgav Igf2bp3

Plunpotency							
Rhox5	Mt2	Asns	Taf7	Folrl	Sox2	Grhpr	Chmp4c
Tdgfl	Ube2a	Aldoa	Nudt4	Gm7325	Jam2	Higdla	Hsf2bp
Utfl	Khdc3	Tdh	Cox5a	Agtrap	Fkbp3	Rpp25	Polr2e
Mkrnl	Pycard	Gjb3	Sod2	Sppl	Cox7b	Rbpms	Blvrb
Dppa5a	Hsp90aal	Rbpms2	S100al3	Hells	Ash21	Mmp3	Ldhb
Uppl	Prrcl	Prpsl	Fkbp6	Dppa4	Dut	Apobec3	Apocl
ChchdlO	Hatl	Fam25c	Rhox9	Gabarapl 2	Dtymk	Spc24	Syngrl
Klf2	Calcoco2	Eif2s2	Gdf3	Rhox6	Gpx4	Xlr3a	Bexl
Trap 1a	Impa2	Cenpm	2700094K13Rik	Rhoxl	Eif4ebpl	Reel 14	Nr2c2ap
Mylpf	Saa3	Nanog	Fmrlnb	Cdc51	Morel	Mtf2	
1700013H16Rik	Ooep	Ndufa412	Hmgn2	Texl9. 1	Fabp3	Snrpn	
AA467197	Bnip3	Syce2	Ubald2	Trim28	Zfp428	Gml3580	
Dhxl6	Mtl	Gml325 1	Lactb2	Atp5gl	Aqp3	Gmnn	

Cell cycle							
Mcm4	Lbr	Cdk1	Ndc80	Cdca2	Rrm2	Hjurp	Rpa2
Smc4	Cenpf	Slbp	Mcm6	Nasp	Tipin	Tacc3	Gins2
Gtse1	Birc5	Aurkb	Rrm1	Gmnn	Casp8ap2	Mcm5	E2f8
Ttk	Dtl	Kif11	Mlf1ip	Cdc6	Tubb4b	Anp32e	Cdc25c
Rangap1	Dscc1	Cks1b	Top2a	Pold3	Kif23	Dlgap5	Nek2
Ccnb2	Cbx5	Blm	Hmgb2	Ckap2l	Exo1	Ect2	Cdc20
Cenpa	Usp1	Msh2	Ccne2	Fam64a	Rfc2	Nuf2	Rad51ap1
Cenpe	Hmmr	Gas2l3	G2e3	Ubr7	Pola1	Cdc45	
Cdca8	Wdr76	Tyms	Tmpo	Fen1	Mki67	Ckap5	
Ckap2	Ung	Hjurp	Nusap1	Bub1	Tpx2	Ctcf	
Rad51	Hn1	Hells	Ncapd2	Brip1	Aurka	Clspn	

Pcna	Cks2	Priml	Mcm2	Atad2	Anln	Cdca7	
Ube2c	Kif20b	Uhrfl	Kif2c	Psrcl	Chaflb	Cdca3	

ER Stress							
Nck2	Chac1	Creb3	Itpr1	Os9	Stt3b	Dnajb9	Crebrf
Ankzf1	Pdia3	Sec61b	Edem1	Ddit3	Rnf185	Tmx1	Bak1
Dnajb2	Bcl2l11	Erp44	Bbc3	Erlin2	Xbp1	Jkamp	Rnf5
Rhbdd1	Ddrgk1	AI314180	Psmc4	Ppp2cb	Erlec1	Sel1l	Atf6b
Bcl2	Tmx4	Jun	Bax	Ubxn8	Stc2	Psmc1	Bag6
Ubxn4	Trib3	Casp9	Ppp1r15a	Casp3	Trp53	Atxn3	Flot1
Yod1	H13	Fbxo6	Vimp	Pik3r2	Alox15	Derl1	Eif2ak2
Ppp1r15b	Edem2	Fbxo2	Rnf121	Amfr	Derl2	Rnf139	Pmaip1
Fam129a	Cebpb	Ube4b	Anks4b	Herpud1	Trim25	Foxred2	Tmx3
Edem3	Ptpn1	Ube2j2	Ern2	Aars	Cdk5rap3	Pla2g6	Syvn1
Atf6	Vapb	Psmc2	Atp2a1	Selk	Ccdc47	Atf4	Erlin1
Ufc1	Srpx	Tmub1	Brsk2	Ero1l	Psmc5	Ep300	
Atf3	Aifm1	Tmem129	Ins2	Psmc6	Ern1	Tmbim6	
Man1b1	Ubqln2	Wfs1	Ccnd1	Trim13	Nploc4	Txndc11	
Tor1a	Mbtps2	Ube2k	Map3k5	Dnajc3	P4hb	Sdf2l1	
Hspa5	Usp13	Tbl2	Nrbf2	Casp4	Txndc5	Ufd1l	
Dab2ip	Ufm1	Get4	Derl3	Casp12	Faf2	Eif2b5	
Nfe2l2	Serp1	Bhlha15	Ube2g2	Scamp5	Ubqln1	Nrros	
Dnajc10	Creb3l4	Creb3l2	Tmem259	Pml	Atg10	Pdia5	
Psmc3	Tmem67	Pdia4	Creb3l3	Parp16	Thbs4	Gsk3b	
Creb3l1	Ufl1	Eif2ak3	Hsp90b1	Nck1	Col4a3bp	Park2	
Thbs1	Ube2j1	Rnf103	Apaf1	Uba5	Pik3r1	Stub1	
Eif2ak4	Vcp	Aup1	Ifng	Usp19	Pdia6	Pdia2	

Epithelial Identity							
Cdhl	Cldn3	Cldn7	Ocln	Crb3	Krtl9	Dsp	Pkpl
Tgml	Cldn4	Cldnl 1	Epcam	Krt8	Pkp3		

ECM Rearrangement							
Sulf1	Creb3l1	B4galt1	Mia	Atxn1l	Adamts2	Tnfrsf11b	Cyp1b1
Col19a1	Hsd17b12	Reck	Spint2	Crispld2	Wnt3a	Col14a1	Fshr
Col3a1	Wt1	Tgfbr1	Aplp1	Foxf1	Mfap4	Has2	Mkx
Col5a2	Grem1	Col27a1	Hpn	Foxc2	Serpinf2	Ptk2	Lox
Fn1	Spint1	P3h1	Klk4	Agt	Vtn	Scx	Hpse2
Ihh	Cst3	Hspg2	Acan	Exoc8	Nf1	Fbln1	Kazald1
Col4a4	Fkbp1a	Vwa1	Serpinh1	Ero1l	Col1a1	Adamts20	Nfkb2
Col4a3	Mmp9	Dnajb6	Apbb1	Lgals3	Ramp2	Col2a1	
Serpinb5	Sulf2	Emilin1	Ilk	Ripk3	Gfap	Myh11	
Fmod	Atp7a	Mpv17	Ric8	Loxl2	Sox9	Ccdc80	
Elf3	Nox1	Apbb2	Muc5ac	Lcp1	Ero1lb	Abi3bp	
Lamc1	Col4a6	Pdgfra	Ctgf	Mmp13	Nid1	App	
Tnr	Prdx4	Ambn	Nr2e1	Mmp20	Foxf2	Serac1	
Dpt	Gpm6b	Dmp1	Nepn	Col5a3	Foxc1	Plg	
Ddr2	Egfl6	Ibsp	P4ha1	Smarca4	Ripk1	Smoc2	
Olfml2b	Postn	Tfip11	Spock2	Aplp2	Tfap2a	Has1	
Tgfb2	Rxfp1	Eln	Adamts14	Mpzl3	Ecm2	Noxo1	
Itga8	Sfrp2	Plod3	Mmp11	Thsd4	B4galt7	Col11a2	
Adamtsl2	Hapln2	Col1a2	Col18a1	Anxa2	Tgfbi	Tnxb	
Col5a1	Ctss	Ndnf	Myf5	Myo1e	Pxdn	Tnf	
Pomt1	Adamtsl4	Vhl	Col4a1	Nphp3	Smoc1	2300002M23Rik	
Eng	St7l	Mfap5	Csgalnact1	Dag1	Ltbp2	Flot1	
Lmx1b	Col11a1	Ercc2	Comp	Lamb2	Flrt2	Hsp90ab1	

Gsn	Npnt	Bcl3	Gfod2	Kif9	Fbln5	Washl	
01fml2a	Cyr61	Tgfbl	Has3	Sh3pxd2b	Eglam	Vit	

Apoptosis							
Ercc5	Procr	Slc35d1	Ldhb	Zfp365	Zbtb16	Sphk1	Abcc5
Serpinb5	Blcap	Plk3	Lrmp	Prmt2	Rps27l	Rhbdf2	Trp63
Inhbb	Ada	Rnf19b	Tm7sf3	Mknk2	Mapkapk3	Baiap2	Fam162a
Steap3	Fgf13	Sfn	Tgfb1	Dram1	Ip6k2	Dcxr	App
Btg2	Irak1	Fuca1	Sertad3	Apaf1	Tcn2	Hist1h1c	Rab40c
Phlda3	Tspyl2	Epha2	Cebpa	Btg1	Lif	Ninj1	Bak1
Tnni1	Sat1	Wrap73	Klk8	Mdm2	Upp1	Nol8	Def6
Rgs16	Zmat3	Mxd4	Bax	Ddit3	Ccng1	F2r	Cdkn1a
Ier5	Hspa4l	Rchy1	Ppp1r15a	Gls2	Cyfip2	Ankra2	Tap1
Slc19a2	Slc7a11	Iscu	Rpl18	Dgka	Gnb2l1	Plk2	Ier3
Adck3	Tm4sf1	Triap1	Aen	Cdkn2aip	Hint1	Sdc1	Polh
Ephx1	Rap2b	Prkab1	Rrp8	Hmox1	Gm2a	Gpx2	Ccnd3
Ptpn14	Fbxw7	Trafd1	Ccp110	Rrad	Hist3h2a	Zfp36l1	Hbegf
Atf3	S100a4	Pom121	Nupr1	Cdh13	Alox8	Fos	Hdac3
Notch1	S100a10	Pdgfa	Ptpre	Osgin1	Trp53	Ccnk	Rad9a
Rxra	Txnip	Gadd45a	Hras	Cgrrf1	Tax1bp3	Jag2	Ctsf
Ralgds	Nhlh2	Vamp8	Eps8l2	Abhd4	Traf4	Ndrg1	Slc3a2
Ak1	Dnttip2	Retsat	Ctsd	Kif13b	Cdk5r1	Pmm1	Fas
Stom	Clca2	Tprkb	Cd81	Rb1	Ppm1d	Plxnb2	
Ddb2	Wwp1	Tgfa	Perp	Nudt15	Rad51c	Vdr	
Cd82	Klf4	Mxd1	Rps12	Tsc22d1	Tob1	Csrnp2	
Il1a	Ikbkap	Sec61a1	Tpd52l1	Casp1	Krt17	Acvr1b	
Pcna	Cdkn2a	Xpc	Sesn1	St14	Hexim1	Sp1	
Bmp2	Cdkn2b	Ccnd2	Foxo3	Ei24	Fdxr	Abat	

Trib3	Jun	H2afj	Ddit4	Vwa5a	Itgb4	Socsl	

SASP							
Il6	Cxcl2	Csf2	Fgf7	Igfbp4	Mmp14	Icam3	Egfr
Il7	Cxcl3	Mif	Vegfa	Igfbp6	Timp2	Tnfrsf11b	Fn1
Il1a	Ccl8	Areg	Ang	Igfbp7	Serpine1	Tnfrsf1a	
Il1b	Ccl13	Ereg	Kitl	Mmp1	Serpinb2	Tnfrsf1b	
Il13	Ccl3	Nrg1	Cxcl12	Mmp3	Plat	Tnfrsf10b	
Il15	Ccl20	Egf	Pigf	Mmp10	Plau	Fas	
Cxcl15	Ccl16	Fgf2	Igfbp2	Mmp12	Ctsb	Plaur	
Cxcl1	Ccl26	Hgf	Igfbp3	Mmp13	Icam1	Il6st	

Neural Identity							
Vtn	Zeb2	Sox1	Pax6	Sox2	Msx1	Atoh1	Tubb3
Ednrb	Hes5	Neurod1	Cdh2	Id2	Msi1	Rbfox3	
Sox21	Fabp7	Pax3	Sox9	Hoxb1	Msi2	Map2	

Placental Identity							
4933433p14rik	Dusp9	Pkp2	Tnfrsf23	Serpinb9d	Krt18	1600014k23rik	Hapln3
Esx1	H19	9630050e16rik	Sos1	Plekhh1	Nrn1l	Tbrg1	Fam176a
Afap1	Tmem37	Pvrl2	Dlx3	2210011c24rik	Sfi1	Slit1	Pdlim1
Zfyve21	Mmp15	Zfp568	Ippk	Cd320	Tlr5	A730090h04rik	Ube2q2
Erv3	Fam101b	Vtcn1	Htr2b	Ccnjl	Rhou	4931406p16rik	Au018091
Atg12	Phf16	Il6ra	Dusp16	Entpd2	Arhgef6	Opn3	Bdkrb2
Las1l	4930422n03rik	Foxo4	Cdc73	Il1r2	Tmem185b	Pdia4	E130203b14rik
Rbp1	Ada	Hsp90b1	1700025g04rik	Sfmbt2	Tram2	B930054o08	S100g
Prl2b1	Mmp1a	Prl7c1	Prl4a1	1700011m02rik	Cited1	1700031f05rik	4933402e13rik
Prl3d1	Gpr126	Prl6a1	Zfp655	Plekha7	Cited2	Inhba	Dapk2

Rnf2	Arf2	Cdh5	Slcl3a4	Sfrp5	Zfand2a	Inhbb	Gml l985
Set	Tinagll	Fgd6	Ceacaml4	Ppplr3f	Krt25	Helz	Fndc3b
Mrgprg	Mfi2	Cysltr2	Ceacaml5	Obsll	Klk4	Sele	Twsgl
Aa763515	Rpn2	Rhox6	Trap l a	Slc23a3	Tnfrsfl lb	Pdia6	Aldhla3
Tfpi	Abhd2	Cdh3	Ceacaml2	Tmem87b	2010204kl3iHk	Pdia5	Lnx2
Etosl	Hrctl	Spp2	Gml65 15	Epasl	Torlaip2	Creb3	Taf7
Slc5a6	Adm	Ziml	Ceacaml3	Ccdc68	Fmrlnb	Efnal	Ai844869
1600025ml7Hk	Abhd6	Flnb	4930447f24rlk	Kdelr2	Ctsr	Dlg5	Clecl2b
Gm9	Slc7al	Rbbp7	Gzmd	Pramefl2	Ctsq	Procr	Prkcsh
Creb312	Tead4	Map3k7	Foxj2	Lrp8	Prl8a2	Fgfrl	Lama5
Bbx	Mbnl3	Rhox9	Fbxll9	Pard6b	Ctsm	Gnb4	Tchh
Prl3cl	Gprl	Whsclll	Gzmc	Peg 10	Prl8al	23 10030g06|k	Lamal
Mta3	2900057e 15rik	Slc38al	Gzmf	N4bp2	Ctsj	Gcml	Rps6ka6
Prl2al	Ldocl	1600012pl7|k	Gzme	Pla2g4e	Mpzll	Psgl8	Vhl
Gm9 112	Adaml9	Adra2b	Gzmg	Fam78b	Stra6	Goltlb	Eps812
Afap 112	Rybp	Pgf	Patl2	Arrdc3	Bcap3 1	Psgl9	Polg
Erlin2	Col4al	1200009i06rjk	3830417al3itk	Pla2g4d	Cregl	Psgl6	
Pard3	Fndc3cl	Mfsd7c	Tspanl4	Rassf8	Tcfap2c	Slc2al	
Aifll	Col4a2	Esam	Handl	Au015836	Prl7bl	Psgl7	
Dmrtcla	4930502e 18rik	Gprl07	AtxnlO	Csnkle	Ghrh	Htra3	
4932442108rik	Pkn2	AuO 15791	Mgat4a	Stagl	4930486124rlk	Klhll3	
Gjb2	Rlim	Arhgap8	Unc50	Vnnl	Neurog2	Ets2	
Gjb5	1600015il0r|k	Ankrdl7	I12rb	Tchhll	5430425j l2rlk	Nppc	
Slco5al	Afp	Cul7	Ceacaml 1	Plala	Prl7al	Tgml	
Wdr61	Tmeml40	23 10067p03|k	Plekhgl	Slc45a4	Prl7a2	Tmeml08	
Kitl	Fstl3	Irs3	Prl3bl	Tex264	Mirl l99	Usp53	
9430027b09i)ik	Ing4	Prl5al	Folrl	Pcdhl2	TbcldlOa	Mark3	
Tfrc	Taf71	Fntb	A830080d0 lHk	Ctr9	Ralbpl	Cbx8	

Slc6a2	Sultlel	Tceanc	Blzfl	Ccrlll	Pdgfra	Hspa5	
Wdr45	Olrl	Lepr	Zfp667	Htatsfl	Morc4	Spats2	
Zxda	2610019f03rlk	Tnfrsf9	Fltl	9030409gl 1rik	Rarres2	Limk2	
Prdx4	F1l	Papola	Usp27x	Tspan9	Arid3a	Mkl2	
Faml22b	Fbxw8	Srd5al	Hdac4	Rassf6	Lifr	Shroom4	
Zxdb	Sema4c	Clqtnfl	Itgb3	463 1402f24rlk	Shisa3	Shrooml	
Zxdc	Ctnnbipl	Slc38a4	Sri	A2m	Uevld	Pou2f3	
Pip5kla	Tfpi2	Angpt4	Sema3f	Rimklb	Scnnlb	Acvr2b	
Placl	ZbtblO	Ctla2a	Prl3al	Locl005045$9	Dnajbl2	Rbms2	
Igf2as	Mitf	9930012kl 1rik	Bahdl	Apob	Brwd3	Atg4b	
Usp9x	Gpr50	Mical3	Sin3b	Tmeml50a	Hhipll	Pappa2	
Psg28	Hic2	Apoa4	Gm2a	9130404d08k	Fbln7	Rbm25	
Bmp8b	Tpbpb	Cul4b	Serpinb9g	Prl8a6	Maspl	Gm4793	
Fnl	Slc9a6	3632454122rik	Bend4	Cts6	Nrk	Nidi	
Psg23	Prl7dl	Psg-psl	Bend5	Prl8a8	Pvr	Uba6	
Bmp8a	Tpbpa	Lcor	Serpinb9b	Prl8a9	Atp2cl	Lamcl	
Psg21	Slco2al	Tnfrsf22	Serpinb9c	Cts3	Amot	Slc40al	

X reactivation							
Gm21950	Slc9a7	Rhox3h	Slitrk4	Fam47c	Zdhhc15	Bhlhb9	Samt1
Gm21364	Rp2	Rhox2h	Ctag2	Gm7173	1700121L16Rik	Gprasp2	4921511M17Rik
Gm14346	Jade3	Rhox5	4930447F04Rik	Mageb16	Magee2	Arxes2	Gm10057
Gm14345	Rgn	Rhox6	Slitrk2	Gm26775	Pbdc1	Arxes1	Gm15140
Gm14351	Ndufb11	Rhox7a	1700036O09Rik	Tmem47	Magee1	Bex2	4930524N10Rik
Gm3701	Rbm10	Rhox8	Gm1140	4930595M18Rik	5330434G04Rik	Nxf3	Samt4
Gm3706	Uba1	Rhox7b	Gm14692	Dmd	Cypt2	Bex4	Samt2
Gm14347	Cdk16	Rhox9	4933436I	Tsga8	Fgf16	Tceal8	Cldn34b1
Gm10921	Usp11	Btg1-ps1				Tceal5	

Gml0922 Gm3750 Gm3763 Mycs Gml4374 Nudtl l AU022751 NudtlO Bmp15 Shroom4 Dgkk Ccnb3 Akap4 Clcn5 Usp27x Ppplr3f Ppplr3fos Foxp3 Ccdc22 Cacnalf Syp Gml4703 Prickle3 Plp2 Magix Gpkow Wdr45 RP23-	Araf Synl Timpl Cfp Elkl Uxt Zfpl82 Spaca5 Zfp300 Ssxal Gm21876 4930453H23Rik Gm6938 Gm26593 Agtr2 Slc6al4 Gm28269 Gm28268 K1M13 Wdr44 Gm4907 Gm4985 Gm27192 Gm5934 Gm4297 Gm5935 Gm5 169	Btgl-ps2 RhoxlO Rhoxl l Rhoxl2 Rhoxl3 Zbtb33 Tmem255a Atplb4 Lamp2 Gm7598 Cul4b Mctsl Clgaltlcl Gml4565 6030498E09Rik Cyptl5 Cyptl4 Gria3 Thoc2 Xiap Stag2 Gm43337 Sh2dla Tenml Gm362 Dcafl212 Dcafl211	OlRik Fmrlos Fmrl Fmrlnb Gml4698 Gm6812 Gml4705 Aff2 17001 11N16Rik 1700020N15Rik Ids 1110012L19Rik 4930567H17Rik BC023829 Mamldl Mtml Mtmrl Cd9912 Gml6189 Hmgb3 Gpr50 Vma21 Gml l41 Prrg3 Fatel Cnga2	Fthll7a Tab3 Gk Gml4764 Gml4762 5430427019Rik Samt3 NrObl Mageb4 Illrapll Gm27000 Pet2 4932429P05Rik 4930415L06Rik Gm44 Gml4773 Mageb2 Gm5072 Gm8914 1700084M14Rik Gml4781 Mageb5 Magebl Magebl8 Gm5941 1700003E24Rik	Atrx Magtl Cox7b Atp7a Tlrl3 Pgkl Taf9b Fnd3c2 Fndc3cl Cysltrl Gm5 127 Zcchc5 Lpar4 P2ryl0 A630033H20Rik Gprl74 Itm2a Tbx22 2610002M06Rik Fam46d Gm732 Gm379 Brwd3 Hmgn5 Sh3bgrl Gm6377 RP23-	Bexl Tceal7 Wbp5 Ngfrapl Kir3dl2 Kir3dll Tceal3 Tceall Morf412 Glra4 Plpl Rab9b H2bfm Tmsbl51 Tmsbl5b2 Tmsbl5bl Slc25a53 Zcchcl8 Faml99x Esxl Illrapl2 Texl3a Nrk Serpina7 4930513006Rik 4933428M09Rik Mumlll	Magea6 Magea3 Magea8 Magea2 Magea5 Mageal Cldn34b2 Satl Acot9 Prdx4 Ptchdl Gml5 156 Gml5 155 Phex Sms Mbtps2 Yy2 Smpx Gml5 169 K1M34 Cnksr2 Rps6ka3 Eiflax Map7d2 A830080DOlRik Sh3kbpl Map3kl5

109E24. 10 Praf2 Cede 120 Tfe3 Gṙ pap1 Kcndl Otud5 Pim2 Slc35a2 Pqbpl Timml7b Gml0491 Gml0490 Pcskln Eras Hdac6 Gatal Glod5 Gml4820 Suv39hl Was Wdrl3 Rbm3 Rbm3os Tbcld25 Ebp Porcn	Gml993 E330010L02Rik Gm5 168 Gm2012 Gm2030 Six Gml4525 Gm6121 Gml0230 Gm2101 Gml0058 Gm21 17 Gm4836 Gml0147 Gm2165 Gml0096 Gm2200 Gm26818 Gm3669 Gml0488 E330016L19Rik Gml4632 Gm7437 Gml4974 Gml0487 Gm21447 Spin2f	Prr32 4930515L19Rik Actrtl Gm29242 Smarcal Ocrl Apln Xpnpep2 Sash3 Zdhhc9 Utpl4a 9530027J09Rik Bcorll Elf4 Aifml Rab33a Zfp280c Slc25al4 Gprl 19 Rbmx2 Gm595 Enox2 Gml4696 Gml4697 Arhgap36 Olfrl320 01frl321	Magea4 Gabre MagealO Gabra3 Gabrq Cetn2 Nsdhl Gml4684 Zfpl85 Pnma5 Pnma3 Xlr4a Xlr3a Xlr5a Gml4685 DXBayl8 Xlr5b Spin2d Xlr3b Xlr4b F8a Xlr4c Xlr3c Xlr5c RP23-95K12. 13 Zfp275 Gml8336	BC061 195 Arx Polal Pcytlb Pdk3 AU015836 Gml4798 Zfx Eif2s3x K1M15 Fam90alb Apoo Gml4827 Magedl Gspt2 Zxdb RP23-9K14.6 Gm26617 Spin4 Arhgef9 Amerl Asbl2 Zc4h2 Zc3hl2b 1700010DOlRik Las 11	240M8.2 Pou3f4 Cylcl Gml01 12 Rps6ka6 Hdx RP23-466J17.3 Tex 16 4933403008Rik Apool Satll 2010106ElORik Zfp71 1 Poflb Gml4936 Chm Dach2 K1M4 Ube2dnll Ube2dnl2 4930555B12Rik Cpxcrl H2afb2 Gml4920 Gm28579 Tgif21x2	Trap la D330045A20Rik Rnfl28 Tbcld8b Gml5013 Ripply 1 Cldn2 Morc4 Rbm41 Nup62cl Pihlh3b Gml5046 Frmpd3 Prpsl Tsc22d3 Mid2 Eif2c5 Tex 13 Vsigl PsmdlO Atg4a Col4a6 Col4a5 Irs4 Gml5295 Gml5294 Gml5298	Pdhal Adgrg2 Gml5241 Phka2 Gml5243 Ppefl Rsl Cdkl5 Gja6 Scml2 Gml5262 Rai2 Scmll Gml5205 Nhs Gml5202 Reps2 Rbbp7 Txlng Syapl Ctps2 SlOOg Grpr Rnfl38rtl Apls2 Zrsr2 Car5b Siahlb

Ftsj l Slc38a5 SsxblO Ssxb9 Ssxbl Ssxb2 Gml4459 Ssxb6 Ssxb3 Ssxb8 Ssx9 Ssxb5 Gm6592 Gm575 1 B630019 K06Rik Fthll7b Fthll7c Fthll7d Fthll7e Fthll7f 4930402K 13Rik Lancl3 Gml4862 Xk 1700012L 04Rik Gml4501	Gm2784 Gm2777 Gm21883 Spin2e Gm21608 Gm21637 Gm21645 Gm2799 Gmclll Gm5926 Gm2195 1 Gm21657 Gm21789 Gm2825 Spin2-ps6 Gm2863 Gm2854 Gm2913 Gm2927 Gm2933 Gm2964 Gm21870 Gm21681 Spin2g Gm21699 Gml4552 Gml0486 Gm2309	Igsfl 01frl322 01frl323 01frl324 Stk26 Frmd7 Rap2c Mbnl3 Hs6st2 Usp26 17000800 16Rik Gpc4 Gpc3 Gml4582 A630012P 03Rik Cede 160 Phf6 Hprt Gm28730 Placl Faml22b Faml22c Mospdl Etd Gml4597 Cxxlc Cxxla	Gm26726 Zfp92 Trex2 Haus7 Bgn Atp2b3 Dusp9 Pnck Slc6a8 Bcap3 1 Abcdl Plxnb3 Sϕ k3 Idh3g Ssr4 Pdzd4 L Icam Arhgap4 Avpr2 NaalO Renbp Hcfcl Iraki Mecp2 Opnlmw Tex28 Tktll Flna	Msn F630028O lORik Vsig4 Hsf3 Heph Gprl65 Pgrl51 Eda2r Ar Ophnl Yipf6 Stard8 Efnbl Gml4812 Gml4809 Gml4808 Pjal Tmem28 Eda Awat2 Otud6a Igbpl Dgat216 Awatl P2ry4 Arr3 Pdzdl l	Tgif21xl Gml4929 Pabpc5 Pcdhl 1x H2afb3 Napll3 Gml7521 Cldn34cl Astx6 Srsx Gml7577 Gml495 1 Astx2 Gml7412 Cldn34c2 Gml4950 Gml7467 Cldn34c3 Astx5 Vmn2rl2 1 Astxla Gml7584 Astx4a Gml7469 Astx4b Astxlb Gml7361	Gucy2f Nxt2 Kcnell Acsl4 Tmeml64 Ammecrl Rgagl Chrdll Pak3 Capn6 Dcx A730046J 19Rik Algl3 Tϕς 5 Tϕ c5os Zcchcl6 Lhfpll Amot Htr2c I113ra2 Lrch2 Gml5 128 Gml5080 Gml5 107 Gml5 114 Gm8334 Gml5 127	Tmem27 Ace2 Bmx Pir Figf Piga Asbl l Asb9 Mospd2 Fancb Gml7604 Glra2 Gemin8 Gpm6b Ofdl Trappc2 Rab9 Tceanc Egfl6 Gml5226 Gml720 Gml5230 Gm8817 Gml5232 Gml5228 Tmsb4x Tlr8 Tlr7

Cybb Gm5 132 Dynlt3 Hypm 4930557A 04Rik Sytl5 Srpx Rpgr Otc Tspan7 Gml0489 Midlipl Gml4493 Gml4483 Gml4474 Gml4477 Gml4476 Gml4484 Gml4479 Gml4482 Gml4478 Gml4475 Gm4906 Bcor Gml4635 Atp6ap2 18100300 07Rik	Gml4553 Gml4819 Dockl l I113ral Zcchcl2 Lonrf3 Gm6268 Gml4569 Pgrmcl Akapl7b Slc25a43 Slc25a5 Gml4549 23 10010 G23Rik C330007 P06Rik Ube2a Nkrf Gml5008 43349 Sowahd Rpl39 Upf3b Nkap Akapl4 Ndufal Rnfl l3al Gm9	Cxxlb 4930502E 18Rik 1700013H 16Rik Zfp3613 Xlr Gml6405 Gml6430 Slxll 3830403N 18Rik Gm773 1600025 M17Rik Zfp449 Gm2155 Smiml012 a Gm2174 Ddx26b Gml0477 Gm648 Mmgtl Slc9a6 Fhll Mtap7d3 Adgrg4 Brs3 Htatsfl Vglll	Emd RpllO Dnaselll Taz Atp6apl Gdil Fam50a Plxna3 Lage3 Ubl4a Slcl0a3 Fam3a Ikbkg G6pdx Gm6880 01M326- psl 01frl325 Gm5640 Gm6890 Gm5936 Gab3 Dkcl Mppl Smim9 F8 Fundc2 Cmc4	Kif4 Gdpd2 Gml4902 Dlg3 Tex 11 Slc7a3 Snxl2 Foxo4 Gm614 Gm20489 I12rg Medl2 Nlgn3 Gjbl Zmym3 Nono Itgblbp2 Tafl Ogt Cxcr3 Gm4779 8030474K 03Rik Nhsl2 Rgag4 Pin4 Ercc61 Rps4x	Gm21616 Astx4c Gml7693 Astxlc Gml7522 Astx4d Gml7267 Astx3 493241 IN 23Rik Gm382 492151 1C 20Rik Cldn34c4 4930558G 05Rik Diaph2 Pcdhl9 Gm2685 1 Tnmd Tspan6 Srpx2 Sytl4 Cstf2 Noxl Xkrx Arll3a Trmt2b Tmem35	Luzp4 Gml5099 Ott Gml5092 Gml5093 Gml5 100 Gml5085 Gml5086 Gml0439 Gml5097 Gml5091 Gml5 104 Tmem29 Apex2 Alas2 Pfkfbl Tro Maged2 Gm27191 Gnl31 Fgdl Tsr2 Gml5 138 Wnk3 A230072 ElORik Faml20c Phf8	Prps2 Gml5239 Frmpd4 Msl3 Arhgap6 Gml5261 Amelx Hccs Gml5245 Midi 4933400A I1Rik Gml5726 Gml5247 Gm21887 Asmt

Medl4 Usp9x 2010308F 09Rik Ddx3x Nyx Cask Gpr34 Gpr82 Gm5382 Gml4505 Drrl Cyptl Maoa Maob Ndp Efhc2 Fundcl Dusp21 Kdm6a 4930578C 19Rik Gm26652 BC04970 2 Chst7	Rhoxl Rhox2a Rhox3a Rhox4a Rhox3a2 Rhox4a2 Rhox2b Rhox4b Rhox2c Rhox3c Rhox4c Rhox2d Rhox4d Rhox2e Rhox3e Rhox4e Rhox2f Rhox3f Rhox4f Rhox3g Rhox2g Rhox4g	Gml4718 Cd401g Arhgef6 Rbmx Gm364 GprlOl Zic3 4930550L 24Rik Fgfl3 F9 Mcf2 Atpl 1c Gm7073 Gml4661 Sox3 Gml4662 Gml4664 Cdrl Ldocl 4933402E 13Rik 493 14000 07Rik 1700019B 21Rik Gm6760 3830417A 13Rik	Mtcpl Brcc3 Vbpl Gml5384 Rab39b Gml5063 Pls3 Gml4715 Gml4707 Gml4717 Cldn34b3 Cldn34b4 Cldn34d Tbllx Prkx Gml4742 Pbsn Gml4744 5430402E lORik Obpla Gm5938 Obplb Gml4743 4930480E 11Rik Prrgl	Citedl Hdac8 Phkal Gm91 12 Dmrtclb Dmrtclcl Dmrtclc2 170003 IF 05Rik Dmrtcla 170001 1 M02Rik Napll2 Cdx4 Chicl Gm26952 Tsx Gm26992 Tsix Xist Jpx Ftx Zcchcl3 Slcl6a2 Rlim C77370 Abcb7 Uprt	Cenpi Drp2 Taf71 Timm8al Btk Rpl36a Gla Hnrnph2 Armcx4 Armcxl Armcx6 Armcx3 Armcx2 Nxf2 Zmatl Gml5023 Tceal6 Pramel3 Gm5 128 Gm7903 AV32080 1 Nxf7 Prame Tcpl lx2 Tmsbl5a Armcx5 Gpraspl	Huwel Hsdl7bl0 Ribcl Smcla Iqsec2 Kdm5c Kantr Tspyl2 Gprl73 Cldn34a Shroom2 Gprl43 Usp5 1 Magehl Foxr2 Rragb Klf8 Ubqln2 Cypt3 Kctdl2b RP23- 106P7.5 22100130 21Rik Spin2c	

XEN							
Dab2	Pdgfra	Gata6	Fxyd3	Soxl7	Lamal	Gata4	Krt8
Fst	Pthlr	Foxql	Tet3	Foxa2	Lambl		

Trophoblast							
Ascl2	Cdx2	Esrrb	Grn	Lipg	Smad3	Tfap2c	Gata3
Bmp4	Elf5	Ets2	Igf2	Pcsk6	Snai1	Vav1	Krt7
Bmp8b	Eomes	Fgfr2	Jade1	Ptpra	Tead4	Yap1	Krt18

Trophoblast progenitors							
Rhox6 Rhox9 3830417A13Rik Gjb3 Gm9112 Hspb1 Nup62cl Ldoc1 Hspe1 Rhox12 Tex19.1 Gjb5 Sin3b 1700086L19Rik Ldhb Krt19 Hmgn5	Hmgn2 Odc1 Klhl13 Ncl Tyms Prss8 Atp5g3 Dusp9 Gmnn Rrm2 Tbrg1 Cct3 Nhp2 Ppid Ccna2 Anp32b Cacybp Chchd2	Tuba1b Cenpw Cct7 Sfn Fkbp4 Ndufb6 Snrpe Cenph Rad51 Set Cd164 Cox6b1 Hnrnpdl Lsm2 Exoc3l4 Dut Pramef12	Immt Smagp Hnrnpa2b1 Cox7b Snx10 Stip1 Rnf4 Gm648 Cct6a Snrpd2 Psmg2 Tk1 Rps5 Mtx2 Phb Hspa8 mt-Nd5	Rps21 Pdlim2 Rpl24 Asf1a Eif4a3 Ssb Timm17a Mrpl18 Cenpk Dcakd Hikeshi U2af1 Acp1 Tipin Fkbp3 Cdca3 Tubb4b Mycbp	Ccnd3 Rpl5 Nip7 Psma5 Spc24 Mdh2 Cep164 Cs Zc3h15 Pea15a Tsen15 Ippk Thoc3 Pithd1 Pak1ip1 1110038B12Rik Wdr18	Mrpl54 Rps26 Ndufb9 Arpc1a Rps28 Prpf31 Mrpl12 Epop Cct5 Pdap1 Ezh2 Gpbp1 Psme3 Ube2c Cbx1 Gata2 Nxf7 Smc4	Ruvbl2 Ndufv1 Polr2l Asns Prkrip1 1700021F05Rik Aimp1 Rps7 Tra2b Cox17 Mrpl19 Chchd4 Polr1d Ubfd1 2410015M20Rik Tbcb Chchd1

Trap1a
Plac1
Cdkn1c
Bex1
Fthl17a
Dbi
Ube2a
Dnaja1
Phactr1
Phlda2
Hand1
Selenoh
Rhox5
Atp5g1
Hmgn1
Hat1
Plet1
Gm9
Rbbp7
Hspd1
Mrfap1
Krt7
Esam
Krt8
Fstl3
Ghrh
Ranbp1
Npm1

Phb2
Snrpf
Ran
Gale
mt-Nd4
Birc5
Tpm2
Hsd17b4
Rpl22l1
Snrpd1
Hspa14
Wfdc2
Rfc4
Rgcc
Mfsd2a
Cct8
Ubxn1
Ddt
Dtymk
C430049B03Rik
Magoh
Calm2
Mrps22
Impdh2
Brd3
Fscn1
2610528J1IRik

Cd320
Snrpd3
Psmb7
Mcm7
Taf1d
H2afz
Ndufb2
Lyar
Rbms2
Eif5b
Rbm8a
Dynll1
Stmn1
Got2
Cox7c
Lsm6
Ccne2
Sap18
Liph
Pa2g4
Slc38a4
Irx3
Srsf3
Dpy30
Hmgcl
Cenpa
Mgll
Eef1g

Orc6
Dctpp1
Sugt1
Wdr77
Suclg1
Ddx39
Polr2f
Rpl38
Rpa2
Fmr1nb
Gng12
Tuba1c
Aasdhppt
Pfdn6
Hspa9
Eif1a
Pop5
Nasp
Xlr4b
Snrpb2
Nop58
Uqcrc2
Cfdp1
Hn1l
Tsn
Psma6
Ssrp1

Apip
Mdk
Rpl14
Cox7a2
Hnrnpc
Sdr39u1
Slc25a3
Psma7
Psmd12
Cyc1
Apex1
Rad23b
C1qbp
Cox6c
Txn1
Med19
Slirp
G3bp1
Ak2
Krt18
Rsl1d1
Csrp1
1600025M17Rik
Rpp30
Mrpl38
Emg1
Cebpzos

Nol7
Tomm70a
Snu13
Psma2
Eif2s2
Usmg5
Eif3e
Cops5
Mrpl3
Mybbp1a
Elp2
1110004F10Rik
St13
Tbca
Snrpa1
H2afv
Mcm5
Tcp1
Atp1b1
Aprt
Nup37
Hebp1
Lsm8
Mbd3
Gtf3c6
Rpa3
Cdc34

Tfap2c
Creb3
Clns1a
1810022K09Rik
Eif2b1
Idh3a
Sae1
Eif5a
Fhl2
Lap3
Ncbp2
Eps8l2
Cdk4
Rfc3
Cdk1
Mrps25
Coq3
Med10
Emd
Ptrh2
Mrps18c
Med4
Fam133b
Crip2
Ndufaf3
Thap4
Mrps16

Serbp1
Hsph1
Xpo1
2310033P09Rik
Prpf19
Apoo
Hagh
Ndufa9
Mrpl2
Ndufb7
Psmb1
Txndc9
Hnrnpa1
Ndufs7
Farsb
Cycs
Tmem11
Rps17
Mrpl14
Diablo
Cox4i1
Pkp2
Psmc2
Psmc1
Slc25a4
Eloc
Vma21

H19	Zwint	Atp5cl	Acaala	Nsmce4a	Ndufb8	Uchl3	Mif
Sdcl	Tmem37	Imp4	Rpf2	Cct2	Naplll	Meal	Timml3
Rps41	Ndufa5	Cks2	Lgalsl	Rpsl6-ps2	Adgrf5	Psma3	
mt-Ndl	Eif2sl	Rnd2	Psmd6	Ruvbll	Ptges3	TimmlO	
Hsp90aal	Hsdl7b2	Knstrn	Aplm2	Arppl9	Polr2j	Rrml	
Mbnl3	Galkl	Atp5fl	Plppl	Rpl27	Ndufal2	Hnmpd	
Htatsfl	Cct4	Skpla	Ndufaf2	Dcunld5	Cyb5b	Tomm22	
Hsp90abl	Cox5a	Igf2bpl	Cull	Rpll8	Tmod3	Ndufabl	
Las 11	Dkkll	Mrpl21	Ndufal 1	Mrpll5	Ndufv2	Aifml	
Ptma	Hmgb2	Srsf7	mt-Col	Psmal	Ash21	Tfam	
mt-Cytb	Tubb5	Psipl	Tomm4 0	Baspl	Spc25	Rrpl5	
Snrpg	Med21	Llph	Ndufs8	Tead2	Dnajc2	Rps2	
Fdxl	Nmel	Erdrl	Derl3	Prmtl	4921524J 17Rik	Tinf2	
Glrx5	Cdca8	Atp5k	mt-Nd2	Esfl	Gins4	Lypla2	
Alpl	Tsen34	Rmdn3	Ckslb	Banfl	Naa38	Ppmlg	
Elf3	Oaf	Peg 10	Eif3g	Pinl	Pole3	Dars	
Ndufa4	Ccnbl	Ccnel	Nop 16	Mta3	Nucb2	Ingl	
Dynll2	Ascl2	Rps271	Itpa	Priml	Tomm7	Psmb2	
Hsp25-psl	Lsm4	Ezr	Mat2a	Ppih	Erh	Fcfl	
	Ahsal	Psmd7	Gnl3	Eif3i	Rps8	Rpl30	
			Pdcd5		Samm50		

Spiral Artery Trophpblast Giant Cells							
Car2	Psg22	Rgsl7	Psipl	Eif31	Got2	Rpsl8	Cct6a
Set	K1M13	Mpzl2	Tnfaip8	Fscnl	Hnmpa2b 1	Actr3	Nectin 2
1500009L16Rik	Ldocl	Liph	Trap 1a	Ehdl	Prl7dl	Anxa7	Grhpr
Serpinb9e	Galkl	Ddbl	Tubalc	Pramefl2	1110008P14 Rik	Cfll	Cct7

Prl2al	Aφς lb	Irs3	Cd82	Eiflb	Rackl	Gtf2e2	Chordc l
S100a6	Anxa4	Bexl	Gjb5	Mxd4	Rps7	Parva	Vma21
Plac8	Cdx2	Lysmd2	Seφ íηe2	Rap l a	Pdcd5	Eeflg	Rpl39
Serpinb9g	Tpm4	Rpl2211	Tubala	Borcs7	Cct4	Cct2	Ccnbl
Prl6al	Anxa2	Rhox5	Txnl	Torlaip2	Mif	Rpl9	Gm2000
Lgals9	Seφ inb9b	23 10030G06:llik	Ralbpl	Krtl9	Csφ ï	0610007P14Iik	Snrpf
Prl7bl	Derl3	Pdlim2	C430049B03I¾k	Avpil	Cox5a	Nmrkl	Aamp
Ada	Tfap2c	Nostrin	H2afz	Actgl	Rpl27	Eny2	Smarcb l
Aldhla3	Baspl	Glrx5	Pdcd4	Cdkn2aipnl	Npml	Epop	Prelid1
Serpinb6b	Rbbp7	Tpml	Jup	Bex3	Ppdpf	Ran	Paklip1
Sri	Caldl	Cnn2	Morf412	Dnajc8	Ets2	Krtl8	Hmbs
Fstl3	Laspl	Grb2	Pfnl	Ubfdl	Nrk	Kat7	Polr2j
Serpinb9d	Hmgn5	Fbliml	Actnl	Cfap20	Gga2	Exosc8	Calm3
Prl2c5	Spata21	Uppl	Aifll	Zwint	Krt7	Rpl23a	Ezr
H19	Tbrgl	Ppplrl4b	Cdh5	Rps4x	Ranbpl	Rps8	Rps3a1
Aprt	Dusp9	Cdknlc	Eif4ebpl	Mycbp	Rps41	Rps3	Elovl5
Seφ íub9c	TmsblO	Tfpi	Erccl	Ndufaf3	Ywhab	Rrm2	Rpsl7
Ascl2	Dynll2	Fermt2	Mvp	As3mt	Fkbpla	Dtymk	Rps5
Placl	Ctnnbipl	Palm	Ndufal l	Hatl	Pdcl3	RpllOa	
Mt2	Sin3b	Tubb5	Ugp2	Rps20	Rpsl6	Actr2	
Fthll7a	Igfbp7	SlOOal l	Prmt5	Myl6	Gnai3	Olal	
Tφ 53i11	Mpzll	Krt8	1700086L19Ilik	Pygl	Eif4e3	Cklf	
Mrfapl	Olrl	Zyx	1600025M17I¾l	Rpp21	Rpll2	Cfdpl	
Phactrl	Mbnl3	Alad	Aφς 2	K1M22	Tipin	RpslO	
Tnfrsf9	Myll2a	Faml62a	Abracl	Cetn3	Aφς 5	Rpl36a	
Lgalsl	Nek6	AA467197	Vasp	I12rg	Eif2sl	Rpsl9	
Pitrml	Sbsn	Rps271	Gngl2	Pletl	Chpl	Snφ g	
Ncmap	Copz2	Ncaml	Sqstml	Gm91 12	Cepl64	Clqtnf6	

	Eif2s2	Dcakd	Tpm2	Eifla	Rpsa	Atpifl	

Spongiotrophoblasts							
Phlda2	Cs	Pttg1	Cops5	Lsm8	Impa2	Drg1	Mrto4
Dio3	Lgalsl	Trappc5	Psmd12	Gadd45g	2010107E04Rik	Nae1	Rnf128
Dkkl1	Hagh	Eif3g	Panx1	Med7	Ndufb5	Hspa8	Wdr77
Hspb1	Npm1	Gpx4	Dld	2310033P09Rik	0610007P14Rik	Dars	Pepd
Tmem14c	Tex30	Gtf2h5	Ppid	Atp11a	Gtf3c6	Ubald2	Ddx18
Cidea	Mfge8	Magoh	Dnajc2	Skp1a	Dnajc19	Hnrnpk	Lrrfip2
Tfrc	Usp1	Fam50a	Hspd1	Eloc	Atp5k	Idh3a	Psmb7
Batf3	B3gnt7	Cct3	Hmgb2	Nsmce2	Tubb2a	Plekhf2	Erdr1
Sin3b	Mageh1	Srsf3	Uaca	Slc25a3	Slirp	Vps35	Rps28
Prss8	mt-Nd4	Rfc4	Wwtr1	Gadd45b	Phb2	Mrpl47	Fnta
Ldoc1	Emc8	Eif1a	Psmd6	Cfdp1	Psmc1	Birc5	Rtn3
Maoa	mt-Nd5	Marcksl1	Hnrnpc	H2afz	Folr1	Unc50	Idh3b
Cdkn1c	Commd4	Serpinb9e	Mrps23	Ppa1	Bax	Dut	Elob
Lasl1	Dnaja2	Apoo	Nap1l1	Atp5b	Rmdn3	Cdc34	Pfdn6
Rhox6	Tbca	Slc2a1	Tead2	Polr2e	G3bp1	Nabp1	Sugt1
Tex19.1	Ndufb2	Vdac3	Cd164	Clns1a	Trim27	Hadhb	Dstn
2610528J11Rik	Tubb4b	Cox5a	Pparg	Dnajb6	St13	Aimp1	Smarcb1
Gkap1	Sct	Ppp1r3g	Rpl22l1	Rnf181	Slc38a2	Fus	Coq3
Cldn7	Ing2	Cct5	Rhox5	Rnf4	Dusp9	Etfb	Igsf8
Slc22a18	Cd320	Anxa4	Psmd7	Hdac1	Cggbp1	Hnrnpab	Tomm22
Rhox9	Hsd11b2	Nsmce4a	Ndufa4	Prpf19	Ptma	Ndufb4	Hmbs
Mrps6	Vamp8	C430049B03Rik	Ndufb6	Nsmce1	Chchd1	Exosc8	Cyc1
Serpinb9g	Tbrg1	Tmem147	Tma7	Gm11361	Rpl18	Rplp1	Txnl1
Aqp3	mt-Nd2	Pa2g4	Med21	mt-Rnr1	Psmc6	Cox7b	Fam104a
	Gm9		Cox6b1			Mrpl19	Hn1

mt-Cytb	Slc38al	Tyms	Tardbp	Ncbpl	Atp5cl	Nsfllc	Ctnnal
Hsp25-psl	Rbbp7	Eif4al	Uqcrc2	Blvra	Eroll	Timml7 g	Ndufs8
Rdh12	AtxnlO	Snrpe	Psma6	Pϕ sap 1	Hspa9	Pigp	Bsg
Krtl8	Hsp90aal	Smul	Larp7	Ube2el	Anapcl5	Ndufsl	Gskip
Pfdnl	Calml	Tbcb	Ranbpl	S100al6	Rps8	Appbp2	Cnihl
Tulpl	Hspel	Baspl	Mϕ 14	Serbpl	Seϕ inb9d	Zwint	Rbm8a
Selenoh	Faml36a	Fam90alb	Suclgl	RablO	Cotll	Duspl 1	Gm2a
Dynll2	Elf3	Nup85	Pgrmcl	Rala	Ash21	Mcm2	Eif3e
Glrx5	Prkd2	Lonp2	Mdh2	Psmdl3	Arl6ipl	Set	Erh
Slcl6al	mt-Col	Mϕ s22	Rpl5	Pmpca	Borcs7	Scarb2	Naa35
Krt8	Ncl	Lyar	Ndufa5	Serpinb9b	Psmc2	Smc4	Mϕ 13
Tmeml50 a	Hadh	Fermt2	Gucdl	Ppa2	Zcchcl7	Ywhaq	Mapllc3 b
Stx3	Cisdl	Srsf6	Car2	Hebpl	Ncbp2	Cdca8	Tcpl
Gjb2	Snrpg	Nxf7	Dnajc9	Mϕ 115	Psmbl	Hmgcl	SrsflO
Nudt22	Syngrl	Rad23b	Wdrl8	Rrm2	Priml	Tra2a	Psma3
Mbnl3	Chchd2	Fkbp3	Cox7c	Ccnbl	Thoc3	Npepll	Ndcl
Gm9 112	Ubqlnl	Atp5o	Ssb	Gprl37b	Nop58	Med28	Mtch2
Cd9	Fbxll9	Cct8	Ran	Idh3g	Polrld	H2afv	Psmdl 1
Rbpl	Pphlnl	Snx5	Emd	Srsf7	Sap 18	Sdhb	Rpl27
Rps41	Slc25a5	Clqbp	Hsp90ab 1	Slc25a4	Gmfb	Uqcrcl	E2f5
Eif2s2	Ccdc5 1	Bglap3	Hnmpal	Gata2	Lsm4	Nsϕ ĩ	Pitpnb
Ugp2	Mpdul	Atp5fl	Atp5al	Nhp2	Rps5	Snipf	
Zfp655	Eif2sl	ChchdlO	Psmg2	Rars	Cdipt	Snϕ d2	
mt-Ndl	Hspal4	Olrl	Pdcd5	Snx6	Uspl4	Rabif	
Tdϕ	Prkcz	Cenph	Cacybp	Dpy30	Psme3	Commd5	
Urod	Tafld	Uchl3	Lsr	Ube2c	Lamtorl	Smiml l	
Hmgn5	Mipll6	Cenpk	Ttc4	Ahsal	Cycs	Cox4il	
	170002 1F0	Paklipl		Peg 10	Ndufb8		

Car4	5Rik	Gml5536	Cox7a2	Eif3i	Imp4	Cetn3	
Krtl9	Rap2c	Naa38	Lsm6	Mϕ 155	Mϕ s25	Ruvbl2	
Rassf6	Acvr2b	Tϕ ïï	Stmnl	Rfc5	Nop 16	Strap	
Tfeb	Irx3	Psmc5	Ccna2	Cystml	Eif3d	Txnl	
Hbegf	Placl	Got2	Uchl5	Ndufaf2	Sael	Cyb5r3	
Rab9	Abhd5	Syce2	Gadd45g ipl	Cox 14	Uqcrfsl	Szrdl	
Dnajal	Seϕ iηe2	Atp5g3	Epop	Usp39	Ilf2	Eeflg	
Fhl	Snrpd3	Atplbl	Ndufb9	Hatl	Rad5 1	Ndufs7	
Atp6v0dl	Prss36	Maea	Txndc9	Lysmd2	Psmc3	Mϕ 145	
Impdh2	Perp	Psmal	Slc38a4	Psma7	Hnrnpdl	Samm50	
Aplm2	Tmeml09	Ddx39	Rbbp4	Pole3	Brixl	Fdxl	
Sod2	Cct6a	Tmeml 16	Lgalsl	Renbp	Cox6c	Ndufvl	
Slc26a2	3830417A 13Rik	Nasp	Psmfl	Mrpl41	Ddt	Siupal	

Oligodendrocyte precursor cells (OPC)							
Sppl	Mcm3	S100a3	Rassf4	Adam9	Irfl	Col23al	Mmp2
Ccnbl	Pgcp	Creb5	Nt5dcl	Mnsl	Kif20b	Col4a5	Plekhbl
Pdgfra	Neu4	Tram2	Kif23	Bean	Tcn2	Cdldl	Slc7al l
Den	Emp3	Serpinfl	Troap	Zfp3611	Rnfl80	Pcdhga5	Cenpl
Rlbpl	Slc6a20a	Enppl	Slc25a29	Ssfa2	Slc38a3	Gal3stl	1118
Slc6al3	Igf2	Tacc3	Epn2	Tnfrsfl lb	Lgals2	Ddah2	Alpl
Inmt	Kif2c	Spry4	Qpct	Gpr81	17001 12 E06Rik	Alx3	Cede 18
Pnlip	Zcchc24	Loxl3	Gml9705	Tmeml46	Neil3	4921530 L18Rik	Fam35a
Lum	Mxra8	Cyplbl	Timp4	Kctdl2b	2900005J 15Rik	Frmd8	20103 17 E24Rik
Cmbl	Ampd3	Htra3	Jun	Col9a3	Clgn	Gprl46	Fdxr
Pcolce	Ccnb2	Ccl5	Cxcll2	Ostfl	Cercam	Phldb2	Medl8
Postn	Chstl l	Ezh2	Col3al	D2Ertd75 Oe	6720463 M24Rik	Itfg3	MtmrlO
Apod	Kif20a	Agbl2	Rfx4	Fbxo7	LOC6266 93	Trim45	E130309 F12Rik
Ednrb	Musk	Maml2	Ppfibpl	Clecla	Ehd2	Cdk4	111003 11 02Rik
Scrgl	SlOOb	Klhl5	Cyr61	Gpx7	Thbsl	Itga9	Hells
Tmem45a	mt_AK13 1586	Frmd7	Zebl	Atp6v0e		Prtg	Tϕv 4
Fam70b	Efempl	Ccl2	Ppic	Cdkl		Cdk5rap2	

Cspg4
Cacng4
Fabp7
Pbk
1110015018Rik
Emidl
Serpingl
Oligl
Vtn
Prcl
Faml80a
E130306D19Rik
Bgn
Lmcdl
Colla2
Spc25
Calcrl
Itih5
TmemlOO
Adm
Tmeml76a
06 10040JOIRik
Pmel
A930009A15Rik
Cavl
Nuprl
Gstm2
Ckap2
Spryl
Top2a
1190002F15Rik
Ube2c
Ccl7
Cp

Gpc5
Tmeml76b
Shc4
Gm2a
SlOOal
Galnt3
S100al6
Clqtnf6
Afap 112
Lbp
Cdkn2c
Vipr2
Chst5
Gpx8
Pdpn
Lims2
Mavs
Aurka
Empl
01ig2
Aox3
Mytl
Fignll
Pcdhgc3
Gpsm2
Mir568
Cd9
Fanci
Fam64a
Zic4
Cd40
Meoxl
Ect2
Rcn3
Cyp2j9
1190002H23Rik
Wipfl

Fam70a
Abtb2
Fkbp9
Cenpe
Slc2al2
Slc22a8
Ladl
Clqtnf2
Ccndl
Lamal
Smc4
Adamtsl3
Vegfc
S100a6
Kankl
Irak4
Sh3bp4
Btd
Mc5r
Rnf43
Collal
Bcasl
Plkl
Notchl
Angptll
Cdca8
Mc4r
Gpt2
mt_AKl43357
Hapln3
Lpo
Hpsl
Boll
Sema3d
S100al3
Nuf2
Ggt5
Meisl

Rhoc
Abhd2
Traf4
Tspan4
Cpxml
SoxlO
E1301 14P18Rik
Mfsd2a
Lrp4
Fos
Tpx2
Cenpi
Lame 3
Mapk7
Lama2
Fosb
Susd5
Dpyd
Uhrfl
Plekho2
Tmc6
Apobec3
Faml l4al
Birc5
B3gnt5
Itgb8
Stonl
Kcnj 1O
363245 1006Rik
Socs3
Tmeml44
Ptgfr
Slcl6al2
Chaflb
Dbi
Gfral
Cdca2

Pcyoxll
Caprin2
Pabpc5
Fzd6
Gm5089
Cenpf
Mmpl l
Rasa3
Gsn
Gm9839
Sall3
1810034E14Rik
Gpr3711
Tril
Jam2
Evi51
Dna2
Serpina3n
Cdc20
Sulfl
P2rx7
Map3kl
Dab2
Clqtnf7
Kif22
Xlr3b
Kifl8a
Zfp3612
S100a4
Seel
A330041J22Rik
Plat
Fam71f2
Smocl
Sox8
Hmgb2
Bmp6

Cd302
Coll5al
Plekhg6
Creb313
Map3k8
Timp3
Akapl3
Arhgap29
Melk
Antxr2
Bmp7
Rabl3
Tsgal4
Smpd2
Abca6
Gatm
Slitrk6
Snx22
Mpzll
Prkcq
4933425H06Rik
Gprc5a
Pcca
Prelp
Gnb4
Cyp2j6
Ctdspl
Rab34
Fzd9
Msh6
Cep72
Otos
Anxa2
Ftsjdl
Saal
Sh3tc2
Rnpepll
Atpla2

Arhgapl9
4930517El IRik
Rasll la
Tubalc
Islr
Prrxl
Rrm2
Pars2
Cftr
Slcl3a5
Lgals3bp
Cklf
Col4a2
Vamp5
Rassf8
Faml32a
Rftn2
Dill
Caldl
A430107013Rik
Fam82al
Tcirgl
Nusapl
Gprl82
Serpindl
Mcm7
Sgk3
Lekrl
Srpx2
Gpldl
1700013G23Rik
Icaml
Jam3
mt_AK159184
Coblll
Trafl

Cyp20al
Col4al
Antxrl
Aldhlal
Gabl
1300014106Rik
9930021D14Rik
Tmem220
Rhpnl
Tmeml98b
Ebfl
Ssl8
E2f8
Faml l1a
Tgfbr3
Sema5b
Ifitm3
Gdpd2
Cfh
Nnat
D930014E17Rik
Mcm9
Gins2
Slcla5
Ptgds
Tnpol
Ifitm2
Notch2
Luzp2
Mure

Vcan	Poldl	Cenpn	Gpr82	Pomtl	Pion	Mmd2	
Ugdh	1810010H 24Rik	Spsb4	Nhsll	Orail	Ppplrl4b	Sulf2	
Mdk	Cdcl4a	Cks2	Zfp41	Frrsl	Myll2a	Cnn2	
Gprl7	Tgfa	Fkbp7	Cyp4v3	Shmtl	Ndc80	Ror2	
Tnfrsfla	Tnr	Pmp22	Mtssll	Plscrl	mt_AK14 0174	Rsul	
Ptpizl	Phxr4	Cdca3	Slc22a6	Car8	AI8545 17	1700018 G05Rik	
Cdc25c	Pllp	Frk	Derl3	Srebfl	Matn4	Rab3 1	
Pcdhl5	Arhgap3 1	Kcnj 16	Limal	Plekha2	Foxcl	Dynltlc	
Ckap21	Kcnh8	Ltbpl	Ecil	Txlna	Vcaml	Sfmbt2	
Pdgfrl	Tbxl8	Cdol	Selenbpl	Epasl	Cpa4	Nkiras2	
Lhfpl3	Seϕ iηe2		Stk32a	4933406J lORik	Mdfic	Wnt7a	
Ogn					Cspg5	Mpzl2	
Itih2							

Astrocytes							
Gjal	Gramd3	Slc7al l	Btd	Zfyve21	Aldh6al	Alpl	Neu4
Gjb6	Slc7al0	Phkal	Gpldl	Lgr4	Pou3f4	Gludl	Ugtla2
CldnlO	3110082J 24Rik	Id4	Ccdcl41	Tmeml7 6a	Clmn	Tsc22d3	BC01352 9
F3	Hsd3b7	Agmo	cx_tRNA-Ala-GCG	Sycp2	Timp3	Ccbl2	Zfp783
Slcla3	Mtl	Fermt2	Tomlll	Cptla	Slc6a20a	Tnfaip8	Fjxl
Slc39al2	Bean	Crot	Scrgl	Mettll lb	Mif4gd	Zfp438	Rasl2-9-ps
Sdc4	Appl2	Elovl2	Smpd2	Loxl3	Plscr2	Hesl	Suclg2
Acsbgl	Chi311	FkbplO	Bdh2	Abhd4	Pnp	A130022J 15Rik	GdflO
Mfge8	Adhfel	MegflO	Elovl5	Papss2	Btbdl7	Slcl3a3	Atp6v0e
Ntsr2	Pxmp2	AA38788 3	Cd38	Pdgfrl	Pdk4	Cklf	Csgalnact 1
Lcat	Tlr3	Oaf	Ttyhl	Retsat	Fzd2	Egfr	1700003 M07Rik
Cml5	Vcaml	1118	Ccdc90a	Tcf712	Slc7a2	Ghr	Pyroxd2
Aqp4	Ctso	Pmp22	Crlf3	Sema4b	Tubb2b	Slc25a35	Efemp2
Pla2g7	Agxt211	Fabp7	Slc26a6	Rnasel2	Rapgef3	Ephx2	Afap 112
Ppap2b	AI46413 1	Faml63a	Lxn	Fgfrl	Prkdl	Rbpl	Dbi
Ppplr3c	Maob	Satl	Pcsk6	Igf2	Adora2b	Pdlim5	Gml073 1
Slprl	Rfx4	Kirrel2	Paqr8	Nat2	Aoxl	Cdc42epl	11900051 06Rik
Slc25al8	Acat3	Serhl	Luzp2	Mirl l92	Hist2h3cl	Qk	Abhdl4b
Plcd4	Mmd2	Gstkl	Egfl6	Dcxr	Cyp7bl	Faϕ 1	
Chrdll	Ugtla6a	Zfp3612	Fgd6	Apln	Arsk	2210417 K05Rik	
Faml07a	Gdpd2	Arhgef26	Hgf	Nrarp	Dhrsl l		
Dio2					S100al3		

Gpr3711
Mt2
Entpd2
Gstml
Cbs
Tst
Prodh
Slcolcl
Gfap
Tlcdl
Mlcl
Apoe
C030018K13Rik
Slc38a3
Aldoc
Timp4
Cyp2d22
Slcl5a2
Htral
Atpl3a4
Atpla2
Prdx6
2010002N04Rik
Fgfr3
Pdpn
Sox9
Fxydl
Itih3
Faml76a
Cyp4fl5
Gldc
Cml3
Ndp
Cyp2j9
Slcl4al
E1301 14P18Rik

Bmprlb
Prelp
Pon2
Tril
Gpc5
Nat8
C030037D09Rik
Cyp4fl4
Nkain4
Gml l627
Slc27al
Natl
Mertk
Fmol
2900052NOlRik
Cth
TmemlO0
Cideb
Cmll
Efempl
Mdk
Kcnj 16
Daam2
Scara3
Mfsd2a
1700084COlRik
Rftn2
Prex2
Dhrs3
Grm3
1700019G17Rik
Hepacam
Pgcp
Clu
Smpdl3a

Slc4a4
Cyp4fl3
Emp2
Gm973
Agt
Lixl
Uppl
Naaa
Nfe212
Steap3
Ptpizl
Cd63
Cmtm5
Gabrgl
Phkgl
Gasl
Selenbpl
Gpx8
Soatl
SlOOal
Thrsp
A330048O09Rik
Sc4mol
Rfx2
Phgdh
Hopx
Naprtl
Ndrg2
Acaa2
Slcla2
B230209KOlRik
S100al6
Pbxipl
Spatal7
Lpar4
Gpr56
Aass

Cibl
Hspb8
Acssl
Acsl6
Pion
Notch2
Ppil6
Tcn2
Renbp
Pax6
Cyr61
Gpam
Klfl5
Swap70
Slc6al l
Lgals4
Psd2
Pnpla7
Sall3
MyolO
Elmod3
Histlh2bc
Smox
Ndel
A330076C08Rik
2610034M16Rik
Gml303 1
Enho
Tnfsfl3
Plxnbl
Cdkn2c
Gem
Tmeml76b
Nudt7
E030003E18Rik
Cnn3

S100a4
Sfxn5
Dok7
Plscrl
Den
Ddo
1810014BOlRik
Nwdl
Ugp2
Myo6
Gpt
Cst3
01fr287
Kctdl4
Zbtb20
Ddhdl
ZnrG
Olfmll
Rmst
Tmem5 1
Hsdl lbl
Rdh5
Eyal
Odf311
Kankl
Paqr6
Utpl4b
Histlh4h
Lpcat3
Aldhla2
Lum
A2m
Rpe65
Rcn3
Gnal3
Cyp2j6
Fpgs
Plodl

Histlh2bq
Histlh2br
Gng5
Acsl3
Sultlal
Maml2
Echdc2
Tmem229a
c2_tRNA-Ala-GCG
Notchl
Slcl2a4
Agpat5
Rlbpl
LOC433374
Kctdl2b
Ecil
Tex 11
Lmcdl
Cbr3
Zic5
Calr4
Lhx2
Atplb2
Sox21
Gjb2
Dera
Hsdl2
Lpin3
Vgll4
Zcchc24
Slc22a4
Kcnj 1O
Vav3
Gli3
Akt2
Eps8
Nfia

Arapl
Calml4
Chst2
Emx2
Slc22a6
Parp3
Gml0052
Cede 18
Tifa
Triml2a
Serpine2
Mro
Vcl
Per3
Taf4b
I113ral
1190002H23Rik
Gypc
Kcnj 13
Gabrbl
Cmtm3
Itga7
Angptll
Stkl7b
Hacll
01fr288
Faml81b
Ccdc77
D630033O1 lRik
Phxr4
Nek3
1700084J12Rik
Asrgll
Gprc5d
Decrl
Lonrf3

Trip6
Lama2
Gml7660
Rin2
Fndc4
Slc30al0
Scg3
Abcd4
C230035I16Rik
Ptplad2
Rasa2
Acadl
Lrrc9
1700040N02Rik
Zfp521
Prkcd
Ranbp31
Npcl
Hif3a
Pfkfbl
Fcgr2b
Rdml
Mmpl4
Grtpl
Wnt7b
Trp53bp2
C2
Lgals3bp

Pdlim4 Aldhlll Mgstl Dbx2 Ezr Slc9a3rl	Fam20a Gm5083 Abhd3 Ednrb St3gal4 Rarres2 Glul Faml98a Gm5089	Hadh Acotl 1 Pax6osl Ttpa Gstt3 Cdhl9 Nrlh3	4932438H23Rik Lrp4 Id3 Aqp9 Histlh4i Tdo2 Gstm5 Slcolb2	Fgfr2 Dockl Frrsl Fads2 Seppl Trp63	Tsc22d4 Lrrc5 1 Grhll Tnfrsfl9 Adrbk2 2810055G20Rik	Rnfl82 Mmgt2 Paqr7 Haplnl Cox6b2 Sohlh2 Nphp3 Idh2 Btgl	

Cortical Neurons							
Nos1 Fam84a Unc5d Rnd2 Pou3f2 Pdzrn3 Hs3st1 Sstr2 Pcp4 Meis2 Lrrc16b Plekhf2 Sorl1 Ppp2r2b Trim9 Pou3f1 Frmd4b Mllt3 Plcb1 Ppp2r1b Lsamp Enc1 Robo2	Scrt2 Cdh4 Slc17a6 Osbpl6 Sema3c Kif21b Wnt7b Tbr1 Chga Tenm4 Lmo1 Tsc22d1 Igfbpl1 Nrn1 Wbscr17 Itpk1 Sox5 Prex1 Rcor2 Kctd4 Cited2 Epha3 Palmd Bcar1	Neurod2 Srrm4 Adgrl2 Jarid2 Pou3f3 Cttnbp2 X6330403K07Rik Nav2 Pantr1 Lrpap1 Trim2 Nek6 Ldhb Lhx2 Tagln3 Mn1 Vopp1 Gm17750 Nfib Neurod6 Rasgef1b Hs6st2 Insm1 Tmem178	Serpinil Ttc28 Epha5 Ankrd6 Tmeml58 Plxna4 Nfasc F2r Fmnl2 Cbfa2t2 Lztsl Sorbs2 Frmd4a Plxna2 Foxgl Cdknlb Luzp2 Dpyl911 Rbfox3 Cd24a Cdldl Cyth2 Negri Hist3h2ba	Nedd4l Fam114a2 Cux1 Mta2 Acly Baz2b Phf21b Phip Tmeff1 Ddah2 Grina Smim18 Rbfox1 Sncaip Lrp8 Avl9 Nfix Tnrc18 Znrf2 Adgrg1 Abracl Mpped1 Gria2 Zbtb18	Gstm7 Emx1 Tmem108 Dbn1 Myt1l Cul1 H1f0 Kif21a Ilf2 Rpf1 Ing4 Hist3h2a Bcl7a Hivep3 Hbb.bs Gdap1l1 Fam107b Podxl2 Setbp1 Wbp1 Ip6k2 Igsf3 Gm14964 Nrp1	Elavl4 Scg5 Scrn1 Ptprs Midn Kdm2b Laptm4a Fam49a Acin1 G3bp2 Mdk Sbk1 Auts2 Kdm5b Ap3s1 Basp1 Tmem57 Peli1 Cux2 Ttc9b Rundc3a Mpped2 Mkrn1 Akap9	Cdk2ap1 Cplx2 Efnb2 Klhdc2 Ccng2 Parp6 Nipsnap1 Tax1bp3 Ezr Nol4 Elavl2 Arhgef2 Nsg2 Pbx1 43346 Zfp462

RadialGlia-Id3							
Id3 Idl Foxj l Mtl Mt2 Pla2g7 Hes5 Hesl Mia Egrl Metrn Fos Tmem47 Ednrb Tppp3 Clu Serpine2 Riiad1 Gfap Sparcll Apoe Slcla3 Nlrxl Selm Ttyhl Gstml Lxn Cyr61 Fbxo2 Mlcl Enkur Mlfl Mgstl Slc9a3r1 Bean	Heyl Aldoc Anxa2 Atplb2 Ncan Atpla2 Cybrdl Tmeml07 Lgalsl Slcl4a2 Rhoq Tlcdl Rhoc Sox9 Ccndl X1500015OlORik Bhlhe40 Zfp3611 Ddit41 Nimlk Nme5 Lfng Tagln2 Mfge8 Stom Pbxipl Empl Mpp6 Pdpn S100al6 Tspan33 Aldhlll Fam212b Fzd9 Pdlim5 Eepdl Ier3	Efcabl Nes Mest Slc6al l Glul Faml81b Camk2d Zfp3612 Gjal X2810459M l lRik Spry2 Vim Acadl Igfbp2 Ckb Paqr8 Gng5 Hspa2 Lrigl Erf Zic5 X1810037I17Rik Bcl2 Ier2 Vcaml Ptn Nkdl Trim47 Ptprzl Krccl Scd2 Tnfrsfl9 Zfp36 Idil Serpinhl Ntrk2	Add3 Lrp4 Ifitm3 Tspanl5 Slc27al Gludl Timp3 Hopx Cav2 Arl4a Chptl Fhll Tst Plpp3 Spal7 Tomlll 43352 Msn Pttgl Ninj l Fkbp9 Ctsc Rrbpl Prkcdbp Gnai2 Nr3cl Ldha Slc38a3 Zcchc24 ZnrG Akrlbl0 Hadh Myo6 Kcnj lO Acadm	Morn2 Nafl Cripl GrblO Itm2c Sparc Mmd2 Mcm3 Acyp2 Adcyaplrl S100al3 Eif4ebpl Irsl Cibl Afapll2 Ttyh3 Notch2 S100a6 X2610301B20Rik Magtl Itgb5 Kbtbdl l SlOOal Mif4gd Tnfaip8 Pcx Dnajc3 Dagl Rgs20 Tapbp Hmgcsl Nudt4 Mlec Degsl Abhd4 Sp3os Sashl	Slc25a25 Pmp22 B9dl Purb Ctso Axl Dhcr24 Tppl Stxbp6 Rasa3 Cbfb Pacsin2 Gcsh Parva Zebl Nkain4 Snx5 Ormdl2 Adgrvl Stard4 Car2 Sox21 Slprl Slcl2a4 Hacdl Cd9 Wwpl Jun K1M13 Gabrbl Msi2 B2301 18H07Rik Eef2kmt Nr2c2ap Dpcd I16st Rgcc	Pex7 Galkl Hsdl7b7 Anxa5 Ift22 Sgcb 43358 Tmem218 Slcla2 Rbpl Arhgef26 Dnajcl5 Pmml Cfap36 Etfa Pidl Ctdspl Ecil Plxnbl Klf6 X I500009L16Rik Emc7 Dennd2a Zdhhc21 Plcel Oat MyolO Phyhipl Maml2 Irs2 Msmol Mras Mtssll Asrgll Faml95a Socs2 Fadsl	X2810417H13Rik Extl Tancl Lhfp Amot F3 Pmfl Stat3 Ppplrla Gprc5b Dhfr Lyrm5 Cdk2 Nfkbia Cntln Gasl Pfnl Prdxl Golph3 Cy stall Kcnip3 Prdx4 Rad23a Traml Dclkl Hspa5 Gm2a Smo Spcs3 AI8545 17 Flna Csrpl Gpt2 Ift74 Sytl l Clicl 1118

Fabp7 Dbi Emp2 Ppp1r3c Igfbp5 Wis Tpbg Fgfr3 Hepacam Aqp4 Oligl Tnc Mt3 Slc4a4 Gngl2 Pacrg Rspo3 Phgdh Tril Qk Ccdc80 Aard Plat 01ig2 Rfx4 Cmtm5 Id4 Socs3 Scdl	Fbln2 Junb Peal5a Kcnell Etv4 Rampl Sfxn5 Egfr Klf4 Gpx8 Cpne2 ChchdlO Ndrg2 Rmst Nebl Jam2 Acsbgl Pon2 Fosb Smpdl3a Fatl Sema6a Gdpd2 Tsc22d4 Sall3 Gsta4 Cspg5 Neatl	Suclg2 Metrnl Rgma Rcnl Axin2 Klf9 Klfl5 Npas3 Satl Chst2 Paqr4 Cd63 Spryl Dkk3 Bmprla Epdrl Yapl Adamtsl Mnsl Aldoa Ccnd2 Slcla4 Nog SlOOal l Itga6 Fgfbp3 Duspl X3110082J24Rik X1700088E04Rik	Psph Psatl Prrxl Tns3 Slc39al Itgav Gm5617 Ccpglos Notchl Prrl8 Cbs Rest Anxa6 Insigl Nrarp Emc2 Thrsp Efemp2 Acotl Bphl Nr4al Ppic Cxxc5 Ill lral Gins2 Rorb Sox2 Rabl3 Nacc2 Ung	Fjxl Uhrfl Slcl5a2 Cenpw X I 110004E09Rik Cebpb Tspanl2 Tribl Pcgf5 Pnp Faml20a Gmnn Polr3h Creb5 Pygb Trim9 Ppargcla Grm5 Rab3 1 Grhpr Btg2 Gale Tjpl Cnp Donson Cst3 Hspa41 Cln5	Rnftl Rasll la Ak3 Echdcl Nr2f6 Vamp3 Arhgef40 Ifngrl Phxr4 Tm7sf2 Mvk Dnajc24 Hsdl2 Bola3 Wwtrl **Traf3** Spata24 Bakl Tspan7 Lppos Nab2 Mcee Chsyl Dusp6 Midlipl Cetn2 Dtd2 Trpsl	Trip6 Rexo2 Ptgfrn Sri Nfe212 X23 10022B05Rik Snx3 Thbs3 PcdhlO Elofl Tctexld2 Fgfr2 43345 Betl Spsb4 Lss Phlda3 E2f5 Nrcam Ddahl Klhdc8b Plin3 KlflO Klf3 Gltp Ccdc8 Speccl X4933434E20Rik	Myll2a Scrgl Nphpl Proml Ctnnal Pde4b Ligl Itgb8 Sox8

RadialGlia-GdflO							
GdflO	Assl	Pdpn	Arhgef26	Gmnn	Ligl	Rfcl	Msi2
Id3	Htral	Dkk3	Rcnl	Pdcd4	Prps2	Glol	Tyms
Tesc	X2810459M1lRik	Col9a3	Noval	Cdl64	Gstm5	Tpx2	Spg20

Thrsp
Tnfrsfl9
Frzb
Idl
Sdpr
Emidl
E330013P0 4Rik
Hspb8
Pdlim3
Den
Gfap
X1500015 OlORik
Mt2
Lefl
Rmst
Gasl
Tst
Mgll
Zic5
Sp5
Hopx
Prex2
Eyal
X0610040J OlRik
Cavl
Mtl
Adamtsl9
Wnt8b
Nme7
Cripl
Zfp3611
Cyplbl
Lhx9
Vim
Rgs20
Hes5

Bcl2112
Gjal
E1301 14P1 8Rik
Nkdl
Ninj l
Enpp2
Fzdl
Selm
Hadh
Psph
Sfxn5
Aard
Lrrcl
Dbi
Frasl
Slc9a3rl
Ltbpl
Dmrta2os
Notchl
Lhfp
Emx2
Bcl2
Axin2
Etv4
Sez61
Efcabl
Fos
Mro
Tnc
Rhoc
Rfx4
Rgma
GrblO
Ung
Atpla2
St3gal4
X2700046A 07Rik

Mgstl
Lrp4
Foxol
Dmd
Entpd2
Dmrt3
Chst2
Gpx8
Tsc22d4
Isocl
FkbplO
X 1110015 018Rik
Gngl2
Epdrl
Cpne2
Ptgfrn
Mt3
Zicl
Lmcdl
Notch2
Id4
Msn
Mlcl
Qk
Smco4
Eepdl
Myl9
Cdkn2c
Tspan7
Cd9
Gabra4
Dtl
Gnai2
Plpp3
Cenpf
Klf9
Faml67a
Gldc

Appl2
Mki67
Phxr4
Anxa6
Nr2f6
Gli3
Tgifl
Pygb
Tspanl5
Sdc2
Tspanl2
Fatl
Zfp3612
Hells
Hmgb2
Cdca8
Cst3
Aifll
Itga6
Lockd
Gstml
Acotl
Ube2c
Pttgl
Lixl
Btg3
Otxl
Cbfb
Pnp
Tgif2
Cks2
Pbk
Rpa2
Limdl
Idil
Cyba
Top2a
Sesn3

Maml2
Scrgl
Kcnmb4
Ccna2
Kbtbdl l
Lap3
Knstrn
Gng5
Chptl
Snx5
4335 1
Slit2
Itgb8
Mcm3
Prdx4
Litaf
Ctdsp2
Kcnipl
Hnll
Gcsh
Hs2stl
Cdkl
Slcla4
Dhcr24
Arl4a
Dhfr
Shisa4
Tmeml07
Pcx
Ldha
Slc39al
Serpinhl
Tcfl9
Bola3
Ndel
E2f5
Camk2d
Cdk2

