A comprehensive library of human transcription factors for cell fate engineering

Alex H. M. Ng1,2,3,14, Parastoo Khoshakhlagh1,2,3,14, Jesus Eduardo Rojo Arias4,12, Giovanni Pasquini4, Kai Wang5,6, Anka Swiersy4, Seth L. Shipman7, Evan Appleton1,2,3, Kiavash Kiae1,2,3, Richie E. Kohman1,2, Andyna Vernet2, Matthew Dysart1,2, Kathleen Leeper1,2, Wren Saylor1,2, Jeremy Y. Huang1,2, Amanda Graveline2, Jussi Taipale8,9,10, David E. Hill1,11, Marc Vidal11, Juan M. Melero-Martin5,6, Volker Busskamp6,12✉ and George M. Church1,2,3✉

Human pluripotent stem cells (hPSCs) offer an unprecedented opportunity to model diverse cell types and tissues. To enable systematic exploration of the programming landscape mediated by transcription factors (TFs), we present the Human TFome, a comprehensive library containing 1,564 TF genes and 1,732 TF splice isoforms. By screening the library in three hPSC lines, we discovered 290 TFs, including 241 that were previously unreported, that induce differentiation in 4 days without alteration of external soluble or biomechanical cues. We used four of the hits to program hPSCs into neurons, fibroblasts, oligodendrocytes and vascular endothelial-like cells that have molecular and functional similarity to primary cells. Our cell-autonomous approach enabled parallel programming of hPSCs into multiple cell types simultaneously. We also demonstrated orthogonal programming by including oligodendrocyte-inducible hPSCs with unmodified hPSCs to generate cerebral organoids, which expedited in situ myelination. Large-scale combinatorial screening of the Human TFome will complement other strategies for cell engineering based on developmental biology and computational systems biology.

A longstanding goal of stem cell, tissue and organoid engineering is the creation of any cell type or tissue in a facile, controlled manner. This is foundational for accurate cellular modeling in fundamental science, disease modeling, drug discovery and regenerative medicine. hPSCs have a virtually unlimited expansion capability and the potential to differentiate into any cell type. However, reliable protocols to generate most cell lineages are lacking, and existing protocols are often limited by long time scales or modest efficiencies. Many established protocols attempt to mimic the complexity of developmental biology, with its protracted timelines. During development, the process of cell-type specification is intertwined with other regulatory events that spatiotemporally position the proper cell types in defined population sizes but are not relevant to in vitro protocols that aim to rapidly generate large numbers of cells. Most current protocols also depend on external signals with known roles in development, such as soluble factors or mechanical cues, which often converge on TFs to control cell-type-specific genetic programs. This reliance on external cues, which are often incompatible among different specialized cell types, makes it challenging to generate multiple lineages in the same culture.

An alternative approach is direct activation of TFs, which can both shorten the time for cell conversion and isolate it from other developmental events. TF induction has been shown to transdifferentiate cells between lineages1, reprogram somatic cells to pluripotency2 and differentiate stem cells3. Previous studies on applying TFs for cell fate engineering have used either experimental testing of a limited set of developmentally relevant TFs4–8 or computational prediction of TFs inferred from genome-scale datasets with modest accuracy9–11. However, the field has lacked a global experimental study of the estimated 1,600 TFs in the human genome11. The largest screen to our knowledge explored 481 TFs at once11.

After constructing the Human TFome, we screened each TF for its ability to induce differentiation of three human induced pluripotent stem cell (hiPSC) lines. Of the 290 hits, we found that 241 had not been previously associated with cell differentiation. We characterized four of the 290 TFs in depth—ATOH1, NKX3-1, ETV2 and SOX9—which programmed hiPSCs within 4 days into induced neurons, induced fibroblasts, vascular endothelial-like cells and induced oligodendrocytes, respectively. We also demonstrated parallel programming of hiPSCs into two or three defined cell types simultaneously in the same culture without lineage-specific cues. Finally, we developed an orthogonal programming approach by including SOX9-engineered, oligodendrocyte-producing hiPSCs at the genesis of developmentally inspired cerebral organoids, which accelerated myelination.

1Department of Genetics, Blavatnik Institute, Harvard Medical School, Boston, MA, USA. 2Wyss Institute for Biologically Inspired Engineering at Harvard University, Boston, MA, USA. 3GC Therapeutics, Inc, Cambridge, MA, USA. 4Technische Universität Dresden, Center for Molecular and Cellular Bioengineering (CMCB), Center for Regenerative Therapies Dresden (CRTD), Dresden, Germany. 5Department of Surgery, Harvard Medical School, Boston, MA, USA. 6Gladstone Institutes and University of California, San Francisco, San Francisco, CA, USA. 7Department of Biochemistry, University of Cambridge, Cambridge, UK. 8Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden. 9Laboratory of Neural Development and Plasticity, Department of Developmental and Molecular Biology, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA. 10Department of Neurology, Massachusetts General Hospital, Boston, MA, USA. 11Department of Ophthalmology, Medical Faculty, University of Bonn, Bonn, Germany. 12Present address: Wellcome-MRC Cambridge Stem Cell Institute, Jeffrey Cheah Biomedical Centre, Cambridge Biomedical Campus, University of Cambridge, Cambridge, UK. 13These authors contributed equally: Alex H. M. Ng, Parastoo Khoshakhlagh.

✉e-mail: volker.busskamp@ukbonn.de; gchurch@genetics.med.harvard.edu
290 TFs individually induce stem cell differentiation. We generated a library containing 1,564 human TFs, including 1,732 splice-isoform-level open reading frames (ORFs) (Supplementary Fig. 1a,b, Supplementary Table 1 and Methods), annotated based on a rigorous curation\(^1\). Cloning from complementary DNA (cDNA) is a bottleneck in generating ORF collections owing to low or tissue-confined TF expression in the source material. We overcame this de novo synthesis of the 273 ORFs that were not available in existing large ORF collections\(^{15-18}\) (Supplementary Fig. 1c) and merged them to create the Human TFome.

To express the Human TFome in hiPSCs, we first cloned the library as one pool into an all-in-one doxycycline-inducible (Tet-On), puromycin-selectable lentiviral vector (Supplementary Fig. 1d–h). We ensured that cells received, at most, one TF by transducing them with the pooled library at a low multiplicity of infection (MOI) of 0.1, and we confirmed the induction of TF overexpression (Supplementary Fig. 1i). We transduced, in triplicates, three hiPSC lines (PGP1, CRTD5 and ATCC-DYS0100) that have similar characteristics (that is, all reprogrammed from male fibroblasts using Sendai virus) to reduce effects arising from the variability among hiPSC lines. Greater than 87% TF coverage was achieved in all three hiPSC lines (Supplementary Fig. 1j) by transducing a sufficient number of cells.

We used loss of pluripotency as a readout and enriched two cell populations by fluorescence-activated cell sorting (FACS): 1) ‘differentiated’ cells, which we define here as cells that have downregulated expression of pluripotency markers but have not been evaluated for expression of cell-type-specific markers and 2) cells that remained pluripotent (Fig. 1a and Supplementary Fig. 2a). Four days post TF induction (dpi), we stained cells for the pluripotency marker TRA-1-60 (refs. \(^{2,22}\)) and sorted for differentiated cells (low TRA-1-60 signal) and pluripotent cells (high TRA-1-60 signal) (Supplementary Fig. 2b,c). To identify TFs enriched in the two populations, we amplified the integrated TFs using universal primers for Illumina sequencing and sequenced them. A score threshold defined by the ratio of TF counts in TRA-1-60 low versus high gates was used to identify TFs responsible for differentiation (Supplementary Fig. 2d and Methods).

Our screen identified 290 TFs that individually induced differentiation in at least two hiPSC lines and 65 TFs in all three lines (Fig. 1b, Supplementary Fig. 2e and Supplementary Tables 2 and 3), despite the variability among hiPSC lines\(^{11,22}\). A PubMed query indicated that 241 of the 290 TFs (83.1%) and 54 of the 65 TFs (83.0%) have not been previously reported as programming or differentiation factors upon overexpression (Fig. 1b and Supplementary Table 4). The hits included known differentiation-inducing TFs, such as NEUROG1 (ref. \(^{20}\)) and ASCL1 (ref. \(^{21}\)), confirming the biological relevance of our screen. For validation, we selected the top TFs in each of the four overlapping groups of hits from the three lines (Supplementary Fig. 2e), for a total of 16 TFs. For screening, we used lentiviruses owing to their highly titratable transduction, which enabled us to perform a controlled Human TFome screen at a single copy per cell. However, to validate individual TFs, we used PiggyBac transposons, both to avoid the multi-step lentiviral production, which impedes high-throughput cell line engineering, and to reduce the silencing and variable expression associated with lentiviral vectors. Indeed, we observed increased differentiation efficiency using PiggyBac compared to lentivirus for a positive-control TF, NEUROG1 (Fig. 1c). All 16 TFs were successfully validated in the PiggyBac system based on the loss of NANOG, OCT4/POU5F1 or SOX2 expression at 4 dpi (Fig. 1d and Supplementary Fig. 2f).

Next, we characterized the identity and functionality of the differentiated cells induced by four TFs: ATOH1, NKX3-1, ETV2 and SOX9.

ATOH1 programs hiPSCs into induced neurons. ATOH1 was the most efficient driver of differentiation in all three hiPSC lines (Fig. 1d). This TF has been implicated in the development of inner ear hair cells\(^{24}\). It has also been used to induce neuronal differentiation in stem cells in conjunction with soluble neurogenesis cues and culturing techniques\(^{25,26}\). We tested whether ATOH1 alone is sufficient to program hiPSCs into induced neurons without specialized culturing conditions. Indeed, we observed rapid 4-day neuronal programming without changing the media composition, as 99±1% of ATOH1-induced cells were positive for the neuronal marker NCAM (Fig. 2a), and cells exhibited neuronal morphology based on TUBB3 (Fig. 2b) and NF200 protein expression (Supplementary Fig. 3a). ATOH1-programmed cells and primary neurons were transcriptomically similar based on unbiased principal component analysis (PCA) (Fig. 2c) and upregulation of key neuronal genes (Fig. 2d and Supplementary Table 5). This was corroborated by CellNet network-based transcriptomic analysis (Supplementary Fig. 3b) and by the enrichment of neuronal gene ontologies in the upregulated genes (Supplementary Fig. 3c). Furthermore, the downregulation of pluri potency genes was confirmed transcriptionally (Supplementary Fig. 3d). We categorized these cells as ‘induced’, which we define here as cells with concurrent expression of functional markers of a specific lineage and residual levels of pluripotency markers.

To assess the function of the induced neurons, after 4 dpi, we cultured them without further doxycycline administration in neuronal maturation media and performed whole-cell patch-clamp electrophysiology. We observed single action potentials in response to current injection at 7 dpi (Supplementary Fig. 3e), trains of action potentials at 14 dpi (Fig. 2e) and spontaneous action potentials at 21 dpi (Supplementary Fig. 3f; summarized in Fig. 2f). Altogether, these results show that ATOH1 alone can program hiPSCs into induced neurons that remain functional in long-term cultures.

