Differentiation of Human Cells and Tissues Using a Comprehensive Human Transcription Factor Library

Citation
Ng, Hon Man Alex. 2018. Differentiation of Human Cells and Tissues Using a Comprehensive Human Transcription Factor Library. Doctoral dissertation, Harvard University, Graduate School of Arts & Sciences.

Permanent link
http://nrs.harvard.edu/urn-3:HUL.InstRepos:41129209

Terms of Use
This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA

Share Your Story
The Harvard community has made this article openly available. Please share how this access benefits you. Submit a story.

Accessibility
Differentiation of Human Cells and Tissues using a
Comprehensive Human Transcription Factor Library

A dissertation presented □
by □

Hon Man Alex Ng □

to □

the Committee on Higher Degrees in Systems Biology □
in partial fulfillment of the requirements □
for the degree of □
Doctor of Philosophy □
in the subject of □
Systems Biology □

Harvard University □
Cambridge, Massachusetts □
April 2018 □
Abstract

- Functional cells and tissues differentiated from human pluripotent stem cells (hPSCs) have potential applications for disease modeling, drug discovery and regenerative therapy. Current differentiation methods, however, remain complex and time-consuming. For the purpose of rapid generation of pure population of cells, protocols typically result in heterogeneous mixtures with low yields. In the case of organoid tissue differentiation, key cell types are missing. Recently, transcription factor (TF)-mediated conversion of hPSCs into differentiated cells and tissues have improved these differentiation methods.

- First, in order to systematically explore the capacity of TF-mediated cell conversion, we created the human TFome expression library, the first comprehensive human TF library of over 1,700 open reading frames (ORFs) representing 1,576 TFs. We screened the human TFome to identify TFs that could individually induce loss of pluripotency in hiPSCs. We found 243 hits that could differentiate multiple hiPSC lines, even in pluripotency-reinforcing conditions after only four days. This suggests a widespread ability for individual TFs to induce differentiation. Twenty-five top ranking hits were validated and preliminary characterization suggests they induce diverse cell fates. Lastly, we present two novel cell conversions from hiPSCs into
functional neurons and stromal fibroblast cells with nearly complete conversion efficiency. Together, these results illustrate the power of the human TFome library to produce many differentiated cell types with high efficiency and speed.

We then applied TF-mediated cell conversion to improve organoid tissue differentiation. The interior of many organoids are necrotic due to a lack of vasculature. Previous attempts to vascularize cerebral organoids with human umbilical vein endothelial cells (HUVECs) resulted in their spontaneous separation away from developing organoids. Here, we devised a TF-mediated strategy where alternative fates are induced in the same media conditions. We identified a splice isoform of ETV2 that potently converts hiPSCs into functional endothelial cells even in neural-inducing conditions. This enabled simultaneous production of TF-driven endothelial cells alongside media-driven neural cells and led to the vascularization of cerebral organoids. This hybrid differentiation approach could be generalized for the incorporation of missing cell types in other organoids and tissues.
# Table of Contents

Abstract

Table of Contents

List of Figures

Acknowledgements

Chapter 1: Introduction

1.1 Human induced pluripotent stem cells

1.2 Culture and Differentiation of hPSCs

1.3 Transcription factors for cell conversion

1.4 Challenges in selecting transcription factors for cell type conversion

1.5 Organoids

1.6 Major questions addressed in this dissertation

Chapter 2: Differentiating diverse human cell types using a comprehensive transcription factor library

2.1 Chapter summary

2.2 Introduction

2.3 Results

2.4 Discussion

2.5 Methods

Chapter 3: Vascularization of cerebral organoids using transcription factor-mediated orthogonal differentiation

3.1 Chapter summary

3.2 Introduction

3.3 Results

3.4 Discussion

3.5 Methods

Chapter 4: Discussion and future directions

References
# List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1.</td>
<td>Three broad approaches to differentiate stem cells.</td>
</tr>
<tr>
<td>Figure 2.1.</td>
<td>Construction of the human TFome library.</td>
</tr>
<tr>
<td>Figure 2.2.</td>
<td>Loss of pluripotency marker TRA-1-60 as a cell-agnostic readout.</td>
</tr>
<tr>
<td>Figure 2.3.</td>
<td>A pooled FACS-based screen reveals 243 hits for hiPSC differentiation.</td>
</tr>
<tr>
<td>Figure 2.4.</td>
<td>PiggyBac transposons enable rapid engineering of inducible hiPSC lines.</td>
</tr>
<tr>
<td>Figure 2.5.</td>
<td>Fifteen hits are validated to induce differentiation with diverse morphologies.</td>
</tr>
<tr>
<td>Figure 2.6.</td>
<td>Transcriptomic characterization of eleven differentiation-inducing hits.</td>
</tr>
<tr>
<td>Figure 2.7.</td>
<td>NKX3-2 induces stromal fibroblast-like cells.</td>
</tr>
<tr>
<td>Figure 2.8.</td>
<td>NEUROG3 induces neuronal, not endocrine, differentiation.</td>
</tr>
<tr>
<td>Figure 3.1.</td>
<td>Orthogonal differentiation strategy for vascularizing cerebral organoids.</td>
</tr>
<tr>
<td>Figure 3.2.</td>
<td>A splice isoform of ETV2 induces potent endothelial differentiation.</td>
</tr>
<tr>
<td>Figure 3.3.</td>
<td>Highly efficient endothelial differentiation in neural-inducing conditions.</td>
</tr>
<tr>
<td>Figure 3.4.</td>
<td>Endothelial cell phenotype is stable after doxycycline withdrawal.</td>
</tr>
<tr>
<td>Figure 3.5.</td>
<td>ETV2 isoform t has transcriptomic signature of endothelial cells.</td>
</tr>
<tr>
<td>Figure 3.6.</td>
<td>Differentiated endothelial cells are functional.</td>
</tr>
<tr>
<td>Figure 3.7.</td>
<td>Vascularization of cerebral organoids.</td>
</tr>
</tbody>
</table>
Acknowledgements

First and foremost, I would like to thank my advisor, Professor George Church. He has inspired, encouraged and supported me to think ambitiously and creatively. He has cultivated a lab environment where it feels like you can achieve anything you set your mind to. It has been a privilege to work under his mentorship.