Naa50
Sypl
Krccl
Eci2
Jam2
Cisd3
Fezf2
Lhfpl2
Mcm5
Nadk
Tjpl
Cxxc5
Proml
Pacsin 3
Pankl
Dennd 2a
Rdml
Uspl
Cmc2
Nit2
Adgrb 1
Nme4
Echdc 1
Apoe
Mcm6
Smc2
Dclkl
Dtymk
Jam3
Pax6
Paqr4
Stard4
Elavil
Vcan
Histlh 1e

Atxn7
Cenpw
Ddahl
Proxlos
Torlb
Asahl
Ndufc2
Bmprla
Crip2
Cpne3
Lysmd2
Sat2
Abhd4
Faml20a
Rcn3
Ckslb
Kpna2
Evi5
Pmfl
Dpysl4
Ifitm2
Bach2
Slc35a4
Kcnell
Cdol
Sival
Pcna
Efemp2
Cntln
X23 10022 B05Rik
Acadm
Ier2
Cdc42sel
Adrbk2
Mvk
Rragd
D8Ertd82e
Nudt4

Fut9
Proxl
Pmp22
Ccdc34
Sntal
Cdv3
Tmem2 56
Ssl8
Aamdc
43345
Sox6
Arhgap5
Paics
Snap23
Scd2
Ctdspl
Gsr
Fkbp9
X49334 31E20Ri k
Atplbl
Exosc5
Mettll
Atplal
Syce2
Ost4
Actnl
Rangrf
Hmgn3
Nrarp
Carnmtl
Hmbs
Rnftl
Sytl l
Fuz
Tspanl8
Fam96a

Tpbg Slcla2 Aldoc Slcla3 Psatl Ttyhl Hesl Tspan33 Cpne8 Hepacam Sox9 Vcaml Ccndl Tmem47 Gludl Snedl Ccdc80 Fbxo2 Lfng Tfap2c Ndrg2 Cthrcl Cav2 Mmd2 Phgdh	Fbln2 Vephl Tmeml32c Dmrta2 Col2al Emp2 Nimlk Loxll Pbxipl Mfge8 Rest Trip6 Gabrbl Fgfr3 Pon2 Tns3 Tgfb2 Fam49b Prkcdbp Cspg5 Zcchc24 Slc27al Sashl Gas6 Adgrvl	Paqr8 Rftn2 Stxbp6 X23 10009 B15Rik Gins2 Uhrfl Ephbl Clu Lrrc4c Gsap X2810417 H13Rik Cdca3 Socs2 Adcyaplrl Ptn Yapl Cbs Sparc Cenpm Cyr61 Prdx6 Vatll Sox2 Ttyh3	Csrpl Tancl Erf Sox8 Tex9 Map3kl Fignll Sirpa Spc24 Dnajcl Ephb3 Atplb2 Mif4gd Heyl K1M5 Birc5 Sapcd2 Tead2 Ecil Chd7 Npas3 Cenpa Hrspl2 Klf4	Ccnb2 SlOOal l Tmem97 Rabl lfip2 Eefld Mcm4 Suclg2 Gem Ehbpl Insigl Pdk3 Amot Smo A730017 C20Rik Vamp3 Ramp2 Arhgef40 Epsl5 Wwtrl Rnf26 Vgll4 Rexo2 Btgl Cdon	Tulp3 Mcee Nudt5 Ptprg Histlh 2ap Decrl Higdl a Ift74 Lsm2 Ldlrad 3 Cachd 1 Ppplr 1a Histlh 4i Acadl Mcm2 Nacc2 Prdxl Fxyd6 Nr2el Itgb3b P Ckap2 Vldlr Tipin Homer 2 Kctdl 2 Dagl Rpe	Csad Purb Rpl2211 Fjxl Mpp6 Bcl7c Stx4a Mgatl 43358 X2810004 N23Rik X150001 1 K16Rik Anp32b Rpal Spredl Hspa41 Crot Tmeml67 Echdc2 Caldl Lhx2 Nek6 Lyrm5 Toporsos Arl6	Dennd5 a Nudcd2 Dnphl Ybx3 Speccl Tpil Akr7a5

RadialGlia-Neurog2							
Neurog2	Kif26b	Wasf2	Dnajb2	Echdcl	Asahl	Hyal2	Ndufaf7
Eomes	Tmem98	Ecil	Asnsdl	Elavil	B230354 K17Rik	Nrnl	Gm8730
Gadd45g	Fam53b	Mmpl4	Zbed3	Akr7a5	Acadvl	Shmt2	Dexi

Rhbdl3
Ptgds
Btbd17
Snhg18
Lima1
Tfap2c
Mfng
Btg2
Myo10
Csrp1
Tead2
Pax6
Celsr1
Gm29260
Chd7
Acads
Heg1
Dll1
Gamt
Kcne1l
Tox3
Rcn1
Gfap
Igfbp5
Hes6
Efhd2
Inppl1
Lrrn3
Sfrp1
Nme4
Sox21
Loxl1
Fam210b
Dbi
Tgif2
Ccnd2
Vim
Mfap4

Dhx32
Abcd2
Lzts1
Dll3
Aif1l
Cbs
X1500015O10Rik
Gpx8
Cmc1
Slc1a2
Bcl2l12
Rnaseh2b
Mcm2
Ezr
Gng5
Tank
Apool
Spsb4
Hrsp12
Cd63
Ccdc136
Ddit4
Grb10
Pttg1
Nr2e1
Tmem218
Btg3
Zeb1
Eef1d
Sstr2
Thrsp
Sema5a
Gas1
Slco1c1
Rcn3
Ctnna1
F2r
Zfp703

Ckb
Gadd45gip1
Ddah1
Glo1
Ccs
Ift74
Slc25a5
Sfxn5
B230118H07Rik
Pam
Lzts2
Hmgn2
Ddr1
Ninj1
Srek1ip1
Adk
Snx5
Acot1
Zfand1
X2610301B20Rik
Serpinh1
Cib1
Fbln1
Syne2
Nrg1
Ncald
Elavl2
Phgdh
Ly6e
Insm1
Abca1
Slc1a3
Ttc8
Phyh
Ccdc167
Dnajc15

Vps37b
Fubp3
Dcaf8
Tbrg1
Ufm1
Wscd1
Lta4h
Idh2
Gstm5
Sema5b
Hadh
Ftsj3
Pyurf
Eci2
Paqr8
Fam96a
Atf5
Rps18.ps3
Cdca7
Rexo2
X2810004N23Rik
Prdx1
Efs
Golph3l
Echs1
Ormdl2
Exosc3
Ccdc58
Anp32b
Cul1
Sox6
Hdac1
Tmem33
Limd1
Tor1aip1
Por
Adcyap1r1

Ift22
Ctnnb1
Azi2
Ece2
Pmepa1
Bphl
Fundc2
RP23.207N5.2
Paics
Rbpj
Rangrf
Rpl22l1
Ptbp1
Nedd4
Aco1
Flna
Nkain4
Rprm
AI854517
Polr3k
Hsd17b4
Trap1
Mcee
Npc2
D10Jhu81e
Mettl1
Dazap2
Ino80b
Rbbp9
Prdx6
Elp4
H1f0
Exosc5
Sipa1l1
Sesn1
Gm14305
Pbdc1

Cnih4
Yif1a
Ift52
Srsf6
Hibadh
Foxp4
Gnpda2
Cpne3
Lamp2
Itgb3bp
Rcor2
Cplx2
Cadm3
Ankrd6
Myl12a
Lman2
Cnpy2
Mrpl17
Trp53
Mrps14
Fars2
Serinc2
Prdx3
Fam162a
Atp5g2
Sp3os
Mettl5
Clic4
Twf1
Lap3
Creb5
Emx1
Rrs1
Cdkn2c
Rps27l
Ebpl
Timm21
Nsmce4a

Zfp62
Svip
Ubxn2a
Rad23a
Golim4
Scrn1
Vrk3
Urod
Taf10
Pdcd4
Rbfox3
Mphosph10
Emg1
Smarcad1
Rrp15
Ldha
Ppib
Cdk4
X1500011K16Rik
Tmed1
Cdk5rap3
Acly
Lyrm4
Slc48a1
Mt2
X1110012L19Rik
Fam174b
X1810037I17Rik
Hnrnpf
Tpm4
Mt1
Acvr2b
Gcsh
Ift57
X2310039H08Rik

Pno1
Gspt1
Fxn
Snhg6
Ccdc86
Bola3
Kti12
Pou2f1
Mrpl24
Rit1
Lztfl1
X1810058I24Rik
Swt1
Eif3i
Spata2
Tef
Vamp3
Ift43
Guf1
Gm10020
X2310011J03Rik
Setbp1
Rnf13
Mccc1
Akr1b3
Hspe1
Ralgds
Hmgn5
Immp1l
Carnmt1
Iscu
Isca2
Tspan3
Gkap1
Actl6a
Pdia6
Ppie

Mdk Notchl Gem Magil Corolc Mfap2 E1301 14 P 18Rik Dleu7 Ascll Igdcc4 Tmeml3 2b Myo6 Uaca Slc30al0 Gml l62 7 Pdlim4 Zhx2 Jam3 Zfp423 Cdl64 Pgpepl Dhrs4 Igsf8	Mdgal Inhbb Pnpla2 Zfp3611 Sufu Smco4 Rab8b Dmrta2 Ndrg2 Cdk2apl Ehbpl Echdc2 Egrl Hs3stl Msn Hmg20b Cbfa2t2 Rgs3 Elavl4 Aldh2 Chn2 Rabl3 Fdxl Mfge8	Lyrm5 Smpd2 Litaf Nudt5 Krccl Scp2 Ube2g2 Betl Trappc6a Tsc22d4 Actr3b Dnajc24 Sdc3 Sox2 Fezf2 Gtf3c6 Emidl Pcmtd2 Aldh6al Prmt8 Smiml l Kdm7a Qsoxl Nrarp Pex7 B9dl	Cyba Hadha Teadl Calu Ndufc2 Etfa Dync21il TmedlO Snapin Lrp8 Hdhd2 Cdk6 Ssl8 Ctage5 Pcbd2 Fam58b Qars Tfdp2 Aldh7al Kat6b Nit2 Tcf3 Adgrgl Acadm Glrx2	Wdr61 Adgra3 Pabpcl Llgll Clicl X2210016 F16Rik Draxin Ginml Ddx52 Msi2 Zfp219 Ppp2r3c Rcn2 Arl6ip6 Tmed4 Stx4a Klf3 Ivd Fgd4 Bbx Ssbpl Hadhb X2810006 K23Rik Bckdha Efnbl	Dhx40 Mmd2 Rhoc Ppp2r3d Spire 1 H2afv Mϕ 154 Tlel Tpcnl Igbpl Dszf5 Sec23b Chracl Smim20 Gpil Pts Plagll Rcbtb2 MϕIIO Pgap2 Zmizl Slc35b2 Morn2 Zfp664	Rpe Zbtb38 Crnkll Aamdc Gnpat Pfkl Gml0073 Mybbpla Capn2 Eiflb Ntrk2 Pga m l Josd2 Tϕ c4aρ Ctsz Ubxn4 Lengl Tmem230 Tmeml78 Sat2 Cd320 Dermd5a Ost4 Nabp2 Nudcd2 Faml20a Mrfapl	Sod2 Odcl Fucal Polr3c Med9 Pex2

Long-term MEFs							
Rps3a3 Timpl Bexl Rhox5 Gml545 9 S100a6 Gml032	Ckslb Pinl Ccngl Tpil Eif4ebpl Tubb6 Txnl4a Cdkn2a	Utfl Trappc4 Vdac2 Mϕ s6 Gml0039 Srupe Ruvbl2 Txnrdl	Crabpl Pfdnl Atp5b Hspa9 Nedd8 Ube2a Nsmce 1 Rpl23a	Nop 16 Tacc3 Ncl Naca Hintl Rcn2 Pgd MϕII I	Manf Psmc2 Dnlz Rps25 Pdrgl Steapl Snx5 Rtn4	Rplp 1 Srsf3 Psma 5 Polr2 e Eif31 Srupa	Cox6al Ppmlg Nosip Olal Gtf2f2 Hprt Sec 13 Ndufs6

0 Gsto1 Gm11942 S100a4 Gm10260 Mif Esd Gm15772 Anxa1 Ctgf Rps27l Pkm Bex3 Txn1 Tagln2 Tnfrsf12a Ldha Selenoh Serpinb2 Gm28438 Tex19.1 Gm10263 Tubb5 Birc5 Ran Anxa2 Gsta4 Nme1 Trap1a Rrm2 Prdx1 Il11 Tm4sf1 Tuba1c Tuba1b	Npm1 Cenpa Tagln Lgals1 Tmsb4x Hmgn1 Atp5g3 Acot7 Ranbp1 Plaur Vim Cnih4 Anxa3 Tnfrsf11b Dctpp1 Cnn2 Eif5a Ass1 Krt18 Cdc20 Psma6 Ccnb2 Prelid1 AA465934 Cct8 Ppia Bola2 Eef1b2 Dut Ap1s1 Rpsa-ps10 Psma2 Cct4 Hmga2 Psmd8 Pclaf Snrpd1	Actb Snrpa1 Mrto4 Abracl Pgk1 Ngf Cct3 Hbegf Rack1 S100a11 Eno1b Cox5a Timm17a Eloc Mtch2 Fkbp3 2810025M15Rik Slc25a3 Rps13 Rpl7a Gm11273 Pa2g4 Thyn1 Cdk4 Eif1ax Serpine1 Psma1 Cct7 Btf3 Hspd1 Gng2 Mtpn Tomm40 Ccnb1 Slc25a5 Psmb3 Tyms Rpl13a	Psmd12 Dynll1 Rps20 Rhoc Pdlim1 Cct5 Phf5a Glrx3 Sh3bgrl3 Pomp Nudcd2 Apoc1 Nmd3 Rpl19 Cacybp Ddx39 Hnrnpc Spp1 Cstb Cox7b Tes Lxn Nasp Atp5o Rpl39 Eif4a3 Gars Gjb3 Mrpl20 Elob Ptgr1 Acta2 Eif3d Bdnf Cops6 Pno1 Fam162a Hnrnpab	Rps17 Ftl1 Strap Atp5f1 Idh3a Ctxn1 Avpi1 Rps8 Stip1 Cdca8 Mdm2 Eif2b3 Arl6ip1 Rps3 Capg Hspe1 Edf1 Calr Spc24 Rps24-ps3 Prdx2 Shmt2 2810004N23Rik Lamtor1 2010107E04Rik Yrdc Commd3 Pebp1 Ccna2 Perp Tmem126a Rps5 Fcf1 Atp6v1g1 Dars Lsm5 Tpm4	Csnk2b Nab2 Hcfc1r1 Eif1a Cap1 Fhl2 Pam16 Psmb5 Chchd1 Dtymk Bud31 Rassf1 Rbm8a Snu13 Snrpd2 Mthfd2 Gins2 Hsd17b12 Rplp0 Bzw1 Psmd13 Denr Atpif1 Cox7a2 Ptrh2 Mybbp1a Nsun2 Mrpl30 Aimp1 Emc6 Arpp19 Snx3 Coq7 Tmco1 Rars Phb2 1810022K09Rik Apex1	Rps4x Farsa Rpl17 Mrps15 Cisd1 Eif2s2 Arpc5 Mrpl42 Noct Txndc9 Mrpl35 Nt5c Snrpg Eif3i Rpl7l1 Tgif1 Rab1a Nip7 Plp2 Vps29 Dph3 Ndufb6 Lap3 Naa38 Zyx Sae1 Rpl30 Tpm2	Eif3g Brix1 Timm10 Mrps14 Sf3b4 Prps1 Emc8 Ndufs4 Uba3 Srm Gtf2h5 Mrpl17 Selenof Praf2 Med7 Tuba1a Tspan4 Degs1 Rps26 Ppil3 Dnaja2 Itgb1bp1 Cldn4 Commd2 Nol7 Cops5 Txndc17 Txn2 Prdx4 Wdr12 Prdx5 Vta1 Alad Imp4 Exosc8 Mrpl39 Rpl22 Nras

Enol Cks2 Psatl Ube2c Cldn3 Fabp3 Hatl Mrpll2 Eif2sl Cfl l Myll2a Tubb4b Clicl Cdkl Aprt Gm4366 Hmgal Vmpl Crlfl Gapdh Banfl Rpll8 Galkl	Bax Rpl27 Inhba Psph Gml673 Naplll Pttgl Eeflel Sϕ 14 Psmdl4 Bri3bp Asns RpslO Clqbp Cnihl Rpll2 Nhp2 Cct2 Cdkn2b Rpl2211	Tbca Sgkl Aldoa Mtap Actgl Rps41 Gmnn Prdx6 Med21 Dnphl Pfdn4 1110008F13 Rik Lsm2 Pfnl Slcl6a3 Psmc6 Capzb Txnll Uqcrq	Mϕ 113 Rpsl 2 Rpll l Fkbpla Eefld Rplp2 Nme4 Aurka Aaas Fosll Ndufb8 Lsm8 Timm50 Hnl 2200002D0 IRik Serbpl Ankrdl Rbxl Itga5	Cct6a Rpl34-psl Mϕ 128 Ssscal Hspbl Rgsl6 Rpl9 Paics Ciapinl Mϕ 151 Elofl Mϕ s18a Tcpl Tkl Phlda3 Zwint Rheb Chmp6 Ndufa7 Cox6bl	Tpml Rslldl Rϕ 9 Psmb6 Bag2 Psmcl Nup35 Psmbl Prss23 Ndufa8 Akl Bcap3 1 Sigmarl Ak6 1500009L 16Rik Tipin Slirp Snx7 Pmfl	Uqcrb Cede58 Rpl6 Gpxl Ppplrl l Thoc7 Cdc37 Polr2f Nradd Aϕς2 Mϕ ĩ57 Gnl3 Vbpl Pmm1 Rpsl5a Mob4 Atxn10 Usp39 Zfp593 Hikeshi Tars Rpl28 Erh Rpsl5 Phgdh Krt8	

						Cox17 Fez2 Tbpl1 Arhgdia Dda1	

Embryonic mesenchyme							
Matn4 Matn1 Col9a1 Col9a3 Cnmd Asb4 Col9a2 Wwp2 Sox9 Col2a1 Nnat Hapln1 Cytl1 Cd24a Mest Mia Bex2 Mpz Cdkn1c Papss2 Stmn1 Ldha Plod2 Cdk4 Slc26a2 Bex3 Epyc Pdia6 Ss18l2	S100b Crabp1 Fibin Siva1 Gpc3 Cthrc1 Tpi1 Hnrnpd Col11a1 Cpe Fgfr3 Eno1 Ccnd1 Rflna Rangap1 Maged2 Mlf2 Snrpa1 H2afx Cacybp Gale Pdrg1 P4hb Ldhb Srm Susd5 Ltv1 Tubb5 Gadd45gi	Hmgn1 1110004F10Rik Gm1673 Psmd6 Ssr2 Sub1 H19 Grb10 Prpf19 Elovl6 Dek Pkm Snrpd3 Ptov1 Psmc4 Nop10 Tial1 Lman1 Tceal9 Hspd1 Eef1g Krtcap2 Snap47 Cks1b Tmem97 Kdelr2 Selenoh Vdac3	Pdap1 Sdhaf2 Hpf1 Rer1 Tmed1 Mif Hnrnpm Gars Capn6 Fus Psma7 Gstm5 Fkbp11 Skp1a Apex1 Papss1 Cct3 Mrpl15 Nsfl1c Anapc11 Mcm7 Npm1 Snhg6 Rnf7 Ssrp1 Cnpy2 Tfg Lrrc59 Mdk	Prelid1 2210013O21Rik Serf1 Pdxdc1 Srsf3 Gnl3 Ndufa4 Meg3 Fkbp4 Rcn1 Itm2a Hsp90b1 Ugdh Ddx39b Hspe1 Sec61b Ptma Atxn10 Ranbp1 Cct6a Mrpl34 Serpinh1 Dcakd Atp5j Tecr Serp1 Nme1 Hnrnpc	Bub3 Psmb6 Thoc3 2310036O22Rik Rpl36al Limd2 Hnrnpa2b1 Snx17 Elp2 Atp5a1 Slirp Atp5k Blmh Nasp Hint1 Ddx39 Ap1m1 Eif5a Galk1 Polr2i Tspan4 Atp5f1 Rpl11 Rpl14 Luc7l3 Ube2e3 Ywhab Akr1a1	Peg3 Atp5g1 Slc25a4 Nop58 Chchd2 Arf1 Ier3ip1 Rps27a Calr Swi5 Rps9 Cox5a Rpl18 Ndrg2 Usmg5 Rps2 Tmem258 Serbp1 Rps13 Elob Dad1 Rpsa Gapdh Gnas	Rpl31 Rps11 mt-Nd1 Rpl10 Rps5 Rpl26 Rps8 Rps15a Rplp0 Rpl13 Rps25 Rpl18a Rps14 Dlk1 Rpl41

	p1 Ccnd2	Srsf2 *Srpl2*	K1M13	Atp5o Siupa	Rps26 Ndufcl	Tsc22d1 Igf2 Id3 Cfll Hsp90abl Rpsl7	

Cxcl12 co-expressed							
Il1r1 Col3a1 Col5a2 Igfbp5 Sned1 Ifi203 Nenf Pfkfb3 1110008P14Rik Lcn2 Serping1 Ube2l6 Fibin B2m Eid1 Fgf7 Cpxm1 Ism1 Cst3 Lbp Wisp2 Zbp1 Srpx	Il13ra1 Apln Hs6st2 Bgn Slc16a2 Capn6 Gpm6b Cp Dclk1 Mme Ptx3 Tbx15 Slc16a1 Vcam1 Penk Svep1 Ugcg Plpp3 Podn Hivep3 Col8a2 Nbl1 Mfap2 Dhrs3	H6pd Isg15 Steap4 Emilin1 Htra3 Nsg1 Sod3 Pdgfra Cxcl5 Cxcl1 Plac8 Spp1 Pkd2 Tgfbr3 Oasl2 Col1a2 Ptn Rarres2 Tmem176a Loxl3 Cyp26b1 Antxr1 Slc6a6 Cxcl12	C1ra C1s1 P3h3 Fxyd1 Rcn3 Fcgrt Saa3 Prss23 P2ry6 Adm Il4ra Ifitm2 H19 Igf2 Rspo3 Bicc1 Col6a1 Aes Igf1 Dram1 Dcn Lum Ndufa4l2 Lrp1	Gas6 Sfrp1 Slc7a2 Comp Bst2 Rnf150 Ier2 Nfix Junb Mmp2 Mt2 Mt1 Cdh11 Hp Stc1 Pdlim2 Slc39a14 Tsc22d1 Mmp13 Mmp3 Clmp Nnmt Islr Loxl1	Itga11 Col12a1 Selm Ebf1 Slfn2 Col1a1 Igfbp4 Mrc2 Timp2 Lgals3bp Sfrp4 Aspn Ogn S1pr3 Cxcl14 Gas1 Vcan Pik3r1 Il6st Stxbp6 Hif1a Zfp36l1 Npc2 Ltbp2	Serpina3g Serpina3n Ghr Osmr Lifr Snhg18 Ly6e A4galt Fbln1 Pdzrn4 Rtp4 Mylk Fstl1 Nfkbiz Abi3bp Tmem45a Col8a1 Adamts5 Kcnj15 Fndc1 Sod2 Thbs2 Angptl4 Cyp1b1	Pkdcc Epas1 Colec12 Egr1 Lox Iigp1 Synpo Pdgfrb Efemp2 Pcsk5 Ifit3 Ifit1

Ifitml co-expressed

1500015010 Rik Crocc2 Snedl Fmod	Seϕ iι¾ 1 Cst3 Ptgis Slcl6a2 Fabp5	Cp Gperl Gngl l Cemip Adm	Ifitm2 Ifitml H19 Akapl2 Gjal	1500009L1 6Rik Scara5 Zic5 Mmpl3 Clmp	Ctsh Zicl Zic4 Ebfl Sfϕ 4	Tgfbi Hifla Aspg Fblnl Kng2	Apod Abi3 bp Epha 3 Smoc 2 Thbs 2 Epasl Prdm6

Matn4 co-expressed							
Spats2l Igfbp5	Kcns1 Matn4	Penk Mfap2	Eln Cpxm2	Pdgfrl Igfbp3	Mfap4	Igfbp4	Nov

2-cell								
Tcl1b1 Dusp7 Zbed3 Tcl1b2 Gm839 NA.139 91 Gm1965 Phf1 Tcl1b3 Siah2 Tcl1b4 Phc2 Tcl1 Tbx19 Obox3 NA.685 5 Gm1278 9 Wee2 Bcl2l10 Rph3a	Pxt1 Smad3 B4galt6 X7420426 K07Rik Creld1 Lbx1 Gad2 Mn1 Ccdc69 Pak7 Stradb Rfpl4 Fam43b Gli3 Grm2 Parp12 D6Ertd474e Reep2 Btbd2 Gpr68 Slc45a3	Omt2b Obox5 Itga9 Ptprr NA.15153 Hmces Mfsd2a Tgfb2 Plekhg1 Mcu Myo3a Gm11131 Zscan4d Bmp2k Btg4 Fyn NA.13288 Pik3cd Adcy5 Smpd3 Pld1 NA.80	Inpp4a NA.151 03 Mllt3 Mcc Slc15a5 Fam167 a Pip5k1b Bmp5 NA.150 72 Oosp1 Vil1 NA.220 7 Bcorl1 Zfp513 Plxnc1 F2r Kcnk18 Klhl8 Cby3	Stbd1 NA.135 79 Man1c1 Sh3bp1 Kit Nos1ap Mvb12b Prr5l Adm2 Igsf11 Aida Rimkla Jazf1 Tshz1 Gng3 Dpysl3 Gfod1 Tesc Oosp2 Syt11 Tmcc3	Ampd3 NA.15121 Angel2 Sipa1l1 Gm21762 NA.9588 Gm13023 Olfr288 Gm12735 H2.Q6 NA.15138 Wasf3 Polm Man2a2 Gm9125 Usp21 Tmc8 Ccdc92 Lrrc4 NA.10324 Sipa1l2 Nlrp4e	Stk36 Sytl4 Tmem92 Akt3 X9130023 H24Rik Hoxa7 Coro2b NA.15065 Ctdspl AU015836 Cngb1 NA.10579 Usp46 Cdc42se2 Gyg Igdcc3 Plag1 Arntl2 Fbxw14 Catsperg1 Itpk1	Rnf182 NA.124 07 Ptpre Zcchc2 Tcstv1 Spesp1 Ppp1r3d Grip1 Hsd17b 13 Tet3 Wdr25 Mapkbp 1 Fchsd2 Fam19a 2 Ssh1 Errfi1 Fbxw22 Ajap1 Gm2076 7

Gm6507
Th
Musk
NA. 10366
Tmcc2
Fa2h
Spry4
Tbxa2r
Rimsl
NA.4062
Papd7
NA. 14200
NA.7294
Gml 827
NA.5539
NA.3541
Uspl71b
Bmpl5
Tfap2e
Rbm38
Zdhhc8
Lztsl
Tcllb5
Slco3al
Dclk2
Tulp3
NA. 1891
NA. 15 124
Rgsl7
Zfp352
NA. 10433
Cmya5
Iqca
Tubg2
Kcnhl
X2210019I11Rik
Accsl
X2010107G23Rik
B4galt2
AC126035.1
Uspl71c
Rab3d
NA. 10463
Eif4e3
Prkaca
NA. 12521
Mmp2
Axin2
Fzd2
Cbx2
Fmnl3
Hpcall
Prrgl
Sebox
Oboxl
Zfp957
Taar2
Rassf5
Afapll2
Tmeml84b
Omt2a
Trim75
Pcdh9
Foxj2
Tmtcl
Prkdl
Ppmlh
NA.95 12
AUO 16765
Oasld
Gml775 1
Krt84
Uncl3c
Fmn2
Angptl2
X9530082P21Rik
Pdgfrl
Rasd2
Per3
Smiml4
Hipk2
Slc24a3
AA415398
St6gall
Ctdspl
Adarb2
Foxml
Adamtsll
Arhgap20os
Lingo2
Tox3
Bmp6
Fsdl
Gm21818
Tcf20
E330012B07Rik
Tob2
X4933427D06Rik
Dnah7c
Angel 1
Prlr
Ccdc6
Shb
NA.7047
Cpal
Sbkl
Zscan4c
Slcla4
Ablim2
Manse 1
NA. 15 114
Peakl
Colgalt2
Zfp30
Rapgef5
Ctif
Eif4elb
Ifitm6
Cobl
Zfp46
Ppplr9b
Mypop
Mlltl 1
Cdh4
Ccnjl
Midn
Tspan5
Gbas
Ttbkl
B4galnt4
Gml 1381
Rragc
Nrpl
AU022751
Ncehl
Lrrcl6a
Oosp3
Faml99x
Myadml
Elavl2
Plek
Spocdl
Dennd3
Lip lb
Pcdhl5
Nav2
NA. 10749
D6Ertd527e
Timd4
Efna5
Rspo2
Mamll
LsmlO
Slc6a7
Gml5668
Lrrc8a
Txndc2
Gm28784
Efcabl2
Tef
Nhsll
Glis3
Mark2
Apela
Adam33
Cacnalh
AI854703
Zfp703
Creb314
Fzd7
Mmpl9
Khdclb
Prrx2
Kmt2d
Gja3
Ramp3
Orail
Sufu
Lefl
NA. 15 19
Nav3
Gstm5
Smox
X4933404012Rik
Vps9dl
Sortl
Shank2
X4933415A04Rik
Faml 17a
Jade2
Ptcra
Dpfl
Pld6
Ets2
Elmod3
Acot3
Apol7b
Pacs2
Tmeml08
Dmwd
Ubash3b
X23 10061104Rik
Fbxw24
Ccno
Acox3
BC147527
NA.3893
Eef2k
Farpl
E330034G19Rik
Prss46
Spire 1
Nlgnl
Dbnddl
A630095E13Rik
Nr2el
Gml3 103
Lhx8
Nrep
Pla2g4c
Rasa4
AI987944
NA. 12447
Prmt2
Dact3
Magil
Gml3 191
Emilin2
Smagp
Spinl
Tbcld8
Gphn
Synm
Tmem72
Fkbp5
Clvs2
Rnf220
Platr22
B4galt4
Sgms2
Aicda
Glisl
E330021D16Rik
Oogl
Sh3rf3
Ttyh3
C330021F2
Epha3
DpplO
Slc30a3
Gm28078
Itga8
NA. 15 123
Taf9b
Plxna4
Mfsd6
Pou4fl
Fgfrll
Evl
Gdf9
Dnasell3
Shroom4
Fbxo43
Unci 3b
Scg3
Fgf7
C87499
Tubb3
NA.232
Limdl
Esytl
AF067061
Trakl
Slc22a23

Cdr2 Mfap2 Gnal2 Cntnapl NA. 10280 Mesp2 Vrtn ParplO Fam222a Pkd212 SamdlO Tbx4	Nrsn2 Trim60 Slc25a48 Snph Antxrl B020004C17Rik Derl3 Ahdcl	Ybx2 Kifl7 Lmxla Pou2f2 Ninj l Cables 1 Meis2	2 Ms4al Diras2 Pde4c Pptc7 D13Ertd608e Gml6050 Faml3 1a Obox7 Cythl Rnf26 Nobox	Prss45 Trim7 117 Sbf2 Tcf7 Ksrl Rundc3b NA. 1579 Lmol	Fbxwl8 Kpna7 NA.613 1 Tbcld2b Fhod3 Pygol Ap3m2	3Rik N4bpl Dcakd Obox2 Gramd2 Tmeml80 Prr32 Ccdc88a	

4-cell							
X1700019E08Rik Gcml Gm26815 Handl Esxl NA. 13936 Mbnl3 Tgfbl NA. 11398 Ltb X1700003E16Rik Pil6 Calm5 Tmem37 01fr836 Map7dl Tceal8 Nfatcl Wbp5 NA.7187 Tcf23	Esam Tmc5 Kcne3 Dnmt3bos Nags Zfp644 Tspan6 Gm9732 Sycpl NA.965 1 AI606181 Foxal Ccdc89 Nrg2 Eidl Rtn4r P4ha3 Cavl NA.7320 Tex 15 Rbml2 Bexl	Otopl Caapl Tc2n Kcnfl Slc38a2 Gm9918 Spata25 Myc C2cd4b Gm595 Rbm41 NA. 1261 1 Cacng7 Jakmipl NA.5 175 Zswim5 Obox8 Syne3 Lrrcl5 Iraklbpl Kcnk5 Pdlim3	NA. 15084 Eif4e Ttc30al Ccr4 Hoxb9 Tmem5 Zfp273 Nabpl Adaml9 Ythdc2 Gramdla Rnfl l AC133 10 3.1 Ctsl Crabpl Uhrf2 NA.556 Faml22b Cbfb Lpar6 Gm6871	Tmem210 Pdlim4 Lamp2 X1810034E14Rik Pcolce2 NA.55 1 Pgm211 Chicl Trim40 Rmdn2 Ddit4 Tramlll Ptprcap Epm2a H3f3b Agbl2 Igfbp3 Upk3b X6030443J06Rik Robo4	E030044B06Rik Arrdc3 Spink2 Rhoq Ddx60 Cdkn2a Psma8 Best2 Gml5 128 Dppa2 Mettl20 Ei24 Nr2c2 D930016D06Rik X4930503E14Rik Soxl5 Six4 Ramp2 NA.44	Ptdss2 Vmnlr90 Cracr2b P3h4 Gm26632 Clec2g Gml6302 Elf4 Slc25a46 Tmem47 Sowahc Mxra7 Apls3 Hfml Ccdc57 Wipfl NA. 11442 Wdr5b Plin5 Dixdcl Gml l23 Brwd3	NA.9870 Toporsl Mlfl Gm26745 X1700092M07Rik Akapl2 Cnnml Tmem63a 01fr815 Tacr2 Adamtsl4 RdhlO Pxdcl Cyr61 Prpf4b X1700123IOlRik NA. 1350 NA.9846 Unc5cl Zfp948 NA. 13261

Noto
Pet2
Nuprl
43353
Myh7
Zfp457
Nxf2
Prdml4
Dlx3
X4930502E18Rik
X1700065O20Rik
WntlOb
Bbsl2
Lrrcl9
Phyhipl
Pla2g4a
Tceal7
Siahla
Trim56
Magea8
Hesl
Btgl
Zfp239
Gml0226
P2iy4
Usp9y
Gm5930
Sox21
Selenbpl
Gm6526
NA. 15085
X1700049G17Rik
Gm53
Mycn
Gml5097
NA. 10436

NA.8609
Gml 1961
Fgr
X3110021N24Rik
X9030407P20Rik
Tbcldl2
NA. 15089
NA.7248
Abcb5
Sphkl
Hivep2
Beanl
Spsb4
NA.9430
Armcx4
Zfp758
Tnfrsfl la
NA.5916
NA. 15077
Pkdll3
Hicl
Chrnd
NA.407
Magea5
X1700019B21Rik
Pm20d2
Sec 16b
Mastl
NA. 1742
Nrxn2
Acsl4
B230219D22Rik
Gml55 18
Ptprzl
NA. 15 112
A930017

Mat2a
Gml4443
Klfl7
Lixll
Trpd5213
Gml4124
Fscnl
Platr25
Trim2
Tuba3b
Wnk3
Map7d2
Morc4
Kalrn
NA.93 16
Platr3
Cyplal
Sox30
X3222401L13Rik
Gml6185
NA.264
Gml7056
Hsdl7bl4
Tmem229b
Usp44
Crybal
Gbxl
Gm8126
Nufip2
Ubaly
Irf2bpl
Aim2
NA.4044
Ranbp6
Id4
Platr23
Spic

Gml6010
Ahil
Spaca6
Ube2e3
Xcrl
Zfp874a
Cenpq
NA.3213
Ggt7
Zfp85
Ctsk
Gm28043
Ctag2
01frl43
Mier3
Isll
Pank3
Ap4b l
Pik3c2a
Capn9
Foxfl
Tnfsfl3b
NA. 1494
Rnftl
Notch4
Gml23 15
Aebpl
Tex37
Rhox9
X4930432K21Rik
Soat2
Hesxl
Vatl
Nlrp6
Hrk
Prrtl
Zfp40
Argl

Ddias
Gml5389
Lame2
Calb2
NA.337
Mtmr6
Fam65c
Lrifl
Ehd2
Chmbl
Cpz
Prep
Slc24a4
Zfp950
Mesdcl
Zfp729a
Gm8 104
NA.539
NA. 15064
Hmhal
Wdr54
Jrkl
Pax6
Etnkl
Cebpa
Hsf3
Fzd4
Hkdcl
CldnlO
SmimlOl1
Gm26782
Zfp945
Slc26al0
Gm6268
NA. 180
Cardl4

Gm5773
Slcl2a2
Slc35f5
Lbhdl
H2afx
Arl4c
NA. 10058
FkbplO
Krt28
Set
Cbx3
Sdc3
Cyp2j6
Endog
X943002OKOlRik
Atp2c2
Gml0550
Coll7al
Wsbl
Slcl9al
Rsph9
Zfand5
Seppl
Relb
Gm2399
Atg3
Prss36
NA.222
Elovl3
Npas2
Nme5
Mysml
C130026I21Rik
NA.6224
Lrrc58
NA.7446

Amigo2
NA.5634
AC125 149.1
Ppwdl
Gm26522
Rasgefla
Zfp874b
Cyb561dl
Ttc29
Gm7334
NA. 15 101
Uaca
NA.8430
Obox6
Nanos2
X4930505A04Rik
Trpc5os
Rnpc3
A930003A15Rik
Pnn
NA.4962
Hnrnpll
NA. 186
Ctsb
NA. 10139
X4930447C04Rik
NA. 10456
Gabra4
Col5a3
Pbld2
Cd81
Lrrc46
Gm7073
Fam228b
Ctsc
Mrap

Tdpoz4
Zfp799
Nafl
NA.990 1
NA.7995
Gml0509
Gm28875
Rnd2
Nudtl6
Rsrpl
Uty
Vgf
NA. 12375
NA.2730
Unc45b
Pigw
D730003I15Rik
Gm4285
Slfn9
Edaradd
Slc5a3
L3mbtl3
Pin
Gml 1508
NA.4305

Fbnl Adgrbl Klf2 Fam212a Fgf3 Tcpl ll2 Sema6b	K l lRik NA.4501 Mbnll B3gnt8 Gm29087 Dsc3 Irf7 Plek2	Gml7404 Chadl Cede 152 01fm3 NA. 12133 Ffar4	Man2clos Gm5532 Hnrnpal Tnfrsfla E112 Dszf5	Rimklb Zfp953 Fgf4 Tenm3 Mirl7hg Ambn Btbd3 Fbln2 Per2	Bhlhb9 Mplkip Sparcll NA.7433 Cfap73 Gml416 8 Slcl6al4 Avl9 Ogn X170001 9G24Rik	Grikl Rblccl NA.7081 Dgat2 AC133 10 3.5 Lcat NA.4426	

8-cell							
NA.71 10 Cyp2d9 Ackr3 Perp Cstl3 NA.9215 Cpne3 Dok2 Cd28 Phlda3 Cartpt Cthrcl Msc Stxbp6 NA.810 Stfa211 Pdzd3 Gm27204 Anxa3 NA. 1015 Vrk2 Npy Tspanl Stard4 Lectl	Xist Arhgefl6 NA.689 Kcnv2 Fkbp9 Gas6 H60b Gm26692 Slcl2a7 Plagll Ppmlk Ppfibp2 Gml2705 Vavl NA.8401 Pla2g7 Dkkl Sbp Hsdl7bl Rragd Tmem8 1 H60c Svil Pramel5 Irf5	Lif Qpct NA.88 Nr4al Grinl Nup62cl TrmtlOb Exoc314 I830077J0 2Rik NA.7942 Hsh2d Cd300a Ptpn6 Gm6020 Siglecg Prrg3 Zfp932 Gm21060 X1010001 N08Rik Rnfl38 Sync Xkr9 Gml7655	BC052040 Ly6a Prdx6 Chmp4c X2410141 K09Rik Fbxl20 Tyms Eps812 A230083G 16Rik Prkra Gm9776 Laspl Cstf3 Akrlc21 Hoxa9 Ecell NA.4219 X9430060 I03Rik Mocos Slc6al4 Smpdl3a Nudtl l Krt7	Zfp936 NA.5874 Vpreb3 Vsxl Kctdl Ccdc84 Gstal Zfp275 Hopx NA.3556 NA.3384 Vgll4 Ptdssl NA.6297 Plcdl Gm265 1 4 NA.4998 NA.7408 Gml650 3 NA. 1047 9 Plxnb2 Slcl0a4 Salll	Slc7a7 Gml4582 Adgrg3 NA.6826 Rpl39 Nog Gm26584 Fbp2 Clcnka Gml4401 Mef2d Myo15b Cdc42ep3 NA.2700 Hhex Gml2289 Hmga2 Zfp429 Pou5fl 4335 1 AdgrG Faml98b Hprt NA.71 1 Grk6	NA. 13976 Arfip2 NA.9630 Pmaipl Gcfc2 Gml305 1 Gml9667 NA. 10925 NA.5489 Lrpapl Regl Golga7 Chordcl I122ra2 Gml l630 Ehdl Pkp2 Pdcd6 Efnal Ttc39b Cyba NA. 14015 Cd209e NA.9466 Gm205 15	NA.3445 Plekhfl Cd59a Tfcp211 Gml321 2 Parpl6 Nln NA. 1527 NA.4804 NA.3235 Esrp2 Ly96 X903062 4J02Rik NA.3453 Mfs d8 Slc45a4 Urgcp Igbpl Lgals8 NA.4193 Atp6v0e 2 Chptl NA.588

Gyltllb
Nxpe5
Dynap
Gml5446
Zfp934
PlatrlO
Amot
Id3
Amotl2
Gm26740
Abcbla
Diaph2
Akrlcl4
Ciyab
1133
Slcl9a2
Epasl
NA. 1618
Pcdhbl6
Bex4
Tmem64
Bmp8b
Gml0139
Gpc4
Vnnl
Rbmsl
Apob
X9330185 C12Rik
Camk4
NA.559
Mpped2
Poflb
Papss2
Tbx20
Gng2
Nr2f2
Rarb
Gml0772

Deaf 1211
Gm413 1
X4930550 L24Rik
Zfp52
NA.3646
X4930522 L 14Rik
Slco2al
Gm26836
Ap3b2
NA.41 12
NA. 10665
Tmem245
Pik3r6
Tsix
Hsdl7bl 1
Zfp354a
Gml 11O
Bves
Xlr
AI467606
Mtml
Ccngl
Arhgdib
Faml24a
Slc52a3
Gml3 154
Suox
NA.2957
Fgfl3
Parva
Casc4
X9230009 I02Rik
F12
X22 10404 O09Rik
SlOOal 1
X5430403

Eno2
Amph
Cede 150
Cdc42epl
NA.4813
Eda2r
Hes2
Etl4
Vangll
Atp8b4
Cav2
Slc29a3
Nradd
Tmem253
NA. 1630
Ddahl
Ano9
Acp5
B2303 12C 02Rik
LITC23
Cux2
NA.9543
Gm6712
NA.7720
Faml29a
NA.2889
Gml0324
Slc29a4
NA.2540
Gml25 14
Cd53
Msmol
Rampl
Postn
Havcrl
Ttpa
Gjb3
Ahsg

NA.5 168
Oimdll
NA.4188
Hspa8
Rassf7
Star
Pkdlll
D930020B 18Rik
Arhgapl8
Ppp2r2c
Denndlb
BC05 1665
Diiall
Klf8
Gml3235
B4galtl
NA.5 135
NA. 1892
Ckslbrt
LiTc37a
Krt27
Wnt3a
Smocl
Igsfl
NA.5696
Kcnh7
Gm 13242
Sema5b
NA.9923
NA.5 13
GrM3
Lparl
NA.3947
Isl2
Fes
Nap 112
Sh3glb2
Nck2

NA. 1214 8
C3arl
Gml306 2
Fndc3cl
Dpyl912
Ano2
NA. 1390 0
Iqgap3
Sh3d21
43160
Akp3
Glt28d2
Gm
NA.2621
Cwh43
NA.7337
Sh3tcl
Pinlrtl
C030039 L03Rik
Caldl
Akap2
I113ral
NA.9845
Sbpl
NA. 1027
NA.3 116
Alcam
NA. 1390 6
Immt
Card 11
Asap2
Smim22
Sycn
Ak7
Nprl2

Atp2bl
Satl
Fam2 17b
Etohil
G430049 J08Rik
Fam83b
Pde7a
NA.4566
Cldn4
Foxf2
Pank4
B930036 NIORik
NA.7030
Gm26668
Gabrd
Tbx3
X943000 2A10Rik
Ctsf
NA.6
Gm27206
Rnf208
Bhmt2
NA.293 1
NA.691
Adam21
Serine 1
NA. 1264 9
Mybpc2
Runx1
Vtn
Fancb
KlflO
Gm26624
NA. 1030 3
NA.7385

A530040E 14Rik
NA.443 1
Rnf32
Ly6g6e
Ldbl
Gml l541
Gm2366
Prrl9
Cmtm5
Tmem45a
NA.9621
NA.336
Gml0687
Zfp4 18
Gml976
NA. 1763
NA.7085
Acyp2
Oxctl
Pigz
Tpd52
NA.47
Mllt6
Plcgl
Pnpla2
Gml5 137
Dnajc6
X2410018 L13Rik
Actnl
NA.223
Rbks
Nrtn
Fut9
Ednrb
Zfp458
Itpkb
NA. 11397

Adam4
Zfp607
Atp6v0a 4
Arhgap2 7
Cdh1
Il17re
NA.3823
NA.4035
NA.4009
Lpin1
Atg4c
Alg13
Rad23a
Gm2653 8
Prr151
NA.7290
Upf3b
Slco4c1
NA.5912
Emilin1
NA.5335
Tmem14 4
Zfp599

Zfp157	G16Rik Steap3 Matn3 Slc22al3 Fgd4	Strada Reepl Ncf2 NA.4991	Gata6 Slc36a3os NA. 14579 Bok	Zfp422 Alg6 Npnt NA.424 Psrcl Sfrpl Ace2	NA.487 NA.2929 Rdh5 NA.5637 Vps33b	NA. 1522 NA.991 1 NA.2756	

16-cell							
Gm2245 Fabp5 Gml7067 Apoal Stat6 Capn6 Abcal Gml4305 Eomes Zfp3611 Sox2 Sh3bp5 Ptgdr As3mt Pmaipl Dokl Slc37a2 Tinagll Aldhlb1 Mafb Lypd8 BC048679 Gml4412 Otx2 NA. 186	H2afy Rhob Trip6 Tmsb4x Slc6al3 Plk5 Col4al Shkbpl Mgst2 Cdcl23 Dsg2 Mpzl2 Glrx Frrsll Gss Hebpl Sox7 Cbx4 Fbxo3 Pnma2 Fam92a Ddx3y Wfdc2 Msx2 X5730507COlRik Herpudl Hspalb AdamtslO	Khdc3 X4930558J18Rik Gml4409 Top2b Ank2 NudtlO Pvrll Anxa9 Hal Slc2al Acaa2 Lyrm9 Slprl Pgapl E130012A19Rik Xbpl Zcchcl6 Mapt Arl6ip5 Pou2fl Cited4 Tbxl Zip119b X1700086P04Rik Cstal Efnbl Hemkl	Tbca Mycl Phlppl Sqstml Hbegf Serpinb6a Acpl Nanog Reml Sppl Texl9.2 Pdzklipl X1700095A21Rik Camkl Gml4327 Bend7 Alg8 Napll3 Vpsl3c Epcam Dpysl4 Fas Tgfbr2 Dmcl Ctgf Sult4al Zfp459 Zfp688	Erlec 1 Slc7al5 Vcpkmt Trim47 Bcl91 Evpl Actgl AU021092 Cdkl8 Dok2 Cldn23 Nsmaf Cpxml Impadl Crip2 Lamcl NA.61 14 Eps811 Camk2d Alcam Assl Mospd2 Lip 11 Trim21 Slc24a5 Csf3r 43352 Lrrc75b NA. 13 142	Adam9 Pomtl Gjb3 Acad 12 Tmeml35 X2610528Jl lRik BC029214 Them5 Atp8al Psmg2 Sik2 Wnt6 Bre Elf3 Pigz Itga7 Lrmp Vapb Bhlhal5 Gml0605 Hsp90aal Nsdhl Sdcbp2 Faml32a X2700068H02Rik Kbtbdl3	NA. 12986 Egfl7 Ormdll B3gnt3 BC052040 Paqr5 Pfn2 Gml4403 Vmn2r29 Gstpl Gml7087 Slc5a2 NA.73 16 Npcl Pmsl Sccpdh Spcs3 NA.499 Slc4a2 Gatadl Atp2a3 Fancb Rac3 Mthfsd Acadvl NA. 10404 Tfcp211	Nipal Tppl Gm4673 Slc35al NA.5230 Hdac3 Whamm Gpx2 Trappc 1 Tmeml98 NA.4039 X3110052M02Rik Adprh Thrsp NA. 10775 Gm26578 NA.385 1 Aasdhppt Pkp2 Plgrkt NA. 14210 Itm2b Duspl 1 Lgals9 Sdhaf4 Emp2 Idhl

6	Mdhl	X4930522L14Rik	Cgrefl	Map2k3os	NA. 102	NA. 1896	Zfp850
Oxt	Rhoc	Hormad2	NA.92	Prkce	Gimap9	X101000lB22Rik	Txndcl7
BC05 1142	Ier2	Cd82	Naal l	X4930563D23Rik	Gm4262	Erf	Apeh
Kcnn4	Slfn3	Map3kl	NA.388	Ank	NA.6479	Slc28a3	Gml0439
Zfp93 1	Zfp759	X I500009L16Rik	Tdrp	Dact2	Ralb	Junb	NA. 1925
Pletl	B3galt2	Phfl 1d	Pcbdl	Pacsin3	Tmeml7	Zfpl l9a	Cnn3
Ppl	Laccl	NA. 13623	Slco2al	Hmcn2	NA. 1999	Perp	Mmpl5
Chpf	Tnsl	Trim38	Cyb5rl	Eef2kmt	Leprot	NA.369	Cxcr6
Tspan3	Tmem45b	Vps29	Magea2	Chchd7	Ube2q2	Calcoco2	Foxb2
Hyal2	Tapl	Tbllx	Prokrl	Zfp248	Lmfl	Gm28085	Lama5
Fstl3	Slc38a4	Lsr	Mbnl2	NA. 10780	Tmeml47	Clqa	X1700080O16Rik
Slfn2	D10Jhu81e	I117rc	Mex3b	Clecl 1a	Sh3bgrl3	NA. 1618	Gml6136
Dusp6	Srxnl	Aqp3	Gml6712	Sgpll	Tradd	Zfp81	Asap3
Cat	Spata9	Zfp429	Zfp395	Xlr3a	111 Orb	Ntf5	Syngr4
Nppb	Pmepal	Ggtl	Krt8	Msc	X1700086O06Rik	Oaslg	Zdhhcl5
Tpcn2	Gm26853	Tcea2	Tceall	Zfp442	Sdhaf3	Appl2	Fam83b
Cede 169	Pfkfb4	Gm5 141	Gata3	Gml4418	Galnt9	Gnal5	Rnase4
Elovl5	Zfp266	Tmem5 1	Serinc2	Usp25	Ogdhl	Gm6169	Fbxl21
NA. 12239	Cdc42ep5	Stx7	Rgsl4	Ntpcr	Pearl	Cmal	Hdx
Zfp326	Magea3	A530017D24Rik	Mocsl	Prosl	Fezfl	Lrrn2	
AI3 17395	Chrna3	X1700003M07Rik	Tmeml3 1	Lpp	Svbp	Acot6	
AA467197	Gm26624	Ladl	Vps45	Trp53il l	Larplb	Dmrta2	
NA. 11335	Elovl7	Hint2	Plpp2	X2610008E1 lRik	A730015C16Rik	Skidal	
Ptges	Nkx6.2	Exph5	Mogat2	Akrlel	Gm26779	Ccngl	
Smiml	Ctam	Sfrpl	NA. 12035	Pla2g7	Cryzll	Trabd	
Kirrel	Nfkbiz	Hspel	B2301 18H07Rik	NA.4703	Stl4	X2410022Ml lRik	
Gbp9	Cyp4fl4	X9430065F17Rik	Serpinb6c	Gmpr2	Egr4	Tet2	
Ckap4	Tnfrsflb	Ahcy	Fos	StardlO	Hmgal.rs1	Cetn3	
Naps a	Dsp	Magee2	P2ry2	Enpep	Lcpl	Sri	
Gjb5	Khnyn	Mageb4	Lgal s4	Prss35	Hadh	Vill	
Clic3	Rndl	Gm7325	Epb4111	NA.2001	Sec 1414	Msantd4	
Marcks	Hnf4a	Tmem266	Snrk	Eml2	Txndcl2	Abhdl4a	
NA.7249	Adat2	Txnl	X2410018L13Rik	Ghdc	NA.7425	Gm413 1	
Scd2	X2200002DOlRik	Rec8	Rims4	X2610301B20Rik	Histlh2bc	Pnpla6	
	Gabarapll	Tgm2	Gchfr	Pdzd3	P2rx3	NA.413 1	
	NA. 12352		Nrgl	Gm5424		Smapl	
	Shc2		Skil			Lysmd2	

Adgre5 Faml29 b Pycr2 Dcafl21 1 Barx2 I14ra		Xkr6 Egln3 Man2al			Arhgef5 Sfmbt2 Btg2 Ndufc2	Xrcc4	

32-cell							
Lrp2 Fhl2 Capn2 Sppl BC0533 93 Hspb8 Cdx2 Krtl8 Enpep Elf3 Vgll3 Wnt7b Akrlb8 C2cd4a Bglap3 Rab17 Serpinb 9b Bmyc Cmbl Klf6 Krt8 Nppb Tppl Tmem9 Dppal Rhox5 Gm5424	Ezr Fam213b Xbpl CeacamlO NA.5461 Msn Frmd4b Glrx Gapdh Gstpl Seφ iub6c Epb4111 NA. 123 12 Lgalsl Ptges D10Jhu81e StardlO Apoal Cela2a Tuba4a H2.K1 Hint2 Cubn Rnfl28 Dusp4 Ogdhl X1500009L 16Rik Tet2	Oc90 Mapre3 Gm364 Gstol Nanog Eml2 Lsr Stl4 Nfic B2301 18H 07Rik Gm6169 Gm7325 Gm26917 Zfp93 1 Rp2 Tat Epcam Rnfl30 Gml4403 Tmeml39 Pycr2 Plscrl Mfi2 Adad2 Dsp Mbp Chrnb4 Tfcp211	Ptprn Gpr4 Ptgrl 43352 Nrl Optn Slc25al3 Dqxl Gm26579 Tmeml25 Cmip Gml4325 Dtd2 Tspan3 Srxnl Huslb Slc6al3 Adaml5 Vill Sult6bl Mecp2 Tarml Camkl Mgl2 Chstl3 Myhl3 Barx2 X1810030 O07Rik	Baiap211 Cdc42ep5 Etfb Gml2169 Mdhl Pletl Wdrl Zfp37 Histlh3c H2afy NA. 148 X1700042 G15Rik Adrb3 Gml4399 Fthll7a H2.D1 Cat NA. 1550 Fgfbpl Lgals4 Trim50 Prkcdbp Tφπι6 NA. 1546 Cidea Nagk Slc38a4 Serinc2	Plod2 Phfl 1d Pdgfa SlOOalO Tpm4 Pgm2 Gml4326 Xrcc5 Esd Actr3b D630003 M21Rik Ppplrl4d Mkrn3 Adgrl2 NA. 101 14 Sox6 Tnsl Emp2 Col4al Ndrgl Dap3 Capzb Fhl4 Wfdc2 Anp32a X23 10015 AlORik Hist3h2a	Tcn2 Rnaset2b Aldh2 Dab2ip Actb Cck Efhd2 Pank4 Arvcf Gml4327 Wdr6 Abcg2 Mgstl Aldh3a2 Omd Chrnal Tdpl Sgpll Ttf2 Faml29b Emc9 Tmeml7 NA. 102 Vps29 AU02109 2 Pard6g Kcnkl2 X8030474	Fez2 Rap2 b Prkce Gm2 381 Gucy lb2 NA.7 242 Histl hie Gmpr Pla2g 6 NA.2 972 NA.7 262 Anxa 6 Fthll 7e Cdc4 2ep3 Tradd Sccpd h Xlr3b Figla NA.1 4180 Dap

Id2
Gjb5
Nek6
Oasla
Scd2
Atp12a
Gstp2
Ngfa p1
Pycard
Pafah2
Cstal
Fam213a
Binl
Gm694
Dsg2
Assl
Gm4737
Slc38al
Slc38al1
Camk2d
Bex2
Sdc4
Rfx4
NA.7440
Tinagll
Col7al
Kng2
Adgre5
Tnfrsf9
Mmell
Lgals9
Texl9.2
Gata3
Atxn711
Txndc12
Clcnkb

Chmp2b
Lama3
Fbxo3
Elovl7
Patl2
Cede13
Col4a2
Acaa2
Acaala
Apbblip
Tmx4
Snai2
AI662270
Sox9
Tes
Trim38
Cryz
Anxa2
Sft2d2
NA.388
X2610528J11Rik
Gsn
Hadh
X0610009O20Rik
Plp2
Abcc4
Lcpl
Actgl
Fam25c
Xk
NA.92
Fabp3.ps1
Ube216
Nsmaf
Cited4
Fabp3
As3mt

Exph5
Rcanl
X9530059014Rik
Eef2kmt
Mucl
Efcab5
Nyniin
Gm26603
N1ф 4c
Susd2
Tst
Khdc3
Plbl
NA.5999
Tdф
Gale
Gml4322
Cpxml
Tmprssl2
SlOOal1
Hoxd3osl
A230005M16Rik
Hnf4a
Histlh3d
Bdnf
Ppp4rl
Lta4h
Dpysl4
Tmeml02
Trhr2
Tbllx
Kremen2
D130040H23Rik
Cyp4f39
Tmem266
NA.5910

Ccdc43
Ppmlm
Slc24a5
Xlr3a
Tmeml98
BC051019
Erbb2
Cnpy2
Idh3a
Dab2
Mksl
Gimap9
NA. 1892
Hk2
Marcks
Gm773
NA.83
Cdk5
Gstm6
AtxnlO
Smco2
Enolb
Pir
Gpx2
Csf3r
Atg4c
Uhrfl
Clic3
Gstm7
Coasy
Tmem256
NA.529
Tmem45b
Krt23
Mpzl2
Sqstml
Zfp780b

Rgsl4
Tpil
Gstzl
Ggtl
Insig2
Ly6a
X2310039H08Rik
NA.2957
Carl2
F2rll
Zfp454
Eci3
Gjb3
Ly6f
Pηlipф 2
Praf2
Gml4393
Abcb8
Mras
Gml4444
Bckdhb
NA.9436
Tbxl5
Acsf2
Slcl8al
Hdx
Apocl
Seф iub6a
Zyx
Rec8
Ppplrl8
Cyb5a
Fblnl
Dpyl911
Tpml
Gdf3
Map2k6

Slc37a2
Gml4418
Hsdl7b4
Sergef
Psme2b
Ill lral
Tpcn2
Sh3bgrl2
Asic3
Lurapll
Plau
Fam83h
Tф 53i11
AA467197
Gml4409
NA.513
Mettl7al
Clic4
Acol
Sh3bp5
NA. 1866
Gm4779
Cbr4
NA.6249
MyhlO
Crip2
Psmb9
Gm4926
I117rc
Sdhaf4
Dokl
Slc25a39
Ccdc42
Aф 8aї
Echsl
Akrlel
Nudtl 1
Gcat

K03Rik
Atplbl
A330050F15Rik
Hdac3
Ftx
Fthll7d
NA.4386
Arl2
Apeh
Slc2al2
Zfp850
Iftl40
Slc2al
Prkx
X1700086O06Rik
Cox7b
Faml36a
Pwwp2b
Cyb5r3
Mapt
Vpsl3c
Abcal
Hibch
Micall
Adat2
Lpp
Srebfl
Arhgap9
NA. 14050
Tctnl
Tubalb
Whamm
Smyd4
Cbfa2t3
Arhgef25
Nbll
Mgat4b

Hspd1
Efcab10
Tubb2a
Gprc5d
Smim12
Mtmr7
Gsta3
Skida1
Idh1
Hlf
Tcea3
Znrd1as
Pkm
Map3k15
Ak4
Gm12828
Myo1e
Slc4a5
Slc2a3
Sdr42e1
Slc7a6
Snx19
Nduf af3
Plin2
Gipc1
Pla2g4f

Trp53bp 2		Gss				Adh4	

[0251] In a nutshell, and further discussed below, we identified notable features within the landscape, including sets of cells classified as pluripotent-, epithelial-, trophoblast-, neural-, and stromal-like based on strong expression of signatures related to these cell types and a set of cells (FIG. 24E, purple) that appeared poised to undergo a mesenchymal-to-epithelial transition (MET) following withdrawal of dox (FIG. 24E, orange). The relative proportions of these subsets at different times differed between serum and 2i conditions (FIG. 24G).

[0252] Using Waddington-OT, we calculated the ancestor and descendant distributions for all cells and determined the trajectories to/from various cell sets (FIG. 24F, arrows). Briefly, the time course began with MEFs at day 0 in the lower right, proceeded leftward to day 2, and then upward over the subsequent week toward two destinations: the MET Region and the Stromal Region. The cells in the MET Region were predicted to give rise to the pluripotent-, epithelial-, trophoblast-, and neural-like cells, with this last class seen in serum but not 2i conditions. By contrast, the Stromal Region appeared to be terminal: cells entered the region, but our model predicted that they did not leave (FIG. 3IE).

[0253] The optimal-transport analysis provided insights into when cell fates emerged. As early as 1.5 days, cells' fates began to concentrate toward either the MET Region or Stromal Region, and the distinction sharpened over the next several days (FIG. 25G). The fate of pluripotent-, epithelial-, trophoblast-, and neural-like cells did not appear to be determined until after withdrawal of dox on day 8. That was, the ancestor distributions of these cell types were indistinguishable on and before day 8.

[0254] The model was predictive and robust

[0255] Before analyzing the cell sets and trajectories in greater detail, we assessed the accuracy and robustness of our model. Because current experimental approaches for tracing cell lineage did not provide a rich description of the full transcriptional state of a cell set's ancestors, we developed a computational approach to test the model. Specifically, we used optimal transport between the distribution of cells at times t1 and t3 to predict the distribution of cells at an intermediate time t2 and compared this prediction to the observed distribution at t2.

[0256] Our predicted trajectories were accurate, such that the distance between the computational prediction and experimental observation at t2 was similar in magnitude to the distance between the two experimental replicates taken at t2, confirming that the prediction is roughly as good as could be expected given experimental variation (FIG. 24H, FIGs. 30A-30G, Methods).

[0257] The optimal-transport analysis was also robust to perturbations of the data and parameter settings. We down-sampled the number of cells at each time point, down-sampled the number of reads in each cell, perturbed our initial estimates for cellular growth and death rates, and perturbed the parameters for entropic regularization and unbalanced transport. In all cases, we found that the interpolation results above are stable across wide range of perturbations (STAR Methods).

[0258] In initial stages of reprogramming, cells progressed toward stromal or MET fates

[0259] Reprogramming began with all cells exhibiting rapid changes. By day 1, cells showed an increase in cell-cycle signatures and a decrease in MEF identity. MEF identity continued to fall through day 3, by which point nearly all cells showed lower signatures than the vast majority of MEFs at day 0 (FIG. 24D). Over time, cells assumed either Stromal or MET identities (FIGs. 25A-25H).

[0260] Cells in the Stromal Region showed distinctive signatures, which fully emerged after withdrawal of dox at day 8; these signatures included a secretory phenotype (SASP), extracellular matrix (ECM) rearrangement, senescence, and cell cycle inhibitors (FIG. 25A). By contrast, the MET Region contained cells with increased proliferation and loss of fibroblast identity (FIG. 25E).

[0261] Mapping signatures of distinct stromal cell types obtained across mouse tissues from a mouse cell atlas (Han et al., 2018) showed that the most widely expressed stromal signatures corresponded to embryonic mesenchyme and long-term cultured MEFs (FIG. 31A). Yet, the Stromal Region did not simply reflect "MEF reversion." The gene expression profiles were distinct from (FIG. 3IF) and more heterogeneous than day 0 MEFs, with clusters of cells with signatures that more closely correspond to other stromal cell types, such as those found in neonatal muscle and neonatal skin (p-values < 0.01) at levels 20- to 30-fold higher than day 0 MEFs.

[0262] The proportion of stromal cells peaks several days after dox withdrawal (at -64% of cells at day 10.5 in 2i conditions and day 11 in serum conditions) and then declines through day 18, consistent with the low proliferation signature relative to other cells in the landscape (FIG. 24G). A subset of stromal cells expresses an apoptosis signature starting on day 9, which peaks at day 14.5 in -14% of stromal cells in serum conditions and at day 13 in -3% in 2i conditions.

[0263] Our trajectory analysis allowed us to trace how these fates were gradually established: we found that the ancestor distributions of cells in the Stromal and MET Regions differred by 30%» at day 3 and by 60%> at day 6 (FIG. 25H). A powerful predictor of a cell's fate was its expression level of the OKSM transgene, with high values predictive of MET fate and low values predictive of stromal fate (FIG. 31C); the expression level statistically explained ~50%> of the variance in the logarithm of the fate ratio (MET Region fate probability divided by Stromal Region fate probability) by day 2 and ~75%> by day 5 (FIG. 31C). Importantly, the divergence was gradual and could not be described by a simple graph with a sharp (that was, zero-dimensional) branch point. Indeed, our optimal-transport analysis indicated that a significant minority of cells that were on the trajectory to the MET region continues to switch to the trajectory to the Stromal Region (FIG. 25G).

[0264] Regulatory analysis identified TFs associated with the two trajectories. Three TFs (Dmrtc2, Zic3, and Pou3fl) were induced in all cells (from undetectable levels at day 0), but showed higher expression along the trajectory to the MET Region (FIG. 25E, 25F). Zic3 was required for maintenance of pluripotency (Lim et al., 2007), Pou3fl was required for self-renewal of spermatogonial stem cells (Wu et al., 2010), and Dmrtc2 was involved in germ cell development (Gegenschatz-Schmid et al., 2017; Yamamizu et al., 2016). Four TFs (Id3, Nfix, Nfic, and Prrxl) were upregulated in all cells (from basal levels at day 0) but showed higher expression in cells with a stromal fate (FIGs. 25E, 25F). (Analysis of subsequent time points showed that, following withdrawal of dox, these genes maintained high expression in stromal cells but shut off in cells along the trajectory to iPSCs.) Nfix was reported to repress embryonic expression programs in early development, while Nfic and Prrxl were associated with mesenchymal programs (Froidure et al., 2016; Messina et al., 2010; Ocana et al., 2012). Id3 was known to inhibit transcription through formation of nonfunctional dimers that were incapable of binding to DNA. Higher expression of Id3 along the trajectory toward stromal cells may seem somewhat surprising, because forced expression of Id3 was shown to increase reprogramming efficiency (Hayashi et al., 2016; Liu et al., 2015). However, Id3 might cause increased efficiency via its activity in stromal cells, which secreted factors that enhance iPSC reprogramming (Mosteiro et al., 2016) (see below), or via activity in non-stromal cells, in which it was expressed through day 8, albeit at lower levels.

[0265] There has been much interest in finding early markers of successful reprogramming—namely, genes whose early expression was correlated with a cell's descendants being enriched for iPSCs. Our analysis suggested that it would be more precise to define "early markers of successful MET", because the iPSC, trophoblast and neural fates did not appear to be established until after withdrawal of dox at day 8.

[0266] Trajectory analysis revealed early markers of successful MET, including known markers such as Fut9 (which synthesizes the glyco-antigen SSEA-1) and novel candidates such as Shisa8. Shisa8 was the most differentially expressed gene at day 1.5. When we sorted cells based on the ratio of their likelihood of transition to the MET Region vs Stromal Region, we found Shisa8 expressed in 50% of the top quartile but only 5% of cells in the bottom quartile. (Table 16). Shisa8 was a little-studied mammalian-specific member of the Shisa gene family in vertebrates, which encoded single-transmembrane proteins that played roles in development and are thought to serve as adaptor proteins (Pei and Grishin, 2012; Polo et al., 2012). (Analysis of subsequent time points showed that Shisa8 and Fut9 also showed similar patterns following dox withdrawal: both were expressed strongly in cells along the trajectory toward successful reprogramming, and lowly expressed in other lineages (FIG. 31D).)

Table 16 - Differential genes between top ancestors of MET vs. top ancestors of stromal cells.

Differential genes between top ancestors of MET vs. Stromal cells at D1.5					
Gene	**p-value**	**Average logFC**	**Fraction expressed in top ancestors of MET**	**Fraction expressed in top ancestors of stromal cells**	**Adjusted p-value**
Shisa8	2.37E-56	0.439583976	0.505	0.051	4.52E-52
Anpep	1.24E-44	0.399501581	0.548	0.141	2.37E-40
Gch1	5.09E-37	0.381008072	0.607	0.245	9.71E-33
Gpm6b	1.24E-29	0.275486032	0.538	0.209	2.37E-25

Npnt	3.61E-30	0.382743398	0.714	0.395	6.89E-26
Dsp	9.36E-34	0.290320422	0.389	0.072	1.79E-29
Rbl	1.12E-25	0.280506707	0.616	0.315	2.13E-21
Dgat2	5.18E-28	0.349298687	0.524	0.225	9.88E-24
Carl2	1.06E-23	0.299588702	0.552	0.254	2.02E-19
Lrp4	9.73E-27	0.247967802	0.405	0.11	1.86E-22
Clql3	2.93E-26	0.325323868	0.45	0.155	5.60E-22
Sgol2a	1.65E-25	0.33023125	0.685	0.395	3.16E-21
Gm26737	2.93E-25	0.534938533	0.656	0.368	5.59E-21
Lepr	1.15E-22	0.588193067	0.695	0.417	2.19E-18
Nol4l	1.78E-21	0.374175462	0.65	0.374	3.40E-17
Gm29666	1.49E-20	0.279383915	0.511	0.237	2.84E-16
Pfkp	8.34E-30	0.316216243	0.796	0.524	1.59E-25
RP23-4H17.3	4.98E-21	0.441940336	0.695	0.425	9.51E-17
Ralgps2	4.40E-22	0.217741022	0.38	0.117	8.40E-18
Xafl	1.12E-18	0.328905337	0.564	0.307	2.14E-14
Zdhhc2	2.08E-17	0.200585787	0.519	0.264	3.97E-13
Ppmlk	1.38E-22	0.307219164	0.658	0.411	2.63E-18
McmlO	1.99E-16	0.230302782	0.593	0.348	3.80E-12
Gml3075	1.33E-27	0.861118262	0.771	0.528	2.53E-23
Repl5	2.80E-18	0.29626083	0.658	0.423	5.34E-14
Pola2	3.37E-23	0.311939681	0.748	0.519	6.44E-19
Trim37	7.52E-17	0.218079056	0.583	0.358	1.44E-12
Rtkn	3.27E-18	0.287996995	0.382	0.16	6.24E-14
Ppif	1.58E-21	0.252798031	0.767	0.548	3.02E-17
Rsfl	2.84E-15	0.229977128	0.591	0.374	5.42E-11
Ptcra	5.85E-13	0.417578437	0.413	0.2	1.12E-08
Nmrkl	4.51E-13	0.528279491	0.554	0.344	8.61E-09
Perp	4.55E-65	0.656396496	0.963	0.753	8.69E-61
Chmp2b	1.29E-30	0.335057338	0.849	0.64	2.46E-26
Pcgf2	5.58E-15	0.541239697	0.591	0.387	1.07E-10
Gmcll	4.30E-14	0.523834071	0.544	0.344	8.21E-10
Pacsl	1.50E-18	0.251074727	0.785	0.587	2.87E-14
Wdr35	3.75E-14	0.224471336	0.656	0.464	7.15E-10
Ppat	2.16E-16	0.243243284	0.708	0.517	4.13E-12
Slamfl	5.19E-11	0.228267013	0.468	0.28	9.90E-07
Homer2	6.66E-14	0.236094482	0.624	0.438	1.27E-09

Cenph	7.86E-14	0.206088745	0.72	0.538	1.50E-09
B930036N1ORik	2.34E-10	0.518225771	0.544	0.368	4.46E-06
Hpcall	8.65E-13	0.208476389	0.613	0.438	1.65E-08
H2-T23	8.64E-11	0.235054556	0.337	0.164	1.65E-06
Sgoll	2.01E-16	0.266408936	0.853	0.683	3.83E-12
Ccdcl37	2.58E-20	0.287870449	0.793	0.624	4.93E-16
Exosc2	9.42E-37	0.652481854	0.933	0.765	1.80E-32
Gkapl	1.74E-23	0.397791708	0.781	0.613	3.31E-19
Agl	1.58E-16	0.495744367	0.798	0.63	3.01E-12
Ckap2	8.06E-12	0.205735226	0.796	0.632	1.54E-07
Nt5dc3	1.29E-10	0.200909668	0.638	0.481	2.46E-06
Tapbpl	7.86E-09	0.226071905	0.315	0.164	0.000150089
Shoc2	9.21E-15	0.231434184	0.751	0.601	1.76E-10
Faap24	3.98E-11	0.2159197	0.642	0.495	7.60E-07
Haus8	2.63E-16	0.634579918	0.744	0.599	5.01E-12
Cenpf	7.61E-11	0.214446511	0.908	0.763	1.45E-06
Mrpsll	3.66E-41	0.430516438	0.906	0.763	6.99E-37
Aldh3al	8.14E-08	0.221022512	0.456	0.313	0.001554728
Gm7120	8.12E-08	0.306764672	0.311	0.168	0.001550761
Lpgatl	4.28E-16	0.244225687	0.806	0.665	8.17E-12
Topbpl	5.86E-12	0.224664357	0.734	0.593	1.12E-07
Mrps6	3.39E-43	0.396132536	0.939	0.798	6.47E-39
1700047117Rik2	5.69E-09	0.200128893	0.521	0.382	0.000108639
Myc	4.08E-26	0.347729368	0.898	0.763	7.80E-22
TimmlO	4.34E-14	0.223178202	0.845	0.71	8.28E-10
Mrpl9	9.74E-09	0.222293218	0.503	0.368	0.000185972
Famll4a2	2.19E-18	0.23879583	0.83	0.697	4.18E-14
Rm3	1.49E-11	0.228168673	0.724	0.591	2.84E-07
Dcafl7	2.63E-08	0.521823548	0.487	0.354	0.00050265
Asph	2.31E-14	0.224904909	0.787	0.656	4.42E-10
Abcblb	6.60E-40	0.441369564	0.947	0.818	1.26E-35
Ctnnbll	2.19E-11	0.207192935	0.777	0.648	4.18E-07
Slbp	1.84E-15	0.374861946	0.873	0.748	3.52E-11
TexlO	3.22E-15	0.251420666	0.8	0.677	6.14E-11
Dennd5b	3.94E-11	0.298384346	0.755	0.632	7.52E-07
Lrrc42	3.19E-14	0.250507008	0.748	0.626	6.09E-10

Paip2b	6.60E-09	0.233070859	0.691	0.571	0.000126059
1700037H04Rik	3.73E-13	0.21591323	0.777	0.663	7.12E-09
Noal	1.13E-34	0.490924229	0.9	0.787	2.17E-30
Gtf2hl	5.71E-19	0.253937461	0.843	0.738	1.09E-14
Ndcl	4.28E-18	0.25208573	0.89	0.785	8.16E-14
Ddx42	1.64E-13	0.213024231	0.83	0.726	3.13E-09
Golga3	9.43E-07	0.495832978	0.595	0.491	0.018003133
Pop5	1.28E-28	0.301595886	0.949	0.847	2.44E-24
Tgfbi	1.63E-09	0.200070657	0.828	0.726	3.11E-05
Hells	3.70E-13	0.222587886	0.949	0.851	7.06E-09
Plk4	1.42E-23	0.57479234	0.922	0.826	2.72E-19
Ezh2	1.90E-18	0.236909466	0.906	0.81	3.64E-14
Naa20	8.41E-18	0.270587809	0.806	0.714	1.61E-13
Epnl	1.54E-14	0.209191303	0.902	0.812	2.94E-10
Smnl	9.92E-38	0.401700379	0.941	0.853	1.89E-33
Mcm7	1.42E-16	0.229113377	0.955	0.867	2.72E-12
Enah	1.19E-12	0.207086155	0.828	0.742	2.27E-08
Mrps25	2.24E-16	0.238478878	0.863	0.783	4.27E-12
Carnmtl	7.08E-15	0.213768504	0.871	0.791	1.35E-10
Zfpl06	4.55E-12	0.206955912	0.943	0.863	8.69E-08
Hmgb3	4.37E-16	0.244565953	0.879	0.802	8.34E-12
PsmblO	8.45E-25	0.305887579	0.937	0.861	1.61E-20
Scp2	7.16E-12	0.211532788	0.883	0.808	1.37E-07
Histlh2ap	1.60E-27	0.599321987	0.978	0.904	3.05E-23
Limk2	1.79E-12	0.34639987	0.81	0.738	3.42E-08
Dbf4	5.21E-15	0.209332579	0.922	0.851	9.95E-11
Bazla	2.09E-20	0.276857187	0.881	0.812	4.00E-16
Ifrd2	4.47E-21	0.25780276	0.908	0.84	8.53E-17
Ccdc50	1.00E-25	0.293196782	0.955	0.888	1.92E-21
Pbdcl	3.94E-14	0.228782894	0.875	0.808	7.52E-10
Wdr45b	8.91E-11	0.203638926	0.832	0.769	1.70E-06
Noc2l	8.02E-21	0.235002625	0.951	0.89	1.53E-16
Ruvbll	3.88E-11	0.20097654	0.828	0.767	7.41E-07
Prmt5	1.96E-13	0.20762784	0.888	0.832	3.74E-09
Tmem245	1.26E-32	0.731436804	0.963	0.908	2.40E-28
Pnol	1.18E-22	0.284205102	0.894	0.84	2.25E-18
Chchd7	1.97E-33	0.376522958	0.92	0.867	3.76E-29

Yiflb	2.51E-12	0.204286063	0.91	0.857	4.80E-08
Nip7	1.61E-09	0.317643192	0.896	0.843	3.07E-05
Stmnl	7.91E-13	0.214767905	0.926	0.875	1.51E-08
Rtcb	3.23E-21	0.248019171	0.933	0.885	6.16E-17
Nmt2	9.69E-54	0.59549564	0.988	0.941	1.85E-49
Fnta	2.30E-11	0.208830016	0.824	0.779	4.40E-07
Snhg9	4.41E-41	0.578853339	0.971	0.928	8.42E-37
Taxlbpl	1.04E-11	0.20563376	0.855	0.812	1.98E-07
Cdk6	9.45E-13	0.216050004	0.935	0.896	1.80E-08
Tcofl	3.45E-31	0.302647593	0.965	0.928	6.58E-27
Cebpz	1.09E-16	0.237798069	0.939	0.902	2.09E-12
Loxl2	1.30E-17	0.571139295	0.89	0.857	2.48E-13
Rangapl	2.34E-40	0.369409656	0.984	0.953	4.46E-36
Dek	1.64E-18	0.231074803	0.996	0.967	3.12E-14
Nolcl	9.61E-30	0.309060428	0.986	0.959	1.83E-25
Mybbpla	1.01E-15	0.209760443	0.969	0.943	1.92E-11
Uchl3	4.63E-23	0.291386824	0.963	0.937	8.83E-19
Mt2	2.21E-46	0.647830277	0.982	0.959	4.21E-42
Fam 177a	7.40E-29	0.318947806	0.965	0.943	1.41E-24
Ak2	2.85E-38	0.322110667	0.992	0.971	5.45E-34
Pdcdll	1.06E-26	0.317776644	0.994	0.973	2.03E-22
Clnsla	7.78E-15	0.200963226	0.955	0.935	1.49E-10
Nsun2	4.46E-23	0.25780744	0.965	0.947	8.51E-19
Eiflax	6.10E-25	0.259171146	0.998	0.982	1.17E-20
Utplll	2.11E-21	0.247732591	0.978	0.963	4.03E-17
Nifk	4.74E-16	0.25794523	0.973	0.959	9.06E-12
Mrpl36	8.39E-15	0.203735334	0.963	0.949	1.60E-10
Chchd4	3.75E-49	0.406592072	0.99	0.978	7.15E-45
Mtl	1.69E-19	0.330543022	0.99	0.98	3.23E-15
Mcm6	5.05E-14	0.203330997	0.93	0.92	9.64E-10
2810004N23Rik	2.73E-25	0.282539829	0.982	0.973	5.21E-21
Lmo4	1.74E-66	0.775349512	0.992	0.986	3.31E-62
Sms	1.65E-36	0.313663566	0.992	0.986	3.15E-32
Tmem5	7.44E-27	0.31509393	0.949	0.943	1.42E-22
Abcfl	4.64E-25	0.277959491	0.992	0.988	8.85E-21
Sfxnl	6.98E-21	0.212944289	0.984	0.98	1.33E-16
Gml6286	8.21E-20	0.224472114	0.988	0.984	1.57E-15

Cox7a2l	1.45E-19	0.200215258	0.994	0.99	2.77E-15
Psatl	2.81E-16	0.206124692	0.994	0.99	5.37E-12
Zfosl	5.30E-16	0.206256512	0.992	0.988	l.OlE-11
Nhp2ll	9.94E-34	0.239069695	1	0.998	1.90E-29
Txn2	8.06E-23	0.202261807	0.994	0.992	1.54E-18
Dctppl	1.40E-22	0.221067567	0.992	0.99	2.67E-18
Eif3jl	8.55E-20	0.270419381	0.992	0.99	1.63E-15
Nhp2	3.24E-68	0.348934627	1	1	6.19E-64
Txnl4a	6.38E-49	0.36485702	0.99	0.99	1.22E-44
Naplll	1.10E-46	0.276547552	1	1	2.10E-42
Srm	1.22E-45	0.356879476	0.992	0.992	2.32E-41
Tomm5	1.65E-43	0.313429107	1	1	3.15E-39
Dnajc2	4.24E-40	0.373302174	0.988	0.988	8.10E-36
Ddx21	2.72E-35	0.383841731	0.996	0.996	5.18E-31
Ncl	6.24E-31	0.351868277	1	1	1.19E-26
Serbpl	1.10E-27	0.22648657	1	1	2.11E-23
Naal5	1.44E-20	0.281257486	0.982	0.982	2.75E-16
Maplb	1.99E-11	0.211674236	0.949	0.949	3.79E-07
Gngl2	3.44E-45	0.336166251	0.994	0.996	6.58E-41
Bola2	1.95E-33	0.243627002	0.998	1	3.72E-29
Ddxl8	1.13E-20	0.236133065	0.994	0.996	2.15E-16
Calml	4.37E-20	0.209338392	0.998	1	8.35E-16
Llph	2.37E-16	0.207946587	0.994	0.996	4.52E-12
Hnrnpm	1.63E-15	0.211499543	0.99	0.992	3.11E-11
NoplO	2.74E-32	0.258763009	0.996	1	5.23E-28
Wdr43	1.46E-25	0.286052346	0.992	0.996	2.80E-21
mt-Nd3	2.70E-23	0.241501548	0.994	0.998	5.15E-19
Knopl	1.42E-22	0.257948217	0.992	0.996	2.71E-18
Dpy30	1.40E-15	0.206386698	0.971	0.975	2.67E-11
Dph3	1.25E-33	0.288444631	0.982	0.988	2.38E-29
Anp32b	6.68E-20	0.23155113	0.99	0.996	1.28E-15
Odcl	2.58E-14	0.212362532	0.988	0.996	4.92E-10

[0267] iPSCs emerge through a tight bottleneck from cells in the MET Region

[0268] Trajectory analysis showed that cells from the MET region subsequently gained a broad epithelial identity and began to rapidly diverge to give rise the iPS-, epithelial-, trophoblast-, and neural-like cells (FIG. 26A). Importantly, the ancestor distributions of these classes were not distinguishable before the withdrawal of dox at day 8, suggesting that the cells' fates did not appear yet to be determined at that point (FIG. 26B).

