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

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

(43) **International Publication Date**
02 May 2024 (02.05.2024)

WIPO | PCT

(10) **International Publication Number**
WO 2024/091801 A2

(51) **International Patent Classification:**
Not classified

(21) **International Application Number:**
PCT/US2023/076715

(22) **International Filing Date:**
12 October 2023 (12.10.2023)

(25) **Filing Language:** English

(26) **Publication Language:** English

(30) **Priority Data:**
63/415,729 13 October 2022 (13.10.2022) US

(71) **Applicants: PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US). **THE REGENTS OF THE UNIVERSITY OF CALIFORNIA** [US/US]; 1111 Franklin Street, 5th Floor, Oakland, California 94607-5200 (US).

(72) **Inventors: APPLETON, Evan**; c/o Harvard University, Smith Center - Suite 727, 1350 Massachusetts Avenue, Cambridge, Massachusetts 02138 (US). **CHURCH, George M.**; c/o Harvard University, Smith Center, Suite 727, 1350 Massachusetts Avenue, Cambridge, Massachusetts 02138 (US). **TAO, Jenhan**; 151 Tremont Street, Suite 11S, Boston, Massachusetts 02111 (US).

(74) **Agent: CHANCELLOR, K. Blake** et al.; Wolf, Greenfield & Sacks, P.C., 600 Atlantic Avenue, Boston, Massachusetts 02210-2206 (US).

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

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

Published:
— *without international search report and to be republished upon receipt of that report (Rule 48.2(g))*
— *with sequence listing part of description (Rule 5.2(a))*

(54) **Title:** METHODS AND COMPOSITIONS FOR INDUCING CELL DIFFERENTIATION

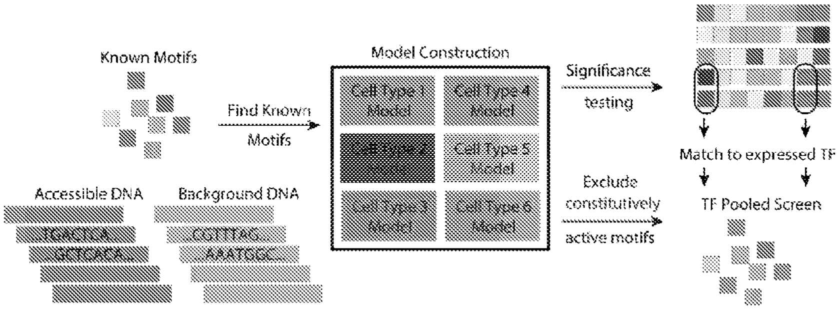


FIG. 1C

(57) **Abstract:** Provided herein are methods and compositions for differentiating induced pluripotent stem cells into one or more cell types by overexpressing one or more transcription factors.

METHODS AND COMPOSITIONS FOR INDUCING CELL DIFFERENTIATION

CROSS-REFERENCE TO RELATED APPLICATIONS

This application claims the benefit under 35 U.S.C. § 119(e) of U.S. Provisional Application No. 63/415,729, filed October 13, 2022, which is hereby incorporated by reference in its entirety.

GOVERNMENT LICENSE RIGHTS

This invention was made with government support under W911NF-17-2-0089 awarded by U.S. Army Research Office and under DK091183 awarded by National Institutes of Health. The government has certain rights in the invention.

REFERENCE TO AN ELECTRONIC SEQUENCE LISTING

The contents of the electronic sequence listing (H049870763WO00-SEQ-KVC.xml; Size: 31,902 bytes; and Date of Creation: October 10, 2023) is herein incorporated by reference in its entirety.

BACKGROUND

Stem cell and cell-fate engineering are promising research areas that may accelerate the development of therapeutics for a number of diseases. However, current approaches to stem cell and cell-fate engineering are laborious and costly. Efficient methods for differentiating stem cells into specific cell types that may be used for therapeutic disease intervention remain elusive.

SUMMARY

The present disclosure relates, at least in part, to methods and compositions for generating astrocyte-like cells (iAstIIs), cytotoxic T-cell-like cells (iCytoTs), hepatocyte-like cells (iHeps), regulatory T-cell-like cells (iTRegs), B cell-like cells (iBCells), and/or microglia-like cells (iMicroglia) from pluripotent stem cells. The present disclosure also relates, at least in part, for identifying transcription factors that improve differentiation efficiency of pluripotent stem cells into astrocyte-like cells, cytotoxic T-cell-like cells, hepatocyte-like cells, regulatory T-cell-like cells, B cell-like cells, and/or microglia-like cells.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding ERG, EGR1, FLI1, FOSB, or any combination thereof. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ERG. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding EGR1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding FLI1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding FOSB.

In some embodiments, the PSC expresses or overexpresses ERG, EGR1, FLI1, FOSB, or any combination thereof.

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.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: a protein selected from ERG, EGR1, FLI1, and FOSB, wherein the protein is overexpressed. In some embodiments, the PSC expresses or overexpresses: ERG, EGR1, FLI1, FOSB, or any combination thereof. 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 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ERG, EGR1, FLI1, and FOSB.

Aspects of the present disclosure relate to a composition comprising: a population of any one of the PSCs described herein. In some embodiments, the population comprises at least 2500/cm2 of the PSC.

Aspects of the present disclosure relate to 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 ERG, EGR1, FLI1, and FOSB to produce astrocyte-like cells. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ERG. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding EGR1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding FLI1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding FOSB.

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

In some embodiments, the population comprises $1x10^2$ -$1x10^7$ PSCs. In some embodiments, the population of PSCs is cultured for at least 1 day. In some embodiments, the population of PSCs is cultured for about 3-6 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the astrocyte-like cells are CD44+ and A2B5+.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ZBTB1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RUNX3. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RELA. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding NRF1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ERF. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding SP4. In some embodiments, the PSC expresses or overexpresses ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

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.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: a protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4, wherein the protein is overexpressed. In some embodiments, the PSC expresses or overexpresses: ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof. 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 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises 8-

10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4.

Aspects of the present disclosure relate to a composition comprising: a population of any one of the PSCs described herein. In some embodiments, the population comprises at least 2500/cm2 of the PSC.

Aspects of the present disclosure relate to 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 ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 to produce cytotoxic T-cell-like cells. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ZBTB1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RUNX3. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ERF. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding SP4.

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

In some embodiments, the population comprises $1\text{x}10^2$ -$1\text{x}10^7$ PSCs. In some embodiments, the population of PSCs is cultured for at least 1 day. In some embodiments, the population of PSCs is cultured for about 3-6 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the cytotoxic T-cell-like cells are CD3+ and CD8+.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding HNF4G, TEAD4, RFX3, or any combination thereof. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding HNF4G. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding TEAD4. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RFX3. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding HNF4A. In some embodiments, the PSC expresses or overexpresses HNF4G, TEAD4, RFX3, or any combination thereof. In some embodiments, the PSC further expresses or overexpresses HNF4A.

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.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: a protein selected from HNF4G, TEAD4, and RFX3, wherein the protein is overexpressed. In some embodiments, the PSC expresses or overexpresses: HNF4G, TEAD4, RFX3, or any combination thereof. 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 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from HNF4G, HNF4A, TEAD4, and RFX3. In some embodiments, the PSC comprises 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from HNF4G, TEAD4, and RFX3.

Aspects of the present disclosure relate to a composition comprising: a population comprising any one of the PSCs described herein. In some embodiments, the population comprises at least 2500/cm2 of the PSC.

Aspects of the present disclosure relate to 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 HNF4G, TEAD4, and RFX3 to produce hepatocyte-like cells. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding HNF4G. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding TEAD4. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RFX3. In some embodiments, the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding HNF4A.

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

In some embodiments, the population comprises $1x10^2$ -$1x10^7$ PSCs. In some embodiments, the population of PSCs is cultured for at least 1 day. In some embodiments, the population of PSCs is cultured for about 3-6 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the hepatocyte-like cells are CD184+ and ASGPR1+.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof. In some embodiments, the engineered polynucleotide comprises comprising an open reading frame encoding ETS1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ETV3. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding GABPA. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding KLF9. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding NFKB1. In some embodiments, the PSC expresses or overexpresses ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof.

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.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: a protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, wherein the protein is overexpressed. In some embodiments, the PSC expresses or overexpresses: ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof. 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 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1.

Aspects of the present disclosure relate to a composition comprising: a population of any one of the PSCs described herein. In some embodiments, the population comprises at least 2500/cm2 of the PSC.

Aspects of the present disclosure relate to 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 ETS1, ETV3, GABPA, KLF9, and NFKB1 to produce regulatory T-cell-like cells. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ETS1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ETV3. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding GABPA. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding KLF9. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NFKB1.

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

In some embodiments, the population comprises $1x10^2$ -$1x10^7$ PSCs. In some embodiments, the population of PSCs is cultured for at least 1 day. In some embodiments, the population of PSCs is cultured for about 3-6 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the regulatory T-cell-like cells are CD3+ and CD25+.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding EBF1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ZBTB1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RELA. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding NRF1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding REL. In some embodiments, the PSC expresses or overexpresses EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof.

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.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: a protein selected from EBF1, ZBTB1, RELA, NRF1, and REL, wherein the protein is overexpressed. In some embodiments, the PSC expresses or overexpresses: EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof. 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 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from EBF1, ZBTB1, RELA, NRF1, and REL. In some embodiments, the PSC comprises 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from EBF1, ZBTB1, RELA, NRF1, and REL.

Aspects of the present disclosure relate to a composition comprising: a population of any one of the PSCs described herein. In some embodiments, the population comprises at least 2500/cm2 of the PSC.

Aspects of the present disclosure relate to 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 EBF1, ZBTB1, RELA, NRF1, and REL to produce B cell-like cells. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding EBF1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ZBTB1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding REL.

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

In some embodiments, the population comprises 1x10^2 -1x10^7 PSCs. In some embodiments, the population of PSCs is cultured for at least 1 day. In some embodiments, the population of PSCs is cultured for about 3-6 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the B cell-like cells are CD19+ and CD27+.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: an engineered polynucleotide comprising an open reading frame encoding SPI1, ZBTB1, RELA, STAT2, or any combination thereof. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding SPI1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding ZBTB1. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding RELA. In some embodiments, the engineered polynucleotide comprises an open reading frame encoding STAT2.

In some embodiments, the PSC expresses or overexpresses SPI1, ZBTB1, RELA, STAT2, or any combination thereof.

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.

Aspects of the present disclosure relate to a pluripotent stem cell (PSC) comprising: a protein selected from SPI1, ZBTB1, RELA, and STAT2, wherein the protein is overexpressed.

In some embodiments, the PSC expresses or overexpresses: SPI1, ZBTB1, RELA, STAT2, or any combination thereof. 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 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from SPI1, ZBTB1, RELA, and STAT2. In some embodiments, the PSC comprises 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from SPI1, ZBTB1, RELA, and STAT2.

Aspects of the present disclosure relate to a composition comprising: a population comprising any one of the PSCs described herein. In some embodiments, the population comprises at least 2500/cm2 of the PSC.

Aspects of the present disclosure relate to 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 SPI1, ZBTB1, RELA, and STAT2 to produce microglia-like cells. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding SPI1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ZBTB1. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding STAT2.

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

In some embodiments, the population comprises $1\text{x}10^2$ -$1\text{x}10^7$ PSCs. In some embodiments, the population of PSCs is cultured for at least 1 day. In some embodiments, the population of PSCs is cultured for about 3-6 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the microglia-like cells are CD11b+ and CX3CR1+.

Aspects of the present disclosure relate to a method, comprising: (i) analyzing epigenetics data for a target cell type to identify genomic sites that are available for binding of a transcription factor and generating a first pool of transcription factors; (ii) analyzing transcriptomic data for the target cell type to identify expression levels of the transcription factors associated with the genomic sites that are available for binding identified in step (i) and generating a second pool of transcription factors; (iii) using a first statistical method to filter background data and identify transcription factors that are present in the first pool of transcription factors and the second pool of transcription factors and generating a third pool of transcription factors, wherein the third pool of transcription factors comprises transcription factors that are in both the first pool and the second pool; (iv) using a second statistical method to determine the statistical significance of the transcription factors in the third pool of transcription factors; and (v) repeating steps (i)-(iv) one or more times to iteratively refine the third pool of transcription factors.

In some embodiments, the epigenetics data provides information related to whether genomic chromatin is open or closed. In some embodiments, the epigenetics data is produced by DNAse-seq, ATAC-seq, or ChIP-seq.

In some embodiments, the transcriptomic data provides information related to whether there are more transcripts of the transcription factor in the target cell type than in a non-target cell type. In some embodiments, the transcriptomic data is produced by RNA-seq.

In some embodiments, the first statistical method is linear regression algorithm. In some embodiments, the first statistical method is a logistic regression algorithm. In some embodiments, the first statistical method is a L1-regularized logistic regression model (LASSO).

In some embodiments, the background data is associated with transcription factors that are not expressed in the target cell type at a higher expression level than in the non-target cell type.

In some embodiments, the second statistical method is a log-likelihood ratio test.

In some embodiments, the method further comprises transfecting transcription factors of the third pool into a stem cell. In some embodiments, the method further comprises inducing differentiation of the stem cell into the target cell type. In some embodiments, the method further comprises analyzing the target cell type to identify additional transcription factors associated with the target cell type. In some embodiments, the method further comprises using data from the target cell type to further refine the previous steps.

In some embodiments, the target cell type is an astrocyte, a cytotoxic T-cell, a hepatocyte, a regulatory T-cell, a B cell, or a microglial cell. In some embodiments, differentiation of stem cells using one or more of the transcription factors in the third pool results in production of the target cell type in no more than 6 days.

Aspects of the present disclosure relate to a method for generating a transcription factor screening pool comprising: using at least one computer hardware processer to perform: accessing at least one statistical model relating one or more input transcription factors to differentiation efficiency of a cell having the one or more input transcription factors; obtaining differentiation efficiency information for the one or more input transcription factors; generating, using the at least one statistical model and the differentiation efficiency information, a transcription factor pool having transcription factors that are predicted to differentiate the cell into a target cell type in accordance with the differentiation efficiency information.

In some embodiments, the at least one statistical model correlates chromatin accessibility data and transcriptomics data to make initial predictions relating the one or more input transcription factors to differentiation efficiency of the cell having the one or more input transcription factors. In some embodiments, the at least one statistical model distinguishes open chromatin data from background data.

In some embodiments, the open chromatin data is associated with the target cell type.

In some embodiments, the method further comprises identifying an initial set of transcription factor motifs positively correlated with the open chromatin data by using a statistical coefficient trained to distinguish the open chromatin data from the background data.

In some embodiments, the differentiation efficiency information corresponds to a mode of a distribution of differentiation efficiency data used to train the at least one statistical model.

In some embodiments, wherein the at least one statistical model was trained using measured differentiation efficiency values having a multimodal distribution with modes, and the differentiation efficiency information corresponds to a mode of the multimodal distribution with the highest value.

In some embodiments, the transcription factors of the transcription factor pool have predicted differentiation efficiency within a distribution centered at the mode of the multimodal distribution with the highest value. In some embodiments, the differentiation efficiency information corresponds to a Gaussian distribution centered at a mode of a distribution for differentiation efficiency data used to train the at least one statistical model. In some embodiments, the differentiation efficiency information corresponds to a high differentiation efficiency component of a distribution of differentiation efficiency values for transcription factors. In some embodiments, generating the transcription factor pool further comprises: generating an initial pool of transcription factors; using transcription factors in the initial pool as input to the at least one statistical model to obtain values for differentiation efficiency; selecting, based on the values for differentiation efficiency and the differentiation efficiency information, one or more of the transcription factors in the initial pool to include in the transcription factor pool.

In some embodiments, the at least one statistical model comprises at least one regression model. In some embodiments, the at least one statistical model comprises at least one neural network. In some embodiments, the at least one statistical model has a recurrent neural network architecture. In some embodiments, the at least one statistical model comprises a L1-regularized logistic regression model (LASSO). In some embodiments, the at least one statistical model comprises a log-likelihood ratio test.

Aspects of the present disclosure relate to a system comprising: at least one hardware processor; and at least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one hardware processor, cause the at

least one hardware processor to perform: accessing at least one statistical model relating one or more input transcription factors to differentiation efficiency of a cell having the one or more input transcription factors; obtaining differentiation efficiency information for transcription factors, wherein the differentiation efficiency information corresponds to a mode of a distribution for differentiation efficiency data used to train the at least one statistical model; and

generating, using the at least one statistical model and the differentiation efficiency information, a transcription factor pool having transcription factors with predicted differentiation efficiency in accordance with the differentiation efficiency information. In some embodiments, the target cell type is a Type II astrocyte, cytotoxic T-cell, regulatory T-cell, hepatocyte, B cell, or microglial cell.

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.

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:

FIGs. 1A-1E show an overview of the machine-guided experimental workflow with the human TFome to identify and optimize transcription factor (TF) conversion combinations. (**FIG. 1A**) *CellCartographer* considers only TFs with known binding motifs that are found to be highly specific to target cell identity. (**FIG. 1B**) The *CellCartographer* workflow uses epigenetics and transcriptomics NGS data to determine TF pools for screening with the TFome (dashed box). Iterative rounds of screening are refined with ML and engineered induced pluripotent stem cell (iPSC) lines with sufficient differentiation undergo clonal isolation to isolate high-efficiency clones. (**FIG. 1C**) TF-binding motifs and chromatin accessibility data are used to train a classifier model to determine TFs that are associated with the cell types of interest and then filtered with RNAseq data to get a finalized sub-library of TFs for pooled screening. (**FIG. 1D**) iPSCs are nucleofected with TF-cassette pools that are integrated randomly into the genome where any one cell may receive some combination of these factors in either multiple copies or not at all. (**FIG. 1E**) *In silico* validation of screening lists — for four cell types with previously validated TF-overexpression differentiation factors, our model accurately re-identifies these factors (shaded) in the top TFs that would be put into a screen.

FIGs. 2A-2D show computational analysis of 34 cell types with *CellCartographer*. (**FIG. 2A**) Multidimensional scaling of the similarity in gene expression between different cell types. (**FIG. 2B**) Multidimensional scaling of the similarity in TFs correlated with open chromatin. (**FIG. 2C**) Motifs correlated and anti-correlated with open chromatin vary across 34 cell types analyzed. (**FIG. 2D**) Highly ranked motifs correlated with open chromatin for cell types derived from yolk sac (microglia), endoderm (hepatocyte), mesoderm (B cell, T-cell, regulatory T-cell), and ectoderm (astrocyte).

FIGs. 3A-3F show primary pooled screens for cell types originating from each germ layer. For each cell type, a negative antibody stain for iPSCs without TFs (LEFT), the cell population with induced TFs (MIDDLE), and the barcoded TF appearance frequency in the transcriptome of double-positive cell populations (RIGHT) is shown. (**FIG. 3A**) Type II Astrocytes (ectoderm). (**FIG. 3B**) Microglia (yolk sac). (**FIG. 3C**) CD8-positive T-cells (mesoderm). (**FIG. 3D**) B cells (mesoderm). (**FIG. 3E**) Regulatory T-cells (mesoderm). (**FIG. 3F**). Hepatocytes (endoderm).

FIGs. 4A-4H. show iteratively engineered poly-clonal and mono-clonal cell lines. (**FIGs. 4A-4D**) For each cell type, we show percent double-positive for fluorescence-activated cell sorting (FACS) analysis of canonical markers for non-clonal and mono-clonal (dashed box) cell lines, and an iPSC + media control (solid box) differentiated for six days in cell-type-specific media + DOX for (**FIG. 4A**) Type II Astrocytes (iAstIIs) (**FIG. 4B**) Cytotoxic T-cells (iCytoTs) (**FIG. 4C**) Hepatocytes (iHeps) and (**FIG. 4D**) Regulatory T-cells (iTRegs). (**FIG. 4E**) Differential gene expression (quantified by Z-score) for all genes for two replicates of each differentiated cell type in both media conditions. (**FIG. 4F**) Principal component analysis of all genes for each cell type in each media condition and a primary cell control. (**FIG. 4G**) Differential gene expression (quantified by Reads Per Kilobase Million (RPKM)) for key marker genes across target cell types and iPSCs. (**FIG.**

4H) Metascape analysis of gene enrichment of high-efficiency clones for genes that were upregulated in these lines and differentiation conditions compared to iPSCs. Analysis of select highly significant GO Terms from TOP 50 for each differentiated cell type and condition is shown (-log10(P) ≥ 3).

FIGs. 5A-5L show functional validation of iAstIIs, iHeps, iCytoTs, and iTRegs. (**FIGs. 5A-5C**) Stimulation of Type II astrocytes over 10 min with small molecules with (**FIG. 5A**) 100μM ATP, (**FIG. 5B**) 100μM glutamate, and (**FIG. 5C**) 30mM KCl. (LEFT) Relative fluorescence of six individual astrocytes. Astrocyte cell population shown before (TOP) and after (BOTTOM) addition of small molecule. (**FIG. 5D**) BF image of induced iHeps prior to hepatotoxicity testing. iHeps, primary hepatocytes and iPCs titrated with (**FIG. 5E**) Nefazodone, (**FIG. 5F**) Acetaminophen, and (**FIG. 5G**) Troglitazone for 24h and assayed for percent viability (survival rate normalized to each cell type without toxins applied). (**FIG. 5H**) Brightfield imaging of T-cell populations (LEFT to RIGHT): Primary CD8 T-cells, iTRegs, Primary CD8 T-cells + activation beads, iCytoTs + activation beads. (**FIG. 5I**) Suppression assay for iTRegs co-cultured with activated primary CD8 T-cells. (**FIG. 5J**) Calculated percent suppression with titrated dosing of iTRegs in suppression assay. Primary T-cells have been shown to suppress in the range of 20/30/40% respectively. (**FIG. 5K**) Activation assay for iCytoTs. (**FIG. 5L**) Percent of proliferating primary CD8 T-cells and iCD8 cells post-activation.

FIGs. 6A-6B show cell line differentiation for double-positive surface markers after 6 days of differentiation in cell-type-specific growth medium with doxycycline (DOX). For B cells (**FIG. 6A)** and microglia (**FIG. 6B)**, percent double-positive for FACS analysis of canonical markers for iPSC only (solid box), non-clonal and mono-clonal (dashed box) cell lines is shown.

FIGs. 7A-7F show cell line differentiation for either one or both cell surface markers after 6 days of differentiation in cell-type-specific growth medium with DOX. For Type II astrocytes (**FIG. 7A**), CD8-positive T-cells (**FIG. 7B**), microglia (**FIG. 7C**), regulatory T-cells (**FIG. 7D**), hepatocytes (**FIG. 7E**), and B cells (mesoderm) (**FIG. 7F**), percent differentiated for FACS analysis of canonical markers for non-clonal, mono-clonal (dashed box) cell lines, and iPSC+ media control (solid box) is shown.

DETAILED DESCRIPTION

Cell types generated by differentiation of stem cells have the potential to accelerate therapeutic discoveries for a variety of diseases. The present disclosure relates, at least in part, to methods and compositions for generating astrocyte-like cells (iAstIIs), cytotoxic T-cell-like cells (iCytoTs), hepatocyte-like cells (iHeps), regulatory T-cell-like cells (iTRegs), B cell-like cells (iBCells), and/or microglia-like cells (iMicroglia) from pluripotent stem cells (e.g., induced pluripotent stem cells (iPSCs)). The present disclosure also relates, at least in part, to methods for identifying transcription factors that improve differentiation efficiency of pluripotent stem cells into astrocyte-like cells, cytotoxic T-cell-like cells, hepatocyte-like cells, regulatory T-cell-like cells, B cell-like cells, and/or microglia-like cells.

Astrocyte-Like Cells

Some aspects of the present disclosure provide astrocyte-like cells and methods of producing such cells. Astrocytes are specialized glial cells in the brain that regulate neuronal synapses and play an important role in the neuroimmune system. An astrocyte-like cell is a cell that exhibits phenotypic characteristics of astrocytes. For example, an astrocyte-like cell may express one or more biomarkers expressed by an astrocyte or exhibit one or more functions exhibited by an astrocyte. Astrocytes are broadly classified as Type I astrocytes and Type II astrocytes. In some embodiments, the astrocyte-like cell of the present disclosure are Type I astrocytes or exhibit phenotypic characteristics of Type I astrocytes. Phenotypic characteristics of Type I astrocytes include, for example, a protoplasmic presentation with short astrocytic processes. In some embodiments, the astrocyte-like cell of the present disclosure are Type II astrocytes or exhibit phenotypic characteristics of Type II astrocytes. Phenotypic characteristics of Type II astrocytes include, for example, a fibrous presentation with long astrocytic processes.

Native astrocytes typically express the gene Cluster of Differentiation 44 (CD44) and the gene A2B5, two putative marker genes for astrocytes. In some embodiments, an astrocyte-like cell expresses CD44 (i.e., is CD44-positive (CD44+)). In some embodiments, an astrocyte-like cell expresses A2B5 (i.e., is A2B5+). Thus, in some embodiments, the astrocyte-like cells produced by the methods provided herein are CD44+/A2B5+ astrocyte-like cells. (i.e., cells that express the CD44 protein and the A2B5 protein).

Other biomarkers of astrocyte identity include high levels of glial fibrillary acidic protein (GFAP) (e.g., ≥80% of cells of an iPSC-derived population express GFAP) and low levels of neuronal class III beta-tubulin (TUJ1) (e.g., ≤15% of cells of an iPSC-derived population express TUJ1). Additional biomarkers of astrocyte identity include high levels of expression of aquaporin-4 (AQP4) (see, e.g., Jurga AM , Paleczna M, Kadluczka J, Kuter

KZ. Beyond the GFAP-Astrocyte Protein Markers in the Brain. *Biomolecules*. 2021; 11(9):1361.).

Cytotoxic T-cell-Like Cells

Some aspects of the present disclosure provide cytotoxic T-cell-like cells and methods of producing such cells. Cytotoxic T cells are a type of immune cell associated with the innate immune system. Cytotoxic T-cells are T lymphocytes that kill foreign cells and pathogens in the body. A cytotoxic T-cell-like cell is a cell that exhibits phenotypic characteristics of cytotoxic T-cells. For example, a cytotoxic T-cell-like cell may express one or more biomarkers expressed by a cytotoxic T-cell or exhibit one or more functions exhibited by a cytotoxic T-cell.

Cytotoxic T-cells that development in the body typically express the gene Cluster of Differentiation 3 (CD3) and the gene Cluster of Differentiation 8 (CD8), two putative marker genes for cytotoxic T-cells. In some embodiments, a cytotoxic T-cell-like cell expresses CD3 (i.e., is CD3-positive (CD3+)). In some embodiments, a cytotoxic T-cell-like cell expresses CD8 (i.e., is CD8+). Thus, in some embodiments, the cytotoxic T-cell-like cells produced herein are CD3+/CD8+ cytotoxic T-cell-like cells (i.e., cells that express the CD3 protein and the CD8 protein).

Hepatocyte-Like Cells

Some aspects of the present disclosure provide hepatocyte-like cells and methods of producing such cells. Hepatocytes are specialized epithelial cells in the liver that play an important role in metabolism, detoxification, and protein synthesis. Hepatocytes also participate in the innate immune response by secreting immune proteins in response to invading cells, pathogens, and microorganisms. A hepatocyte-like cell is a cell that exhibits phenotypic characteristics of hepatocytes. For example, a hepatocyte-like cell may express one or more biomarkers expressed by a hepatocyte or exhibit one or more functions exhibited by a hepatocyte.

Hepatocytes produced in the body express the gene Cluster of Differentiation 184 (CD184) and the gene Asialoglycoprotein Receptor 1 (ASGPR1), two putative marker genes for hepatocytes. In some embodiments, a hepatocyte-like cell expresses CD184 (i.e., is CD184-positive (CD184+)). In some embodiments, a hepatocyte-like cell expresses ASGPR1 (i.e., is ASGPR1+). Thus, in some embodiments, the hepatocyte-like cells produced by the methods provided herein are CD184+/ASGPR1+ hepatocyte-like cells (i.e., cells that express the CD184 protein and the ASGPR1 protein).

Regulatory T-cell-Like Cells

Some aspects of the present disclosure provide regulatory T-cell-like cells and methods of producing such cells. Regulatory T-cells are a specialized type of T-cell that act to suppress the immune response and maintain homeostasis in the body. A regulatory T-cell-like cell is a cell that exhibits phenotypic characteristics of regulatory T-cells. For example, a regulatory T-cell-like cell may express one or more biomarkers expressed by a regulatory T-cell or exhibit one or more functions exhibited by a regulatory T-cell.

Regulatory T-cells produced in the body express the gene Cluster of Differentiation 3 (CD3) and the gene Cluster of Differentiation 25 (CD25), two putative marker genes for regulatory T-cells. In some embodiments, a regulatory T-cell-like cell expresses CD3 (i.e., is CD3-positive (CD3+)). In some embodiments, a regulatory T-cell-like cell expresses CD25 (i.e., is CD25+). Thus, in some embodiments, the regulatory T-cell-like cells produced by the methods provided herein are CD3+/CD25+ regulatory T-cell-like cells (i.e., cells that express the CD3 protein and the CD25 protein).

B Cell-Like Cells

Some aspects of the present disclosure provide B cell-like cells and methods of producing such cells. B cells are a specialized type of white blood cells that make antibodies. B cells are a part of the immune system and develop from stem cells in the bone marrow. A B cell-like cell is a cell that exhibits phenotypic characteristics of B cells. For example, a B cell-like cell may express one or more biomarkers expressed by a B cell or exhibit one or more functions exhibited by a B cell.

B cells produced in the body express the gene Cluster of Differentiation 19 (CD19) and the gene Cluster of Differentiation 27 (CD27), two putative marker genes for B cells. In some embodiments, a B cell-like cell expresses CD19 (i.e., is CD19-positive (CD19+)). In some embodiments, a B cell-like cell expresses CD27 (i.e., is CD27+). Thus, in some embodiments, the B cell-like cells produced by the methods provided herein are CD19+/CD27+ B cell-like cells (i.e., cells that express the CD19 protein and the CD27 protein).

Microglia-Like Cells

Some aspects of the present disclosure provide microglia-like cells and methods of producing such cells. Microglia are a specialized type glial cells that function as macrophages in the central nervous system. A microglia-like cell is a cell that exhibits phenotypic characteristics of microglia. For example, a microglia-like cell may express one or more biomarkers expressed by a microglial cell or exhibit one or more functions exhibited by a microglial cell.

Microglia produced in the body express the gene Cluster of Differentiation 11b (CD11b) and the gene C-X3-C Motif Chemokine Receptor 1 (CX3CR1), two putative marker genes for microglia. In some embodiments, a microglia-like cell expresses CD11b (i.e., is CD11b-positive (CD11b+)). In some embodiments, a microglia-like cell expresses CX3CR1 (i.e., is CX3CR1+). Thus, in some embodiments, the microglia-like cells produced by the methods provided herein are CD11b+/CX3CR1+ microglia-like cells (i.e., cells that express the CD11b protein and the CX3CR1 protein).

Pluripotent Stem Cells

The astrocyte-like cells, cytotoxic T-cell-like cells, hepatocyte-like cells, regulatory T-cell-like cells, B cell-like cells, and microglia-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.

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. *See, e.g.*, Ye L *et al. Curr Cardiol Rev.* 2013 Feb; 9(1): 63–72, incorporated herein by reference.

The PSCs provided herein are engineered to differentiate into a particular cell type of interest by overexpressing one or more proteins (e.g., transcription factors) in the PSCs. A transcription factor is a protein that controls the rate of transcription. In some embodiments, a protein is expressed in a PSC 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, for example, an endogenous (baseline) level. A cell "expresses" 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. In some embodiments, a control level of protein expression is an endogenous (baseline) level of expression of that same protein, for example, in a naturally-occurring pluripotent stem cell.

The term "protein" encompasses full length functional proteins as well as full-length or truncated functional variants of a protein, unless stated otherwise. Thus, the term "protein" encompasses full length functional transcription factors as well as full-length or truncated functional variants of the transcription factors, unless stated otherwise. A variant protein may comprise an amino acid sequence that has, for example, at least 80% identity to the amino acid sequence of a corresponding wild-type or reference protein and still exhibits the same function(s) as the corresponding wild-type or reference protein. In some embodiments, a variant protein has at least 85%, at least 90%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% identity to a wild-type or reference protein. In some embodiments a global alignment is used to determine the percent identity between two proteins (e.g., the alignment spanning the entire length of both proteins). In other embodiments, such as those involving truncated variants, a local alignment may be used to determine the percent identity between regions of similarity between the two proteins (e.g., the alignment spanning the entire length of the truncated proteins but not the entire length of the wild-type or reference protein).

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 (e.g., induced pluripotent stem cells) into one or more cell fate selected from, for example, astrocyte-like cells, cytotoxic T-cell-like cells, hepatocyte-like cells, regulatory T-cell-like cells, B cell-like cells, and microglia-like cells.

Differentiation of Astrocyte-Like Cells

Some aspects of the present disclosure provide a PSC (e.g., iPSC, such as a human iPSC) that is engineered to differentiate into an astrocyte-like cell. In some embodiments, the PSC comprises: a (one or more, e.g., 1, 2, 3, or 4) protein selected from ETS Transcription Factor ERG (ERG), Early Growth Response 1 (EGR1), friend leukemia integration 1 transcription factor (FLI1), and FBJ murine osteosarcoma viral oncogene homolog B (FOSB), wherein the protein is overexpressed. In some embodiments, the PSC comprises: one or more proteins selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises: two or more proteins selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises: three or more proteins selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises: one protein selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises: two proteins selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises: three proteins selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises: ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises and/or overexpresses ERG. In some embodiments, overexpression refers to an expression level above the expression level in a control cell. In some embodiments, the PSC comprises and/or overexpresses EGR1. In some embodiments, the PSC comprises and/or overexpresses FLI1. In some embodiments, the PSC comprises and/or overexpresses FOSB. In some embodiments, the PSC comprises and/or overexpresses ERG and EGR1; ERG and FLI1; ERG and FOSB; EGR1 and FLI1; EGR1 and FOSG; or FLI1 and FOSB. In some embodiments, the PSC comprises and/or overexpresses any preceding combination and one (at least one) additional protein selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises and/or overexpresses ERG, EGR1, FLI1, and FOSB.

A PSC, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ERG. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EGR1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding

FLI1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding FOSB. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ERG and an engineered polynucleotide comprising an open reading frame encoding EGR1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ERG and an engineered polynucleotide comprising an open reading frame encoding FLI1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ERG and an engineered polynucleotide comprising an open reading frame encoding FOSB. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EGR1 and an engineered polynucleotide comprising an open reading frame encoding FLI1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EGR1 and an engineered polynucleotide comprising an open reading frame encoding FOSG. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding FLI1 and an engineered polynucleotide comprising an open reading frame encoding FOSB. In some embodiments, the PSC comprises any preceding combination of engineered polynucleotides and one (at least one) additional engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ERG, EGR1, FLI1, and FOSB. In some embodiments, the PSC comprises an engineered polynucleotide comprising an open reading frame encoding ERG, an engineered polynucleotide comprising an open reading frame encoding EGR1, an engineered polynucleotide comprising an open reading frame encoding FLI1, and an engineered polynucleotide comprising an open reading frame encoding FOSB.

Differentiation of Cytotoxic T Cell-Like Cells

Some aspects of the present disclosure provide a PSC (e.g., iPSC, such as a human iPSC) that is engineered to differentiate into an cytotoxic T-cell-like cell. In some embodiments, the PSC comprises: a (one or more, e.g., 1, 2, 3, or 4) protein selected from Zinc finger and BTB domain containing 1 (ZBTB1), Runt-related transcription factor 3 (RUNX3), REL-associated protein (RELA), Nuclear respiratory factor 1 (NRF1), ETS2 Repressor Factor (ERF), and Sp4 transcription factor (SP4), wherein the protein is overexpressed. In some embodiments, the PSC comprises: one or more proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: two or more proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: three or more proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: four or more proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: five or more proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: one protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: two proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: three proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: four proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: five proteins selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises: ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises and/or overexpresses ZBTB1. In some embodiments, overexpression refers to an expression level above the expression level in a control cell. In some embodiments, the PSC comprises and/or overexpresses RUNX3. In some embodiments, the PSC comprises and/or overexpresses RELA. In some embodiments, the PSC comprises and/or overexpresses NRF1. In some embodiments, the PSC comprises and/or overexpresses ERF. In some embodiments, the PSC comprises and/or overexpresses SP4. In some embodiments, the PSC comprises and/or overexpresses ZBTB1 and RUNX3; ZBTB1 and RELA; ZBTB1 and NRF1; ZBTB1 and SP4; RUNX3 and RELA; RUNX3 and NRF1; RUNX3 and ERF; RUNX3 and SP4; RELA and NRF1; RELA and ERF; RELA and SP4; NRF1 and ERF; NRF1 and SP4; or ERF and SP4. In some embodiments, the PSC comprises and/or overexpresses any preceding combination and one (at least one) additional protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises and/or overexpresses ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4.