Dr. Volker Busskamp took me under his wing when I was a rotation student and continued after I joined the lab. He taught me everything with the utmost precision and detail, and encouraged me to be brave in the face of the unknown. After he left to start his own lab, he has been one of my close mentors and collaborators throughout graduate school. I’d also like to thank Dr. Anka Kempe from his lab who offered invaluable technical assistance.

Dr. Parastoo Khoshakhlagh and I worked closely during my last two years in the lab. Where I helped her with genetics and molecular biology, my efforts were multiplied and returned to me in the form of tissue engineering expertise and entrepreneurial advice. She has also been an amazing colleague and friend. Dr. Evan Appleton, Songlei Liu and Jeremy Huang also joined the TFome team. It has been great having people who share similar goals to discuss and collaborate with. Dr. John Aach has provided sagely advice throughout the years. Dr. Mark Skylar-Scott has been a great collaborator on the organoid work and I have learned much from him as well. I’m also humbled to be entrusted with mentoring and working with talented research assistants: Matthew Dysart, Kathleen Leeper, Chrystal Mavros, Wren Saylor and Renata Martin.

The brilliant scientists of 233 have been a constant source of support both scientifically and socially. Thanks Drs. Reza Kalhor, Henry Lee, Nili Ostrov, Javier Fernandez-Juarez, Ben Stranges and Jun Teramoto. Fellow graduate student Anik Debnath ensured that at least one of us
would keep the post-docs on their toes. Thanks also go to honorary members of 233: Yoav Mayshar, Susan Byrne, Seth Shipman, Raj Chari and Prashant Mali.

Members of the Church lab as a whole deserve credit, as they are too numerous to enumerate. In particular I’d like to thank Emma Taddeo and Vonda Shannon for their support and for the lab’s smooth operation behind the scenes. I’ve also learned a lot about translating science outside academia from Jessica McDonough, James Culverwell, Leandro Vetcher, Su Chiang, Nick Conway and Mary Tolikas.

I’d also like to thank members of my dissertation committee, Professors Andrew Murray, Ed Boyden and Adam Cohen, who are inspiring scientists and engineers and also members of my dissertation defense committee, Professors Andrew Murray, Doug Melton and Ron Weiss. I am thankful for their time and for the chance to discuss my work with them.

The Systems Biology PhD Program has been amazing, and I’d like to thank Professors Andrew Murray and Timothy Mitchison for directing the program, and Samantha Reed and Elizabeth Pomerantz for coordinating everything so smoothly. My classmates John Ingraham*, Thomas Hiscock*, Peter Koch*, Martin Lukacisin and Jose Reyes have been great friends and I enjoyed our banter over beers, boardgames and biking (* indicates roommates).

I’m also grateful to my friends outside of science: Johnson Lam, Jessica Zou, Kenny Chan, Nian Mei Lin, Eva Diep, Eddie Chan and Henry Mei. Without them, I may have spent all of my waking hours in the lab.

I would also like to thank my early scientific mentors, Professors Thibault Mayor, Joerg Gsponer, Eric Lagally and Honglin Luo. Honglin hosted me as a high school summer research assistant, which got me hooked into science despite not fully realizing what was really going on. Eric hosted me as part of the undergrad iGEM team and introduced me to synthetic biology.
Thibault and Joerg hosted me in their labs for most of my college years and solidified my desire to pursue graduate school.

Thanks to the Canadian Natural Sciences and Engineering Research Council (NSERC) and the Peter Lynch Foundation for their financial support.
To my parents and sister for supporting me all these years.
Chapter 1: Introduction

Stem cells are defined by their ability to differentiate into mature cell types and self-renew. Pluripotent stem cells possess the ability to become all cell types in the body and have potential applications in regenerative medicine, drug screening, disease modeling and fundamental research. While there have been significant advances in methods to differentiate pluripotent stem cells, most protocols remain complex and time-consuming. For applications where pure populations of cells are needed, most differentiation protocols result in heterogeneous populations with low efficiency. Where complex three-dimensional tissues are desired, organoids typically lack the full repertoire of cell types. Recently, transcription factors (TFs) have been used to convert pluripotent stem cells into differentiated cell types with higher yields and simpler protocols. This thesis explores the use of a comprehensive TF library to differentiate human induced pluripotent stem cells (hiPSCs) into functional cell types, and the use of TFs to improve organoid differentiation by adding missing vascular endothelial cells to cerebral organoids.
1.1 Human induced pluripotent stem cells

Mouse embryonic stem cells (mESCs) were first cultured by Martin Evans, Matthew Kaufman and Gail Martin from the inner cell mass of blastocysts (1,2). They showed these mESCs to be pluripotent by their ability to spontaneously differentiate into cells derived from all three germ layers *in vitro* and *in vivo* by injection into mice that resulted in a teratoma.

