



(51) International Patent Classification:

C12N 15/10 (2006.01) *G01N 1/28* (2006.01)
C12Q 1/68 (2006.01) *G01N 1/30* (2006.01)

(21) International Application Number:

PCT/US2016/025077

(22) International Filing Date:

30 March 2016 (30.03.2016)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/140,438 30 March 2015 (30.03.2015) US

(71) Applicants: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, MA 02138 (US). **ALLEN INSTITUTE** [US/US]; 615 Westlake Ave N., Seattle, Washington 98109 (US).

(72) Inventors: **JANG, Sumin**; 5 Cowperthwaite Street, Apt. 211, Cambridge, MA 02138 (US). **THOMSEN, Elliot, R.**; 1529 Queen Anne Avenue N, Apt. 103, Seattle, WA 98109 (US). **DOYLE, Adele, M.**; 209 Wavcrest Court, Goleta, CA 93117 (US). **LEVI, Boaz, P.**; 5627 6th Avenue NW, Seattle, WA 98107 (US). **RAMANATHAN, Sharad**; 41 Myrtle Street, Belmont, MA 02479 (US).

(74) Agent: **TREVISAN, Maria, A.**; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, MA 02210-2206 (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, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JP, KE, KG, KN, KP, KR, 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, 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))

(54) Title: METHODS TO ISOLATE HIGH QUALITY RNA FROM FIXED SAMPLES

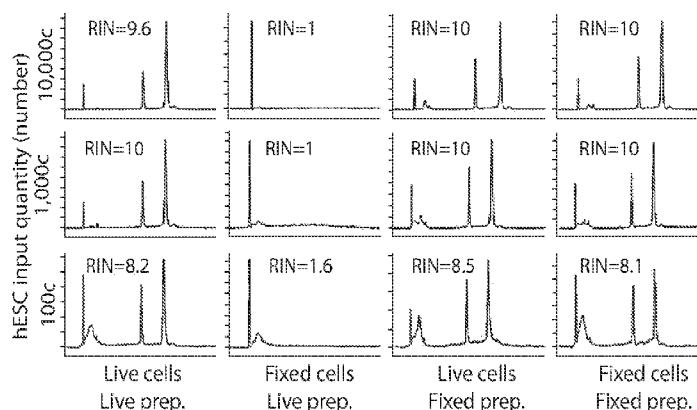


Figure 1A

(57) Abstract: Some aspects of this disclosure provide compositions, methods, and kits for isolating RNA from fixed biological samples. Such compositions, methods, and kits are useful for isolating intact (e.g., not degraded or fragmented) RNA, for example, RNA having an RNA integrity number (RIN) ≥ 8.0 .

- 1 -

METHODS TO ISOLATE HIGH QUALITY RNA FROM FIXED SAMPLES

RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. §119(e) of U.S. Provisional Application No. 62/140,438 filed March 30, 2015, entitled “METHODS TO ISOLATE HIGH QUALITY RNA FROM FIXED SAMPLES”, the entire contents of which are incorporated by reference herein.

FEDERALLY SPONSORED RESEARCH

This invention was made with U.S. Government support under 5DP1 MH099906 awarded by the National Institutes of Health. The U.S. Government has certain rights in the invention.

BACKGROUND OF INVENTION

Accurate measurement of RNA expression in cells and tissues is needed for many clinical and research studies. Formaldehyde fixation of cells is a common method to preserve the intracellular nucleic acid and protein content prior to analysis. However, recovering RNA from cells that have been formaldehyde fixed results in degraded and chemically modified fragments of RNA, rather than a distribution of full length RNA typical of un-fixed cells. These degraded fragments of RNA can be biased (i.e., unrepresentative of true abundance levels in the cell), difficult to identify (when fragments are short or map to multiple regions), and difficult to quantify (due to incomplete or biased recovery and lack of mapping confidence). Accordingly, improved methods for recovering RNA from fixed cells and tissues are needed.

SUMMARY OF INVENTION

Aspects of the present disclosure relate to the recognition that currently available methods for isolating (or extracting, as the terms are used interchangeably herein) RNA from formaldehyde or formalin fixed cells and tissues, while sometimes providing adequate yields of total RNA, do not readily and reliably produce high yields of intact (e.g., un-fragmented) total RNA. For example, RNA obtained from fixed cells using current methodology typically has an RNA integrity number (RIN) between 1.0 and 4.0, where RIN is based on a scale from 1.0-10.0, with 1.0 referring to totally degraded RNA and 10.0 referring to totally intact RNA. The

- 2 -

methods and kits provided herein represent improvements over existing technology. The methods of this disclosure surprisingly yet reproducibly result in isolated RNA from fixed cells having a RIN of greater than 8.0, and often greater than 9.0.

In some aspects, the methods of this disclosure differ from prior art methods in the length and temperature of certain incubations. For example, conventional methods typically require high temperature (e.g., 80°C), limited duration incubation of a fixed biological sample. Such high temperature incubations were considered necessary in the art to reverse cross-linking induced by fixation. However, these conventional methods, as described herein, generally produce samples of fragmented RNA. Unexpectedly, certain aspects of the methods provided herein utilize lower temperatures (e.g., about 56°C) for prolonged periods and thereby produce high yields of intact RNA, without the need for additional high temperature (e.g., 80°C) incubations. The method and kits described herein may also be used to isolate RNA from small numbers of cells, even single cells. Using such high quality, intact RNA in downstream research and clinical applications can, inter alia, reduce measurement bias, improve detection and mapping accuracy, and enable quantification of expressed RNA transcripts from fixed cells and tissues.

The methods of this disclosure can be applied at the single cell level. This is the first reported method having this capability. This is a significant advantage because of the single cell nature of many biological applications including but not limited to single cancer cell isolation and analysis, stem cell isolation and analysis, and the like.

The methods of this disclosure can be applied to archived samples. These methods can be used to extract high-quality RNA at a later time from properly archived and stored samples. These methods are therefore not dependent on the use of freshly harvested tissue. This is a significant advantage particularly in low-resource clinical or research settings where immediate processing is not available or possible.

The methods of this disclosure also provide the ability to deeply profile rare cell populations, as exemplified herein. For example, the methods can be used for transcriptome profiling of sorted cells, including fixed, stained and sorted cells from for example primary tissue such as but not limited to fetal tissue. As demonstrated herein, such analysis may be performed on single cells such as single neural progenitors and single cortical progenitors such as but not limited to [PAX6+ SOX2+ TBR2+] and [PAX6+ SOX2+ TRB2-] cells. This technique can be used to enrich rare populations such as those identified by an antibody or other

- 3 -

binding partner. As provided herein including in the Examples, an infrequent (1%) progenitor population is harvested and analyzed, leading to the discrete populations therein such as for example a roughly 0.4% population and a roughly 0.6% population of the total cells.

Thus, in one aspect, this disclosure provides a method for purifying RNA from a fixed
5 biological sample, comprising: (a) incubating the sample in an aqueous lysis buffer at a temperature between about 50-60°C for a time between about 2-4 hours; (b) contacting the sample with a DNase for a time sufficient to digest DNA; (c) contacting the sample with a substrate that binds RNA, and (d) eluting bound RNA from the substrate.

In some embodiments, the fixed biological sample is a formalin-fixed cell sample or a
10 formalin-fixed tissue sample. In some embodiments, the fixed biological sample is a formalin-fixed paraffin-embedded (FFPE) sample, and optionally the method further comprises an initial deparaffinization step.

In some embodiments, step (a) comprises incubating the sample in the aqueous lysis
buffer at a temperature between about 52-58°C. In some embodiments, step (a) comprises
15 incubating the sample in the aqueous lysis buffer at a temperature between about 54-57°C. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C +/- 2°C. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C.

In some embodiments, step (a) comprises incubating the sample in the aqueous lysis
20 buffer for about 2-3 hours. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer for about 3-4 hours. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer for about 3 +/- 0.5 hours. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer for about 3.0 hours. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis
25 buffer at a temperature of about 56°C for about 3.0 hours.

This disclosure provides a variation of the foregoing method in which step (a) comprises incubating the sample in the aqueous lysis buffer for about 0.5 to 3 hours, including 1-3 hours, 1-2 hours, or about 1 hour.

In some embodiments, the method does not comprise incubating the sample at a
30 temperature above 60°C. In some embodiments, the method does not comprise incubating the sample at a temperature of about 80°C or higher.

- 4 -

In some embodiments, the lysis buffer comprises a buffering agent and a detergent. In some embodiments, the buffering agent is selected from the group consisting of Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof. In some embodiments, the detergent is selected from the group consisting of non-ionic, cationic, anionic and zwitterionic detergents. In some embodiments, the lysis buffer further comprises one or more substances selected from the group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers. In some embodiments, the lysis buffer has a pH in a range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

In some embodiments, a proteolytic agent either is added to the mixture in step (a) in addition to the lysis buffer, or a proteolytic agent is already comprised in the lysis buffer, the proteolytic agent being selected from the group consisting of proteases and non-enzymatic proteolytic compounds. In some embodiments, the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocyanide, recombinant Bacillus proteases, Lysozyme, and a mixture thereof.

In some embodiments, once eluted, RNA has an RNA integrity number (RIN) between about 5.0-10.0. In some embodiments, the eluted RNA has a RIN of at least 5.0, at least 5.5, at least 6.0, at least 6.5, at least 7.0, at least 7.5, at least 8.0, at least 8.5, at least 9.0, or at least 9.5. In some embodiments, the method produces an amount of eluted RNA of at least 0.2 µg, at least 2 µg, or at least 20 µg. In some embodiments, the RNA once eluted has a purity of at least 1.8 as assessed by the 260 nm/280 nm ratio. In some embodiments, the RNA once eluted comprises no more than 0.01% or 0.001% (w/w) of DNA.

In some embodiments, the fixed biological sample comprises at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , at least 10^9 , at least 10^{10} , at least 10^{11} , or at least 10^{12} cells. In some embodiments, the fixed biological sample comprises embryonic stem cells, pluripotent stem cells, somatic cells, germ cells, or gametes. In some embodiments, the fixed biological sample comprises human cells or tissue.

In another aspect, this disclosure provides a kit for purifying RNA from a fixed biological sample, comprising an aqueous lysis buffer and instructions, wherein the instructions provide that a fixed biological sample should not be incubated in the lysis buffer at a temperature above 60°C.

- 5 -

In some embodiments, the instructions provide that a fixed biological sample should not be incubated in the lysis buffer at a temperature of 80°C or higher. In some embodiments, the instructions provide that a fixed biological sample is incubated in the aqueous lysis buffer at a temperature between about 50-60°C, for a time between about 2-4 hours.

5 In some embodiments, the lysis buffer comprises a buffering agent and a detergent. In some embodiments, the buffering agent is selected from the group consisting of Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof. In some embodiments, the detergent is selected from the group consisting of anionic and zwitterionic detergents. In some embodiments, the lysis buffer further comprises one or more substances selected from the
10 group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers. In some embodiments, the lysis buffer has a pH in the range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

 In some embodiments, the kit further comprises a proteolytic agent selected from the group consisting of proteases and non-enzymatic proteolytic compounds. In some
15 embodiments, the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocysteine, recombinant Bacillus proteases, Lysozyme, and a mixture thereof.

 In some embodiments, the kit further comprises one or more of a DNase, RNA binding substrate, and elution buffer.

20 In yet another aspect, this disclosure provides a method for purifying RNA from a fixed biological sample, comprising: (a) incubating the sample in an aqueous lysis buffer at a temperature between about 50-60°C for a time between about 0.5-3 hours; (b) contacting the sample with a substrate that binds RNA, and (c) eluting bound RNA from the substrate.

 In some embodiments, the fixed biological sample is a formalin-fixed cell sample or a
25 formalin-fixed tissue sample. In some embodiments, the fixed biological sample is a formalin-fixed paraffin-embedded (FFPE) sample, and optionally the method further comprises an initial de-paraffinization step.

 In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature between about 52-58°C. In some embodiments, step (a) comprises
30 incubating the sample in the aqueous lysis buffer at a temperature between about 54-57°C. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C +/- 2°C. In some embodiments, step (a) comprises incubating the

- 6 -

sample in the aqueous lysis buffer at a temperature of about 56°C. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer for about 0.5-2 hours. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer for about 1 +/- 0.5 hours. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer for about 1.0 hour. In some embodiments, step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C for about 1.0 hour.

In some embodiments, the method does not comprise incubating the sample at a temperature above 60°C. In some embodiments, the method does not comprise incubating the sample at a temperature of about 80°C or higher.

In some embodiments, the lysis buffer comprises a buffering agent and a detergent. In some embodiments, the buffering agent is selected from the group consisting of Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof. In some embodiments, the detergent is selected from the group consisting of non-ionic, cationic, anionic and zwitterionic detergents. In some embodiments, the lysis buffer further comprises one or more substances selected from the group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers. In some embodiments, the lysis buffer has a pH in a range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

In some embodiments, a proteolytic agent either is added to the mixture in step (a) in addition to the lysis buffer, or a proteolytic agent is already comprised in the lysis buffer, the proteolytic agent being selected from the group consisting of proteases and non-enzymatic proteolytic compounds. In some embodiments, the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocyanide, recombinant Bacillus proteases, Lysozyme, and a mixture thereof.

In some embodiments, the method further comprises evaluating the RNA integrity of the RNA by gene expression analysis of housekeeping genes or other markers that are consistent with that observed in live cells.

In some embodiments, the method produces an amount of eluted RNA of at least 0.1 pg, at least 1 pg, at least 10 pg, at least 100 pg, or at least 1000 pg.

In some embodiments, the fixed biological sample comprises 1-10 cells, or at least 10, at least 50, at least 100, at least 10^3 , at least 10^4 , or at least 10^5 cells. In some embodiments, the fixed biological sample comprises 1, 2, 3, 4, 5, 6, 7, 8, 9, or 10 cells.

- 7 -

In some embodiments, the fixed biological sample comprises embryonic stem cells, pluripotent stem cells, somatic cells, germ cells, or gametes. In some embodiments, the fixed biological sample comprises human cells or tissue.

These and other advantages, features, and uses of the invention will be apparent from the entirety of this disclosure.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1A-1E show high quality and yield RNA isolated from fixed and sorted cells. Figure 1A presents representative bioanalyzer traces showing column purified total RNA from 10,000, 1,000, and 100 live or fixed, permeabilized and DAPI-stained H1 hESCs with or without reverse-crosslinking (Fixed prep. and Live prep respectively). Data are from two independent experiments. Samples were diluted or concentrated so 66 cell equivalents were loaded per lane. Figures 1B and 1C show RNA integrity and RNA yield quantified from H1 hESCs. The first three columns of each graph correspond to live cells using the live cell preparation; the second three columns correspond to fixed cells using the live cell preparation; the third three columns correspond to live cells using the fixed cell preparation; the last three columns correspond to fixed cells using the fixed cell preparation. Figure 1D shows a comparison of 30-Ct from 82 genes with 11 and 12 replicates of 10,000 live or fixed hES cell samples (respectively) processed with the fixed cell protocol. Figure 1E shows a comparison of Δ Ct from qRT-PCR data between total brain RNA (white rows; second, fourth and sixth rows from the top), and fixed H1 hESCs (grey shaded rows; first, third, and fifth rows from the top) purified using RNeasy columns or dT25 beads. Data combined from 3 biological replicates.

Figures 2A-2F show single cell RNA-Seq from fixed cells. Figure 2A presents representative Bioanalyzer traces of SmartSeq2 amplified cDNA from single hESCs (Live or Fixed) processed by the standard protocol (Live) or with reverse crosslinking and purification using DYNABEADS® Oligo(dT)₂₅ (Fixed). Figure 2B shows cDNA quantity measured after 19 cycles of SmartSeq2 amplification. Data are from two independent experiments except the ERCC only and empty well controls. Error bars indicate standard deviation, number of single cell samples are indicated on bars. Data are from two biological replicates. Figure 2C shows HiSeq RNA-Seq data for each cell processed by the four listed conditions. Single-cell libraries were sequenced to a moderate depth (7.0M reads +/- 0.9M). 100% is all barcoded reads

- 8 -

independent of mapping. Figure 2D shows the number of unique genes detected by RNA-Seq per cell. Figure 2E shows a heatmap with hierarchical clustering for single hES cells showing pairwise correlation. Column colors indicate experimental replicate and rows indicate protocol. Figure 2F presents a table showing number of differentially expressed genes between the
 5 different conditions calculated by DESeq.

Figures 3A-3D show prospective isolation and profiling of single differentiating hES cells shows concordance between protein expression and gene expression. Figure 3A depicts a sorting schematic showing differentiated single cells sorted from undifferentiated cells using PAX6 intracellular antibody staining. Six single PAX6+ and PAX6- cells were sorted and
 10 processed for RNA-Seq by the SmartSeq2 protocol. The PAX6- cells within the region of the trace indicated by "P6" are collected in tubes 1-6, depicted below. The PAX6+ cells within the region of the trace indicated by "P5" are collected in tubes 7-12, depicted below. Figure 3B presents normalized MiSeq RNA-Seq data showing substantial expression of three common housekeeping genes from each sorted cell. Data for each of the cells (1cpos1-1cpos6 and
 15 1cneg1-1cneg6) consists of three columns: on the left, GAPDH; in the middle, ACTB; and on the right, PPIA. Figure 3C presents normalized MiSeq RNA-Seq data showing population-specific expression of PAX6+ and POU5F1. Data for each of the cells (1cpos1-1cpos6 and 1cneg1-1cneg6) consists of two columns: on the left, PAX6 and on the right, POU5F1. Figure 3D shows substantial POU5F expression is only highly detected in three of the six PAX6- cells. FACS
 20 analysis performed with those cells stained with antibodies against OCT4 (expressed by the POU5F1 gene) and PAX6 shows half of PAX6- cells and very few PAX6+ cells express substantial OCT4. Green boxes show PAX6+ cells and orange boxes highlight PAX6- cells. Data are from one biological experiment.

Figure 4 provides representative BioAnalyzer profiles of RNA extracted according to the
 25 methods of this disclosure (left panel) versus control (right panel).

Figures 5A-B show that human cortical progenitors are diverse and intermixed during development. (Figure 5A) Model of the progenitor compartment shows a mixture of radial glial cells (RG), outer RGs (oRGs), intermediate progenitors (IPCs) and other unmarked cortical cell types. Known markers for each cell type are shown below. RGs identified by antibody staining
 30 are called SP (SOX⁺, PAX⁺, TBR2⁻), and IPCs are called SPT (SOX⁺, PAX⁺, TBR2⁺). (Figure 5B) Immunocytochemistry images of 18 PCW germinal zones. *Top left*: Low magnification images stitched together (*left*), or individual micrograph (*right*) of the VZ, the iSZ, and oSZ stained by

- 9 -

DAPI, TBR2, PAX6, and SOX2 with the colors indicated. Scale bars are 100 μ m. High magnification for VZ/iSZ region (*middle*) and oSZ region (*bottom*) show SP (arrow with filled arrowhead) and SPT (arrowhead) cells. Many cells the VZ, iSZ, and oSZ lack progenitor markers and are unknown cell types (arrows with empty arrowhead). Colors are as in top, scale bars are 50 μ m.

Figure 6 depicts a schematic for an experiment comparing live and fixed cells and preparations.

Figures 7A-7G. FSC-Seq of human cortical progenitors identifies genetic signatures for vRGs and oRGs. (Figure 7A) Schematic shows single-cell transcriptomic profiling of primary SP and SPT cells. Color scheme is followed throughout Figure, and scale is 1cm. (Figure 7B) FACS plots show gating strategy and frequency of gated cells compared to total cortical cells. *Left* plot includes only G₀-G₁ cortical cells and *right* plot shows only cells gated on the *left* plot. (Figure 7C) Barplots show the frequency of indicated genes detected per cell. Only cells that had >50,000 mapped reads are included (n = 86 SP and n = 22 SPT cells). (Figure 7D) Correlation between cells based on the module eigengene values derived from WGCNA analysis. (Figure 7E) Heatmap showing the correlation of individual fixed singles to the indicated module eigengenes. Module 3 and 4 eigengenes are enriched in mostly non-overlapping sets of SP cells (Figure 7F) Heatmap shows correlation of single cell-derived eigengenes with 15 PCW (*top*) and 21PCW (*bottom*) gene expression data from the BrainSpan Atlas of the Developing Human Brain ⁷. Abbreviations are: subgranular zone (SG), subplate (SP), intermediate zone (IZ), outer subventricular zone (oSZ), inner subventricular zone (iSZ), and VZ (ventricle zone). (Figure 7G) Heatmap showing normalized gene expression of genes differentially expressed between modules vRG and oRG in the single cells (*left*) and in 21 PCW human tissues (*right*). Both in the single cell data and the LCM data, vRG and oRG genes largely do not overlap cells or regions. All heat maps show normalized expression with red indicating high expression, blue indicating low expression; and single cell heatmaps in Figure 7E and 7G are clustered as in Figure 7D.

Figures 8A-I shows confirmation of markers for vRGs and oRGs, and model for the lineage progression of radial glial progenitors. (Figure 8A) DESeq analysis of vRG versus oRG single cell data show that HOPX and ANXA1 (in red) are two of the most differentially expressed genes between these groups. All genes with pval < 0.001 and Log₂ fold change differential expression of >2 or <-2 are shown. (Figure 8B) Greyscale (*left*) micrographs of the

- 10 -

entire cortical depth showing DAPI, SOX2, and either vRG marker ANXA1 or oRG marker HOPX identified from gene expression, and co-stained (*right*) micrographs of only the germinal zone showing co-staining SOX2 (red) and ANXA1 or HOPX (green). Tissue is 16 PCW cortex, scales are 100 μ m. ANXA1 staining overlaps with SOX2 only in the VZ while HOPX bright cells extensively co-stain with SOX2 in the SZ but rarely in the VZ. Abbreviations are: marginal zone (MZ), cortical plate (CP), subplate/intermediate zone (SP/IZ), outer subventricular zone (oSZ), inner subventricular zone (iSZ), and ventricle zone (VZ). (Figure 8C) Micrograph of germinal zone of 19PCW cortex showing co-staining SOX2 (red) and HOPX (green), scales are 100 μ m. (Figure 8D) High resolution images of ANXA1 (top) and HOPX (bottom) in the oSZ (left) and VZ/iSZ region (right). Scale is 50 μ m. Examples of co-expression with SOX2 (arrowhead) and SOX2-only cells (arrow) are indicated. A rare apical endfoot in a HOPX expressing cell is indicated (asterisk). (Figure 8E) High resolution micrograph of a mitotic cell in the oSZ co-stained with SOX2, HOPX and phospho-VIM. Scale bar is 20 μ m. (Figures 8F-H) Quantitation of SOX2⁺, HOPX⁺ or HOPX⁻ cells in oSZ that are Ki67⁺ (f); percentage of SOX2⁺ cells that are also HOPX⁺ (Figure 8G); and the percentage of HOPX⁺ cells that are also SOX2⁺ (Figure 8H). Data is mean $\pm$ St. Dev. for five human tissue aged between 14 and 19 PCW; *** p<0.001 by homoscedastic TTEST. Between 200-500 cells evaluated per brain. (Figure 8I) Network analysis of vRG (blue) and oRG (green) co-expressed genes. Strength of co-expression is shown (blue bars), and top putative regulators of gene expression (yellow) were identified through enrichment of transcription factor binding near co-expressed genes in existing epigenetic data sets ¹⁸. Only binding sites accessible to DNase1 in human fetal forebrain were considered ¹⁷. Genes regulated by HOPX, SRF and ELK1 were identified in previous studies ^{19,20,24}.

Figure 9 shows high quality and yield RNA from fixed and sorted cells. Total recovered RNA yield quantified from indicated numbers of H1 hESCs. Error bars indicate standard deviation. Sample names are described previously (Figure 6), sample numbers were as in Figure 1B. Quantities for 100 cell samples are not shown since low molecular weight broad peak that appeared after concentration (not RNA specific) interfered with quantitation.

Figures 10A-10B show single cell FSC-Seq with hESCs. (Figure 10A) Bar plot showing the number of mapped reads per cell. Single-cell libraries were sequenced to a moderate depth (7.0M reads $\pm$ 0.9M). 100% is all barcoded reads independent of mapping. (Figure 10B)

- 11 -

Alignment statistics for HiSeq RNA-Seq data for cells processed by the four listed conditions. * $p < 0.05$ and ** $p < 0.001$ by homoscedastic TTEST.

Figures 11A-11B shows gating strategy for prospective isolation and profiling of human cortical progenitor cells. (Figure 11A) A representative image of the entire gating strategy for the isolation of the SP and SPT cells. *Top row*: cells are gated for morphology, then single cells, and last for the presence of DAPI. Only cells in the G₀-G₁ phase of the cell cycle were sorted. *Bottom row*: Progenitors were first identified by high expression of SOX2 and PAX6, and then for the presence (orange) or absence (blue) of TBR2. *Bottom right*: Projection of progenitor populations on all cells gated by the strategy on the top row and stained for TBR2 and DCX show that SP and SPT progenitors are negative for DCX. (Figure 11B) Representative SP and SPT cells were gated including all phases of the cell cycle. SP and SPT were subsequently evaluated for DAPI fluorescence. Quantitation of cells in S-G₂-M from three independent prenatal tissues (between 18 and 19 PCW) shows a significant, 2-fold, difference between SP and SPT cells.* = $p < 0.05$, paired TTEST.

Figures 12A-12F show MiSeq read statistics for human prenatal SP and SPT cells. (Figure 12A) Representative Bioanalyzer traces show cDNA amplification of one SP and one SPT cell. (Figure 12B) Histogram showing the distribution of SP and SPT cells with respect to the number of mapped reads. Cells with fewer than 50,000 mapped reads were excluded (average $119,252 \pm 55,746$ reads). (Figure 12C) Mapping percent for each cell with more than 50,000 mapped reads; 100% is all barcoded reads independent of mapping. (Figure 12D) Correlation between cells based on high variability genes and (Figure 12E) all ERCCs reads only. High variability genes but not the ERCCs (spiked into each sample at defined concentrations) segregate cells into blocks based on the cell phenotype of the sorted cell. (Figure 12F) Bar plot for SP and SPT cells show the number of unique genes detected at each expression threshold. Error bar indicates standard deviation.

Figure 13 shows data from human cortical progenitors prepared by FSC-Seq clustered appropriately with independent data from live single cells. Data from fixed SP and SPT cells were compared to an existing data set generated from live prenatal human neuronal single cell by hierarchical clustering. Clustering was based on the common 482 of the top 500 genes contributing to the first principal component that was identified in a previous study¹³. The two previously published IPCs cluster with the fixed SPT cells, and previously published RGs cluster with our fixed SP cells. This clustering robustly categorizes cells by cell type despite the

- 12 -

two data sets differing in: sample age, sample donor, dissociation, cell capture, cDNA amplification, and using live or fixed cells. NPCs are neural precursor cells generated from hESCs¹³. Lower color bar annotations were based on annotation from the earlier study¹³.

Figure 14. Dot plots show Principal Component Analysis (PCA) for the first four principal components based on the genes with the highest variances across all prenatal cell samples with more than 50,000 mapped reads.

Figure 15. Heatmaps of the genes that comprise each of the 6 WGCNA modules are shown. Heatmap colors show gene expression normalized between cells of a given gene. Red indicates high expression, blue indicates low expression. SP and SPT cells are noted by the color bar and in the legend, and are organized based on clustering of WGCNA eigengenes (Figure 7D).

Figure 16 shows modules 3 and 4 are enriched in non-overlapping sets of SP cells. FACS analysis of SP and SPT cells stained with TBR2 and isotype (top) or CYR61 (bottom). Frequencies of events in quadrant gates are indicated (right). Most CYR61⁺ cells are in the SP but not the SPT cells indicating CYR61 highlights a subpopulation of RG but not IPC cells.

Figures 17A-17C. (Figure 17A) Heatmaps show the enrichment of the module eigengene expression pattern on Allen Brain Atlas Human prenatal cortical gene expression data. Regions that correlate better to any given eigengene are red, while poorer correlation is blue. (Figure 17B) Heatmaps shows specific marker genes expressed in the BrainSpan Atlas of the Developing Human Brain⁷. Genes were differentially-expressed between module 3 and 4 enriched cell, though they may not be in the original modules (Figure 8A). Regions with strong correlation to a given gene are red, while poor correlation is blue. (Figure 17C) Expression of HOPX in macaque though development. Plot shows expression of HOPX in the visual cortex (V1) of the NIH Blueprint Non-Human Primate Atlas. Each column shows different ages from embryonic day 40 (E40), through 48 months after birth (48M). Rows are micro-dissected regions. Red color indicates high expression. Abbreviations for prenatal time points are: marginal zone (MZ), outer cortical plate (oCP), inner cortical plate (iCP), subplate (SP), intermediate zone (IZ), outer subventricular zone (oSZ), inner subventricular zone (iSZ), and VZ (ventricle zone). Abbreviations for postnatal time points are: Layer 1-6 (L1-L6), and white matter (WM). Expression of HOPX is initiated in V1 at E70, when the oSZ is formed, and while first detected in the entire germinal zone, is subsequently enriched in the oSZ.