[0269] By day 11.5-12.5, the iPS-like cells began to show a clear signature of pluripotency, including canonical marker genes such as Nanog, Sox2, Zfp42, Otx2, Dppa4, and an elevated cell-cycle signature (FIGs. 26C, 26D). In 2i conditions, these iPS-like cells accounted for 12% of cells by day 11.5 and 80-90% from days 15 through 18. In serum conditions, the trend was similar, but the process was delayed by roughly one day and was far less efficient: the pluripotency signature was found in 3.5% of cells by day 12.5 and peaked at just 10-15% from days 15.5 through 18 (FIG. 24G). Notably, we found substantial heterogeneity among the iPSC-related cells. Recent studies reported that a small subset of cells in 2i conditions showed a signature characteristic of the embryonic 2-cell (2C) stage (Falco et al., 2007; Kolodziejczyk et al., 2015; Macfarlan et al., 2012). Scoring our iPS-like cells with signatures based on profiles from 2 cell-, 4 cell-, 8 cell-, 16 cell-, and 32 cell-stage embryos (Goolam et al., 2016) (Table 15, FIG. 32A, 32B), -20% of cells in both 2i and serum conditions showed a 2C, 4C, 8C, 16C, or 32C signature (with roughly half showing signatures for two consecutive stages).

[0270] Trajectory analysis suggested that successfully reprogrammed cells passed through a tight bottleneck in days 10-1 1. The ancestral distribution of iPSCs spanned -40% of all cells at day 8.5. It falls to -10% of cells at day 10 in 2i conditions and only -1% at day 11 in serum conditions. These results suggested that only a small and distinct subset of cells transitioning out of the MET Regions toward various fates had the potential to become iPS cells (below). These iPSC progenitors did not yet fully acquired the pluripotency signature but were changing rapidly toward this fate. They resided along certain thin 'strings' in the FLE representation (FIG. 24F, white arrow and 4C, green). iPSC ancestors then rose to -40% at day 14 in 2i (and 10% on day 14 in serum), reflecting rapid expansion of pluripotent precursors (FIG. 26C, yellow).

[0271] By clustering genes according to similar expression trends along the trajectories to successful reprogramming in 2i and serum conditions, we found induction of various groups of genes involved in regulation of pluripotency, and repression of genes involved in certain metabolic changes and RNA processing (FIG. 32C). Among the upregulated genes, 24 were preferentially expressed in the late stage of reprogramming on successful trajectories and were mostly absent from other cell types; these included Ooep, Fmrlnb, Lncencl, and Tell (FIG. 32C, Table 17). These genes can be candidate markers for fully reprogrammed cells.

Table 17 - List of genes for 15 groups of genes along the successfully reprogrammed trajectory reported in FIG. 32A

Gene sets related to FIG. 32A							
1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Sbspon	Terf1	Lypla1	Lactb2	Pnkd	Rpl7	Tcea1	Il1rl1
Dst	1700007K13Rik	Tceb1	Igfbp2	Ptma	Rpl31	Mcm3	Fhl2
Nrp2	Ass1	Dnpep	Trip12	Dtymk	H3f3a	Sgol2a	Col3a1
Eef1b2	Mdk	Tfcp2l1	Marc2	Dbi	Rpl7a	Psmd1	Col5a2
Serpine2	Chchd5	Kdm5b	Gm13580	Snrpe	Rpl12	R3hdm1	Sdpr
Ephx1	Praf2	Swt1	Hat1	Cacybp	Zfos1	Mcm6	Fn1
Nudt5	Timm17b	Atp1b1	Tfpi	Ndufs2	Pcsk1n	Dhx9	Col6a3
Commd3	Hdac6	Phyh	Platr3	F11r	Rpl10	Gm2000	Gpc1
Ndufa8	Ndufb11	Wdr5	Scand1	Atp5c1	Bex2	Prrc2c	Serpinb2
Ccdc34	Uxt	Odf2	Platr27	Tubb4b	Ndufb5	Parp1	Ubxn4
Nop10	Klhl13	Rif1	Fthl17c	Spc25	Rps3a1	Nvl	Klhdc8a
Knstrn	Slc25a5	AA467197	Usp9x	2700094K13Rik	Apoa1bp	Lbr	Ptgs2
Dtd1	Ube2a	Slc24a5	Ndufa1	Cd59a	Txnip	Enah	Rgs16
Rbck1	Upf3b	Mrps5	Gm9	Eif3m	Gstm1	Cenpf	Ier5
Nnat	Rhox6	Eif2s2	Rhox1	Rad51	Rpl34	Dtl	Soat1
Rbm3	Rhox9	Mybl2	Rhox5	Spint1	Rps20	Yme1l1	Copa
Hmgb3	Mcts1	Gtsf1l	Thoc2	Hypk	Gm11808	Set	Grem2
Fundc2	Bcap31	Wfdc2	Rbmx2	Dut	Rps6	Prrc2b	Col5a1
Slc7a3	Idh3g	Ncoa3	Usp26	1700037H04Rik	Rps8	Rpl35	Angptl2
Hmgn5	Lage3	Sall4	Hprt	Tpx2	Laptm5	Hnrnpa3	Hspa5
2210013O21Rik	Pbdc1	Tfap2c	1700013H16Rik	Ube2c	Rpl11	Nusap1	Gorasp2
Rnf13	Bex4	Ebp	Fmr1nb	Aurka	Rpl22	Mga	Creb3l1
Cks1b	Bex1	Atp6ap2	Dusp9	Ppdpf	Rpl9	Zfp106	Rcn1
Psmb4	Wbp5	Nono	Ssr4	Plp2	Rpl5	Myef2	Bdnf

Bolal	Ngfrapl	Algl3	Dkcl	NaalO	Rpl21	Xrn2	Thbsl
Gstm5	Trapla	Gm8797	Vbpl	Pdhal	Gapdh	Csnk2al	Fgf7
Psrcl	Hsdl7bl0	Tpd52	Pdk3	Exosc8	Rps9	Ubal	Dstn
Cth	Rab9	Chmp4c	Lasll	Smc4	Cox6b2	Gnl3l	Rrbpl
Ndufb6	Dnajcl9	Lrrc31	Ogt	Pmfl	Rpl28	Huwel	Thbd
Cdc26	Lamtor2	Actl6a	Pin4	Rab25	Rps5	Smcla	Srxnl
Psipl	Fdps	Fxrl	Atrx	Anp32e	Rpsl9	Sms	Chmp4b
Cdkn2a	Psmd4	Sox2	Magtl	Atp5fl	Rpsl6	Midi	Procr
Lltdl	Acp6	Noct	Cox7b	Stoml2	Eif3k	1810022K09Rik	Dlgap4
Tmem59	Hadh	PlatrlO	Pgkl	Ctnnall	Spint2	Ndufcl	Ptpnl
Hspbll	Acer2	Hiatl	Rpl36a	Nasp	Cox6bl	Slc39al	Pmepa1
Uqcrh	Slc2al	Elovl6	Prpsl	Cdc20	Rpll3a	Ilf2	Slco4a1
Ptprf	Gjb5	Acadm	Fgdl	Ppih	Rpll8	Larp7	Pgrmc1
Eif3i	Hdacl	Zfp292	Prdx4	Cdca8	Idh2	Tet2	Bgn
Atpifl	Hscb	Aqp3	A830080D0IRik	Zbtb8os	Rps3	Fubpl	Itm2a
Stmnl	Ung	Klf4	Rbbp7	Rpa2	Rpl27a	Anp32b	Fndc3b
Enol	Cldn4	Echdc2	Zrsr2	Hmgn2	Rpsl3	Smc2	Sec62
Fgfbpl	Cldn3	Gjb3	Ttcl4	Miip	Rpsl5a	Zfp462	Postn
Shisa3	Atp6vlf	Fabp3	Jadel	Apitdl	Uqcrc2	Puml	Faml98b
Scarb2	Mkrnl	Rps6kal	Vangll	Park7	Ypel3	Srrml	SlOOa7a
Cops4	Cct7	Rsrpl	Ak4	Tyms	Ifitm3	Rcc2	Crctl
Gltp	Nful	Tcea3	Fbliml	Cenpa	Rplp2	Gm26825	Ngf
Pop5	Slc2a3	Usp48	Zfp600	Qdpr	Mrpl23	Tomm7	Rhoc
Pebpl	Fkbp4	Alpl	Gml3251	Med28	Rpsl2	4930548H24Rik	Csfl
Rpl6	Ldhb	Gml3154	2610305D13Rik	Paics	Rpsl5	Rfcl	Collla1
Ran	AU018091	Agtrap	Fbxo6	G3bp2	Rpl6l	Grsfl	F3
Mospd3	Ligl	Insigl	Rbpj	Hnrnpdl	Naca	Hnrnpd	Ostc
Hmgbl	Beam	Dnajb6	Crlf2	Cit	Rps26	Golga3	Cyr61
Ndufa4	Exosc5	Yesl	Ppplcc	Rfc5	Ndufal3	Mcm7	Bel10

Podxl	Gmfg	Lap3	Arf5	Chchd2	Rpll8a	Luc7l2	Glipr2
Akrlb3	Map4kl	Kit	Stra8	Rfc2	Bst2	Cbx3	Sec61b
Hnrnpa2bl	Ppplrl4a	Rest	Ube2s	Atp5j2	Cox4il	Immt	Tnc
Lsm3	Tbcb	Sppl	Zfp787	Lsm5	Rpll3	TmsblO	Eval b
Trh	Gpil	Mtf2	Tmeml60	Tcf7ll	Rpll5	Dqxl	Errfil
Mgstl	Etfb	Pxmp2	Calm3	Suclgl	Rps24	Mcm2	Ost4
Trappc6a	Ucp2	Ulkl	Zfp428	Tpil	Rpl23a-ps3	Ptms	Ugdh
Dmrtc2	Folrl	Medl3l	Plekha4	Cdca3	Rpll3-ps3	Aebp2	Apbb2
Fbl	Mrpll7	Tbx3	Arrdc4	Lockd	Rps25	Fam60a	Igfbp7
Krtdap	Arl6ipl	Sbnol	Eif3f	Peg3	Fxyd6	Trim28	Cxcl5
Prmtl	Aldoa	Cops6	Septl	Gltscr2	Rpll0-ps3	Hnrnpl	Ppbp
Bax	Pycard	Slc25al3	Ctbp2	Sael	Rpl4	Polr2i	Cxcl3
Ldha	Bnip3	Asns	Sycp3	Lsr	Gsta4	Sema4b	Cxcll
Tm2d3	Utfl	Trim24	Nudt4	Ruvbl2	Eeflal	Prcl	Cxcl2
l7Rn6	Ifitm2	Zc3havl	Sap30	Bcat2	Rpl29	Blm	Ereg
Ndufc2	Cenpw	Ezh2	Gm2694	Snrpn	Rpsa	RP23-4H17.3	U90926
Ndufabl	Ddit4	Tra2a	Fam25c	Coq7	Rpll4	Bclafl	Rsrc2
Tmem219	Cisdl	Gdf3	Sapl8	Plkl	Rps27a	Ptges3	Denr
Vkorcl	Ddt	Dppa3	Klf5	Spnsl	Gnb2ll	Arglul	Ubc
Mki67	ChchdlO	Nanog	Khdc3	Dctppl	Rpl26	Mcm5	Serpinel
Glrx3	Pfkl	Lpcat3	Ooep	Fbxo5	Rpl23	Smarca5	Pcolce
Cd81	Polr2e	Cd9	Higdla	Sf3b5	Rpll9	Cnotl	Kdelr2
Perp	Gpx4	2810474019Rik	Mrps24	Cdkl	Rpl27	Rps26-psl	Cavl
Mif	Cirbp	Apocl	Eif4al	Lsm7	Dcxr	Aars	Fine
Atp5d	1500009L16Rik	Apoe	Clqbp	Eef2	Rps23	Ankrdll	Ptn
Ndufs7	Priml	Pvrl2	Suzl2	Mrpl42	Btf3	Wapl	Capg
Uqcrll	Eif4ebpl	Cox7al	AI662270	Cct2	Rps7	Rpgripl	Rab7
Oazl	Ankrd37	Tdrdl2	Dynll2	Atp5b	Wdr89	Suptl6	Fbln2
Slc25a3	Cope	Tead2	E130012A19Rik	Ormdl2	Rpl30	Zc3hl3	Sec13
Ndufal2	Sin3b	Gtf2hl	Gnal3	Sarnp	Gml0020	Uchl3	Cxcll 2
Cnpy2	Syce2	Spty2dl	Snhg20	Hmgb2	Rpl8	Anapcl3	Tspan9

Nabp2	Asnal	Mfge8	Texl9.1	Lsm4	Rpl3	Gnai2	Arhgdib
Slc25a4	Mtl	Ticrr	Pfkp	Tecr	Rpl35a	UqcrlO	1111
Apela	2700060E02Rik	Zfand6	Tubb2b	Orc6	Gm9843	Actr2	Ehd2
Isynal	Mrpsl6	Eed	H2afy	Nudt21	Sodl	Canx	Pvr
Mrpl34	Tkt	Tmem41b	Cox7c	Cdhl	Psmbl	Alkbh5	Plaur
Ndufb7	Mphosph8	Gga2	Lncencl	Psmb5	NdufblO	Ncorl	Psmd8
Prdx2	Esco2	Nfatc2ip	Nampt	Dhrs4	RpslO	Pfas	Fxyd5
Pllp	Bnip3l	Mylpf	Ifi27	Cdca2	RpllOa	Naa38	Rcn3
Got2	Sugtl	Echsl	Tell	Spc24	Ddah2	Xafl	Klfl3
PsmblO	Pigyl	Ifitml	Papola	H2afx	Gm26917	Ywhae	Vimp
Rab4a	Psma4	Taldol	Apobec3	Slc35f2	AY036118	Tafl5	Lrrc32
Dnajc9	Cox5a	Fgf4	Smclb	Pkm	2410015M2ORik	Npepps	Map6
Itm2b	Morf4ll	Akapl2	Pim3	Anp32a	Rpl27-ps3	Top2a	Adm
Atp5l	H2afv	Sgkl	Rpl39l	Snapc5	Gml0036	Acly	Mical2
Cadml	Commdl	Tetl	Eif4a2	Tipin	Prelid2	Bptf	Tgfbli1
Crabpl	Pttgl	Spic	Adprh	Ccnb2	Rpsl4	Fasn	Rnhl
2810417H13Rik	Psmb6	Csrp2	Dppa4	Cox7a2	Rpll7	Slcl6a3	H19
Rps27l	Psmdl2	Baz2a	Dppa2	Gpxl	Gm6133	Dek	Igf2
Gtf2a2	Atp5h	Ash2l	Cggbpl	Impdh2	Fau	Rbm25	Cttn
Hmgn3	Galkl	Zfp42	Morc3	Ndufaf3	Cox8a	Dnajc21	Rgsl7
Nf2	Psma2	Tmeml92	Brwdl	Uqcrcl	Eeflg	MyolO	Ctgf
Ramp3	Acotl3	Nr2c2ap	Tmeml81a	Zmat5	Gm9493	Rad21	Sarla
Mdhl	Uqcrb	Klf2	Dynltla	Pold2	Rpl9-ps6	Stl3	Col6a2
Hintl	Cetn3	AnapclO	Mpcl	Snrnp25	Gstol	Limal	Pofut2
Aldh3al	Dhfr	Dnase2a	Pgp	Npml	Rpsl2-ps3	Usp7	Pttglip
Poldip2	Mycn	Mt2	Gfer	Hmmr	mt-Co2	Etv5	Bsg
Krtl9	Psma6	Gabarapl2	Piml	Cdkn2aipnl		Tfrc	Timp3
Krtl7	Fkbp3	Kat6b	Myolf	Tmeml07		Gsk3b	Btgl
Itgb4	Atp6vld	Hesxl	Dhxl6	Cldn7		Coxl7	Atp2bl
Secl4ll	Brixl	Zfhx2	Dazl	Atp5gl		Gm8186	Raplb
Tkl	Cox6c	Rnaseh2b	Vapa	Cbxl		Srpkl	Ndufa412
Stard3nl	Eif3e	Tdh	Ralbpl	Psmb3		Stk38	Myl6

Histlhlb	Tonsl	Rgcc	Arll4epl	Jup		Brd4	Hmoxl
Histlhle	Gcat	Zbtb44	Prrcl	Dcakd		Gm42418	Junb
Uqcrfsl	Syngrl	Rpp25	Fbxol5	Sumo2		Uhrfl	Mmp2
Eci2	Cenpm	Rbpms2	Gstp2	Birc5		Khsrp	Gm22
Ndufs6	Ndufa6	U2surp	D030056L22Rik	Stral3		Birc6	Actal
Mrps36	Atp5g2	Slc25a36		Histlh2ae		Erdrl	Nrpl
Id2	Paml6	Amt		Gmnn		Matr3	Vcl
Rtnl	Pigx	Arih2		Cks2		Stipl	Arf4
Sival	Ndufb4	Slc25a20		Higd2a		Incenp	Selk
Ahnak2	Dynltlf	Tdgfl		Ccnbl		Tmem258	Mustnl
Nudtl4	Thoc6	Trim71		Rrm2		Hells	Spcsl
Crip2	Tceb2	Uppl		Misl8bpl		Scd2	Fermt2
Ptp4a3	Ccnf	Cct4		Mthfdl		Eif3a	Gjb2
Ly6a	Ndufv3	Skpla		Cct5		mt-Ndl	Ubl5
Eefld	Ndufa7	Vdacl		Cycl			Col5a3
Tst	Tubb5	Gm2a		Eif3l			Cnnl
HlfO	Rpp21	Mpdul		Tubalb			Oaf
Pmml	Znrdl	Tmem256		Krt8			Thyl
Samm50	Oardl	Scpepl		Hnrnpal			Trappc4
Eif4b	Ndufv2	Igf2bpl		Mrpl40			Ncaml
2610318N02Rik	Tgifl	Calcoco2		Rfc4			Wdr61
Dgcr6	Cebpzos	Dnajc7		Bbx			Cspg4
Fetub	Mta3	Slc25a39		Ezr			Sema7a
Atp5o	Pfdnl	Grn		Acat2			Loxll
Agpat4	Impa2	Ccdc43		Cldn6			Mapk6
Nme4	Smc3	Ttyh2		Ppill			Col 12a1
Mapkl3		Wbp2		U2afl			Amotl2
Cd320		Ubald2		Pfdn6			Selm
Ly6g6c		Jarid2		Lsm2			Xbpl
Ly6g6f		Ubxn2a		Polrlc			Aebpl
Dnphl		1110008L16Rik		Ndufall			Ykt6
Cox7a2l		Esrrb		Crb3			Tns3

Pigf		Ckb		Myl12b			Sec61g
Ecscr		Atxn10		Dpy30			Sertad2
Cyb5a		Slc25a1		Epcam			Rtn4
Rnaseh2c		Morc1		Paip2			Adam19
Trmt112		Jam2		Lmnb1			Sqstm1
Carnmt1		Wtap		Atp5a1			Sparc
Avpi1		Sod2		Ndufs8			Kctd11
Ndufb8		Rnf5		Rbm4b			Gabarap
Cuedc2		Zfp57		Banf1			Cxcl16
Sfr1		Cdc5l		Mrpl49			Tax1bp3
		Slc29a1		Arl2			Pafah1b1
		Gm7325		Fkbp2			Serpinf1
		Ccnd3					Ift20
		Ppm1b					Ccl2
		Msh2					Ccl5
		Msh6					Vmp1
		Cystm1					Col1a1
		Taf7					Copz2
		Dcp2					Igfbp4
		Snx2					Eif1
		Cndp2					Timp2
		Chka					Klf6
		Ubxn1					Inhba
		Klf9					Serpinb6a
		Scd1					Card19
		mt-Co1					Pdlim7
							Tmed9
							Smim15
							Plk2
							Rhob
							Nfkbia
							Arf6

							Frmd6
							Actn1
							Ltbp2
							Dlk1
							Tnfaip2
							Crip1
							Snhg18
							Cthrc1
							Ext1
							Has2
							Wisp1
							Myh9
							Lgals1
							Kdelr3
							Atf4
							Tuba1c
							Itga5
							Vasn
							Col8a1
							Ier3
							Ppp1r11
							Vegfa
							Ltbp1
							Crim1
							Fez2
							Cdc42ep3
							Zfp36l2
							Hbegf
							Yipf5
							Lox
							Ier3ip1
							Efemp2
							Ehbp1l1

							Ehd1
							Fads3
							Ankrd1
							Dusp5

Table 17 (Cont'd)

Gene sets related to FIG. 32A						
9	**10**	**11**	**12**	**13**	**14**	**15**
Map4k4	Snhg6	Ptp4a1	Bag2	Sdhaf4	Imp4	Eif5b
Bzw1	Mpzl1	Actr1b	Mrpl30	Sumo1	Tuba4a	Nop58
Raph1	Creg1	Hspd1	Hspe1	Aamp	Ncl	Rpl37a
Arpc2	Uap1l1	Bok	Acadl	Eif4e2	Ssna1	Myeov2
Tmbim1	Ptges	Tsn	Stk16	Timm17a	Surf2	Sept2
Lrrfip1	Serf2	Nucks1	Adipor1	Ufc1	Urm1	Ddx18
Ube2f	Slc20a1	Tpr	Phlda3	Pfdn2	Ppp2r4	B930036N10 Rik
Hdlbp	Cst3	Uck2	Prdx6	Hspa14	Dpm2	Nmt2
Nifk	Gss	Hnrnpu	Mpc2	Edf1	Arpc5l	Sptan1
Actr3	Sdc4	Eprs	Mgst3	Dnlz	Timm10	Exosc2
Csrp1	Adrm1	Smyd2	Cnih4	1110008P14 Rik	Ssrp1	Dync1i2
Arpc5	Lamp2	Rbm17	Aida	Tor2a	Snrpb	Psmc3
Qsox1	Renbp	Agpat2	St6galnac4	Psmb7	Ppid	Usp50
Prrx1	S100a1	Fbxw2	Pdia3	Dnmt3b	Gpatch4	Cse1l
Tmco1	S100a13	Mtx2	Mrps26	Tgif2	Jtb	Atp5e
Tagln2	Cnn3	Caprin1	Naa20	Rpn2	Nras	Rps21
Wdr26	Atp6v1g1	1500011K16 Rik	Fkbp1a	Tceal8	Gar1	Plk4
Degs1	Tm2d1	Nop56	Id1	Morf4l2	Cenpe	Naa15
Capn2	Atp6v0b	Snx5	Dynlrb1	Fabp5	Sep15	Rps27
Rrp15	Snhg12	Raly	Romo1	Car2	Ebna1bp2	Mrpl9
Hacd1	Sh3bgrl3	1110008F13 Rik	Samhd1	Selt	Svbp	Sars

Surf4	Pdpn	Srsf6	Topi	Cct3	Mrpsl5	Agl
Ptrhl	Smiml4	Sysl	Pfdn4	Ssr2	Thrap3	Ccne2
Faml29b	Coxl8	Rael	Gnas	Rbm8a	Ak2	Otud6b
Gsn	Hspb8	Ddx3x	Ctsz	1810037117Rik	Tmem234	Vcp
Rbmsl	Tmeml2Oa	Vma21	Slmo2	Ube2d3	Zcchcl7	TexlO
Grbl4	Arpclb	Ccna2	Fhll	Dnajal	Hnrnpr	Tmem245
Zak	Gpnmb	Tpm3	G6pdx	Clta	Ddost	Lepr
Nfe2l2	Malsul	Atplal	Xist	Prdxl	Mrto4	Ccdcl63
Nckapl	Pole4	Csdel	Sh3bgrl	Psmb2	Sdhb	Ybxl
Zc3hl5	Chmp2a	Eif4e	Tmem35	Marcksll	Szrdl	Gml3075
Itgav	Vasp	Ddahl	Ammecrl	Trnaulap	Mrpl20	Noc2l
Cd44	Rabacl	Rad23b	Eiflax	Nude	Aurkaipl	Faml33b
Emc7	Blvrb	Ndcl	Stmn2	Sfn	Lrpapl	Abcblb
Eif3jl	Capnsl	Ctps	Lhfp	Tmem60	Mrfapl	Dhxl5
B2m	Dkkll	Pabpc4	Tm4sfl	Ppplcb	Lyar	Noal
Fbnl	Nuprl	Mycbp	Mbnll	Slbp	Dynlll	Atp5k
Prnp	Snx3	Sfpq	Lxn	Plac8	Cox6al	Pdapl
H13	Psap	Ptp4a2	Hdgf	Anapc5	Arl6ip4	Ndufa5
Pdrgl	Cstb	Ythdf2	Mex3a	Por	Mrpsl7	Rbm28
Maprel	Gadd45b	Srm	S100al6	Ywhag	Eif4h	Pdia4
Eif6	Aril	Gnbl	SlOOalO	Capza2	Mdh2	Serbpl
Myl9	Ddit3	Nadk	Mrps21	Gstkl	Fisl	Hk2
Ywhab	Cd63	Dbf4	Phgdh	Ruvbll	Znhitl	Paip2b
Timpl	Ifi30	Dnajc2	Camk2d	Arpc4	Fscnl	Snrpg
Hs6st2	Hsbpl	Abcf2	Cisd2	Hnrnpf	Arpcla	Gmcll
Flna	Mapllc3b	Rheb	Fam92a	M6pr	Pomp	Wbpll
Msn	Cyba	Ppmlg	Tmem55a	Mlf2	2610001J05Rik	Dennd5b
Satl	Tomm20	Iscu	Ggh	Cops7a	Cycs	Ndufa3
Sh3kbpl	Ghitm	Mlec	Tomm5	Goltlb	Vamp8	Cnot3
Anxa5	Psme2	RnflO	Txnl	Clptml	Faml 36a	U2af2
Ufml	Ctsb	Atp2a2	Nfib	Psmc4	Cnbp	Iqgapl

Dclkl	Srpr	Gnb2	Scp2	Nup62	Hmces	Ipo7
Wwtrl	Tbrgl	Eif3b	Ktil2	Mesdc2	Chchd4	Teadl
Serpl	Hexa	Fam220a	Akrlal	Ppp4c	Emgl	1110004F10 Rik
Ssr3	Rablla	Cczl	Macfl	Bccip	Phb2	Knopl
Crabp2	Spg21	Bri3	Utplll	Phlda2	Mrpl51	Bola2
Lmna	Ppib	Gtf3a	Wasf2	Ltvl	Tsen34	Fus
S100a4	Rhoa	Hsphl	Mtfrll	Zwint	Napa	Hras
SlOOall	Pdlim4	Mat2a	Id3	Ube2n	Mrpsl2	Polr2l
Vcaml	Cd68	Mthfd2	Hspg2	Myl6b	Nudtl9	Ap2a2
Snx7	Ggnbp2	H2afj	Minosl	Fam32a	EmclO	Amdl
Ppp3ca	Nidi	Strap	Acot7	Ddx39	Grwdl	Ddx21
Pdlim5	Ninjl	Bcatl	Atad3a	Ier2	Snrpal	Cdc34
Lmo4	Ctsl	Slcla5	Cdk6	Calr	Mrpsll	Metap2
Sh3glbl	Gml0116	Tomm40	Sri	Cneplrl	Aen	PetlOO
Gng5	Glrx	Eif4g2	Mrpl33	MM	Clnsla	Timm44
Wis	Twistnb	Gdel	Grpell	Ciapinl	Tufm	Haus8
Chchd7	Npc2	Mettl9	Limchl	Gcsh	Ino80e	Gfod2
Impadl	Dap	Eif3c	Ociadl	Emc8	Bckdk	Nip7
Rab2a	Ndrgl	Kcnqlotl	Ociad2	Chmpla	Bub3	2810004N23 Rik
Ndufaf4	Cyb5r3	Rwddl	Septll	Gnpnatl	Urah	Gnl3
Ube2jl	Tmbim6	Ppal	Anxa3	Bmp4	Napll4	Nisch
Tpm2	Litaf	Mbd3	Pdgfa	Dadl	Snrpd3	Ktnl
Tlnl	Hacd2	Abhdl7a	Racl	Tsc22dl	Sumo3	Mrpl52
Plin2	Hcfclrl	Map2k2	Kpna7	Aasdhppt	Timml3	Loxl2
Mtap	Atp6v0e	Aes	Polrld	Rpusd4	Thopl	Gml0076
Jun	Ostfl	Rtcb	Shfml	Oaz2	Dohh	Tafld
Jakl	Pdliml	Naplll	Lsm8	Fam96a	Yeats4	Gm26737
Mast2		Cs	1810058I24R ik	Rsl24dl	Cdk4	Arppl9
Elovll		Dlcl	Gngl2	Rnf7	Pa2g4	Rps27rt
Txlna		Abcel	Aupl	Rbpl	Lsml	Limk2
Clic4		Dnaja2	Bola3	Rrp9	Fkbp8	Nudcd3

Cdc42		E2f4	Actg2	Nme6	Ccdcl24	Hnrnpab
Nppb		Psmd7	Arl6ip5	Ewsrl	Ddal	Larpl
Pgd		Dcunld5	Foxpl	Arfl	Rbmxll	Mybbpla
Cgrefl		Rp9	Rhnol	Trp53	Lsm6	Ap2bl
Ywhah		Ei24	Magohb	Car4	D8Ertd738e	Cite
Gml673		Rdx	Ybx3	Slc35bl	2310036022 Rik	Nfe2ll
Wdrl		Imp3	Epnl	H3f3b	Cmc2	Pcgf2
Pcdh7		Polr2m	Sepwl	Gaa	Aprt	Nmtl
Tpst2		Cdv3	Gemin7	Anapcll	Vdac2	Ddx5
Corolc		Map4	Egln2	Dusll	Apexl	Rpl38
Tmed2		G3bpl	Tmeml47	Paklipl	Nedd8	Srsf2
Aplsl		Srsfl	Pdcd5	Emb	N6amt2	Prpf4b
Fam20c		Lrrc59	Josd2	Pdia6	Reep4	HnrnpaO
Actb		Snf8	Aktlsl	Ywhaq	Pinl	Nsa2
Cyth3		Kpnbl	Igflr	Max	Tmedl	Smnl
Slc7al		Psme3	Serpinhl	Eif2sl	Ecsit	Rps29
Colla2		Lsml2	Rrml	Srsf5	Elofl	Slirp
Tes		Faml 04a	Prkcdbp	Ahsal	Hmbs	2010107E04 Rik
Calu		Prpsapl	Parva	Subl	Manf	Rpl37
Caldl		Gpsl	Tspan4	Mcrsl	Tma7	Wdr70
Mtpn		Gdi2	Ccndl	Tarbp2	Ccdcl2	Polr2k
Zyx		Rala	Epb41l2	Copzl	Cld	Rangapl
Tex261		Ssrl	Marcks	Glyrl	Nhp2	Hesl
Cyp26bl		B230219D22 Rik	Cd24a	Ube2v2	Uqcrq	Son
Sec61al		Cxcll4	Gjal	Ap2ml	Atoxl	Snhg9
Brkl		Hnrnpk	Arid5b	Dnajbll	Gukl	Hnrnpm
Ltbr		Nsun2	Plpp2	Cct8	Rangrf	Rps28
Gabarapll		RablO	Snrpf	Tcpl	Eif5a	Abcfl
Empl		Smc6	Atxn7l3b	Rabllb	Tmem97	Ptcra
Erccl		Odcl	Shmt2	Mrpsl8b	Nmel	Sgoll
Cd3eap		Srp54b	Lrpl	Meal	Mrpl27	Wdr43
Axl		Glrx5	Col4a2	Calm2	Phb	Cebpz

Actn4		Eif5	Ckap2	Polr2d	Coa3	Epb41l4aos
2200002D01 Rik		Pabpcl	Vps36	Eifla	Ictl	Ndufa2
Atf5		Ly6e	Fgfrl	BC031181	Hnl	Rbm22
Emp3		Pcbp2	Nrgl	Pgaml	Mrps7	Tcofl
Prss23		Rslldl	Uba52	Xpnpepl	1810043H04 Rik	Nars
Rrp8		Gsptl	Pgls	mt-Co3	Mrpll2	Ddbl
Ilk		Mapkl	Scoc	mt-Nd4	Tmeml4c	Nmrkl
Rras2		Eif4gl	Nfix		Nopl6	Usmg5
Pik3c2a		Ppplr2	Arl2bp		Prelidl	Pdcdll
Itpripl2		0610012G03 Rik	Gml0073		Lman2	mt-Nd2
Tnrc6a		Naa50	Zfhx3		Ddx46	mt-Atp8
Cdipt		Tomm70a	2310022B05 Rik		2010111I01R ik	mt-Nd3
Abracl		Srrm2	Ube2el		Mrpl36	mt-Nd4l
Col6al		Kif5b	Dph3		Sf3b6	mt-Nd5
Slcl9al		Etfl	Anxa8		Sptssa	
Ube2g2		Hspa9	Cnihl		Erh	
Cnn2		Ube2d2a	Lgals3		TmedlO	
Nfic		Psatl	Tptl		Snwl	
Ncln		Npm3	Mbnl2		Zfp706	
Txnrdl			Smco4		9130401M01 Rik	
Ckap4			Rexo2		Chracl	
Elk3			Cryab		Polr2f	
Phldal			Anxa2		Tomm22	
Llph			Nedd4		Adsl	
Hmga2			Cdl09		Rbxl	
Tmem5			Iraklbpl		Phf5a	
Col4al			Syncrip		Nhp2ll	
Tm2d2			Pcolce2		Rrp7a	
Rwdd4a			Mras		Tubala	
Cpe			Pcbp4		Ranbpl	
Tpm4			Ifrd2		Hmgnl	

Dnajbl			Cmtm7		Tmem242	
Piezol			Purb		Mrpll8	
Tcf25			GrblO		Rnpsl	
Itgbl			Sptbnl		Ube2i	
Flnb			Ccngl		Stubl	
Gchl			Chd3		Mrpl28	
Pnp			Pfnl		Srsf3	
Mmpl4			Txndcl7		Glol	
Esd			Emc6		Mrpll4	
Kctdl2			Nxn		Srsf7	
Dnajc3			Timm22		Snrpdl	
Ipo5			Ccl7		Hdac3	
Amotll			Duspl4		Cdk2ap2	
Tag In			Nme2		Corolb	
Pafahlb2			Spop		Ppplca	
Rcn2			FkbplO		Mrplll	
Csk			Ptrf		Sf3b2	
Tpml			Becnl		Eiflad	
Bnip2			Vatl		Cfll	
Tmed3			Limd2		Ssscal	
Plscrl			Syngr2		Polr2g	
Rassfl			Faml95b		Tmeml09	
Prkar2a			Histlh2ap		Prpfl9	
Crtap			Fa ml 20a		Rcll	
Slc35e4			Gadd45g		Nolcl	
Ccm2			Sfxnl		Zdhhc6	
Anxa6			Cltb		mt-Cytb	
Mprip			Serfl			
Map2k3			Mast4			
Pitpna			Sdcl			
Myolc			Soxll			
FamlOlb			Bzw2			
Tnfaipl			Bazla			
Mmd			Fa ml 77a			
Ccdcl37			Timm9			

P4hb			Synj2bp			
Arhgdia			Calml			
Sox4			Meg3			
Tubb2a			Aktl			
Pxdcl			Oxctl			
Txndc5			Ywhaz			
Bicd2			Eny2			
Tgfbi			Myc			
Pdcd6			Txn2			
Vcan			Polr3h			
Tmeml67			Zcrbl			
Zcchc9			Dazap2			
Maplb			Prrl3			
Gpx8			Carhspl			
Fst			Emp2			
Rock2			Faml 62a			
FamllOc			Fstll			
Ifrdl			Chmp2b			
Cfl2			Cdknla			
Mgat2			Clicl			
Flrt2			Mydgf			
Fbln5			Memol			
Ddx24			Srpl9			
Kiel			Reep5			
Ghr			Dpysl3			
Baspl			Ap3sl			
Mtdh			Ppic			
Plec			Gml6286			
Rpsl9bpl			Txnl4a			
Desil			Gstpl			
Tspo			Prdx5			
Slc48al			Famllla			
Fkbpll			Ak3			
Comt						
Vps8						
Lpp						

Ccdc50						
Senp5						
Ccdc80						
Phldb2						
Cldndl						
App						
Tnfrsfl2a						
Uqcc2						
Slc39a7						
Ppplrl8						
Myll2a						
Lbh						
Cyplbl						
Mcfd2						
Slc39a6						
Binl						
Egrl						
Smim3						
Tubb6						
1810055G02 Rik						
Fosll						
Neatl						
Rps6ka4						
Ppplrl4b						
Ahnak						
Fthl						
Ccdc86						
Anxal						
Acta2						
Myof						
Tm9sf3						

[0272] In particular, regulatory analysis identified a series of TFs that were upregulated in cells along the trajectory to iPSCs and predictive of the expression of the pluripotency programs (FIG. 26D). The earliest predictive TFs were expressed at day 9 (including Nanog, Sox2, Mybl2, Elf3, Tgifl, Klf2, Etv5, and Cdc51) and additional predictive TFs were induced at day 10 (including Klf4, Esrrb, Spic, Zfp42, Hesxl, and Msc). Of these 14 TFs, 9 had previously described roles in regulation of pluripotency (Nanog, Sox2, Mybl2, Klf2, Cdc51, Klf4, Esrrb, Zfp42, and Hesxl) (Aaronson et al., 2016; Boheler, 2009; Buganim et al., 2012; Hu et al., 2009; Jeon et al., 2016; Li et al., 2015; Shi et al., 2006). A further wave of predictive TFs was upregulated in the iPSC trajectory between day 12 and 14, including Obox6, Sohlh2, Ddit3, and Bhlhe40. Among these late TFs, Obox6 and Sohlh2 were particularly notable, because they were not induced in the trajectories to any other cell fate. Obox6 and Sohlh2 had not previously been reported to be involved in regulation of pluripotency, but both had been implicated in maintenance and survival of germ cell development (Park et al., 2016; Rajkovic et al., 2002).

[0273] An important change known to occur in the late stages of successful reprogramming was the reversal of X-chromosome inactivation in female cells. Our trajectory analysis identified the correct order of events as previously reported, but without the need for specialized experiments. Specifically, a study based on microscopy of cells labeled with antibodies to specific pluripotency proteins and RNA FISH for Xist (Pasque et al., 2014) showed that Xist downregulation preceded X-chromosome reactivation and positioned these events relative to the appearance of four pluripotency-associated proteins in Nanog-positive cells. Consistently, in our model, along the trajectory to successful reprogramming (but not elsewhere), cells at day 10 showed strong downregulation of Xist but did not yet display a signature of X-reactivation (FIGs. 26E, 26F, Methods). X-reactivation was complete at day 18, with the signature score having risen from 1.05 at day 10 to -1.95 at day 18, consistent with the expected increase in X-chromosome expression (FIG. 26F) (Pasque et al., 2014).

[0274] Development of extra-embryonic-like cells during reprogramming

[0275] Our trajectories showed that another subset of cells emerges from the MET Region, gained a strong epithelial signature by day 9, and went on to express a clear trophoblast signature (FIG. 27A, 27B). The trophoblast signature was detectable by day 10.5 and peaked by day 12.5, when such cells accounted for -20% of all cells in both serum and 2i conditions (FIG. 24G). Trophoblast and pre-implantation programs had previously been observed late in human reprogramming (Cacchiarelli et al., 2015)

[0276] The cells spanned a spectrum of developmental programs associated with specific trophoblasts subsets. Briefly, in normal development the extraembryonic trophoblast progenitors (TPs) gave rise to the chorion, which formed labyrinthine trophoblasts (LaTBs), and the ectoplacental cone, which gave rise to various types of spongiotrophoblasts (SpTBs) and trophoblast giant cells (TGCs), including spiral artery trophoblast giant cells (SpA-TGCs). We scored our cells with signatures we derived from placental scRNA-seq (Nelson et al., 2016) for TP, SpT, TG and SpA-TGCs (Table 15), as well as three well-characterized markers (Msx2, Gcml and Cebpa) of LaTBs (Simmons et al., 2008; Ueno et al., 2013), for which no data were available to derive signatures (FIG. 33A). A substantial number of cells expressed TP, SpTB or SpATG signatures in serum conditions and TP or SpTB signatures in 2i conditions, at 10% FDR (Figure 5C). We also observed a cluster of -200 trophoblasts cells that expressed the three LaTBs markers (in 2i but not serum), which were largely separate from those expressing signatures of ectoplacental derivatives. In addition to trophoblast-like cells, -125 cells expressed a signature (Lin et al., 2016) for the primitive endoderm (XEN-like cells), the other cell type that contributes to extraembryonic tissue (FIG. 33B, FDR 0.1%). Notably, these cells were seen only in a single replicate at a single time point (day 15.5) in serum conditions only. Two previous studies reported the generation of XEN-like cells during OKSM-induced reprogramming to iPSCs (Parenti et al., 2016, Zhao et al., 2018).

[0277] Regulatory analysis associated various TFs with the trajectory from the MET Region to the overall set of trophoblasts (FIG. 27B). TFs at day 10.5 that were predictive of subsequent trophoblast fates included several involved in trophoblast self-renewal (Gata3, Elf5, Mycn, Mybl2) (Kidder and Palmer, 2010) and early trophoblast differentiation (Ovol2, Ascl2) (Latos and Hemberger, 2016), as well as others expressed in trophoblasts but without known roles in trophoblast differentiation (Rhox6, Rhox9, Batf3 and Elf3).

[0278] Trajectory and regulatory analysis also identified TFs that were predictive of specific cell subsets. Ancestors of cells with the TP signature expressed Gata3, Pparg, Rhox9, Mytll, Hnflb, and Prdml 1. Gata3 was involved for trophoblast progenitor differentiation (Ralston et al., 2010) and Pparg was involved for trophoblast proliferation and differentiation of labyrinthine trophoblasts (Parast et al., 2009). The other TFs were known to be expressed in placenta, but their roles in cellular differentiation had not been well characterized. Ancestors of cells with the

SpTB or LaTB signature expressed Gata2, Gcml, Msx2, Hoxdl3, and Nrlh4. Gata2 was known to be involved for regulation of specific trophoblast programs (Ma et al., 1997). Gcml and Msx2 had specific roles in LaTB differentiation, EMT and trophoblast invasion (Liang et al., 2016; Simmons and Cross, 2005), respectively. Nrlh4 was detected in placental tissue, but its role in trophoblast differentiation had not been characterized. Ancestors of cells with the SpA-TGC signature expressed Handl, Bbx, Rhox6, Rhox9, and Gata2. Handl was known to be necessary for trophoblast giant cell differentiation and invasion (Scott et al., 2000). Bbx was a core trophoblast gene known to induced by upstream TFs Gata3 and Cdx2 (Ralston et al., 2010) **(FIGs. 33A-33E)**.

[0279] Neural-like cells also emerged from the MET Region during reprogramming in serum conditions.

[0280] Only in serum conditions, a third subset of cells emerged from the MET Region, gained a strong epithelial signature, and went on to develop clear neural signatures (FIGs. 27D-27F). These cells were not seen in 2i conditions, presumably due to the differentiation inhibitors in this condition. Compared to the trophoblast-like cells, the signature for neural identity emerged more slowly, by roughly two days (FIG. 24G). The ancestors of neural like cells diverged from the ancestors of trophoblasts and iPSCs by day 9 (FIG. 26B), and then underwent a rapid transition at day 12.5, losing their epithelial signatures and gaining neural signatures (FIGs. 27D, 27E). The signature was maintained through day 18, when such cells comprised 21.5% of all cells in serum conditions.

[0281] In normal neural development, neuroepithelial cells lost their epithelial identity and upregulated glial factors, transforming into radial glial cells (Florio and Huttner, 2014; Ming and Song, 201 1). Radial glial cells gave rise to astrocytes and oligodendrocytes, and in the CNS also served as progenitors for many neurons (Ming and Song, 201 1). To probe these identities, we used scRNA-Seq data from mouse brain to derive signatures that distinguished different cell types and differentiation states (Table 15). These included signatures of (i) astrocytes, oligodendrocyte precursor cells (OPCs), and neurons in adult brain from in the Allen Brain Atlas (http://www.brain-map.org), and (ii) three unlabeled clusters of radial glial cells in E18 mouse brain (Han et al., 2018), each distinguished by high expression of a different gene (Id3, GdflO, and Neurog2, respectively).

[0282] Cells in the landscape spanned multiple stages of neuronal differentiation. Cells near the base of the "neural spike" in the landscape (day 12.5-18) expressed radial glial and neural stem-cell markers (including Pax6 and Sox2) and cells further out along the spike (day 15-18) expressed markers of neuronal differentiation (including Neurog2 and Map2. About 70% of the neural-like cells had significant expression (at 10% FDR) of at least one of the six signatures (FIG. 27G). Cells with the three radial glial signatures appeared first, concurrent with the loss of epithelial identity and first gained of neural lineage identity by day 12.5 (FIG. 27F). Cells expressing the signatures derived from adult neurons and glia emerged around day 14 in the neural spike and grew in abundance for the duration of the time course. Their ancestors were concentrated in the radial glial populations on day 13.5, with a particular concentration in the GdflO RG subpopulation. While the glial populations overlapped substantially, the neurons form a distinct population with substantial substructure. The subset of cells with signatures of adult neurons included cells with canonical markers for excitatory and inhibitory neurons (Slcl7a6 and Gadl, respectively). Expression signatures that distinguished these two classes of cells showed strong, albeit incomplete, overlapped with respective programs of excitatory and inhibitory neurons in the Allen Brain Atlas (FIG. 27G, Methods).

[0283] Regulatory analysis identified TFs predictive of the overall neural-like cell population, with the top TFs all known to have roles in various stages of neurogenesis. These TFs included those known to promote early neurogenesis (Rarb, Foxp2, Emxl, Pou3f2, Nr2fl, Mytll, Neurod4), regulated late neurogenesis (Scrt2, Nhlh2, Pou2f2), regulated differentiation and survival of neural subtypes (Onecutl, Tal2, Barhll, Pitx2), and played roles in neural tube formation (Msxl, Msx3).

[0284] <u>The developmental landscape highlighted potential paracrine signals</u>

[0285] As the reprogramming landscape included a substantial and under-appreciated diversity of differentiating cell subsets, including stromal, epithelial, neural and trophoblast cells, we asked how they might affect each other as they undergo dynamic processes concurrently. In particular, paracrine signaling played a key role in normal development and had also been shown to affect reprogramming, with secretion of inflammatory cytokines enhancing reprogramming efficiency (Mosteiro et al., 2016). Accordingly, we systematically cataloged the contemporaneous occurrence of ligand-receptor pairs across cell subsets in the developmental landscape. We defined an interaction score based on the product of (1) fraction of cells of type A expressing ligand X and (2) the fraction of cells of type B expressing the cognate receptor Y, at the same time t (FIGs. 28A, 28B and 34B, Methods). We examined 180 individual cognate ligand-receptor pairs, as well as an aggregate score across all pairs between cell clusters (FIG. 34A) and across those pairs related to the SASP signature.

[0286] The landscape revealed rich potential for paracrine signaling (FIG. 28B, FIG. 34B, Table 18). In particular, we observed high interaction scores for several SASP ligands in stromal cells with receptors expressed in iPSCs, such as Gdf9 with Tdgfl (Polo et al., 2012) and Cxcll2 with Dpp4 (FIGs. 28C, 28F, 34C).

Table 18 - Potential ligand-receptor pairs between stromal cells and iPSCs, neural-like cells, and trophoblast cells ranked by standardized interaction scores

Ligand: Stromal cells. Receptor: iPSCs				**Ligand: Stromal cells. Receptor: Neural-like cells**				**Ligand: Stromal cells. Receptor: Trophoblast cells**		
Ligand-Receptor Pair	**Maximal standardized interaction score**	**Peak Score Day**		**Ligand-Receptor Pair**	**Maximal standardized interaction score**	**Peak Score Day**		**Ligand-Receptor Pair**	**Maximal standardized interaction score**	**Peak Score Day**
Gdf9.Tdgfl	55.83015277	14		Crlfl.Cntfr	76.16064491	16.5		Csfl.Csflr	111.8151997	18
Cxcll2.Dpp4	42.40247659	12.5		Fgf2.Vtn	66.31283077	18		Cxcl5.Cxcr2	102.1031447	18
Ngf.Ngfr	26.79815659	12		Clcfl.Cntfr	52.04021271	15.5		Cxcll.Cxcr2	85.46017232	18
Cclll.Dpp4	23.75254375	14		Vegfa.Vtn	39.99828338	18		II6.II6ra	70.79780689	18
Kitl.Kit	20.48156022	17.5		Bdnf.Ntrk2	38.24132006	17		Cxcl2.Cxcr2	68.04261554	18
Ccl5.Dpp4	20.22465038	12.5		Tgfb2.Vtn	37.9492686	18		Cxcl3.Cxcr2	62.67646817	17.5
Inhba.Acvr2b	18.91224205	17		Tgfbl.Vtn	37.71506462	18		II7.II2rg	57.89558657	17
Fgf7.Fgfr4	18.88448993	12		Tgfb3.Tgfbrl	32.86035119	17		Vegfa.Fltl	52.30228603	18
Nppc.Nprl	17.71660947	16.5		Bdnf.Sortl	29.14910223	17		Tg.Lrp2	45.35387653	9.5
Fgf7.Fgfr2	17.2915253	9		Ill6.Grin2a	27.83837935	13.5		Ccl2.Ackr2	44.70456305	17
Grn.Cryl	17.25111965	17		Inhba.Acvr2b	25.85377693	15.5		Sppl.Itgbl	44.39437623	18
Fgf2.Fgfr3	17.18398331	15.5		Apln.Aplnr	23.46381586	14		Ill5.II2rg	43.96702273	18
Sppl.F2	16.91745599	17		Bmpl.Adrala	21.99556814	17.5		Ccl7.Ackr2	42.35095481	17
Tgfb3.Tgfbrl	15.80306191	9		Ill6.Grin2b	21.85263644	18		Tnfsf9.Tnfrsf9	41.80288631	15.5
Bdnf.Ntrk2	15.73929703	12		Vegfa.Ephb2	21.76727834	17		Cxcll5.Cxcr2	41.37975891	18
Avp.Avprlb	15.6652861	15		Tgfbl.Tgfbrl	21.71078611	17		Vegfb.Fltl	40.59359924	18
Inhbb.Acvr2b	15.22902239	18		Ngf.Sortl	21.55867193	16.5		Fgf2.Fgfrl	40.1892017	18
Tnfsf8.Tnfrsf8	14.9661866	17.5		Ereg.Erbb4	21.23888338	17		Ill5.II2rb	37.23349427	18

Ucn2.Crhr2	14.66104887	14		Cxcll2.Cxcr4	20.66598418	16.5		II2.II2rg	34.72049417	17
Sst.Sstr3	14.53946813	12.5		Nov.Notch 1	20.64844205	17		Illrn.Illr2	34.60876011	18
Cxcll2.Cxcr4	13.99702972	9.5		Inhbb.Acvr2b	20.20541981	15.5		Bmp4.Bmpr2	33.37381523	18
Fgfl.Fgfr4	13.23808582	14		Egf.Vtn	20.11367671	14.5		Ppbp.Cxcr2	33.31119733	17
Gdf6.Bmprlb	13.23695383	11.5		Fgf7.Fgfr2	19.85021209	9		Flt3I.FIt3	31.32026205	17
Gdf9.Bmprlb	12.81536347	11.5		FgflO.Fgfr2	19.77063453	12		Inhba.Acvr2b	31.21420166	16.5
Gdf5.Acvr2b	12.41295756	17.5		Fgf2.Fgfr3	19.20901825	18		II2.II2rb	31.17852066	17
Cxcl3.Cxcr2	12.28144255	9		Inhba.Igsfl	19.00415822	13.5		Inhbb.Acvrlb	31.08869402	18
CxcllO.Dpp4	12.0118101	16.5		Pomc.Vtn	18.61879864	14		Inhba.Acvrlb	30.95069812	18
Tnfsfll.Tnfrsflla	11.98501062	18		Tgfb2.Tgfbrl	18.40997602	17		Ccl8.Ackr2	30.92303758	17
Tnfsfll.Med24	11.31495458	17		Gdf9.Tdgfl	18.12847923	10.5		Pgf.Fltl	28.55965416	17
Bdnf.Inpp5k	11.02760154	17		Gdnf.Gfral	17.94758176	18		Tgfb3.Tgfbrl	28.48415966	18
Cxcl5.Cxcr2	10.76725496	9		Ednl.Ednrb	17.81157803	17		Inhba.Tgfbr3	27.97080183	18
Bmp2.Bmprlb	10.52856679	11.5		Gdfll.Acvr2b	16.93911315	15.5		Inhbb.Acvr2b	27.64710304	18
Inhba.Acvrlb	10.45689595	15.5		Gdf5.Bmprlb	16.87028377	17		Ccl3.Ackr2	27.17947452	14.5
Fgfl.Fgfr3	9.904359216	14		Gdf5.Acvr2b	16.68587549	15.5		Tgfb3.Sdc4	26.70563028	18
Tgfb3.Eng	9.606914311	18		Igfl.Igflr	16.40043325	17.5		Inhba.Acvrll	24.8733331	16.5
Crlfl.Cntfr	9.491489628	9		Ngf.Ngfr	16.1554284	9		Wnt5a.Fzd5	24.08669584	18
Tg.Lrp2	9.311152429	9.5		Cxcl5.Ackrl	15.81074369	17		Egf.Erbb3	22.88090865	18
Nppa.Nr5a2	9.196846339	15.5		Tg.Lrp2	15.56587296	9.5		Gdf5.Acvr2b	22.79535492	16.5
Sppl.Itgbl	9.094293313	9		Ill6.KcnjlO	15.40280917	15		Tgfbl.Itgb6	22.73325122	18
Tgfb3.Sdc4	8.962618473	18		Ccl2.Ackrl	14.80314224	17		Vegfc.Flt4	22.64781847	18
Avp.Avpr2	8.816318411	16		Illrn.Illr2	14.70537108	17		Vegfa.Kdr	21.61880314	13
Bmp4.Bmprlb	8.789458439	11.5		Wnt5a.Fzd2	14.59368545	16.5		Ill8.Ill8rap	21.45320636	18
Gdfll.Acvr2b	8.657009643	17.5		Inhbb.Igsfl	14.56070266	13.5		Tgfb2.Tgfbr3	21.43696896	12.5
Ctgf.Egfr	8.474450513	9		Ccll2.Ackrl	14.48343455	15		Fgf7.Fgfr2	21.27556999	9
Nov.Notch 1	7.853128492	9.5		Ccl7.Ackrl	14.45732094	17		Ccll2.Ackr2	20.65465765	15
Cxcll.Cxcr2	7.825570863	9		Fgfl.Fgfr3	13.98128161	14		Tgfbl.Tgfbr3	19.07802333	18
Pomc.Mc5r	7.803289928	13		Cort.Sstr2	13.83366019	14.5		Cclll.Ackr2	19.06812091	16.5
Inhba.Acvr2a	7.697312114	10		Vegfa.Kdr	13.52841955	17		Ccl28.Ackr2	19.0608243	16.5
Ill6.Cd4	7.691300029	16		Bmp4.Bmprlb	13.17024743	17		Kitl.Kit	18.32774459	10
Hcrt.Npffr2	7.611421106	14.5		Igfl.Igsfl	13.1615924	13.5		Gdfll.Acvr2b	17.1611013	16.5
Nppa.Nprl	7.327171012	15.5		Inhba.Acvr2a	12.86079359	15.5		Bdnf.Inpp5k	16.94541624	18
Fgf2.Fgfrl	6.935257539	18		Gdnf.Gfra2	12.82585678	18		Ccl5.Ackr2	16.65970084	10.5
Inhbb.Acvrlb	6.8878958	15.5		Ntf3.Ntrk2	12.69375513	14		Ngf.Ngfr	16.41502139	9
Ccll7.Ccr4	6.846358767	17		Cxcll.Ackrl	12.64243264	17		Igfl.Igflr	16.27850014	18

Ill6.Grin2b	6.789839819	14.5	Fgf2.Fgfrl	12.31083274	18	Bmp2.Bmpr2	15.99972954	18
Bdnf.Sortl	6.67375428	9	Vegfa.Nrp2	12.23441434	18	Tgfbl.Acvrll	15.96504429	16.5
Tgfb2.Tgfbrl	6.519268162	9	Bmp6.Acvr2b	12.1758211	13.5	Gdf5.Bmpr2	15.58998037	16.5
Ntf3.Ntrk2	6.438685726	12	Hbegf.Erbb4	12.00500039	14.5	Tgfb2.Tgfbrl	15.53065603	18
Ccl3.Ccr5	6.407610415	12.5	Vegfc.Kdr	11.97527882	18	Tgfbl.Tgfbrl	15.49109459	18
Ptn.Plxnb2	6.364004505	9	Ccll7.Ackrl	11.93535268	16	Inha.Tgfbr3	14.94814105	18
Egf.Erbb3	6.33209249	17	Cxcl3.Cxcr2	11.79741482	9	Ccl27a.Ackr2	14.35654443	17
Fgf9.Fgfr3	6.17049013	15.5	Wnt2.Fzd9	11.76547196	14.5	Pf4.Ldlr	13.49144052	17.5
Ntf3.Ntrk3	6.071479576	12.5	Tnfsfll.Med24	11.58428169	17	Vegfc.Kdr	13.42241254	12.5
Wnt5a.Fzd5	6.049412152	17.5	Cxcll5.Ackrl	11.39063421	16	FgflO.Fgfr2	12.93211376	12
Ill6.Kcnj4	5.956600472	9	Cxcl5.Cxcr2	10.81475088	9	Pdgfc.Pdgfra	12.7181284	18
FgflO.Fgfr2	5.735961453	10	Sppl.Itgbl	10.57557893	9	Ccl25.Ackr2	12.58225578	10.5
Csf3.Csf3r	5.660332275	18	Ccl8.Ackrl	10.24654012	18	Crlfl.Cntfr	12.56270017	9
Ngf.Sortl	5.631416895	9	Gdf5.Acvr2a	9.947335355	16.5	Inhba.Acvrl	12.49512116	18
Wnt2.Fzd9	5.625683619	13	Inhbb.Acvr2a	9.83065505	17.5	Inhbb.Acvrl	12.17571989	18
Ngf.Ntrkl	5.482536008	18	Bmp2.Bmprlb	9.823905055	17	Bmp4.Bmprla	12.13592365	18
Ccl2.CcrlO	5.204305876	9	Ngf.Ntrkl	9.765431603	15.5	Hgf.Met	11.85706092	18
Gdf5.Bmprlb	5.164323069	11.5	Ctgf.Egfr	9.510948488	9	Avp.Avprlb	11.8443167	12.5
Ccl7.CcrlO	5.03794601	9	Ill6.Grin2c	9.210664243	16.5	Wnt5a.Lrp6	11.2866016	18
Inhba.Igsfl	4.652799622	16.5	Igf2.Vtn	9.08515341	15.5	Illrn.Illrl	11.21386458	18
Igfl.Igsfl	4.623901723	16.5	Fgf9.Fgfr3	8.929720296	13	Npff.Npffr2	11.12680175	12.5
Kitl.Epor	4.572546653	9	Ucn2.Crhr2	8.529535163	10	Gpil.Amfr	11.09557616	18
Bmp6.Bmprlb	4.21969712	11.5	Gdf9.Bmprlb	8.458633534	12.5	Ccl2.Ccr5	10.87678026	17
Ill6.Grin2a	4.182303182	12	Cxcll.Cxcr2	8.317259429	9	Inhba.Acvr2a	10.71764165	18
Tgfbl.Tgfbrl	4.165309406	9	Pnoc.Oprll	8.170486417	13	Inhbb.Acvr2a	10.62573575	18
Hmgbl.Pgr	4.162814163	9.5	Inha.Acvr2a	8.005902758	15.5	Ccll7.Ccr4	10.22222634	11.5
Tnfsfl3b.Tnfrsfl7	4.077062584	16.5	Inhba.Acvrlb	7.58971181	9.5	Vegfa.Lyvel	9.978529316	11.5
Ill6.Grin2c	3.818702923	17	Fgf7.Fgfr4	7.313765731	16	Lif.Lifr	9.836393324	16.5
Crh.Crhr2	3.804963778	14	Ptn.Plxnb2	7.174330257	9	II25.Ill7rb	9.820316363	16
Tgfbl.Eng	3.789167413	17	Btc.Erbb4	7.130596933	14.5	Ccl8.Ccr5	9.277471947	16.5
Ccl5.Ccr5	3.765684384	10.5	Grn.Cryl	7.038337946	16.5	Ill6.KcnjlO	9.099847388	14.5
Ccl3.Ackr4	3.748657973	12.5	Ill6.Kcnj2	7.031491551	18	Bdnf.Ntrk2	9.027486627	12.5
Ccl2.Ccr5	3.746070011	12.5	Ednl.Ednra	6.737910303	17.5	Ednl.Ednrb	8.719812556	14
Gdf5.Acvr2a	3.726614996	16	Avp.Oxtr	6.701328931	16.5	Cxcll2.Cxcr4	8.696493411	17
Npff.Npffr2	3.71584242	14.5	Tgfb3.Sdc4	6.648807091	9	Fgf9.Fgfrl	8.617860569	18
Inhbb.Igsfl	3.660059949	16.5	Ill6.Kcnj4	6.296091418	9	Sppl.F2	8.219496273	13.5
Bmp6.Acvr2b	3.613241885	13.5	Sppl.F2	6.250718711	14.5	Ptn.Plxnb2	8.085698538	9

Lif.Lifr	3.59302184	12.5		Adm.Calcrl	6.127364131	18		Tnfsfll.Med24	8.080587047	18
Inhbb.Acvr2a	3.573362535	16		Artn.Gfra3	6.100580729	18		Ctgf.Egfr	8.025815916	9
Tgfb2.Eng	3.493150482	18		Ccl5.Ackrl	6.08281121	16		Ghrl.Ptger3	7.831218363	15
Tnfsfl3b.Tnfrsfl3b	3.485242199	14		Tgfb3.Eng	6.075334099	9		Ctfl.Lifr	7.478421588	18
Bmp2.Bmprla	3.421538818	9		Gdf6.Bmprlb	5.814695498	17.5		Pdgfd.Pdgfrb	7.440471865	18
Bmp2.Eng	3.277644443	12		Hmgbl.Pgr	5.524547346	9.5		Gdf5.Acvr2a	7.437486529	17.5
Pf4.Ldlr	3.252582504	11.5		Wnt5a.Lrp6	5.416442742	15		Cxcll2.Dpp4	7.386223592	12.5
Ntf5.Ngfr	3.228481212	12		Vegfa.Lyvel	5.365931818	16.5		Cclll.Ccr5	7.344244377	16.5
Ccl5.Ccr4	3.054614918	17		Ccll7.Ccr4	5.313995351	9.5		Gdf5.Bmprla	7.242141121	17.5
Pgf.Nrp2	3.013909017	9		Sst.Sstr2	4.993026408	12.5		Artn.Gfra3	6.624252893	16
Fgf8.Fgfr4	3.01220056	14		Vegfa.Fltl	4.860449031	13.5		Ill8.Illrl2	6.470340015	18
Artn.Gfra3	3.008145345	16		Bmp6.Bmprlb	4.604550067	16.5		Inha.Acvr2a	6.410004454	18
				Egf.Erbb3	4.487189494	10.5		Gdf6.Bmpr2	6.362677796	18
				Kitl.Epor	4.470894246	9		Ntf3.Ntrk2	6.34714587	12.5
				Gdf9.Acvr2a	4.461925767	12.5		Gdf5.Acvrl	6.33836936	18
				Ccl2.CcrlO	4.287535378	9		Tslp.Prnp	6.263327318	18
				Fgf9.Fgfr2	4.104799154	11		Gdf9.Tdgfl	6.170602382	10.5
				Ill6.Cd4	4.102677906	15.5		Bdnf.Sortl	5.94172272	9
				Ccl2.Ccr5	4.06128803	18		Bmp2.Acvrl	5.90978443	18
				Ntf3.Ntrkl	4.045425855	15.5		Bmp6.Acvr2b	5.871545931	13.5
				Bmp2.Bmprla	4.007512362	9		Tnfsfll.Tnfrsflla	5.868170248	15.5
				Pdgfc.Pdgfra	4.000578173	18		II6.II6st	5.857031136	18
				Bmp4.Bmprla	3.973107083	17		Kitl.Epor	5.493268145	14
				Ghrl.Ptger3	3.959803347	15		Hmgbl.Pgr	5.439455664	9.5
				Illl.Illlral	3.931542903	16.5		Gdf9.Bmpr2	5.301534907	17.5
				Ccl7.CcrlO	3.86216627	9		Ngf.Sortl	5.181692923	9
				Gdf5.Bmprla	3.812514632	16.5		Tnfsfl3b.Tnfrsfl3b	5.166928123	15.5
				Ntf5.Ntrk2	3.800422565	15.5		Ucn2.Crhr2	5.15524664	9
				Ntf3.Ntrk3	3.791204113	13		Fgfl.Fgfrl	5.090269326	18
				Ccl8.Ccr5	3.6877203	18		Pdgfa.Pdgfra	4.960203778	18
				Vegfb.Fltl	3.67289066	13.5		Fgf7.Fgfr4	4.959156503	12
				Ccl5.Ccr4	3.652617678	9.5		Nov.Notch1	4.944351734	9.5
				Inhba.Acvrl	3.386360757	18		Bmp2.Bmprla	4.828229043	18
				Inhbb.Acvrl	3.330148881	18		Fgf2.Fgfr3	4.718080894	13.5
				Wntl.Fzd9	3.30422519	12.5		Grn.Cryl	4.629614942	9
				Npff.Npffrl	3.243049647	16		Tgfb3.Eng	4.541775835	9

								TnfsflO.TnfrsflOb	**4.456880919**	**16.5**
								Hcrt.Hcrtrl	**4.407762506**	**14.5**
								Ccl5.Ccr5	**4.218364077**	**16**
								H16.Kcnj4	**4.184296843**	**9**
								Ghrl.Ptgir	**4.00490292**	**15**
								Cxcll6.Cxcr6	**3.995533009**	**18**
								Ccl3.Ccr5	**3.825939759**	**12.5**
								Ill6.Grin2c	**3.804620341**	**14**
								Ccl5.Ccr4	**3.700028296**	**13**
								Ill7b.Ill7rb	**3.43715641**	**10.5**
								Hmgbl.Ar	**3.425935882**	**11**
								Ntf3.Ntrkl	**3.384388196**	**13**
								Ngf.Ntrkl	**3.213785377**	**13**
								Ccll2.Ccr5	**3.032941015**	**16**

[0287] Analysis of the neural-like cells revealed particularly interesting interaction scores involving Cntfr (FIGs. 28D, 28G, 34D), an 116-family co-receptor whose activation played critical roles in neural differentiation and survival (Elson et al., 2000; Nakashima et al., 1999). On day 11.5 in serum conditions, one day before the early neuronal signatures appear, neural ancestors upregulated expression of Cntfr; expression was 4.6-fold higher in epithelial cells that were neural ancestors versus those that were not. Just before, on day 10.5, stromal cells began expressing three activating ligands for Cntfr (Crlfl, Lif, Clcfl). We speculated that these events may help trigger the program of neural differentiation among a subset of epithelial cells in serum conditions. The analysis also revealed a potential interaction involving the ligand-receptor pair Bdnf-Ntrk2, which had been implicated in promoting neuronal development, maturation and survival (Chen et al., 2015; Jukkola et al., 2006; Yun et al., 2008) (FIGs. 28D, 28G, 34D). The same ligand-receptor interactions were seen in 2i conditions, but the MEK inhibitor in 2i medium would be expected to block Cntfr signaling and subsequent neural differentiation.

[0288] Trophoblast-like cells also showed notable interaction scores, including Csfl and Csflr (FIGs. 28E, 28H). In early placental development, Csfl was expressed in maternal columnar epithelial cells and Csflr was expressed in fetal trophoblasts, suggesting a functional role of this interaction in trophoblast development and differentiation. Many of the other top-ranked interactions were between a single receptor in trophoblast cells (Cxcr2) and multiple members of the same ligand family (Cxcl5, Cxcll, Cxcl2, Cxcl3, and Cxcll5) (FIGs. 24E, 24H, 34E). Cxcr2 had been shown to be necessary for trophoblast invasion in human trophoblast cells (Vandercappellen et al., 2008; Wu et al., 2016).

[0289] RNA expression revealed genomic aberrations in stromal and trophoblast-like cells

[0290] We hypothesized that some cell types might harbor detectable genomic aberrations. In particular, trophoblasts were known to undergo endocycles of replication in vivo (Edgar et al., 2014), resulting in selective amplification of specific genomic regions containing functionally important genes (Hannibal and Baker 2016). Additionally, our stromal cells exhibited signs of stress and cell death which may be associated with genomic aberrations.

[0291] To identify potential genomic aberrations, we scored the scRNA-Seq data for large regions showing coherent increases or decreases in gene expression, following successful approaches we developed to identify aberrant regions in individual tumor cells in a patient (Patel et al., 2014). We searched copy-number variations at the level of whole chromosomes and subchromosomal regions spanning 25 consecutive housekeeping genes (median size 25 Mb) (STAR Methods). To evaluate the detection of subchromosomal events, we analyzed scRNA-Seq data from oligodendroglioma (Tirosh et al. 2016): the method had high specificity, but sensitivity to detect only about one-third of events.

[0292] Whole-chromosome aneuploidies were detected in 4.0% of trophoblast cells and 2.1%o of stromal cells, compared to only 1.1% of all other cells across the landscape. Most whole-chromosome events were consistent with loss or gain of a single copy of the chromosome (FIG. 281). Subchromosomal events were detected in 6.9%> of trophoblast cells and 3.2% of stromal cells, compared to only 1.2% in most other cells types and 0.4% in neural cells (Figure 6J); the true proportions are likely to be about 3-fold higher, given the estimated sensitivity.

[0293] Trophoblast-like cells showed recurrent events at a higher frequency than stromal cells. Among trophoblast cells harboring aberrations, 8.6% were detected as carrying a recurrent event involving apparent duplication (50% higher expression) of a region containing 74 genes (FIG. 28K). Among the genes are Wnt7b, which was required for normal placental development (Parr et al., 2001); Prr5, which mediates Pdfgb signaling required for development of labyrinthine cells (Ohlsson et al., 1999; Woo et al., 2007); and several genes identified as 'core trophoblast genes' (Cyb5r3, Cenpm, Srebf2, and Pmml). The top 15 recurrent events also included the amplification of the prolactin gene cluster on chromosome 13 in 1% of cells. These observations suggested that the trophoblast-associated mechanisms of genomic alteration may be expressed, to some extent, in our trophoblast-like cells.

[0294] In the stromal cells with evidence of genomic aberration, the most common recurrent events had lower frequency. Notably, however, the most frequently amplified region contained cell cycle inhibitors Cdkn2a, Cdkn2b, and Cdkn2c, while the most frequently lost region contained Cdkl3, which promotes cell cycling, and Mapk9, loss of which promotes apoptosis. These observations suggested that genomic alterations in these regions may contribute to development stromal cells.

[0295] Forced expression of Obox6 enhanced reprogramming

[0296] Finally, we explored whether some of the new TFs identified by regulatory analysis along the trajectory to iPSCs might provide ways to increase reprogramming efficiency. In principle, TFs could increase the efficiency of reprogramming in several ways, including increasing the transition frequency to iPSC precursors, boosting the growth rate of iPSC precursors, reducing alternative fates of other epithelial-related fates, or increasing supportive paracrine signaling from non-iPS cells.

[0297] We focused on Obox6, which our regulatory analysis discovered as the TF most strongly correlated with reprogramming success, among those not previously implicated in the process. Obox6 (oocyte-specific homeobox 6) is a homeobox gene of unknown function that is preferentially expressed in the oocyte, zygote, early embryos and embryonic stem cells (Rajkovic et al., 2002). (Although Obox6 was the only Obox family member detected in our experiment, we note that a better-studied oocyte-specific homeobox Oboxl has been shown to enhance reprogramming efficiency, promote MET, and be able to substitute for Sox2 in reprogramming (Wu et al., 2017)). While Obox6 was expressed only in a small fraction of cells (<1%) before day 12, cells expressing Obox6 during day 5.5 to day 8 are highly biased toward the MET Region, with 94% being in the top 50% of cells with respect to the proportion of descendants in this region (FIG. 29A).

[0298] We tested whether expressing Obox6 together with OKSM during days 0-8 can boost reprogramming efficiency. We infected our secondary MEFs with a Dox-inducible lentivirus carrying either Obox6, the known pluripotency factor Zfp42 (Rajkovic et al., 2002; Shi et al., 2006), or no insert as a negative control. Both Obox6 and Zpf42 increased reprogramming efficiency of secondary MEFs by ~2-fold in 2i and even more so in serum, with the result confirmed in multiple independent experiments (FIGs. 29B, 29C, and 36A-36F). Assays in primary MEFs showed similar increases in reprogramming efficiency (FIGs. 26A-36F).

[0299] Together, these computational and experimental results suggested that the role of Obox6 in reprogramming merits further study.

[0300] In addition, we identified GDF9 that can significantly booster reprogramming efficiency. We added GDF9 to the medium from day 8. We observed more Oct4-GFP positive colonies (iPSCs) **(FIG. 37).** We also confirmed that we saw more iPSCs after adding GDF9 by scRNA sequencing.

[0301] **FIG. 38** shows adding GDF9 to the medium resulted in more iPSCs.

[0302] **Discussion**

[0303] Understanding the trajectories of cellular differentiation was important for studying development and for regenerative medicine. Large-scale, single-cell profiling had dramatically advanced progress toward this goal. However, the challenge of turning snapshots from single-cell profiling into accurate movies of cellular differentiation had not yet been fully solved. Here, we described two resources for the scientific community: a new analytical approach to reconstructing trajectories, and a massive dataset of 315,000 cells from time courses of classic reprogramming from fibroblasts to iPSCs under two conditions. By applying the approach to the dataset, we shed new light on this well-studied problem, and provide a template for future studies in other systems.

[0304] An optimal transport framework to model cell differentiation

[0305] Waddington-OT provided an inherently probabilistic approach that described transitions between time points in terms of stochastic couplings, derived from a modified version of the mathematical method of optimal transport. The approach yielded a natural concept of trajectories in terms of ancestor and descendant distributions for any set of cells at a given time point. This allowed us gracefully to recover, for example, branching events (by the emergence of bimodality in the descendant distribution) or shared vs. distinct ancestry between two cell sets (by convergence of the ancestor distributions) (FIGs. 23C-23E). The trajectories can then be used to study differentiation between classes of cells at different times, including creating regulatory models to infer TFs involved in activating specific gene-expression programs. Our model did not impose strict structural constraints a priori on the nature of these processes, allowing for gradual changes over time rather than sharp discrete transitions. Moreover, OT can be applied to even a single pair of time points (if the transition is expected to be sufficiently smooth) and thus can be helpful even for a small experimental scheme. Indeed, we validated Waddington-OT by testing its ability to accurately infer cellular distributions at held-out intermediate time points and by showing that its results are robust across wide variation in parameters.

[0306] Waddington-OT differred from previous approaches because it (i) did not attempt to force cells onto a simple branching graph, (ii) made explicit use of temporal information, and (iii) allowed for cell growth and death. We also found that Waddington-OT appeared to perform better than several graph-based methods, at least for studying cellular reprogramming from fibroblasts to iPSCs (FIGs. 35A-35B, Methods). Specifically, the widely and successfully used program Monocle2 (Qiu et al., 2017) generated trajectories that a) were inconsistent with known information about time (day 18 stromal cells give rise to essentially all cells after day 0), and b) placed neural and iPS together as one terminal state. The recently developed program URD (Farrell et al., 2018) could avoid the latter problem by finding trajectories to specific cell sets of interest, but a) it generated trajectories which contradicted the gradual MET/Stromal fate specification we saw in our data (in URD, the stromal branch completely diverges at day 0.5), and b) the binary nature of the URD tree could not capture the multifurcation of neural, iPS, trophoblast and epithelial cells from MET.