A PSC, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RUNX3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding SP4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding RUNX3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding ERF. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding SP4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RUNX3 and an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RUNX3 and an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RUNX3 and an engineered polynucleotide comprising an open reading frame encoding ERF. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RUNX3 and an engineered polynucleotide comprising an open reading frame encoding SP4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA and an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA and an engineered polynucleotide comprising an open reading frame encoding ERF. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA and an engineered polynucleotide comprising an open reading frame encoding SP4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding NRF1 and an engineered polynucleotide comprising an open reading frame encoding ERF. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding NRF1 and an engineered polynucleotide comprising an open reading frame encoding SP4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ERF and an engineered polynucleotide comprising an open reading frame encoding SP4. In some embodiments, the PSC comprises any preceding combination of engineered polynucleotides and one (at least one) additional engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4. In some embodiments, the PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1, an engineered polynucleotide comprising an open reading frame encoding RUNX3, an engineered polynucleotide comprising an open reading frame encoding RELA, an engineered polynucleotide comprising an open reading frame encoding NRF1, an engineered polynucleotide comprising an open reading frame encoding ERF, and an engineered polynucleotide comprising an open reading frame encoding SP4.

Differentiation of Hepatocyte-Like Cells

Some aspects of the present disclosure provide a PSC (e.g., iPSC, such as a human iPSC) that is engineered to differentiate into an hepatocyte-like cell. In some embodiments, the PSC comprises: a (one or more, e.g., 1, 2, 3, or 4) protein selected from Hepatocyte Nuclear Factor 4 Alpha (HNF4A), Hepatocyte Nuclear Factor 4 Gamma (HNF4G), TEA Domain Transcription Factor 4 (TEAD4), and Regulatory Factor X3 (RFX3), wherein the protein is overexpressed. In some embodiments, the PSC comprises: one or more proteins selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises: two or more proteins selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises: three or more proteins selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises: one protein selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises: two proteins selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises: three proteins selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises: HNF4A, HNF4G, TEAD4, and RFX3.

In some embodiments, the PSC comprises and/or overexpresses HNF4A. In some embodiments, overexpression refers to an expression level above the expression level in a control cell. In some embodiments, the PSC comprises and/or overexpresses HNF4G. In some embodiments, the PSC comprises and/or overexpresses TEAD4. In some embodiments, the PSC comprises and/or overexpresses RFX3. In some embodiments, the PSC comprises and/or overexpresses HNF4A and HNF4G; HNF4A and TEAD4; HNF4A and RFX3; HNF4G and TEAD4; HNF4G and RFX3; TEAD4 and RFX3; In some embodiments, the PSC comprises and/or overexpresses any preceding combination and one (at least one) additional protein selected from HNF4A, HNF4G, TEAD4, and RFX3.In some embodiments, the PSC comprises and/or overexpresses HNF4A, HNF4G, TEAD4, and RFX3.

A PSC, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from HNF4A, HNF4G, TEAD4, and RFX3.In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4A. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4G. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding TEAD4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RFX3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4A and an engineered polynucleotide comprising an open reading frame encoding HNF4G. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4A and an engineered polynucleotide comprising an open reading frame encoding TEAD4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4A and an engineered polynucleotide comprising an open reading frame encoding RFX3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4G and an engineered polynucleotide comprising an open reading frame encoding TEAD4. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4G and an engineered polynucleotide comprising an open reading frame encoding RFX3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding TEAD4 and an engineered polynucleotide comprising an open reading frame encoding RFX3. In some embodiments, the PSC comprises any preceding combination of engineered polynucleotides and one (at least one) additional engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from HNF4A, HNF4G, TEAD4, and RFX3. In some embodiments, the PSC comprises an engineered polynucleotide comprising an open reading frame encoding HNF4A, an engineered polynucleotide comprising an open reading frame encoding HNF4G, an engineered polynucleotide comprising an open reading frame encoding TEAD4, and an engineered polynucleotide comprising an open reading frame encoding RFX3.

Differentiation of Regulatory T-Like Cells

Some aspects of the present disclosure provide a PSC (e.g., iPSC, such as a human iPSC) that is engineered to differentiate into a regulatory T-cell-like cell. In some embodiments, the PSC comprises: a (one or more, e.g., 1, 2, 3, or 4) protein selected from

ETS Proto-Oncogene 1, Transcription Factor (ETS1), ETS Variant Transcription Factor 3 (ETV3), GA Binding Protein Transcription Factor Subunit Alpha (GABPA), and Krueppel-like factor 9 (KLF9), Nuclear Factor Kappa B Subunit 1 (NFKB1), wherein the protein is overexpressed. In some embodiments, the PSC comprises: one or more proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: two or more proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: three or more proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: four or more proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1.In some embodiments, the PSC comprises: one protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: two proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: three proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: four proteins selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises: ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises and/or overexpresses ETS1. In some embodiments, overexpression refers to an expression level above the expression level in a control cell. In some embodiments, the PSC comprises and/or overexpresses ETV3. In some embodiments, the PSC comprises and/or overexpresses GABPA. In some embodiments, the PSC comprises and/or overexpresses KLF9. In some embodiments, the PSC comprises and/or overexpresses NFKB1. In some embodiments, the PSC comprises and/or overexpresses ETS1 and ETV3; ETS1 and GABPA; ETS1 and KLF9; ETV3 and GABPA; ETV3 and KLF9; ETV3 and NFKB1; GABPA and KLF9; GABPA and NFKB1; orKLF9 and NFKB1.In some embodiments, the PSC comprises and/or overexpresses any preceding combination and one (at least one) additional protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1.In some embodiments, the PSC comprises and/or overexpresses ETS1, ETV3, GABPA, KLF9, and NFKB1.

A PSC, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETS1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETV3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding GABPA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding KLF9. In some embodiments, a

PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETS1 and an engineered polynucleotide comprising an open reading frame encoding ETV3. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETS1 and an engineered polynucleotide comprising an open reading frame encoding GABPA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETS1 and an engineered polynucleotide comprising an open reading frame encoding KLF9. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETS1 and an engineered polynucleotide comprising an open reading frame encoding NFKB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETV3 and an engineered polynucleotide comprising an open reading frame encoding GABPA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETV3 and an engineered polynucleotide comprising an open reading frame encoding KLF9. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETV3 and an engineered polynucleotide comprising an open reading frame encoding NFKB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding GABPA and an engineered polynucleotide comprising an open reading frame encoding KLF9. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding GABPA and an engineered polynucleotide comprising an open reading frame encoding NFKB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding KLF9 and an engineered polynucleotide comprising an open reading frame encoding NFKB1. In some embodiments, the PSC comprises any preceding combination of engineered polynucleotides and one (at least one) additional engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1. In some embodiments, the PSC comprises an engineered polynucleotide comprising an open reading frame encoding ETS1, an engineered polynucleotide comprising an open reading frame encoding ETV3, an engineered polynucleotide comprising an open reading frame encoding GABPA, an engineered polynucleotide comprising an open reading frame encoding KLF9, and an engineered polynucleotide comprising an open reading frame encoding NFKB1.

Differentiation of B Cell-Like Cells

Some aspects of the present disclosure provide a PSC (e.g., iPSC, such as a human iPSC) that is engineered to differentiate into an B cell-like cell. In some embodiments, the PSC comprises: a (one or more, e.g., 1, 2, 3, or 4) protein selected from Zinc finger and BTB domain containing 1 (ZBTB1), EBF Transcription Factor 1 (EBF1), REL-associated protein (RELA), Nuclear respiratory factor 1 (NRF1), and REL-associated protein (REL), wherein the protein is overexpressed. In some embodiments, the PSC comprises: one or more proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: two or more proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: three or more proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: four or more proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: one protein selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: two proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: three proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: four proteins selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises: ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises and/or overexpresses ZBTB1. In some embodiments, overexpression refers to an expression level above the expression level in a control cell. In some embodiments, the PSC comprises and/or overexpresses EBF1. In some embodiments, the PSC comprises and/or overexpresses RELA. In some embodiments, the PSC comprises and/or overexpresses NRF1. In some embodiments, the PSC comprises and/or overexpresses REL. In some embodiments, the PSC comprises and/or overexpresses ZBTB1 and EBF1; ZBTB1 and RELA; ZBTB1 and NRF1; EBF1 and RELA; EBF1 and NRF1; EBF1 and REL; RELA and NRF1; RELA and REL; or NRF1 and REL. In some embodiments, the PSC comprises and/or overexpresses any preceding combination and one (at least one) additional protein selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises and/or overexpresses ZBTB1, EBF1, RELA, NRF1, and REL.

A PSC, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EBF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding EBF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding REL. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EBF1 and an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EBF1 and an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding EBF1 and an engineered polynucleotide comprising an open reading frame encoding REL. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA and an engineered polynucleotide comprising an open reading frame encoding NRF1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA and an engineered polynucleotide comprising an open reading frame encoding REL. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding NRF1 and an engineered polynucleotide comprising an open reading frame encoding REL. In some embodiments, the PSC comprises any preceding combination of engineered polynucleotides and one (at least one) additional engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ZBTB1, EBF1, RELA, NRF1, and REL. In some embodiments, the PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1, an engineered polynucleotide comprising an open reading frame encoding EBF1, an engineered polynucleotide comprising an open reading frame encoding RELA, an engineered polynucleotide comprising an open reading frame encoding NRF1, and an engineered polynucleotide comprising an open reading frame encoding REL.

Differentiation of Microglia-Like Cells

Some aspects of the present disclosure provide a PSC (e.g., iPSC, such as a human iPSC) that is engineered to differentiate into an microglia-like cell. In some embodiments, the PSC comprises: a (one or more, e.g., 1, 2, 3, or 4) protein selected from Zinc finger and BTB domain containing 1 (ZBTB1), Spi-1 Proto-Oncogene (SPI1), REL-associated protein (RELA), and Signal Transducer And Activator Of Transcription 2 (STAT2), wherein the protein is overexpressed. In some embodiments, the PSC comprises: one or more proteins selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises: two or more proteins selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises: three or more proteins selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises: one protein selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises: two proteins selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises: three proteins selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises: ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises and/or overexpresses ZBTB1. In some embodiments, overexpression refers to an expression level above the expression level in a control cell. In some embodiments, the PSC comprises and/or overexpresses SPI1. In some embodiments, the PSC comprises and/or overexpresses RELA. In some embodiments, the PSC comprises and/or overexpresses STAT2. In some embodiments, the PSC comprises and/or overexpresses ZBTB1 and SPI1; ZBTB1 and RELA; ZBTB1 and STAT2; SPI1 and RELA; SPI1 and STAT2; or RELA and STAT2. In some embodiments, the PSC comprises and/or overexpresses any preceding combination and one (at least one) additional protein selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises and/or overexpresses ZBTB1, SPI1, RELA, and STAT2.

A PSC, in some embodiments, comprises an engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding SPI1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding STAT2. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding SPI1. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1 and an engineered polynucleotide comprising an open reading frame encoding STAT2. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding SPI1 and an engineered polynucleotide comprising an open reading frame encoding RELA. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding SPI1 and an engineered polynucleotide comprising an open reading frame encoding STAT2. In some embodiments, a PSC comprises an engineered polynucleotide comprising an open reading frame encoding RELA and an engineered polynucleotide comprising an open reading frame encoding STAT2. In some embodiments, the PSC comprises any preceding combination of engineered polynucleotides and one (at least one) additional engineered polynucleotide comprising an open reading frame encoding a (one or more) protein selected from ZBTB1, SPI1, RELA, and STAT2. In some embodiments, the PSC comprises an engineered polynucleotide comprising an open reading frame encoding ZBTB1, an engineered polynucleotide comprising an open reading frame encoding SPI1, an engineered polynucleotide comprising an open reading frame encoding RELA, and an engineered polynucleotide comprising an open reading frame encoding STAT2.

Engineered Transcription Factors - 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 initial 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 is it 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 physicallyregulated 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.

An engineered polynucleotide comprising an open reading frame encoding ERG (e.g., UniprotKB Accession No. P11308). In some embodiments, the protein comprises the sequence of:

```
MASTIKEALSVVSEDQSLFECAYGTPHLAKTEMTASSSSDYGQTSKMSPRVPQQDWLSQPPARVTIKMECNPSQV
NGSRNSPDECSVAKGGKMVGSPDTVGMNYGSYMEEKHMPPPNMTTNERRVIVPADPTLWSTDHVRQWLEWAVKEY
GLPDVNILLFQNIDGKELCKMTKDDFQRLTPSYNADILLSHLHYLRETPLPHLTSDDVDKALQNSPRLMHARNTG
GAAFIFPNTSVYPEATQRITTRPDLPYEPPRRSAWTGHGHPTPQSKAAQPSPSTVPKTEDQRPQLDPYQILGPTS
SRLANPGSGQIQLWQFLLELLSDSSNSSCITWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDK
NIMTKVHGKRYAYKFDFHGIAQALQPHPPESSLYKYPSDLPYMGSYHAHPQKMNFVAPHPPALPVTSSSFFAAPN
PYWNSPTGGIYPNTRLPTSHMPSHLGTYY*  (SEQ ID NO: 1)
```

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 1.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding EGR1 (e.g., UniprotKB Accession No. P18146). In some embodiments, the protein comprises the sequence of:

MTTLKEAVTFKDVAVVFTEEELRLLDLAQRKLYREVMLENFRNLLSVGHQSLHRDTFHFLKEEKFWMMETATQRE
GNLGGKIQMEMETVSESGTHEGLFSHQTWEQISSDLTRFQDSMVNSFQFSKQDDMPCQVDAGLSIIHVRQKPSEG
RTCKKSFSDVSVLDLHQQLQSREKSHTCDECGKSFCYSSALRIHQRVHMGEKLYNCDVCGKEFNQSSHLQIHQRI
HTGEKPFKCEQCGKGFSRRSGLYVHRKLHTGVKPHICEKCGKAFIHDSQLQEHQRIHTGEKPFKCDICCKSFRSR
ANLNRHSMVHMREKPFRCDTCGKSFGLKSALNSHRMVHTGEKRYKCEECGKRFIYRQDLYKHQIDHTGEKPYNCK
ECGKSFRWASGLSRHVRVHSGETTFKCEECGKGFYTNSQRYSHQRAHSGEKPYRCEECGKGYKRRLDLDFHQRVH
RGEKPYNCKECGKSFGWASCLLNHQRIHSGEKPFKCEECGKRFTQNSQLYTHRRVHSGEKPFKCEECGKRFTQNS
QLYSHRRVHTGVKPYKCEECGKGFNSKFNLDMHQRVHTGERPYNCKECGKSFSRASSILNHKRLHGDEKPFKCEE
CGKRFTENSQLHSHQRVHTGEKPYKCEKCGKSFRWASTHLTHQRLHSREKLLQCEDCGKSIVHSSCLKDQQRDQS
GEKTSKCEDCGKRYKRRLNLDTLLSLFLNDT* (SEQ ID NO: 2)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 2.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding FLI1 (e.g., UniprotKB Accession No. Q01543). In some embodiments, the protein comprises the sequence of:

MDGTIKEALSVVSDDQSLFDSAYGAAAHLPKADMTASGSPDYGQPHKINPLPPQQEWINQPVRVNVKREYDHMNG
SRESPVDCSVSKCSKLVGGGESNPMNYNSYMDEKNGPPPPNMTTNERRVIVPADPTLWTQEHVRQWLEWAIKEYS
LMEIDTSFFQNMDGKELCKMNKEDFLRATTLYNTEVLLSHLSYLRESSLLAYNTTSHTDQSSRLSVKEDPSYDSV
RRGAWGNNMNSGLNKSPPLGGAQTISKNTEQRPQPDPYQILGPTSSRLANPGSGQIQLWQFLLELLSDSANASCI
TWEGTNGEFKMTDPDEVARRWGERKSKPNMNYDKLSRALRYYYDKNIMTKVHGKRYAYKFDFHGIAQALQPHPTE
SSMYKYPSDISYMPSYHAHQQKVNFVPPHPSSMPVTSSSFFGAASQYWTSPTGGIYPNPNVPRHPNTHVPSHLGS
YY* (SEQ ID NO: 3)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 3.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding FOSB (e.g., UniprotKB Accession No. P53539). In some embodiments, the protein comprises the sequence of:

MFQAFPGDYDSGSRCSSSPSAESQYLSSVDSFGSPPTAAASQECAGLGEMPGSFVPTVTAITTSQDLQWLVQPTL
ISSMAQSQGQPLASQPPVVDPYDMPGTSYSTPGMSGYSSGGASGSGGPSTSGTTSGPGPARPARARPRRPREETL
TPEEEEKRRVRRERNKLAAAKCRNRRRELTDRLQAETDQLEEEKAELESEIAELQKEKERLEFVLVAHKPGCKIP
YEEGPGPGPLAEVRDLPGSAPAKEDGFSWLLPPPPPPPLPFQTSQDAPPNLTASLFTHSEVQVLGDPFPVVNPSY
TSSFVLTCPEVSAFAGAQRTSGSDQPSDPLNSPSLLALWIHPAFLY (SEQ ID NO: 4)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 4.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding ZBTB1 (e.g., UniprotKB Accession No. Q9Y2K1). In some embodiments, the protein comprises the sequence of:

MAKPSHSSYVLQQLNNQREWGFLCDCCIAIDDIYFQAHKAVLAACSSYFRMFFMNHQHSTAQLNLSNMKISAECF
DLIILQFMYLGKIMTAPSSFEQFKVAMNYLQLYNVPDCLEDIQDADCSSSKCSSSASSKQNSKMIFGVRMYEDTVA
RNGNEANRWCAEPSSTVNTPHNREADEESLQLGNFPEPLFDVCKKSSVSKLSTPKERVSRRFGRSFTCDSCGFGF
SCEKLLDEHVLTCTNRHLYQNTRSYHRIVDIRDGKDSNIKAEFGEKDSSKTFSAQTDKYRGDTSQAADDSASTTG

SRKSSTVESEIASEEKSRAAERKRIIIKMEPEDIPTDELKDFNIIKVTDKDCNESTDNDELEDEPEEPFYRYYVE
EDVSIKKSGRKTLKPRMSVSADERGGLENMRPPNNSSPVQEDAENASCELCGLTITEEDLSSHYLAKHIENICAC
GKCGQILVKGRQLQEHAQRCGEPQDLTMNGLGNTEEKMDLEENPDEQSEIRDMFVEMLDDFRDNHYQINSIQKKQ
LFKHSACPFRCPNCGQRFETENLVVEHMSSCLDQDMFKSAIMEENERDHRRKHFCNLCGKGFYQRCHLREHYTVH
TKEKQFVCQTCGKQFLRERQLRLHNDMHKGMARYVCSICDQGNFRKHDHVRHMISHLSAGETICQVCFQIFPNNE
QLEQHMDVHLYTCGICGAKFNLRKDMRSHYNAKHLKRTL (SEQ ID NO: 5)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 5.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding RUNX3 (e.g., UniprotKB Accession No. Q13761).

In some embodiments, the protein comprises the sequence of:

MASNSIFDSFPTYSPTFIRDPSTSRRFTPPSPAFPCGGGGGKMGENSGALSAQAAVGPGGRARPEVRSMVDVLAD
HAGELVRTDSPNFLCSVLPSHWRCNKTLPVAFKVVALGDVPDGTVVTVMAGNDENYSAELRNASAVMKNQVARFN
DLRFVGRSGRGKSFTLTITVFTNPTQVATYHRAIKVTVDGPREPRRHRQKLEDQTKPFPDRFGDLERLRMRVTPS
TPSPRGSLSTTSHFSSQPQTPIQGTSELNPFSDPRQFDRSFPTLPTLTESRFPDPRMHYPGAMSAAFPYSATPSG
TSISSLSVAGMPATSRFHHTYLPPPYPGAPQNQSGPFQANPSPYHLYYGTSSGSYQFSMVAGSSSGGDRSPTRML
ASCTSSAASVAAGNLMNPSLGGQSDGVEADGSHSNSPTALSTPGRMDEAVWRPY* (SEQ ID NO: 6)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 6.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding RELA (e.g., UniprotKB Accession No. Q04206).

In some embodiments, the protein comprises the sequence of:

MDELFPLIFPAEPAQASGPYVEIIEQPKQRGMRFRYKCEGRSAGSIPGERSTDTTKTHPTIKINGYTGPGTVRIS
LVTKDPPHRPHPHELVGKDCRDGFYEAELCPDRCIHSFQNLGIQCVKKRDLEQAISQRIQTNNNPFQVPIEEQRG
DYDLNAVRLCFQVTVRDPSGRPLRLPPVLSHPIFDNRAPNTAELKICRVNRNSGSCLGGDEIFLLCDKVQKEDIE
VCPQASTPALSLYVIPEHHQL* (SEQ ID NO: 7)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 7.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding NRF1 (e.g., UniprotKB Accession No. Q16656). In some embodiments, the protein comprises the sequence of:

MEEHGVTQTEHMATIEAHAVAQQVQQVHVATYTEHSMLSADEDSPSSPEDTSYDDSDILNSTAADEVTAHLAAAG
PVGMAAAAAVATGKKRKRPHVFESNPSIRKRQQTRLLRKLRATLDEYTTRVGQQAIVLCISPSKPNPVFKVFGAA
PLENVVRKYKSMILEDLESALAEHAPAPQEVNSELPPLTIDGIPVSVDKMTQAQLRAFIPEMLKYSTGRGKPGWG
KESCKPIWWPEDIPWANVRSDVRTEEQKQRVSWTQALRTIVKNCYKQHGREDLLYAFEDQQTQTQATATHSIAHL
VPSQTVVQTFSNPDGTVSLIQVGTGATVATLADASELPTTVTVAQVNYSAVADGEVEQNWATLQGGEMTIQTTQA
SEATQAVASLAEAAVAASQEMQQGATVTMALNSEAAAHAVATLAEATLQGGGQIVLSGETAAAVGALTGVQDANG
LFMADRAGRKWILTDKATGLVQIPVSMYQTVVTSLAQGNGPVQVAMAPVTTRISDSAVTMDGQAVEVVTLEQ*
(SEQ ID NO: 8)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 8.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding ERF (e.g., UniprotKB Accession No. P50548). In some embodiments, the protein comprises the sequence of:

MKTPADTGFAFPDWAYKPESSPGSRQIQLWHFILELLRKEEYQGVIAWQGDYGEFVIKDPDEVARLWGVRKCKPQ
MNYDKLSRALRYYYNKRILHKTKGKRFTYKFNFNKLVLVNYPFIDVGLAGGAVPQSAPPVPSGGSHFRFPPSTPS
EVLSPTEDPRSPPACSSSSSSLFSAVVARRLGRGSVSDCSDGTSELEEPLGEDPRARPPGPPDLGAFRGPPLARL
PHDPGVFRVYPRPRGGPEPLSPFPVSPLAGPGSLLPPQLSPALPMTPTHLAYTPSPTLSPMYPSGGGGPSGSGGG
SHFSFSPEDMKRYLQAHTQSVYNYHLSPRAFLHYPGLVVPQPQRPDKCPLPPMAPETPPVPSSASSSSSSSSSPF
KFKLQPPPLGRRQRAAGEKAVAGADKSGGSAGGLAEGAGALAPPPPPPQIKVEPISEGESEEVEVTDISDEDEED
GEVFKTPRAPPAPPKPEPGEAPGASQCMPLKLRFKRRWSEDCRLEGGGGPAGGFEDEGEDKKVRGEGPGEAGGPL
TPRRVSSDLQHATAQLSLEHRDS* (SEQ ID NO: 9)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 9.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding SP4 (e.g., UniprotKB Accession No. Q02446). In some embodiments, the protein comprises the sequence of:

MSDQKKEEEEEAAAAAAMATEGGKTSEPENNNKKPKTSGSQDSQPSPLALLAATCSKIGTPGENQATGQQQIIID
PSQGLVQLQNQPQQLELVTTQLAGNAWQLVASTPPASKENNVSQPASSSSSSSSSNNGSASPTKTKSGNSSTPGQ
FQVIQVQNPSGSVQYQVIPQLQTVEGQQIQINPTSSSSLQDLQGQIQLISAGNNQAILTAANRTASGNILAQNLA
NQTVPVQIRPGVSIPLQLQTLPGTQAQVVTTLPINIGGVTLALPVINNVAAGGGTGQVGQPAATADSGTSNGNQL
VSTPTNTTTSASTMPESPSSSTTCTTTASTSLTSSDTLVSSADTGQYASTSASSSERTIEESQTPAATESEAQSS
SQLQPNGMQNAQDQSNSLQQVQIVGQPILQQIQIQQPQQQIIQAIPPQSFQLQSGQTIQTIQQQPLQNVQLQAVN
PTQVLIRAPTLTPSGQISWQTVQVQNIQSLSNLQVQNAGLSQQLTITPVSSSGGTTLAQIAPVAVAGAPITLNTA
QLASVPNLQTVSVANLGAAGVQVQGVPVTITSVAGQQQGQDGVKVQQATIAPVTVAVGGIANATIGAVSPDQLTQ
VHLQQGQQTSDQEVQPGKRLRRVACSCPNCREGEGRGSNEPGKKKQHICHIEGCGKVYGKTSHLRAHLRWHTGER
PFICNWMFCGKRFTRSDELQRHRRTHTGEKRFECPECSKRFMRSDHLSKHVKTHQNKKGGGTALAIVTSGELDSS
VTEVLGSPRIVTVAAISQDSNPATPNVSTNMEEF* (SEQ ID NO: 10)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 10.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding HNF4A (e.g., UniprotKB Accession No. P41235). In some embodiments, the protein comprises the sequence of:

MRLSKTLVDMDMADYSAALDPAYTTLEFENVQVLTMGNDTSPSEGTNLNAPNSLGVSALCAICGDRATGKHYGAS
SCDGCKGFFRRSVRKNHMYSCRFSRQCVVDKDKRNQCRYCRLKKCFRAGMKKEAVQNERDRISTRRSSYEDSSLP
SINALLQAEVLSRQITSPVSGINGDIRAKKIASIADVCESMKEQLLVLVEWAKYIPAFCELPLDDQVALLRAHAG
EHLLLGATKRSMVFKDVLLLGNDYIVPRHCPELAEMSRVSIRILDELVLPFQELQIDDNEYAYLKAIIFFDPDAK
GLSDPGKIKRLRSQVQVSLEDYINDRQYDSRGRFGELLLLLPTLQSITWQMIEQIQFIKLFGMAKIDNLLQEMLL
GGPCQAQEGRGWSGDSPGDRPHTVSSPLSSLASPLCRFGQVA* (SEQ ID NO: 11)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 11.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding HNF4G (e.g., UniprotKB Accession No. Q14541). In some embodiments, the protein comprises the sequence of:

MDMANYSEVLDPTYTTLEFETMQILYNSSDSSAPETSMNTTDNGVNCLCAICGDRATGKHYGASSCDGCKGFFRR
SIRKSHVYSCRFSRQCVVDKDKRNQCRYCRLRKCFRAGMKKEAVQNERDRISTRRSTFDGSNIPSINTLAQAEVR
SRQISVSSPGSSTDINVKKIASIGDVCESMKQQLLVLVEWAKYIPAFCELPLDDQVALLRAHAGEHLLLGATKRS
MMYKDILLLGNNYVIHRNSCEVEISRVANRVLDELVRPFQEIQIDDNEYACLKAIVFFDPDAKGLSDPVKIKNMR
FQVQIGLEDYINDRQYDSRGRFGELLLLLPTLQSITWQMIEQIQFVKLFGMVKIDNLLQEMLLGGASNDGSHLHH
PMHPHLSQDPLTGQTILLGPMSTLVHADQISTPETPLPSPPQGSGQEQYKIAANQASVISHQHLSKQKQL*
(SEQ ID NO: 12)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 12.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding TEAD4 (e.g., UniprotKB Accession No. Q15561). In some embodiments, the protein comprises the sequence of:

MYGRNELIARYIKLRTGKTRTRKQVSSHIQVLARRKAREIQAKLKDQAAKDKALQSMAAMSSAQIISATAFHSSM
ALARGPGRPAVSGFWQGALPGQAGTSHDVKPFSQQTYAVQPPLPLPGFESPAGPAPSPSAPPAPPWQGRSVASSK
LWMLEFSAFLEQQQDPDTYNKHLFVHIGQSSPSYSDPYLEAVDIRQIYDKFPEKKGGLKDLFERGPSNAFFLVKF
WADLNTNIEDEGSSFYGVSSQYESPENMIITCSTKVCSFGKQVVEKVETEYARYENGHYSYRIHRSPLCEYMINF
IHKLKHLPEKYMMNSVLENFTILQVVTNRDTQETLLCIAYVFEVSASEHGAQHHIYRLVKE* (SEQ ID NO: 13)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 13.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding RFX3 (e.g., UniprotKB Accession No. P48380). In some embodiments, the protein comprises the sequence of:

MQTSETGSDTGSTVTLQTSVASQAAVPTQVVQQVPVQQQVQQVQTVQQVQHVYPAQVQYVEGSDTVYTNGAIRTT
TYPYTETQMYSQNTGGNYFDTQGSSAQVTTVVSSHSMVGTGGIQMGVTGGQLISSSGGTYLIGNSMENSGHSVTH
TTRASPATIEMAIETLQKSDGLSTHRSSLLNSHLQWLLDNYETAEGVSLPRSTLYNHYLRHCQEHKLDPVNAASF
GKLIRSIFMGLRTRRLGTRGNSKYHYYGIRVKPDSPLNRLQEDMQYMAMRQQPMQQKQRYKPMQKVDGVADGFTG
SGQQTGTSVGQTVIAQSQHHQQFLDASRALPEFGEVEISSLPDGTTFEDIKSLQSLYREHCEAILDVVVNLQFSL
IEKLWQTFWRYSPSTPTDGTTITESRSESTSFPIHFHG* (SEQ ID NO: 14)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 14.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding ETS1 (e.g., UniprotKB Accession No. P14921). In some embodiments, the protein comprises the sequence of:

MKAAVDLKPTLTIIKTEKVDLELFPSPDMECADVPLLTPSSKEMMSQALKATFSGFTKEQQRLGIPKDPRQWTET
HVRDWVMWAVNEFSLKGVDFQKFCMNGAALCALGKDCFLELAPDFVGDILWEHLEILQKEDVKPYQVNGVNPAYP
ESRYTSDYFISYGIEHAQCVPPSEFSEPSFITESYQTLHPISSEELLSLKYENDYPSVILRDPLQTDTLQNDYFA
IKQEVVTPDNMCMGRTSRGKLGGQDSFESIESYDSCGQEMGKEEKQT* (SEQ ID NO: 15)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 15.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding ETV3 (e.g., UniprotKB Accession No. P41162). In some embodiments, the protein comprises the sequence of:

MKAGCSIVEKPEGGGGYQFPDWAYKTESSPGSRQIQLWHFILELLQKEEFRHVIAWQQGEYGEFVIKDPDEVARL
WGRRKCKPQMNYDKLSRALRYYYNKRILHKTKGKRFTYKFNFNKLVMPNYPFINIRSSGKIQTLLVGN (SEQ ID NO: 16)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 16.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding GABPA (e.g., UniprotKB Accession No. Q06546). In some embodiments, the protein comprises the sequence of:

MTKREAEELIEIEIDGTEKAECTEESIVEQTYAPAECVSQAIDINEPIGNLKKLLEPRLQCSLDAHEICLQDIQL
DPERSLFDQGVKTDGTVQLSVQVISYQGIEPKLNILEIVKPADTVEVVIDPDAHHAESEAHLVEEAQVITLDGTK
HITTISDETSEQVTRWAAALEGYRKEQERLGIPYDPIQWSTDQVLHWVVWVMKEFSMTDIDLTTLNISGRELCSL
NQEDFFQRVPRGEILWSHLELLRKYVLASQEQQMNEIVTIDQPVQIIPASVQSATPTTIKVINSSAKAAKVQRAP
RISGEDRSSPGNRTGNNGQIQLWQFLLELLTDKDARDCISWVGDEGEFKLNQPELVAQKWGQRKNKPTMNYEKLS
RALRYYYDGDMICKVQGKRFVYKFVCDLKTLIGYSAAELNRLVTECEQKKLAKMQLHGIAQPVTAVALSTASLQT
EKDNL (SEQ ID NO: 17)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 17.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding KLF9 (e.g., UniprotKB Accession No. Q13886). In some embodiments, the protein comprises the sequence of:

MSAAAYMDFVAAQCLVSISNRAAVPEHGVAPDAERLRLPEREVTKEHGDPGDTWKDYCTLVTIAKSLLDLNKYRP
IQTPSVCSDSLESPDEDMGSDSDVTTESGSSPSHSPEERQDPGSAPSPLSLLHPGVAAKGKHASEKRHKCPYSGC
GKVYGKSSHLKAHYRVHTGERPFPCTWPDCLKKFSRSDELTRHYRTHTGEKQFRCPLCEKRFMRSDHLTKHARRH
TEFHPSMIKRSKKALANALL (SEQ ID NO: 18)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 18.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding NFKB1 (e.g., UniprotKB Accession No. P19838). In some embodiments, the protein comprises the sequence of:

MAEDDPYLGRPEQMFHLDPSLTHTIFNPEVFQPQMALPTADGPYLQILEQPKQRGFRFRYVCEGPSHGGLPGASS
EKNKKSYPQVKICNYVGPAKVIVQLVTNGKNIHLHAHSLVGKHCEDGICTVTAGPKDMVVGFANLGILHVTKKKV
FETLEARMTEACIRGYNPGLLVHPDLAYLQAEGGGDRQLGDREKELIRQAALQQTKEMDLSVVRLMFTAFLPDST
GSFTRRLEPVVSDAIYDSKAPNASNLKIVRMDRTAGCVTGGEEIYLLCDKVQKDDIQIRFYEEEENGGVWEGFGD
FSPTDVHRQFAIVFKTPKYKDINITKPASVFVQLRRKSDLETSEPKPFLYYPEIKDKEEVQRKRQKLMPNFSDSF
GGGSGAGAGGGGMFGSGGGGGGTGSTGPGYSFPHYGFPTYGGITFHPGTTKSNAGMKHGTMDTESKKDPEGCDKS
DDKNTVNLFGKVIETTEQDQEPSEATVGNGEVTLTYATGTKEESAGVQDNLFLEKAMQLAKRHANALFDYAVTGD
VKMLLAVQRHLTAVQDENGDSVLHLAIIHLHSQLVRDLLEVTSGLISDDIINMRNDLYQTPLHLAVITKQEDVVE
DLLRAGADLSLLDRLGNSVLHLAAKEGHDKVLSILLKHKKAALLLDHPNGDGLNAIHLAMMSNSLPCLLLLVAAG
ADVNAQEQKSGRTALHLAVEHDNISLAGCLLLEGDAHVDSTTYDGTTPLHIAAGRGSTRLAALLKAAGADPLVEN
FEPLYDLDDSWENAGEDEGVVPGTTPLDMATSWQVFDILNGKPYEPEFTSDDLLAQGDMKQLAEDVKLQLYKLLE
IPDPDKNWATLAQKLGLGILNNAFRLSPAPSKTLMDNYEVSGGTVRELVEALRQMGYTEAIEVIQAASSPVKTTS

QAHSLPLSPASTRQQIDELRDSDSVCDSGVETSFRKLSFTESLTSGASLLTLNKMPHDYGQEGPLEGKI* (SEQ ID NO: 19)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 19.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding EBF1 (e.g., UniprotKB Accession No. Q9UH73). In some embodiments, the protein comprises the sequence of:

MFGIQESIQRSGSSMKEEPLGSGMNAVRTWMQGAGVLDANTAAQSGVGLARAHFEKQPPSNLRKSNFFHFVLALY
DRQGQPVEIERTAFVGFVEKEKEANSEKTNNGIHYRLQLLYSNGIRTEQDFYVRLIDSMTKQAIVYEGQDKNPEM
CRVLLTHEIMCSRCCDKKSCGNRNETPSDPVIIDRFFLKFFLKCNQNCLKNAGNPRDMRRFQVVVSTTVNVDGHV
LAVSDNMFVHNNSKHGRRARRLDPSEGTPSYLEHATPCIKAISPSEGWTTGGATVIIIGDNFFDGLQVIFGTMLV
WSELITPHAIRVQTPPRHIPGVVEVTLSYKSKQFCKGTPGRFIYTALNEPTIDYGFQRLQKVIPRHPGDPERLPK
EVILKRAADLVEALYGMPHNNQEIILKRAADIAEALYSVPRNHNQLPALANTSVHAGMMGVNSFSGQLAVNVSEA
SQATNQGFTRNSSSVSPHGYVPSTTPQQTNYNSVTTSMNGYGSAAMSNLGGSPTFLNGSAANSPYAIVPSSPTMA
SSTSLPSNCSSSSGIFSFSPANMVSAVKQKSAFAPVVRPQTSPPPTCTSTNGNSLQAISGMIVPPM* (SEQ ID NO: 20)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 20.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding REL (e.g., UniprotKB Accession No. Q04864). In some embodiments, the protein comprises the sequence of:

MASGAYNPYIEIIEQPRQRGMRFRYKCEGRSAGSIPGEHSTDNNRTYPSIQIMNYYGKGKVRITLVTKNDPYKPH
PHDLVGKDCRDGYYEAEFGQERRPLFFQNLGIRCVKKKEVKEAIITRIKAGINPFNVPEKQLNDIEDCDLNVVRL
CFQVFLPDEHGNLTTALPPVVSNPIYDNRAPNTAELRICRVNKNCGSVRGGDEIFLLCDKVQKDDIEVRFVLNDW
EAKGIFSQADVHRQVAIVFKTPPYCKAITEPVTVKMQLRRPSDQEVSESMDFRYLPDEKDTYGNKAKKQKTTLLF
QKLCQDHVNFPERPRPGLLGSIGEGRYFKKEPNLFSHDAVVREMPTGVSSQAESYYPSPGPISSGLSHHASMAPL
PSSSWSSVAHPTPRSGNTNPLSSFSTRTLPSNSQGIPPFLRIPVGNDLNASNACIYNNADDIVGMEASSMPSADL
YGISDPNMLSNCSVNMMTTSSDSMGETDNPRLLSMNLENPSCNSVLDPRDLRQLHQMSSSSMSAGANSNTTVFVS
QSDAFEGSDFSCADNSMINESGPSNSTNPNSHGFVQDSQYSGIGSMQNEQLSDSFPYEFFQV* (SEQ ID NO: 21)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 21.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding SPI1 (e.g., UniprotKB Accession No. P17947). In some embodiments, the protein comprises the sequence of:

MLQACKMEGFPLVPPQPSEDLVPYDTDLYQRQTHEYYPYLSSDGESHSDHYWDFHPHHVHSEFESFAENNFTELQ
SVQPPQLQQLYRHMELEQMHVLDTPMVPPHPSLGHQVSYLPRMCLQYPSLSPAQPSSDEEEGERQSPPLEVSDGE
ADGLEPGPGLLPGETGSKKKIRLYQFLLDLLRSGDMKDSIWWVDKDKGTFQFSSKHKEALAHRWGIQKGNRKKMT
YQKMARALRNYGKTGEVKKVKKKLTYQFSGEVLGRGGLAERRHPPH* (SEQ ID NO: 22)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 22.

