



(51) International Patent Classification:

C12N 5/0797 (2010.01) C12N 5/10 (2006.01)
C12N 5/0793 (2010.01)

(21) International Application Number:

PCT/US2017/051122

(22) International Filing Date:

12 September 2017 (12.09.2017)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

62/393,324	12 September 2016 (12.09.2016)	US
62/411,728	24 October 2016 (24.10.2016)	US
62/492,552	01 May 2017 (01.05.2017)	US
62/517,307	09 June 2017 (09.06.2017)	US

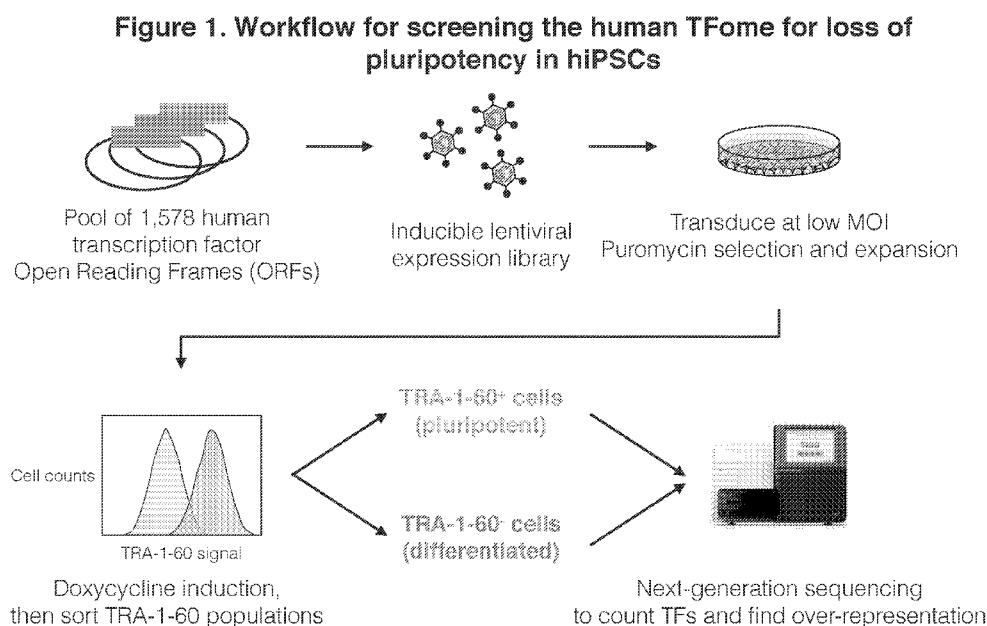
(71) Applicant: **PRESIDENT AND FELLOWS OF HARVARD COLLEGE** [US/US]; 17 Quincy Street, Cambridge, Massachusetts 02138 (US).

(72) Inventors: **NG, Alex H.M.**; 11 Sumner Road, Cambridge, Massachusetts 02138 (US). **CHURCH, George M.**; 218 Kent Street, Brookline, Massachusetts 02446 (US). **BUSSKAMP, Volker**; Pressestelle der TU Dresden, 01062 Dresden (DE).

(74) Agent: **IWANICKI, John P.**, Banner & Witcoff, LTD., 28 State Street, Suite 1800, Boston, Massachusetts 02109 (US).

(81) Designated States (*unless otherwise indicated, for every kind of national protection available*): AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BN, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DJ, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IR, IS, JO, JP, KE, KG, KH, KN, KP,

(54) Title: TRANSCRIPTION FACTORS CONTROLLING DIFFERENTIATION OF STEM CELLS



(57) **Abstract:** Forced expression of a handful of transcription factors (TFs) can induce conversions between cell identities; however, the extent to which TFs can alter cell identity has not been systematically assessed. Here, we assembled a human TFome, a comprehensive expression library of 1,578 human TF clones with full coverage of the major TF families. By systematically screening the human TFome, we identified 77 individual TFs that induce loss of human-induced-pluripotent-stem-cell (hiPSC) identity, suggesting a pervasive ability for TFs to alter cell identity. Using large-scale computational cell type classification trained on thousands of tissue expression profiles, we identified cell types generated by these TFs with high efficiency and speed, without additional selections or mechanical perturbations. TF expression in adult human tissues only correlated with some of the cell lineage generated, suggesting more complexity than observation studies can explain.

KR, KW, KZ, LA, LC, LK, LR, LS, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PA, PE, PG, PH, PL, PT, QA, RO, RS, RU, RW, SA, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW.

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

Published:

- with international search report (Art. 21(3))
- before the expiration of the time limit for amending the claims and to be republished in the event of receipt of amendments (Rule 48.2(h))

TRANSCRIPTION FACTORS CONTROLLING DIFFERENTIATION OF STEM CELLS

RELATED APPLICATION DATA

[01] This application claims priority to U.S. Provisional Application No. 62/393,324 filed on September 12, 2016 which is hereby incorporated herein by reference in its entirety for all purposes.

STATEMENT OF GOVERNMENT INTERESTS

[02] This invention was made with government support under 5P50HG005550 and HG008525 awarded by the National Institutes of Health. The government has certain rights in the invention.

TECHNICAL FIELD OF THE INVENTION

[03] This invention is related to the area of cell fates. In particular, it relates to induction of differentiated cells on the one hand and to the maintenance of stem cells on the other hand.

BACKGROUND OF THE INVENTION

[04] The discovery of mouse embryonic stem cells (ESCs), human ESCs and mouse and human induced pluripotent stem cells (hiPSC) has expanded the working modes in biology; unlike studying biological phenomena such as differentiation, gaining and maintaining cellular identity *in vivo*, we are now theoretically able to mimic some of these processes in a dish. The use of hiPSCs facilitates studying the genesis of human cell types in an ethically approved setting, and enables production of medically relevant cell types for research and medicine. However, exploiting the full differentiation potency of stem cells is only possible with few differentiated cell types. So far, stem cell differentiation protocols are multifaceted and tailored to individual cell types. These

protocols often yield highly heterogeneous populations, which may mask the cell type of interest. Whereas the initial triggers that drive stem cells out of pluripotency are known, we know very little about subsequent molecular events that occur during differentiation. For example, initial triggers include the application of differentiation media, the addition of small molecules, application of growth factors or 3D culturing techniques. These stimuli act via cellular signaling cascades and converge on transcription factors (TFs), which alter gene expression by activation and/or repression. Some of these cascades may recapitulate *in vivo* development, for example retinogenesis in 3D retinal organoids whereas monolayer cultures seem to follow alternative molecular differentiation routes although final cell types can have high similarity with *in vivo* cell types. Examples of cell identity transitions, both during the course of natural development *in vivo* and in synthetic *in vitro* systems, demonstrate the importance of controlling these transcriptional programs. The ability to control cell identity has led to the current preoccupation of the stem cell field, as facile access to primary-like human cells would enable many applications in disease modeling, drug screening and regenerative medicine. Fundamentally, the ability to find and control transitions between cell identities would greatly enhance our understanding of cell identity.

- [05] Upon obtaining a cellular identity, its maintenance is a central feature of multicellular organisms. This comprises terminally differentiated cells but also physiological stem cells, which can give rise to specific cell types in the body. To maintain the stem cell pool, transcriptional programs are active in large part by transcription factors that prevent these cells from differentiating. *In vitro*, maintaining the pluripotency of ESCs or iPSCs is essential to expand and maintain these cells for downstream applications. Labor-intensive and delicate culturing techniques have been developed to maintain pluripotency and to avoid spontaneous differentiation. Advances in the formulation of culture media have significantly improved the robustness of stem cell maintenance, made feeder cell co-cultures dispensable and opened the usage of stem cells to a larger scientific community.
- [06] Due to the dependence on transcriptional programs to maintain or convert cellular identities it would be desirable to gain direct control at the transcriptomic level. Detailed

study of the transcriptional control of the lac operon was the first example that a certain class of DNA-binding proteins, transcription factors, initiates, enhances or represses transcription. It was shown that one could obtain transcriptional control on cell fate conversion within the same germ layer by ectopic TF expression, which was pioneered by overexpressing MyoD in fibroblasts, which subsequently converted to myoblasts, and C/EBP-alpha to convert B cells into macrophages. Forced TF induction can also convert cells arising from different germ layers, pioneered by the three TFs Brn2, Ascl1 and Myt1l (BAM) that convert fibroblasts into neurons. It has also been shown that certain sets of TF can change the identity of neurons in living animals. Another striking example is the generation of iPSCs by overexpressing Oct3/4, Sox2, Klf4 and c-Myc ("Yamanaka factors") in fibroblasts. TF choices were mostly "biologically inspired;" the ones known from *in vivo* development were tested *in vitro* manually or proposed by computational approaches. This biased selection of ectopic TFs led to the successful generation of a handful of cell types. Notably, the failure rate of selected TFs to induce desired cell types from stem cells is relatively high, either not working at all or resulting in unexpected cell fates. Due to experimental and technical differences, for example gene delivery routes, transient or inducible TF expression or differences in starting cell types, we cannot easily troubleshoot these results. Since the forced TF expression in human stem cells can be very efficient, we wondered how one could confer this differentiation route systematically on producing other cell types. Obviously, one needs to standardize the technical aspects of TF delivery, expression, screening and analysis. In contrast to "biologically-inspired" TFs, an unbiased and systematic TF library screen to identify novel TFs that convert stem cells to other cell types or which reinforce pluripotency would be desirable and complementary.

- [07] So far, there have not been unbiased, systematic open reading frame (ORF) screens for converting cell identity. More recently, CRISPR-based activator screens have been performed but not for cell conversion. CRISPR-based activators for cell conversion appear to be insufficient to overcome barriers for cell identity conversion. Systematic, RNAi-based screens have revealed important pluripotency factors to understand stem cell

biology, but over-expression to maintain pluripotency has not been systematically performed.

- [08] Cell types derived from human induced pluripotent stem cells (hiPSC) have high relevance for biomedical research and medicine, but robust, efficient and rapid protocols are lacking for many cell type: could we generate new protocols to generate cell types?
- [09] There is a continuing need in the art for methods of changing cell fate and maintaining cell fate so that cell populations available for cellular transplantation and drug screening can better reflect the diverse cell types in the human body.

SUMMARY OF THE INVENTION

- [10] One aspect of the invention is a method of inducing differentiation of induced pluripotent stem cells. A nucleic acid comprising an open reading frame (ORF) encoding a transcription factor, the transcription factor (TF) protein, or an activator of transcription of the gene encoding the transcription factor, is delivered to the induced pluripotent stem cells. As a consequence, the amount of the transcription factor in the induced pluripotent stem cells is increased, and the induced pluripotent stem cells differentiate to form differentiated cells.
- [11] Another aspect of the invention is a method of maintaining pluripotency of induced pluripotent stem cells. A nucleic acid comprising an open reading frame encoding the protein, the protein, or an activator of transcription of an open reading frame encoding the protein is delivered to the induced pluripotent stem cells. The protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells. As a consequence of the delivery, the induced pluripotent stem cells maintain expression of keratin sulfate cell surface antigen, TRA-1-60, which is a marker of stem cell identity.

- [12] Another aspect of the invention is an engineered human differentiated cell that comprises one or more nucleic acids comprising an open reading frame. The one or more ORFs encodes a transcription factor selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MTF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1.
- [13] Yet another aspect of the invention is an engineered human differentiated cell which comprises a nucleic acid comprising an open reading frame encoding a protein. The protein is selected from the group consisting of ZNF70; ZNF461; TFAP2B; ZNF426; MTF; CDX2; MEOX2; AKNA; NKX2-8; NKX3-2; NKX2-3; ZNF16 / HZF1; ETV2; TFPD3; RELA / p65; NEUROG1; ID4; HES7; MXD4; SOX14; FOXP1; E2F2; NEUROG3; ZNF148 / ZBP89; GRLF1 / ARHGAP35; BOLA2; ZNF616; MAX; ATOH1; PRDM5; LHX5; ZNF273; MAFK; HOXA1; HIF2A / EPAS1; MAFB; E2F3; PRDM7; ZNF44; HMGA1; NRL; BATF3; MYOG; KLF15; LMX1A; HOXB6; DMRTB1; ATF7; SCRT2; ZNF593; HES2; ZSCAN2; MSX2; ID3; SOX12; GLI1; DPRX; SMAD3; ZBED3; CAMTA2; MSC; ASCL3; BARX1; DMRT1; HOXA10; TSC22D3; ZNF837; MXD3; ZNF692; NHLH2; ZNF626; THAP3; SRY; WT1; SHOX; ZNF43; and ZSCAN1.
- [14] Still another aspect of the invention is an induced pluripotent stem cell which comprises a nucleic acid comprising an open reading frame encoding a protein. The protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells. The induced pluripotent stem cell expresses keratin sulfate cell surface antigen, TRA-1-60. The transcription factor may be one or more of the group consisting of AEBP2; AIRE; ARNT2; ARNTL; ATF6; BACH1; BARX2; CASZ1; CLOCK; CREBRF; CTCFL; CXXC1; DLX2; E2F1; EBF1; EGR1; ELF4; ELF5; EMX2; ETS1; ETV1; ETV7; FEZF1; FOXJ3; FOXL2; FOXN2; FOXO1; FOXP3; FOXS1; GATA1; GCM1; HMBOX1; HOXB9; HOXC10; HOXC6; HOXD10; HOXD3;

HOXD8; HSF2; HSFY1; IRF5; IRF9; IRX2; KCMF1; KLF1; KLF13; KLF16; KLF7; KLF8; LBX1; LCORL; MAF; MEF2A; MEIS2; MNT; MYBL2; MYF6; MYRF; NANOG; NEUROD1; NFE2L2; NFE2L3; NFIB; NFIX; NKRF; NOTO; NR0B1; NR1B1; NR1B3; NR1C2; NR1C3; NR1F3; NR1I2; NR2B2; NR2C1; NR3A1; NR3A2; NR3C4; NR5A2; OSR1; PAX6; PAX7; PAX9; PHOX2B; PKNOX1; PLAGL1; PREB; PRRX1; RBPJ; RC3H2; RFX5; SATB1; SMAD2; SMAD4; SP1; SP4; SP6; SREBF1; STAT3; TAL2; TBX14; TBX18; TBX21; TBX22; TBX5; TEAD2; TERF1; TERF2; TFAP2A; TFDP2; TFEB; THAP1; THAP10; TP63; TSHZ3; YEATS2; ZBED2; ZBTB12; ZBTB14; ZBTB17; ZBTB21; ZBTB35; ZBTB9; ZEB1; ZFH3; ZFPM2; ZFY; ZFYVE26; ZIC4; ZKSCAN1; ZKSCAN12; ZKSCAN14; ZKSCAN19; ZKSCAN24; ZKSCAN4; ZKSCAN5; ZMAT2; ZNF114; ZNF138; ZNF146; ZNF227; ZNF253; ZNF266; ZNF280A; ZNF280C; ZNF282; ZNF296; ZNF311; ZNF317; ZNF337; ZNF34; ZNF35; ZNF350; ZNF366; ZNF396; ZNF398; ZNF41; ZNF415; ZNF460; ZNF485; ZNF497; ZNF511; ZNF512; ZNF517; ZNF543; ZNF550; ZNF586; ZNF613; ZNF615; ZNF619; ZNF641; ZNF644; ZNF645; ZNF648; ZNF655; ZNF658; ZNF662; ZNF664; ZNF669; ZNF704; ZNF740; ZNF754; ZNF75A; ZNF774; ZNF789; ZNF8; ZNF80; ZNF84; ZSCAN16; ZSCAN22; ZSCAN32; ZXDA; ZXDC; and ZZZ3.

- [15] Another aspect of the invention is a method of inducing differentiation of induced pluripotent stem cells. Increased expression is induced of a gene selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MITF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1. The induced pluripotent stem cells thereby differentiate to form differentiated cells.
- [16] Yet another aspect of the invention is a method of inducing differentiation of induced pluripotent stem cells. Expression of a gene is induced. The gene is selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG,

FOXC1, MITF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1. The induced pluripotent stem cells thereby differentiate to form differentiated cells.

- [17] These and other embodiments which will be apparent to those of skill in the art upon reading the specification provide the art with tools and methods for controlling the cell fate of stem cell populations.

BRIEF DESCRIPTION OF THE DRAWINGS

- [18] Fig. 1. Workflow for screening the human TFome (a comprehensive expression library of 1,578 human transcription factor (TF) clones with full coverage of the major TF families) for loss of pluripotency in hiPSCs (human-induced-pluripotent-stem-cell)
- [19] Fig. 2. General strategy for high-throughput screening for cell type conversion
- [20] Fig. 3. The Human TFome expression library
- [21] Fig. 4. Expression vectors for delivery of human TFome
- [22] Fig. 5. Hits from high-throughput screening for stem cell differentiation [loss of stem cell identity (TRA-1-60) in human induced pluripotent stem cells]
- [23] Fig. 6. Hits are enriched for developmental genes and protein domains
- [24] Fig. 7. Selected transcription factors that induce stem cell differentiation
- [25] Fig. 8. ATOH1 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [26] Fig. 9. NEUROG3 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

- [27] Fig. 10. ETV2 induces endothelial differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [28] Fig. 11. MYOG induces muscle differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [29] Fig.12. FOXC1 induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [30] Fig. 13. MITF induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [31] Fig. 14. SOX14 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [32] Fig. 15. ZSCAN1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [33] Fig. 16. WT1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [34] Fig. 17. NEUROGENIN1 and NEUROGENIN2 induce neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [35] Fig. 18. NEUROGENIN1 and NEUROGENIN2 and EMX1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [36] Fig. 19. NEUROGENIN1 and NEUROGENIN2 and EMX2 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

- [37] Fig. 20. NEUROGENIN1 and NEUROGENIN2 and TBR1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [38] Fig. 21. NEUROGENIN1 and NEUROGENIN2 and FOXG1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations
- [39] Fig. 22. Selected transcription factors (first nine shown) are expressed at a low level or not at all, compared to housekeeping genes (last three shown) in hiPSCs
- [40] Fig. 23. Average copy number of TF ORFs integrated into the hiPSC genome

DETAILED DESCRIPTION OF THE INVENTION

- [41] The inventors have developed techniques for inducing differentiation of stem cells into particular cell lineages, as well as techniques for inducing stem cells to continue to proliferate as stem cells. Rather than using special growth conditions, small molecules, or growth factors, the techniques deploy transcription factors (TFs) to trigger differentiation programs or stem cell renewal programs that prevent differentiation.
- [42] As determined, certain transcription factors are able to induce stem cells to differentiate to particular lineages. For example, MITF causes stem cells to form melanocytes. Similarly, CDX2 causes stem cells to form placental cells. MYOG causes stem cells to form smooth muscle cells. An initial indicator of differentiation can be the loss of stem cell specific markers. For example, the keratin sulfate cell surface antigen, TRA-1-60, can be assayed in the cells, as it is lost early in the process of differentiation. Using loss of stem cell markers as an indicator also provides a general means for detecting relevant transcription factors, rather than looking for acquisition of a marker that is newly expressed during differentiation by a particular cell lineage. Loss-of-marker screening may be applied to cell types other than stem cells, for instance, to identify TFs that directly convert fibroblasts into cell types of interest. In some aspects, combination of

transcription factors may be used to achieve differentiation to a particular cell lineage. The combination may achieve a cell type or a cell sub-type that is not achieved by either transcription factor alone. Alternatively, the combination may achieve the same cell type as one of the transcription factors alone, but may achieve it more efficiently.

