

**METHODS AND COMPOSITIONS FOR PRODUCING GRANULOSA-LIKE CELLS****CROSS-REFERENCE TO RELATED APPLICATIONS**

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 63/326,640, filed April 1, 2022, and U.S. Provisional Application No. 63/444,108, filed February 8, 2023, each of which is hereby incorporated by reference in its entirety.

**REFERENCE TO AN ELECTRONIC SEQUENCE LISTING**

The contents of the electronic sequence listing (H049870758WO00-SEQ-KVC.xml; Size: 7,426 bytes; and Date of Creation: March 30, 2023) is herein incorporated by reference in its entirety.

**BACKGROUND**

Granulosa cells are specialized cells that, upon maturation, provide critical steroids and growth hormones to the developing oocyte. Granulosa cells surround developing oocytes and participate in maintaining a potential pregnancy by modulating hormone levels. Granulosa cell dysfunction forms the basis of many forms of human female infertility, yet efficient methods for generating granulosa *in vitro* remain elusive.

**SUMMARY**

The present disclosure relates, at least in part, to methods and compositions for generating granulosa *in vitro* from pluripotent stem cells (PSCs). The present disclosure provides experimental data demonstrating, unexpectedly, that overexpression of certain transcription factors, for example, Nuclear Receptor Subfamily 5 Group A Member 1 (NR5A1) and a Runt-Related Transcription Factor (RUNX) family member (e.g., RUNX1 and/or RUNX2), is sufficient to generate granulosa (e.g., AMHR2+, CD82+, FOXL2+, and/or EPCAM– granulosa-like cells) from iPSCs in as few as 5 to 7 days.

Some aspects of the present disclosure provide a PSC comprising: an engineered polynucleotide comprising an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein.

In some embodiments, the PSC comprises the engineered polynucleotide comprising an open reading frame encoding NR5A1.

In some embodiments, the PSC comprises the engineered polynucleotide comprising an open reading frame encoding a RUNX family protein.

In some embodiments, the RUNX family protein is RUNX1.

In some embodiments, the RUNX family protein is RUNX2.

In some embodiments, the PSC expresses or overexpresses: NR5A1; RUNX1; RUNX2; NR5A1 and RUNX1; NR5A1 and RUNX2; or NR5A1, RUNX1, and RUNX2.

In some embodiments, the PSC further comprises an engineered polynucleotide comprising an open reading frame encoding a Transcription factor 21 (TCF21) protein.

In some embodiments, the PSC expresses or overexpresses TCF21.

In some embodiments, the PSC comprises the engineered polynucleotide comprising an open reading frame encoding GATA Binding Protein 4 (GATA4).

In some embodiments, the PSC expresses or overexpresses GATA4.

In some embodiments, the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

In some embodiments, the heterologous promoter is an inducible promoter.

Other aspects of the present disclosure provide a PSC comprising: a protein selected from NR5A1 and a RUNX family protein, wherein the protein is overexpressed.

In some embodiments, the PSC expresses or overexpresses: NR5A1; RUNX1; RUNX2; NR5A1 and RUNX1; NR5A1 and RUNX2; or NR5A1, RUNX1, and RUNX2.

In some embodiments, the PSC further comprises a TCF21 protein.

In some embodiments, the PSC expresses or overexpresses TCF21.

In some embodiments, the PSC further comprises a GATA4 protein.

In some embodiments, the PSC expresses or overexpresses GATA4.

In some embodiments, the PSC is a human PSC.

In some embodiments, the PSC is an induced PSC (iPSC).

In some embodiments, the PSC comprises 1-20, optionally 8-10, copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2).

Further aspects of the present disclosure provide a composition comprising: a population of the PSC of any one of the preceding paragraphs or described elsewhere herein.

In some embodiments, the population comprises at least 10,000/cm<sup>2</sup> of the PSC.

Some aspects of the present disclosure provide a method, comprising: culturing, in culture media, a population of PSCs to produce an expanded population of PSCs; and

expressing in PSCs of the expanded population a protein selected from NR5A1 and a RUNX family protein to produce granulosa-like cells.

In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NR5A1.

In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding a RUNX family protein.

In some embodiments, the RUNX family protein is RUNX1.

In some embodiments, the RUNX family protein is RUNX2.

In some embodiments, the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding a TCF21 protein.

In some embodiments, the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding a GATA4 protein.

In some embodiments, the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

In some embodiments, the heterologous promoter is an inducible promoter.

In some embodiments, the population comprises  $1 \times 10^2$  -  $1 \times 10^7$  PSCs.

In some embodiments, the population of PSCs is cultured for about 4-10 days. For example, the population of PSCs may be cultured for about 6 days.

In some embodiments, the granulosa-like cells are AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and/or EPCAM<sup>-</sup>.

Other aspects of the present disclosure provide a method comprising: (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein;

(b) culturing the PSCs in feeder-free, serum-free culture media to produce an expanded population of PSCs; and (c) culturing PSCs of the expanded population in a series of induction media comprising an inducing agent to produce AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and/or EPCAM<sup>-</sup> granulosa-like cells.

In some embodiments, the method comprises delivering to PSCs (i) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding NR5A1 and (i) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a RUNX family protein.

In some embodiments, the RUNX family protein is RUNX1.

In some embodiments, the RUNX family protein is RUNX2.

In some embodiments, the engineered polynucleotide is a transposon and the delivering further comprises delivering a transposase to the PSCs.

In some embodiments, the inducible promoter is a chemically-inducible promoter, optionally a doxycycline-inducible promoter.

In some embodiments, the feeder-free, serum-free culture media of (b) comprises a solubilized basement membrane preparation extracted from the Engelbreth-Holm-Swarm (EHS) mouse sarcoma.

In some embodiments, the solubilized basement membrane preparation comprises extracellular matrix (ECM) proteins and growth factors.

In some embodiments, the ECM proteins are selected from Laminin, Collagen IV, heparan sulfate proteoglycans, and entactin/nidogen.

In some embodiments, the feeder-free, serum-free culture media of (b) comprises growth factors selected from recombinant human basic fibroblast growth factor (rh bFGF) and recombinant human transforming growth factor  $\beta$  (rh TGF $\beta$ ).

In some embodiments, the culturing of (b) is for about 6-24 hours.

In some embodiments, the PSCs of the expanded population of (c) are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 20,000 cells/cm<sup>2</sup>.

In some embodiments, the culturing of (c) comprises culturing the PSCs in a first induction media and culturing the PSCs in a second induction media.

In some embodiments, the first induction media comprises one or more of L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Dulbecco's Modified Eagle Medium (DMEM)/F-12, Advanced RPMI (Roswell Park Memorial Institute) 1640 Medium, a glycogen synthase kinase (GSK) 3 inhibitor, a small molecule or protein inhibitor of the BMP signaling pathway, a small molecule ROCK inhibitor, and an inducing agent (e.g., doxycycline).

In some embodiments, the culturing the PSCs in a first induction media is for about 36 to about 60 hours, optionally about 48 hours.

In some embodiments, the second induction media comprises one or more of L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Advanced RPMI 1640 Medium, DMEM/F-12, and an inducing agent (e.g., doxycycline).

In some embodiments, the culturing the PSCs in a second induction media is for about 96 to about 144 hours, optionally about 120 hours.

In some embodiments, the second induction media is removed and replaced with fresh second induction media at about 24-hour intervals.

Further aspects of the present disclosure provide a granulosa-like cell produced by the method of any one of the preceding claims.

The details of one or more embodiments of the invention are set forth in the description below. Other features or advantages of the present invention will be apparent from the following drawings and detailed description of several embodiments, and also from the appended claims.

Some aspects provide an ovarian organoid comprising granulosa-like cells of any one of the preceding claims and human primordial germ cell-like cells (hPGCLCS).

In some embodiments, a method of any one of the preceding paragraphs further comprising combining the granulosa-like cells with hPGCLCS to form an ovarian organoid.

### BRIEF DESCRIPTION OF THE DRAWINGS

The accompanying drawings are not intended to be drawn to scale. In the drawings, each identical or nearly identical component that is illustrated in various figures is represented by a like numeral. For purposes of clarity, not every component may be labeled in every drawing. In the drawings:

**FIG. 1** shows barcode enrichment screening for TFs activating FOXL2 expression. TF expression plasmids were introduced into FOXL2-tdTomato reporter iPSCs. Three TF pools were evaluated: C5, containing 5 fmol total of all TFs (equimolar mix); B5, containing 5 fmol total of a subset of TFs; and B50, containing 50 fmol total of the same subset. After nucleofection and puromycin selection, cells were treated with doxycycline to induce TF expression, either in pluripotency-supporting mTeSR™ Plus medium or mesoderm-inducing medium. After 5 days, FOXL2+ cells were isolated by FACS and DNA was extracted. Barcode frequencies were compared between FOXL2+ cells and the initial population of cells before TF expression.

**FIG. 2** shows a TF combinatorial screen. Various combinations of TF expression plasmids were introduced into FOXL2-tdTomato reporter iPSCs. The iPSCs were then plated in differentiation medium (DMEM/F12, 10% KSR, 1 µg/mL doxycycline). During the first 2 days the cells were additionally treated with 3 µM CHIR99021 and 10 µM Y-27632. After 5 days of differentiation the cells were dissociated and analyzed by flow cytometry to measure the proportion of AMHR2+ FOXL2+ CD82+ EPCAM– granulosa-like cells. A second replicate was treated with androstenedione (500 ng/mL) and FSH (0.15 IU/mL) from days 4 to 6 of differentiation. Estradiol production was measured by ELISA.

**FIG. 3** shows transcriptional characterization of granulosa-like cells. FOXL2+ cells generated by TF-mediated differentiation were analyzed by RNA-seq. TPM values for known markers of gonadal/granulosa, adrenal, and pluripotent cells were compared between male fetal gonad, primordial and primary granulosa cells, sorted FOXL2+ cells, COV434 ovarian tumor cells, and iPSCs. Values are the averages of at least two biological replicates.

**FIG. 4** shows that combination of top TFs generates granulosa-like cells in high yield. Expression plasmids for top TFs (NR5A1, RUNX1, RUNX2, TCF21) were introduced into iPSCs, and single colonies were picked to generate monoclonal lines. Clone 1F has NR5A1, RUNX1, and TCF21 expression plasmids integrated (confirmed by PCR). The 5-day differentiation protocol with doxycycline-induced TF expression resulted in a near-uniform population of FOXL2+ CD82+ granulosa-like cells. In comparison, spontaneous differentiation without doxycycline resulted in only a small number of granulosa-like cells.

**FIGS. 5A-5B** show hormonal signaling by granulosa-like cells. **FIG. 5A** shows that granulosa-like cells produce estradiol in the presence of androstenedione and either FSH or forskolin (FK). Results are shown from nine monoclonal populations of granulosa-like cells ( $n = 2$  biological replicates for each of 9 clones, error bars are 95% CI), as well as the COV434 and KGN human ovarian cancer cell lines (controls), HGL5 immortalized primary human granulosa cells, and primary adult mouse granulosa cells. Asterisks mark lines where FSH production significantly (2-tailed t-test,  $p < .05$ ) increased upon stimulation. Exact p-values are given in the Source Data. Monoclonal lines of iPSCs bearing integrated TFs (NR5A1, TCF21, GATA4, RUNX1) were generated by picking single colonies from a polyclonal population. These lines were then subjected to the 5-day granulosa differentiation protocol. Cells were subsequently treated with androstenedione (500 ng/mL) for 24 hours in the presence or absence of FSH (0.25 IU/mL) and/or forskolin (100  $\mu$ M). Estradiol production was measured by ELISA. Three out of four lines showed hormone-responsive estradiol production, whereas the fourth line (found by PCR to be lacking RUNX1) constitutively produced high levels of estradiol. **FIG. 5B** shows that ovaroids produce both estradiol and progesterone. Estradiol production requires androstenedione and is stimulated by FSH. Results are shown for ovaroids formed with six different monoclonal samples of granulosa-like cells ( $n = 1$  sample per ovaroid per condition), at 3 days post aggregation.

**FIG. 6** shows a protocol for inducing granulosa-like cells. hiPSCs containing integrated TF expression plasmids are cultured in mTeSR™ Plus medium on Corning® Matrigel® Matrix. For induction of granulosa-like cells, hiPSCs are dissociated to single cells and plated on Corning® Matrigel® Matrix at a density of 10,000 – 20,000 per  $\text{cm}^2$  in DK10

medium (DMEM/F12 with GlutaMAX™ Supplement and 10% Knockout Serum Replacement) plus 3  $\mu$ M CHIR99021, 10  $\mu$ M Y-27632, and 1  $\mu$ g/mL doxycycline to induce TF expression. 48 hours later, and subsequently at 24 hour intervals, the medium is changed and replaced with fresh DK10 + 1  $\mu$ g/mL doxycycline. After a total of 120 hours, the granulosa-like cells are ready to use for downstream experiments.