Human embryonic stem cell lines were first derived by James Thomson and colleagues in a similar fashion from *in vitro* fertilized human embryos (3). This method requires the destruction of human embryos, which is an ethical concern, and does not allow an adult to generate their own pluripotent stem cells. John Gurdon and Shinya Yamanaka won the Nobel Prize in Physiology or Medicine in 2012 for their discovery that adult cells can be induced back into pluripotency. By using somatic cell nuclear transfer, John Gurdon demonstrated that when the nuclei of an intestinal epithelial cell is transferred into an enucleated egg, the differentiated cell’s nuclei could be induced into embryonic development that results in viable tadpoles (4). His findings show that genetic information is not lost after differentiation, and that factors within he egg’s cytoplasm are sufficient to induce pluripotency.

Takahashi and Yamanaka defined the genetic factors that induced mouse somatic cells into pluripotency (5). They discovered that four TFs - OCT4, SOX2, C-MYC, and KLF4 - converted mouse fibroblasts into induced pluripotent stem cell cells (iPSCs). James Thomson and colleagues later demonstrated induced pluripotency from human fibroblasts using a different combination of TFs: OCT4, SOX2, NANOG, and LIN28 (6).

hiPSCs have been shown to be pluripotent by a number of methods, including the expression of pluripotency markers, ability to differentiate into cells arising from the three germ layers *in vitro* and the formation of teratomas upon injection into mice. The ability of hiPSCs to
contribute to or develop into a viable human has not been proven due to legal bans and ethical concerns on human cloning. Nevertheless, mouse iPSCs have been shown to contribute to the germ line as chimeras by blastocyte injection (7) and to develop into viable mice by tetraploid blastocysts (8).

1.2 Culture and Differentiation of hPSCs

Methods to culture hPSCs have advanced significantly during the past decades. hiPSCs were first cultured on irradiated fibroblasts as feeder cells to maintain pluripotency (6). Feeder-free, pluripotency-reinforcing conditions were later defined using recombinant basic fibroblast growth factor (bFGF) and Matrigel, a mixture of extracellular matrix (ECM) proteins derived from mouse sarcoma whose main composition includes laminin (9, 10). Defined ECM components for hPSC culture have since been developed, including Laminin-521 (11). Improvements in passaging and cloning efficiency were achieved using Y-27632, a Rho-kinase inhibitor (12). These key advances have made hiPSC culture more reliable, reproducible and scalable.

hPSCs were first differentiated spontaneously using embryoid bodies to demonstrate pluripotency (3). But due to the uncontrolled nature of spontaneous differentiation, many undesired lineages are generated and limited their use for downstream applications. Directed differentiation towards specific lineages have been developed using knowledge from developmental biology. This has informed what steps to take in the branching, multi-stage process towards progenitors and into terminally differentiated cell types.

Using differentiation of hPSCs into neurons as an example, a major advancement was the report of the dual-SMAD inhibition protocol (13). This protocol eliminated the need for undefined factors secreted by stromal cells, removed the need for embryoid body formation and
sped up the process compared to withdrawal of pluripotency factors. This protocol yields >80% Pax6-positive neural cells in 7 days that are capable of later differentiating into mature, functional neurons. SMAD inhibitors were chosen based on observations that inhibitors of the bone morphogenic protein (BMP) pathway are released by the Spemann organizer during embryonic development. One disadvantage of this protocol is that it requires attention to plating density. Lower densities will produce neural crest cells, which are typically undesired.

Manipulating tissue micro-environment can also induce hPSC differentiation. Mechanical properties of the ECM, such as matrix stiffness, have important roles in stem cell differentiation and maintenance (14). Differentiation of hPSCs into neurons has been achieved using soft synthetic matrices without additional growth factors, and do not occur on a stiff matrix (15). The use of a soft matrix was inspired by the relative softness of brain tissues as compared to other tissues such as muscles and bones. Regardless of which method is used for differentiation, they will eventually converge on transcription factors and changes in gene expression to alter cell identity.
Figure 1.1. Three broad approaches to differentiate stem cells.

Developmentally inspired soluble factors, such as morphogens, bind receptors and induce a signaling cascade to induce TF expression. Matrix properties, such as stiffness, are sensed by G protein-coupled receptors that transduce signals for eventual TF expression. Genetic perturbations, such as TF over-expression, induce expression of downstream genes.
1.3 Transcription factors for cell conversion

Transcription factors (TFs) are defined as proteins that both bind DNA in a sequence-specific manner and regulate transcription (16). Based on their work on the bacterial lac operon, François Jacob and Jacques Monod proposed the first model of gene regulation. In this model, trans-acting factors control transcription by binding specific DNA sequences near genes (17). The first sequence-specific human TF, Sp1, was discovered by William Dynan and Robert Tjian (18). It is estimated that there are ~1,600 TFs encoded in the human genome (19, 16). The largest family of approximately 750 TFs has C₂H₂ zinc finger domains, some of which are involved in repressing transposable elements in the genome (20). Roughly 250 TFs contain homeodomains and another 100 TFs contain bHLH domains. Both of these domains are associated with developmental processes (19).

Forced expression of TFs can convert cell identity. Robert Davis, Harold Weintraub and Andrew Lassar discovered the first example of this, where over-expression of MyoD cDNA converted mouse fibroblasts into myoblast-like cells (21). Typically, a combination of TFs is necessary for direct conversion from one somatic cell type into another in a process called transdifferentiation. Conversions from fibroblasts into neurons, cardiomyocytes, hepatocytes and blood progenitors have used combinations of TFs (22, 23, 24, 25, 26). The most dramatic example of TF-mediated cell conversion is the reprogramming of somatic cells back to pluripotency into cells called induced pluripotent stem cells (5, 6). The first reported conversion of mouse fibroblasts to pluripotency used Oct3/4, Sox2, c-Myc, and Klf4. These four TFs are known as the Yamanaka factors. The first study to do this from human fibroblasts used Oct4, Sox2, Nanog and Lin28. TFs have also recently been used to convert hPSCs into differentiated...
cells, such as neurons and endothelial cells, typically with higher yields and faster speeds (27, 28, 29, 30).