- [43] Stem cells may be sorted from differentiated cells or differentiating cells by various means, typically based on differential gene expression. For example, fluorescence activated cell sorting may be used to separate cells on the basis for their expression of TRA-1-60. Alternatively, certain differentiated cells may be sorted from other differentiated cells and from cells on the basis of their expression of a lineage-specific cell surface antigen. Yet another means is by assessing expression at the RNA level, by single cell RNA sequencing without any sorting or pre-selection step. Such techniques are known in the art and may be used as is suitable and convenient for a particular application.
- [44] Any means known in the art for increasing the amount of a transcription factor in a stem cell can be used. This may involve delivery of either a nucleic acid comprising an open reading frame encoding the transcription factor, delivery of the transcription factor itself, or delivery of an activator of the transcription factor or its expression. Any technique known in the art for such delivery may be used. For example, for delivery of a cDNA, a viral or plasmid vector may be used. The open reading frame (encoding any isoform of the TF) may be inducible or repressible for control, to achieve a suitable level of expression. The nucleic acid comprising the open reading frame may be a cDNA, an mRNA, or a synthetic or engineered nucleic acid. Some transcription factors may require a critical amount of expression to effectively induce differentiation, such as the equivalent of at least 5, 10, 15, 20, 25, or 50 copies of the ORF per cell. Other factors may require less than a certain threshold of expression due to possible toxicity at high levels, such as less than 20, 10, or 5 copies per cell. Increased levels of expression may also be achieved by increasing the copy number of an ORF, for example, by using a higher copy number vector or by using a transposon. In some embodiments, nuclease-null or “dead” Cas9 variants may be used to activate the transcription of a desired transcription factor. See,

e.g., Chavez et al., *Nature Methods* 13:563-569, 2016. In other embodiments, modified RNAs – RNAs that encode the transcription factor, but use synthetic nucleotides that improve stability and reduce degradation – may be used. In some cases, use of culture media adapted for a particular cell type may increase the expression of the ORF that induces expression of that cell type. Expression of an ORF can be increased from a non-expressed gene, from a gene expressed at a low level, or from a gene expressed at a robust level. Overexpression is expression at level that is higher than the level that is expressed before induction from a gene that is expressed at a low, medium or high level.

- [45] An exogenous open reading frame is typically an open reading frame that differs from the similar gene or mRNA in the cell. It may be engineered to have a different control sequence or sequences, such as promoter, operator, enhancer, terminator, etc. It may be engineered to have no introns. It may be engineered to be fused to a second open reading frame to which it is not fused in the human genome.
- [46] The differentiated cells that can be produced using the methods described here will have multiple applications. They can be used for regenerative medicine, such as transplanting the cells into a recipient in need of a certain type of cell. They can be used for drug testing, both in cell culture as well as after transplantation. The cells may be used to deliver a product to a part of a body, for example, if they naturally produce or are engineered to produce and secrete the product.
- [47] Drug testing in the cells may use substances that are known or unknown to have a certain biological activity. The substances may be elements, compounds or mixtures, whether natural or synthetic. The cells may be used to determine a desirable activity of a potential drug or conversely to determine undesirable effects of a substance or lack of such effects. The contacting of the substance with the cells may be in culture or in a human or animal body. The activity or side effects of the substance may be determined in vitro or in vivo, irrespective of where the contacting occurred.

- [48] Changes that are observed in the cells being tested are not limited. The cells can be observed for effects on cell growth, apoptosis, secreted products, expression of particular products, etc. The genome of these cells may be edited to match mutations found in patients with disease. Any type of assay known in the art for such changes may be used, including but not limited to immunological assays, morphological observations, histochemical stains, reverse transcription polymerase chain reaction, protein blots, mass spectrometry, hybridization assays, electrophysiology, etc.
- [49] The stem cells may be obtained from any source. One particularly useful source is human induced pluripotent stem cells. Mouse induced pluripotent stem cells and mouse embryonic stem cells may also be used, as well as such cells from other animals. The use of human embryonic stem cells may be regulated or ethically undesirable, but these may be used as well.
- [50] Differentiated cells may be identified by any property or set of properties that is characteristic or defining of that type of differentiated cell. For example, different cell types have a unique transcriptome. The transcriptome may be used as a means of matching and identifying an unknown cell type to a known cell type. The transcriptome may be used qualitatively or quantitatively. Similarly a proteome may be used a means of identifying an unknown differentiated cell type. Some cell types may be identifiable based on morphology, growth habit, secretion products, enzymatic activity, cellular function, and the like. Any means known in the art for identifying cells may be used.
- [51] The above disclosure generally describes the present invention. All references disclosed herein are expressly incorporated by reference. A more complete understanding can be obtained by reference to the following specific examples, which are provided herein for purposes of illustration only, and are not intended to limit the scope of the invention.

EXAMPLE 1

- [52] We decided to compose a complete TF ORF library by combining factors from existing resources. Absent TFs were obtained by *de novo* gene synthesis. This “human TFome”, comprises 1,578 canonical human TFs driven by an inducible promoter system within an all-in-one Tet-ON lentiviral vector backbone for stable integration in hiPSCs. By screening the human TFome we found over 70 TFs that induce loss of hiPSC identity, suggesting pervasive potency for TFs to alter cell identity. This resource, the entire library as well as individual TFs, will be publicly available at Addgene (non-profit plasmid repository) and a subset of hiPSC lines in which certain TFs can be induced will be available. We applied the human TFome as single TFs per hiPSCs to identify individual TFs that convert stem cells to other cell identities or reinforce pluripotency. High-throughput fluorescence cell sorting (FACS) in combination with next generation sequencing and subsequent bioinformatic analyses was performed to screen for TFs in two hiPSC lines. We show that expression levels are critical and can be elevated by using transposon-mediated integration of expression cassettes. The application of the human TFome in hiPSCs resulted in the validation of over ten TFs cell fate converting TFs. Furthermore, we discovered over one hundred TFs that reinforce pluripotency, which may help to improve the robustness of stem cell cultures.

EXAMPLE 2

Generating and applying the human TFome

- [53] Systematic and comprehensive TF-wide induction screening in human stem cells requires the availability of a TF-expression library. Notably, only partial human libraries, for example TFs included in the ORFeome, were accessible. Therefore, we decided to assemble TFs from available resources or by *de novo* gene synthesis to compile the “human TFome”, an expression library of 1,578 TFs representing all canonical human transcription factors (Vaquerizas et al. 2009) and further curated.

- [54] The pLIX403 lentiviral vector was chosen because it allows for genomic transgene integration, doxycycline-inducible TF expression from a Tet-On system and puromycin selection of transduced cells. The individual TFs in shuttling vectors (pENTR) were cloned into pLIX403 by pooled gateway cloning. To discriminate the ectopic TFs from intrinsic expression, the TFs were marked by a V5 epitope tag translated on the backbone downstream of the Gateway recombination site on the C-terminus of the TF. The pLIX403 vector comprises a second ubiquitous promoter cassette driving the rTA3 gene needed to activate the TetOn promoter in the presence of the small molecule doxycycline and bicistronically a puromycin selection marker. About 98% of the TFs were detectable by next-generation sequencing after subcloning in the DNA plasmid library, which subsequently was used to produce lentiviral particles.
- [55] We transduced the PGP1 from the Personal Genome Project and ATCC DYS0100 human iPSC lines with the human TFome at a low multiplicity of infection (MOI = 0.1), such that cells would receive a single lentiviral integration at most. In total, we transfected both cell lines each with a complete human TFome pool as well as with two subpools that we have six independent transductions. To obtain sufficient coverage of the library, we ensured that on average each TF was present in at least one hundred cells after lentiviral transduction. Genes that confer resistance to the antibiotics bleomycin, blasticidin, and hygromycin, which should not induce differentiation, were also transduced as negative controls. Transduced cells were then selected for TF-integration using puromycin, expanded and doxycycline was added for four days for continuous TF induction. FACS separated the stained population into the differentiated TRA-1-60^{low} and the pluripotent TRA-1-60^{high} population. For each of these populations, the integrated genes were amplified using universal primer PCR, and sequenced.

EXAMPLE 3

Transcription factors-wide screening for alterations in pluripotency in human induced pluripotent stem cells

- [56] To identify TFs that induce stem cells into *any* differentiated cell, we devised a strategy to screen for individual TFs that potentially cause loss of pluripotency rather than enriching for cells with any specific cell type marker. We used a fluorescence-activated cell sorting (FACS) approach, which enables multiplexed assessment of the human TFome followed by TF identification by using next-generation sequencing. To measure the loss of pluripotency, we stained for the keratin sulfate cell surface antigen TRA-1-60, which is rapidly lost upon exit from pluripotency. Importantly, because we aimed to identify potent inducers of differentiation, we performed the screen in the standard mTeSR1 stem cell media.
- [57] We first aimed to comprehensively identify single TFs that would induce loss of pluripotency. To score each TF for its differentiation potential, we computed a ratio of each TF by dividing the number of normalized reads sequenced in the differentiated TRA-1-60^{low} population versus the pluripotent TRA-1-60^{high} population. We expect a TF that induces differentiation to be highly enriched in the TRA-1-60^{low} population compared to the pluripotent TRA-1-60^{high} population, and hence have a high ratio. Conversely a TF that reduces spontaneous differentiation, for instance by maintaining pluripotency, would have fewer reads in the differentiated population compared to the pluripotent population. We set a threshold for a TF having no effect based on the score of the truncated fluorescent proteins. TFs that increase cell proliferation without affecting pluripotency would be expected to have differentiation scores similar to these truncated fluorescent proteins.
- [58] Based on this differentiation scoring metric, we identified over 70 TFs that induce loss of pluripotency (“differentiation-inducing TFs”) and over 100 TFs that reinforce pluripotency (“pluripotency-reinforcing TFs”), as compared to the truncated fluorescent protein controls. To assess the quality of this screen, we conducted gene set enrichment analysis (GSEA) on the differentiation-inducing TFs. We discovered that differentiation-inducing TFs are enriched for developmental processes and system development, which is consistent with their ability to induce loss of pluripotency in our screen. In terms of protein domains, basic helix-loop-helix DNA-binding domains, which are often present

in TFs involved in development were enriched, whereas, zinc fingers and Küppel-associated box (KRAB) domains were depleted. Overall, these results suggest that our screen in human stem cells recovers developmentally important TFs that have been identified in model organisms *in vivo*.

EXAMPLE 4

Transcription factors induce rapid and efficient differentiation when expressed at high levels

[59] To craft genetic recipes that rapidly and efficiently induce hiPSCs into cell types, we engineered inducible single-TF expressing cell lines that could produce cell types of interest on demand. We originally aimed to mimic closely the conditions of the screen, namely single-copy lentiviral integration. Our initial efforts successfully validated the screen, however the differentiation efficiencies were weak (~10%), similar to CRISPR-Cas9 activator-based differentiation experiments. To overcome this challenge while keeping with our goal for simple differentiation protocols without additional growth factors, mechanical steps or selections, we wondered whether low TF expression levels of the integrated lentiviral vector was the bottleneck of highly efficient differentiation. We surmised that certain crucial target genes may be occluded with a low probability of TF binding and activation of transcription; thus a higher expression would in theory increase the probability of a successful binding and hence transcriptional event that may activate positive feedback loops that induce exit from pluripotency.

[60] To express high levels of TF, we first tested transduction with high titer lentiviral particles, which resulted in massive cell death. To improve the transduction efficiency and the viability of hiPSCs, we constructed a PiggyBac transposon-based vector with similar doxycycline-inducible TF expression and the ability to use puromycin to select for transduced cells. Critically, this transposon system allows for facile high-copy integration of TFs; in our hands, we average ~15 integrated copies per genome, as assessed by digital droplet PCR. Due to this integration efficiency, it is challenging to set copy numbers to a single TF per cell. Therefore, we decided not to repeat the screen with the lentiviral

system but selected TFs that had a high differentiation score for subcloning into the PiggyBac vectors.

- [61] Using this high expression system, we assessed differentiation efficiency. By bright-field microscopy, we observed a rapid loss of stem cell morphology, specifically migration away from colonies and adoption of distinct cell morphologies. We quantified loss of stem cell identity by intracellular flow cytometry for NANOG, OCT4 and SOX2. Uninduced stem cells had >90% NANOG⁺ OCT4⁺ SOX2⁺ cells, whereas doxycycline-induced cells had <10% NANOG⁺ OCT4⁺ SOX2⁺ cells.

EXAMPLE 5

Systematic classification of TF-induced lineages

- [62] Next, we needed an approach of identifying *a priori* the cell lineage being generated – we needed a method analogous to BLAST, but for tissue profiles. Several studies have been published that generally compare gene expression profiles to infer common drug targets, disease mutation effects, etc; however, the broad range of genes did not allow for rigorous and unambiguous identification of cell lineage against expression profiles catalogued in the Gene Expression Omnibus. A BLAST-like approach also requires an extensive reference panel to compare to; however the mechanistic gene regulatory network-based algorithm CellNet requires many samples per tissue, which is not currently available. We systematically use thousands of RNA-seq datasets from many tissues as training data. We adapted KeyGenes, a machine-learning-based algorithm, to systematically BLAST our transcriptomes against high-quality tissue expression datasets.
- [63] The TFs that induce loss of stem cell identity may be caused by conversion into differentiated cell types, or general loss of cell identity without a specific identity. To systematically determine what lineage these TF may be inducing, we first needed a systematic approach to classify cell types. We curated a set of large-scale human tissue expression profile studies to train a machine learning classifier for cell types. This curation comprises RNA-seq samples representing tissues from the GTEx study, Human

Protein Expression Atlas and Illumina Human Body Map. We then applied this as training data for KeyGenes, LASSO regression-based classifier for cell type identification.

- [64] To determine what cell lineages are being generated by these TFs, we performed RNA-seq on each cell population and used them as query for the expanded KeyGenes classifier.
- [65] Interestingly, expression of these TFs in adult human tissue did not predict which lineages were produced in hiPSCs. For instance, ETV2 induces endothelial differentiation but is most highly expressed in the testis, ATOH1 activated a neuronal cell identity but is highly expressed in the colon and small intestine, and CDX2 induced a placental identity, but is also highly expressed in the colon and small intestine. Thus observational studies alone do not appear to predict which TFs can induce which lineages, and suggest that synthetic methods of converting cell type do not necessarily recapitulate *in vivo* development.
- [66] Expression of specific markers for these cell lineages was present from the RNA-seq data). To enhance differentiation into these lineages, we induced TF expression in the presence of standard culturing conditions for those lineages, and assess conversion efficiency for lineage-specific markers by flow cytometry. Morphologies and lineage-specific marker expression were tested by immunohistochemical stainings. Together, these data indicate that TF-derived cells can be potently and rapidly generated.

EXAMPLE 6

EXPERIMENTAL METHODS

Annotation and manual curation of the TFome

- [67] Canonical human transcription factors were previously annotated by Vaquerizas et al. (2009) based on a computation search for genes that bind DNA in a sequence-specific manner, but are not enzymatic and do not form part of the core initiation complex, resulting in 1,591 TFs (classified as “a” and “b”, which have experimental evidence for

regulatory function, and “other” as probable TFs with undefined DNA-binding domains; class “x” was excluded as having promiscuous DNA-binding domains). We further included TFs that were predicted but did not have experimental evidence (class “c”). The following major TF families were filled in according to HUGO Gene Nomenclature Committee (HGNC): zinc finger (includes C2H2-containing domains), homeodomain (includes LIM, POU, TALE, HOXL, NKL, PRD sub-families), basic helix-loop-helix and forkhead. Pseudogenes as annotated by HGNC were removed. Duplicated and unmapped genes were removed, and all genes were converted to approved gene names using the HGNC multi-symbol checker. The final target set of TFs in the human TFome contains 1,578 genes.

Assembly and quality control of the human TFome

- [68] All TFs are cloned in pDONR-series standardized Gateway-compatible vector, building on Yang et al. 2011, Jolma et al., transOMIC technologies cDNAs, ORFs (transOMIC Inc), DNA Repository at Arizona State University, and codon-optimized synthesis by Gen 9.

Pooled cloning of the human TFome into pLIX403 viral vector

- [69] To perform pooled LR cloning into the pLIX403 viral expression vector (Addgene 41395), each 96-well plate of DNA was combined into its own sub-pool and quantified using Qubit dsDNA Broad Range Assay Kit (Invitrogen Q32853). 75ng of pENTR-TFome subpool was used for a LR Clonase II (Invitrogen 11791100) reaction overnight and digested with Proteinase K. 1μL of the reaction was transformed into One Shot Stbl3 chemically competent cells (Invitrogen C737303), plated onto a 50 cm² LB Agar plate with 100μg/mL Carbenicillin and incubated overnight at 30°C. Serial dilutions were performed to estimate the number colonies per plate. Approximately 10,000 to 20,000 colonies grew per plate, representing 100 to 200-fold coverage per 96-member sub-pool.
- [70] Colonies from each sub-pool plate were scraped and DNA was extracted using QIAGEN Plasmid Plus Midi Kit (QIAGEN 12943) to generate pLIX403-TFome subpools. To

cloning efficiency and library coverage, pENTR and pLIX403 subpools were pooled to generate pENTR-TFome and pLIX403-TFome. 5µg of each pool was sheared to 200bp on a Covaris S2, and 1µg was used for library preparation using NEBNext Ultra DNA Library Prep Kit for Illumina (NEB E7370L).

Lentiviral production and transduction

- [71] Lentiviruses were produced as before using 45µg DNA for two 15cm dishes. hiPSCs were transduced with TFome lentiviruses at a multiplicity of infection (MOI) = 0.1.