[0307] Tracking cell differentiation trajectories and fates in a diverse reprogramming landscape

[0308] Although the reprogramming of fibroblasts to iPSCs had been intensively studied since it was discovered by Yamanaka, our study shedded new light on the process - providing insights that could only be obtained from large-scale single-cell profiles across dense time courses matched with appropriate analytical methods.

[0309] First, single-cell profiling with large numbers of cells along a dense time course revealed remarkable and unappreciated diversity in the reprogramming landscape, with large classes of cells having distinct biological programs, related to distinct states and tissues (pluripotency, trophoblasts, neural tissue, epithelium and stroma). In earlier studies based on bulk RNA analysis, we and others had detected expression of individual genes characteristic of various lineages during reprogramming. (Mikkelsen et al., 2008; O'Malley et al., 2013; Parenti et al., 2016). Studying these classes in greater detail, we found a tremendous richness of cells expressing distinct gene-expression programs associated with specific cell types in vivo. Examples included: (i) within iPSC-like cells, programs associated with 2-, 4-, 8-, 16-, and 32-cell stage embryos; (ii) within extra-embryonic-like cells, programs associated with several distinct types of trophoblasts and programs associated with primitive endoderm (at one time point); (iii) within neural-like cells, programs associated with astrocytes, oligodendrocytes, and neurons, as well as specific subprograms associated with excitatory and inhibitory neurons; and (iv) within stromal-like cells, distinct programs associated with a wider range of stromal cells than simply MEFs. Further work will be needed to determine the extent to which these cell types adopt the full identity of natural cell types that they resemble.

[0310] This dramatic diversity raised several key questions that Waddington-OT has helped us begin to address, including: (1) What are the differentiation and fate trajectories that span these cell subsets? When do they diverge, from which ancestors, and to which cells do they give rise? (2) What cell intrinsic regulatory mechanisms may drive each fate, especially transcription factors? (3) What might be the role of cells of different types at cross-communicating and supporting across differentiation trajectories and fates in general, and for the iPSC fate in particular?

[0311] First, our trajectory and regulatory analysis allowed us to build a model that synthesizes a comprehensive view of the differentiation and fate trajectories in the landscape (FIG. 29D). We highlighted several key fate decisions, in a manner that allowed us to understand their gradual and continuous nature. During the initial phase of reprogramming, cells began to diverge in two alternative directions: toward stromal cells or toward an MET state (FIG. 29D, blue and purple). In the MET direction this divergence was not sharp: although some ancestors exhibited biases in cell fate as early as day 1.5, cells continued to 'switch' their fate preference from MET to Stromal up to day 8 (FIGs. 29A-29D, arrows from purple to blue zones). In contrast, the Stromal Region was terminal, and the reverse phenomenon was not seen by our model. Following withdrawal of dox at day 8, the cells in the MET state gave rise to iPSC-, trophoblast-, neural-, and epithelial-like cells. We found no evidence that particular cells had biases towards any of these fates before this point, whereas our analysis clearly distinguished the biases that arise once dox was withdrawn. The ancestors that would lead to iPSCs were distinguished early after withdrawal (day 9), and they passed through a narrow bottleneck towards iPSC. Conversely, other cells in the MET region first assumed an epithelial-like state, with ancestors leading to trophoblasts vs. neural cells (in serum) becoming distinguished a few days later. Within neural cells (in serum) and trophoblast-like cells (in both conditions), there was substantial additional divergence, which we could at times trace to additional divergence between ancestors at later time point. For example, the radial glial population expressing GdflO RG at day 13.5 was enriched for ancestors of later emerging neuron-like cells.

[0312] Second, by characterizing events that occurred along the trajectory toward any cell class, we identified TFs that might drive subsequent fates (FIG. 29D). Along the path toward pluripotency, we readily rediscovered known TFs, validating our approach, but also identified several new TFs not previously implicated in the process. We tested one such new TF, Obox6, which was associated with a strong bias toward MET early and toward pluripotency late; we found that forced expression of Obox6 increased reprogramming efficiency. Along paths to other fates, we similarly rediscovered TFs known to play a role in differentiation of the corresponding cells in vivo, as well as identified TFs that were expressed in the target cell type but had not been implicated in differentiation per se.

[0313] Third, contemporaneous expression of receptor-ligand pairs across cell subsets highlighted potential paracrine interactions between the stromal cells and the iPSC-like, neural-like and trophoblast-like cells, which might play key roles in the initial differentiation and maintenance of these cell types. If many of these potential interactions could be validated by experimental assays, it would suggest that efficient reprogramming requires alternative cell types, or the exogenous replacement of the factors they supply. Additionally, single-cell expression revealed likely regions of genomic aberration; the frequency of such events was significantly higher in our trophoblast and stromal cells, consistent with known biological properties of these cell types.

[0314] Prospects for models and studies of differentiation and development

[0315] Our method captured several key aspects of cellular differentiation and, importantly, can be extended to capture additional features. First, the framework currently assumed that a cell's trajectory depended only on its current gene-expression levels. As it became possible to perform single-cell profiling simultaneously for gene expression and epigenomic states, one can readily incorporate both types of information. Second, our framework for learning regulatory models assume that trajectories are cell autonomous, but may be extended to incorporate intercellular interactions, such as the potential paracrine signaling postulated here, by using optimal transport for interacting particles (Ambrosio et al., 2008; Santambrogio, 2015) (STAR Methods). Third, various methods are being developed for obtaining lineage information about cells, based on the introduction of barcodes at discrete time points or even continuously (Frieda et al., 2017; McKenna et al., 2016). Barcodes can be used to recognize cells that descend from a recent common ancestor cell, but do not currently directly reveal the full gene-expression state of the ancestral cell. However, they can be incorporated into our optimal-transport framework to improve the inference of ancestral cell states. Finally, our method can be refined to analyze multiple time points simultaneously, rather than just pairs of consecutive time points; this can be particularly useful for situations where the number of cells at different time points varies significantly.

[0316] In summary, our findings indicated that the process of reprogramming fibroblasts to iPSCs unleashed a much wider range of developmental programs and subprograms than previously characterized.

[0317] **References**

- Aaronson, Y., Livyatan, L, Gokhman, D., and Meshorer, E. (2016). Systematic identification of gene family regulators in mouse and human embryonic stem cells. Nucleic Acids Research 44, 4080-4089.
- Daniel et al., (2018). A revised airway epithelial hierarchy includes CFTR-expressing ionocytes. Nature 2018, accepted.
- Ambrosio, L., Gigli, N., and Savare, G. (2008). Gradient flows: in metric spaces and in the space of probability measures (Springer Science & Business Media).
- Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. Icwsm, 8:361-362.

- Bendall, S.C., Davis, K.L., Amir, E.-a.D., Tadmor, M.D., Simonds, E.F., Chen, T.J., Shenfeld, D.K., Nolan, G.P., and Pe'er, D. (2014). Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157, 714-725.
- Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S., and Li, M. S. (2015). Package FNN.
- Boheler, K.R. (2009). Stem cell pluripotency: a cellular trait that depends on transcription factors, chromatin state and a checkpoint deficient cell cycle. Journal of cellular physiology 221, 10-17.
- Briggs, J.A., Weinreb, C , Wagner, D.E., Megason, S., Peshkin, L., Kirschner, M.W., and Klein, A.M. (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science.
- Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150, 1209-1222.
- Cacchiarelli, D., Trapnell, C , Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., Ho Sui, S.J., Wu, Z., Akopian, V., Gifford, C.A., Doench, J., Rinn, J.L., Daley, G.Q., Meissner, A., Lander, E.S., and Mikkelsen, T. (2015). Integrative Analyses of Human Reprogramming Reveal Dynamic Nature of Induced Pluripotency. Cell 162.
- Cannoodt, R., Saelens, W., Sichien, D., Tavernier, S., Janssens, S., Guilliams, M., Lambrecht, B.N., De Preter, K., and Saeys, Y. (2016). SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. bioRxiv.
- Chen, E.Y., Tan, CM., Kou, Y., Duan, Q., Wang, Z., Meirelles, G.V., Clark, NR., and Ma'ayan, A. (2013). Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 14, 128.
- Chen, Q., Zhang, M., Li, Y , Xu, D., Wang, Y , Song, A., Zhu, B., Huang, Y., and Zheng, J.C. (2015). CXCR7 Mediates Neural Progenitor Cells Migration to CXCL12 Independent of CXCR4. Stem cells (Dayton, Ohio) 33, 2574-2585.

- Chizat, L., Peyre, G., Schmitzer, B., and Vialard, F.-X. (2017). Scaling algorithms for unbalanced transport problems. arXiv preprint arXiv:160705816v2.
- Coppe, J.-P., Desprez, P.-Y., Krtolica, A., and Campisi, J. (2010). The senescence-associated secretory phenotype: the dark side of tumor suppression. Annual Review of Pathological Mechanical Disease 5, 99-1 18.
- Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transport. Paper presented at: Advances in neural information processing systems.
- Elson, G.C., Lelievre, E., Guillet, C , Chevalier, S., Plun-Favreau, H., Froger, J., Suard, I., de Coignac, A.B., Delneste, Y., and Bonnefoy, J.-Y. (2000). CLF associates with CLC to form a functional heteromeric ligand for the CNTF receptor complex. Nature neuroscience 3, 867.
- Falco, G , Lee, S.L., Stanghellini, I., Bassey, U.C., Hamatani, T., and Ko, M.S. (2007). Zscan4: a novel gene expressed exclusively in late 2-cell embryos and embryonic stem cells. Developmental biology 307, 539-550.
- Farrell, J.A, Wang, Y., Riesenfeld, S.J., Shekhar, K., Regev, A , and Schier, A.F. (2018). Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. Science.
- Fincher, C.T., Wurtzel, O., de Hoog, T., Kravarik, K.M., and Reddien, P.W. (2018). Cell type transcriptome atlas for the planarian <em>Schmidtea mediterranea</em>. Science.
- Florio, M., and Huttner, W.B. (2014). Neural progenitors, neurogenesis and the evolution of the neocortex. Development 141, 2182-2194.
- Fonseca, E.T.d., Man?anares, A.C.F., Ambr®,Esio, C.E., and Miglino, M.A.I. (2013). Review point on neural stem cells and neurogenic areas of the central nervous system. Open Journal of Animal Sciences Vol.03No.03, 6.
- Frieda, K.L., Linton, J.M., Hormoz, S., Choi, J., Chow, K.-H.K., Singer, Z.S., Budde, M.W., Elowitz, M.B., and Cai, L. (2017). Synthetic recording and in situ readout of lineage information in single cells. Nature 541, 107.
- Froidure, A., Mar chal-Duval, E., Ghanem, M., Gerish, L., Jaillet, M., Crestani, B., and Mailleux, A. (2016). Mesenchyme associated transcription factor PRRXl: A key regulator of IPF fibroblast. European Respiratory Journal 48.

- Gegenschatz-Schmid, K., Verkauskas, G., Demougin, P., Bilius, V., Dasevicius, D., Stadler, M.B., and Hadziselimovic, F. (2017). DMRTC2, PAX7, BRACHYURY/T and TERT Are Implicated in Male Germ Cell Development Following Curative Hormone Treatment for Cryptorchidism-Induced Infertility. Genes 8, 267.
- Goolam, M., Scialdone, A., Graham, S.J.L., Macaulay, I.C., Jedrusik, A., Hupalowska, A., Voet, T., Marioni, J.C., and Zernicka-Goetz, M. (2016). Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. Cell 165, 61-74.
- Gouti, M., Briscoe, J., and Gavalas, A. (201 1). Anterior Hox genes interact with components of the neural crest specification network to induce neural crest fates. Stem cells (Dayton, Ohio) 29, 858-870.
- Haghverdi, L., Buettner, F., and Theis, F.J. (2015). Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989-2998.
- Haghverdi, L., Buettner, M., Wolf, F.A., Buettner, F., and Theis, F.J. (2016). Diffusion pseudonyme robustly reconstructs lineage branching. bioRxiv, 041384.
- Han, X., Wang, R., Zhou, Y., Fei, L., Sun, H., Lai, S., Saadatpour, A., Zhou, Z., Chen, H., Ye, F., et al. (2018). Mapping the Mouse Cell Atlas by Microwell-Seq. Cell 172, 1091-1107.el017.
- Hayashi, Y., Hsiao, E.C., Sami, S., Lancero, M., Schlieve, C.R., Nguyen, T., Yano, K., Nagahashi, A., Ikeya, M., Matsumoto, Y., et al. (2016). BMP-SMAD-ID promotes reprogramming to pluripotency by inhibiting pl6/INK4A-dependent senescence. Proceedings of the National Academy of Sciences of the United States of America 113, 13057-13062.
- Hou, P., Li, Y., Zhang, X., Liu, C , Guan, J., Li, H., Zhao, T., Ye, J., Yang, W., Liu, K., et al. (2013). Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds. Science 341, 651-654.
- Hu, G , Kim, J., Xu, Q., Leng, Y., Orkin, S.H., and Elledge, S.J. (2009). A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes & development 23, 837-848.
- Hussein, S.M., Puri, M.C., Tonge, P.D., Benevento, M., Corso, A.J., Clancy, J.L., Mosbergen, R., Li, M., Lee, D.-S., and Cloonan, N. (2014). Genome-wide characterization of the routes to pluripotency. Nature 516, 198.

- Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one 9, e98679.
- Jeon, H., Waku, T., Azami, T., Khoa le, T.P., Yanagisawa, J., Takahashi, S., and Ema, M. (2016). Comprehensive Identification of Kruppel-Like Factor Family Members Contributing to the Self-Renewal of Mouse Embryonic Stem Cells and Cellular Reprogramming. PloS one 11, e0150715.
- Jukkola, T., Lahti, L., Naserke, T., Wurst, W., and Partanen, J. (2006). FGF regulated gene-expression and neuronal differentiation in the developing midbrain-hindbrain region. Developmental biology 297, 141-157.
- Kan, L., Israsena, N., Zhang, Z., Hu, M., Zhao, L.R., Jalali, A., Sahni, V., and Kessler, J.A. (2004). Soxl acts through multiple independent pathways to promote neurogenesis. Developmental biology 269, 580-594.
- Kantorovitch, L. (1958). On the Translocation of Masses. Management Science 5, 1-4.
- Kester, L., and van Oudenaarden, A. (2018). Single-Cell Transcriptomics Meets Lineage Tracing. Cell Stem Cell.
- Kidder, B.L., and Palmer, S. (2010). Examination of transcriptional networks reveals an important role for TCFAP2C, SMARCA4, and EOMES in trophoblast stem cell maintenance. Genome Res 20, 458-472.
- Kim, D.H., Marinov, G.K., Pepke, S., Singer, Z.S., He, P., Williams, B., Schroth, G.P., Elowitz, M.B., and Wold, B.J. (2015). Single-cell transcriptome analysis reveals dynamic changes in IncRNA expression during reprogramming. Cell stem cell 16, 88-101.
- Klein, A.M., Mazutis, L., Akartuna, L, Tallapragada, N., Veres, A., Li, V., Peshkin, L., Weitz, D.A., and Kirschner, M.W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187-1201.
- Kolodziejczyk, Aleksandra A., Kim, Jong K., Tsang, Jason C , Ilicic, T., Henriksson, J., Natarajan, Kedar N., Tuck, Alex C , Gao, X., Buhler, M., Liu, P., et al. (2015). Single Cell RNA-Sequencing of Pluripotent States Unlocks Modular Transcriptional Variation. Cell Stem Cell 17, 471-485.

- Kumar, R.M., Cahan, P., Shalek, A.K., Satija, R., Jay DaleyKeyser, A., Li, H., Zhang, J., Pardee, K., Gennert, D., Trombetta, J.J., et al. (2014). Deconstructing transcriptional heterogeneity in pluripotent stem cells. Nature 516, 56.
- Latos, P.A., and Hemberger, M. (2016). From the stem of the placental tree: trophoblast stem cells and their progeny. Development 143, 3650-3660.
- Lattin, J.E., Schroder, K., Su, A.I., Walker, J.R, Zhang, J., Wiltshire, T., Saijo, K., Glass, C.K., Hume, D.A., Kellie, S., et al. (2008). Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome research 4, 5.
- Lazarov, O., Mattson, M.P., Peterson, D.A., Pimplikar, S.W., and van Praag, H. (2010). When neurogenesis encounters aging and disease. Trends in neurosciences 33, 569-579.
- Le ́onard, C. (2014). A survey of the schro ̈dinger problem and some of its connections with optimal transport. Discrete and Continuous Dynamical Systems - Series A (DCDS-A), 34(4):1533-1574.
- Li, R., Liang, J., Ni, S., Zhou, T., Qing, X., Li, H., He, W., Chen, J., Li, F., Zhuang, Q., et al. (2010). A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell Stem Cell 7, 51-63.
- Li, W.-Z., Wang, Z.-W., Chen, L.-L., Xue, H.-N., Chen, X., Guo, Z.-K., and Zhang, Y. (2015). Hesxl enhances pluripotency by working downstream of multiple pluripotency-associated signaling pathways. Biochemical and Biophysical Research Communications 464, 936-942.
- Liang, H., Zhang, Q., Lu, J., Yang, G , Tian, N., Wang, X., Tan, Y , and Tan, D. (2016). MSX2 Induces Trophoblast Invasion in Human Placenta. PloS one 11, e0153656.
- Lim, L.S., Loh, Y.H., Zhang, W., Li, Y., Chen, X , Wang, Y , Bakre, M., Ng, H.H, and Stanton, L.W. (2007). Zic3 is required for maintenance of pluripotency in embryonic stem cells. Molecular biology of the cell 18, 1348-1358.
- Lin, J., Khan, M., Zapiec, B., and Mombaerts, P. (2016). Efficient derivation of extraembryonic endoderm stem cell lines from mouse postimplantation embryos. Scientific reports 6, 39457.

- Liu, J., Han, Q., Peng, T., Peng, M , Wei, B., Li, D., Wang, X., Yu, S., Yang, J., Cao, S., et al. (2015). The oncogene c-Jun impedes somatic cell reprogramming. Nature cell biology 17, 856-867.
- Liu, L.L., Brumbaugh, J., Bar-Nur, O., Smith, Z., Stadtfeld, M., Meissner, A., Hochedlinger, K., and Michor, F. (2016). Probabilistic Modeling of Reprogramming to Induced Pluripotent Stem Cells. Cell reports 17, 3395-3406.
- Ma, G.T., Roth, M.E., Groskopf, J.C., Tsai, F.Y., Orkin, S.H., Grosveld, F., Engel, J.D., and Linzer, D.I. (1997). GATA-2 and GATA-3 regulate trophoblast-specific gene expression in vivo. Development 124, 907-914.
- Macfarlan, T.S., Gifford, W.D., Driscoll, S., Lettieri, K., Rowe, H.M., Bonanomi, D., Firth, A., Singer, O., Trono, D., and Pfaff, S.L. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. Nature 487, 57-63.
- Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N , and Martersteck, E.M. (2015). Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202-1214.
- Marco, E., Karp, R.L., Guo, G , Robson, P., Hart, A.H., Trippa, L., and Yuan, G.C. (2014). Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proceedings of the National Academy of Sciences of the United States of America 111, E5643-5650.
- Matsumoto, H., and Kiryu, H. (2016). SCOUP: a probabilistic model based on the Ornstein-Uhlenbeck process to analyze single-cell expression data during differentiation. BMC Bioinformatics 17, 232.
- McKenna, A., Findlay, G.M., Gagnon, J.A., Horwitz, M.S., Schier, A.F., and Shendure, J. (2016). Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science 353, aaf7907.
- Mertins, P., Przybylski, D., Yosef, N., Qiao, J., Clauser, K., Raychowdhury, R., Eisenhaure, T.M., Maritzen, T., Haucke, V., Satoh, T., et al. (2017). An Integrative Framework Reveals Signaling-to-Transcription Events in Toll-like Receptor Signaling. Cell reports 19, 2853-2866.

- Messina, G., Biressi, S., Monteverde, S., Magli, A., Cassano, M., Perani, L., Roncaglia, E., Tagliafico, E., Starnes, L., Campbell, C.E., et al. (2010). Nfix regulates fetal-specific transcription in developing skeletal muscle. Cell 140, 554-566.
- Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49.
- Ming, G.L., and Song, H. (201 1). Adult neurogenesis in the mammalian brain: significant answers and significant questions. Neuron 70, 687-702.
- Mosteiro, L., Pantoja, C , Alcazar, N., Marion, R.M., Chondronasiou, D., Rovira, M., Fernandez-Marcos, P.J., Munoz-Martin, M., Blanco-Aparicio, C , and Pastor, J. (2016). Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. Science 354, aaf4445.
- Nakashima, K., Wiese, S., Yanagisawa, M., Arakawa, H., Kimura, N , Hisatsune, T., Yoshida, K., Kishimoto, T., Sendtner, M., and Taga, T. (1999). Developmental requirement of gpl30 signaling in neuronal survival and astrocyte differentiation. The Journal of neuroscience : the official journal of the Society for Neuroscience 19, 5429-5434.
- Nelson, A.C., Mould, A.W., Bikoff, E.K., and Robertson, E.J. (2016). Single-cell RNA-seq reveals cell type-specific transcriptional signatures at the maternal-foetal interface during pregnancy. Nat Commun 7, 11414.
- O'Malley, J., Skylaki, S., Iwabuchi, K.A., Chantzoura, E., Ruetz, T., Johnsson, A., Tomlinson, S.R., Linnarsson, S., and Kaji, K. (2013). High resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature 499, 88.
- Ocana, O.H., Corcoles, R , Fabra, A., Moreno-Bueno, G., Acloque, H., Vega, S., Barrallo-Gimeno, A., Cano, A., and Nieto, M.A. (2012). Metastatic colonization requires the repression of the epithelial-mesenchymal transition inducer Prrxl. Cancer cell 22, 709-724.
- Parast, M.M., Yu, H., Ciric, A , Salata, M.W., Davis, V., and Milstone, D.S. (2009). PPARgamma regulates trophoblast proliferation and promotes labyrinthine trilineage differentiation. PloS one 4, e8055.

- Parenti, A., Halbisen, M.A., Wang, K., Latham, K., and Ralston, A. (2016). OSKM induce extraembryonic endoderm stem cells in parallel to induced pluripotent stem cells. Stem cell reports 6, 447-455.
- Park, M., Lee, Y., Jang, H., Lee, O.H., Park, S.W., Kim, J.H., Hong, K., Song, H , Park, S.P., Park, Y.Y., et al. (2016). SOHLH2 is essential for synaptonemal complex formation during spermatogenesis in early postnatal mouse testes. Scientific reports 6, 20980.
- Pasque, V., Tchieu, J., Karnik, R , Uyeda, M., Dimashkie, A.S., Case, D., Papp, B., Bonora, G., Patel, S., and Ho, R. (2014). X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. Cell 159, 1681-1697.
- Patel, A.P., Tirosh, L, Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science (New York, NY) 344, 1396-1401.
- Pei, J., and Grishin, N.V. (2012). Unexpected diversity in Shisa-like proteins suggests the importance of their roles as transmembrane adaptors. Cellular signalling 24, 758-769.
- Plass, M., Solana, J., Wolf, F.A., Ayoub, S., Misios, A., Glazar, P., Obermayer, B., Theis, F.J., Kocks, C , and Rajewsky, N. (2018). Cell type atlas and lineage tree of a whole complex animal by single-cell transcriptomics. Science.
- Polo, J.M., Anderssen, E., Walsh, R.M., Schwarz, B.A., Nefzger, CM., Lim, S.M., Borkent, M., Apostolou, E., Alaei, S., and Cloutier, J. (2012). A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617-1632.
- Porpiglia, E., Samusik, N., Van Ho, A. T., Cosgrove, B. D., Mai, T., Davis, K. L., Jager, A., Nolan, G. P., Bendall, S. C , Fantl, W. J., et al. (2017). High-resolution myogenic lineage mapping by single-cell mass cytometry. Nature Cell Biol., 19:558-567.
- Qiu, X , Mao, Q., Tang, Y , Wang, L., Chawla, R., Pliner, H., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv, 110668.
- Rajkovic, A., Yan, C , Yan, W., Klysik, M., and Matzuk, M.M. (2002). Obox, a Family of Homeobox Genes Preferentially Expressed in Germ Cells. Genomics 79, 711-717.

- Ralston, A., Cox, B.J., Nishioka, N., Sasaki, H., Chea, E., Rugg-Gunn, P., Guo, G., Robson, P., Draper, J.S., and Rossant, J. (2010). Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development 137, 395-403.
- Ramskold, D., Luo, S., Wang, Y.-C, Li, R., Deng, Q., Faridani, O.R., Daniels, G.A., Khrebtukova, I., Loring, J.F., Laurent, L.C., et al. (2012). Full-Length mRNA-Seq from single cell levels of RNA and individual circulating tumor cells. Nature biotechnology 30, 777-782.
- Rashid, S., Kotton, D.N., and Bar-Joseph, Z. (2017). TASIC: determining branching models from time series single cell data. Bioinformatics 33, 2504-2512.
- Richard Jordan, D. K. and Otto, F. (1998). The variational formulation of the fokker. SIAM J. Math. Anal., 29(1): 1-17.
- Rostom, R., Svensson, V., Teichmann, S., and Kar, G. (2017). Computational approaches for interpreting scRNA-seq data. FEBS letters.
- Sakakibara, S., Nakamura, Y., Satoh, H., and Okano, H. (2001). Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. The Journal of neuroscience : the official journal of the Society for Neuroscience 21, 8091-8107.
- Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L., and Nolan, G. P. (2016). Automated mapping of phenotype space with single-cell data. Nature methods, 13:493-496.
- Sansom, S.N., Griffiths, D.S., Faedo, A., Kleinjan, D.J., Ruan, Y., Smith, J., van Heyningen, V., Rubenstein, J.L., and Livesey, F.J. (2009). The level of the transcription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis. PLoS genetics 5, el00051 1.
- Santambrogio, F. (2015). Optimal transport for applied mathematicians. Birkauser, NY, 99-102.
- 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.
- Scott, I.C., Anson-Cartwright, L., Riley, P., Reda, D., and Cross, J.C. (2000). The HANDl basic helix-loop-helix transcription factor regulates trophoblast differentiation via multiple mechanisms. Molecular and cellular biology 20, 530-541.

- Setty, M., Tadmor, M.D., Reich-Zeliger, S., Angel, O., Salame, T.M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'er, D. (2016). Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature biotechnology 34, 637-645.
- Shalek, A.K., Satija, R., Adiconis, X., Gertner, R.S., Gaublomme, J.T., Raychowdhury, R., Schwartz, S., Yosef, N., Malboeuf, C , Lu, D., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature 498, 236.
- Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y., and Pei, D. (2006). Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. J Biol Chem 281, 23319-23325.
- Shu, I , Wu, C , Wu, Y , Li, Z., Shao, S., Zhao, W., Tang, X., Yang, H., Shen, L., Zuo, X., et al. (2013). Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell 153, 963-975.
- Simmons, D.G., and Cross, J.C. (2005). Determinants of trophoblast lineage and cell subtype specification in the mouse placenta. Developmental biology 284, 12-24.
- Simmons, D.G, Natale, D.R., Begay, V., Hughes, M., Leutz, A., and Cross, J.C. (2008). Early patterning of the chorion leads to the trilaminar trophoblast cell structure in the placental labyrinth. Development 135, 2083-2091.
- Stadtfeld, M., Maherali, N., Borkent, M., and Hochedlinger, K. (2010). A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nature methods 7, 53-55.
- Street, K., Risso, D., Fletcher, R.B., Das, D., Ngai, J., Yosef, N , Purdom, E., and Dudoit, S. (2017). Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. bioRxiv.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell 126, 663-676.
- Takahashi, K., and Yamanaka, S. (2016). A decade of transcription factor-mediated reprogramming to pluripotency. Nature Reviews Molecular Cell Biology 17, 183.
- Takaishi, M., Tarutani, M., Takeda, J., and Sano, S. (2016). Mesenchymal to Epithelial Transition Induced by Reprogramming Factors Attenuates the Malignancy of Cancer Cells. PloS one 11, e0156904.

- Tanay, A., and Regev, A. (2017). Scaling single-cell genomics from phenomenology to mechanism. Nature 541, 331-338.
- Tang, F., Barbacioru, C , Wang, Y., Nordman, E., Lee, C , Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A., et al. (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nature Methods 6, 377.
- Tasic, B., Menon, V., Nguyen, T.N., Kim, T.K., Jarsky, T., Yao, Z., Levi, B., Gray, L.T., Sorensen, S.A., Dolbeare, T., et al. (2016). Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. NatNeurosci 19, 335-346.
- Tirosh, I., Venteicher, A.S., Hebert, C , Escalante, L.E., Patel, A.P., Yizhak, K., Fisher, J.M., Rodman, C , Mount, C , and Filbin, M.G. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature 539, 309-313.
- Tonge, P.D., Corso, A.J., Monetti, C , Hussein, S.M., Puri, M.C., Michael, LP., Li, M., Lee, D.-S., Mar, J.C., and Cloonan, N. (2014). Divergent reprogramming routes lead to alternative stem-cell states. Nature 516, 192-197.
- 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.
- Ueno, M., Lee, L.K., Chhabra, A., Kim, Y.J., Sasidharan, R., Van Handel, B., Wang, Y., Kamata, M., Kamran, P., Sereti, K.-L, et al. (2013). c-Met-dependent multipotent labyrinth trophoblast progenitors establish placental exchange interface. Developmental cell 27, 373-386.
- Vandercappellen, J., Van Damme, J., and Struyf, S. (2008). The role of CXC chemokines and their receptors in cancer. Cancer letters 267, 226-244.
- Villani, C. (2008). Optimal transport: old and new, Vol 338 (Springer Science & Business Media).
- Waddington, C.H. (1936). How animals develop (New York).
- Waddington, C.H. (1957). The strategy of the genes; a discussion of some aspects of theoretical biology (London, Allen & Unwin [1957]).
- Wagner, A., Regev, A., and Yosef, N. (2016). Revealing the vectors of cellular identity with single-cell genomics. Nat Biotech 34, 1145-1 160.

- Wagner, D.E., Weinreb, C , Collins, Z.M., Briggs, J.A., Megason, S.G., and Klein, A.M. (2018). Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. Science.
- Watanabe, Y., Stanchina, L., Lecerf, L., Gacem, N., Conidi, A., Baral, V., Pingault, V., Huylebroeck, D., and Bondurand, N. (2017). Differentiation of Mouse Enteric Nervous System Progenitor Cells Is Controlled by Endothelin 3 and Requires Regulation of Ednrb by SOX10 and ZEB2. Gastroenterology 152, 1139-1 150.el 134.
- Weinreb, C , Wolock, S., and Klein, A. (2016). SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. bioRxiv.
- Weinreb, C , Wolock, S., Tusi, B.K., Socolovsky, M., and Klein, A.M. (2017). Fundamental limits on dynamic inference from single cell snapshots. bioRxiv.
- Welch, J.D., Hartemink, A.J., and Prins, J.F. (2016). SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. Genome Biology 17, 106.
- Whiteman, EX., Fan, S., Harder, J.L., Walton, K.D., Liu, C.J., Soofi, A., Fogg, V.C., Hershenson, M.B., Dressier, G.R., Deutsch, G.H., et al. (2014). Crumbs3 is essential for proper epithelial development and viability. Molecular and cellular biology 34, 43-56.
- Wu, D., Hong, H., Huang, X., Huang, L., He, Z., Fang, Q., and Luo, Y. (2016). CXCR2 is decreased in preeclamptic placentas and promotes human trophoblast invasion through the Akt signaling pathway. Placenta 43, 17-25.
- Wu, L., Wu, Y., Peng, B., Hou, Z., Dong, Y., Chen, K., Guo, M., Li, H., Chen, X., Kou, X., et al. (2017). Oocyte-Specific Homeobox 1, Oboxl, Facilitates Reprogramming by Promoting Mesenchymal-to-Epithelial Transition and Mitigating Cell Hyperproliferation. Stem Cell Reports 9, 1692-1705.
- Wu, X., Oatley, J.M., Oatley, M.J., Kaucher, A.V., Avarbock, M.R., and Brinster, R.L. (2010). The POU domain transcription factor POU3F1 is an important intrinsic regulator of GDNF-induced survival and self-renewal of mouse spermatogonial stem cells. Biology of reproduction 82, 1103-1 111.
- Yamamizu, K., Sharov, A.A., Piao, Y., Amano, M., Yu, H , Nishiyama, A., Dudekula, D.B., Schlessinger, D., and Ko, M.S. (2016). Generation and gene expression profiling of 48 transcription-factor-inducible mouse embryonic stem cell lines. Scientific reports 6, 25667.

- Ying, Q.-L., Wray, J., Nichols, J., Batlle-Morera, L., Doble, B., Woodgett, J., Cohen, P., and Smith, A. (2008). The ground state of embryonic stem cell self-renewal. Nature 453, 519.
- Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917-1920.
- Yun, C , Mendelson, J., Blake, T., Mishra, L., and Mishra, B. (2008). TGF-beta signaling in neuronal stem cells. Disease markers 24, 251-255.
- Zhao, T., Fu, Y., Zhu, J., Liu, Y., Zhang, Q., Yi, Z., Chen, S., Jiao, Z., Xu, X., Xu, J., Duo, S., Bai, Y., Tang, C , Li, C , and Deng, H. (2018). Single-Cell RNA-Seq Reveals Dynamic Early Embryonic-like Programs during Chemical Reprogramming. Cell Stem Cell 23, 1-15.
- Zunder, E.R., Lujan, E., Goltsev, Y., Wernig, M., and Nolan, G.P. (2015). A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell 16, 323-337.
- Zwiessele, M., and Lawrence, N.D. (2016). Topslam: Waddington Landscape Recovery for Single Cell Experiments. bioRxiv.

[0318] **Key resources**

[0319] Key resources used in this study are shown below.

REAGENTS or RESOURCE	SOURCE	IDENTIFIER
Recombinant DNA		
FUW Tet-On vector	Addgene	#20323
Zfp42 cDNA	Origene	MG203929
Obox6 cDNA	Origene	MR215428
Chemicals, Peptides, and Recombinant Proteins		
leukemia inhibitory factor (LIF)	Millipore	ESG1107
PD0325901	Sigma	PZ0162-25MG
CHIR99021	Sigma	PZ0162-25MG
Critical Commercial Kits		
Chromium™ Single Cell 3' Reagent	10X genomics	PN-120230, PN-120231, PN-

Kits v1		120232
Chromium™ Single Cell 3' Reagent Kits v2	10X genomics	PN- 120237
Fugene HD reagent	Promega	E2311
Cloning Reagents		
Gibson Assembly	NEB	E2611S
Sequence-Based Reagents		
Deposited Data		
Single cell RNA-seq raw data (pilot study) Single cell RNA-seq raw data	NCBI Gene Expression Omnibus NCBl Gene Expression Omnibus	GSE106340 GSE115943
Experimental Models: Organisms/Strains		
OKSM secondary MEFs	Konrad Hochedlinger lab	OKSM x B6.Cg- *Gt(ROSA)26Sor*$^{Jml(rtTA*M2)Jae}$/J x B*6;l29S4-Pou5f l*^{tm2Jae}/J
Primary MEFs	Rudolf Jaenisch lab	B6.Cg- *Gt(ROSA)26Sof*$^{M1<rtTA*M2)Jae}$lJ x B6;129S4-/*ᴾou5f*7ᵉ ^{2Ja}7 J
Software and Algorithms		
Waddington-OT	This paper	https ://github .com/broadinstitute /wot
Scaling algontlim for unbalanced transport	(Chizat et al., 2016)	
CeilRanger	10X genomics	v2.0.0
ForceAtlas2	Gephi	vO.9.2
Seurat		v2.1.0
Scanpy		v0.2.8
Monocle2	(Qiu et al. 2017)	v2.8.0
URD	(Farrell et al 2018)	vl.O

[0320] Method Details

[0321] I. Modeling developmental processes with optimal transport

[0322] We developed a method to model development based on Optimal Transport. Section 1 reviews the concept of gene expression space and introduces our probabilistic framework for time series of expression profiles. Section 2 introduces our key modeling assumption to infer temporal couplings over short time scales. Section 3 shows how we can compute an optimal coupling between adjacent time points by solving a convex optimization problem, and how we can leverage an assumption of Markovity to compose adjacent time points and estimate temporal couplings over longer intervals. Section 4 describes how to interpret transport maps. Specifically, Section 4.1 shows how to compute ancestors and descendants of cells, Section 4.2 describes an interesting physical interpretation of entropy-regularization, and Section 4.3 shows how we learn gene regulatory networks to summarize the trajectories.

[0323] 1. Developmental processes in gene expression space

[0324] A collection of mRNA levels for a single cell is called an *expression profile* and is often represented mathematically by a vector in *gene expression space.* This is a vector space that has dimension equal to the number of genes, with the value of the ith coordinate of an expression profile vector representing the number of copies of mRNA for the ith gene. Note that real cells only occupy an integer lattice in gene expression space (because the number of copies of mRNA is an integer), but we pretended that cells can move continuously through a real-valued G dimensional vector space.

[0325] As an individual cell changes the genes it expresses over time, it moves in gene expression space and describes a trajectory. As a population of cells develops and grows, a *distribution* on gene expression space evolves over time. When a single cell from such a population is measured with single cell RNA sequencing, we obtained a noisy estimate of the number of molecules of mRNA for each gene. We represented the measured expression profile of this single cell as a sample from a probability distribution on gene expression space. This sampling captured both (a) the randomness in the single-cell RNA sequencing measurement process (due to subsampling reads, technical issues, etc.) and (b) the random selection of a cell from the population. We treated this probability distribution as *nonparametric* in the sense that it wsa not specified by any finite list of parameters.

[0326] In the remainder of this section we introduced a precise mathematical notion for a *developmental process* as a generalization of a stochastic process. Our primary goal was to infer the ancestors and descendants of subpopulations evolving according to an unknown developmental process. This information was encoded in the *temporal coupling* of the process, which is lost because we kill the cells when we perform scRNA-Seq. We claimed it was possible to recover the temporal coupling over short time scales provided that cells don't change too much. Therefore we could make inferences about which cells go where. We showed in the remainder of this section how to do this with *optimal transport.*

[0327] 1.1 A mathematical model of developmental processes

[0328] We began by formally defining a precise notion of the developmental trajectory of an individual cell and its descendants. Intuitively, it was a continuous path in gene expression space that bifurcated with every cell division. Formally, we defined it as follows:

[0329] Definition 1 (single-cell developmental trajectory). Consider a cell $x(0) \in \mathbb{R}^G$. Let k(t) ≥ 0 specifiy the number of descendants at time t, where k(0) = 1. A single-cell development trajectory is a continuous function

$$x : [0, T) \to \underbrace{\mathbb{R}^G \; x \; \mathbb{R}^G \; \mathrm{x} \ldots \mathrm{x} \; \mathbb{R}^G}_{k(t) \text{ times}} .$$

This means that x(t) is a k(t)-tuple of cells, each represented by a vector in $\mathbb{R}^G$*:*

$$\mathrm{ar(i)} = (x_1(t), \ldots, ¾_{(t)}(t)) \; .$$

We referred to the cells x\(t), . . . , xk(t)(t) as the descendants of x(0).

[0330] Note that we could not directly measure the temporal dynamics of an individual cell because scRNA-Seq was a destructive measurement process: scRNA-Seq lysed cells so it was possible to measure the expression profile of a cell at a single point in time. As a result, it was not possible to directly measure the descendants of that cell, and the full trajectory was unobservable. However, one can learn something about the probable trajectories of individual cells by measuring snapshots from an evolving population.

[0331] Published methods typically represent the aggregate trajectory of a population of cells by means of a graph structure. While this recapitulates the branching path traveled by the descendants of an individual cell, it may over-simplify the stochastic nature of developmental processes. Individual cells have the potential to travel through different paths, but any given cell travels one and only one such path. Our goal was to assign a likelihood to the set of possible paths, which in general were not finite and therefore cannot be a represented by a graph.

[0332] We defined a developmental process to be a time-varying probability distribution on gene expression space. One simple example of a distribution of cells is that we can represent a set of cells

XI, . . . , x_n by the distribution

$$\mathbb{P} = \frac{1}{n}\sum_{i=1}^{n}\delta_{x_i}.$$

[0333] Similarly, we could represent a set of single-cell trajectories $x_1(t)$, . . . , $x_n(t)$ with a distribution over trajectories. This was a special case of a developmental process, which we defined as follows:

Definition 2 (developmental process). A developmental process Pt is a time-varying distribution (i.e. stochastic process) on gene expression space.

[0334] Recall that a stochastic process was determined by its temporal dependence structure. This was specified by the coupling (i.e. joint distribution) between random variables at different time points. Given that a cell had a particular expression profile y at time t_2, where did it come from at time t_{i}? This was the information lost by not tracking individual cells overtime.

[0335] **Definition 3** (temporal coupling). Let Ptbe a developmental process and consider two time points s<t. Let Xt ~ Pt denote the expression profile of a random cell at time t and let X_s denote the expression profile of the cell of origin at times.

[0336] The temporal coupling $\gamma_{s,t}$ is defined as the law of the joint distribution:

$$\gamma_{s,t} = \mathcal{L}(X_s, X_t).$$

Equivalently,

$$\int_{x\in A}\int_{y\in B}\gamma_{s,t}(x,y)dxdy = \Pr\{X_s \in A, X_t \in B\}$$

for any seis $A, B \subset$ S^{G}-

[0337] The temporal coupling $\gamma_{s,t}$ was not technically a coupling of P_s and P? in the standard sense because it does not necessarily have marginals P_s and P?:

$$\int \gamma_{s,t}(x,y)dx = \mathbb{P}_t(y), \quad \text{but} \quad \int \gamma_{s,t}(x,y)dy \neq \mathbb{P}_s(x).$$

[0338] Biologically, this was the case when cells grow at different rates. Then proliferative cells from the earlier time point were over-represented when we look for the origin of cells at the later time point. In the following definition, we introduced a relative growth rate function to describe the relationship between the expression profile of a cell and the average number of living descendants it gave rise to after certain amount of time.

[0339] **Definition 4.** A relative growth rate function associated with a temporal coupling is a function g(x)

satisfying

$$\int \gamma_{s,t}(x,y)dy = \mathbb{P}_s(x)\frac{g(x)^{t-s}}{\int g(x)^{t-s}d\mathbb{P}_s(x)}.$$

[0340] The integral on the left-hand side represented the amount of mass coming out of *x* and going to any *y*. The term P(x) on the right hand side accounted for the abundance of cells with expression profile x, and the function *g(x)* represented the exponential increase in mass per unit time.

[0341] Having defined the notion of developmental processes and temporal couplings, we now turned to estimating these from data.

[0342] 2. The optimal transport principle for developmental processes

[0343] Single-cell RNA-Seq allowed us to sample cells from a developmental process at various time points, but it did not give any information about the coupling between successive time points. Without making any assumptions, it was impossible to recover the temporal coupling even given infinite data in the form of the full distributions P_s and Pi. However, we claimed that it was reasonable to assume that cells don't change expression by large amounts over short time scales. This assumption allowed us to estimate the coupling and infer which cells go where.

[0344] We began with a simple one-dimensional example to build intuition.

[0345] **Example 1.** Let Xo ~ N (0, σ^2) and Xi ~ N (μ, σ^2) be one dimensional Gaussian variables representing the location of a particle at time 0 and at time 1. One simple heuristic to estimate γˆ is to minimize the squared distance that the particle moves from time 0 to time 1:

$$\hat{\gamma} \leftarrow \arg\min_{\pi} \mathbb{E}_\pi \|X_0 - X_1\|^2.$$

[0346] We minimized over all couplings π with marginals (0, σ^2) and (μ, σ^2). One can check that the optimal joint distribution is a two dimensional Gaussian with the following dependence structure:

$X_1 = X_0 + \mu.$

[0347] This heuristic to couple marginals was called *optimal transport* (OT). If *c(x, y)* denoted the cost of transporting a unit mass from *x* to *y*, and the amount we transferred from *x* to *y* is *π(χ, y),* then the total cost of transporting mass according to such a transport plan π is given by

$$\iint c(x, y)\pi(x, y)dxdy.$$

[0348] In this study we focused on the cost defined by the squared-Euclidean distance

$$c(x, y) = \|x - y\|^2,$$

[0349] on an appropriate input space. We made this choice to focus on Wasserstein-2 transport because of the many attractive theoretical properties it enjoyed over Wasserstein- 1 transport (Villani, 2008).

[0350] The optimal transport plan minimized the expected cost subject to marginal constraints:

$$\begin{aligned} \pi(\mathbb{P}, \mathbb{Q}) = \underset{\pi}{\text{minimize}} \quad & \iint c(x, y)\pi(x, y)dxdy \\ \text{subject to} \quad & \int \pi(x, \cdot)dx = \mathbb{Q} \qquad (1) \\ & \int \pi(\cdot, y)dy = \mathbb{P}. \end{aligned}$$

[0351] Note that this was a linear program in the variable π because the objective and constraints were both linear in π. The optimal objective value defined the *transport distance* between P and Q (it was also called the Earthmover' s distance or Wasserstein distance). Unlike many other ways to compare distributions (such as KL-divergence or total variation), optimal transport took the geometry of the underlying space into account. For example, the KL-Divergence was infinite for any two distributions with disjoint support, but the transport distance depended on the separation of the support. For a comprehensive treatment of the rich mathematical theory of optimal transport, we refer the reader to (Villani, 2008).

[0352] 2.1 The optimal transport principle for developmental processes.

[0353] We proposed to use optimal transport to estimate the temporal coupling of a developmental process. We made two modifications to classical optimal transport to adapt it to our biological setting.

[0354] 1. Classical optimal transport had conservation of mass built into the constraints (1). We accounted for growth by rescaling the distribution **Pi** before applying OT.

[0355] 2. The coupling identified by classical optimal transport was purely deterministic in the sense that each point was transported to a single point. However, for cells whose fates were not completely determined, the true coupling should have a degree of entropy to it. We therefore added a term to the objective to promote entropy in the transport coupling.

[0356] Injecting a small amount of entropy also made sense even for a population of cells with truly deterministic descendant distribution. When we sampled finitely many cells at time t_2, the true descendants of any given *t* cell were not captured. Therefore entropy in the transport map could be used to represent our statistical uncertainty in the inferred descendant distribution.

[0357] In order to state the optimal transport principle, we first introduced some notation. Let **Pi** denote a developmental process with temporal coupling $y_S t$ and with relative growth function *g(x).* Let Q*s* denote the distribution obtained by rescaling $\mathbf{P}_s$ by the relative growth rate:

$$\mathbb{Q}_s(x) = \mathbb{P}_s(x)\frac{g^{t-s}(x)}{\int g^{t-s}(z)d\mathbb{P}_s(z)}.$$

[0358] Finally, let $\pi_{s,t}(\epsilon)$ denote the entropy-regularized optimal transport coupling of Q_S and **Pi,** defined as the solution to the following optimization problem

$$\begin{aligned} \pi_{s,t}(\epsilon) = \underset{\pi}{\text{minimize}} \quad & \int\int c(x,y)\pi(x,y)dxdy - \epsilon\int \pi(x,y)\log\pi(x,y)dxdy \\ \text{subject to} \quad & \int \pi(x,\cdot)dx = \mathbb{Q}_s \\ & \int \pi(\cdot,y)dy = \mathbb{P}_t. \end{aligned} \tag{2}$$

[0359] We now stated the optimal transport principle for developmental process

$$s \approx t \implies \pi_{s,t}(\epsilon) \approx \gamma_{s,t}.$$

[0360] In words, over short time scales, the true coupling was well approximated by the OT coupling. In section 3, we show how to estimate $\pi_{«_i*}(\epsilon)$ **from** data (we occasionally omit the dependence on ϵ and write $\pi_{s,t}$). This in turn gives us an estimate of $y_{s,t}$.

[0361] 3. Inferring temporal couplings from empirical data

[0362] In this section we showed how to estimate the temporal couplings of a developmental process from data.

[0363] **Definition 5** (developmental time series). A developmental time series was a sequence of samples from a developmental process Pt on R^. This was a sequence of sets Si, . . . , $S_T \subset R^G$ collected at times ti, . . . ,ïτ GR. Each Si is a set of expression ′profiles in R^G drawn independently from P_t .

[0364] From this input data, we formed an empirical version of the developmental process. Specifically, at each time point *t*, we formed the empirical probability distribution supported on the data* *S*,. We summarize this in the following definition:

[0365] **Definition 6** (Empirical developmental process). An empirical developmental process $\hat{P}$t is a time vary- ing distribution constructed from a developmental time course Si, . . . , ST :

$$\hat{\mathbb{P}}_{t_i} = \frac{1}{|S_i|} \sum_{x \in S_i} \delta_x. \tag{3}$$

[0366] The empirical developmental process was undefined for t ∉ {ti, . . . ,ïτ }.

[0367] In order to estimate the coupling from time *t*\ to time t_2, we first constructed an initial estimate the growth rate function g(x). In practice, we form an initial estimate gˆ(x) as the expectation of a birth-death process on gene expression space with birth-rate β(χ) and death rate δ(χ) defined in terms of expression levels of genes involved in cell proliferation and apoptosis. We ultimately leveraged techniques from unbalanced transport (Chizat et al., 2017) to refine this initial estimate to learn cellular growth and death rates automatically from data.

[0368] We then form the rescaled empirical distribution

$$\hat{\mathbb{Q}}_{t_1}(x) = \hat{\mathbb{P}}_{t_1}(x) \frac{\hat{g}(x)^{t_1 - t_2}}{\int \hat{g}(z)^{t_1 - t_2} d\hat{\mathbb{P}}_{t_1}(z)},$$

and compute the optimal transport map $\hat{\pi}_{t_1,t_2}$ between $\hat{Q}$*tl* and $\hat{P}$*t2*

[0369] 3.1 Estimating couplings between adj acent time points

[0370] In order to identify an optimal transport plan connecting Qˆt1 and Pˆt2 , we solved an optimization problem with a matrix-valued optimization variable. In the classical zero-entropy setting (2) with « = 0 was a linear program. While the classical optimal transport linear program could be difficult to solve for large numbers of points, fast algorithms have been recently developed (Cuturi, 2013) to solve the entropically regularized version of the transport program. Entropic regularization speeded up the computations because it made the optimization problem strongly convex, and gradient ascent on the dual could be realized by successive diagonal matrix scalings called Sinkhorn iterations (Cuturi, 2013). These were very fast operations.

[0371] The scaling algorithm for entropically regularized transport had also been extended to work in the setting of unbalanced transport (Chizat et al., 2017), where the equality constraints were relaxed to bounds on the marginals of the transport plan (in terms of KL-divergence or total variation or a general f-divergence). In our application this was very attractive from a modeling perspective for the following reasons:

[0372] 1. We may have specified the growth rate function gˆ(x). Unbalanced transport adjusted the input growth rate in order to reduce the transport cost. This allowed us to automatically learn growth rates from scratch.

[0373] 2. Even if the growth rates were completely uniform, the random sampling could introduce what looked like growth. For example, suppose there was a rare subpopulation of cells consisting of 5% of the total. If at one time point, we randomly sampled fewer of these cells so that they comprised 4% of the total, and at the next time point we sample 6%, then it would look like this population had increased by 50%. Unbalanced transport could automatically adjust for this apparent growth.

[0374] We used both entropic regularization and unbalanced transport. To compute the transport map between the empirical distributions of expression profiles observed at time *t,* and *ti+i,* we solved the following optimization problem

$$\begin{aligned} \hat{\pi}_{t_i,t_{i+1}} &= \arg\min_{\pi} \sum_{x\in S_i}\sum_{y\in S_{i+1}} c(:r, \ y)\pi(x, ¾') \sim \epsilon \ I^{\ v} (x * y)\}\log \pi(x, y)dxdy \\ \text{sutrject to} \quad & \mathrm{KL}\left[\sum_{x\in S_i} \pi(x, y) \middle\| d\hat{\mathbb{P}}_{t_{i+1}}(y)\right] \leq \frac{1}{\lambda_1} \\ & \mathrm{KL}\left[\sum_{y\in S_{i+1}} \pi(x, y) \middle\| d\hat{\mathbb{Q}}_{t_i}(x)\right] \leq \frac{1}{\lambda_2} \end{aligned} \tag{4}$$

[0375] where ϵ, λ\ and λ_2 are regularization parameters.

[0376] This is a convex optimization problem in the matrix variable $\pi \in \mathbb{R}^{N_i \times N_{i+1}}$, where N_i = |s,| i_s i_s th_e number of cells sequenced at time ti. It takes about 5 seconds to solve this unbalanced transport problem using the scaling algorithm of (Chizat et al., 2017) on a standard laptop with Ni ≈ 5000.

[0377] Note that by default the densities (on the discrete set Si) of the empirical distributions specified in equation (3) are simply $d\hat{\mathbb{P}}_{t_i}(x) = \frac{1}{¾}$. However, in principle one could use nonuniform empirical distributions (e.g., if one wanted to include information about cell quality).

[0378] To summarize: given a sequence of expression profiles *S*i, . . . , *ST*, we solved the optimization problem (4) for each successive pair of time points *S*, *s* +i. For the pair of time-points *(ti, t,+\)*, this gave us a transport map^ $\hat{¾}_{,i_{i+1}}$. With enough data, this may be a good estimate of $\pi_{t_i,t}$**W** -because it is well known that transport maps are consistent in the sense that

$$\operatorname{!im}_{N_i,N_{i+1}\to\infty} \hat{\pi}_{t_i,t_{i+1}} = \pi_{t_i,t_{i+1}}.$$

[0379] Taken together with the optimal transport principle:

$$\pi_{t_i,t_{i+1}} \approx \gamma_{t_i,t_{i+1}};$$

[0380] We therefore could estimate $\gamma_{t_{i,¾+1}}$ f_{rom} $\hat{\pi}_{t_i,t_{i+1}}$ when Ni is large enough.

[0381] 3.2 Estimating long-range couplings

[0382] We relied on an assumption of Markovity (or memorylessness) in order to estimate couplings over longer time intervals. Recall that a stochastic process was Markov if the future was independent of the past, given the present. Equivalently, it was fully specified by the couplings between pairs of time points. We defined Markov developmental processes in a similar spirit:

[0383] **Definition 7** (Markov developmental process). A Markov developmental process Pt is atime-varying distribution onR^thatiscompletely specified by couplings between pairs of time points in the following sense. For any three time points $s < t < \tau$, the long-range coupling $\gamma_{s,\tau}$ was equal to the composition of short-range couplings:

$$\gamma_{t,\tau} \circ \gamma_{s,t} = \gamma_{s,\tau}.$$

[0384] Note that the optimal transport maps $\hat{\pi}$s,t did not have this compositional property. Composing the OT coupling from time s to t and then from t to τ was not the same as optimally transporting from s directly to τ. In general, we do not recommend computing OT maps directly between non-adjacent time points. We leveraged the Markovity assumption to estimate couplings over long time intervals by composing estimates over shorter intervals. Formally, for any pair of time points ti, ti+k, we estimate the coupling $\hat{\gamma}_{t_i,t_{i+k}}$ by composing as follows:

$$\hat{\gamma}_{t_i,t_{i+k}} = \hat{\pi}_{t_i,t_{i+1}} \circ \hat{\pi}_{t_{i+1},t_{i+2}} \circ \ldots \circ \hat{\pi}_{t_{i+k-1},t_{i+k}}.$$

[0385] These compositions were computed via ordinary matrix multiplication.

[0386] It is an interesting question to what extent developmental processes are Markov. On gene expression space, they were likely not strictly Markov because, for example, the history of gene expression could influence chromatin modifications, which may not themselves be fully reflected in the observed expression profile but could still influence the subsequent evolution of the process. However, it was possible that developmental processes could be considered Markov on some augmented space. Note that our core technique for estimating a single temporal coupling over a short time interval does **not** rely on any Markov assumption.

[0387] 4. Interpreting transport maps

[0388] In the previous section we introduced the principle of optimal transport for time series of gene expression profiles. Given a time series of expression profiles $S_1, \ldots, ST$, we used this principle to compute a sequence of transport maps between subsequent time slices. In this section we define the *ancestors* and *descendants* of any subset of cells from this sequence of transport maps in section 4.1. Then, in section 4.2 we explain an intuitive physical interpretation of entropy-regularization. Finally, in section 4.3 we describe a connection between optimal transport, gradient flows, and Waddington's landscape.

[0389] 4.1 Defining ancestors, descendants and trajectories

[0390] We defined the descendants and ancestors of subgroups of cells evolving according to a Markov (i.e. memoryless) developmental process.

[0391] Our definition of ancestors and descendants relies on a notion of *pushing* sets of cells through a trans- port map. Before defining ancestors and descendants, we introduce this terminology. As a distribution on the product space R^G ×R^G, a coupling *y* assigns a number *y(A, B)* to any pair of sets *A*, *B* ⊂*R*^{G}

$$\gamma(A, B) = \int_{x \in A} \int_{y \in B} \gamma(x, y) dx dy.$$

[0392] This number π(A, *B)* represented the amount of mass coming from *A* and going to *B*. When we did not specify a particular destination, the quantity *y{A,*) specified the full distribution of mass coming from *A*. We referred to this action as *pushing A* through the transport plan *y*. More generally, we could also push a *distribution* μ forward through the transport plan *y* via integration

$$\mu \mapsto \int \gamma(x, \cdot) d\mu(x).$$

[0393] We refer to the reverse operation as pulling a set *B* back through *y*. The resulting distribution $\gamma(\cdot, B)$ encodes the mass ending up at *B*. We can also pull distributions μ back through *y* in a similar way:

$$\mu \mapsto \int \gamma(\cdot, y) d\mu(y).$$

[0394] We sometimes refer to this as *back-propagating* the distribution μ (and to pushing μ forward as *forward propagation).*

[0395] Equipped with this terminology, we define ancestors and descendants as follows:

[0396] **Definition 8** (descendants in a Markov developmental process). Consider a set of cells $C \subset$]ll·^{G} which lived at time ti were part of a population of cells evolving according to a Markov developmental process Pt. Let γ_{t_1, t_2} denote the coupling from time ti to time t_2. The descendants of C at time t_2 are obtained by pushing C through γ.

[0397] **Definition 9** (ancestors in a Markov developmental process). Consider a set of cells C **c** $\mathbb{R}^{G}$, which lived at time t_2 and were part of a population of cells evolving according to a Markov developmental process Pt. Let π denote the transport map for P_t from time t_2 to time ti. The ancestors of C at time ti were obtained by pulling C back through γ.

[0398] **Trajectories:** We defined to the *ancestor trajectory* to a set C as the sequence of ancestor distributions at earlier time points. Similarly, we refer to the *descendant trajectory* from a set C as the sequence of descendant distributions at later time points.

[0399] 4.2 A physical interpretation of entropy regularized optimal transport

[0400] In this section we explain an interesting physical interpretation of entropy-regularized optimal transport. Consider a collection of N indistinguishable particles undergoing Brownian motion with diffusion coefficient ϵ. Suppose we observe the N particle positions at time 0 and at time 1. If $N=l$, the distribution on paths connecting the starting and ending point is called a Brownian bridge. For $N > 1$, the distribution over paths involves two components:

[0401] 1. A coupling of the particles specifying which particle goes where (because the particles are indistinguishable, this is not uniquely specified by the observations).

[0402] 2. Given a matching, the distribution on paths for each matched pair is a Brownian bridge.

[0403] The coupling was a random permutation that matched points at time 0 to points at time 1. The distribution of this random permutation depends on the variance of the Brownian motion. It turned out that the expected (i.e. average) coupling could be computed by maximum entropy optimal transport. These ideas could be traced back to Schrodinger's 1932 work in statistical electrodynamics (Schrodinger, 1932), but the connection to optimal transport was not made explicit until recently (Le´onard, 2014). We summarize this in the following theorem:

[0404] **Theorem 1.** Entropy regularized optimal transport gives the expectation of the distribution over cou- plings induced by Brownian motion (when the diffusion coefficient of the Brownian motion is equal to the entropy regularization parameter).

[0405] 4.3 Gradient flow and Waddington' s landscape

[0406] In this section we show how optimal transport can be interpreted as a gradient flow in gene expression space (capturing cell-autonomous processes) or in the space of distributions (capturing cell-nonautonomous processes). For a full treatment of the rich OT theory of gradient flows, we refer the reader to (Ambrosio et al., 2005; Santambrogio, 2015).

[0407] We began by considering the simple setting described by Waddington's landscape, which described a gradient flow in gene expression space and is a special case of what we could capture with optimal transport. Mathematically, Waddington's landscape defined a potential function Φ assigning potential energy Φ(χ) to a cell with expression profile *x*. The cells roll eddownhill according to the gradient of Φ to describe a trajectory *x(t)* satisfying the differential equation

$$\frac{dx}{dt} = -\nabla\Phi(x). \qquad (5)$$

[0408] This equation governing the trajectory of individual cells induced a flow in the distribution of the population of cells:

$$\frac{d\mathbb{P}_t}{dt} = \mathrm{div}!V\Phi\{x\text{-})\mathrm{F}_t]. \qquad (6)$$

[0409] Intuitively, this equation stated that the change in mass for each small volume of space (on the left-hand side) was equal to the flux of mass in and out (given by the divergence on the right hand side).

[0410] Optimal transport can capture this type of potential driven dynamics: the true coupling specified by (5) is close to the optimal transport coupling over short time scales. To motivate this, we appeal to a classical theorem establishing a dynamical formulation of optimal transport.

[0411] **Theorem 2** (Benamou and Brenier, 2001). The optimal objective value of the transport problem (1) is equal to the optimal objective value of the following optimization problem

$$\begin{array}{ll} \underset{\rho,v}{\text{minimize}} & \int_0^1 \int_{\mathbb{R}^G} \|v(t,x)\|^2 \rho(t,x)\,dt\,dx \\ \text{subject to} & \rho(0,\cdot) = \mathbb{P}, \quad p(l,\cdot) = \mathbb{Q} \quad\quad . \\ & \nabla\cdot(\rho v) = \dfrac{\partial\rho}{\partial t} \end{array} \qquad (7)$$

[0412] In this theorem, v was a vector-valued velocity field that advected the distribution *p* from P to Q, and the objective value to be minimized was the kinetic energy of the flow (mass x squared velocity). In our setting, the two distributions were snapshots P*s* and P_t of a developmental process at two time points, and the theorem showed that the transport map $\pi_{s,t}$

could be seen as a point-to-point summary of a least-action continuous time flow, according to an unknown velocity field. In the special case when the velocity field was the gradient of a potential Φ (i.e. Waddington landscape), the theorem implied that the coupling (5) achieved the optimal transport cost. In other words, OT could capture potential driven dynamics. In addition, optimal transport could also describe much more general settings. This velocity field could change over time and also depended on the entire distribution of cells, so optimal transport could describe very general developmental processes including those with cell-cell interactions, as described below.

[0413] We showed that the evolution (6) was a special case of a *Wasserstein gradient flow* to minimize the linear energy functional

$$E(\mathbb{P}) = \int \Phi(x) d\mathbb{P}(x).$$

[0414] We then described non-linear gradient flows, which can capture cell-cell interactions. To understand gradient flows, we started with the familiar notion of gradient descent:

$$x_{k+1} = -\eta \nabla E(x_k) + x_k.$$

[0415] This was rewritten as *aproximal procedure,* where one seeks to minimize E over all *x* in the proximity of¾

$$¾ \;_{+1} = \arg\min_{x} \quad \boldsymbol{E} \{x) + \frac{1}{2\eta} \| s - x_k \|^2. \tag{8}$$

[0416] We performed a similar proximal procedure in the space of distributions, replacing the Euclidean norm $\| \cdot \|^2$ with the Wasseerstein distance:

$$\mathbb{P}_{k+1} = \arg\min_{p} E(\rho) + \frac{1}{2\eta} W_2^2(\rho, \mathbb{P}_k). \tag{9}$$

[0417] This produced a sequence of iterates Po, Pi, . . . , P$_k$. The gradient flow was the limit obtained as we shrink the step-size $\eta \downarrow 0$. In (Richard Jordan and Otto, 1998), it's proven that for the linear energy functional

$$E(\mathbb{P}) = \int \Phi(x) d\mathbb{P}(x),$$

[0418] the limiting gradient flow converges to a solution of (6).

[0419] Going beyond the linear energy functional associated with Waddington' s landscape, one could describe cell-cell interactions with an interaction energy of the form

$$E(\mathbb{P}) = \iint I(x, y) d\mathbb{P}(x) d\mathbb{P}(y).$$

[0420] Gradient flows for interaction potentials are discussed in chapter 7 of (Santambrogio, 2015).

[0421] **Learning models of gene regulation** Motivated by this interpretation of optimal transport as a gradient flow according to an unknown vector field, we described a strategy to estimate such a vector field from data in **Waddington-OT: Concepts and Implementation .** We interpreted the vector field as a model of gene regulation - it predicted gene expression at later time points as a function of transcription factor expression at current time points. We assumed that the vector field did not change over time, and described a cell-autonomous flow, but we do not assume that it comes from a potential function.

[0422] **II. WADDINGTON-OT : Concepts and Implementation**

[0423] Building on the theoretical foundations developed in Modeling developmental processes with optimal transport, we developed WADDINGTON-OT: our method for computing ancestor and descendant trajectories, interpolating developmental processes, inferring gene regulatory models, and visualizing developmental landscapes. We begin with an overview in Section 1, and we then describe the specific details in Sections 2 - 8.

[0424] 1. Overview

[0425] To apply WADDINGTON-OT to a new dataset. The code is available on GitHub: https://github .com/broadinstitute/wot/

[0426] In the sections below we describe our procedures for computing transport maps, computing trajectories to cell sets, fitting local and global regulatory models, visualizing the developmental landscape, interpolating the distribution of cells at held-out time points.

[0427] To keep the focus here general-purpose, we deferred all reprogramming-specific details to the subsequent sections Methods.

[0428] **Input data:** The input to our suite of methods was a temporal sequence of single cell gene expression matrices, prepared as described in **Preparation of expression matrices.**

[0429] **Computing** transport **maps:** Waddington-OT calculated transport maps between consecutive time points and automatically estimated cellular growth and death rates. In Section 2 below we provide guidelines for defining the cost function, selecting regularization parameters and (optionally) providing an initial estimate of growth and death rates.

[0430] Ancestors, descendants, and trajectories: We describe in Section 3 how we computed trajectories plot trends in gene expression. Briefly, the *developmental trajectory* of a subpopulation of cells refers to the sequence of ancestors coming before it and descendants coming after it. Using the transport maps, we calculated the forward or backward transport probabilities between any two classes of cells at any time points. For example, we took successfully reprogrammed cells at day 18 and use back-propagation to infer the distribution over their precursors at day 17.5. We then propagated this back to day 17, and so on to obtain the ancestor distributions at all previous time points. This was the developmental trajectory to iPS cells. We plotted trends in gene expression over time.

[0431] Fitting regulatory models: We describe our method to fit a regulatory model to the transport maps in Section 4. Transcription factors (TFs) that appeared to play important roles along trajectories to key destinations were identified by two approaches. The first approach involved constructing a global regulatory model. Pairs of cells at consecutive time points were sampled according to their transport probabilities; expression levels of TFs in the cell at time t were used to predict expression levels of all non-TFs in the paired cell at time t + 1, under the assumption that the regulatory rules are constant across cells and time points. (TFs were excluded from the predicted set to avoid cases of spurious self-regulation). The second approach involved local enrichment analysis. TFs were identified based on enrichment in cells at an earlier time point with a high probability (> 80%) of transitioning to a given fate vs. those with a low probability (< 20%).

[0432] Visualizing **the developmental landscape** To visualize the developmental landscape, we first reduced the dimensionality of the data with diffusion components, and then embedded the data in two dimensions with force-directed graph visualization (as described in Section 5). While alternative visualization methods, such as t-distributed Stochastic Neighbor Embedding (t-SNE), were well suited for identifying clusters, they did not preserve global structures relevant to studying trajectories across a time course. FLE better reflected global structures by including repulsive forces between dissimilar points. In particular, these repulsive forces seemed to do a good job of splaying out the spikes present in the diffusionmap embedding.

[0433] **Geodesic interpolation:** To validate the temporal couplings, Waddington-OT could interpolate the distribution of cells at a held-out time point. The method wsa performing well if the interpolated distribution was close to the true held-out distribution (compared to the distance between different batches of the held-out distribution). Otherwise, it was possible that the method requires more data or finer temporal resolution.

[0434] Section 6 describes our method to interpolate the distribution of cells at a held-out time point. Our validation results for IPS reprogramming are presented in the subsequent section on **Validation by geodesic interpolation .** We performed extensive sensitivity analysis to show that our temporal couplings produce valid interpolations over a wide range of parameter settings perturbations to the data (down sampling cells or reads). See **QUANTIFICATION AND** STATISTICAL **ANALYSIS** for this sensitivity analysis.

[0435] 2. Computing transport maps

[0436] Recall that for any pair of time points we computed a transport plan that minimizes the expected cost of re-distributing mass, subject to constraints involving the relative growth rate (see **Modeling developmental processes with optimal transport** for a precise statement of the optimization problem). To compute these transport matrices, we needed to specify a cost function, numerical values for the regularization parameters, and (optionally) an initial estimate for the relative growth rate.

[0437] 2.1 Cost function

[0438] To compute the cost of transporting each individual point *x* from time *t* to position^ at time t_2, we first performed principal components analysis (PCA) on the data from this pair of time points to reduce to 30 dimensions. This dimensionality reduction was performed separately for each pair of adjacent time points. We defined the cost function to be squared Euclidean distance in this 'local -PCA space' .

[0439] Finally, we normalized the cost matrix by dividing each entry by the median cost for that time interval. Here the cost matrix was the matrix with entries Cy = c(xi, $_{y\ j}$) for each xi form time ti and $_{y\ j}$ at time t_2. This rescaling of the cost allowed us to refer to specific numerical values of the regularization parameters, without worrying about the global scale of distances.

[0440] 2.2 Regularization parameters

[0441] The optimization problem (4) involved three regularization parameters:

[0442] 1. The *entropy* parameter *E* controlled the entropy of the transport map. An extremely large entropy parameter gave a maximally entropic transport map, and an extremely small entropy parameter gave a nearly deterministic transport map. The default value was 0.05.

[0443] 2. λ\ controlled the degree to which transport was unbalanced along the rows. Large values of λ\ imposed stringent constraints related to relative growth rates. Small values of λ\ gave the algorithm more flexibility to change the relative growth rates in order to improve the transport objective. The default value was 1. To visually inspect the degree of unbalancedness, we recommend plotting the input row-sums vs the output row-sums of the transport map (See FIGs. 30A-30G).

[0444] 3. λι controlled the degree to which transport is unbalanced along the columns. The default value was $\lambda_2 = 50$. This large value essentially imposed equality constraints for the column marginals. A smaller value of λ_2 would allow different amounts of mass to transport to some cells at time t_2. We recommend keeping a large value for λ_2 so that the results are balanced along the columns. To visually inspect the degree of unbalancedness, one can plot the input column-sums vs the output column-sums of the transport map.

[0445] As we demonstrate in **QUANTIFICATION AND STATISTICAL ANALYSIS,** our validation results were stable over a wide range of values for *E* and λ\.

[0446] 2.3 Estimating relative growth rates

[0447] Our method solved the optimization problem (4) several times, using the output row-sums of the optimal transport map πˆï 1,ï2 as a new estimate for the relative growth rate function gˆ(x). By default, we initialize with *g(x)* = 1, so that all cells growed at the same rate. With some prior knowledge of growth rates (e.g. based on gene signatures of proliferation and apoptosis), this could be incorporated in the initial estimate for gˆ(x). For our reprogramming data, we showed how we formed an initial estimate for relative growth rates in **Estimating growth and death rates and computing transport maps**.

[0448] 3 Ancestors, descendants, and trajectories

[0449] Recall that the transport map $\hat{\pi}_{t}1, t2$ connecting cells from time *t*\ to cells from time t_2 has a row for each cell *x* at time *t*\ and a column for each cell *y* at time t_2. Each row specifies the *descendant distribution* of a single cell x from time t_1. The descendant mass is the sum of all the entries across a row. This row-sum was proportional to the number of descendants that x would contribute to the next time point. Intuitively, the descendant distribution specified which cells at time t_2 were likely to be descendants of x (see section 4.1 of **Modeling developmental processes with optimal transport** for the formal definition of descendants in a developmental process).

[0450] Similarly, each column specified the ancestor distribution of a cell y from time t_2. The ancestor mass was usually the same for each cell y. The ancestor distribution told us which cells at time t_1 were likely to give rise to the cell y.

[0451] Given a set of cells C, we computed the descendant distribution of the entire set by adding the descendant distributions of each cell in the set. This was computed efficiently via matrix multiplication as follows: Let S_1 donote all the cells from time point tl, and let

$$p\{x\} = \begin{cases} 1 & x \in C \\ 0 & \text{otherwise} \end{cases}$$

[0452] denote the uniform distribution on $C \subset S$. The descendant distribution of C was given by πˆï 1,ï2 p. One could compute ancestor distributions in a similar way

[0453] After computing the trajectory to or from a cell set C (in the form of a sequence of ancestor and descendant distributions), we computed trends in expression for any gene or gene signature along the trajectory. For each time point, we simply computed the mean expression weighting each cell according to the probability distribution defined by the ancestor or descendant distribution.

[0454] 4. Learning gene regulatory models

[0455] In this section we describe two strategies to summarize the transport maps by learning models of gene regulation. The first model we describe is a simple local enrichment analysis to identify transcription factors (TFs) enriched in ancestors of a set of cells. The second model is motivated by the dynamical systems formulation of optimal transport, as described above in Section 4.3.

[0456] 4.1 Local model: TF enrichment analysis of top ancestors

[0457] We performed local enrichment analysis as follows. Given a set of cells C at time t_2, we first computed the ancestor distribution of C at an earlier time t_1, as described in Section 3 above. We then selected cells contributing the most mass to the ancestor distribution, until a certain amount of mass was accounted for (e.g. 30% of the ancestor mass). We referred to these as the *top ancestors* at time t_1 of the cell set C. Finally, we compared the top ancestors to a null set of cells from the same time point. For example, this null cell set could be:

[0458] all cells except for the top ancestors,

[0459] the bottom *ancestors* (defined to be all cells except for the top ancestors of a less-strict cut-off),

[0460] the bottom ancestors restricted to a specialized subset (e.g. all other trophoblasts when C is a specific subset of trophoblasts like spongiotrophoblasts).

[0461] 4.2 Global model: learning a cell-autonomous gradient flow

[0462] To learn a simple description of the temporal flow, we assumed that a cell's trajectory was cell-autonomous and, in fact, depended only on its own internal gene expression. We knew this was wrong as it ignored paracrine signaling between cells, and we returned to discuss models that include cell-cell communication at the end of this section. However, this assumption is powerful because it exposes the time-dependence of the stochastic process P_t as arising from pushing an initial measure through a differential equation:

$$\dot{x} = f(x). \qquad (10)$$

[0463] Here f was a vector field that prescribes the flow of a particle x (see FIG. 4 for a cartoon illustration of a distribution flowing according to a vector field). Our biological motivation for estimating such a function f was that it encoded information about the regulatory networks that created the equations of motion in gene-expression space.

[0464] We set up a regression to learn a regulatory function f that models the fate of a cell at time $t,_{+1}$ as a function of its expression profile at time $t,$. Our approach involved sampling pairs of points using the couplings from optimal transport:

[0465] For each pair of time points $t,,$ $t,_{+1}$, we sampled pairs of cells $(X_{t_i}, X_{t_{i+1}})$ from the joint distribution specified by the transport map $\hat{¾}$ $_{,t_{i+1}}$.

[0466] Using the training data generated in the first step, we set up the followingregression:

$$\min_{f \in \mathcal{F}} \quad \mathbb{E}_{\hat{\pi}_{t_i,t_{i+1}}} \left\| X_{t_{i+1}} - f(X_{t_i}) \right\|^2,$$

[0467] where ^was a rectified-linear function class defined in terms of a specific generalized logistic function ℓ :¾ $\mapsto \mathbb{R}$:

$$\ell(x; k, b, y_0, XQ) = \frac{k y_0}{y_0 + (k - y_0)e^{-b(x-x_0)}},$$

[0468] where $k, b, y_0, x_0 \in \mathbb{R}$ were parameters of the generalized logistic function $\ell(x)$.

[0469] We define a function class -^consisting of functions $f : \mathbb{R}^G \to$ ¾^G of the form

$$f(x) = \boldsymbol{Ui(WTx)},$$

[0470] where i was applied entry-wise to the vector $WTx \in \mathbb{R}^M$ to obtain a vector that we multiplied against $U \in$ ¾$^{G \times M}$. Here $T \in$ ¾$^{G_{TF} \times G}$ denoted a projection operator that selected only the coordinated of x that were transcription factors, and GTF was the number of transcription factors. This gave a set of low-rank, linear functions with sparse factors. Each rank-1 component was interpreted as a regulatory module of transcription factors acting on a module of regulated genes.

[0471] We set up the following optimization over matrices

[0472]

$$\min_{U,W} \mathbb{E}_r \left\| \frac{X_{t_i} - X_{t_{i+1}}}{\Delta_t} - U\ell(WTX_{t_i}) \right\|^2 + \eta_1\|U\|_1 + \eta_2\|W\|_{1,} + \eta_3\|W\|_2^2 \qquad (11)$$
$$\text{s.t.} \quad U \geq 0.$$

[0473] where $(X_{ti}\ , Xti+\backslash\ )$ is a pair of random variables distributed according to the normalized transport map r, and $\|U\|_1$ denotes the sparsity-promoting/i norm of U, viewed as a vector (that is, the sum of the absolute value of the entries of U). Each rank one component (row of U or column of W) gives us a group of genes controlled by a set of transcription factors. The regularization parameters η_1 and η_2 control the sparsity level (i.e. number of genes in these groups).

[0474] **Implementation:** We designed a stochastic gradient descent algorithm to solve (11). Over a sequence of epochs, the algorithm sampled batches of points $(X t_i\ , Xti+\backslash\ )$ from the transport maps, computed the gradient of the loss, and updates the optimization variables U and x W. The batch sizes were determined by the Shannon diversity of the transport maps: for each pair of consecutive time points, we computed the Shannon diversity S of the transport map, then randomly sampled max(S 10^{-5}, 10) pairs of points to add to the batch. We ran for a total of 10, 000 epochs.

[0475] **Cell non-autonomous processes:** We concluded our treatment of gene regulatory networks by discussing an approach to cell-cell communication. Note that the gradient flow (10) only made sense for cell autonomous processes. Otherwise, the rate of change in expression $\dot{x}$ was not just a function of a cell's own expression vector x*(t),* but also of other expression vectors from other cells. We accommodated cell non-autonomous processes by allowing *f* to also depend on the full distribution P?:

$$\frac{dx}{dt} = f(x, \mathbb{P}_t). \qquad (12)$$

[0476] Concretely, we could allow *f* to depend on the mean expression levels of specific genes (expressed by any cell) encoding, for example, secreted factors or direct protein measurements of the factors themselves.

[0477] 5. Geodesic interpolation

[0478] Optimal transport provided an elegant way to interpolate distribution-valued data, analogous to how linear regression can be used to interpolate numerical or vector-valued data. Given two numerical data- points, a simply way to interpolate was to connect them with a line; this was the shortest path connecting the observed data. Given two distributions, we interpolated by finding the shortest path in the space of distributions. To do this we needed a notion of distance between distributions, and for this we use the metric induced by optimal transport. This metric space was called Wasserstein space, and this form of interpolation was called geodesic interpolation (Villani, 2008).