NKX3-1 programs hiPSCs into induced fibroblasts. NKX3-1 was the second most efficient hit from our validation studies (Fig. 1d). This TF is essential in prostate differentiation\(^{27}\) and can replace OCT4/POU5F1 in fibroblast reprogramming into iPSCs\(^{28}\). However, NKX3-1 has not been reported for hPSC differentiation. We used CellNet\(^{29}\) to ascertain the identity of NKX3-1–induced cells based on transcriptomic network analysis: NKX3-1–programmed cells were serendipitously identified as fibroblasts (Supplementary Fig. 4a).

We overexpressed NKX3-1 in hiPSCs with no additional soluble differentiation cues. At 4 dpi, 79±2% of the cells expressed the fibroblast marker VIM as revealed by flow cytometry (Fig. 2g and Supplementary Fig. 4b). NKX3-1-induced cells were also positive for ALCAM and HSP47 by immunostaining (Fig. 2h). The cells had a transcriptomic signature similar to that of primary fibroblasts based on unbiased PCA (Fig. 2i) and upregulation of key fibroblast genes as shown by RNA sequencing (RNA-seq) (Fig. 2j and Supplementary Table 5). Gene Ontology analysis showed upregulation of fibroblast-associated genes (Supplementary Fig. 4c) and downregulation of the pluripotency genes NANOG and OCT4/POU5F1 (Supplementary Fig. 4d).

A hallmark of fibroblasts is their function in wound healing and their ability to remodel the extracellular matrix. These can be assessed in scratch\(^{30}\) and collagen contraction assays\(^{31}\) in vitro. At 4 dpi, without switching to fibroblast-specific media and without further doxycycline administration, after we scratched the monolayer of NKX3-1–induced cells, the cells migrated to seal the wound, whereas non-induced cells did not (Fig. 2k,l). In a separate functional assay, collagen-embedded NKX3-1–induced cells were able to contract and reduce the surface area of the collagen gel (Supplementary Fig. 4e,f). Hence, our data demonstrated that NKX3-1 alone is a potent driver of differentiation into induced fibroblasts, with functionality in vitro.
ETV2 isoform 2 programs hiPSCs into vascular endothelial-like cells. Our differentiation approach, which does not use lineage-specific media, might facilitate the generation of multiple cell types in the same culture. Given the importance of vascularization in building synthetic tissues, along with the indispensable role of ETV2 in vascular differentiation from PSCs, we aimed to determine ETV2’s programming ability without lineage-specific media. We queried ETV2’s rank in our screen, where it was among the top hits in one of the three hiPSC lines. As TF splice isoforms play important roles in cell differentiation, and only ETV2 isoform 1 was included in the screen, we aimed to determine the effect of all four of ETV2’s annotated splice isoforms (Supplementary Fig. 5a) on cell-autonomous vascular cell conversion. The ETV2 isoform used in a previous study for hiPSC differentiation could not be readily ascertained, and the contribution of these isoforms to vascular programming without lineage-specific media has not been examined.

We tested all four isoforms of ETV2 independently and found that isoforms can have a major effect on programming efficiency. Using ETV2 isoform 2 (ENST00000402764.6), 95 ± 0.2% of cells expressed the endothelial cell marker VE-cadherin (CDH5) at 4 dpi (Fig. 3a), compared to only 48 ± 1% using the longer ETV2 isoform 1 (Fig. 3a). The other two isoforms had low to no ability to induce endothelial conversion (isoform 3: 21 ± 2% and isoform 4: 0 ± 0%; Fig. 3a), despite similar integrated vector copy numbers for all four isoforms (Supplementary Fig. 5b). Only isoform 2 induced nearly complete expression of VE-cadherin, with its characteristic cobblestone morphology (Fig. 3b). We examined the functionality of these four cell populations in angiogenesis assays. Only isoform 2 had high tubulogenesis capability (Fig. 3c).

We focused on ETV2 isoform 2, the most potent splice variant, for deeper analysis. At 4 dpi, these cells had a transcriptomic signature similar to that of primary endothelial cells based on unbiased PCA (Fig. 3d), upregulation of key vascular endothelial genes (Fig. 3e and Supplementary Table 5) and a transcriptomic comparison of highly variable genes (Supplementary Fig. 5c). They also had an endothelial cell type classification based on CellNet network-based transcriptomic analysis (Supplementary Fig. 5d) and on Gene Ontology analysis of the upregulated genes (Supplementary Fig. 5e). Furthermore, we detected increased expression of PLVAP, PECAM1 and CDH5 (VE-cadherin) in ETV2 isoform 2-induced cells, along with downregulation of pluripotency genes by single-cell RNA sequencing (scRNA-seq) (Fig. 3f). Downregulation of pluripotency genes was further confirmed by bulk RNA-seq (Supplementary Fig. 5f) and flow cytometry (Supplementary Fig. 5g). Although these cells expressed endothelial functional markers, they continued to express residual amounts of the pluripotency marker OCT4 at day 4. We, therefore, defined them as ‘vascular endothelial-like’ cells. Notably, the cells remained stably programmed after the induction of exogenous ETV2 isoform 2 was stopped (Supplementary Fig. 5h).
We assessed the angiogenic ability of the cells. Using an in vitro tube formation assay, we observed lumens by transmission electron microscopy (TEM) (Fig. 3g) with diameters of 4.3 ± 0.7 µm (Fig. 3h) and tight junctions (Supplementary Fig. 5i), similar to that of capillaries in vivo45. To assess function in vivo, we transplanted ETV2-induced cells subcutaneously into nude mice. Seven days after transplantation, we observed mature blood vessels derived from the cells, as evidenced by human-specific CD31-lined lumens supported by mouse α-smooth muscle actin (SMA)-positive perivascular cells (Fig. 3i). The blood vessels integrated with the host circulatory system and were perfused, as shown by the presence of host red blood cells within their lumens in a serial tissue section (Fig. 3j). We observed 40 ± 6 vessels per mm² on average (Fig. 3k).

**Parallel cell programming in co-cultures.** With the aim of constructing complex tissues, we investigated the concept of parallel programming. In this approach, multiple lineages are generated simultaneously in the same dish in lineage-independent media (Fig. 4a). We mixed inducible neuronal, endothelial and fibroblast hiPSCs in a pairwise fashion with identical ratios (that is, ETV2 + NKX3-1, ATOH1 + NKX3-1 and ATOH1 + ETV2 hiPSCs) and activated TF expression for 4 d. The final cell populations were assessed by immunostaining. All three cell-type-specific markers—MAP2, ALCAM and VE-cadherin—were expressed in the co-cultures as expected (Fig. 4b). We quantified the populations by flow cytometry and observed similar outcomes, with approximately equal proportions of each cell type (Fig. 4c). These data suggest that cell types can be differentiated in parallel, in the same media and without additional soluble factors.

Next, we combined the three cell types, inducible ATOH1, NKX3-1 and ETV2 hiPSCs, in one culture and induced differentiation. The resulting cells expressed markers for fibroblasts, neurons and vascular endothelial cells (Fig. 4b,c). The proportions of each cell type by flow cytometry were 15.9 ± 0.6% NCAM⁺, 22.6 ± 1.1% VIM⁺ and 38.1 ± 1.6% VE-cadherin⁺. We also performed scRNA-seq to assess whether the transcriptomic signatures of the three cell lineages were retained. Therefore, a more sensitive scRNA-seq method was desired rather than maximizing the number of cells analyzed. This was achieved using FACs-based single-cell isolation and Smart-seq2 library preparation, which detects more molecules per cell than microfluidic-based methods46. We observed three distinct populations, with each one correlating to one induced TF (Fig. 4d). In each population, the TF matched its expected cell-type-specific expression signature based on unbiased marker selection because some canonical genes showed weak expression and were not readily detected by scRNA-seq (Fig. 4e). We detected strong cell-type–specific transcriptomic signatures in the few cells that were captured. To test parallel programming in three-dimensional cell culture, we performed similar mixing experiments and generated spheroids. All three lineages appeared, as assessed by microscopy and flow cytometry (Supplementary Fig. 6a,b).

**SOX9 programs hiPSCs into induced oligodendrocytes.** Organoid technology is limited by the long timelines required for the emergence of certain cell types. In cerebral organoids, for instance, it takes 103–210 d for mature myelin to form48,49. There are no reported protocols, to our knowledge, to differentiate hiPSCs to myelin-producing oligodendrocytes in lineage-independent media using TFs48,49. To address this need with the Human TFome library, we identified 15 TFs involved in oligodendrocyte differentiation and maturation in vivo using prior knowledge from developmental biology (Supplementary Table 6). We queried their rank in our screen, which yielded SOX9 as the top hit. SOX family members have broad roles in the development of various tissues48,49, and SOX9 has been used in combination with NFIB and soluble factors to induce astrocyte differentiation48. However, SOX9 has not been reported, individually or in a combination, to program hiPSCs into oligodendrocytes.

At day 4 after SOX9 induction, hiPSCs were converted to oligodendrocyte progenitors without additional lineage-specifying cues (Fig. 5a). The cells expressed the hallmark oligodendrocyte progenitor marker O4 (82 ± 6%) and were positive for NG2, a glial progenitor marker, by immunostaining (Fig. 5b). Their transcriptomic signatures were similar to primary oligodendrocytes based on unbiased PCA (Fig. 5c), upregulation of key oligodendrocyte genes (Fig. 5d and Supplementary Table 5) and transcriptomic comparison of highly variable genes (Supplementary Fig. 7a). We also detected expression of the oligodendrocyte lineage markers CSPG4 (NG2) and MYRF by scRNA-seq (Fig. 5e). Pluripotency genes were downregulated in both scRNA-seq (Fig. 5e) and bulk RNA-seq (Supplementary Fig. 7b).

To assess the myelination potential of SOX9-induced cells, we applied parallel programming to produce an oligo-neuronal co-culture. We combined inducible SOX9 hiPSCs along with our previously described, fully characterized hiPSC-derived inducible neurons, which project long axons upon differentiation50. We then activated TF expression. At 3 dpi, without additional external culture-specific factors, we observed oligodendrocytes contacting
axons and beginning the ensheathment process (Supplementary Fig. 7c). Robust myelin sheaths around axons were visualized by TEM after 30 d of co-culture in photo-micropatterned microchannels\(^{43,44}\) (Fig. 5f). We computed G-ratios, a metric for compact myelin, to be 0.56 ± 0.02 (Fig. 5g), which are similar to those of the human brain (0.5–0.8)\(^{45}\). These results confirmed the in vitro myelination functionality of SOX9-induced oligodendrocytes.