Flow cytometry analysis and Fluorescence Activated Cell Sorting (FACS)

- [72] Cells were trypsinized, washed and resuspended in FACS buffer (PBS with 10% FBS). For surface antigens, live cells were stained with fluorophore-conjugated antibodies and viability dye CellTrace Calcein Blue, AM (Life Technologies, C34853) at 1×10^7 cells/mL for 30 minutes on ice in the dark. For intracellular staining, cells were fixed using BD Cytofix fixation buffer (BD Biosciences, 554655) at 1×10^7 cells/mL for 20 minutes, washed with BD Perm/Wash buffer (BD Biosciences, 554723), and permeabilized in BD Perm/Wash buffer for 10 minutes, then stained with antibodies and DAPI in the dark for 30 minutes. Stained cells were washed twice with FACS buffer, filtered into a strainer-capped tube (Falcon, 352235) and run on a BD LSRFortessa. Compensation for spectral overlap was determined by staining AbC Total Antibody Compensation Beads (Life Technologies, A10497) with single fluorophore-conjugated antibodies.

Transcriptome library preparation and sequencing

- [73] TRIzol (Life Technologies, 15596-018) was added directly to cells and incubated for 3 minutes and used for RNA extraction using Direct-zol RNA MiniPrep (Zymo Research, R2050). RNA was quantified by Qubit RNA HS Kit (Molecular Probes, Q32852). 1µg was used for Poly(A) isolation using Poly(T) beads (Bioo Scientific, 512979) and used for RNA-seq library preparation with unique molecular identifiers (UMIs) using NEXTflex Rapid Directional qRNA-seq kit (Bioo Scientific, 5130-02D). Libraries were

amplified on a LightCycler real-time quantitative PCR machine (Roche) by spiking in SYBR Gold (Life Technologies, S11494). Mid-logarithm amplified libraries were collected and purified using AMPure XP beads (Agencourt, A63881). Libraries were quality controlled by TapeStation (Agilent) and quantitative PCR (KAPA Biosystems).

Transduction and generation of stable cell lines

- [74] For individual PiggyBac transductions, cells 500,000 to 800,000 cells were nucleofected using Nucleofector P3 solution (Lonza, V4XP-3032) using the Nucleofector X-Unit. Puromycin-resistant cells were selected and expanded.

WE CLAIM:

1. A method of inducing differentiation of induced pluripotent stem cells, comprising:
delivering to the induced pluripotent stem cells a nucleic acid comprising an open reading frame encoding a transcription factor, the transcription factor protein, or an activator of transcription of the open reading frame encoding the transcription factor, whereby the amount of the transcription factor in the induced pluripotent stem cells is increased, and the induced pluripotent stem cells differentiate to form differentiated cells.
2. The method of claim 1 wherein the transcription is selected from the group consisting of: ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MITF, WT1, CDX2, SOX14, TFPD3, SMAD3, and ZSCAN1.
3. The method of claim 1 wherein the transcription factor is selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 & NEUROG2}, {NEUROG1 & NEUROG2 & EMX1}, {NEUROG1 & NEUROG2 & EMX2}, {NEUROG1 & NEUROG2 & TBR1}, {NEUROG1 & NEUROG2 & FOXG1} and neuronal cells are formed.
4. The method of claim 3 wherein the transcription factor is selected from the group consisting of {NEUROG1 & NEUROG2}, {NEUROG1 & NEUROG2 & EMX1}, {NEUROG1 & NEUROG2 & EMX2}, {NEUROG1 & NEUROG2 & TBR1}, {NEUROG1 & NEUROG2 & FOXG1} and cortical neuronal cells are formed.
5. The method of claim 1 wherein the transcription factor is ETV2 and endothelial cells are formed.
6. The method of claim 1 wherein the transcription factor is MYOG and muscle cells are formed.
7. The method of claim 1 wherein the transcription factor is FOXC1 and cardiac muscle progenitor cells are formed.

8. The method of claim 1 wherein the transcription factor is MITF and neural crest and melanocyte cells are formed.
9. The method of claim 1 wherein the transcription factor is WT1 and kidney progenitor cells are formed.
10. The method of claim 1 wherein the transcription factor is CDX2 and placental cells are formed.
11. The method of claim 1 wherein a combination of two or more of said nucleic acids comprising an open reading frame encoding a transcription factor, two or more open reading frames encoding a transcription factor, two or more of said transcription factor proteins, or two or more of said activators of transcription of an open reading frame encoding a transcription factor, are delivered to the induced pluripotent stem cells.
12. The method of claim 1 further comprising:
 - contacting the differentiated cells with a test substance and observing a change in the differentiated cells induced by the test substance.
13. The method of claim 1 further comprising:
 - transplanting the differentiated cells into a patient.
14. The method of claim 13 wherein the induced pluripotent stem cells are derived from somatic cells of the patient.
15. The method of claim 13 further comprising:
 - contacting the differentiated cells with a test substance and observing a change in the differentiated cells induced by the test substance.
16. The method of claim 15 wherein the contacting is performed before the transplanting.
17. The method of claim 15 wherein the contacting is performed after the transplanting.
18. The method claim 1 wherein the transcription factor is selected from the group consisting of ZNF70; ZNF461; TFAP2B; ZNF426; MITF; CDX2; MEOX2; AKNA; NKX2-8; NKX3-2; NKX2-3; ZNF16 / HZF1; ETV2; TFDP3; RELA / p65; NEUROG1; ID4; HES7; MXD4; SOX14; FOXP1; E2F2; NEUROG3; ZNF148 / ZBP89; GRLF1 / ARHGAP35; BOLA2; ZNF616; MAX; ATOH1; PRDM5; LHX5; ZNF273; MAFK; HOXA1; HIF2A / EPAS1; MAFB; E2F3; PRDM7; ZNF44; HMGA1; NRL; BATF3; MYOG; KLF15; LMX1A; HOXB6; DMRTB1; ATF7; SCRT2; ZNF593; HES2;

ZSCAN2; MSX2; ID3; SOX12; GLI1; DPRX; SMAD3; ZBED3; CAMTA2; MSC; ASCL3; BARX1; DMRT1; HOXA10; TSC22D3; ZNF837; MXD3; ZNF692; NHLH2; ZNF626; THAP3; SRY; WT1; SHOX; ZNF43; and ZSCAN1.

19. The method of claim 18 further comprising the step of sorting the differentiated cells using fluorescence activated cell sorting.
20. The method of claim 19 further comprising :
 - assaying the differentiated cells and determining a set of transcribed genes;
 - comparing the set of transcribed genes of the differentiated cells to one or more reference sets of transcribed genes from one or more reference tissues or cells; and
 - identifying a match between the differentiated cells and a reference tissue or cell.
21. The method of claim 19 further comprising :
 - assaying the differentiated cells and determining amounts of a set of transcribed genes;
 - comparing the amounts of the transcribed genes of the differentiated cells to one or more reference sets of amounts of transcribed genes from one or more reference tissues or cells; and
 - identifying a match between the differentiated cells and a reference tissue or cell.
22. The method of claim 19 further comprising the step of identifying differentiated cells as a type of differentiated cell by assaying morphological features of the differentiated cells and matching the morphological features to a reference tissue or cell's morphological features.
23. The method of claim 19 further comprising the step of identifying differentiated cells as a type of differentiated cell by assaying protein marker expression of the differentiated cells and matching the protein marker expression to a reference tissue or cell's protein marker expression.
24. The method of claim 19 further comprising the step of identifying differentiated cells as a type of differentiated cell by assaying a function and matching the function to a function of a reference tissue or cell.
25. The method of claim 19 wherein the differentiated cells are sorted on the basis of loss of expression of a stem cell marker.

26. The method of claim 25 wherein the stem cell marker is keratin sulfate cell surface antigen, TRA-1-60.
27. The method of claim 25 wherein the loss of expression of a stem cell marker is determined in a medium adapted for growth of stem cells.
28. A method of maintaining pluripotency of induced pluripotent stem cells, comprising:
delivering to the induced pluripotent stem cells a nucleic acid comprising an open reading frame encoding a protein, the protein, or an activator of transcription of an open reading frame encoding the protein, wherein the protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells; whereby the induced pluripotent stem cells maintain expression of keratin sulfate cell surface antigen, TRA-1-60.
29. The method of claim 28 wherein the transcription factor is selected from the group consisting of: AEBP2; AIRE; ARNT2; ARNTL; ATF6; BACH1; BARX2; CASZ1; CLOCK; CREBRF; CTCFL; CXXC1; DLX2; E2F1; EBF1; EGR1; ELF4; ELF5; EMX2; ETS1; ETV1; ETV7; FEZF1; FOXJ3; FOXL2; FOXN2; FOXO1; FOXP3; FOXS1; GATA1; GCM1; HMBOX1; HOXB9; HOXC10; HOXC6; HOXD10; HOXD3; HOXD8; HSF2; HSFY1; IRF5; IRF9; IRX2; KCMF1; KLF1; KLF13; KLF16; KLF7; KLF8; LBX1; LCORL; MAF; MEF2A; MEIS2; MNT; MYBL2; MYF6; MYRF; NANOG; NEUROD1; NFE2L2; NFE2L3; NFIB; NFIX; NKRF; NOTO; NR0B1; NR1B1; NR1B3; NR1C2; NR1C3; NR1F3; NR1I2; NR2B2; NR2C1; NR3A1; NR3A2; NR3C4; NR5A2; OSR1; PAX6; PAX7; PAX9; PHOX2B; PKN0X1; PLAGL1; PREB; PRRX1; RBPJ; RC3H2; RFX5; SATB1; SMAD2; SMAD4; SP1; SP4; SP6; SREBF1; STAT3; TAL2; TBX14; TBX18; TBX21; TBX22; TBX5; TEAD2; TERF1; TERF2; TFAP2A; TFD2P2; TFEB; THAP1; THAP10; TP63; TSHZ3; YEATS2; ZBED2; ZBTB12; ZBTB14; ZBTB17; ZBTB21; ZBTB35; ZBTB9; ZEB1; ZFHX3; ZFPM2; ZFY; ZFYVE26; ZIC4; ZKSCAN1; ZKSCAN12; ZKSCAN14; ZKSCAN19; ZKSCAN24; ZKSCAN4; ZKSCAN5; ZMAT2; ZNF114; ZNF138; ZNF146; ZNF227; ZNF253; ZNF266; ZNF280A; ZNF280C; ZNF282; ZNF296; ZNF311; ZNF317; ZNF337; ZNF34; ZNF35; ZNF350; ZNF366; ZNF396; ZNF398; ZNF41; ZNF415; ZNF460; ZNF485; ZNF497; ZNF511; ZNF512; ZNF517; ZNF543; ZNF550; ZNF586; ZNF613; ZNF615; ZNF619;

ZNF641; ZNF644; ZNF645; ZNF648; ZNF655; ZNF658; ZNF662; ZNF664; ZNF669; ZNF704; ZNF740; ZNF754; ZNF75A; ZNF774; ZNF789; ZNF8; ZNF80; ZNF84; ZSCAN16; ZSCAN22; ZSCAN32; ZXDA; ZXDC; and ZZZ3.

30. The method of claim 1 wherein a cDNA is delivered and expression of the cDNA is inducible.
31. The method of claim 28 wherein a cDNA is delivered and expression of the cDNA is inducible.
32. The method of claim 1 wherein a cDNA is delivered such that multiple copies of the cDNA per genome are integrated in the induced pluripotent stem cells.
33. The method of claim 28 wherein a cDNA is delivered such that multiple copies of the cDNA per genome are integrated in the induced pluripotent stem cells.
34. An engineered human differentiated cell comprising a nucleic acid comprising an open reading frame that encodes a transcription factor selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MITF, WT1, CDX2, SOX14, TFPD3, SMAD3, and ZSCAN1, wherein the open reading frame is intronless.
35. The engineered human differentiated cell of claim 34 wherein the cell is a neuronal cell and the one or more the open reading frames encode a transcription factor selected from the group consisting of: ATOH1, NEUROG3, {NEUROG1 & NEUROG2}, {NEUROG1 & NEUROG2 & EMX1}, {NEUROG1 & NEUROG2 & EMX2}, {NEUROG1 & NEUROG2 & TBR1}, and {NEUROG1 & NEUROG2 & FOXG1}.
36. The engineered human differentiated cell of claim 34 wherein the cell is a cortical neuronal cell and the one or more open reading frames encode a transcription factor selected from the group consisting of: {NEUROG1 & NEUROG2}, {NEUROG1 & NEUROG2 & EMX1}, {NEUROG1 & NEUROG2 & EMX2}, {NEUROG1 & NEUROG2 & TBR1}, {NEUROG1 & NEUROG2 & FOXG1}.
37. The engineered human differentiated cell of claim 34 wherein the cell is selected from the group consisting of (1) an endothelial cell, (2) a muscle cell, (3) a cardiac muscle

progenitor cell, (4) a neural crest cell or melanocyte, (5) a kidney progenitor cell, or (6) a placental cell, and the one or more open reading frames encodes a transcription factor selected from the group consisting of: (1) ETV2, (2) MYOG, (3) FOXC1, (4) MITF, (5) WT1, and (6) CDX2, respectively.

38. An engineered human differentiated cell, comprising:

a nucleic acid comprising an open reading frame encoding a protein, wherein the protein is selected from the group consisting of ZNF70; ZNF461; TFAP2B; ZNF426; MITF; CDX2; MEOX2; AKNA; NKX2-8; NKX3-2; NKX2-3; ZNF16 / HZF1; ETV2; TFDP3; RELA / p65; NEUROG1; ID4; HES7; MXD4; SOX14; FOXP1; E2F2; NEUROG3; ZNF148 / ZBP89; GRLF1 / ARHGAP35; BOLA2; ZNF616; MAX; ATOH1; PRDM5; LHX5; ZNF273; MAFK; HOXA1; HIF2A / EPAS1; MAFB; E2F3; PRDM7; ZNF44; HMGA1; NRL; BATF3; MYOG; KLF15; LMX1A; HOXB6; DMRTB1; ATF7; SCRT2; ZNF593; HES2; ZSCAN2; MSX2; ID3; SOX12; GLI1; DPRX; SMAD3; ZBED3; CAMTA2; MSC; ASCL3; BARX1; DMRT1; HOXA10; TSC22D3; ZNF837; MXD3; ZNF692; NHLH2; ZNF626; THAP3; SRY; WT1; SHOX; ZNF43; and ZSCAN1, wherein the open reading frame is intronless.

39. An induced pluripotent stem cell, comprising:

a nucleic acid comprising an open reading frame encoding a protein, wherein the protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells; wherein the induced pluripotent stem cell expresses keratin sulfate cell surface antigen, TRA-1-60, wherein the open reading frame is intronless.

40. The induced pluripotent stem cell of claim 39 wherein the transcription factor is selected from the group consisting of: AEBP2; AIRE; ARNT2; ARNTL; ATF6; BACH1; BARX2; CASZ1; CLOCK; CREBRF; CTCFL; CXXC1; DLX2; E2F1; EBF1; EGR1; ELF4; ELF5; EMX2; ETS1; ETV1; ETV7; FEZF1; FOXJ3; FOXL2; FOXN2; FOXO1; FOXP3; FOXS1; GATA1; GCM1; HMBOX1; HOXB9; HOXC10; HOXC6; HOXD10; HOXD3; HOXD8; HSF2; HSFY1; IRF5; IRF9; IRX2; KCMF1; KLF1; KLF13; KLF16;

KLF7; KLF8; LBX1; LCORL; MAF; MEF2A; MEIS2; MNT; MYBL2; MYF6; MYRF; NANOG; NEUROD1; NFE2L2; NFE2L3; NFIB; NFIX; NKRF; NOTO; NR0B1; NR1B1; NR1B3; NR1C2; NR1C3; NR1F3; NR1I2; NR2B2; NR2C1; NR3A1; NR3A2; NR3C4; NR5A2; OSR1; PAX6; PAX7; PAX9; PHOX2B; PKNOX1; PLAGL1; PREB; PRRX1; RBPJ; RC3H2; RFX5; SATB1; SMAD2; SMAD4; SP1; SP4; SP6; SREBF1; STAT3; TAL2; TBX14; TBX18; TBX21; TBX22; TBX5; TEAD2; TERF1; TERF2; TFAP2A; TFDP2; TFEB; THAP1; THAP10; TP63; TSHZ3; YEATS2; ZBED2; ZBTB12; ZBTB14; ZBTB17; ZBTB21; ZBTB35; ZBTB9; ZEB1; ZFHX3; ZFPM2; ZFY; ZFYVE26; ZIC4; ZKSCAN1; ZKSCAN12; ZKSCAN14; ZKSCAN19; ZKSCAN24; ZKSCAN4; ZKSCAN5; ZMAT2; ZNF114; ZNF138; ZNF146; ZNF227; ZNF253; ZNF266; ZNF280A; ZNF280C; ZNF282; ZNF296; ZNF311; ZNF317; ZNF337; ZNF34; ZNF35; ZNF350; ZNF366; ZNF396; ZNF398; ZNF41; ZNF415; ZNF460; ZNF485; ZNF497; ZNF511; ZNF512; ZNF517; ZNF543; ZNF550; ZNF586; ZNF613; ZNF615; ZNF619; ZNF641; ZNF644; ZNF645; ZNF648; ZNF655; ZNF658; ZNF662; ZNF664; ZNF669; ZNF704; ZNF740; ZNF754; ZNF75A; ZNF774; ZNF789; ZNF8; ZNF80; ZNF84; ZSCAN16; ZSCAN22; ZSCAN32; ZXDA; ZXDC; and ZZZ3.

41. The engineered human differentiated cell of claim 34 wherein the nucleic acid comprising one or more exogenous open reading frames is selected from the group consisting of a cDNA, a synthetic nucleic acid, or an mRNA.
42. The engineered human differentiated cell of claim 35 wherein the nucleic acid comprising one or more exogenous open reading frames is selected from the group consisting of a cDNA, a synthetic nucleic acid, or an mRNA.
43. The engineered human differentiated cell of claim 36 wherein the nucleic acid comprising one or more exogenous open reading frames is selected from the group consisting of a cDNA, a synthetic nucleic acid, or an mRNA.
44. The engineered human differentiated cell of claim 37 wherein the nucleic acid comprising one or more exogenous open reading frames is selected from the group consisting of a cDNA, a synthetic nucleic acid, or an mRNA.