- 13 -

Figure 18. Greyscale (*left*) micrographs of the entire cortical depth showing DAPI, SOX2, and HOPX, and co-stained (*right*) micrographs showing co-staining SOX2 (red) and HOPX (green). Tissue is 19 (*top*), and 15 (*bottom*) PCW cortex, scales are 100 μ m. Insets of the oSZ (i, ii, and iv) and the VZ (iii, v) are shown to right. Substantial overlap of HOPX and SOX2 staining in the oSZ (i, ii, and iv), some overlap in the 15PCW VZ (v), and no overlap in the 19 PCW VZ (iii). Abbreviations are: marginal zone (MZ), cortical plate (CP), subplate/intermediate zone (SP/IZ), outer subventricular zone (oSZ), inner subventricular zone (iSZ), and VZ (ventricle zone).

Figure 19 shows greyscale and merged micrographs of a representative region of oSZ stained for SOX2, HOPX, and Ki67. White arrows are SOX⁺, HOPX⁻, Ki67⁺ cells and arrowheads are SOX⁺, HOPX⁺, Ki67⁺ cells.

The foregoing information is additionally apparent by reference to the color version of the foregoing Figures, available upon request.

DETAILED DESCRIPTION OF INVENTION

Fixation is a chemical process by which cells and tissues are preserved from decay, thereby preventing autolysis or putrefaction. Fixation terminates any ongoing biochemical reactions, and may also increase the mechanical strength or stability of the treated cells or tissues. Fixation theoretically preserves cells and tissues as close to their natural state as possible. This is desirable if the cells or tissues are to be examined, such as by gene expression and histological studies. A common fixative agent is formaldehyde, which is typically employed as a saturated water solution that contains about 40% formaldehyde by volume or 37% by mass (referred to as “100% formalin”). Formalin is a commercial name for formaldehyde (i.e., the active fixative in formalin is formaldehyde). Formaldehyde preserves (or fixes) cells or tissues by inducing cross-linking of primary amino groups in proteins with other nearby nitrogen atoms in proteins or nucleic acids. While formaldehyde fixed cells or tissues are preserved for analysis (which can be performed immediately or even years later), the molecular cross-linking that underlies fixation is believed to cause significant challenges to certain applications, including for example to studies involving ribonucleic acids (RNA), such as messenger RNA (mRNA) and microRNA (miRNA).

- 14 -

For applications that require the isolation of RNA from fixed cells or tissues such as gene expression studies (e.g., using RT-PCR, microarray analysis, etc.), the fixation process, at least in part, typically results in the isolation of heavily fragmented and chemically (e.g., formaldehyde) modified RNAs that are not accurately reflective of the natural RNA pool that existed in the cell just prior to fixation. As such, these degraded fragments of RNA can be biased (e.g., unrepresentative of true abundance levels in the cell), difficult to identify (e.g., when fragments are short or map to multiple regions), and difficult to quantify (e.g., due to incomplete or biased recovery and lack of mapping confidence).

Various studies have attempted to increase the quality and quantity of RNA from formaldehyde fixed cells or tissues. These studies have analyzed the effects of temperature, pH and the presence of RNases (Specht et al., "Quantitative Gene Expression Analysis in Microdissected Archival Formalin-Fixed and Paraffin-Embedded Tumor Tissue." *Am J Pathology*, 2001, 158(2):419-429); the effect of base composition on biased recovery of RNA (Doleshal et al., "Evaluation and validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues." *J Mol Diagn*. 2008, 10:203-11); incomplete recovery of RNA from the insoluble material and fragmentation due to incomplete lysis or cross-links remaining after heating and protease incubation (Park et al., "Detection of Hepatitis C Virus RNA Using Ligation-Dependent Polymerase Chain Reaction in Formalin-Fixed, Paraffin-Embedded Liver Tissues." *Am J Pathol*. 1996, 149: 1485-1491; Srinivasan et al., "Effect of fixatives and tissue processing on the content and integrity of nucleic acids." *Am J Pathology*. 2002, 161:6.); chemical modifications on recovered RNA that disrupt physical recognition sites necessary for reverse transcription or sequence amplification (e.g., mono-methylol additions to amino groups, N-CH₂OH, or methylene bridges formed by condensation of amino bases and N-methylols, N-CH₂-N; Masuda et al., "Analysis of chemical modification of RNA from formalin-fixed samples and optimization of molecular biology applications for such samples." *Nuc Acids Res*. 1999, 27:4436-43.); and oxidation on exposed surfaces and microbe/fungal infections affecting RNA in samples stored for several years (von Ahlfen et al., "Determinants of RNA Quality from FFPE Samples." *PLoS ONE*. 2007, 2:e1261). The few studies that have quantified formaldehyde modifications on RNA suggest that formaldehyde modifies RNA strands in a random and as-yet-unmeasurable way (Masuda, *Nuc Acids Res*. 1999, 27:4436-43). Notwithstanding the analyses that have been carried out to date, the typical result is RNA having RINs no greater than about 4. This disclosure provides such a

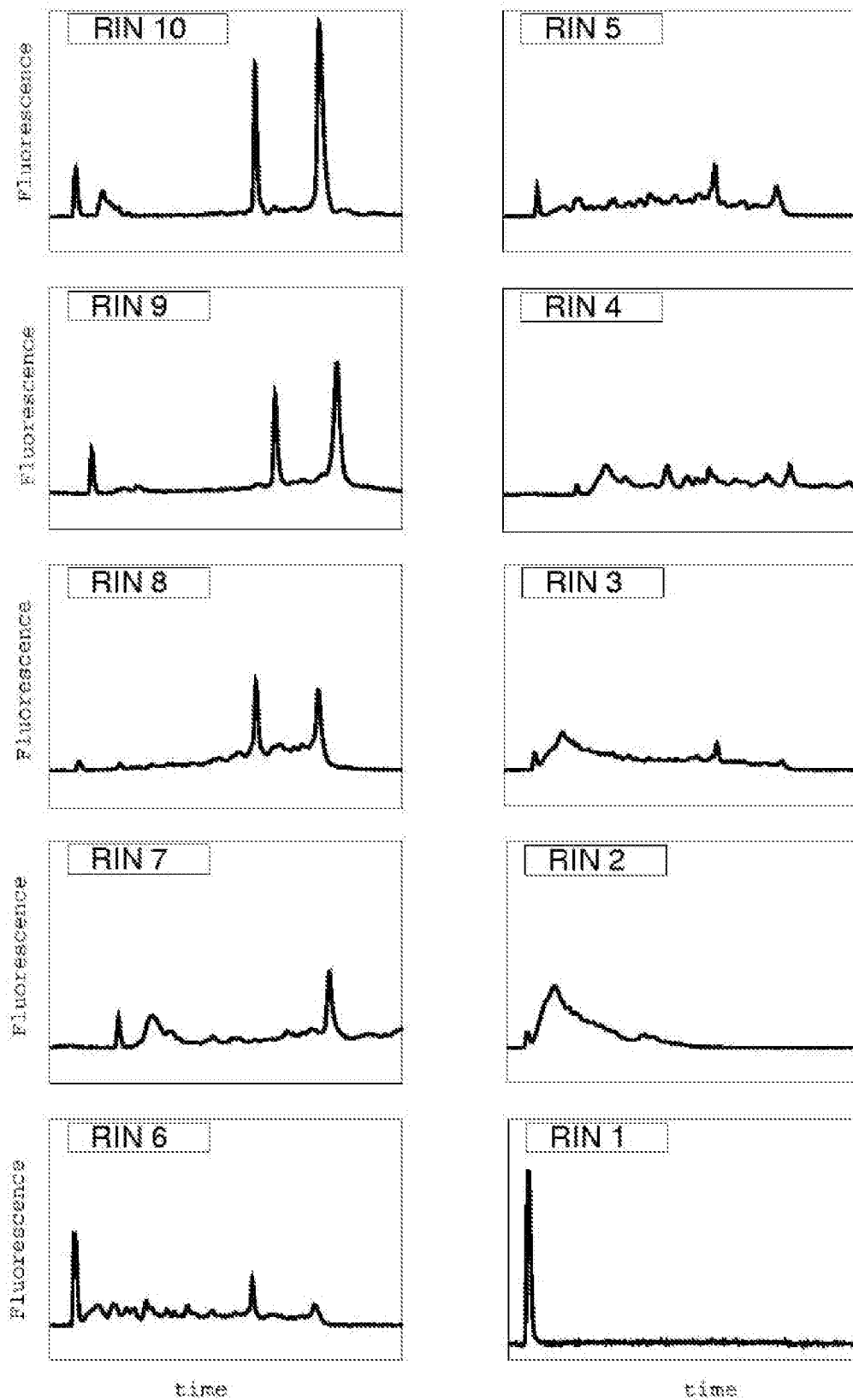
- 15 -

method, and variations thereof, and in doing so provides a reproducible and robust process for isolating high quality RNA suitable for a variety of applications.

RNA quantity and quality is typically classified according to its purity, concentration, and the relative distribution of fragment sizes (reviewed by Fleige and Pfaffl, "RNA integrity and the effect on the real-time qRT-PCR performance." *Mol Aspects Med.* 2006 Apr-Jun;27(2-3):126-39). Spectrophotometric methods, such as Nanodrop (Thermo Fisher Scientific, Waltham, Massachusetts) and Qubit (Life Technologies, Carlsbad, California), measure the concentration of nucleic acids and their purity, or the contribution of nucleic acids relative to non-nucleic acids to the absorbance spectrum, but generally do not measure RNA integrity (e.g., quality of RNA and/or the intact nature of the RNA). Determination of RNA integrity may be performed by measuring the distribution of RNA fragment sizes using the capillary electrophoresis-based Agilent Bioanalyzer (Agilent Technologies, Santa Clara, California). In this assay, the time for RNA to migrate through a microfluidic channel is proportional to fragment size. Longer fragments take a longer time to migrate and be detected. The abundance of fragments present is inferred from the intensity of a fluorescent dye that binds nucleic acids. Full length, intact total RNA measured on a Bioanalyzer produces a size distribution profile characterized by two abundant and narrow peaks at high molecular weight representing the 28S (larger) and 18S ribosomal RNA fragments (see RIN 10 from Figure 2 of Schroeder et al., "The RIN: an RNA integrity number for assigning integrity values to RNA measurements." *BMC Molecular Biology.* 2006, 7:3., reproduced below). The lower molecular weight peaks apparent as small jagged peaks to the left of the 18S and 28S peaks are indicative of RNA degradation and contribute to lower RIN scores.

An RNA integrity number, or RIN, is calculated based on the peak profile of the electropherogram trace using an algorithm designed by Agilent for use with the Bioanalyzer (Schroeder, 2006). RINs range from 1 (degraded RNA) to 10 (intact RNA). Below is shown the overall size distribution, characteristic rRNA peaks (see large peaks in RIN 10 plot), and RINs for a range of good to poor quality RNA samples, as reported by Schroeder et al. From these plots, it is apparent that higher RIN samples have distinctive 28S and 18S peaks and that these peaks diminish in lower RIN samples and lower molecular weight peaks become more apparent.

- 16 -



The industry standard to isolate RNA from formaldehyde fixed cells is to use a commercially available kit, such as those provided by Qiagen (Hilden, Germany; e.g.,

- 17 -

miRNeasy FFPE RNA Isolation Kit, cat. no. 217504), Life Technologies (Carlsbad, California; e.g., RecoverAll™ Total Nucleic Acid Isolation Kit, cat. No. AM1975), Agilent Technologies (Santa Clara, California), NuGEN Technologies (San Carlos, California), and Promega (Fitchburg, Wisconsin; e.g., ReliaPrep™ FFPE Total RNA Miniprep System, cat. no. Z1001).

5 However, these kits do not reproducibly isolate intact RNA. For example, certain suppliers of these kits interpret a RIN of ~4 as successful RNA isolation from formaldehyde fixed cells, despite the fact that this RIN level still implies a degraded pool of RNA. Comparisons of RNA quality obtained from different suppliers have observed no marked differences in electropherogram traces, from which RIN scores are calculated (Doleshal et al., "Evaluation and

10 validation of total RNA extraction methods for microRNA expression analyses in formalin-fixed, paraffin-embedded tissues." *J Mol Diagn.* 2008, 10:203-11; Boeckx, et al., "Expression analysis on archival material: comparison of 5 commercially available RNA isolation kits for FFPE material." *Diagn Mol Pathol.* 2011 Dec;20(4):203-11), indicating that intact RNA is rarely obtained using conventional methodology. As a result, manufacturers of commonly used

15 kits often recommend further processing of the isolated RNA for certain downstream applications, such as RT-PCR gene expression analysis. For example, preparing RNA (e.g., mRNA) from un-fixed cells for RT-PCR analysis typically involves use of oligo-dT primers to amplify cDNA from all mRNA in an unbiased fashion (due to intact mRNAs having a poly-A tail). However, fragmented mRNA will not be consistently amplified due to loss of the poly-A

20 tail. As such, certain manufacturers recommend using gene-specific primers or random nonamers to amplify the fragmented RNA generated according to their methodology, as opposed to conventional oligo-dT primers employed with intact RNA. In contrast, the methods described herein, which provide for isolation of intact RNA from fixed cells having a RIN comparable to that of RNA obtained from fresh (un-fixed) cells, allow for and enable the use of the RNA

25 without the need for further modification or processing, as is common with fragmented RNA. This is desirable, as it reduces potential bias associated with methods for amplifying fragmented RNA.

Methods for isolating intact RNA from fixed cells

30 According to one embodiment, methods for isolating (e.g., extracting) RNA from a fixed biological sample are provided. The methods are useful, *inter alia*, for obtaining high yields of

- 18 -

RNA having RINs corresponding to a sample of RNA having largely intact RNA (e.g., RIN $\geq$ 8.0), from small numbers of cells and also from single cells.

The fixed biological sample is any biological sample containing RNA, which was fixed, e.g., subjected to a fixation condition. Such fixed biological samples include fixed cells
5 including embryonic stem cells, pluripotent stem cells, somatic cells, germ cells, and gametes. In some embodiments, the fixed biological sample is a fixed tissue, such as connective tissue (e.g., areolar tissue, reticular tissue, adipose tissue, bone, cartilage, blood), epithelial tissue (e.g., simple epithelial tissue, secretory epithelial tissue), muscle tissue (e.g., skeletal muscle tissue, smooth muscle tissue, cardiac muscle tissue), and nervous tissue (e.g., brain tissue, spinal cord
10 tissue, peripheral nervous tissue).

In some embodiments, the cells or tissue are obtained from a subject, such as a human subject. In some embodiments, the subject is a non-human mammal. In some embodiments, the subject is a non-human primate. In some embodiments, the subject is a rodent. In some
15 embodiments, the subject is a sheep, a goat, a cattle, a cat, or a dog. In some embodiments, the subject is a vertebrate, an amphibian, a reptile, a fish, an insect, a fly, or a nematode. In some embodiments, the subject is a research animal. In some embodiments, the subject, cells or source of the cells is genetically engineered, e.g., a genetically engineered non-human subject. The subject may be of either sex and at any stage of development.

Typically, a fixed biological sample is fixed by exposing the sample to a fixative agent,
20 such as formaldehyde (commercially available as "formalin"). Other fixative agents include, but are not limited to, para-formaldehyde, glutaraldehyde, methanol, mercurials (e.g., B-5 and Zenker's fixative), picrates, and HOPE fixative (Hepes-glutamic acid buffer-mediated organic solvent protection effect fixative).

The methods comprise incubating a fixed biological sample with an aqueous lysis buffer
25 ("lysis buffer"). The lysis buffer is any buffer that lyses the fixed cells and tissues, thereby releasing the RNA within. Such aqueous lysis buffers are known in the art, and include, for example, those commercially available from vendors such as Qiagen: Buffer PKD (Proteinase K digest buffer, cat. no. 1034963); Life Technologies: Digestion Buffer (e.g., from RecoverAll™ Total Nucleic Acid Isolation Kit, cat. No. AM1975), Agilent Technologies: Proteinase K
30 Digestion Buffer (e.g., from Absolutely RNA FFPE Kit, cat. no. 400809); and Promega: Lysis Buffer (LBA, cat. no. Z101A). Other lysis buffers include those described in U.S. published

- 19 -

patent applications US 2013/0280787, US 2013/0280728, and US 2013/0164825, the entire contents of each are hereby incorporated by reference.

In some embodiments, lysis buffer comprises a buffering agent and a detergent.

Buffering agents are well known, and include, without limitation, Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof.

Detergents are also well known, and include, without limitation, non-ionic, cationic, anionic and zwitterionic detergents. In some embodiments, the detergents are selected from anionic or zwitterionic detergents. In some embodiments, the lysis buffer may comprise anionic detergents, such as sodium dodecyl sulfate (SDS). In some embodiments, nonionic surfactants, such as substituted phenol or sugar polyethoxylates, commercially available for instance as Triton X-114 (Dow Chemical Co., Midland, Mich., USA), Triton X-100 (Dow Chemical Co., Midland, Mich., USA) or Tween 20 (Merck, Darmstadt, Germany) may be used, as well as cationic surfactants, such as quaternary ammonium surfactants, e.g. cetyl trimethylammonium bromide (CTAB) or tetradecyl trimethylammonium bromide (TTAB).

In some embodiments, the lysis buffer comprises a proteolytic agent, which digests proteinaceous components of the cells and/or tissue, helping to break down the cellular components and allowing RNA to be released. Proteolytic agents are known in the art, and include, without limitation, proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocyanide, recombinant Bacillus proteases, and lysozyme. In some embodiments, the lysis buffer comprises proteinase K. In some embodiments, the proteolytic agent (e.g., proteinase K) is not comprised in the buffer, but is added to the sample, e.g., after the lysis buffer has been added to the sample.

In some embodiments, the lysis buffer further comprises one or more of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers. In some embodiments, the lysis buffer comprises a nucleophilic reagent, such as those specifically provided in paragraphs [0040]-[0052] of U.S. published patent application US 2013/0280728.

In some embodiments, the lysis buffer has a pH in a range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

In some embodiments, the lysis buffer comprises guanidine salts. In some embodiments, a binding buffer, which may be used after proteinase K digestion and cross-linking reversal, comprises guanidine salts. In some embodiments, the lysis buffer comprises one or more

- 20 -

nucleophilic reagents such as nucleophilic reagents that induce cross-linking reversal. Examples may be found in published US application US 2013/0280728.

In some embodiments, the method comprises incubating the fixed biological sample in an aqueous lysis buffer at a temperature that is sufficient to free RNA and allow for the isolation
5 of intact RNA.

In some embodiments, "intact RNA" is defined as RNA having a RIN of at least 5.0, at least 5.5, at least 6.0, at least 6.5, at least 7.0, at least 7.5, at least 8.0, at least 8.5, at least 9.0, or at least 9.5. In some embodiments, intact RNA has a $RIN \geq 8.0$. Methods for determining RINs of RNA are known, and include those described by Schroeder et al., "The RIN: an RNA
10 integrity number for assigning integrity values to RNA measurements." *BMC Molecular Biology*. 2006; 7:3 (the entire contents of which are hereby incorporated by reference), e.g., using the Agilent 2100 Bioanalyzer.

In some embodiments, "intact RNA" is defined as RNA that, when assessed by gene expression, shows substantial reproducible expression of housekeeping genes. Examples of
15 housekeeping genes include, without limitation, GAPDH, ACTB, and PPIA.

Conventional methods generally require the incubation of the fixed sample at a temperature of 80°C, e.g., to reverse cross-linking induced by fixation. However, the present inventors have discovered that conventional methods having such an incubation step produce fragmented RNA. Without being bound by any particular theory, the present inventors believe
20 high temperatures (e.g., 60-80°C) contribute to the fragmentation of RNA. Thus, in some embodiments, the methods provided herein do not involve incubating the sample at a temperature above 60°C. In some embodiments, the methods provided herein do not comprise incubating the sample at a temperature above 65°C, 70°C, 75°C, 80°C, or higher. In some embodiments, the methods do not comprise incubating the sample at a temperature of 80°C or
25 higher.

Accordingly, in some embodiments, the methods comprise incubating a fixed biological sample at a temperature of about 50-60°C. In some embodiments, the methods comprise incubating a fixed biological sample at a temperature of 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, or 60°C. In some embodiments, the methods comprise incubating a
30 fixed biological sample at a temperature of 56°C +/- 2°C. In some embodiments, the methods comprise incubating a fixed biological sample at a temperature of 56°C.

- 21 -

In some embodiments, the methods further comprise incubating a fixed biological sample in an aqueous lysis buffer for a time and at a temperature that is sufficient to free RNA and allow for the isolation of intact RNA. Conventional methods generally require incubations for times of less than one hour, for example 15 minutes at a temperature between 50-60°C, followed by a brief (e.g., ~15 minute) incubation at a higher temperature, for example 80°C. However, the inventors have discovered that such incubation times and temperatures generally produce fragmented isolated RNA. Unexpectedly, by increasing the incubation period at a lower temperature range (e.g., 50-60°C) and not incubating the sample at a higher temperature (e.g., 80°C, generally believed in the art to be required to reverse cross-linking), it was discovered that intact, high RIN RNA (e.g., having $RIN \geq 8.0$) could be isolated from a fixed biological sample.

Accordingly, in some embodiments, the methods comprise incubating a fixed biological sample in a lysis buffer (e.g., as described herein) for a time of about two (2) to four (4) hours at 50-60°C. In some embodiments, the methods comprise incubating a fixed biological sample for a time of about 2-3 hours, 3-4 hours, or for about 3.0 hours +/- 0.5 hours. In some embodiments, the methods comprise incubating a fixed biological sample for about 3 hours (e.g., at a temperature as described herein, including a temperature in the range of 50-60°C such as but not limited to 56°C).

In other embodiments, the methods comprise incubating a fixed biological sample in a lysis buffer (e.g., as described herein) for a time of about 0.5 to 3 hours at 50-60°C. In some embodiments, the methods comprise incubating a fixed biological sample for a time of about 0.5-2 hours, 0.5-1 hours, or for about 1 hour +/- 0.5 hour. In some embodiments, the methods comprise incubating a fixed biological sample for about 1 hour (at a temperature as described herein, including a temperature in the range of 50-60°C such as but not limited to 56°C).

In some embodiments, the methods further comprise contacting a fixed biological sample with a DNase. Suitable DNase enzymes and other enzymes affecting e.g., non-target biomolecules (RNAs) as well as buffers and solutions that can be used for performing a respective enzymatic treatment are well known in the art. For example, suitable DNases are commercially available from the vendors described herein, and include RNase-free DNase I available from Qiagen (cat. no. 79254). Typically, the sample is contacted with a DNase following the incubation in lysis buffer, so as to increase the enzyme's ability to digest DNA that has been released as a result of the lysis/digestion step. In some embodiments, an additional

- 22 -

buffering agent may be added to increase the efficiency of the DNase treatment (e.g., Qiagen DNase Booster Buffer).

In some embodiments, the methods further comprise contacting the sample with a substrate that binds RNA. In some embodiments, the substrate is referred to as a “membrane,” and is comprised in a column suitable for centrifugation (e.g., in a microcentrifuge). In some
5 embodiments, such a column comprising a substrate (e.g., membrane) that binds RNA is referred to as a spin column. Such spin columns are well known in the art, and include those provided by the vendors described herein. For example suitable RNA binding substrates include the Qiagen RNEASY® MINELUTE® Spin Columns (e.g., from the miRNeasy FFPE RNA
10 Isolation kit, cat. no. 217504), Filter Cartridge columns from Life Technologies (e.g., from RecoverAll™ Total Nucleic Acid Isolation Kit, cat. No. AM1975), RNA-Binding Spin Cups from Agilent Technologies (e.g., from Absolutely RNA FFPE Kit, cat. no. 400809), and Promega ReliaPrep™ FFPE Binding Columns (e.g., from the ReliaPrep™ FFPE Total RNA Miniprep System, cat. no. Z1001). With such spin columns, the sample is added to the column,
15 and is centrifuged so as to pass the sample through the RNA-binding substrate, thereby binding the RNA to the substrate.

In other embodiments, the substrate is referred to as a “bead” and binds to nucleic acid (e.g., RNA). Such beads may be separated from a solution by centrifugation, filtration, use of a magnet, or any other method known in the art. Beads compatible for use in the methods
20 described herein include those provided by the vendors described herein. For example, suitable RNA binding beads include, without limitation, DYNABEADS® Oligo(dT)₂₅ beads from Life Technologies (e.g., Life Technologies cat. no. 61002) and MagJET Magnetic Beads from Thermo Scientific (e.g., from MagJet RNA Kit, cat. Not. #K2731). In methods using such beads, the beads may be washed and then are added to a sample. The beads and bound RNA are
25 separated from the sample, for example by centrifugation, and the supernatant containing contaminants and other undesired components is removed.

In some embodiments, before the sample is contacted with a substrate that binds RNA, a binding buffer is added to the sample to enhance the binding of the RNA to the substrate. Such binding buffers are well known in the art, and include those contained in commercially available
30 kits, such as those described herein.

In some embodiments, the method further comprises eluting the RNA. Methods and reagents for eluting bound RNA from an RNA-binding substrate are well known. For example,

- 23 -

RNA may be eluted from an RNA-binding substrate using water (e.g., RNase-free water). In some embodiments, RNA may be eluted using a commercially available elution buffer, such as those provided in commercially available kits described herein. In some embodiments, prior to eluting the RNA, ethanol is added to the sample. In some embodiments, prior to eluting the RNA, and after liquids have been spun through (out) of the column, the substrate is washed to remove contaminants, for example using a commercially available wash buffer, such as those contained within commercially available kits described herein. In some embodiments, such as those utilizing RNA-binding substrates comprised in a spin column, the RNA is eluted by contacting the substrate (e.g., membrane) with water or an elution buffer, and centrifuging the column so as to pass the water or elution buffer through the substrate, thereby removing the RNA from the substrate and collecting the RNA in a collection tube. Once eluted, the integrity of the RNA can be assessed, as described herein.

The methods and kits described herein may be used to isolate high quality RNA from biological samples that contain any cell or cell type known in the art. Examples of cells include, without limitation, embryonic stem cells, pluripotent stem cells, somatic cells, germ cells, gametes, or a combination thereof. In some embodiments, the biological sample may contain at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , at least 10^9 , at least 10^{10} , at least 10^{11} , or at least 10^{12} cells. In other embodiments, the methods and kits may be used to isolate high quality RNA from biological samples that contain a small number of cells or even a single cell. In some embodiments, the biological sample may contain 1-10 cells, or at least 10, at least 50, at least 100, at least 10^4 , or at least 10^5 cells. In some embodiments, the biological sample may contain 1, 2, 3, 4, 5, 6, 7, 8, 9 or 10 cells. In some embodiments, the cells are from fetal tissue, including human fetal tissue.

Kits

According to another embodiment, kits for isolating RNA are provided. In some embodiments, the kits comprise reagents necessary for performing the methods as provided herein. For example, in some embodiments, the kit comprises an aqueous lysis buffer, e.g., as provided herein. In some embodiments, the lysis buffer comprises a buffering agent and a detergent, as provided herein. In some embodiments, the lysis buffer further comprises one or more substances selected from the group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers. In

- 24 -

some embodiments, the lysis buffer has a pH in the range of from about 4 to 11, from about 7 to 10, or from about 8 to 9. In some embodiments, the kit comprises a proteolytic agent selected from the group consisting of proteases and non-enzymatic proteolytic compounds. In some embodiments, the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocysteine, recombinant Bacillus proteases, Lysozyme, and a mixture thereof. In some embodiments, the kit comprises one or more of a DNase, RNA binding substrate, and elution buffer.

The kit typically comprises instructions. The instructions may be written on one of the containers of the kit, including for example a box or other housing within which all components are contained. Alternatively, the instructions may be written on paper or other tangible form provided with the kit.

In some embodiments, the instructions provide that a fixed biological sample should not be incubated in the lysis buffer at a temperature above 60°C, above 65°C, above 70°C, above 75°C, or above 80°C. In some embodiments, the instructions provide that a fixed biological sample should not be incubated in the lysis buffer at a temperature of 80°C or higher. In some embodiments, the instructions provide that a fixed biological sample is incubated in the aqueous lysis buffer at a temperature between about 50-60°C, for a time between about 2-4 hours. In some embodiments, the instructions provide that a fixed biological sample is incubated in the aqueous lysis buffer at a temperature between about 50-60°C, for a time between about 0.5-3 hours. In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer at a temperature of 50°C, 51°C, 52°C, 53°C, 54°C, 55°C, 56°C, 57°C, 58°C, 59°C, or 60°C. In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer at a temperature of 56°C +/- 2°C. In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer at a temperature of 56°C. In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer for a time of about 2-3 hours, 3-4 hours, or for about 3.0 hours +/- 0.5 hours. In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer for a time of about 0.5-2 hours, 0.5-1 hours, or for about 1 hour +/- 0.5 hour. In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer for about 3 hours (e.g., at about 56°C). In some embodiments, the instructions provide that a fixed biological sample is incubated in lysis buffer for about 1 hour (e.g., at about 56°C).

- 25 -

EXAMPLES

In order that the invention described herein may be more fully understood, the following Examples are set forth. It should be understood that these Examples are for illustrative purposes only and are not to be construed as limiting this disclosure in any manner.

5

Example 1

Isolation of intact RNA from fixed cells

Materials

In this example, the materials used are commercially available from Qiagen, as part of
10 the miRNeasy FFPE RNA Isolation kit (Cat# 217504). The materials include:

- RNEASY® MINELUTE® Spin Columns (each in a 2 ml Collection Tube)
- Collection Tubes (1.5 ml, 2.0 ml)
- Buffer RBC
- Buffer PKD
- 15 -Proteinase K
- RNase-Free DNase I
- DNase Booster Buffer
- Buffer RPE
- RNase-free water

20 The cells used in this example were mouse embryonic stem cells that were formaldehyde fixed, permeabilized, intracellular antibody stained, and FACs sorted prior to isolating RNA according to the protocol below.