To evaluate the in vivo engraftability and functionality of SOX9-induced oligodendrocytes, we transplanted the cells using established methods\(^{39,46}\) into Shiverer mice. This congenitally hypomyelinated mouse model has a knockout in the myelin basic protein (MBP) gene. We detected engraftment of SOX9-induced cells in immunostained brain sections 10 weeks after transplantation based on the presence of MBP (Fig. 5h), which can be expressed only in the donor cells. No MBP was observed in the control group. Furthermore, based on TEM of brain cross-sections, we observed compact myelin in mice transplanted with SOX9-induced cells but rarely in the control group (Fig. 5i). The number of myelinated axons in mice transplanted with cells was significantly higher than in the control group (Fig. 5j; \(P < 3.1 \times 10^{-5}\)).
Taken together, our results demonstrate that induction of SOX9 alone was sufficient to program hiPSCs cell autonomously into induced oligodendrocytes. The molecular profiles of the cells with continuous doxycycline administration were similar to those of primary oligodendrocytes, and the cells could form compact myelin in a co-culture and in hypomyelinated mice.

Orthogonal programming accelerates the myelination of cerebral organoids. To accelerate myelination in cerebral organoids and construct more accurate models of human brain tissue, we harnessed the SOX9-induced oligodendrocytes to introduce the concept of orthogonal cell programming (Fig. 5k). In this approach, cell-autonomous TF overexpression is used in conjunction with a differentiation method that uses external cues (Supplementary Fig. 7d). To this end, we combined inducible SOX9 hiPSCs with unmodified hiPSCs and induced cerebral organoid formation. After 4 d, we added doxycycline to induce SOX9 expression in the orthogonally programmed group and compared it with the control group where SOX9 was not induced. At 40 dpi, we observed myelin oligodendrocyte glycoprotein (MOG) in immunostained cross-sections of orthogonally programmed organoids (Fig. 5l) but not in the controls. TEM on these cross-sections showed robust
resourced by Nature Biotechnology

Discussion

The Human TFome enables systematic investigation of TF-based programming for cell engineering. We conducted a cell-type-agnostic screen of the complete library in three hiPSC lines using cell-type-specific media components for the initiation of differentiation, even if TFs are also used. The ATOH1-induced neurons are similar to those described in our previous work, which were generated using TFs of the Neurogenin family, in that they do not require neuronal-induction culturing conditions, which is a major requirement for neuronal-induction culturing conditions, which is a major...
Fig. 5 | SOX9 induces oligodendrocytes that engraft and form compact myelin in vivo and in cerebral organoids. a, SOX9 rapidly and efficiently programs hiPSCs into induced oligodendrocytes at 4 dpi. Bar plot of flow cytometry for O4 oligodendrocyte marker compared to non-induced cells. Mean ± s.e.m., n = 3 biologically independent samples per group, two-sided Student’s t-test. b, SOX9-induced cells exhibit O4 and NG2 oligodendrocyte protein marker expression at 4 dpi compared to non-induced cells using immunofluorescent staining. Scale bar, 100 μm. Experiments were performed independently at least three times with similar results. c, SOX9-induced cells are transcriptionally similar to primary oligodendrocytes. PCA of RNA-seq samples from SOX9-induced cells (purple) overlap with samples from primary mature oligodendrocytes (OLs; orange), with similarity to oligodendrocyte progenitor cells (OPCs; red) and are distinctly separated from newly formed oligodendrocytes (OLs; brown) and PGPi hiPSCs (gray). d, SOX9-induced cells show similar upregulation of oligodendrocyte markers as primary oligodendrocytes. Heat map of oligodendrocyte gene expression profiles. e, SOX9-induced cells form a homogeneous cluster that expresses oligodendrocyte markers and loses pluripotency markers. UMAP plot of scRNA-seq samples of SOX9-induced cells 4 dpi (right cluster) and hiPSCs (left cluster) showing gray (low) to red (high) for the level of gene expression of the indicated oligodendrocyte marker. f, SOX9-induced cells exhibit O4 and NG2 oligodendrocyte protein marker expression at 4 dpi compared to non-induced cells using immunofluorescent staining. Scale bar, 100 μm. Experiments were performed independently at least three times with similar results. g, Quantification of G-ratio for compact myelin is within the physiological range. Mean ± s.e.m., n ≥ 3 biologically independent samples. *** P < 0.001. Exact P values are provided in Supplementary Table 7. h, Transplanted SOX9-induced cells engraft and express MBP in Shiverer (MBP knockout) mice at 2.5 months. Immunofluorescent staining of a brain tissue section for the MBP myelin marker after PBS injection or SOX9-induced cell transplantation into Shiverer mice. Scale bar, 200 μm. Experiments were performed independently at least three times with similar results. i, Transplanted SOX9-induced cells form compact myelin in Shiverer mice. TEM of a cross-section from Shiverer mice after PBS injection or SOX9-induced cell transplantation. Scale bar, 200 nm. Experiments were performed independently at least three times with similar results. j, Shiverer mice with transplanted SOX9-induced cells have significantly more myelinated axons than PBS-injected animals. Quantification of the number of myelinated axons from Shiverer mice after PBS injection or SOX9-induced cell transplantation. Mean ± s.e.m., n = 12 micrographs taken at distinct locations derived from two PBS-injected animals; n = 21 micrographs taken at distinct locations derived from three species with cell transplantation, two-sided Student’s t-test. k, Schematic of orthogonal programming where engineered hiPSCs for TF-inducible differentiation are incorporated at the genesis of developmentally inspired cerebral organoids to synthetically accelerate myelination. l, Orthogonal programming of inducible SOX9 cells within cerebral organoids accelerated expression of the MOG myelin marker. Immunofluorescent staining of a cerebral organoid section for MOG myelin marker and NeuN neuronal marker in orthogonally induced versus non-induced organoids. Scale bar, 100 μm. Experiments were performed independently at least three times with similar results. m, SOX9 orthogonally programmed organoids for compact myelin. TEM of myelin in a cerebral organoid. Yellow region magnified on the right. Scale bar, 200 nm. Inset scale bar, 100 nm. Experiments were performed independently at least three times with similar results. n, Quantification of G-ratio for compact myelin in cerebral organoids shows physiological resemblance. Mean ± s.e.m., n = 3 biologically independent samples. ** P < 0.01. Exact P values are provided in Supplementary Table 7.
advantage. Our NKX3-1-induced fibroblast protocol represents an advance in that no TF-based differentiation protocol for fibroblasts has been reported, and current differentiation methods require embryoid body formation, collagen embedding and replating over 3 weeks of culture with unknown conversion efficiency\(^{48}\). Our ETV2 isoform 2 protocol for vascular endothelial-like cells generates 95% VE-cadherin\(^+\) cells compared to ~40% in previous work\(^{49-51}\), emphasizing the importance of isoforms in cell programming. Our SOX9-induced oligodendrocyte protocol is more streamlined and facile than a previous TF-based method\(^{14}\), which required multi-step, cell-type-specific media conditions to bring hiPSCs to the neural progenitor stage, followed by glial-specific soluble factors to promote oligodendrocyte differentiation.

We found that PiggyBac transposons, which were optimized for high-throughput individual cell line engineering, significantly improved differentiation efficiency compared to lentiviral gene transfer. We demonstrated that long-term TF induction is not required to differentiate cells into a stable cell fate. ATOH1-induced neurons and NKX3-1-induced fibroblasts were functional in vitro (Fig. 2e, i), and ETV2-induced vascular endothelial-like cells were functional in vivo (Fig. 3i–k) without continuous TF induction. The practical benefit is the alleviation of long-term replenishment of TF-induction molecules to achieve differentiation and the possibility of excising PiggyBac transposons from the genomes of engineered cells\(^{52}\).

A major challenge in tissue engineering is the generation of multiple cell types in one culture, especially supporting cells such as fibroblasts or endothelial cells\(^{12}\). We showed that our lineage-independent differentiation approach can help address this issue. Previously, liver bud-like tissues containing multiple lineages were produced through heterogeneous expression of *Gata6* in hiPSCs in the same culture\(^{12}\). However, control over cell identities and the ratios of each cell type were not attained owing to the lack of genetic switches to tune these parameters. Parallel programming using TFs discovered from our screen allowed us to achieve high-resolution control over cell-type composition and proportion. This enabled co-differentiation of hiPSCs into induced neurons, fibroblasts and vascular endothelial-like cells without lineage-specific cues. We also used parallel programming of induced oligodendrocytes and neurons to rapidly construct robust in vitro myelination models, featuring physiological cell–cell interactions.

Further research is needed to understand how cell types differentiated in parallel compare with homogeneously differentiated cells. Although our four protocols could override pluripotency networks to achieve nearly complete differentiation in 4 d, the resulting cells were not transcriptomically equivalent to their primary cell counterparts, despite having the expected functionality. We have not explored whether our approach can mitigate the effects of different media compositions or the variability in differentiation propensity between hiPSC lines\(^{12,122}\). Another important avenue for future research is to investigate how many specialized cell types can be generated by single or combinatorial TF expression.

As demonstrated here, we envisage that the Human TFome will be a powerful asset in conjunction with other approaches for cell and tissue engineering. As with other genome-scale methods, interpreting the results of Human TFome screens requires information from developmental biology, computational systems biology, cell atlases, functional assays or other sources. In this study, to understand the specific markers for the four analyzed cell types; to learn that the top hit from our screen, ATOH1, is involved in neuronal differentiation; to identify ETV2 as a driver of vascular endothelial programming; to select 15 TFs involved in oligodendrocyte development; and to validate 49 of our hits as previously associated with differentiation, we referred to the developmental biology literature. The systematic, empirical nature of the Human TFome approach is well suited to integration with computational approaches, such as CellNet\(^{1}\), to iteratively refine cell programming. We used CellNet here to identify the phenotype of our second most efficient hit, NKX3-1.

For organoid engineering, our orthogonal programming strategy complements existing methods such as assembloids\(^{53-55}\), genetic approaches\(^{51,56}\), and in vivo approaches\(^{57}\). We overcame the slow emergence of oligodendrocytes in cerebral organoids by including SOX9-induced hiPSCs at the beginning of organoid production. Current protocols for myelinated organoids rely on external developmental signaling and require 103–210 d for myelin formation\(^{46,57}\). Orthogonal programming reduced the timeline by more than half, to 40 dpi, to achieve accelerated myelination with mature myelin markers and compact myelin formation with the correct cytoarchitecture. Orthogonal programming might also allow for the incorporation of missing cell types from diverse lineages, distribution of cells more homogeneously within the tissue and fine control over cell proportions, which would all facilitate the adoption of more physiological tissue architecture.

The Human TFome is synergistic with mammalian cell atlases\(^{56,60-65}\). Whereas cell atlas projects ‘read’ cell types and states, the Human TFome enables TF screening aimed at ‘writing’ the cellular programs and has the potential to produce cell types and states that are currently inaccessible. The concepts presented here enable cross-pollination of tissue engineering and genomics to discover recipes for each cell type and state throughout development and aging.

### Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41587-020-0742-6.