**FIG. 7** shows the fraction of OCT4<sup>+</sup> and DAZL<sup>+</sup> cells relative to the total (DAPI<sup>+</sup>) over time in human ovaroids and mouse xeno-ovaroids. Counts were performed at eleven time points on images from two replicates of human ovaroids (F66/N.R1.G.F #4 and F66/N.R2 #1 granulosa-like cells + hPGCLCs) and one replicate of mouse xeno-ovaroids.

**FIGs. 8A-8D** show scRNA-seq analysis of ovaroids (F66/N.R1.G.F #4 granulosa-like cells + hPGCLCs). Data from all samples (days 2, 4, 8, and 14) were combined for joint dimensionality reduction and clustering. **FIG. 8A** shows expression (log2 CPM) of selected granulosa (FOXL2), stroma/theca (NR2F2), and germ cell (PRDM1) markers. **FIG. 8B** shows Leiden clustering shows four main clusters; the expression (log2 CPM) of marker genes is plotted for each. **FIG. 8C** shows mapping of cells onto a human fetal ovary reference atlas and assignment of cell types. **FIG. 8D** shows the proportion of somatic cell types, germ cells, DAZL<sup>+</sup> cells, and DDX4<sup>+</sup> cells in ovaroids from each day.

## DETAILED DESCRIPTION

Ovaria granulosa cells are important for many aspects of female reproduction, including the support of oocyte development and hormonal signaling during the menstrual cycle. Current *in vitro* models, including primary human or mouse granulosa cells and granulosa cell tumor lines, are inadequate for studying these processes (Havelock et al. 2004). For example, the COV434 ovarian tumor line, commonly used as a model for granulosa cells (Zhang 2000), lacks transcriptional and phenotypic characteristics of *bona fide* granulosa cells. Previously reported protocols for differentiating induced pluripotent stem cells (iPSCs) into granulosa-like cells are either low-yielding (Lan et al. 2013; Lipskind et al. 2018) or only applicable to mouse cells (Yoshino et al. 2021). Aspects of the present disclosure relate to a robust, scalable method of using direct transcription factor overexpression to mediate differentiation of iPSCs to produce granulosa-like cells that are AMHR2<sup>+</sup> (Anti-Mullerian Hormone Receptor Type 2), CD82<sup>+</sup> (Cluster of Differentiation 82), FOXL2<sup>+</sup> (Forkhead Box L2), and/or EPCAM<sup>+</sup> (Epithelial Cellular Adhesion Molecule) in about 4 to about 10 days. It should be understood that the term “granulosa-like cells” encompasses cells that express

granulosa-specific markers, such as AMHR2, CD82, and/or FOXL2 and/or do not express EPCAM, and exhibit other characteristics of naturally-occurring granulosa.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding a protein selected from Nuclear Receptor Subfamily 5 Group A Member 1 (NR5A1) and a Runt-Related Transcription Factor (RUNX) family protein. In some embodiments, the engineered polynucleotide comprising an open reading frame encoding NR5A1. In some embodiments, the engineered polynucleotide comprising an open reading frame encoding a RUNX family protein. In some embodiments, the RUNX family protein is Runt-Related Transcription Factor 1 (RUNX1). In some embodiments, the RUNX family protein is Runt-Related Transcription Factor 2 (RUNX2).

### **Granulosa-Like Cells**

Some aspects of the present disclosure provide granulosa-like cells and methods of producing such cells. A granulosa cell or follicular cell is a somatic cell of the sex cord that is closely associated with the developing female gamete (oocyte/egg) in the ovary of mammals. In the primordial ovarian follicle, and later in follicle development (folliculogenesis), granulosa cells advance to form a multilayered cumulus oophorus surrounding the oocyte in the preovulatory or antral (Graafian) follicle. The major functions of granulosa cells include the production of sex steroids, as well as myriad growth factors thought to interact with the oocyte during its development. The sex steroid production begins with follicle-stimulating hormone (FSH) from the anterior pituitary, stimulating granulosa cells to convert androgens (coming from the thecal cells) to estradiol by aromatase during the follicular phase of the menstrual cycle. After ovulation the granulosa cells turn into granulosa lutein cells that produce progesterone. The progesterone may maintain a potential pregnancy and causes production of a thick cervical mucus that inhibits sperm entry into the uterus.

There are two types of granulosa cells: cumulus cells (CC) and mural granulosa cells (MGC). Cumulus cells surround the oocyte. They provide nutrients to the oocyte and influence the development of the oocyte in a paracrine fashion. Mural granulosa cells line the follicular wall and surround the fluid-filled antrum. The oocyte secretes factors that determine the functional differences between CCs and MGCs. CCs primarily support growth and development of the oocyte whereas MGCs primarily serve an endocrine function and

support the growth of the follicle. Cumulus cells aid in oocyte development and show higher expression of SLC38A3, a transporter for amino acids, and Aldoa, Eno1, Ldh1, Pfkfb, Pkm2, and Tpi1, enzymes responsible for glycolysis. MGCs are more steroidogenically active and have higher levels of mRNA expression of steroidogenic enzymes such as cytochrome P450. MGCs produce an increasing amount of estrogen which leads to the LH surge. Following the LH surge, cumulus cells undergo cumulus expansion, in which they proliferate at a ten-fold higher rate than MGCs in response to FSH. During expansion CCs also produce a mucified matrix required for ovulation.

Granulosa cells express a number of different biomarkers that can be used to distinguish granulosa and granulosa-like cells from other cell types. For example, granulosa cells are typically positive for Anti-Mullerian Hormone Receptor Type 2 (AMHR2) (AMHR2+), CD82 molecule (CD82+), Forkhead Box L2 (FOXL2+), and negative for

Epithelial Cell Adhesion Molecule (EPCAM-). Thus, in some embodiments, the granulosa-like cells produced by the methods provided herein are AMHR2+, CD82+, FOXL2+, and/or EPCAM- granulosa-like cells (i.e., cells that express AMHR2, CD82, and/or FOXL2 protein but do not express detectable levels of EPCAM).

There are other characteristics of granulosa-like cells that distinguish them from non-granulosa-like cells including, but not limited to, expression of combinations of adhesion proteins such as adherens (B-catenin, a,catenin, N-cadherin, Nectins 1-3) as well as tight junctions (JAM-A, cingulin), desmosomes (dsg2, Dsc2) and linkers (afadin, ZO-1,2 and ZONAB). Furthermore, granulosa-like cells are distinguished by biosynthesis of combinations of estradiol, progesterone, and AMH.

### **Pluripotent Stem Cells**

The granulosa-like cells provided herein are differentiated from pluripotent stem cells. Pluripotent stem cells are cells that have the capacity to self-renew by dividing, and to develop into the three primary germ cell layers of the early embryo (e.g., ectoderm, endoderm, and mesoderm), and therefore into all cells of the adult body, but not extra-embryonic tissues such as the placenta (Shi et al. 2017).

Non-limiting examples of pluripotent stem cells include induced pluripotent cell (iPSCs), “true” embryonic stem cell (ESCs) derived from embryos, embryonic stem cells made by somatic cell nuclear transfer (ntESCs), and embryonic stem cells from unfertilized eggs (parthenogenesis embryonic stem cells, or pESCs). In some embodiments, a pluripotent cell is a human pluripotent cell.

In some embodiments, a pluripotent stem cell is an embryonic stem cell, such as a human embryonic stem cell. “Embryonic stem cell” is a general term for pluripotent stem cells that are made using embryos or eggs, rather than for cells genetically reprogrammed from the body. As used herein, “ESCs” encompass true ESCs, ntESCs, and pESCs.

In other embodiments, a pluripotent stem cell is an induced pluripotent stem cell, such as a human induced pluripotent stem cell. iPSCs may be derived from skin or blood cells that have been reprogrammed back into an embryonic-like pluripotent state that enables the development of an unlimited source of any type of human cell.

Some aspects of the present disclosure provide a PSC comprising: a protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2), wherein the protein is expressed or overexpressed. In some embodiments, the PSC further comprises a TCF21 protein. In some embodiments, the protein is expressed at a level that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, or at least 100% higher than a control level. In some embodiments, the PSC further comprises a GATA4 protein. In some embodiments, the protein is expressed at a level that is at least 5%, at least 10%, at least 15%, at least 20%, at least 25%, at least 50%, or at least 100% higher than a control level. In some embodiments, a control level is an endogenous level of the protein, for example in a naturally-occurring pluripotent stem cell. In some embodiments, a PSC comprises NR5A1. In some embodiments, a PSC expresses or overexpresses NR5A1. In some embodiments, a PSC comprises a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, a PSC expresses or overexpresses a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, a PSC comprises RUNX1. In some embodiments, a PSC expresses or overexpresses RUNX1. In some embodiments, a PSC comprises RUNX2. In some embodiments, a PSC expresses or overexpresses RUNX2. In some embodiments, a PSC comprises TCF21. In some embodiments, a PSC expresses or overexpresses TCF21. In some embodiments, a PSC comprises GATA4. In some embodiments, a PSC expresses or overexpresses GATA4.

Data provided herein shows that combinatorial expression of NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2) results in a 2-15 fold increase in efficiency of granulosa cell-like production, relative to a control, optionally wherein the control is efficiency of granulosa cell-like production in a PSC expressing only one of NR5A1 or a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, a PSC comprises NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, a PSC expresses or overexpresses NR5A1 and a RUNX family protein (e.g.,

RUNX1 and/or RUNX2). In some embodiments, a PSC comprises NR5A1 and RUNX1. In some embodiments, a PSC expresses or overexpresses NR5A1 and RUNX1. In some embodiments, a PSC comprises NR5A1 and RUNX2. In some embodiments, a PSC expresses or overexpresses NR5A1 and RUNX2. In some embodiments, a PSC further comprises TCF21. In some embodiments, a PSC further expresses or overexpresses TCF21. In some embodiments, a PSC further comprises GATA4. In some embodiments, a PSC further expresses or overexpresses GATA4.

### **Transcription Factors**

The granulosa-like cells provided herein are differentiated from pluripotent stem cells, in some embodiments, by expressing one or more (e.g., 2, 3, 4, 5, 6, 7, 8, or 9) transcription factors (i.e., a protein that controls the rate of transcription). Differentiation is the process by which an uncommitted cell or a partially committed cell commits to a specialized cell fate. Aspects of the present disclosure relate to the differentiation of uncommitted pluripotent stem cells into a granulosa-like cell fate.

In some embodiments, the transcription factors are selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress NR5A1. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress RUNX1. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress RUNX2. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress TCF21. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress GATA4. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress NR5A1 and RUNX1. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to express or overexpress NR5A1 and RUNX2. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to further express or overexpress TCF21. In some embodiments, pluripotent stem cells, such as hPSCs or iPSCs, are engineered to further express or overexpress GATA4.

A cell “expressed” a particular protein if the level of the protein in the cell is detectable (e.g., using a known protein assay). A cell “overexpresses” a particular protein (e.g., engineered polynucleotide encoding the protein) if the level of the protein is higher than (e.g., at least 5%, at least 10%, or at least 20% higher than) the level of the protein expressed from an endogenous, naturally-occurring polynucleotide encoding the protein.

### **Engineered Polynucleotides and Polypeptides**

The pluripotent stem cells of the present disclosure, in some embodiments, comprise engineered polynucleotides. An engineered polynucleotide is a nucleic acid (e.g., at least two nucleotides covalently linked together, and in some instances, containing phosphodiester bonds, referred to as a phosphodiester backbone) that does not occur in nature. Engineered polynucleotides include recombinant nucleic acids and synthetic nucleic acids. A recombinant nucleic acid is a molecule that is constructed by joining nucleic acids (e.g., isolated nucleic acids, synthetic nucleic acids or a combination thereof) from two different organisms (e.g., human and mouse). A synthetic nucleic acid is a molecule that is amplified or chemically, or by other means, synthesized. A synthetic nucleic acid includes those that are chemically modified, or otherwise modified, but can base pair with (bind to) naturally occurring nucleic acid molecules. Recombinant and synthetic nucleic acids also include those molecules that result from the replication of either of the foregoing.

An engineered polynucleotide may comprise DNA (e.g., genomic DNA, cDNA or a combination of genomic DNA and cDNA), RNA or a hybrid molecule, for example, where the nucleic acid contains any combination of deoxyribonucleotides and ribonucleotides (e.g., artificial or natural), and any combination of two or more bases, including uracil, adenine, thymine, cytosine, guanine, inosine, xanthine, hypoxanthine, isocytosine and isoguanine.