Mouse ES cell preparation, fixation, permeabilization, staining and sorting

- 25 1. Mouse embryonic stem cells were differentiated to obtain populations enriched for early germ layer progenitor cell types as well as intermediate cell types, as defined by the presence and/or absence of a combination of transcription factor markers.
- a. Cells were seeded at a density of 10^6 cells per 100mm diameter tissue culture dish with 10mL of N2B27 media
 - 30 b. ES cells were maintained in LIF+2i (Ying et al., Nature 2008)

- 26 -

c. Differentiated populations were obtained by manipulation of Fgf, Wnt and/or Bmp signaling for 1 to 5 days

2. Following differentiation, cells were trypsinized with 0.01% trypsin, then quenched with FBS and collected in a 15mL conical tube.
- 5 3. Cells were pelleted at 200rcf for 5 minutes and supernatant was aspirated.
4. Cells were resuspended in 4% paraformaldehyde in PBS containing 2U/uL RNasin (Promega) at a density of $\sim 10^7$ cells/mL, and incubated on ice for 15 minutes
5. Following fixation, cells were pelleted at 3000rcf for 1 minute at 4°C and supernatant was aspirated (subsequent changes of buffer leading up to FACS, including washes, all
10 involve this step unless stated otherwise).
6. Cells were washed three times with 1mL PBS containing 2U/ μ L RNasin, after which they were permeabilized with 1mL of ice-cold 100% methanol and transferred to 1.5mL non-stick tubes.
7. Cells were then washed once in blocking buffer (5% secondary antibody host – donkey –
15 serum) containing 0.4U/ μ L RNasin to get rid of residual methanol and then blocked with blocking buffer containing 2U/ μ L RNasin at 4°C for 30 minutes with gentle agitation (1~2Hz on standard titer plate shaker).
8. During this time, primary antibody solutions were prepared in FACS buffer (1% BSA in PBS) and pre-incubated with 2U/uL RNasin for 10 minutes at RT.
- 20 9. Following blocking, cells were resuspended in primary antibody solution at a density of $\sim 10^7$ cells/mL and incubated overnight at 4°C with gentle agitation. The staining targets were primarily transcription factors to allow for identification and FACS-purification of different cell types found within heterogeneous populations.
10. The following morning, secondary antibody solutions were prepared in FACS buffer and
25 pre-incubated with 2U/uL RNasin for 10 minutes at RT
11. Cells were transferred to secondary antibody solution at a density of $\sim 10^7$ cells/mL and incubated for 1 hour in the dark at 4°C with gentle agitation.
12. Cells were transferred to fresh FACS buffer containing 2U/uL RNasin at a density of $\sim 10^7$ cells/mL and sorted via FACS (BD FACS Aria) into 1.5mL tubes pre-coated with
30 FACS buffer to minimize sticking and containing 100 μ L to 500 μ L FACS buffer with 2U/uL RNasin.

- 27 -

Protocol

The protocol, in this Example, represents an improvement over the protocol recommended by the manufacturer. The improved method allowed microgram quantities of RNA having a RIN >9.0 to be obtained.

5 Following FACs sorting of stained cells, roughly 10^6 cells were pelleted via centrifugation in a 1.5 ml tube. The protocol then proceeded as follows:

- 1) 150 μ l Buffer PKD was added to the pelleted cells and mixed by vortexing.
- 2) 10 μ l Proteinase K was then added to the sample, and mixed gently by pipetting up and down.
- 10 3) The sample was incubated at 56°C for three (3) hours. At this point, an optional 'pause' step can be implemented, during which the sample can be stored at 4°C overnight, without adversely affecting yield or quality of eluted RNA.
- 4) The sample was then incubate on ice for three (3) minutes, then centrifuged for 15 minutes at 20,000 x g (13,500 rpm on microcentrifuge).
- 15 5) The supernatant was then transferred to a new microcentrifuge tube taking care not to disturb the pellet.
- 6) DNase Buffer Booster equivalent to a tenth of the total sample volume (approximately 16 μ l) and 10 μ l DNase I stock solution was added to the sample. The sample was mixed gently by inverting the tube.
- 20 7) The sample was then incubated at room temperature for 15 minutes.
- 8) 320 μ l Buffer RBC was then added to the sample, and mixed thoroughly.
- 9) 1120 μ l of ethanol (100%) was then added to the sample, and was mixed well by pipetting.
- 10) 700 μ l of the sample was immediately transferred to an RNEASY® MINELUTE® spin
25 column placed in a 2 ml collection tube, and centrifuged for 15 seconds at 8000 x g (10,000 rpm). The flow through was discarded.
- 11) Step 10 was repeated until the entire sample had passed through the spin column.
- 12) 500 μ l Buffer RPE was then added to the spin column, and centrifuged for 15 seconds at 8000 x g (10,000 rpm). The flow through was discarded.
- 30 13) Step 12 was repeated.
- 14) The spin column was then placed in a new 2 ml collection tube, and was centrifuged at full speed for five (5) minutes.

- 28 -

15) The spin column was then placed in a new 1.5 ml collection tube. 20 µl of RNase-free water was added directly to the spin column membrane. The sample was then centrifuged at full speed for one (1) minute to elute the RNA.

5 Results

Depending on the size of the purified population obtained via FACS, a total of 0.2~2ug of RNA was obtained.

The RNA quality was RIN $\sim 9.2 \pm 1.2$ (mean $\pm$ standard deviation), as compared to RIN of $\sim 6.2 \pm 2.8$ for RNA that was not extracted according to this protocol.

10 Representative profiles are provided in Figure 4 (left panel is the profile of RNA achieved using the methods disclosed herein, and right panel is the control profile).

Example 2

Isolation of intact RNA from fixed cells (small cell numbers to single cells)

15 Table 1: Reagents

Reagent:	Company:	Product Number:
RNasin Plus	Promega	N2615
BSA fraction V	Gemini Bioproducts	700-106P
Triton X100	Sigma	T8787
Tween20	Sigma	P9416
10x PBS	Ambion	AM9625
DYNABEADS®		
Oligo(dT) ₂₅ (dT 25 beads)	Life Technologies	61002
miRNeasy FFPE Kit	Qiagen	217504
16% PFA aq	EMS	15710
20X SSPE	Ambion	AM9767
RNase Free H2O	Ambion	AM9932

25 Table 2: Buffer Compositions

PB (Permeabilization Buffer)	1xPBS/0.1%TX100/1%BSA/1:400RNasin
SB (Staining Buffer)	1xPBS/1%BSA/1:400RNasin
PBST	1xPBS+0.05%Tween20
CS (Cell Capture Solution)	Proteinase K Solution in PKD Buffer (1:16) (components supplied with Qiagen miRNeasy FFPE Kit)
HB (Hybridization Buffer)	2XSSPE/0.05%Tween20/1:400RNasin

- 29 -

Equipment

RNase-free Hood: AirClean Systems - AirClean600 PCR Workstation

Centrifuge: Beckman Coulter - Allegra X14R 120V - Swing Bucket Rotor

5 Flow Cytometer: BD FACS ARIA

Fixation and Storage

Dissociated, live, cultured, human cells were in single-cell suspension in 1xPBS within either a 15ml conical tube (>1M cells) or a 1.7ml tube (<1M cells). All steps other than
10 centrifugation were performed in an RNase-free hood (Airclean600, AirClean Systems). Cells were centrifuged for 3min at 335xG to pellet then the supernatant was aspirated. The cells were resuspended in 4% PFA in PBS (0.5ml per million cells) and incubated on ice for 15 minutes. The cells were pelleted at 524xG and supernatant removed. The cells were washed one time in 2ml Staining Buffer (SB). Aliquots of two million cells were then resuspended in SB (200µL
15 per 2 million cells) and stored at -80°C in 1.7ml tubes.

Permeabilization and Staining

All steps other than centrifugation were performed within an RNase-free hood. The cells were thawed by hand warming then immediately put on ice. 1ml Permeablization Buffer was
20 added to cells and they were incubated 10 minutes on ice. The cells were then centrifuged for 3min at 524xG to pellet, and the supernatant was removed. An antibody cocktail in SB was added to the cells to stain them for various antigens, as well as other markers including DAPI. This mix was incubated 30 minutes on ice. The cells were washed in 1ml of SB, as described above. After the wash the cells were resuspended in SB (750µl per million cells). The cells
25 were strained through a 70µm filter. Cells were then be sorted based on their expression of extracellular or intracellular antigens.

Cell Sorting into Collection Buffer

Cells of the designated profile were sorted into collection tubes.

30 For samples of 1-100 sorted cells, cells were sorted directly into PCR tubes containing 5µl of Capture Solution (CS). For samples of 101-1000 cells, cells were sorted directly into PCR tubes containing 100µl of CS, while for samples of greater than 1000 cells were sorted

- 30 -

directly into 1.7 ml tubes containing 100µl of SB. After collection, cells were centrifuged for 3 minutes at 524xG, the supernatant was removed and 100µl of CS was added. Cells were frozen at -80 until ready to harvest RNA.

Samples of more than 1000 cells were not tested with the dT25 bead purification system described below. These samples may be purified by starting at step 10 of “Purification of Total RNA from microdissected FFPE Tissue Sections” protocol from the miRNeasy FFPE Kit (Qiagen – 217504).

Reverse Crosslinking

Cells in CS were thawed by hand warming, vortexed for 5sec at maximum speed, and spun down quickly without forming a pellet for ~1sec in microcentrifuge. The cells were then incubated 1 hour at 56°C in a thermal cycler.

RNA Purification (using dT25 beads)

For purification of samples with fewer than 100 cells, 10µl of dT25 beads were washed 3 times in 1x Hybridization Buffer (HB). The beads were finally resuspended in 2xHB in half of the original volume to create a 2x bead concentrate.

For purification of samples with 101-1000 cells - 25µl of beads per sample were washed 3x in 100µl HB and then resuspended in 100µl 2xHB to create a more dilute bead solution.

For samples of 100 cells or fewer cells, 5µl of prewashed 2x dT25 beads were added 1:1 to the samples in 5µl of CS. For samples 101-1000 cells, 100µl of the diluted dT25 beads were added 1:1 to make a final volume of 200µl. The beads were mixed until uniform by flicking. Samples were then incubated 1min 56C in a thermal cycler, then 10min at room temperature with mixing, then 1min on ice. The beads were then washed three times in 100µl 1xHB then once in 100µl PBS. These washes were performed by resuspending the beads by flicking, centrifuging the tubes, and then using a plate magnet (Life Technologies) to pellet the beads. After the last wash all excess PBS was removed from the beads by pipet. Then 2.8µl of RNase free water was added. The beads were resuspended by gentle flicking while avoiding the liquid sticking to the sides and cap. The tubes were immediately incubated at 80°C in a thermal cycler with a 90°C hot lid for 2 minutes. The contents of the tubes were quickly spun down in a microcentrifuge, placed on the magnetic rack, and the mRNA containing supernatant was transferred to a clean PCR tube. The pure mRNA was stored at -80°C.

Results

To improve recovery to RNA from fixed cells, particularly at the single cell level, we compared the yield of RNA purified using Qiagen MinElute columns with dT25 magnetic beads. RNA was purified from 1, 10 or 100 H1 hES cells or 10, 100, or 1000pg of Total Brain (TB) RNA using both methods, and then compared by qRT-PCR (Figure 1E). We detected, on average, about 2 fold more RNA from cells after purifying with dT25 beads. We did not detect improved mRNA detection for most genes using dT25 beads to purify TB RNA. This shows that dT25 beads can more readily detect many cell-associated mRNAs.

We tested if mRNA purified from single fixed, permeabilized, stained and sorted single H1 hESCs would be suitable for genome wide transcriptional profiling. Single live, or fixed permeabilized and DAPI-stained cells were subjected to full length polyadenylated transcript amplification using the SmartSeq2 method (Picelli, Bjorklund et al. 2013). Prior to amplification, cells were either applied directly to the cell lysis buffer, or were reverse crosslinked and mRNA purified with dT25 beads. Comparable amounts of cDNA were amplified in all conditions except for fixed cell sorted directly into the cell lysis buffer (Fixed/Live) (Figures 2A and B). cDNA libraries were generated by Tagmentation (Nextera XT Kit) and sequenced to a moderate depth (7.0M reads +/- 0.9M). Cells from all conditions except fixed cells processed by the live cell protocol gave comparable frequencies of mRNA genome only, and ERCC spike in control, and unmapped reads (Figure 2C). In contrast, cells processed with the fixed cell protocol gave significantly fewer mitochondrial, ribosomal RNA, and genome-specific reads than live cells processed with the live cell protocol. About 10,000 unique genes were identified from each condition except the Fixed/Live samples which only detected ~6000 unique genes per cell (Figure 2D). Each single cell from all procedures but the Fixed/Live cell showing good correlation, clustering independent of protocol (Figure 2E), and showed few or no statistically significant differentially expressed genes by DESeq (Anders and Huber 2010) (Figure 2F). These data show that this protocol allows unbiased transcriptional profiling from single fixed, permeabilized, and stained H1 hESCs.

We further confirmed that multiple cell populations we detected by gene expression from human embryonic stem cells differentiation toward a neural fate (Figure 3A). Upon sequencing, we observed that all single cells contained high expression of housekeeping genes (Figure 3B) but showed high expression of PAX6 transcripts only in the PAX6+ sorted cells, and high

- 32 -

expression of POU5F1 transcripts only in half of the PAX6- cells (Figure 3C). Importantly, high expression of POU5F1 protein was only detected in half of the PAX6- cells.

Example 3

5

Identifying networks governing human radial glial diversity and lineage with fixed single-cell transcriptomics.

10 The diversity of cell types that populate the human neocortex is derived from cortical progenitors and is thought to contribute to higher cerebral function. Yet, distinct types of radial glial progenitor cells (RGs) populating the germinal zone of the developing human neocortex have largely been distinguished by position and morphology alone. Molecular markers that identify different progenitor types are unknown and the genetic pathways that control development, fate and function of RGs are poorly understood. To determine the molecular

15 diversity of human RGs we developed and validated a new method for transcriptome profiling of individual fixed, stained, and sorted cells. Using this method and differential gene expression analysis, we show that RGs can be classified into ventricle zone (VZ)-enriched RGs (vRGs) and outer subventricular zone (oSZ)-localized RGs (oRG). The oRG cells uniquely express HOPX, a tumor suppressor that can inhibit immediate early genes (IEGs) specifically expressed in vRGs.

20 Computational analysis of our data in conjunction with other large data sets suggests that the expression of vRG markers are maintained through the binding of SRF transcription factors while oRG-specific genes are repressed by the REST complex. Expression of HOPX inhibits the SRF complex leading to the decreased expression IEGs in vRG cells, and transition to oRG identity. This network and lineage progression is further supported by our computationally

25 defined lineage analysis, and spatial and temporal expression analysis in Macaque and Human gene expression atlases. Our findings provide an important step towards a molecular understanding of the gene expression networks that control lineage and function of recently-evolved human neocortical progenitors including oRGs.

30

Methods

Cell isolation from fetal cortex

Tissue was provided by the Birth Defects Research Laboratory (BDRL) at the University of Washington, who obtained appropriate written informed consent and provided available non-identifying information for each sample. Cortical pieces were divided into one half for sectioning and the other half for cell isolation. The half for sectioning was fixed in 4% PFA in PBS overnight at 4°C, then cryoprotected in 30% sucrose in PBS for 48-72 h, rinsed briefly with PBS and embedded and frozen in OCT. The other half (approx. 0.25 - 0.5 mL volume) was minced into small pieces with #5 forceps (Fine Science Tools, Foster City CA) in Ca²⁺- and Mg²⁺-free HBSS (14175-095, Life Technologies, Chicago IL, Chicago IL). Minced pieces were treated with 2 mL trypsin solution for 20 min at 37°C (Ca²⁺- and Mg²⁺-free HBSS, 10 mM HEPES, 0.5 mM EDTA, 0.25 mg/ml bovine pancreatic trypsin (EMD Millipore, Billerica MA), 10 µg/ml DNase I (Roche, Basel, Switzerland), pH 7.6). Digestion was quenched with 6 mL of ice-cold quenching buffer (440 ml Leibovitz L-15 medium, 50 ml water, 5 ml 1M HEPES pH 7.3–7.4, 5 ml 100x Pen-Strep, 20 ml 77.7 mM EDTA pH 8.0 [prepared from Na₂H₂EDTA], 1g bovine serum albumin [A7030, Sigma, St. Louis MO]) containing 100 µg/ml trypsin inhibitor (T6522, Sigma) and 10 µg/ml DNase I (Roche). Samples were then pelleted (220xg, 4 min, 4°C), resuspended with 1 mL of quenching buffer and triturated on ice with a P1000 pipette set to 1 mL, using 25 gentle cycles up and down without forming bubbles. The cell suspension was then diluted to 30-40 mL in quenching buffer, filtered through a 45 micron cell filter, pelleted (220xg, 10 min, 4°C), resuspended in 5 mL Staining Medium, and counted on a hemocytometer (typically ~30-50 million live cells isolated per cortical piece at ~50% viability).

Cell isolation from culture

H1 hESCs (WiCell, Madison WI) were maintained on Matrigel (Corning) in mTESR1 media (StemCell Technologies, Vancouver BC). Adherent cell cultures were dissociated with StemPro Accutase Cell Dissociation Reagent (Life Technologies, Chicago IL). The cells were centrifuged (220xg, 3 min) and the dissociation solution was removed. Cells were washed then resuspended in RNase-free Staining Buffer (SB) (1XPBS pH 7.4, 1% BSA (Gemini Bioproducts, West Sacramento, CA), and RNasin Plus at 1/400 (Promega, Madison WI)). Cells were placed on ice until fixation or sorting.

- 34 -

Fixation

The single cell suspension was fixed with 4% PFA (Electron Microscopy Sciences, Hatfield PA) in PBS on ice for 15 minutes, then pelleted (335xg, 3 min, 4°C), washed once with 1 mL SB, then resuspended in SB at 10 million cells/mL, and frozen at -80°C in aliquots.

Staining

Cells were thawed and permeabilized by resuspending and incubating cells 10 min on ice in 1XPBS (Ambion), 0.1% Triton X-100 (Sigma), 1%BSA, 1/400 RNasin Plus. One million cells were incubated in staining buffer (SB) (1XPBS, 1%BSA, 1/400 RNasin Plus) with primary antibodies for 30 min at 4°C. Antibodies used were: Alexa488 or PE-conjugated anti-PAX6 (O18-1330; BD Biosciences), PE-conjugated anti-DCX (30/Doublecortin; BD Biosciences), PerCP-Cy5.5-conjugated anti-SOX2 (O30-678; BD Biosciences), eFluor660-conjugated anti-EOMES (WD1928; eBiosciences), Alexa647-conjugated anti-OCT3/4 (40/Oct-3; BD Biosciences), rabbit anti-CYR61 (D4H5D; Cell Signaling Technology, Danvers, MA) followed by PE-conjugated goat anti-rabbit IgG (Life Technologies, Chicago IL). Cells were washed in SB, resuspended in SB containing 1ug/ml DAPI (Life Technologies), and filtered prior to sorting.

Sorting

Sorting was carried out on a BD FACS ARIA-II SORP (BD Biosciences, San Jose CA) using a 130um nozzle. Single cells were sorted into strip tubes containing 5µl of PKD Buffer (Qiagen, Germantown MD) with 1:16 Proteinase K Solution (Qiagen, Germantown MD). Populations up to 1000 cells were sorted into 1.5 ml tubes with 100 ul of PKD solution. 10,000 cell samples were sorted into 1.5 ml tubes containing 500µl of SB, pelleted, supernatant removed, and resuspended in 100µl PKD solution.

RNA extraction

Total RNA was purified using either the standard methods from the RNeasy Micro Kit (Live prep.) or a modified protocol using the miRNeasy FFPE Kit (Qiagen, Germantown, MD) (Fixed prep.). For the Fixed prep., cells in 100µl of PKD/Proteinase K solution were thawed at room temperature (RT), mixed, and incubated at 56°C for 1 h. Samples were then centrifuged at 20,000xg for 20 min. The supernatant was transferred to a new tube and 10µl of DNase Booster

- 35 -

Buffer and 10µl of DNase I stock solution were added with mixing by inversion. Samples were incubated 15 min at RT then 320µl of buffer RBC was added. 1120µl of ethanol was added and the samples were mixed. Finally, the samples were applied to MinElute columns followed by two washes with 500µl RPE Buffer. RNA was eluted into 15µl water. RNA was evaluated and
5 quantified on a Bioanalyzer 2100 (Agilent, Santa Clara CA) using the RNA Pico Kit. To normalize input, 1µl was used directly from 1,000 cells, 1µl of 10-fold diluted RNA from 10,000 cells, and 1µl of 10-fold concentrated RNA from 100 cells (Savant DNA 120 SpeedVac, Thermo Scientific).

To purify RNA with dT25 beads, samples were thawed at RT, mixed, and then incubated
10 at 56°C for 1 h in a thermal cycler with the lid set at 66°C (Bio-Rad, Hercules CA). Cells were vortexed for 10 sec, spun down, and placed on ice. Oligo dT25 magnetic beads (Life Technologies, Chicago IL) were prepared with 3 washes of 1X Hybridization Buffer (2XSSPE, 0.05%Tween20, 1:400 RNasin Plus) and then resuspended in half of the original volume of 2X Hybridization Buffer. Samples up to 100 cells or Total Brain RNA (Life Technologies) up to
15 1000pg require 5µl washed dT25 beads (0.05mg of beads per reaction).

Beads were added to reverse crosslinked samples and they were heated to 56°C for 1 min, incubated at RT 10 min to allow mRNA hybridization, then placed on ice. Beads were washed three times in 100µl of ice cold Hybridization Buffer, followed by a subsequent wash using ice cold 1X PBS. All PBS was removed and 2.8µl of RNase free water was added, the
20 beads resuspended, and the mixture was incubated at 80°C for two min then immediately pelleted on a room temperature magnet. The supernatant containing mRNA was rapidly removed and transferred to a new tube and stored at -80°C.

Smart Seq 2

25 Sequencing libraries prepared as previously reported¹¹. After reverse transcription and template switching cDNA was amplified KAPA HotStart HIFI 2X ReadyMix (Kapa Biosystems, Wilmington MA) for 19 or 22 cycles for RNA from single hESC or prenatal cells, respectively. PCR products are purified using Ampure XP beads (Beckman Coulter, Indianapolis IN). cDNA was quantified with either High Sensitivity DNA Chip (Agilent, Santa
30 Clara CA) on a Bioanalyzer 2100, or with Quant-iT PicoGreen dsDNA Assay Kit (Life Technologies; Grand Island NY) on an Enspire plate reader (PerkinElmer, Waltham MA). 1ng of cDNA used to generate RNA-Seq libraries using the Nextera XT library prep system

- 36 -

(Illumina, San Diego CA). Single cells from primary tissue appeared to contain lower cellular RNA content compared to H1 hESCs, requiring a reduction in ERCC spike-in RNAs by 10-fold and addition of three extra PCR cycles. Sequencing was carried out on Illumina MiSeq using 31 base paired-end reads. For sequencing on the HiSeq we used 50 base paired-end reads. We saw
5 few global differences in read statistics between MiSeq runs and HiSeq runs for the same samples.

RNA-Seq data analysis

Raw read data was aligned to GRCh37 (hg19) using the RefSeq annotation gff file
10 downloaded on 4/23/2013. Transcriptome alignment was performed first using RSEM²⁷, unmapped reads were then aligned to hg19 using Bowtie²⁸, and remaining unmapped reads were aligned to the ERCC sequences. Using a custom script we calculated the read mapping % of each cell to mRNA (RefSeq), mitochondrial RNA (mtRNA), noncoding RNA (mRNA, and rRNA), genome, and ERCC only. Principle component and clustering analysis were performed
15 using transcripts per million (TPM) values per gene (log2 transformed). Only high variance genes with adjusted p value smaller than 0.05 and present (mapped reads > 0) in more than 10 cells were selected.

Computational analysis

20 *WGCNA analysis:* To avoid gene modules due to co-expression within a single cell, we sub-sampled half of the cells 10 times, and calculated the topological overlap matrix for each sub-sampling. The final similarity matrix was taken as the minimum of all the topological overlap measure (TOM) similarity matrices. WGCNA clustering was performed using soft power of 4, cut height of 0.995, and minimum module size of seven. Clustering of cells was performed
25 using hclust based on module eigengenes inferred by WGCNA and “ward” distance measure. Cells were grouped into four clusters.

We detected differentially expressed genes between every pair of cell clusters using the Bioconductor limma package. The union of all such genes was defined as cluster-specific markers.

- 37 -

Human Atlas Comparisons: Comparison with LCM microarray data was performed using only cluster-specific marker genes. The microarray expression values were log transformed, then subtracted by gene means. The eigengene expression in LCM datasets were computed using gene modules inferred from the single-cell datasets. For all heatmap plots, gene vectors were
5 normalized by subtracting means.

Network Analysis: We predicted regulators of vRG and oRG as transcription factors with enriched ChIP-Seq binding sites in the vRG specific genes set or the oRG specific gene set. All ChIP-Seq datasets were downloaded from ENCODE website (ENCODE Data Coordination
10 Center)¹⁸. To address the concern that these ChIP-Seq experiments were conducted in different cell lines, we only use the peaks that overlap with DNase hypersensitive sites in fetal brain, obtained from the epigenetic roadmap database¹⁷. We then defined active transcription start sites (TSS) as TSSs with H3K4me3 sites in fetal brain. Each fetal brain DNase1 accessible sites was associated with the closest active TSS, or are within +/-5K of the active TSSs. For a given ChIP-
15 Seq dataset, we test for enrichment of overlapping ChIP-Seq at fetal brain accessible sites associated with the given gene set, relative to the overlap with all fetal brain accessible sites using hypergeometric test at pvalue cutoff 0.001. We required that the enriched ChIP-Seq peaks associated with the given gene set show enrichment against the alternative gene set based on similar hypergeometric test at pvalue cutoff 0.05. Based on this test, Elk and SRF are selected as
20 the top enriched ChIP-Seq peaks for vRG gene sets, and NRSF for oRG gene sets. To confirm whether the ChIP-Seq peaks are due to direct binding of the transcription factor, we confirmed the presence of known motifs for the factors within the ChIP-Seq peaks.

Tissue immunocytochemistry

25 Cortical tissue sections were immersion fixed in 4% paraformaldehyde in PBS for 24 h at 4°C. Tissue was washed in PBS, transferred to 30% sucrose in PBS, stored in sucrose (Sigma) for 2 days (or until tissue sank). Tissue was embedded in Tissue-Tek O.C.T (Sakura Finetek, Torrance, CA) and stored at -80°C. For staining, 25µm coronal sections were cut, adhered to slides, and stored at -80°C. Slides were thawed and dried for 15 min at RT. PBS was
30 dripped onto slide to wash away OCT and slides were then placed in 4% paraformaldehyde fixative for 20 min at RT. Slides were washed with PBS, and antigens were unmasked with 10mM pH 6.0 sodium citrate buffer at 95°C for 10 min. Slides were then washed with PBS and

- 38 -

blocked in 10% normal donkey serum (Jackson Immunoresearch), 0.1% triton X-100 in PBS for 1 hour at RT. Primary antibodies were diluted in blocking solution and applied to slides overnight at 4°C. Slides were washed in PBS and incubated in secondary antibodies diluted 1:400 in blocking solution for 1 hour at RT. Slides were washed with PBS and counterstained with 1µg/mL DAPI (Life Technologies) for 1 min at RT and mounted with Fluormount G (Southern Biotech). Primary antibodies used were: Goat anti-SOX2 (Santa Cruz Biotechnology; SC17320), Rabbit anti-PAX6 (BioLegend; prb-278p); Rat anti-EOMES (eBioscience; WD1928); Rabbit anti-HOPX (Santa Cruz Biotechnology; FL-73), and Mouse anti-Lipocortin-1 (ANXA1) (BioLegend; 74/3). Secondary antibodies were: Alexa488 donkey anti-rat IgG (H+L); Alexa568 donkey anti-rabbit IgG (H+L), and Alexa647 donkey anti-goat IgG (H+L) (Life Technologies).

Confocal images were acquired on a Leica TCS SP8 confocal microscope. Images were acquired using sequential acquisition of individual channels to prevent bleed through. Representative maximum intensity projections were created from four adjacent optical sections from the middle of a z-stack. Epifluorescence images of the full cortical thickness were acquired on a Nikon Eclipse Ti (Melville, NY) with a motorized stage using a 10X objective, and the images were stitched together using MetaMorph (Sunnyvale, CA) software. Brightness and contrast adjustments were made using Adobe Photoshop CS6 and assembled using Adobe Illustrator CS6 (San Jose, CA).