Received: 3 November 2019; Accepted: 19 October 2020; Published online: 30 November 2020

### References

1. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **31**, 987–1000 (1987).
2. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
3. Zhang, Y. et al. Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* **78**, 785–798 (2013).
4. Parekh, U. et al. Mapping cellular reprogramming via pooled overexpression screens with paired fitness and single-cell RNA-sequencing readout. *Cell Syst.* **7**, 548–555 (2018).
5. Tsunemoto, R. et al. Diverse reprogramming codes for neuronal identity. *Nature* **557**, 375–380 (2018).
6. Pritsker, M., Ford, N. R., Jenq, H. T. & Lemischka, I. R. Genomewide systematic analyses of cell states perturbed by inducing single transcription factors. *Cell* **175**, 598–599 (2018).
7. Theodorou, E. et al. A high throughput embryonic stem cell screen identifies Oct-2 as a bifunctional regulator of neuronal differentiation. *Genes Dev.* **23**, 575–588 (2009).
8. Yamamizu, K. et al. Identification of transcription factors for lineage-specific ESC differentiation. *Stem Cell Rep.* **1**, 545–559 (2013).
9. Cahan, P. et al. CellNet: network biology applied to stem cell engineering. *Cell* **158**, 903–915 (2014).
10. Rackham, O. J. et al. A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* **48**, 331–335 (2016).
11. D’Allessio, A. C. et al. A systematic approach to identify candidate transcription factors that control cell identity. *Stem Cell Rep.* **5**, 763–775 (2015).
12. Lambert, S. A. et al. The human transcription factors. *Cell* **175**, 598–599 (2018).
13. Nakatake, Y. et al. Generation and profiling of 2,135 human ESC lines for the systematic analyses of cell states perturbed by inducing single transcription factors. *Cell Rep.* **31**, 107655 (2020).
14. Vaipierzas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Lascombe, N. M. A census of human transcription factors: function, expression and evolution. *Nat. Rev. Genet.* **10**, 252–263 (2009).
15. Jolma, A. et al. DNA-binding specificities of human transcription factors. Cell 152, 327–339 (2013).
16. Seiler, C. Y. et al. DNASeq plasmid and PSI:Biology-Materials repositories: resources to accelerate biologic research. Nucleic Acids Res. 42, D1253–D1260 (2014).
17. Yang, X. et al. A public genome-scale lentiviral expression library of human ORFs. Nat. Methods 8, 659–661 (2011).
18. Wismann, S. et al. The ORFeome Collaboration: a genome-scale human ORF-clone resource. Nat. Methods 13, 191–192 (2016).
19. Adewumi, O. et al. Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. Nat. Biotechnol. 25, 803–816 (2007).
20. Russkamp, V. et al. Rapid neurogenesis through transcriptional activation in human stem cells. Mol. Syst. Biol. 10, 760 (2014).
21. Choi, J. et al. A comparison of genetically matched cell lines reveals the equivalence of human iPSCs and ESCs. Nat. Biotechnol. 33, 1173–1181 (2015).
22. Cahan, P. & Daley, G. Q. Origins and implications of pluripotent stem cell variability and heterogeneity. Nat. Rev. Mol. Cell Biol. 14, 357–368 (2013).
23. Chanda, S. et al. Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem Cell Rep. 3, 282–296 (2014).
24. Bermingham, N. A. et al. Math1: an essential gene for the generation of inner ear hair cells. Science 284, 1837–1841 (1999).
25. Sagal, J. et al. Pronuclear transcription factor Atoh1 drives highly efficient differentiation of human pluripotent stem cells into dopaminergic neurons. Stem Cells Transl. Med. 3, 888–898 (2014).
26. Xue, Y. et al. Synthetic mRNAs drive highly efficient iPSC cell differentiation to dopaminergic neurons. Stem Cells Transl. Med. 8, 112–123 (2019).
27. Dutta, A. et al. Identification of an NXX3-1-G9a-UTY transcriptional regulatory network that controls prostate differentiation. Science 352, 1576–1580 (2016).
28. Mai, T. et al. NXX3-1 is required for induced pluripotent stem cell reprogramming and can replace OCT4 in mouse and human iPSC induction. Nat. Cell Biol. 20, 900–908 (2018).
29. Radley, A. H. et al. Assessment of engineered cells using CellNet and RNA-seq. Nat. Protoc. 12, 1089–1102 (2017).
30. Liang, C. C., Park, A. Y. & Guan, J. L. In vitro scratch assay: a convenient and inexpensive method for analysis of cell migration in vitro. Nat. Protoc. 2, 329–333 (2007).
31. Bell, E., Ivarsson, B. & Merrill, C. Production of a tissue-like structure by contraction of collagen lattices by human fibroblasts of different proliferative potential in vitro. Proc. Natl Acad. Sci. USA 76, 1274–1278 (1979).
32. Lee, D. et al. ER71 acts downstream of BMP, Notch, and Wnt signaling in blood and vessel progenitor specification. Cell Stem Cell 2, 497–507 (2008).
33. Baralle, F. E. & Giudice, J. Alternative splicing as a regulator of development and tissue identity. Nat. Rev. Mol. Cell Biol. 18, 437–451 (2017).
34. Potter, R. F. & Groom, A. C. Capillary diameter and geometry in cardiac and skeletal muscle studied by means of corrosion casts. Microvasc. Res. 25, 68–84 (1983).
35. Schaum, N. et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature 526, 367–372 (2018).
36. Madhavan, M. et al. Induction of myelinating oligodendrocytes in human cortical spheroids. Nat. Methods 15, 700–706 (2018).
37. Marton, R. M. et al. Differentiation and maturation of oligodendrocytes in human three-dimensional neural cultures. Nat. Neurosci. 22, 484–491 (2019).
38. Garcia-Leon, J. A. et al. SOX10 single transcription factor-based fast and efficient generation of oligodendrocytes from human pluripotent stem cells. Stem Cell Rep. 10, 655–672 (2018).
39. Ehrlich, M. et al. Rapid and efficient generation of oligodendrocytes from human induced pluripotent stem cells using transcription factors.Proc. Natl Acad. Sci. USA 114, E2243–E2252 (2017).
40. Sarkar, A. & Hochfeldinger, K. The sox family of transcription factors: versatile regulators of stem and progenitor cell fate. Cell Stem Cell 12, 15–30 (2013).
41. Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R. & de Crombrugghe, B. Sox9 is required for cartilage formation. Nat. Genet. 22, 85–89 (1999).
42. Canals, I. et al. Rapid and efficient induction of functional astrocytes from human pluripotent stem cells. Nat. Methods 15, 693–696 (2018).
43. Khoshakhlagh, P., Sivakumar, A., Pace, L. A., Sazer, D. W. & Moore, M. J. Methods for fabrication and evaluation of a 3D microengineered model of myelinated peripheral nerve. J. Neural Eng. 15, 064001 (2018).
44. Khoshakhlagh, P. & Moore, M. J. Photoresponsive interpenetrating network of hyaluronic acid and PuraMatrix as a selectively tunable scaffold for neurite growth. Acta Biomater. 16, 23–34 (2015).
45. Mohammadi, S. et al. Whole-brain in-vivo measurements of the axonal G-ratio in a group of 37 healthy volunteers. Front. Neurosci. 9, 441 (2015).
46. Windrem, M. S. et al. Fetal and adult human oligodendrocyte progenitor cell isolates myelinate the congenitally dysmyelinated brain. Nat. Med. 10, 93–97 (2004).
47. Lancaster, M. A. et al. Cerebral organoids model human brain development and microcephaly. Nature 501, 373–379 (2013).
48. Togo, S. et al. Differentiation of embryonic stem cells into fibroblast-like cells in three-dimensional type I collagen gel cultures. In Vitro Cell. Dev. Biol. Anim. 47, 114–124 (2011).
49. Elcheva, I. et al. Direct induction of haematopoietical programs in human pluripotent stem cells by transcriptional regulators. Nat. Commun. 5, 4372 (2014).
50. Morita, R. et al. ETS transcription factor ETV2 directly converts human fibroblasts into functional endothelial cells. Proc. Natl Acad. Sci. USA 112, 160–165 (2015).
51. Cakir, B. et al. Engineering of human brain organoids with a functional vascular-like system. Nat. Methods 16, 1169–1175 (2019).
52. Woltjen, K. et al. piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature 458, 766–770 (2009).
53. Ronaldson-Bouchard, K. & Vunjak-Novakovic, G. Organs-on-a-Chip: a fast track for engineered human tissues in drug development. Cell Stem Cell 22, 310–324 (2018).
54. Guye, P. et al. Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6. Nat. Commun. 7, 10243 (2016).
55. Bagley, J. A., Reumann, D., Bian, S., Levi-Strauss, J. & Knoblich, J. A. Fused cerebral organoids model interactions between brain regions. Nat. Methods 14, 743–751 (2017).
56. Birey, F. et al. Assembly of functionally integrated human forebrain spheroids. Nature 545, 54–59 (2017).
57. Xi, X. et al. Fusion of regionally specified hPSC-derived organoids models human brain development and interneuron migration. Cell Stem Cell 21, 383–398 (2017).
58. Cederquist, G. Y. et al. Specification of positional identity in forebrain organoids. Nat. Biotechnol. 37, 436–444 (2019).
59. Mansouz, A. et al. An in vivo model of functional and vascularized human brain organoids. Nat. Biotechnol. 36, 432–441 (2018).
60. Rozenblatt-Rosen, O., Stubbington, M. J. T., Regev, A. & Teichmann, S. A. The Human Cell Atlas: from vision to reality. Nature 550, 451–453 (2017).
61. Han, X. et al. Construction of a human cell landscape at single-cell level. Nature 581, 303–309 (2020).
62. Cusanovich, D. A. et al. A single-cell atlas of in vivo mammalian chromatin accessibility. Cell 174, 1309–1324 (2018).
63. Moss, J. et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. Nat. Commun. 9, 5068 (2018).

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.
© The Author(s), under exclusive licence to Springer Nature America, Inc. 2020
Methods
Annotation of human TFs. A starting set of 1,591 TFs was used, based on their evidence codes “a” and “b” (confirmed experimental evidence), “c” (prediction only) and “other” (probable TFs with undefined DNA-binding domains). Additional TFs curated by the Human Genome Organization's HGNC Gene Nomenclature Committee (HGNC) were added: zinc fingers (including those containing C2H2 domains), homeodomains (including LIM, POU, TALE, HOXL, NKI and PRD sub-families) and basic helix-loop-helix and forkhead TFs. Pseudogenes, as annotated by HGNC or Ensembl, were removed. All genes were converted to gene names using the HGNC multi-synonym checker. The final set of TFs included in the Human TFome contained 1,564 genes (Supplementary Table 1).

Construction of the Human TFome expression library. Gateway-compatible ORFs for the target set of 1,564 genes were requested from the Vidal lab ORFcollection (http://orfcollection.dna.nus.edu.sg) and transOMIC.com. For the missing ORFs, Uniprot protein sequences of 273 genes were reverse-translated and codon-optimized for synthesis by Gen9. Where multiple isoforms exist, the one designated as “canonical” by Uniprot was selected. If the isoform was longer than 4 kb, the longest isoform below 4 kb was chosen owing to synthesis constraints. Synthesized genes were cloned into pDONR221 using BP Clonase II (Invitrogen, 11789020). All pDONR plasmids are available on Addgene, DNASU or transOMIC. TF Sources are summarized in Supplementary Fig. 1a. To determine the tissue expression of synthesized TFs, GTEx65 version 6 median tissue FPKMs were downloaded. For each TF, median expression across tissues was computed. TFs that were synthesized were compared to TFs that were cloned from cDNAs.