[0479] We derived a modified version of geodesic interpolation that took into account cell growth. Ordinarily, an interpolating distribution was computed by first computing a transport map between the distributions, and then connecting each point in the first distribution to points in the second according to the transport map. Finally, an interpolating point cloud was produced by from the midpoints of those line segments. (More generally, instead of taking just midpoints, one could also construct a family of interpolations that sweep from the first distribution to the second). We extended this framework to accommodate growth by changing the mass of the point we placed at the midpoint (to account for the fact that cells would have a different number of descendants at time *t* than they would at time *ti).*

[0480] Specifically, to interpolate at time $s \in (t_1, t_2)$ we first renormalize the rows of the transport map so they sum to roughly $\frac{\hat{g}(x)^{s-t_1}}{\int \hat{g}(x)^{s-t_1} d\hat{\mathbb{P}}_{t_1}}$ instead of $\frac{\hat{g}(x)^{t_2-t_1}}{\int \hat{g}(x)^{t_2-t_1} d\hat{\mathbb{P}}_{t_1}(x)}$ *–This took* into account the descendant mass each cell would have by time s instead of by time t_2. We then sampled points zi, . . . , z_N as follows:

[0481] 1. Sampling a pair of points (x, y) from the joint distribution specified by the transport map.

[0482] 2. Identifying the point

$$z = \alpha x + (1 - \alpha)y$$

along the line segment connecting x and y. Here a is given by $s = ah + (1 - \alpha)t_2$.

[0483] By repeating the steps above, we accumulate a point-cloud of points zi, . . . , z_N . Finally, we define the interpolating distribution as

$$\hat{\mathbb{P}}(s) = \frac{1}{N}\sum_{i=1}^{N}\delta_{z_i}.$$

[0484] Equipped with this notion of interpolation, we tested the performance of optimal transport by comparing the interpolated distribution to held-out time points. Using the data from time ti and ti+2, we interpolated to estimate the distribution Pti+1 . We then computed the Wasserstein distance between the interpolated distribution and the observed distribution. We compared this distance to a null model generated from the independent coupling where we sample pairs (x, y) independently $x \sim \hat{\mathbb{P}}_{t_i}$ and $y \sim \hat{\mathbb{P}}_{t_{i+2}}$ in step 1 above. We also compared the interpolated distance to distance between batches of $\mathbb{P}_{t_{i+1}}$. Optimal transport was performing well if the interpolated point cloud was as close to the batches of the held out time point as the batches were to each other, and the null-interpolated point cloud was farther away.

[0485] **Bibliography**

- Ambrosio, L., Gigli, N., and Savare, G. (2005). Gradient Flows: In Metric Spaces and in the Space of Probability Measures. Lectures in Mathematics. ETH Zu¨rich. Birkha¨user Basel.
- Bastian, M., Heymann, S., Jacomy, M., et al. (2009). Gephi: an open source software for exploring and manipulating networks. Icwsm, 8:361-362.
- Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D., Li, S., and Li, M. S. (2015). Package FNN.
- Chizat, L., Peyre', G , Schmitzer, B., and Vialard, F.-X. (2017). Scaling algorithms for unbalanced transport problems. Mathematics of Computation.

- Cuturi, M. (2013). Sinkhorn distances: Lightspeed computation of optimal transportation distances. In
- Neural Information Processing Systems (NIPS).
- Jacomy, M., Venturini, T., Heymann, S., and Bastian, M. (2014). Forceatlas2, a continuous graph layout algorithm for handy network visualization designed for the gephi software. PloS one, 9:e98679.
- Le ́onard, C. (2014). A survey of the schro ̈dinger problem and some of its connections with optimal transport. Discrete and Continuous Dynamical Systems - Series A (DCDS-A), 34(4):1533-1574.
- Porpiglia, E., Samusik, N., Van Ho, A. T., Cosgrove, B. D., Mai, T., Davis, K. L., Jager, A., Nolan, G. P., Bendall, S. C , Fantl, W. J., et al. (2017). High-resolution myogenic lineage mapping by single-cell mass cytometry. Nature Cell Biol., 19:558-567.
- Richard Jordan, D. K. and Otto, F. (1998). The variational formulation of the fokker. SIAM J. Math. Anal., 29(1): 1-17.
- Samusik, N., Good, Z., Spitzer, M. H., Davis, K. L., and Nolan, G. P. (2016). Automated mapping of phenotype space with single-cell data. Nature methods, 13:493-496.
- Santambrogio, F. (2015). Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling. Progress in Nonlinear Differential Equations and Their Applications. Springer Inter- national Publishing.
- Schrodinger, E. (1932). Sur la theorie relativiste de l'electron et l’interpretation de la mecanique quan- tique. Ann. Inst. H. Poincare, 2:269-310.
- Villani, C. (2008). Optimal Transport Old and New. Springer.
- Zunder, E. R., Lujan, E., Goltsev, Y., Wernig, M., and Nolan, G. P. (2015). A continuous molecular roadmap to ipse reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell, 16:323-337.

[0486] **III Experimental methods**

[0487] 1. Derivation of secondary MEFs

[0488] OKSM secondary Mouse embryonic fibroblasts (MEFs) were derived from E13.5 female embryos with a mixed B6;129 background. The cell line used in this study was homozygous for ROSA26-M2rtTA, homozygous for a polycistronic cassette carrying *Oct4, Klf4,* *Sox2,* and *Myc* at the *Collal* locus and homozygous for an EGFP reporter under the control of the *Oct4* promoter (Stadtfeld et al., 2010). Briefly, MEFs were isolated from E13.5 embryos from timed-matings by removing the head, limbs, and internal organs under a dissecting microscope. The remaining tissue was finely minced using scalpels and dissociated by incubation at 37°C for 10 minutes in trypsin-EDTA (Thermo Fisher Scientific). Dissociated cells were then plated in MEF medium containing DMEM (Thermo Fisher Scientific), supplemented with 10% fetal bovine serum (GE Healthcare Life Sciences), non-essential amino acids (Thermo Fisher Scientific), and GlutaMAX (Thermo Fisher Scientific). MEFs were cultured at 37°C and 4% CO2 and passaged until confluent. All procedures, including maintenance of animals, were performed according to a mouse protocol (2006N000104) approved by the MGH Subcommittee on Research Animal Care.

[0489] 2. Derivation of Primary MEFs

[0490] Primary MEFs were derived from E13.5 embryos with a B6.Cg-Gt(ROSA)$^{26Sortml(rtTA*M2)Ja}$7JxB6;129S4-Pou5fl ^{tm2Jae}/J background. The cell line was homozygous for ROSA26-M2rtTA, and homozygous for an EGFP reporter under the control of the Oct4 promoter. MEFs were isolated as mentioned above.

[0491] 3. Reprogramming assay

[0492] For the reprogramming assay, 20,000 low passage MEFs (no greater than 3-4 passages from isolation) were seeded in a 6-well plate. These cells were cultured at 37°C and 5% CO2 in reprogramming medium containing KnockOut DMEM (GIBCO), 10% knockout serum replacement (KSR, GIBCO), 10% fetal bovine serum (FBS, GIBCO), 1% GlutaMAX (Invitrogen), 1% nonessential amino acids (NEAA, Invitrogen), 0.055 mM 2-mercaptoethanol (Sigma), 1% penicillin-streptomycin (Invitrogen) and 1,000 U/ml leukemia inhibitory factor (LIF, Millipore). Day 0 medium was supplemented with 2 μg/mL doxycycline Phase-l(Dox) to induce the polycistronic OKSM expression cassette. Medium was refreshed every other day. At day 8, doxycycline was withdrawn, and cells were transferred to either serum-free 2i medium containing 3 μM CHIR99021, 1 μM PD0325901, and LIF (Phase-2(2i)) (Ying et al., 2008) or maintained in reprogramming medium (Phase-2(serum)). Fresh medium was added every other day until the final time point on day 18. Oct4-EGFP positive iPSC colonies should start to appear on day 10, indicative of successful reprogramming of the endogenous Oct4 locus.

[0493] 4. Sample collection

[0494] We profiled a total of 315,000 cells from two time-course experiments across 18 days in two different culture conditions: in the first we profiled ~65,000 cells collected over 10 time points separated by ~48 hours; in the second we profiled ~250,000 cells collected over 39 time points separated by ~12 hours across an 18-day time course (and every 6 hours between days 8 and 9). In the larger experiment, duplicate samples were collected at each time point. Cells were also collected from established iPSCs cell lines reprogrammed from the same MEFs, maintained either in Phase-2(2i) conditions or in Phase-2(serum) medium. For all time points, selected wells were trypsinized for 5 mins followed by inactivation of trypsin by addition of MEF medium. Cells were subsequently spun down and washed with IX PBS supplemented with 0.1% bovine serum albumin. The cells were then passed through a 40 micron filter to remove cell debris and large clumps. Cell count was determined using Neubauer chamber hemocytometer to a final concentration of 1000 cells/μl.

[0495] 5. Single-cell RNA-seq

[0496] ScRNA-seq libraries were generated from each time point using the 10X Genomics Chromium Controller Instrument (10X Genomics, Pleasanton, CA) and Chromium™ Single Cell 3' Reagent Kits v1 (~65,000 cells experiment) and v2 (~250,000 experiment) according to manufacturer's instructions. Reverse transcription and sample indexing were performed using the CIOOO Touch Thermal cycler with 96-Deep Well Reaction Module. Briefly, the suspended cells were loaded on a Chromium controller Single-Cell Instrument to first generate single-cell Gel Bead-In-Emulsions (GEMs). After breaking the GEMs, the barcoded cDNA was then purified and amplified. The amplified barcoded cDNA was fragmented, A-tailed and ligated with adaptors. Finally, PCR amplification was performed to enable sample indexing and enrichment of the 3' RNA-Seq libraries. The final libraries were quantified using Thermo Fisher Qubit dsDNA HS Assay kit (Q32851) and the fragment size distribution of the libraries were determined using the Agilent 2100 BioAnalyzer High Sensitivity DNA kit (5067-4626). Pooled libraries were then sequenced using Illumina Sequencing. All samples were sequenced to an average depth of 87 million paired-end reads per sample (see Experimental Methods), with 98 bp on the first read and 10 bp on the second read. In the larger experiment, we profiled 259,155 cells to an average depth of 46,523 reads per cell.

[0497] 6. Lentivirus vector construction and particle production

[0498] To test whether transcription factors (TFs) improve late-stage reprogramming efficiency, we generated lentiviral constructs for the top candidates *Zfp42,* and *Obox6.* cDNAs for these factors were ordered from Origene *(ZJp42*-MG203929, and *Obox6-MR2* 15428) and cloned into the FUW Tet-On vector (Addgene, Plasmid #20323) using the Gibson Assembly (NEB, E2611S). Briefly, the cDNA for each TF was amplified and cloned into the backbone generated by removing *Oct4* from the FUW-Teto-Oc*t4* vector. All vectors were verified by Sanger sequencing analysis. For lentivirus production, HEK293T cells were plated at a density of 2.6xlO cells/well in a 10cm dish. The cells were transfected with the lentiviral packaging vector and a TF-expressing vector at 70-80% growth confluency using the Fugene HD reagent (Promega E2311), according to the manufacturer's protocols. At 48 hours after transfection, the viral supernatant was collected, filtered and stored at -80°C for future use.

[0499] 7. Reprogramming efficiency of secondary MEFS together with individual TFs

[0500] We sought to determine the ability of the candidate TFs to augment reprogramming efficiency in secondary MEFs; the use of secondary MEFs for reprogramming overcomes limitations associated with random lentiviral integration events at variable genomic locations. Briefly, secondary MEFs were plated at a concentration of 20,000 cells per well of a 6-well plate. Cells were infected with virus containing Zfp42, Obox6, or an empty vector and maintained in reprogramming medium as described above. At day 8 after induction, cells were switched to either Phase-2(2i) or Phase-2(serum). On day 16, reprogramming efficiency was quantified by measuring the levels of the EGFP reporter driven by the endogenous *Oct4* promoter. FACS analyses was performed using the Beckman Coulter CytoFLEX S, and the percentage of Oct4-EGFP^{+} cells was determined. Triplicates were used to determine average and standard deviation.

[0501] 8. Reprogramming efficiency of primary MEFS with individual TFs and OKSM

[0502] We also independently tested the performance of TFs in primary MEFs. To this end, lentiviral particles were generated from four distinct FUW-Teto vectors, containing Oct4, Sox2, Klf4, and Myc, previously developed in the Jaenisch lab. MEFs from the background strain B6.Cg-Gt(ROSA)26Sor $_{\text{tmi(rtTA*M2)}^{\text{j}}\text{ae}}$/j _ B6;129S4-Pou5fl $^{\text{tm2Jae}}$/J were infected with these lentiviral particles, together with a lentivirus expressing tetracycline-inducible Zfp42, Obox6 or no insert. Infected cells were then induced with 2 μg/mL doxycycline in ESC reprogramming medium (day 0). At day 8 after induction, cells were switched to either Phase-2(2i) or Phase-2(serum). On day 16, the number of Oct4-EGFP$^+$ colonies were counted using a fluorescence microscope. Triplicates for each condition used to determine average values and standard deviation.

[0503] **IV. Preparation of expression matrices**

[0504] To compute an expression matrix from scRNA-Seq data, we aligned sequenced reads to obtain a matrix U of UMI counts, with a row for each gene and a column for each cell. To reduce variation due to fluctuations in the total number of transcripts per cell, we divide the UMI vector for each cell by the total number of transcripts in that cell. Thus we define the expression matrix E in terms of the UMI matrix U via:

$$E = \frac{U_{ij}}{\sum_{i=1}^{G} U_{ij}} \times 10^4.$$

[0505] In our subsequent analysis, we make use of two variance-stabilizing transforms of the expression matrix E. In particular, we define

1. $\tilde{E}$ to be the log-normalized expression matrix. The entries of $\tilde{E}$ are obtained via
$$\tilde{E} = logiEi_j + 1)$$
2. $\bar{E}$ to be the truncated expression matrix. The entries of $\bar{E}$ are obtained by capping the entries of $\bar{E}$ at the 99.5% quantile.

[0506] When we refer to an expression profile, by default we refer to a column of $\tilde{E}$ unless otherwise specified.

[0507] 1. Aligning reads

[0508] The 98 bp reads were aligned to the UCSC mmlO transcriptome, and a matrix of UMI counts was obtained using Cellranger from the 10X Genomics pipeline (v2.0.0) with default parameters (https: //support. lOxgenomics.com/single-cell-gene-expression/software/pi pelines/latest/installati on). Quality control metrics about barcoding and sequencing such as the estimated number of cells per collection and the median number of genes detected across cells are summarized in Table 14. To estimate expression of exogenous OKSM factors from OKSM cassette, we extracted RBGpA sequence (839 bp) from the OKSM cassette FASTA file, and generated a reference using the mkref function from the Cellranger pipeline.

[0509] 2. Downsampling and filtering expression matrix

[0510] The expression matrix was downsampled to 15,000 UMIs per cell. Cells with less than 2000 UMIs per cell in total and all genes that were expressed in less than 50 cells were discarded, leaving 251,203 cells and G= 19,089 genes for further analysis. The elements of expression matrix were normalized by dividing UMI count by the total UMI counts per cell and multiplied by 10,000 i.e. expression level is reported as transcripts per 10,000 reads.

[0511] 3. Selecting variable genes

[0512] We used the function MeanVarPlot from the Seurat package (v2.1.0) (Satija et al., 2015) to select 1479 variable genes. First, we divided genes into 20 bins based on their average expression levels across all cells. Second, we computed Fano factor of gene expression in each bin and then z-scored. The Fano factor, defined as the variance divided by the mean, was a measure of dispersion. Finally, by thresholding the z-scored dispersion at 1.0, we obtained a set of 1479 variable genes. After selecting variable genes, we created a variable gene expression matrix by renormalizing as described above.

[0513] **V. Visualization; force-directed layout embedding**

[0514] In this section we introduced our two dimensional visualization technique based on force-directed layout embedding (FLE) (Bastian et al., 2009; Jacomy et al., 2014). FLE was large-scale graph visualization tool which simulated the evolution of a physical system in which connected nodes experience attractive forces, but unconnected nodes experience repulsive forces. It better captured global structures than tSNE. Initial FLE algorithms used simple electrostatic and spring forces, but modern FLE algorithms allowed for more elaborate interactions that could depend on the degree of nodes or included gravity terms that attracted all nodes to the center (this was especially important for disconnected graphs, which would otherwise fly apart). Starting from a random initial position of vertices, the network of nodes evolved in such a manner that at any iteration a new position of vertices was computed from the net forces acting on them.

[0515] We applied FLE to visualize the nearest neighbor graph generated from our data.

[0516] **Implementation:** Our visualization took as input the expression matrix of highly-variable genes, selected as described in the previous section of the STAR Methods. First, we reduced to 100 dimensions by computing a 100 dimensional diffusion component embedding of the dataset using SCANPY (vO.2.8) with default parameters. Second, for each cell we computed its 20 nearest neighbors in 100-dimensional diffusion component space to produce a nearest neighbor graph. For this step, we used the approximate k-NN algorithm Annoy from the R package RCPPANNOY (vO.0.10). Finally, we computed the force-directed layout on the k-NN graph using the ForceAtlas2 algorithm (Jacomy et al., 2014) from the Gephi Toolkit (vO.9.2) (Bastian et al., 2009).

[0517] **VI. Creating gene signatures and cell sets**

[0518] 1. Gene signatures

[0519] We then constructed curated gene signatures from various databases of gene signatures. Given a set of genes, we scored cells based on their gene expression. In particular, for a given cell we computed the z-score for each gene in the set. We then truncated these z-scores at 5 or -5, and defined the signature of the cell to be the mean z-score over all genes in the gene set.

[0520] The table below summarizes the sources from which we obtained signatures. In two cases (neural identity and epithelial identity), we constructed signatures manually using marker genes. A pluripotency gene signature was determined in this work using the pilot dataset. We performed differential gene expression analysis between two groups of cells: mature iPSCs and cells along the time course DO to D16 and took the top 100 genes with increased expression in mature iPSCs. A proliferation gene signature was obtained by combining genes expressed at Gl/S and G2/M phases.

[0521] In several places, we also computed gene signatures based on co-expression with a given gene of interest. For instance, in the stromal region we noticed several genes *(Cxcll2, Ifitml,* and *Matn4)* with expression patterns that were distinct from a signature of long-term cultured MEFs (FIG. 31D). For each gene, we computed a co-expression signature by finding the set of genes with expression levels in stromal cells that were >\5% correlated with the gene of interest. We found that these gene signatures were significantly overlapping (p-value < 0.01, hypergeometric test) with signatures of stromal cells in neonatal muscle and neonatal skin in the Mouse Cell Atlas. Similarly, in the neural region we derived signatures of genes co-expressed with *Gadl* and with *Slcl7a6* (FIG. 33C). These signatures significantly overlapped signatures of inhibitory and excitatory neurons, respectively, derived from the Allen Brain Atlas.

Gene Signature	Source
MEF identity	(Chen et al., 2013; Han et al., 2018; Lattin et al., 2008)
Pluripotency	This work.
Proliferation	(Tirosh et al., 2016)
ER stress	GO:0034976, Biological Process Ontology
Epithelial identity	This work. Marker genes: (Li et al., 2010; Takaishi et al., 2016; Whiteman et al., 2014)
ECM rearrangement	GO:0030198, Biological Process Ontology
Apoptosis	Hallmark P53 Pathway, MSigDB
Senescence	(Coppe et al., 2010)
Neural identity	This work. Marker gene sources: (Fonseca et al., 2013; Gouti et al., 2011; Kan et al., 2004; Lazarov et al., 2010; Sakakibara et al., 2001; Sansom et al., 2009; Watanabe et al., 2017)
Trophoblast	(Han et al., 2018)
X reactivation	chromosome X
XEN	(Lin et al., 2016)
Trophoblast progenitors	(Han et al., 2018)
Spiral Artery Trophpblast Giant Cells	(Han et al., 2018)
Oligodendrocyte precursor cells (OPC)	(Tasic et al., 2016)
Astrocytes	(Tasic et al., 2016)
Cortical Neurons	(Tasic et al., 2016)
RadialGlia-ld3	(Han et al., 2018)
RadialGlia-GdflO	(Han et al., 2018)
RadialGlia-Neurog2	(Han et al., 2018)
Long-term MEFs	(Han et al., 2018)

Embryonic mesenchyme	(Han et al., 2018)
Cxcl l 2 co-expressed	This work.
Ifitm l co-expressed	This work.
Matn4 co-expressed	This work.
2,4,8, 16,32-cell	(Goolam et al., 2016)

[0522] 2. Cell sets

[0523] Using the gene signatures described above, we created coarse cell sets defining the broad regions of the landscape (iPSC, Trophoblast, Neural, Stromal, Epithelial, and MET), and cell subtype sets defining different cell types within a region (stromal, trophoblast, and neural subtypes, along with 2- through 32-cell stages).

[0524] To define the coarse cell sets, we first computed a rough partitioning of the landscape by clustering cells using the Louvain method of spectral clustering to obtain 65 cell clusters using k=5 nearest neighbors (FIG. 34A). By examining signature score activity levels over clusters, we grouped several clusters to form cell sets for the iPSC, Stromal and Neuronal regions. Because our densely sampled data did not always segregate into distinct clusters, we defined some additional coarse cell sets by signature scores. We defined the trophoblast cell set to include all cells with Trophoblast signature greater than 0.7. We defined the epithelial cell set to include all cells with epithelial identity signature greater than 0.8, minus all cells included in other cell sets (mostly removing the trophoblasts with epithelial signature). Finally, we defined the MET Region as the ancestors of iPS, Trophoblast, Neural and Epithelial cells. In particular, we computed the top ancestors of each major cell set, then merged these cell sets and removed the cells *in* each major cell set.

[0525] Within the Stromal, Trophoblast, Neural and iPSC cell sets, we then conducted more sensitive statistical tests for cell subtype signatures. We did this by calculating empirical p-values for the subtype signature score for each (region-specific) subtype in each cell. In each of 100,000 permutation trials, we randomly and independently shuffled the expression levels of each gene across the cells within a region. In each cell, we then computed signature scores in the permuted data, and generated p-values by determining the frequency at which the permuted score was greater than the original score. While the results shown in figures and discussed in the main text were based on shuffling genes across cells, we similarly permuted the expression levels within each cell, and found consistent results. Finally, we controlled for multiple hypothesis testing by calculating FDR q-values, and used a threshold FDR of 10% to define cell subtype sets.

[0526] **VII. Estimating growth and death rates and computing transport maps**

[0527] 1. Initial estimate of growth rates

[0528] We formed an initial estimate of the relative growth rate as the expectation of a birth-death process on gene expression space with birth-rate $\beta(\chi)$ and death rate $\delta(\chi)$ defined in terms of expression levels of genes involved in cell proliferation and apoptosis. Multi-state birth-death processes had been used before to model growth, death, and transitions in iPS reprogramming (Liu et al., 2016). A birth-death process was a classical model for how the number of individuals in a population could vary over time. The model was specified in terms of a birth rate β and death rate δ: During a time interval At, the probability of a birth was $\beta\Delta$ï and the probability of a death was $\delta\Delta$ï. The doubling time for a birth death process was defined as follows. Starting with N(0) = n, the time τ it would take to get to an expected population size of $EN(t) = 2n$ is

$$\tau = \frac{\text{In } 2}{\beta - \delta}$$

[0529] The half-life could be computed in a similar way. We applied a sigmoid function to transform the proliferation score into a birth rate. The sigmoid function smoothly interpolated between maximal and minimal birth rates. We specified the maximal birth rate to be $\beta_{MAX} = 1.7$. Therefore, the fastest cell doubling time is

$$\frac{\ln 2}{1.7} \approx 0.41\ days \approx 9.6\ hours,$$

by the doubling time equation above. We defined the minimal birth rate as $\beta_{\text{MîN}} = 0.3$. Therefore the slowest cell doubling time is

$$\frac{\ln 2}{0.3} = 2.3\ days = 55\ hours.$$

[0530] Similarly, we transformed the apoptosis signature into an estimate of cellular death rates by applying a sigmoid function to smoothly interpolate between minimal and maximal allowed death rates. We defined the minimal death rate parameter to be $\delta_{MIN} = 0.3$, and the maximal death rate parameter as $\delta_{MAX} = 1.7$. By the calculations above, these correspond to half-lifes of 55 and 9.6 hours respectively.

[0531] 2. Learning growth rates and computing transport maps

[0532] Using the growth rates defined in the previous section as an initial estimate, we computed transport maps and automatically improved these growth rates using the Waddington-OT software package (see Section _Computing transport maps_). For the cost function, we used squared Euclidean distance in 30 dimensional local PCA space computed on the variable gene data from the relevant pair of time points. We used the following parameter settings:

$$e = 0.05,\ \lambda_1 = 1, \lambda_2 = 50,\ \text{growth\_iters} = 3.$$

[0533] The parameters $\lambda_1$and λ_2 control the degree to which the row-sums and column-sums were unbalanced. A larger value of λ_1 induced a greater correlation between the input and output growth rates. The Waddington-OT package iterated the procedure of computing transport maps based on input growth rates, and then using the output growth rates as new input growth rates to recompute transport maps. We ran this for growth_iters = 3 total iterations.

[0534] This gave us a set of transport maps between each pair of time points, which could be used to estimate the temporal coupling. From this estimate of the temporal coupling, we computed ancestor and descendant distributions to each of the major cell sets defined in the previous section.

[0535] **VIII. Regulatory analysis**

[0536] We performed regulatory analysis to identify modules of transcription factors regulating modules of genes with our global regulatory model from the Waddington-OT software package, described in Section _Learning gene regulatory models_. The optimization began by specifying the number of gene modules, and establishing an initial estimate for each. We used spectral clustering to initialize the modules: genes were clustered into 50 sets, with one module corresponding to each set, and weights set to 0 for genes outside the set, and 1 for genes within the set.

[0537] We then specified a time lag between TF and gene module expression. In order to test for potential regulatory interactions on different time scales, we computed global regulatory models with three time lags: 6hrs, 48hrs, and 96hrs. This allowed us to identify factors that were predictive several days in advance —for instance, Nanog is a very early predictor of pluripotency and was found to be associated with a pluripotency associated gene expression module in the 96 hour model —as well as those predictive on shorter time scales —for instance, we TFs that were predictive of neural-associated expression modules in the 6 and 48 hour models, but did not find such predictive TFs in the 96 hour model.

[0538] Finally, we set regularization and stochastic block size parameters. Default values available in the code online were used in this study. Briefly, regularization parameters were tuned on small training datasets to enforce sparsity (l1 penalties) and reduce model complexity (l2 penalty) while still achieving a good fit (>60% correlation between predicted and observed expression) in training data. These parameters may be specifically tuned in new datasets. The stochastic block size and number of epochs were set according to available hardware resources.

[0539] **IX. Validation by geodesic interpolation**

[0540] We validated Waddington-OT by demonstrating that we could accurately interpolate the distribution of cells at held out time points. We applied geodesic interpolation (described in **Waddington-OT; Concepts and Implementation**) to our reprogramming data to predict the distribution of cells at each time point, using only the data from the previous and next time points. In other words, we sought to predict the distribution P_{t_2} at time t_2 from the distributions at neighboring time points: P_{t_1} and P_{t_3} (FIGs. 24H, 30D). To determine a baseline for performance, we examined the distance between the two different batches of the held-out distribution (FIGs. 24H, 30D).

[0541] To compute the optimal transport coupling from P_{t_1} to P_{t_3}, we used the Waddington-OT package with default parameters. For the cost function we computed 30 dimensional local PCA coordinates using only the points from time $t_1$and t_3. We then embedded the data from time t_2 into the 30 dimensional local PCA space which was computed using only the data from time $t_1$and t_3. Finally, we used Wasserstein-2 distance to compute distance between point clouds.

[0542] **X. Paracrine signaling**

[0543] To characterize potential cell-cell interactions between contemporaneous cells during reprogramming, we first collected a list of ligands and receptors found in the GO database. The set of ligands (415 genes) was a union of three gene sets from the following GO terms:

1) *cytokine activity* (GO:0005125),
2) *growthfactor activity* (GO: 0008083), and
3) *hormone activity* (GO: 0005 179).

[0544] The set of receptors (2335 genes) was defined by the GO term *receptor activity* (GO: 0004872). Next, we used a curated database of mouse protein-protein interactions (Mertins et al., 2017) and identified 580 potential ligand-receptor pairs.

[0545] First, we defined an interaction score /A;B;X;Y;t as the product of (1) the fraction of cells (*E*A;X;t) in cell-set A expressing ligand X at time t and (2) the fraction of cells (i¾;Y;t) in cell-set B expressing the cognate receptor Y at time *t*. We define the aggregate interaction score lA;B;t as a sum of the individual interaction scores across all pairs:

$$I_{A;B;t} = \sum_{All\ X\text{-}Y\ pairs} I_{A;B;X;Y;t} = \sum_{All\ X\text{-}Y\ pairs} F_{A;X;t}\ F_{B;Y;t}$$

[0546] We depicted the aggregate interaction scores for all combinations of cell clusters in FIGs. 28B, 34B.

[0547] Second, we sought to explore individual ligand-receptor pairs at a given day and condition between cell ancestors of interest. For this purpose we defined the interaction score ^A;B;X;Y;t as the product of (1) the average expression of the ligand X in ancestors at time t of a cell set A and (2) the average expression of the cognate receptor Y in ancestors at time t of a cell set B. Values of the interaction scores /A;B;X;Y ;t are high for ubiquitously expressed ligands and receptors at a given day and may be nonspecific to a pair of cell ancestors of interest. Thus, we used permutations to generate an empirical null distribution of interaction scores. In each of the 10,000 permutations, we randomly shuffled the labels of cells and calculated the interaction score *I*s A;B;X;Y ;t. We then standardized each ligand-receptor interaction score by taking the distance between the interaction score /A;B;X;Y;t and the mean interaction score in units of standard deviations from the permuted data

((/A;B;X;Y ;t - mean(/ SA;B;X;Y;t))/sd(/SA;B;X;Y;t)).

[0548] We depicted examples of standardized interaction scores ranked by their values in FIGs. 28C-28E and 34C-34E. Replacement of the average expression of the ligand with the total expression of the ligand in the calculation of the standardized interaction score did not affect the results.

[0549] **XI. Classification of differential genes along the trajectory to iPSCs**

[0550] To identify differential genes along the successful trajectory to iPSCs we computed the average expression (TPM) of all 19,089 genes in ancestors of iPSCs. The average expression values were log2 transformed and we filtered out genes for which the difference between maximal and minimal expression value between day 0 and day 18 was less than 1, leaving 2311 genes for further analysis. The genes were classified into 15 groups by k-means clustering as implemented in the R package stats. To identify the number of clusters we applied a gap statistic (Tibshirani et al. 2001) using the function clusGap from R package cluster v2.0.6.

[0551] We performed functional enrichment analysis on the identified gene clusters using the findGO.pl program from the HOMER suite (Hypergeometric Optimization of Motif Enrichment, v4.9. 1) (Heinz et al. 2010) with Benjamini and Hochberg FDR correction for multiple hypothesis testing (retaining terms at FDR < 0.05). All genes that passed quality-control filters were used as a background set.

[0552] **XII. Identifying large chromosomal aberrations**

[0553] We have previously developed methods to identify copy number variations (CNVs) in scRNA-Seq data from tumor samples (Patel et al., 2014; Tirosh et al., 2016). That analysis differed from our current study in two key aspects: (1) the data were based on full length scRNA-seq (SMART-Seq2), and sequenced to greater depth in each cell, and (2) there we could rely on the clonal expansion of CNVs to make it easier to identify recurring chromosomal aberrations.

[0554] We performed three types of analysis to detect aberrant expression in large chromosomal regions. First, we searched cells with significant up- or down-regulation at the level of entire chromosomes. Second, we ran a coarse analysis to identify cells with significant net aberrant expression across windows spanning 25 broadly-expressed genes. Focusing on regions that were enriched for cells with significant aberrations found by this coarse filter, we then performed a more sensitive test to compute the significance of aberrations in each window in each cell.

[0555] Empirical p-values and false discovery rates (FDRs) for both analyses were computed by randomly permuting the arrangement of genes in the genome, as described below. Permutations for both types of analysis were done as follows. In each of 100,000 permutations we randomly shuffled the labels of genes in the entire dataset, while preserving the genomic coordinates of genes (with each position having a new label each time) and the expression levels in each cell (so that each cell has the same expression values, but with new labels). We then computed either whole chromosome or subchromosomal aberration scores for each cell.

[0556] To identify whole-chromosome aberrations scores in each cell, we began by calculating the sum of expression levels in 25Mbp sliding windows along each chromosome, with each window sliding IMbp so that it overlapped the previous window by 24Mbp. For each window in each cell, we then calculated the Z-score of the net expression, relative to the same window in all other cells. We then counted the fraction of windows on each chromosome with an absolute value Z-score > 2. This fraction served as the whole-chromosome aberration score for each chromosome in each cell. To assign a p-value to the whole-chromosome score for cell(i) chromosome(j), we calculated the empirical probability that the score for cell(i) chromosome(j) in the randomly permuted data was at least as large as the score in the original data.

[0557] Subchromosomal aberration scores were computed as follows. We began by identifying the 20% of genes with the most uniform expression across the entire dataset. This was done by calculating the Shannon Diversity $e^{-¾\ E_{gc} \ln E_{gc}}$ for each gene g (where $E_{g\ c}$ was the expression matrix as defined above in **<u>Preparation of expression matrices</u>)**, and taking the 20% of genes with the largest values. Using these genes, we subset the expression matrix and renormalized by TPM, and then computed in each cell the sum of expression in sliding windows of 25 consecutive genes, with each window sliding by one gene and overlapping the previous window (on the same chromosome) by 24 genes. In each window, we calculated the Z-score relative to all cells at day 0. The net (coarse filter) subchromosomal aberration score for a cell was calculated as the 12-norm of the Z-scores across all windows. To assign a p-value to the subchromosomal aberration score for cell(i), we calculated the empirical probability that the score for cell(i) in the randomly permuted data was at least as large as the score in the original data.

[0558] Finally, to identify the specific region(s) of genomic aberrations in each cell, we conducted a more sensitive test using just the cells in the stromal and trophoblast regions. Again using 25 housekeeping gene windows, we computed the average z-score of gene expression for genes in each window in each cell. We then compared the scores in all windows in all cells to similar scores computed for each cell in 100,000 random permutation trials, and then assigned p-values based on the frequency of extremely high (gain) or low (loss) expression values.

[0559] For each of the aberration scores and associated p-values described above, we controlled for multiple hypothesis testing by calculating FDR q-values, using a false discovery threshold of 10%.

[0560] QUANTIFICATION AND STATISTICAL ANALYSIS

[0561] **I. Analyzing the stability of optimal transport**

[0562] To test the stability of our optimal transport analysis to perturbations of the data and parameter settings, we downsampled the number of cells at each time point, downsampled the number of reads in each cell, perturbed our initial estimates for cellular growth and death rates, and perturbed the parameters for entropic regularization and unbalanced transport. We found that our geodesic interpolation results are stable to a wide range of perturbations, summarized in the following table:

Number of cells per batch	**Number of UMIs Per cell**	**Max Growth** β_{MAX}	**Min Growth** β_{MIN}	**Max Death** δ_{MAX}	**Min Death** δ_{MIN}	**Entropy regularization** ϵ	**Unbalanced transport** λ
Down to: 200	Down to: 1000	33 hrs to 5.5 hrs	None to 9.5 hrs	33 hrs to 5.5 hrs	None to 9.5hrs	5×10^{-5} to 0.5	0.1 to 32

[0563] To generate this table, we ran geodesic interpolation with all but one of these settings fixed to default values. The default parameter values that we used were:

$$e = 0.05,\ \lambda_1 = 1, \lambda_2 = 50,\ \beta_{MAX} = 1.7,\ \delta_{MAX} = 1.7,\ \beta_{MIN} = 0.3,\ \delta_{MIN} = 0.3.$$

[0564] Moreover, by default we used all reads per cell and all cells per batch.

[0565] **II. Performance of other methods**

[0566] 1. Monocle2

[0567] Monocle2 fitted the data into a graph without using prior information of the number of potential fates (Qiu et al., 2017).

[0568] We ran Monocle2 (v2.8.0) with default parameters on a subset of our dataset containing 1,000 cells per time point. Running on our full dataset would require more RAM than we had access to.

[0569] In our data, Monocle2 failed to distinguish iPS, neuronal-like, and trophoblast-like cells as distinct destinations (FIG. 35A-35B). It put together day 18 stromal cells and day 0 MEFs at the root of the tree, and placed iPS, neural-like and trophoblast-like cells on a different branch from cells in the MET Region. Moreover, because the program could incorporate temporal information, it returned a trajectory that was inconsistent with the measured temporal progression. The output of the program implied that day 0 MEF cells gave rise to day 18 stromal cells, which in turn gave rise to everything else.

[0570] 2. URD

[0571] URD identified trajectories from a user-specified root to a set of user-specified tips by performing random walks according to a Markov diffusion kernel.

[0572] We ran URD (vl.O) with default parameters on a subset of our dataset containing 1,000 cells per time point. Running on our full dataset would require more RAM than we had access to.

[0573] In our data, URD predicted that all fates diverge extremely early, with stromal cells diverging from other cells soon after day 0; trophoblast-like cells diverging from neural-like and iPS cells as early as day 1; and neural-like and iPS cells diverging at day 2 (FIGs. 35A-35B). Additionally, URD failed to assign over half (51%) of the cells to any trajectory.

[0574] Comparing the two branches for iPS and neural (FIGs. 35A-35B - segments 6 and 7) revealed no distinctive pattern between the supposedly divergent trajectories from day 3 - 8. The divergent trajectories appeared to be an artifact of the fact that the method requires a distinct branch point.

[0575] Moreover, because the method did not incorporate growth rates, the transitions to iPS and Neural come disproportionately from stromal cells.

[0576] **HI. Pilot study**

[0577] In our pilot study, we collected 65,000 expression profiles over 16 days at 10 distinct time points (and 9 in serum). We compared results from the larger study to the pilot study in FIGs. 30A-30G, where we showed trends in expression along trajectories to each major cell set:

iPSCs, Neural-like, Trophoblast-like (placenta-like in pilot), and Stromal. We found that the expression trends were reasonably similar. Moreover, by comparing the ancestor divergence plots for the two studies, we found that in both studies the stromal population gradually diverged early in the time course and there was a sharp divergence of iPSC from Neural and Trophoblast just after removal of Dox at day 8.

[0578] Data and Software Availability

[0579] We have uploaded our data to NCBI Gene Expression Omnibus. The identification numbers are:

Single cell RNA-seq raw data (pilot study)	GSE106340
Single cell RNA-seq raw data	GSE1 15943

[0580] Our software package is available on GitHub: https://github.com/broadinstitute/wot

[0581] S

[0582] Reference Cited

[0583] 1. C. H. Waddington, How animals develop. (New York, 1936).

[0584] 2. C. H. Waddington, The strategy of the genes; a discussion of some aspects of theoretical biology. (London, Allen & Unwin [1957], 1957).

[0585] 3. E. Z. Macosko et al., Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. Cell 161, 1202-1214 (2015).

[0586] 4. A. M. Klein et al., Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. Cell 161, 1187-1201 (2015).

[0587] 5. G. X. Zheng et al., Massively parallel digital transcriptional profiling of single cells. Nature communications 8, 14049 (2017).

[0588] 6. A. Tanay, A. Regev, Scaling single-cell genomics from phenomenology to mechanism. Nature 541, 331-338 (2017).

[0589] 7. A. Wagner, A. Regev, N. Yosef, Revealing the vectors of cellular identity with single-cell genomics. Nat Biotech 34, 1145-1 160 (2016).

[0590] 8. S. C. Bendall et al., Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157, 714-725 (2014).

[0591] 9. C. Trapnell et al., The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nature biotechnology 32, 381-386 (2014).

[0592] 10. M. Setty et al., Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature biotechnology 34, 637-645 (2016).

[0593] 11. E. Marco et al., Bifurcation analysis of single-cell gene expression data reveals epigenetic landscape. Proceedings of the National Academy of Sciences of the United States of America 111, E5643-5650 (2014).

[0594] 12. J. M. Polo et al., A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617-1632 (2012).

[0595] 13. Y. Buganim et al., Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150, 1209-1222 (2012).

[0596] 14. S. M. Hussein et al., Genome-wide characterization of the routes to pluripotency. Nature 516, 198 (2014).

[0597] 15. P. D. Tonge et al., Divergent reprogramming routes lead to alternative stem-cell states. Nature 516, 192-197 (2014).

[0598] 16. J. O'Malley et al., High resolution analysis with novel cell-surface markers identifies routes to iPS cells. Nature 499, 88 (2013).

[0599] 17. X. Qiu et al., Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv, 110668 (2017).

[0600] 18. S. C. Bendall et al., Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development. Cell 157, 714-725 (2014).

[0601] 19. R. Rostom, V. Svensson, S. Teichmann, G. Kar, Computational approaches for interpreting scRNA-seq data. FEBS letters, (2017).

[0602] 20. L. Haghverdi, F. Buettner, F. J. Theis, Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics 31, 2989-2998 (2015).

[0603] 21. L. Haghverdi, M. Buttner, F. A. Wolf, F. Buettner, F. J. Theis, Diffusion pseudotime robustly reconstructs lineage branching. Nat Meth 13, 845-848 (2016).

[0604] 22. K. Campbell, C. Yau, Ouija: Incorporating prior knowledge in single-cell trajectory learning using Bayesian nonlinear factor analysis. bioRxiv, (2016).

[0605] 23. R. Cannoodt et al., SCORPIUS improves trajectory inference and identifies novel modules in dendritic cell development. bioRxiv, (2016).

[0606] 24. J. D. Welch, A. J. Hartemink, J. F. Prins, SLICER: inferring branched, nonlinear cellular trajectories from single cell RNA-seq data. Genome Biology 17, 106 (2016).

[0607] 25. K. Street et al., Slingshot: Cell lineage and pseudotime inference for single-cell transcriptomics. bioRxiv, (2017).

[0608] 26. H. Matsumoto, H. Kiryu, SCOUP: a probabilistic model based on the Ornstein-Uhlenbeck process to analyze single-cell expression data during differentiation. BMC Bioinformatics 17, 232 (2016).

[0609] 27. S. Rashid, D. N. Kotton, Z. Bar-Joseph, TASIC: determining branching models from time series single cell data. Bioinformatics 33, 2504-2512 (2017).

[0610] 28. M. Zwiessele, N. D. Lawrence, Topslam: Waddington Landscape Recovery for Single Cell Experiments. bioRxiv, (2016).

[0611] 29. C. Weinreb, S. Wolock, B. K. Tusi, M. Socolovsky, A. M. Klein, Fundamental limits on dynamic inference from single cell snapshots. bioRxiv, (2017).

[0612] 30. C. Villani, Optimal transport: old and new. (Springer Science & Business Media, 2008), vol. 338.

[0613] 31. M. Cuturi, in Advances in neural information processing systems. (2013), pp. 2292-2300.

[0614] 32. L. Chizat, G. Peyre, B. Schmitzer, F.-X. Vialard, Scaling algorithms for unbalanced transport problems. arXiv preprint arXiv: 1607.05816, (2016).

[0615] 33. J. H. Levine et al., Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. Cell 162, 184-197 (2015).

[0616] 34. K. Shekhar et al., Comprehensive Classification of Retinal Bipolar Neurons by Single-Cell Transcriptomics. Cell 166, 1308-1323.el330 (2016).

[0617] 35. R. R. Coifman et al., Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proceedings of the National Academy of Sciences of the United States of America 102, 7426-7431 (2005).

[0618] 36. M. Jacomy, T. Venturini, S. Heymann, M. Bastian, ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. PloS one 9, e98679 (2014).

[0619] 37. E. R. Zunder, E. Lujan, Y. Goltsev, M. Wernig, G. P. Nolan, A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell 16, 323-337 (2015).

[0620] 38. C. Weinreb, S. Wolock, A. Klein, SPRING: a kinetic interface for visualizing high dimensional single-cell expression data. bioRxiv, (2016).

[0621] 39. K. Takahashi, S. Yamanaka, Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. cell 126, 663-676 (2006).

[0622] 40. J. Yu et al., Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917-1920 (2007).

[0623] 41. J. Shu et al., Induction of pluripotency in mouse somatic cells with lineage specifiers. Cell 153, 963-975 (2013).

[0624] 42. P. Hou et al., Pluripotent Stem Cells Induced from Mouse Somatic Cells by Small-Molecule Compounds. Science 341, 651-654 (2013).

[0625] 43. D. H. Kim et al., Single-cell transcriptome analysis reveals dynamic changes in IncRNA expression during reprogramming. Cell stem cell 16, 88-101 (2015).

[0626] 44. A. Parenti, M. A. Halbisen, K. Wang, K. Latham, A. Ralston, OSKM induce extraembryonic endoderm stem cells in parallel to induced pluripotent stem cells. Stem cell reports 6, 447-455 (2016).

[0627] 45. T. S. Mikkelsen et al., Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49 (2008).

[0628] 46. M. Stadtfeld, N. Maherali, M. Borkent, K. Hochedlinger, A reprogrammable mouse strain from gene-targeted embryonic stem cells. Nature methods 7, 53-55 (2010).

[0629] 47. Z. D. Smith, I. Nachman, A. Regev, A. Meissner, Dynamic single-cell imaging of direct reprogramming reveals an early specifying event. Nat Biotechnol 28, 521-526 (2010).

[0630] 48. J. Pei, N. V. Grishin, Unexpected diversity in Shisa-like proteins suggests the importance of their roles as transmembrane adaptors. Cellular signalling 24, 758-769 (2012).

[0631] 49. M. Meyyappan, H. Wong, C. Hull, K. T. Riabowol, Increased expression of cyclin D2 during multiple states of growth arrest in primary and established cells. Molecular and cellular biology 18, 3163-3172 (1998).

[0632] 50. J.-P. Coppe, P.-Y. Desprez, A. Krtolica, J. Campisi, The senescence-associated secretory phenotype: the dark side of tumor suppression. Annual Review of Pathological Mechanical Disease 5, 99-1 18 (2010).

[0633] 51. L. Mosteiro et al., Tissue damage and senescence provide critical signals for cellular reprogramming in vivo. Science 354, aaf4445 (2016).

[0634] 52. Q.-L. Ying et al., The ground state of embryonic stem cell self-renewal. Nature 453, 519 (2008).

[0635] 53. 1. Tirosh et al., Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature 539, 309-313 (2016).

[0636] 54. S. C. Andrews et al., Cdknlc (p57 Kip2) is the major regulator of embryonic growth within its imprinted domain on mouse distal chromosome 7. BMC Developmental Biology 7, 53 (2007).

[0637] 55. N. Barker et al., Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449, 1003-1007 (2007).

[0638] 56. G. C. Elson et al., CLF associates with CLC to form a functional heteromeric ligand for the CNTF receptor complex. Nature neuroscience 3, 867 (2000).

[0639] 57. A. Fowden, C. Sibley, W. Reik, M. Constancia, Imprinted genes, placental development and fetal growth. Hormone Research in Paediatrics 65, 50-58 (2006).

[0640] 58. A. Ralston et al., Gata3 regulates trophoblast development downstream of Tead4 and in parallel to Cdx2. Development 137, 395-403 (2010).

[0641] 59. G. Burton, H.-W. Yung, T. Cindrova-Davies, D. Charnock-Jones, Placental endoplasmic reticulum stress and oxidative stress in the pathophysiology of unexplained intrauterine growth restriction and early onset preeclampsia. Placenta 30, 43-48 (2009).

[0642] 60. V. Pasque et al., X chromosome reactivation dynamics reveal stages of reprogramming to pluripotency. Cell 159, 1681-1697 (2014).

[0643] 61. K. Tomoda et al., Derivation conditions impact X-inactivation status in female human induced pluripotent stem cells. Cell stem cell 11, 91-99 (2012).

[0644] 62. Q. Bai et al., Dissecting the first transcriptional divergence during human embryonic development. Stem Cell Reviews and Reports 8, 150-162 (2012).

[0645] 63. A.-H. Monsoro-Burq, E. Wang, R. Harland, Msxl and Pax3 cooperate to mediate FGF8 and WNT signals during Xenopus neural crest induction. Developmental cell 8, 167-178 (2005).

[0646] 64. L. Pevny, M. Placzek, SOX genes and neural progenitor identity. Current opinion in neurobiology 15, 7-13 (2005).

[0647] 65. V. Y. Wang, H. Y. Zoghbi, Genetic regulation of cerebellar development. Nature reviews. Neuroscience 2, 484 (2001).

[0648] 66. Y. Liu, A. W. Helms, J. E. Johnson, Distinct activities of Msxl and Msx3 in dorsal neural tube development. Development 131, 1017-1028 (2004).

[0649] 67. M. Bergsland et al., Sequentially acting Sox transcription factors in neural lineage development. Genes Dev 25, 2453-2464 (201 1).

[0650] 68. K. Achim et al., The role of Tal2 and Tall in the differentiation of midbrain GABAergic neuron precursors. Biology open 2, 990-997 (2013).

[0651] 69. A. Domanskyi, H. Alter, M. A. Vogt, P. Gass, I. A. Vinnikov, Transcription factors Foxal and Foxa2 are required for adult dopamine neurons maintenance. Frontiers in cellular neuroscience 8, 275 (2014).

[0652] 70. K. Takebayashi-Suzuki, A. Kitayama, C. Terasaka-Iioka, N. Ueno, A. Suzuki, The forkhead transcription factor FoxBl regulates the dorsal-ventral and anterior-posterior patterning of the ectoderm during early Xenopus embryogenesis. Developmental biology 360, 11-29 (201 1).

[0653] 71. G. Hu et al., A genome-wide RNAi screen identifies a new transcriptional module required for self-renewal. Genes & development 23, 837-848 (2009).

[0654] 72. W.-Z. Li et al., Hesxl enhances pluripotency by working downstream of multiple pluripotency-associated signaling pathways. Biochemical and Biophysical Research Communications 464, 936-942 (2015).

[0655] 73. W. Shi et al., Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. J Biol Chem 281, 23319-23325 (2006).

[0656] 74. A. Rajkovic, C. Yan, W. Yan, M. Klysik, M. M. Matzuk, Obox, a Family of Homeobox Genes Preferentially Expressed in Germ Cells. Genomics 79, 711-717 (2002).

[0657] [SI) Villani C. Optimal Transport Old and New. Springer; 2008.

[0658] [S2] Chizat L, Peyre' G, Schmitzer B, Vialard FX. Scaling Algorithms for Unbalanced Transport Problems. Mathematics of Computation. 2017;.

[0659] [S3] Cuturi M. Sinkhorn Distances: Lightspeed Computation of Optimal Transportation Distances. In: Neural Information Processing Systems (NIPS); 2013. .

[0660] [S4] https://support. 10xgenomics.com/single-cell-gene-expression/ software/pipelines/latest/installation.

[0661] [S5] Coifman RR, Lafon S, Lee AB, Maggioni M, Nadler B, Warner F, et al. Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps. Proc Natl Acad Sci U S A. 2005;102:7426-7431.

[0662] [S6] Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics. 2015;31:2989-2998.

[0663] [S7] Haghverdi L, Buettner M, Wolf FA, Buettner F, Theis FJ. Diffusion pseudotyme robustly recon- structs lineage branching. bioRxiv. 2016;p. 041384.

[0664] **[S8]** Angerer P, Haghverdi L, Bu¨ttner M, Theis FJ, Marr C, Buettner F. destiny: diffusion maps for large-scale single-cell data in R. Bioinformatics. 2015;32:1241-1243.

[0665] [S9] Moignard V, Woodhouse S, Haghverdi L, Lilly AJ, Tanaka Y, Wilkinson AC, et al. Decoding the regulatory network of early blood development from single-cell gene expression measurements. Nature Biotechn. 2015;33:269-276.

[0666] [S10] SettyM,TadmorMD,Reich-ZeligerS, AngelO, SalameTM, KathailP, et al. Wishbone identifies bifurcating developmental trajectories from single-cell data. Nature Biotechn. 2016;34:637-645.

[0667] [SI 1] Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nature Biotechn. 2015;33:495-502.

[0668] [S12] HeinzS,BennerC,SpannN,BertolinoE,LinYC,LasloP,etal.Simplecombinationso flineage- determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol cell. 2010;38:576-589.

[0669] [S13] Bastian M, Heymann S, Jacomy M, et al. Gephi: an open source software for exploring and manipulating networks. Icwsm. 2009;8:361-362.

[0670] [S14] Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algo- rithm for handy network visualization designed for the Gephi software. PloS one. 2014;9:e98679.

[0671] [S15] Beygelzimer A, Kakadet S, Langford J, Arya S, Mount D, Li S, et al.. Package FNN;.

[0672] [SI 6] Zunder ER, Lujan E, Goltsev Y, Wernig M, Nolan GP. A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry. Cell Stem Cell. 2015;16:323-337.

[0673] S17 Porpiglia E, Samusik N, Van Ho AT, Cosgrove BD, Mai T, Davis KL, et al. High-resolution myogenic lineage mapping by single-cell mass cytometry. Nature Cell Biol. 2017;19:558-567.

[0674] S18 Samusik N, Good Z, Spitzer MH, Davis KL, Nolan GP. Automated mapping of phenotype space with single-cell data. Nature methods. 2016;13:493-496.

[0675] S19 Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theor Exp. 2008;2008:P10008.

[0676] S20 Levine JH, Simonds EF, Bendall SC, Davis KL, El-ad DA, Tadmor MD, et al. Data-driven phenotypic dissection of AML reveals progenitor-like cells that correlate with prognosis. Cell. 2015;162:184-197.

[0677] S21 Shekhar K, Lapan SW, Whitney IE, Tran NM, Macosko EZ, Kowalczyk M, et al. Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics. Cell. 2016;166:1308- 1323.

[0678] S22 Csardi G, Nepusz T. The igraph software package for complex network research. InterJournal, Complex Systems. 2006;1695:1-9.

[0679] S23 Friedman J, Hastie T, Tibshirani R. Sparse inverse covariance estimation with the graphical lasso. Biostatistics. 2008;9:432-441.

[0680] S24 Rosvall M, Bergstrom CT. Maps of random walks on complex networks reveal community struc- ture. Proc Natl Acad Sci U S A. 2008;105:1 118-1 123.

[0681] S25 Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner H, et al. Reversed graph embedding resolves complex single-cell developmental trajectories. bioRxiv. 2017;p. 110668.

[0682] S26 Qiu X, Hill A, Packer J, Lin D, Ma YA, Trapnell C. Single-cell mRNA quantification and differ- ential analysis with Census. Nature methods. 2017;14:309-315.

[0683] S27 Mao Q, Wang L, Goodison S, Sun Y. Dimensionality reduction via graph structure learning. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. ACM; 2015. p. 765-774.

[0684] S28 Rashid S, Kotton DN, Bar-Joseph Z. TASIC: determining branching models from time series single cell data. Bioinformatics. 2017;p. btxl73.

[0685] S29 Lattin JE, Schroder K, Su AI, Walker JR, Zhang J, Wiltshire T, et al. Expression analysis of G Protein-Coupled Receptors in mouse macrophages. Immunome Res. 2008;4:5.

[0686] S30 Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.

[0687] S31 Tirosh I, Venteicher AS, Hebert C, Escalante LE, Patel AP, Yizhak K, et al. Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. Nature. 2016;539:309-313.

[0688] S32 Li R, Liang J, Ni S, Zhou T, Qing X, Li H, et al. A mesenchymal-to-epithelial transition initiates and is required for the nuclear reprogramming of mouse fibroblasts. Cell stem cell. 2010;7:51-63.]

[0689] S33 Whiteman EL, Fan S, Harder JL, Walton KD, Liu CJ, Soofi A, et al. Crumbs3 is essential for proper epithelial development and viability. Mol Cell Biol. 2014;34:43-56.

[0690] S34 Takaishi M, Tarutani M, Takeda J, Sano S. Mesenchymal to Epithelial Transition Induced by Re- programming Factors Attenuates the Malignancy of Cancer Cells. PloS one. 2016;l l:e0156904.

[0691] S35 Hewitt KJ, Agarwal R, Morin PJ. The claudin gene family: expression in normal and neoplastic tissues. BMC cancer. 2006;6:186.

[0692] S36 Coppe´ JP, Desprez PY, Krtolica A, Campisi J. The senescence-associated secretory phenotype: the dark side of tumor suppression. Annu Rev Pathol. 2010;5:99-1 18.

[0693] S37 da Fonseca ET, Mane ^nares ACF, Ambro ´sio CE, Miglino MA. Review point on neural stem cells and neurogenic areas of the central nervous system. Open J Anim Sci. 2013;3:242.

[0694] S38 Sakakibara Si, Nakamura Y, Satoh H, Okano H. Rna-binding protein Musashi2: developmentally regulated expression in neural precursor cells and subpopulations of neurons in mammalian CNS. J Neurosci. 2001;21:8091-8107.

[0695] S39 Gouti M, Briscoe J, Gavalas A. Anterior Hox genes interact with components of the neural crest specification network to induce neural crest fates. Stem cells. 201 1;29:858-870.

[0696] S40 Watanabe Y, Stanchina L, Lecerf L, Gacem N, Conidi A, Baral V, et al. Differentiation of Mouse Enteric Nervous System Progenitor Cells Is Controlled by Endothelin 3 and Requires Regulation of Ednrb by SOX10 and ZEB2. Gastroenterology. 2017; 152: 1139— 1150.

[0697] S41 Sansom SN, Griffiths DS, Faedo A, Kleinjan DJ, Ruan Y, Smith J, et al. The level of the tran- scription factor Pax6 is essential for controlling the balance between neural stem cell self-renewal and neurogenesis. PLoS Genetics. 2009;5:el00051 1.

[0698] S42 SKan L, Israsena N, Zhang Z, Hu M, Zhao LR, Jalali A, et al. Soxl acts through multiple inde- pendent pathways to promote neurogenesis. Dev Biol. 2004;269:580-594.

[0699] S43 Lazarov O, Mattson MP, Peterson DA, Pimplikar SW, van Praag H. When neurogenesis encoun- ters aging and disease. Trends Neurosci. 2010;33:569-579.

[0700] S44 Tibshirani R, Walther G, Hastie T. Estimating the number of clusters in a data set via the gap statistic. J R Stat Soc Series B Stat Methodol. 2001;63:41 1-423.

[0701] S45 Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell. 2012;151(7): 1617—1632.

[0702] S46 Mertins P, Przybylski D, Yosef N, Qiao J, Clauser K, Raychowdhury R, et al. An Integrative Framework Reveals Signaling-to-Transcription Events in Toll-like Receptor Signaling. Cell re- ports. 2017;19(13):2853-2866.

[0703] S47 ChoiJ, HuebnerAJ, ClementK, WalshRM, SavolA, LinK, etal. Prolonged Mekl/2suppression impairs the developmental potential of embryonic stem cells. Nature. 2017;548:219-223.

[0704] S48 Parenti A, Halbisen MA, Wang K, Latham K, Ralston A. OSKM induce extraembryonic endo- derm stem cells in parallel to induced pluripotent stem cells. Stem cell reports. 2016;6(4):447- 455.

[0705] [S49] Lin J, Khan M, Zapiec B, Mombaerts P. Efficient derivation of extraembryonic endoderm stem cell lines from mouse postimplantation embryos. Scientific reports. 2016;6.

[0706] [S50] Edgar R, Mazor Y, Rinon A, Blumenthal J, Golan Y, Buzhor E, et al. LifeMap Discovery?: the embryonic development, stem cells, and regenerative medicine research portal. PloS one. 2013;8(7):e66629.

***

[0707] Various modifications and variations of the described methods, pharmaceutical compositions, and kits of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific embodiments, it will be understood that it is capable of further modifications and that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention that are obvious to those skilled in the art are intended to be within the scope of the invention. This application is intended to cover any variations, uses, or adaptations of the invention following, in general, the principles of the invention and including such departures from the present disclosure come within known customary practice within the art to which the invention pertains and may be applied to the essential features herein before set forth.

CLAIMS

What is claimed is:

1. A method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Obox6 into a target cell to produce an induced pluripotent stem cell.

2. The method of claim 1, further comprising introducing into the target cell at least one nucleic acid encoding a reprogramming factor selected from the group consisting of: Gdf9, Oct3/4, Sox2, Soxl, Sox3, Soxl5, Soxl7, Klf4, Klf2, c-Myc, N-Myc, L-Myc, Nanog, Lin28, Fbxl5, ERas, ECAT15-2, Tell, beta-catenin, Lin28b, Sall l, Sal 14, Esrrb, Nr5a2, Tbx3, and Glisl.

3. The method of claim 1, further comprising introducing into the target cell at least one nucleic acid encoding a reprogramming factor selected from the group consisting of: Oct4, Klf4, Sox2 and Myc.

4. The method of claim 1, wherein the nucleic acid encoding Obox6 is provided in a recombinant vector.

5. The method of claim 4, wherein the vector is a lentivirus vector.

6. The method of claim 2, where the nucleic acid encoding the reprogramming factor is provided in a recombinant vector.

7. The method of claim 1, further comprising a step of culturing the cells in reprogramming medium.

8. The method of claim 1, further comprising a step of culturing the cells in the presence of serum.

9. The method of claim 1, further comprising a step of culturing the cells in the absence of serum.

10. The method of claim 1, wherein the induced pluripotent stem cell expresses at least one of a surface marker selected from the group consisting of: Oct4, SOX2, KLf4, c-MYC, LIN28, Nanog, Glisl , TRA-160/TRA-1-81/TRA-2-54, SSEA1, SSEA4, Sal4, and Esrbbl.

11. The method of claim 1, wherein the target cell is a mammalian cell.

12. The method of claim 1, wherein the target cell is a human cell or a murine cell.

13. The method of claim 1, wherein the target cell is a mouse embryonic fibroblast.

14. The method of claim 1, wherein the target cell is selected from the group consisting of: fibroblasts, B cells, T cells, dendritic cells, keratinocytes, adipose cells, epithelial cells, epidermal cells, chondrocytes, cumulus cells, neural cells, glial cells, astrocytes, cardiac cells, esophageal cells, muscle cells, melanocytes, hematopoietic cells, pancreatic cells, hepatocytes, macrophages, monocytes, mononuclear cells, and gastric cells, including gastric epithelial cells.

15. A method of producing an induced pluripotent stem cell comprising introducing at least one of Obox6, Spic, Zfp42, Sox2, Mybl2, Msc, Nanog, Hesxl and Esrrb into a target cell to produce an induced pluripotent stem cell.

16. A method of producing an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell to produce an induced pluripotent stem cell.

17. A method of increasing the efficiency of production of an induced pluripotent stem cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell.

18. A method of increasing the efficiency of production of an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell to produce an induced pluripotent stem cell.

19. An isolated induced pluripotential stem cell produced by the method of claim 1, 15, or 16.

20. A method of treating a subject with a disease comprising administering to the subject a cell produced by differentiation of the induced pluripotent stem cell produced by the method of claim 1, 15, or 16.

21. A composition for producing an induced pluripotent stem cell comprising Obox6 in combination with reprogramming medium.

22. A composition for producing an induced pluripotent stem cell comprising one or more of the factors identified in or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6 in combination with reprogramming medium.

23. Use of Obox6 for production of an induced pluripotent stem cell.

24. Use of a factor identified in or one or more of the factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6 for production of an induced pluripotent stem cell.

25. A method of increasing the efficiency of reprogramming a cell comprising introducing Obox6 into a target cell to produce an induced pluripotent stem cell.

26. A method of increasing the efficiency of reprogramming a cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5 and Table 6, into a target cell to produce an induced pluripotent stem cell.

27. A computer-implemented method for mapping developmental trajectories of cells, comprising:

generating, using one or more computing devices, optimal transport maps for a set of cells from single cell sequencing data obtained over a defined time course;

determining, using one or more computing devices, cell regulatory models, and optionally identifying local biomarker enrichment, based on at least the generated optimal transport maps;

defining, using the one or more computing devices, gene modules; and

generating, using the one or more computing devices, a visualization of a developmental landscape of the set of cells.

28. The method of claim 27, wherein determining cell regulatory models comprise sampling pairs of cells at a first time and a second time point according to transport probabilities.

29. The method of claim 28, further comprising using the expression levels of transcription factors at the earlier time point to predict non-transcription factor expression at the second time point.

30. The method of claim 27, wherein identifying local biomarker enrichment comprises identifying transcription factors enriched in cells having a defined percentage of descendants in a target cell population.

31. The method of claim 30, wherein the defined percentage is at least 50% of mass.

32. The method of claim 27, wherein defining gene modules comprises partitioning genes based on correlated gene expression across cells and clusters.

33. The method of claim 32, wherein partitioning comprises partitioning cells based on graph clustering.

34. The method of claim 33, wherein graph clustering further comprises dimensionality reduction using diffusion maps.

35. The method of claim 27, wherein the visualization of the developmental landscape comprises high-dimensional gene expression data in two dimensions.

36. The method of claim 33, wherein the visualization is generated using force-directed layout embedding (FLE).

37. The method of claim 27, wherein the visualization provides one or more cell types, cell ancestors, cell descendants, cell trajectories, gene modules, and cell clusters from the single cell sequencing data.

38. A computer program product, comprising:

a non-transitory computer-executable storage device having computer-readable program instructions embodied thereon that when executed by a computer cause the computer to execute the methods of anyone of claims 27 to 37.

39. A system comprising:

a storage device; and

a processor communicatively coupled to the storage device, wherein the processor executes application code instructions that are stored in the storage device and that cause the system to executed the methods of any one of claims 27 to 37.

40. A method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Gdf9 into a target cell to produce an induced pluripotent stem cell.

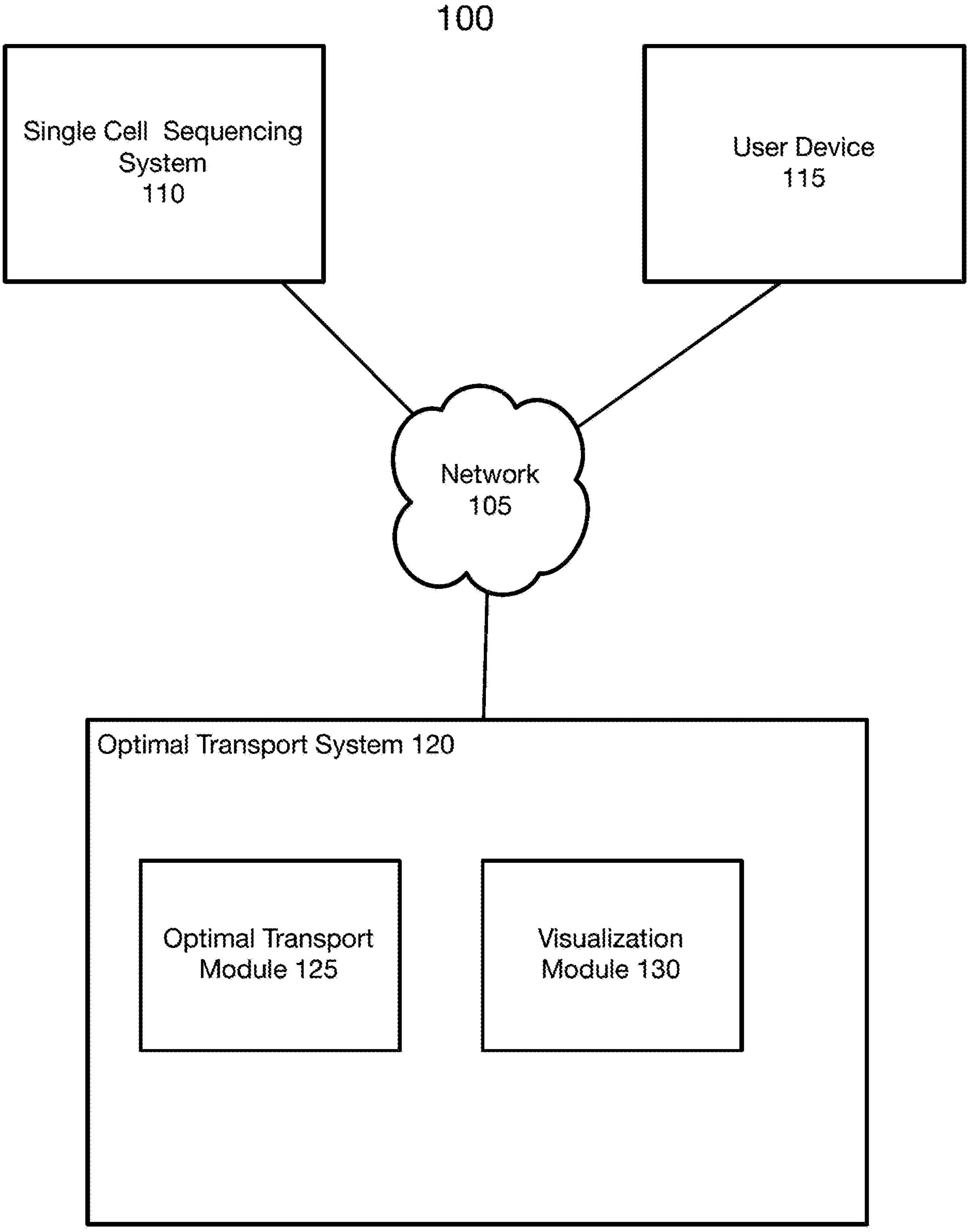

100
Single Cell Sequencing System 110
User Device 115
Network 105
Optimal Transport System 120
Optimal Transport Module 125
Visualization Module 130

FIG. 1

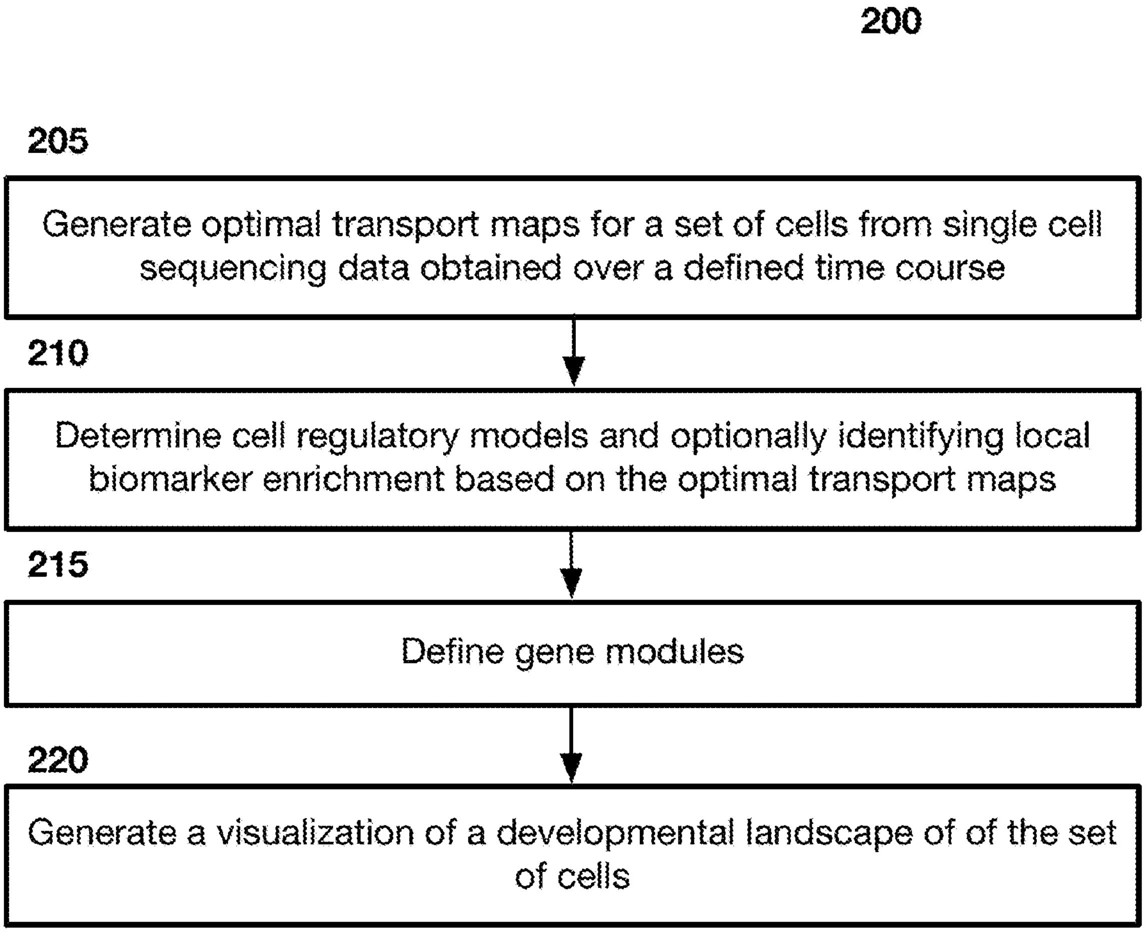

200
205
Generate optimal transport maps for a set of cells from single cell sequencing data obtained over a defined time course
210
Determine cell regulatory models and optionally identifying local biomarker enrichment based on the optimal transport maps
215
Define gene modules
220
Generate a visualization of a developmental landscape of of the set of cells

FIG. 2

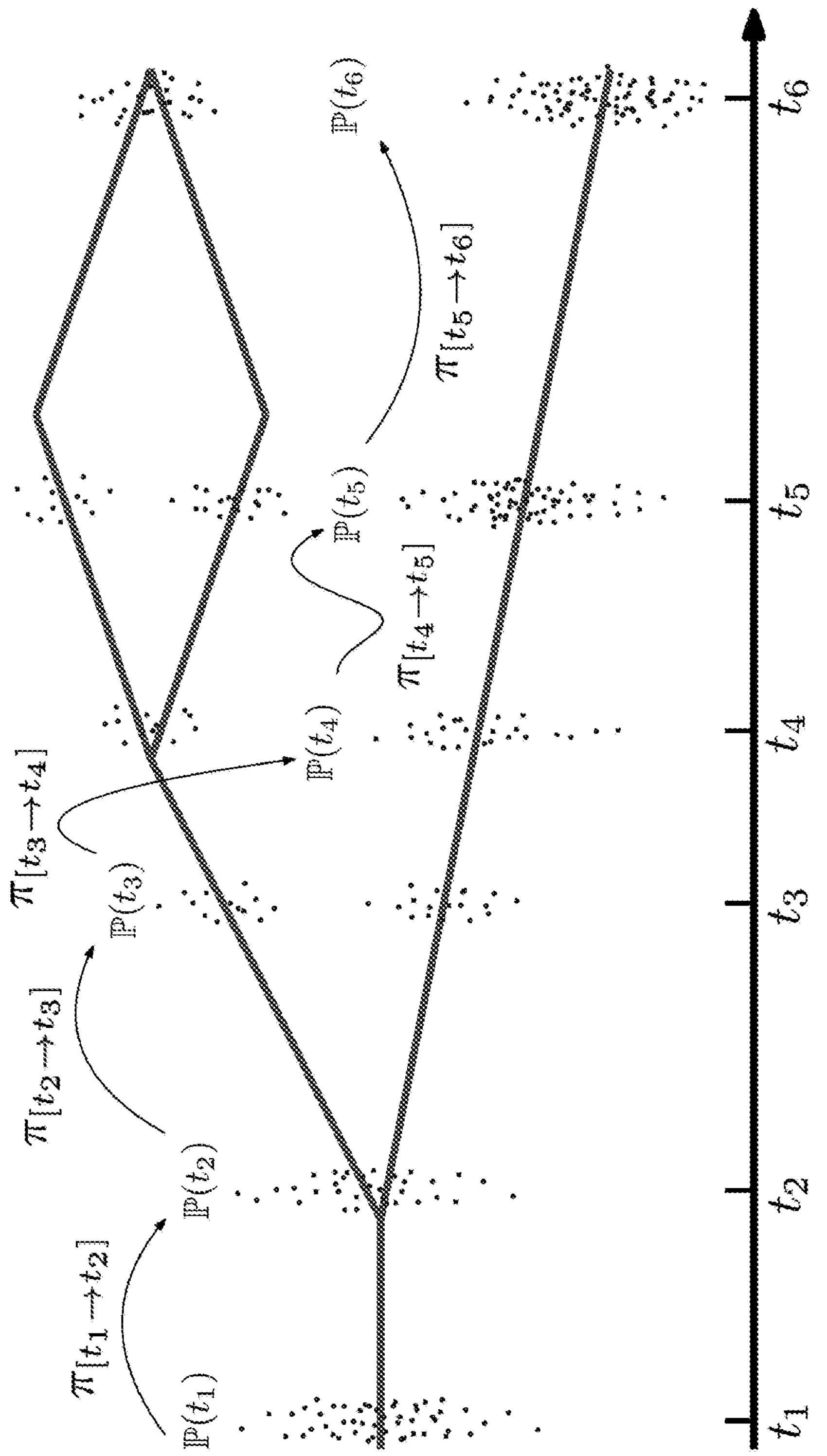

$\mathbb{P}(t_1)$
$\pi_{[t_1 \to t_2]}$
$\mathbb{P}(t_2)$
$\pi_{[t_2 \to t_3]}$
$\mathbb{P}(t_3)$
$\pi_{[t_3 \to t_4]}$
$\mathbb{P}(t_4)$
$\pi_{[t_4 \to t_5]}$
$\mathbb{P}(t_5)$
$\pi_{[t_5 \to t_6]}$
$\mathbb{P}(t_6)$
t_1
t_2
t_3
t_4
t_5
t_6