45. The engineered human differentiated cell of claim 38 wherein the nucleic acid comprising one or more exogenous open reading frames is selected from the group consisting of a cDNA, a synthetic nucleic acid, or an mRNA.
46. The induced pluripotent stem cell of claim 39 wherein the nucleic acid comprising one or more exogenous open reading frames is selected from the group consisting of a cDNA, a synthetic nucleic acid, or an mRNA.
47. A method of inducing differentiation of induced pluripotent stem cells, comprising:
inducing increased expression of a gene selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MTF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1, wherein the induced pluripotent stem cells differentiate to form differentiated cells.
48. The method of claim 47 wherein the gene is expressed weakly in the induced pluripotent stem cells prior to said step of inducing.
49. The method of claim 47 wherein the gene is induced to express the gene at levels greater than the expression levels endogenous to an induced pluripotent stem cell.
50. A method of inducing differentiation of induced pluripotent stem cells, comprising:
inducing expression of a gene selected from the group consisting of ATOH1, NEUROG3, {NEUROG1 and NEUROG2}, {NEUROG1 and NEUROG2 and EMX1}, {NEUROG1 and NEUROG2 and EMX2}, {NEUROG1 and NEUROG2 and TBR1}, {NEUROG1 and NEUROG2 and FOXG1}, ETV2, MYOG, FOXC1, MTF, SOX14, WT1, TFPD3, CDX2, SMAD3, and ZSCAN1, wherein the induced pluripotent stem cells differentiate to form differentiated cells.
51. The method of claim 50 wherein the gene is not expressed in the induced pluripotent stem cells prior to said step of inducing.

Figure 1. Workflow for screening the human TFome for loss of pluripotency in hiPSCs

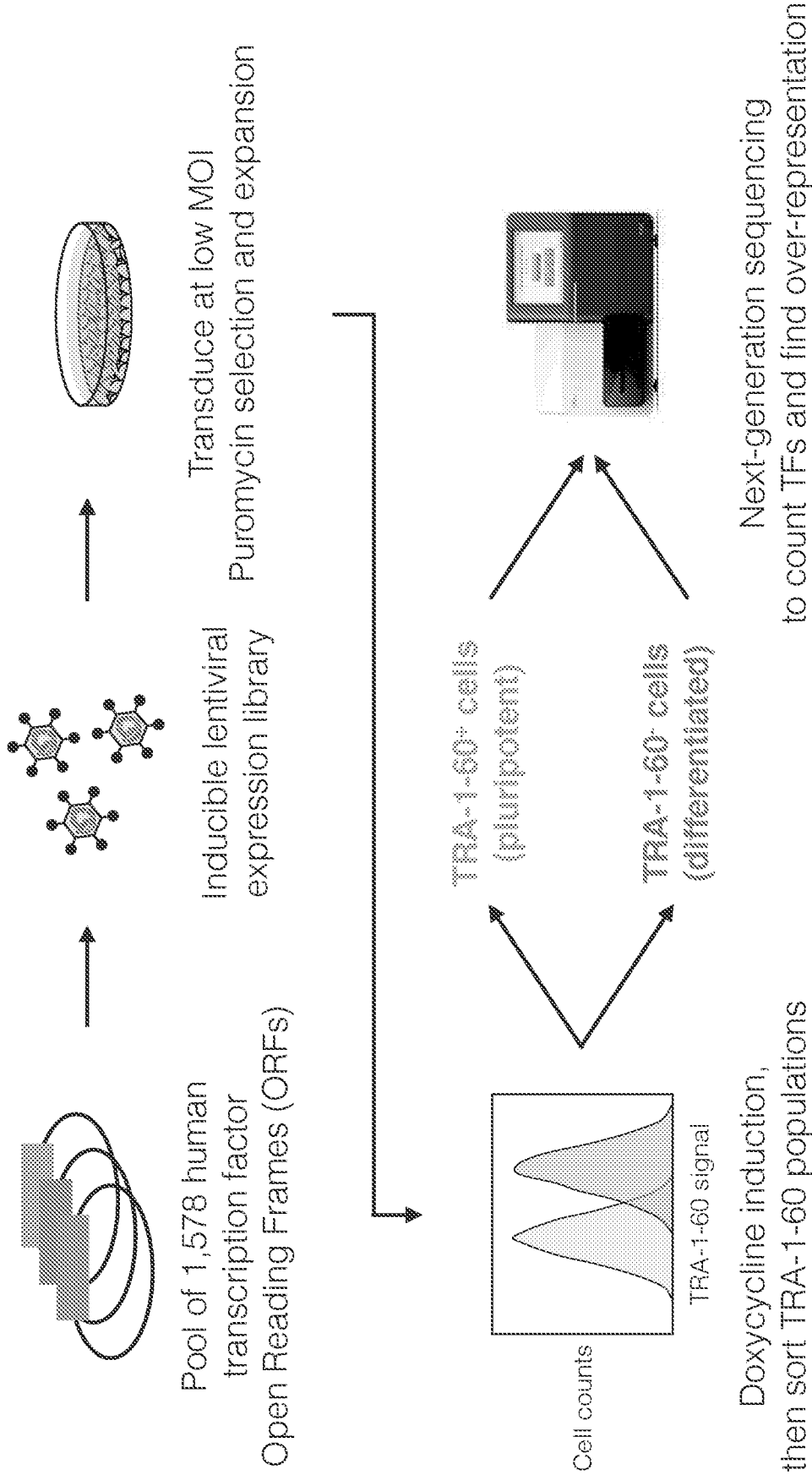


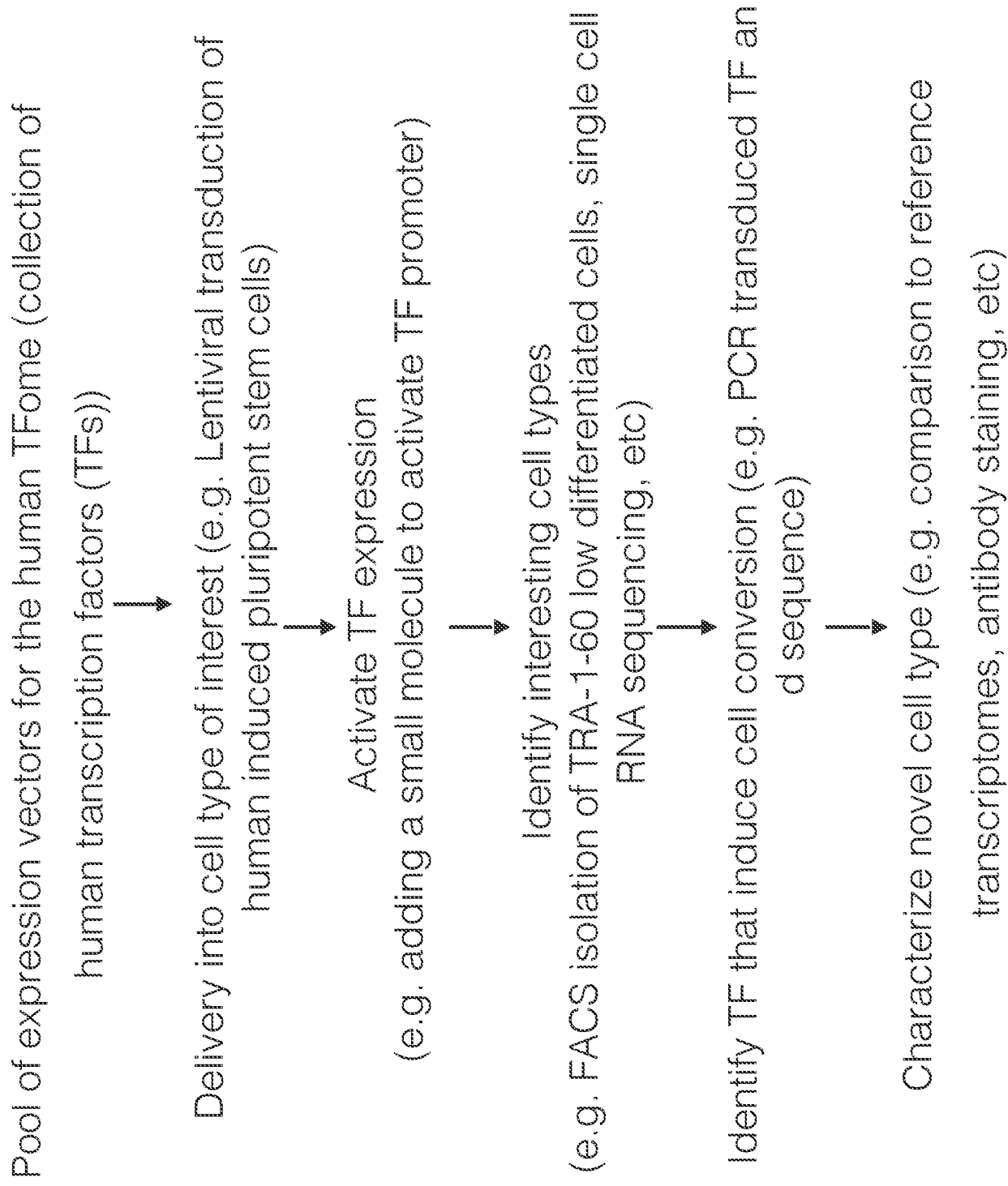
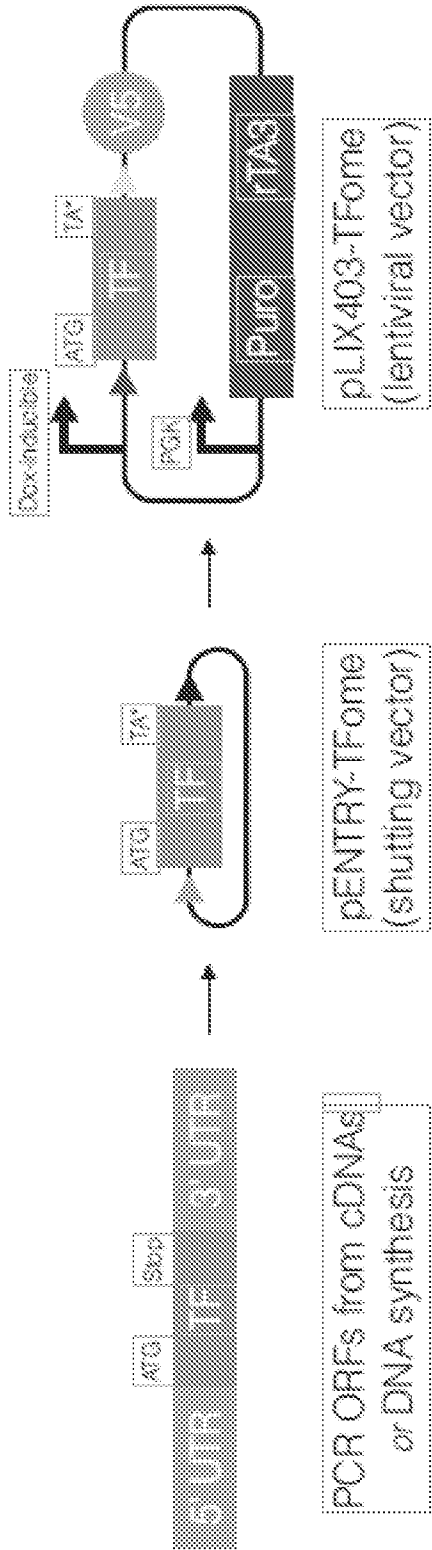
Figure 2. General strategy for high-throughput screening for cell type conversion

Figure 3. The Human TFome expression library



Size distribution of TFs

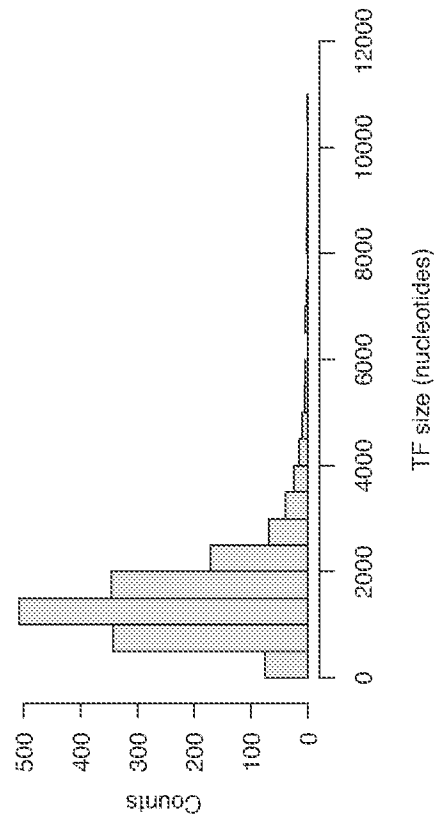
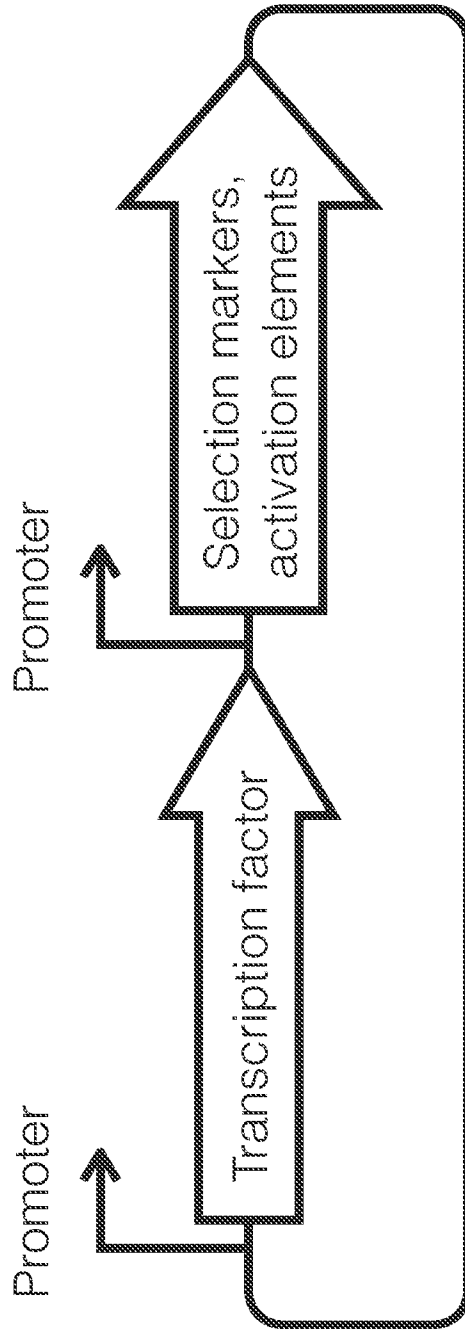


Figure 4. Expression vectors for delivery of human TFome



Example promoters:

Constitutive expression (e.g. EF1alpha promoter)

Inducible expression (e.g. small molecule, heat-shock, light) Cell type-dependent expression (e.g. neuronal promoter)

Example delivery vectors:

Transient expression (e.g. plasmid with no mammalian integration ability)

Viral vectors (e.g. lentiviral, retroviral, AAV)

Transposon-based (e.g. PiggyBac)

Figure 5. Hits from high-throughput screening for stem cell differentiation [loss of stem cell identity (TRA-1-60) in human induced pluripotent stem cells]

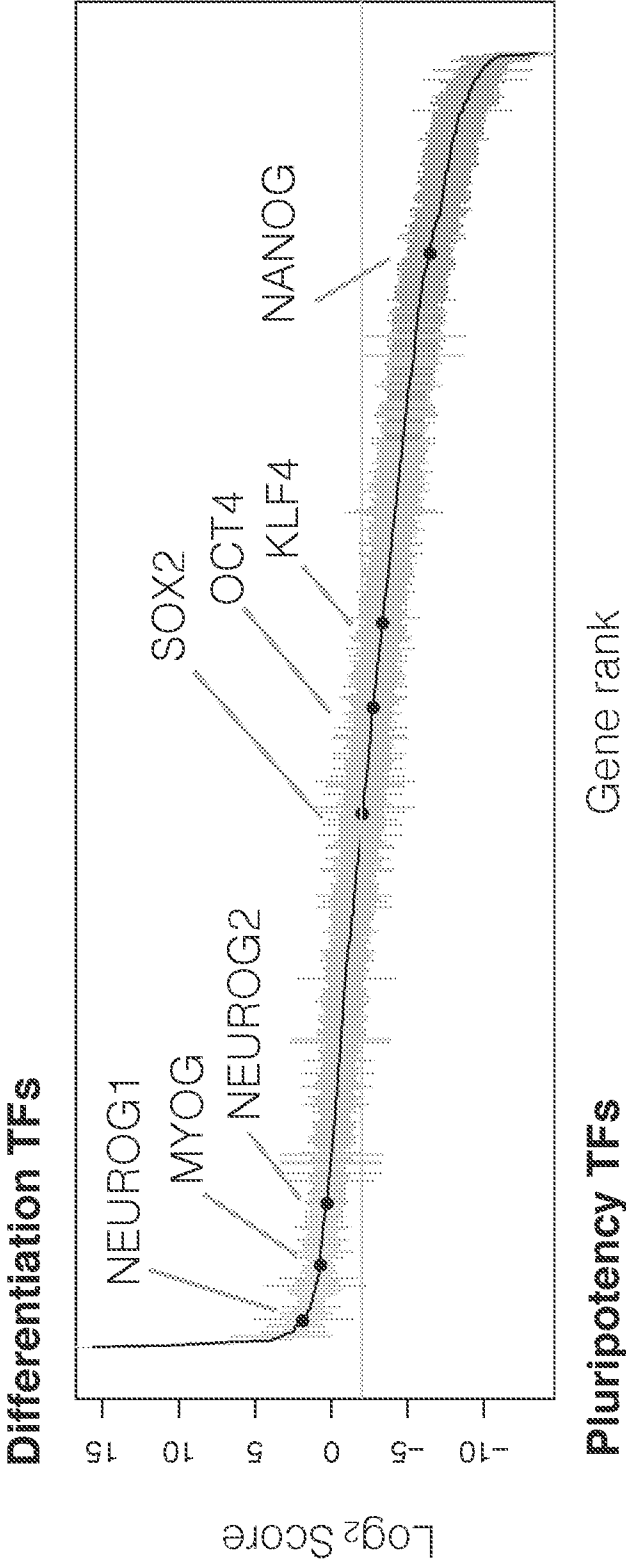
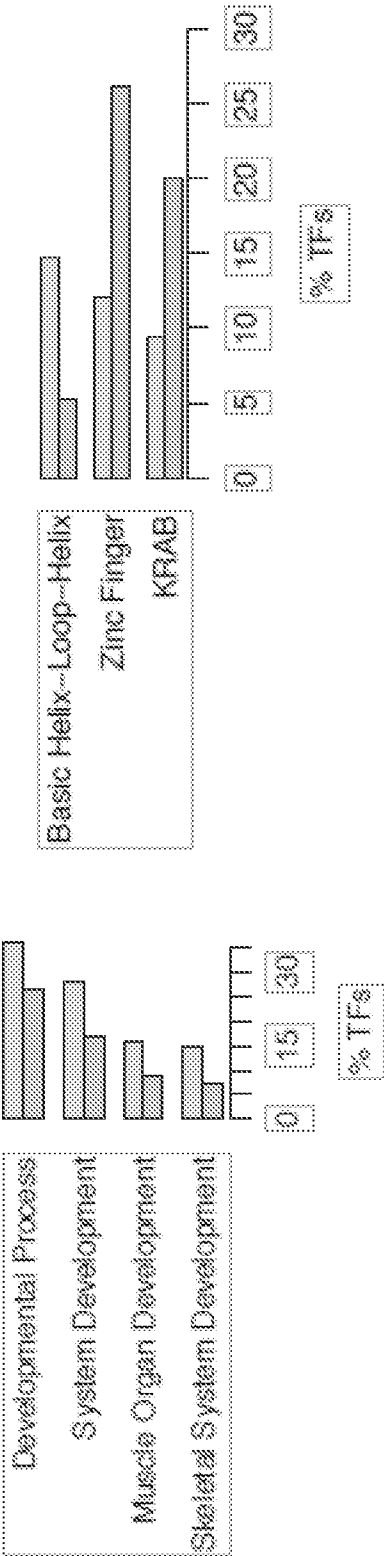