1.4 Challenges in selecting transcription factors for cell type conversion

The challenge in TF-mediated cell type conversion is identifying which TFs can induce the desired conversion. In the examples described previously, candidate TFs were selected based on their importance in the development of the target cell type, or based on their selective expression in the target cell type. After selecting a candidate pool of typically ten to thirty TFs, several rounds of screening narrows down the minimal set of necessary and sufficient TFs for cell conversion.

More recently, computational algorithms have been developed to predict TFs that can induce cell conversion. These include mining gene expression data to identify antagonistic TF pairs (31), reconstructing gene regulatory networks to predict TFs that push the network towards the desired cell type (32, 33), scoring TF based on tissue specificity (34), and using chromatin state with expression profiles to predict TFs that induce cell conversion (35). Most of the predicted combinations for cell conversions remain to be validated experimentally.

Unbiased screening of TFs takes a reverse approach in identifying TFs for cell conversion. Systematic studies for the differentiation of mESCs using several hundred TFs or using cDNA libraries have been conducted. Two studies used the activation of a neuronal reporter to select TFs that differentiate mESCs into that lineage (36, 37). One study used depletion of TFs due to reduced proliferation of differentiated cells compared to mESCs as a proxy for differentiation (38). Three studies from the same laboratory used gene expression microarrays to infer differentiation of mESCs (39, 40, 41). A comprehensive study of all human TFs and their ability to differentiate hPSCs into any cell type remains unexplored.
1.5 Organoids

Organoids are aggregates of cells derived from stem cells that self-organize into three-dimensional structures to resemble organs (42, 43). In contrast to homogenous populations of a single cell type, organoids are considered to be more physiological due to the presence of multiple cell types that function together in a three-dimensional environment. Intestinal crypt organoids generated by Toshiro Sato, Hans Clevers and colleagues were the first organoids generated from adult stem cells (44). Their discovery that Lgr5-positive cells marked intestinal stem cells enabled them to isolate these adult stem cells. They embedded single stem cells within a Matrigel matrix and used appropriate signaling molecules to promote their self-organization into crypt-villus structures. Mototsugu Eiraku, Yoshiki Sasai and colleagues generated the first organoids from ESCs and produced retinal organoids with optic-cup morphologies (45). These retinal organoids formed structures with apical-basal polarity as seen by the presence of an outer retinal pigment epithelial layer and an inner neural retinal layer.

Since these initial reports, many protocols have been developed to produce organoids resembling a variety of organs. These include organs derived from ectoderm (e.g. whole brain, cerebellum and hippocampus), mesoderm (e.g kidney), and endoderm (e.g. stomach, lung and liver) (46, 47, 48, 49, 50, 51, 52). In addition to recapitulating development in vitro to understand fundamental biology, organoids have also been used to model diseases such as microcephaly, Zika virus, autism spectrum disorder, and cystic fibrosis (53, 54, 55, 56, 57, 58).

Transcription factors have also been used to generate organoids. Thyroid organoids were generated from mESCs by aggregation and transient over-expression of NKX2-1 and PAX8, and addition of thyroid-stimulating hormone (59). Liver bud-like organoids were produced by
transient, heterogenous expression of GATA6 in a minimal media, leading to sorting and emergence of hepatocytes, vascular cells and stromal cells (60).

Although organoids generate multiple cell types with spatial self-organization, many organoids lack the full repertoire of cell types found in vivo. Important cell types such as vascular cells, mesenchymal cells, stromal cells, immune cells and neural cells, are often missing (61). For instance in brain organoids, vascular endothelial cells, pericytes and microglia have not been observed to differentiate endogenously (62). As these cell types are derived from germ layers other than ectoderm, other methods of incorporating them will need to be explored.

1.6 Major questions addressed in this dissertation

In this dissertation, I develop methods to address two broad questions in the field: How can we comprehensively determine which TFs are capable of inducing cell type conversion? How can we use TFs to incorporate additional cell types within organoids?

In Chapter 2, we created the human TFome library, the first comprehensive human TF open reading frame (ORF) expression library. We used the human TFome library to query virtually all human TFs for their ability to differentiate hiPSCs into any differentiated cell type. We found 243 hits that induced differentiation in multiple hiPSC lines in four days under pluripotency-maintaining conditions. Most of these TFs are novel for their ability to induce differentiation. We then validated 15 high ranking TFs to confirm their capacity to induce differentiation. To retrospectively identify the differentiated cell type, we used RNA sequencing and classification algorithms. Lastly, we describe two novel cell conversions from hiPSCs into functional stromal fibroblast-like cells and functional neuron-like cells.

In Chapter 3, we developed a strategy called orthogonal differentiation to incorporate additional cell types to organoids. We applied this strategy to vascularize cerebral organoids.
using TFs. We identified a potent splice isoform of ETV2, which enabled this differentiation approach. This isoform converted hiPSCs into endothelial cells in both pluripotency-maintaining and neural-inducing conditions. These endothelial cells have a stable cell identity, have transcriptomic signatures of endothelial cells and are functional by their ability to form tubes and lumens that resemble capillaries. Together, this enabled orthogonal differentiation into two alternative cell types in the same media condition, which led to the successful incorporation of vasculature within cerebral organoids.
Chapter 2: Differentiating diverse human cell types using a comprehensive transcription factor library

2.1 Chapter summary

Ectopic expression of transcription factors (TFs) can reprogram somatic cells to pluripotency and convert cells into other cell types. The full extent to which TFs can alter cell identity is unknown. Here, we constructed the human TFome library, the first comprehensive human expression library that compasses all 1,576 annotated human TFs with 1,748 open reading frames (ORFs). We screened the human TFome library for differentiation of human induced pluripotent stem cells (hiPSCs) into any lineage. We identified 243 hits that individually caused loss of pluripotency in multiple hiPSC lines, even in pluripotency-maintaining conditions after only four days. This suggests a widespread capacity of individual TFs to alter cell identity. We validated 15 top ranking hits and show they differentiated hiPSCs into diverse cell fates. Two cell conversions were functionally characterized to produce stromal fibroblast-like cells and neuronal-like cells. We anticipate the human TFome library will accelerate efforts to differentiate many human cell types.