In some embodiments, an engineered polynucleotide comprises an open reading frame encoding STAT2 (e.g., UniprotKB Accession No. P52630). In some embodiments, the protein comprises the sequence of:

MAQWEMLQNLDSPFQDQLHQLYSHSLLPVDIRQYLAVWIEDQNWQEAALGSDDSKATMLFFHFLDQLNYECGRCS
QDPESLLLQHNLRKFCRDIQPFSQDPTQLAEMIFNLLLEEKRILIQAQRAQLEQGEPVLETPVESQQHEIESRIL
DLRAMMEKLVKSISQLKDQQDVFCFRYKIQAKGKTPSLDPHQTKEQKILQETLNELDKRRKEVLDASKALLGRLT
TLIELLLPKLEEWKAQQQKACIRAPIDHGLEQLETWFTAGAKLLFHLRQLLKELKGLSCLVSYQDDPLTKGVDLR
NAQVTELLQRLLHRAFVVETQPCMPQTPHRPLILKTGSKFTVRTRLLVRLQEGNESLTVEVSIDRNPPQLQGFRK
FNILTSNQKTLTPEKGQSQGLIWDFGYLTLVEQRSGGSGKGSNKGPLGVTEELHIISFTVKYTYQGLKQELKTDT
LPVVIISNMNQLSIAWASVLWFNLLSPNLQNQQFFSNPPKAPWSLLGPALSWQFSSYVGRGLNSDQLSMLRNKLF
GQNCRTEDPLLSWADFTKRESPPGKLPFWTWLDKILELVHDHLKDLWNDGRIMGFVSRSQERRLLKKTMSGTFLL
RFSESSEGGITCSWVEHQDDDKVLIYSVQPYTKEVLQSLPLTEIIRHYQLLTEENIPENPLRFLYPRIPRDEAFG
CYYQEKVNLQERRKYLKHRLIVVSNRQVDELQQPLELKPEPELESLELELGLVPEPELSLDLEPLLKAGLDLGPE
LESVLESTLEPVIEPTLCMVSQTVPEPDQGPVSQPVPEPDLPCDLRHLNTEPMEIFRNCVKIEEIMPNGDPLLAG
QNTVDEVYVSRPSHFYTDGPLMPSDF* (SEQ ID NO: 23)

In some embodiments, the protein comprises a sequence that has at least 80%, at least 85%, at least 90%, or at least 95% identity to the amino acid sequence of SEQ ID NO: 23.

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, and 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. In some embodiments, a PSC comprises fewer than 25 copies of an engineered polynucleotide. For example, a PSC may comprise fewer than 20, fewer than 15, or fewer than 10 copies of an engineered polynucleotide. Greater than 20 copies are also contemplated herein.

Methods of Producing Cells for Stem Cells

Some aspects of the present disclosure relate to a method of using direct transcription factor overexpression in conjunction with growth factor culturing to induce CD44$^+$ and A2B5$^+$ astrocyte-like cells from stem cells (e.g., iPSCs) in fewer than 6 days (e.g., about 3 to about 5 days, e.g., about 3, about 4, or about 5 days). The methods of producing astrocyte-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) (e.g., iPSCs, such as human iPSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from ERG, EGR1, FLI1, and FOSB to produce astrocyte-like cells. In some embodiments, the method comprises expressing ERG in PSCs of the expanded population. In some embodiments, the method comprises expressing ERG1 in PSCs of the expanded population. In some embodiments, the method comprises expressing FLI1 in PSCs of the expanded population. In some embodiments, the method comprises expressing FOSB in

PSCs of the expanded population. In some embodiments, the method comprises expressing ERG and ERG1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ERG and FLI1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ERG and FOSB in PSCs of the expanded population. In some embodiments, the method comprises expressing ERG1 and FLI1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ERG1 and FOSB in PSCs of the expanded population. In some embodiments, the method comprises expressing FLI1 and FOSB in PSCs of the expanded population. In some embodiments, the method comprises expressing any preceding combination and a (at least one) protein selected from ERG, EGR1, FLI1, and FOSB in PSCs of the expanded population. In some embodiments, the method comprises expressing ERG, EGR1, FLI1, and FOSB in PSCs of the expanded population.

Other aspects of the present disclosure relate to a method of using direct transcription factor overexpression in conjunction with growth factor culturing to induce CD3$^+$ and CD8$^+$ cytotoxic T-cell-like cells from stem cells in fewer than 6 days (e.g., about 3 to about 5 days, e.g., about 3, about 4, or about 5 days). The methods of producing cytotoxic T-cell-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) (e.g., iPSCs, such as human iPSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 to produce cytotoxic T-cell-like cells. In some embodiments, the method comprises expressing ZBTB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing RUNX3 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ERF in PSCs of the expanded population. In some embodiments, the method comprises expressing SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and RUNX3 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and ERF in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing RUNX3 and RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing RUNX3 and NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing RUNX3 and ERF in PSCs of the expanded population. In some embodiments, the method comprises expressing RUNX3 and SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA and NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA and ERF in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA and SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing NRF1 and ERF in PSCs of the expanded population. In some embodiments, the method comprises expressing NRF1 and SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing ERF and SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing any preceding combination and a (at least one) protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 in PSCs of the expanded population.

Yet other aspects of the present disclosure relate to a method of using direct transcription factor overexpression in conjunction with growth factor culturing to induce CD184$^+$ and ASGPR1$^+$ hepatocyte-like cells from stem cells in fewer than 6 days (e.g., about 3 to about 5 days, e.g., about 3, about 4, or about 5 days). The methods of producing hepatocyte T-cell-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) (e.g., iPSCs, such as human iPSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from HNF4A, HNF4G, TEAD4, and RFX3 to produce hepatocyte T-cell-like cells. In some embodiments, the method comprises expressing HNF4A in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4G in PSCs of the expanded population. In some embodiments, the method comprises expressing TEAD4 in PSCs of the expanded population. In some embodiments, the method comprises expressing RFX3 in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4A and HNF4G in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4A and TEAD4 in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4A and RFX3 in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4G and TEAD4 in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4G and RFX3 in PSCs of the expanded population. In some embodiments, the method comprises expressing TEAD4 and RFX3 in PSCs of the expanded population. In some embodiments, the method comprises expressing any preceding combination and a (at least one) protein selected from HNF4A, HNF4G, TEAD4, and RFX3, in PSCs of the expanded population. In some embodiments, the method comprises expressing HNF4A, HNF4G, TEAD4, and RFX3 in PSCs of the expanded population.

Still other aspects of the present disclosure relate to a method of using direct transcription factor overexpression in conjunction with growth factor culturing to induce CD3+ and CD25+ regulatory T-cell-like cells from stem cells in fewer than 6 days (e.g., about 3 to about 5 days, e.g., about 3, about 4, or about 5 days). The methods of producing regulatory T-cell-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) (e.g., iPSCs, such as human iPSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1 to produce regulatory T-cell-like cells. In some embodiments, the method comprises expressing ETS1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ETV3 in PSCs of the expanded population. In some embodiments, the method comprises expressing GABPA in PSCs of the expanded population. In some embodiments, the method comprises expressing KLF9 in PSCs of the expanded population. In some embodiments, the method comprises expressing NFKB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ETS1 and ETV3 in PSCs of the expanded population. In some embodiments, the method comprises expressing ETS1 and GABPA in PSCs of the expanded population. In some embodiments, the method comprises expressing ETS1 and KLF9 in PSCs of the expanded population. In some embodiments, the method comprises expressing ETS1 and NFKB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ETV3 and GABPA in PSCs of the expanded population. In some embodiments, the method comprises expressing ETV3 and KLF9 in PSCs of the expanded population. In some embodiments, the method comprises expressing ETV3 and NFKB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing GABPA and KLF9 in PSCs of the expanded population. In some embodiments, the method comprises expressing GABPA and NFKB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing KLF9 and NFKB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing any preceding combination and a (at least one) protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, in PSCs of the expanded population. In some embodiments, the method comprises expressing ETS1, ETV3, GABPA, KLF9, and NFKB1, in PSCs of the expanded population.

Further aspects of the present disclosure relate to a method of using direct transcription factor overexpression in conjunction with growth factor culturing to induce CD19+ and CD27+ B cell-like cells from stem cells in fewer than 6 days (e.g., about 3 to about 5 days, e.g., about 3, about 4, or about 5 days). The methods of producing B cell-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) (e.g., iPSCs, such as human iPSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from ZBTB1, EBF1, RELA, NRF1, and REL to produce B cell-like cells. In some embodiments, the method comprises expressing ZBTB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing EBF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing REL in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and EBF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and REL in PSCs of the expanded population. In some embodiments, the method comprises expressing EBF1 and RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing EBF1 and NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing EBF1 and REL in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA and NRF1 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA and REL in PSCs of the expanded population. In some embodiments, the method comprises expressing NRF1 and REL in PSCs of the expanded population. In some embodiments, the method comprises expressing any preceding combination and a (at least one) protein selected from ZBTB1, EBF1, RELA, NRF1, and REL in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1, EBF1, RELA, NRF1, and REL in PSCs of the expanded population.

Additional aspects of the present disclosure relate to a method of using direct transcription factor overexpression in conjunction with growth factor culturing to induce CD11b+ and CX3CR1+ microglia-like cells from stem cells in fewer than 6 days (e.g., about 3 to about 5 days, e.g., about 3, about 4, or about 5 days). The methods of producing microglia-like cells provided herein, in some aspects, comprises culturing, in culture media, a population of pluripotent stem cells (PSCs) (e.g., iPSCs, such as human iPSCs) to produce an expanded population of PSCs; and expressing in PSCs of the expanded population a protein selected from ZBTB1, SPI1, RELA, and STAT2 to produce microglia-like cells. In some embodiments, the method comprises expressing ZBTB1 in PSCs of the expanded population. In some embodiments, the method comprises expressing SPI1 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing STAT2 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and SPI1 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1 and STAT2 in PSCs of the expanded population. In some embodiments, the method comprises expressing SPI1 and RELA in PSCs of the expanded population. In some embodiments, the method comprises expressing SPI1 and STAT2 in PSCs of the expanded population. In some embodiments, the method comprises expressing RELA and STAT2 in PSCs of the expanded population. In some embodiments, the method comprises expressing any preceding combination and a (at least one) protein selected from ZBTB1, SPI1, RELA, and STAT2 in PSCs of the expanded population. In some embodiments, the method comprises expressing ZBTB1, SPI1, RELA, and STAT2 in PSCs of the expanded population.

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, in some embodiments, about $1x10^2$-$1x10^{10}$, about $1x10^2$-$1x10^9$, about $1x10^2$-$1x10^8$, or about $1x10^2$-$1x10^7$ PSCs. In some embodiments, the population comprises about $1x10^3$-$1x10^8$ or about $1x10^3$-$1x10^7$ PSCs. In some embodiments, the population comprises about $1x10^4$-$1x10^7$ or about $1x10^5$-$1x10^6$ PSCs. In some embodiments, the population comprises about $1x10^1$ PSCs, about $1x10^2$ PSCs, about 1x10^3 PSCs, about 1x10^4 PSCs, about 1x10^5 PSCs, about 1x10^6 PSCs, about 1x10^7 PSCs, about 1x10^8 PSCs, about 1x10^9 PSCs, or about 1x10^{10} PSCs.

In some embodiments, the population of PSCs is cultured for about 1 day to about 10 days. In some embodiments, the population of PSCs is cultured for no more than 10 days. For example, the population of PSCs may be cultured for no more than 9 days, no more than 8 days, no more than 7 days, no more than 6 days, or no more than 5 days. In some embodiments, the population of PSCs is cultured for no more than 6 days.

In some embodiments, the population of PSCs is cultured for about 2 to about 6 days, about 2 to about 5 days, about 2 to about 4 days, about 3 to about 6 days, about 3 to about 5 days, or about 3 to about 4 days. In some embodiments, the population of PSCs is cultured for about 2 days, about 3 days, about 4 days, about 5 days, or about 6 days.

In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 10 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein (e.g., selected from (a) ERG, EGR1, FLI1 and FOSB to produce astrocyte-like cells; (b) ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 to produce cytotoxic T-cell-like cells; (c) HNF4G, TEAD4, and RFX3 to produce hepatocyte-like cells; (d) ETS1, ETV3, GABPA, KLF9, and NFKB1 to produce regulatory T-cell-like cells; (e) EBF1, ZBTB1, RELA, NRF1, and REL to produce B cell-like cells; or (f) SPI1, ZBTB1, RELA, and STAT2 to produce microglia-like cells). In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 9 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 8 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 7 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 6 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 5 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 4 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 3 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein. In some embodiments, a differentiated cell type is produced from a PSC (or a population of PSCs) within 2 days of expressing (e.g., inducing expression of) an engineered (exogenous) nucleic acid encoding a transcription factor provided herein.

Some methods of the present disclosure comprise (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a (one or more) protein selected from ERG, EGR1, FLI1, and FOSB; (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 astrocyte-like cells (e.g.,CD44+/A2B5+ astrocyte-like cells). In some embodiments, the series of induction media comprises a first, a second, a third, and a fourth induction media.

Some methods of the present disclosure comprise (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a (one or more) protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4; (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 cytotoxic T-cell-like cells (e.g., CD3+/CD8+ cytotoxic T-cell-like cells). In some embodiments, the series of induction media comprises a first, a second, a third, and a fourth induction media.

Some methods of the present disclosure comprise (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a (one or more) protein selected from HNF4A, HNF4G, TEAD4, and RFX3; (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 hepatocyte-like cells (e.g., CD184+/ASGPR1+ hepatocyte-like cells). In some embodiments, the series of induction media comprises a first, a second, a third, and a fourth induction media.

Some methods of the present disclosure comprise (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a (one or more) protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1; (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 regulatory T-cell-like cells (e.g., CD3+/CD25+ regulatory T-cell-like cells). In some embodiments, the series of induction media comprises a first, a second, a third, and a fourth induction media.

Some methods of the present disclosure comprise (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a (one or more) protein selected from EBF1, ZBTB1, RELA, NRF1, and REL; (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 B cell-like cells (e.g., CD19+/CD27+ B cell-like cells). In some embodiments, the series of induction media comprises a first, a second, a third, and a fourth induction media.

Some methods of the present disclosure comprise (a) delivering to PSCs an engineered polynucleotide comprising an inducible promoter operably linked to an open reading frame encoding a (one or more) protein selected from SPI1, ZBTB1, RELA, and STAT2; (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 microglia-like cells (e.g., CD11b+/CX3CR1+ microglia-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, 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 5x10^3 PSCs. For example, the expanded population (e.g., at the time of induction) may comprise at least 1x10^4, at least 1x10^5, at least 1x10^6, or at least 1x10^7 PSCs. In some embodiments, the expanded population of PSCs comprises about 5x10^3 PSCs to about 1x10^7 PSCs.

In some embodiments, PSCs of the expanded population are cultured at a density of about 2,000 cells/cm^2 to about 3,000 cells/cm^2. In some embodiments, PSCs of the expanded population are cultured at a density of about 500/cm^2 - 10000/cm^2 PSCs. In some embodiments, the PSCs of the expanded population are cultured at a density of about 1000/cm^2 - 9500/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 1500/cm^2 - 9000/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 2000/cm^2 - 8500/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 2500/cm^2 - 8000/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 3000/cm^2 - 7500/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 3500/cm^2 - 7000/cm^2 PSCs. In some embodiments, the population comprises 4000/cm^2 - 6500/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 4500/cm^2 - 6000/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of about 5000/cm^2 - 5500/cm^2 PSCs. In some embodiments, PSCs of the expanded population are cultured at a density of at least 500/cm^2 PSCs, at least 1000/cm^2 PSCs, at least 1500/cm^2 PSCs, at least 2000/cm^2 PSCs, at least 2500/cm^2 PSCs, at least 3000/cm^2 PSCs, at least 3500/cm^2 PSCs, at least 4000/cm^2 PSCs, at least 4500/cm^2 PSCs, at least 5000/cm^2 PSCs, at least 5500/cm^2 PSCs, at least 6000/cm^2 PSCs, at least 6500/cm^2 PSCs, at least 7000/cm^2 PSCs, at least 7500/cm^2 PSCs, at least 8000/cm^2 PSCs, at least 8500/cm^2 PSCs, at least 9000/cm^2 PSCs, at least 9500/cm^2 PSCs, or at least 10000/cm^2 PSCs.

In some embodiments, PSCs of the expanded population are cultured for 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 2 to about 8 days, about 2 to about 7 days, about 2 to about 6 days, about 2 to about 5 days, about 2 to about 4 days, about 3 to about 8 days, about 3 to about 7 days, about 3 to about 6 days, about 3 to about 5 days, or about 3 to about 4 days. In some embodiments, PSCs of the expanded population are cultured for about 2 days, about 3 days, about 4 days, about 5 days, about 6 days, about 7 days, or about 8 days.

In some embodiments, PSCs of the expanded population are cultured in a first induction media for about 6 to about 36 hours. For example, the PSC may be cultured in a first induction media for about 6 to about 24 hours, about 6 to about 18 hours, about 6 to about 12 hours, 12 to about 36 hours, about 12 to about 24 hours, about 12 to about 18 hours, 18 to about 36 hours, or about 18 to about 24 hours. In some embodiments, the PSCs are cultured in a first induction 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, about 24 hours, about 25 hours, about 26 hours, about 27 hours, about 28 hours, about 29 hours, or about 30 hours.

In some embodiments, PSCs of the expanded population are cultured in a second induction media for about 6 to about 36 hours. For example, the PSC may be cultured in a second induction media for about 6 to about 24 hours, about 6 to about 18 hours, about 6 to about 12 hours, 12 to about 36 hours, about 12 to about 24 hours, about 12 to about 18 hours, 18 to about 36 hours, or about 18 to about 24 hours. In some embodiments, the PSCs are cultured in a second induction 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, about 24 hours, about 25 hours, about 26 hours, about 27 hours, about 28 hours, about 29 hours, or about 30 hours.

In some embodiments, PSCs of the expanded population are cultured in a third induction media for about 6 to about 36 hours. For example, the PSC may be cultured in a third induction media for about 6 to about 24 hours, about 6 to about 18 hours, about 6 to about 12 hours, 12 to about 36 hours, about 12 to about 24 hours, about 12 to about 18 hours, 18 to about 36 hours, or about 18 to about 24 hours. In some embodiments, the PSCs are cultured in a third induction 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, about 24 hours, about 25 hours, about 26 hours, about 27 hours, about 28 hours, about 29 hours, or about 30 hours.

In some embodiments, PSCs of the expanded population are cultured in a fourth induction media for about 6 to about 36 hours. For example, the PSC may be cultured in a fourth induction media for about 6 to about 24 hours, about 6 to about 18 hours, about 6 to about 12 hours, 12 to about 36 hours, about 12 to about 24 hours, about 12 to about 18 hours, 18 to about 36 hours, or about 18 to about 24 hours. In some embodiments, the PSCs are cultured in a fourth induction 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, about 24 hours, about 25 hours, about 26 hours, about 27 hours, about 28 hours, about 29 hours, or about 30 hours.

In some embodiments, PSCs are incubated for at least 6 hours. In some embodiments, after incubation, the media is removed from the plate and the plate is washed with DMEM/F12.

Some aspects provide a method of producing astrocyte-like cells , comprising: (a) delivering 1-20 copies, for example, 8-10 copies, of an engineered nucleic acid to a population of human stem cells, wherein the engineered nucleic acid comprises an inducible promoter operably linked to an open reading frame encoding a transcription factor selected from ERG, EGR1, FLI1 and FOSB; (b) inducing activation of the inducible promoter; and (c) culturing the population of human stem cells for no more than 10 days, preferably no more than 7 days, in a series of induction media (e.g., a first, second, third and/or fourth induction media as described herein) to produce astrocyte-like cells (e.g.,CD44+/A2B5+ astrocyte-like cells).

In some embodiments, the method of producing astrocyte-like cells comprises delivering to the human stem cells 1-20 copies of each of the following: (i) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ERG, (ii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding EGR1, (iii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding FLI1, and (iv) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding FOSB.

Some aspects provide a method of producing cytotoxic T-cell-like cells, comprising: (a) delivering 1-20 copies, for example, 8-10 copies, of an engineered nucleic acid to a population of human stem cells, wherein the engineered nucleic acid comprises an inducible promoter operably linked to an open reading frame encoding a transcription factor selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4; (b) inducing activation of the inducible promoter; and (c) culturing the population of human stem cells for no more than 10 days, preferably no more than 7 days, in a series of induction media (e.g., a first, second, third and/or fourth induction media as described herein) to produce cytotoxic T-cell-like cells (e.g., CD3+/CD8+ cytotoxic T-cell-like cells).

In some embodiments, the method of producing cytotoxic T-cell-like cells comprises delivering to the human stem cells 1-20 copies of each of the following: (i) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ZBTB1, (ii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding RUNX3, (iii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding RELA, (iv) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding NRF1, (v) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ERF, and (vi) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding SP4.

Some aspects provide a method of producing hepatocyte-like cells, comprising: (a) delivering 1-20 copies, for example, 8-10 copies, of an engineered nucleic acid to a population of human stem cells, wherein the engineered nucleic acid comprises an inducible promoter operably linked to an open reading frame encoding a transcription factor selected from HNF4G, TEAD4, and RFX3; (b) inducing activation of the inducible promoter; and (c) culturing the population of human stem cells for no more than 10 days, preferably no more than 7 days, in a series of induction media (e.g., a first, second, third and/or fourth induction media as described herein) to produce hepatocyte-like cells (e.g., CD184+/ASGPR1+ hepatocyte-like cells).

In some embodiments, the method of producing hepatocyte-like cells comprises delivering to the human stem cells 1-20 copies of each of the following: (i) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding HNF4G, (ii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding TEAD4, and (iii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding RFX3.

Some aspects provide a method of producing regulatory T-cell-like cells, comprising: (a) delivering 1-20 copies, for example, 8-10 copies, of an engineered nucleic acid to a population of human stem cells, wherein the engineered nucleic acid comprises an inducible promoter operably linked to an open reading frame encoding a transcription factor selected from ETS1, ETV3, GABPA, KLF9, and NFKB1; (b) inducing activation of the inducible promoter; and (c) culturing the population of human stem cells for no more than 10 days, preferably no more than 7 days, in a series of induction media (e.g., a first, second, third and/or fourth induction media as described herein) to produce regulatory T-cell-like cells (e.g., CD3+/CD25+ regulatory T-cell-like cells).

In some embodiments, the method of producing regulatory T-cell-like cells comprises delivering to the human stem cells 1-20 copies of each of the following: (i) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ETS1, (ii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ETV3, (iii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding GABPA, (iv) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding KLF9, and (v) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding NFKB1.

Some aspects provide a method of producing B cell-like cells, comprising: (a) delivering 1-20 copies, for example, 8-10 copies, of an engineered nucleic acid to a population of human stem cells, wherein the engineered nucleic acid comprises an inducible promoter operably linked to an open reading frame encoding a transcription factor selected from EBF1, ZBTB1, RELA, NRF1, and REL; (b) inducing activation of the inducible promoter; and (c) culturing the population of human stem cells for no more than 10 days, preferably no more than 7 days, in a series of induction media (e.g., a first, second, third and/or fourth induction media as described herein) to produce B cell-like cells (e.g., CD19+/CD27+ B cell-like cells).

In some embodiments, the method of producing B cell-like cells comprises delivering to the human stem cells 1-20 copies of each of the following: (i) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding EBF1, (ii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ZBTB1, (iii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding RELA, (iv) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding NRF1, and (v) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding REL.

Some aspects provide a method of producing microglia-like cells, comprising: (a) delivering 1-20 copies, for example, 8-10 copies, of an engineered nucleic acid to a population of human stem cells, wherein the engineered nucleic acid comprises an inducible promoter operably linked to an open reading frame encoding a transcription factor selected from SPI1, ZBTB1, RELA, and STAT2; (b) inducing activation of the inducible promoter; and (c) culturing the population of human stem cells for no more than 10 days, preferably no more than 7 days, in a series of induction media (e.g., a first, second, third and/or fourth induction media as described herein) to produce microglia-like cells (e.g., CD11b+/CX3CR1+ microglia-like cells).

In some embodiments, the method of producing microglia-like cells comprises delivering to the human stem cells 1-20 copies of each of the following: (i) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding SPI1, (ii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding ZBTB1, (iii) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding RELA, and (iv) an engineered nucleic acid comprising an inducible promoter operably linked to an open reading frame encoding STAT2.

In some embodiments, the engineered nucleic acid(s) is/are integrated into the genome of the human stem cells.

In some embodiments, the human stem cells are pluripotent stem cells (PSCs), for example, induced pluripotent stem cells (iPSCs).

In some embodiments, the inducing step comprises delivering a chemical inducing agent, such as doxycycline or tetracycline, that activates the inducible promoter (e.g., a doxycycline-inducible promoter or a tetracycline-inducible promoter).

In some embodiments, the human stem cells are first expanded in a feeder-free, serum-free culture media, prior to delivery of the engineered nucleic acid(s).

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, kanemycin, 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 75 to 150 µl per 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., 80ng/ml to 120ng/ml) and recombinant human transforming growth factor β (rh TGFβ) (e.g., 20 to 25pM). In some embodiments, culture media further comprises rh bFGF and rh TGFβ. In some embodiments, culture media comprises mTeSR™ media (STEMCELL Technologies).

In some embodiments, a first induction media comprises one or more of (e.g., 2, 3, 4, or more of) B-27 Supplement (e.g., 90X to110X), L-alanyl-L-glutamine (e.g., 1.8 mM to 2.2 mM), an inducing agent (e.g., doxycycline (e.g., 50 ng/ml to 2000 ng/ml)), Activin A (e.g., 50 ng/ml to 150 ng/ml), a glycogen synthase kinase (GSK) 3 inhibitor (e.g., 2.8 µM to 3.2 µM), a selective FGFR1 and FGFR3 inhibitor (e.g., 90 nM to 110 nM), and a small molecule ROCK inhibitor (e.g., 8 µM to 12 µM). In some embodiments, a first induction media comprises B-27, L-alanyl-L-glutamine, an inducing agent (e.g., doxycycline), Activin A, a glycogen synthase kinase (GSK) 3 inhibitor, and a selective FGFR1 and FGFR3 inhibitor. For example, the first induction media may comprise aRB27 Media, doxycycline, Activin A, CHIR99021, and PD 173074.

In some embodiments, the second induction media comprises one or more of (e.g., 2, 3, 4, or more of) B-27 Supplement (e.g., 90X to 110X), an inducing agent (e.g., doxycycline (e.g., 50 ng/ml to 2000 ng/ml)), a small molecule inhibitor of tankyrase (TNKS) (e.g., 0.9 µM to 1.1 µM), and a human bone morphogenic protein 4 (hBMP4) (e.g., 20 ng/ml to 250 ng/ml). In some embodiments, the second induction media comprises B-27, an inducing agent (e.g., doxycycline), a small molecule inhibitor of tankyrase (TNKS), and a human bone morphogenic protein 4 (hBMP4). For example, the second induction media may comprise aRB27 Media, doxycycline, XAV939, and human bone morphogenic protein 4 (hBMP4).

In some embodiments, the third induction media comprises one or more of (e.g., 2, 3, 4, or more of) B-27, an inducing agent (e.g., doxycycline), a small molecule inhibitor of tankyrase (e.g., 0.9 μM to 1.1 μM), stem cell factor (SCF) (e.g., 25 ng/ml to 200 ng/ml), and epidermal growth factor (EGF) (e.g., 25 ng/ml to 100 ng/ml). In some embodiments, the third induction media comprises B-27 Supplement (e.g., 90X to 110X), an inducing agent (e.g., doxycycline (e.g., 50 ng/ml to 2000 ng/ml)), a small molecule inhibitor of tankyrase (e.g., 0.9 μM to 1.1 μM), stem cell factor (SCF) (e.g., 25ng/ml to 200ng/ml), and epidermal growth factor (EGF) (e.g., 25 ng/ml to 100 ng/ml). For example, the third induction media may comprise aRB27 Media, doxycycline, XAV939, SCF, and EGF.

In some embodiments, the fourth induction media comprises one or more of (e.g., 2, 3, 4, or more of) B-27 Supplement (90-110X), an inducing agent (e.g., doxycycline (e.g., 50 ng/ml to 2000 ng/ml)), a small molecule inhibitor of tankyrase (e.g., 0.9 μM to 1.1 μM), hBMP4 (e.g., 20 ng/ml to 250 ng/ml), SCF (e.g., 25 ng/ml to 200 ng/ml), and EGF (e.g., 25 ng/ml to 100 ng/ml). In some embodiments, the fourth induction media comprises B-27, an inducing agent (e.g., doxycycline), a small molecule inhibitor of tankyrase, hBMP4, SCF, and EGF. For example, the fourth induction media may comprise aRB27 Media, doxycycline, XAV939, hBMP4, SCF, and EGF.

The ‘aRB27 Media’ used herein comprises Advanced RPMI, B-27™ Supplement, minus vitamin A (Thermo Fisher) or plus vitamin A, GlutaMAX™ Supplement (Thermo Fisher), non-essential amino acids (NEAA), Primocin® (a broad-spectrum antibiotic), and Y-27632 (a small molecule ROCK inhibitor).

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

Activin-A is a dimeric glycoprotein, which belongs to the transforming growth factor-β (TGF-β) family.

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

PD 173074 is a selective FGFR1 and FGFR3 inhibitor (IC50 values are ~5 nM, ~21.5 nM, ~100 nM, ~17600 nM and ~19800 nM for FGFR3, FGFR1, VEGFR2, PDGFR and c-Src respectively, and > 50000 nM for EGFR, InsR, MEK and PKC).

XAV939 is a potent, small molecule inhibitor of tankyrase (TNKS) 1 and 2 (IC_{50} = 11 and 4 nM, respectively) (Huang et al.). By inhibiting TNKS activity, XAV939 increases the protein levels of the axin-GSK3β complex and promotes the degradation of β-catenin in SW480 cells (Huang et al.), thereby inhibiting WNT pathway downstream actions.

Therapeutic Compositions and Method of Use

The present disclosure provides, in some embodiments, therapeutic compositions comprising the astrocyte-like cells, cytotoxic T-cell-like cells, hepatocyte-like cells, regulatory T-cell-like cells, B cell-like cells, and/or microglia-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. The subject may have a disease, disorder, or symptoms of a disease associated with astrocyte dysfunction, cytotoxic T-cell dysfunction, hepatocyte dysfunction, regulatory T-cell, B cell, or microglia dysfunction.

The compositions may be administered to a subject in a therapeutically effective amount. The term " "therapeutically effective amount" refers to the amount of cell material 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.

Methods of Identifying Transcription Factors for Differentiation of Stems Cells into Target Cell Types

Aspects of the present disclosure relate to a method for identifying transcription factors that are able to differentiate a stem cell into a target cell type (e.g., an astrocyte, a cytotoxic T-cell, a hepatocyte, a regulatory T-cell, a B cell, or a microglial cell), comprising: (i) analyzing epigenetics data for a target cell type to identify genomic sites that are available for binding of a transcription factor and generating a first pool of transcription factors; (ii) analyzing transcriptomic data for the target cell type to identify expression levels of the transcription factors associated with the genomic sites that are available for binding identified in step (i) and generating second pool of transcription factors; (iii) using a first statistical method to filter background data and identify transcription factors that are present in the first pool of transcription factors and the second pool of transcription factors and generating a third pool of transcription factors, wherein the third pool of transcription factors; (iv) using a second statistical method to determine the statistical significance of the transcription factors in the third pool of transcription factors; and (v) repeating steps (i)-(iv) one or more times to iteratively refine the third pool of transcription factors.

In some embodiments, the epigenetics data provides information related to whether genomic chromatin is open or closed. In some embodiments, the epigenetics data is produced by DNAse-seq, ATAC-seq, or ChIP-seq.

In some embodiments, the transcriptomic data provides information related to whether there are more transcripts of the transcription factor in the target cell type than in a non-target cell type. In some embodiments, the transcriptomic data is produced by RNA-seq.

In some embodiments, the first statistical method is linear regression algorithm. In some embodiments, the first statistical method is a logistic regression algorithm. In some embodiments, the first statistical method is a L1-regularized logistic regression model (LASSO). In some embodiments, the background data is associated with transcription factors that are not expressed in the target cell type at a higher expression level than in the non-target cell type. In some embodiments, the second statistical method is a log-likelihood ratio test.

In some embodiments, the method further comprises transfecting transcription factors of the third pool into a stem cell. In some embodiments, the method further comprises inducing differentiation of the stem cell into the target cell type. In some embodiments, the method further comprises analyzing the target cell type to identify additional transcription factors associated with the target cell type. In some embodiments, the method further comprises using data from the target cell type to further refine the steps of the method.

In some embodiments, the target cell type is an astrocyte, a cytotoxic T-cell, a hepatocyte, a regulatory T-cell, a B cell, or a microglial cell. In some embodiments, differentiation of stem cells using one or more of the transcription factors in the third pool results in production of the target cell type in no more than 6 days.

Aspects of the present disclosure relate to a method for generating a transcription factor screening pool comprising: using at least one computer hardware processer to perform: accessing at least one statistical model relating one or more input transcription factors to differentiation efficiency of a cell having the one or more input transcription factors; obtaining differentiation efficiency information for the one or more input transcription factors; generating, using the at least one statistical model and the differentiation efficiency information, a transcription factor pool having transcription factors that are predicted to differentiate the cell into a target cell type in accordance with the differentiation efficiency information.

In some embodiments, the at least one statistical model correlates chromatin accessibility data and transcriptomics data to make initial predictions relating the one or more input transcription factors to differentiation efficiency of the cell having the one or more input transcription factors. In some embodiments, the at least one statistical model distinguishes open chromatin data from background data. In some embodiments, the open chromatin data is associated with the target cell type.

In some embodiments, the method further comprises identifying an initial set of transcription factor motifs positively correlated with the open chromatin data by using a statistical coefficient trained to distinguish the open chromatin data from the background data.

In some embodiments, the differentiation efficiency information corresponds to a mode of a distribution of differentiation efficiency data used to train the at least one statistical model.

In some embodiments, the at least one statistical model was trained using measured differentiation efficiency values having a multimodal distribution with modes, and the differentiation efficiency information corresponds to a mode of the multimodal distribution with the highest value.

In some embodiments, the transcription factors of the transcription factor pool have predicted differentiation efficiency within a distribution centered at the mode of the multimodal distribution with the highest value.

In some embodiments, the differentiation efficiency information corresponds to a Gaussian distribution centered at a mode of a distribution for differentiation efficiency data used to train the at least one statistical model. In some embodiments, the differentiation efficiency information corresponds to a high differentiation efficiency component of a distribution of differentiation efficiency values for transcription factors.

In some embodiments, generating the transcription factor pool further comprises: generating an initial pool of transcription factors; using transcription factors in the initial pool as input to the at least one statistical model to obtain values for differentiation efficiency; selecting, based on the values for differentiation efficiency and the differentiation efficiency information, one or more of the transcription factors in the initial pool to include in the transcription factor pool.

In some embodiments, wherein the at least one statistical model comprises at least one regression model. In some embodiments, the at least one statistical model comprises at least one neural network.

In some embodiments, the at least one statistical model has a recurrent neural network architecture. In some embodiments, the at least one statistical model comprises a L1-regularized logistic regression model (LASSO). In some embodiments, the at least one statistical model comprises a log-likelihood ratio test.