Figure 6. Hits are enriched for developmental genes and protein domains



Differentiation TFs
All TFs

Figure 7. Selected transcription factors that induce stem cell differentiation

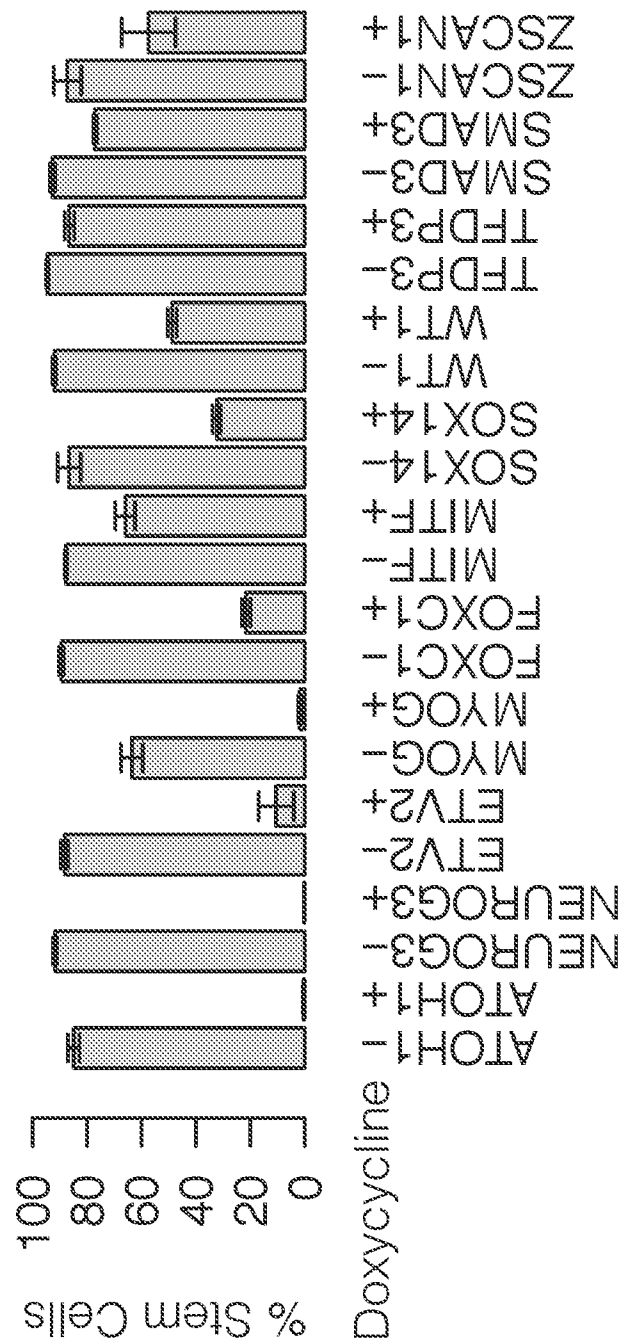


Figure 8. ATOH1 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

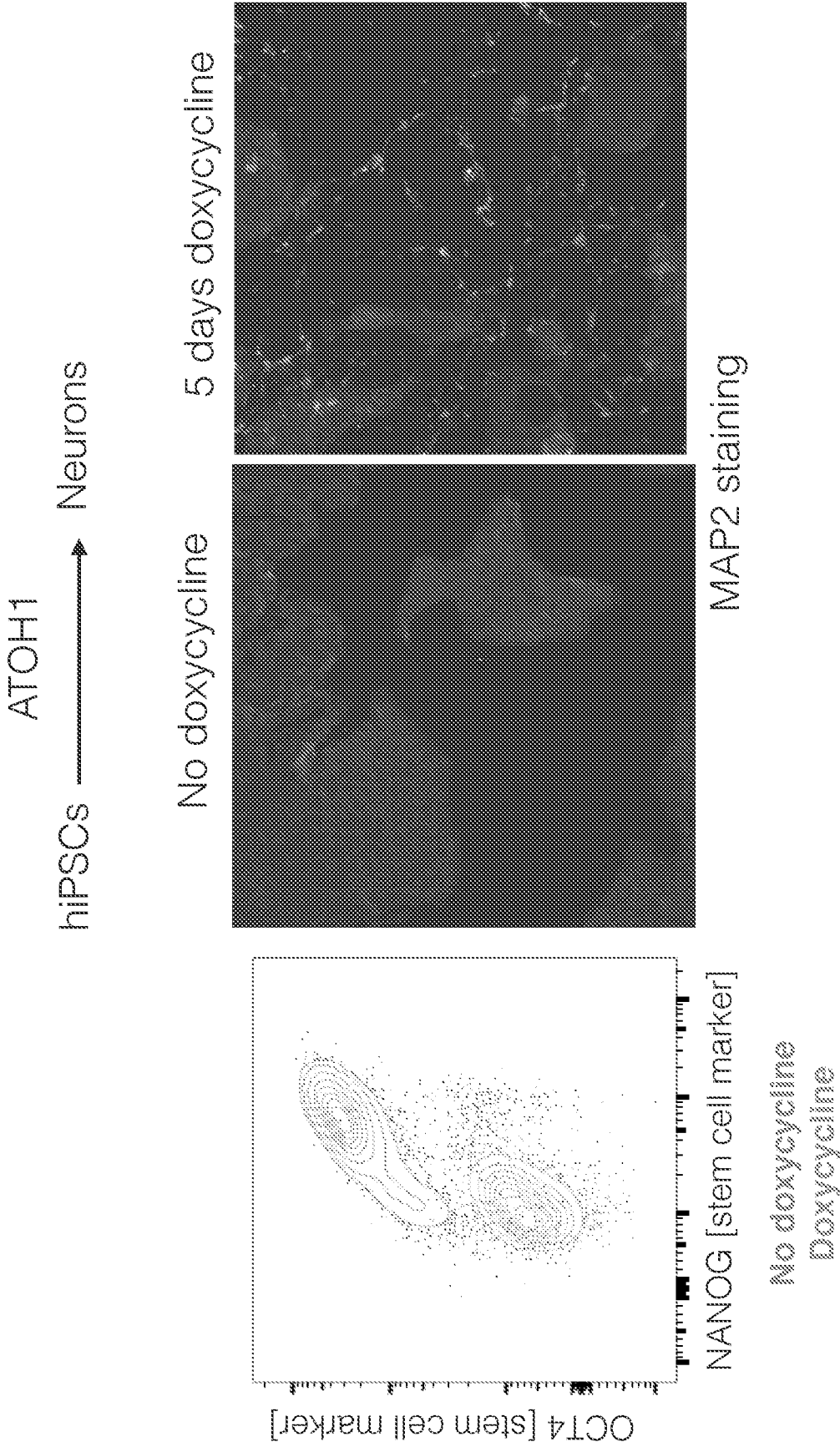


Figure 9. NEUROG3 induces neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

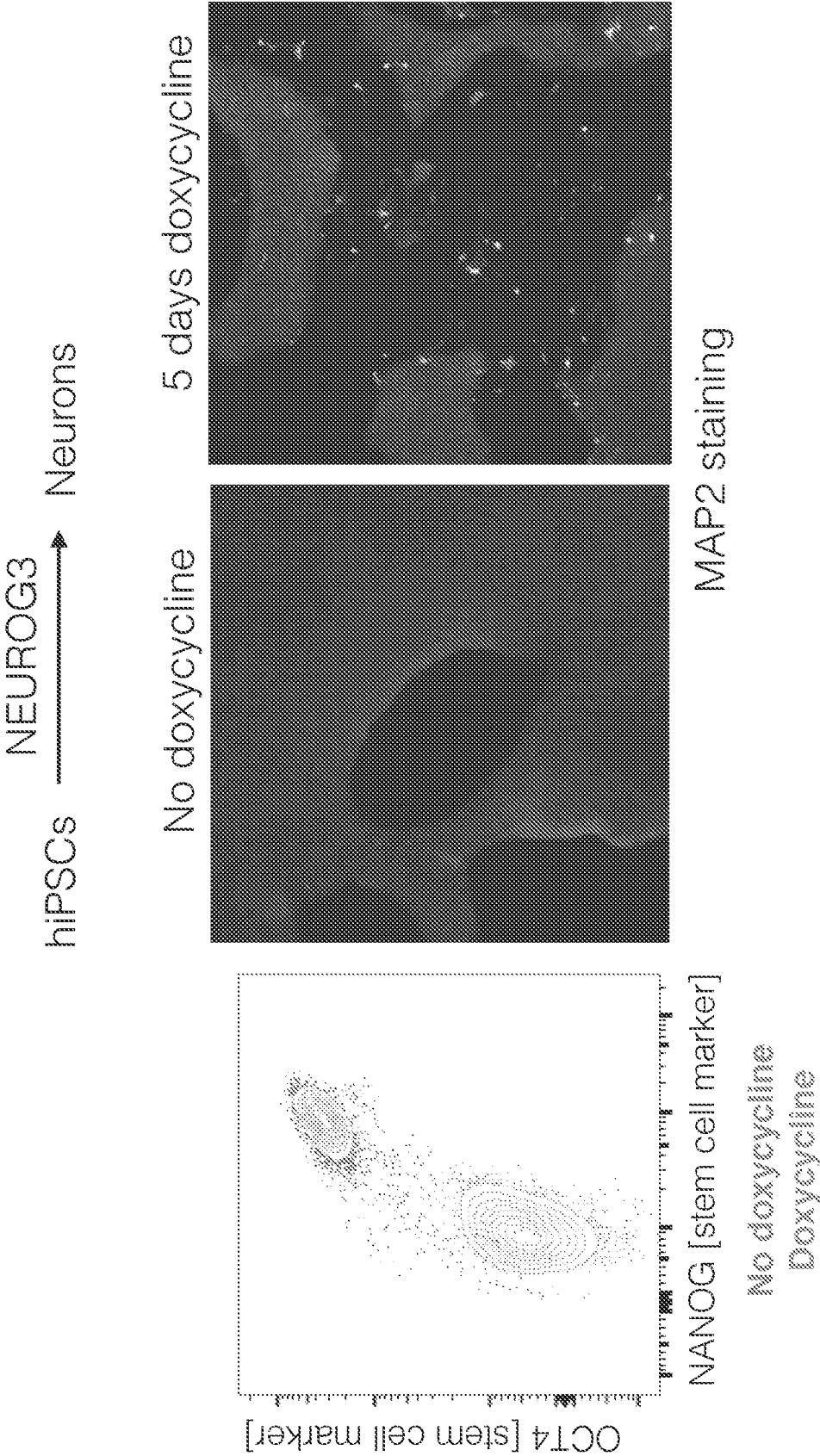
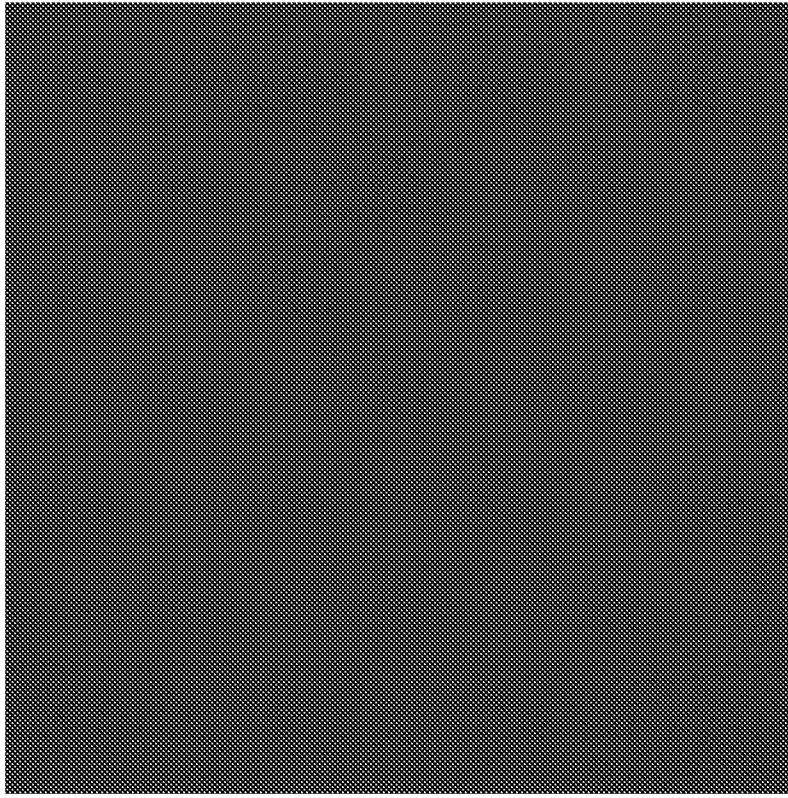
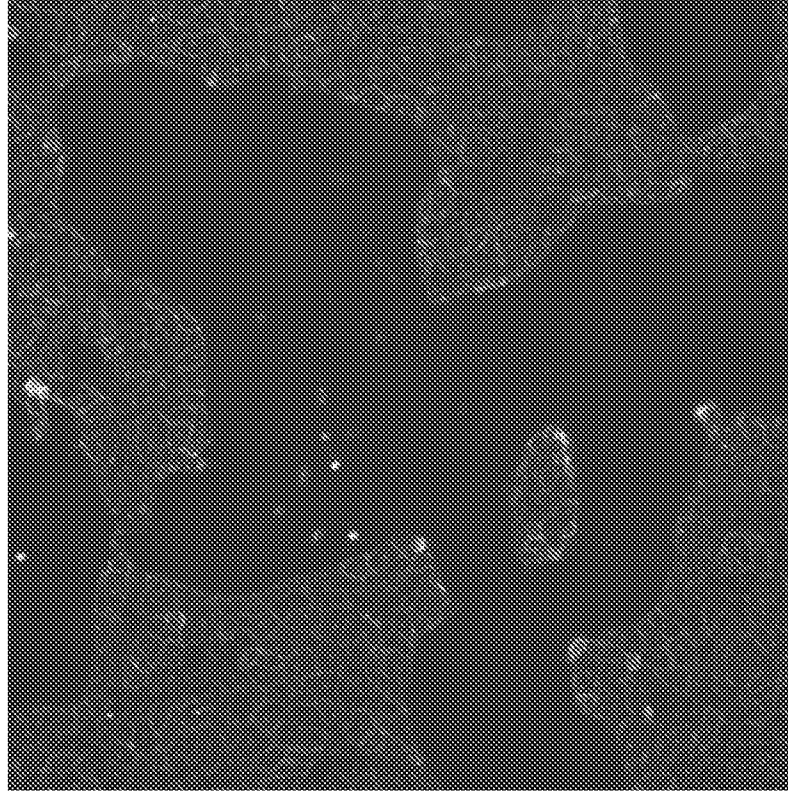


Figure 10. ETV2 induces endothelial differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

No doxycycline



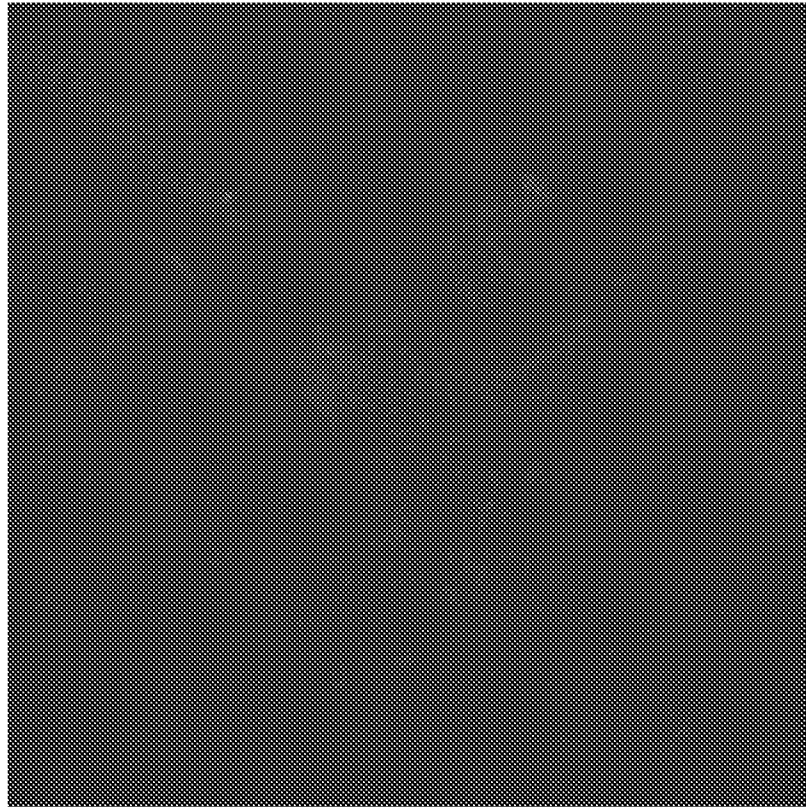
6 days doxycycline



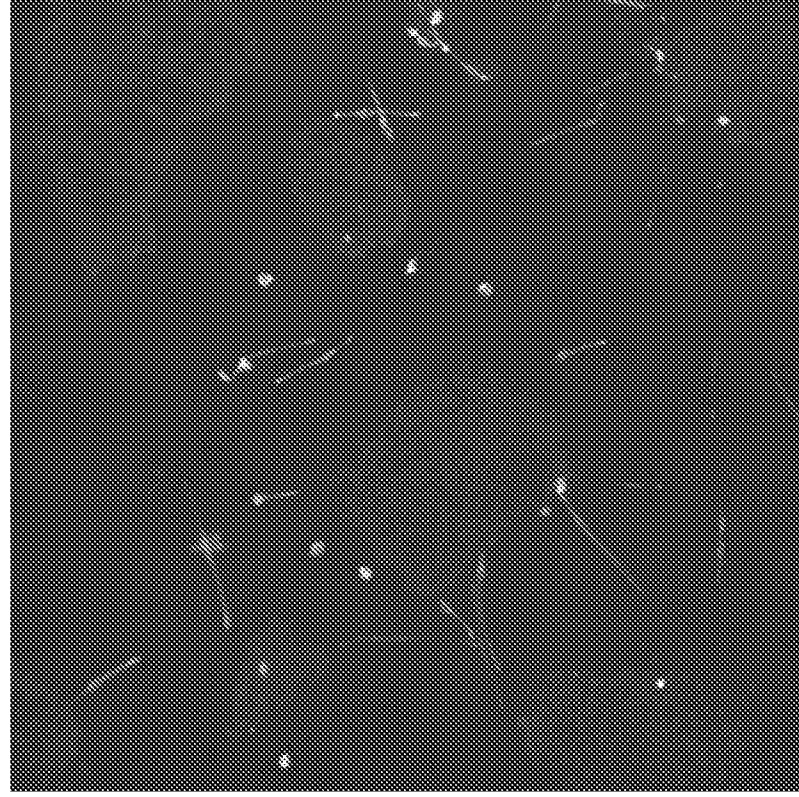
VE-Cadherin staining

Figure 11. MYOG induces muscle differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