In some embodiments, a polynucleotide is a complementary DNA (cDNA). cDNA is synthesized from a single-stranded RNA (e.g., messenger RNA (mRNA) or microRNA (miRNA)) template in a reaction catalyzed by reverse transcriptase.

Engineered polynucleotides of the present disclosure may be produced using standard molecular biology methods (see, e.g., *Green and Sambrook, Molecular Cloning, A Laboratory Manual*, 2012, Cold Spring Harbor Press). In some embodiments, nucleic acids are produced using GIBSON ASSEMBLY® Cloning (see, e.g., Gibson, D.G. *et al. Nature Methods*, 343–345, 2009; and Gibson, D.G. *et al. Nature Methods*, 901–903, 2010, each of which is incorporated by reference herein). GIBSON ASSEMBLY® typically uses three enzymatic activities in a single-tube reaction: 5′ exonuclease, the 3′ extension activity of a

DNA polymerase and DNA ligase activity. The 5' exonuclease activity chews back the 5' end sequences and exposes the complementary sequence for annealing. The polymerase activity then fills in the gaps on the annealed domains. A DNA ligase then seals the nick and covalently links the DNA fragments together. The overlapping sequence of adjoining fragments is much longer than those used in Golden Gate Assembly, and therefore results in a higher percentage of correct assemblies. Other methods of producing engineered polynucleotides may be used in accordance with the present disclosure.

In some embodiments, an engineered polynucleotide comprises a promoter operably linked to an open reading frame. A promoter is a nucleotide sequence to which RNA polymerase binds to initiate transcription (e.g., ATG). Promoters are typically located directly upstream from (at the 5' end of) a transcription initiation site. In some embodiments, a promoter is a heterologous promoter. A heterologous promoter is not naturally associated with the open reading frame to which it is operably linked.

In some embodiments, a promoter is an inducible promoter. An inducible promoter may be regulated *in vivo* by a chemical agent, temperature, or light, for example. Inducible promoters enable, for example, temporal and/or spatial control of gene expression. Inducible promoters for use in accordance with the present disclosure include any inducible promoter described herein or known to one of ordinary skill in the art. Examples of inducible promoters include, without limitation, chemically/biochemically-regulated and physically-regulated promoters such as alcohol-regulated promoters, tetracycline-regulated promoters (e.g., anhydrotetracycline (aTc)-responsive promoters and other tetracycline responsive promoter systems, which include a tetracycline repressor protein (tetR), a tetracycline operator sequence (tetO) and a tetracycline transactivator fusion protein (tTA)), steroid-regulated promoters (e.g., promoters based on the rat glucocorticoid receptor, human estrogen receptor, moth ecdysone receptors, and promoters from the steroid/retinoid/thyroid 25 receptor superfamily), metal-regulated promoters (e.g., promoters derived from metallothionein (proteins that bind and sequester metal ions) genes from yeast, mouse and human), pathogenesis-regulated promoters (e.g., induced by salicylic acid, ethylene or benzothiadiazole (BTH)), temperature/heat-inducible promoters (e.g., heat shock promoters), and light-regulated promoters (e.g., light responsive promoters from plant cells). In some embodiments, the inducible promoter is a tetracycline-inducible promoter. In some embodiments, the inducible promoter is a doxycycline-inducible promoter. In other embodiments, a promoter is a constitutive promoter (active *in vivo*, unregulated).

An open reading frame is a continuous stretch of codons that begins with a start codon (e.g., ATG), ends with a stop codon (e.g., TAA, TAG, or TGA), and encodes a polypeptide, for example, a protein. An open reading frame is operably linked to a promoter if that promoter regulates transcription of the open reading frame.

Vectors used for delivery of an engineered polynucleotide include minicircles, plasmids, bacterial artificial chromosomes (BACs), and yeast artificial chromosomes. Transposon-based systems, such as the piggyBac™ system (e.g., Chen et al. Nature Communications. 2020; 11(1): 3446), is also contemplated herein.

A pluripotent stem cells, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, the engineered polynucleotide comprises an open reading frame encoding NR5A1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RUNX1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RUNX2. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding TCF21. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding GATA4.

In some embodiments, a pluripotent stem cell comprises an engineered polynucleotide comprising an open reading frame encoding NR5A1 and an engineered polynucleotide comprising an open reading frame encoding a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, a pluripotent stem cell comprises an engineered polynucleotide comprising an open reading frame encoding NR5A1 and an engineered polynucleotide comprising an open reading frame encoding RUNX1. In some embodiments, a pluripotent stem cell comprises an engineered polynucleotide comprising an open reading frame encoding NR5A1 and an engineered polynucleotide comprising an open reading frame encoding RUNX2. In some embodiments, a pluripotent stem cell further comprises an engineered polynucleotide comprising an open reading frame encoding TCF21. In some embodiments, a pluripotent stem cell further comprises an engineered polynucleotide comprising an open reading frame encoding GATA4.

An engineered polynucleotide encoding comprising an open reading frame encoding Nuclear Receptor Subfamily 5 Group A Member 1 (NR5A1) (e.g., UniprotKB Accession No. Q13285), in some embodiments, encodes a protein comprising the sequence of:

MDYSYDEDLDELCPVCGDKVSGYHYGLLTCECKGFFKRTVQNNKHYTCTESQSKIDKTQRKRCPFCRFQKCLT  
 VGMRLA VRADMRGRNKF GPMYKRDRALKQKKAI RANGFKLETGPPMGVPPPPPPADYVLPSSLHGPEPK  
 GLAAGPPAGPLGDFGAPALPMAVPGAHGPLAGYLYPAFPGRAIKSEYPEPYASPPQPGLPYGYPEPFGSGPNVPE  
 LILQLLQLEPDEDQVRARILGCLQEPTKSRPDQPAAGLLCRMADQTFISIVDWARRCMVFKELEVADQMTLLQN  
 CWSELLVFDHIYRQVQHGEKSILLVTGQVELTTVATQAGSLHSLVLRQELVLQLLALQLDRQEFVCLKFII  
 LFSLDLKFLLNNHILVKDAQEKANAALLDYTLCHYPHCGDKFQQLLLCLVEVRALSMQAKEYLYHKHLGNEMPRNN  
 LLIEMLQAKQT (SEQ ID NO: 1)

An engineered polynucleotide encoding comprising an open reading frame encoding Runt-Related Transcription Factor 1 (RUNX1) (e.g., UniprotKB Accession No. Q01196), in some embodiments, encodes a protein comprising the sequence of:

MRIPVDASTSRRFTPPSTALSPGKMSEALPLGAPDAGAALAGKLRS GDRSMVEVLADHPGELVRTDSPNFLCSVL  
 PTHWRCNKTLPIAFKVVALGDVPDGLTVTMAGNDENYSAELRNATAAMKNQVARFNDLRFVGRSGRGKSFTLTII  
 TVFTNPPQVATYHRAIKITVDGPREPRRHRQKLDDQTKPGSLSFSERLSELEQLRRTAMRVSPHHPAPTNPTRAS  
 LNHSTAFNPQPQSOMQDTRQIQPSPPSYDQSYQYLGSIASPSVHPATPISPGRASGMTTLSAELSSRLSTAPDL  
 TAFSDPRQFPALPISIDPRMHYPGAFTYSPTPVTSGIGIGMSAMGSATRYHTYLP PPYPGSSQAQGGPFQASSPS  
 YHLYYGASAGSYQFSMVGGERSPPRILPPCTNASTGSALLNPSLPNQSDVVEAEGSHSNSTPNMAPSARLEEAVW  
 RPY (SEQ ID NO: 2)

An engineered polynucleotide encoding comprising an open reading frame encoding Runt-Related Transcription Factor 2 (RUNX2) (e.g., UniprotKB Accession No. Q13950), in some embodiments, encodes a protein comprising the sequence of:

MASNSLFSVTTPCQQNFFWDPSTSRRFSPSSSLQPGKMSDVSPVVAQQQQQQQQQQQQQQQQQQQEEAAA  
 AAAAAAAAAAAAAA VPRLRPPHDNRTMVEIIADHPAELVRTDSPNFLCSVLPSHWRCNKTLPAFAFKVVALGEVPD  
 GTVVTVMAGNDENYSAELRNASAVMKNQVARFNDLRFVGRSGRGKSFTLTITVFTNPPQVATYHRAIKVTVDGPR  
 EPRRHRQKLDDSKPSLFS DRLSDLGRIPHP SMRVGVPPQNPRPSLNSAPSPFNPQGQSQITDPRQAQSSPPWSYD  
 QSYPSYLSQMTSPSIHSTPLSSTRGTGLPAITDVPRIISDDDTATSDFCLWPSTLSKKSQAGASELGPFS DPRQ  
 FPSISSLTESRFSNPRMHYPATFTYTPPVTSGMSLGM SATTHYHTYLP PPYPGSSQSQSGPFQTSSTPYLYGTS  
 SGSYQFPMPVGGDRSPSRMLPPCTTTNGSTLLNPNLPNQNDGVDADGSHSSSPTVLNSSGRMDES VWRPY  
 (SEQ ID NO: 3)

An engineered polynucleotide encoding comprising an open reading frame encoding Transcription factor 21 (TCF21) (e.g., UniprotKB Accession No. O43680), in some embodiments, encodes a protein comprising the sequence of:

MSTGSLSDVEDLQEVEMLECDGLKMDSNKEFVTSNESTEESNCENGSPQKGRGGLGKRRKAPT KKSPLSGVSQE  
 GKQVQRNAANARERARMRVLSKAFSRKLTLPWPVPDTKLSKLDTLRLASSYIAHLRQILANDKYENGYIHPVNL  
 TWPFMVAGKPESDLKEVVTASRLCGTTAS (SEQ ID NO: 4)

An engineered polynucleotide encoding comprising an open reading frame encoding GATA Binding Protein 4 (GATA4) (e.g., UniprotKB Accession No. P43694), in some embodiments, encodes a protein comprising the sequence of:

MYQSLAMAANHGPYPGAYEAGGPGAFMHGAGAASSPVYVPTPRVPSSVLGLSYLQGGGAGSASGGASGGSSGGAA  
 SGAGPGTQQGSPGWSQAGADGAAYTPPVSPRFSFPGTTGSLAAAAAAAAREAAAYSSGGGAAGAGLAGREQYG  
 RAGFAGSYSSPYPAYMADV GASWAAAAASAGPFDSPVLHSLPGRANPAARHPNLD MFDDFSEGRECVNCGAMST  
 PLWRRDGTGHYLCNACGLYHKMNGINRPLIKPQRRLSASRRVGLSCANCQTTTTTLWRRNAEGEPVCNACGLYMK  
 LHGVPRPLAMRKEGIQTRKRKPKNLNKS KTPAAPSGSESLPPASGASSNSSNATTSSSEEMRPIKTEPGLSSHYG  
 HSSSVSQTFVSAMSGHGPSIHPVLSALKLSPQGYASFPVSQSPQTSSKQDSWNSLVLADSHGDIITA (SEQ ID  
 NO: 5)

The number of copies of an engineered polynucleotide delivered to a PSC may vary. In some embodiments, a PSC comprises 1-20 copies of an engineered polynucleotide. For example, a PSC may comprise 1-15, 1-10, 2-10, 2-15, 2-10, 5-20, 5-15, or 5-10 copies of an engineered polynucleotide. In some embodiments, a PSC comprises 8-10 copies of an engineered polynucleotide. Greater than 20 copies are also contemplated herein.

### Methods of Producing Granulosa-Like Cells

The methods of producing granulosa-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2) to produce granulosa-like cells. In some embodiments, the methods further comprising expressing in PSCs of the expanded population a TCF21 protein. In some embodiments, the methods further comprising expressing in PSCs of the expanded population a GATA4 protein.

In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NR5A1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding a RUNX family protein (e.g., RUNX1 and/or RUNX2). In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RUNX1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RUNX2. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding TCF21. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding GATA4.

In some embodiments, the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

In some embodiments, the heterologous promoter is an inducible promoter, non-limiting examples of which are provided elsewhere herein.

The population a starting population comprises about  $1 \times 10^2$ - $1 \times 10^{10}$ , about  $1 \times 10^2$ - $1 \times 10^9$ , about  $1 \times 10^2$ - $1 \times 10^8$ , or about  $1 \times 10^2$ - $1 \times 10^7$  PSCs. In some embodiments, the population comprises about  $1 \times 10^3$ - $1 \times 10^8$  or about  $1 \times 10^3$ - $1 \times 10^7$  PSCs. In some embodiments, the population comprises about  $1 \times 10^4$ - $1 \times 10^7$  or about  $1 \times 10^5$ - $1 \times 10^6$  PSCs. In some embodiments, the population comprises about  $1 \times 10^1$  PSCs, about  $1 \times 10^2$  PSCs, about  $1 \times 10^3$  PSCs, about  $1 \times 10^4$  PSCs, about  $1 \times 10^5$  PSCs, about  $1 \times 10^6$  PSCs, about  $1 \times 10^7$  PSCs, about  $1 \times 10^8$  PSCs, about  $1 \times 10^9$  PSCs, or about  $1 \times 10^{10}$  PSCs.