Results

Describing the diversity of radial glial (RG) progenitors and the complex intracellular molecular networks that generate them is essential for understanding human brain development. RGs are present in the ventricular zone (VZ) and are bi-polar with an attached end foot at the ventricular surface (Figure 5A)¹. In contrast, outer subventricular zone (oSZ) RG cells (oRGs) can be distinguished by their position in the oSZ, a lack of an apical endfoot, and a basal process that can extend to the pial surface²⁻⁴. The oRGs are thought to be important in the cortical expansion observed in gyrified brains and the huge increase in size and complexity of the human brain. (Figure 5A)^{1,5}. Since RGs comprise only a small subset of cells in the human cortex, analysis of these cells has largely been limited to morphological descriptions with antibodies for only a couple markers to confirm cell identity²⁻⁴, molecular characterization of micro-dissected tissue which contains an unknown variety of cell types^{6,7}, or cells sorted based on cell surface

- 39 -

markers^{8,9} (Figure 5B). Because of these issues, we still lack specific molecular markers of oRG cells and have only a limited understanding of human progenitor diversity.

Understanding the full diversity of RG progenitors requires transcriptomic profiles of larger numbers of single cells, ideally from targeted subpopulations. Our strategy to study the diversity of progenitor cells in a developing human brain was to isolate individual progenitor cells from human fetal tissue based on the expression patterns of canonical transcription factors SOX2 and PAX6, profile the transcriptomes of single cells to identify the RG diversity, and computationally determine the gene expression networks controlling progenitor biology and lineage.

Although recent advances have allowed the profiling of fixed cells labeled by intracellular markers¹⁰, single-cell profiling from fixed, stained and sorted cells was not thought possible due to degradation and mRNA loss during fixation, staining and sorting. To overcome this challenge, we first developed a protocol to extract intact RNA from populations of fixed human embryonic stem cells (hESC), and determined its efficacy in comparison to extraction of RNA from live cells (Figure 6). Analysis of total RNA showed high quality RNA integrity numbers (RINs) from 10,000 cells to 100 fixed cells using our protocol (Figures 1A-1B), with no loss of total RNA (Figure 9). Fixation also did not introduce bias in mRNA detection as assessed by quantitative reverse transcriptase polymerase chain reaction (qRT-PCR); high correlations between live and fixed cells were detected for every gene tested (Figure 1D).

Using this new mRNA recovery method, we further adapted our technique to single cells (Figure 1E). Live or fixed H1 hESC RNA-Seq libraries were prepared transcriptomic profiling using the full length mRNA amplification method SmartSeq2¹¹. The “Fixed” prep amplified comparable amounts of cDNA from individual fixed and live cells, and those cDNAs were prepared into RNA-Seq libraries and sequenced (Figures 2A-2B and Figure 6). There were no significant differences in frequencies of reads mapping to different classes of RNA between live and fixed cells the “Fixed” prep (Figure 2C and Figure 10A). Furthermore, Pearson correlation, and differential gene expression did not distinguish live from fixed cells (Figures 2D and 2F). Less than thirty genes were differentially detected between live and fixed cells; and many of those genes are known to lack or have differential polyadenylation (Table 3)¹². These data show that our method faithfully recovers single-cell RNA from permeabilized, stained, and sorted cells that can be used for transcriptomic profiling.

- 40 -

Table 3 consists of differentially expressed genes in hESCs between conditions. Differential expression is based on DESeq. Significantly differentially expressed genes ($\text{padj} < 0.01$) are shown with their fold change. The “LL LF DESeq genes” tab compares Live/Live versus Live/Fixed, and “LL LF DESeq genes” tab compares Live/Live versus Fixed/Fixed.

5 Mitochondrial, Histone, rRNA genes are indicated, as well as genes previously shown to undergo differential polyadenylation¹². Fixed/Live comparisons were omitted since there were too many differentially expressed genes, and no differentially expressed genes were detected between Fixed/Live, and Fixed/Fixed at the indicated threshold.

Next we used our method to characterize progenitor diversity in the developing human
10 cortex. The germinal zones of the human cortex, including the VZ, inner, and outer subventricular zones (iSZ and oSZ), have a mixture of progenitor and other cell types (Figure 5A), making it difficult to resolve cell type-specific gene expression signatures by position alone⁴. RGs in the VZ and oRGs in the oSZ have been reported to co-express PAX6 and SOX2, but lack the intermediate progenitor cell (IPC) marker TBR2^{3,4}. Antibody staining of human
15 cortical tissue harvested at 19 post-conceptual weeks (PCW) confirmed intermingled populations of presumed early IPCs (SOX2⁺, PAX6⁺, TBR2⁺ (“SPT”)), presumed RGs (SOX2⁺, PAX6⁺, TBR2⁻ (“SP”) cells, more mature IPCs (TBR2⁺, SOX2⁻), and cells unlabeled by any of our markers (Figures 5A-5B). In order to profile RGs and early IPCs, we isolated SP and SPT cells from a dissociated human cortex at 18 PCW (Figure 7A). We only profiled cells in G₀-G₁
20 phase to reduce cell-cycle dependent gene expression variation. Only $3.2 \pm 0.9\%$ of cells were SOX2⁺ PAX6⁺, and this population was further divided into TBR2⁺ cells (SPT) ($2.1 \pm 0.6\%$) and TBR2⁻ cells (SP) ($1.0 \pm 0.3\%$) (Figure 7B, Figure 11A, Table 4). Single-cell transcriptomic profiles were generated from these sorted cells (96 SP and 24 SPT cells from one 18 PCW sample); mRNA was harvested using our method, and cDNA was amplified with SmartSeq2¹¹
25 (Figures 12A-12F). Data from single cells with more than 50,000 mapped reads each and detected an average of 3500-4000 unique genes per progenitor (Figure 12F). A similar quantity of unique genes was detected from live prenatal human single cells from an independent study¹³. Gene expression corresponded to the sorted populations; PAX6 and SOX2 mRNAs were expressed in most single cell libraries (89% and 81%, respectively), and TBR2 mRNA was
30 enriched in the SPT versus the SP population (82% and 26%, respectively) (Fig. 7C). Lastly, our cells clustered with 12 previously identified RGs, and two IPCs profiled from an earlier

- 41 -

independent study (Figure 13)¹³. These data demonstrate SP and SPT cells strongly resemble independently profiled RGs and IPCs, despite substantially different processing methods.

Our method allowed us to sample an order of magnitude more RGs and IPCs than had been previously analyzed by genome-wide transcriptomics¹³, and to identify progenitor

5 heterogeneity. Principal component analysis (PCA) on high variance gene set was sufficient to separate most of the RGs from the IPCs (Figure 14). In addition, we identified six modules of co-expressed genes by weighted gene co-expression network analysis (WGCNA)¹⁴ (Figure 15). Module eigengenes were used to cluster cells into four major classes: three SP and one SPT (Figure 7D). Modules 1 and 2 reflected the division of the majority of SP from SPT cells, and
10 contained canonical markers of RGs including VIM, and IPCs including EOMES (TBR2), HES6, and NEUROG1 (Figure 7E, Figure 15). Modules 5 and 6 contained many cell cycle-related genes, and were enriched in SPT cells. Consistent with this, analysis of DNA content showed twice as many SPT cells are in the S-G₂-M phases of the cell cycle as SP cells (Figure 11B). Module 3 revealed an RG subpopulation significantly enriched for the immediate early
15 genes (IEGs) EGR1 and FOS¹³. SP cells expressing EGR1 and FOS still had some diversity based on the variable expression patterns of CXCL12, ANXA1 and CYR61 (data not shown). FACS analysis confirmed FSC-Seq data that SP, but not the SPT cells could be divided into two subpopulations based on the expression of CYR61 protein (Figure 16). The module 4 eigengene was complementary to module 3 in the SP cells (Figure 7E), and is composed of genes such as
20 FAM107A, HOPX, and SLCO1C1 (Figure 7G). Many genes in module 4 have no previously known roles in cortical development, and FAM107A and ANXA1 are genes uniquely expressed in human but not mouse RG cells¹⁵. Module 4 genes were enriched in the oSZ compartment relative to the VZ in 21 but not 15 or 16 PCW cortical samples based on analysis of the BrainSpan Atlas of the Developing Human Brain, while genes in module 3 were always
25 enriched in the VZ (Figure 7F, Figures 17A-17C)⁷. Strong HOPX expression is only observed at embryonic day 70 in according to the NIH Blueprint Non-Human Primate Atlas (macaque), coinciding with the emergence of the oSZ (Figure 17C)¹⁶. Thus, our data analysis showed that RGs enriched for module 3 genes are vRGs, and those enriched for module 4 genes are oRGs^{3,4}.

To test our single cell transcriptomics analysis, we next conducted differential expression
30 analysis to identify a set of genes enriched in vRG and oRG cells. We identified nine genes enriched in our vRG single cells and 15 genes enriched in our oRG cells (Figure 8A). Using antibody staining we confirmed that the genes marked vRG and oRG cells. ANXA1 showed

- 42 -

signal only in the VZ (except for vascular staining), while oRG marker HOPX showed high expression in the oSZ and iSZ, and very faint signal in vRG cells of a 16 PCW cortex (Figure 8B). At later stages such as PCW 19 and 21, HOPX was almost never detected in the VZ, and is exclusively in the oSZ (Figure 8C, 17B, and 18). HOPX⁺SOX2⁺ cells in the oSZ displayed oRG morphology with basally directed phospho-VIM⁺ processes, and rarely exhibited apical process anchored at the ventricular surface (Figure 8D-8E). HOPX⁺ progenitors in the oSZ were proliferative since more than 15% of HOPX⁺SOX2⁺ cells co-stained with Ki67, while HOPX-SOX2⁺ cells (presumably IPCs) were significantly more proliferative (Figure 8F). HOPX was a very specific marker of oRG progenitors: more than 80% of SOX2⁺ cells in the oSZ co-stained with HOPX, and less than 1% of HOPX⁺ cells in the oSZ are SOX2⁻ (Figure 8G-8H). These data confirm that genes in modules 3 and 4 mark vRG and oRG cells respectively, and that HOPX is a precise molecular marker of human oRG cells.

To define molecular networks that control lineage and biology of vRG and oRG cells, we analyzed the differentially expressed gene list to identify a transcription factor network. We identified DNase1 hypersensitive sites from fetal human brain¹⁷ proximal to vRG- or oRG-specific genes. Transcription factor-specific ChIP data from the Encode project (ENCODE Data Coordination Center)¹⁸ was used to identify transcription factor occupancy at open sites. There was a statistically significant enrichment in SRF and ELK binding sites proximal to vRG genes, and NRSF/REST binding sites proximal to oRG specific genes (Figure 8I). It is known that SRF and ELK regulate FOS and EGR1^{19,20}, loss of SRF leads to higher expression of SOX2 in the mouse germinal zone²¹, and that HOPX inhibits the function of SRF in developing cardiac^{22,23}. These findings point to a model in which SRF maintains expression of the vRG marker, and the activation of HOPX leads to the inhibition of SRF and the depletion of vRG genes. This inhibition leads to the de-repression of oRG genes through inactivation of the NRSF/REST repressor complex and the basal migration of the nascent oRG cells. Consistent with the prediction, HOPX expression was shown to inhibit vRG genes FOS, JUN, EGR1, and CYR61 in colorectal cancer²⁴. Furthermore, the expression of NRSF/REST inhibits neural progenitor migration during mouse neurogenesis²⁵, and has been implicated in microcephaly²⁶. Our model is also consistent with human and macaque gene expression data: HOPX expression is initiated as the oSZ begins to expand, it is first expressed in the VZ and the oSZ, and at later times is enriched in the oSZ (Figure 17C). Together, our data along with the computational

- 43 -

analysis suggest that the epithelial vRG progenitors, marked by IEGs, give rise to oRGs through the expression of HOPX and results in suppression of vRG genes.

In summary, the method we developed to study the diversity of human cortical progenitors is a new technology that enables targeted purification and efficient single-cell
5 profiling of low frequency cell populations from primary tissue or other sources, obviating the need for transgenesis or viral labeling of specific cell types, and providing RNAseq data comparable to that obtained from live cells. Using this technique, we were able to obtain a molecular characterization of the oRG and vRG progenitor classes formed in the mid-gestational human cortex, and identified new marker genes for subtypes of RGs including the first specific
10 molecular marker of oRG cells. Finally, we used our transcriptional data to direct the computational analysis of existing epigenetic data and propose molecular network that controls the identity and lineage of vRGs and oRGs.

15

20

25

30

- 44 -

Table 3. LL LF DESeq Genes and LL FF DESeq Genes

LL LF DESeq Genes:

	log2 Fold Change	pvalue	padj	Mitochondrial	Histone Cluster	rRNA	Highlighted in Yang et al
C8orf77	3.5872	2.07E-08	9.29E-06				
DPY19L4	2.7358	5.31E-05	0.00823				
ERH	-1.13342	2.64E-05	0.00435				
HIST1H2AD	6.4709	7.13E-07	0.0002		X		
HIST1H3D	8.52317	1.27E-05	0.00244		X		
JAK3	2.24639	8.77E-06	0.00184				X
MALAT1	2.88095	7.02E-19	6.29E-16				X
MIDN	-2.51419	4.81E-06	0.00111				
MTRNR2L1	2.93652	3.31E-23	4.44E-20				
MTRNR2L2	2.4768	5.00E-21	5.03E-18				
MTRNR2L6	3.21156	1.77E-14	1.34E-11				
MTRNR2L8	2.51701	2.81E-11	1.51E-08				
NRSN1	3.75355	1.02E-13	6.83E-11				
PAR-SN	2.00452	3.55E-07	0.00012				
PLEKHM3	2.30361	5.31E-05	0.00823				
TVAS5	2.72347	4.01E-37	3.23E-33	X			
ZNF429	2.9932	1.32E-05	0.00247				
c12_SSU-rRNA_Hsa	6.15776	1.58E-26	3.18E-23			X	
c17_LSU-rRNA_Hsa	6.26435	8.53E-13	5.29E-10			X	
c19_LSU-rRNA_Hsa	4.82865	1.83E-14	1.34E-11			X	
c21_LSU-rRNA_Hsa	4.3541	6.41E-05	0.00975			X	
c2_LSU-rRNA_Hsa	4.40804	2.98E-21	3.43E-18			X	
c5_LSU-rRNA_Hsa	6.13066	1.31E-11	7.52E-09			X	
c5_SSU-rRNA_Hsa	6.17057	1.12E-27	3.01E-24			X	
c7_SSU-rRNA_Hsa	7.08359	3.89E-28	1.57E-24			X	
c8_LSU-rRNA_Hsa	2.63919	3.91E-06	0.00093			X	
cx_LSU-rRNA_Hsa	2.7576	2.08E-06	0.00051			X	
cy_SSU-rRNA_Hsa	5.72161	2.49E-26	4.02E-23			X	

- 45 -

Table 3 continued. LL FF DESeq Genes:

	log2 Fold Change	pvalue	padj	Mitochondrial	Histone Cluster	rRNA	Highlighted in Yang et al
C8orf77	3.5872	2.07E-08	9.29E-06				
ERH	-1.13342	2.64E-05	0.00435				
HIST1H2AD	6.4709	7.13E-07	0.0002		X		
HIST1H3D	8.52317	1.27E-05	0.00244		X		
MALAT1	2.88095	7.02E-19	6.29E-16				X
MCM3	0.15471	0.53483	0.99978				
MIDN	-2.51419	4.81E-06	0.00111				
MTRNR2L1	2.93652	3.31E-23	4.44E-20				
MTRNR2L2	2.4768	5.00E-21	5.03E-18				
MTRNR2L6	3.21158	1.77E-14	1.34E-11				
MTRNR2L8	2.51701	2.81E-11	1.51E-08				
NRSN1	3.75355	1.02E-13	6.83E-11				
PAR-SN	2.00452	3.55E-07	0.00012				
RPS2	-0.7072	0.00013	0.01583				
SLC2A14	1.86264	0.00114	0.07952				
TMEM136	0.60586	0.29914	0.99978				
TVAS5	2.72347	4.01E-37	3.23E-33	X			
c12_SSU-rRNA_Hsa	6.15776	1.58E-26	3.18E-23			X	
c17_LSU-rRNA_Hsa	6.26435	8.53E-13	5.29E-10			X	
c19_LSU-rRNA_Hsa	4.82865	1.83E-14	1.34E-11			X	
c2_LSU-rRNA_Hsa	4.40804	2.98E-21	3.43E-18			X	
c5_LSU-rRNA_Hsa	6.13066	1.31E-11	7.52E-09			X	
c5_SSU-rRNA_Hsa	6.17057	1.12E-27	3.01E-24			X	
c7_SSU-rRNA_Hsa	7.08359	3.89E-28	1.57E-24			X	
cy_SSU-rRNA_Hsa	5.72161	2.49E-26	4.02E-23			X	

Table 4. Table describing primary tissues used.

Brain ID	Days post conception	PMI at arrival (approx)	Sex	Sample ID	ICC	Single Cell RNA-Seq	FACS analysis
4010	125 days	4 hrs	Male	H14.4010.A	N	Y	Y
4013	115 days	4 hrs	Male	H14.4013.A	Y	N	N
4019	132 days	unknown	?	H14.4019.A	Y	Y	Y
4020	127 days	~ 5 hrs	?	H14.4020.A	Y	N	N
4022	130 days	~ 4 hrs	?	H14.4022	Y	N	N
4023	96 days	~ 4 hrs	?	H14.4023.A	Y	N	Y
4029	108 days	~ 4 hrs	Female	H14.4029	Y	N	N
4010	127 days	~ 4 hrs	?	H15.4010	Y	N	N

- 46 -

REFERENCES

1. Lui, J. H., Hansen, D. V. & Kriegstein, A. R. Development and evolution of the human neocortex. *Cell* **146**, 18-36, doi:10.1016/j.cell.2011.06.030 (2011).
2. Betizeau, M. *et al.* Precursor diversity and complexity of lineage relationships in the outer subventricular zone of the primate. *Neuron* **80**, 442-457, doi:10.1016/j.neuron.2013.09.032 (2013).
3. Fietz, S. A. *et al.* OSVZ progenitors of human and ferret neocortex are epithelial-like and expand by integrin signaling. *Nature neuroscience* **13**, 690-699, doi:10.1038/nn.2553 (2010).
4. Hansen, D. V., Lui, J. H., Parker, P. R. & Kriegstein, A. R. Neurogenic radial glia in the outer subventricular zone of human neocortex. *Nature* **464**, 554-561, doi:10.1038/nature08845 (2010).
5. Dehay, C., Kennedy, H. & Kosik, Kenneth S. The Outer Subventricular Zone and Primate-Specific Cortical Complexification. *Neuron* **85**, 683-694, doi:10.1016/j.neuron.2014.12.060 (2015).
6. Fietz, S. A. *et al.* Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. *Proceedings of the National Academy of Sciences of the United States of America* **109**, 11836-11841, doi:10.1073/pnas.1209647109 (2012).
7. Miller, J. A. *et al.* Transcriptional landscape of the prenatal human brain. *Nature* **508**, 199-206, doi:10.1038/nature13185 (2014).
8. Florio, M. *et al.* Human-specific gene ARHGAP11B promotes basal progenitor amplification and neocortex expansion. *Science*, doi:10.1126/science.aaa1975 (2015).
9. Johnson, M. B. *et al.* Single-cell analysis reveals transcriptional heterogeneity of neural progenitors in human cortex. *Nature neuroscience* advance online publication, doi:10.1038/nn.3980 (2015).
10. Molyneaux, B. J. *et al.* DeCoN: genome-wide analysis of in vivo transcriptional dynamics during pyramidal neuron fate selection in neocortex. *Neuron* **85**, 275-288, doi:10.1016/j.neuron.2014.12.024 (2015).
11. Picelli, S. *et al.* Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nature methods* **10**, 1096-1098, doi:10.1038/nmeth.2639 (2013).

12. Yang, L., Duff, M. O., Graveley, B. R., Carmichael, G. G. & Chen, L. L. Genomewide characterization of non-polyadenylated RNAs. *Genome Biol* **12**, R16, doi:10.1186/gb-2011-12-2-r16 (2011).

13. Pollen, A. A. *et al.* Low-coverage single-cell mRNA sequencing reveals cellular
5 heterogeneity and activated signaling pathways in developing cerebral cortex. *Nature biotechnology* **32**, 1053-1058, doi:10.1038/nbt.2967 (2014).

14. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17, doi:10.2202/1544-6115.1128 (2005).

15. Lui, J. H. *et al.* Radial glia require PDGFD-PDGFRbeta signalling in human but
10 not mouse neocortex. *Nature* **515**, 264-268, doi:10.1038/nature13973 (2014).

16. Smart, I. H., Dehay, C., Giroud, P., Berland, M. & Kennedy, H. Unique morphological features of the proliferative zones and postmitotic compartments of the neural epithelium giving rise to striate and extrastriate cortex in the monkey. *Cerebral cortex* **12**, 37-53 (2002).

17. Roadmap Epigenomics, C. *et al.* Integrative analysis of 111 reference human
15 epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

18. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74 (2012).

19. Clarkson, R. W. E., Shang, C. A., Levitt, L. K., Howard, T. & Waters, M. J.
20 Ternary Complex Factors Elk-1 and Sap-1a Mediate Growth Hormone-Induced Transcription of Egr-1 (Early Growth Response Factor-1) in 3T3-F442A Preadipocytes. *Molecular Endocrinology* **13**, 619-631, doi:doi:10.1210/mend.13.4.0266 (1999).

20. Latinkic, B. V., Zeremski, M. & Lau, L. F. Elk-1 Can Recruit SRF to Form a Ternary Complex Upon the Serum Response Element. *Nucleic Acids Research* **24**, 1345-1351,
25 doi:10.1093/nar/24.7.1345 (1996).

21. Lu, P. P. & Ramanan, N. Serum response factor is required for cortical axon growth but is dispensable for neurogenesis and neocortical lamination. *The Journal of neuroscience : the official journal of the Society for Neuroscience* **31**, 16651-16664, doi:10.1523/JNEUROSCI.3015-11.2011 (2011).

22. Chen, F. *et al.* Hop is an unusual homeobox gene that modulates cardiac development. *Cell* **110**, 713-723 (2002).

- 48 -

23. Shin, C. H. *et al.* Modulation of cardiac growth and development by HOP, an unusual homeodomain protein. *Cell* **110**, 725-735 (2002).

24. Yamashita, K., Katoh, H. & Watanabe, M. The homeobox only protein homeobox (HOPX) and colorectal cancer. *International journal of molecular sciences* **14**, 23231-23243, doi:10.3390/ijms141223231 (2013).

25. Mandel, G. *et al.* Repressor element 1 silencing transcription factor (REST) controls radial migration and temporal neuronal specification during neocortical development. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 16789-16794, doi:10.1073/pnas.1113486108 (2011).

26. Yang, Y. J. *et al.* Microcephaly gene links trithorax and REST/NRSF to control neural stem cell proliferation and differentiation. *Cell* **151**, 1097-1112, doi:10.1016/j.cell.2012.10.043 (2012).

27. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* **12**, 323, doi:10.1186/1471-2105-12-323 (2011).

28. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25, doi:10.1186/gb-2009-10-3-r25 (2009).

EQUIVALENTS

While several inventive embodiments have been described and illustrated herein, those of ordinary skill in the art will readily envision a variety of other means and/or structures for performing the function and/or obtaining the results and/or one or more of the advantages described herein, and each of such variations and/or modifications is deemed to be within the scope of the inventive embodiments described herein. More generally, those skilled in the art will readily appreciate that all parameters, dimensions, materials, and configurations described herein are meant to be exemplary and that the actual parameters, dimensions, materials, and/or configurations will depend upon the specific application or applications for which the inventive teachings is/are used. Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific inventive embodiments described herein. It is, therefore, to be understood that the foregoing embodiments are presented

- 49 -

by way of example only and that, within the scope of the appended claims and equivalents thereto, inventive embodiments may be practiced otherwise than as specifically described and claimed. Inventive embodiments of the present disclosure are directed to each individual feature, system, article, material, kit, and/or method described herein. In addition, any
5 combination of two or more such features, systems, articles, materials, kits, and/or methods, if such features, systems, articles, materials, kits, and/or methods are not mutually inconsistent, is included within the inventive scope of the present disclosure.

All definitions, as defined and used herein, should be understood to control over dictionary definitions, definitions in documents incorporated by reference, and/or ordinary
10 meanings of the defined terms.

All references, patents and patent applications disclosed herein are incorporated by reference with respect to the subject matter for which each is cited, which in some cases may encompass the entirety of the document.

The indefinite articles “a” and “an,” as used herein in the specification and in the claims,
15 unless clearly indicated to the contrary, should be understood to mean “at least one.”

The phrase “and/or,” as used herein in the specification and in the claims, should be understood to mean “either or both” of the elements so conjoined, i.e., elements that are conjunctively present in some cases and disjunctively present in other cases. Multiple elements listed with “and/or” should be construed in the same fashion, i.e., “one or more” of the elements
20 so conjoined. Other elements may optionally be present other than the elements specifically identified by the “and/or” clause, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, a reference to “A and/or B”, when used in conjunction with open-ended language such as “comprising” can refer, in one embodiment, to A only (optionally including elements other than B); in another embodiment, to B only (optionally
25 including elements other than A); in yet another embodiment, to both A and B (optionally including other elements); etc.

As used herein in the specification and in the claims, “or” should be understood to have the same meaning as “and/or” as defined above. For example, when separating items in a list, “or” or “and/or” shall be interpreted as being inclusive, i.e., the inclusion of at least one, but also
30 including more than one, of a number or list of elements, and, optionally, additional unlisted items. Only terms clearly indicated to the contrary, such as “only one of” or “exactly one of,” or, when used in the claims, “consisting of,” will refer to the inclusion of exactly one element of

- 50 -

a number or list of elements. In general, the term “or” as used herein shall only be interpreted as indicating exclusive alternatives (i.e. “one or the other but not both”) when preceded by terms of exclusivity, such as “either,” “one of,” “only one of,” or “exactly one of.” “Consisting essentially of,” when used in the claims, shall have its ordinary meaning as used in the field of patent law.

As used herein in the specification and in the claims, the phrase “at least one,” in reference to a list of one or more elements, should be understood to mean at least one element selected from any one or more of the elements in the list of elements, but not necessarily including at least one of each and every element specifically listed within the list of elements and not excluding any combinations of elements in the list of elements. This definition also allows that elements may optionally be present other than the elements specifically identified within the list of elements to which the phrase “at least one” refers, whether related or unrelated to those elements specifically identified. Thus, as a non-limiting example, “at least one of A and B” (or, equivalently, “at least one of A or B,” or, equivalently “at least one of A and/or B”) can refer, in one embodiment, to at least one, optionally including more than one, A, with no B present (and optionally including elements other than B); in another embodiment, to at least one, optionally including more than one, B, with no A present (and optionally including elements other than A); in yet another embodiment, to at least one, optionally including more than one, A, and at least one, optionally including more than one, B (and optionally including other elements); etc.

It should also be understood that, unless clearly indicated to the contrary, in any methods claimed herein that include more than one step or act, the order of the steps or acts of the method is not necessarily limited to the order in which the steps or acts of the method are recited.

In the claims, as well as in the specification above, all transitional phrases such as “comprising,” “including,” “carrying,” “having,” “containing,” “involving,” “holding,” “composed of,” and the like are to be understood to be open-ended, i.e., to mean including but not limited to. Only the transitional phrases “consisting of” and “consisting essentially of” shall be closed or semi-closed transitional phrases, respectively, as set forth in the United States Patent Office Manual of Patent Examining Procedures, Section 2111.03.

- 51 -

What is claimed is:

CLAIMS

1. A method for purifying RNA from a fixed biological sample, comprising:
 - (a) incubating the sample in an aqueous lysis buffer at a temperature between about 50-
5 60°C for a time between about 2-4 hours;
 - (b) contacting the sample with a DNase for a time sufficient to digest DNA;
 - (c) contacting the sample with a substrate that binds RNA, and
 - (d) eluting bound RNA from the substrate.
- 10 2. The method of claim 1, wherein the fixed biological sample is a formalin-fixed cell sample or a formalin-fixed tissue sample.
3. The method of claim 1, wherein the fixed biological sample is a formalin-fixed paraffin-embedded (FFPE) sample, and further comprising an initial deparaffinization step.
- 15 4. The method of any one of claims 1-3, wherein step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature between about 52-58°C.
5. The method of claim 4, wherein step (a) comprises incubating the sample in the aqueous
20 lysis buffer at a temperature between about 54-57°C.
6. The method of claim 5, wherein step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C +/- 2°C.
- 25 7. The method of claim 5, wherein step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C.
8. The method of any one of claims 1-7, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 2-3 hours.

- 52 -

9. The method of any one of claims 1-7, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 3-4 hours.

10. The method of any one of claims 1-7, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 3 +/- 0.5 hours.

11. The method of any one of claims 1-7, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 3.0 hours.

12. The method of claim 1, wherein step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C for a time about 3.0 hours.

13. The method of any one of claims 1-12, wherein the method does not comprise incubating the sample at a temperature above 60°C.

14. The method of any one of claims 1-13, wherein the method does not comprise incubating the sample at a temperature of about 80°C or higher.

15. The method of any one of claims 1-14, wherein the lysis buffer comprises a buffering agent and a detergent.

16. The method of claim 15, wherein the buffering agent is selected from the group consisting of Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof.

17. The method of claim 15 or 16, wherein the detergent is selected from the group consisting of non-ionic, cationic, anionic and zwitterionic detergents.

18. The method of any one of claims 1-17, wherein the lysis buffer further comprises one or more substances selected from the group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers.