FIG. 3

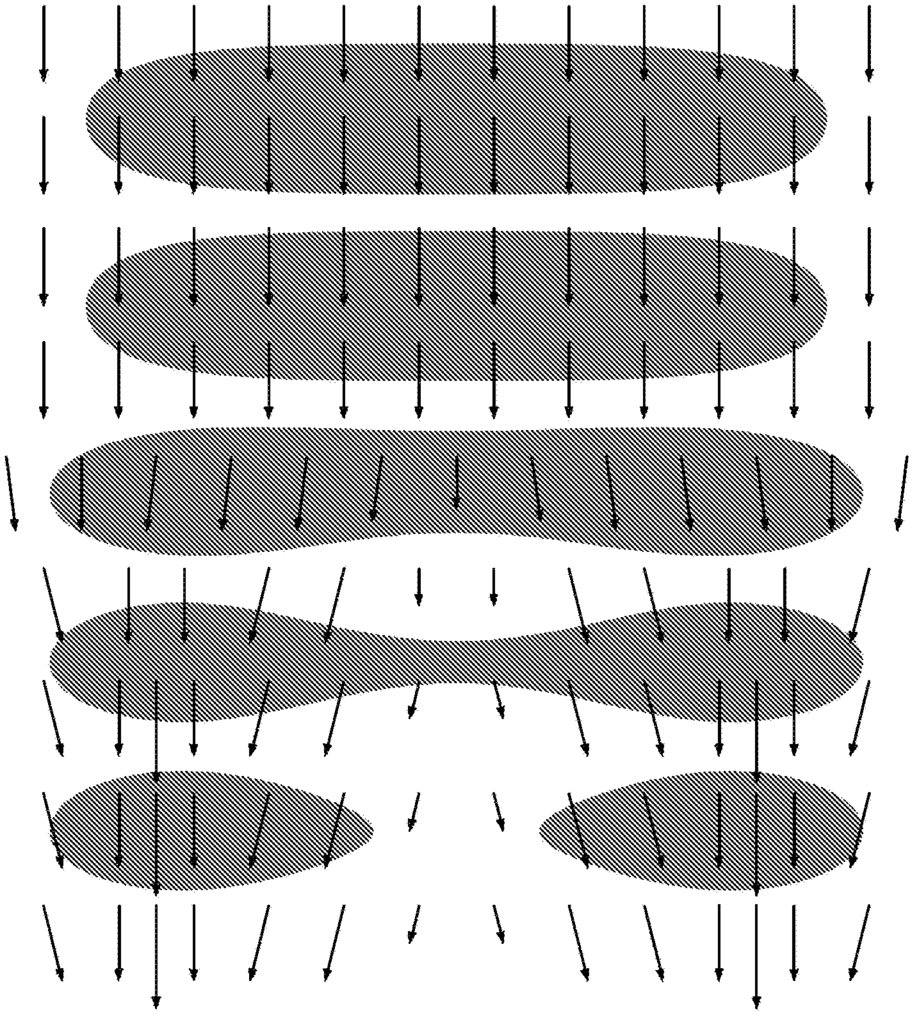


FIG. 4

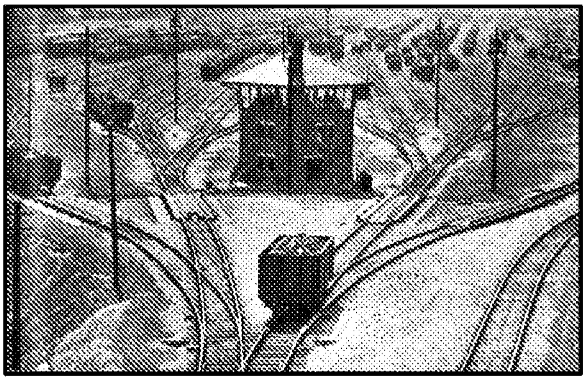


FIG. 5A

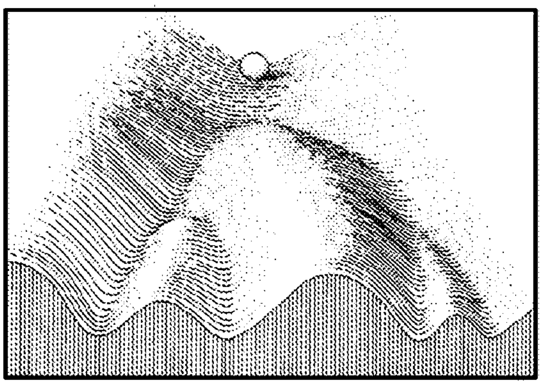


FIG. 5B

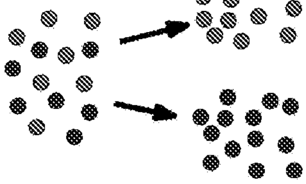


FIG. 5C

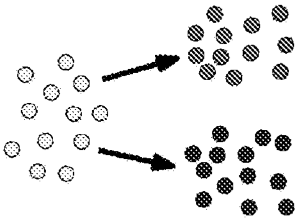


FIG. 5D

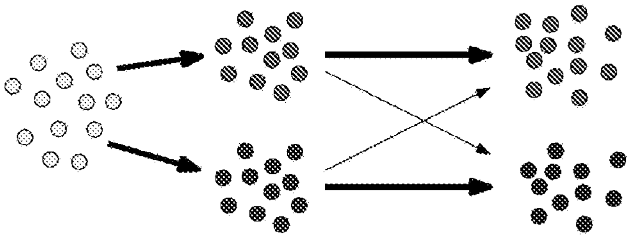


FIG. 5E

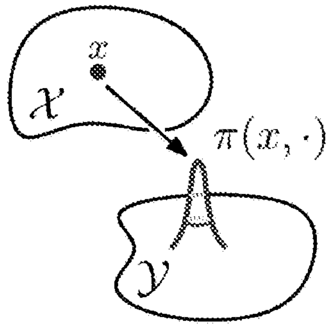


FIG. 5F

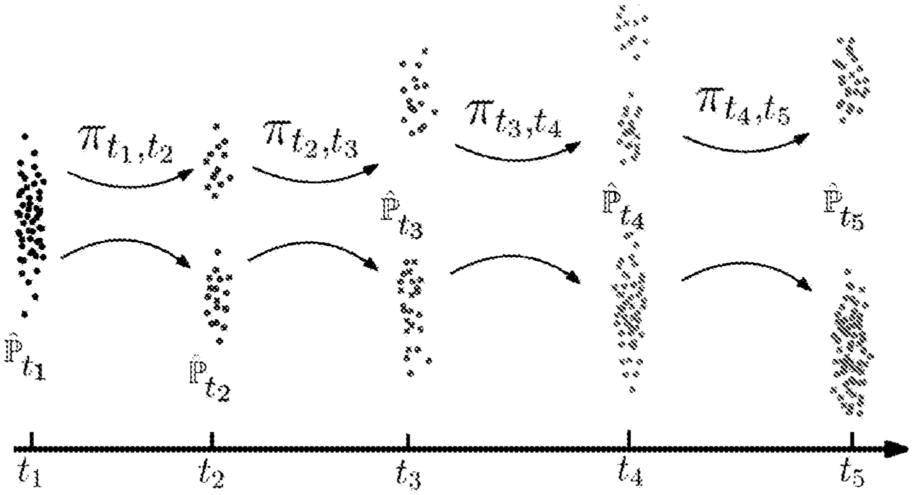


FIG. 5G

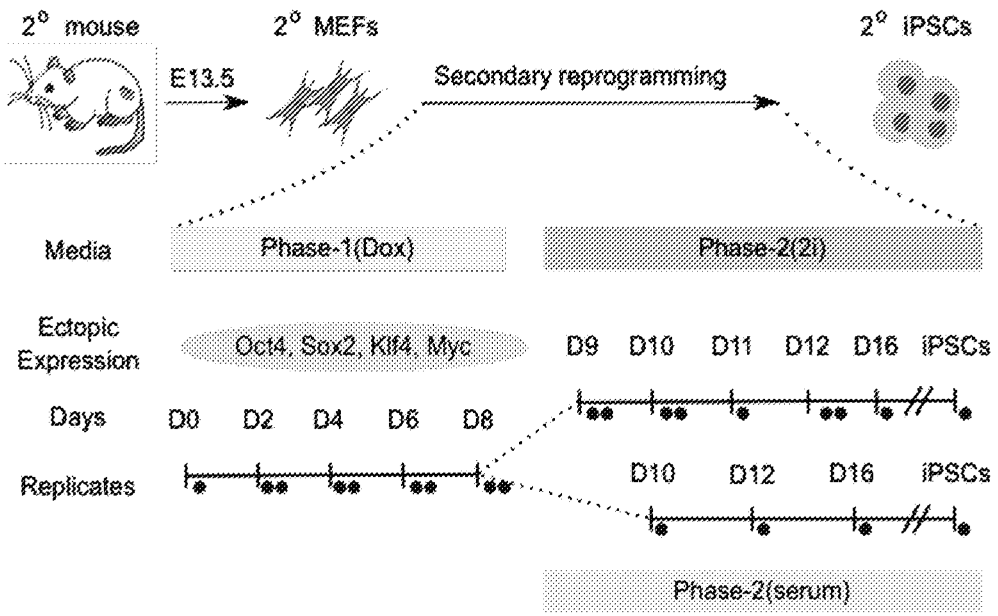


FIG. 6A

		Sample	Cell Number
Phase-1(Dox)		D0	4,060
		D2-1	2,890
		D2-2	2,729
		D4-1	2,882
		D4-2	3,962
		D6-1	3,198
		D6-2	3,168
		D8-1	2,142
		D8-2	2,625
Phase-2	(2i)	D9-1	2,441
		D9-2	2,174
		D10-1	2,878
		D10-2	2,619
		D11	1,529
		D12-1	5,139
		D12-2	2,155
		D16	4,500
		iPSCs	2,916
	(serum)	D10	2,088
		D12	2,895
		D16	3,703
		iPSCs	3,088
		Total	**65,781**

FIG. 6B

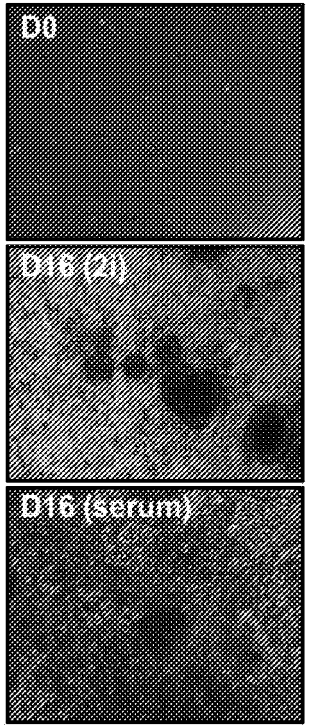


FIG. 6C

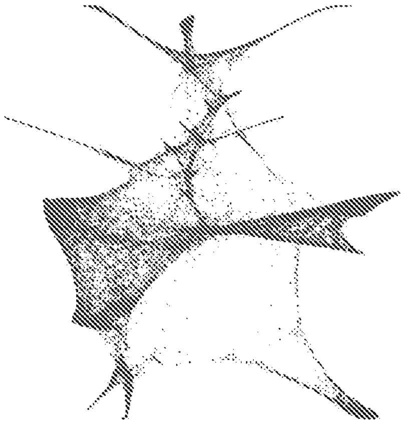


FIG. 7A

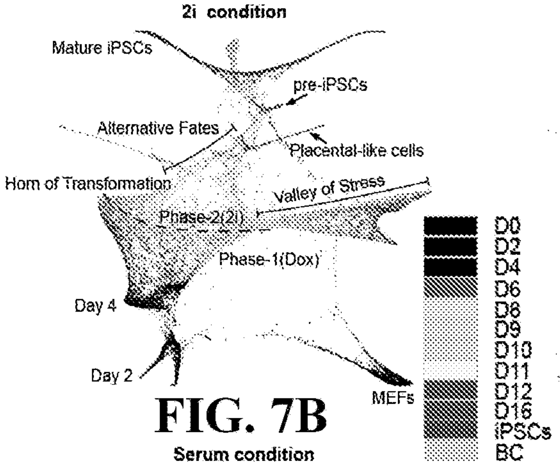


FIG. 7B

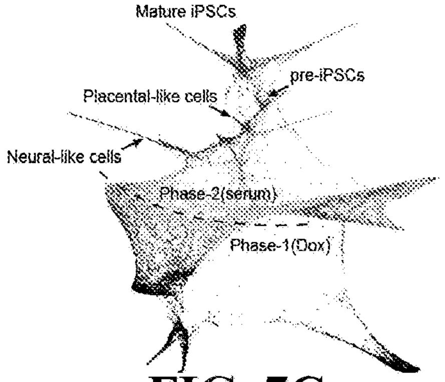


FIG. 7C

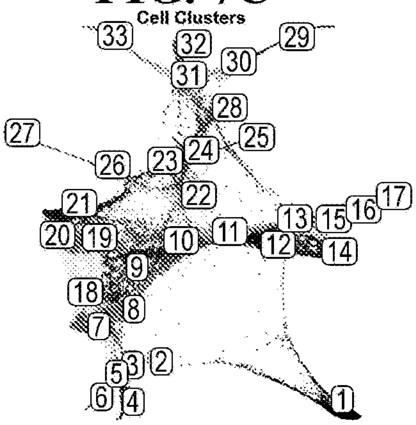


FIG. 7D

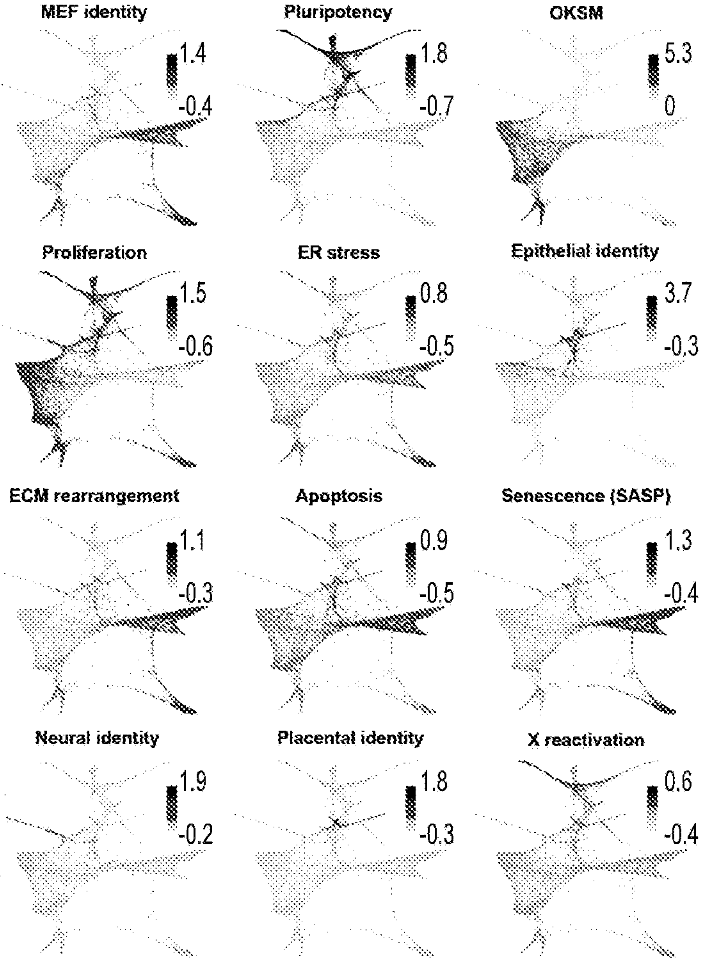


FIG. 7E

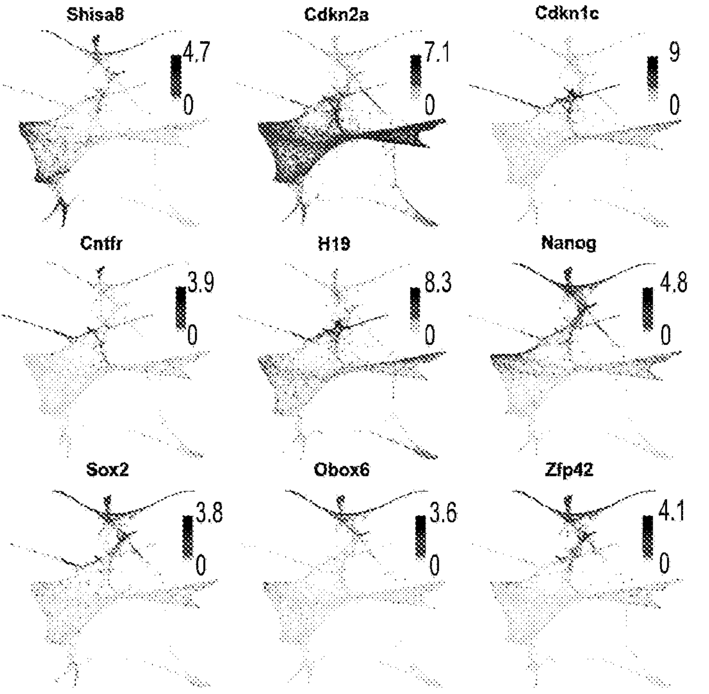


FIG. 7F

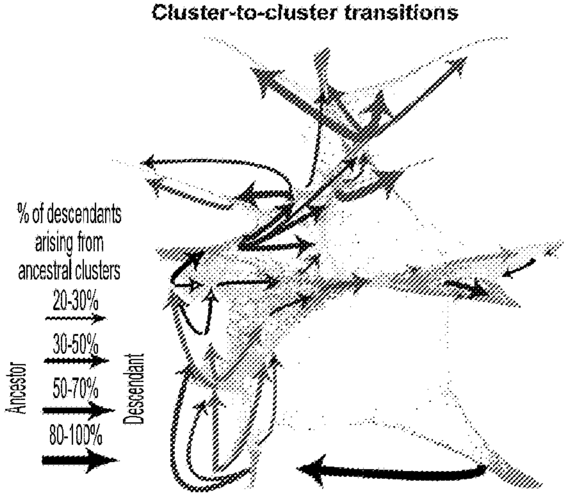

Cluster-to-cluster transitions
% of descendants arising from ancestral clusters
20-30%
30-50%
50-70%
80-100%
Ancestor
Descendant

FIG. 8A

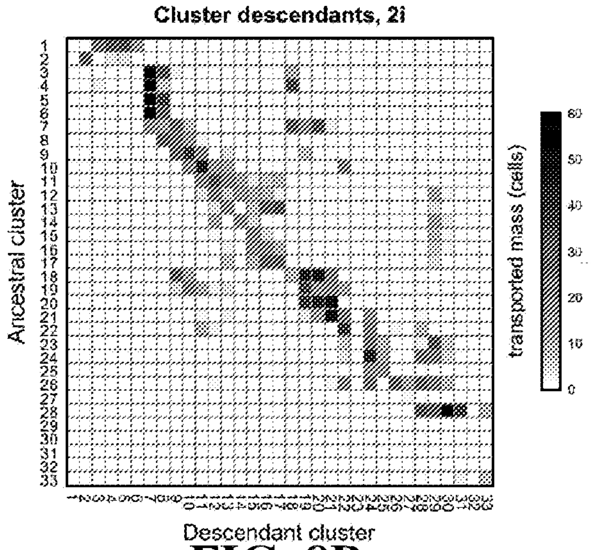


FIG. 8B

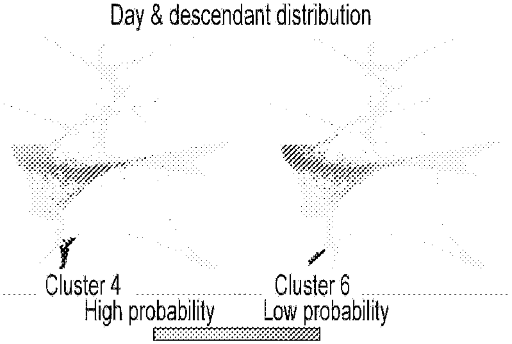


FIG. 8C

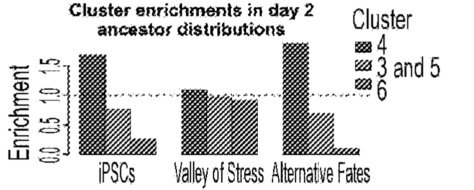


FIG. 8D

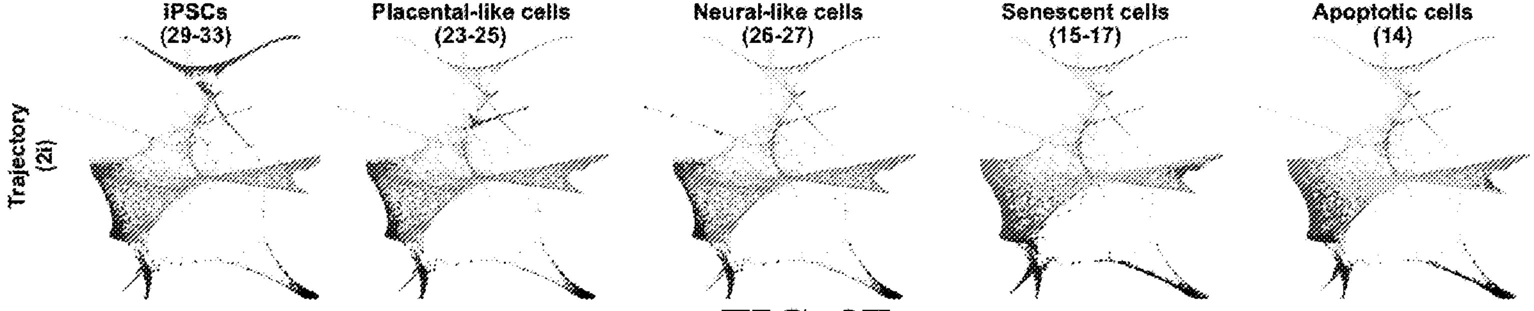


FIG. 8E

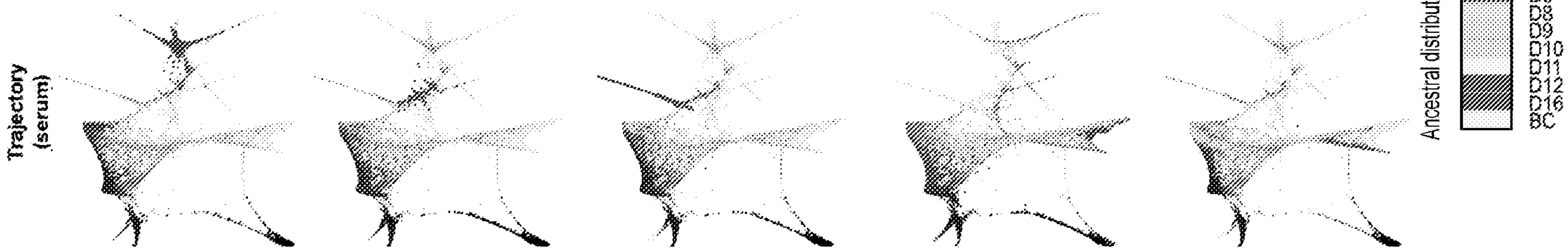


FIG. 8F

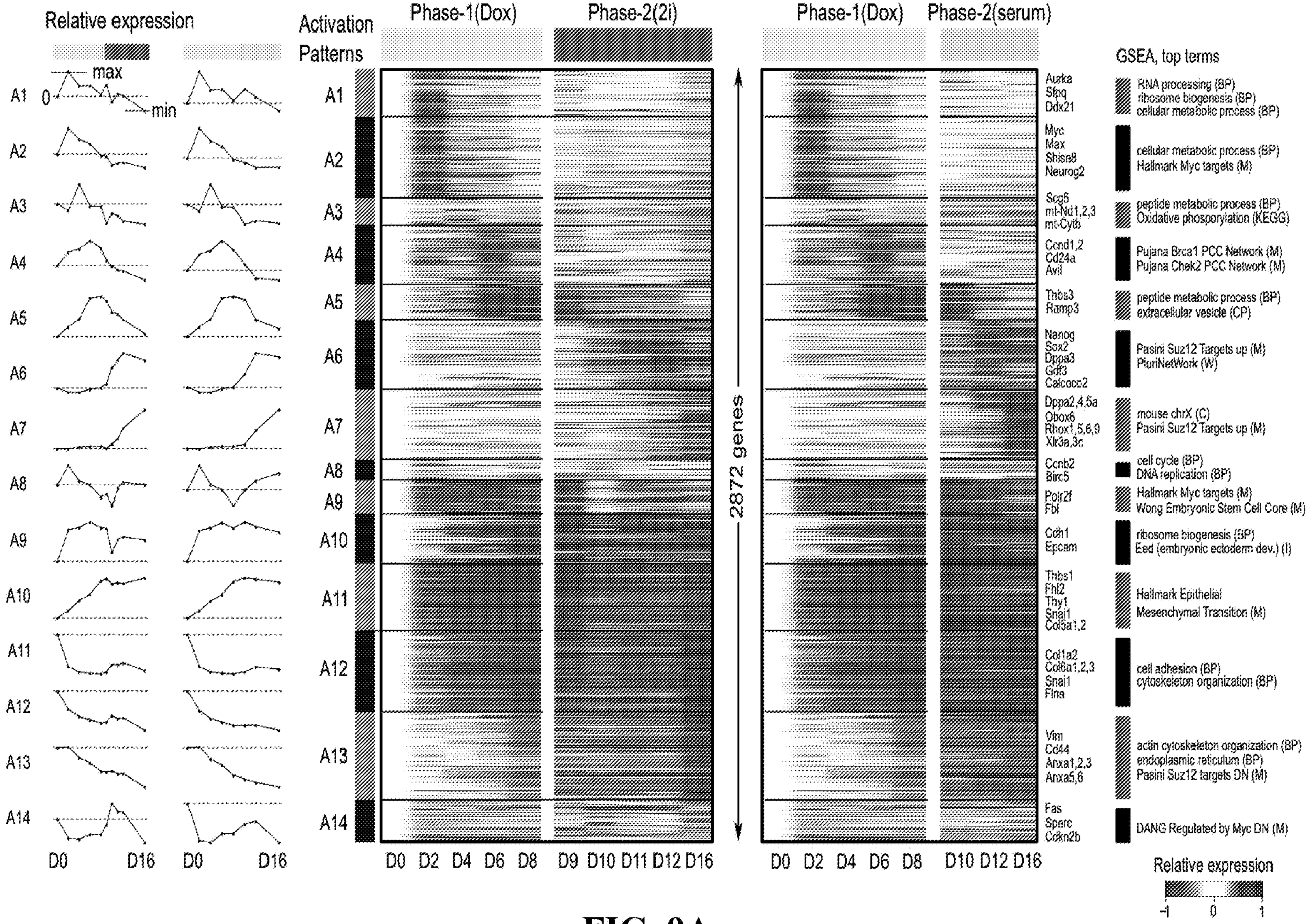

Relative expression
max
0
min
A1
A2
A3
A4
A5
A6
A7
A8
A9
A10
A11
A12
A13
A14
D0
D16
D0
D16
Activation Patterns
Phase-1(Dox)
Phase-2(2i)
D0 D2 D4 D6 D8
D9 D10 D11 D12 D16
2872 genes
Phase-1(Dox)
Phase-2(serum)
D0 D2 D4 D6 D8
D10 D12 D16
Aurka
Sfpq
Ddx21
Myc
Max
Shisa8
Neurog2
Scg5
mt-Nd1,2,3
mt-Cytb
Ccnd1,2
Cd24a
Avil
Thbs3
Ramp3
Nanog
Sox2
Dppa3
Gdf3
Calcocs2
Dppa2,4,5a
Obox6
Rhox1,5,6,9
Xlr3a,3c
Ccnb2
Birc5
Polr2f
Fbl
Cdh1
Epcam
Thbs1
Fhl2
Thy1
Snai1
Col5a1,2
Col1a2
Col6a1,2,3
Snai1
Flna
Vim
Cd44
Anxa1,2,3
Anxa5,6
Fas
Sparc
Cdkn2b
GSEA, top terms
RNA processing (BP)
ribosome biogenesis (BP)
cellular metabolic process (BP)
cellular metabolic process (BP)
Hallmark Myc targets (M)
peptide metabolic process (BP)
Oxidative phosporylation (KEGG)
Pujana Brca1 PCC Network (M)
Pujana Chek2 PCC Network (M)
peptide metabolic process (BP)
extracellular vesicle (CP)
Pasini Suz12 Targets up (M)
PluriNetWork (W)
mouse chrX (C)
Pasini Suz12 Targets up (M)
cell cycle (BP)
DNA replication (BP)
Hallmark Myc targets (M)
Wong Embryonic Stem Cell Core (M)
ribosome biogenesis (BP)
Eed (embryonic ectoderm dev.) (I)
Hallmark Epithelial
Mesenchymal Transition (M)
cell adhesion (BP)
cytoskeleton organization (BP)
actin cytoskeleton organization (BP)
endoplasmic reticulum (BP)
Pasini Suz12 targets DN (M)
DANG Regulated by Myc DN (M)
Relative expression
-1
0
1

FIG. 9A

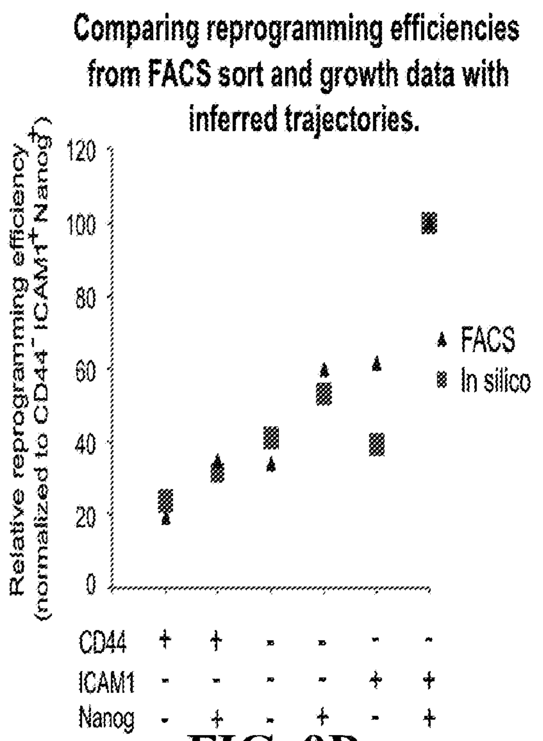


FIG. 9B

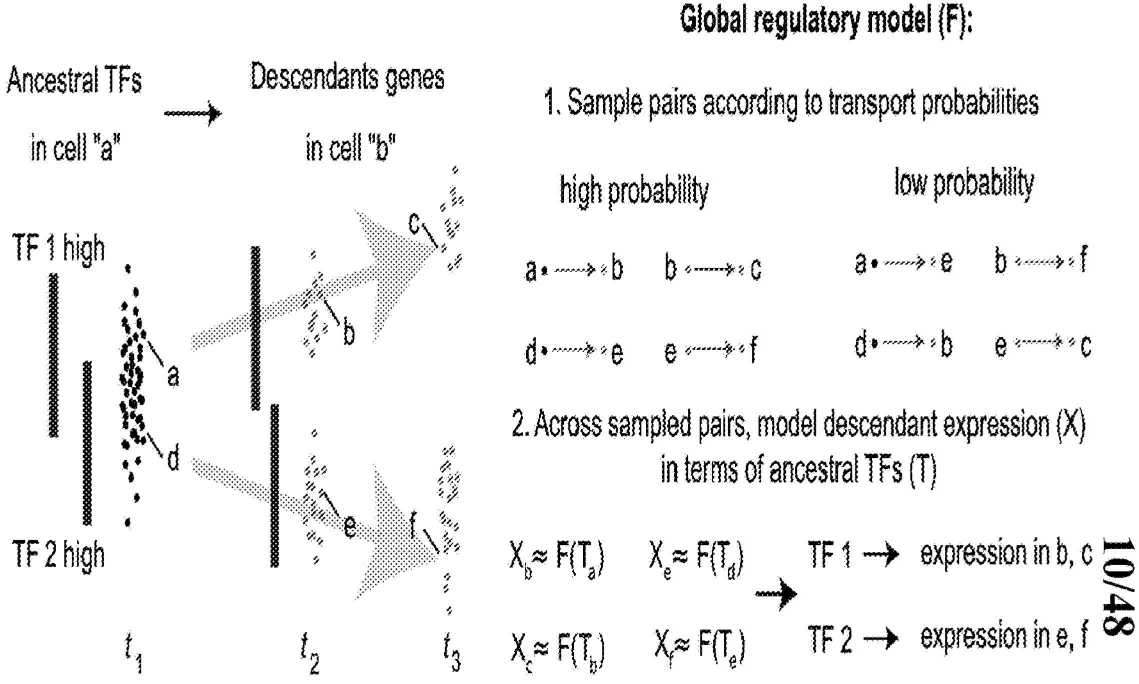

Ancestral TFs in cell "a"
Descendants genes in cell "b"
TF 1 high
TF 2 high
a
d
b
e
c
f
t_1
t_2
t_3
Global regulatory model (F):
1. Sample pairs according to transport probabilities
high probability
low probability
a → b
b → c
a → e
b → f
d → e
e → f
d → b
e → c
2. Across sampled pairs, model descendant expression (X) in terms of ancestral TFs (T)
$X_b \approx F(T_a)$
$X_e \approx F(T_d)$
$X_c \approx F(T_b)$
$X_f \approx F(T_e)$
TF 1 → expression in b, c
TF 2 → expression in e, f

FIG. 9C

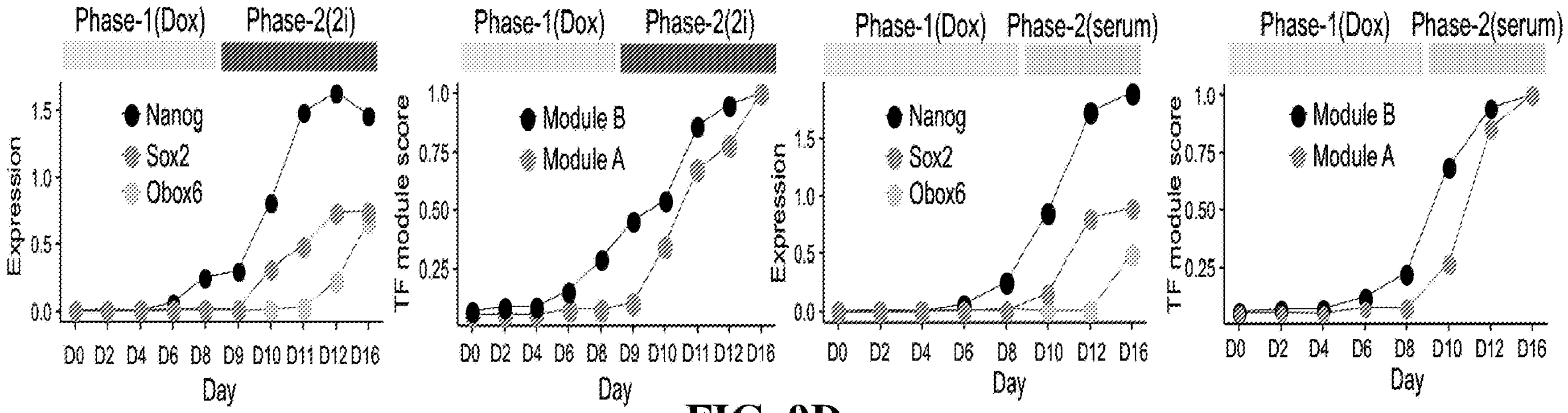


FIG. 9D

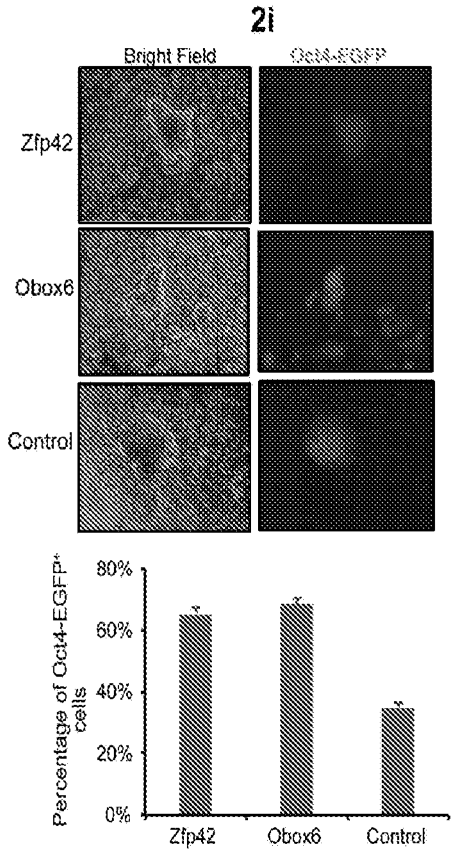


FIG. 10A

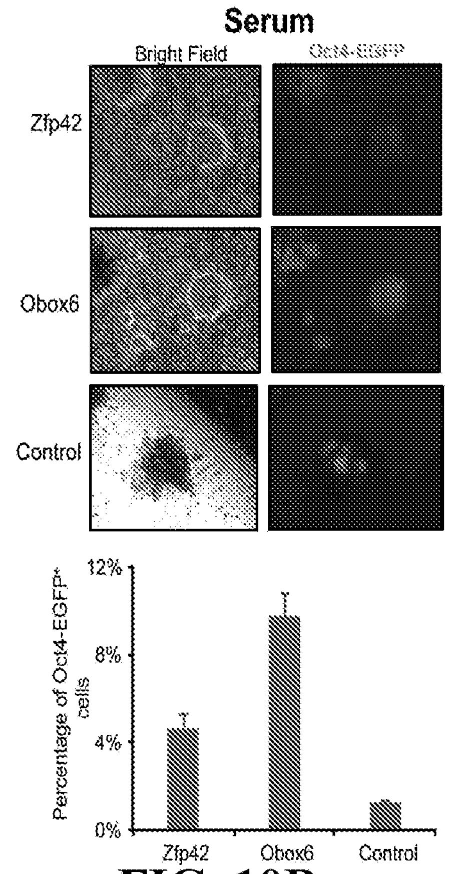


FIG. 10B

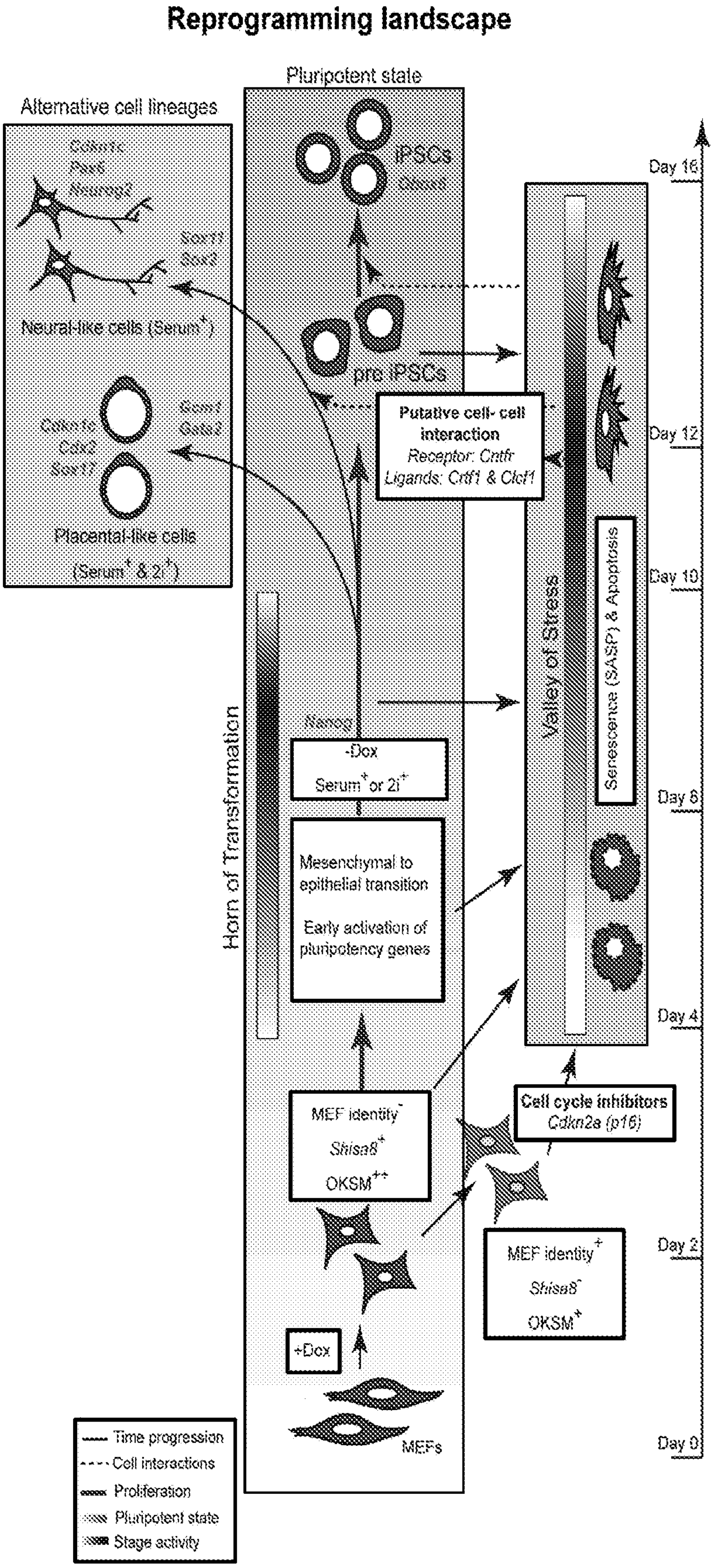


FIG. 10C

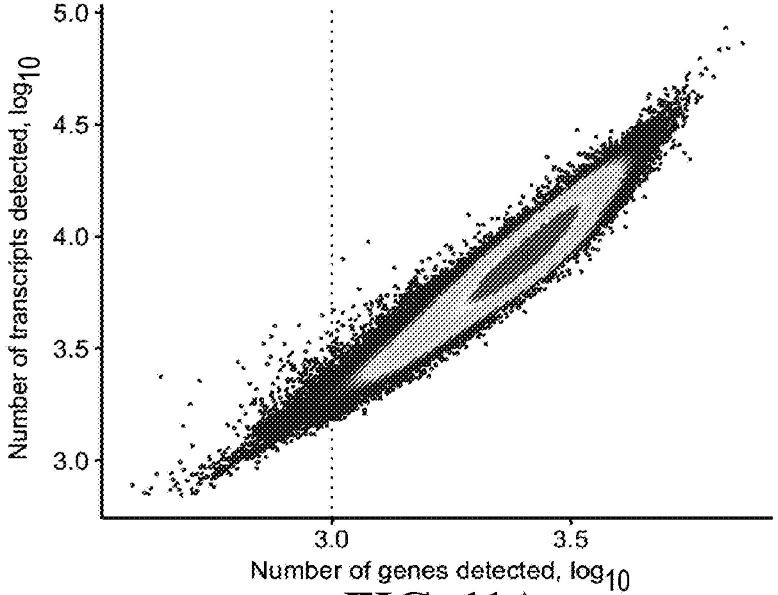


FIG. 11A

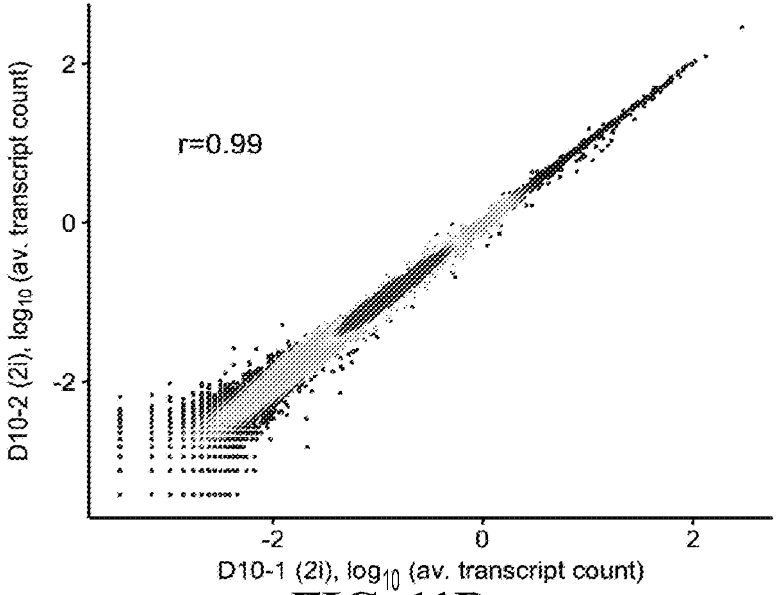


FIG. 11B

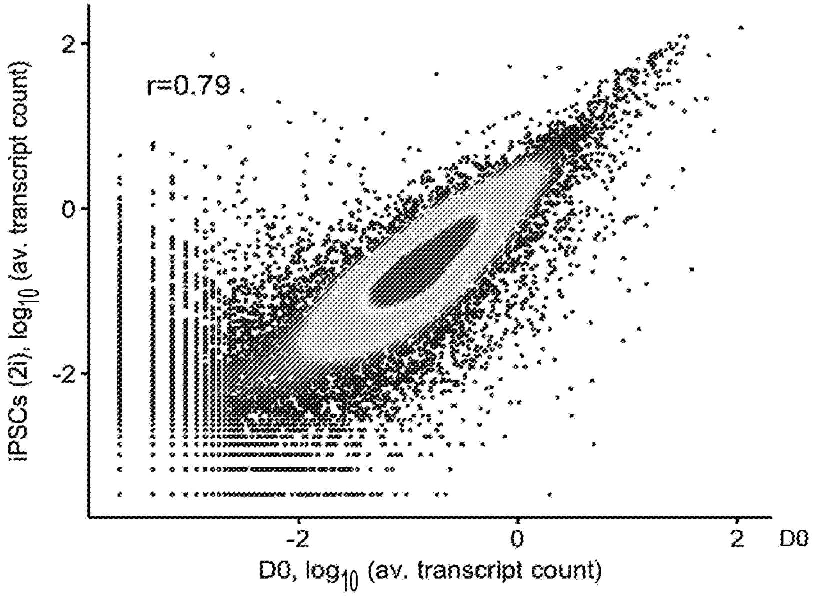


FIG. 11C

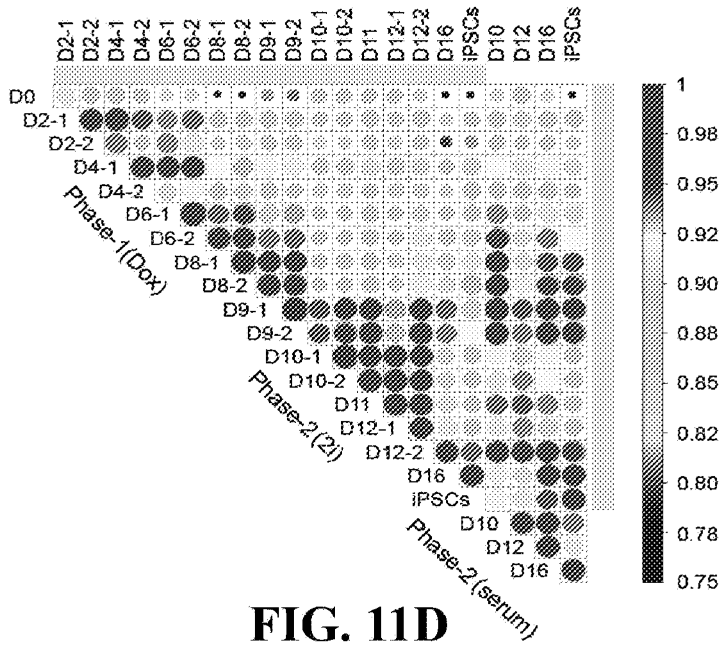


FIG. 11D

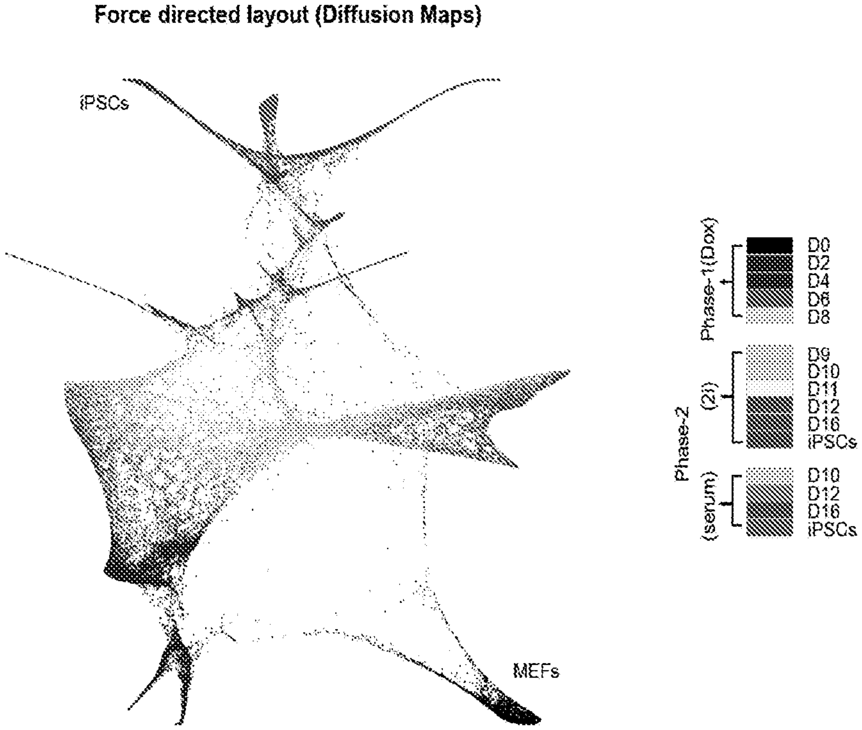
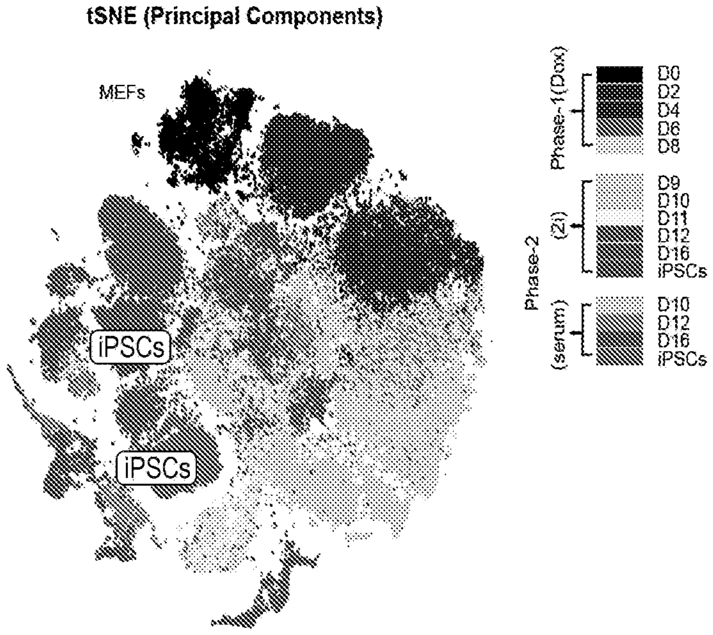


FIG. 12A



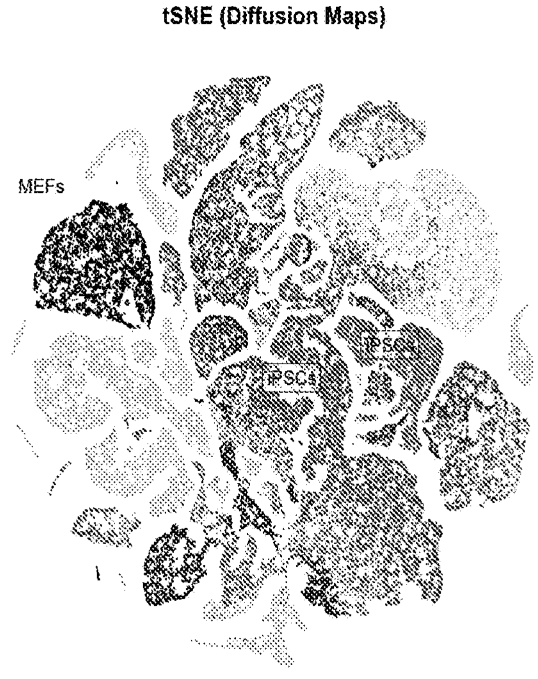


FIG. 12B

FIG. 12C

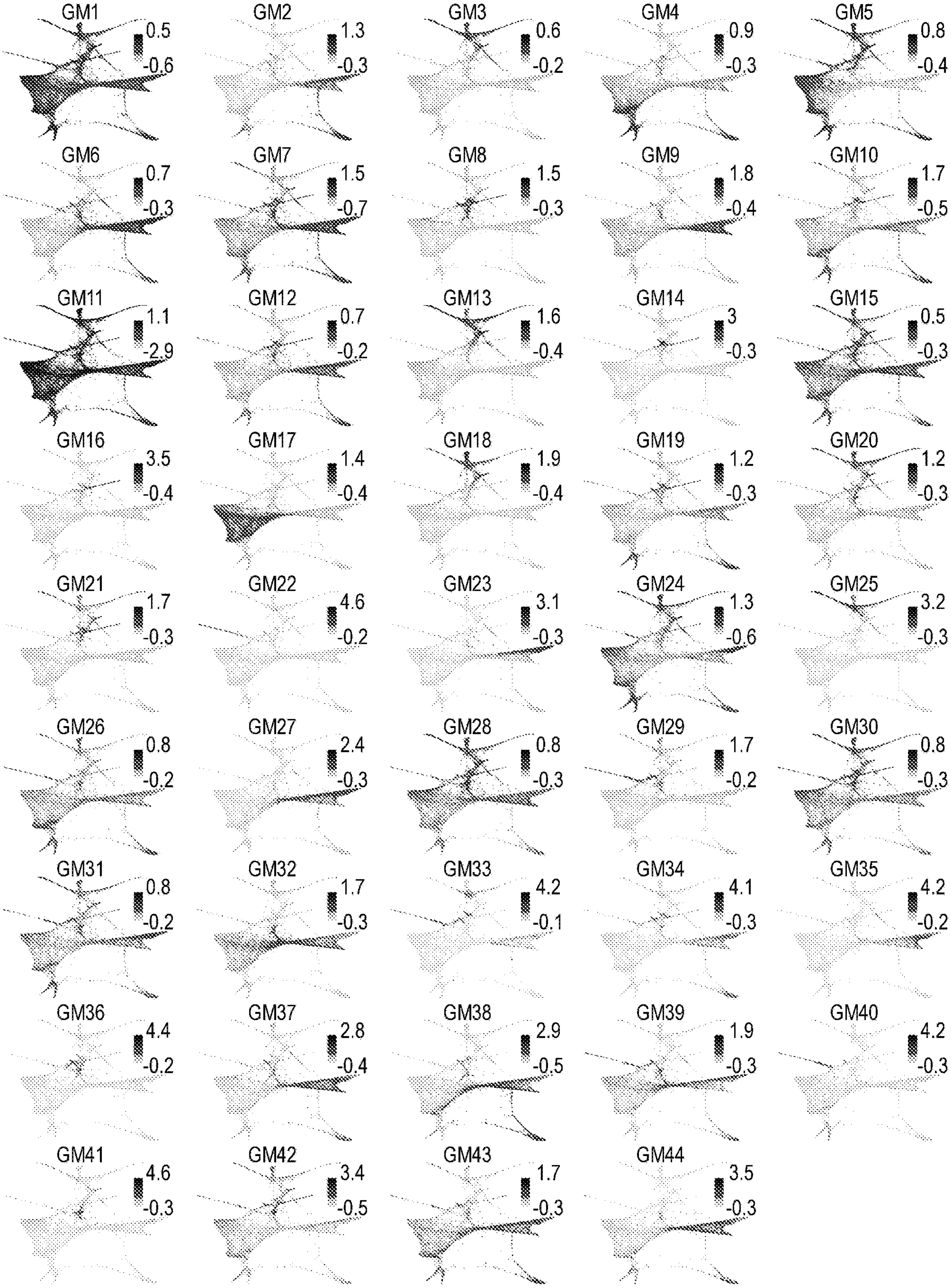

GM1
0.5
-0.6
GM2
1.3
-0.3
GM3
0.6
-0.2
GM4
0.9
-0.3
GM5
0.8
-0.4
GM6
0.7
-0.3
GM7
1.5
-0.7
GM8
1.5
-0.3
GM9
1.8
-0.4
GM10
1.7
-0.5
GM11
1.1
-2.9
GM12
0.7
-0.2
GM13
1.6
-0.4
GM14
3
-0.3
GM15
0.5
-0.3
GM16
3.5
-0.4
GM17
1.4
-0.4
GM18
1.9
-0.4
GM19
1.2
-0.3
GM20
1.2
-0.3
GM21
1.7
-0.3
GM22
4.6
-0.2
GM23
3.1
-0.3
GM24
1.3
-0.6
GM25
3.2
-0.3
GM26
0.8
-0.2
GM27
2.4
-0.3
GM28
0.8
-0.3
GM29
1.7
-0.2
GM30
0.8
-0.3
GM31
0.8
-0.2
GM32
1.7
-0.3
GM33
4.2
-0.1
GM34
4.1
-0.3
GM35
4.2
-0.2
GM36
4.4
-0.2
GM37
2.8
-0.4
GM38
2.9
-0.5
GM39
1.9
-0.3
GM40
4.2
-0.3
GM41
4.6
-0.3
GM42
3.4
-0.5
GM43
1.7
-0.3
GM44
3.5
-0.3

FIG. 13

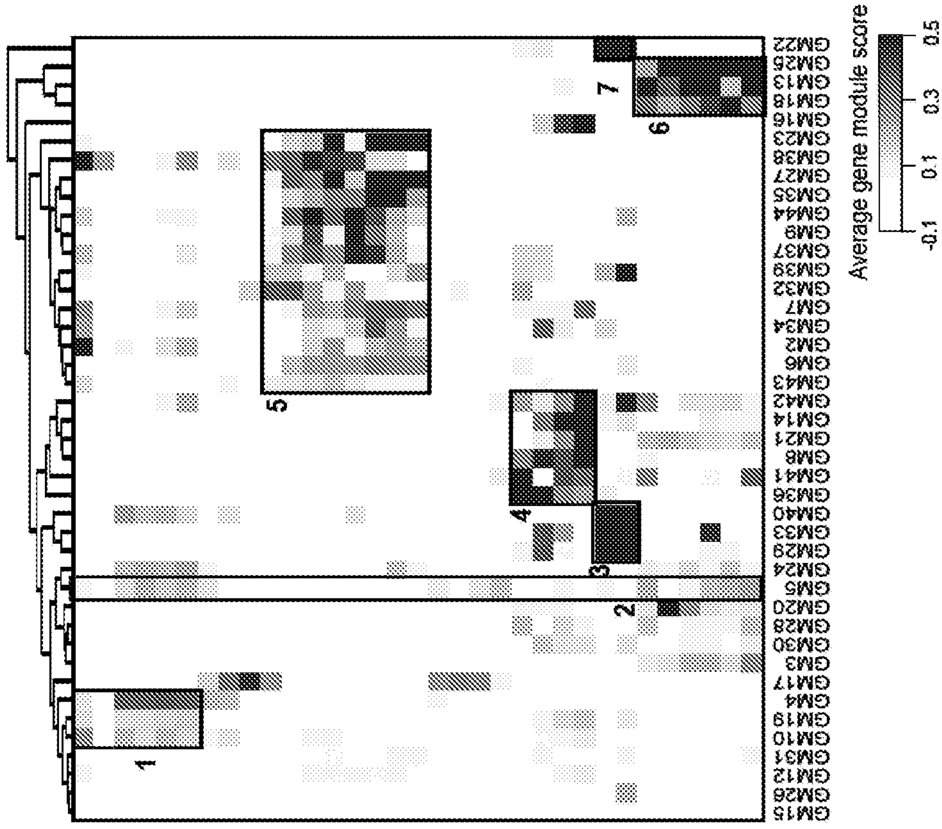


FIG. 14B

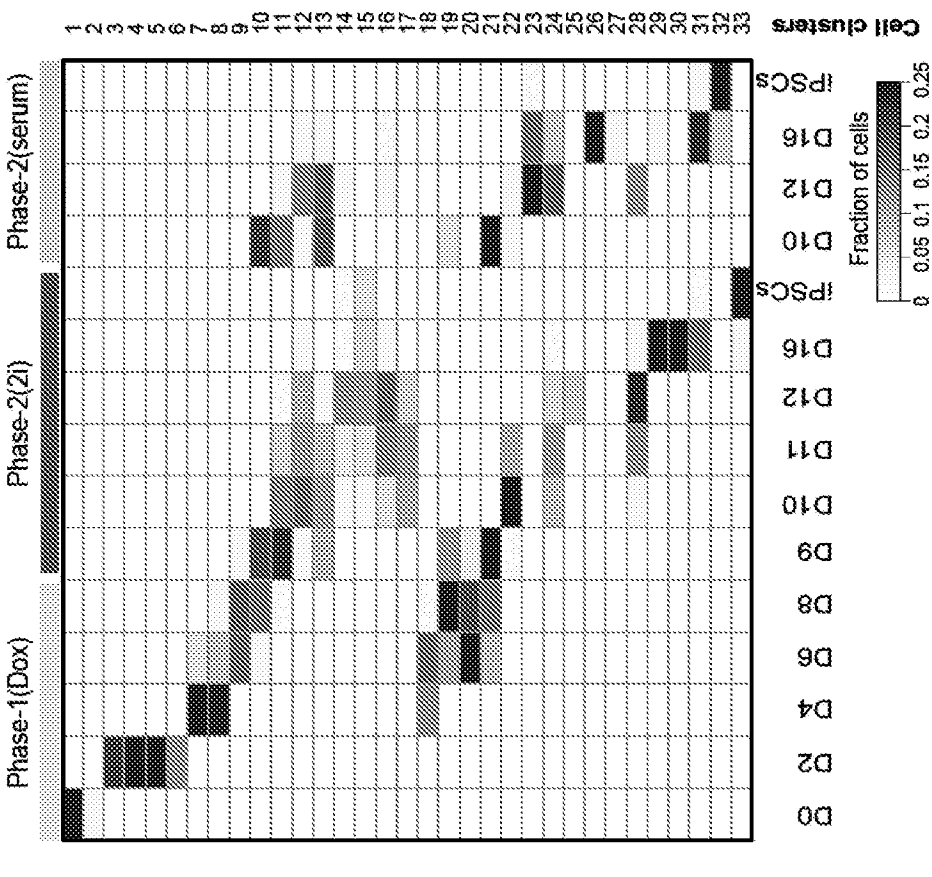


FIG. 14A

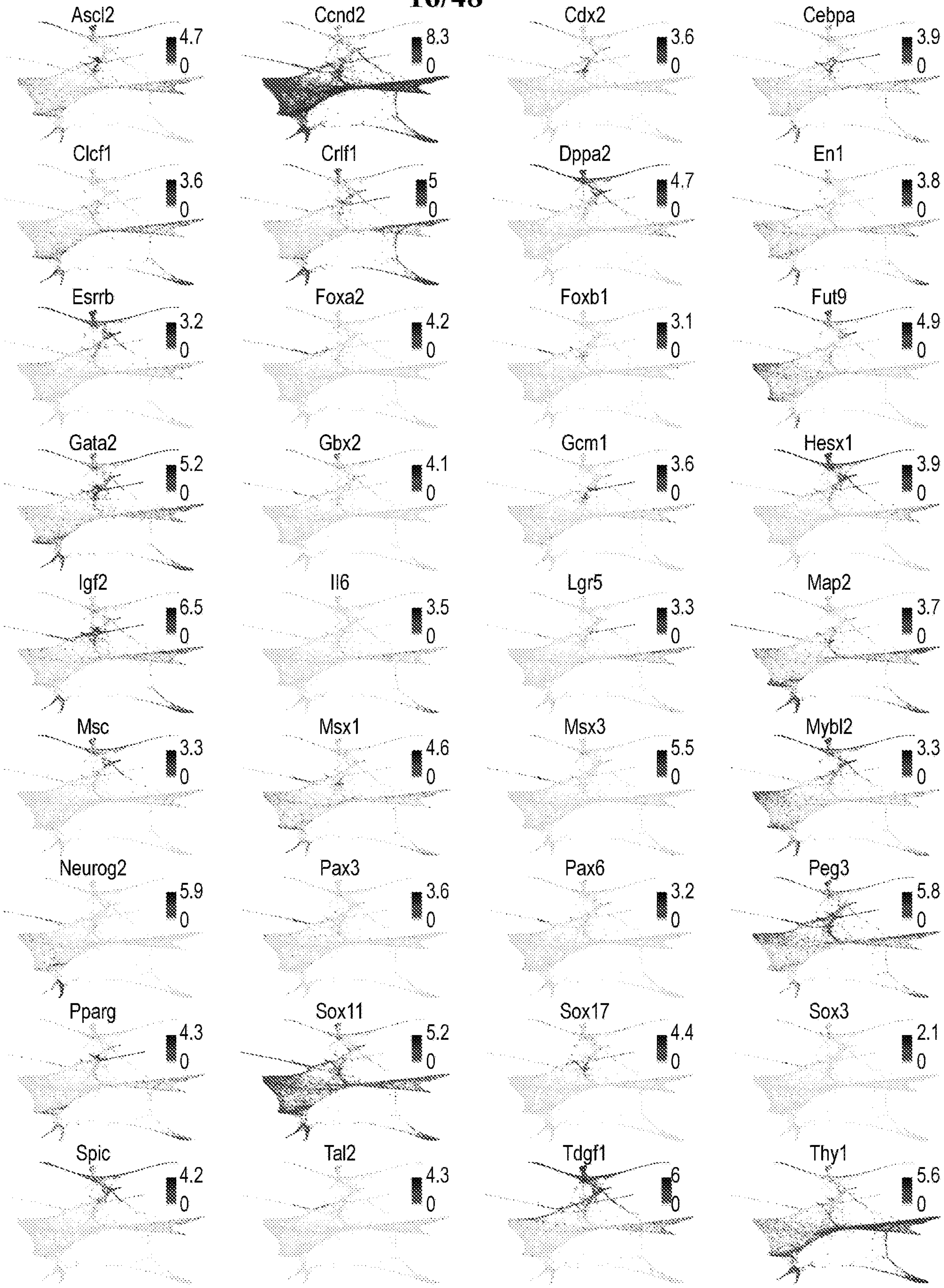

Ascl2
4.7
0
Ccnd2
8.3
0
Cdx2
3.6
0
Cebpa
3.9
0
Clcf1
3.6
0
Crlf1
5
0
Dppa2
4.7
0
En1
3.8
0
Esrrb
3.2
0
Foxa2
4.2
0
Foxb1
3.1
0
Fut9
4.9
0
Gata2
5.2
0
Gbx2
4.1
0
Gcm1
3.6
0
Hesx1
3.9
0
Igf2
6.5
0
Il6
3.5
0
Lgr5
3.3
0
Map2
3.7
0
Msc
3.3
0
Msx1
4.6
0
Msx3
5.5
0
Mybl2
3.3
0
Neurog2
5.9
0
Pax3
3.6
0
Pax6
3.2
0
Peg3
5.8
0
Pparg
4.3
0
Sox11
5.2
0
Sox17
4.4
0
Sox3
2.1
0
Spic
4.2
0
Tal2
4.3
0
Tdgf1
6
0
Thy1
5.6
0