Pooled library cloning into lentiviral expression vector. All pDONR-TFs were pooled and cloned into pLIX_403, the doxycycline-inducible lentiviral vector (a gift from David Roth; Addgene plasmid no. 41395), using LR Clonase II (Invitrogen, 11789100), transfected into 293T cells containing pMD2G (Addgene, plasmid no. 12259), psPAX2 (Addgene, plasmid no. 12260) and pLIX_403-TF pools were prepared for next-generation sequencing using the NEBNext Ultra DNA Library Prep Kit for Illumina (New England Biolabs, E7370L). Libraries were quantified using the KAPA Real-Time Library Amplification Kit (Roche, KK2702) and loaded onto an Illumina MiSeq v3 150-cycle kit (MS-102-3001). About 97% of the TFs fit within the pLIX_403 vector cargo limit of approximately 4 kb (Supplementary Fig. 1e). The TF library was subcloned into the lentiviral expression vector with even representation and high coverage (Supplementary Fig. 1f–h).

Data processing and analysis for library cloning. To align reads to the TFome library, reference TF sequences were first indexed using the STAR aligner (Cell Ranger v2.2.2) according to the STAR command, counted using bash scripts and plotted in R.

Cell culture. The PGP1 hiPSC line without integrated Y amanaka factors was seeded with Y-27632 ROCK inhibitor (Millipore, 68801) for 1 d. Cells were selected with 1 μM puromycin (Gibco, A1113803). hiPSCs were passaged using TrypLE Express (Life Technologies, 12604013) and seeded with Y-27632 ROCK inhibitor (Millipore, 68801). Colony formation was performed on the CytoTune 2.0 Sendai Reprogramming Kit (ATCC, DYS0100), both using the CytoTune 2.0 Sendai Reprogramming Kit (ATCC, DYS0100) and pLIX_403-TFs into 293T cells. Media were exchanged after 24 h, and supernatants were harvested at 48 h and 72 h after transfection. The supernatants were filtered (0.45-μm PES filter, Corning 431220) and combined. Viral particles were resuspended with PBS in 1/100 of the supernatant volume (100 μl aliquots into 1.5-ml screw-cap tubes, snap-frozen on dry ice and stored as 50-μl aliquots) and stored at −80 °C. Viral particles were diluted 10× with PBS to achieve a final MOI of 1.0 for transduction. The lentiviral particles were resuspended with PBS in 1/100 of the supernatant volume (100 μl aliquots) and stored at −80 °C. Titrations were performed by quantitative polymerase chain reaction (PCR) as previously described. The titers ranged from 10^4 to 10^6 IU/ml. To transduce the entire library to achieve single-copy integrations, we used a large hiPSC population so that, on average, every TF was theoretically represented by more than 100 cells. Next, 750,000 PGP1, ATCC-DYS0100 and CRTD5 hiPSCs were each transduced with 75,000 lentiviral particles (MOI = 0.1) in mTeSR1 Technologies, C34853) at 1 × 10^7 cells per ml for 30 min on ice in the dark. For intracellular staining, cells were fixed using BD Cytofix fixation buffer (BD Biosciences, 554669) at 1 × 10^7 cells per ml for 20 min, washed with BD Perm/ Wash buffer (BD Biosciences, 554723) and permeabilized in BD Perm/Wash buffer for 10 min and then stained with fluorophore-conjugated antibodies and DAPI in the dark for 30 min. Stained cells were washed twice with FACS buffer, filtered into a strainer-capped tube (Falcon, 352235) and run on a BD LSRFortessa. FACS was performed on a BD FACS Aria or Beckman Coulter MoFlo Astrios. Spectral overlap was compensated using single fluorophore-conjugated ABC Total Antibody Compensation Beads (Life Technologies, A10497) with single fluorophore-conjugated antibodies. All antibodies are listed in the table below. Flow cytometry data were analyzed using FlowJo 10.2.

Antibodies used in this study.

| Name                                    | Vendor        | Catalog number |
|-----------------------------------------|---------------|----------------|
| PE anti-TRA-1-60                         | BD Biosciences| 560193         |
| PE anti-NANOG                            | BD Biosciences| 560483         |
| PerCP-Cy5.5 anti-Oct3/4                  | BD Biosciences| 560794         |
| Alexa 647 anti-SOX2                      | BD Biosciences| 560294         |
| BV421 anti-NCAM                          | BioLegend     | 318328         |
| APC anti-NCAM                            | BD Biosciences| 555518         |
| FITC anti-VE-cadherin                    | BD Biosciences| 560411         |
| Alexa 647 anti-VE-cadherin               | BD Biosciences| 561567         |
| Alexa 488 anti-vimentin                   | BD Biosciences| 562338         |
| APC anti-O4                              | R&D Systems   | F81326A        |
| Anti-NG2                                 | BD Biosciences| 562415         |
| Anti-MBP                                | Millipore     | MAB836         |
| Anti-MOG                                | Abcam         | ab15597        |
| Anti-vimentin                            | R&D Systems   | AF1205         |
| Anti-HP54                                | R&D Systems   | MAB89161-100   |
| Anti-ALCAM                               | R&D Systems   | AF1172-SP      |
| Anti-TUB83                               | Millipore     | A89354         |
| Anti-NF200                               | Sigma-Aldrich | N4142          |
| Anti-NeuN                                | Synaptic Systems | 266 004       |
| Anti-V5                                 | Thermo Fisher Scientific | R960-25     |
| Anti-VE-cadherin                         | Cell Signaling| 2500P          |
| Human-specific anti-CD31                 | Dako          | M0823          |
| Anti-α-smooth muscle actin               | Abcam         | ab65694        |
| Anti-MAP2                                | Abcam         | ab92434        |
| Alexa 488 anti-mouse IgG                 | Invitrogen    | A21202         |
| Alexa 647 anti-mouse IgG                 | Invitrogen    | A31571         |
| Alexa 568 anti-goat IgG                  | Invitrogen    | A11057         |
| Alexa 647 anti-goat IgG                  | Invitrogen    | A21447         |
| Alexa 488 anti-rabbit IgG                | Invitrogen    | A21208         |
| Alexa 555 anti-rabbit IgG                | Invitrogen    | A31572         |
| C3 anti-chicken IgG                      | Jackson Immunoresearch | 703-165-155   |
| Biotinylated anti-raf IgG                | Vector Laboratories | BA-9401     |

Lentiviral production and transduction. Lentiviral particles were produced as previously described as one pool containing the complete Human TFome library. Particles were produced by transfecting polyethylenimine (PolyScience, 24765), PMD2G (Addgene, plasmid no. 12259), pSPAX2 (Addgene, plasmid no. 12260) and pLIX_403-TFs into 293T cells. Media were exchanged after 24 h, and supernatants were harvested at 48 h and 72 h after transfection. The supernatants were filtered (0.45-mM PES filter, Corning 431220) and combined. Viral particles were then precipitated at 4 °C overnight using PEG solution (BioCat, K904-50-BV). The lentiviral particles were resuspended with PBS in 1/100 of the supernatant volume (100 μl aliquots) and stored at −80 °C. Titrations were performed by quantitative polymerase chain reaction (PCR) as previously described. The titers ranged from 10^4 to 10^6 IU/ml. To transduce the entire library to achieve single-copy integrations, we used a large hiPSC population so that, on average, every TF was theoretically represented by more than 100 cells. Next, 750,000 PGP1, ATCC-DYS0100 and CRTD5 hiPSCs were each transduced with 75,000 lentiviral particles (MOI = 0.1) in mTeSR1...
media (two six-well plates per pool). The low MOI ensures that most cells are not transduced (and are, thus, selected against by puromycin), and the cells that are transduced are most likely to receive only a single TF. Culture medium was exchanged daily. Forty-eight hours after transduction, 3 µg ml⁻¹ of puromycin was added to the medium to eliminate non-transduced PGPi cells. Cells were maintained and propagated as mentioned before.

**Tissue screen using FACS.** Each independently transduced hiPSC population was expanded into three 10-cm dishes, representing three replicates (Supplementary Fig. 2a). Each dish was seeded with 2 million cells in mTeSR1 with ROCK inhibitor, 0.5 µg ml⁻¹ of doxycycline and 1 µg ml⁻¹ of puromycin. The next day and daily thereafter until 4 dpi, the media were replenished with mTeSR1 containing doxycycline. As hiPSCs became confluent at 4 dpi, this day was chosen for dissociation and staining. Cells were not passaged for longer culture because of the faster proliferation rate of hiPSCs, which might out-compete differentiated cells. Cells were then dissociated using TrypLE Express and counted using an automated cell counter (Thermo Fisher Scientific, Countess II, AMQAX1000). Typically, 10 million live cells were harvested per 10-cm dish, stained for PE anti-TRA-1-60 and Calcein AM (Life Technologies, C1430) in mTeSR1 media at 10⁷ cells per ml for 30 min in the dark, washed and then filtered for debris using a single-cell strainer (Falcon, 352253). Cells were sorted on a BD FACSAria or a Beckman Coulter MoFlo Astrios. After gating Calcein AM-positive cells, cells in the bottom 10% and the middle 50% of TRA-1-60 expression were designated as TRA-1-60⁻/⁻ (dissociation gate) and TRA-1-60⁺/⁻ (stem cell gate), respectively. Approximately 0.8 million hiPSCs (per 7-cm dish) for 4 million TRA-1-60⁺/⁻ cells were collected for each replicate in the whole plate-screen. All cells were subsequently spun down into the cell pellet and the cell pellet was frozen at −20 °C.

**Genomic extraction, PCR and library preparation for sequencing.** Genomic DNA was extracted from sorted cells using the DNAeasy Blood & Tissue Kit (Qiagen, 69506) or on a QIAcube (Qiagen). More than 5 µg of genomic DNA was used for PCR and subsequently purified using the QIAquick PCR Purification Kit (Qiagen, 28106) on a QIAcube. Purified DNA was quantified using the Qubit dsDNA broad range quantitation kit (Thermo Fisher Scientific, Q32833). One microgram of DNA was sheared to an average of 200 bp on a Covaris sonicator E220 and used for library preparation using NEBNext Ultra DNA Library Prep (New England Biolabs, E7370). Samples were sequenced on an Illumina NextSeq 500 on high-output mode (Supplementary Table 2) with high coverage (Supplementary Fig. 1j).