In some embodiments, the population of PSCs is cultured for about 4 to about 10 days, about 4 to about 9 days, about 4 to about 8 days, about 4 to about 7 days, about 4 to about 6 days, about 5 to about 10 days, about 5 to about 9 days, about 5 to about 8 days, about 5 to about 7 days, or about 5 to about 6 days. In some embodiments, the population of PSCs is cultured for about 4 days, about 5 days, about 6 days, about 7 days, about 8 days, about 9 days, or about 10 days.

Some methods of the present disclosure provide methods comprising (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2); (b) culturing the PSCs in feeder-free, serum-free culture media to produce an expanded population of PSCs; and (c) culturing PSCs of the expanded population in a series of induction media comprising an inducing agent to produce AMHR2+, CD82+, FOXL2+, and/or EPCAM- granulosa-like cells. In some embodiments, the series of induction media comprises a first, a second, a third, and a fourth induction media. In some embodiments, an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a TCF21 protein is also delivered to the PCSs (e.g., in step (a)). In some embodiments, an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a GATA4 protein is also delivered to the PCSs (e.g., in step (a)).

In some embodiments, the PSCs are cultured in feeder-free, serum-free culture media for about 6 to about 24 hours. For example, the PSC may be cultured in feeder-free, serum-free culture media for about, 6 to about 12 hours. In some embodiments, the PSCs are cultured in feeder-free, serum-free culture media for about 6 hours, about 7 hours, about 8 hours, about 9 hours, about 10 hours, about 11 hours, about 12 hours, about 13 hours, about

14 hours, about 15 hours, about 16 hours, about 17 hours, about 18 hours, about 19 hours, about 20 hours, about 21 hours, about 22 hours, about 23 hours, or about 24 hours.

In some embodiments, the expanded population of PSCs comprises at least  $5 \times 10^3$  PSCs. For example, the expanded population (e.g., at the time of induction) may comprise at least  $1 \times 10^4$ , at least  $1 \times 10^5$ , at least  $1 \times 10^6$ , or at least  $1 \times 10^7$  PSCs. In some embodiments, the expanded population of PSCs comprises about  $5 \times 10^3$  PSCs to about  $1 \times 10^7$  PSCs.

In some embodiments, PSCs of the expanded population are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 30,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 25,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 20,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 15,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of about 15,000 cells/cm<sup>2</sup> to about 30,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of about 15,000 cells/cm<sup>2</sup> to about 25,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of about 15,000 cells/cm<sup>2</sup> to about 20,000 cells/cm<sup>2</sup>. In some embodiments, PSCs of the expanded population are cultured at a density of at least 10,000/cm<sup>2</sup>, at least 15,000/cm<sup>2</sup>, at least 20,000/cm<sup>2</sup>, at least 25,000/cm<sup>2</sup>, or at least 30,000/cm<sup>2</sup>.

In some embodiments, PSCs of the expanded population are cultured for no longer than 10 days, no longer than 9 days, no longer than 8 days, no longer than 7 days, no longer than 6 days, no longer than 5 days, or no longer than 4 days. For example, PSCs of the expanded population may be cultured for about 4 to about 10 days, about 4 to about 9 days, about 4 to about 8 days, about 4 to about 7 days, about 4 to about 6 days, about 5 to about 10 days, about 5 to about 9 days, about 5 to about 8 days, about 5 to about 7 days, or about 5 to about 6 days. In some embodiments, PSCs of the expanded population are cultured for about 4 days, about 5 days, about 6 days, about 7 days, about 8 days, about 9 days, or about 10 days.

In some embodiments, PSCs of the expanded population are cultured in a first induction media for about 36 to about 60 hours. For example, the PSC may be cultured in a first induction media for about 36 to about 54 hours, about 36 to about 48 hours, about 42 to about 60 hours, about 42 to about 54 hours, about 42 to about 48 hours, about 48 to about 60 hours, or about 48 to about 54 hours. In some embodiments, the PSCs are cultured in a first

induction media for about 36 hours, about 42 hours, about 48 hours, about 54 hours, or about 60 hours.

In some embodiments, PSCs of the expanded population are cultured in a second induction media for about 96 to about 144 hours. For example, the PSC may be cultured in a second induction media for about 96 to about 132 hours, about 96 to about 120 hours, about 96 to about 108 hours, about 108 to about 144 hours, about 108 to about 132 hours, about 108 to about 120 hours, about 120 to about 144 hours, or about 120 to about 132 hours. In some embodiments, the PSCs are cultured in a second induction media for about 96 hours, about 108, about 120 hours, about 132 hours, or about 144 hours.

Culturing in the second induction media comprises, in some embodiments, several (one or more) media changes. For example, the second induction media may be removed and replaced with new (fresh) second indication media every (about) 12 hours, every 24, hours, every 36 hours, or every 48 hours. In some embodiments, the second induction media is changed every (about) 24 hours.

### **Transfection Methods**

The engineered polynucleotide of the present disclosure may be delivered to a PSC using any one or more transfection method, including chemical transfection methods, viral transduction methods, and electroporation.

In some embodiments, an engineered polynucleotide is delivered on a vector. A vector is any vehicle, for example, a virus or a plasmid, that is used to transfer a desired polynucleotide into a host cell, such as a PSC. In some embodiments, the vector is a viral vector. In some embodiments, a viral vector is not a naturally occurring viral vector. The viral vector may be from adeno-associated virus (AAV), adenovirus, herpes simplex virus, lentiviral, retrovirus, varicella, variola virus, hepatitis B, cytomegalovirus, JC polyomavirus, BK polyomavirus, monkeypox virus, Herpes Zoster, Epstein-Barr virus, human herpes virus 7, Kaposi's sarcoma-associated herpesvirus, or human parvovirus B 19. Other viral vectors are encompassed by the present disclosure.

In some embodiments, a viral vector is an AAV vector. AAV is a small, non-enveloped virus that packages a single- stranded linear DNA genome that is approximately 5 kb long and has been adapted for use as a gene transfer vehicle (Samulski, RJ et al., *Annu Rev Virol.* 2014;1(1):427-51). The coding regions of AAV are flanked by inverted terminal repeats (ITRs), which act as the origins for DNA replication and serve as the primary packaging signal (McLaughlin, SK et al. *Virol.* 1988;62(6): 1963-73; Hauswirth, WW et al.

1977;78(2):488-99). Thus, an AAV vector typically includes ITR sequences. Both positive and negative strands are packaged into virions equally well and capable of infection (Zhong, L et al. Mol Ther. 2008;16(2):290-5; Zhou, X et al. Mol Ther. 2008;16(3):494- 9; Samulski, RJ et al. Virol. 1987;61(10):3096-101). In addition, a small deletion in one of the two ITRs allows packaging of self-complementary vectors, in which the genome self-anneals after viral uncoating. This results in more efficient transduction of cells but reduces the coding capacity by half (McCarty, DM et al. Mol Ther. 2008;16(10): 1648-56; McCarty, DM et al. Gene Ther. 2001;8(16): 1248-54).

In some embodiments, a polynucleotide is delivered to a cell using a transposon/transposase system. For example, the piggyBac™ transposon system may be used. A piggyBac™ transposon is a mobile genetic element that efficiently transposes between vectors and chromosomes via a “cut and paste” mechanism (Woodard et al. 2015). During transposition, the piggyBac™ transposase recognizes transposon-specific inverted terminal repeat sequences (ITRs) located on both ends of the transposon vector and efficiently moves the contents from the original sites and integrates them into TTAA chromosomal sites. The piggyBac™ transposon system facilitates efficient integration of a polynucleotide into a cell genome.

Thus, in some embodiments, the method further comprises delivering to a PSC a transposon comprising an engineered polynucleotide and also delivering a transposase.

In some embodiments, an engineered polynucleotide is delivered to a cell using electroporation. Electroporation is a physical transfection method that uses an electrical pulse to create temporary pores in cell membranes through which the engineered polynucleotide can pass into cells. *See, e.g., Chicaybam L et al. Front. Bioeng. Biotechnol.*, 23 January 2017.

Following transfection, the engineered polynucleotides may be integrated into the genome of a PSC. In some embodiments, an engineered polynucleotide may further comprise an antibiotic resistance gene to confer resistance to an antibiotic used in an antibiotic drug selection process. In this way, a ‘pure’ population of cells comprising an integrated engineered polynucleotide may be obtained. In some embodiments, a population of cells comprising an integrated engineered polynucleotide are selected using antibiotic drug selection. Antibiotic drug selection is the process of treating a population of cells with an antibiotic so that only cells that are capable of surviving in the presence of said antibiotic will remain in the population. Non-limiting examples of antibiotics that may be used for antibiotic drug selection include: puromycin, blasticidin, geneticin, hygromycin, mycophenolic acid, zeocin, carbenicillin, kanamycin, ampicillin, and actinomycin.

### Culture Media

The methods provided herein, in some embodiments, comprise culturing PSCs in a feeder-free, serum-free culture media. Culture media may comprise, for example, a solubilized basement membrane preparation extracted from the Engelbreth-Holm-Swarm (EHS) mouse sarcoma (e.g., Corning® Matrigel® Matrix) (coated at  $\sim 75 \mu\text{l}/\text{cm}^2$  to  $\sim 150 \mu\text{l}/\text{cm}^2$  of lot-based diluted suspension). In some embodiments, the solubilized basement membrane preparation comprises one or more extracellular matrix (ECM) protein and one or more growth factor. For example, the ECM proteins may be selected from Laminin, Collagen IV, heparan sulfate proteoglycans, and entactin/nidogen.

In some embodiments, culture media further comprises one or more growth factor, for example, selected from recombinant human basic fibroblast growth factor (rh bFGF) (e.g., 80 ng/ml to 120 ng/ml) and recombinant human transforming growth factor  $\beta$  (rh TGF $\beta$ ) (e.g., 20 pM to 25 pM). In some embodiments, culture media further comprises rh bFGF and rh TGF $\beta$ . In some embodiments, culture media comprises mTeSR™ Plus medium (STEMCELL Technologies).

In some embodiments, a first induction media comprises one or more of (e.g., 2, 3, 4, or more of) the first induction media comprises one or more of L-alanyl-L-glutamine (e.g., 1.8 mM to 2.2 mM), antibiotic (e.g., penicillin and/or streptomycin) (e.g., 45 U/ml to 50 U/ml), Dulbecco's Modified Eagle Medium (DMEM)/F-12 (e.g., 15 mM HEPES, no glutamine), Advanced RPMI (Roswell Park Memorial Institute) 1640 Medium (with non-essential amino acids and sodium pyruvate), a glycogen synthase kinase (GSK) 3 inhibitor (e.g., 3  $\mu\text{M}$  to 10  $\mu\text{M}$ ), a protein inhibitor (e.g. noggin, 2 ng/mL to 20 ng/mL) or small molecule inhibitor (e.g., dorsomorphin, e.g., 100 nM to 400 nM) of the BMP signaling pathway, a small molecule ROCK inhibitor (e.g., 9  $\mu\text{M}$  to 11  $\mu\text{M}$ ), and an inducing agent (e.g., doxycycline (e.g., 50 ng/ml to 2000 ng/ml)). For example, the first induction media may comprise DK10 medium, CHIR99021, Y-27632 (a small molecule ROCK inhibitor), and doxycycline.

In some embodiments, the second induction media comprises one or more of (e.g., 2, 3, 4, or more of), L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Advanced RPMI 1640 Medium (with non-essential amino acids and sodium pyruvate), DMEM/F-12, and an inducing agent (e.g., doxycycline). For example, the second induction media may comprise DK10 medium and doxycycline.

Advanced RPMI 1640 (Thermo Fisher Scientific) is manufactured with glucose, non-essential amino acids, sodium pyruvate (without L-glutamine and HEPES), vitamins, inorganic salts, proteins (e.g., AlbuMAX® II, human transferrin, and insulin recombinant full chain), and trace elements.

The DK10 medium used herein comprises KO-SR, glutamine, penicillin, streptomycin, and Dulbecco's Modified Eagle Medium (DMEM)/F-12.

The KnockOut™ Serum Replacement (KO-SR) used herein is a serum-free formulation used as a replacement for fetal bovine serum.

GlutaMAX™ Supplement comprises L-alanyl-L-glutamine, which is a dipeptide substitute for L-glutamine.