2.2 Introduction

On-demand production of any human cell type in vitro has the potential to transform many areas of basic research and medicine. However, we lack the ability to generate most of the thousands of estimated human cell types necessary for disease modeling, drug discovery, and tissue engineering (63, 64). Current approaches to produce specific cell types include recapitulating developmental biology with soluble factors (65), manipulating tissue micro-
environments through synthetic scaffolds (14), and programming cell identity with genetic elements (66). These strategies all ultimately converge on a common effect: altering the transcriptional state of the cell.

Transcription factors are key mediators of the transcriptional state of the cell, and over-expressing them can alter cell identity: somatic cells can be reprogrammed into pluripotent stem cells (5, 6) and trans-differentiated into myoblasts (21), neurons (22), cardiomyocytes (23), macrophages (67), hepatocytes (24, 25), and blood progenitors (26). The selection of TFs for these conversions typically focuses on what candidate TFs to screen for what desired cell type. A group of 10 to 30 candidate TFs is typically chosen based on developmental studies, or more recently from computational predictions, for a given cell type (31, 32, 34, 33, 35). Larger scale studies using cDNA libraries or hundreds of TFs have been conducted (36, 38, 37, 41), but a full survey of the ~1,500 human TFs with respect to their ability to induce cell type conversion remains incomplete. Such a systematic and comprehensive study has been impossible due to the lack of a comprehensive TF expression library.

Here, we have assembled the human TFome library, the first comprehensive human TF expression library representing all 1,576 annotated human TFs with 1,748 ORFs. Using this library, we queried TFs for their ability to differentiate hiPSCs into any lineage. 243 hits were individually able to induce differentiation in multiple hiPSC lines, suggesting a widespread for individual TFs to alter cell identity. 15 top ranking hits were validated and induced diverse cell fates. We found that NKX3-2 differentiates hiPSCs into functional stromal fibroblast-like cells and NEUROG3 induced neuronal, not endocrine, conversion. These demonstrate the power of the human TFome library for large-scale cell type conversion.
2.3 Results

2.3.1 Construction of a comprehensive human transcription factor library

We define the human TFome as the collection of all annotated TF genes in the human genome. Current annotations vary between 1,500 to 1,600 human TFs (16, 19). A complete human TFome expression library does not currently exist, thus our first goal was to construct such a library. We manually curated existing TF annotations and aimed to obtain at least one open reading frame (ORF) clone for the 1,576 TF genes. We consolidated 1,475 ORFs from various sources including the most complete ORF collection, the Broad-CCSB ORFeome v8.1 (Figure 2.1A) (68, 69, 70). For the remaining 273 ORFs that were not available in these collections, we synthesized them to produce an ensemble of 1,748 isoform-level clones that represent all 1,576 TFs. To explain the absence of those 273 ORFs from current repositories, we hypothesized that these genes may be expressed at lower levels, and hence difficult to clone from cDNA libraries or not available as cDNA clones from the Mammalian Gene Collection (71). Indeed, we observed these genes to be expressed at lower levels compared to genes that were successfully cloned from cDNA (Figure 2.1B). Consolidation of existing ORF collections together with the synthesis of hundreds of genes enabled the creation of the human TFome library.

To deliver and express the human TFome library in hiPSCs, we verified several features. More than 97% of the TF ORFs are smaller than 4kb, which made it possible to package them into lentiviral particles for gene transfer and genomic integration in hiPSCs (Figure 2.1C). We performed pooled cloning of the library from Gateway entry vectors into an all-in-one Tet-On
lentiviral vector (Figure 2.1D) (72). We achieved high coverage of the entire library and observed high expression in hiPSCs (Figure 2.1E,F).
Figure 2.1. Construction of the human TFome library.

(A) Table of the source of ORFs in the Human TFome ORF library. (B) Boxplot showing ORFs cloned from cDNAs originated from genes that are expressed in tissues higher than ORFs that were not available from cDNA, and therefore were synthesized. For each TF, a median tissue expression value was computed and the TFs that were synthesized (n = 273) were compared to those that were cloned from cDNAs (n = 1303). Medians are marked by the thick line. Boxes contain the inner 50% of the data. Whiskers show first quartile – 1.5x interquartile range (IQR) and third quartile + 1.5x IQR. Outliers are excluded from the plot for clarity. Statistical significance was determined using the Wilcoxon rank-sum test. (C) Distribution of TF ORF size. Size constraints of adeno-associated virus (AAV), lentivirus and PiggyBac transposon vectors are indicated. (D) Schematic of Gateway shuttling vectors pooled and then cloned into a lentiviral expression vector containing the doxycycline-inducible promoter (Tet-On) with a V5 C-terminal tag and constitutive PGK promoter driving the puromycin resistance gene (PuroR) and the reverse tetracycline-controlled transactivator (rtTA). (E) Correlation between TFs in the pDONR Gateway shuttling vector and after pooled cloning into the pLIX_403 lentiviral vector. FPKM, fragments per kilobase per million reads. (F) Ectopic TF expression after lentiviral transduction into hiPSCs, as shown by immunostaining for common V5 tag at the 3’ end of TFs sub-cloned into pLIX_403. *** P < 0.001. Scale bar, 300µm.
2.3.2 Loss of pluripotency as a cell-type-agnostic readout for differentiation

To identify TFs that can differentiate hiPSCs into any lineage with high efficiency and speed, we performed our experiments with two important parameters. First, based on our previous work using NEUROG1/2 to induce neuronal differentiation with >90% efficiency in only four days in pluripotency-maintaining conditions (28), we hypothesized the existence of other TFs that are equally potent at differentiating cells under the same conditions. Additionally, to generalize the screen for differentiation into any lineage, we did not use a lineage-specific marker; instead, we used loss of pluripotency as a cell-type-agnostic marker for differentiation. Specifically, we stained for TRA-1-60, a surface keratan sulfate antigen that is expressed on pluripotent cells and is rapidly lost upon differentiation (28, 73).