**Analysis of sequencing data from Tissue screening.** Sequencing reads of amplified TFs were aligned to reference sequences using STAR aligner v2.5.2a. The ratio of cells in TRA-1-60 high versus low gates was used as an indicator of loss of pluriotyposity above basal spontaneous differentiation for each TF, as the TRA-1-60 high fraction served as an internal control. A log₂(TRA-1-60⁻/TRA-1-60⁺) score was computed using DESeq2 (ref. 2). A statistical threshold for defining TFs as hits was determined based on maximizing enrichment compared to random sampling. Hits were identified as those TFs scored by their log₂(TRA-1-60⁻/TRA-1-60⁺) score. Then, an enrichment score was computed for each candidate threshold. The enrichment score is computed by the number of TFs above that threshold in at least two cell lines minus the number of TFs above that threshold in a randomized dataset (n = 10,000 random samples). The enrichment scores for a set of candidate thresholds are plotted in Supplementary Fig. 2d. A threshold of 25% had the highest enrichment score.

**Literature search to determine novelty of hits.** To determine if a TF hit was novel for its ability to induce differentiation or forward programming upon overexpression, the name of each TF hit and the terms ‘differentiation’ or ‘programming’ or ‘overexpression’ were queried in PubMed on April 26, 2020. TFs were considered ‘known’ if at least one paper reported that the overexpression of ‘programming’ or ‘overexpression’ were queried in PubMed on April 26, 2020. TFs were considered ‘known’ if at least one paper reported that the overexpression of these terms. TF hits and mouse OPCs, only homologous genes between the two species annotated in the Mouse Genome Database were filtered and measured for gene expression using the identical pipeline for gene expression quantification. Gene expression profiles of ATOH1-induced neurons were compared to human brain samples from the ENCODE Project (ENCSDR239KFG for human cortex and enCSR239KFG for human brain). SOX9-induced fibroblasts were compared to newly formed oligodendrocytes, myelinating oligodendrocytes and oligodendrocyte precursor cells collected from mouse brain (GSF52564). NKX3-1-induced fibroblasts were compared to fibroblasts (GSF31211). ETV2-induced endothelial cells were compared to human umbilical cord endothelial cells (HUVs) (GSF31511). Gene counts were calculated by the R package gene expression differences EnvAnnotate9 (http://www.informatics.jax.org/downloads/reports/HOM_MouseHumanSequence.rpt) were retained. A DESeq2 (ref. 1) object was created providing the cell line information in the design formula. PCA was performed on vst transformed and SVA batch corrected TPM values using the precomprob function. Removal of hiPSCs before PCA did not affect the separation between differentiated and primary cells.

**Electrophysiology.** Assessment of neuronal function was performed as previously described in mice with minor modifications. Rat astrocytes (Gibco, N7745100) were seeded onto Matrigel-coated coverslips and cultured in DMEM with GlutaMAX, 10% FBS and N2 supplement on 35-mm dishes (17520484). One day before seeding induced neurons, astrocytes were cultured in Neurobasal media with GlutaMAX and B27 supplement (Gibco, 17504044). ATOH1 hiPSCs were labeled by lentiviral transduction of constitutively expressed green fluorescent protein (GFP) (FUGW, Addgene no. 14883). Next, 500,000 GFP-labeled ATOH1 hiPSCs were seeded in a six-well tissue culture plate and induced with doxycycline for 3 days on Matrigel-coated coverslips and then dissociated using TrypLE Express. Cells were counted, and 1 million induced neurons were plated in pre-conditioned media with astrocytes. Media were changed twice weekly. Electrophysiological recordings were carried out at 20–25 °C on an upright Olympus BX51WI microscope. Cells were bathed in 1% BSA/PBS and artificial cerebrospinal fluid containing (in mM) 119 NaCl, 2.5 KCl, 4 CaCl₂, 4 MgSO₄, 1 NaH₂PO₄, 26.2 NaHCO₃, and 11 glucose and saturated with 95% O₂/5% CO₂. Intracellular recordings were obtained using 3- to 5-MΩ glass micropipettes filled with an internal solution containing (in mM) 136 KMeSO₄, 17.8 HEPS, 0.6 MgCl₂, 1 EGTA, 4 Mg-ATP and 0.3 Na-GTP. Traces were collected using a Multichannel 700B amplifier (Molecular Devices), filtered with a 2-kHz Bessel filter, digitized at 10 kHz and stored on a DigData 1440 (Molecular Devices). Traces were analyzed using customized procedures in IGOR Pro (WaveMetrics). Cells were assessed for the

(Raw text continues...
presence of spontaneous action potentials in current-clamp mode. In the absence of spontaneous action potentials, cells were assayed by the injection of a set of 0.6-μs current steps, ranging from −1.40 to 400 pA in 60-pA increments.

Wound healing scratch assay. Wound healing scratch assays were performed as previously described3. NNX3-1 cells grown for 4 d in mTeSR1 either with or without doxycycline were scratched using a 200-μl tip. Scratches were imaged by a Nikon Eclipse Ti using a Nikon Plan Fluor 40× objective with a Zeiss AxioCam IC1 camera on days 0, 1 and 2 after scratching. Scratch areas were quantified in an unbiased and automated manner in Fiji using the Wound Healing plugin (http://dev.mri.cnrs.fr/attachments/download/1992/MRI_Wound_Healing_Tool.mlf). Five scratches were made per well, and the wound areas were averaged to generate one data point per well.

Collagen contraction assay. Collagen contraction assays were performed as previously described3. Briefly, NNX3-1 cells grown for 4 d in mTeSR1 either with or without doxycycline were dissociated and counted. Next, 400,000 cells in 300 μl of mTeSR1 was loaded onto a Chromium chip. Cells were grown for 4 d in mTeSR1 with or without doxycycline, and wells were filled up to 400 μl with RT buffer

The reverse transcriptase was heat inactivated at 70 °C for 15 min, and the cDNA (GE Healthcare) was resuspended in a buffer consisting of 10 mM Tris, 20 mM EDTA, 50 mM NaCl, and 0.1 mM dithiothreitol. cDNA concentration using 0.1 μl of a primer mix was added (5 mM dNTP (Invitrogen), 0.5 μM dT-primer, 0.5 μM UP-primer (AAGCAGTGGTATCAACGCAGAGT)). 9 U of RNase Inhibitor and 10 U of KAPA HiFi HotStart ReadyMix (Peqlab) at a final 1× of fibrinogen (Sigma-Aldrich, cat. no. F8630) and 50 μl of bovine collagen I (Trevigen, cat. no. 3442-050-2) on the neighbor graph computed on the first ten principal components. Spike-in control NNX3-2 cells were removed based on clustering for further analysis and visualization. For the SOX9 dataset, the two datasets arising from different sequencing runs were concatenated. Low-quality samples were filtered if they had fewer than 200 genes expressed, fewer than 50,000 read counts or more than 0.1% of reads mapped to mitochondrial genes. Clusters were computed by the Leiden algorithm (resolution = 0.2) on the neighbor graph computed on the first seven principal components.

Angiogenesis assay and TEM. For endothelial angiogenesis assays, ETV2 isoform 2 hiPSCs were induced to differentiate with 100 μg/ml of Matrigel and 2 μl of doxycycline solution and eluted in 12 μl of fibrinogen-free whole blood on a Reichert Ultracut-S microtome, ultrathin sections (~80 nm) were cut. Sections were then washed in 0.1 M cacodylate buffer and post-fixed with 1% osmium tetroxide solution, incubated in 1% OsO4 in 0.2 M Tris HCl pH 7.4 for 30 min, and dehydrated in a graded series of ethanol: 50%, 70%, 90% and once in 100% (10 min each) ethanol. Samples were then placed in propylene oxide for 1 h and infiltrated overnight in a 1:1 mixture of propyleneoxide and TAAB Epon (Marivac Canada). The next day, samples were embedded in TAAB Epon and polymerized at 60 °C for 24 h. Ultrathin sections (~80 nm) were cut on a Reichert Ultracet-S microtome, picked up onto copper grids stained with lead citrate and examined using a Tecnai 1200EX transmission electron microscope. Images were recorded with an AMT 2k CCD camera.

In vivo vascular network-forming assay. Six-week-old nude mice were purchased from Jackson Laboratories and housed in compliance with Boyce Children’s Hospital animal care guidelines. All animal and related protocols were approved by the Institutional Animal Care and Use Committee. ETV2 cells for implantation were expanded in 7 d in vitro after differentiation and mixed with human mesenchymal stromal cells at a 1:1 ratio for a total of 2 × 10⁶ cells per mouse in 200 μl of pH-neutral pre-gel solution containing 3 mg/ml of bovine collagen I ( Trevigen, cat. no. 3442-050-01), 2 mg/ml of fibrinogen (Sigma-Aldrich, cat. no. F8630) and 50 μl of Matrigel (Corning, cat. no. 354234). After carboxyrimidate, mice were injected first with 50 μl of 10 μl/mouse thrombin (Sigma-Aldrich, cat. no. T4648) and then with 200 μl of cell-laden pre-gel solution into the same site. All experiments were performed using a Tecan Infinite M Nano microplate reader (Miltonyi Biotec) in duplicates. Cells were counted using the Smart-seq2 workflow, as

Only those with a transcript support level of 1 were retained. Annotations with identical consensus coding DNA sequences were collapsed into one isoform. Mark genes were computed by Student's t-test between each cluster and all the other cells.

Analysis of scRNA-seq data. Analysis of single-cell expression matrices was performed on Python Notebooks using custom functions and the package Scapy in v.1.4.3 (ref. 80). Raw matrices were processed with filtering steps to obtain cells with high-quality reads (see below). Expression values were then normalized to 10 (counts for Smart-seq2 data) and log-transformed. Highly variable genes were then selected for downstream visualization and clustering analysis. Biological cell-to-cell variation was linearly regressed out based on counts per cell and percentage of mitochondrial genes, and count values were scaled to a maximum value of 10.

After dimensionality reduction, processed matrices were used for cluster analysis and visualization. Mark genes were computed by Student's t-test between each cluster and all the other cells.

For ETV2 and SOX9 datasets, the 10x Chromium pipeline was used. Reads were mapped to the human genome (GRCh38), and count matrices were generated using CellRanger v2.0.0. For the ETV2 dataset, low-quality cells were filtered out if they had fewer than 200 genes expressed, fewer than 50,000 read counts or more than 0.1% of reads mapped to mitochondrial genes. Clusters were computed by the Leiden algorithm (resolution = 0.2) on the neighbor graph computed on the first ten principal components.

For the parallel programming dataset of ATOH1, ETV2 and NNX3-1 cells co-cultured in two dimensions, the data were analyzed using a pipeline developed in house. Briefly, reads were quality checked by FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and mapped using STAR v2.5.4a to the Ensembl GRCh38v94 human genome reference sequence69. The Picard v2.9.0 tool CollectRNASeqMetrics was used to inspect the abundance of ribosomal RNA transcripts (https://broadinstitute.github.io/picard/). The expression index of cells by genes was then computed using featureCount67 for genes annotated on the Ensembl GRCh38 v94 genome (ftp://ftp.ensembl.org/pub/release-94/gtf/homo_sapiens/Homo_sapiens.GRCh38.94.gtf.gz). Low-quality cells were filtered out if they had fewer than 2000 genes expressed, fewer than 50,000 counts or more than 0.3% of reads mapped to mitochondrial genes. Cell clusters were computed with the Leiden algorithm (resolution = 0.2) on the neighbor graph computed on the first seven principal components.
were carried out in five mice, and explants were harvested after 1 week. Human mesenchymal stromal cells (hMSCs) were isolated from the white adipose tissue and cultured on uncoated plates using MSCGM (Lonza, cat. no. PT-3001) supplemented with 10% Fencode FBS (Genesi, cat. no. 25-514) and 1x penicillin–streptomycin–glutamine (Thermo Fisher Scientific, cat. no. 10378106). All experiments were carried out using passage 6–10 hMSCs.