- 53 -

19. The method of any one of claims 1-18, wherein the lysis buffer has a pH in a range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

20. The method of any one of claims 1-19, wherein a proteolytic agent either is added to the mixture in step (a) in addition to the lysis buffer, or a proteolytic agent is already comprised in the lysis buffer, the proteolytic agent being selected from the group consisting of proteases and non-enzymatic proteolytic compounds.

21. The method of claim 20, wherein the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocyanide, recombinant Bacillus proteases, Lysozyme, and a mixture thereof.

22. The method of any one of claims 1-21, wherein RNA once eluted has an RNA integrity number (RIN) between about 5.0-10.0.

23. The method of claim 22, wherein the eluted RNA has a RIN of at least 5.0, at least 5.5, at least 6.0, at least 6.5, at least 7.0, at least 7.5, at least 8.0, at least 8.5, at least 9.0, or at least 9.5.

24. The method of any one of claims 1-23, producing an amount of eluted RNA of at least 0.2 µg, at least 2 µg, or at least 20 µg.

25. The method of any one of claims 1-24, wherein the RNA once eluted has a purity of at least 1.8 as assessed by the 260 nm/280 nm ratio.

26. The method of any one of claims 1-25, wherein the RNA once eluted comprises no more than 0.01% or 0.001% (w/w) of DNA.

27. The method of any one of claims 1-26, wherein the fixed biological sample comprises at least 10^5 , at least 10^6 , at least 10^7 , at least 10^8 , at least 10^9 , at least 10^{10} , at least 10^{11} , or at least 10^{12} cells.

- 54 -

28. The method of any one of claims 1-27, wherein the fixed biological sample comprises embryonic stem cells, pluripotent stem cells, somatic cells, germ cells, or gametes.

29. The method of any one of claims 1-28, wherein the fixed biological sample comprises human cells or tissue.

30. A kit for purifying RNA from a fixed biological sample, comprising an aqueous lysis buffer and instructions, wherein the instructions provide that a fixed biological sample should not be incubated in the lysis buffer at a temperature above 60°C.

31. The kit of claim 30, wherein the instructions provide that a fixed biological sample should not be incubated in the lysis buffer at a temperature of 80°C or higher.

32. The kit of claim 30 or 31, wherein the instructions provide that a fixed biological sample is incubated in the aqueous lysis buffer at a temperature between about 50-60°C, for a time between about 2-4 hours.

33. The kit of any one of claims 30-32, wherein the lysis buffer comprises a buffering agent and a detergent.

34. The kit of claim 33, wherein the buffering agent is selected from the group consisting of Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof.

35. The kit of claim 33 or 34, wherein the detergent is selected from the group consisting of anionic and zwitterionic detergents.

36. The kit of any one of claims 30-35, wherein the lysis buffer further comprises one or more substances selected from the group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers.

37. The kit of claim 36, wherein the lysis buffer has a pH in the range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

- 55 -

38. The kit of any one of claims 30-37, further comprising a proteolytic agent selected from the group consisting of proteases and non-enzymatic proteolytic compounds.

5 39. The kit of claim 38, wherein the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-C, bromocyanide, recombinant Bacillus proteases, Lysozyme, and a mixture thereof.

40. The kit of any one of claims 30-39, further comprising one or more of a DNase, RNA
10 binding substrate, and elution buffer.

41. A method for purifying RNA from a fixed biological sample, comprising:
(a) incubating the sample in an aqueous lysis buffer at a temperature between about 50-
60°C for a time between about 0.5-3 hours;
15 (b) contacting the sample with a substrate that binds RNA, and
(c) eluting bound RNA from the substrate.

42. The method of claim 41, wherein the fixed biological sample is a formalin-fixed cell
sample or a formalin-fixed tissue sample.
20

43. The method of claim 41, wherein the fixed biological sample is a formalin-fixed
paraffin-embedded (FFPE) sample, and further comprising an initial deparaffinization step.

44. The method of any one of claims 41-43, wherein step (a) comprises incubating the
25 sample in the aqueous lysis buffer at a temperature between about 52-58°C.

45. The method of claim 44, wherein step (a) comprises incubating the sample in the
aqueous lysis buffer at a temperature between about 54-57°C.

30 46. The method of claim 45, wherein step (a) comprises incubating the sample in the
aqueous lysis buffer at a temperature of about 56°C +/- 2°C.

- 56 -

47. The method of claim 45, wherein step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C.

48. The method of any one of claims 41-47, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 0.5-2 hours.

49. The method of any one of claims 41-47, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 1 +/- 0.5 hours.

50. The method of any one of claims 41-47, wherein step (a) comprises incubating the sample in the aqueous lysis buffer for a time about 1.0 hours.

51. The method of claim 41, wherein step (a) comprises incubating the sample in the aqueous lysis buffer at a temperature of about 56°C for a time about 1.0 hours.

52. The method of any one of claims 41-51, wherein the method does not comprise incubating the sample at a temperature above 60°C.

53. The method of any one of claims 41-52, wherein the method does not comprise incubating the sample at a temperature of about 80°C or higher.

54. The method of any one of claims 41-53, wherein the lysis buffer comprises a buffering agent and a detergent.

55. The method of claim 54, wherein the buffering agent is selected from the group consisting of Trips, Mops, Mes, Hepes, phosphates, borates, carbonates, and combinations thereof.

56. The method of claim 54 or 55, wherein the detergent is selected from the group consisting of non-ionic, cationic, anionic and zwitterionic detergents.

- 57 -

57. The method of any one of claims 41-56, wherein the lysis buffer further comprises one or more substances selected from the group consisting of ethanol, chelating agents, nucleophilic reagents, reducing agents, inorganic salts, guanidine salts, pH indicators, and stabilizers.

5 58. The method of any one of claims 41-57, wherein the lysis buffer has a pH in a range of from about 4 to 11, from about 7 to 10, or from about 8 to 9.

59. The method of any one of claims 41-58, wherein a proteolytic agent either is added to the mixture in step (a) in addition to the lysis buffer, or a proteolytic agent is already comprised in
10 the lysis buffer, the proteolytic agent being selected from the group consisting of proteases and non-enzymatic proteolytic compounds.

60. The method of claim 59, wherein the proteolytic agent is selected from the group consisting of proteinase K, trypsin, chymotrypsin, papain, pepsin, pronase, endoproteinase Lys-
15 C, bromocyanide, recombinant Bacillus proteases, Lysozyme, and a mixture thereof.

61. The method of any one of claims 41-60, further comprising evaluating the RNA integrity of the RNA by gene expression analysis of housekeeping genes.

20 62. The method of any one of claims 41-61, producing an amount of eluted RNA of at least 0.1 pg, at least 1 pg, at least 10 pg, at least 100 pg, or at least 1000 pg.

63. The method of any one of claims 41-62, wherein the fixed biological sample comprises 1-10 cells, at least 10, at least 50, at least 100, at least 10^3 , at least 10^4 , or at least 10^5 cells.

25

64. The method of any one of claims 41-63, wherein the fixed biological sample comprises embryonic stem cells, pluripotent stem cells, somatic cells, germ cells, or gametes.

65. The method of any one of claims 41-64, wherein the fixed biological sample comprises
30 human cells or tissue.

1/38

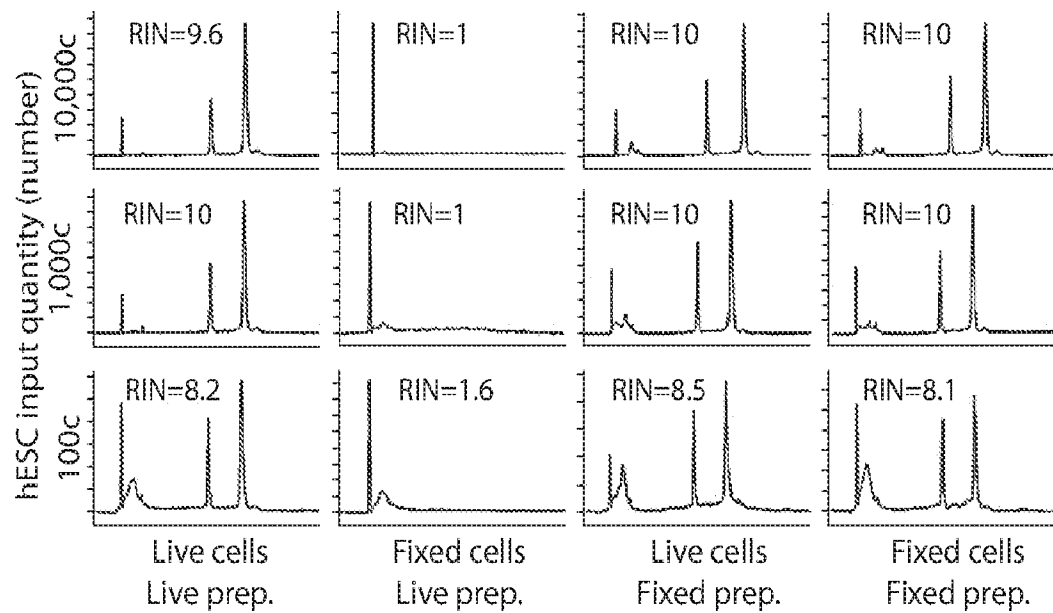


Figure 1A

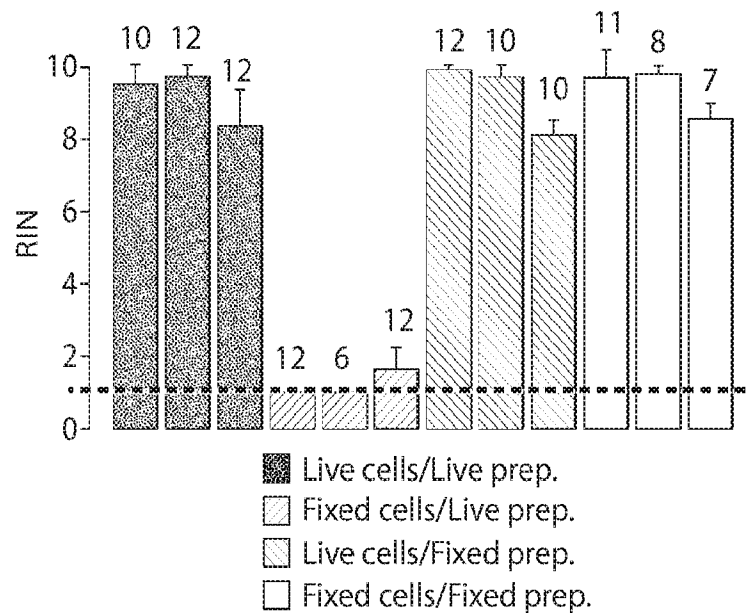


Figure 1B

2/38

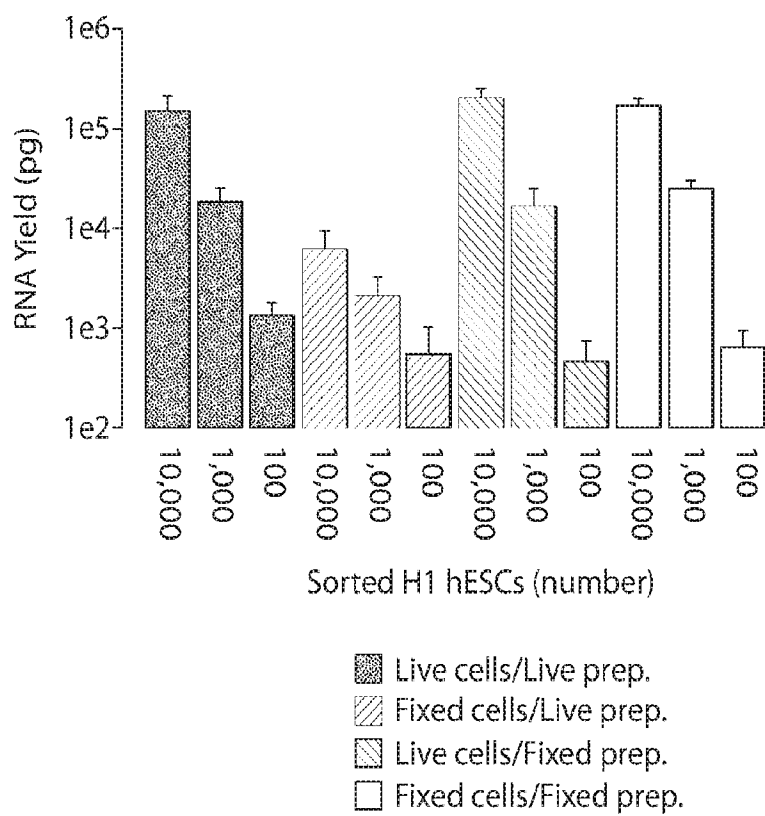


Figure 1C

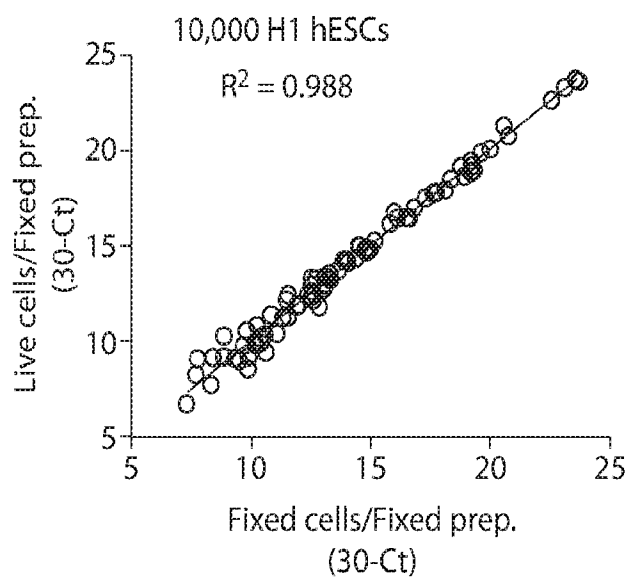


Figure 1D

3/38

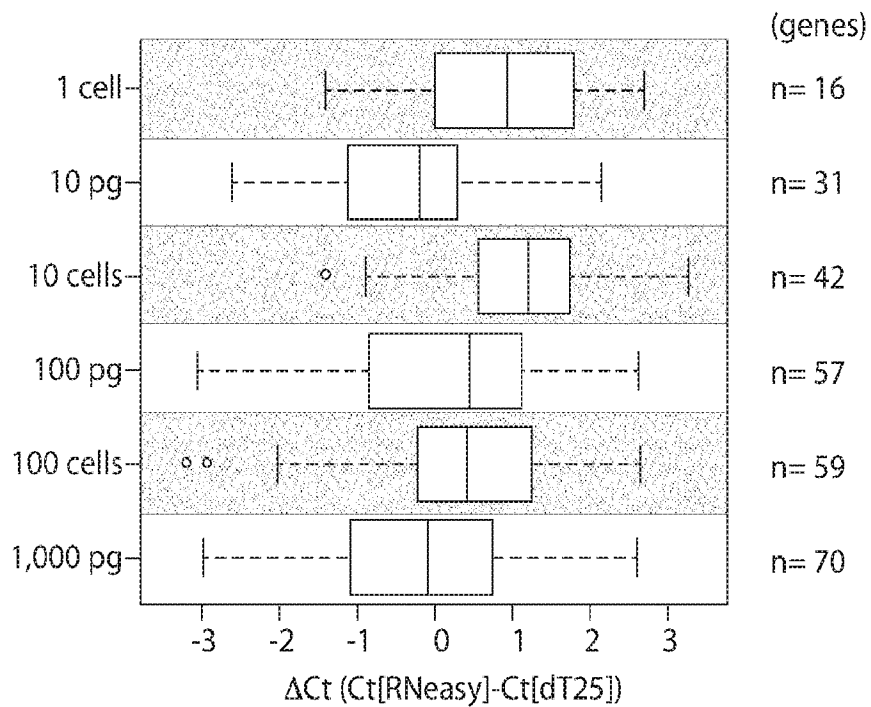


Figure 1E

4/38

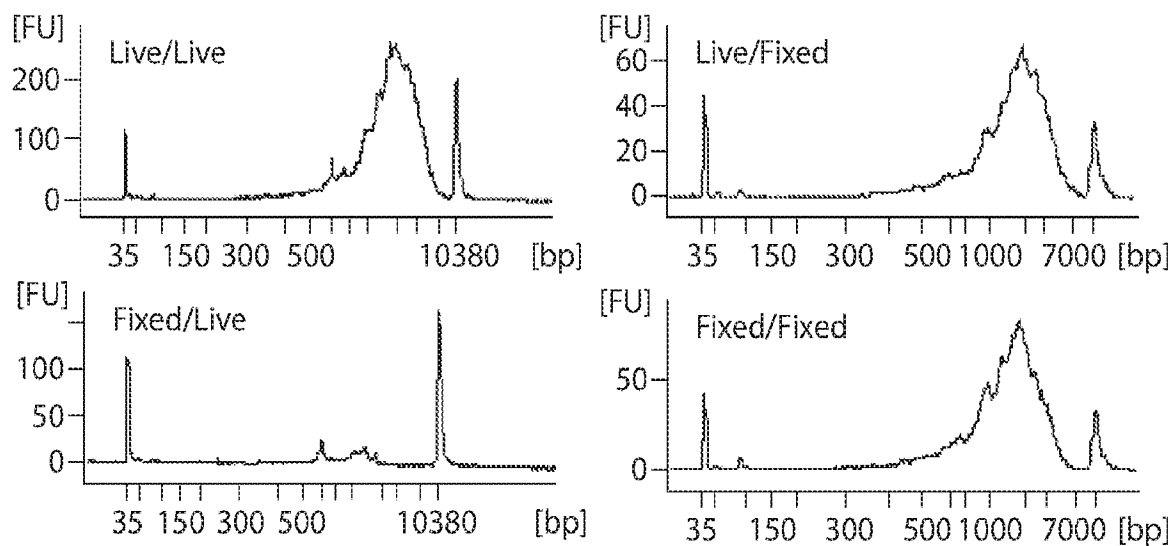


Figure 2A

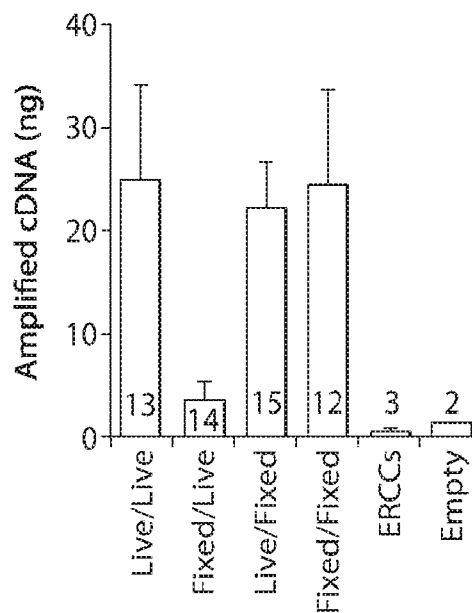


Figure 2B

5/38

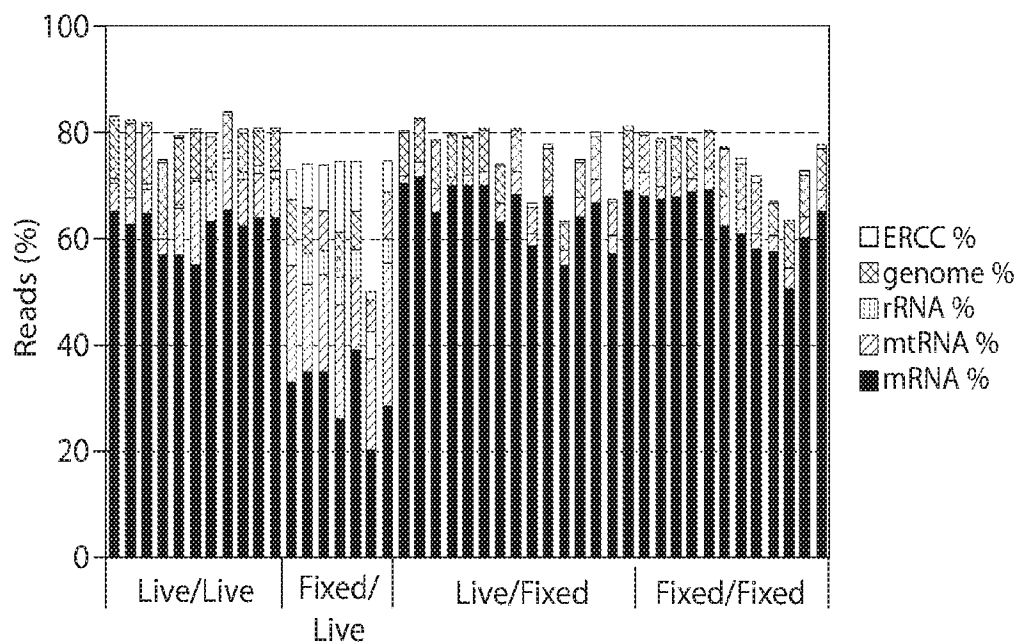


Figure 2C

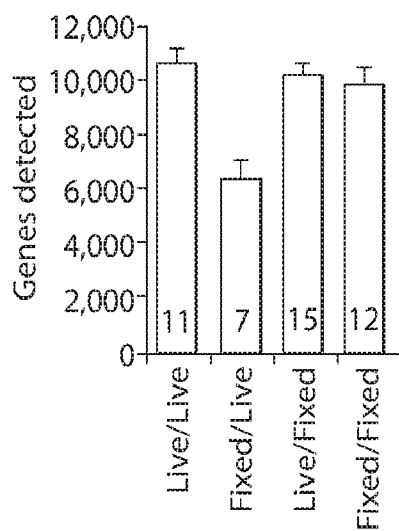


Figure 2D

6/38

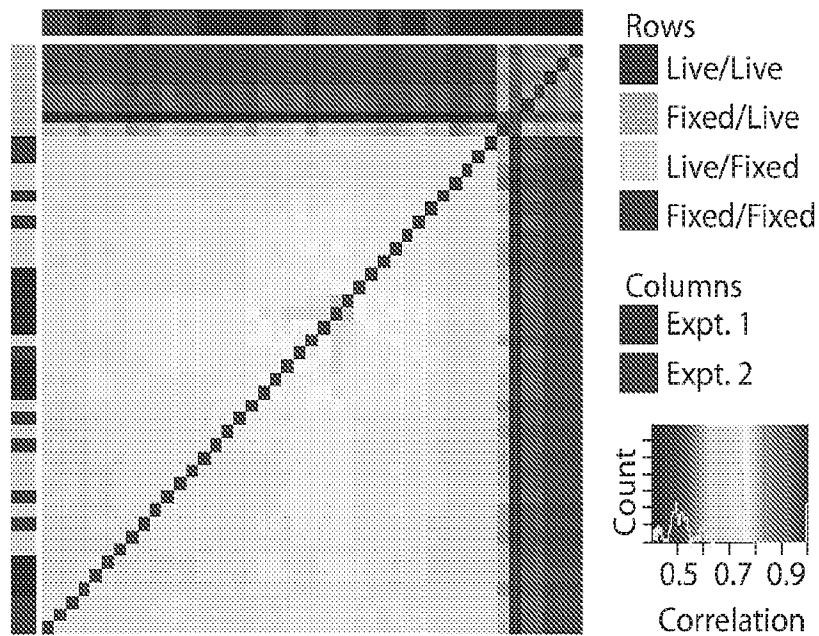


Figure 2E

Differentially expressed genes				
	Live/Live	Fixed/Live	Live/Fixed	Fixed/Fixed
Live/Live	0			
Fixed/Live	2578	0		
Live/Fixed	28	2970	0	
Fixed/Fixed	25	2535	0	0