Aspects of the present disclosure relate to a system comprising: at least one hardware processor; and at least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one hardware processor, cause the at

least one hardware processor to perform: accessing at least one statistical model relating one or more input transcription factors to differentiation efficiency of a cell having the one or more input transcription factors; obtaining differentiation efficiency information for transcription factors, wherein the differentiation efficiency information corresponds to a mode of a distribution for differentiation efficiency data used to train the at least one statistical model; and

generating, using the at least one statistical model and the differentiation efficiency information, a transcription factor pool having transcription factors with predicted differentiation efficiency in accordance with the differentiation efficiency information.

In some embodiments, the target cell type is a Type II astrocyte, cytotoxic T-cell, regulatory T-cell, hepatocyte, B cell, or microglial cell.

Additional Embodiments

Additional embodiments of the disclosure are provided in the following numbered paragraphs:

1. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding ERG, EGR1, FLI1, FOSB, or any combination thereof.

2. The PSC of paragraph 1, comprising the engineered polynucleotide comprising an open reading frame encoding ERG.

3. The PSC of paragraph 1 or 2, comprising the engineered polynucleotide comprising an open reading frame encoding EGR1.

4. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding FLI1.

5. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding FOSB.

6. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses ERG, EGR1, FLI1, FOSB, or any combination thereof.

7. 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.

8. The PSC of paragraph 7, wherein the heterologous promoter is an inducible promoter.

9. A pluripotent stem cell (PSC) comprising: a protein selected from ERG, EGR1, FLI1, and FOSB, wherein the protein is overexpressed.

10. The PSC of paragraph 9, wherein the PSC expresses or overexpresses: ERG, EGR1, FLI1, FOSB, or any combination thereof.

11. The PSC of any one of paragraphs 1-10, wherein the PSC is a human PSC.

12. The PSC of any one of paragraphs 1-11, wherein the PSC is an induced PSC (iPSC).

13. The PSC of any one of the preceding paragraphs, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ERG, EGR1, FLI1, and FOSB, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

14. The PSC of any one of the preceding paragraphs, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected

from ERG, EGR1, FLI1, and FOSB, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

15. A composition comprising: a population of the PSC of any one of the preceding paragraphs.

16. The composition of paragraph 15, wherein the population comprises at least 2500/cm^2 of the PSC.

17. 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 ERG, EGR1, FLI1, and FOSB to produce astrocyte-like cells.

18. The method of paragraph 17, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ERG.

19. The method of paragraph 17 or 18, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding EGR1.

20. The method of any one of paragraphs 17-19, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding FLI1.

21. The method of any one of paragraphs 17-20, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding FOSB.

22. The method of any one of paragraphs 17-21, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

23. The method of paragraph 22, wherein the heterologous promoter is an inducible promoter.

24. The method of paragraph 23, wherein the inducible promoter is a chemically-inducible promoter.

25. The method of paragraph 24, wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

26. The method of any one of the preceding paragraphs, wherein the population comprises $1x10^2$ -$1x10^7$ PSCs.

27. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for at least 1 day.

28. The method of paragraph 27, wherein the population of PSCs is cultured for about 3-6 days,

29. The method of paragraph 28, wherein the population of PSCs is cultured for no more than 6 days.

30. The method of any one of the preceding paragraphs, wherein the astrocyte-like cells are CD44$^+$ and A2B5$^+$.

31. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

32. The PSC of paragraph 31, comprising the engineered polynucleotide comprising an open reading frame encoding ZBTB1.

33. The PSC of paragraph 31 or 32, comprising the engineered polynucleotide comprising an open reading frame encoding RUNX3.

34. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding RELA.

35. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding NRF1.

36. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding ERF.

37. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding SP4.

38. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

39. 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.

40. The PSC of paragraph 39, wherein the heterologous promoter is an inducible promoter.

41. A pluripotent stem cell (PSC) comprising: a protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4, wherein the protein is overexpressed.

42. The PSC of paragraph B11, wherein the PSC expresses or overexpresses: ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

43. The PSC of any one of paragraphs 31-42, wherein the PSC is a human PSC.

44. The PSC of any one of paragraphs 31-43, wherein the PSC is an induced PSC (iPSC).

45. The PSC of any one of the preceding paragraphs, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

46. The PSC of any one of the preceding paragraphs, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

47. A composition comprising: a population of the PSC of any one of the preceding paragraphs.

48. The composition of paragraph B17, wherein the population comprises at least 2500/cm^2 of the PSC.

49. 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 ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 to produce cytotoxic T-cell-like cells.

50. The method of paragraph 49, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ZBTB1.

51. The method of paragraph 49 or 50, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RUNX3.

52. The method of any one of paragraphs 49-51, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RELA.

53. The method of any one of paragraphs 49-52, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NRF1.

54. The method of any one of paragraphs 49-53, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ERF.

55. The method of any one of paragraphs 49-54, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding SP4.

56. The method of any one of paragraphs 49-55, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

57. The method of paragraph 56, wherein the heterologous promoter is an inducible promoter.

58. The method of paragraph 57, wherein the inducible promoter is a chemically-inducible promoter.

59. The method of paragraph 58, wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

60. The method of any one of the preceding paragraphs, wherein the population comprises $1x10^2$ -$1x10^7$ PSCs.

61. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for at least 1 day.

62. The method of paragraph 61, wherein the population of PSCs is cultured for about 3-6 days,

63. The method of paragraph 62, wherein the population of PSCs is cultured for no more than 6 days.

64. The method of any one of the preceding paragraphs, wherein the cytotoxic T-cell-like cells are $CD3^+$ and $CD8^+$.

65. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding HNF4G, TEAD4, RFX3, or any combination thereof.

66. The PSC of paragraph 65, comprising the engineered polynucleotide comprising an open reading frame encoding HNF4G.

67. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding TEAD4.

68. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding RFX3.

69. The PSC of any one of the preceding paragraphs, further comprising an engineered polynucleotide comprising an open reading frame encoding HNF4A.

70. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses HNF4G, TEAD4, RFX3, or any combination thereof.

71. The PSC of paragraph 70, wherein the PSC further expresses or overexpresses HNF4A.

72. 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.

73. The PSC of paragraph 72, wherein the heterologous promoter is an inducible promoter.

74. A pluripotent stem cell (PSC) comprising: a protein selected from HNF4G, TEAD4, and RFX3, wherein the protein is overexpressed.

75. The PSC of paragraph 74, wherein the PSC expresses or overexpresses: HNF4G, TEAD4, RFX3, or any combination thereof.

76. The PSC of any one of paragraphs 65-75, wherein the PSC is a human PSC.

77. The PSC of any one of paragraphs 65-76, wherein the PSC is an induced PSC (iPSC).

78. The PSC of any one of the preceding paragraphs, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from HNF4G, HNF4A, TEAD4, and RFX3, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

79. The PSC of any one of the preceding paragraphs, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from HNF4G, TEAD4, and RFX3, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

80. A composition comprising: a population comprising the PSC of any one of the preceding paragraphs.

81. The composition of paragraph 80, wherein the population comprises at least 2500/cm^2 of the PSC.

82. 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 HNF4G, TEAD4, and RFX3 to produce hepatocyte-like cells.

83. The method of paragraph 82, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding HNF4G.

84. The method of any one of paragraphs 82-83, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding TEAD4.

85. The method of any one of paragraphs 82-84, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RFX3.

86. The method of any one of the preceding paragraphs, wherein the PSCs of the expanded population further comprise an engineered polynucleotide comprising an open reading frame encoding HNF4A.

87. The method of any one of paragraphs 82-86, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

88. The method of paragraph 87, wherein the heterologous promoter is an inducible promoter.

89. The method of paragraph 88, wherein the inducible promoter is a chemically-inducible promoter.

90. The method of paragraph 89, wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

91. The method of any one of the preceding paragraphs, wherein the population comprises $1x10^2$ -$1x10^7$ PSCs.

92. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for at least 1 day.

93. The method of paragraph 92, wherein the population of PSCs is cultured for about 3-6 days,

94. The method of paragraph 93, wherein the population of PSCs is cultured for no more than 6 days.

95. The method of any one of the preceding paragraphs, wherein the hepatocyte-like cells are $CD184^+$ and $ASGPR1^+$.

96. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof.

97. The PSC of paragraph 96, comprising the engineered polynucleotide comprising an open reading frame encoding ETS1.

98. The PSC of paragraph 96 or 97, comprising the engineered polynucleotide comprising an open reading frame encoding ETV3.

99. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding GABPA.

100. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding KLF9.

101. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding NFKB1.

102. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof.

103. 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.

104. The PSC of paragraph 103, wherein the heterologous promoter is an inducible promoter.

105. A pluripotent stem cell (PSC) comprising: a protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, wherein the protein is overexpressed.

106. The PSC of paragraph 105, wherein the PSC expresses or overexpresses: ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof.

107. The PSC of any one of paragraphs 96-106, wherein the PSC is a human PSC.

108. The PSC of any one of paragraphs 96-107, wherein the PSC is an induced PSC (iPSC).

109. The PSC of any one of the preceding paragraphs, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

110. The PSC of any one of the preceding paragraphs, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

111. A composition comprising: a population of the PSC of any one of the preceding paragraphs.

112. The composition of paragraph 111, wherein the population comprises at least 2500/cm^2 of the PSC.

113. 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 ETS1, ETV3, GABPA, KLF9, and NFKB1 to produce regulatory T-cell-like cells.

114. The method of paragraph 113, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ETS1.

115. The method of paragraph 113 or 114, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ETV3.

116. The method of any one of paragraphs 113-115, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding GABPA.

117. The method of any one of paragraphs 113-116, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding KLF9.

118. The method of any one of paragraphs 113-117, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding NFKB1.

119. The method of any one of paragraphs 113-118, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

120. The method of paragraph 119, wherein the heterologous promoter is an inducible promoter.

121. The method of paragraph 120, wherein the inducible promoter is a chemically-inducible promoter.

122. The method of paragraph 121, wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

123. The method of any one of the preceding paragraphs, wherein the population comprises $1x10^2$-$1x10^7$ PSCs.

124. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for at least 1 day.

125. The method of paragraph 124, wherein the population of PSCs is cultured for about 3-6 days,

126. The method of paragraph 125, wherein the population of PSCs is cultured for no more than 6 days.

127. The method of any one of the preceding paragraphs, wherein the regulatory T-cell-like cells are $CD3^+$ and $CD25^+$.

128. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof.

129. The PSC of paragraph 128, comprising the engineered polynucleotide comprising an open reading frame encoding EBF1.

130. The PSC of paragraph 128 or 129, comprising the engineered polynucleotide comprising an open reading frame encoding ZBTB1.

131. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding RELA.

132. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding NRF1.

133. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding REL.

134. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof.

135. 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.

136. The PSC of paragraph 135, wherein the heterologous promoter is an inducible promoter.

137. A pluripotent stem cell (PSC) comprising: a protein selected from EBF1, ZBTB1, RELA, NRF1, and REL, wherein the protein is overexpressed.

138. The PSC of paragraph 137, wherein the PSC expresses or overexpresses: EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof.

139. The PSC of any one of paragraphs 128-138, wherein the PSC is a human PSC.

140. The PSC of any one of paragraphs 128-139, wherein the PSC is an induced PSC (iPSC).

141. The PSC of any one of the preceding paragraphs, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from EBF1, ZBTB1, RELA, NRF1, and REL, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

142. The PSC of any one of the preceding paragraphs, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from EBF1, ZBTB1, RELA, NRF1, and REL, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

143. A composition comprising: a population of the PSC of any one of the preceding paragraphs.

144. The composition of paragraph 143, wherein the population comprises at least 2500/cm^2 of the PSC.

145. 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 EBF1, ZBTB1, RELA, NRF1, and REL to produce B cell-like cells.

146. The method of paragraph 145, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding EBF1.

147. The method of paragraph 145 or 146, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ZBTB1.

148. The method of any one of paragraphs 145-147, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RELA.

149. The method of any one of paragraphs 145-148, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding
NRF1.

150. The method of any one of paragraphs 145-149, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding REL.

151. The method of any one of paragraphs 145-150, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

152. The method of paragraph 151, wherein the heterologous promoter is an inducible promoter.

153. The method of paragraph 152, wherein the inducible promoter is a chemically-inducible promoter.

154. The method of paragraph 153, wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

155. The method of any one of the preceding paragraphs, wherein the population comprises 1x10^2-1x10^7 PSCs.

156. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for at least 1 day.

157. The method of paragraph 156, wherein the population of PSCs is cultured for about 3-6 days,

158. The method of paragraph 157, wherein the population of PSCs is cultured for no more than 6 days.

159. The method of any one of the preceding paragraphs, wherein the B cell-like cells are CD19$^+$ and CD27$^+$.

160. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding SPI1, ZBTB1, RELA, STAT2, or any combination thereof.

161. The PSC of paragraph 160, comprising the engineered polynucleotide comprising an open reading frame encoding SPI1.

162. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding ZBTB1.

163. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding RELA.

164. The PSC of any one of the preceding paragraphs, comprising the engineered polynucleotide comprising an open reading frame encoding STAT2.

165. The PSC of any one of the preceding paragraphs, wherein the PSC expresses or overexpresses SPI1, ZBTB1, RELA, STAT2, or any combination thereof.

166. 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.

167. The PSC of paragraph 166, wherein the heterologous promoter is an inducible promoter.

168. A pluripotent stem cell (PSC) comprising: a protein selected from SPI1, ZBTB1, RELA, and STAT2, wherein the protein is overexpressed.

169. The PSC of paragraph 168, wherein the PSC expresses or overexpresses: SPI1, ZBTB1, RELA, STAT2, or any combination thereof.

170. The PSC of any one of paragraphs 160-169, wherein the PSC is a human PSC.

171. The PSC of any one of paragraphs 160-170, wherein the PSC is an induced PSC (iPSC).

172. The PSC of any one of the preceding paragraphs, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from SPI1, ZBTB1, RELA, and STAT2, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

173. The PSC of any one of the preceding paragraphs, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from SPI1, ZBTB1, RELA, and STAT2, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

174. A composition comprising: a population comprising the PSC of any one of the preceding paragraphs.

175. The composition of paragraph 174, wherein the population comprises at least 2500/cm^2 of the PSC.

176. 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 SPI1, ZBTB1, RELA, and STAT2 to produce microglia-like cells.

177. The method of paragraph 176, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding SPI1.

178. The method of any one of paragraphs 176-177, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding ZBTB1.

179. The method of any one of paragraphs 176-178, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding RELA.

180. The method of any one of the preceding paragraphs, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding STAT2.

181. The method of any one of paragraphs 176-180, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter.

182. The method of paragraph 181, wherein the heterologous promoter is an inducible promoter.

183. The method of paragraph 182, wherein the inducible promoter is a chemically-inducible promoter.

184. The method of paragraph 183, wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

185. The method of any one of the preceding paragraphs, wherein the population comprises 1x10^2 -1x10^7 PSCs.

186. The method of any one of the preceding paragraphs, wherein the population of PSCs is cultured for at least 1 day.

187. The method of paragraph 186, wherein the population of PSCs is cultured for about 3-6 days,

188. The method of paragraph 187, wherein the population of PSCs is cultured for no more than 6 days.

189. The method of any one of the preceding paragraphs, wherein the microglia-like cells are CD11b+ and CX3CR1+.

190. A method, comprising:

(i) analyzing epigenetics data for a target cell type to identify genomic sites that are available for binding of a transcription factor and generating a first pool of transcription factors;

(ii) analyzing transcriptomic data for the target cell type to identify expression levels of the transcription factors associated with the genomic sites that are available for binding identified in step (i) and generating a second pool of transcription factors;

(iii) using a first statistical method to filter background data and identify transcription factors that are present in the first pool of transcription factors and the second pool of transcription factors and generating a third pool of transcription factors, wherein the third pool of transcription factors comprises transcription factors that are in both the first pool and the second pool;

(iv) using a second statistical method to determine the statistical significance of the transcription factors in the third pool of transcription factors; and

(v) repeating steps (i)-(iv) one or more times to iteratively refine the third pool of transcription factors.

191. The method of paragraph 190, wherein the epigenetics data provides information related to whether genomic chromatin is open or closed.

192. The method of paragraph 190 or paragraph 191, wherein the epigenetics data is produced by DNAse-seq, ATAC-seq, or ChIP-seq.

193. The method of any one of the preceding paragraphs, wherein the transcriptomic data provides information related to whether there are more transcripts of the transcription factor in the target cell type than in a non-target cell type.

194. The method of any one of the preceding paragraphs, wherein the transcriptomic data is produced by RNA-seq.

195. The method of any one of the preceding paragraphs, wherein the first statistical method is linear regression algorithm.

196. The method of any one of the preceding paragraphs, wherein the first statistical method is a logistic regression algorithm.

197. The method of any one of the preceding paragraphs, wherein the first statistical method is a L1-regularized logistic regression model (LASSO).

198. The method of any one of the preceding paragraphs, wherein the background data is associated with transcription factors that are not expressed in the target cell type at a higher expression level than in the non-target cell type.

199. The method of any one of the preceding paragraphs, wherein the second statistical method is a log-likelihood ratio test.

200. The method of any one of the preceding paragraphs, further comprising transfecting transcription factors of the third pool into a stem cell.

201. The method of paragraph 200, further comprising inducing differentiation of the stem cell into the target cell type.

202. The method of paragraph 201, further comprising analyzing the target cell type to identify additional transcription factors associated with the target cell type.

203. The method of any one of the preceding paragraphs, further comprising using data from the target cell type to further refine the steps of paragraph 190.

204. The method of any one of the preceding paragraphs, wherein the target cell type is an astrocyte, a cytotoxic T-cell, a hepatocyte, a regulatory T-cell, a B cell, or a microglial cell.

205. The method of any one of the preceding paragraphs, wherein differentiation of stem cells using one or more of the transcription factors in the third pool results in production of the target cell type in no more than 6 days.

206. A method for generating a transcription factor screening pool comprising:
using at least one computer hardware processer to perform:

accessing at least one statistical model relating one or more input transcription factors to differentiation efficiency of a cell having the one or more input transcription factors;

obtaining differentiation efficiency information for the one or more input transcription factors;

generating, using the at least one statistical model and the differentiation efficiency information, a transcription factor pool having transcription factors that are predicted to differentiate the cell into a target cell type in accordance with the differentiation efficiency information.

207. The method of paragraph 206, wherein the at least one statistical model correlates chromatin accessibility data and transcriptomics data to make initial predictions relating the one or more input transcription factors to differentiation efficiency of the cell having the one or more input transcription factors.

208. The method of paragraph 206 or paragraph 207, wherein the at least one statistical model distinguishes open chromatin data from background data.

209. The method of paragraph 208, wherein the open chromatin data is associated with the target cell type.

210. The method of any one of the preceding paragraphs, further comprising identifying an initial set of transcription factor motifs positively correlated with the open chromatin data by using a statistical coefficient trained to distinguish the open chromatin data from the background data.

211. The method of any one of the preceding paragraphs, wherein the differentiation efficiency information corresponds to a mode of a distribution of differentiation efficiency data used to train the at least one statistical model.

212. The method of any one of the preceding paragraphs, wherein the at least one statistical model was trained using measured differentiation efficiency values having a multimodal distribution with modes, and the differentiation efficiency information corresponds to a mode of the multimodal distribution with the highest value.

213. The method of any one of the preceding paragraphs, wherein the transcription factors of the transcription factor pool have predicted differentiation efficiency within a distribution centered at the mode of the multimodal distribution with the highest value.

214. The method of any one of the preceding paragraphs, wherein the differentiation efficiency information corresponds to a Gaussian distribution centered at a mode of a distribution for differentiation efficiency data used to train the at least one statistical model.

215. The method of any one of the preceding paragraphs, wherein the differentiation efficiency information corresponds to a high differentiation efficiency component of a distribution of differentiation efficiency values for transcription factors.

216. The method of any one of the preceding paragraphs, wherein generating the transcription factor pool further comprises:

generating an initial pool of transcription factors;

using transcription factors in the initial pool as input to the at least one statistical model to obtain values for differentiation efficiency;

selecting, based on the values for differentiation efficiency and the differentiation efficiency information, one or more of the transcription factors in the initial pool to include in the transcription factor pool.

217. The method of any one of the preceding paragraphs, wherein the at least one statistical model comprises at least one regression model.

218. The method of any one of the preceding paragraphs, wherein the at least one statistical model comprises at least one neural network.

219. The method of any one of the preceding paragraphs, wherein the at least one statistical model has a recurrent neural network architecture.

220. The method of any one of the preceding paragraphs, wherein the at least one statistical model comprises a L1-regularized logistic regression model (LASSO).

221. The method of any one of the preceding paragraphs, wherein the at least one statistical model comprises a log-likelihood ratio test.

222. A system comprising:

at least one hardware processor; and

at least one non-transitory computer-readable storage medium storing processor-executable instructions that, when executed by the at least one hardware processor, cause the at

least one hardware processor to perform:

accessing at least one statistical model relating one or more input transcription factors

to differentiation efficiency of a cell having the one or more input transcription factors;obtaining differentiation efficiency information for transcription factors, wherein the differentiation efficiency information corresponds to a mode of a distribution for differentiation efficiency data used to train the at least one statistical model; and generating, using the at least one statistical model and the differentiation efficiency information, a transcription factor pool having transcription factors with predicted differentiation efficiency in accordance with the differentiation efficiency information.

223. The method of any one of the preceding paragraphs, wherein the target cell type is a Type II astrocyte, cytotoxic T-cell, regulatory T-cell, hepatocyte, B cell, or microglial cell.

EXAMPLES

Stem cells are the progenitor cells of all differentiating multi-cellular organisms. In principle, it is possible to differentiate these cells into any other type of cell, which can then be used for many different possible therapeutic or diagnostic applications. The creation of induced pluripotent stem cells (iPSCs) has enabled scientists to explore the derivation of many types of cells. While there are diverse general approaches for cell-fate engineering, one of the fastest and most efficient approaches is transcription factor (TF) over-expression. Over-expression of specific combinations of TFs is often a reliable method to differentiate stem cells, but since there are at least 1732 transcription factors in the human genome, selecting the right combination to differentiate iPSCs directly into other cell-types is a difficult task. Here were describe a machine-learning (ML) pipeline, called *CellCartographer*, for using chromatin accessibility data to design multiplex TF pooled-screens for cell type conversions.

We then describe a barcoded bulk RNA-seq method for refining sets of TFs using iterative NGS experiments. We validate this method by differentiating iPSCs into twelve diverse cell types at low efficiency in primary screens and then iteratively refining our differentiation strategy to achieve high efficiency differentiation for six of these cell types, originating from all germ layers in high efficiency in ≤ 6 days. Finally, we functionally characterized engineered iPSC-derived cytotoxic T-cells (iCytoT), regulatory T-cells (iTreg), type II astrocytes (iAstII), and hepatocytes (iHep) to validate fast, robust, and functionally accurate differentiation of stem cells into cell types potentially useful for downstream therapeutic and diagnostic pipelines.

Example 1 - Machine learning for determining TF sub-libraries

It is not known exactly how many human cell types exist, but current estimates put the number in the hundreds, all originating from a single ‘totipotent’ embryonic stem cell. Since the creation of induced pluripotent stem cells (iPSCs), scientists have been trying to recreate differentiation of iPSCs into all of these other types of cells and combine them into tissues or tissue-like structures (a.k.a. ‘cell-fate engineering’). This goal seems feasible given that it has been generally accepted that iPSCs are functionally identical to embryonic stem cells (ESCs). To perform cell-fate engineering, a litany of approaches has been employed that fall into three general categories: (1) application of growth factors into media in either 2D or 3D cell culture, (2) modifications to cell matrix and plate surface conditions, and (3) over-expression of transcription factors (TFs). Generally speaking, the first two categories of approaches have been effective in differentiating many different cell types simultaneously — this makes sense because the general idea is to recapitulate aspects of natural development *in vitro*, where many cell types would differentiate in unison with each other. The drawbacks of these first two approaches are threefold: first, these protocols typically take a long time (often many weeks); second, the efficiency in converting to a single type of cell is often poor; and third, reproducibility across these experiments remains a large challenge. Because TF-based approaches directly manipulate the epigenetic landscape of individual cells, they have proved to address these three issues to a great extent.

While TF-based approaches have been fruitful, the task of identifying the correct TFs for a fast, efficient, and robust 37 cell conversion remains a challenging problem. There are two general ways to go about this research process: (1) an exhaustive literature search for potentially relevant transcription factors for a desired cell type and identify successful combinations via trial-and-error or (2) to use computational tools to predict TFs. While iPSCs were created through a systematic version of the former, this process does not scale — it is very laborious, requires deep expertise of the cell types being converted, and can only account for previously studied TFs associated with specific cell types. The latter approach has been successful in recent years and can be used as a more general approach in minimizing time required to identify effective conversion factors. While these tools have demonstrated some predictive power, they have key limitations: (1) they cannot account for experimental details such as DNA copy count, clonality (i.e. polyclonal v. monoclonal cell lines), expression method, or cell culture conditions; (2) they generally only provide a single combination of TFs for a cell-type conversion that cannot be iteratively revised; (3) most rely on gene expression data exclusively, as opposed to other relevant data types like epigenetics; and (4) it is not easy to employ the tools for new or rare cell types with very limited data. Moreover, while machine-learning (ML) pipelines 49 have yielded impressive results in various areas of molecular biology to date, currently no tools use ML to generate screens for cell-fate engineering.

To address these gaps, we built an epigenetics-based, ML-driven pooled screening tool for engineering cell-fate, called *CellCartographer*. *CellCartographer* uses next generation sequencing based readouts of chromatin accessibility (e.g., DNase-seq, ATAC-seq, ChIP-seq) and transcription (RNA-seq) to predict TFs to be correlated with cell-type identity. Using the predictions made by *CellCartographer*, we can define multiplex pooled-screens of TFs for over-expression, which allows us to explore many experimental variables such as variable stable expression quantities, genomic integration copy count and location, and culture conditions with the option to add more nuance depending on experimental conditions. *CellCartographer* gives outputs agnostic to starting cell type because it has been demonstrated that the same (or similar) TF set can be used to differentiate cells from a variety of originating cell types and because the iterative engineering process from this starting *in silico* screen should be able to accommodate for these differences. We demonstrate how the *CellCartographer* predictions are sufficient for differentiating small sub-populations of cell-surface marker-positive cells for twelve target cell-type samples from all three germ layers. Then, we show how we can use bulk-RNA sequencing to refine the original TF predictions and zoom in on minimal TF combo sets to differentiate stem cells for six cell types from all three germ layers. Once a sufficiently-high percentage of polyclonal cell line differentiation was created, we showed that isolating clones from these populations results in the creation of high-performance clonal lines. Finally, we functionally characterized robust clonal lines of differentiation-inducible iPSC lines for each of the three germ layers: regulatory T-cells (iTreg) and cytotoxic T-cells (iCytoTs) - mesoderm, hepatocytes (iHep) - endoderm, and type-II astrocytes (iAstII) - ectoderm) to validate that the cells are functionally *in vitro* and molecularly accurate. We were able to differentiate four cell types using novel combinations of TFs in as little as 6 days. Importantly, our derivation of iTRegs and iCytoTs may considerably accelerate the investigation of T-cell biology.

As many TFs are controlled for activity by nuclear localization as well as expression, the prevalence of TF motifs at accessible chromatin regions is a better indicator of TF activity and importance for cell identity than abundance of RNA expression. And so, the *CellCartographer* model leverages chromatin accessibility data to make initial predictions of TFs for differentiating towards a given cell type. After initial TF predictions are made, TF transcript levels are used to exclude TFs that are not expressed. The *CellCartographer* pipeline can leverage a variety of assays for chromatin accessibility and transcriptomics to predict a set of TFs for a target cell type, which can then be tested in a pooled screen (FIG. 1B). To broaden the functionality of *CellCartographer*, input data can be either manually uploaded or automatically queried and downloaded from the ENCODE database or GEO. Valid data types for chromatin accessibility include DNase-seq, ATAC-seq, ChIP-seq (H3K27Ac, H3K4me1, and H3K4me3). For transcriptomics data, *CellCartographer* accepts most RNA-seq assays including ribo-depleted, total RNA or polyA RNA seq.

Since the number of TFs in the TFome (1732) with characterized binding sites (891), yields 2^{891} possible outcomes (FIG. 1A). In a full library screen, the chance of observing a correct combination of TFs that differentiate a target cell type in a full library screen with 10^6 starting cells would be unlikely (on the order of 1 in 10^{167}). And so, we reasoned that the number of starting cells and the number of possible combinations formed from the set of transfected TFs should be similar (i.e. 2^{nTFs} m_{cells}). In our case, we nucleofected 10^6 cells per experiment and the screening pools contained approximately 16 plasmids containing a transcription factors driven by a doxycycline-inducible promoter (FIG. 1D). Each TF cassette is integrated randomly within each cell from zero to *n* times, allowing us to explore a large parameter space of DNA integration location and resulting expression amounts of each TF in combination.

In order to make TF predictions for each cell type, we begin by training a logistic regression classifier model to distinguish between open chromatin regions and a set of background genomic loci using the known DNA TF binding motifs drawn from the JASPAR database (FIG. 1C). By using a non-redundant set of motifs we can mitigate the effect of multiple-collinearity on our classifier model. By examining the sign of the model coefficients, we can determine whether the presence of a motif is negatively or positively correlated with open chromatin. From the set of TF motifs positively correlated with chromatin accessibility, we determine the most significant ones using the likelihood ratio test, which is an *in silico* analog of a mutagenesis experiment; the performance of the model is compared to a perturbed model that has been blinded to the presence of a given motif. We select the most significant motifs that are positively correlated with binding and the top 16 corresponding genes that are expressed. Finally, we exclude constitutively active TFs before screening the selected TFs (data not shown). Using publicly available DNase-seq data from ENCODE, we applied our approach on several cell types with simple known combinations of one to two lineage determining TFs and confirmed that these TFs appear in the top TFs predicted by *CellCartographer*. (FIG. 1E).

To computationally validate our model on a larger scale, we applied *CellCartographer* to 34 primary cells types and 29 tissue types. We found that each TF DNA binding motifs strongly correlated with chromatin accessibility had different behaviors in each cell type and tissue type (FIG. 2C). Given that related cell types have similar transcriptional profiles (FIG. 2A), we reasoned that they may also have similar TFs correlated with open chromatin that drive transcriptional profiles. To visualize the similarity between transcriptional profiles, we calculated the pairwise Pearson correlation between the gene expression values of each cell type (log RPKM values) and used multidimensional scaling to embed each cell type in way that respects the pairwise similarity between cell types; using the Spearman correlation model coefficient for each TF, we can also visualize the similarity of TF motifs correlated with open chromatin. We observe that cell types that group together when considering similarities in transcriptional profiles such as adaptive immune cells (e.g., B cells and T-cells) and progenitor cell types (H1-hESC, GM23338, and neural stem progenitor) tend to look similar from the perspective of TF motifs correlated with open chromatin (FIG. 2B).

Example 2 - Primary pooled TF screens for differentiation

To demonstrate that our pooled screening method could be generally applied to any cell type of interest, we identified cell types from each human germ layer and screened TFs combinations to identify populations of cells that came up positive for canonical markers. Specifically, we generated TF pools for: Mesoderm — T-cells (subtypes cytotoxic, delta-gamma, and regulatory), B cells, macrophages, epithelial cells (subtypes kidney, bronchial, and mammary), and osteoblasts; Endoderm — hepatocytes; Ectoderm — type II astrocytes; and Yolk Sac — microglia. For each cell type, we designed two TF pools for each cell type using *CellCartographer* — one pool containing TFs with expression level ≥ 1 RPKM and another containing TFs with expression level ≥ 4 RPKM (data not shown). We then prepared mixed DNA pools of equal concentration of each TF and nucleofected and screened iPSCs (FIG. 1D). We found that the percentage of cells appearing positive in most cases was very small, but ranged from 0.05% (Regulatory T-cells) to 17.64% (B cells), although in almost all cases, the positive population was <1% (FIG. 3). Thus, it appeared that all samples yielded at least a small population of differentiated cells that can be sequenced to determine which TFs from the TFome were present.

From this set of diverse screened cell types, we decided to iteratively refine a set of six that had high clinical relevance — cytotoxic T-cells, Regulatory T-cells, B cells, hepatocytes, type II astrocytes, and microglia. A comparison of the top motifs positively correlated with open chromatin for these six cell types is shown in FIG. 2D; the screening pools for each of these six cell types are not shown. It should be noted that at this step, the selection of specific surface markers biases the downstream analysis and refinement. For example, although TF pools for astrocytes were determined from data based on generic astrocytes (type I or type II), our selection of A2B5 as a surface marker in combination with CD44 selected specifically for type II astrocytes. In the case of the epithelial sub-types, there was some uncertainty of the ideal cell surface markers to use since CD24 was unexpectedly present in the stem cells and stem cells are partially epithelial in quality and express EpCAM to a slightly lesser degree than differentiated epithelial types. Nonetheless, from the pooled screens we were able to sort at least 1000 double-positive cells from each large population for bulk RNA-sequencing. We lysed the sorted cells, prepared sequencing libraries, and amplified the barcoded regions of the TFome cassettes to tell us the relative abundance of TFome cassettes in the double-positive cells (FIGs. 3A-3E). We found that the distributions for each cell type had some variability, but that in general, each cell type had TFs that were represented in the positive population more than others. In fact, only one of the six cell types (cytotoxic T-cells) had all TFs show up in sequencing at least once.

Example 3 - Iterative pooled TF screening and clonal isolation

Using the barcode frequencies, we calculated 3 refined TF pools for each cell type: All TFs that appear in sequencing, TFs that appear greater than average, and TFs that appear one standard deviation or more than average (FIGs. 4A-D. Using the refined TF pools, we performed a second round of differentiation. Given that this round of screening generally limited TF pools to <5 TFs per pool, we built stable cell lines for additional testing and refinement. iPSCs were nucleofected as before, but we selected and stabilized the cell lines before screening differentiation in different settings. Specifically, given the stability of the constructed cell lines (e.g., less cell death), we opted to test them for only six days, and also decided to test their performance in targeT-cell-type growth medium in addition to stem cell medium (data not shown).

In this round, we found broad improvement in differentiation percentage across all six cell types (FIGs. 4A-D). While B cells already had a considerably high differentiation percentage in the primary screening round (17.6%), it improved to an average of greater than 50%. For all other five cell types, the refined lines appeared to improve in differentiation percentage dramatically compared to the populations seen in the primary screen. However, since these populations have mixed identity, it is likely that many of these cells were still only partially differentiated. When we examined the number of cells that were positive for just one (or both) markers, all cell types improved differentiation rates compared to the primary screens (data not shown). When we examined differentiation percentage (both partial and total) in target T-cell-type growth media, we saw even more near-complete differentiation of these cell lines (data not shown). While it was clear that the growth medium is a contributor to differentiation efficiency, the TFs were the major driver of differentiation for all cell types.

Given that our cell lines were clearly making progress towards robust differentiation, but in a limited capacity, we reasoned that perhaps many micro-scale experimental details could be to blame — for example, perhaps cell-cell communication from non-differentiating cells in the population was the issue, or perhaps the details of how many TF cassettes were integrated and in what location was very important. Since we use PiggyBac integrase that integrates variable copies of TF over-expression cassettes in random genomic locations, we hypothesized that perhaps some cells in the cell line population are holding back the rest of the population, and that isolating monoclonal cell lines could improve our differentiation efficiency. Ergo, we sorted random single cells in the population to form monoclonal lines and characterized them. To our satisfaction, for CD8 T-cells, microglia, astrocytes, and hepatocytes, this solved the problem — several clones of each were able to dramatically outcompete the mixed population in differentiation efficiency in all of the aforementioned differentiation conditions (FIGs. 4A-D).

After differentiation of high-performance clones, we performed RNA-sequencing to validate that our clones were generally reflective of target cell types at a molecular level in addition to surface markers. We found that across all genes, our differentiated cells clustered well by cell type in both media conditions (FIG. 4E). Specifically, it was important to see that the molecular characteristics of the T-cell subtypes were in general agreement and were significantly different from all other types. As expected, since these cell types were all from different germ layers (except the T-cell subtypes), the expression profiles were dramatically different across differentiated cell types. This was further reflected in principal component analysis (FIG. 4F) - we observed that our differentiated cell types generally clustered very tightly across both media conditions and that they clustered somewhat well with primary cell types. The clustering of cell types across variable media reinforces that the TF over-expression a more dominant factor than the different media conditions. Next, when we zoom in on key canonical markers for our differentiated cells, they once again cluster as expected and generally show upregulation of expected markers (FIG. 4G). In the case of iAstIIs and iTRegs, there were some interesting, marked difference of key factors across media conditions, suggesting that media formulation may play a key role in the final condition and function of these cells. Finally, when we analyze the complete sets of significantly up-regulated genes ($P < 0.1$) for our high-efficiency clonal lines compared to iPSCs with Metascape, we see enrichment of GO terms that is supportive of cell-type specific features (FIG. 4H).

Example 4 - Functional characterization of differentiated cells

Finally, after refinement of our differentiating cell lines and molecular validation of their identities, we wanted to validate that the cells also functionally perform their intended function for down-stream clinical applications. To this end, we opted to focus on at least one cell type from each germ layer - regulatory T-cells (iTregs), cytotoxic T-cells (iCD8s), type II astrocytes (iAstIIs) and hepatocytes (iHeps). To functionally characterize these cell types, we performed *in vitro* assays based on biological function (FIGs. 5A-5L).