Expiayed grafts were fixed overnight in 10% buffered formalin, embedded in paraffin, and sectioned at a thickness of 7 μm. Microvessel density (vessels per mm²) was defined as the average number of erythrocyte-filled vessels in hematoylin and eosin (H&E)-stained sections collected from the middle of the implants, as previously described10,11. For immunostainings, sections were deparaffinized, and antigen retrieval was carried out for 30 min with boiling citric buffer (6 mM sodium citrate and 0.05% Tween 20, pH 6.0). Tissues were then blocked for 30 min in 5% horse serum and incubated with primary antibodies for 30 min at room temperature. Fluorescent staining was performed using fluorescein-conjugated secondary antibodies, followed by DAPI counterstaining for 30 min at room temperature. Human-specific anti-CD31 (Dako, M0823) was used as a vascular marker. The slides were washed with PBS and incubated with 1:250 biotinylated goat anti-rat antibody. The next day, the slides were blocked with 1% normal donkey serum. Sections were then incubated with primary antibodies in 1% BSA in PBS overnight at 4 °C. The next day, the slides were washed with PBS and incubated with 1:400 rabbit anti-MBP antibodies diluted in 1% BSA in PBS for 1 h at room temperature. Coverslips were then mounted on microscopy slides using AquaMount (VWR, 41799-008) and stored overnight in the dark at room temperature. Thereafter, slides were imaged with an AxioObserver.Z1 microscope with motorized stage (Carl Zeiss Microscopy) using x20 and x40 air objectives. Images were stored in the Live imaging software’s缀制版d and exported to the *.jpg format using ZEN blue software (Carl Zeiss Microscopy).

For flow cytometry and single-cell sorting, ATOH1, ETV2 isoform 2 and NKX3-1 hiPSCs were mixed and seeded in all combinations in mTeSR1 and Y-27632 ROCK inhibitor with or without 0.5 μg ml⁻¹ of doxycycline on Matrigel for 1 week, 30,000 cells were seeded in wells for non-induced controls and 90,000 in wells for doxycycline induction. mTeSR1 with or without doxycycline was replenished daily. After 4 d, cells were dissociated with TrypLE Express, resuspended with 1 ml of PBS per well, transferred to strainer-capped tubes for flow cytometry and spun down for 4 min at 1,400 rpm. Next, the supernatant was aspirated, the stainant with fluorescein-conjugated antibodies against NCAM (1:20), vimentin (1:20) and VE-cadherin (1:20) diluted in 100 μl of 0.2-μm-filtered FACS buffer (1% BSA albumin fraction V, Roth, 8076.2) and 2 mM EDTA, in sterile PBS pH 7.4) for 1 h at 4 °C in the dark. Cells were spun down for 4 min at 1,400 rpm, washed with FACS buffer once, spun down again and re-suspended in 250 μl of FACS buffer. Tubes were kept on ice until analysis within the next hour on a BD FACS Aria II. Additionally, 95 single cells were sorted in a 96-well plate with a BD FACS Aria III (100-μm nozzle) and sent for single-cell sequencing at the CMDB Deep Sequencing Facility.

Parallel cell programming in embryoid bodies. Twenty-four-well AggreWell plates (STEMCELL Technologies, 34450) were prepared following the manufacturer’s protocol. ATOH1, ETV2 isoform 2 and NKX3-1 hiPSCs were seeded in equal ratios in AggreWell EB Formation Medium (STEMCELL Technologies, 05893) with Y-27632 ROCK inhibitor (Millipore, 6880001). A total of 6 × 10⁴ cells were used. Fifty percent AggreWell EB Formation Medium was changed daily. After 4 d, the EBs were harvested and transferred to ultra-low attachment wells (Sigma-Aldrich, LCS3471-24E) using AggreWell EB Formation Medium supplemented with 3 μg ml⁻¹ of puromycin (Life Technologies, A11138-03). TF expression was induced with doxycycline 7 d after seeding.

For flow cytometry analysis, ESs were dissociated using TrypLE Express, washed and resuspended in PBS with 10% FBS. Stained with anti-α-fetoprotein (Sigma-Aldrich, A21181) and anti-CD11b (BD Biosciences, 560802), live cells were stained with fluorescein-conjugated antibodies at 1 × 10⁵ cells per ml for 30 min on ice in the dark. Stained cells were washed twice with FACS buffer, filtered into a strainer-capped tube (Falcon, 352235) and analyzed on a BD LSRII. Flow cytometry data were analyzed using FlowJo 10.2.

For immunofluorescence microscopy, ESs were fixed with 4% PFA (Electron Microscopy Sciences, 15714-S) in PBS for 30 min at room temperature and then washed three times with PBS and kept at 4 °C in PBS. Next, fixed samples were incubated with blocking solution (10% NDS, 1% BSA and 0.5% Triton X-100 in PBS) for 1 h at room temperature. For staining, a buffer similar to the blocking buffer was used but with reduced NDS (3%). Samples were incubated with primary antibodies at room temperature for 1 h, subsequently washed with staining buffer twice, incubated with secondary antibodies and 1 mg ml⁻¹ of DAPI (Roche, 1032672001) at room temperature for 1 h, washed once with staining buffer and then washed twice with PBS. ESs were finally mounted in 1% a-garose on MatTek dishes and immersed in PBS. Stained ESs were imaged on a Zeiss LSM780 upright microscope equipped with a Zeiss Achroplan 40/0.8 water-lensed objective. To maximize signal-to-noise ratios, detector gains were kept between 600 and 900, and laser powers were kept between 1% and 12%. Z-stacks were generated and stored in the manufacturer’s format (.czi). Maximum intensity projections were exported in ZEN blue software.

Orthogonal cell programming in cerebral organoids. Cerebral organoids were generated as previously described12 with minor modifications. To orthogonally program induced oligodendrocytes within cerebral organoids, inducible SOX9 hiPSCs and unmodified hiPSCs were dissociated with TrypLE Express (Life Technologies, 12604013), counted using an automated cell counter (Countess II, AMQAX100, Thermo Fisher Scientific) and mixed at a ratio of 1:1 in AggreWell medium (STEMCELL Technologies, 05893). Next, the single-cell suspension was transferred to AggreWell400 plates (STEMCELL Technologies, 27945) for embryoid body formation. Then, 600,000 cells were seeded into an AggreWell plate containing AggreWell medium with 10 ng ml⁻¹ Y-27632 ROCK inhibitor (Millipore, 6880001). Embryoid bodies were cultured in 96-well plates for 1 week at 37 °C (5% CO₂) and placed in a tissue culture incubator overnight. The next day (day 1 of the protocol), embryoid body culture formation was verified by bright-field microscopy, and the media were changed to neural induction media (DMEM/F12, HEPEs and GlutaMAX (Invitrogen, 11330-032) with N2 supplement (Gibco, A15077-01) and non-essential amino acids (Gibco, 11140-030)). Half of the media was changed daily with neural induction media from days 1–4, 100 μl of N2 in PBS was added to the dishes by pipetting gently with a wide-bore tip to dislodge them from the AggreWell and individually embedded in droplets of undiluted Matrigel (Corning, 354277). To
induce TF expression for orthogonal programming. 0.5 μg ml⁻¹ of doxycycline was added daily into the media starting on day 4. On day 8, media were changed to nervous differentiation media consisting of 1:1 DMEM/F12 containing HEPES and GlutaMAX (Invitrogen, 11330-032) and Neurobasal medium (Invitrogen, 12348-017) with non-essential amino acids (Gibco, 11140-050), N2 supplement (Gibco, A13707-01) and B27 supplement without vitamin A (Gibco, 12587-010). Media were replaced every other day. Organoids were harvested and sliced onto charged glass slides and stored at −20 °C until use. For staining, samples were brought to room temperature and outlined with a wax pen. They were washed three times with MAXWash Washing Medium (ActiveMotif, 15254) to remove any remaining OCT and then blocked using MAXBLOCK Blocking Medium (ActiveMotif, 15252) for 1 h and then washed with MAXWash Washing Medium. Primary antibodies in binding buffer (ActiveMotif, 15251) were added and allowed to stain overnight. Samples were washed three times with wash buffer and then stained for 5 h with secondary antibodies in binding buffer. Samples were washed with wash buffer, stained with DAPI and then mounted for imaging using VECTASHIELD mounting media.

Statistics. All statistics, including statistical tests, sample sizes and types of replicates, are described in the figure legends. Exact P values are listed in Supplementary Table 7. P values less than 0.05 were considered significant.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability. Next-generation sequencing data that support the findings of the study are available in the Gene Expression Omnibus using accession code GSE159786.

Code availability
The code that supports the findings of this study is available from the corresponding authors upon reasonable request.

References
64. Gray, K. A., Yates, B., Seal, R. L., Wright, M. W. & Bruford, E. A. Genenames. org; the HGNC resources in 2015. Nucleic Acids Res. 43, D1079–D1085 (2015).
65. Mele, M. et al. Human genomics. The human transcriptome across tissues and individuals. Science 348, 660–665 (2015).
66. Church, G. M. The personal genome project. Mol. Syst. Biol. 1, 2005.0305 (2005).
67. Kutsche, L. K. et al. Combined experimental and system-level analyses reveal the complex regulatory network of miR-124 during human neurogenesis. Cell Syst. 7, 438–452 (2018).
68. Salmon, P. & Trono, D. Production and titration of lentiviral vectors. in Current Protocols in Human Genetics Ch. 12, Unit 12.10 (Wiley, 2007).
69. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21 (2013).
70. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 530 (2014).
71. Zhang, Y. et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. 34, 11929–11947 (2014).
72. Zhang, J. et al. A genome-wide analysis of human pluripotent stem cell-derived endothelial cells in 2D or 3D culture. Stem Cell Rep. 8, 907–918 (2017).
73. Li, B. & Dewey, C. N. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinf. 12, 323 (2011).
74. Bult, C. J., Blake, J. A., Smith, C. L., Kadin, J. A. & Richardson, J. E. Mouse genome database (MGD). 2019. Nucleic Acids Res. 47, D801–D806 (2019).
75. Anders, S. & Huber, W. Differential expression analysis for sequence count data. Genome Biol. 11, R106 (2010).
76. Leek, J. T. sva: removing batch effects and other unwanted noise from high-throughput sequencing data. Nucleic Acids Res. 42, c161 (2014).
77. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012).
78. Ngo, P., Ramalingam, P., Phillips, J. A. & Furuta, G. T. Collagen gel contraction assay. Methods Mol. Biol. 541, 103–109 (2006).
79. Picelli, S. et al. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. Nat. Methods 10, 1096–1098 (2013).
80. Wolf, F. A., Angerer, P. & Theis, F. J. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 19, 15 (2018).
81. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics 30, 923–929 (2014).
82. Koike, N. et al. Tissue engineering: creation of long-lasting blood vessels. Nature Rev. Mol. Biol. 128, 138–159 (2014).
83. Melerro-Martin, J. M. et al. Engineering robust and functional vascular networks in vivo with human adult and cord blood-derived progenitor cells. Circ. Res. 103, 194–202 (2008).
84. Khoshakhlagh, P. et al. Development and characterization of a bioglass/chitosan composite as an injectable bone substitute. Carbohydrate Polym. 157, 1261–1271 (2017).
85. Khoshakhlagh, P., Bowser, D. A., Brown, J. Q. & Moore, M. J. Comparison of visible and UVA phototoxicity in neural culture systems micropatterned with digital projection photolithography. J. Biomed. Mater. Res. A 107, 134–144 (2019).
86. Douvaras, P. & Fossati, V. Generation and isolation of oligodendrocyte progenitor cells from human pluripotent stem cells. Nat. Protoc. 10, 1143–1154 (2015).