Figure 2F

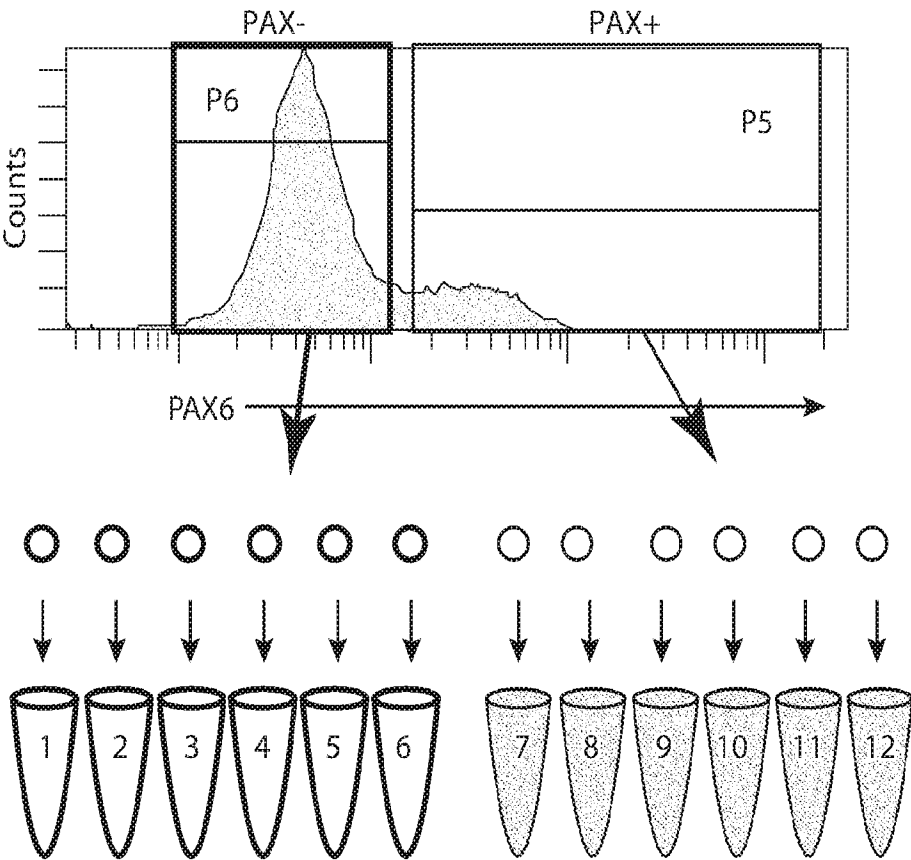


Figure 3A

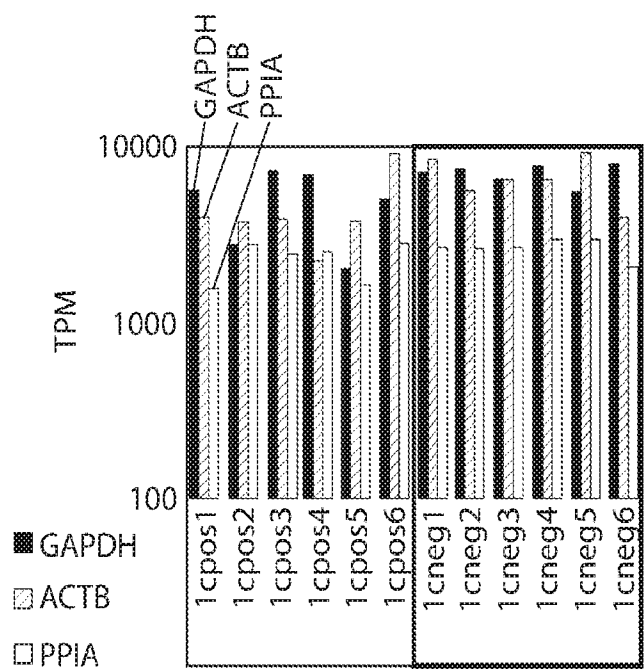


Figure 3B

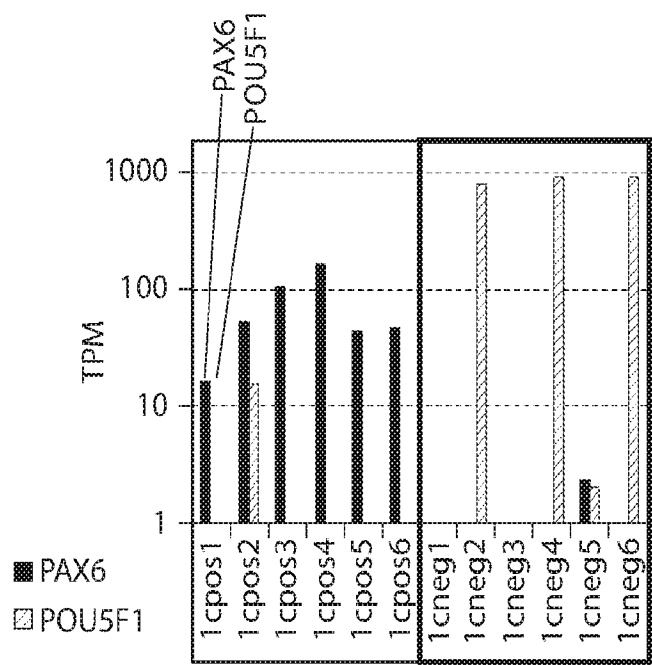


Figure 3C

9/38

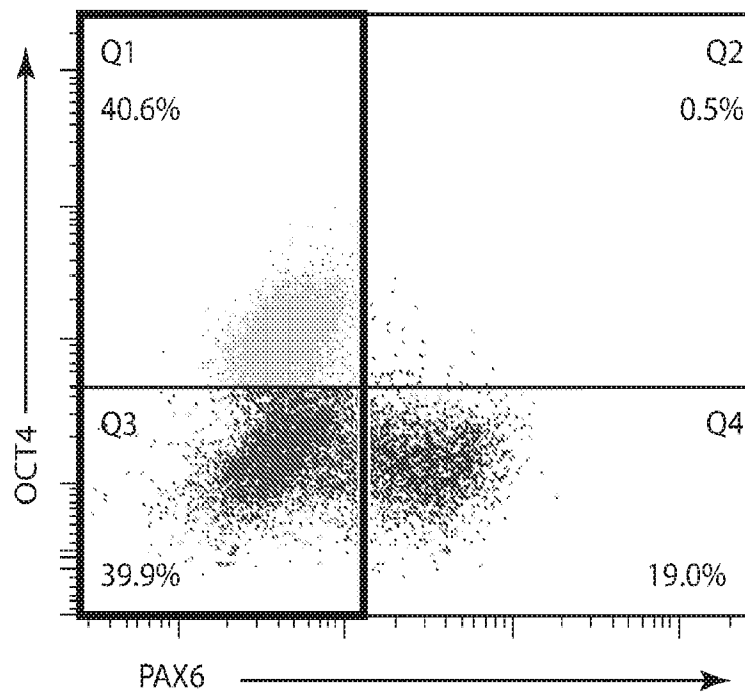


Figure 3D

10/38

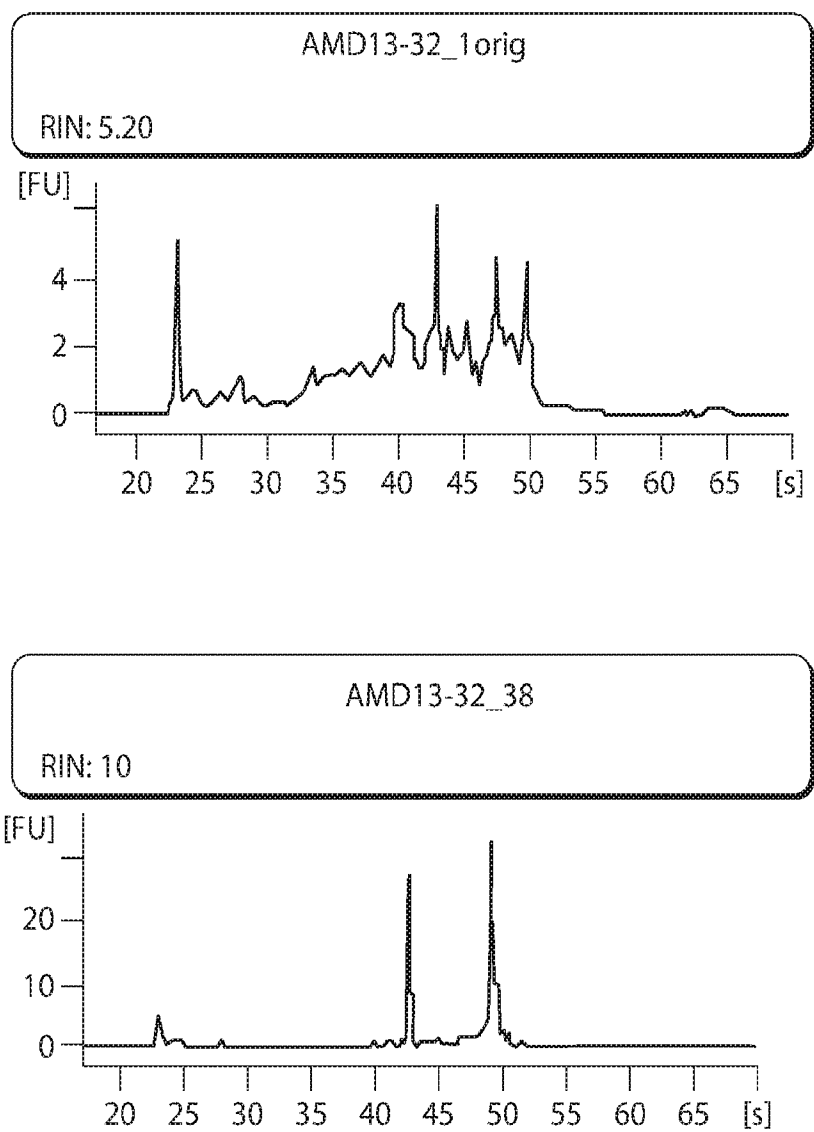


Figure 4

11/38

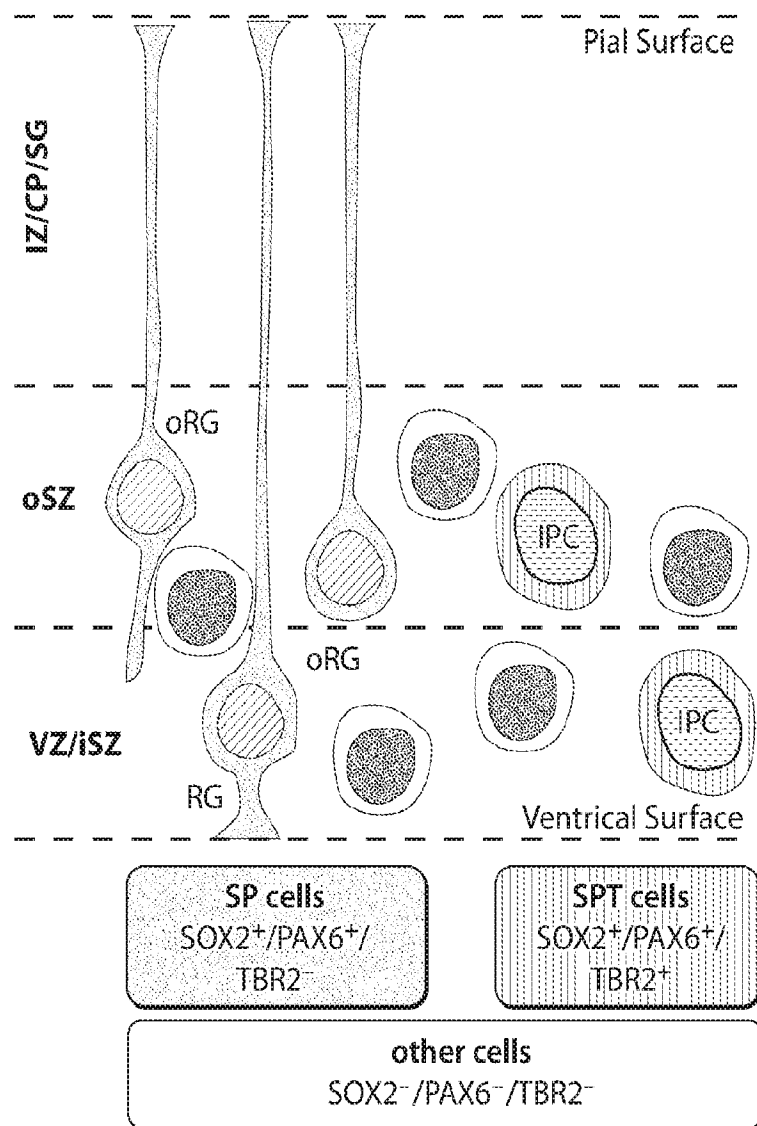


Figure 5A

12/38

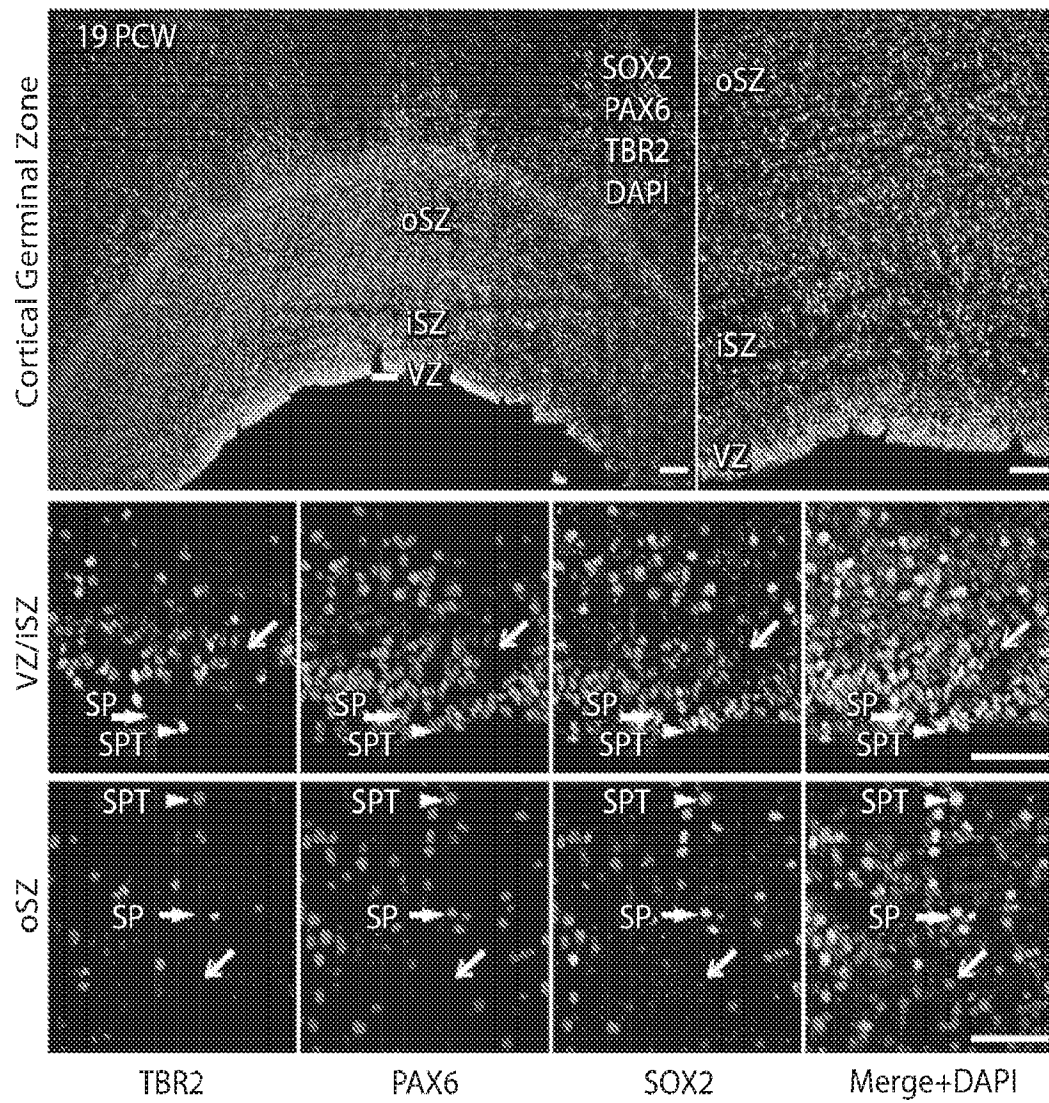


Figure 5B

13/38

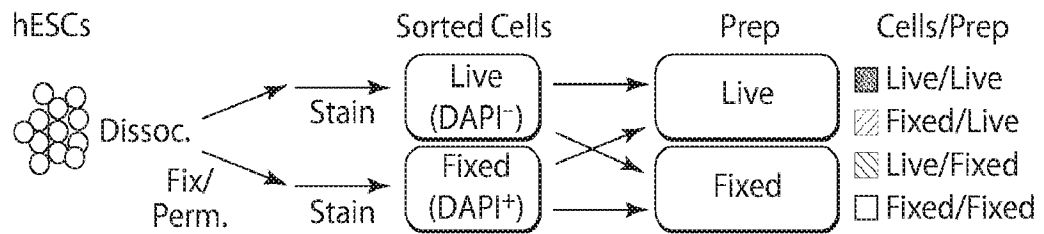


Figure 6

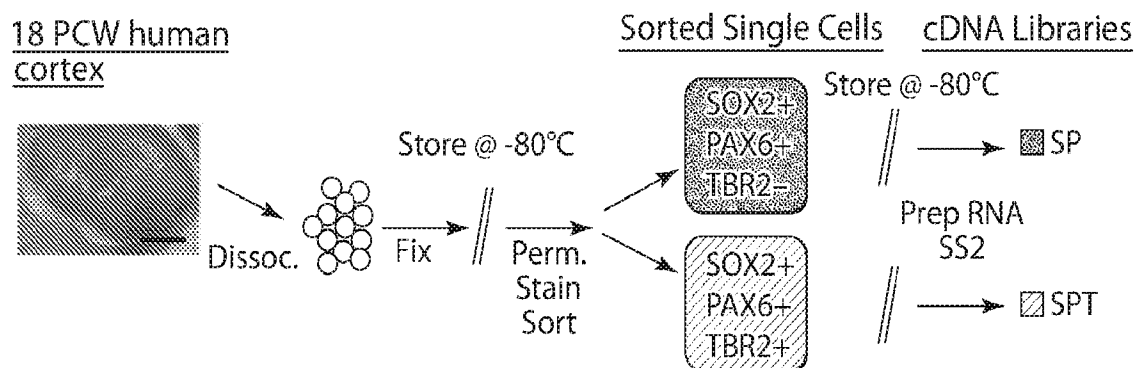


Figure 7A

14/38

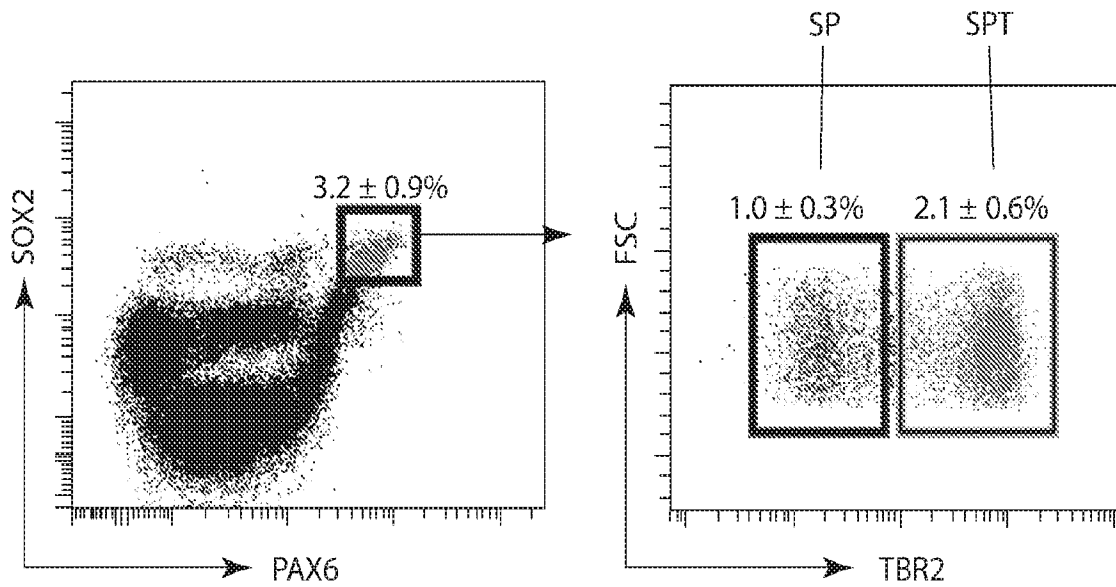


Figure 7B

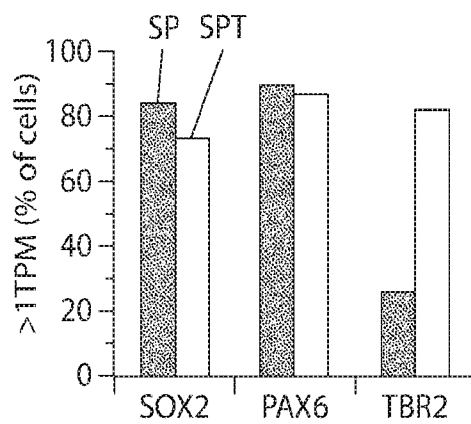


Figure 7C

15/38

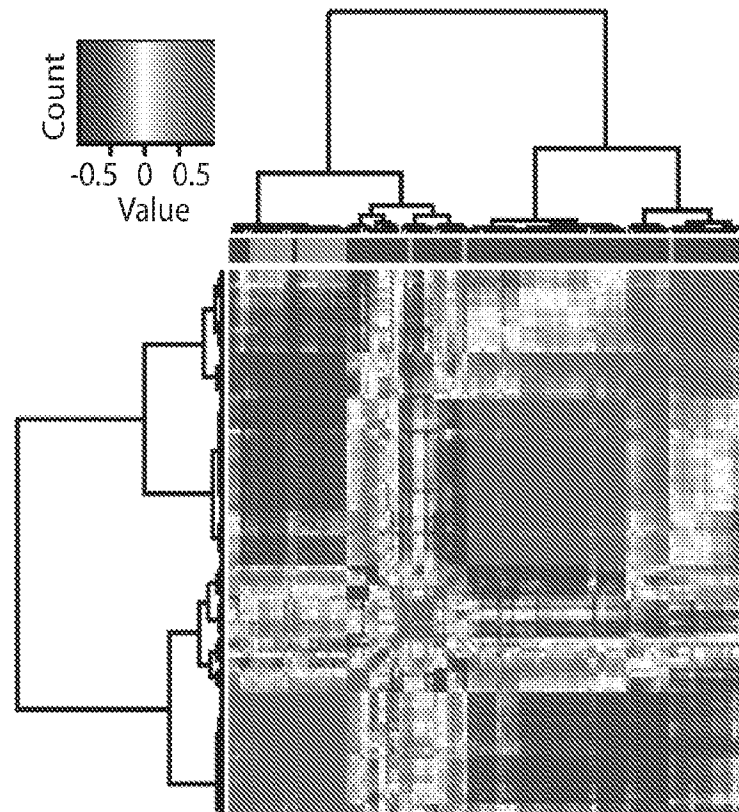


Figure 7D

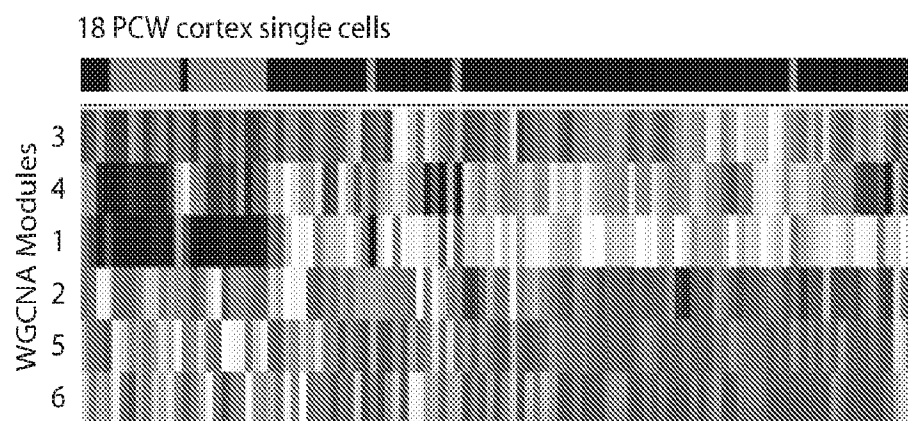


Figure 7E

16/38

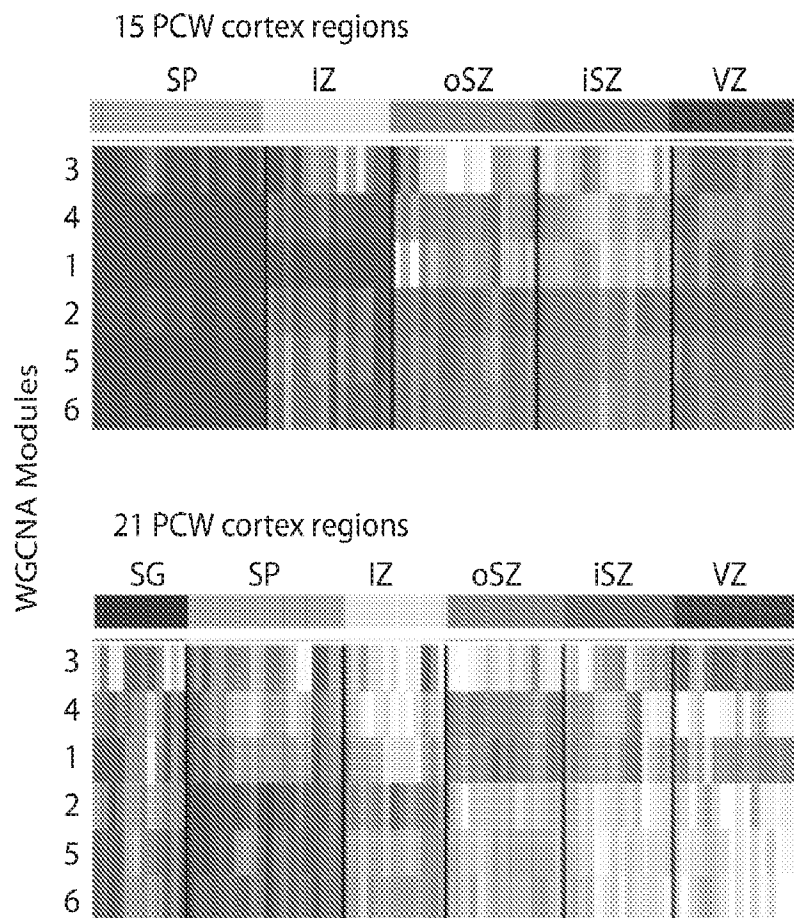


Figure 7F

17/38

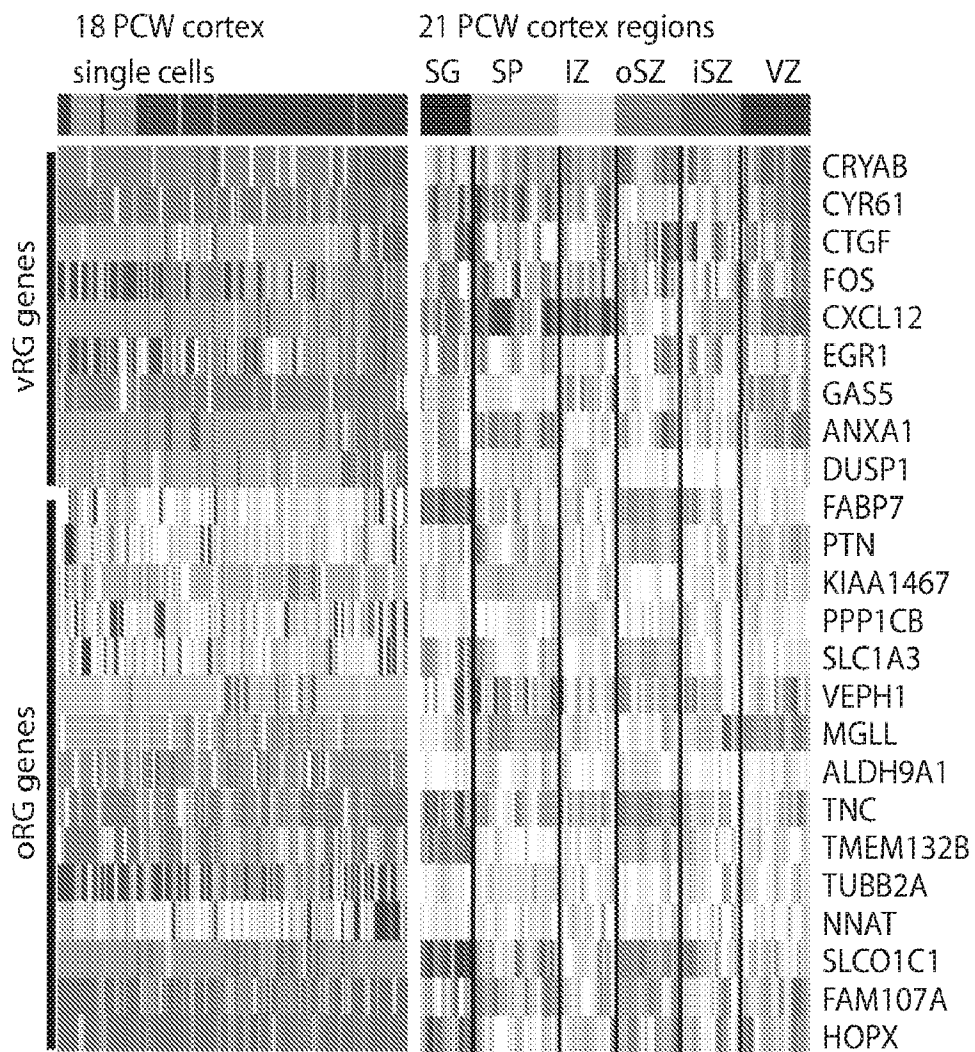


Figure 7G

18/38

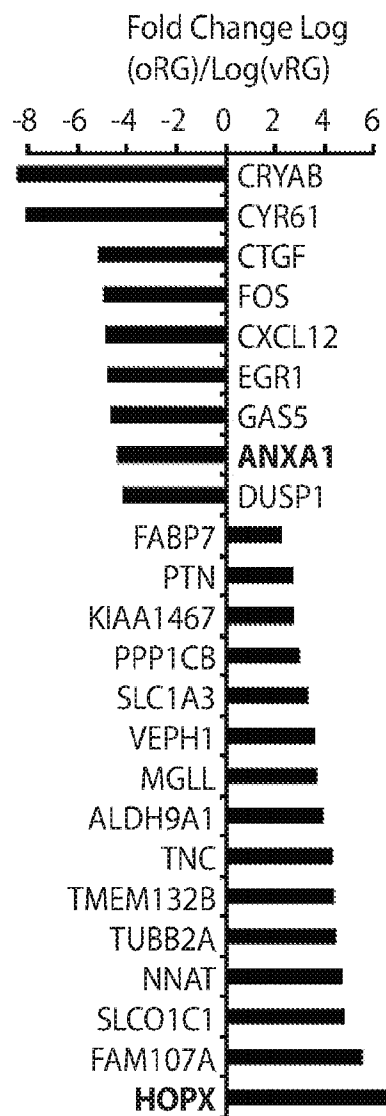


Figure 8A

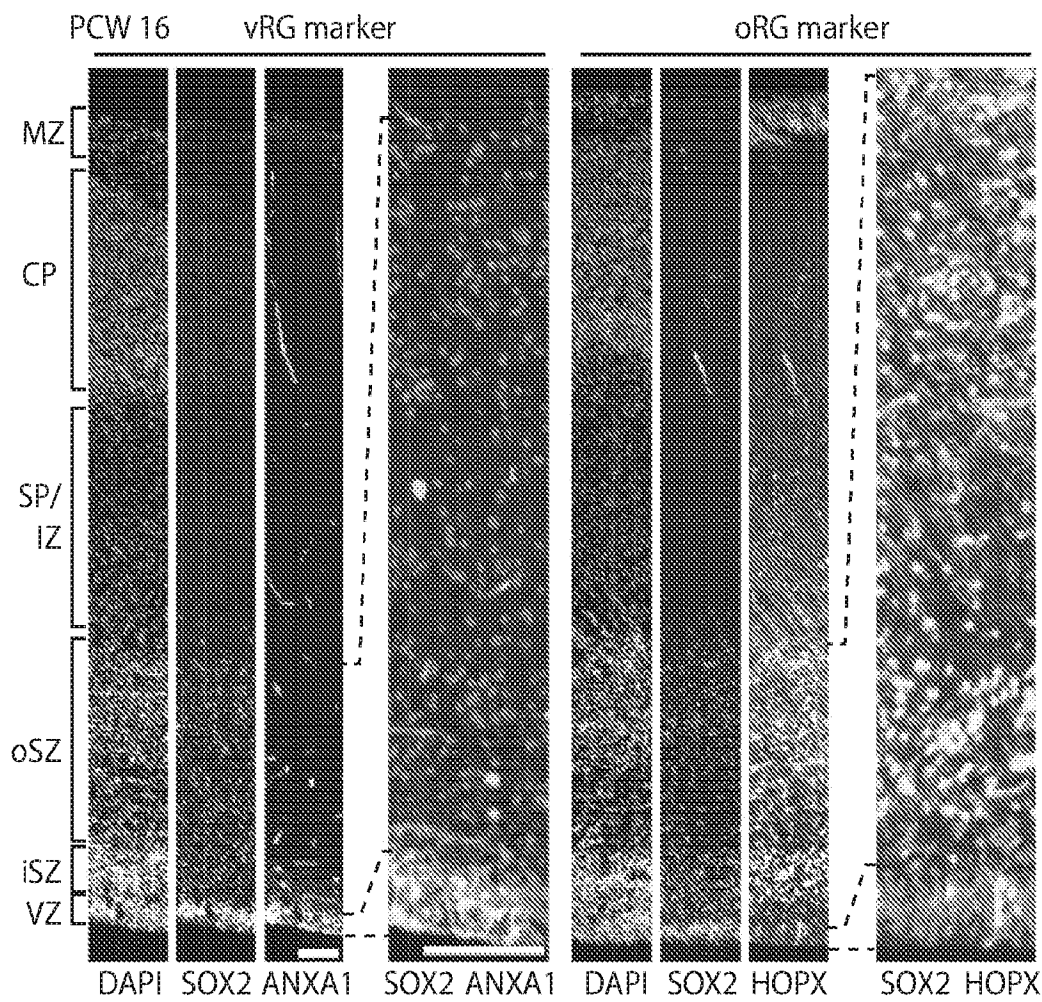


Figure 8B

20/38

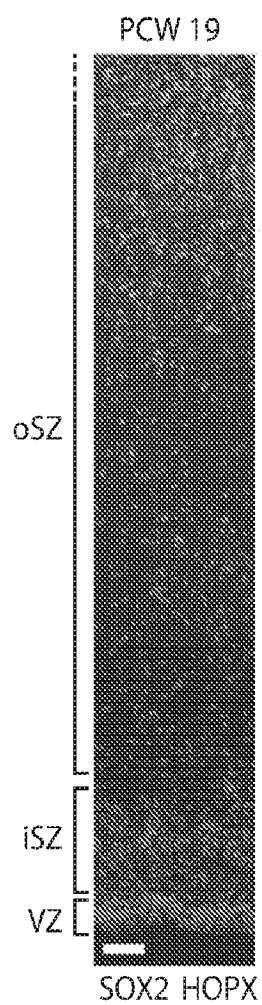


Figure 8C

21/38

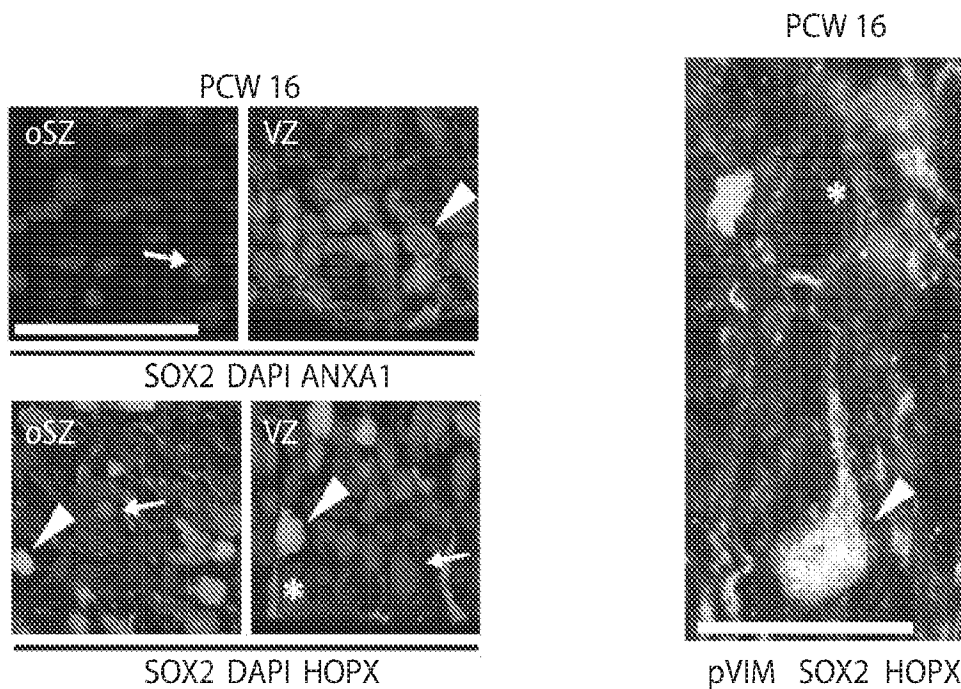


Figure 8D

Figure 8E

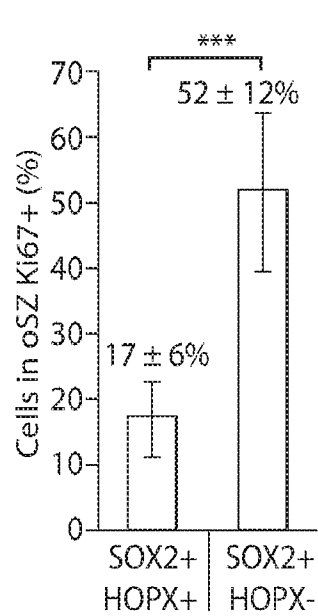


Figure 8F

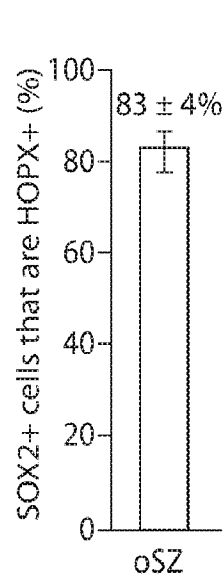


Figure 8G

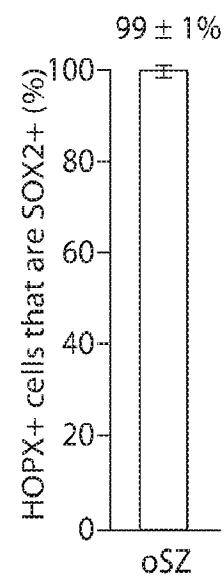


Figure 8H

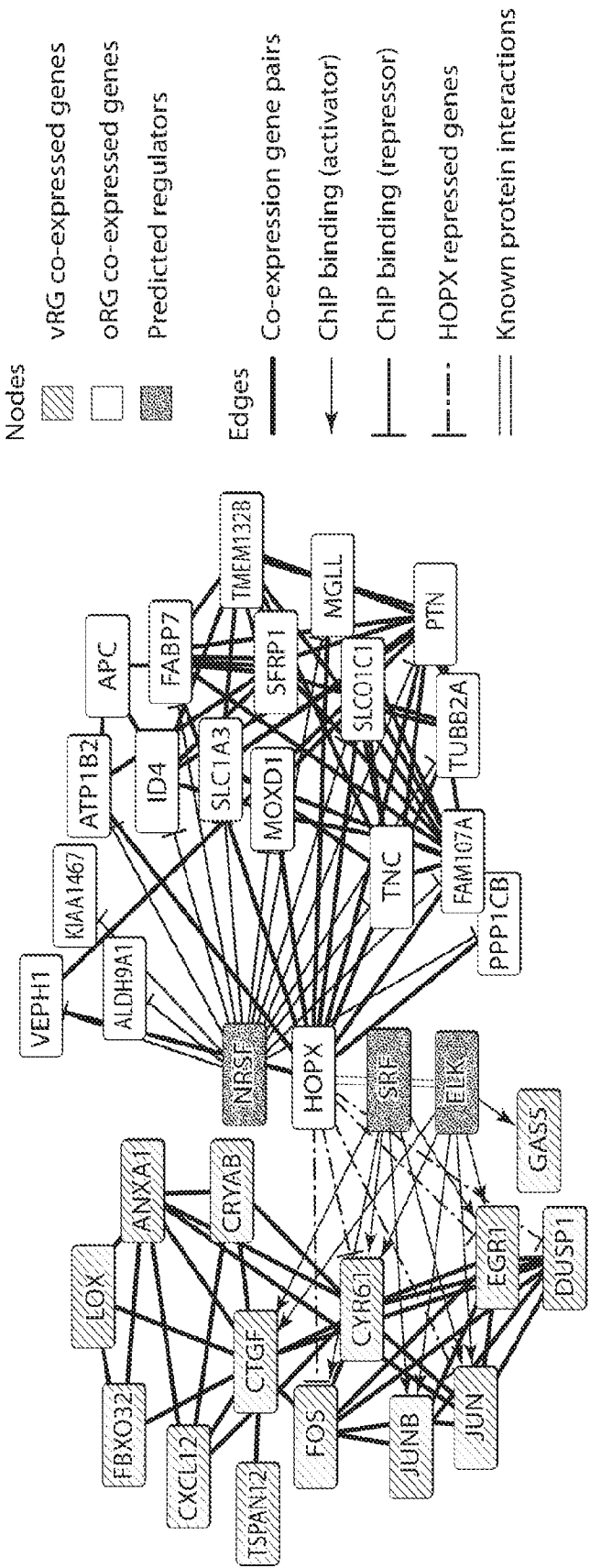


Figure 8I

23/38

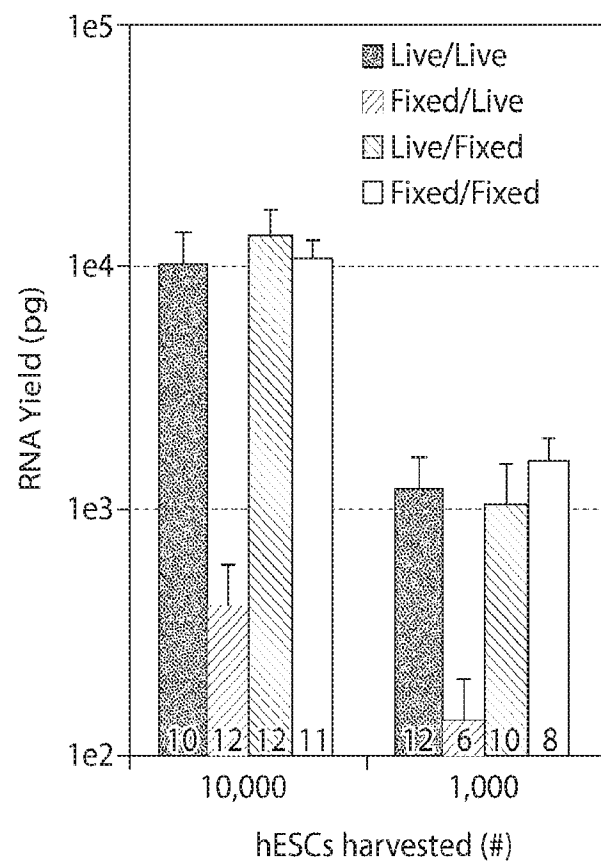


Figure 9

24/38

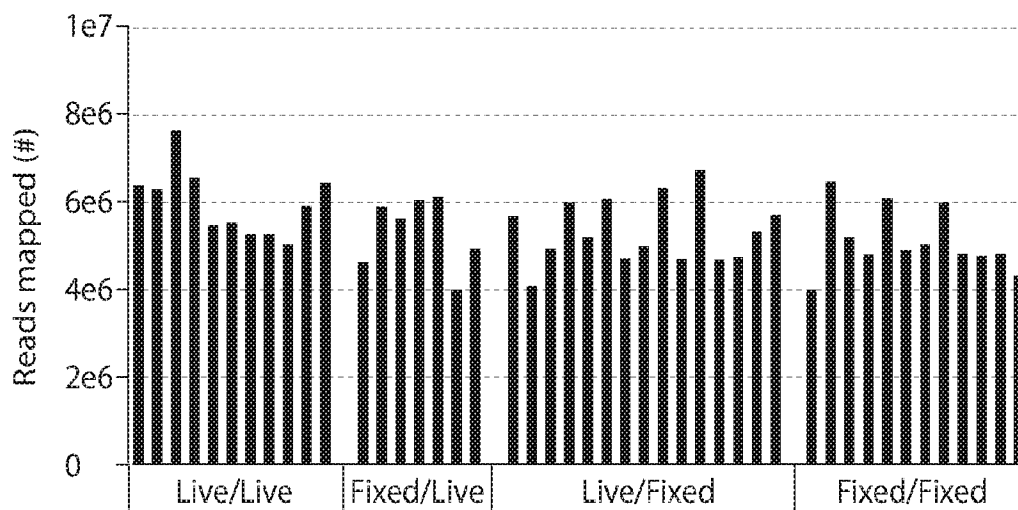


Figure 10A

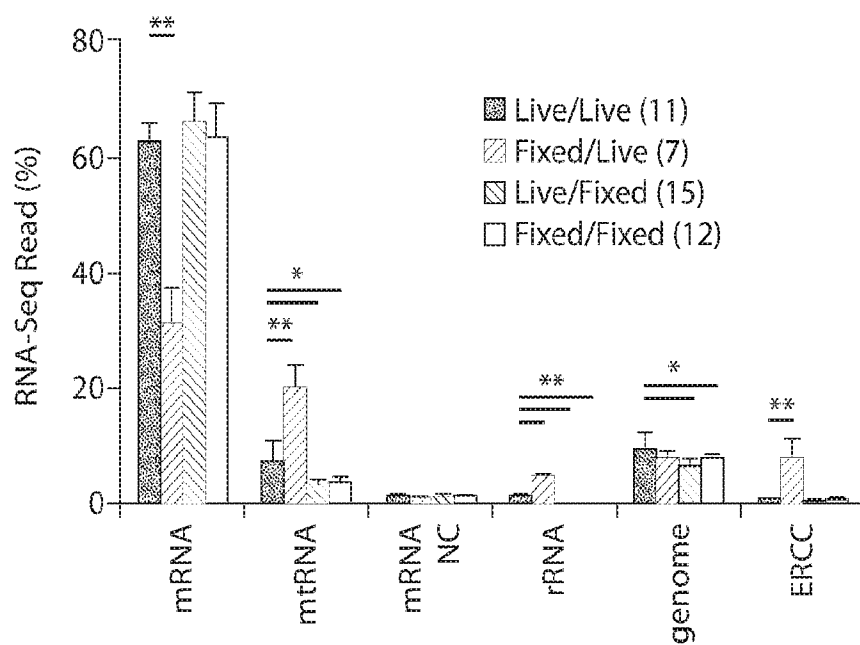


Figure 10B

25/38

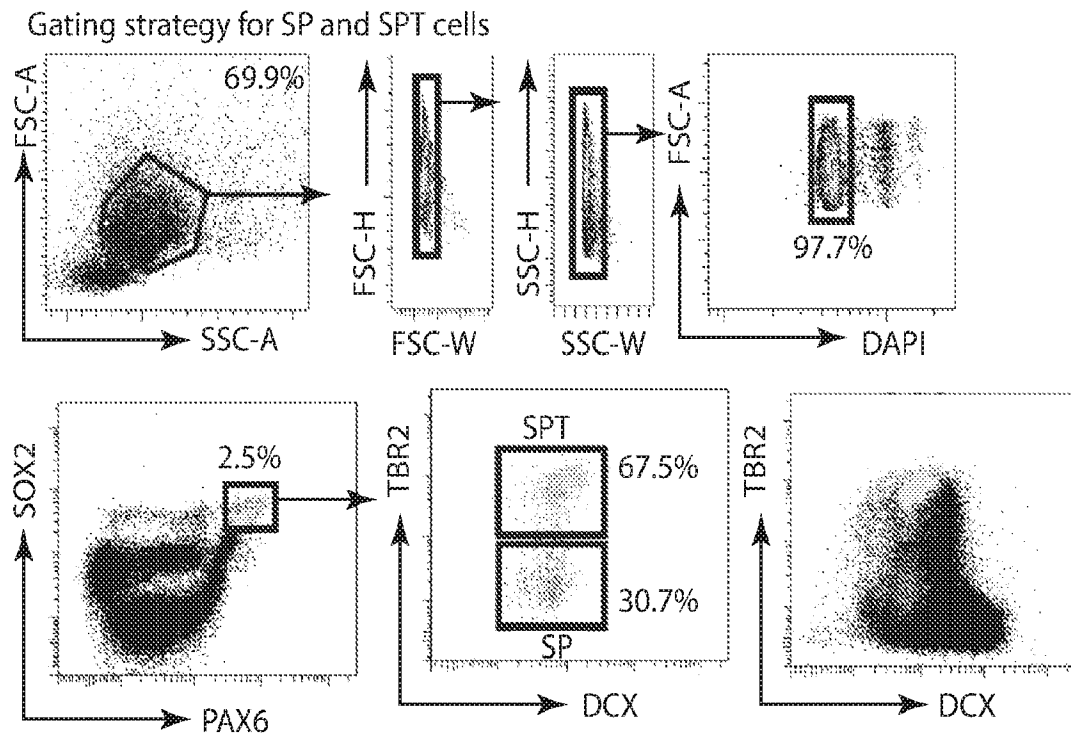


Figure 11A

26/38

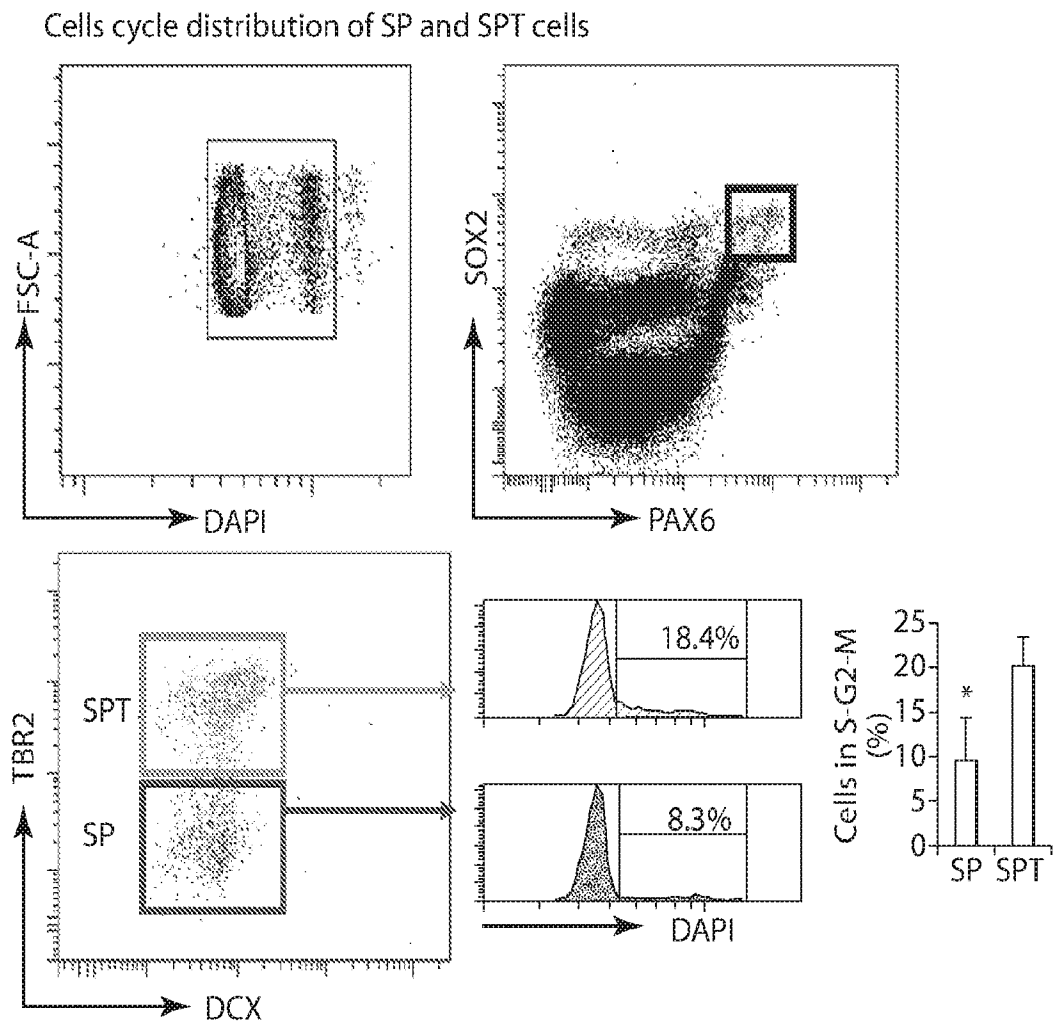