For the iAstIIs, we validated that the morphology was correct and that they were stimulated as expected by certain standard small molecules (FIGs. 5A-C). We observed that at standard concentrations of small-molecules of three classes (glutamate - neurotransmitter, ATP - nucleotide, and KCl - ionic), many plated astrocytes were stimulated. We observed strong increases of relative Fluo-4 fluorescence immediately after induction for individual astrocytes that were both inactive before stimulation and active at times before stimulation. Furthermore, while glutamate and KCl should stimulate both astrocytes and other neuronal types, only astrocytes are stimulated by ATP, confirming that the cells we assayed both had correct astrocyte morphology and exclusive functionality.

For the iHeps, we validated the morphology (FIG. 5D) and compared their viability compared to primary hepatocytes and undifferentiated cells when exposed to hepatotoxins for 24hrs. We observed that our iHeps had highly similar viability to primary hepatocytes after being exposed to Nefazodone (FIG. 5E), Acetaminophen (FIG. 5F), and Troglitozone (FIG. 5G), and demonstrated significantly higher viability compared to undifferentiated iPSCs.

iTRegs were validated by demonstrating that the cells inhibited the expansion of responder T-cells. Before this step, we confirmed that our iTRegs had size and morphology approximately the same as primary cytotoxic responder cells (FIG. 5H). While the size and shape were generally consistent, with both iTRegs and iCytoTs, the primary responder T-cells took on an elongated shape when stimulated, while our iCytoTs did not clearly show this morphological change to stimulus. Responder T-cells were stimulated to activate with IL-2 and CD3+CD28+ beads for three days. After this activation step, responder T-cells were labeled with a fluorescent dye and co-cultured with iTRegs in variable quantities. After 11 days, fluorescence was recorded to validate that the addition of more iTRegs resulted in reduced responder T-cell proliferation (FIGs. 5I, 5J). We observed some reduction in responder T-cell proliferation as we increased the number of iTRegs, albeit modestly compared to prior results with primary regulatory T-cells. Finally, to validate the iCytoTs, we activated them with the same bead-based method used in the regulatory T-cell assay and examined their morphology and interaction with the activator beads (FIG. 5H) and then recorded proliferation. We found that as with the iTRegs, the proliferation was modest, but increased by the number of days the iCytoTs were induced from stem cells prior to the initiation of the assay (FIGs. 5K, 5L).

In summary, we have described how the *CellCartographer* tool and pipeline can guide and refine cell-fate engineering with machine learning and synthetic TF-cassettes from the human TFome. We demonstrated that the primary TF pools for differentiating iPSCs into a diverse set of cell types yields a small population of positive cells for each of the tested types. We then went on to focus on six cell types from each germ layer to show how we can use NGS data from partially engineered cell lines with *CellCartographer* to engineer high-efficiency differentiation-inducible cell lines. Finally, we isolated high-performance clones for four cell types and functionally characterized at least one cell type from each germ layer to validate that our engineered cell lines were functionally accurate *in vitro*.

While *CellCartographer* is not the first software to identify TFs for cell-fate engineering, it presents an advance in three main areas from a software perspective. First, it leverages machine learning to make TF predictions using epigenetics data and enables an iterative pipeline for refining engineered cell lines. We hypothesize that as sequencing technologies continue to improve and more data is generated, *CellCartographer*'s predictions should only improve. Second, *CellCartographer* has a very minimal requirement for producing useful TF pools — it does not require re-training large models for additional cell types, which can prove useful for engineering cell lines for differentiation into exotic cell types with little data available. Furthermore, we were able to successfully engineer iTRegs using TFs determined from *Mus Musculus* data since that was the only epigenetic NGS data available for this cell type, meaning calculations of factors can work cross-species. Finally, the pooled screening philosophy of *CellCartographer*, allows biologists to explore and debug many experimental variables that are generally invisible to software tools — namely synthetic DNA genomic integration location, copy count, and cell culture conditions. Pooled screening and paired ML analysis allows us to screen out these issues. Furthermore, while we use the starting predictions from *CellCartographer* to iteratively refine our cell lines in this study, all of the down-stream tools are compatible with starting predictions from other tools (e.g., another tool could provide the starting prediction and *CellCartographer* and the TFome can still be used downstream), meaning *CellCartographer* can be used to compliment other existing tools.

This work also represents a major advance in terms of identifying four robust TF combinations for differentiation into high-value cell types relevant to therapeutics. At this time, aside from hepatocytes, there are no experimentally established TF combinations for directly differentiating stem cells into type II astrocytes, regulatory T-cells, or cytotoxic T-cells. Furthermore, we demonstrate that this differentiation can be driven in stem cell media in six days or less, meaning that the TF combinations are fast, robust, and solely to credit for the differentiation in these examples. Finally, by performing additional optimizations with specialized media conditions and performing functional assays on iAstIIs, iHeps, iTRegs, and iCytoTs, we show that this strategy should be robust in ultimately obtaining functional clonal cell lines of theoretically any type that can differentiate rapidly, efficiently, and robustly from iPSCs. While the functional qualities of the iAstIIs, and iHeps were more dramatic and complete, the function and viability of the induced T-cells is likely very sensitive to media conditions and could be further improved with additional optimization of growth conditions starting from the stem cell state.

In conclusion, we believe that *CellCartographer* provides a clear benefit to the field of stem cell biology and cell-line engineering. While we have already generated interesting inducibly-differentiating iPSC lines, we strongly believe that this tool can be applied immediately to aid the engineering of other stem cell lines for any number of therapeutic, diagnostic, or other commercial applications.

Example Methods

DNAse-seq and ATAC-seq analysis

Adapters from sequencing reads were trimmed with Homer, using the command: homerTools trim-len 40. Following adapter trimming, reads were aligned using Bowtie2 (with default parameters) and then converted into a Homer tag directory. We called open chromatin regions or peaks with Homer using the following findPeaks command with the following parameters -C 0 -L 0 -fdr 0.9. We then use IDR to identify high confidence open chromatin regions.

Prediction of transcription factors for cell fate engineering

For the set of open chromatin regions for each cell type, we sample from the genome an equivalent number of background peaks that has matching GC content and size. Using a set of non-redundant DNA motifs, which specify the frequency of each nucleotide at each position in the motif, and a background frequency (0.25 at each position), we can calculate a log odds score that indicates how well a sequence matches a motif. For each open chromatin region and background loci, we calculate the highest log odds score for each motif. We standardize the motif scores such that the mean score value is 0 and the variance is 1. Then we train a L1-regularized logistic regression model (LASSO) to discriminate between open chromatin regions and background sites. We assess the importance of each motif using a log-likelihood ratio test where we compare the performance of a perturbed model where a motif has been masked from during the model training procedure and the full model that has observed all motifs. We convert the difference in likelihoods given by the two models to p-values using the chi-squared test. Model coefficients and p-values reported are the average across 5 cross-validation splits. Data processing, model training, and statistical analysis was performed using python and the following packages: pandas, scipy, sklearn, biopython.

Transcriptomics analysis

Adapters from sequencing reads were trimmed with Homer, using the command: homer- Tools trim -len 40. Following adapter trimming, reads were aligned using Bowtie2 (with default parameters) and then converted into a Homer tag directory. We used the Homer analyzeRepeats command to quantify gene expression as RPKM values. Raw read counts at each gene were used as input to DeSeq2 for identifying differentially expressed genes.

Cloning of transcription factors

Transcription factors were cloned into puromycin-resistant cassettes with flanking piggyBac transposon [SystemsBio] genomic integration regions under the control of the mammalian DOX-unducible promoter pTRET. Plasmids for each transcription factor are members of the 'Human TFome' library deposited on Addgene.

Creation of cell lines and cell culture

All differentiating cell lines and differentiation screens were performed on reprogrammed PGP1 fibroblasts using the Sendai-reprogramming-factor virus. PGP1 iPS cells were expanded and nucleofected with P3 Primary cell 4D Nuceleofection kits with pulse code CB150 using 2μg of total DNA for 800,000 cells (1.6 μg TF pool/0.4 μg SPB) [Lonza]. Cells were plated onto Matrigel-cotated plates [Corning] with ROCK-inhibitor [Millipore] and selected with puromycin [Sigma]. Stable cell lines were expanded over several passages using TrypLE [Gibco] in mTeSR1 [StemCell Technologies] and frozen in mFreSR [StemCell Technologies]. Cells were differentiated with 2ng/mL doxycycline [Sigma] at variable conditions as described in (data not shown) in either mTeSR [StemCell Technologies], RPMI-1640 (microglia) [Gibco], Williams' E Medium (hepatocytes) [Gibco], Immunocult-XF T-cell Expansion Media (T-cells) [StemCell Technologies], LGM-3 (B cells) [Lonza], or BrainPhys Media (Astrocytes) [Stem Cell Technologies].

Flow Cytometry and Cell Sorting

Cells were digested in TrypLE [Gibco] and resuspended in growth media before staining with cell surface markers. The following antibodies were used for analysis and cell sorting: [Microglia: CD11b-FITC, CX3CR1-PE]; [CD8-positive T-cells: CD3-PerCP-Cy5.5, CD8-FITC]; [T-Regulatory cells: CD3-PerCP-Cy5.5, CD4-PE-Cy7, FOXP3-PE, CD127-V450]; [B cells: CD19-PE-Cy7, CD27-FITC]; [Hepatocytes: ASGPR1-PE, CD184-APC]; [Astrocytes: CD44-FITC, A2B5-PE]. Cells were sorted and collected on a Sony SH800 FACS for primary screens. For characterization of stable cell lines, cells were stained and analyzed on a BD LSR Fortessa Analyzer flow cytometer. The gating strategy is exemplified in (data not shown).

RNA sequencing

Cells were either collected from FACS (primary screens) or collected directly from culture (refined screens and stable cell line characterization) and were lysed in TRIzol [Invitrogen]. RNA was purified with Direct-zol RNA MicroPrep and RNA MiniPrep kits [Zymo]. Library prep was performed using a SMARTer-seq v2 NGS library prep kit [TARAKA] (primary screens) and NEBNext Ultra II RNA Kits [NEB] (refined screens and stable cell line characterization). Barcodes were amplified from the prepped cDNA using two alternative primer pairs (data not shown). Amplicons were sequenced with a MiSeq kit [Illumina] using Illumina TruSeq indexes.

Astrocyte stimulation assays

iAstIIs were differentiated as described (data not shown) and then transferred to imaging dishes for stimulation as previously described. Briefly, glass bottom dishes [Ibidi 81158] were coated in Poly-d-lysine (0.1 mg/mL) for 2 hours at room temperature, washed twice in PBS [Gibco], and coated overnight in fibronectin (10 μg/mL) [Thermo] at 37°C. Differentiated astrocytes were digested in TrypLE [Gibco] for 7-10 minutes, and 40,000-50,000 cells were transferred to coated dishes and maintained for 2 days before stimulation and imaging. Prior to stimulation and imaging the astrocytes were stained with Fluo-4 (1 μg/mL) [FluoroPure] in BrainPhys without phenol red [StemCell] and incubated in the dark for at least 25 minutes at 37°C. Cells were then washed with fresh media three times and transferred immediately to a Zeiss Axio 3 Inverted Microscope with CO_2 (5%) and temperature control (37°C). After staging, basal activity was measured for at least 2 minutes, after which small molecule stimuli were applied.

Hepatocyte hepatotoxicity assays

iHeps were differentiated as described (data not shown) and then transferred to 96-well plates pre-coated with Matrigel [Corning] and treated with hepatotoxins as previously described. Briefly, after differentiation, 25,000 iHeps, undifferentiated iPSCs, and plateable primary human hepatocytes [ZenBio] were plated in each well and incubated overnight at

37°C. The next day, media was changed to Hepatocyte Medium E (William's E Medium [Gibco], Maintenance Cocktail B [Gibco], and 0.1μM Dexamethasone [Gibco]) for one day. The following day, media was exchanged and supplemented with hepatotoxins (Acetaminophen at [3.125,6.25,12.5,25,50,100] mM [Spectrum], Nefazodone at [1,3,10,30,100,300] μM [Sigma], and Troglitazone at [1,3,10,30,100,300] μM [Sigma]). Cells were incubated again at 37°C for 24 hours, and viability was measured with CellTiter-Glo Luminescent Cell Viability Assay [Promega].

Cytotoxic T-cell activation assays

Primary cytotoxic T-cells (Human Peripheral Blood CD4+CD45RA+ T Cells) [StemCell] and iCytoTs were cultured and activated in the same manner. Briefly, cells were incubated in ImmunoCult- XF T Cell Expansion Medium [StemCell] + IL-2 [R&D Systems] with DYNAL Dynabeads Human T-Activator CD3/CD28 for T Cell Expansion and Activation [Fisher] for 3 days. After this incubation, the cells were stained with Celltrace Violet [Fisher] and moved into new wells at the concentration of 1M cells/well with fresh media (as above) and grown at 37°C for 11 days, changing media every 2-3 days. Finally, cells were analyzed via flow cytometry. Percent activated was determined by gating cells that had diminished fluorescence after proliferation.

Regulatory T-cell proliferation suppression assays

iTRegs were co-cultured with activated primary cytotoxic T-cells in variable quantities. Briefly, iTRegs were differentiated in ImmunoCult-XF T Cell Expansion Medium [StemCell] + IL-2 [R&D Systems] for 4 days and then moved into co-culture with activated and CellTrace Violet [Fisher] stained cytotoxic T-cells and grown at 37°C for 11 days, changing media every 2-3 days. Finally, cells were analyzed via flow cytometry. The percentage of suppression was determined as 100 x [1 - (% of proliferating cells with iTRegs) / (% of proliferating cells without iTregs)] after applying gates for proliferating v. non-proliferating cells and subtracting auto-fluorescence resulting from unstained iTregs.

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 ±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 ERG, EGR1, FLI1, FOSB, or any combination thereof.

2. The PSC of any one of the preceding claims, wherein the PSC expresses or overexpresses ERG, EGR1, FLI1, FOSB, or any combination thereof.

3. The PSC of any one of the preceding claims, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter.

4. A pluripotent stem cell (PSC) comprising: a protein selected from ERG, EGR1, FLI1, and FOSB, wherein the protein is overexpressed.

5. The PSC of any one of the preceding claims, wherein the PSC is a human PSC, optionally an induced PSC (iPSC).

6. The PSC of any one of the preceding claims, comprising 1-20 copies, optionally 8-10 copies, of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ERG, EGR1, FLI1, and FOSB, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

7. A composition comprising: a population of the PSC of any one of the preceding claims, optionally, wherein the population comprises at least 2500/cm^2 of the PSC.

8. 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 ERG, EGR1, FLI1, and FOSB to produce astrocyte-like cells.

9. The method of claim 8, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally an inducible promoter, preferably a chemically-inducible promoter such as a doxycycline-inducible promoter.

10. The method of claim 8 or 9, wherein the population comprises 1x10^2-1x10^7 PSCs.

11. The method of any one of the preceding claims, wherein the population of PSCs is cultured for at least 1 day, about 3-6 days, or no more than 6 days.

12. The method of any one of the preceding claims, wherein the astrocyte-like cells are CD44$^+$ and A2B5$^+$.

13. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

14. The PSC of any one of the preceding claims, wherein the PSC expresses or overexpresses ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

15. The PSC of any one of the preceding claims, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter.

16. A pluripotent stem cell (PSC) comprising: a protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4, wherein the protein is overexpressed.

17. The PSC of claim 41, wherein the PSC expresses or overexpresses: ZBTB1, RUNX3, RELA, NRF1, ERF, SP4, or any combination thereof.

18. The PSC of any one of claims 13-17, wherein the PSC is a human PSC, optionally wherein the PSC is an induced PSC (iPSC).

19. The PSC of any one of the preceding claims, comprising 1-20 copies, optionally 8-10 copies, of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

20. A composition comprising: a population of the PSC of any one of the preceding claims, wherein the population comprises at least 2500/cm^2 of the PSC.

21. 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 ZBTB1, RUNX3, RELA, NRF1, ERF, and SP4 to produce cytotoxic T-cell-like cells.

22. The method of claim 21, wherein the PSCs of the expanded population comprise an engineered polynucleotide comprising an open reading frame encoding SP4.

23. The method of claim 21 or 22, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally .wherein the heterologous promoter is an inducible promoter, further optionally wherein the inducible promoter is a chemically-inducible promoter, further optionally, .wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

24. The method of any one of the preceding claims, wherein the population comprises $1x10^2$ -$1x10^7$ PSCs.

25. The method of any one of the preceding claims, wherein the population of PSCs is cultured for at least 1 day, about 3-6 days, or no more than 6 days.

26. The method of any one of the preceding claims, wherein the cytotoxic T-cell-like cells are $CD3^+$ and $CD8^+$.

27. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding HNF4G, TEAD4, RFX3, or any combination thereof.

28. The PSC of any one of the preceding claims, wherein the PSC expresses or overexpresses HNF4G, TEAD4, RFX3, or any combination thereof.

29. The PSC of claim 28, wherein the PSC further expresses or overexpresses HNF4A.

30. The PSC of any one of the preceding claims, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter.

31. A pluripotent stem cell (PSC) comprising: a one or more proteins selected from HNF4G, TEAD4, and RFX3, wherein the one or more protein or any combination of the one or more proteins is overexpressed.

32. The PSC of any one of claims 27-31, wherein the PSC is a human PSC, optionally wherein the PSC is an induced PSC (iPSC).

33. The PSC of any one of the preceding claims, comprising 1-20 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from HNF4G, HNF4A, TEAD4, and RFX3, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

34. The PSC of any one of the preceding claims, comprising 8-10 copies of the engineered polynucleotide comprising the open reading frame encoding the protein selected from HNF4G, TEAD4, and RFX3, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

35. A composition comprising: a population comprising the PSC of any one of the preceding claims.

36. The composition of claim 35, wherein the population comprises at least $2500/cm^2$ of the PSC.

37. 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 HNF4G, TEAD4, and RFX3 to produce hepatocyte-like cells.

38. The method of claim 37, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter, further optionally wherein the inducible promoter is a chemically-inducible promoter, further optionally wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

39. The method of any one of the preceding claims, wherein the population comprises $1x10^2$-$1x10^7$ PSCs.

40. The method of any one of the preceding claims, wherein the population of PSCs is cultured for at least 1 day, about 3-6 days, or no more than 6 days.

41. The method of any one of the preceding claims, wherein the hepatocyte-like cells are $CD184^+$ and $ASGPR1^+$.

42. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof.

43. The PSC of any one of the preceding claims, wherein the PSC expresses or overexpresses ETS1, ETV3, GABPA, KLF9, NFKB1, or any combination thereof.

44. The PSC of any one of the preceding claims, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter.

45. A pluripotent stem cell (PSC) comprising: a one or more protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, wherein the one or more protein or any combination of the one or more proteins is overexpressed.

46. The PSC of any one of claims 42-45, wherein the PSC is a human PSC, optionally wherein the PSC is an induced PSC (iPSC).

47. The PSC of any one of the preceding claims, comprising 1-20 copies, optionally 8-10 copies, of the engineered polynucleotide comprising the open reading frame encoding the protein selected from ETS1, ETV3, GABPA, KLF9, and NFKB1, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

48. A composition comprising: a population of the PSC of any one of the preceding claims, wherein the population comprises at least 2500/cm^2 of the PSC.

49. 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 ETS1, ETV3, GABPA, KLF9, and NFKB1 to produce regulatory T-cell-like cells.

50. The method of claim 49, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter, further optionally wherein the inducible promoter is a chemically-inducible promoter, further optionally wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

51. The method of any one of the preceding claims, wherein the population comprises $1x10^2$ -$1x10^7$ PSCs.

52. The method of any one of the preceding claims, wherein the population of PSCs is cultured for at least 1 day, about 3-6 days, or no more than 6 days.

53. The method of any one of the preceding claims, wherein the regulatory T-cell-like cells are $CD3^+$ and $CD25^+$.

54. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof.

55. The PSC of any one of the preceding claims, wherein the PSC expresses or overexpresses EBF1, ZBTB1, RELA, NRF1, REL, or any combination thereof.

56. The PSC of any one of the preceding claims, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter.

57. A pluripotent stem cell (PSC) comprising: a one or more protein selected from EBF1, ZBTB1, RELA, NRF1, and REL, wherein the one or more protein or any combination of the one or more proteins is overexpressed.

58. The PSC of any one of claims 54-57, wherein the PSC is a human PSC, optionally wherein the PSC is an induced PSC (iPSC).

59. The PSC of any one of the preceding claims, comprising 1-20 copies, optionally 8-10 copies, of the engineered polynucleotide comprising the open reading frame encoding the protein selected from EBF1, ZBTB1, RELA, NRF1, and REL, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

60. A composition comprising: a population of the PSC of any one of the preceding claims, wherein the population comprises at least 2500/cm^2 of the PSC.

61. 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 EBF1, ZBTB1, RELA, NRF1, and REL to produce B cell-like cells.

62. The method of claim 61, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter, further optionally wherein the inducible promoter is a chemically-inducible promoter, further optionally wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

63. The method of any one of the preceding claims, wherein the population comprises $1x10^2$-$1x10^7$ PSCs.

64. The method of any one of the preceding claims, wherein the population of PSCs is cultured for at least 1 day, about 3-6 days, or no more than 6 days.

65. The method of any one of the preceding claims, wherein the B cell-like cells are CD19$^+$ and CD27$^+$.

66. A pluripotent stem cell (PSC) comprising:

an engineered polynucleotide comprising an open reading frame encoding SPI1, ZBTB1, RELA, STAT2, or any combination thereof.

67. The PSC of any one of the preceding claims, wherein the PSC expresses or overexpresses SPI1, ZBTB1, RELA, STAT2, or any combination thereof.

68. The PSC of any one of the preceding claims, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter.

69. A pluripotent stem cell (PSC) comprising: a one or more protein selected from SPI1, ZBTB1, RELA, and STAT2, wherein the one or more protein or any combination of the one or more proteins is overexpressed.

70. The PSC of any one of claims 66-69, wherein the PSC is a human PSC, optionally wherein the PSC is an induced PSC (iPSC).

71. The PSC of any one of the preceding claims, comprising 1-20 copies, optionally 8-10 copies, of the engineered polynucleotide comprising the open reading frame encoding the

protein selected from SPI1, ZBTB1, RELA, and STAT2, optionally wherein the engineered polynucleotide is integrated into the genome of the PSC.

72. A composition comprising: a population comprising the PSC of any one of the preceding claims, wherein the population comprises at least 2500/cm^2 of the PSC.

73. 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 SPI1, ZBTB1, RELA, and STAT2 to produce microglia-like cells.

74. The method of claim 73, wherein the open reading frame of the engineered polynucleotide is operably linked to a heterologous promoter, optionally wherein the heterologous promoter is an inducible promoter, further optionally wherein the inducible promoter is a chemically-inducible promoter, further optionally wherein the chemically-inducible promoter is a doxycycline-inducible promoter.

75. The method of any one of the preceding claims, wherein the population comprises 1x10^2 -1x10^7 PSCs.

76. The method of any one of the preceding claims, wherein the population of PSCs is cultured for at least 1 day, about 3-6 days, or no more than 6 days.

77. The method of any one of the preceding claims, wherein the microglia-like cells are CD11b+ and CX3CR1+.

78. The method of any one of the preceding claims, wherein the target cell type is a Type II astrocyte, cytotoxic T-cell, regulatory T-cell, hepatocyte, B cell, or microglial cell.

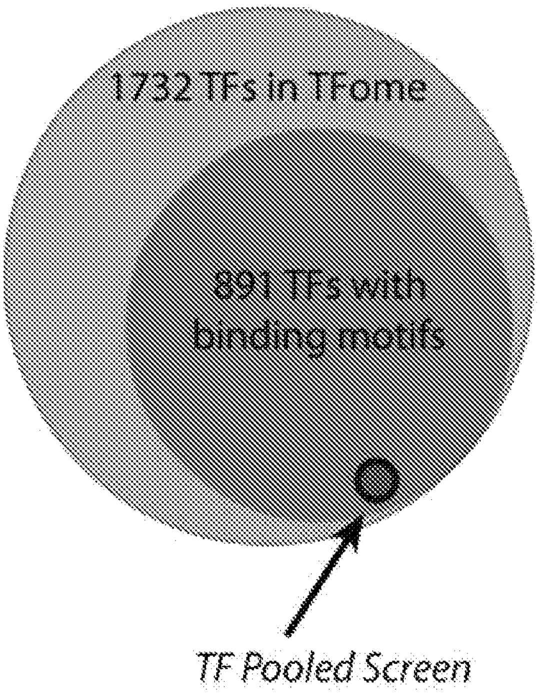

1732 TFs in TFome
891 TFs with binding motifs
TF Pooled Screen

FIG. 1A

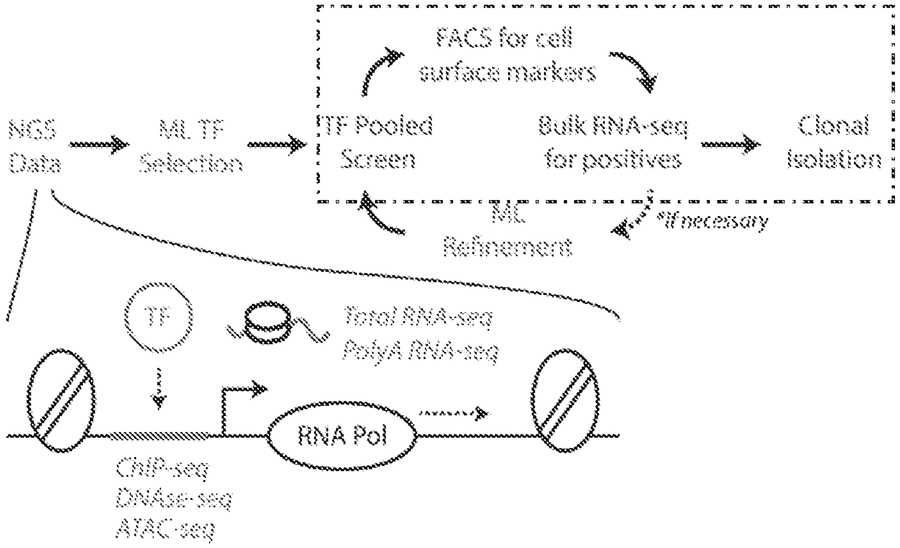

NGS Data
ML TF Selection
TF Pooled Screen
FACS for cell surface markers
Bulk RNA-seq for positives
Clonal isolation
ML Refinement
*if necessary
TF
Total RNA-seq
PolyA RNA-seq
RNA Pol
ChIP-seq
DNAse-seq
ATAC-seq

FIG. 1B

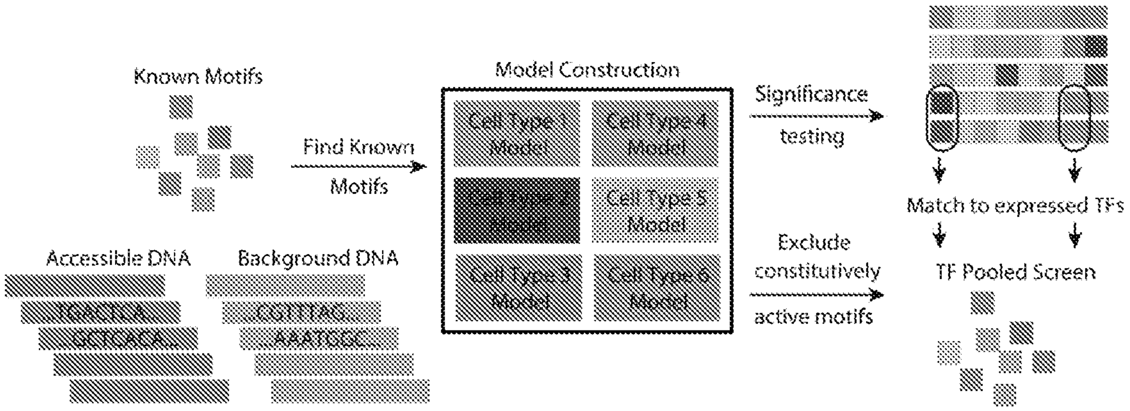

Known Motifs
Find Known
Motifs
Accessible DNA
TGACTCA...
GCTCACA...
Background DNA
CGTTTAG...
AAATGGC...
Model Construction
Cell Type 1 Model
Cell Type 4 Model
Cell Type 2 Model
Cell Type 5 Model
Cell Type 3 Model
Cell Type 6 Model
Significance
testing
Exclude
constitutively
active motifs
Match to expressed TFs
TF Pooled Screen

FIG. 1C

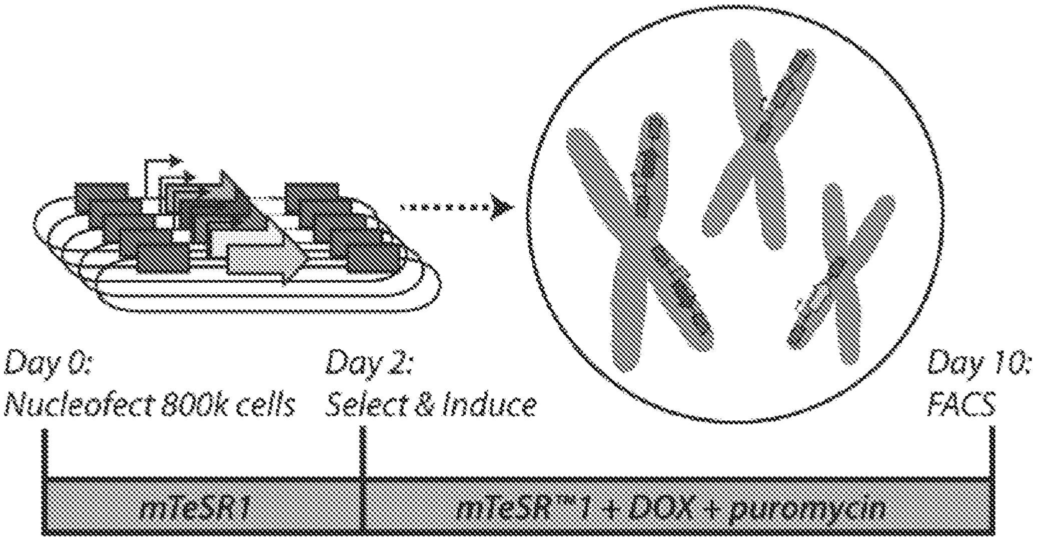

Day 0:
Nucleofect 800k cells
Day 2:
Select & Induce
Day 10:
FACS
mTeSR1
mTeSR™1 + DOX + puromycin

FIG. 1D

TF RANK	h1-ESC	Endothelial Cell	Bipolar Neuron	Macrophage
1	REST	FOSL2	EBF1	[illegible]
2	[illegible]	JUN	KLF14	SPI1
3	[illegible]	JUNB	SP4	[illegible]
4	SP4	JUND	[illegible]	SIX5
5	KLF14	FOS	NEUROD2	JUN
6	TEAD4	FOSL1	[illegible]	JUNB
7	TEAD3	ETV5	OLIG1	JUND
8	TEAD1	[illegible]	OLIG2	FOS
9	NFYB	ERF	OLIG3	FOSL1
10	ZIC4	ERG	[illegible]	FOSL2
11	ZIC1	ETV1	[illegible]	ELF4
12	ZIC3	FEV	BHLHE22	ELF1
13	SIX5	ELK4	BHLHE23	ETV6
14	RFX2	ETS1	MXI1	JDP2
15	RFX5	ELK3	RFX2	NFE2
16	RFX3	ETV3	RFX5	FOSB

FIG. 1E

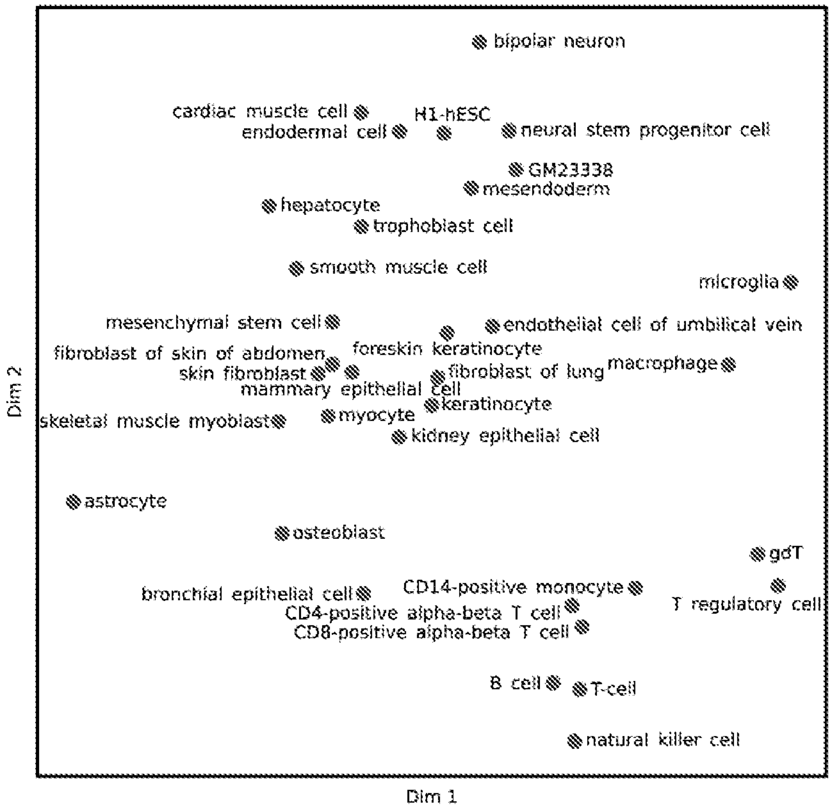

bipolar neuron
cardiac muscle cell
H1-hESC
endodermal cell
neural stem progenitor cell
GM23338
mesendoderm
hepatocyte
trophoblast cell
smooth muscle cell
microglia
mesenchymal stem cell
endothelial cell of umbilical vein
fibroblast of skin of abdomen
foreskin keratinocyte
skin fibroblast
fibroblast of lung
macrophage
mammary epithelial cell
keratinocyte
skeletal muscle myoblast
myocyte
kidney epithelial cell
astrocyte
osteoblast
gdT
bronchial epithelial cell
CD14-positive monocyte
CD4-positive alpha-beta T cell
T regulatory cell
CD8-positive alpha-beta T cell
B cell
T-cell
natural killer cell
Dim 2
Dim 1

FIG. 2A

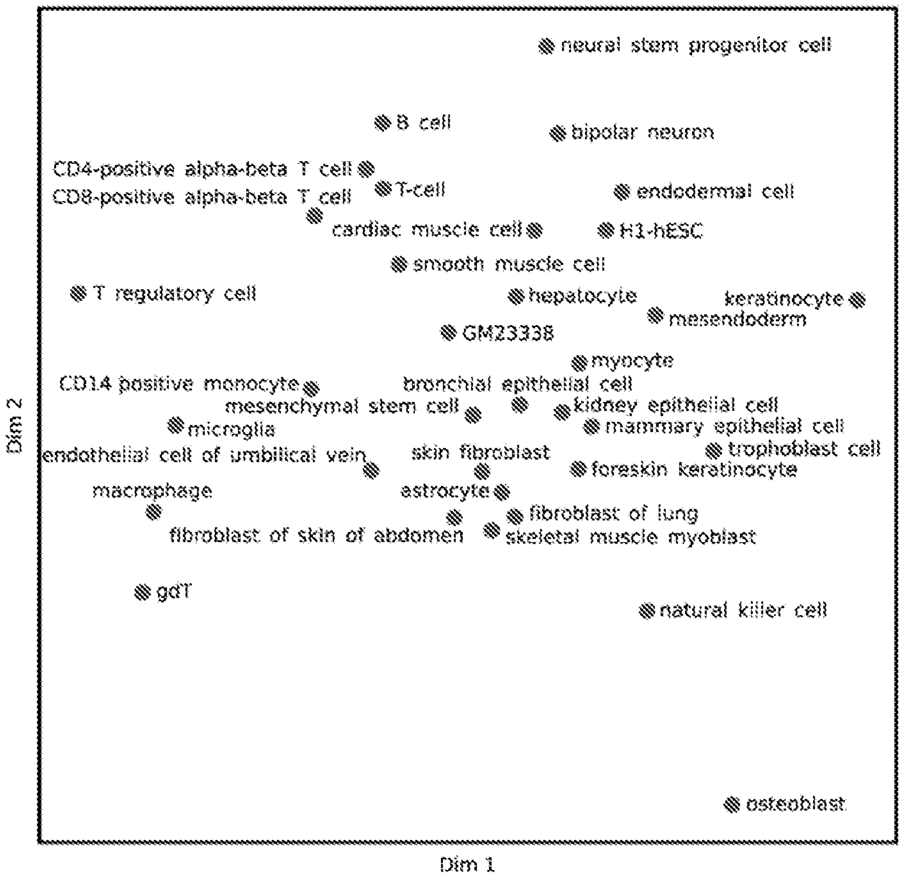

neural stem progenitor cell
B cell
bipolar neuron
CD4-positive alpha-beta T cell
T-cell
CD8-positive alpha-beta T cell
endodermal cell
cardiac muscle cell
H1-hESC
smooth muscle cell
T regulatory cell
hepatocyte
keratinocyte
mesendoderm
GM23338
myocyte
CD14 positive monocyte
bronchial epithelial cell
mesenchymal stem cell
kidney epithelial cell
microglia
mammary epithelial cell
endothelial cell of umbilical vein
skin fibroblast
trophoblast cell
foreskin keratinocyte
macrophage
astrocyte
fibroblast of lung
fibroblast of skin of abdomen
skeletal muscle myoblast
gdT
natural killer cell
osteoblast
Dim 2
Dim 1