CHIR99021 is an aminopyrimidine derivative that is an extremely potent glycogen synthase kinase (GSK) 3 inhibitor, inhibiting both GSK3β (IC<sub>50</sub> = 6.7 nM) and GSK3α (IC<sub>50</sub> = 10 nM). GSK3 is a serine/threonine kinase that is a key inhibitor of the WNT pathway; therefore, CHIR99021 functions as a WNT activator.

Noggin is a protein that binds and inactivates proteins in the BMP family. Dorsomorphin is a small molecule inhibitor of type I BMP receptor serine/threonine kinases. Both these substances are considered inhibitors of the BMP signaling pathway.

### **Therapeutic Compositions and Method of Use**

The present disclosure provides, in some embodiments, therapeutic compositions comprising the granulosa-like cells produced herein. In some embodiments, the compositions further comprise a pharmaceutically-acceptable excipient. The compositions, in some embodiments, are cryopreserved.

Such compositions may be administered to a subject, such as a human subject, using any suitable route of administration. Suitable routes of administration include, for example, parenteral routes such as intravenous, intrathecal, parenchymal, or intraventricular routes. Suitable routes of administration include, for example, parenteral routes such as intravenous, intrathecal, parenchymal, or intraventricular injection.

In some embodiments, a subject is a human subject. Patients suffering from primary ovarian failure or menopause, in which granulosa cell function is compromised may benefit from the use of these cells. Thus, in some embodiments, the subject has been diagnosed with ovarian failure, and the granulosa cells provided herein are used to treat the ovarian failure. In other embodiments, the subject is going through menopause, and the granulosa cells provided herein are used to treat (e.g., alleviate the symptoms associated with) menopause.

Additionally, such compositions may be used to improve the quality and *in vitro* maturation of oocytes and/or embryos, for example, during an *in vitro* fertilization (IVF) process and/or during related assisted reproductive technologies (ART) procedures, such as egg freezing. Such administration methods include for example, co-culture of granulosa-like cells with immature and mature oocytes *in vitro*. In some embodiments, a subject is a human subject. Such subjects will be undergoing oocyte freezing or fertilization through IVF, for example, and may suffer from infertility, age-related oocyte immaturity, polycystic ovarian syndrome (PCOS) and/or ovarian hyperstimulation syndrome (OHSS), which leave many oocytes immature and unusable. Applications of granulosa-like cells *in vitro* may improve the quality and usability of these oocytes.

The compositions may be administered to a subject in a therapeutically effective amount. The term “therapeutically effective amount” refers to the number of granulosa required to confer therapeutic effect on a subject, either alone or in combination with at least one other active agent. Effective amounts vary, as recognized by those skilled in the art, depending on the route of administration, excipient usage, and co-usage with other active agents. The quantity to be administered depends on the subject to be treated, including, for example, the strength of an individual’s immune system or genetic predispositions. Suitable dosage ranges are readily determinable by one skilled in the art and may be on the order of micrograms of the polypeptide of this disclosure. The dosage of the preparations disclosed herein may depend on the route of administration and varies according to the size of the subject.

It is believed that one skilled in the art can, based on the above description, utilize the present invention to its fullest extent. The following specific embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever. All publications cited in the present application are incorporated by reference for the purposes or subject matter referenced in this disclosure.

### **Additional Embodiments**

The disclosure also relates to the additional embodiments set out in the numbered paragraphs below:

1. A pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein.

2. The PSC of paragraph 1, wherein the PSC comprises the engineered polynucleotide comprising an open reading frame encoding NR5A1.
3. The PSC of paragraph 1 or 2, wherein the PSC comprises the engineered polynucleotide comprising an open reading frame encoding a RUNX family protein.
4. The PSC of paragraph 3, wherein the RUNX family protein is RUNX1.
5. The PSC of paragraph 3 or 4, wherein the RUNX family protein is RUNX2.
6. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses: NR5A1; RUNX1; RUNX2; NR5A1 and RUNX1; NR5A1 and RUNX2; or NR5A1, RUNX1, and RUNX2.
7. The PSC of any one of the preceding paragraphs, wherein PSC further comprises an engineered polynucleotide comprising an open reading frame encoding a TCF21 protein.
8. The PSC of paragraph 7, wherein the PSC expresses or overexpresses TCF21.
9. The PSC of any one of the preceding paragraphs, wherein PSC further comprises an engineered polynucleotide comprising an open reading frame encoding a GATA4 protein.
10. The PSC of paragraph 7, wherein the PSC expresses or overexpresses GATA4.
11. The PSC of any one of the preceding paragraphs, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.
12. The PSC of paragraph 11, wherein the heterologous promoter is an inducible promoter.
13. A pluripotent stem cell (PSC) comprising: a protein selected from NR5A1 and a RUNX family protein, wherein the protein is overexpressed.
14. The PSC of paragraph 15, wherein the PSC expresses or overexpresses: NR5A1; RUNX1; RUNX2; NR5A1 and RUNX1; NR5A1 and RUNX2; or NR5A1, RUNX1, and RUNX2.
15. The PSC of paragraph 14, wherein the PSC further comprises a TCF21 protein.
16. The PSC of paragraph 15, wherein the PSC expresses or overexpresses TCF21.
17. The PSC of paragraph of any one of claims 14-16, wherein the PSC further comprises a GATA4 protein.
18. The PSC of paragraph of any one of claims 14-17, wherein the PSC expresses or overexpresses GATA4.

19. The PSC of any one of the preceding paragraphs, wherein the PSC is a human PSC.
20. The PSC of any one of the preceding paragraphs, wherein the PSC is an induced PSC (iPSC).
21. The PSC of any one of the preceding paragraphs, wherein the PSC comprises 1-20, optionally 8-10, copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2).
22. A composition comprising: a population of the PSC of any one of the preceding paragraphs or described elsewhere herein.
23. The composition of paragraph 22, wherein the population comprises at least 10,000/cm<sup>2</sup> of the PSC.
24. A method, comprising: culturing, in culture media, a population of pluripotent stem cells (PSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from NR5A1 and a RUNX family protein to produce granulosa-like cells.
25. The method of paragraph 24, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NR5A1.
26. The method of paragraph 24 or 25, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding a RUNX family protein.
27. The method of paragraph 26, wherein the RUNX family protein is RUNX1.
28. The method of paragraph 26 or 27, wherein the RUNX family protein is RUNX2.
29. The method of any one of paragraph 24-28, wherein the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding a TCF21 protein.
30. The method of any one of paragraph 24-29, wherein the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding a GATA4 protein.
31. The method of any one of the preceding paragraphs, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.
32. The method of any one of the preceding paragraphs, wherein the heterologous promoter is an inducible promoter.

33. The method of any one of the preceding paragraphs, wherein the population comprises  $1 \times 10^2$  -  $1 \times 10^7$  PSCs.
34. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for about 4-10 days.
35. The method of paragraph 34, wherein the population of PSCs is cultured for about 6 days.
36. The method of any one of the preceding paragraphs, wherein the granulosa-like cells are AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and/or EPCAM<sup>-</sup>.
37. The method of paragraph 36, wherein the granulosa-like cells are AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and EPCAM<sup>-</sup>.
38. A method comprising:
- (a) delivering to pluripotent stem cells (PSCs) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein;
  - (b) culturing the PSCs in feeder-free, serum-free culture media to produce an expanded population of PSCs; and
  - (c) culturing PSCs of the expanded population in a series of induction media comprising an inducing agent to produce AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and/or EPCAM<sup>-</sup> granulosa-like cells.
39. The method of paragraph 38 comprising delivering to PSCs (i) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding NR5A1 and (i) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a RUNX family protein.
40. The method of paragraph 38 or 39, wherein the RUNX family protein is RUNX1.
41. The method of any one of paragraphs 38-40, wherein the RUNX family protein is RUNX2.
42. The method of any one of paragraphs 38-41, wherein the engineered polynucleotide is a transposon and the delivering further comprises delivering a transposase to the PSCs.
43. The method of any one of paragraphs 38-42, wherein the inducible promoter is a chemically-inducible promoter, optionally a doxycycline-inducible promoter.

44. The method of any one of paragraphs 38-43, wherein the feeder-free, serum-free culture media of (b) comprises a solubilized basement membrane preparation extracted from the Engelbreth-Holm-Swarm (EHS) mouse sarcoma.
45. The method of paragraph 44, wherein the solubilized basement membrane preparation comprises extracellular matrix (ECM) proteins and growth factors.
46. The method of paragraph 45, wherein the ECM proteins are selected from Laminin, Collagen IV, heparan sulfate proteoglycans, and entactin/nidogen.
47. The method of any one of paragraphs 38-46, wherein the feeder-free, serum-free culture media of (b) comprises growth factors selected from recombinant human basic fibroblast growth factor (rh bFGF) and recombinant human transforming growth factor  $\beta$  (rh TGF $\beta$ ).
48. The method of any one of paragraphs 38-47, wherein the culturing of (b) is for about 6-24 hours.
49. The method of any one of paragraphs 38-48, wherein the PSCs of the expanded population of (c) are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 20,000 cells/cm<sup>2</sup>.
50. The method of any one of paragraphs 38-49, wherein the culturing of (c) comprises culturing the PSCs in a first induction media and culturing the PSCs in a second induction media.
51. The method of paragraph 50, wherein the first induction media comprises one or more of L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Dulbecco's Modified Eagle Medium (DMEM)/F-12, Advanced RPMI (Roswell Park Memorial Institute) 1640 Medium, a glycogen synthase kinase (GSK) 3 inhibitor, a small molecule or protein inhibitor of the BMP signaling pathway, a small molecule ROCK inhibitor, and an inducing agent (e.g., doxycycline).
52. The method of paragraph 50 or 51, wherein the culturing the PSCs in a first induction media is for about 36 to about 60 hours, optionally about 48 hours.
53. The method of any one of paragraphs 50-52, wherein the second induction media comprises one or more of L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Advanced RPMI 1640 Medium, DMEM/F-12, and an inducing agent (e.g., doxycycline).
54. The method of any one of paragraphs 50-53, wherein the culturing the PSCs in a second induction media is for about 96 to about 144 hours, optionally about 120 hours.

55. The method of paragraph 54, wherein the second induction media is removed and replaced with fresh second induction media at about 24-hour intervals.

56. A granulosa-like cell produced by the method of any one of the preceding paragraphs.

57. An ovarian organoid comprising granulosa-like cells of any one of the preceding paragraphs and human primordial germ cell-like cells (hPGCLCS).

58. The method of any one of the preceding paragraphs further comprising combining the granulosa-like cells with hPGCLCS to form an ovarian organoid.

## EXAMPLES

### **Example 1. Directed differentiation of hiPSCs to granulosa-like cells by transcription factor expression**

Ovarian granulosa cells are important for many aspects of female reproduction, including the support of oocyte development and hormonal signaling during the menstrual cycle. Current *in vitro* models, including primary human or mouse granulosa cells and granulosa cell tumor lines, are inadequate for studying these processes. For example, it was found that the COV434 ovarian tumor line, commonly used as a model for granulosa cells, lacks transcriptional and phenotypic characteristics of *bona fide* granulosa cells. Previously reported protocols for differentiating iPSCs into granulosa-like cells are either low-yielding or only applicable to mouse cells. To solve this challenge, a robust, scalable method of producing granulosa-like cells by transcription factor (TF)-mediated differentiation of iPSCs was developed.

To begin, a list of candidate TFs to screen for granulosa induction were identified. To generate this list, 22 TFs that were differentially expressed in granulosa cells compared to hESCs and early mesoderm were identified using previously published datasets for these cell types. Five TFs, identified as important in mouse ovarian development by previous developmental biology studies, were also included. Finally, nine additional TFs were identified that were predicted to be upstream of the others on the list, based on a gene regulatory network analysis taking into account co-expression data as well as binding motifs.

Next, a hiPSC line was engineered with a tdTomato reporter for the granulosa-specific protein FOXL2. Additionally, a barcoded cDNA plasmid library was generated for doxycycline-inducible expression of these TFs. By integrating these plasmids into the FOXL2-tdTomato reporter iPSCs, inducing TF expression under various conditions, and

capturing barcodes from FOXL2+ cells, a set of TFs that were enriched in the granulosa-like population were determined (**FIG. 1**).

Based on these results, combinations of TFs of interest were screened in order to see which TFs could best induce a granulosa-like phenotype (**FIG. 2**). It was found that NR5A1 expression was essential for efficient estradiol production, and that the combination of NR5A1 and a RUNX family member (either RUNX1 or RUNX2) was best for inducing FOXL2 expression and granulosa surface markers. It was further validated that the FOXL2+ granulosa-like cells expressed transcriptional markers of granulosa and gonadal cells, and not markers from the related adrenal lineage (**FIG. 3**).