To illustrate loss of pluripotency as a readout of TF-mediated differentiation, we generated five control stable hiPSC lines that over-express ORFs with known effects at a single copy (Figure 2.2A). After four days of doxycycline induction, the two differentiation-inducing TFs NEUROG1 (28) and ASCL1 (30) induced loss of TRA-1-60 as observed by a large peak of TRA-1-60-low cells (Figure 2.2B). 47±2% and 45±0.2% of NEUROG1- and ASCL1-expressing cells lost TRA-1-60 respectively compared to no doxycycline conditions (Figure 2.2C). By contrast, the antibiotic resistance genes HYGRO<sup>R</sup>, BLA<sup>R</sup> and NEO<sup>R</sup> did not induce loss of TRA-1-60 in either condition (Figure 2.2C). Differentiation (or lack thereof) was immediately apparent by bright-field microscopy, where NEUROG1- and ASCL1-expressing cells migrated away stem cell colonies and exhibited differences in morphology (Figure 2.2D). Cells expressing HYGRO<sup>R</sup>, BLA<sup>R</sup> and NEO<sup>R</sup> remained as stem cell colonies (Figure 2.2D). These results show the use of TRA-1-60 as a cell-type-agnostic readout for differentiation by flow cytometry.
**Figure 2.2. Loss of pluripotency marker TRA-1-60 as a cell-agnostic readout.**

(A) Experimental scheme to generate stable inducible hiPSC lines with single copy TF integration. (B) Flow cytometry for five different hiPSC lines expressing the indicated ORF for four days in pluripotency maintaining conditions and stained for TRA-1-60. (C) Quantification of B (n = 3, Student’s t-test). (D) Bright-field microscopy of five hiPSC lines with or without doxycycline for four days. ** P < 0.01. *** P < 0.001. Scale bar, 100µm.
2.3.3 A pooled, FACS-based TFome screen reveals hundreds of TFs can induce loss of pluripotency

To screen the human TFome for TFs that induce rapid and efficient differentiation, we performed pooled, FACS-based screening using the entire library. We generated a stable mixed hiPSC population capable of inducible expression of the library, then proceeded similarly to the previous experiment (Figure 2.3A). Importantly, after over-expression, staining and FACS, we collected both TRA-1-60-low and TRA-1-60-high populations and computed a ratio for each TF from the massively parallel sequencing data. This was done instead of other comparisons, such as compared to non-induced cells or pre-sorted cells, for a number of important reasons.

Because of the pooled nature of the screen, we first needed to account for differences in the abundance of TFs arising from bacterial propagation, lentiviral packaging and transduction into hiPSCs. By using the ratio of TRA-1-60-low compared to TRA-1-60-high, we could normalize for differences in TF abundance. Second, hiPSCs exhibit basal spontaneous differentiation and could lead to a large number of false positives if only the TRA-1-60-low population was collected. To account for this, we used the ratios of the antibiotic genes as negative controls. Third, the over-expression of certain TFs may increase proliferation or cause cell death. This confound was mitigated by analyzing only the doxycycline-induced population and comparing between TRA-1-60-high and TRA-1-60-low cells. Thus, TFs are over-expressed in the whole population and the only effects on pluripotency are captured.

We conducted a pooled screen in three independent hiPSC lines, with each line transduced three independent times with over >100x representation per TF. We detected 95% of TFs in the input DNA used for lentiviral production (Figure 2.3B). After four days of doxycycline induction and TRA-1-60 sorting, we extracted genomic DNA, performed PCR using

18
universal primers to amplify the genomically integrated TFs and aligned it to the reference sequences. We detected between 87% to 91% of TFs recovered from the three hiPSC lines (Figure 2.3B). Based on the ratio of the negative control genes, we used the top 25% of hits from each hiPSC line and calculated the overlap to determine the top hits (Figure 2.3C). 243 hits were found in at least two of the three cell lines and were statistically significant by permutation testing. NEUROG1 and ASCL1 used as positive controls were detected in this set. We found 56 hits that were found in all three cell lines, which was also statistically significant (Figure 2.3D). We did not detect enrichment of any gene ontologies with either the set of 243 hits or the set of 56 hits compared to all TFs. Overall, this pooled screening in multiple hiPSC lines suggests several hundred candidate TFs that may induce loss of pluripotency in hiPSCs.
Figure 2.3. A pooled FACS-based screen reveals 243 hits for hiPSC differentiation.

(A) Experimental scheme for pooled TFome screening for loss of pluripotency. (B) TFs detected from each sample. (C) Overlap between hits from three hiPSC lines. (D) List of 56 hits found in all three hiPSC lines.
2.3.4 PiggyBac transposons enable rapid engineering of stable hiPSCs for inducible differentiation

To validate these hits, we aimed to generate stable hiPSC lines for each hit and confirm their loss of pluripotency. We began producing lentiviruses for each TF, but found the process to be labor-intensive. We also experienced difficulty with achieving high copy transductions without excessive cell death. For these reasons, we explored the use of PiggyBac transposons to streamline the cell line engineering process. PiggyBac transposons are mobile genetic elements derived from the cabbage looper moth that transposes through a “cut-and-paste” mechanism and have been used in human cells for reprogramming (74).