FIG. 2B

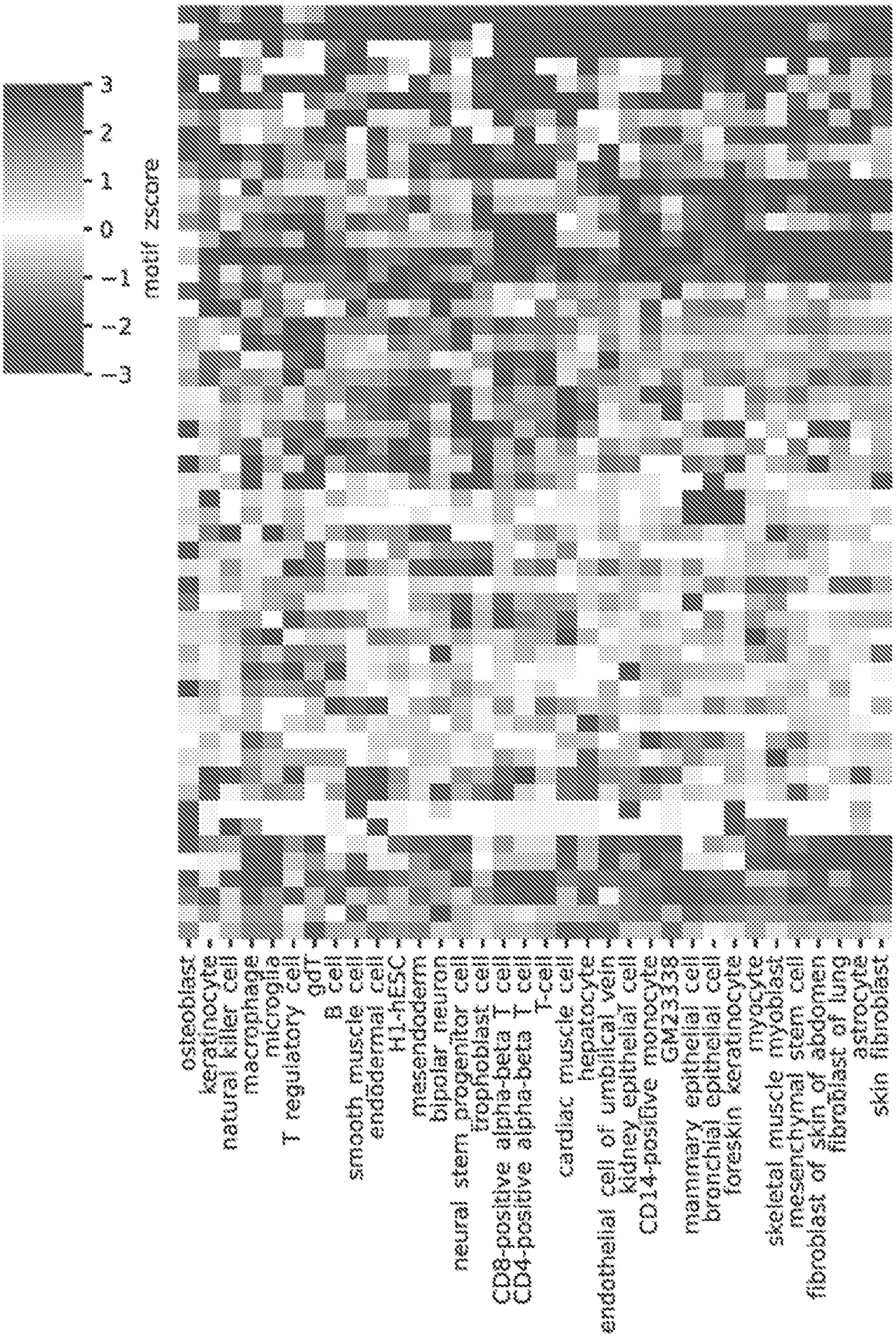

3
2
1
0
-1
-2
-3
motif zscore
osteoblast
keratinocyte
natural killer cell
macrophage
microglia
T regulatory cell
gdT
B cell
smooth muscle cell
endodermal cell
H1-hESC
mesendoderm
bipolar neuron
neural stem progenitor cell
trophoblast cell
CD8-positive alpha-beta T cell
CD4-positive alpha-beta T cell
T-cell
cardiac muscle cell
hepatocyte
endothelial cell of umbilical vein
kidney epithelial cell
CD14-positive monocyte
GM23338
mammary epithelial cell
bronchial epithelial cell
foreskin keratinocyte
myocyte
skeletal muscle myoblast
mesenchymal stem cell
fibroblast of skin of abdomen
fibroblast of lung
astrocyte
skin fibroblast

FIG. 2C

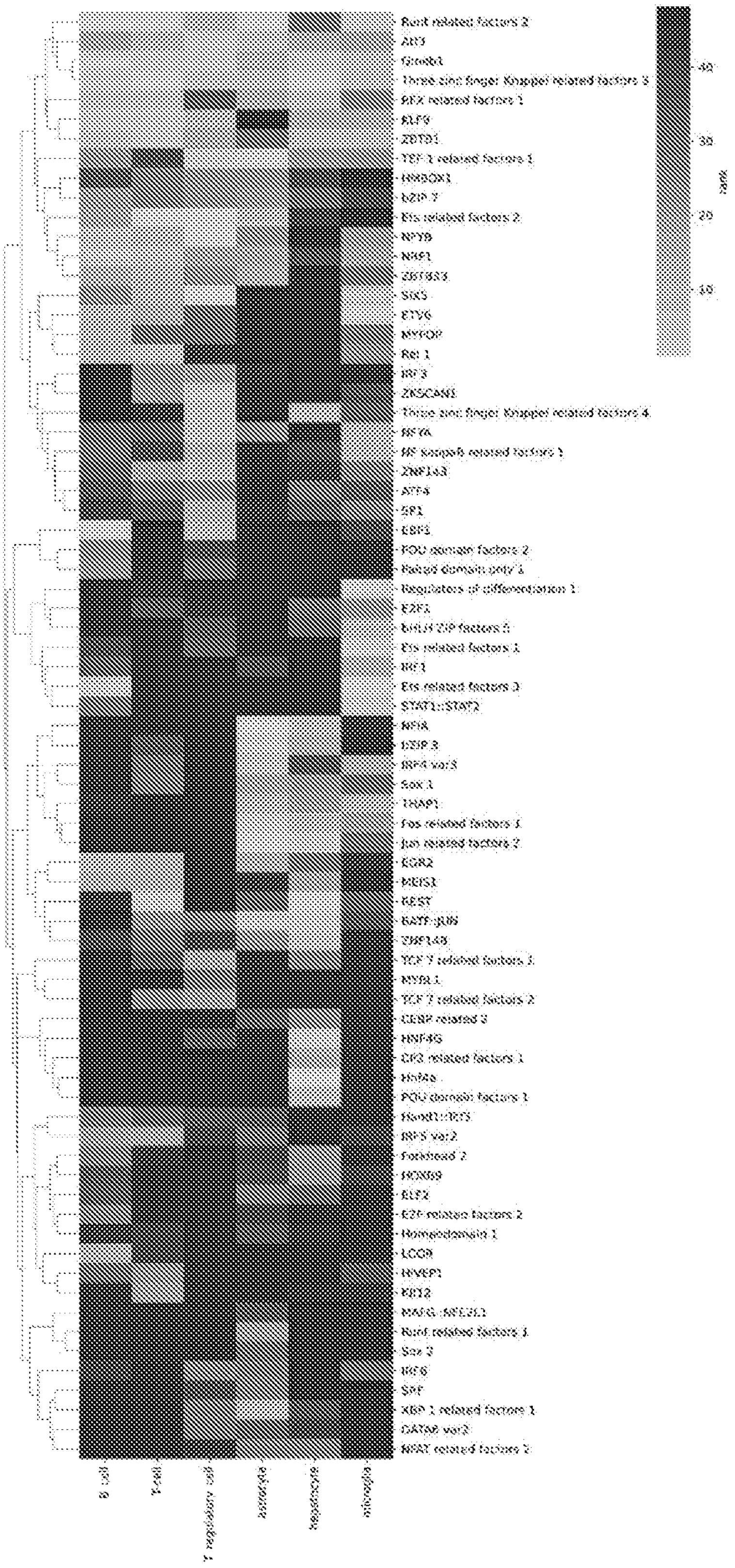

Runt related factors 2
AR3
Gtcebd
Three zinc finger Kruppel related factors 3
RFX related factors 1
KLF9
ZBTB1
TEF 1 related factors 1
HMBOX1
bZIP 7
Ets related factors 2
NFYB
NRF1
ZBTB33
SIX5
ETV6
MYPOP
Rel 1
IRF3
ZKSCAN1
Three zinc finger Kruppel related factors 4
NFYA
NF kappaB related factors 1
ZNF143
ATF4
SP1
EBF1
POU domain factors 2
Paired domain only 1
Regulators of differentiation 1
E2F1
bHLH ZIP factors 5
Ets related factors 1
IRF1
Ets related factors 3
STAT1::STAT2
NFIA
bZIP 3
IRF4 var3
Sox 1
THAP1
Fos related factors 1
Jun related factors 2
EGR2
MEIS1
REST
BATF::JUN
ZNF148
TCF 7 related factors 1
MYBL1
TCF 7 related factors 2
CEBP related 3
HNF4G
CP2 related factors 1
Hnf4a
POU domain factors 1
Hand1::Tcf3
IRF5 var2
Forkhead 2
HOXB9
KLF2
E2F related factors 2
Homeodomain 1
LCOR
HIVEP1
EBF2
MAFG::NFE2L1
Runt related factors 1
Sox 3
IRF6
SPF
XBP 1 related factors 1
GATA6 var2
NFAT related factors 2
B cell
T-cell
T regulatory cell
astrocyte
hepatocyte
microglia
40
30
20
10
rank

FIG. 2D

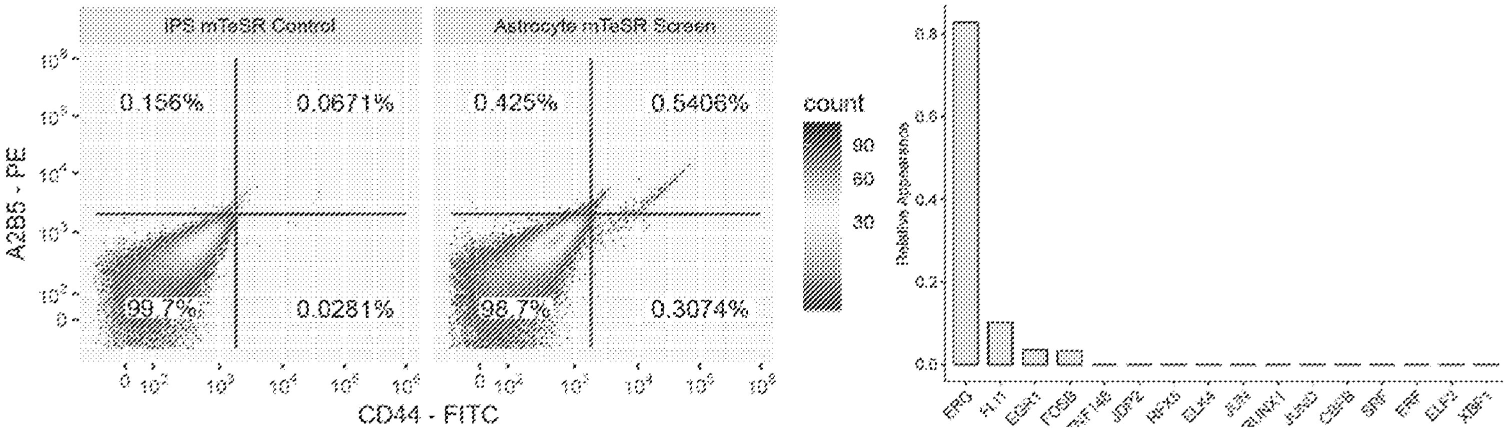

IPS mTeSR Control
Astrocyte mTeSR Screen
A2B5 - PE
CD44 - FITC
0.156%
0.0671%
99.7%
0.0281%
0.425%
0.5406%
98.7%
0.3074%
10^6
10^5
10^4
10^3
10^2
0
count
90
60
30
Relative Appearance
0.8
0.6
0.4
0.2
0.0
ERG
FLI1
EGR1
FOSB
ZNF148
JDP2
HEXS
ELK4
JUN
RUNX1
JUND
CEBPB
SRF
ERF
ELF1
XBP1

FIG. 3A

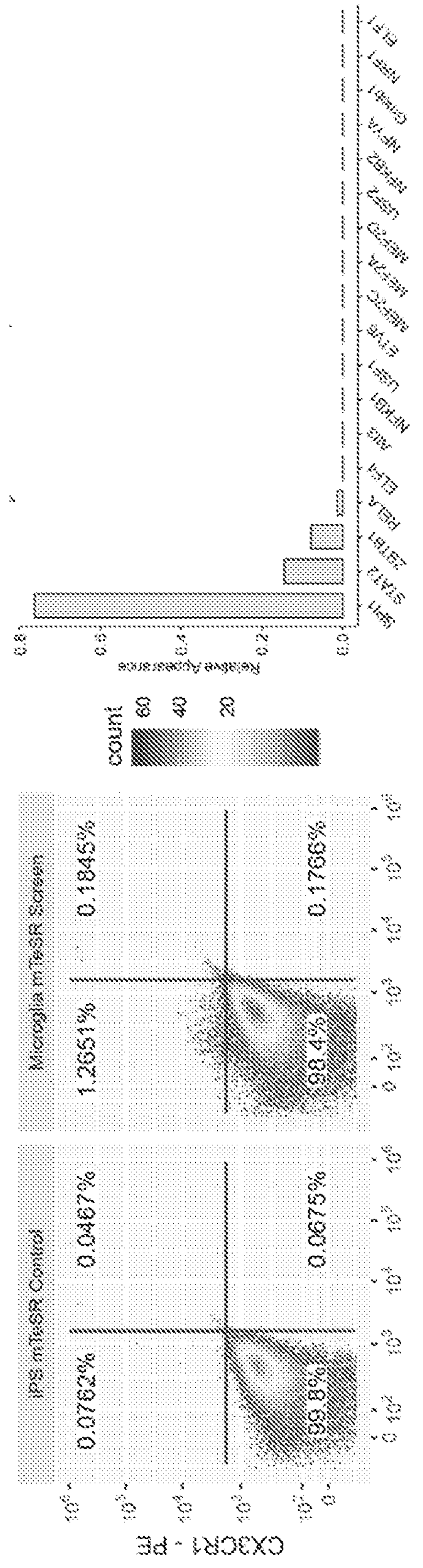

IPS mTeSR Control
0.0762%
0.0467%
99.8%
0.0675%
Microglia mTeSR Screen
1.2651%
0.1845%
98.4%
0.1766%
CX3CR1 - PE
0
10^2
10^3
10^4
10^5
10^6
count
60
40
20
Relative Appearance
0.0
0.2
0.4
0.6
0.8
Spi1
STAT2
STAT1
RELA
ELF4
Atf3
NFKB1
USF1
ETV6
MEF2C
MEF2A
MEF2D
USF2
NFKB2
NFYA
Creb1
Nrf1
Elf1

FIG. 3B

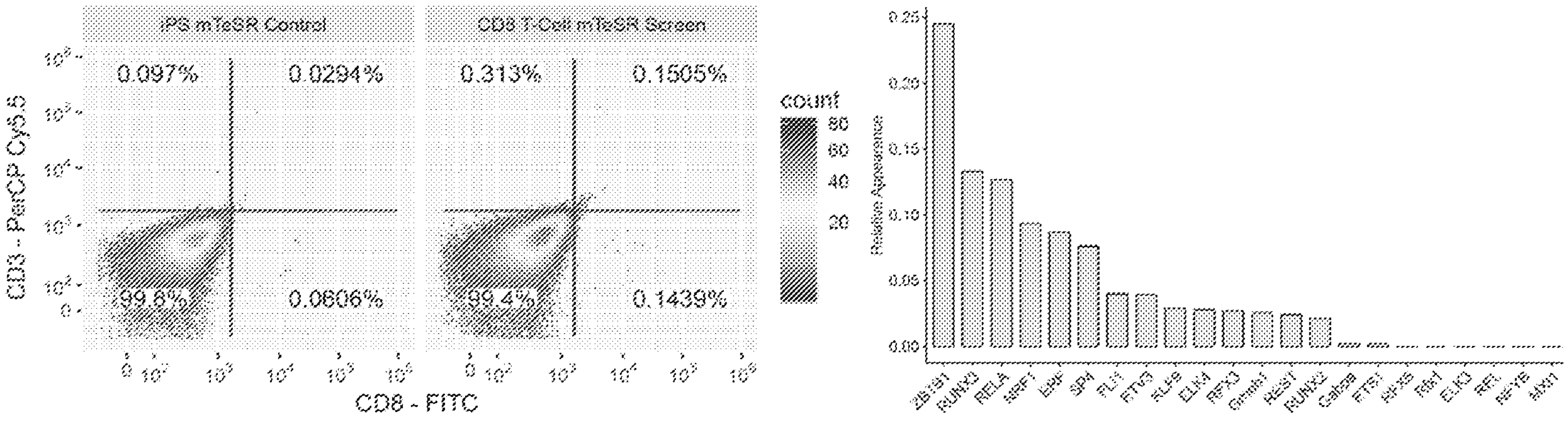

iPS mTeSR Control
CD8 T-Cell mTeSR Screen
CD3 - PerCP Cy5.5
CD8 - FITC
10^5
10^4
10^3
10^2
0
10^2
10^3
10^4
10^5
10^6
0.097%
0.0294%
99.8%
0.0606%
0.313%
0.1505%
99.4%
0.1439%
count
80
60
40
20
Relative Appearance
0.25
0.20
0.15
0.10
0.05
0.00
ZBTB1
RUNX3
RELA
NRF1
EHF
SP4
FLI1
ETV3
KLF9
ELK4
RFX3
Gmeb1
HES1
RUNX2
Gabpa
ETS1
RFX5
Rfx1
ELK3
REL
NFYB
MXI1

FIG. 3C

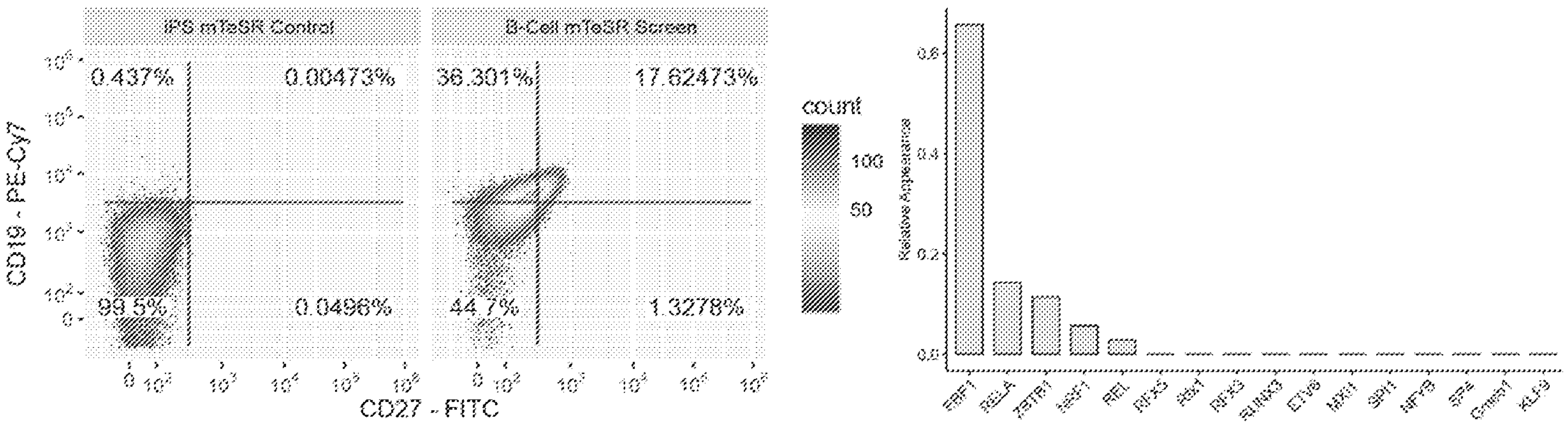

iPS mTeSR Control
B-Cell mTeSR Screen
0.437%
0.00473%
99.5%
0.0496%
36.301%
17.62473%
44.7%
1.3278%
CD19 - PE-Cy7
CD27 - FITC
0
10^2
10^3
10^4
10^5
10^6
count
100
50
Relative Appearance
0.0
0.2
0.4
0.6
EBF1
RELA
ZBTB1
IKZF1
REL
RFX5
Pbx1
RFX3
RUNX3
ETV6
MXI1
SPI1
NFYB
SP2
Gmeb1
KLF9

FIG. 3D

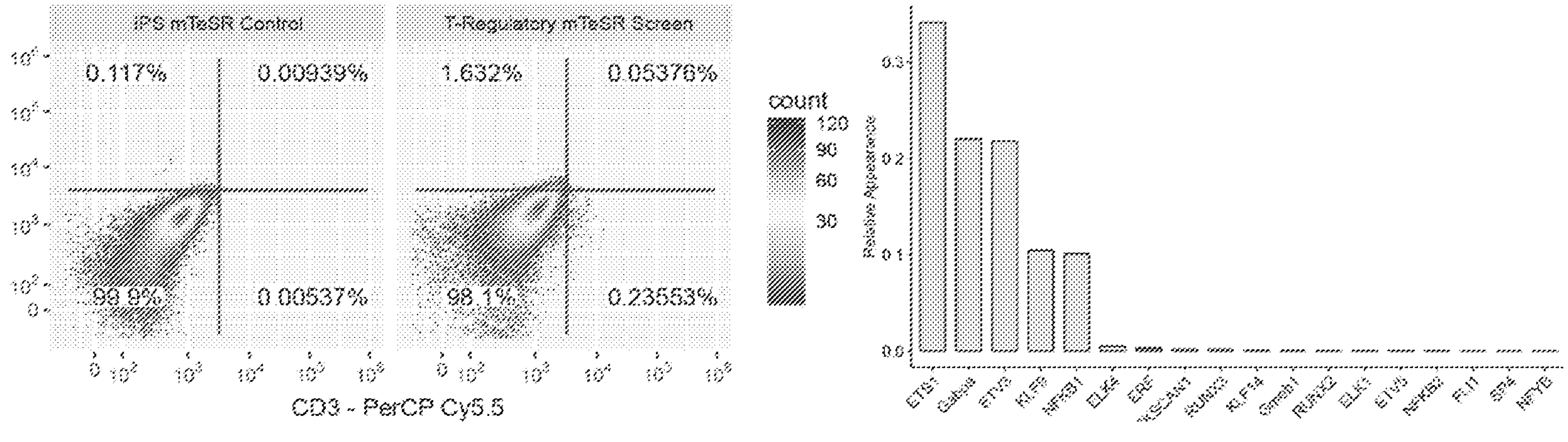

iPS mTeSR Control
T-Regulatory mTeSR Screen
0.117%
0.00939%
99.9%
0.00537%
1.632%
0.05376%
98.1%
0.23553%
CD25 - PE
CD3 - PerCP Cy5.5
10^5
10^4
10^3
10^2
0
count
120
90
60
30
Relative Appearance
0.3
0.2
0.1
0.0
ETS1
Gabpa
ETV6
KLF6
NFKB1
ELK4
ERF
ZKSCAN1
RUNX3
KLF14
Gmeb1
RUNX2
ELK3
ETV5
NFKB2
FLI1
SP4
NFYB

FIG. 3E

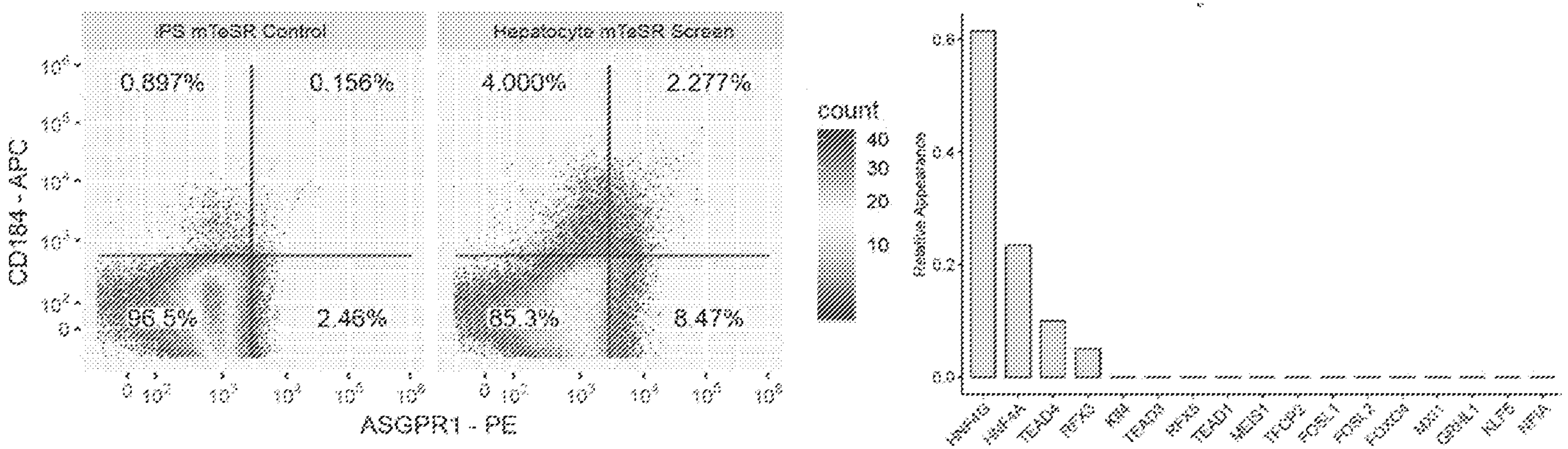
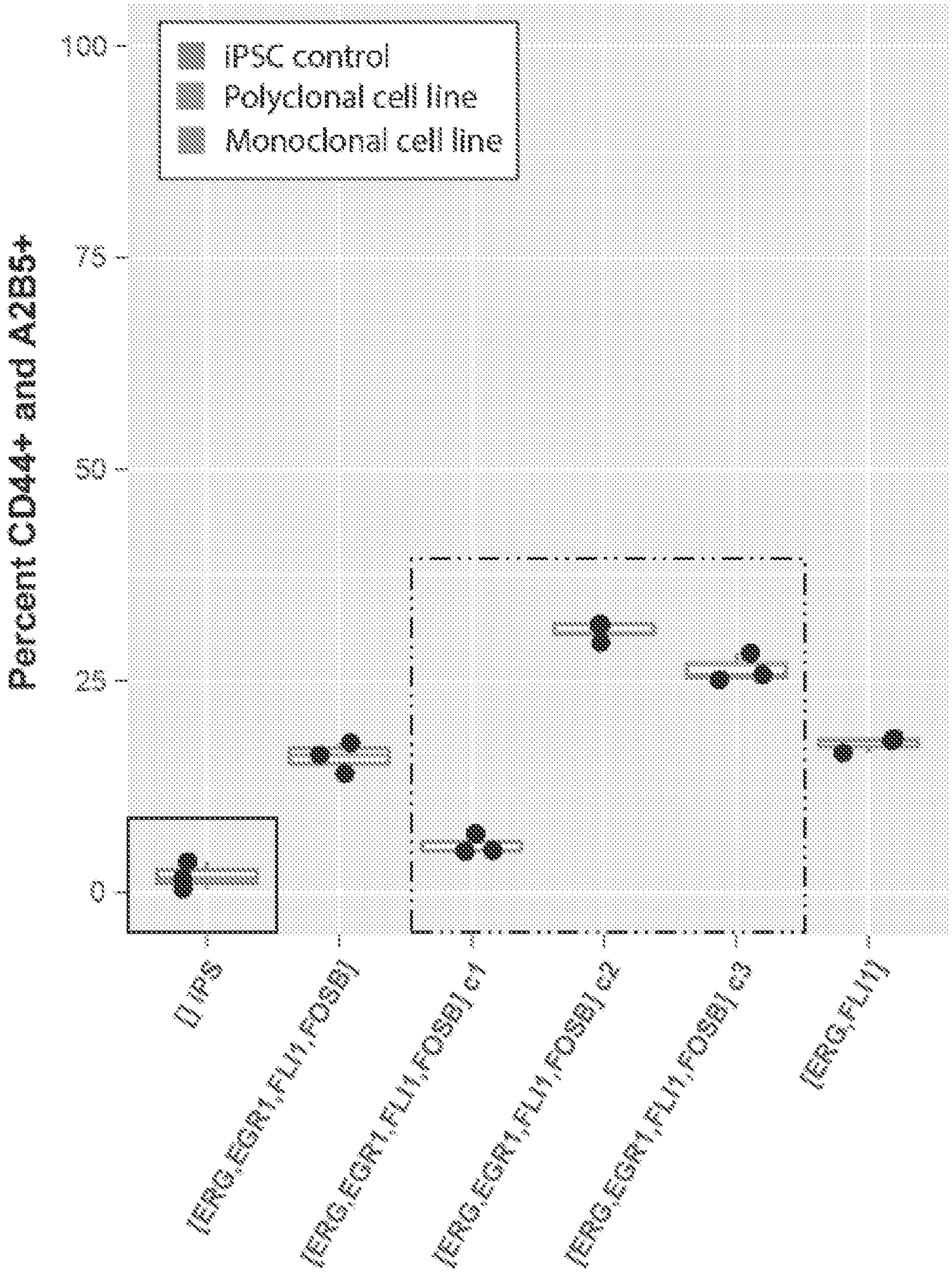

iPS mTeSR Control
Hepatocyte mTeSR Screen
CD184 - APC
10^6
10^5
10^4
10^3
10^2
0
0.897%
0.156%
96.5%
2.46%
4.000%
2.277%
85.3%
8.47%
0
10^2
10^3
10^4
10^5
10^6
ASGPR1 - PE
count
40
30
20
10
Relative Appearance
0.6
0.4
0.2
0.0
HNF4G
HNF4A
TEAD4
RFX3
KI4
TEAD3
RFX5
TEAD1
MEIS1
TFCP2
FOSL1
FOSL2
FOXO4
MXI1
GRHL1
KLF6
NFIA

FIG. 3F



Percent CD44+ and A2B5+
100
75
50
25
0
▨ iPSC control
▨ Polyclonal cell line
▨ Monoclonal cell line
[] iPS
[ERG,EGR1,FLI1,FOSB]
[ERG,EGR1,FLI1,FOSB] c1
[ERG,EGR1,FLI1,FOSB] c2
[ERG,EGR1,FLI1,FOSB] c3
[ERG,FLI1]

FIG. 4A

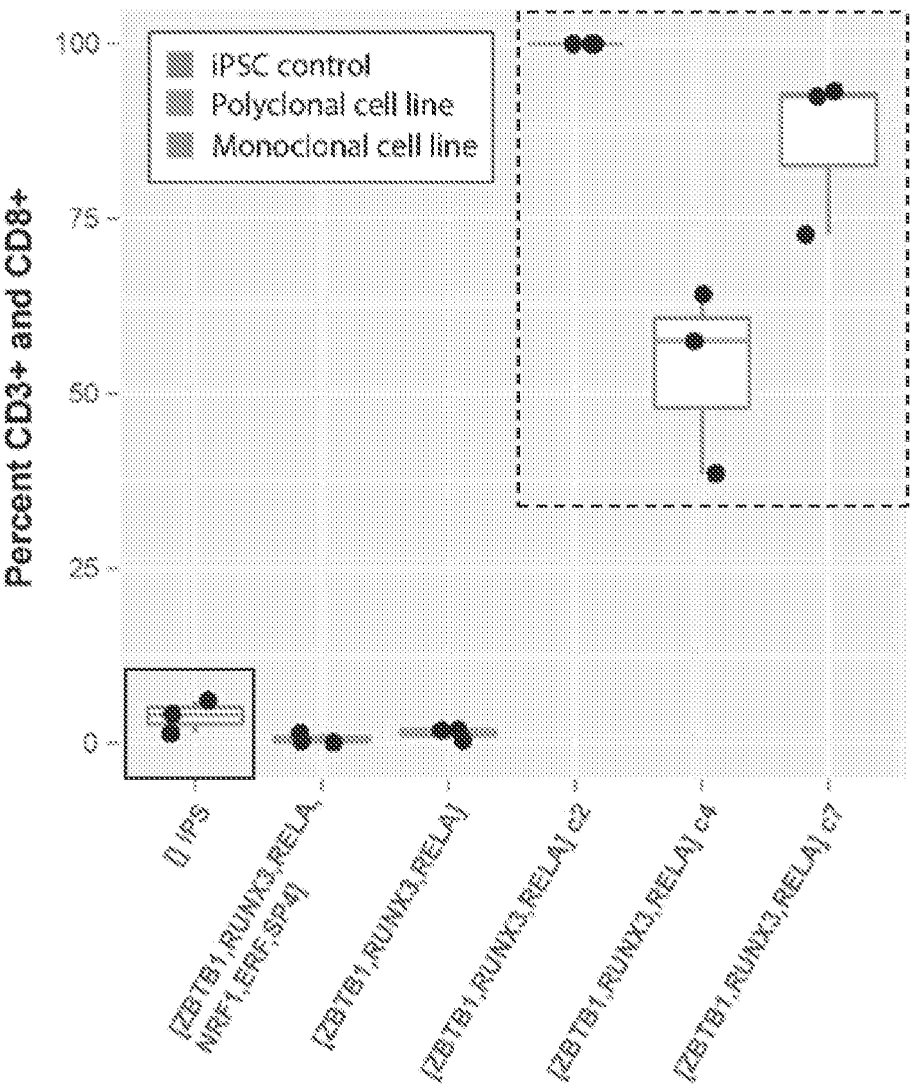

iPSC control
Polyclonal cell line
Monoclonal cell line
Percent CD3+ and CD8+
100
75
50
25
0
[] iPS
[ZBTB1,RUNX3,RELA, NRF1,ERF,SP4]
[ZBTB1,RUNX3,RELA]
[ZBTB1,RUNX3,RELA] c2
[ZBTB1,RUNX3,RELA] c4
[ZBTB1,RUNX3,RELA] c7

FIG. 4B

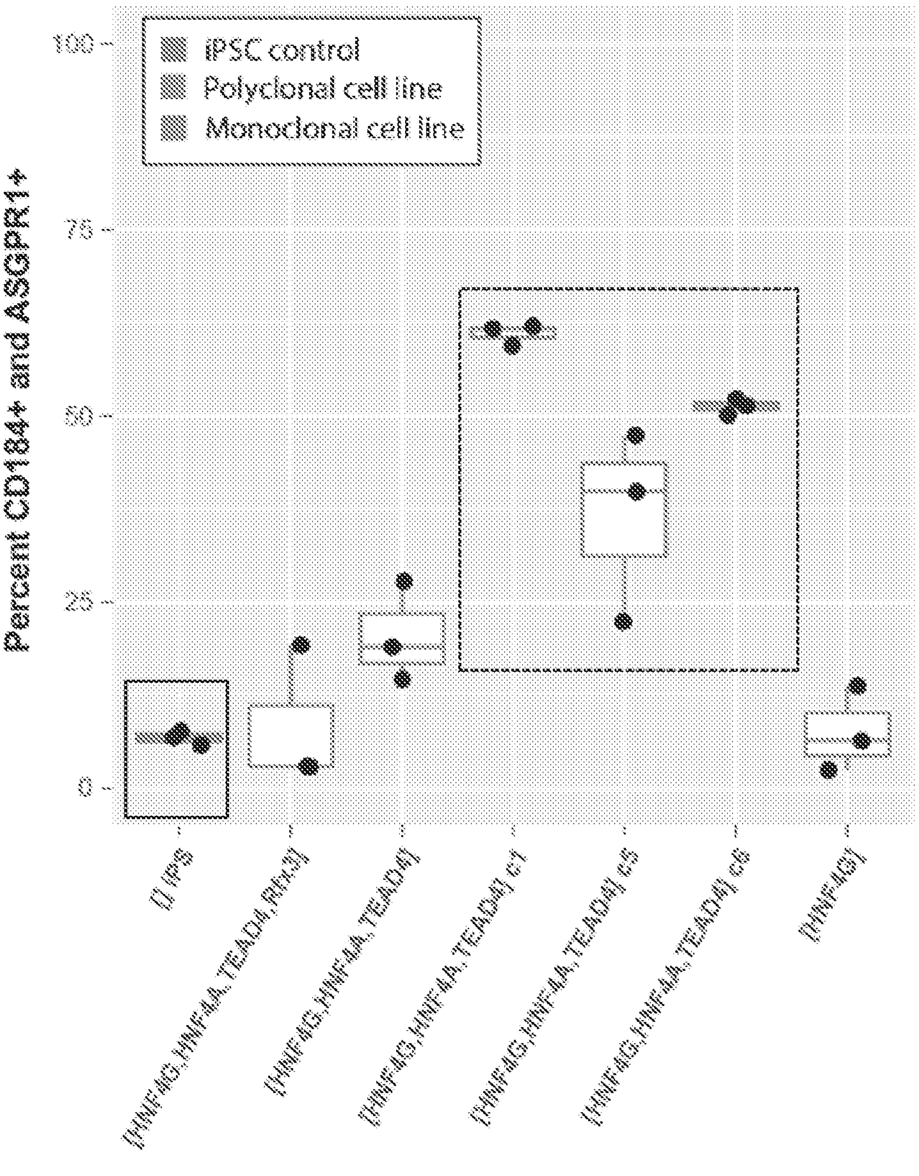

Percent CD184+ and ASGPR1+
100
75
50
25
0
▨ iPSC control
▨ Polyclonal cell line
▨ Monoclonal cell line
[] iPS
[HNF4G,HNF4A,TEAD4,REX3]
[HNF4G,HNF4A,TEAD4]
[HNF4G,HNF4A,TEAD4] c1
[HNF4G,HNF4A,TEAD4] c5
[HNF4G,HNF4A,TEAD4] c6
[HNF4G]