Next, monoclonal hiPSC lines were generated with the top TFs integrated. After choosing lines with efficient induction of granulosa-like cells (**FIG. 4**), their hormonal signaling was validated by measuring estradiol production in response to follicle-stimulating hormone (FSH) or forskolin, which increases levels of the second messenger cAMP. It was found that three lines containing integrated expression plasmids for NR5A1, TCF21, and RUNX1 produced estradiol in response to either FSH or forskolin, as expected for granulosa cells (**FIG. 5**). A fourth line with no RUNX transgene constitutively produced high levels of estradiol, suggesting that RUNX expression is necessary for proper regulation of estradiol production.

The overall protocol for producing granulosa-like cells is depicted in FIG. 6. hiPSCs containing integrated TF expression plasmids are cultured in mTeSR™ Plus medium on Corning® Matrigel® Matrix. For induction of granulosa-like cells, hiPSCs were dissociated to single cells and plated on Corning® Matrigel® Matrix or Collagen I at a density of ~10,000 – 20,000 per cm<sup>2</sup> in DK10 medium (DMEM/F12 with GlutaMAX™ Supplement and 10% Knockout Serum Replacement) plus ~3-4 μM CHIR99021, ~8-12 μM Y-27632, and ~0.5-3 μg/mL doxycycline to induce TF expression. 48 hours later, and subsequently at 24 hour intervals, the medium was changed and replaced with fresh DK10 medium + ~0.5-3 μg/mL doxycycline. After a total of 120 hours, the granulosa-like cells were ready to use for downstream experiments.

In an alternate protocol, iPSCs are cultured on Matrigel or laminin-coated plates and grown to 20 – 40% confluency in mTeSR Plus medium. For induction of granulosa-like cells, the mTeSR Plus is removed, the iPSCs are washed with phosphate-buffered saline (PBS), and the stage 1 medium, containing Advanced RPMI with GlutaMAX™ Supplement, CHIR99021 (~5-8 μM), and either dorsomorphin (~100–400 nM) or noggin (~4-8 ng/mL), is added. After 48 hours the stage 1 medium is removed and replaced with fresh stage 1

medium. After an additional 48 hours the stage 1 medium is removed and replaced with stage 2 medium, containing Advanced RPMI with GlutaMAX™ Supplement and ~0.5-3 µg/mL doxycycline. The stage 2 medium is changed every 24 hours. Granulosa-like cells can be harvested after a total of ~3–4 days in stage 2 medium (~7–8 days total).

The protocol presented herein has several advantages over previously reported methods. It is rapid, producing granulosa-like cells in ~5-8 days, and the resulting cells are sufficiently pure (**FIG. 4**) to not require enrichment by FACS. Due to the monolayer format, it is easily scalable (50 million cells were produced at once and the format is compatible with higher throughput assays). Most importantly, the granulosa-like cells transcriptionally (**FIG. 3**) and phenotypically (**FIG. 5**) display key characteristics of granulosa cells, making them suitable as a model for ovarian biology.

### **Example 2. Granulosa-like cells respond to FSH and perform steroidogenesis**

The ability of the granulosa-like cells to carry out one of the key endocrine functions of granulosa cells (i.e., the production of estradiol) was evaluated. In the ovary, theca cells convert cholesterol to androstenedione, which is the substrate for estradiol production in granulosa cells. The rate-limiting step is oxidative decarboxylation by CYP19A1 (aromatase), producing estrone, which is subsequently reduced to estradiol by enzymes in the HSD17B family, typically HSD17B1 in granulosa cells. *In vivo*, this pathway of estrogen synthesis is stimulated by FSH. The granulosa-like cells were treated with androstenedione, in the presence or absence of FSH or forskolin (which directly increases levels of the FSHR second messenger cAMP). As controls, COV434 and KGN ovarian tumor cells were used, which produce estradiol from androstenedione, as well as immortalized primary human granulosa cells (HGL5) and primary adult mouse ovarian somatic cells. The granulosa-like cells produced estradiol from androstenedione, and in seven out of the nine monoclonal lines tested, this steroidogenic activity significantly increased upon stimulation with FSH or forskolin (**FIG. 5A**). One of the granulosa lines (F3/N.T #5) produced high levels of estradiol in all conditions, and this line, unlike the others, had neither RUNX1 nor RUNX2 expression vectors integrated (data not shown).

The levels of estradiol produced by three FSH-responsive lines were similar to those produced by KGN human granulosa tumor cells, which also showed responses to FSH and forskolin (**FIG. 5A**). In contrast, COV434 cells, which showed no FSHR expression in the RNA seq data (**FIG. 3**), were unresponsive to FSH alone, producing estradiol only in the presence of forskolin. HGL5 immortalized human granulosa cells did not produce estradiol

under any condition. Primary adult mouse ovarian somatic cells produced similar amounts of estradiol to the hiPSC-derived granulosa-like cells (**FIG. 5A**); however, the mouse cells did not show a response to FSH or forskolin, possibly because they were already exposed to FSH *in vivo*. Whether the granulosa-like cells maintained their steroidogenic activity during ovaroid (ovarian organoid) co-culture with hPGCLCs was also investigated. Hormone levels in ovaroid supernatants in the presence or absence of androstenedione and FSH were measured. In addition to estradiol, progesterone was also measured; granulosa cells produce progesterone *in vivo* after ovulation and formation of the corpus luteum. Production of both hormones in five out of six samples (**FIG. 5B**) was observed. Estradiol was produced only in the presence of androstenedione supplementation, and levels increased with FSH treatment. Progesterone was produced in all conditions but was highest in the absence of androstenedione.

### **Example 3. Granulosa-like cells support germ cell development within ovaroids**

Current methods for inducing and culturing human PGC-like cells (hPGCLCs) produce cells corresponding to immature, premigratory PGCs that lack expression of gonadal PGC markers such as DAZL. During fetal development, PGCs mature through interactions with gonadal somatic cells, with DAZL playing a key role in downregulation of pluripotency factors and commitment to gametogenesis. This process has recently been recreated *in vitro* using mouse fetal ovarian somatic cells, which allowed the development of hPGCLCs to the oogonia like stage. *In vitro*-derived human granulosa-like cells could perform a similar role, with the potential for eliminating interspecies developmental mismatches. Therefore, the granulosa-like cells were combined with hPGCLCs to form ovarian organoids, which were termed “ovaroids”.

To generate ovaroids, these two cell types were aggregated in low-binding U-bottom wells, followed by transfer to air-liquid interface Transwell culture. As a comparison, fetal mouse ovarian somatic cells were isolated and aggregate with hPGCLCs. Expression of the mature marker DAZL was observed by immunofluorescence beginning in a subset of OCT4+ hPGCLCs at 4 days of co-culture with hiPSC-derived granulosa-like cells (data not shown). In contrast, robust DAZL expression in co-culture with mouse cells was not observed until day 32 (data not shown), with fainter expression visible at day 26. Similarly, in a previous study using the same hPGCLC line and anti-DAZL antibody, DAZL expression was observed only after 77 days of co-culture with mouse fetal testis somatic cells. The fraction of DAZL<sup>+</sup> cells reached its maximum at day 14 in human ovaroids and day 38 in mouse

ovaroids (**FIG. 7**). In human ovaroids, the fraction of OCT4<sup>+</sup> cells declined after day 8. In mouse ovaroids, the fraction of OCT4<sup>+</sup> cells also declined over time. At day 16 in human ovaroids, DAZL<sup>+</sup> OCT4<sup>-</sup> cells were also apparent (*in situ* images not shown) in addition to DAZL<sup>+</sup> OCT4<sup>+</sup> cells, and past day 38 there were more DAZL<sup>+</sup> cells than OCT4<sup>+</sup> cells in total (**FIG. 7**). The downregulation of OCT4 in DAZL<sup>+</sup> oogonia occurs *in vivo* during the 2<sup>nd</sup> trimester of human fetal ovarian development; however, the transition of DAZL to exclusively cytoplasmic localization that was reported to take place at this stage was not observed. Expression of TFAP2C, an early-stage PGC marker, declined during ovaroid culture and was almost entirely absent by day 8. By contrast, SOX17 expression was still visible on day 8, and OCT4 and DAZL expression continued to day 54. Although this system allowed the rapid development of hPGCLCs to the gonadal stage, the number of germ cells in both hiPSC-derived and mouse-derived ovaroids declined over prolonged culture (**FIG. 7**), indicating that either they were dying or differentiating into other lineages. Unlike mouse-derived ovaroids, the hiPSC-derived ovaroids cultured on Transwells gradually flattened and widened, and by day 38 were largely collapsed.

Nonetheless, in these longer-term experiments, the formation of empty follicle-like structures composed of cuboidal AMHR2<sup>+</sup> FOXL2<sup>+</sup> granulosa-like cells was observed (*in situ* images not shown), suggesting that the TFs could drive folliculogenesis even in the absence of oocytes. Follicle-like structure formation was first visible at day 16, and at day 26 the largest of these structures had grown to 1–2 mm diameter. At day 70, ovaroids had developed follicles of a variety of sizes, mainly small single-layer follicles but also including antral follicles. Cells outside of the follicles stained positive for NR2F2, a marker of ovarian stromal and theca cells. To further examine the gene expression of hPGCLCs and somatic cells in this system, scRNA-seq was performed on dissociated ovaroids at days 2, 4, 8, and 14 of culture, and clustered cells according to gene expression. As expected, the largest cluster (cluster 0) contained cells expressing granulosa markers such as *FOXL2*, *WNT4*, and *CD82* (**FIG. 8** and **8B**). Cells expressing markers of secondary/antral granulosa cells such as *FSHR* and *CYP19A1* were also found within this cluster, although these were much less numerous. A smaller cluster (cluster 1) expressing the ovarian stromal marker *NR2F2* was also present. *NR2F2* is expressed by both stromal and theca cells, but the cells in cluster 1 did not express 17 $\alpha$ -hydroxylase (*CYP17A1*), indicating that they could not produce androgens and were not theca cells.

A cluster of hPGCLCs expressing marker genes such as *CD38*, *KIT*,

*PRDM1*, *TFAP2C*, *PRDM14*, *NANOG*, and *POU5F1* was also observed. Notably, X-chromosomal lncRNAs *XIST*, *TSIX*, and *XACT* were all more highly expressed (an average of ~80-fold, ~20-fold, and ~2900-fold, respectively) in the hPGCLCs relative to other clusters (**FIG. 8B**), suggesting that the hPGCLCs were starting the process of X-reactivation, which in hPGCs is associated with high expression of both *XIST* and *XACT*. The X-chromosomal *HPRT1* gene, known to be more highly expressed in cells with two active X chromosome, was also ~3-fold upregulated.

Next, the *in vitro*-generated ovaroids were compared to a reference atlas of human fetal ovarian development. Scanpy ingest was used to integrate samples into the atlas and annotate each cell with the closest cell type from the *in vivo* data (**FIG. 8C**). The ovaroids consisted mainly of granulosa, gonadal mesenchyme, and pre-granulosa lineages (**FIG. 8D**), with a small fraction of coelomic epithelium. The fraction of granulosa cells increased from day 2 through day 8, potentially representing a maturation of the somatic cell population. As expected, neural, immune, smooth muscle, and erythroid cells, which were present in fetal ovaries, were completely absent from the ovaroids. Epithelial, endothelial, and perivascular cells were detected, but at very low frequency (1% or less), possibly representing a low rate of off-target differentiation.

The overall fraction of germ cells was additionally examined, as well as the fraction of cells expressing the gonadal germ cell markers *DAZL* and *DDX4*, over the course of the experiment (**FIG. 8D**). The germ cell population was defined based on the fetal ovary atlas integration. This population increased from days 2 to 4, but declined thereafter. In comparison, the fraction of *DAZL*<sup>+</sup> and *DDX4*<sup>+</sup> cells also increased from days 2 to 4, but remained roughly constant from days 4 to 14 (**FIG. 8D**). A differential gene expression analysis and gene ontology enrichment were performed on *DAZL*<sup>+</sup> cells relative to *DAZL*<sup>-</sup> cells. Upregulated genes ( $\log_2\text{fc} > 2$ ,  $n = 221$ ) were most highly enriched for terms related to generic developmental processes but also included terms related to adhesion and migration (e.g., “ameboidal-type cell migration”), as well as reproductive system development. Downregulated genes ( $\log_2\text{fc} < -2$ ,  $n = 6451$ ) were strongly related to metabolic processes and mitotic cell division. These data suggest that *DAZL*<sup>+</sup> cells in the ovaroids are

downregulating their metabolism and proliferation, in agreement with the known role of DAZL in suppressing PGC proliferation.