FIG. 15

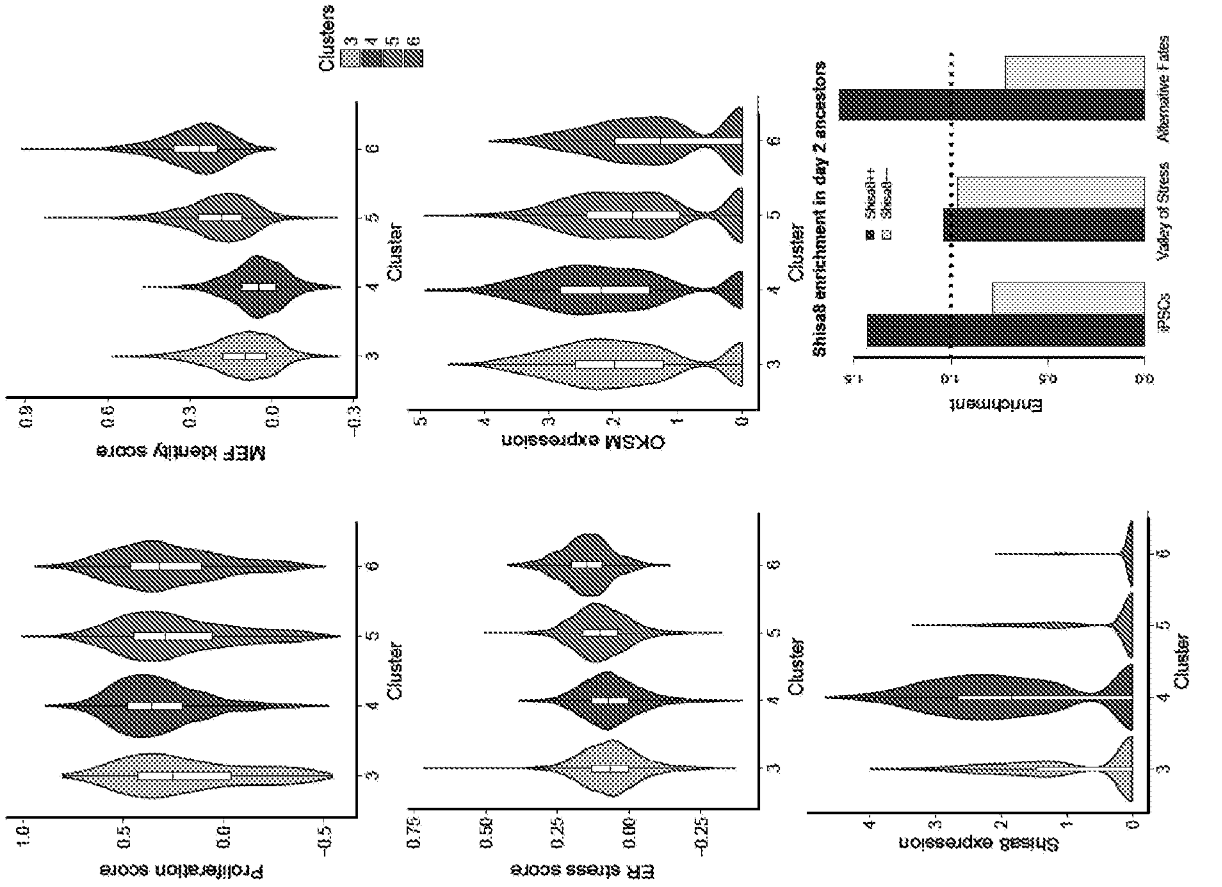


FIG. 16B

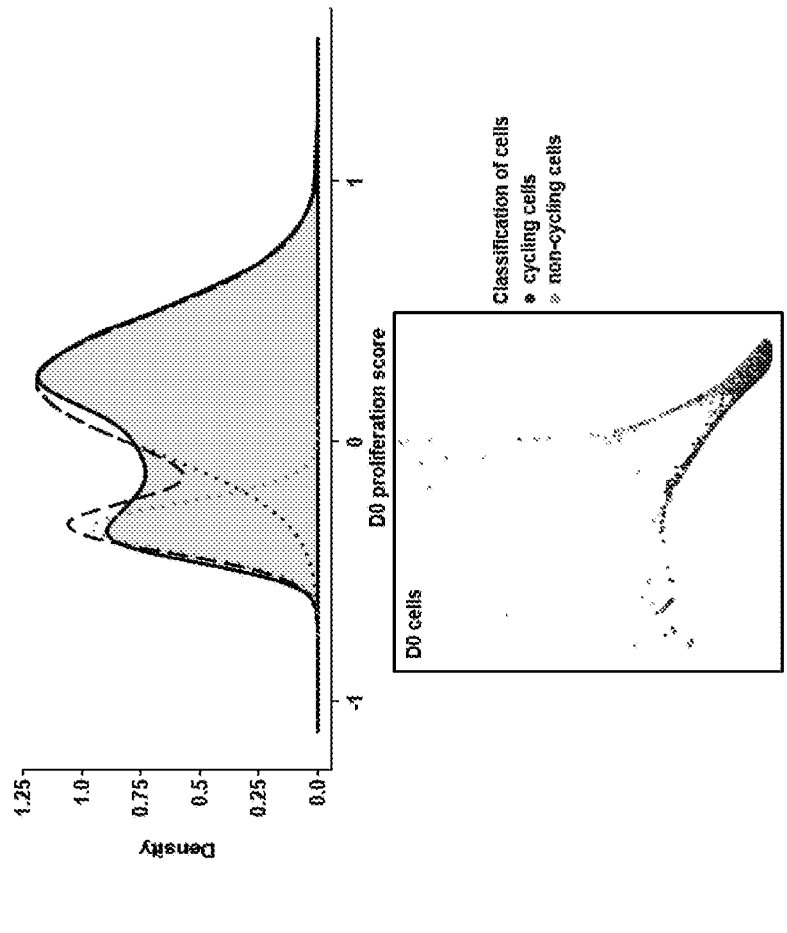


FIG. 16A

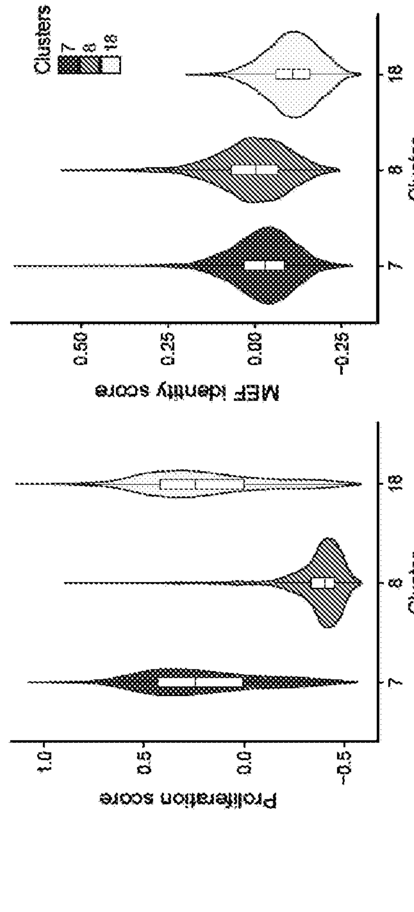


FIG. 16C

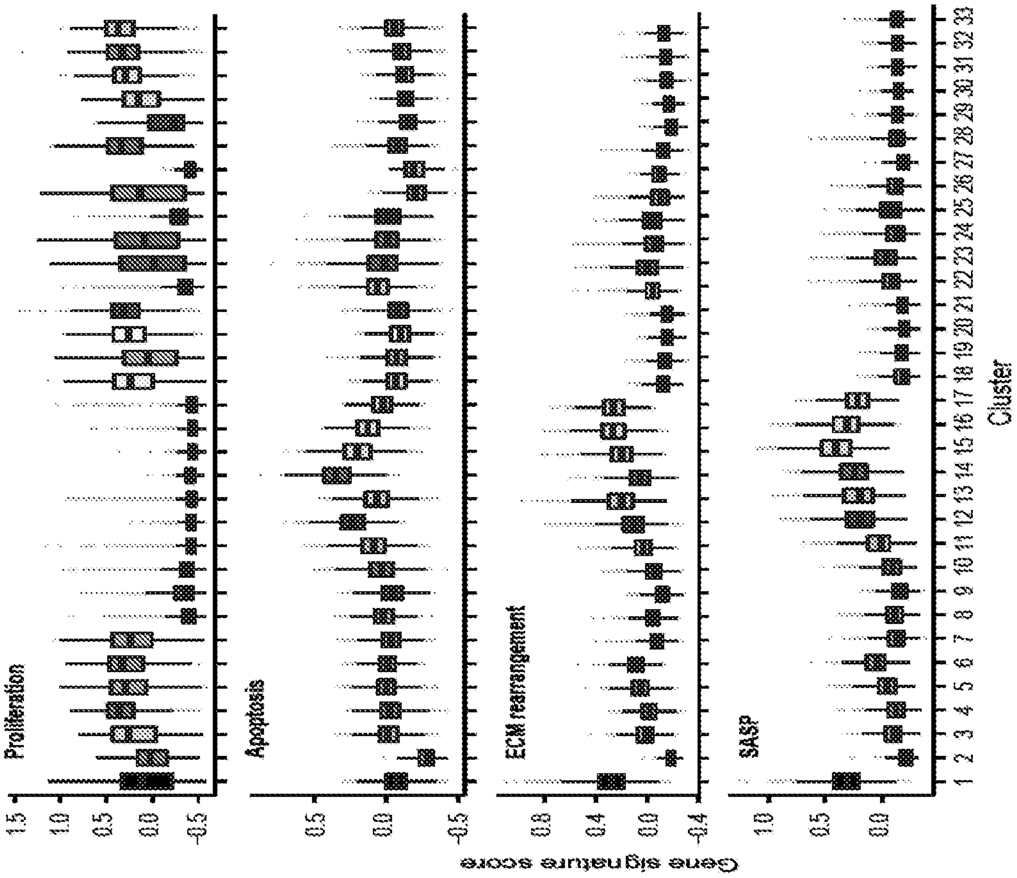


FIG. 16E

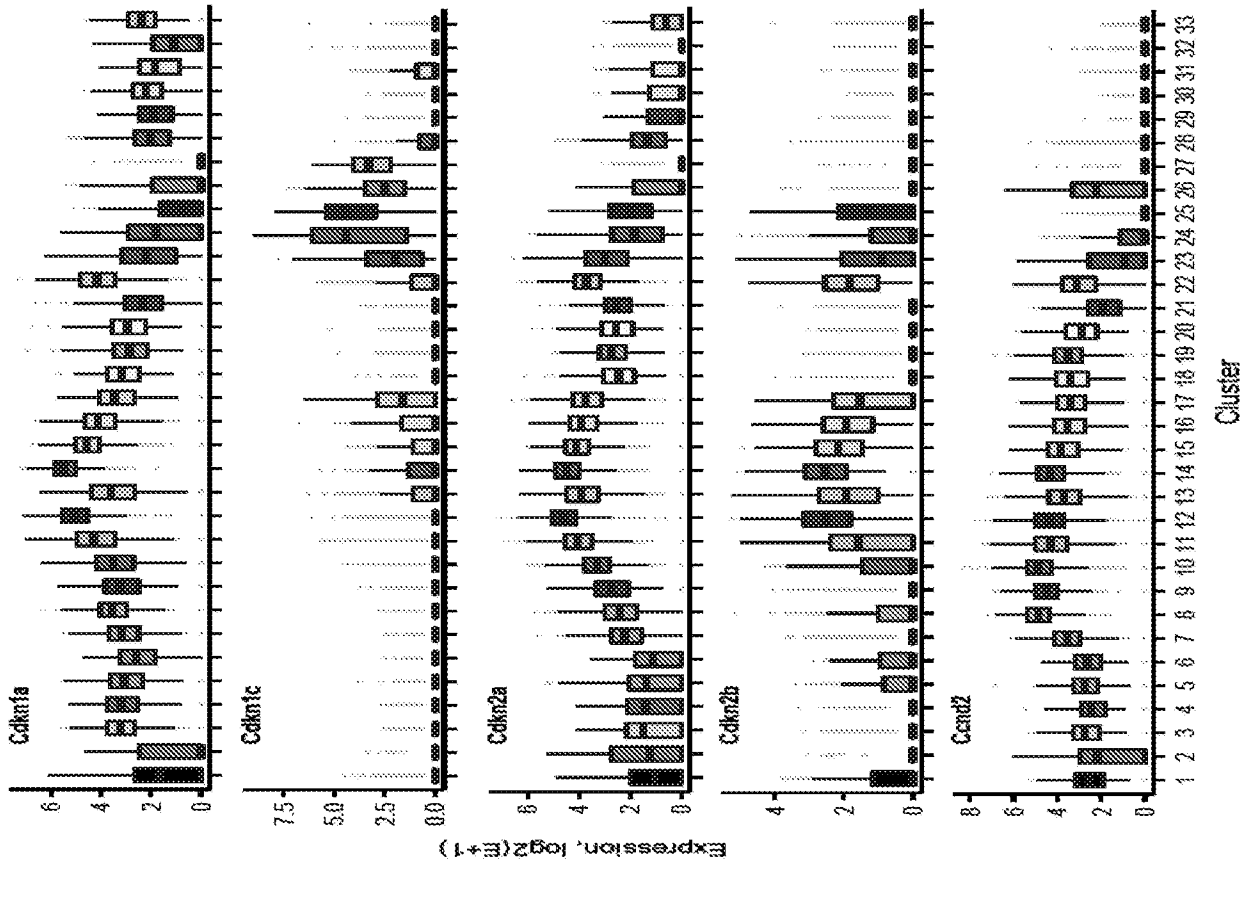


FIG. 16D

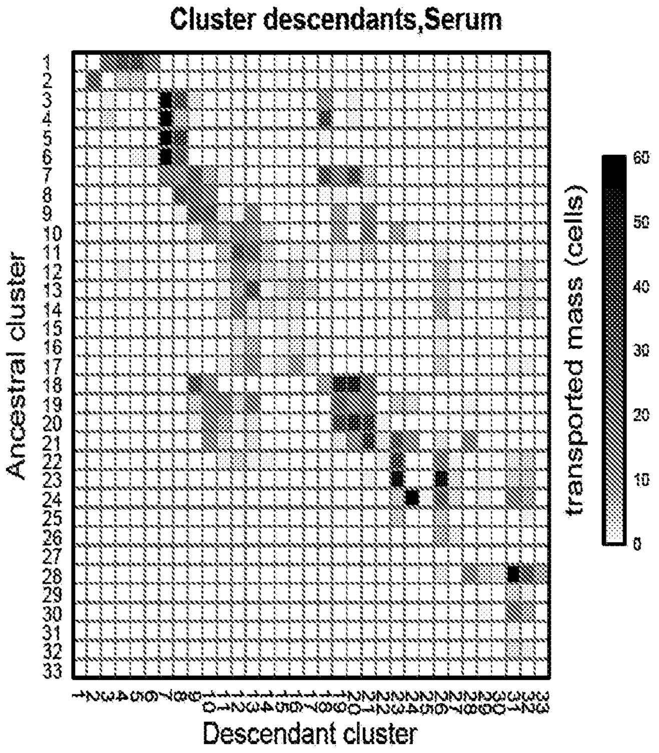


FIG. 17A

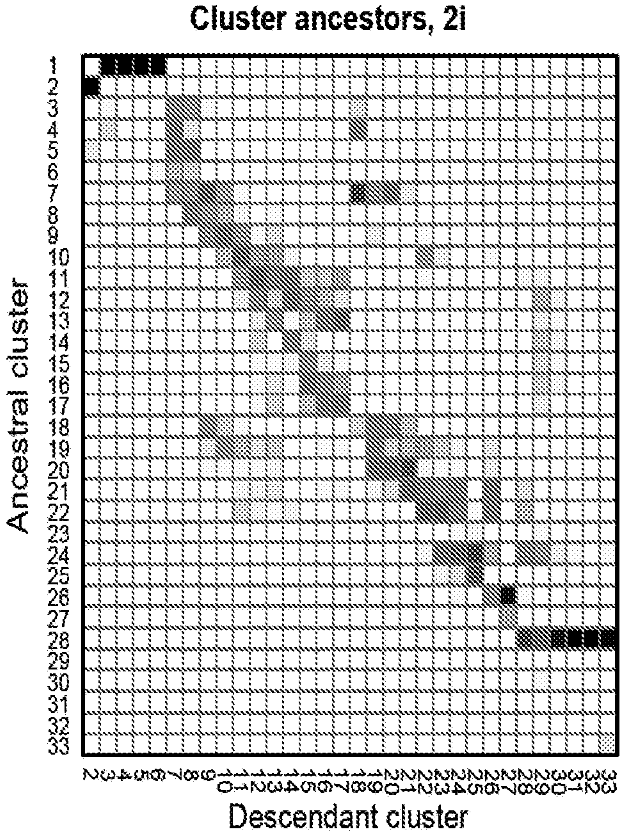


FIG. 17B

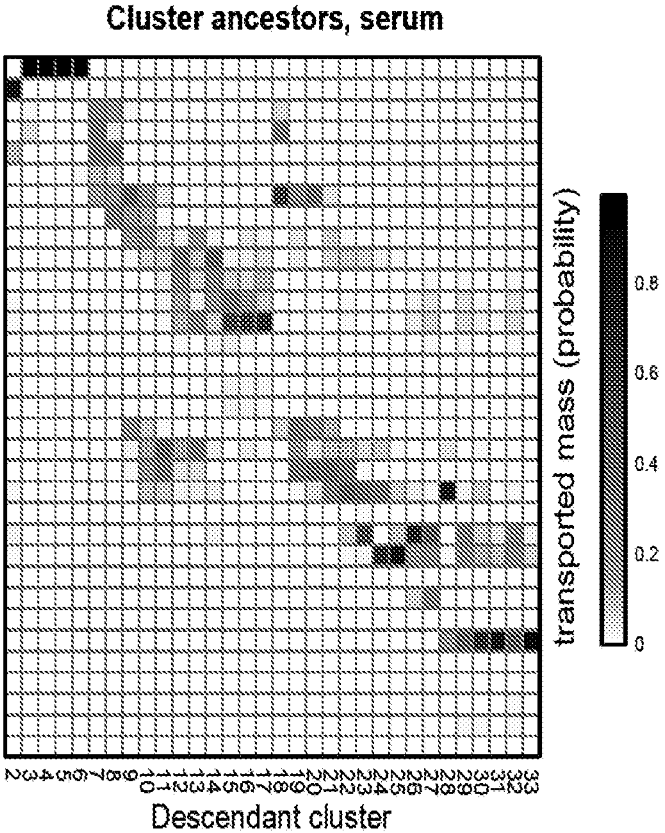


FIG. 17C

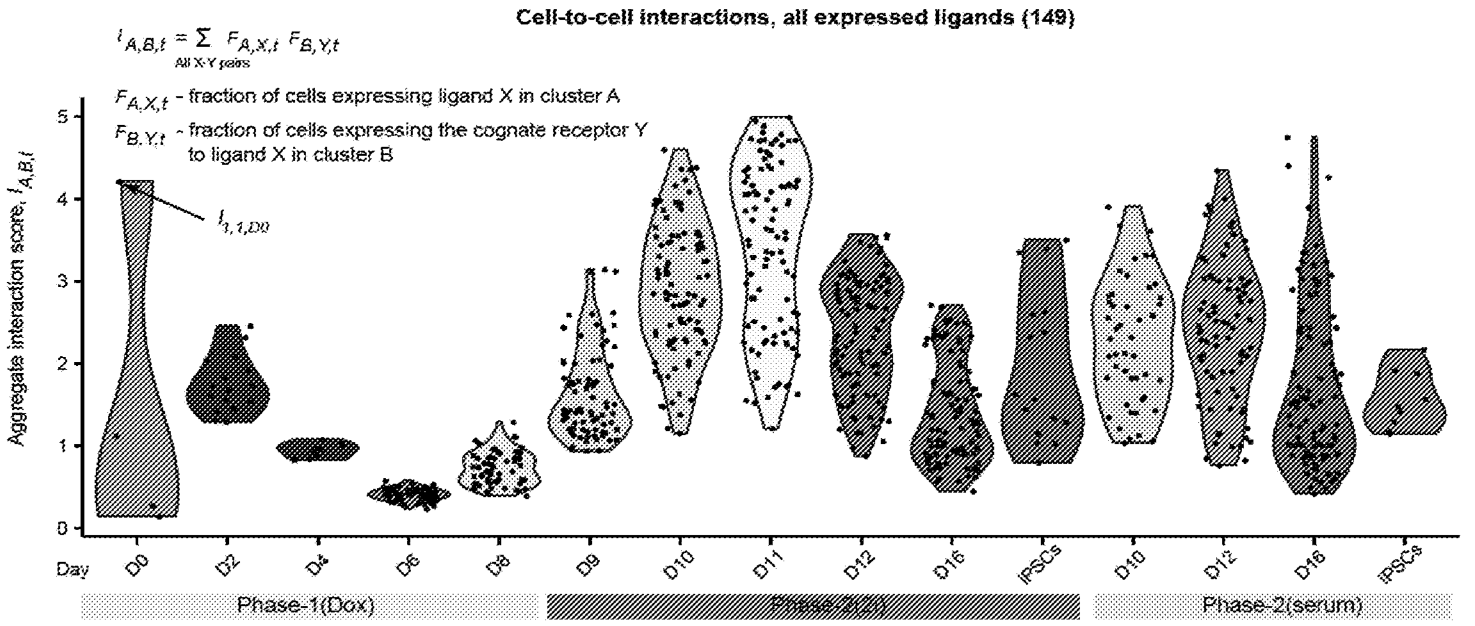


FIG. 18A

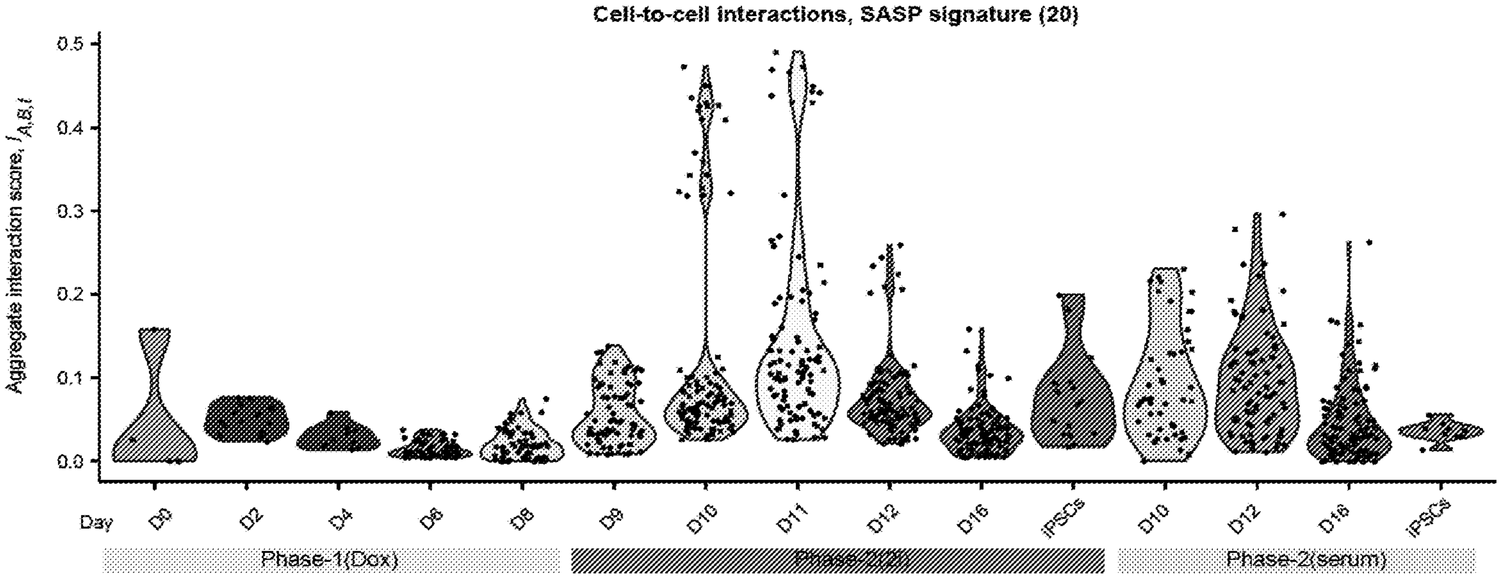


FIG. 18B

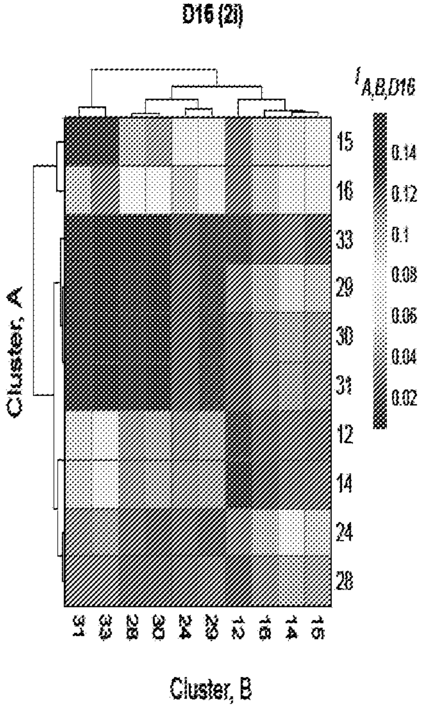


FIG. 18C

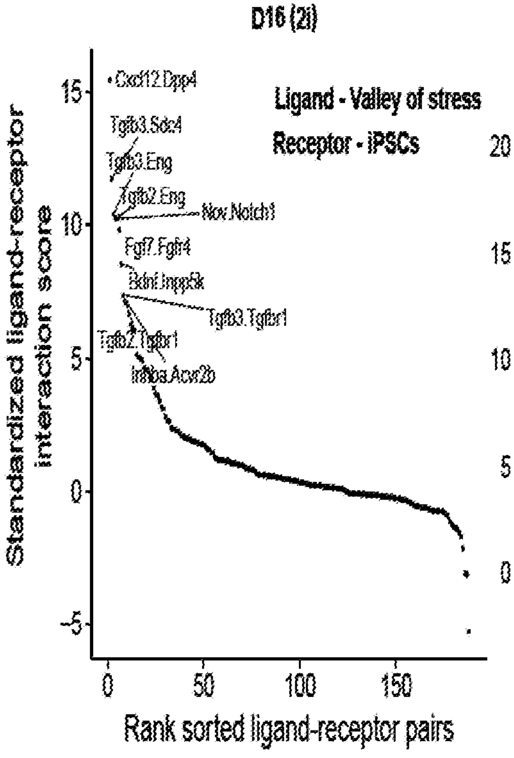


FIG. 18D

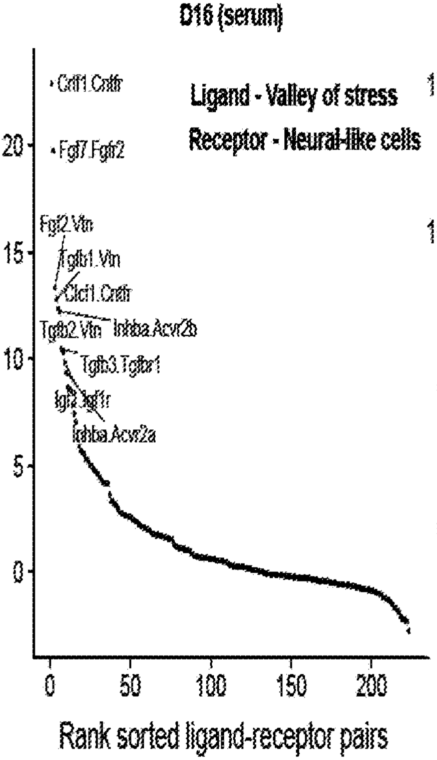


FIG. 18E

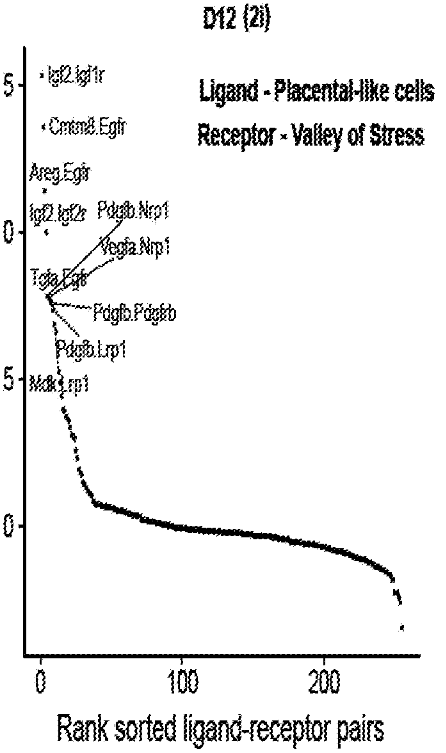


FIG. 18F

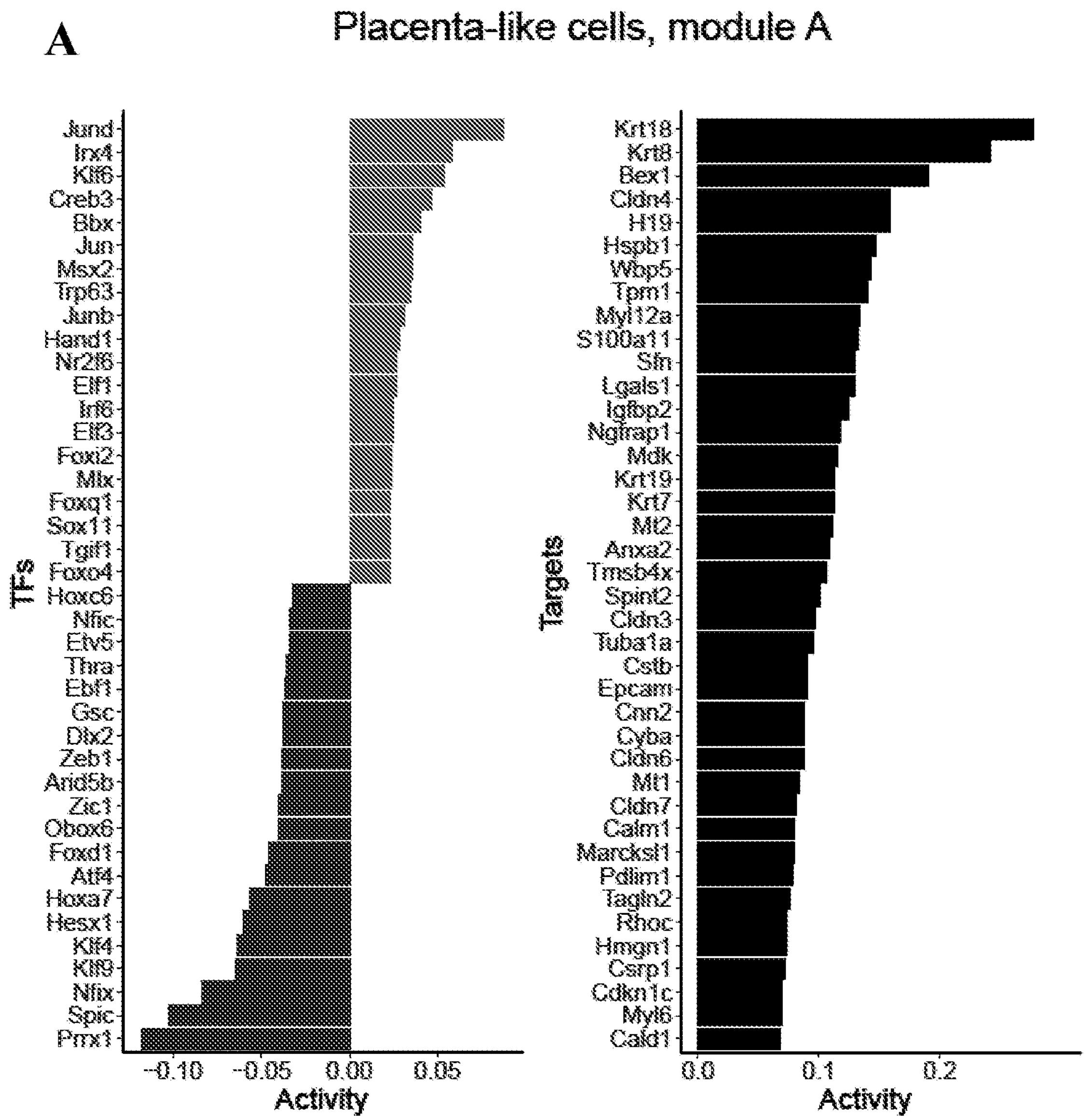
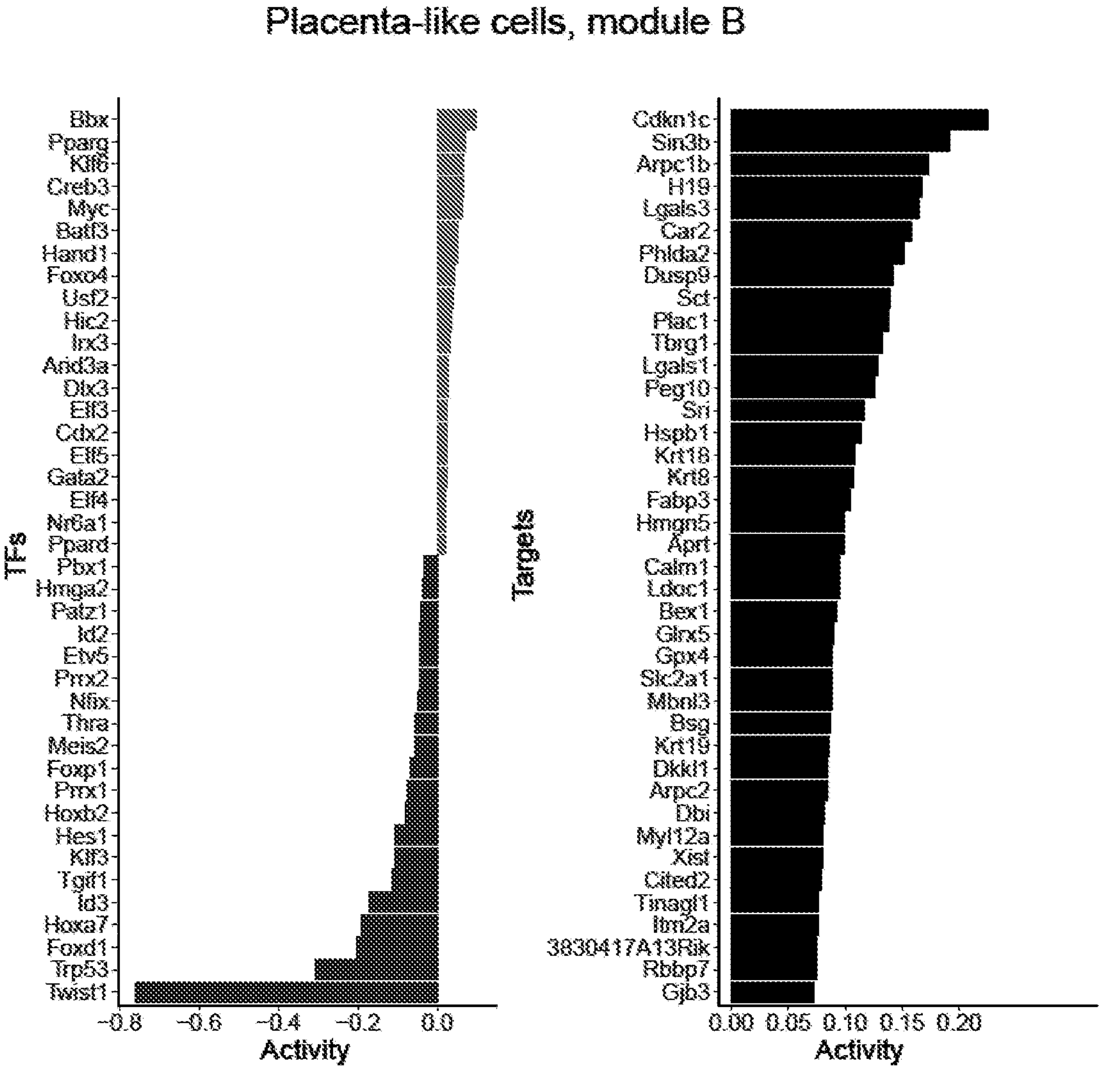

A
Placenta-like cells, module A
TFs
Jund
Irx4
Klf6
Creb3
Bbx
Jun
Msx2
Trp63
Junb
Hand1
Nr2f6
Elf1
Irf6
Elf3
Foxi2
Mlx
Foxq1
Sox11
Tgif1
Foxo4
Hoxc6
Nfic
Etv5
Thra
Ebf1
Gsc
Dlx2
Zeb1
Arid5b
Zic1
Obox6
Foxd1
Atf4
Hoxa7
Hesx1
Klf4
Klf9
Nfix
Spic
Prrx1
−0.10
−0.05
0.00
0.05
Activity
Targets
Krt18
Krt8
Bex1
Cldn4
H19
Hspb1
Wbp5
Tpm1
Myl12a
S100a11
Sfn
Lgals1
Igfbp2
Ngfrap1
Mdk
Krt19
Krt7
Mt2
Anxa2
Tmsb4x
Spint2
Cldn3
Tuba1a
Cstb
Epcam
Cnn2
Cyba
Cldn6
Mt1
Cldn7
Calm1
Marcksl1
Pdlim1
Tagln2
Rhoc
Hmgn1
Csrp1
Cdkn1c
Myl6
Cald1
0.0
0.1
0.2
Activity

FIG. 19A



Placenta-like cells, module B

FIG. 19B

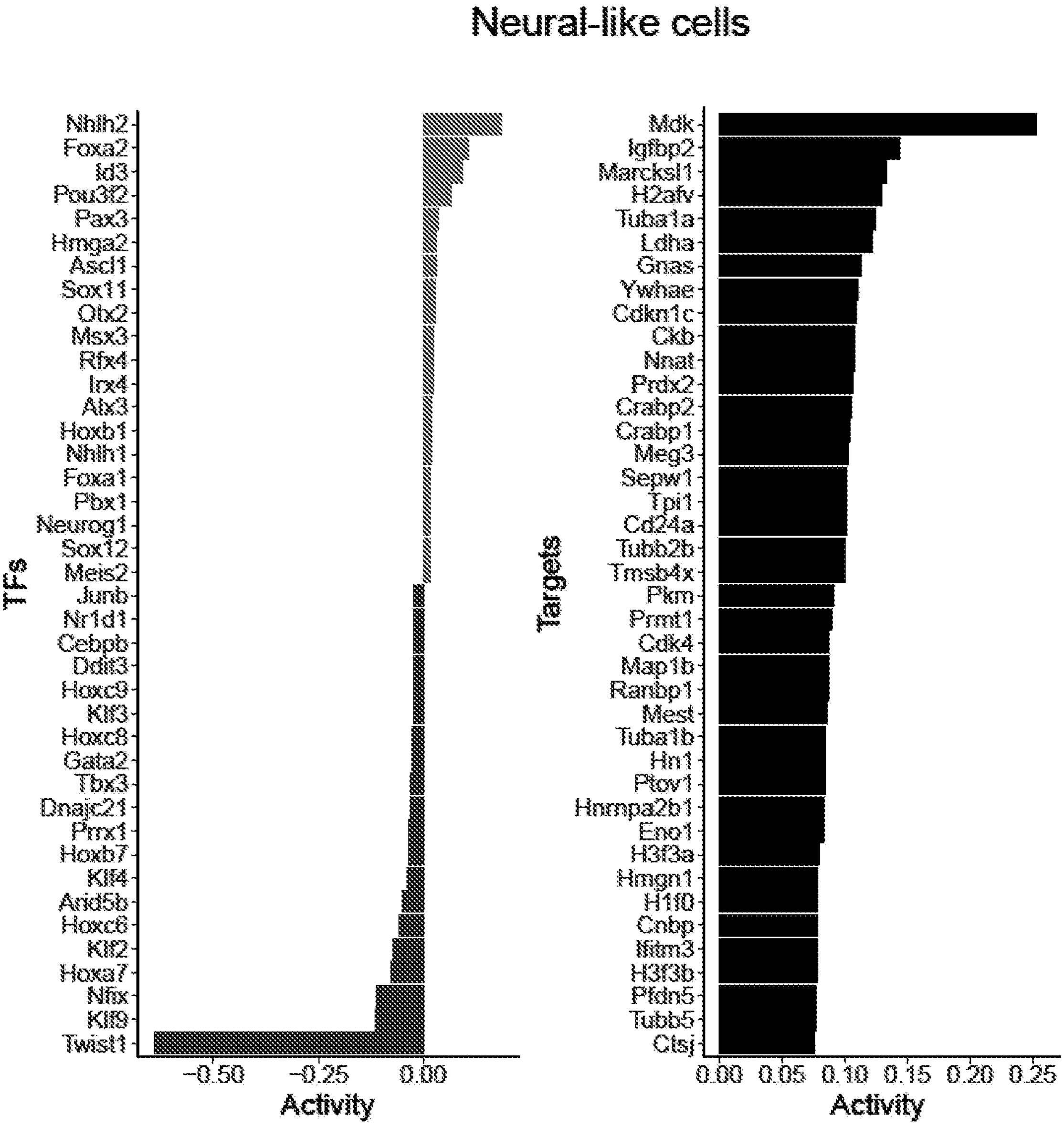

Neural-like cells
TFs
Nhlh2
Foxa2
Id3
Pou3f2
Pax3
Hmga2
Ascl1
Sox11
Otx2
Msx3
Rfx4
Irx4
Alx3
Hoxb1
Nhlh1
Foxa1
Pbx1
Neurog1
Sox12
Meis2
Junb
Nr1d1
Cebpb
Ddit3
Hoxc9
Klf3
Hoxc8
Gata2
Tbx3
Dnajc21
Prrx1
Hoxb7
Klf4
Arid5b
Hoxc6
Klf2
Hoxa7
Nfix
Klf9
Twist1
−0.50
−0.25
0.00
Activity
Targets
Mdk
Igfbp2
Marcksl1
H2afv
Tuba1a
Ldha
Gnas
Ywhae
Cdkn1c
Ckb
Nnat
Prdx2
Crabp2
Crabp1
Meg3
Sepw1
Tpi1
Cd24a
Tubb2b
Tmsb4x
Pkm
Prmt1
Cdk4
Map1b
Ranbp1
Mest
Tuba1b
Hn1
Ptov1
Hnrnpa2b1
Eno1
H3f3a
Hmgn1
H1f0
Cnbp
Ifitm3
H3f3b
Pfdn5
Tubb5
Ctsj
0.00
0.05
0.10
0.15
0.20
0.25
Activity

FIG. 19C

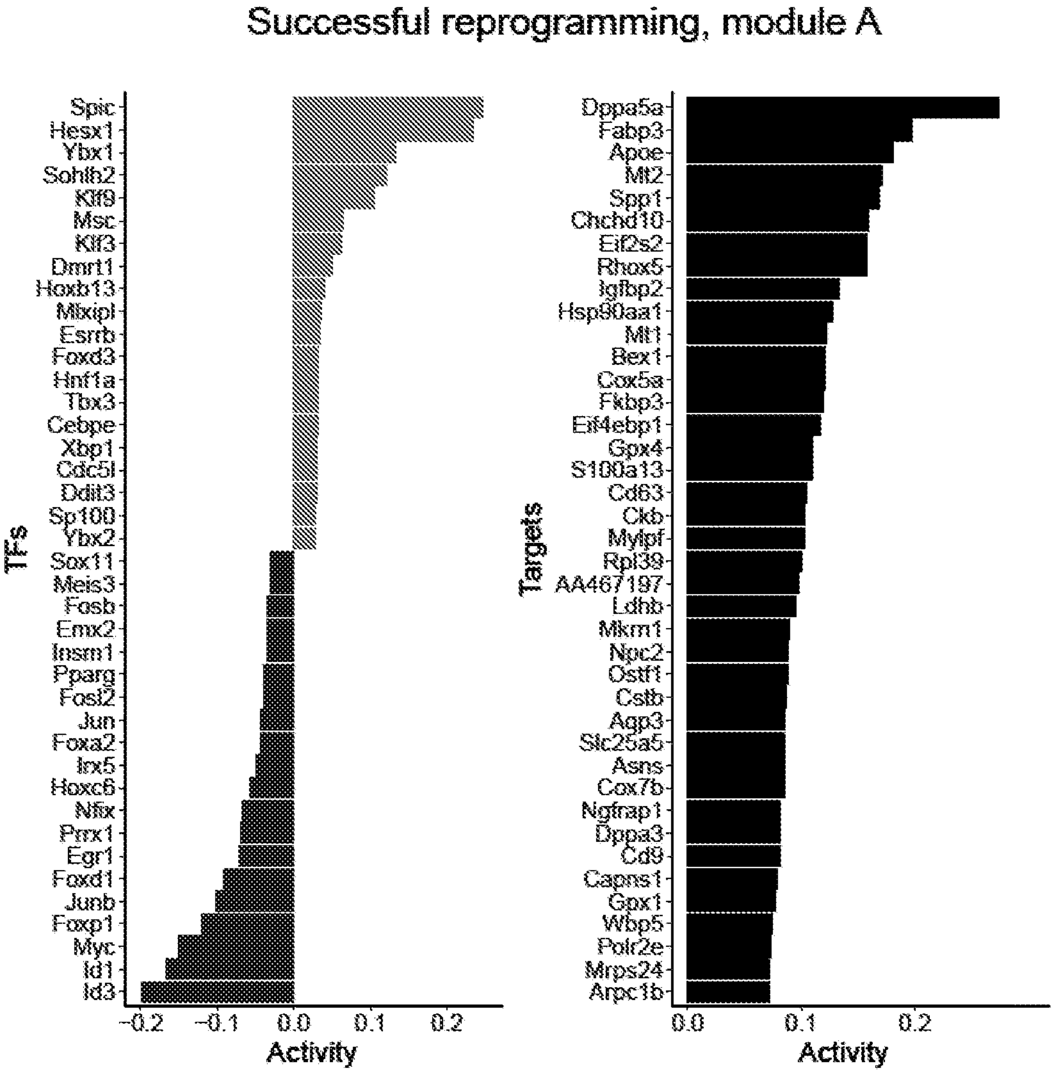

Successful reprogramming, module A
TFs
Spic
Hesx1
Ybx1
Sohlh2
Klf9
Msc
Klf3
Dmrt1
Hoxb13
Mlxipl
Esrrb
Foxd3
Hnf1a
Tbx3
Cebpe
Xbp1
Cdc5l
Ddit3
Sp100
Ybx2
Sox11
Meis3
Fosb
Emx2
Insm1
Pparg
Fosl2
Jun
Foxa2
Irx5
Hoxc6
Nfix
Prrx1
Egr1
Foxd1
Junb
Foxp1
Myc
Id1
Id3
−0.2
−0.1
0.0
0.1
0.2
Activity
Targets
Dppa5a
Fabp3
Apoe
Mt2
Spp1
Chchd10
Eif2s2
Rhox5
Igfbp2
Hsp90aa1
Mt1
Bex1
Cox5a
Fkbp3
Eif4ebp1
Gpx4
S100a13
Cd63
Ckb
Mylpf
Rpl39
AA467197
Ldhb
Mkrn1
Npc2
Ostf1
Cstb
Aqp3
Slc25a5
Asns
Cox7b
Ngfrap1
Dppa3
Cd9
Capns1
Gpx1
Wbp5
Polr2e
Mrps24
Arpc1b
0.0
0.1
0.2
Activity

FIG. 19D

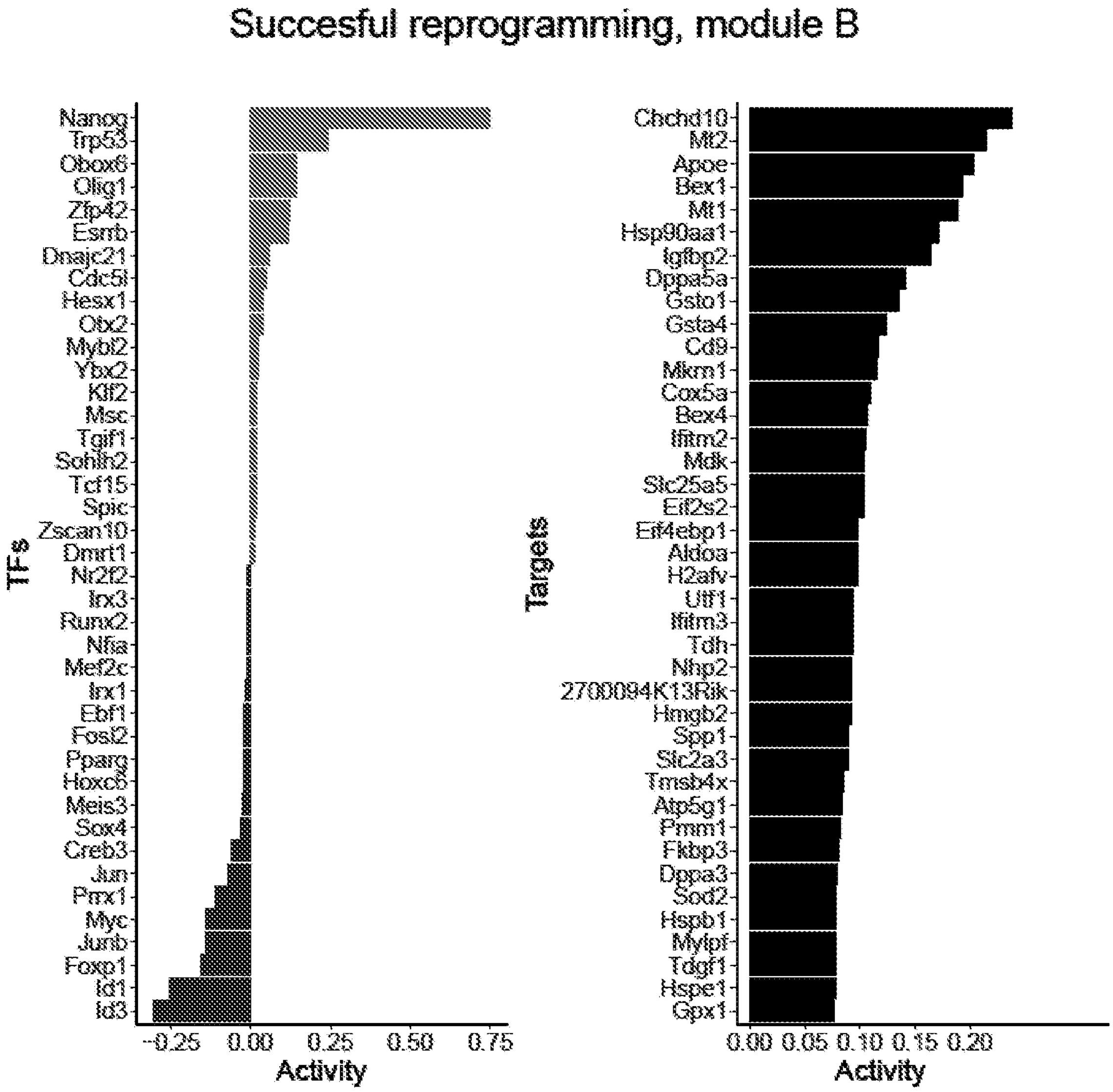

Succesful reprogramming, module B
TFs
Nanog
Trp53
Obox6
Olig1
Zfp42
Esrrb
Dnajc21
Cdc5l
Hesx1
Otx2
Mybl2
Ybx2
Klf2
Msc
Tgif1
Sohlh2
Tcf15
Spic
Zscan10
Dmrt1
Nr2f2
Irx3
Runx2
Nfia
Mef2c
Irx1
Ebf1
Fosl2
Pparg
Hoxc6
Meis3
Sox4
Creb3
Jun
Prrx1
Myc
Junb
Foxp1
Id1
Id3
-0.25
0.00
0.25
0.50
0.75
Activity
Targets
Chchd10
Mt2
Apoe
Bex1
Mt1
Hsp90aa1
Igfbp2
Dppa5a
Gsto1
Gsta4
Cd9
Mkrn1
Cox5a
Bex4
Ifitm2
Mdk
Slc25a5
Eif2s2
Eif4ebp1
Aldoa
H2afv
Utf1
Ifitm3
Tdh
Nhp2
2700094K13Rik
Hmgb2
Spp1
Slc2a3
Tmsb4x
Atp5g1
Pmm1
Fkbp3
Dppa3
Sod2
Hspb1
Mylpf
Tdgf1
Hspe1
Gpx1
0.00
0.05
0.10
0.15
0.20
Activity

FIG. 19E

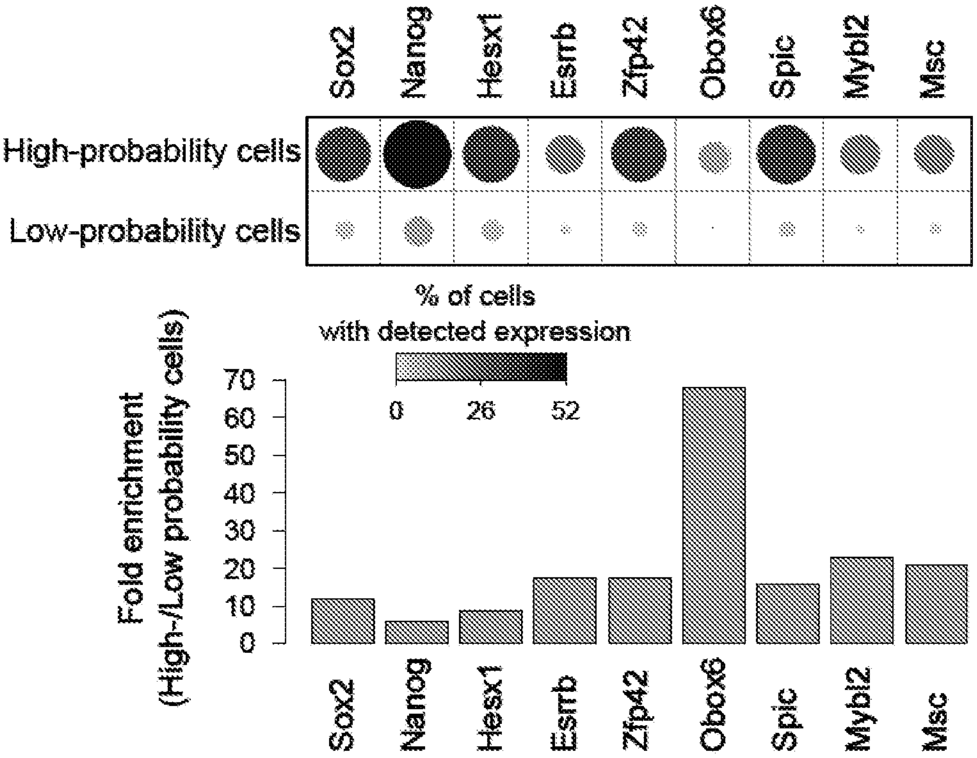

Sox2
Nanog
Hesx1
Esrrb
Zfp42
Obox6
Spic
Mybl2
Msc
High-probability cells
Low-probability cells
% of cells
with detected expression
0
26
52
Fold enrichment
(High-/Low probability cells)
70
60
50
40
30
20
10
0
Sox2
Nanog
Hesx1
Esrrb
Zfp42
Obox6
Spic
Mybl2
Msc

FIG. 19F

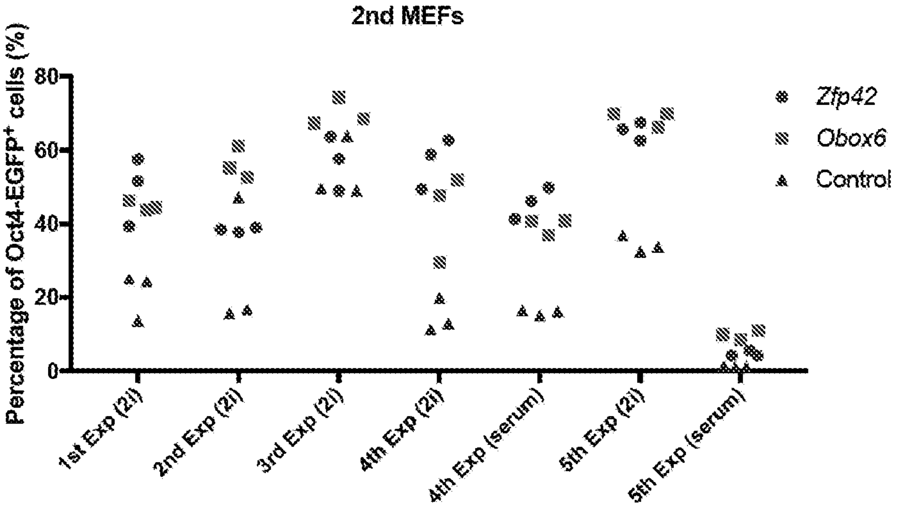


FIG. 20A

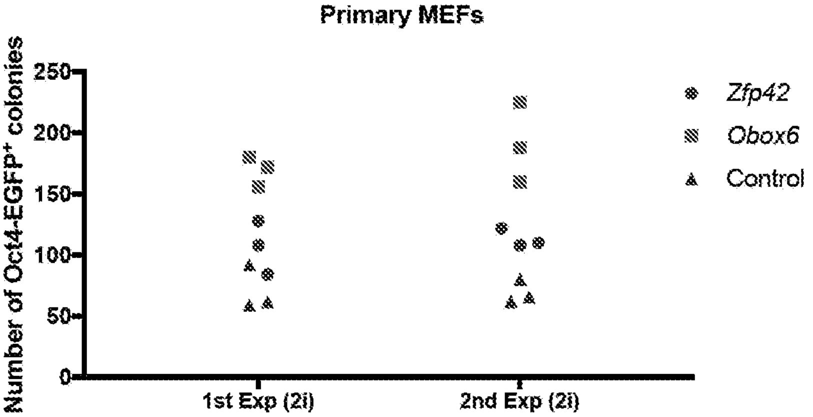


FIG. 20B

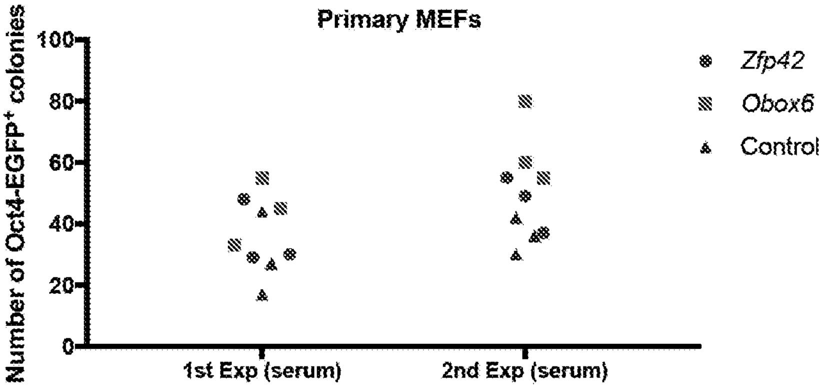


FIG. 20C

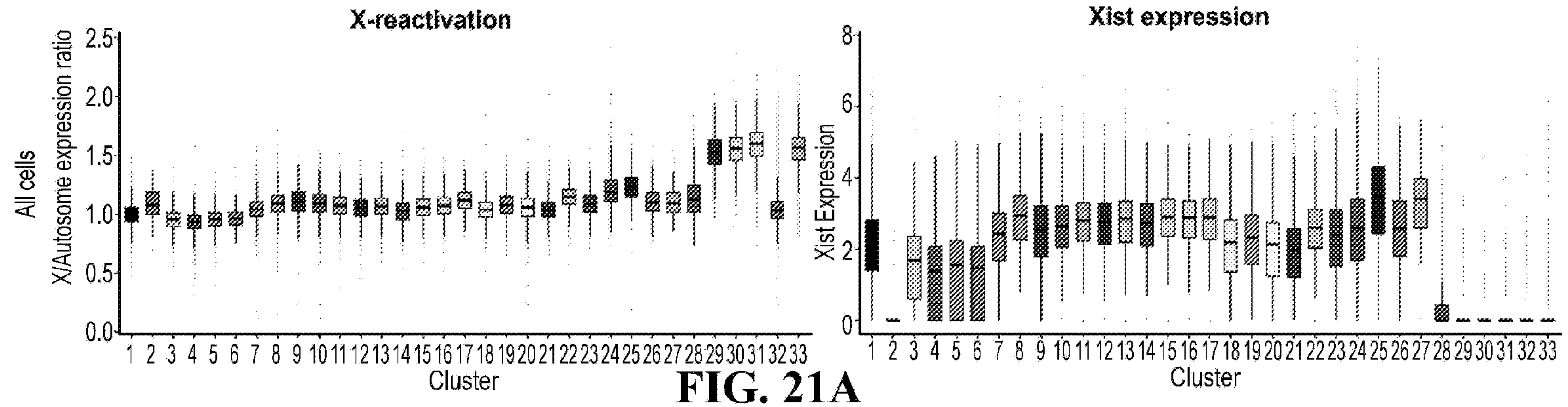


FIG. 21A

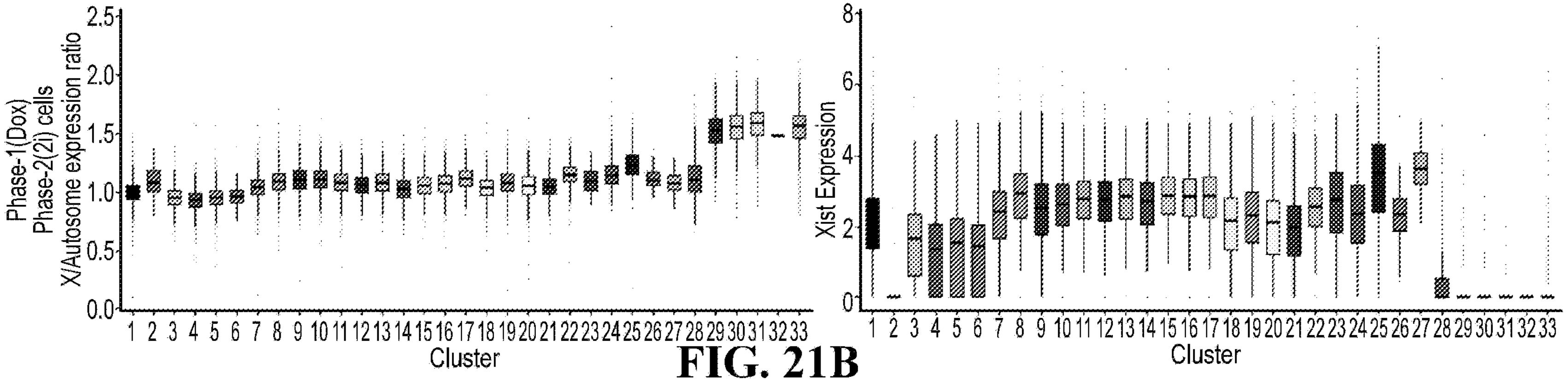


FIG. 21B

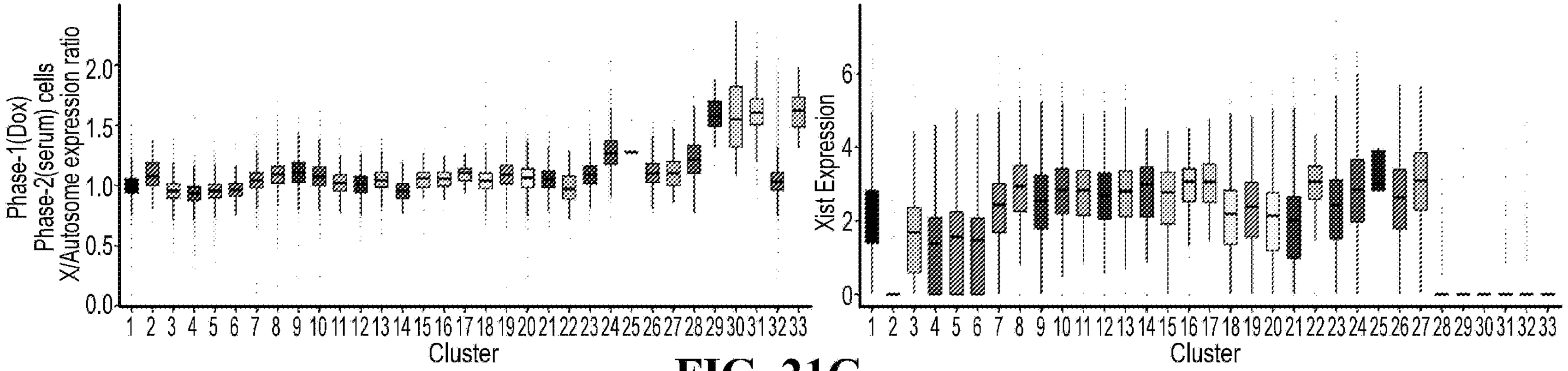


FIG. 21C

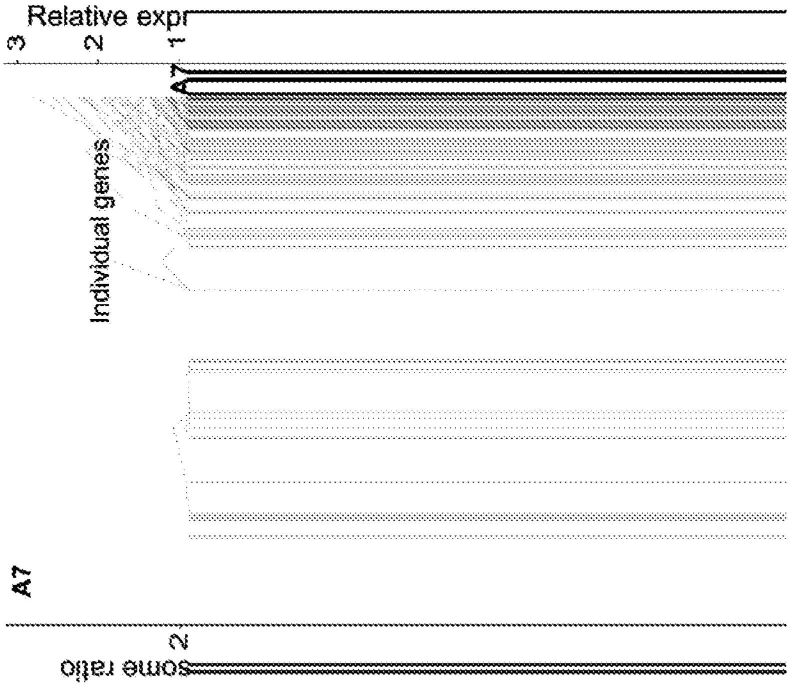


FIG. 21E

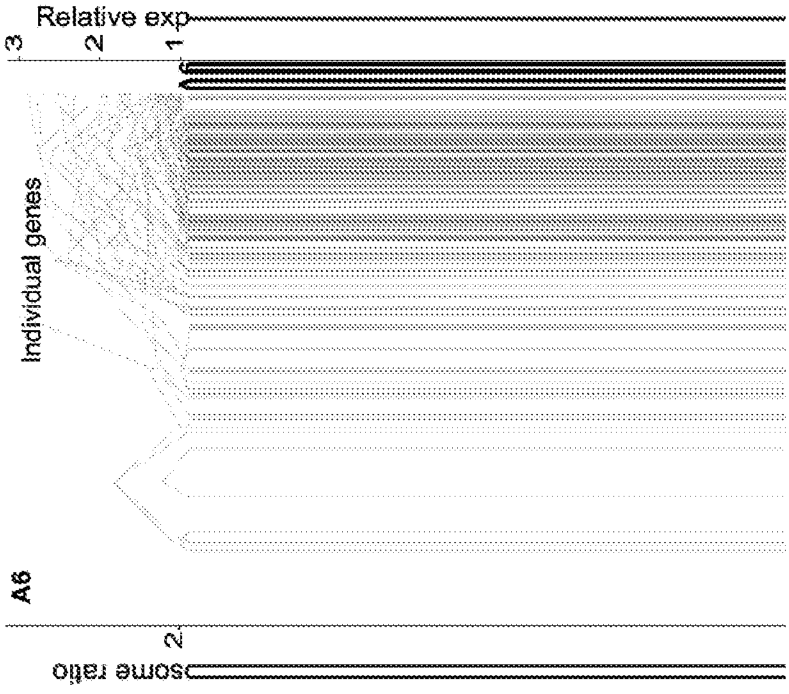


FIG. 21D

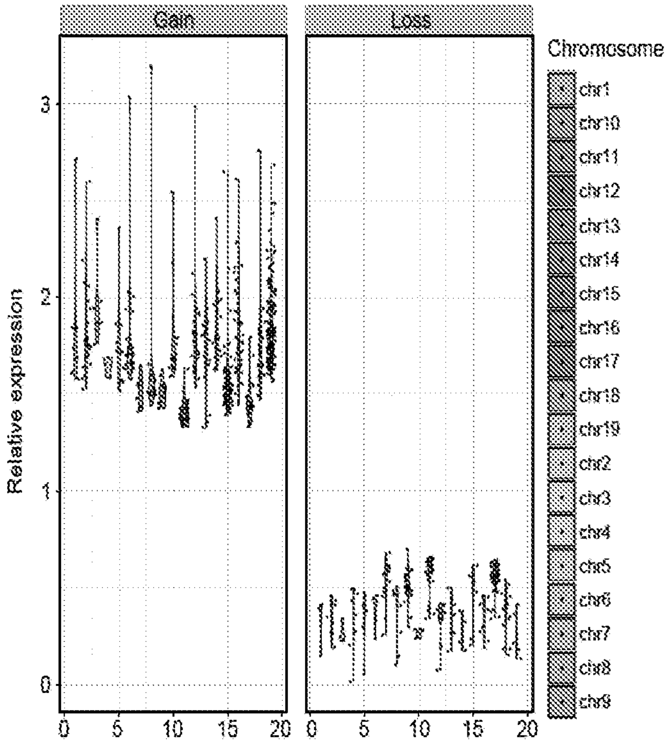


FIG. 22A

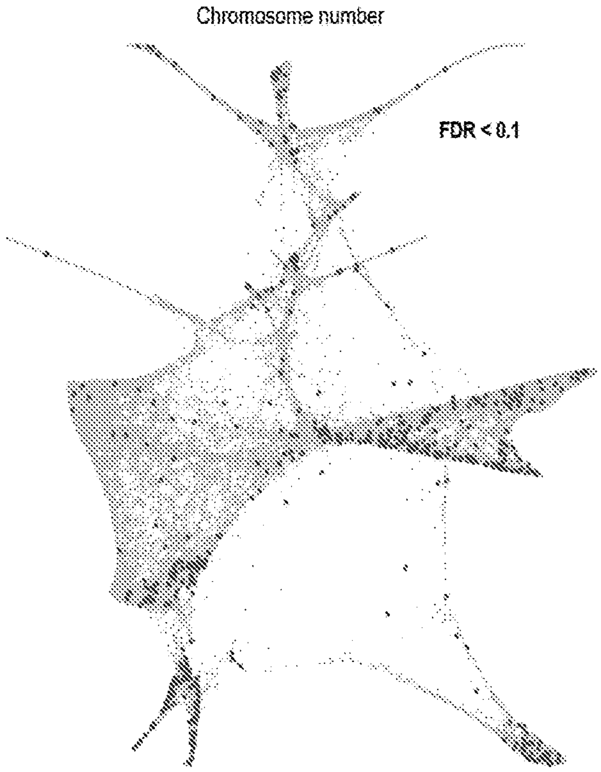


FIG. 22B

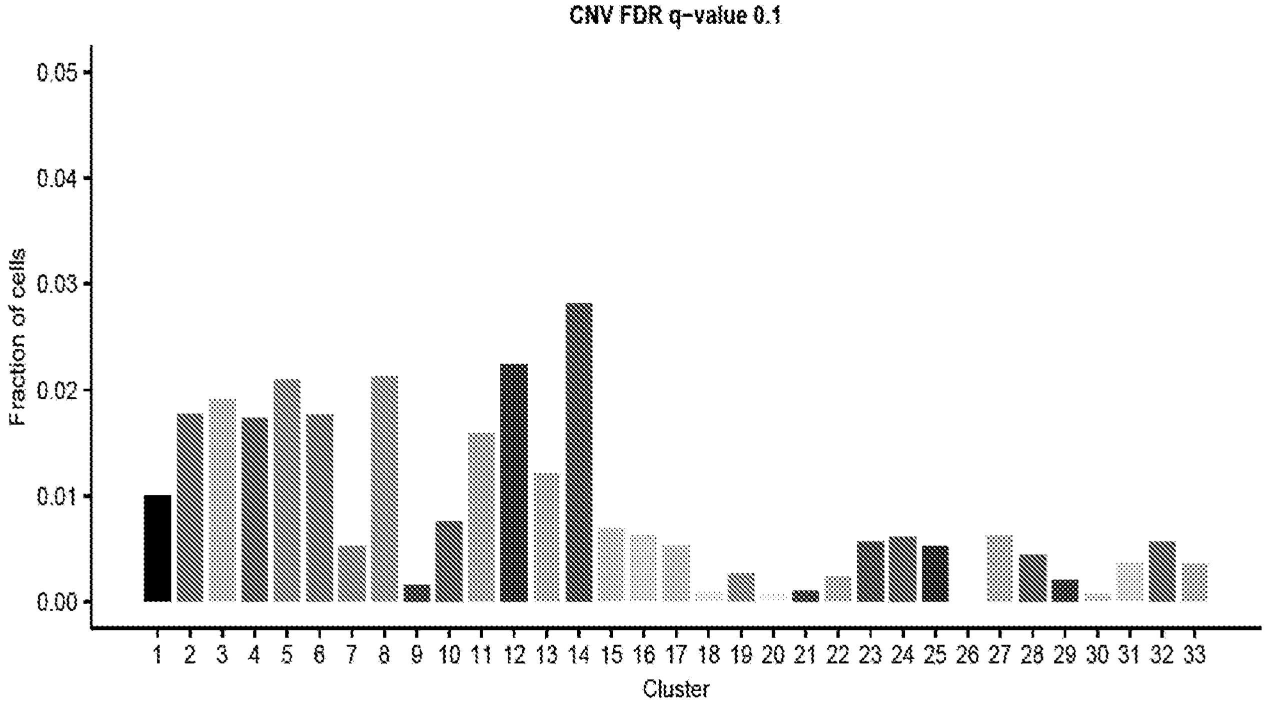


FIG. 22C

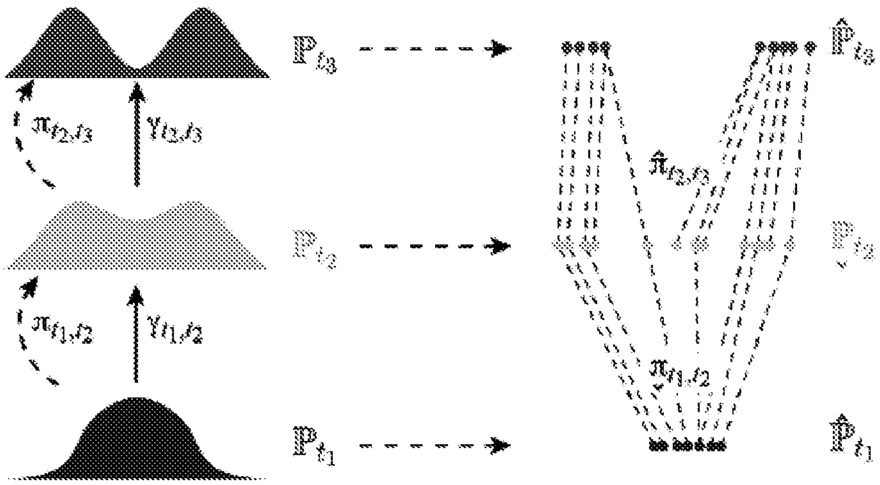


FIG. 23A

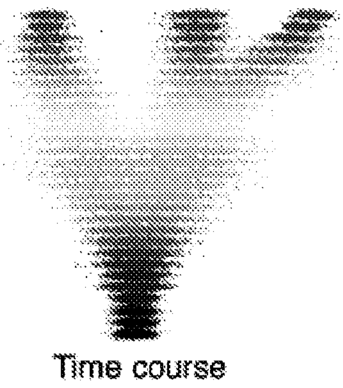
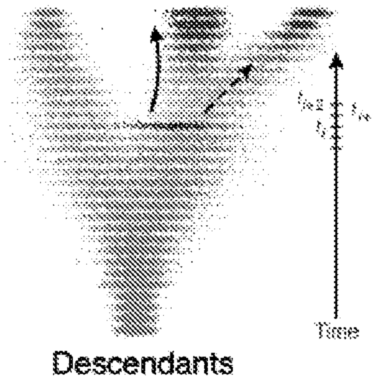
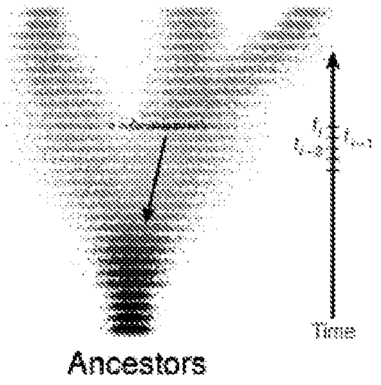

Time course

FIG. 23B



Descendants

FIG. 23C



Ancestors

FIG. 23D

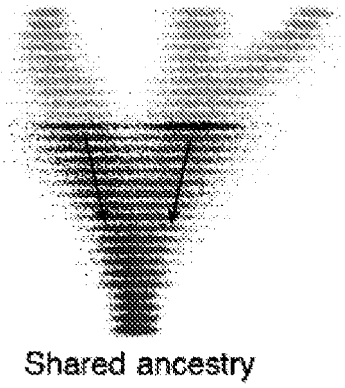
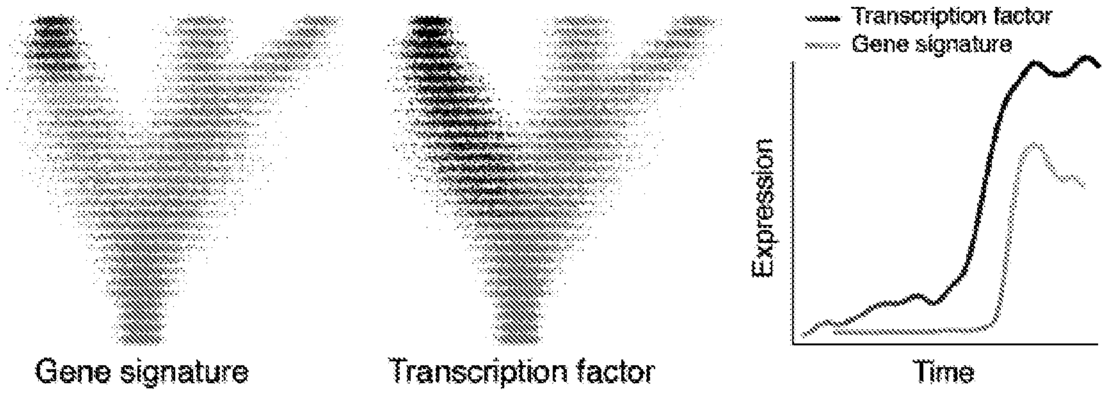