FIG. 4C

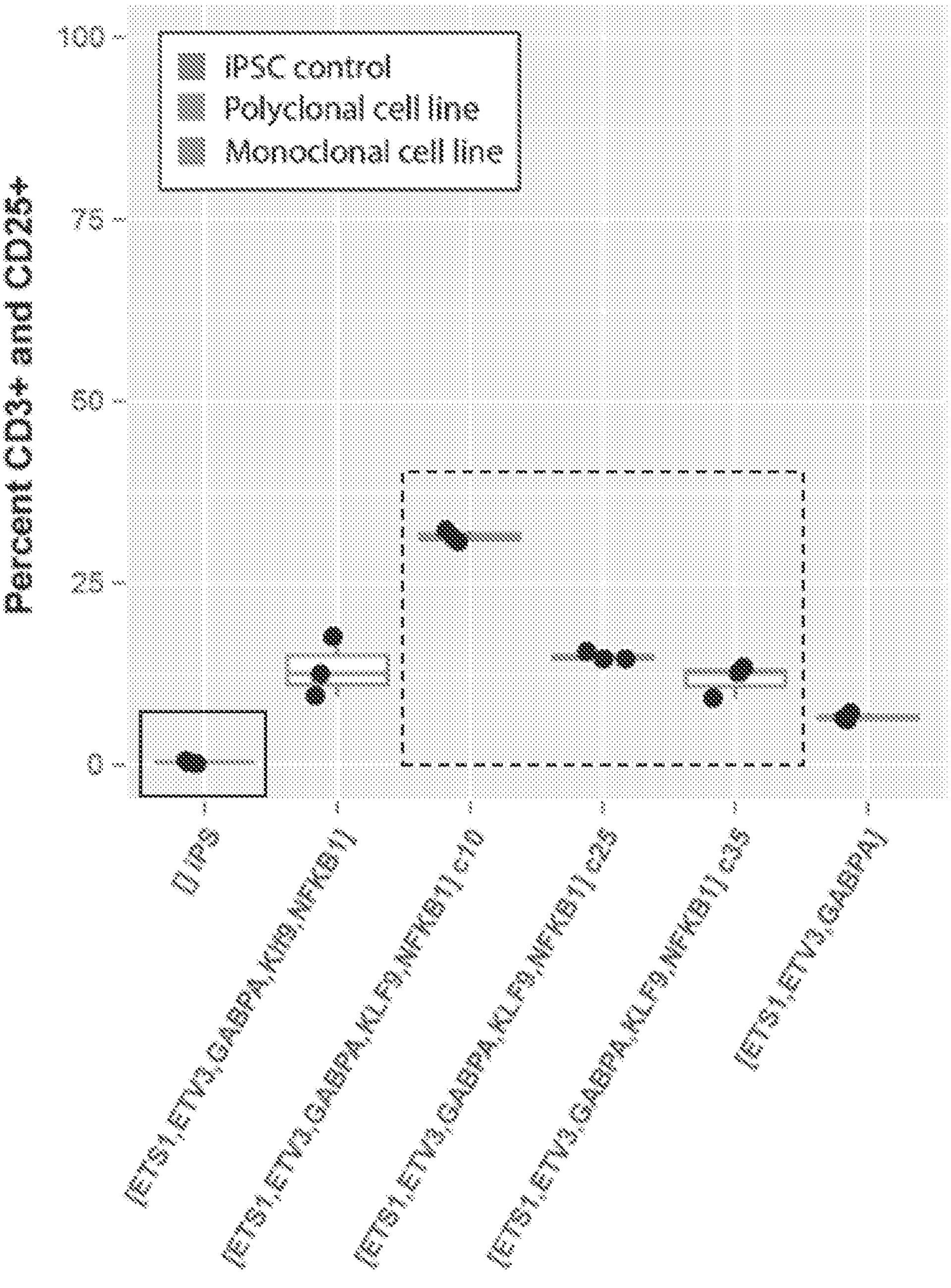

Percent CD3+ and CD25+
100
75
50
25
0
▨ iPSC control
▨ Polyclonal cell line
▨ Monoclonal cell line
[] iPS
[ETS1,ETV3,GABPA,KLF9,NFKB1]
[ETS1,ETV3,GABPA,KLF9,NFKB1] c10
[ETS1,ETV3,GABPA,KLF9,NFKB1] c25
[ETS1,ETV3,GABPA,KLF9,NFKB1] c35
[ETS1,ETV3,GABPA]

FIG. 4D

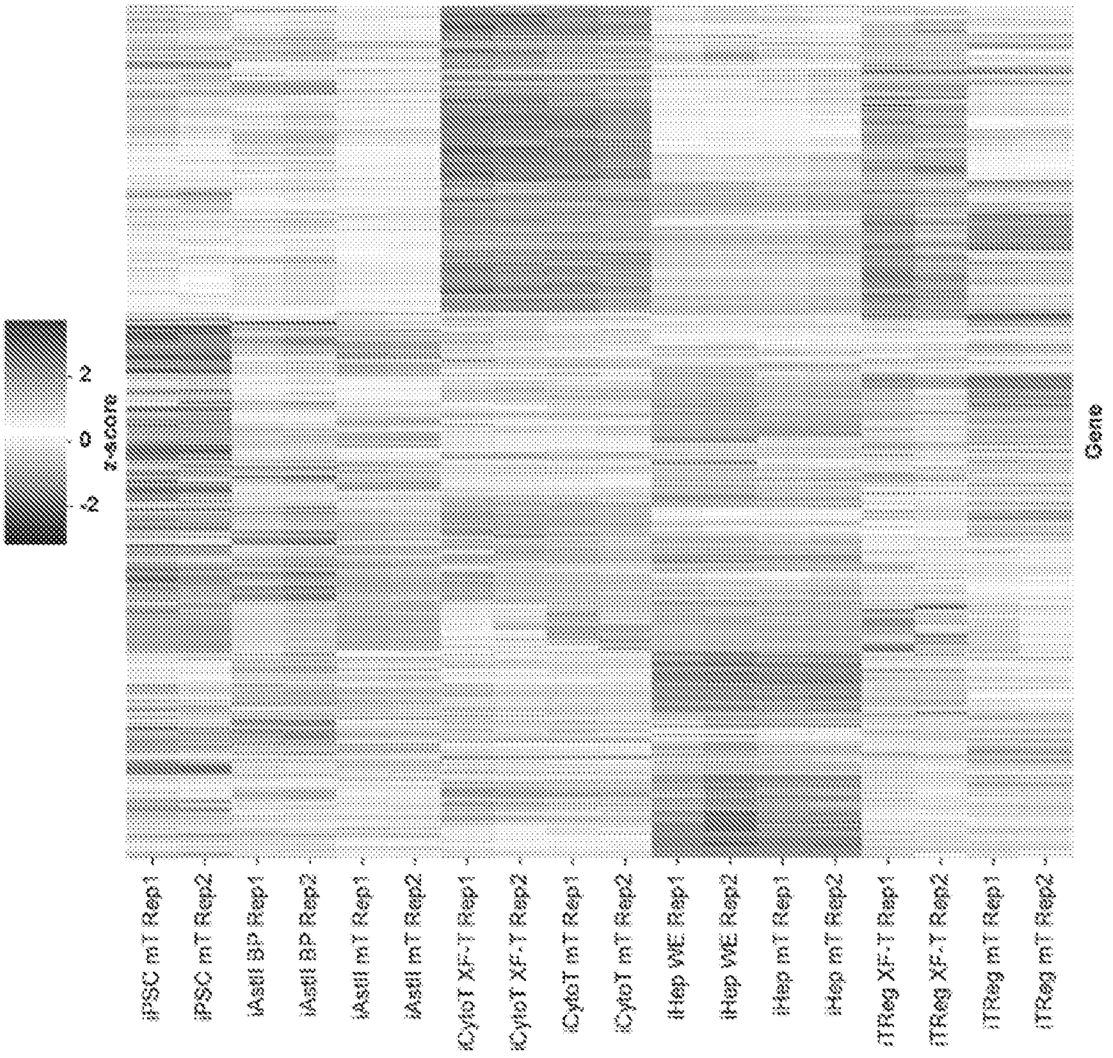

2
0
-2
z-score
Gene
iPSC mT Rep1
iPSC mT Rep2
iAstII BP Rep1
iAstII BP Rep2
iAstII mT Rep1
iAstII mT Rep2
iCytoT XF-T Rep1
iCytoT XF-T Rep2
iCytoT mT Rep1
iCytoT mT Rep2
iHep WE Rep1
iHep WE Rep2
iHep mT Rep1
iHep mT Rep2
iTReg XF-T Rep1
iTReg XF-T Rep2
iTReg mT Rep1
iTReg mT Rep2

FIG. 4E

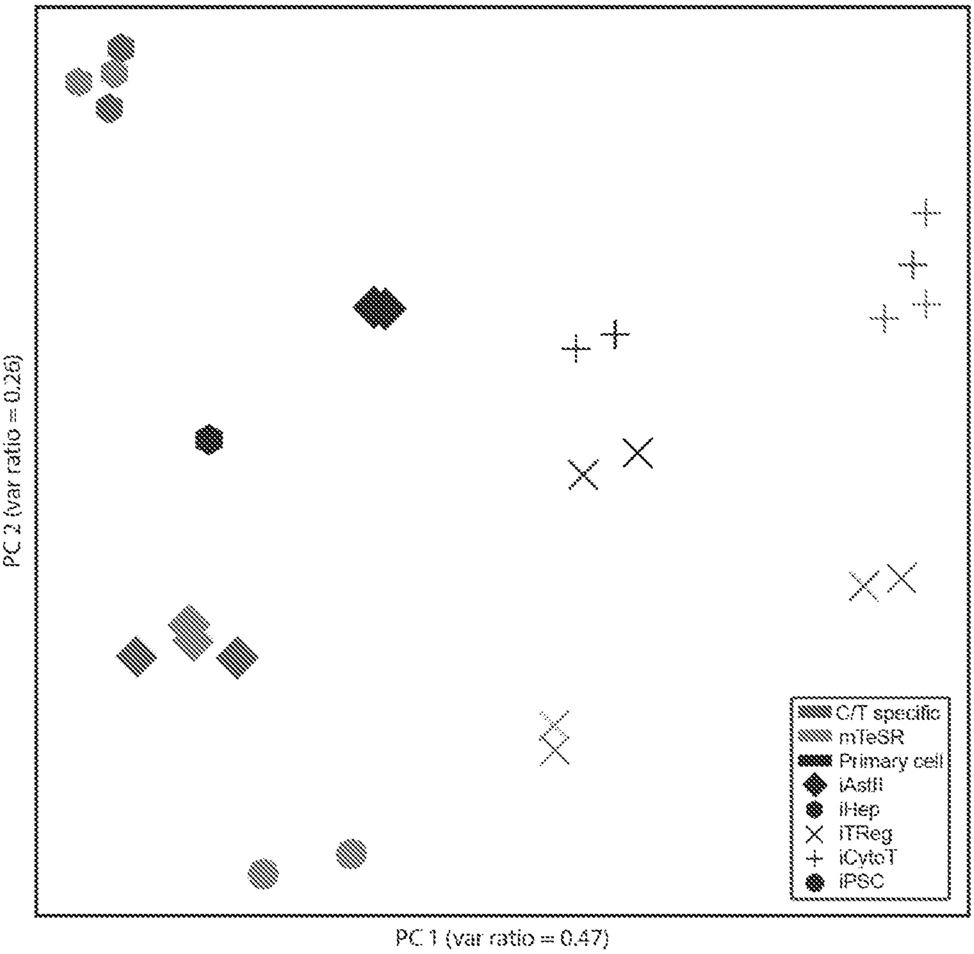

PC 2 (var ratio = 0.26)
PC 1 (var ratio = 0.47)
C/T specific
mTeSR
Primary cell
iAstII
iHep
iTReg
iCytoT
iPSC

FIG. 4F

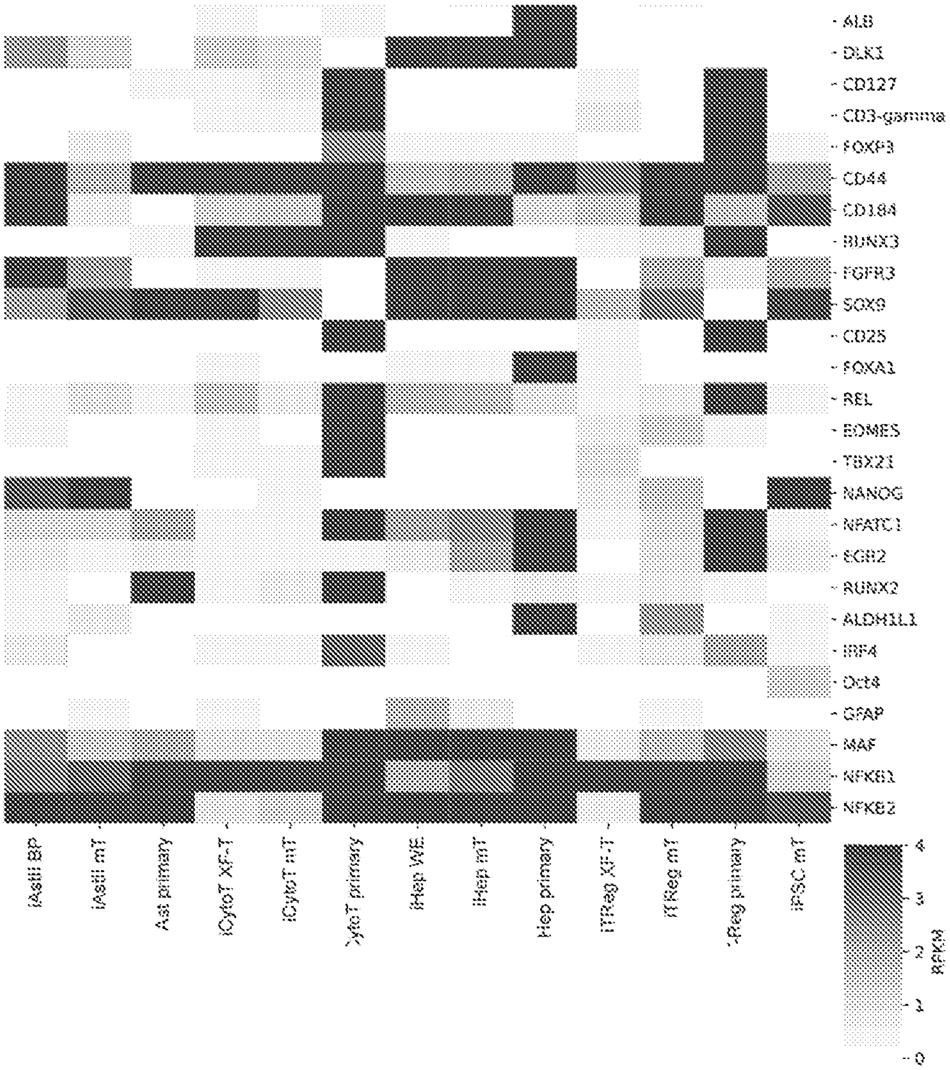
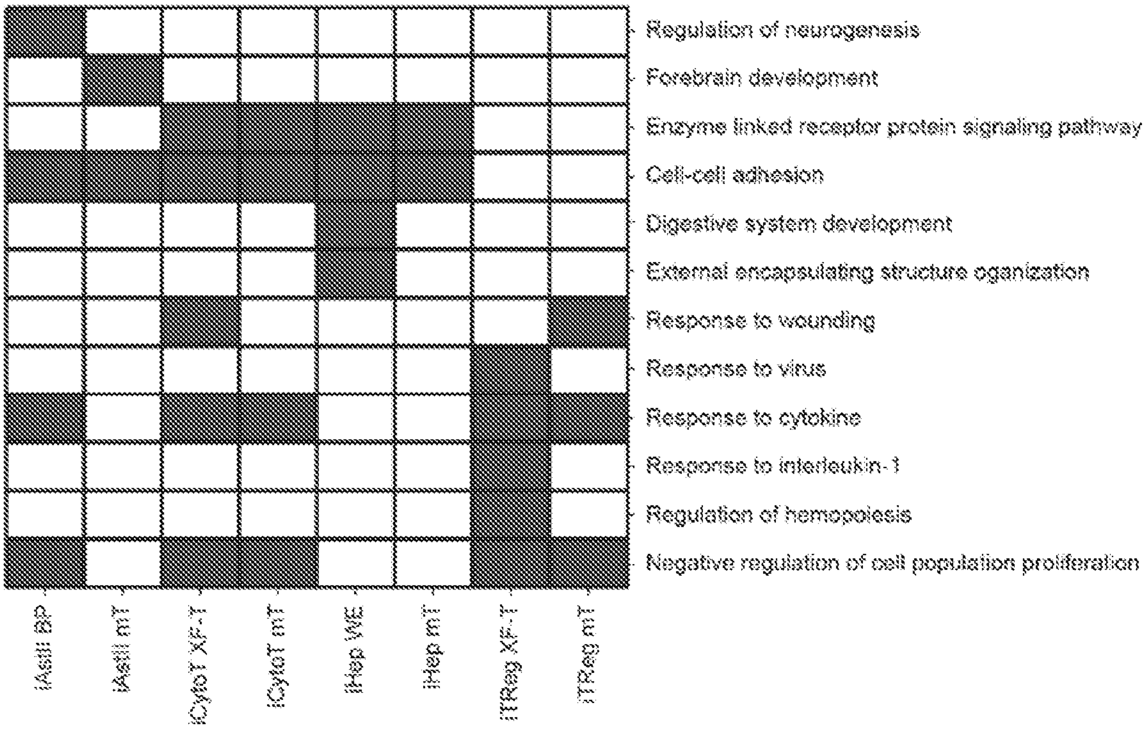

ALB
DLK1
CD127
CD3-gamma
FOXP3
CD44
CD184
RUNX3
FGFR3
SOX9
CD25
FOXA1
REL
EOMES
TBX21
NANOG
NFATC1
EGR2
RUNX2
ALDH1L1
IRF4
Oct4
GFAP
MAF
NFKB1
NFKB2
iAstII BP
iAstII mT
Ast primary
iCytoT XF-T
iCytoT mT
CytoT primary
iHep WE
iHep mT
Hep primary
iTReg XF-T
iTReg mT
TReg primary
iPSC mT
4
3
2
1
0
RPKM

FIG. 4G



	iAstII BP	iAstII mT	iCytoT XF-T	iCytoT mT	iHep WE	iHep mT	iTReg XF-T	iTReg mT
Regulation of neurogenesis	▨							
Forebrain development		▨						
Enzyme linked receptor protein signaling pathway			▨	▨	▨	▨		
Cell-cell adhesion	▨	▨	▨	▨	▨	▨		
Digestive system development					▨			
External encapsulating structure oganization					▨			
Response to wounding			▨					▨
Response to virus							▨	
Response to cytokine	▨		▨	▨			▨	▨
Response to interleukin-1							▨	
Regulation of hemopoiesis							▨	
Negative regulation of cell population proliferation	▨		▨	▨			▨	▨

FIG. 4H

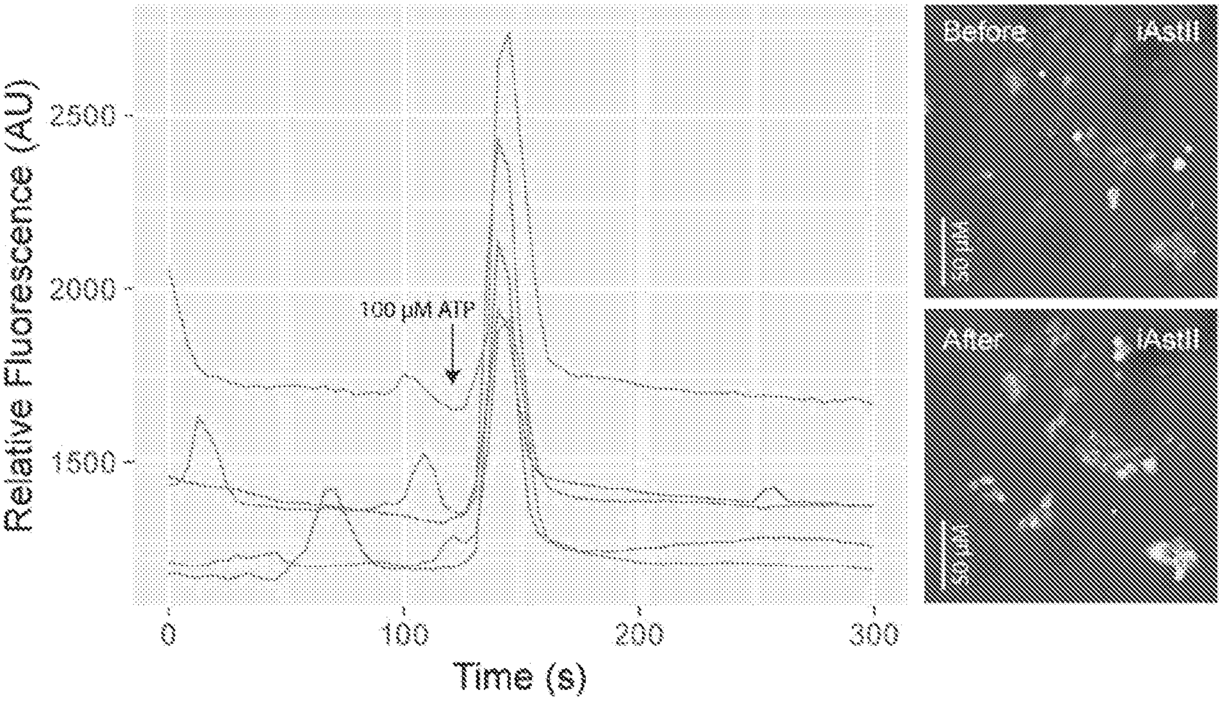

Relative Fluorescence (AU)
2500
2000
1500
100 µM ATP
0
100
200
300
Time (s)
Before
iAstII
50 µM
After
iAstII
50 µM

FIG. 5A

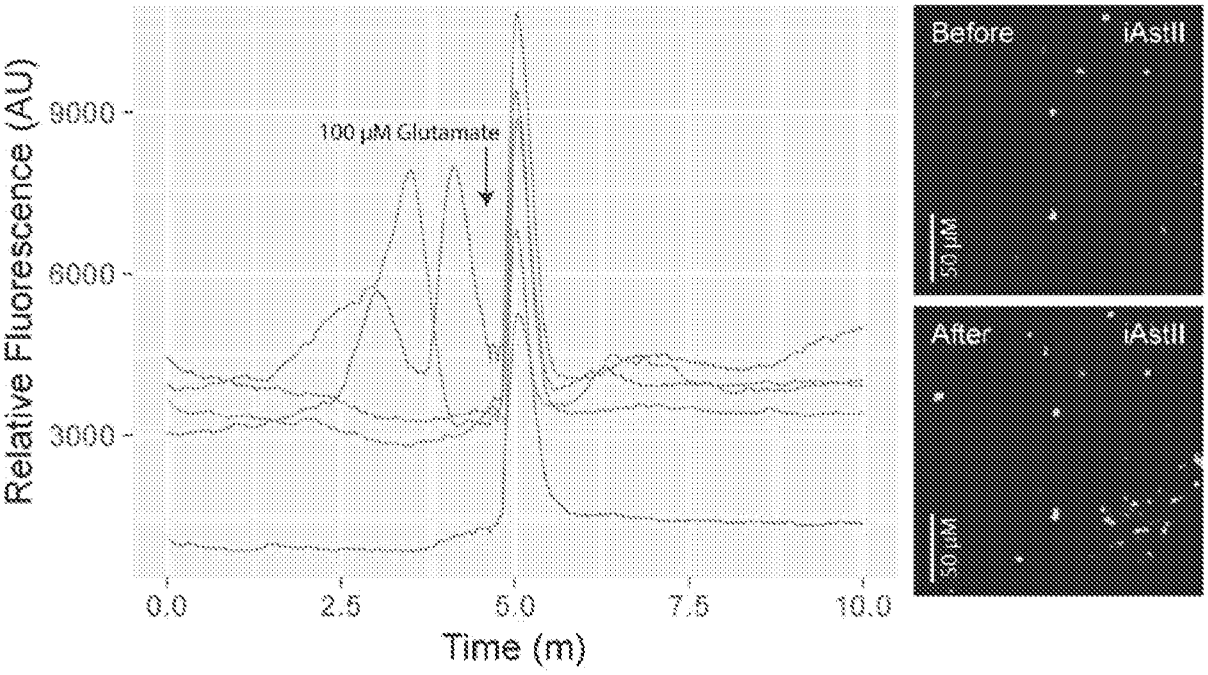

Relative Fluorescence (AU)
9000
6000
3000
100 μM Glutamate
0.0
2.5
5.0
7.5
10.0
Time (m)
Before
iAstII
50 μM
After
iAstII
50 μM

FIG. 5B

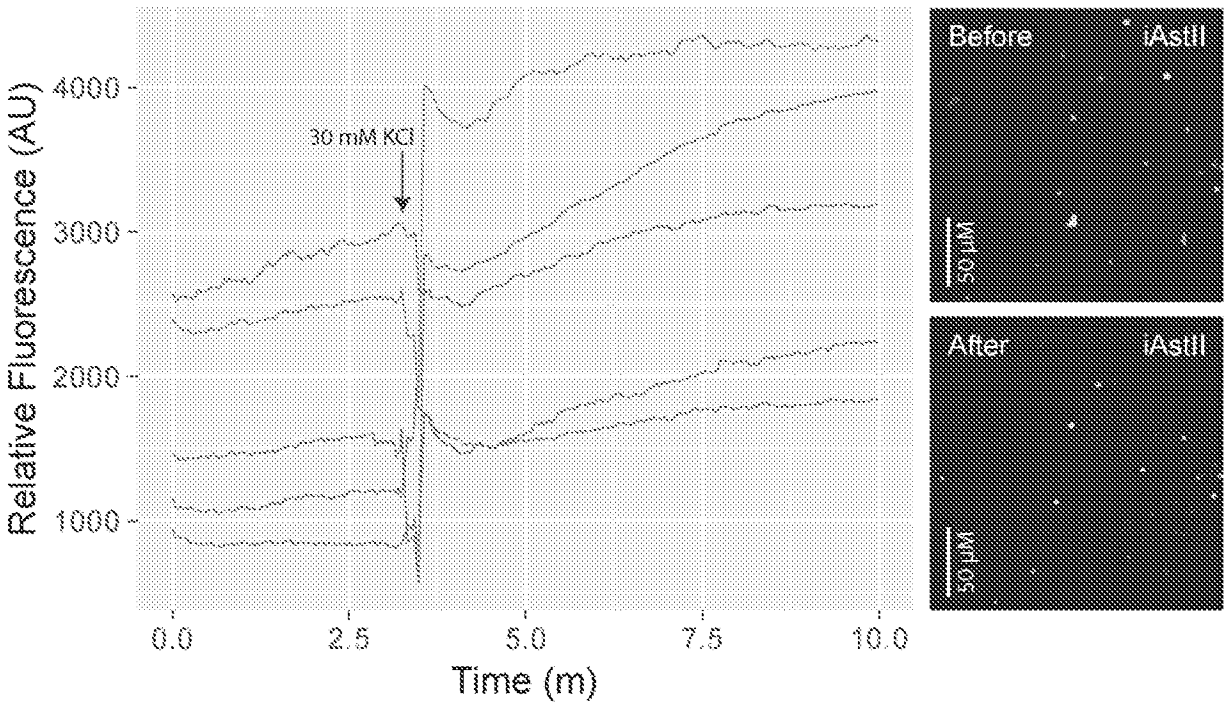

Relative Fluorescence (AU)
4000
3000
2000
1000
30 mM KCl
0.0
2.5
5.0
7.5
10.0
Time (m)
Before
iAstII
50 µM
After
iAstII
50 µM

FIG. 5C

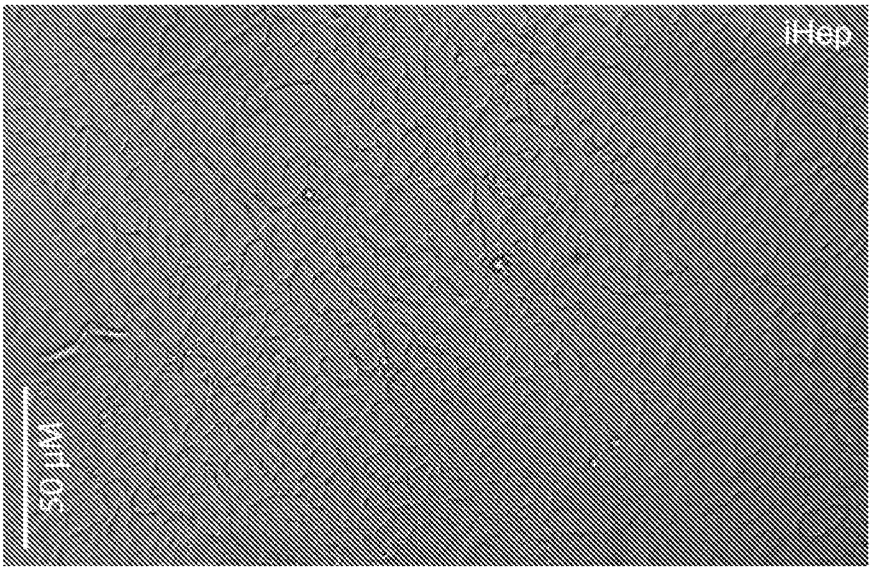

iHep
50 µM

FIG. 5D

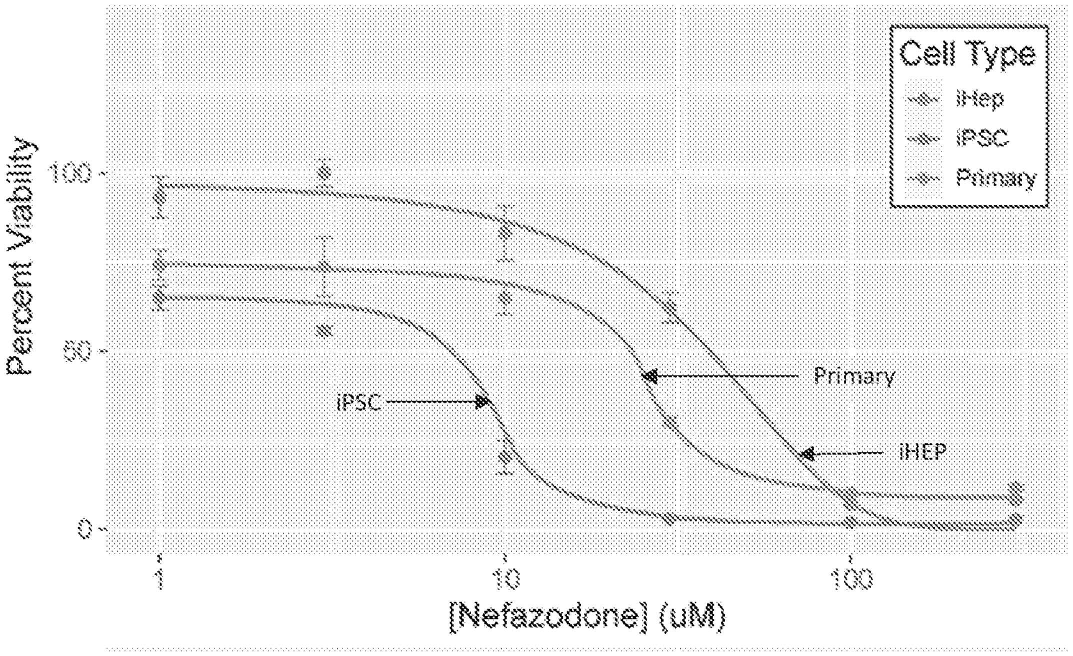

Cell Type
iHep
iPSC
Primary
Percent Viability
100
50
0
1
10
100
[Nefazodone] (uM)
Primary
iPSC
iHEP

FIG. 5E

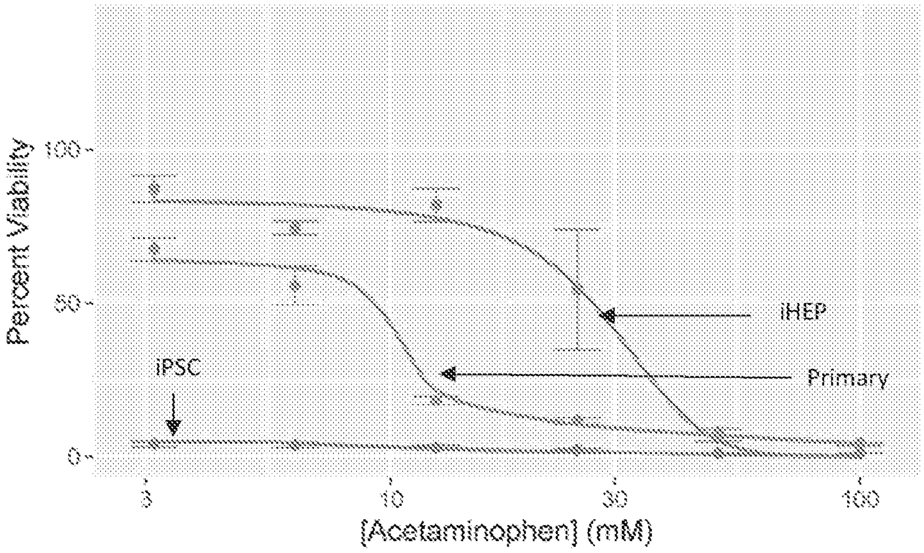

Percent Viability
100
50
0
3
10
30
100
[Acetaminophen] (mM)
iHEP
Primary
iPSC

FIG. 5F

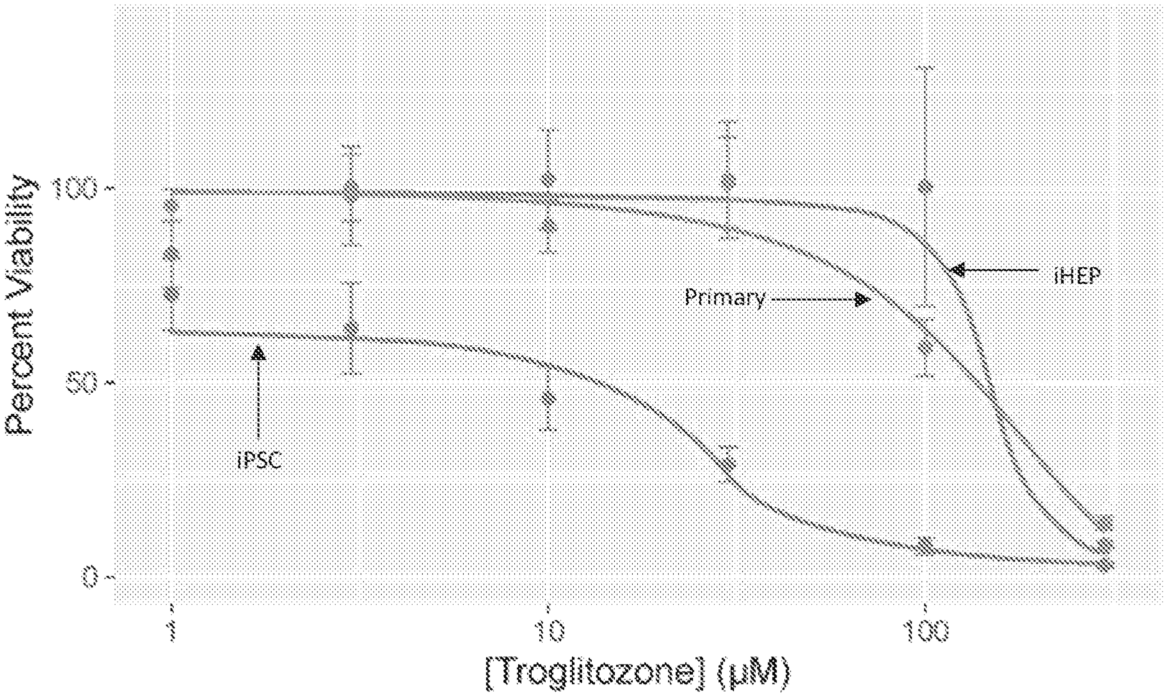

Percent Viability
100
50
0
1
10
100
[Troglitozone] (μM)
iHEP
Primary
iPSC

FIG. 5G

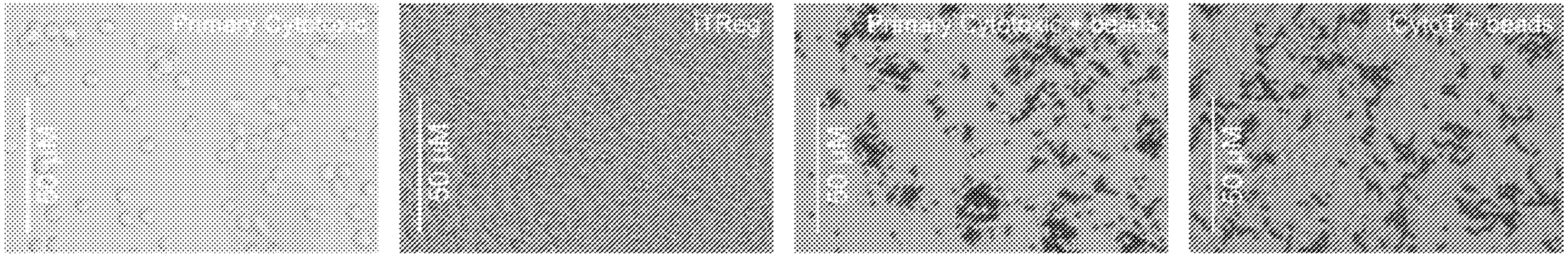

Primary Cytotoxic
50 μM
iTReg
50 μM
Primary Cytotoxic + beads
50 μM
iCytoT + beads
50 μM