## Methods

### *Cell culture*

Two parental hiPSC lines were used in this study: ATCC-BXS0116 female hiPSCs, which are referred to as the F3 line, and the F66 line, an in-house hiPSC line derived from the NIA Aging Cell Repository fibroblast line AG07141 using Epi5 footprint-free episomal reprogramming. The karyotypes of parental lines, as well as engineered reporter lines, were verified by Thermo Fisher Cell ID (SNP-based authentication) + Karyostat, and pluripotency was assessed by Thermo Fisher Pluritest. All lines were identified as normal.

hiPSCs were cultured in mTESR Plus medium (Stemcell Technologies) on standard polystyrene plates coated with hESC-qualified Matrigel (Corning). Medium was changed daily. Passaging was performed using 0.5 mM EDTA, or TRYPLE for experiments requiring single-cell dissociation. hiPSCs were treated with 10  $\mu$ M Y-27632 (Ambeed) for 24 hours after each passage. COV434 cells were cultured in DMEM + 10% FBS + 1X GlutaMax (Gibco). KGN cells (RIKEN, RCB1154) were cultured in DMEM/F12 + 10% FBS + 1X GlutaMax (Gibco). HGL5 cells (ABM cat. T0650) were cultured in Prigrow IV medium (ABM) with 10% FBS. Passaging was performed with TRYPLE (Gibco). hPGCLCs were cultured in S-CM medium, and passaged with Accutase (Stemcell Technologies). Mycoplasma testing was performed by PCR every 3 months; all cells tested negative.

### *Electroporations*

Electroporations were performed using a Lonza Nucleofector with 96-well shuttle, with 200,000 cells in 20  $\mu$ L of P3 buffer. Pulse setting CA-137 was used for all electroporations. Selection with the appropriate agent was begun 48 hours after electroporation and continued for 5 days. For the agents used in this study, this time was sufficient to give a high-purity final cell population.

### *Reporter construction*

Homology arms for *FOXL2* were amplified by PCR from genomic DNA. A targeting plasmid, containing an in-frame C-terminal T2A-tdTomato reporter, as well as a Rox-PGK-PuroTK-Rox selection cassette (Figure 2-Figure Supplement 1A), was constructed by Gibson assembly. The plasmid backbone additionally had an MC1-DTA marker to select against random integration. sgRNA oligos targeting the C-terminal region of *FOXL2* were cloned

into pX330 (Addgene #42230). For generation of the reporter lines, 1 µg donor plasmid and 1 µg sgRNA plasmid were co-electroporated into hiPSCs, which were subsequently plated in one well of a 6- well plate. After selection with puromycin (400 ng/mL), colonies were picked manually with a P20 pipette. The hiPSC lines generated were genotyped by PCR for the presence of wild-type and reporter alleles. Homozygous clones were further verified by PCR amplification of the entire *FOXL2* locus (Figure 2-Figure Supplement 1B) and Sanger sequencing.

To excise the selection cassette, hiPSCs were electroporated with pCAGGS-Dre (1 µg). Selection was performed with ganciclovir (4 µM) and colonies were picked as described above. The excision of the selection cassette was verified by genotyping. Primers used in this study are listed in Supplementary File 1.

#### *TF plasmid construction*

TF cDNAs were obtained from the TFome39 or the ORFeome76 as Gateway entry clones. These were cloned into a barcoded Dox-inducible expression vector (Addgene #175503) using MegaGate cloning<sup>48</sup>. The final expression constructs were verified by Sanger sequencing, which also served to determine the barcode sequences for each TF. Two unique barcodes were used per TF during library pooling. Libraries were pooled using an equimolar quantity of each plasmid (measured using QuBit).

#### *TF screening for granulosa differentiation*

A pooled library of barcoded TF plasmids was electroporated into *FOXL2*-tdTomato reporter hiPSCs, typically at 5 fmol library and 500 ng PiggyBac transposase expression plasmid (Systems Bio). These conditions were chosen to give an average copy number of approximately 5/cell (Figure 2-Figure Supplement 2). Some experiments were also performed at 50 fmol to explore the effects of higher copy numbers. For the screening data presented in Figure 2, two libraries were used: library #1, containing 35 TFs, and library #2, containing 18 TFs. Library #1 was used only at 5 fmol, whereas library #2 was used at both 5 fmol and 50 fmol.

After selection with puromycin (400 ng/mL), hiPSCs were treated with doxycycline (1 µg/mL) in mTESR Plus medium. In additional experiments, hiPSCs were first differentiated to mesoderm following a previously published protocol<sup>77</sup> before doxycycline treatment. In both sets of experiments, doxycycline treatment continued for 5 days, after which the cells were dissociated with TRYPLE and reporter-positive cells were isolated by

FACS. Genomic DNA was extracted (QIAamp DNA Micro kit) from reporter-positive and negative cells, as well as from the initial population before doxycycline treatment.

Barcodes were amplified by PCR (KAPA polymerase), using 10 ng input gDNA per reaction and typically 22 PCR cycles (95°C 15 sec. denature, 58 °C 20 sec. anneal/extend). PCR products were purified using ProNex beads, and a second round of PCR (NEB Q5 polymerase, 6 cycles of 98 °C 5 sec. denature, 61 °C 20 sec. anneal, 72 °C 5 sec. extend, final extension 72 °C 2 min) was performed to add Illumina indices. (Primers are given in Supplementary File 1). These amplicons were again purified using ProNex beads. Samples were normalized and pooled, and barcodes were sequenced on an Illumina MiSeq with 10% PhiX spike-in. To call barcodes, reads were aligned to the set of known barcode sequences. Fold-changes were calculated by comparing barcode frequencies in the sorted FOXL2+ cells to the frequencies in the starting population.

#### *Flow cytometry / Cell sorting*

Cells were dissociated by treatment with TRYPLE for 5 minutes, which was quenched with 4 volumes of ice-cold DMEM + 10% FBS. The suspension was passed through a 70 µm cell strainer. Cells were pelleted (200 g, 5 min) and resuspended in staining buffer (PBS + 3% FBS + antibodies, approx. 100 µL per million cells). Staining continued on ice in the dark for 30 minutes. The suspension was diluted with 9 volumes of PBS + 3% FBS. Cells were pelleted (200 g, 5 min) and resuspended in PBS + 3% FBS + 100 ng/mL DAPI. The suspension was kept on ice in the dark until analysis. Flow cytometry was performed on a BD LSRFortessa, and sorting was performed on a Sony SH800 with 100 µm chip.

Antibody capture beads (BD Biosciences, RRID AB\_10051478), or hiPSCs expressing tdTomato, were used as compensation controls. Antibodies used are given in the Key Resources table. Data analysis was performed using the Cytoflow Python package (version 1.0.0, [github.com/cytoflow/cytoflow](https://github.com/cytoflow/cytoflow))

#### *Protocol for granulosa differentiation*

iPSCs were dissociated with TRYPLE, and plated in DK10 medium (DMEM-F12, 15 mM HEPES, 1X GlutaMax, 10% KSR) with Y-27632 (10 µM), CHIR99021 (3 µM), and doxycycline (1 µg/mL) at a cell density of 12,500/cm<sup>2</sup> on Matrigel-coated polystyrene plates. For 24-well plates the medium volume per well was 0.5 mL; for 6-well plates it was 2 mL. 48 hours after plating, the medium was changed to DK10 + doxycycline (1 µg/mL), and the medium was subsequently changed every 24 hours. Cells were harvested on day 5 unless

otherwise indicated. In the no-TF control differentiation for RNA-seq, the protocol was the same except the cells did not contain TF expression plasmids.

*Additional Protocols for inducing granulosa-like cells*

- hiPSCs containing integrated TF expression plasmids were cultured in mTeSR™ Plus medium on Corning® Matrigel® Matrix. For induction of granulosa-like cells, hiPSCs were dissociated to single cells using TRYPLE and seeded on Corning® Matrigel® Matrix or collagen I coated plates at a density of ~10,000 – 20,000 per cm<sup>2</sup> in DK10 medium (DMEM/F12 with GlutaMAX™ Supplement and 10% Knockout Serum Replacement) plus ~3-5 µM CHIR99021, ~8-12 µM Y-27632, and ~0.5-3 µg/mL doxycycline to induce TF expression. 48 hours later, and subsequently at 24 hour intervals, the medium was changed and replaced with fresh DK10 + ~0.5-3 µg/mL doxycycline. After a total of ~120 hours, the granulosa-like cells were ready to use for downstream experiments.
- hiPSCs containing integrated TF expression plasmids were cultured in mTeSR Plus medium on Matrigel or laminin. When the cultures had reached 20–40% confluency, the medium was removed, the iPSCs were washed with phosphate-buffered saline (PBS), and the stage 1 medium, containing Advanced RPMI with GlutaMAX™ Supplement, CHIR99021 (~6-10 µM), and either dorsomorphin (~100–400 nM) or noggin (~5-8 ng/mL), was added. After 48 hours the stage 1 medium was removed and replaced with fresh stage 1 medium. After an additional 48 hours the stage 1 medium was removed and replaced with stage 2 medium, containing Advanced RPMI with GlutaMAX™ Supplement and ~0.5-3 µg/mL doxycycline. The stage 2 medium was changed every 24 hours. After ~3–4 days in stage 2 medium (~7–8 days total) the granulosa-like cells were ready to use for downstream experiments.

*RNA-seq*

Total RNA was extracted from sorted FOXL2+ granulosa-like cells using the Arcturus PicoPure kit (Thermo Fisher), or from COV434 cells and hiPSCs using the Monarch Total RNA Miniprep kit (NEB). For experiments involving TF overexpression, TF expression plasmids were integrated into hiPSCs as described above (50 fmol / 200,000 cells). After selection with puromycin, TF expression was induced using doxycycline (1000 ng/mL).

Two biological replicates were collected for each sample (iPSC, hiPSC + individual TFs, sorted FOXL2+, no-TF differentiation, KGN, COV434). Libraries were prepared using the NEBNext Ultra II Directional kit following the manufacturer's protocol, and sequenced on an Illumina NextSeq 500 (2 x 75bp paired-end reads). The TPM data shown in Figure 3 were generated using kallisto<sup>78</sup> to pseudoalign reads to the reference human transcriptome (Ensembl GRCh38 v96). Differential expression analysis was performed using DESeq2. PantherDB60 was used to calculate gene ontology enrichment for significantly upregulated ( $\log_2\text{fc} > 3$ ,  $\text{padj} < 0.05$ ) and downregulated ( $\log_2\text{fc} < -3$ ,  $\text{padj} < 0.05$ ) genes for each sample relative to hiPSCs.

#### *TROM Analysis*

The Transcriptome Overlap Measure (TROM) method was employed to identify associated genes that capture molecular characteristics of biological samples and subsequently comparing the biological samples by testing the overlap of their associated genes<sup>58</sup>. TROM scores were calculated as the  $-\log_{10}(\text{Bonferroni corrected } p \text{ value of association})$  on a scale of 0-300. The TROM magnitude is positively correlated with similarity between two independent samples, with a standard threshold of 12 as a generally-accepted indicator of significant similarity.

#### *Ovaroid formation with hPGCLCs and granulosa-like cells*

F2 female hPGCLCs (see Key Resources table) were maintained in long-term culture. Briefly, hPGCLCs were cultured on Matrigel in STO-conditioned medium (GMEM with 13% KSR and 1X NEAA, sodium pyruvate, and GlutaMax, all from Gibco), supplemented with SCF (100 ng/mL, Peprotech), ascorbic acid (50  $\mu\text{g/mL}$ , Gibco), and 2- mercaptoethanol (25  $\mu\text{M}$ , Gibco). hPGCLCs were harvested with Accutase. To form ovaroids, granulosa-like cells were harvested with TRYPLE, counted, and mixed with F2 hPGCLCs. For hormone assays in Figure 3, we used granulosa-like cells from F3/N.R1 #6, F66/N.R1.G.F #4, F66/N.R1.G #7, F66/N.R2 #1, F66/N.R2 #5, and F66/N.R2.G #3. For immunofluorescence experiments in Figures 4 and 5, we used F66/N.R1.G.F #4 and F66/N.R2 #1. For scRNA-seq in Figure 6, we used F66/N.R1.G.F #4.

For each ovaroid, 100,000 granulosa-like cells and 10,000 hPGCLCs were added to each well of a 96-well U-bottom low-bind plate (Corning #7007) in 200  $\mu\text{L}$  of GK15 medium (GMEM, 15% KSR, with 1X GlutaMax, sodium pyruvate, and non-essential amino acids) supplemented with 10 mM Y-27632, 0.1 mM 2-mercaptoethanol, 1  $\mu\text{g/mL}$  doxycycline, 100 ng/mL SCF, and 50  $\mu\text{g/mL}$  primocin. The plate was centrifuged (100 g, 2 min) and incubated (37 °C, 5% CO<sub>2</sub>) for two days. Subsequently, ovaroids were transferred to Transwells

(collagen-coated PTFE, 3  $\mu$ m pore size, 24 mm diameter, Corning #3492) for air-liquid interface culture with aMEM, 10% KSR, 55  $\mu$ M 2-mercaptoethanol, 500 ng/mL doxycycline, and 50  $\mu$ g/mL primocin. Typically, 5-6 ovaroids were cultured on each 6-well Transwell. The medium (1.5 mL) was changed every 2 days.