Acknowledgements
We thank J. Aach, M. O. Karl, R. Kalhor, N. Ostrov and H. Lee for critical feedback and the Church and Busskamp laboratories for support. We acknowledge technical support from the Harvard Biopolymers Facility, the Harvard Division of Immunology Flow Cytometry Core Facility, the Beth Israel Deaconess Medical Center Flow Cytometry Core, the Wyss Flow Cytometry and Microscopy Core, M. Ericsson and P. Coughlin at the Harvard Medical School Electron Microscopy Facility, M. T. Gianatasio at the Dana-Farber/Harvard Cancer Center Specialized Histopathology Core and Rodent Histopathology Core (both supported, in part, by National Cancer Institute Cancer Center Support grant NIH S P30 CA06516) and Harvard Medical School Orchestra Research Computing. We also thank the TU Dresden Center for Molecular and Cellular Bioengineering Advanced Imaging, Deep Sequencing, Flow Cytometry and Stem Cell Engineering core facilities. We would also like to thank J. Gray’s laboratory for electrophysiology support, S. Jeanty and J. Lee (Church lab, Harvard Medical School) for the PGPI Sendai virus hiPSC line, G. Shemyakin and W. Glindmeyer for helpful discussions, A. Jolma, K. Nitta and K. Said for technical assistance and M. Lenieux and J. McDade for their support in depositing the library to Addgene. A.H.M.N. was supported by an NSERC Postgraduate Fellowship and a Peter and Carolyn Lynch Foundation Fellowship. I.E.R.A. was supported by the DIGS-BB program. S.L.S. is a Shurl and Kay Curci Foundation Fellow of the Life Sciences Research Foundation. The Ellison Foundation and Institute Sponsored Research funds from the DFCL Strategic Initiative supported M.V. and D.E.H. The project was supported by the Volkswagen Foundation (Freigeist - A110720), the European Research Council (ERC-SG-678071 - ProNeuroms) and the Deutsche Forschungsgemeinschaft (SPP2127, EXC-2068-390729961 - Cluster of Excellence – Physics of Life at TU Dresden and EXC-2151-390763408 – Cluster of Excellence – ImmunoSensation” at the University of Bonn) to V.B. G.M.C. acknowledges funding from National Human Genome Research Institute grants P50 HG005550 'Center for Causal Variation', RM1 HG008525 'Center for Genomically Engineered Organz', the Simons Foundation for Autism Research Initiative (368485), the Blavatnik Biomedical Accelerator at Harvard University, the FunGCAT program from the Office of the Director of National Intelligence Intelligence Advanced Research Projects Activity, via the Army Research Office, under federal award no. W911NF-17-2-0089 and research funding from R. Merkin and the Merkin Family Foundation.

Author contributions
A.H.M.N., P.K., V.B. and G.M.C. conceived the idea, led the study and designed all experiments. A.H.M.N. and P.K. performed most of the experiments and analyses, with significant technical contributions from J.E.R.A., G.P., K.W., A.S., E.A., E.A., K.K., R.E.K., A.V., M.D., K.L., W.S., J.Y.H., A.G., J.T., D.E.H., M.V. and J.M.M. - M.V. and G.M.C. oversaw the study. A.H.M.N., P.K. and V.B. wrote the manuscript with input and feedback from all authors.

Competing interests
A.H.M.N., P.K. and G.M.C. are inventors on patents filed by the Presidents and Fellows of Harvard College. Full disclosure for G.M.C. is available at https://arp.med.harvard.edu/gmc/tech.html. A.H.M.N., P.K. and G.M.C. are co-founders of and have equity in GC Therapeutics, Inc. No reagents or funding from GC Therapeutics were used in this study.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41587-020-0742-6. Correspondence and requests for materials should be addressed to V.B. or G.M.C.

Reprints and permissions information is available at www.nature.com/reprints.

References
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

- n/a  Confirmed
- □ The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- □ □ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- □ □ The statistical test(s) used AND whether they are one- or two-sided
- □ Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- □ □ A description of all covariates tested
- □ □ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- □ □ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- □ □ For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and P value noted
  Give P values as exact values whenever suitable.
- □ □ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- □ □ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- □ □ Estimates of effect sizes (e.g. Cohen’s d, Pearson’s r), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

| Data collection | Illumina RTA 1.18.54, BD FACSDiva, ZEN 2012 (Blue Edition), QuantaSoft, LightCycler 480 Software |
|-----------------|-------------------------------------------------------------------------------------------------|
| Data analysis   | STAR v2.5.2a, R v3.2.2, DESeq2 v1.8.2, Fiji v1.0, Adobe Illustrator CS6 v16.0.0, FlowJo 10.4, Microsoft Excel, WaveMetrics Igor Pro, Python |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:
- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Next-generation sequencing data will be uploaded to Gene Expression Omnibus (GEO).
Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences  
- Behavioural & social sciences  
- Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

- Sample size: No statistical methods were used to predetermine the sample size.
- Data exclusions: No data was excluded.
- Replication: All findings were reproducible.
- Randomization: Samples were not randomized.
- Blinding: Investigators were not blinded to experiments.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

| n/a | Involved in the study |
|-----|-----------------------|
|     | Antibodies            |
|     | Eukaryotic cell lines |
|     | Palaeontology and archaeology |
|     | Animals and other organisms |
|     | Human research participants |
|     | Clinical data         |
|     | Dual use research of concern |

Methods

| n/a | Involved in the study |
|-----|-----------------------|
|     | ChIP-seq              |
|     | Flow cytometry        |
|     | MRI-based neuroimaging |

Antibodies

This list is also included in the supplementary materials.

- PE anti-TRA 1-60 BD Bioscience 560193
- PE anti-Human NANOQ BD Bioscience 560483
- PerCP-Cy5.5 anti-Oct3/4 BD Bioscience 560794
- Alexa Fluor 647 anti-Sox2 BD Bioscience 560294
- APC Anti-CD56/NCAM BD Bioscience 555518
- FITC Anti-CD144/VE-Cadherin BD Bioscience 560411
- Alexa Fluor 488 anti-Vimentin BD Bioscience 562338
- Anti-Vimentin R&D Systems AF2105
- Anti-HSP47/SERPINE1 R&D Systems MAB91661-100
- Anti-ALCAM/CD166 R&D Systems A1172-5P
- Anti-Beta-III-Tubulin/TUBB3 Millipore AB9354
- Anti-NF-200 Sigma N4142
- Anti-VE Cadherin Cell Signalling Technologies 2500P
- Alexa Fluor488 donkey anti-mouse IgG (H+L) Life Technologies A21202
- Alexa Fluor647 donkey anti-mouse IgG (H+L) Life Technologies A31571
- Alexa Fluor568 donkey anti-goat IgG (H+L) Life Technologies A11057
- Alexa Fluor647 donkey anti-goat IgG (H+L) Life Technologies A21447
- Alexa Fluor488 donkey anti-rabbit IgG (H+L) Life Technologies A21208
- Alexa Fluor555 donkey anti-rabbit IgG (H+L) Life Technologies A31572
- Donkey anti-chicken-Cy3 Jackson Labs Jackson ImmunoResearch 703-165-155

Validation

We used antibodies with validation data from the vendor’s website.
Eukaryotic cell lines

Policy information about **cell lines**

**Cell line source(s)**
The non-integrated PGP1 hiPSC line was generated from adult dermal fibroblasts (Coriell, GM23248) and the CRT05 hiPSC line was generated from foreskin fibroblasts (ATCC CRL-2522) using the CytoTune 2.0 Sendai Reprogramming Kit (Invitrogen, A16517). The ATCC hiPSC (ACS-1019 DYS0100) hiPSC was obtained from ATCC.

**Authentication**
Cell lines were verified by short tandem repeat (STR) profiling (Dana Farber Cancer Institute).

**Mycoplasma contamination**
Cells were tested negative for mycoplasma using the Universal Mycoplasma Detection Kit (ATCC, 30-1012K).

**Commonly misidentified lines**
(See [ICLC register](#))
No commonly misidentified lines were used.

Animals and other organisms

Policy information about **studies involving animals**, [ARRIVE guidelines](#) recommended for reporting animal research

**Laboratory animals**
Six-week-old NOD/SCID mice and Shiverer mice were purchased from Jackson Laboratories.

**Wild animals**
The study did not involve wild animals.

**Field-collected samples**
This study did not involve samples collected from the field.

**Ethics oversight**
Animal study protocols were approved by Harvard Medical School or Boston Children’s Hospital’s Institutional Animal Care and Use Committee.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Flow Cytometry

**Plots**
- Confirm that:
  - The axis labels state the marker and fluorochrome used (e.g., CD4-FITC).
  - The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a ‘group’ is an analysis of identical markers).
  - All plots are contour plots with outliers or pseudocolor plots.
  - A numerical value for number of cells or percentage (with statistics) is provided.

**Methodology**

**Sample preparation**
Cells were dissociated using TrypLE Expression, washed, and resuspended in FACS buffer (PBS with 10% FBS). For surface antigens, live cells were stained with fluorochrome-conjugated antibodies and the viability dye CellTrace Calcein Blue, AM (Life Technologies, C34853) at 1×10^6 cells/ml for 30 minutes on ice in the dark. This viability dye is particularly important when isolating TRA-1-60(low) populations, as loss of TRA-1-60 signal could be the result of dying cell debris. 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) prior to running on flow cytometer or FACS machine.

**Instrument**
Samples for were analyzed on a BD LSRFortessa, and sorted on a BD FACSaria or Beckman Coulter MoFlo Astrios.

**Software**
BD FACSDiva and BC Intellisort software was used during data collection and Flowjo was used for post-collection data analysis.

**Cell population abundance**
“Single cell” purity mode was used to ensure high purity post-sort populations.

**Gating strategy**
Unstained cells and compensation beads were used to inform the gating strategy.

Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary information.