To pilot the use of PiggyBac for inducible differentiation of hiPSCs, we constructed the PBAN all-in-one Tet-On PiggyBac vector, which is similar to the pLIX_403 vector used for the pooled screen (Figure 2.4A). We cloned the control TF NEUROG1 into PBAN and nucleofected hiPSCs at different ratios of the TF vector to the transient vector encoding the transposase. We observed a ratio-dependent increase of differentiation upon doxycycline induction, assessed by staining for loss of NANOG, OCT4/POU5F1 and SOX2 by flow cytometry (Figure 2.4B). We also achieved a higher differentiation efficiency using PiggyBac transposons compared to lentiviral transduction of the same gene with 88±2% and 57±2% differentiation respectively.

We hypothesized that increasing the amount of DNA nucleofected correlated with higher differentiation efficiency because more TFs were being integrated into the genome. To test this, we nucleofected varying amounts of DNA, generated stable hiPSC lines and then extracted genomic DNA to quantify the copy number of integrated vectors by digital droplet PCR. We observed a DNA-dependent increase of integrations with increasing amounts of DNA nucleofected (Figure 2.4C). With lower amounts of DNA, an average of ~2 copies of vector were
integrated per genome whereas higher amounts of DNA resulted in an average of 14 copies of integrated per genome. Together these results indicate that PiggyBac transposons are capable of delivering inducible TFs into hiPSCs to generate stable cell lines.
**Figure 2.4. PiggyBac transposons enable rapid engineering of inducible hiPSC lines.**
(A) All-in-one PiggyBac vector with doxycycline-inducible expression and constitutive puromycin resistance. (B) Flow cytometry analysis of loss of NANOG, OCT4/POU5F1 and SOX2 to compare lentiviral versus PiggyBac delivery of TFs, and different ratios of PiggyBac to the transient vector expressing the transposase (n = 3, t-test). (C) Digital droplet PCR quantification of average integrated PiggyBac copies per genome, with varying amounts of PiggyBac DNA. * P < 0.05. ** P < 0.01. *** P < 0.001. Error bars are standard errors of the mean.
2.3.5 Validation of hits reveal diverse morphologies

We next validated hits using PiggyBac transposons. At this time, we selected the 15 top scoring candidates based on the screen in the PGP1 hiPSC line, and not on their score in the overlap of three hiPSC lines because this additional data was not yet available. We generated 15 individual stable hiPSC lines, and then induced over-expression for four days, and assessed loss of pluripotency by flow cytometry using another panel of pluripotency markers: NANOG, OCT4/POU5F1, and SOX2. In all cases, we observed significant loss of pluripotency, and in some cases close to 100% of cells lost pluripotency (Figure 2.5A).

To corroborate these results, we also observed a loss of stem cell morphology by bright field microscopy (Figure 2.5B). Doxycycline-induced cells no longer appeared as compact stem cell colonies, but non-induced control cells maintained their stem cell morphology. The differentiated cells also appeared to exhibit diverse morphologies, which suggests they were differentiating into cells of different lineages. For instance, ATOH1 and NEUROG3-induced cells appeared to exhibit neurites similar to neuronal cells. NKX3.2 and ETV2-induced cells had large, flatten morphologies. However they were subtly different, as ETV2 cells appeared to be space-filling whereas NKX3.2 cells were larger and in contact with each other. FOXC1, SOX14, HOXB6, ZSCAN1 and MITF appeared flat with comparatively smaller size. Overall, these 15 TFs were validated to induce rapid and efficient stem cell differentiation, potentially into diverse cell lineages.
Figure 2.5. Fifteen hits are validated to induce differentiation with diverse morphologies. (A) Flow cytometry for loss of NANOG, OCT4/POU5F1 and SOX2 for TF hits after four days with or without doxycycline induction. (B) Brightfield microscopy of cells with or without doxycycline. * P < 0.05, ** P < 0.01, *** P < 0.001. Error bars are standard error of the mean. Scale bar, 200µm.
2.3.6 Transcriptomic characterization of hits reveal diverse gene profiles

To determine the lineage that these TFs were inducing, we performed RNA sequencing on the top 11 TFs. To gain a global perspective on the transcriptomic data, we projected the data into three-dimensions using principal component analysis (Figure 2.6A). We observed diverse trajectories for each TF; for instance, ATOH1 and NEUROG3 cells were projected to a similar region; NKX3-2, MYOG and HOXB6 were alone in their respective coordinates and several TFs induced transcriptomes that clustered closely together. The diverse transcriptomic signatures corroborates the diverse morphological data shown previously and support the notion that a variety of cell types are being generated.

These TFs appeared to have powerful effects on the transcriptome. In each case, we found between ~2,000 to ~8,000 statistically significant up-regulated genes and ~4,000 to ~8,000 statistically significant down-regulated genes compared to hiPSCs (Figure 2.6B). The exogenous TFs were over-expressed from ~30 to 70,000 fold over non-induced conditions, indicating the large transcriptomic changes were due to high over-expression of the over-expressed TFs (Figure 2.6C). Together, these transcriptomic data exemplify the diverse lineages being generated by the hits.
Figure 2.6. Transcriptomic characterization of eleven differentiation-inducing hits.