#### *Ovaroid formation with hPGCLCs and mouse fetal ovarian somatic cells*

Fetal ovarian somatic cells were isolated from E12.5 female embryos of CD-1 mice (Charles River) as described by Yamashiro *et al.*<sup>65</sup> For each ovaroid, 50,000 fetal ovarian somatic cells and 5,000 F2 hPGCLCs were combined. Ovaroids were cultured as described above. All mouse experiments were approved by the Harvard Medical School Institutional Animal Care and Use Committee (IACUC).

#### *Immunofluorescence*

Ovaroids were washed with PBS and fixed with 1% PFA overnight at 4 °C. After another PBS wash, ovaroids were detached from the Transwell. In preparation for cryosectioning, ovaroids were transferred to 10% sucrose in PBS. After 24 hr. at 4 °C, the 10% sucrose solution was removed and replaced with 20% sucrose in PBS. After an additional 24 hr. at 4 °C, the ovaroids were embedded in OCT compound and stored at –80 °C until sectioning.

The ovaroids were sectioned to 10  $\mu$ m using a Leica CM3050S cryostat. Sections were transferred to Superfrost Plus slides, which were washed with PBS to remove OCT compound. The slides were washed with PBST (0.1% Triton X-100 in PBS) and sections were circled with a Pap pen. Slides were blocked for 30 min. at room temp. with blocking buffer (1% bovine serum albumin and 5% normal donkey serum [Jackson ImmunoResearch, #017-000-121, lot #152961] in PBST). The blocking buffer was removed and replaced with a solution of primary antibodies in blocking buffer, and the slides were incubated overnight at 4 °C. The antibody solution was removed and the slides were washed with PBST for 3 x 5 min. The slides were incubated with secondary antibody and DAPI solution in blocking buffer for 1 hr. at room temp. in the dark, followed by two 5-minute washes with PBST and one wash with PBS. After staining, samples were mounted in Prolong Gold medium and covered with coverslips. Imaging was performed on a Leica SP5 confocal microscope. Antibodies used are given the Key Resources Table. Images were adjusted for brightness (and only for brightness) in ImageJ (version 2.9.0/1.53t), and cell counts for Figure 4C were performed manually by a researcher who was blinded to the species of the ovaroids.

### *Single Cell RNA Sequencing*

Ovaroids (6 ovaroids per sample, 2 samples per time point) were dissociated using the Miltenyi Embryoid Body Dissociation Kit (Miltenyi #130-096-348). The cells were passed through a 40  $\mu$ m strainer, fixed using the Parse Biosciences fixation kit, and stored at  $-80^{\circ}\text{C}$  until all time points had been collected. Libraries were prepared using the Parse Biosciences WT Mega v1 kit generating libraries of an average of 450bp. The ovaroids took up 8 of the 96 samples; the remaining kit capacity was used for other experiments. The libraries were sequenced on an Illumina NovaSeq 2x150bp S4 flow cell using single index, 6bp, libraries and a 5% PhiX spike-in. Data were demultiplexed into library fastq files and counts matrices were generated using Parse Bioscience's analysis pipeline (version 0.9.6). Downstream data processing, such as doublet filtering, dimensionality reduction, and clustering, was performed using Scanpy (version 1.8.2)<sup>70</sup>. For cell type assignment, the fetal ovarian dataset from the human reproductive cell atlas<sup>69</sup> was used as a reference for scanpy ingest.

### *Collection of primary mouse ovarian somatic cells*

Female BALB/c mice (age 10 – 12 weeks) were confirmed to be in proestrus by visual examination. Mice were killed by CO<sub>2</sub> exposure followed by cervical dislocation, and ovaries were removed by dissection. Ovaries were placed in HEPES-buffered DMEM/F12 with 0.1% bovine serum albumin (2 ovaries per 1.5 mL tube, with 500  $\mu$ L medium) and mechanically disrupted by stabbing with forceps. The cell suspensions were strained through a 40  $\mu$ m strainer to remove oocytes and clumps prior to culture for hormone assays.

### *Steroid hormone assays*

Androstenedione (500 ng/mL) was added to the medium on day 4 of granulosa differentiation. FSH (0.25 IU/mL, BioVision #4781-50 lot 5F07L47810) or forskolin (100  $\mu$ M, Sigma-Aldrich) were also added as indicated. The total medium volume was 0.5 mL per well of 24-well plate. We performed these assays on each of the lines listed in Figure 3-Source Data 2. For controls using human cell lines (COV434, KGN, or HGL5) or mouse primary ovarian somatic cells, 75,000 cells were seeded per well. After 24 hours, the medium was analyzed for estradiol content by ELISA (DRG International, EIA-2693). Concentrations were calculated with a 4- parameter logistic curve fit using the data from the standards provided in the kit. Samples outside the range of the calibration curve were diluted and re-run.

For measuring hormone production in ovarioids, ovarioids were aggregated as described above. Androstenedione (500 ng/mL) and/or FSH (0.25 IU/mL) were added to the aggregation medium (total volume 200  $\mu$ L per ovarioid). After 3 days of culture, the medium was removed and analyzed by ELISA for estradiol (DRG International, EIA-2693) and progesterone (DRG International, EIA-1561). Hormone concentrations were calculated as described above.

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, unless clearly indicated to the contrary, should be understood to mean “at least one.”

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.

The terms “about” and “substantially” preceding a numerical value mean  $\pm 10\%$  of the recited numerical value.

Where a range of values is provided, each value between and including the upper and lower ends of the range are specifically contemplated and described herein.

## CLAIMS

What is claimed is:

1. A pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein.
2. The PSC of claim 1, wherein the PSC comprises the engineered polynucleotide comprising an open reading frame encoding NR5A1.
3. The PSC of claim 1, wherein the PSC comprises the engineered polynucleotide comprising an open reading frame encoding a RUNX family protein.
4. The PSC of claim 3, wherein the RUNX family protein is RUNX1.
5. The PSC of claim 3, wherein the RUNX family protein is RUNX2.
6. The PSC of claim 1, wherein the PSC expresses or overexpresses: NR5A1; RUNX1; RUNX2; NR5A1 and RUNX1; NR5A1 and RUNX2; or NR5A1, RUNX1, and RUNX2.
7. The PSC of claim 1, wherein PSC further comprises an engineered polynucleotide comprising an open reading frame encoding a TCF21 protein.
8. The PSC of claim 7, wherein the PSC expresses or overexpresses TCF21.
9. The PSC of claim 1, wherein PSC further comprises an engineered polynucleotide comprising an open reading frame encoding a GATA4 protein.
10. The PSC of claim 9, wherein the PSC expresses or overexpresses GATA4.
11. The PSC of claim 1, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.
12. The PSC of claim 11, wherein the heterologous promoter is an inducible promoter.
13. A pluripotent stem cell (PSC) comprising: a protein selected from NR5A1 and a RUNX family protein, wherein the protein is overexpressed.
14. The PSC of claim 13, wherein the PSC expresses or overexpresses: NR5A1; RUNX1; RUNX2; NR5A1 and RUNX1; NR5A1 and RUNX2; or NR5A1, RUNX1, and RUNX2.
15. The PSC of claim 14, wherein the PSC further comprises a TCF21 protein.
16. The PSC of claim 15, wherein the PSC expresses or overexpresses TCF21.
17. The PSC of claim 14, wherein the PSC further comprises a GATA4 protein.
18. The PSC of claim 15, wherein the PSC expresses or overexpresses GATA4.

19. The PSC of claim 13, wherein the PSC is a human PSC.
20. The PSC of claim 13, wherein the PSC is an induced PSC (iPSC).
21. The PSC of claim 13, wherein the PSC comprises 1-20, optionally 8-10, copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from NR5A1 and a RUNX family protein (e.g., RUNX1 and/or RUNX2).
22. A composition comprising: a population of the PSC of claim 1.
23. The composition of claim 22, wherein the population comprises at least 10,000/cm<sup>2</sup> of the PSC.
24. A method, comprising: culturing, in culture media, a population of pluripotent stem cells (PSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from NR5A1 and a RUNX family protein to produce granulosa-like cells.
25. The method of claim 24, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NR5A1.
26. The method of claim 24, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding a RUNX family protein.
27. The method of claim 26, wherein the RUNX family protein is RUNX1.
28. The method of claim 26, wherein the RUNX family protein is RUNX2.
29. The method of claim 25, wherein the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding a TCF21 protein.
30. The method of claim 25, wherein the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding a GATA4 protein.
31. The method of claim 25, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.
32. The method of claim 31 wherein the heterologous promoter is an inducible promoter.
33. The method of claim 24, wherein the population comprises 1x10<sup>2</sup>-1x10<sup>7</sup> PSCs.
34. The method of claim 24, wherein the population of PSCs is cultured for about 4-10 days.

35. The method of claim 29, wherein the population of PSCs is cultured for about 6 days.

36. The method of claim 24 wherein the granulosa-like cells are AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and/or EPCAM<sup>-</sup>.

37. The method of claim 36, wherein the granulosa-like cells are AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and EPCAM<sup>-</sup>.

38. A method comprising:

(a) delivering to pluripotent stem cells (PSCs) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a protein selected from NR5A1 and a RUNX family protein;

(b) culturing the PSCs in feeder-free, serum-free culture media to produce an expanded population of PSCs; and

(c) culturing PSCs of the expanded population in a series of induction media comprising an inducing agent to produce AMHR2<sup>+</sup>, CD82<sup>+</sup>, FOXL2<sup>+</sup>, and/or EPCAM<sup>-</sup> granulosa-like cells.

39. The method of claim 38 comprising delivering to PSCs (i) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding NR5A1 and (i) an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a RUNX family protein.

40. The method of claim 38, wherein the RUNX family protein is RUNX1.

41. The method of claim 38, wherein the RUNX family protein is RUNX2.

42. The method of claim 38, wherein the engineered polynucleotide is a transposon and the delivering further comprises delivering a transposase to the PSCs.

43. The method of claim 38, wherein the inducible promoter is a chemically-inducible promoter, optionally a doxycycline-inducible promoter.

44. The method of claim 38, wherein the feeder-free, serum-free culture media of (b) comprises a solubilized basement membrane preparation extracted from the Engelbreth-Holm-Swarm (EHS) mouse sarcoma.

45. The method of claim 44, wherein the solubilized basement membrane preparation comprises extracellular matrix (ECM) proteins and growth factors.

46. The method of claim 45, wherein the ECM proteins are selected from Laminin, Collagen IV, heparan sulfate proteoglycans, and entactin/nidogen.

47. The method of claim 38, wherein the feeder-free, serum-free culture media of (b) comprises growth factors selected from recombinant human basic fibroblast growth factor (rh bFGF) and recombinant human transforming growth factor  $\beta$  (rh TGF $\beta$ ).

48. The method of claim 38, wherein the culturing of (b) is for about 6-24 hours.

49. The method of claim 38, wherein the PSCs of the expanded population of (c) are cultured at a density of about 10,000 cells/cm<sup>2</sup> to about 20,000 cells/cm<sup>2</sup>.

50. The method of claim 38, wherein the culturing of (c) comprises culturing the PSCs in a first induction media and culturing the PSCs in a second induction media.

51. The method of claim 40, wherein the first induction media comprises one or more of L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Dulbecco's Modified Eagle Medium (DMEM)/F-12, Advanced RPMI (Roswell Park Memorial Institute) 1640 Medium, a glycogen synthase kinase (GSK) 3 inhibitor, a small molecule or protein inhibitor of the BMP signaling pathway, a small molecule ROCK inhibitor, and an inducing agent (e.g., doxycycline).

52. The method of claim 50, wherein the culturing the PSCs in a first induction media is for about 36 to about 60 hours, optionally about 48 hours.

53. The method of claim 50, wherein the second induction media comprises one or more of L-alanyl-L-glutamine, antibiotic (e.g., penicillin and/or streptomycin), Advanced RPMI 1640 Medium, DMEM/F-12, and an inducing agent (e.g., doxycycline).

54. The method of claim 50, wherein the culturing the PSCs in a second induction media is for about 96 to about 144 hours, optionally about 120 hours.

55. The method of claim 54, wherein the second induction media is removed and replaced with fresh second induction media at about 24-hour intervals.

56. A granulosa-like cell produced by the method of claim 38.

57. An ovarian organoid comprising granulosa-like cells of claim 56 and human primordial germ cell-like cells (hPGCLCS).

58. The method of claim 38 further comprising combining the granulosa-like cells with hPGCLCS to form an ovarian organoid.