No doxycycline



6 days doxycycline



Myosin heavy chain staining

**Figure 12. FOXC1 induces differentiation of hiPSCs
in stem cell media without embryoid body formation, additional growth
factors or mechanical manipulations**

4 days doxycycline

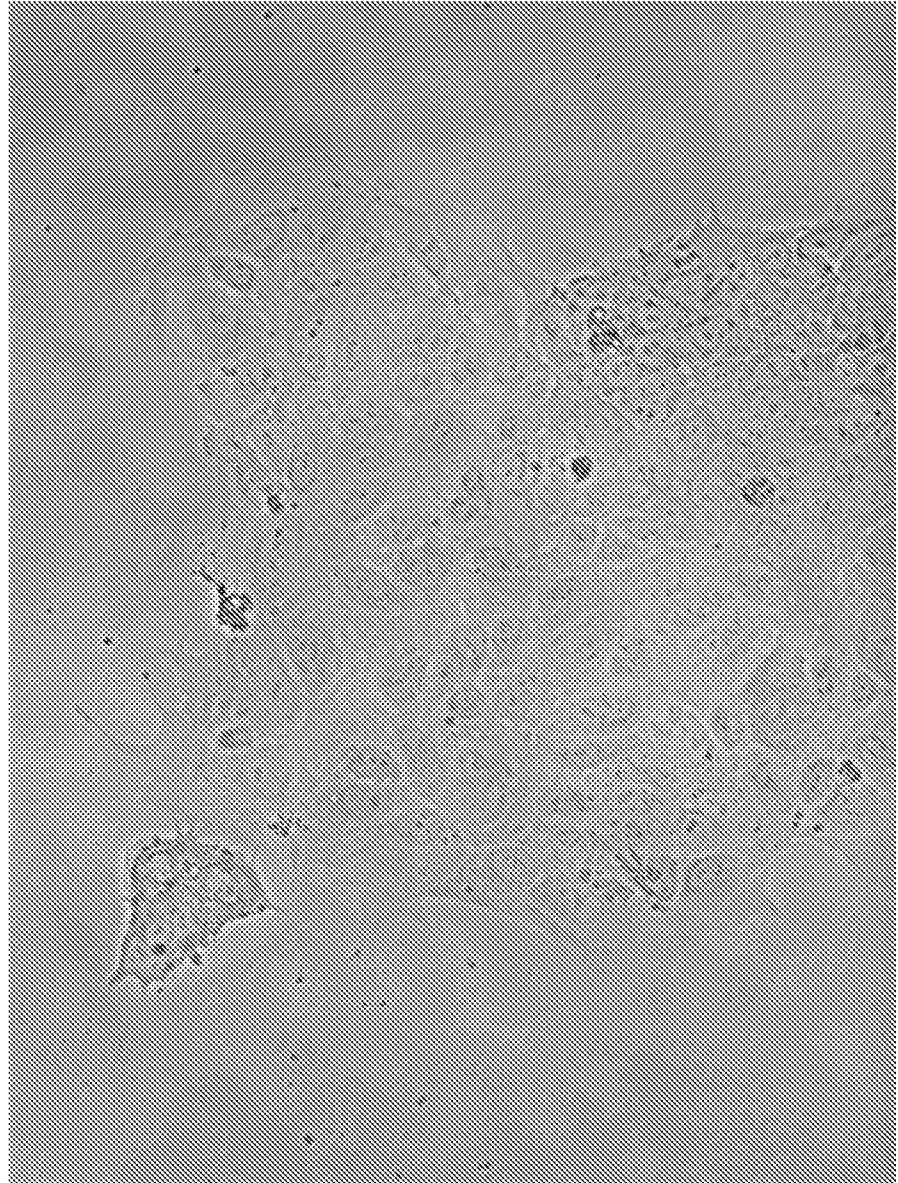


Figure 13. MITF induces differentiation of hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

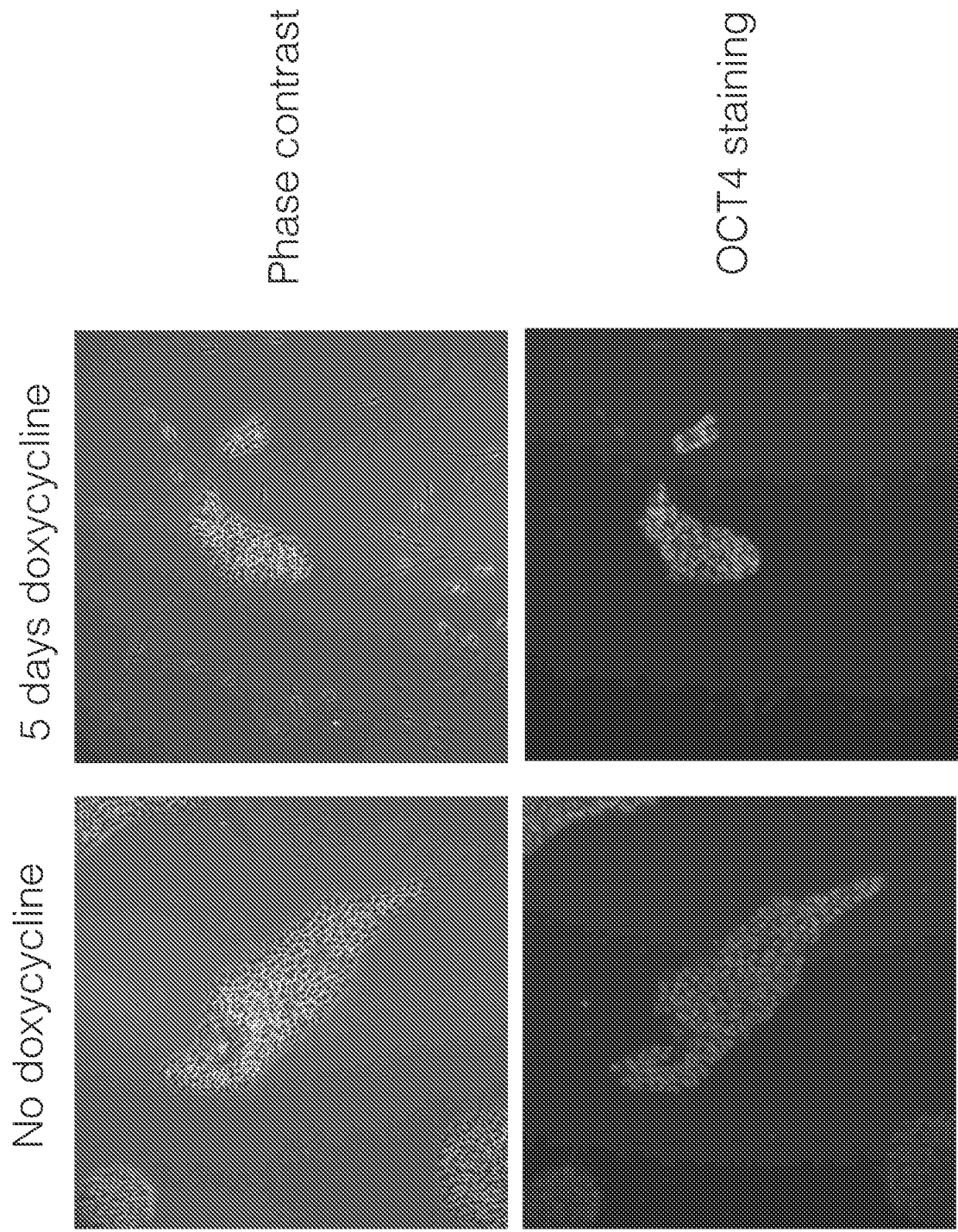


Figure 14. SOX14 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

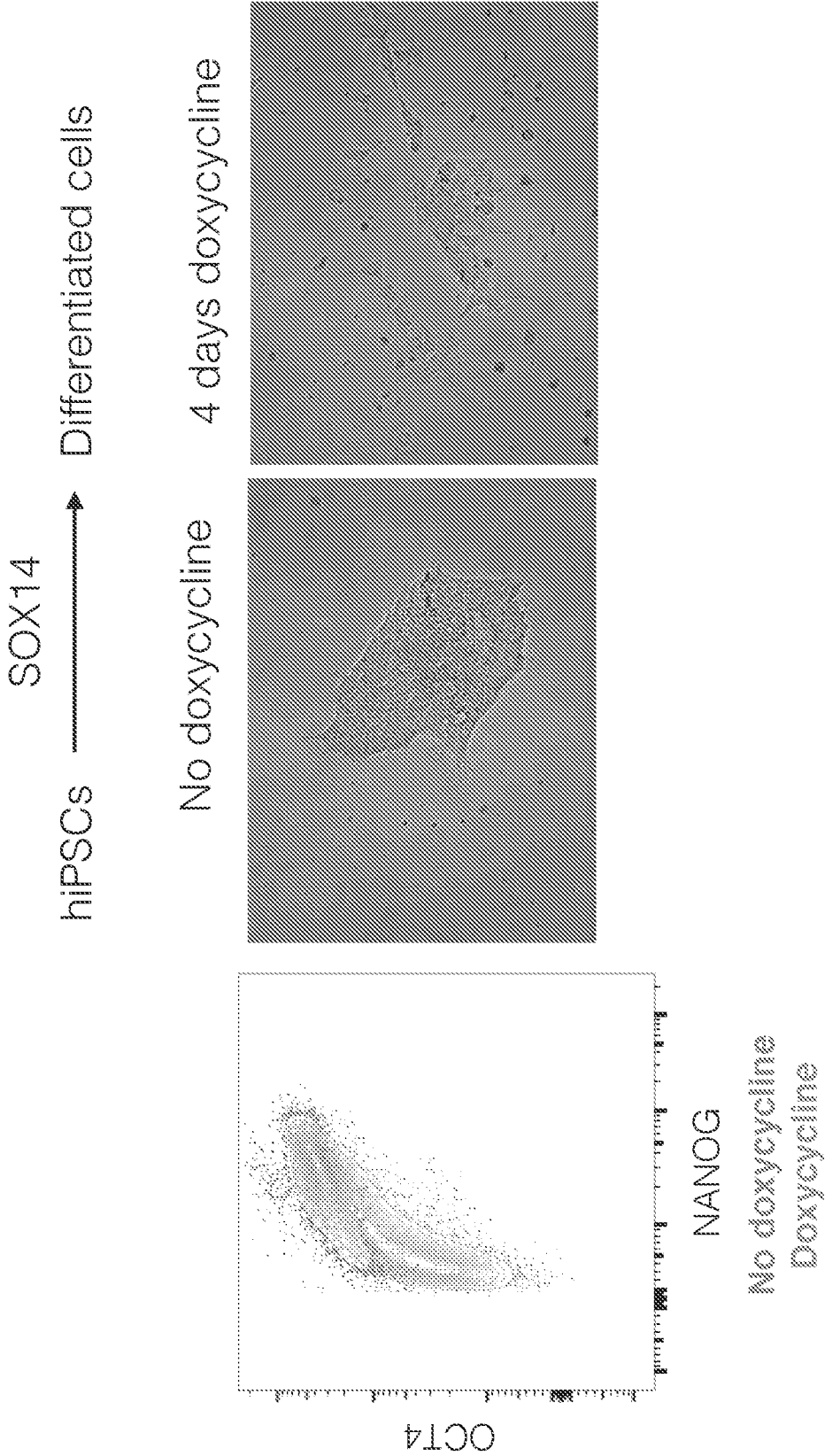


Figure 15. ZSCAN1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

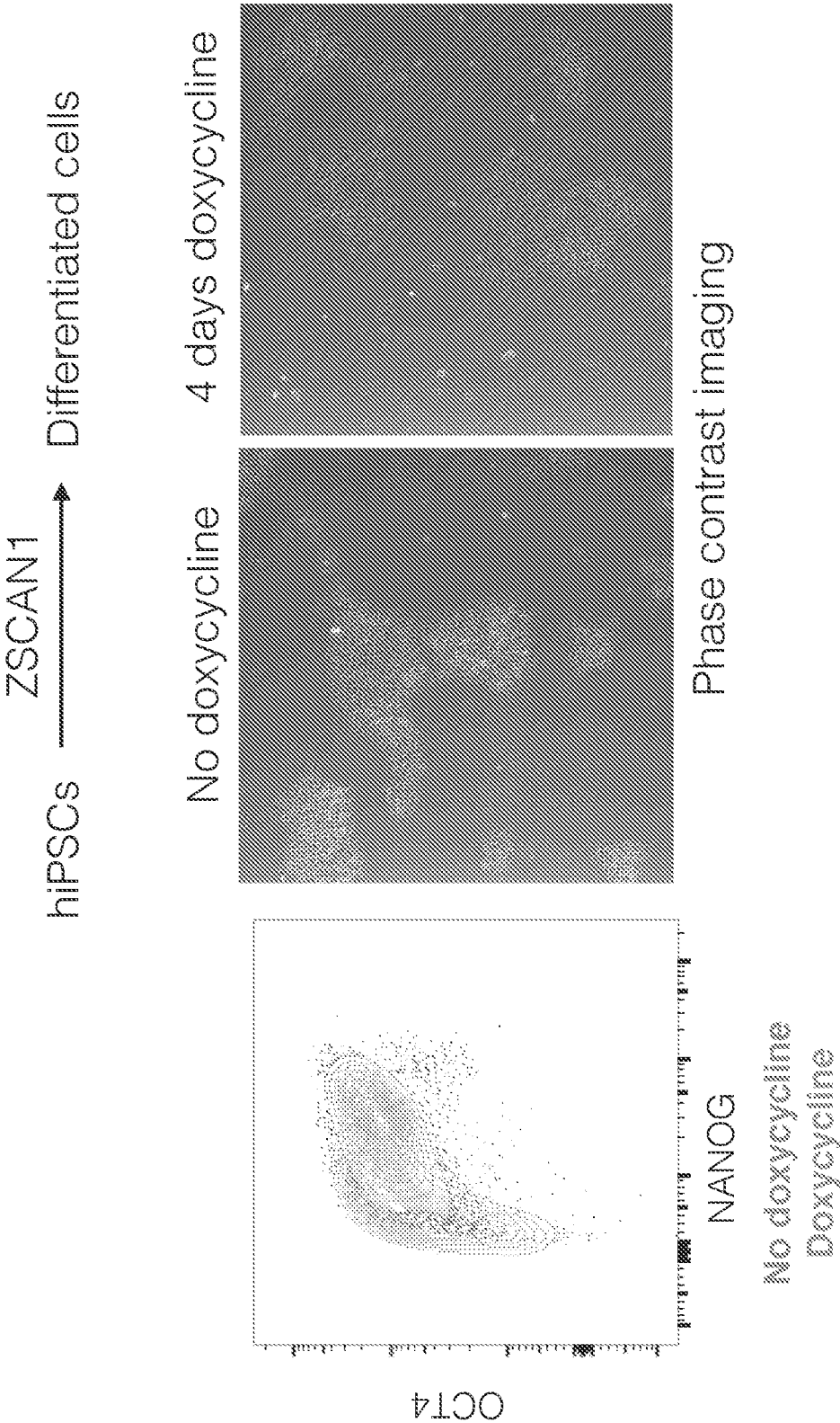


Figure 16. WT1 induces hiPSC differentiation in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

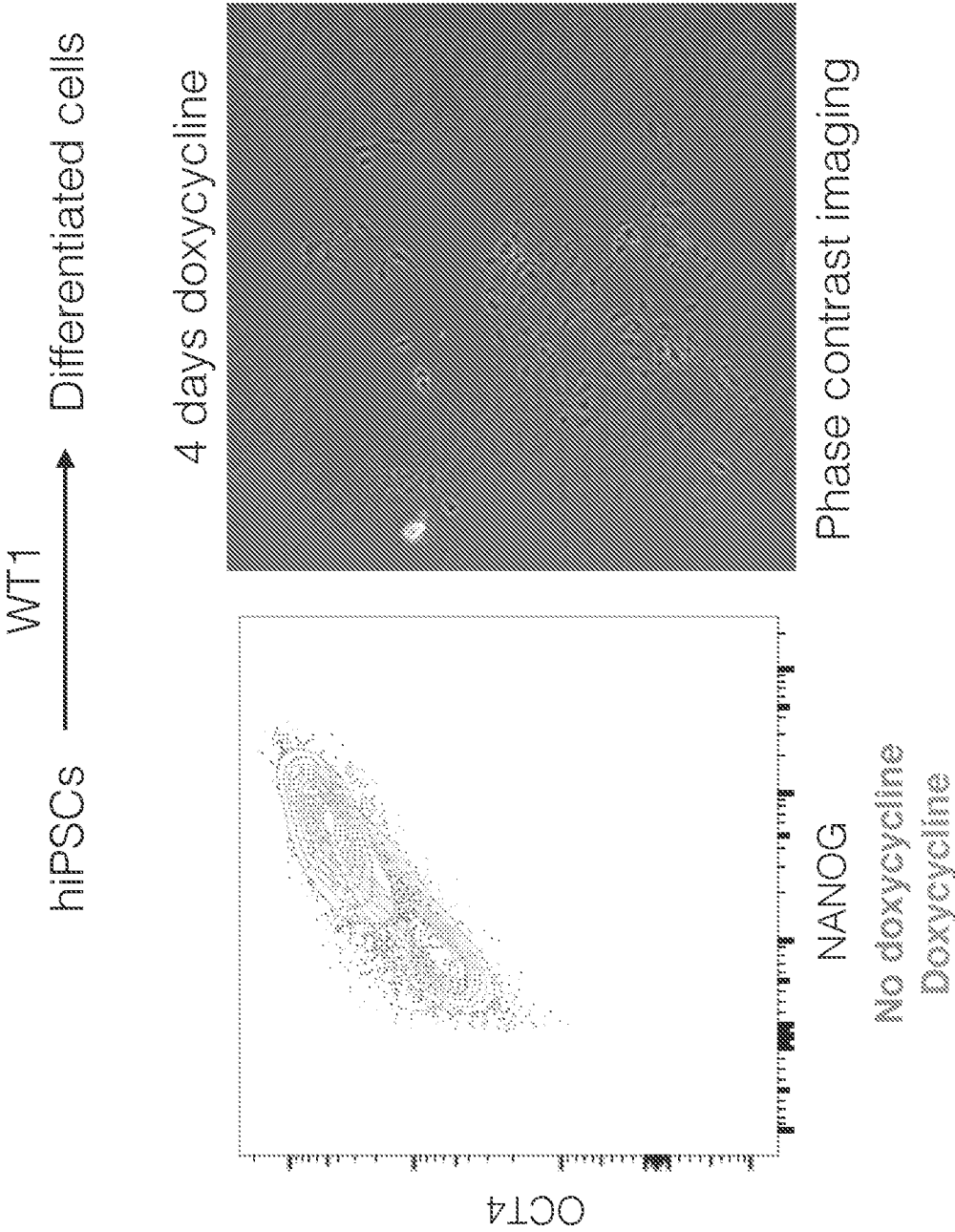


Figure 17. NEUROGENIN1 and NEUROGENIN2 induce neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