Shared ancestry

FIG. 23E



Gene signature

Transcription factor

FIG. 23F

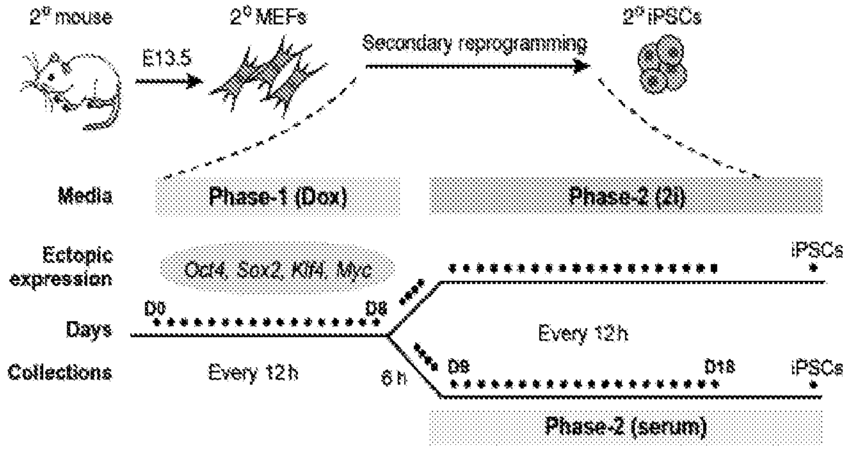


FIG. 24A

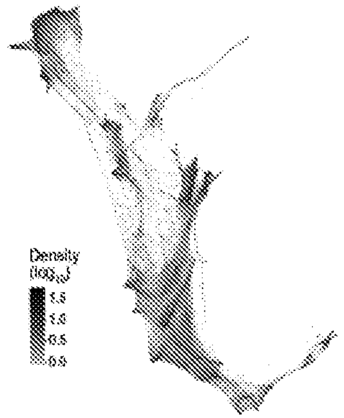


FIG. 24B

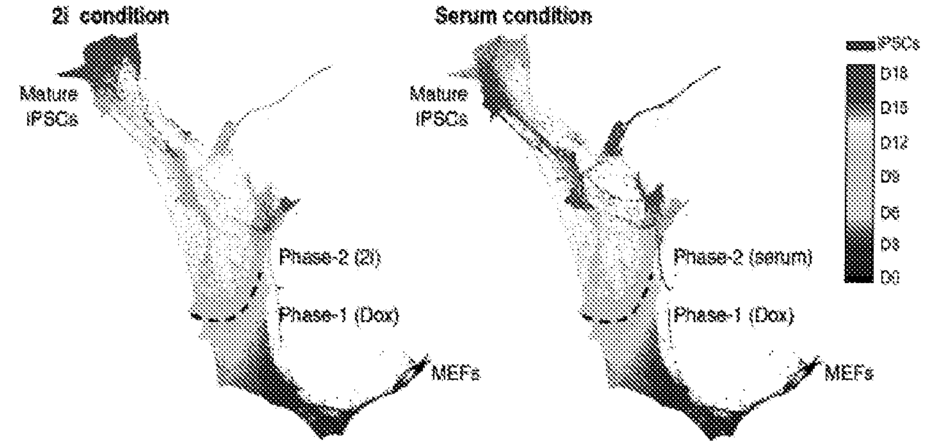


FIG. 24C

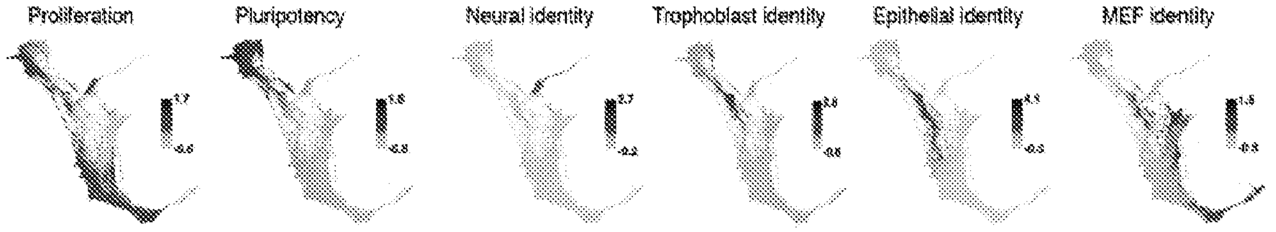


FIG. 24D

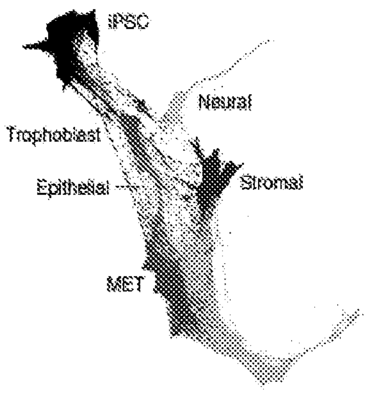


FIG. 24E

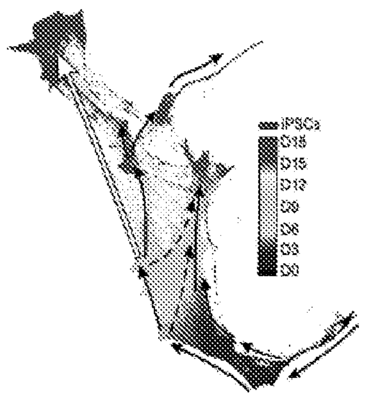


FIG. 24F

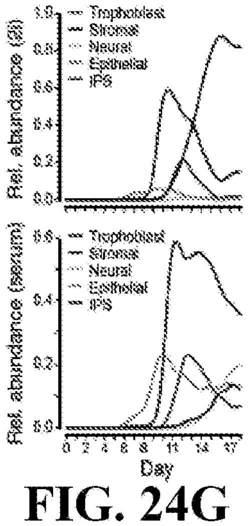


FIG. 24G

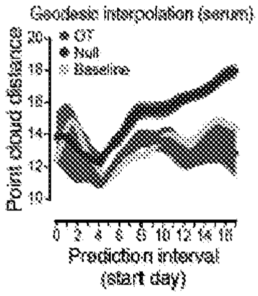


FIG. 24H

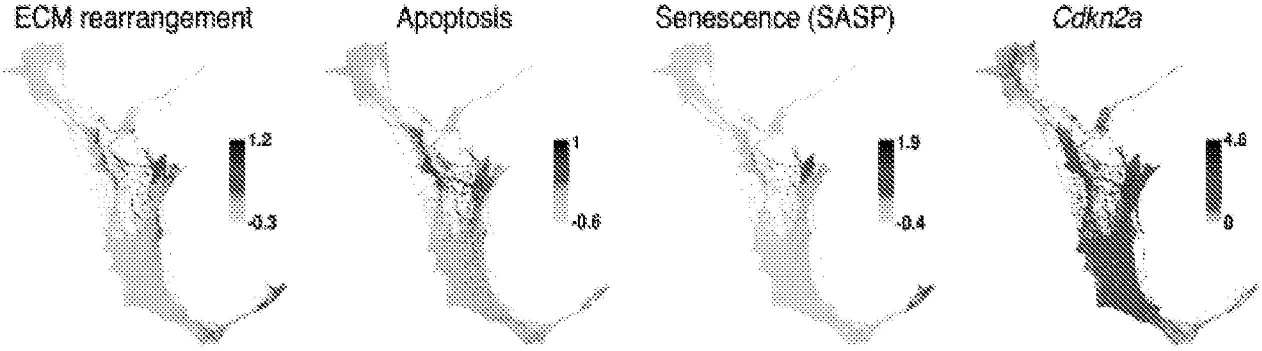


FIG. 25A

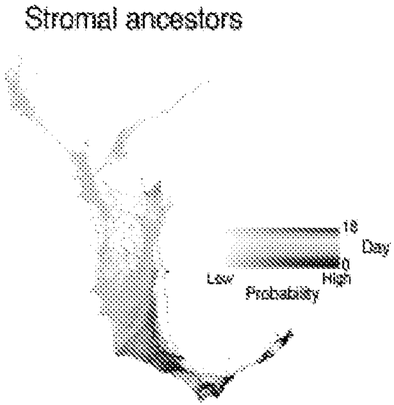


FIG. 25B

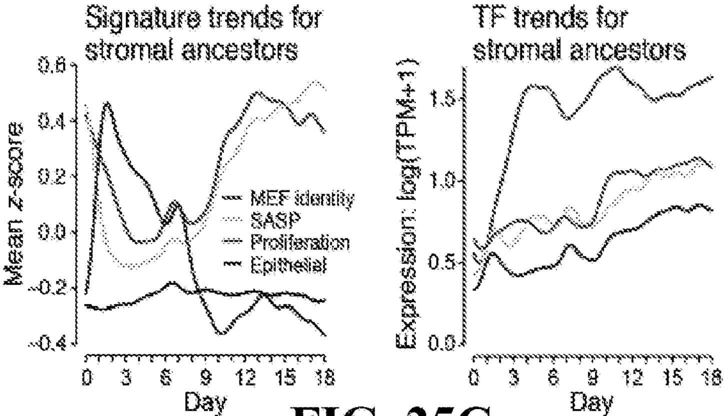


FIG. 25C

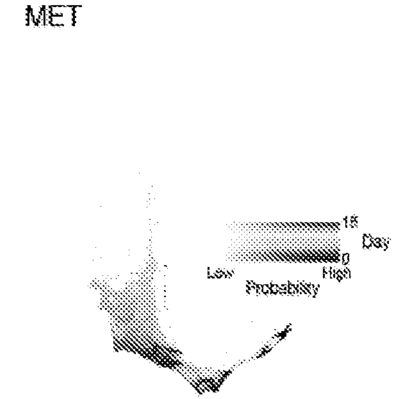


FIG. 25D

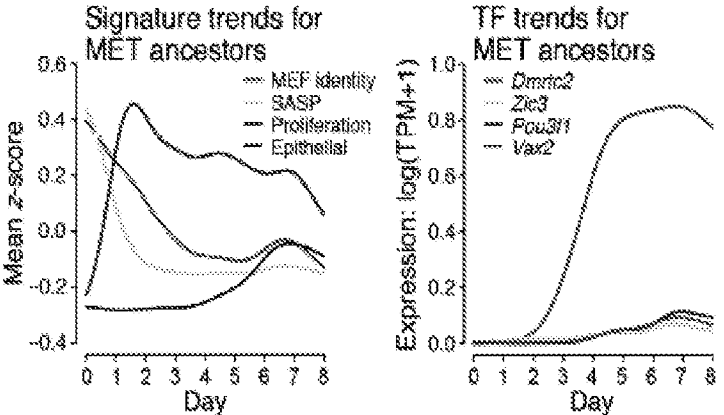


FIG. 25E

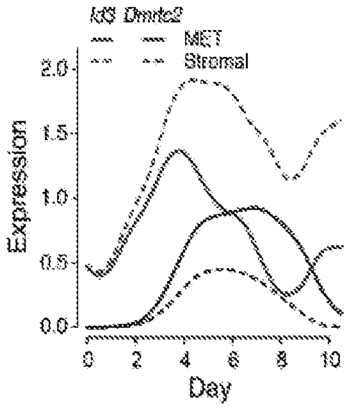


FIG. 25F

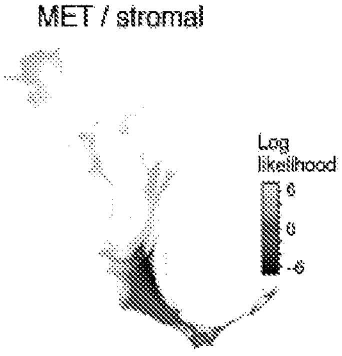


FIG. 25G

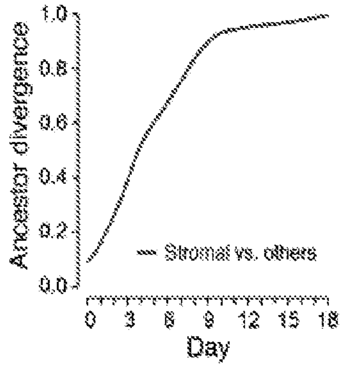


FIG. 25H

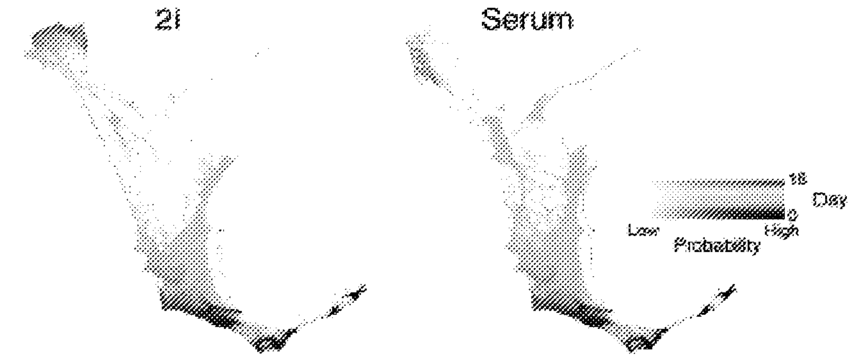


FIG. 26A

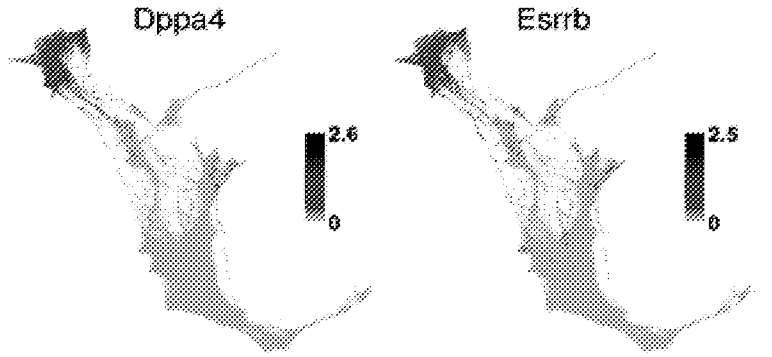


FIG. 26B

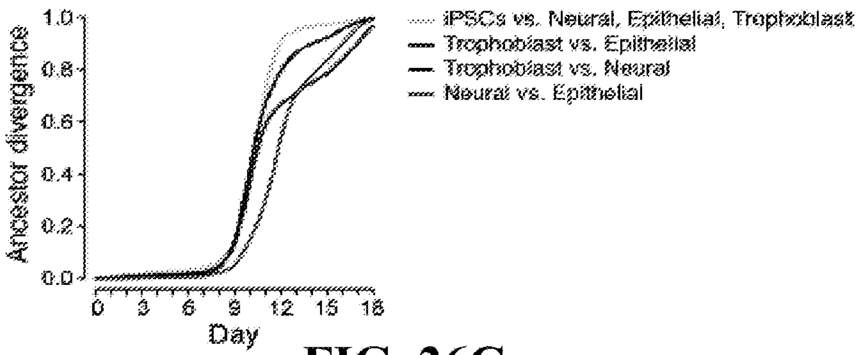


FIG. 26C

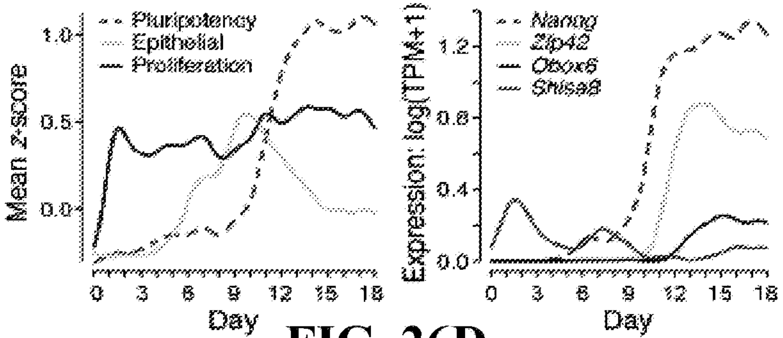


FIG. 26D

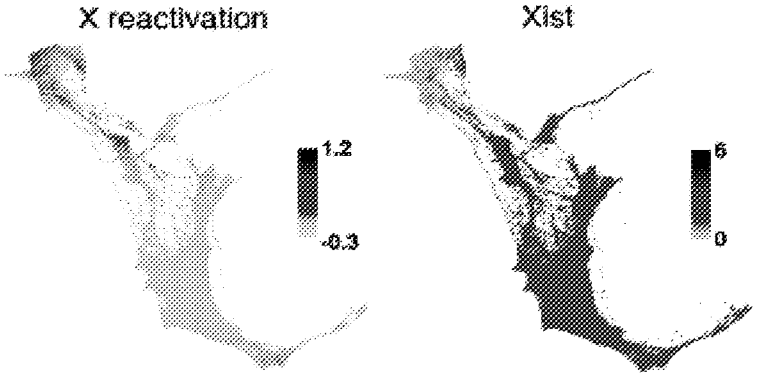


FIG. 26E

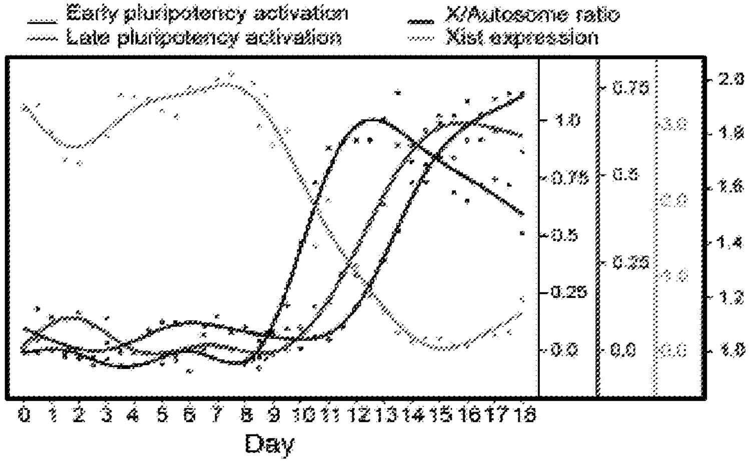


FIG. 26F

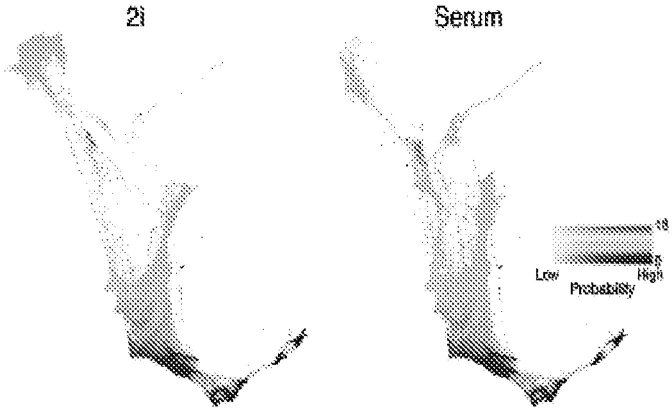


FIG. 27A

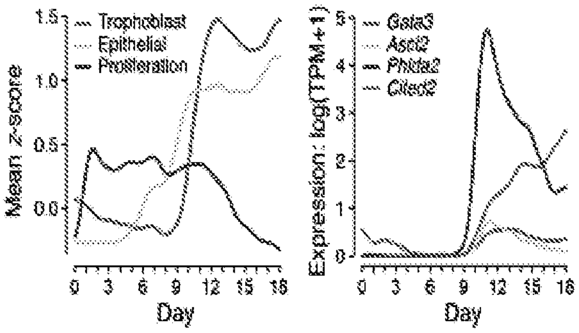


FIG. 27B

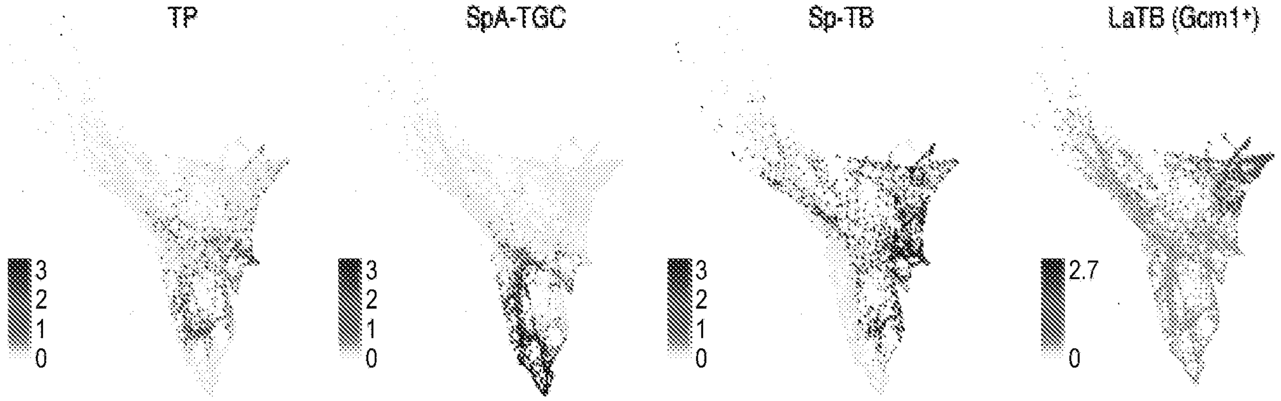


FIG. 27C

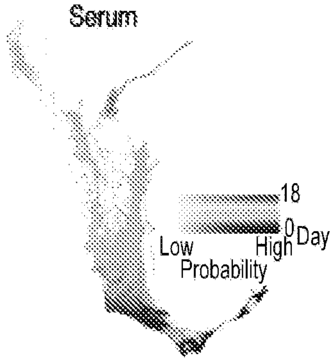


FIG. 27D

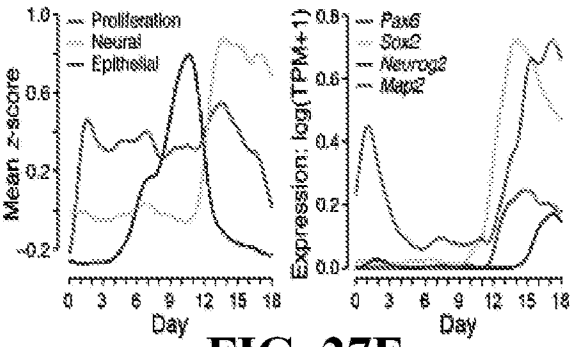


FIG. 27E

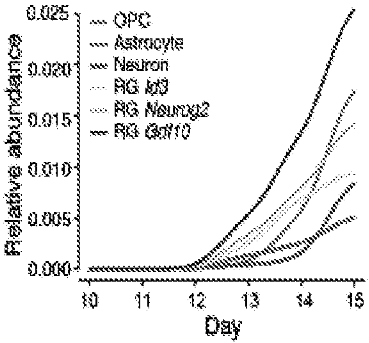


FIG. 27F

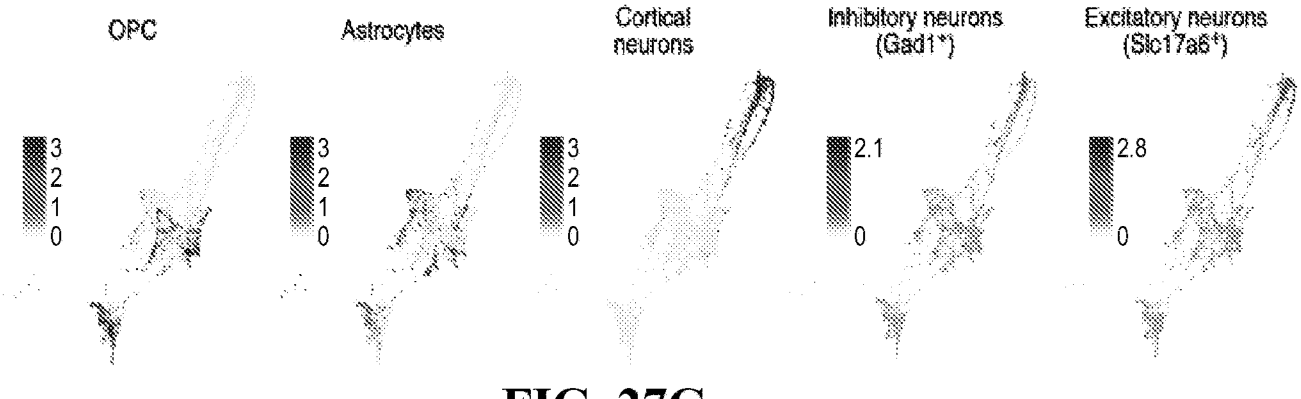


FIG. 27G

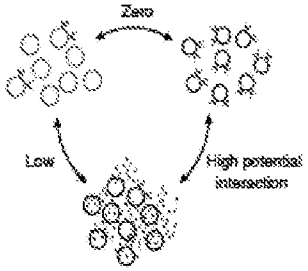


FIG. 28A

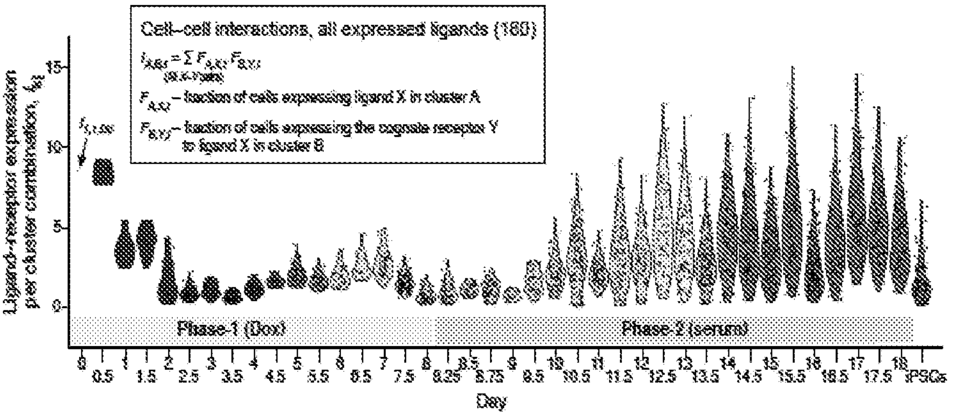


FIG. 28B

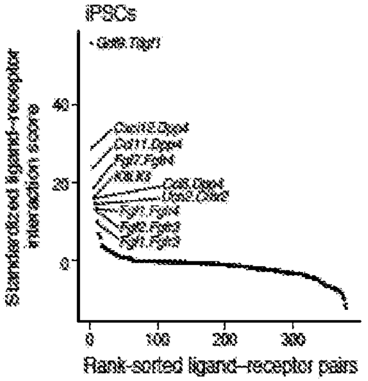


FIG. 28C

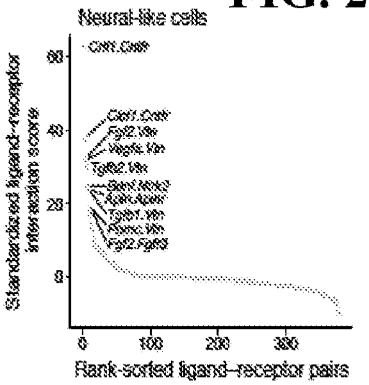


FIG. 28D

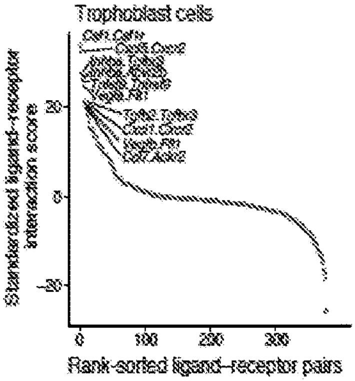


FIG. 28E

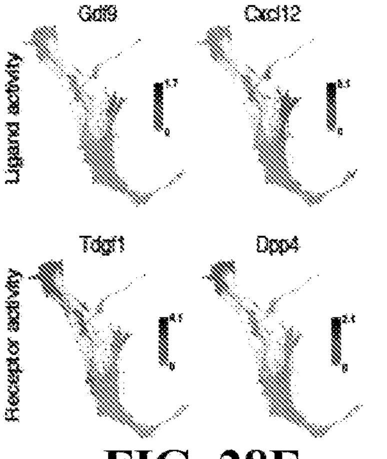


FIG. 28F

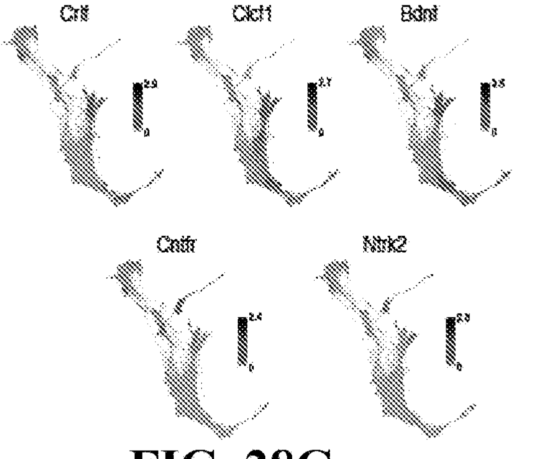


FIG. 28G

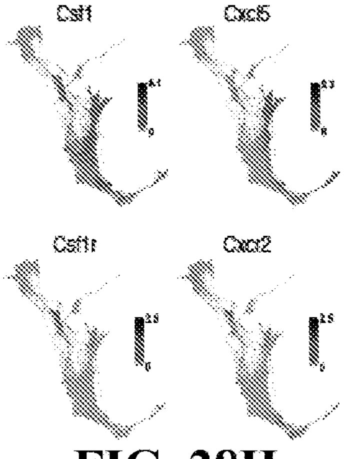


FIG. 28H

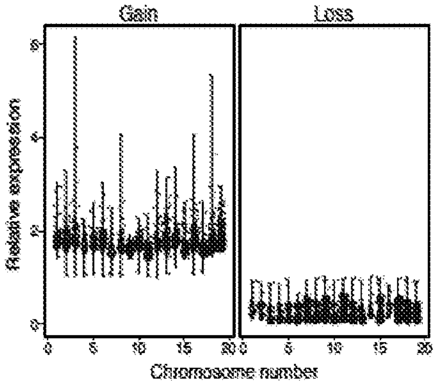


FIG. 28I

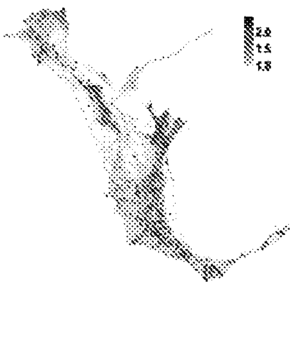


FIG. 28J

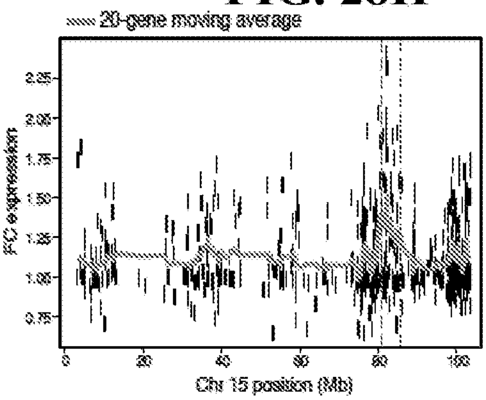


FIG. 28K

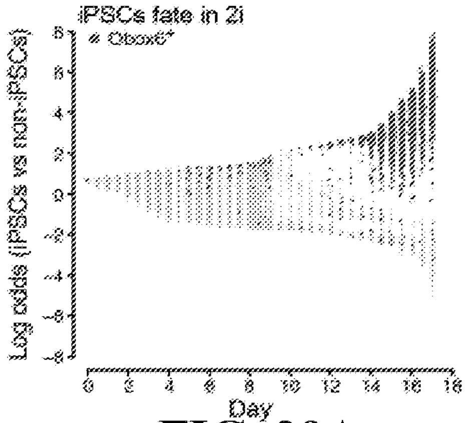


FIG. 29A

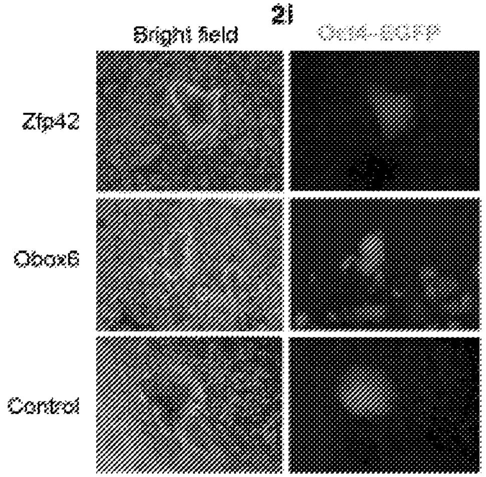


FIG. 29B

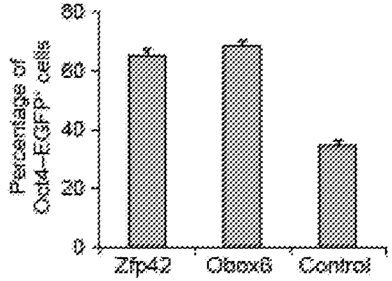


FIG. 29C

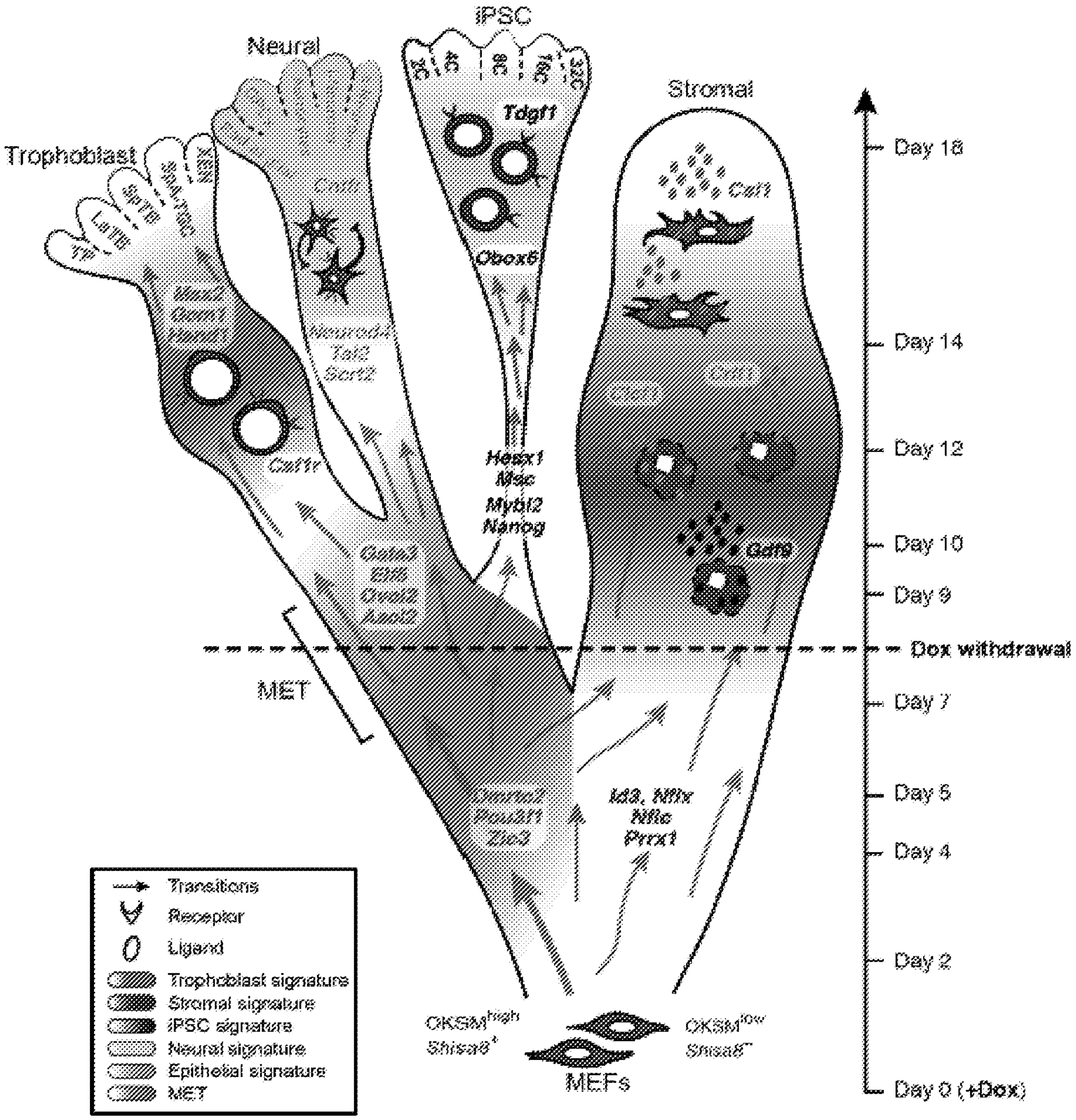


FIG. 29D

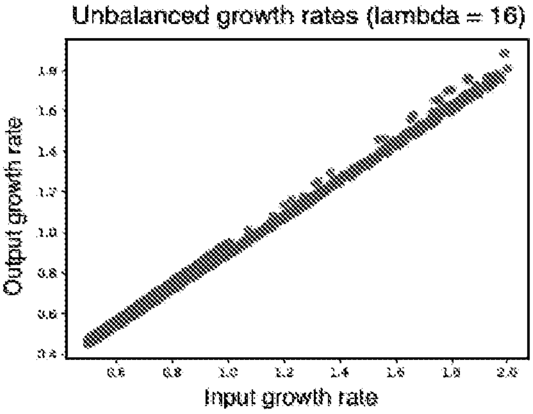


FIG. 30A

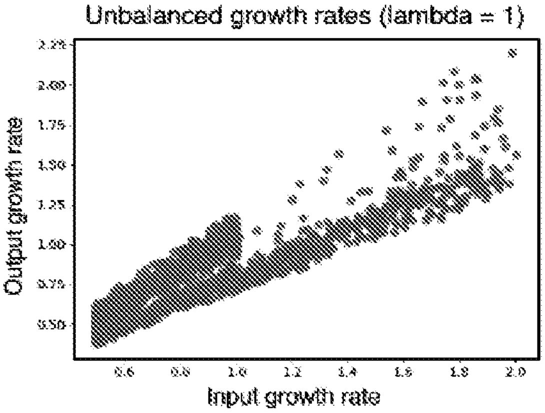


FIG. 30B

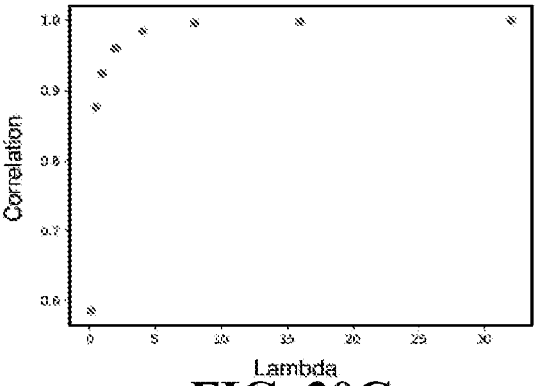


FIG. 30C

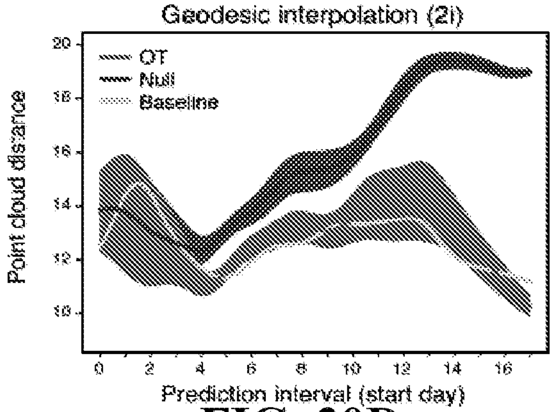


FIG. 30D

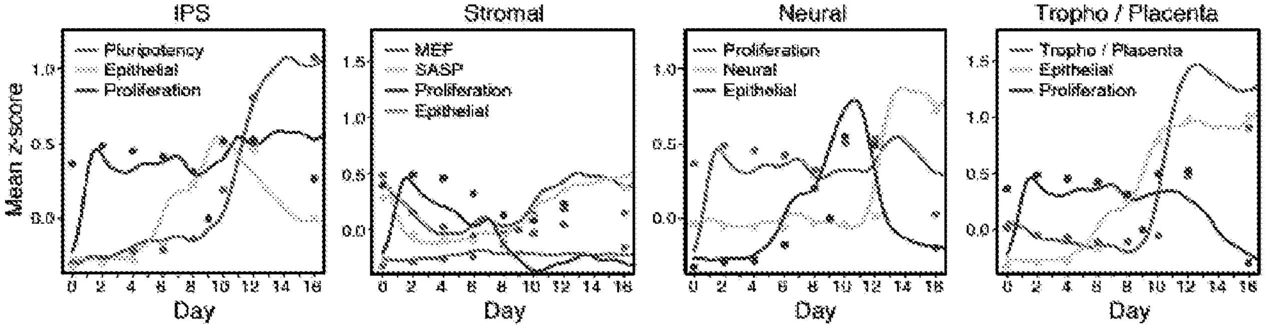


FIG. 30E

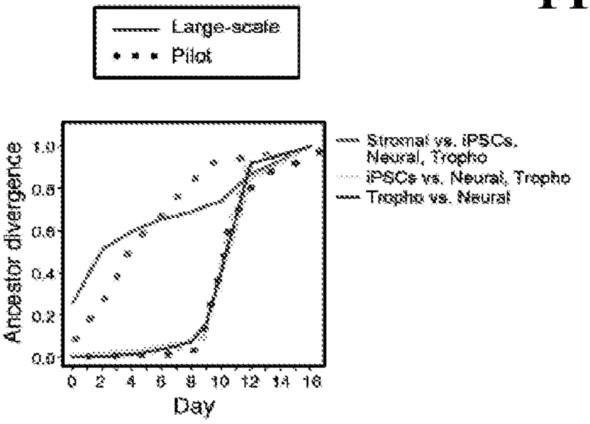


FIG. 30F

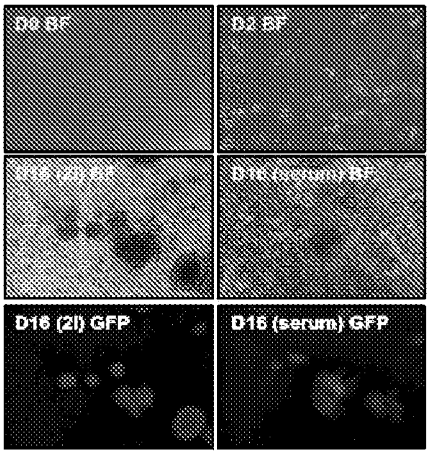


FIG. 30G

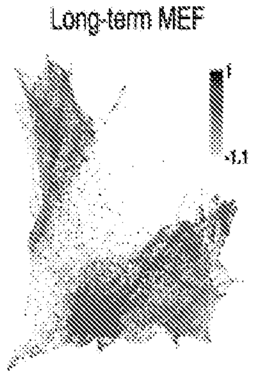


FIG. 31A

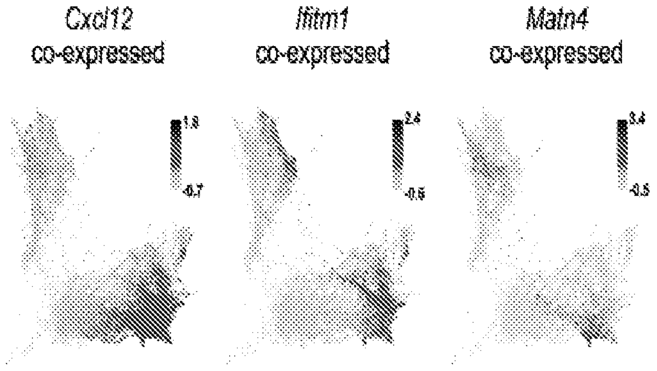


FIG. 31B

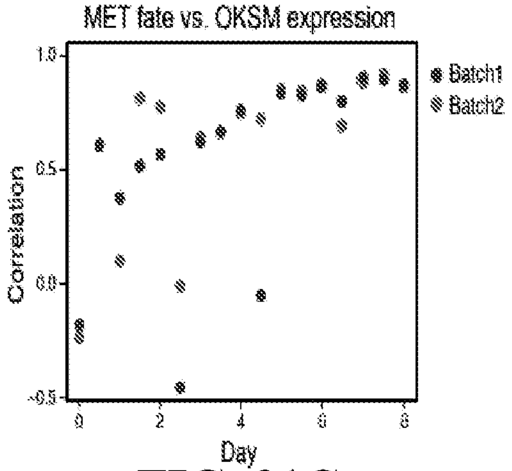


FIG. 31C

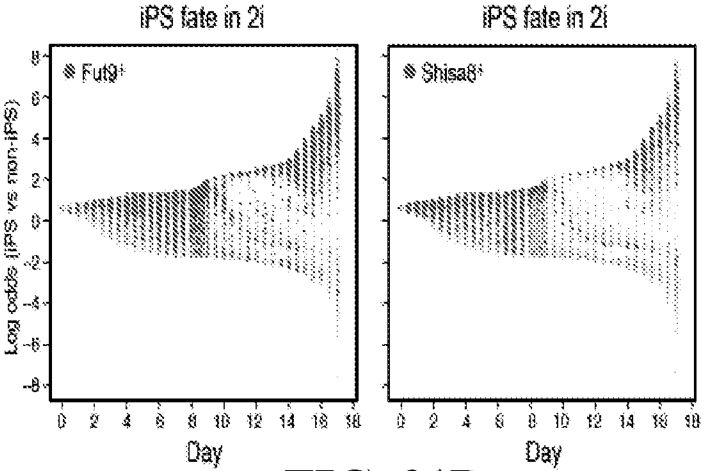


FIG. 31D

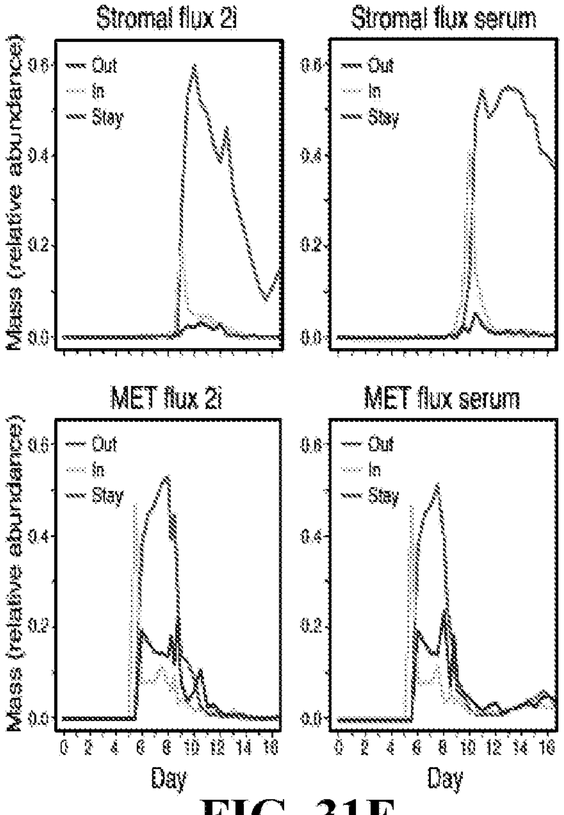


FIG. 31E

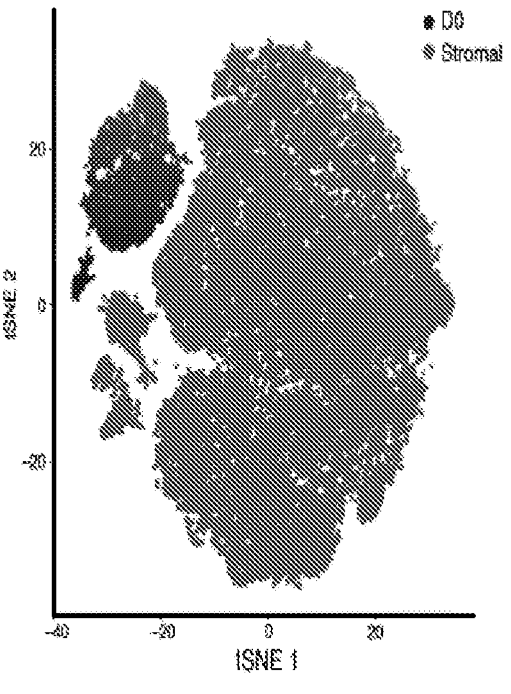


FIG. 31F

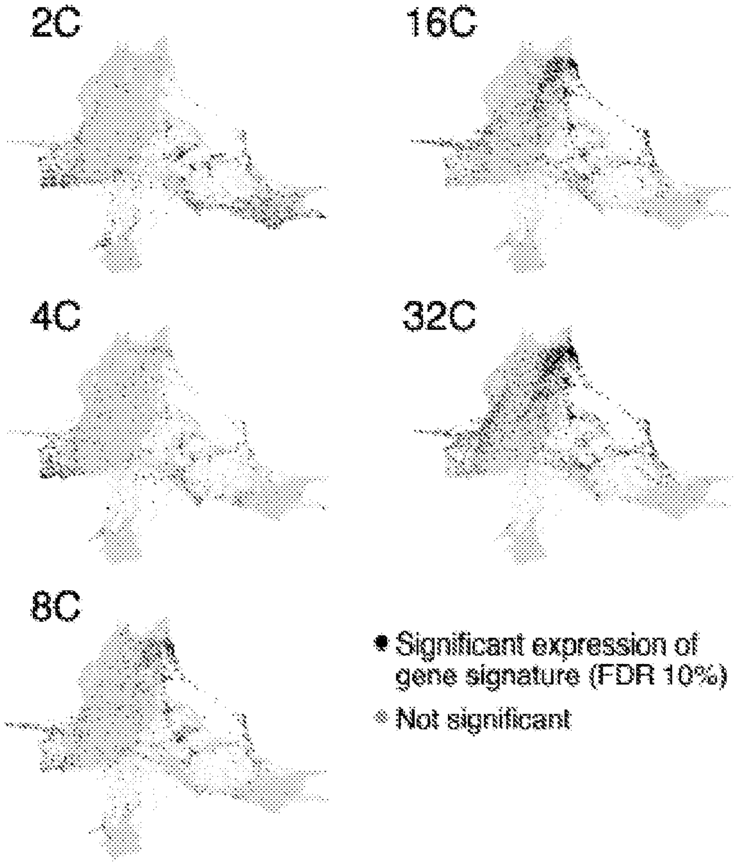


FIG. 32A

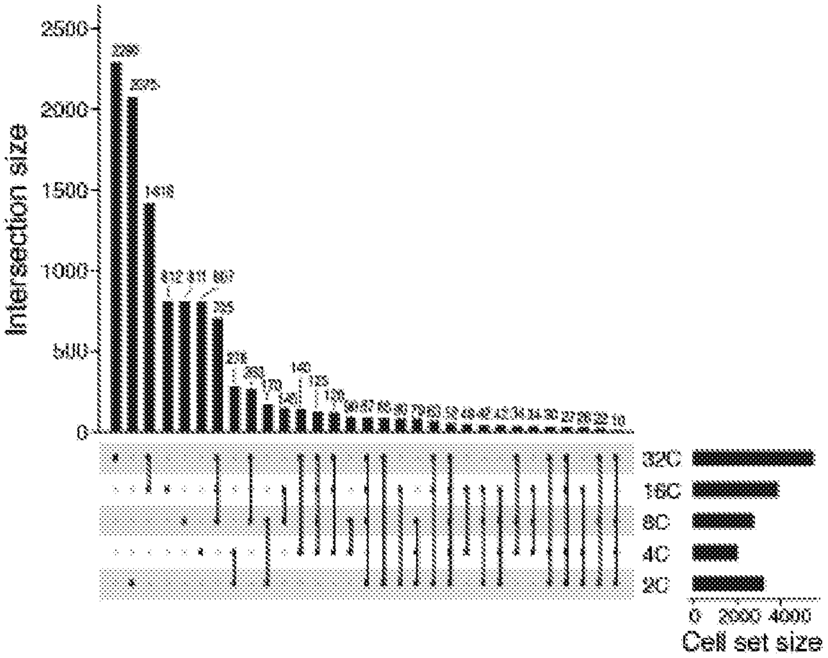


FIG. 32B

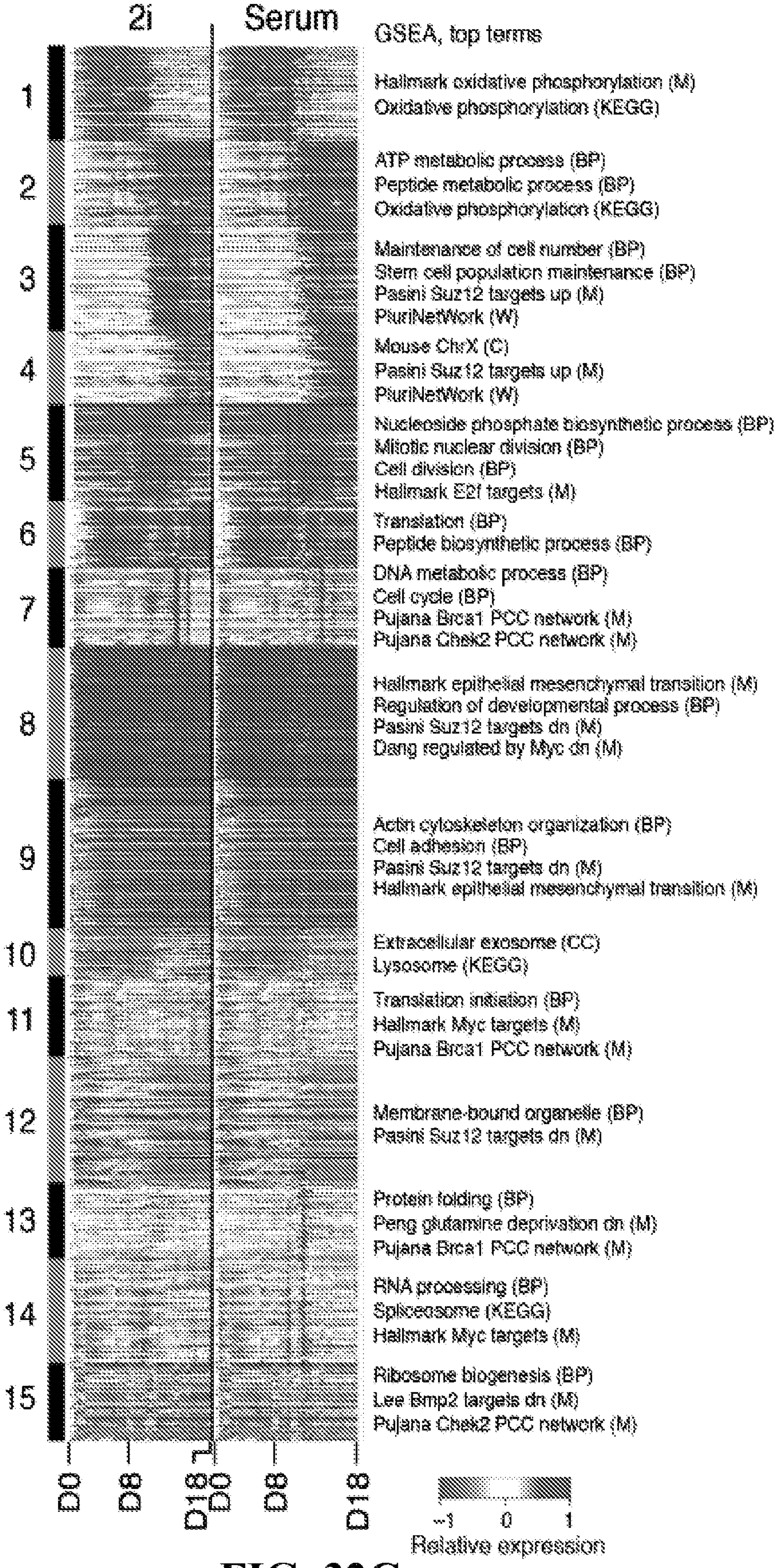

2i
Serum
GSEA, top terms
1
Hallmark oxidative phosphorylation (M)
Oxidative phosphorylation (KEGG)
2
ATP metabolic process (BP)
Peptide metabolic process (BP)
Oxidative phosphorylation (KEGG)
3
Maintenance of cell number (BP)
Stem cell population maintenance (BP)
Pasini Suz12 targets up (M)
PluriNetWork (W)
4
Mouse ChrX (C)
Pasini Suz12 targets up (M)
PluriNetWork (W)
5
Nucleoside phosphate biosynthetic process (BP)
Mitotic nuclear division (BP)
Cell division (BP)
Hallmark E2f targets (M)
6
Translation (BP)
Peptide biosynthetic process (BP)
7
DNA metabolic process (BP)
Cell cycle (BP)
Pujana Brca1 PCC network (M)
Pujana Chek2 PCC network (M)
8
Hallmark epithelial mesenchymal transition (M)
Regulation of developmental process (BP)
Pasini Suz12 targets dn (M)
Dang regulated by Myc dn (M)
9
Actin cytoskeleton organization (BP)
Cell adhesion (BP)
Pasini Suz12 targets dn (M)
Hallmark epithelial mesenchymal transition (M)
10
Extracellular exosome (CC)
Lysosome (KEGG)
11
Translation initiation (BP)
Hallmark Myc targets (M)
Pujana Brca1 PCC network (M)
12
Membrane-bound organelle (BP)
Pasini Suz12 targets dn (M)
13
Protein folding (BP)
Peng glutamine deprivation dn (M)
Pujana Brca1 PCC network (M)
14
RNA processing (BP)
Spliceosome (KEGG)
Hallmark Myc targets (M)
15
Ribosome biogenesis (BP)
Lee Bmp2 targets dn (M)
Pujana Chek2 PCC network (M)
D0
D8
D18
D0
D8
D18
−1
0
1
Relative expression

FIG. 32C

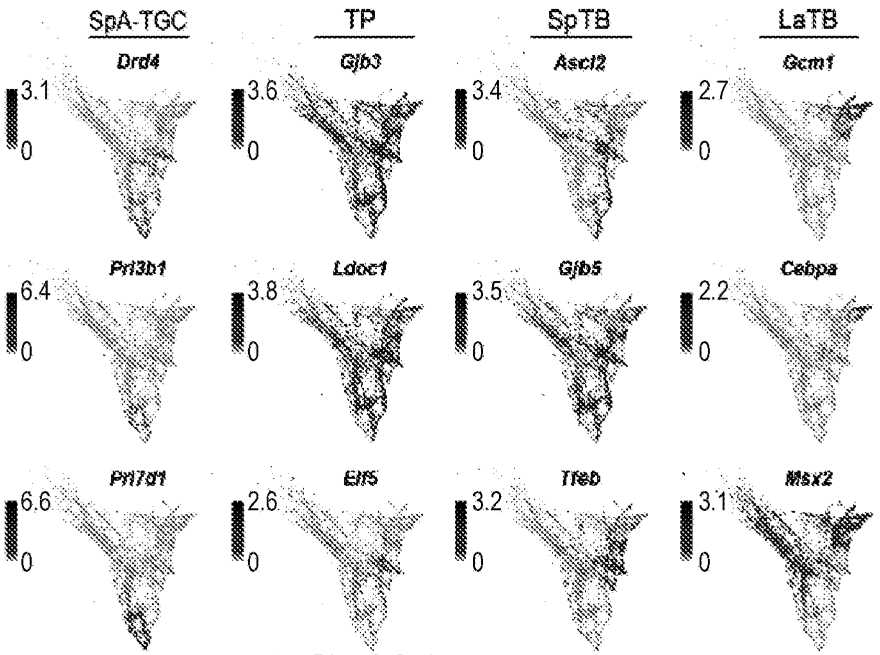


FIG. 33A

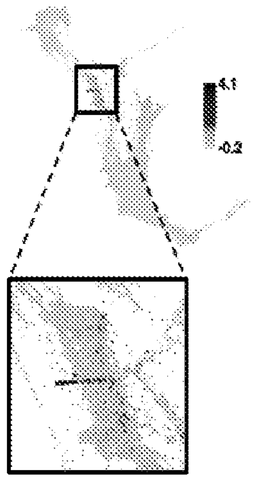


FIG. 33B

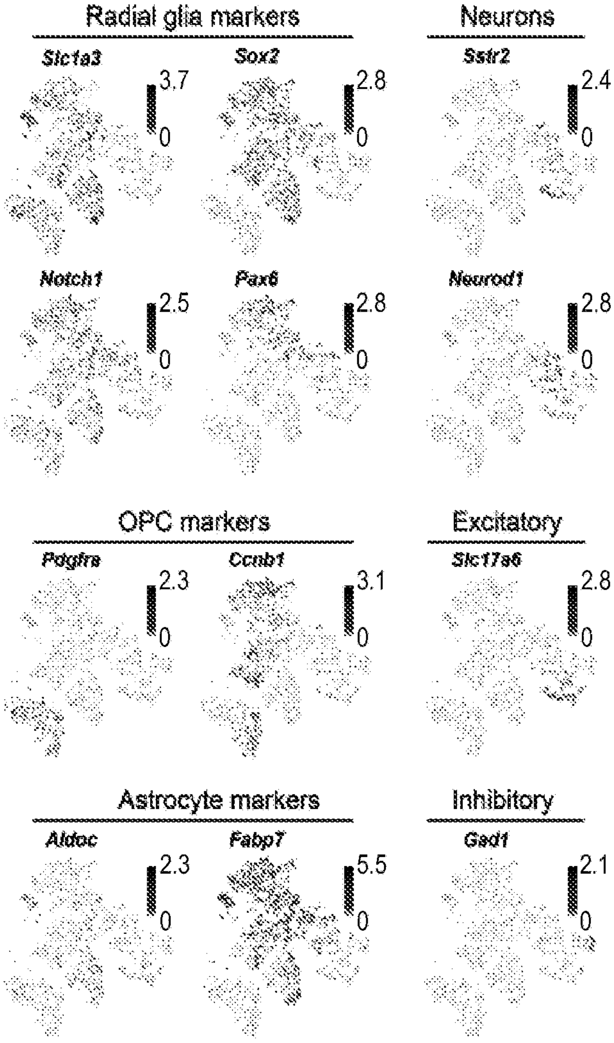


FIG. 33C

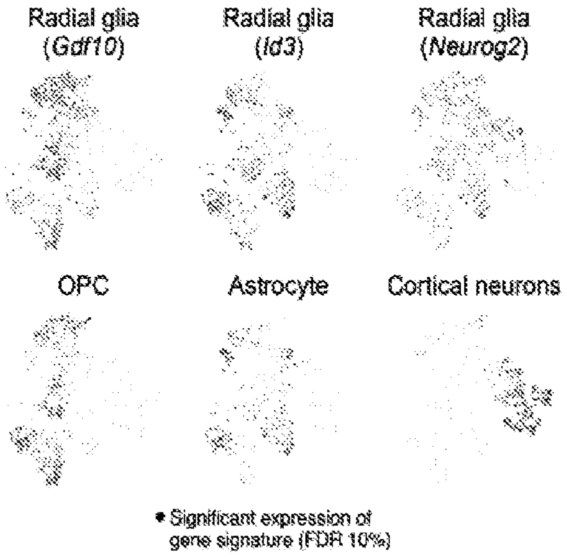


FIG. 33D

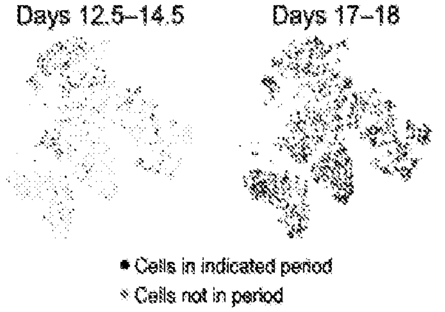


FIG. 33E

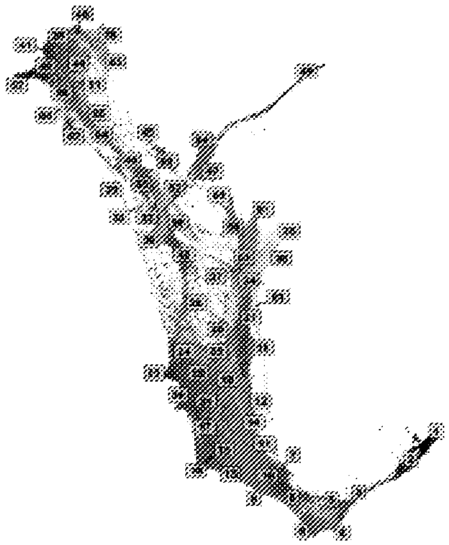


FIG. 34A

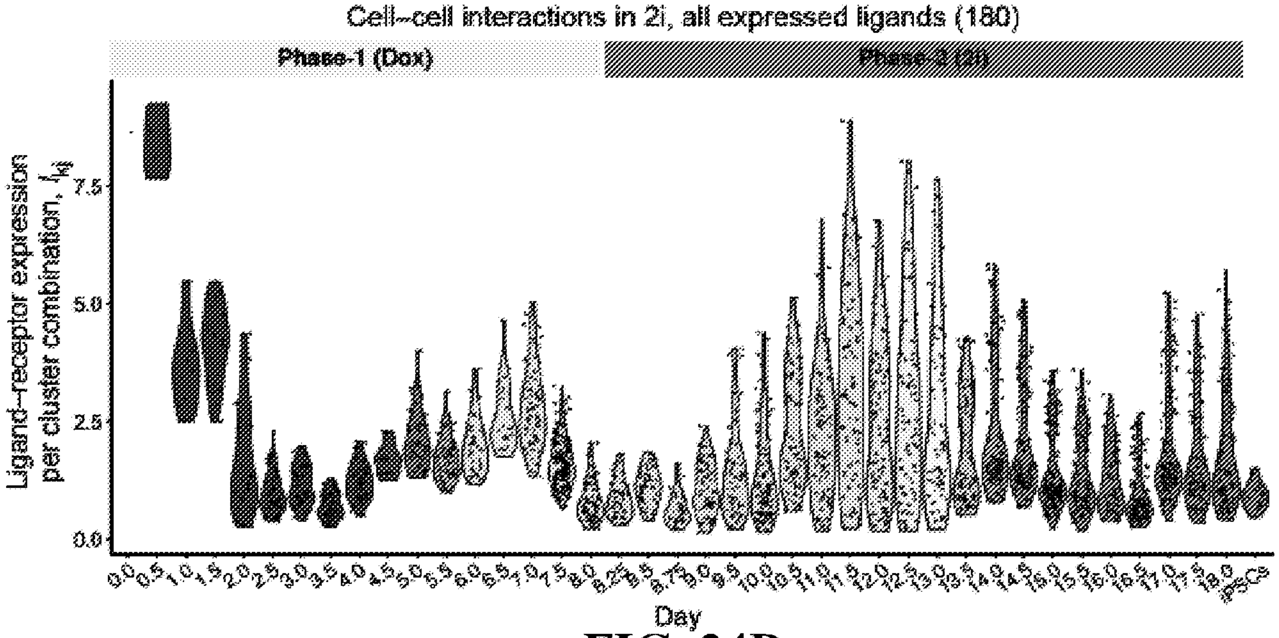


FIG. 34B

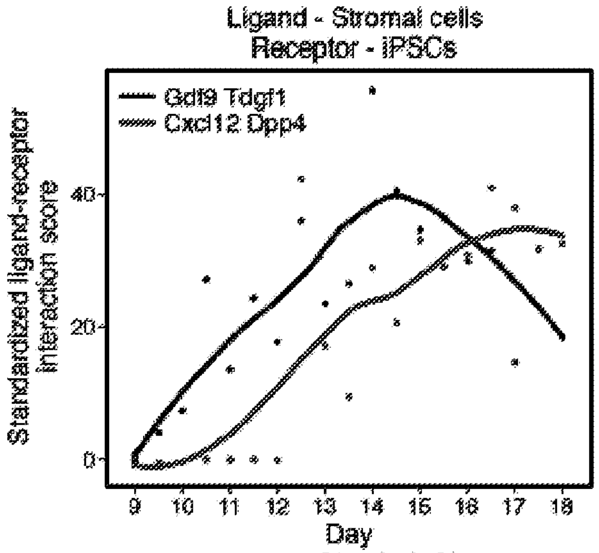


FIG. 34C

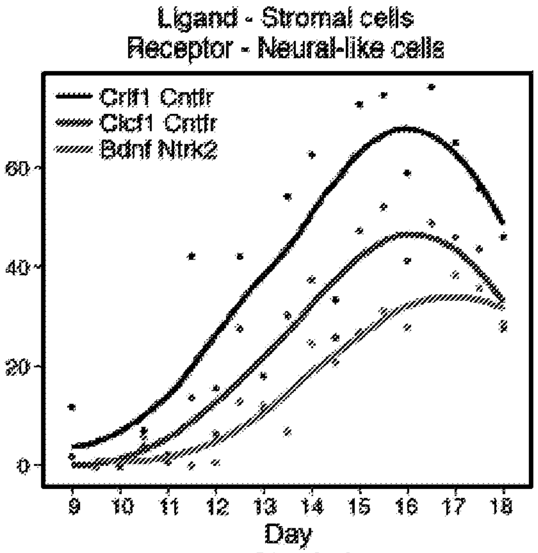


FIG. 34D

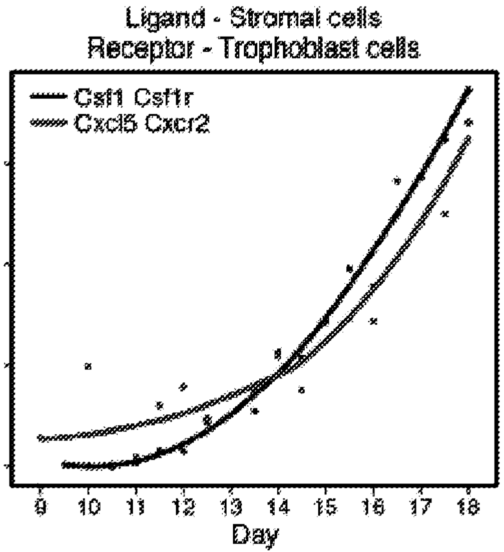


FIG. 34E

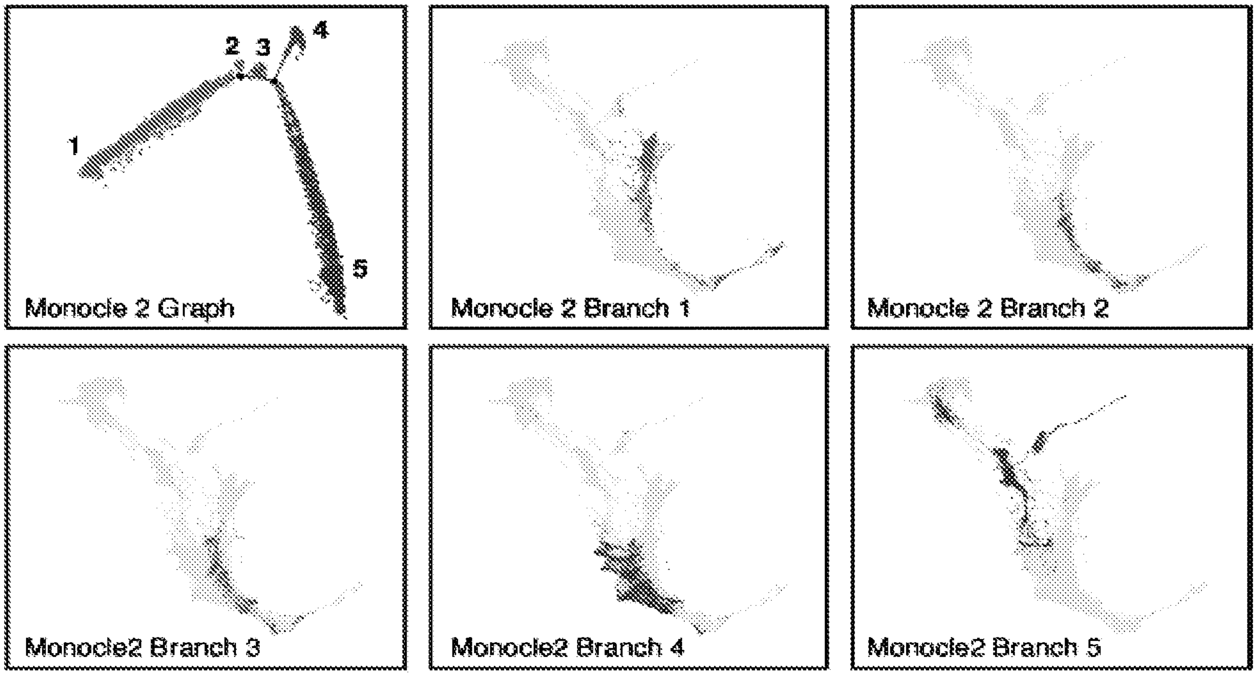


FIG. 35A

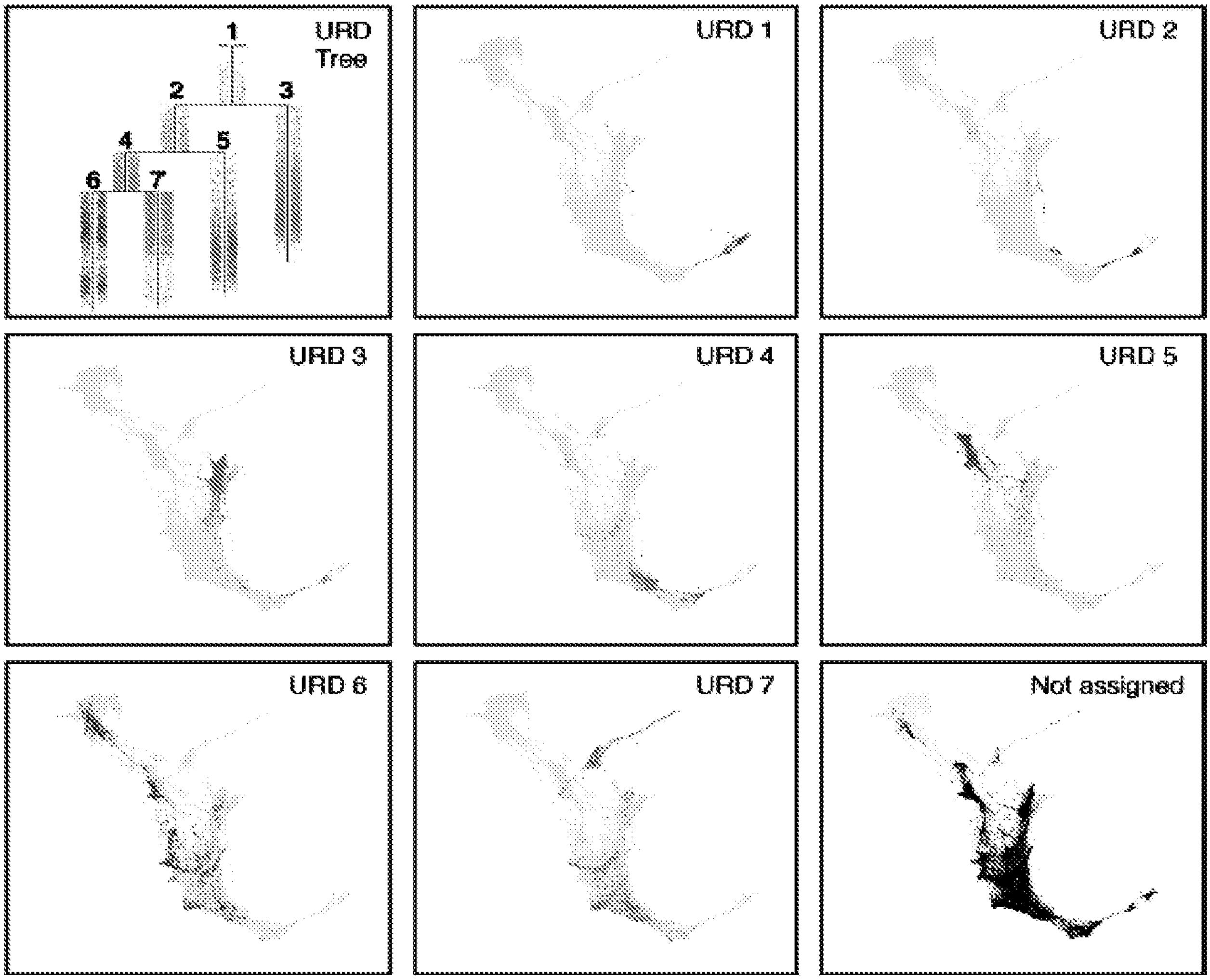


FIG. 35B

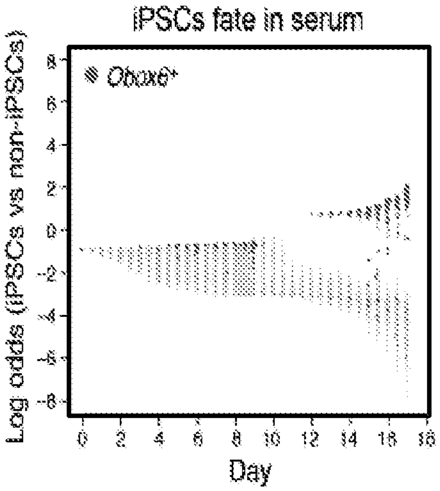


FIG. 36A

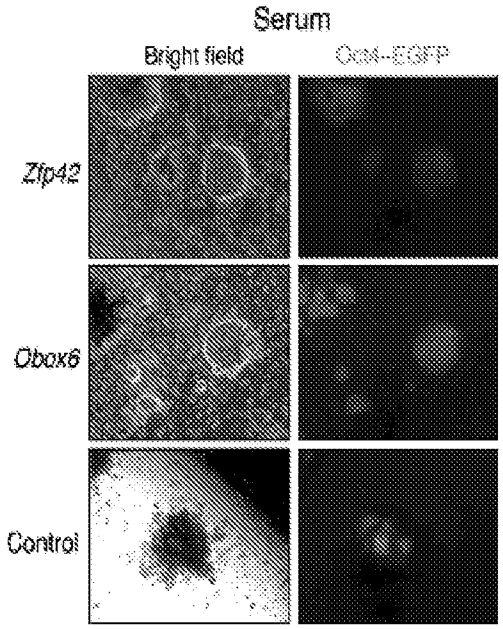


FIG. 36B

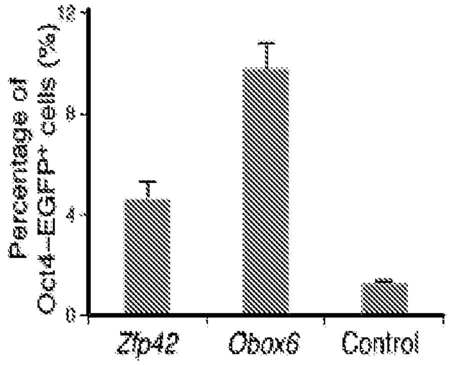


FIG. 36C

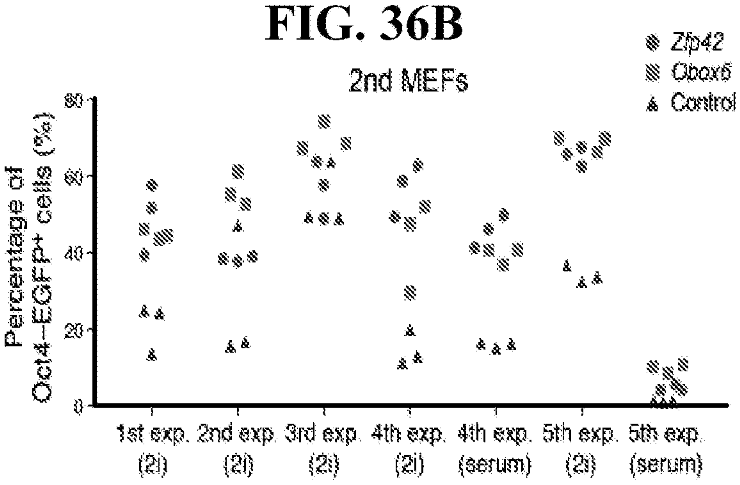


FIG. 36D

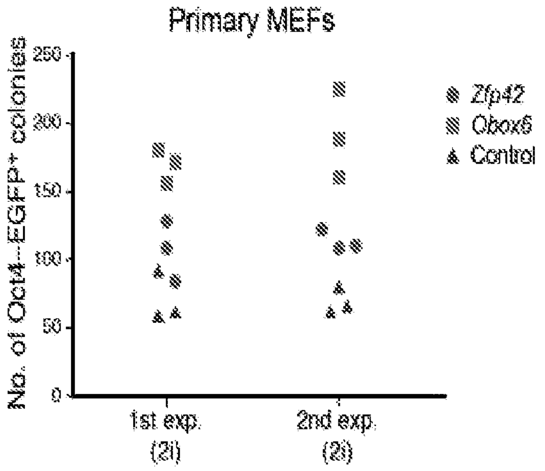


FIG. 36E

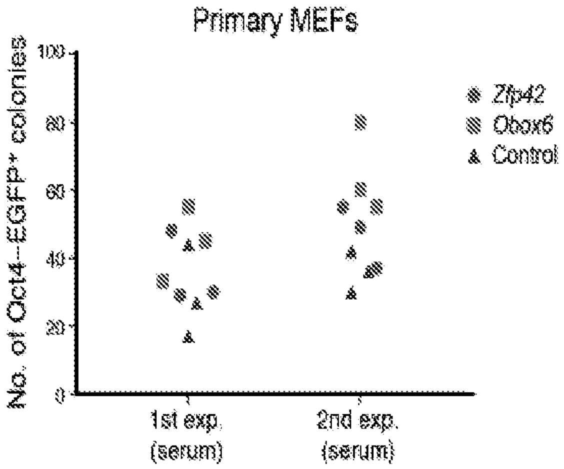


FIG. 36F

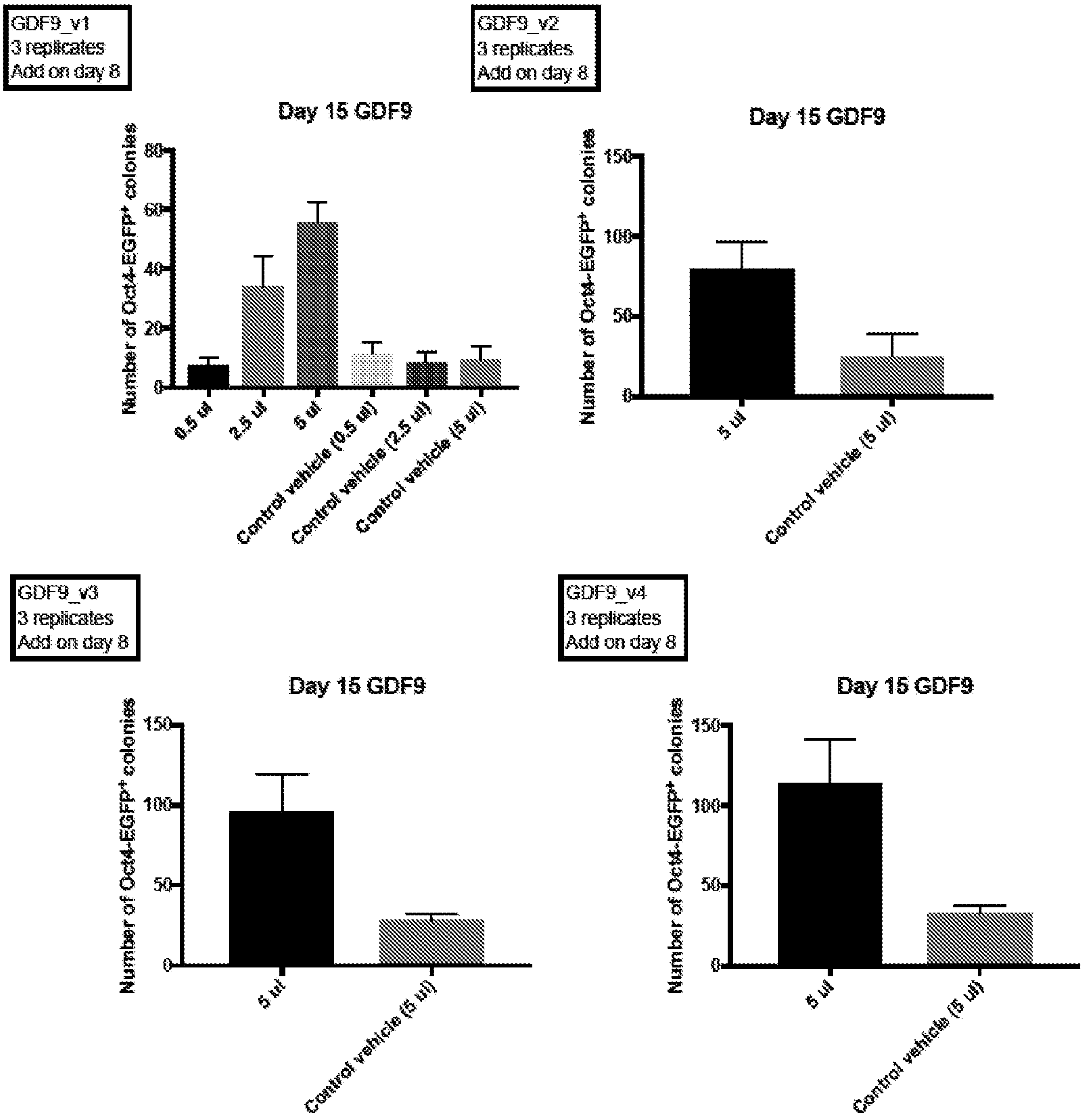

GDF9_v1
3 replicates
Add on day 8
Day 15 GDF9
Number of Oct4-EGFP+ colonies
80
60
40
20
0
0.5 ul
2.5 ul
5 ul
Control vehicle (0.5 ul)
Control vehicle (2.5 ul)
Control vehicle (5 ul)
GDF9_v2
3 replicates
Add on day 8
Day 15 GDF9
Number of Oct4-EGFP+ colonies
150
100
50
0
5 ul
Control vehicle (5 ul)
GDF9_v3
3 replicates
Add on day 8
Day 15 GDF9
Number of Oct4-EGFP+ colonies
150
100
50
0
5 ul
Control vehicle (5 ul)
GDF9_v4
3 replicates
Add on day 8
Day 15 GDF9
Number of Oct4-EGFP+ colonies
150
100
50
0
5 ul
Control vehicle (5 ul)

FIG. 37

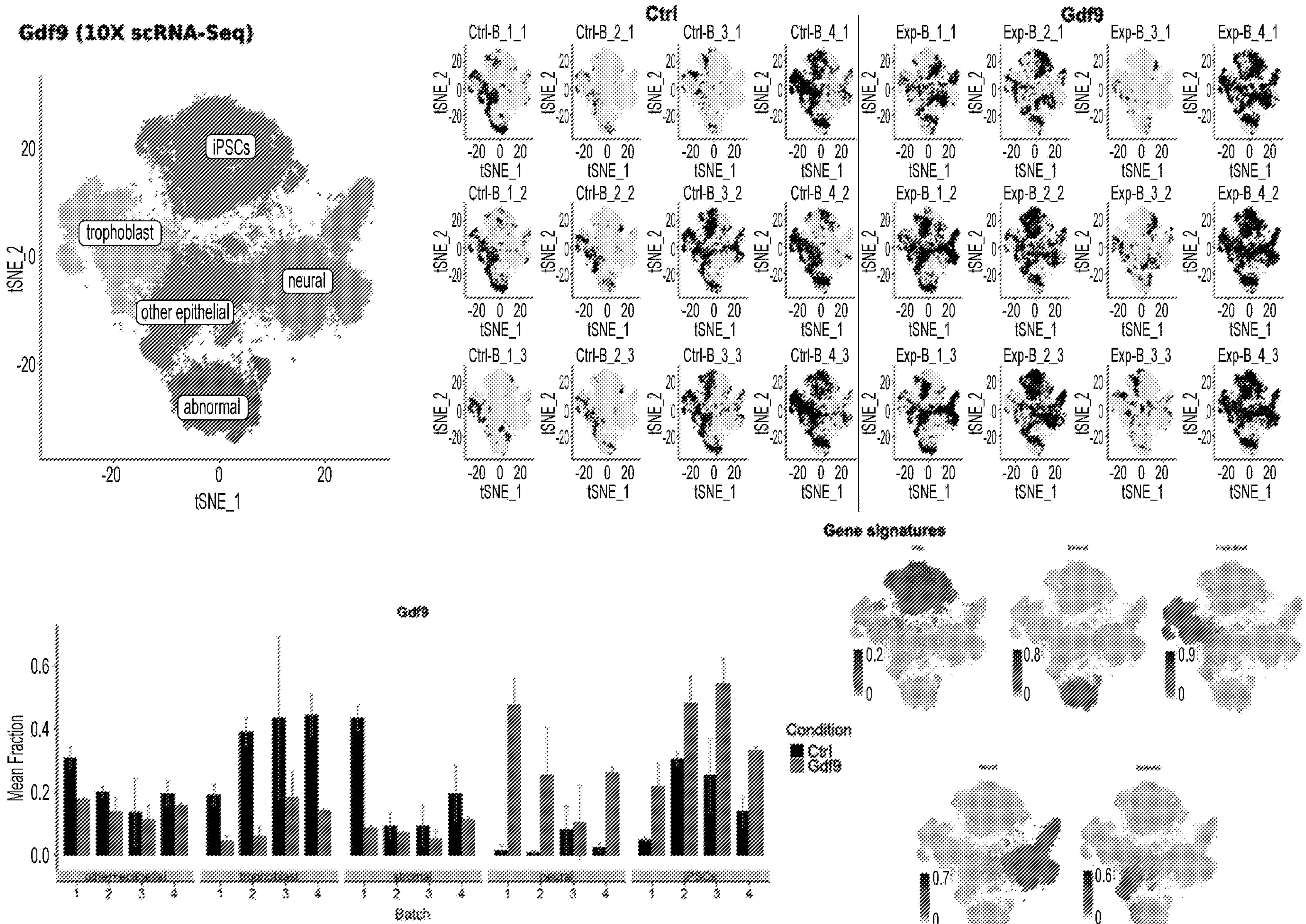

Gdf9 (10X scRNA-Seq)
iPSCs
trophoblast
neural
other epithelial
abnormal
tSNE_2
tSNE_1
20
0
-20
Ctrl
Gdf9
Ctrl-B_1_1
Ctrl-B_2_1
Ctrl-B_3_1
Ctrl-B_4_1
Exp-B_1_1
Exp-B_2_1
Exp-B_3_1
Exp-B_4_1
Ctrl-B_1_2
Ctrl-B_2_2
Ctrl-B_3_2
Ctrl-B_4_2
Exp-B_1_2
Exp-B_2_2
Exp-B_3_2
Exp-B_4_2
Ctrl-B_1_3
Ctrl-B_2_3
Ctrl-B_3_3
Ctrl-B_4_3
Exp-B_1_3
Exp-B_2_3
Exp-B_3_3
Exp-B_4_3
Gene signatures
0.2
0.8
0.9
0.7
0.6
0
Gdf9
Mean Fraction
0.0
0.2
0.4
0.6
other epithelial
trophoblast
neural
iPSCs
1
2
3
4
Batch
Condition
Ctrl
Gdf9

FIG. 38

INTERNATIONAL SEARCH REPORT

International application No.
PCT/UC1 0/51 000

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 5/07, 15/63, 15/85 (2018.01)
CPC - C12N 5/0606, 15/63, 15/85

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
See Search History document

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched
See Search History document

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)
See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X -- Y	(KIM, HM et al.) 'Obox$_4$ regulates the expression of histone family genes and promotes differentiation of mouse embryonic stem cells'; 05 February 2010, FEBS Letters; Volume 584, Issue 3, pages 605-611; abstract; page 606, first column, first, second and third paragraphs; page 607, first column, last paragraph; page 607, second column first paragraph; page 608, figure 2	1-3, 10-12, 15-18, 19/1, 19/15-16, 23-24 ------------------------------ 4-9, 13-14, 21-22
Y	US 2014/0287511 A1 (KO, MSH) 25 September 2014; paragraph [0113]; claims 1, 5, 7-8	4-6
Y	US 2010/0330677 A1 (SMITH, AG) 30 December 2010; paragraphs [0064]-[0066], [0090], [0178]	7-9, 21-22
Y	US 2013/0295579 A1 (XIE, X et al.) 07 November 2013; paragraphs [0007]-[0008], [0012]	13-14

□ Further documents are listed in the continuation of Box C. □ See patent family annex.

*** Special categories of cited documents:**

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts cm piiurily claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date orpriority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" documen , member of the same patent family

Date of the actual completion of the international search 29 October 2018 (29.10.2018)	Date of mailing of the international search report 12 FEB 2019
Name and mailing address of the ISA/ Mail Stop PCT, Attn: ISA/US, Commissioner for Patents P.O. Box 1450, Alexandria, Virginia 22313-1450 Facsimile No. 571-273-8300	Authorized officer Shane Thomas PCT Helpdesk: 571-272-4300 PCT OSP: 571-272-7774

Box No. 1] Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. □ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. □ Claims Nos,;
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

3. □ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. Ill Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

-""-Continued Within the Next Supplemental Box-*" -

1. □ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.

2. □ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.

3. □ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-18, 19/1, 19/15-16, 21-24

Remark on Protest

□ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.

□ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.

□ No protest accompanied the payment of additional search fees.

-"'-Continued from Box No. III Observations where unity of invention is lacking -*** -

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I+, Claims 1-19 (in-part) and 21-24 (in-part) are directed towards a method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Obox6 into a target cell (first exemplary transcription factor).

The method and system will be searched to the extent they encompass a transcription factor of Obox6 (first exemplary transcription factor). Applicant is invited to elect additional transcription factor(s), with specified transcription factor(s) for each, to be searched. Additional transcription factor(s) will be searched upon the payment of additional fees. It is believed that claims 1-19 (in-part) and 21-24 (in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass a method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Obox6 into a target cell (first exemplary transcription factor). Applicants must specify the claims that encompass any additionally elected transcription factor(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be a method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Rhox2a into a target cell (first exemplary elected transcription factor).

Groups I+ share the technical features including a method of producing an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell.

However, these shared technical features are previously disclosed by US 2014/028751 1 A 1 (KO).

Ko discloses a method of producing an induced pluripotent stem cell (method of producing induced stem cells; paragraph [0006]) comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell (method comprises introducing Patl2 (transcription factor identified in Table 4) into somatic (target) cells; paragraphs [0006]-[0007]).

Since none of the special technical features of the Groups I+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Ko reference, unity of invention is lacking.

Group II, Claim 20 is directed towards a method of treating a subject with a disease.

Group III, Claims 25-26 are directed toward a method of increasing the efficiency of reprogramming a cell.

Group IV, Claims 27-39 are directed toward a computer-implemented method for mapping developmental trajectories of cells.

Group V, Claim 40 is directed towards a method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Gdf9.

The inventions listed as Groups I+ and II-V do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Groups I+ include a method of producing an induced pluripotent stem cell comprising introducing at least one of the transcription factors identified in Table 2, Table 3, Table 4, Table 5, and Table 6, into a target cell, which is not present in Groups II-V; the special technical features of Group II include a method of treating a subject with a disease, which is not present in Groups I+ and III-V; the special technical features of Group III include a method of increasing the efficiency of reprogramming a cell, which is not present in Groups I+, II and IV-V; the special technical features of Group IV include a computer-implemented method for mapping developmental trajectories of cells, which is not present in Groups I+, II-III, and V; and the special technical features of Group IV include a method of producing an induced pluripotent stem cell comprising introducing a nucleic acid encoding Gdf9, which is not present in Groups I+ and II-III.

The common technical features of Groups I+, II-III and V are a method of introducing a transcription factor into a target cell.

These common technical features are disclosed by Ko. Ko discloses a method of introducing a transcription factor into a target cell (method comprising introducing Patl2 (transcription factor identified in Table 4) into somatic (target) cells; paragraphs [0006]-[0007]).

Since the common technical features are previously disclosed by Ko, these common features are not special and so Groups I+ and II-IV lack unity.

No technical features are shared between Groups I+ and IV, accordingly, these groups lack unity a priori.

No technical features are shared between Groups II and IV, accordingly, these groups lack unity a priori.

No technical features are shared between Groups III and IV, accordingly, these groups lack unity a priori. No technical features are shared between Groups IV and V, accordingly, these groups lack unity a priori.