FIG. 5H

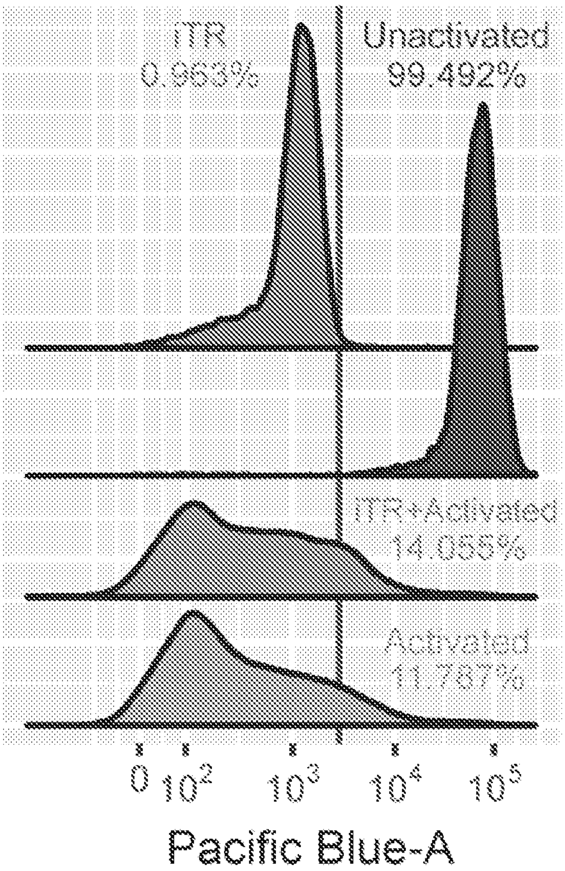

iTR
0.963%
Unactivated
99.492%
iTR+Activated
14.055%
Activated
11.787%
0
10^2
10^3
10^4
10^5
Pacific Blue-A

FIG. 5I

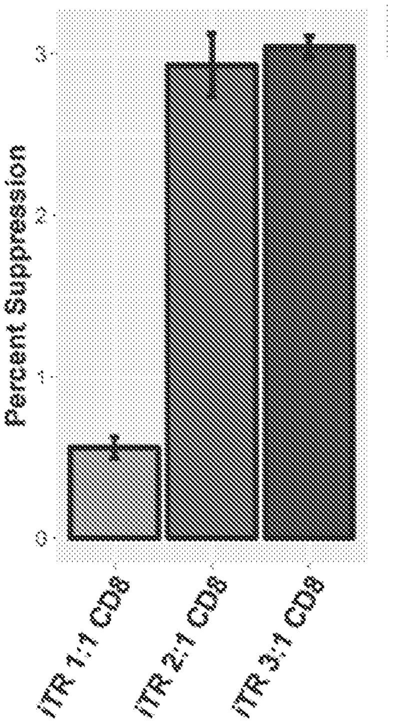


FIG. 5J

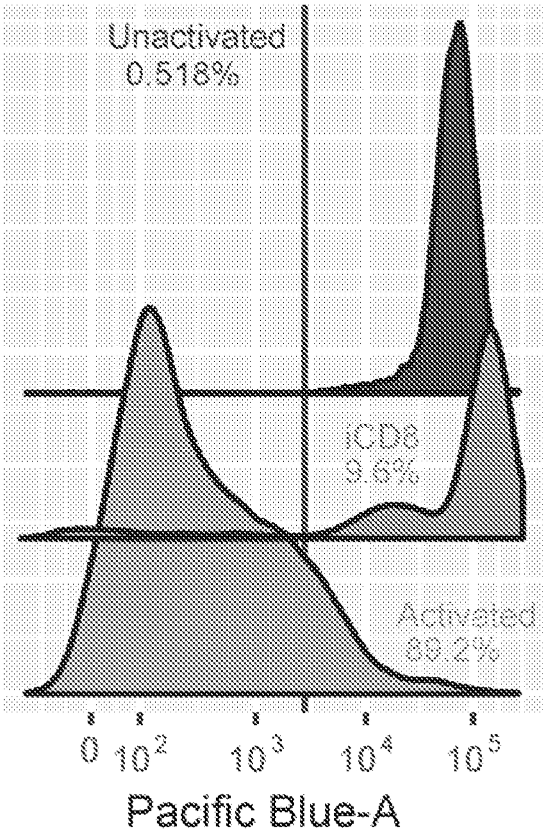

Unactivated
0.518%
iCD8
9.6%
Activated
89.2%
0
10^2
10^3
10^4
10^5
Pacific Blue-A

FIG. 5K

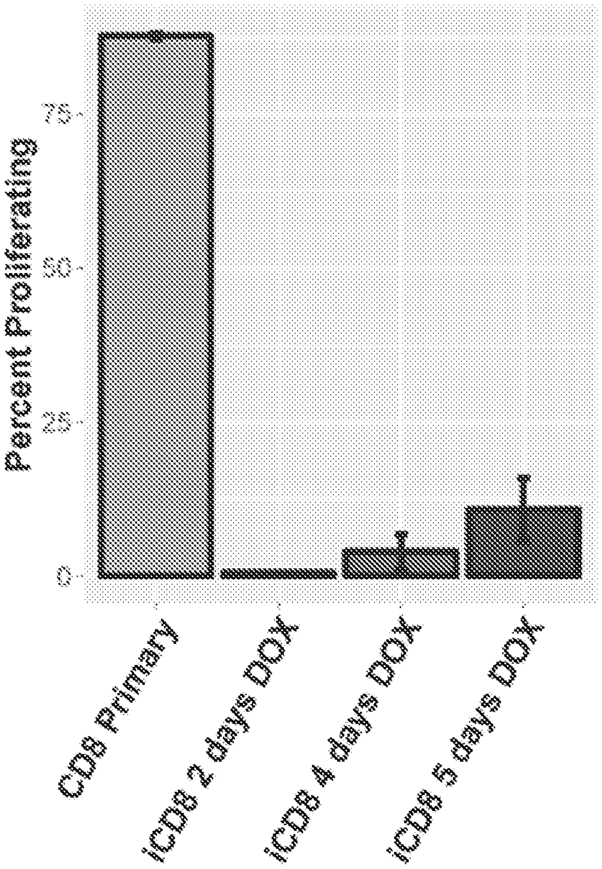

Percent Proliferating
75
50
25
0
CD8 Primary
iCD8 2 days DOX
iCD8 4 days DOX
iCD8 5 days DOX

FIG. 5L

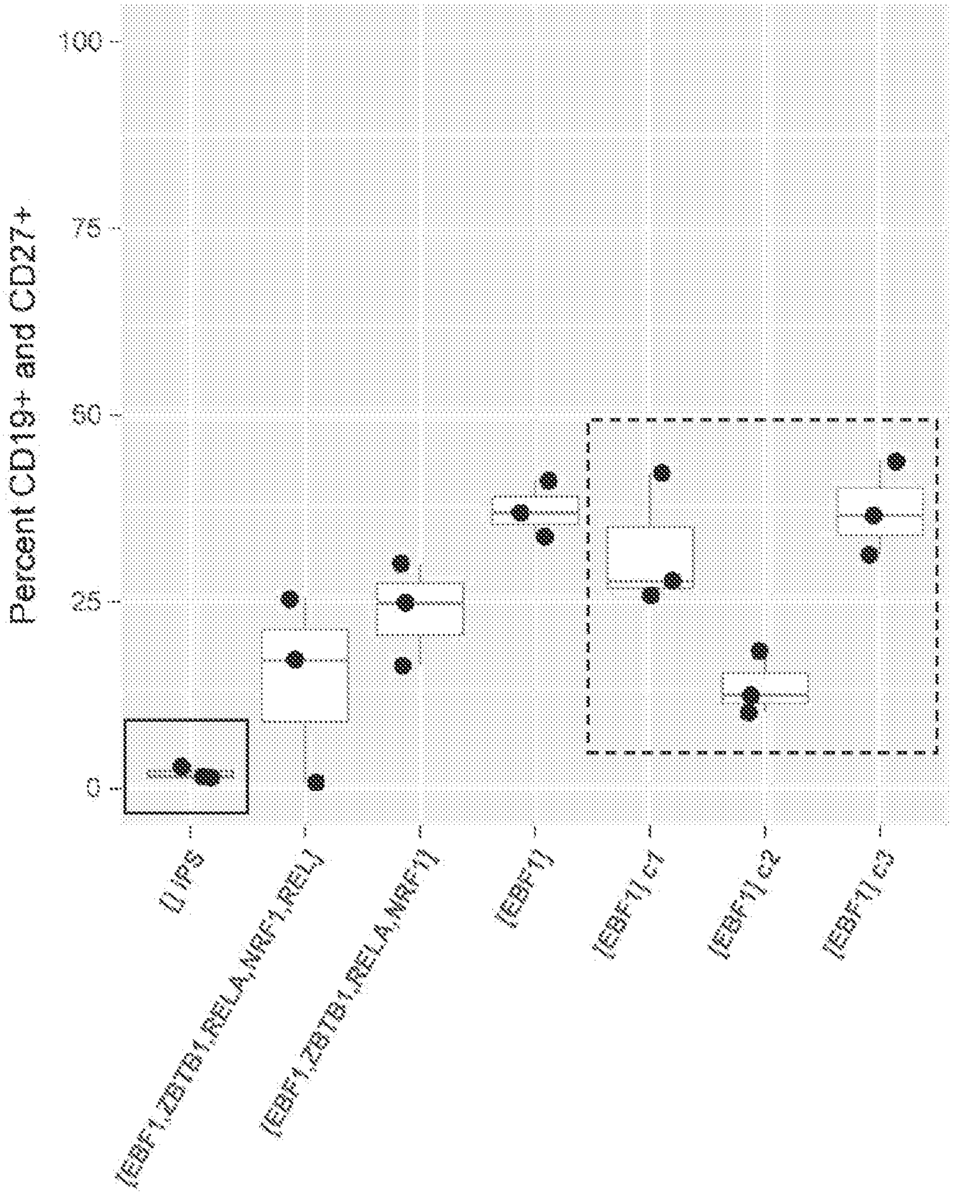

Percent CD19+ and CD27+
100
75
50
25
0
[] iPS
[EBF1,ZBTB1,RELA,NRF1,REL]
[EBF1,ZBTB1,RELA,NRF1]
[EBF1]
[EBF1] c1
[EBF1] c2
[EBF1] c3

FIG. 6A

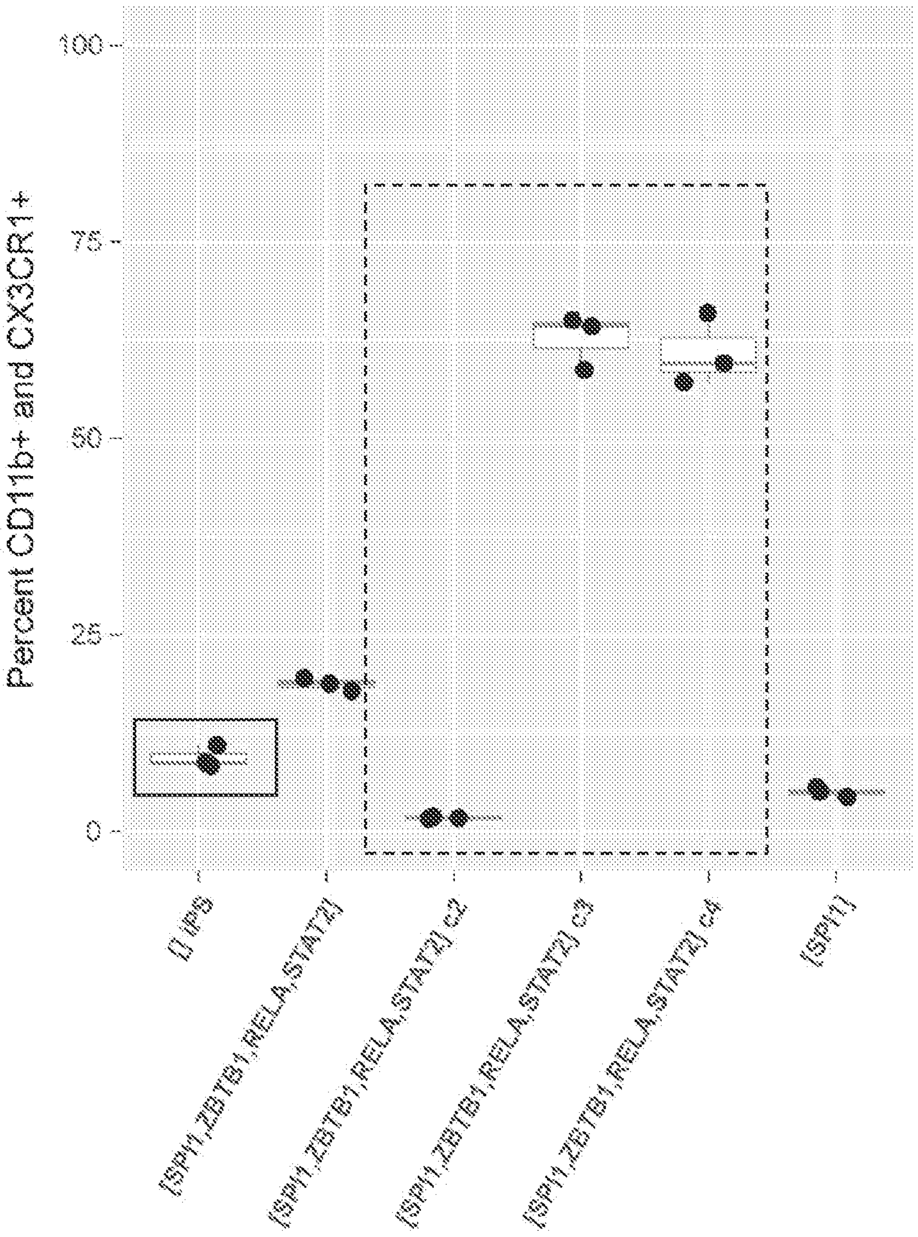

Percent CD11b+ and CX3CR1+
100
75
50
25
0
[] iPS
[SPI1,ZBTB1,RELA,STAT2]
[SPI1,ZBTB1,RELA,STAT2] c2
[SPI1,ZBTB1,RELA,STAT2] c3
[SPI1,ZBTB1,RELA,STAT2] c4
[SPI1]

FIG. 6B

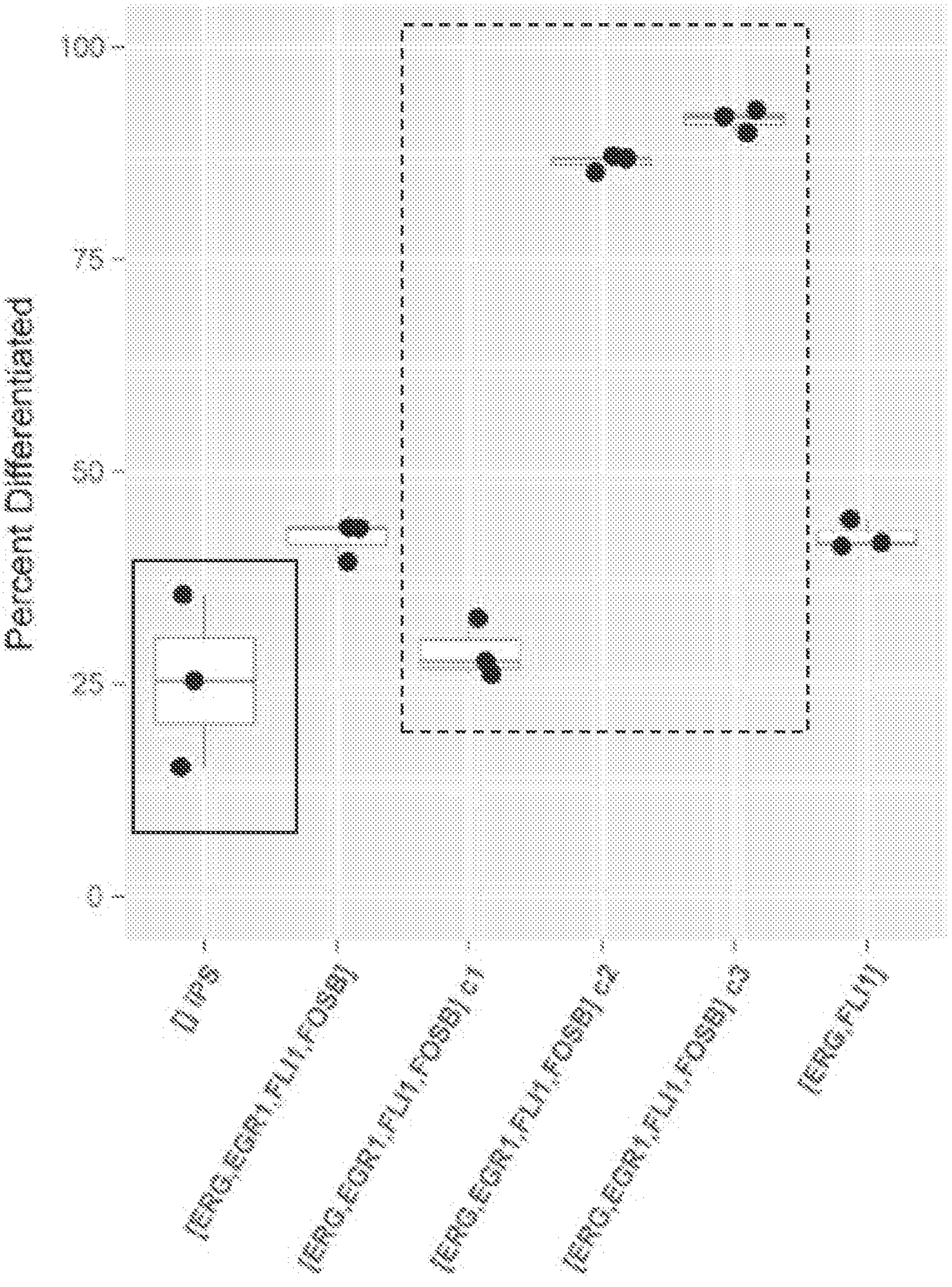

Percent Differentiated
100
75
50
25
0
[] iPS
[ERG,EGR1,FLI1,FOSB]
[ERG,EGR1,FLI1,FOSB] c1
[ERG,EGR1,FLI1,FOSB] c2
[ERG,EGR1,FLI1,FOSB] c3
[ERG,FLI1]

FIG. 7A

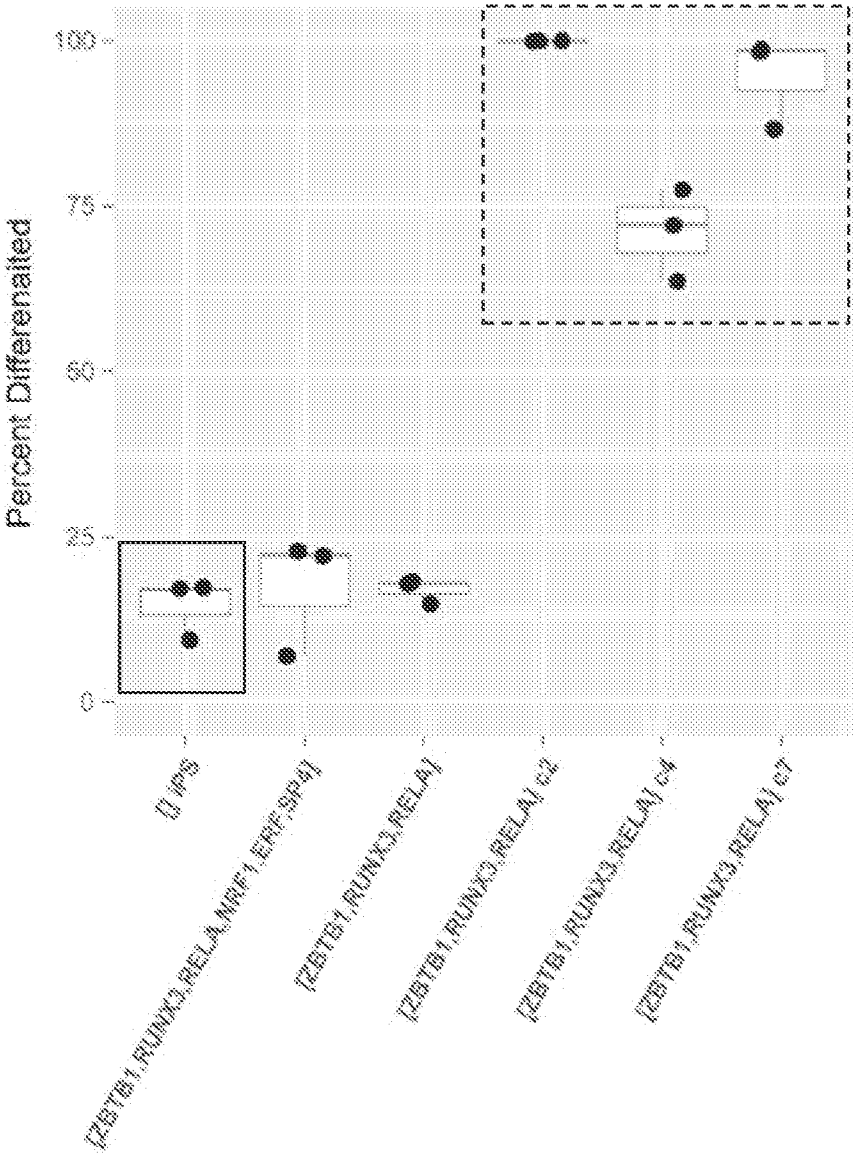

Percent Differenaited
100
75
50
25
0
[] iPS
[ZBTB1,RUNX3,RELA,NRF1,ERF,SP4]
[ZBTB1,RUNX3,RELA]
[ZBTB1,RUNX3,RELA] c2
[ZBTB1,RUNX3,RELA] c4
[ZBTB1,RUNX3,RELA] c7

FIG. 7B

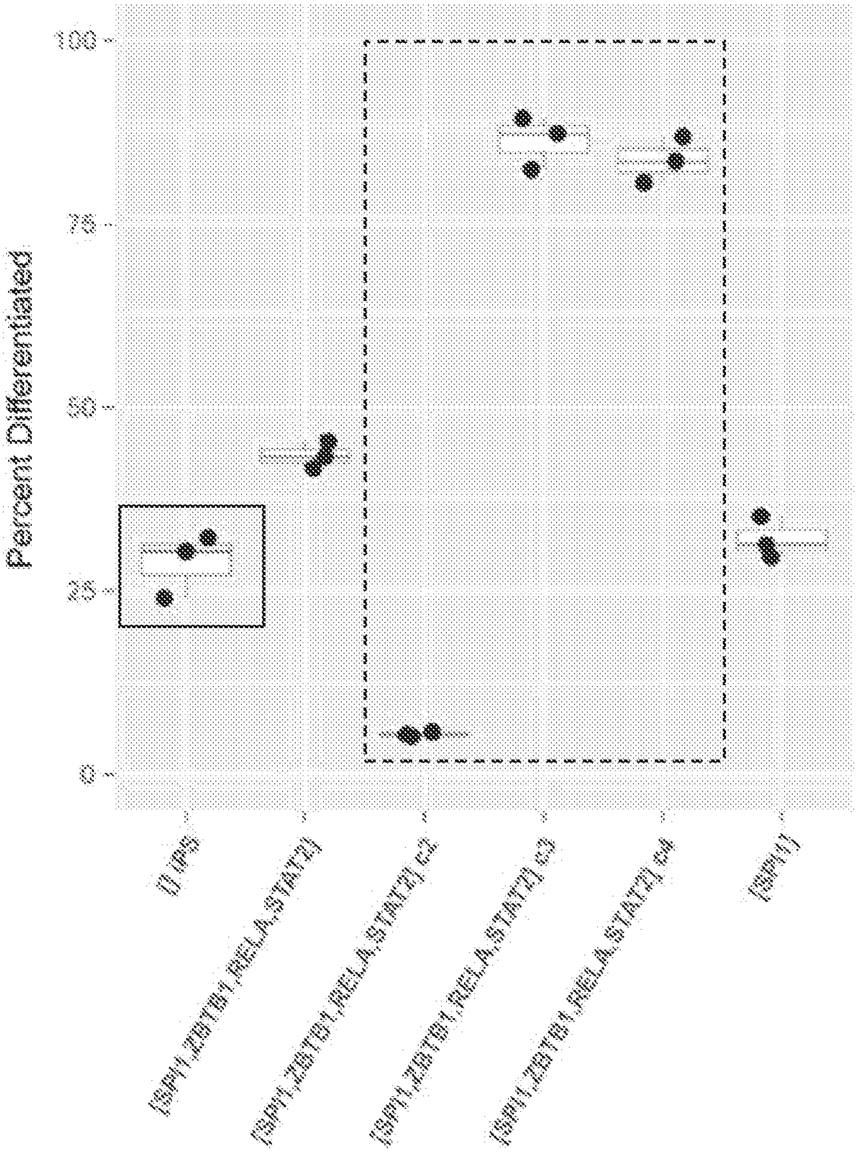

Percent Differentiated
100
75
50
25
0
[] UPS
[SPI1,ZBTB1,RELA,STAT2]
[SPI1,ZBTB1,RELA,STAT2] c2
[SPI1,ZBTB1,RELA,STAT2] c3
[SPI1,ZBTB1,RELA,STAT2] c4
[SPI1]

FIG. 7C

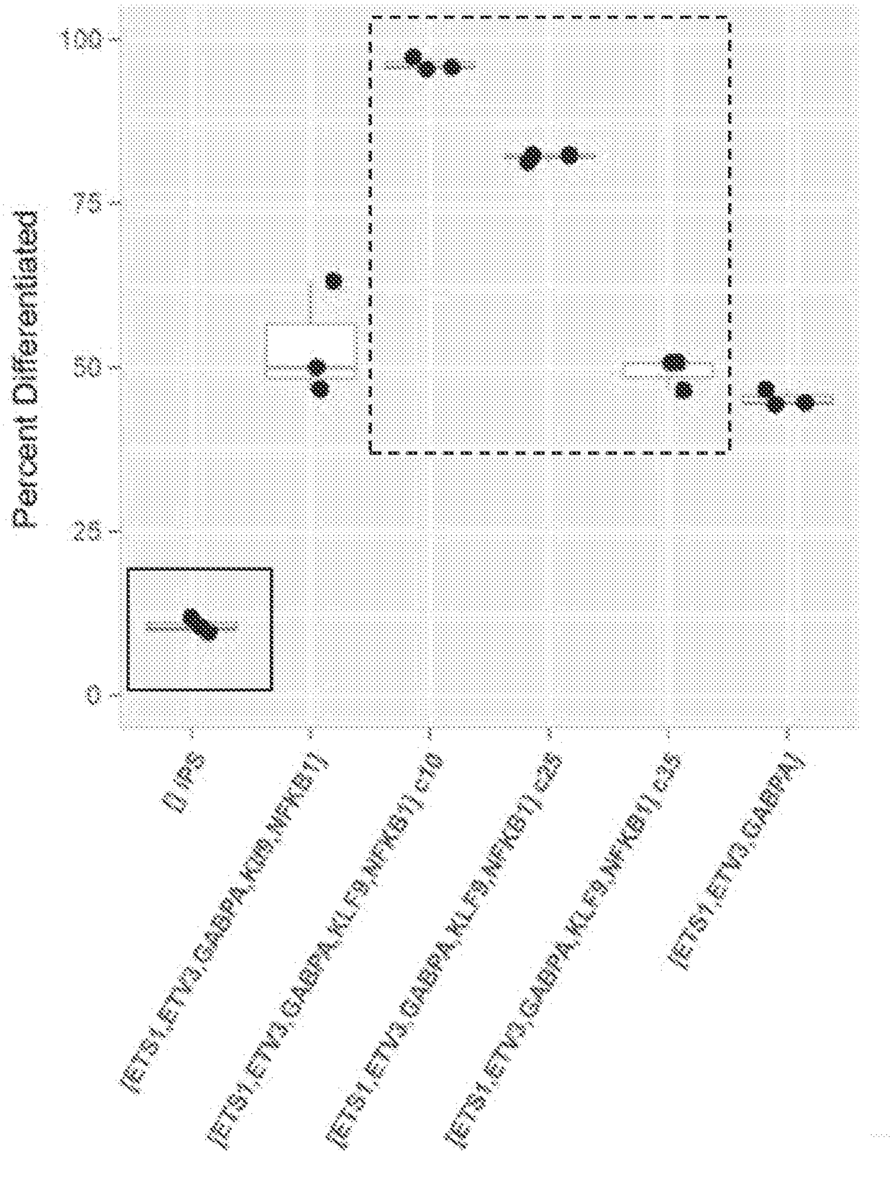

Percent Differentiated
100
75
50
25
0
[] iPS
[ETS1,ETV3,GABPA,KLF9,NFKB1]
[ETS1,ETV3,GABPA,KLF9,NFKB1] c10
[ETS1,ETV3,GABPA,KLF9,NFKB1] c25
[ETS1,ETV3,GABPA,KLF9,NFKB1] c35
[ETS1,ETV3,GABPA]

FIG. 7D

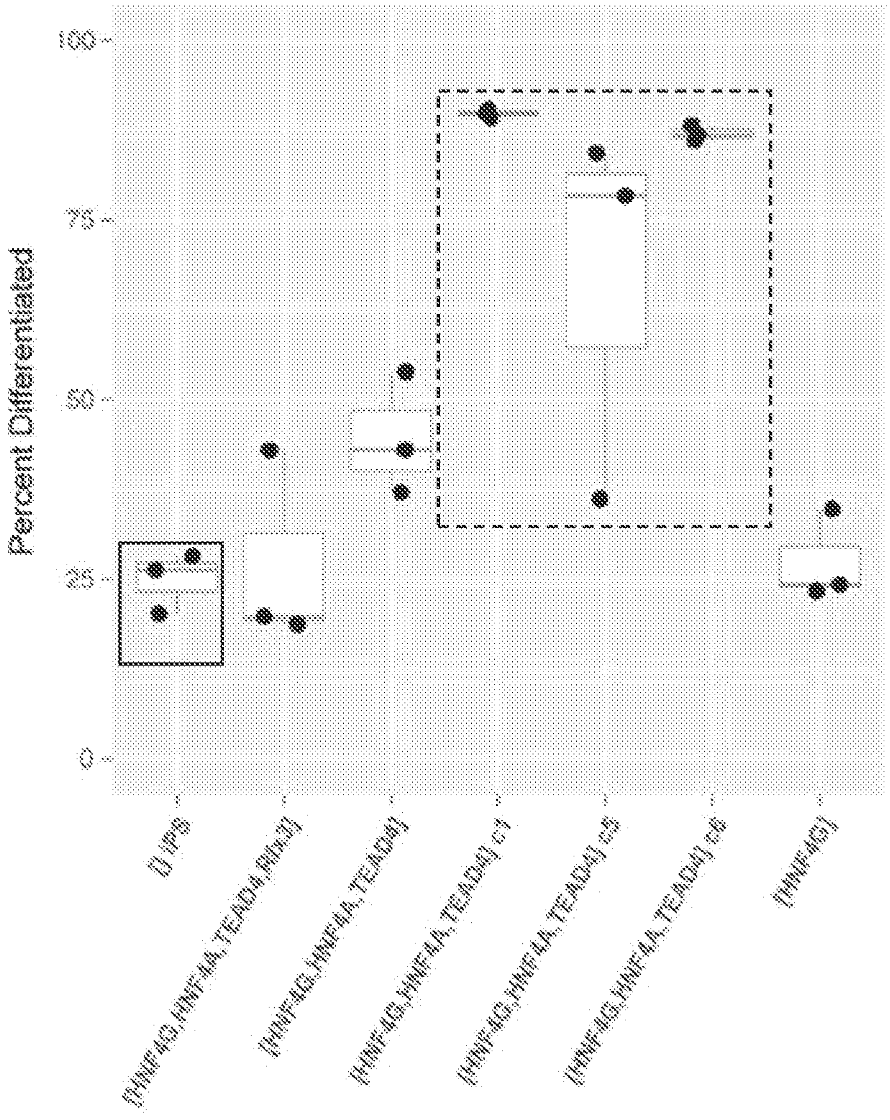

Percent Differentiated
100
75
50
25
0
[] iPS
[HNF4G,HNF4A,TEAD4,RFX3]
[HNF4G,HNF4A,TEAD4]
[HNF4G,HNF4A,TEAD4] c1
[HNF4G,HNF4A,TEAD4] c5
[HNF4G,HNF4A,TEAD4] c6
[HNF4G]

FIG. 7E

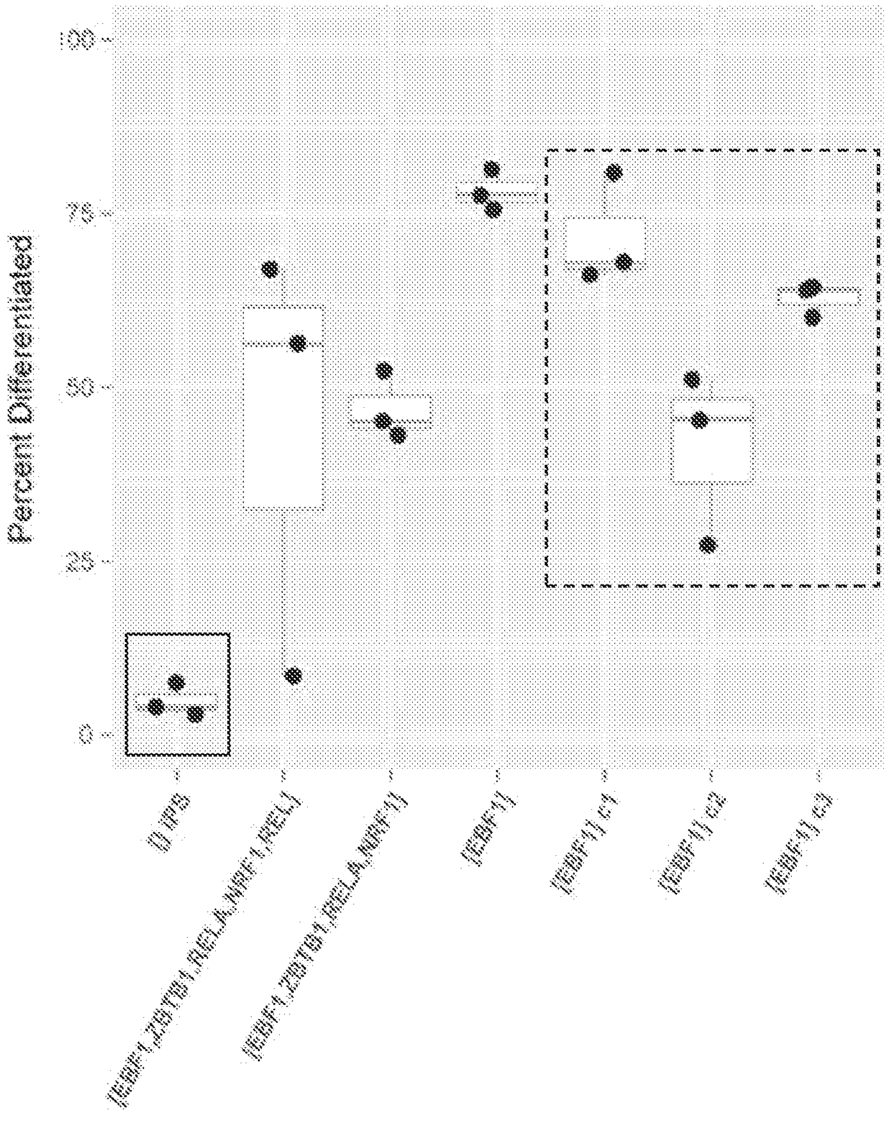

Percent Differentiated
100
75
50
25
0
[] iPS
[EBF1,ZBTB1,RELA,NRF1,REL]
[EBF1,ZBTB1,RELA,NRF1]
[EBF1]
[EBF1] c1
[EBF1] c2
[EBF1] c3

FIG. 7F