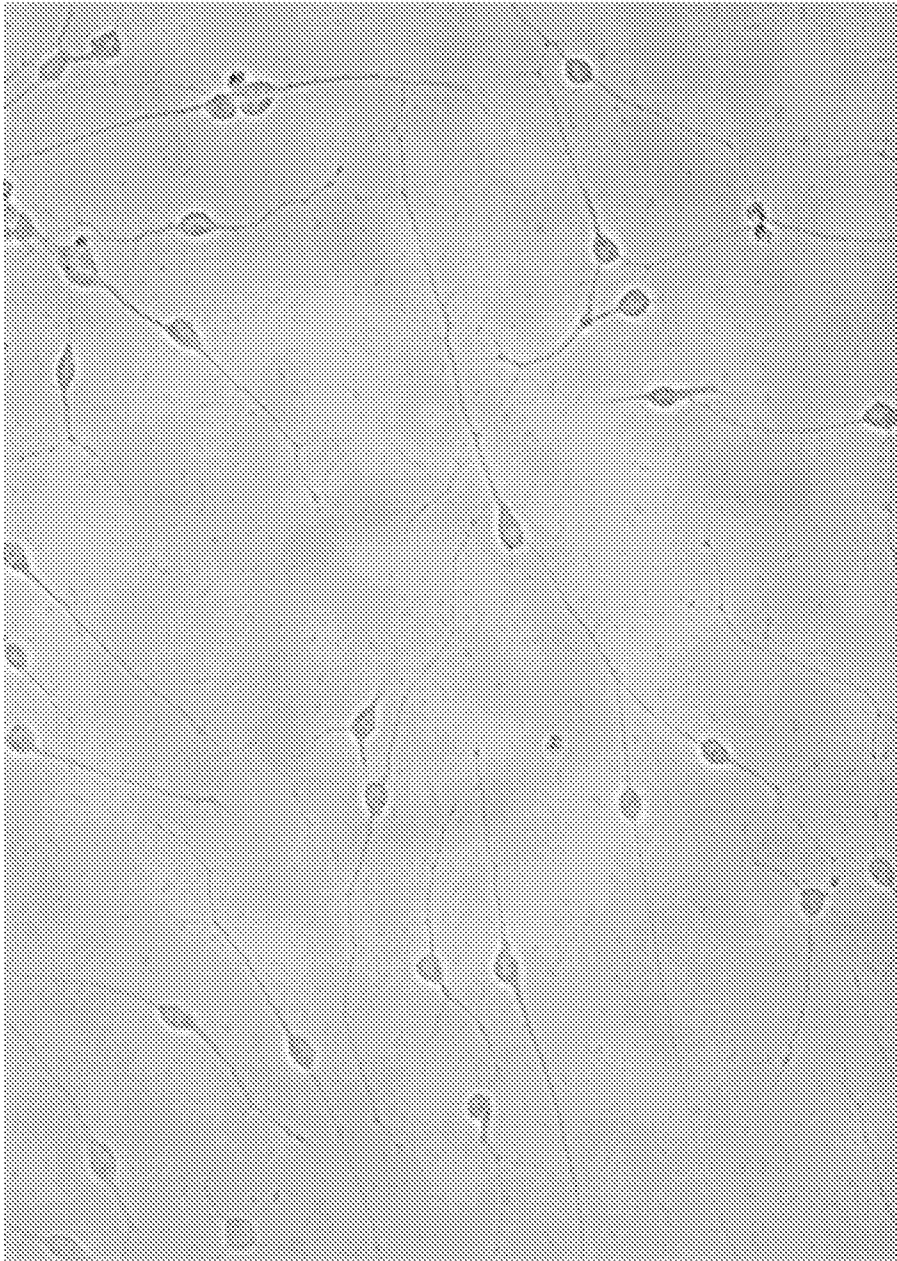
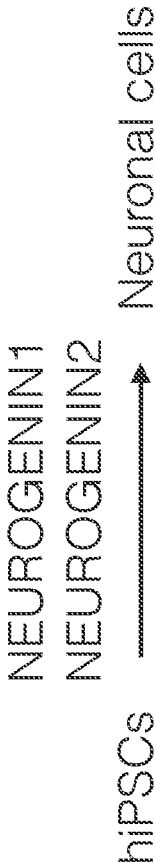


Figure 18. NEUROGENIN1 and NEUROGENIN2 and EMX1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

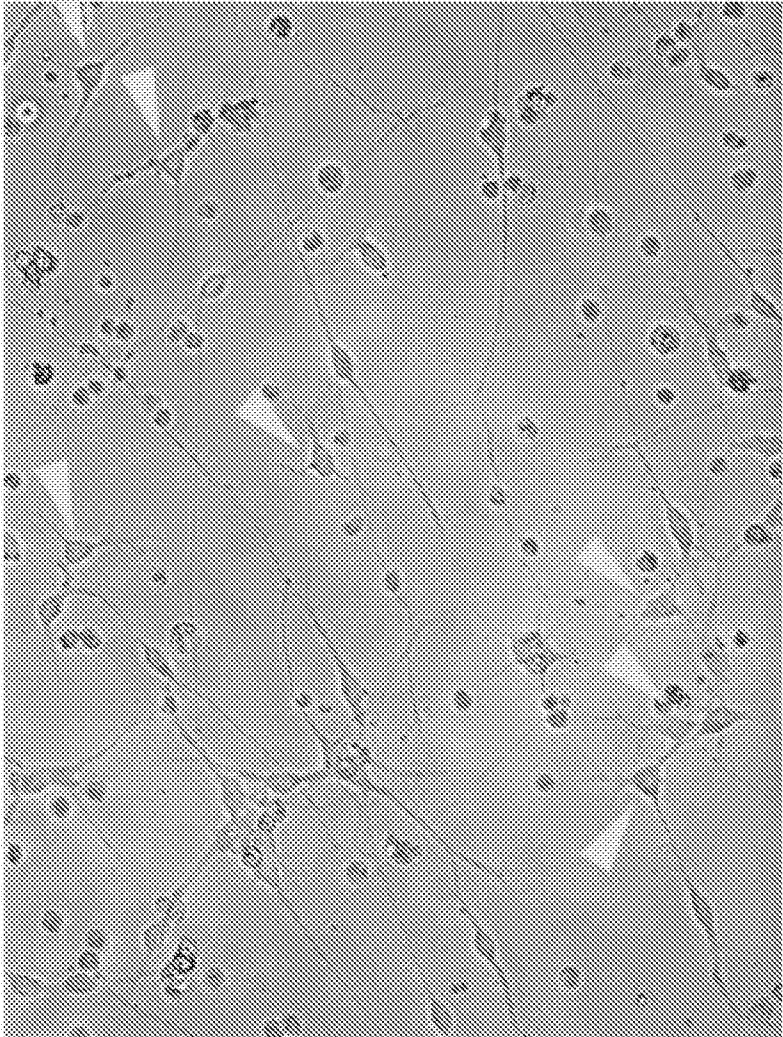
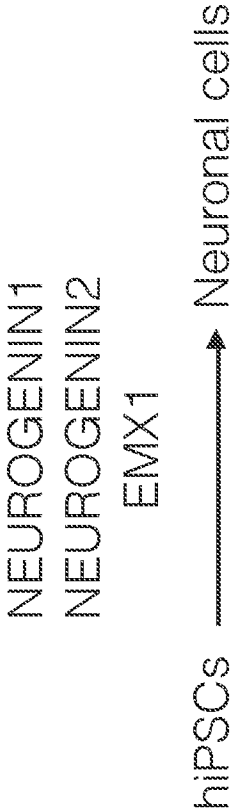


Figure 19. NEUROGENIN1 and NEUROGENIN2 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

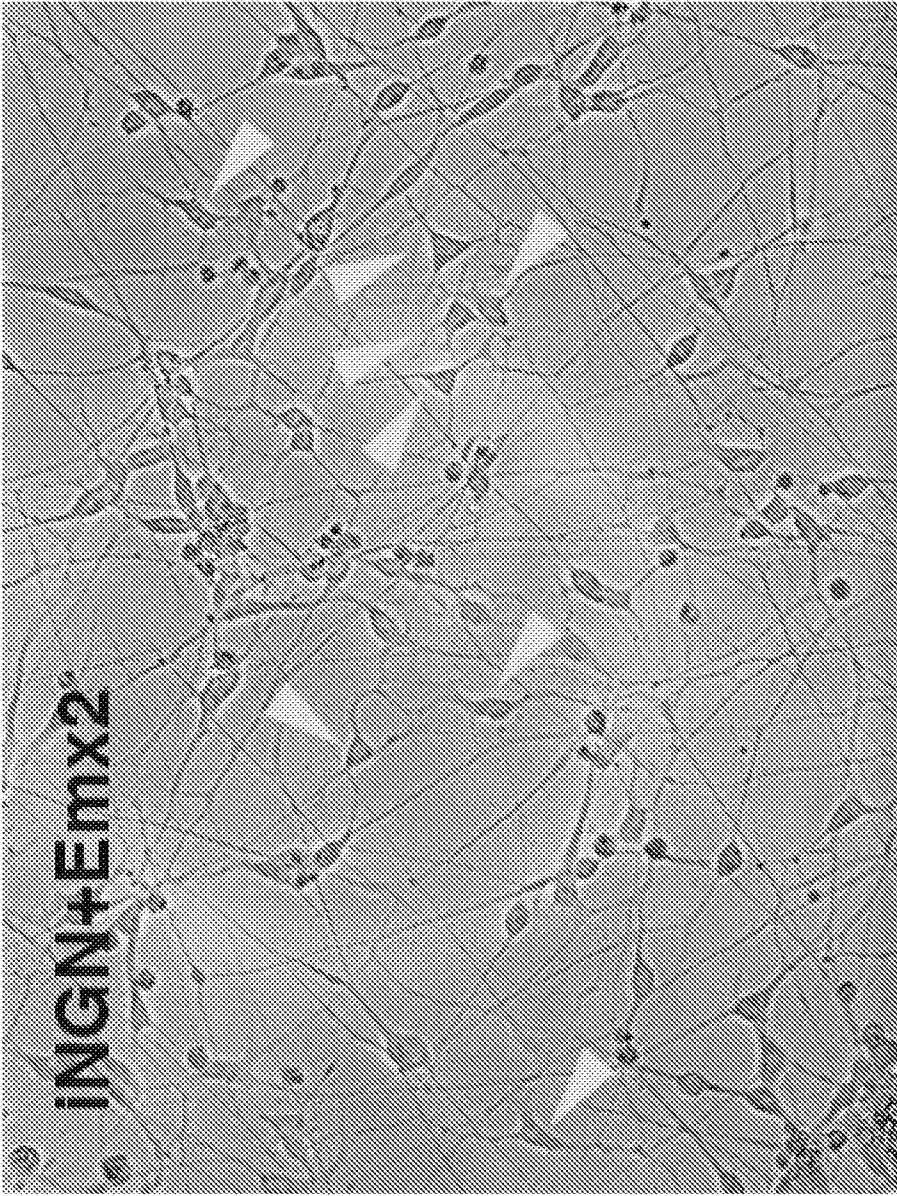


Figure 20. NEUROGENIN1 and NEUROGENIN2 and TBR1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

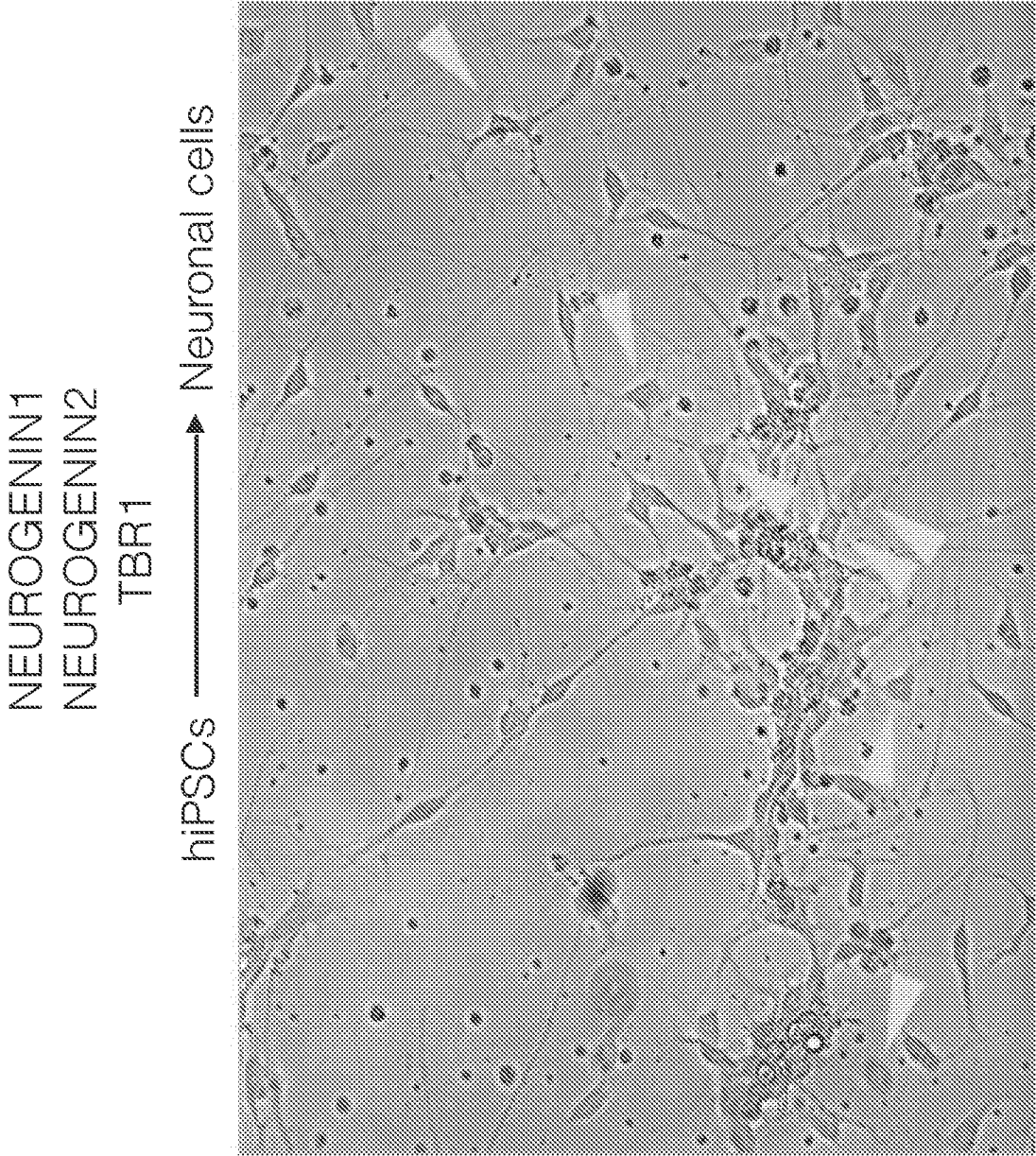


Figure 21. NEUROGENIN1 and NEUROGENIN2 and FOXG1 induce cortical neuronal differentiation from hiPSCs in stem cell media without embryoid body formation, additional growth factors or mechanical manipulations

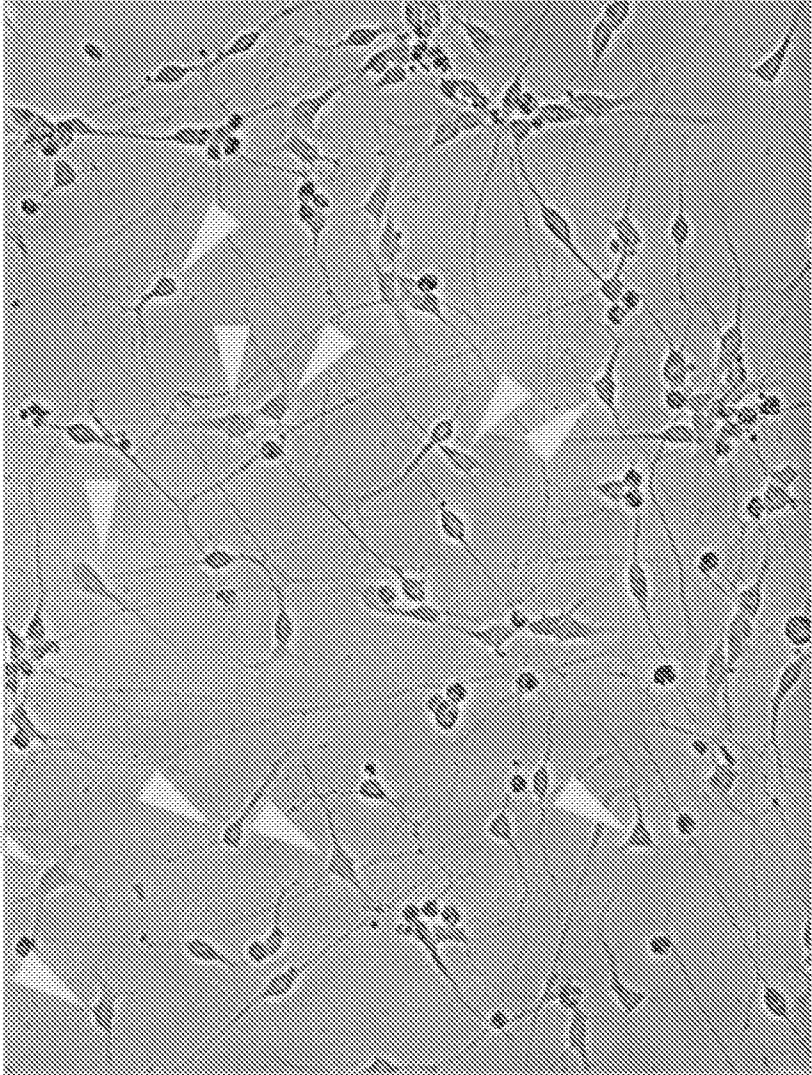
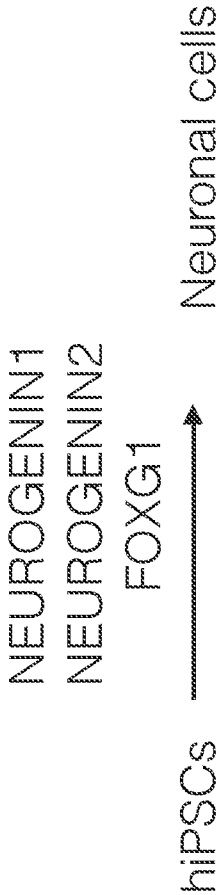