Figure 11B

27/38

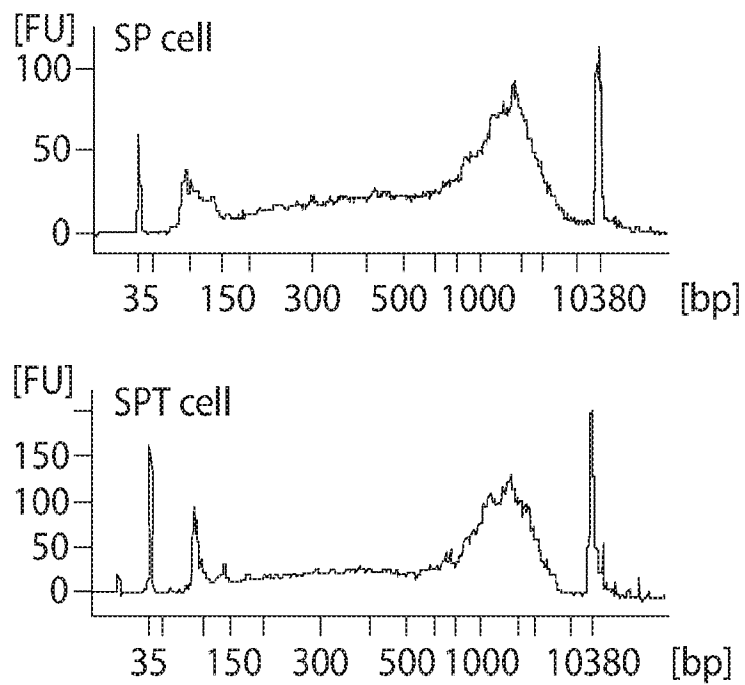


Figure 12A

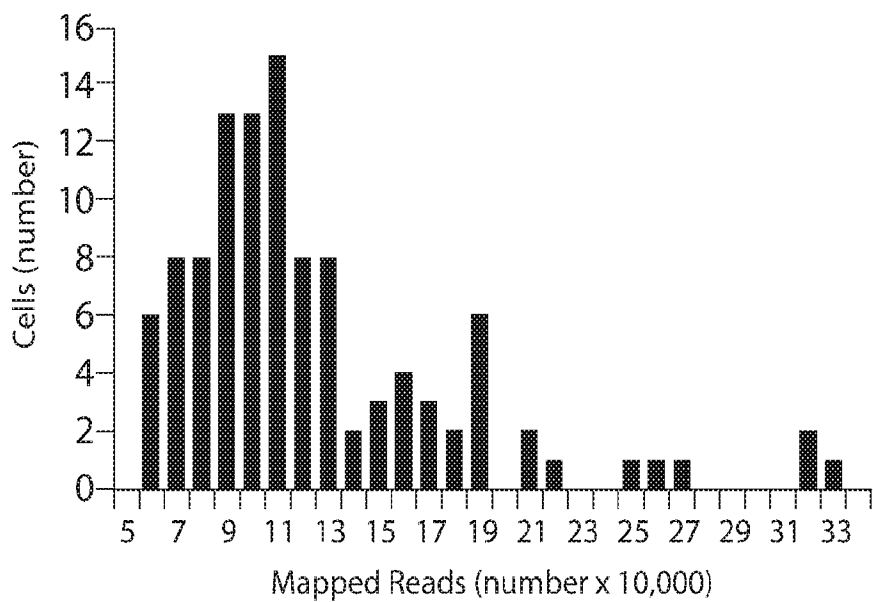


Figure 12B

28/38

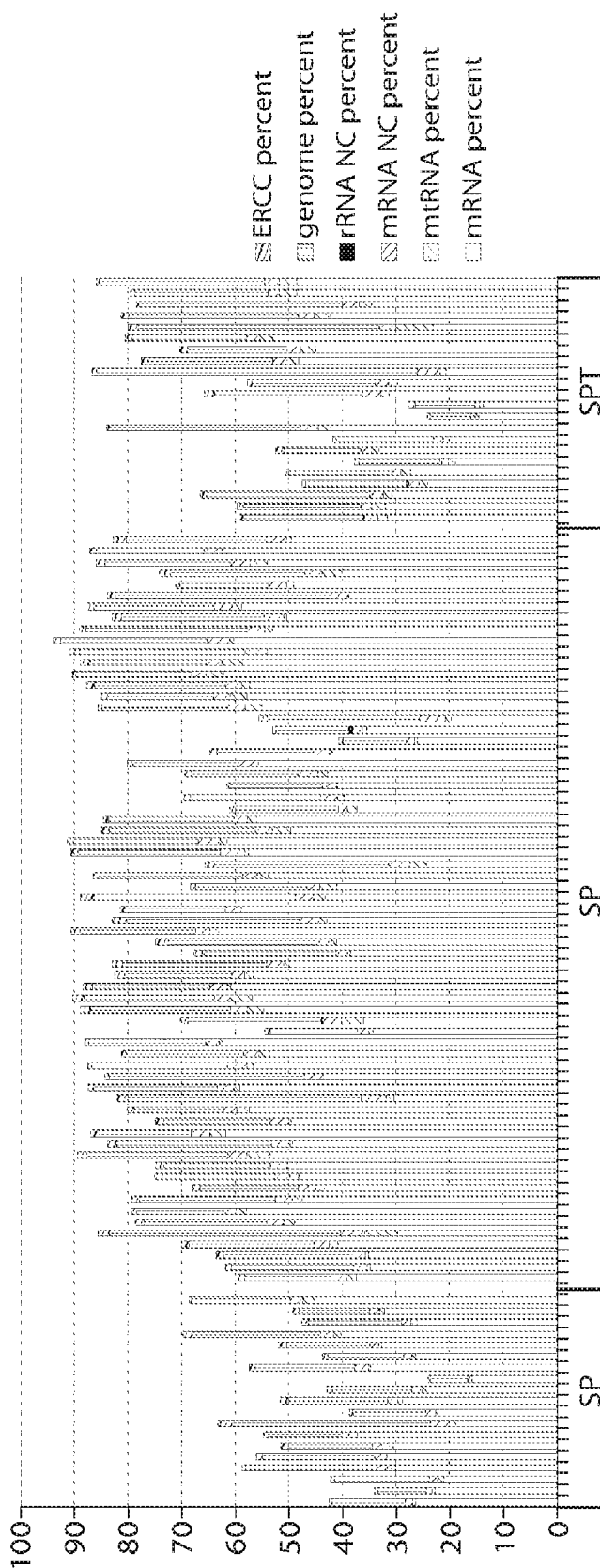


Figure 12C

29/38

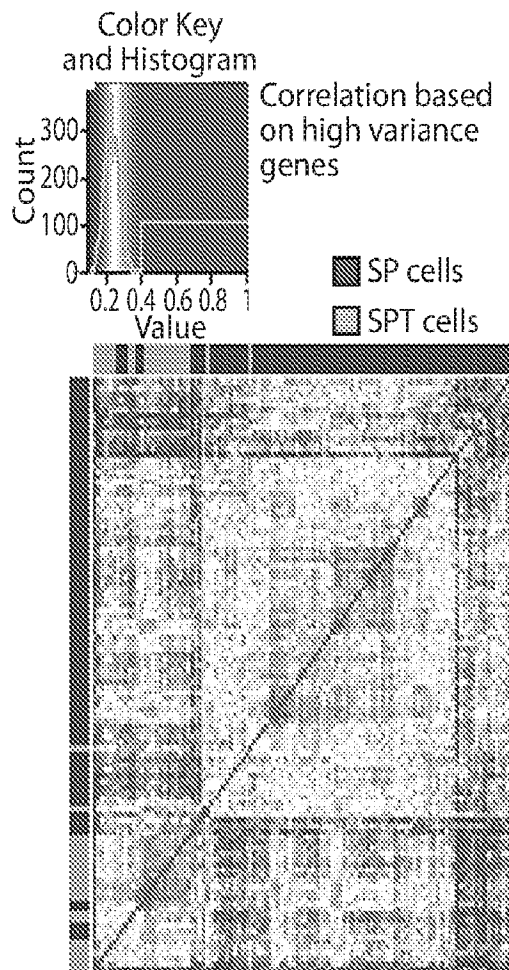


Figure 12D

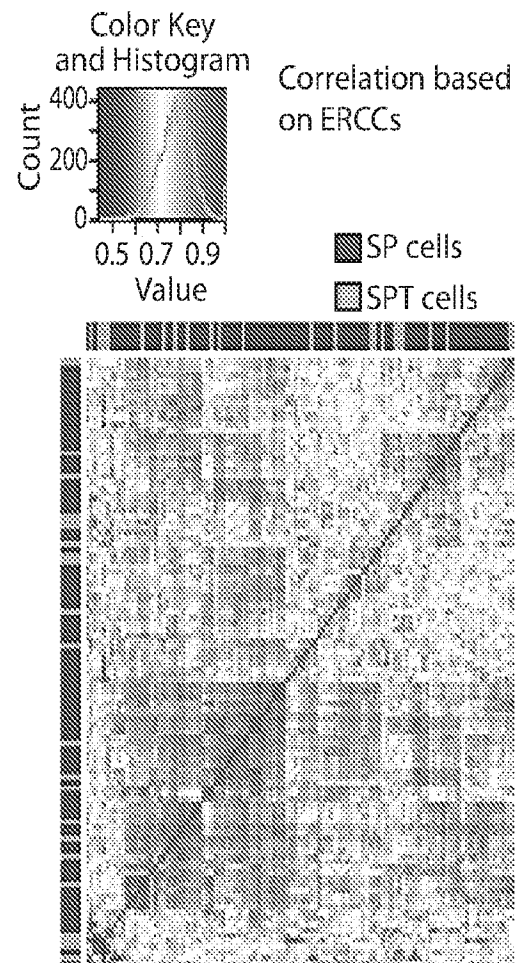


Figure 12E

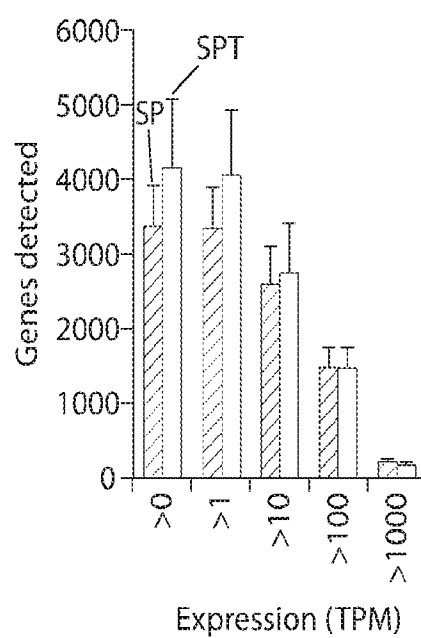


Figure 12F

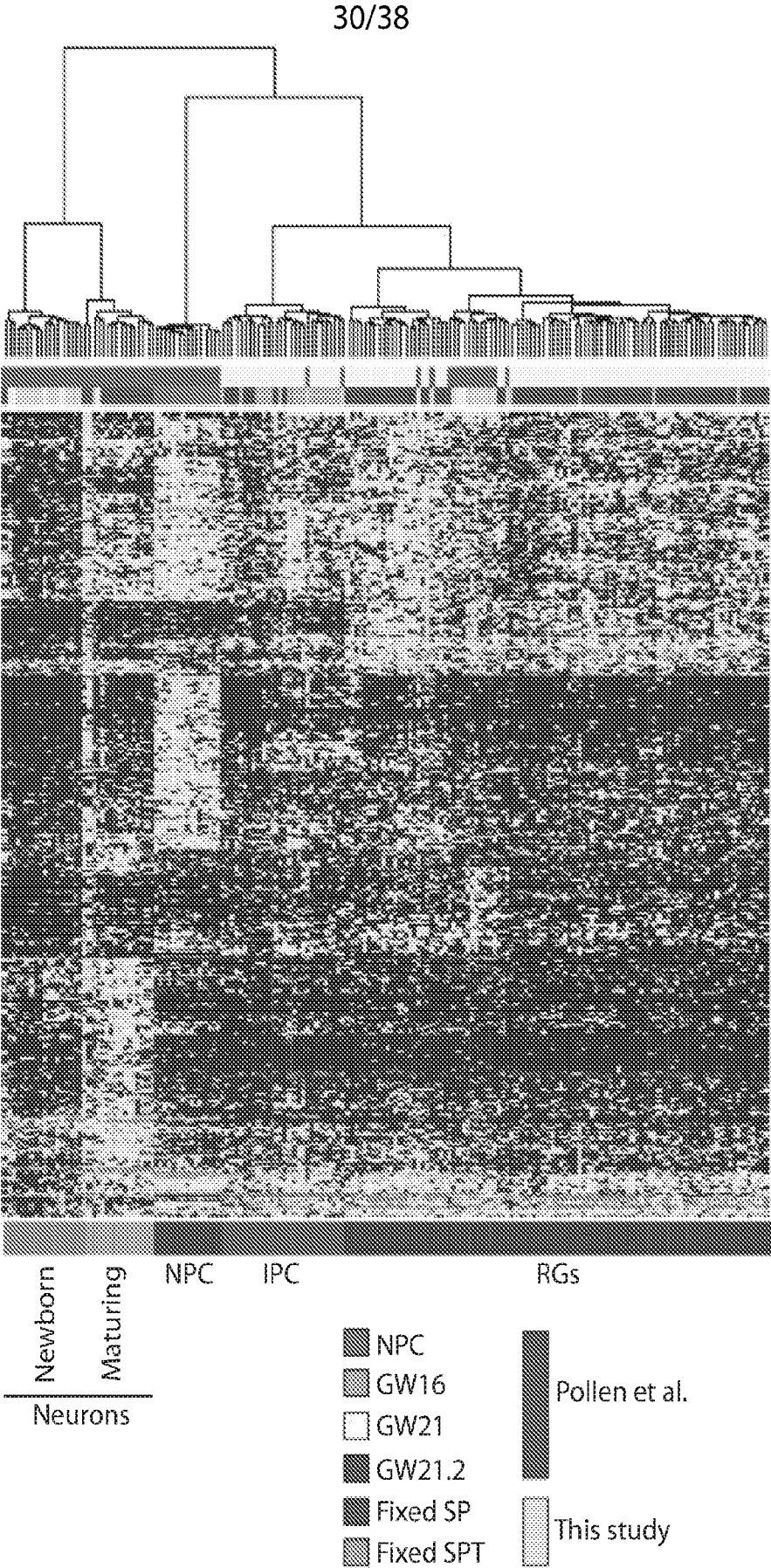


Figure 13

31/38

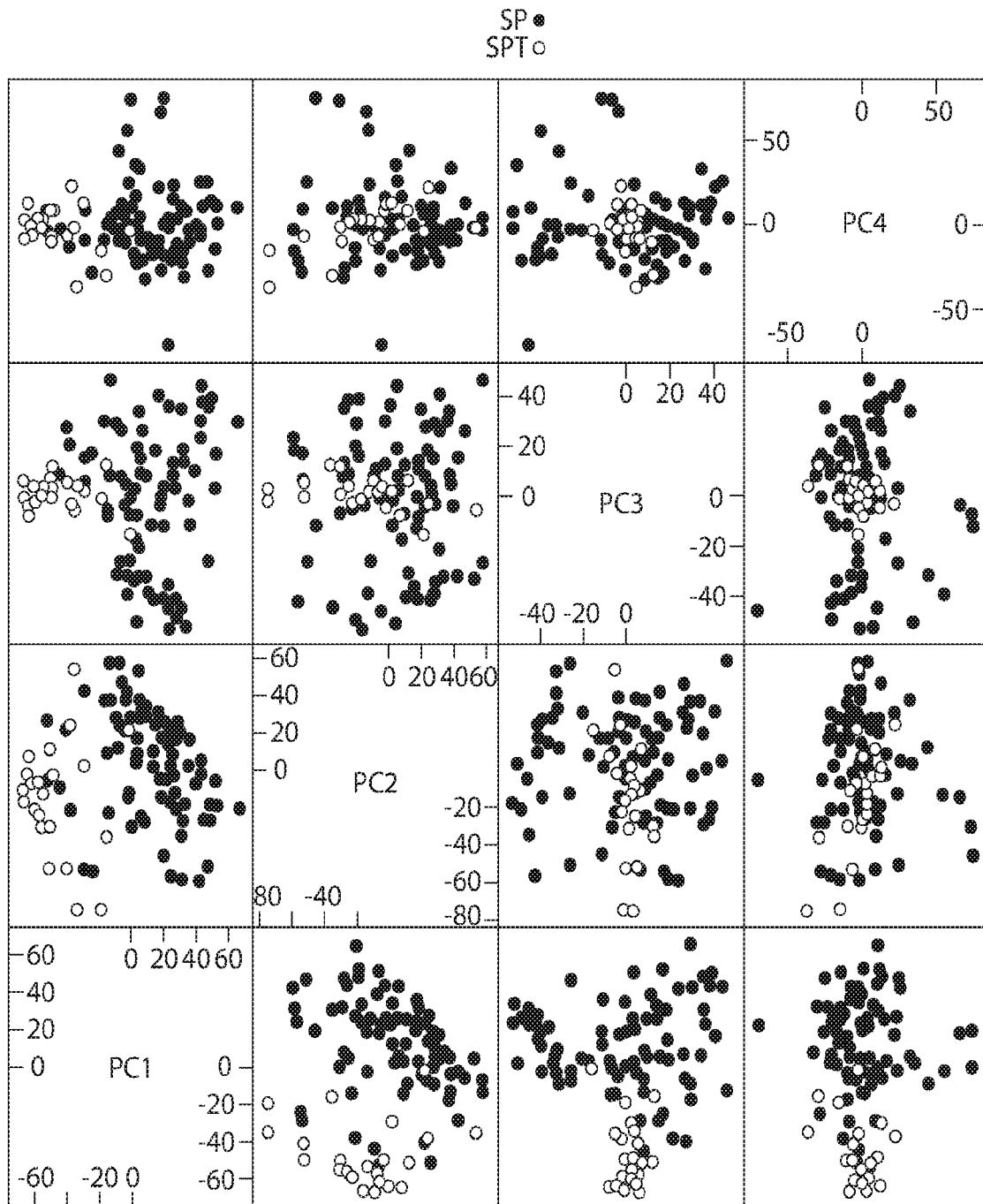


Figure 14

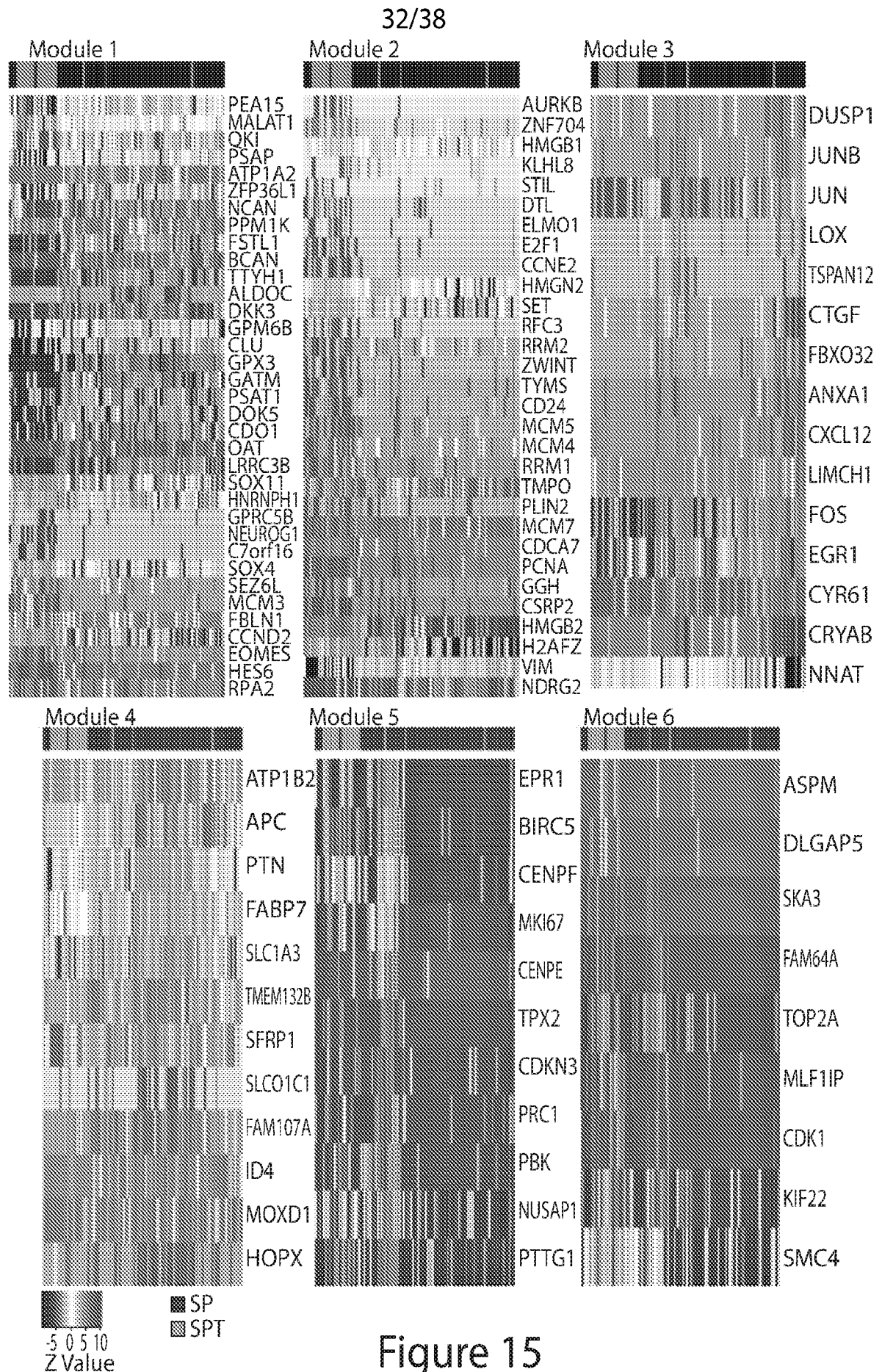


Figure 15

SUBSTITUTE SHEET (RULE 26)

33/38

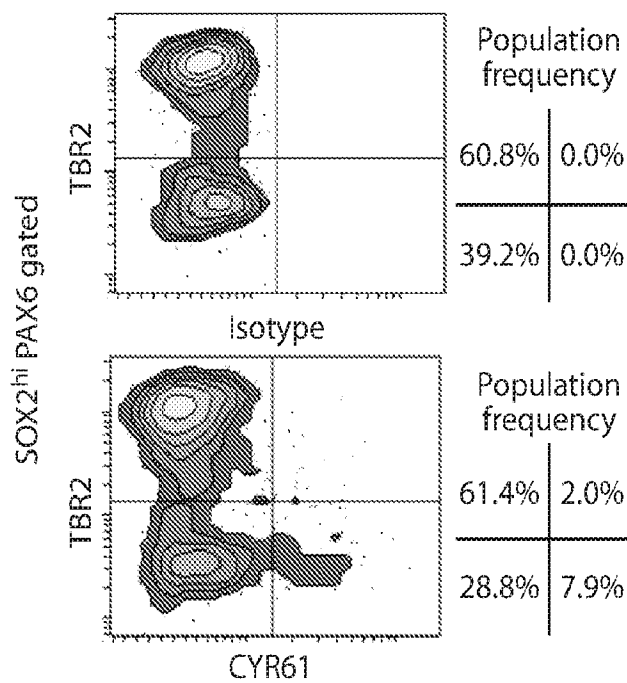


Figure 16

34/37

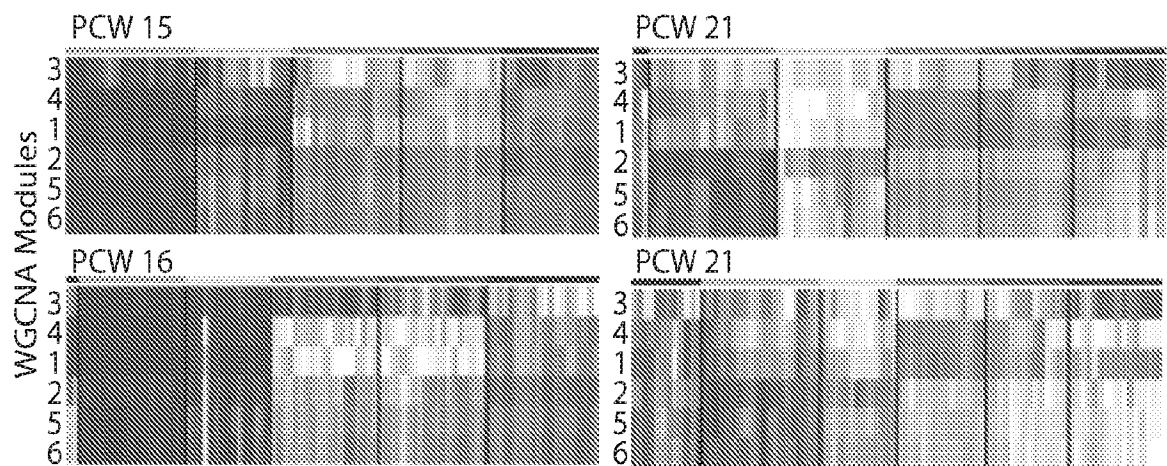


Figure 17A

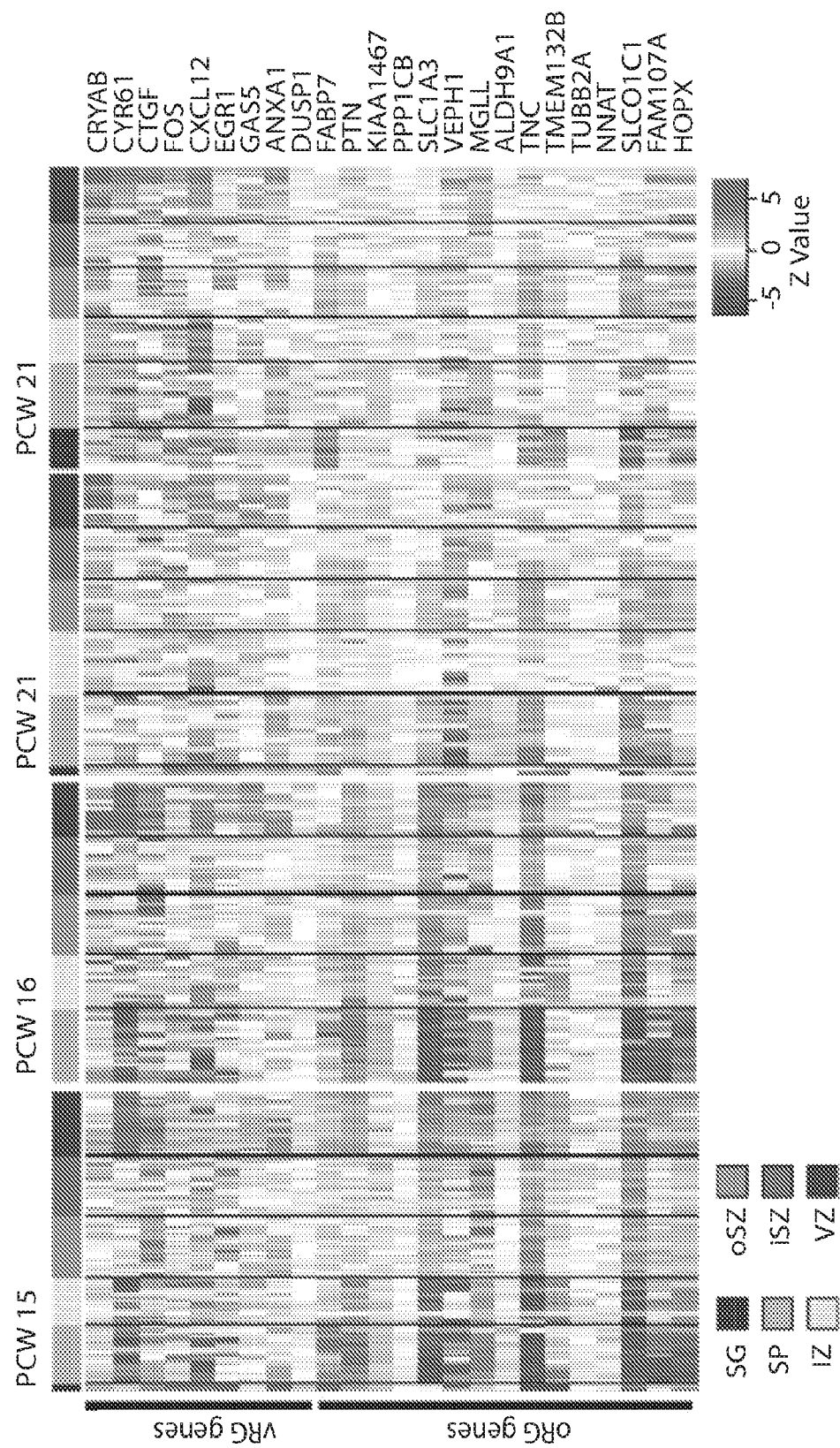


Figure 17B

36/38

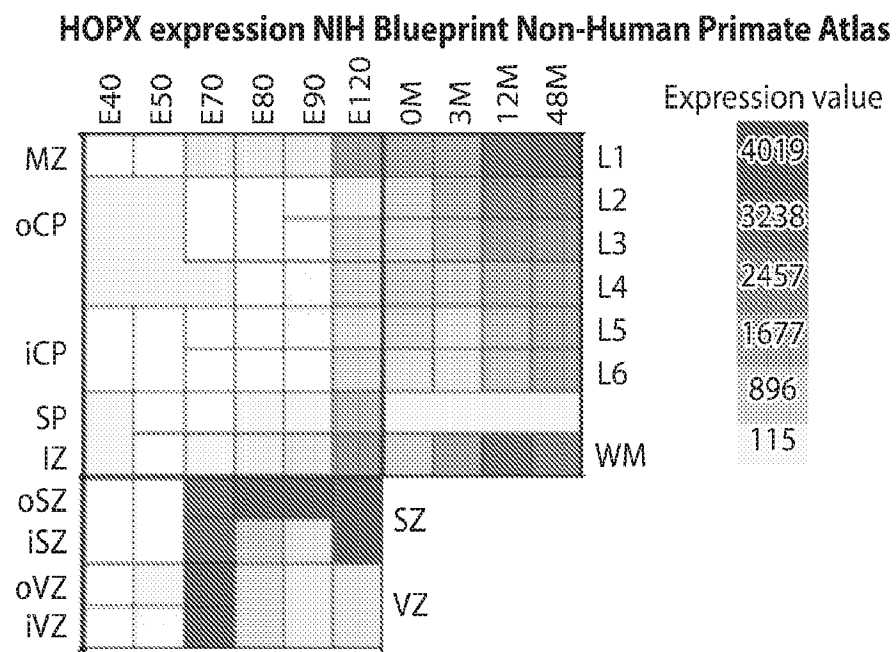


Figure 17C

37/38

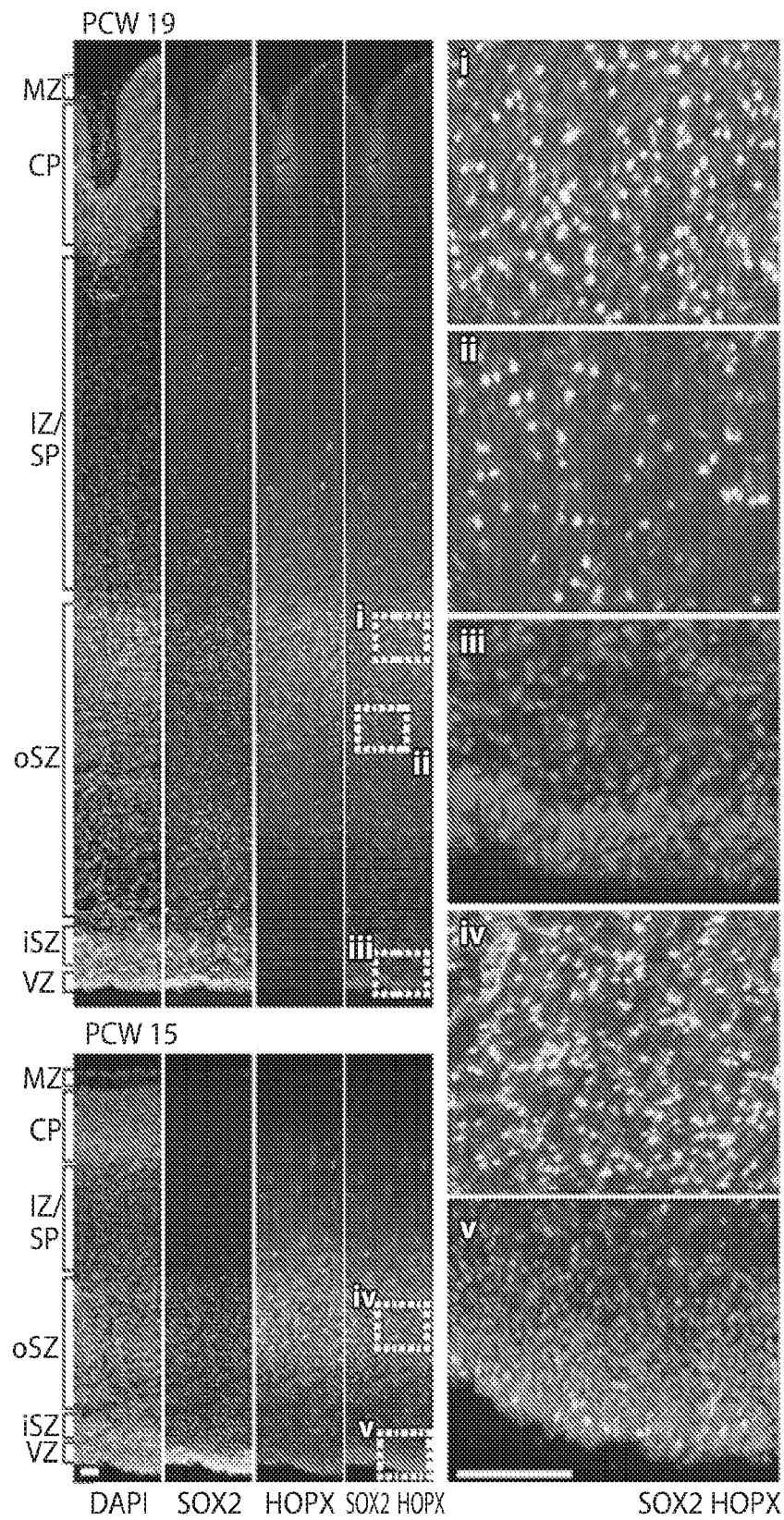


Figure 18

SUBSTITUTE SHEET (RULE 26)

38/38

oSZ 15.5 PCW

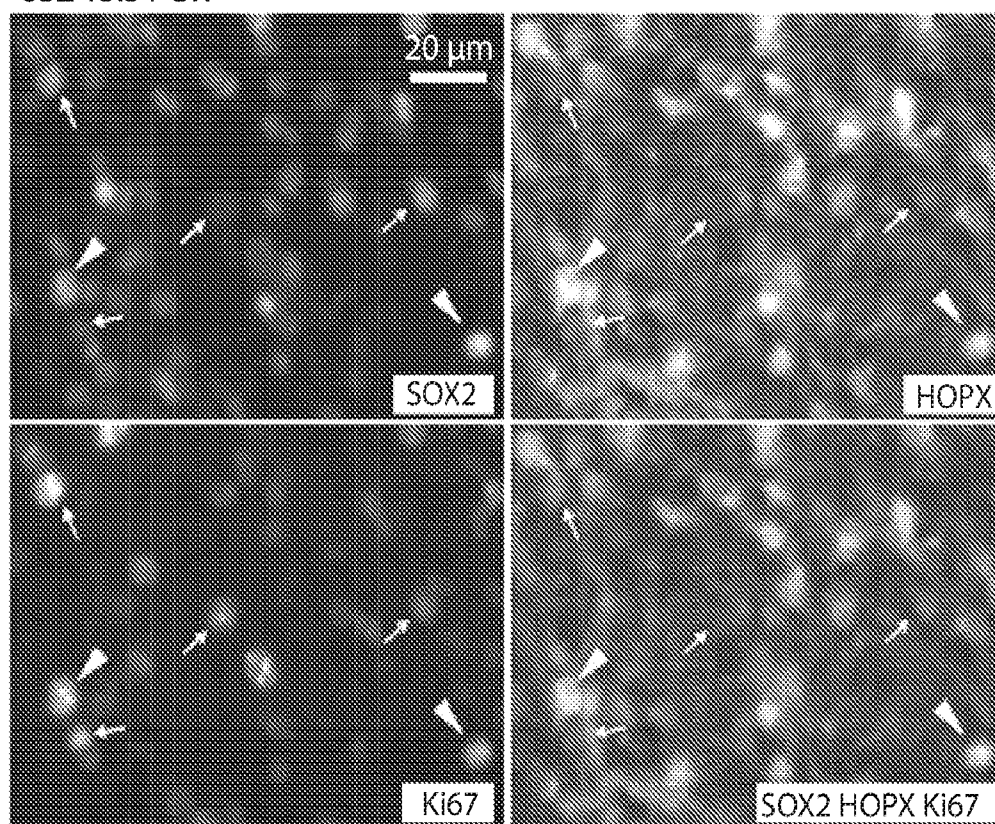


Figure 19

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/25077

A. CLASSIFICATION OF SUBJECT MATTER

IPC(8) - C12N 15/10; C12Q 1/68; G01N 1/28, 1/30 (2016.01)

CPC - C12N 15/10, 15/1003; C12Q 1/6806, 1/68; G01N 1/28, 1/30

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)

IPC(8) - C12N 15/10; C12Q 1/68; G01N 1/28, 1/30 (2016.01)

CPC - C12N 15/10, 15/1003; C12Q 1/6806, 1/68; G01N 1/28, 1/30

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

PatSeer (US, EP, WO, JP, DE, GB, CN, FR, KR, ES, AU, IN, CA); Google Scholar; EBSCO; PubMed; Keywords: purify, isolate, extract, RNA, ribonucleic acid, nucleic acid, fixed biological sample, formalin-fixed, FFPE, lysis buffer, temperature, degrees, Celsius, substrate, membrane, bead, column