Figure 22. Selected transcription factors (light blue) are lowly or not expressed compared to housekeeping genes (gray) in hiPSCs

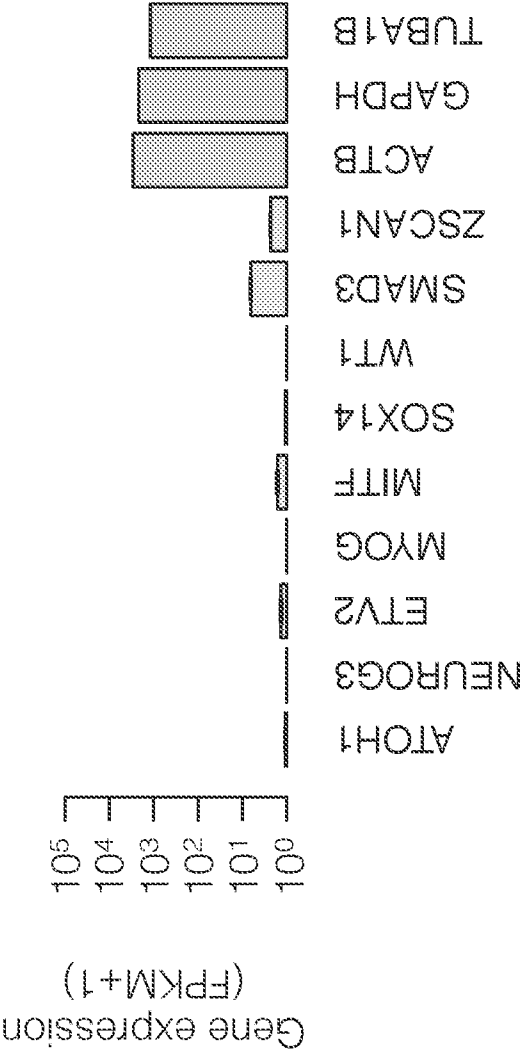
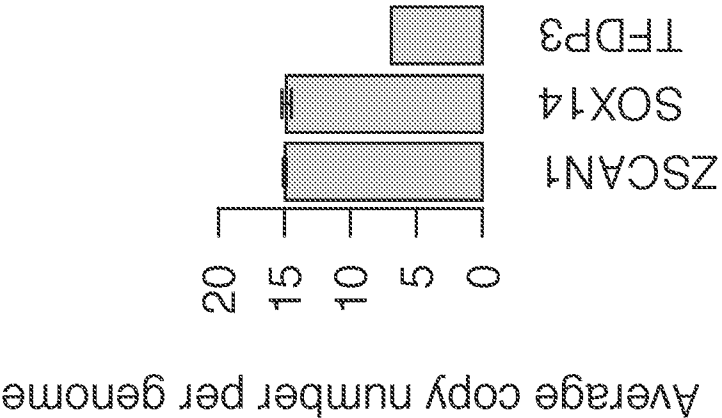


Figure 23. Average copy number of TF ORFs integrated into the hiPSC genome



INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/51122

Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

- Please See the Next Supplemental Page -

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying additional fees, this Authority did not invite payment of additional fees.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Group 1+ Claims 1-3, 11-27, 30, 32, 34-35, 38, 41-42, 45 and 47-51

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest and, where applicable, the payment of a protest fee.
- ☐ The additional search fees were accompanied by the applicant's protest but the applicable protest fee was not paid within the time limit specified in the invitation.
- ☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/51122

A. CLASSIFICATION OF SUBJECT MATTER

IPC - C12N 5/0797, 5/0793, 5/10 (2017.01)

CPC - C12N 5/0695, 5/0696, 5/0627, 5/0619, 5/062, 5/10

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

See Search History document

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

See Search History document

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

See Search History document

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	(SRIPAL, P) Combined Expression of MicroRNAs and Transcription Factors for Promoting Hair Cell Differentiation [online]. Creighton University. 29 July 2013; downloaded from the internet < https://dspace2.creighton.edu/xmlui/handle/10504/47221 > on 27 December, 2017; pages 1-133; abstract; page 5, paragraph 2; page 7, paragraph 2; page 36, paragraph 1; page 64, paragraph 1; page 65, paragraph 2; page 73, paragraph 2	1-3, 11, 18-19, 47, 50
Y	WO 2015/049677 A1 (KADIMASTEM LTD.) 09 April 2015; page 8, lines 18-20; page 35, lines 2-19; claims 1, 10	12-17, 20-27, 30, 32, 34-35, 38, 41-42, 45, 48-49, 51
Y	(CHOI, J et al.) A Comparison of Genetically Matched Cell Lines Reveals the Equivalence of Human iPSCs and ESCs. Nature Biotechnology. 26 October 2015; Vol. 33, No. 11; pages 1-22; page 3, paragraphs 1-3; figures 1C, 2A, 3C; DOI: 10.1038/nbt.3388	12-17
Y	(BUSSKAMP, V et al.) Rapid Neurogenesis through Transcriptional Activation in Human Stem Cells. Molecular Systems Biology. 17 November 2014; Vol. 10; No. 760; pages 1-21; page 3, column 2, paragraph 2 – page 4, column 1, paragraph 1; page 16, column 2, paragraph 2; DOI: 10.15252/msb.20145508.	20-24
Y	(HAENEBALCKE, L et al.) The ROSA26-iPSC Mouse: A Conditional, Inducible, and Exchangeable Resource for Studying Cellular (De)Differentiation. Cell Reports. 07 February, 2013; Vol. 3, No. 2; pages 335-341; figure 3A; page 338, column 2, paragraph 2; DOI: 10.1016/j.celrep.2013.01.016	25-27
Y	US 2011/0154518 A1 (KIM,) 23 June, 2011; paragraphs [0005], [0086]	30
Y		32

☒ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&" document member of the same patent family

Date of the actual completion of the international search

27 December 2017 (27.12.2017)

Date of mailing of the international search report

18 JAN 2018

Name and mailing address of the ISA/

Mail Stop PCT, Attn: ISA/US, Commissioner for Patents
P.O. Box 1450, Alexandria, Virginia 22313-1450

Facsimile No. 571-273-8300

Authorized officer

Shane Thomas

PCT Helpdesk: 571-272-4300
PCT OSP: 571-272-7774

INTERNATIONAL SEARCH REPORT

International application No.

PCT/US17/51122

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	(MULVANEY, J et al.) Atoh1, an Essential Transcription Factor in Neurogenesis and Intestinal and Inner Ear Development: Function, Regulation, and Context Dependency. Journal of the Association for Research in Otolaryngology. 28 February, 2012; Vol. 13, No. 3; pages 281-293; page 282, column 2, paragraph 1; DOI: 10.1007/s10162-012-0317-4	34-35, 38, 41-42, 45
Y	(SAGAL, J et al.) Proneural Transcription Factor Atoh1 Drives Highly Efficient Differentiation of Human Pluripotent Stem Cells Into Dopaminergic Neurons. Stem Cells Translational Medicine. 05 June, 2014; Vol. 3, No. 8; pages 888-898; page 889, column 2, paragraph 2; DOI: 10.5966/sctm.2013-0213	48-49, 51

INTERNATIONAL SEARCH REPORT
Information on patent family members

International application No.

PCT/US17/51122

- Continued from Box III: Observations where unity of invention is lacking -

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Groups I+, Claims 1-27, 30, 32, 34-38, 41-45, 47-51 and ATOH1 are directed toward methods of inducing differentiation of induced pluripotent stem cells; and associated engineered differentiated cells.

The methods and engineered cells will be searched to the extent they encompass a transcription factor encompassing ATOH1 (transcription factor). Applicant is invited to elect additional transcription factor(s) to be searched. Additional transcription factor(s) will be searched upon the payment of additional fees. It is believed that claims 1, 2 (in-part), 3 (in-part), 11-17, 18 (in-part), 19 (in-part), 20 (in-part), 21 (in-part), 22 (in-part), 23 (in-part), 24 (in-part), 25 (in-part), 26 (in-part), 27 (in-part), 30, 32, 34 (in-part), 35 (in-part), 38 (in-part), 41 (in-part), 42 (in-part), 45 (in-part), 47 (in-part), 48 (in-part), 49 (in-part), 50 (in-part) and 51 (in-part) encompass this first named invention and thus these claims will be searched without fee to the extent that they encompass ATOH1 (transcription factor). Applicants must specify the claims that encompass any additionally elected transcription factor(s). Applicants must further indicate, if applicable, the claims which encompass the first named invention, if different than what was indicated above for this group. Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) will result in only the first claimed invention to be searched/examined. An exemplary election would be a transcription factor encompassing NEUROG3 (transcription factor).

Groups II+, Claims 28, 29, 31, 33, 39, 40, 46 and AEBP2 are directed toward a method of maintaining pluripotency of induced pluripotent stem cells; and associated induced pluripotent stem cells.

The method and induced pluripotent stem cells can be searched to the extent they encompass a transcription factor encompassing AEBP2 (transcription factor). Applicant is invited to elect additional transcription factor(s) to be searched. Additional transcription factor(s) can be searched upon the payment of additional fees. It is believed that claims 28, 29 (in-part), 31, 33, 39 (in-part), 40 (in-part) and 46 encompass this first named invention of Groups II+ and thus these claims can be searched with payment of a fee for the search of Groups II+, to the extent that they encompass AEBP2 (transcription factor). Failure to clearly identify how any paid additional invention fees are to be applied to the "+" group(s) can result in only the first claimed invention of Groups II+ to be searched/examined. An exemplary election would be a transcription factor encompassing AIRE (transcription factor).

The inventions listed as Groups I+ and II+ do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons: the special technical features of Groups I+ include a method of inducing differentiation of induced pluripotent stem cells, not present in any of Groups II+; the special technical features of Groups II+ include a method of maintaining pluripotency of induced pluripotent stem cells, not present in any of Groups I+.

Groups I+ and II+ share the technical features including: delivering to induced pluripotent stem cells a nucleic acid comprising an open reading frame encoding a transcription factor; and cell comprising a nucleic acid comprising an open reading frame encoding a protein, wherein the protein is a transcription factor; wherein the open reading frame is intronless.

However, these shared technical features are previously disclosed by US 2014/0234971 A1 to Wisconsin Alumni Research Foundation (hereinafter 'Wisconsin').

Wisconsin discloses delivering to induced pluripotent stem cells (delivering to induced pluripotent stem cells; paragraphs [0009], [0040]) a nucleic acid comprising an open reading frame encoding a transcription factor (modified mRNA encoding a transcription factor; paragraphs [0004], [0009]); and a cell comprising a nucleic acid comprising an open reading frame encoding a protein (a cell comprising a nucleic acid comprising an open reading frame encoding a transcription factor (protein); paragraphs [0004], [0009]), wherein the protein is a transcription factor (wherein the protein is a transcription factor; paragraph [0004]); and wherein the open reading frame is intronless (wherein the nucleic acid comprises modified mRNA (wherein the open reading frame is intronless); paragraph [0009]).

No technical features are shared between the transcription factors of Groups I+ and, accordingly, these groups lack unity a priori. Additionally, even if Groups I+ were considered to share the technical features of including: a method of inducing differentiation of induced pluripotent stem cells, comprising: delivering to the induced pluripotent stem cells a nucleic acid comprising an open reading frame encoding a transcription factor, the transcription factor protein, or an activator of transcription of the open reading frame encoding the transcription factor, whereby the amount of the transcription factor in the induced pluripotent stem cells is increased, and the induced pluripotent stem cells differentiate to form differentiated cells; and an engineered human differentiated cell comprising a nucleic acid comprising an open reading frame that encodes a transcription factor, wherein the open reading frame is intronless; these shared technical features are previously disclosed by Wisconsin, as above.

Wisconsin discloses a method of inducing differentiation of induced pluripotent stem cells (a method of inducing differentiation of induced pluripotent stem cells; paragraphs [0034], [0040], [0041]), comprising: delivering to the induced pluripotent stem cells a nucleic acid comprising an open reading frame encoding a transcription factor (comprising: delivering to the induced pluripotent stem cells modified mRNA (a nucleic acid comprising an open reading frame) encoding a transcription factor; paragraphs [0009], [0040], [0041]), whereby the amount of the transcription factor in the induced pluripotent stem cells is increased (whereby the amount of the transcription factor in the induced pluripotent stem cells is forced (increased); paragraphs [0004], [0041]), and the induced pluripotent stem cells differentiate to form differentiated cells (the induced pluripotent stem cells differentiate to form differentiated cells; paragraph [0034]); and an engineered human differentiated cell (a human hemogenic endothelial cell with pan-myeloid potential (an engineered human differentiated cell); paragraphs [0004], [0041]) comprising a nucleic acid comprising an open reading frame that encodes a transcription factor (comprising modified mRNA (a nucleic acid comprising an open reading frame) encoding a transcription factor; paragraphs [0009], [0040], [0041]), wherein the open reading frame is intronless (wherein the nucleic acid comprises modified mRNA (wherein the open reading frame is intronless); paragraph [0009]).

-Continued Within the Next Supplemental Box-

****-Continued from Previous Supplemental Page: ****-

No technical features are shared between the transcription factors of Groups II+ and, accordingly, these groups lack unity a priori. Additionally, even if Groups II+ were considered to share the technical features of including: a method of maintaining pluripotency of induced pluripotent stem cells, comprising: delivering to the induced pluripotent stem cells a nucleic acid comprising an open reading frame encoding a protein, the protein, or an activator of transcription of an open reading frame encoding the protein, wherein the protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells; whereby the induced pluripotent stem cells maintain expression of keratin sulfate cell surface antigen, TRA-1-60; and an induced pluripotent stem cell, comprising: a nucleic acid comprising an open reading frame encoding a protein, wherein the protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells; wherein the induced pluripotent stem cell expresses keratin sulfate cell surface antigen, TRA-1-60, wherein the open reading frame is intronless; these shared technical features are previously disclosed by US 2010/0093092 A1 to Bamdad, C.C. et al. (hereinafter 'Bamdad') in view of Wisconsin, as above, and US 2013/0029423 A1 to Yamanaka, S. et al. (hereinafter 'Yamanaka').

Bamdad discloses a method of maintaining pluripotency of induced pluripotent stem cells (a method of maintaining pluripotency of induced pluripotent stem cells; abstract, paragraph [0010]), comprising: delivering to cells a nucleic acid comprising an open reading frame encoding an activator of transcription of an open reading frame encoding a protein (comprising contacting cells (delivering to cells) with a gene (a nucleic acid comprising an open reading frame) encoding a transcription factor (an activator of transcription) that increases MUC1* expression (of an open reading frame encoding a protein); paragraphs [0010], [0038]); wherein the protein is a transcription factor (wherein MUC1* (the protein) is a transcription factor; paragraph [0065]); whereby the induced pluripotent stem cells maintain expression of keratin sulfate cell surface antigen, TRA-1-60 (whereby the induced pluripotent stem cells maintain expression of keratin sulfate cell surface antigen, TRA-1-60; paragraphs [0010], [0073]); and an induced pluripotent stem cell (an induced pluripotent stem cell; paragraph [0010]), comprising: a nucleic acid comprising an open reading frame encoding a protein (a gene (a nucleic acid comprising an open reading frame) encoding a transcription factor (a protein); paragraphs [0010], [0038]), wherein the protein is a transcription factor (wherein the protein is a transcription factor; paragraphs [0010], [0038], [0065]); wherein the induced pluripotent stem cell expresses keratin sulfate cell surface antigen, TRA-1-60 (wherein the induced pluripotent stem cell expresses keratin sulfate cell surface antigen, TRA-1-60; paragraphs [0010], [0073]).

Bamdad does not disclose: delivering to induced pluripotent stem cells a nucleic acid; wherein the protein is a transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of a library of transcription factors to a population of induced pluripotent stem cells; and wherein the open reading frame is intronless.

Wisconsin discloses performing gain-of-function genetic screens to identify combinations of transcription factors that guide the acquisition of specific properties (performing gain-of-function genetic screens to identify combinations of transcription factors that guide the acquisition of specific properties; paragraph [0034]); and transducing induced pluripotent stem cells with an open reading frame that is intronless (transducing induced pluripotent stem cells with modified mRNA (an open reading frame that is intronless); paragraphs [0009], [0040]).

Yamanaka discloses preparing a transcription factor library for screening somatic cell reprogramming factors (preparing a transcription factor library for screening somatic cell reprogramming factors; Figure 2, paragraph [0010]) for induction of pluripotent stem cells (for induction of pluripotent stem cells; paragraphs [0108], [0110]).

It would have been obvious to a person of ordinary skill in the art at the time of the invention was made to have modified the disclosure of Bamdad to have delivered one or more transcription factor-encoding open reading frames in the form of modified, intronless mRNA, as disclosed by Wisconsin, in order to better enable the production of cells capable of expressing the one or more transcription factors, such as transcription factors from a library, as disclosed by Yamanaka, to identify transcription factors that improve the maintenance of iPSCs, as disclosed by Bamdad, either by enhancing expression or activity of a specific transcription factor, such as MUC1*, as disclosed by Bamdad, or another transcription factor that is found in a high proportion of stem cells relative to differentiated cells after delivery of the library of transcription factors, in order to further enhance iPSC maintenance by the production of one or more transcription factors useful therefor in the cells.

Since none of the special technical features of the Groups I+ and II+ inventions is found in more than one of the inventions, and since all of the shared technical features are previously disclosed by the Wisconsin reference, and by a combination of the Bamdad, Wisconsin, and Yamanaka references, unity of invention is lacking.