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2013/0280787 A1 (MUELLER, M et al.) 24 October 2013; paragraphs [0012], [0015]-[0021], [0029]-[0031], [0036]-[0044]	1-3, 4/1-3, 5/4/1-3, 6/5/4/1-3, 7/5/4/1-3, 12, 30-31, 32/30-31, 41-43, 44/41-43, 45/44/41-43, 46/45/44/41-43, 47/45/44/41-43, 51
A	US 2007/0026432 A1 (KE, SH et al.) 01 February 2007; entire document	1-3, 4/1-3, 5/4/1-3, 6/5/4/1-3, 7/5/4/1-3, 12, 30-31, 32/30-31, 41-43, 44/41-43, 45/44/41-43, 46/45/44/41-43, 47/45/44/41-43, 51
A	WO 2014/052551 A1 (CEPHEID) 03 April 2014; entire document	1-3, 4/1-3, 5/4/1-3, 6/5/4/1-3, 7/5/4/1-3, 12, 30-31, 32/30-31, 41-43, 44/41-43, 45/44/41-43, 46/45/44/41-43, 47/45/44/41-43, 51

☐ 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 on priority 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 or priority 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

"&" document member of the same patent family

Date of the actual completion of the international search

27 May 2016 (27.05.2016)

Date of mailing of the international search report

27 JUN 2016

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

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US16/25077

Box No. II 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.: 8-11, 13-29, 33-40, 48-50 and 52-65
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III 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:

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.:

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.