(A) Principal component analysis of RNA-seq data from the 11 hiPSC lines over-expressing the indicated TF. The two TFs highlighted in red were chosen for further characterization. (B) The number of statistically significant differentially expressed genes are shown for each TF. (C) The fold-change in over-expression of the exogenous TF is shown. Error bars are standard errors of the mean.
2.3.7 NKX3-2 induces stromal fibroblast-like cells

Based on their highly divergent transcriptome data, we chose NKX3-2 and NEUROG3 for deeper characterization. NKX3-2 over-expression in mesenchymal cells can induce chondrocyte-like phenotypes and homozygous inactivating mutations of NKX3-2 in humans lead to a rare form of skeletal dysplasia (75, 76). To determine what cell type NKX3-2 induces from hiPSCs, we used CellNet, a machine learning classifier that uses transcriptomic data and compares it to reference datasets (32). CellNet reported a mixed fibroblast and stem cell signature, suggesting a fibroblast identity (Figure 2.7A). To further support this, we subjected the transcriptome data to gene ontology analysis (77, 78). The top ontology was endomembrane system organization, suggesting activation of pathways for secretion of extracellular matrix (ECM) proteins typically found in stromal cells (Figure 2.7B).

To support the notion that NKX3-2 induces differentiation of stromal fibroblast-like cells, we assessed the expression of ECM genes (collagen types I, III, and fibronectin), and the stromal cell markers (VIM, ALCAM, S100A4, and CD34). These genes were all statistically significantly up-regulated compared to stem cells and upon the top up-regulated genes (Figure 2.7C). We did not observe statistically significant up-regulation of chondrocyte markers, such as SOX9 or CHI3L1, and we observed only weak up-regulation of aggrecan and collagen type II (data not shown). To determine if NKX3-2 induced stromal fibroblast differentiation with high efficiency, we quantified vimentin expression and observed 99±0.1% of cells stained positive by flow cytometry (Figure 2.7D). Furthermore, we performed antibody staining for vimentin, ALCAM and SERPHINH1 and showed protein expression in the induced stromal fibroblast-like cells but not in stem cells (Figure 2.7E). Our characterization shows that NKX3-2 induces hiPSCs into stromal fibroblast-like cells.
A hallmark of stromal fibroblasts is the ability to remodel ECM (79). This can be functionally assessed by the contraction of a 3D collagen matrix (79, 80). To prove that NKX3-2-induced stromal fibroblasts cells are functional, we embedded either stem cells or NKX3-2-induced stromal cells into a collagen matrix for twenty-four hours. We observed contraction for the induced stromal cells to 1cm$^2$, but not the stem cells which remained at 1.5cm$^2$ (Figure 2.7F). Together, these results show that NKX3-2 converts hiPSCs into functional stromal fibroblast cells with high efficiency. 
Figure 2.7. NKX3-2 induces stromal fibroblast-like cells.
(A) CellNet classification of NKX3-2 induced cells to fibroblast and embryonic stem cell profiles. (B) Top gene ontology showing the most up-regulated genes are enriched for endomembrane system organization. (C) Heatmap of selected stromal fibroblast genes. (D) Flow cytometry for vimentin stromal fibroblast marker. (E) Antibody staining for stromal fibroblast markers. Scale bar, 300µm. Insert width, 100µm. (F) Collagen contraction assay. Scale bar, 4mm. (G) Quantification of surface area from the collagen contraction assay. Error bars are standard errors of the mean. * P < 0.05. ** P < 0.01. *** P < 0.001.
2.3.8 NEUROG3 induces neuronal differentiation

We selected NEUROG3 as another TF for deeper characterization. NEUROG3 is a master regulator of pancreatic islet differentiation and is required for the development of endocrine cells in the pancreas (81, 82). We were intrigued by the possibility that could NEUROG3 could efficiently and rapidly produce pancreatic cells in vitro. We classified NEUROG3-induced cells with CellNet and observed a strong neuronal signature (Figure 2.8A). Furthermore, the top gene ontology was neuron projection morphogenesis, which suggested that pathways for neuronal development were activated (Figure 2.8B). Both of these pieces of data suggest that NEUROG3 may act similar to NEUROG1/2 in inducing neuronal differentiation.

To determine whether NEUROG3 induces neuronal but not endocrine differentiation, we evaluated the expression of a panel of neuronal and endocrine markers. Neuronal markers such as neuron-specific structural genes (MAP2, TUBB3, NEFH), synaptic and neuronal adhesion proteins (DLG4 and NCAM1), and neuronal transcription factors (DCX, PAX6, MYT1L) were up-regulated (Figure 2.8C). REST, a transcriptional repressor to prevent neuronal differentiation (83), was also down-regulated. In contrast, most endocrine markers were largely unchanged or down-regulated (Figure 2.8D). This indicates that NEUROG3 differentiates hiPSCs into neuronal-like cells, but not endocrine cells.

To determine if NEUROG3 induced highly efficient neuronal differentiation, we stained NEUROG3-induced cells for NCAM. We observed 88±1% of cells were NCAM-positive by flow cytometry (Figure 2.8E). We further stained for NeuN, NF200, TUBB3 and MAP2, and showed protein expression as well as correct localization of the protein in the nucleus for NeuN and neurites for the other genes (Figure 2.8F). To assess whether NEUROG3-induced cells were functional, we performed whole-cell patch clamp electrophysiology. We observed single action
potentials upon current injection seven days after maturation in neuronal media as a co-culture with astrocytes, trains of action potentials after fourteen days, and trains of spontaneous action potentials within twenty-one days (Figure 2.8G-J). Together, these results demonstrate that NEUROG3 induces hiPSCs into functional neuronal-like cells with high efficiency.
Figure 2.8. NEUROG3 induces neuronal, not endocrine, differentiation.

(A) CellNet classification of NEUROG3-induced cells matching to neurons. (B) Top gene ontology showing the most up-regulated genes are enriched for neuron projection morphogenesis. (C) Heatmap of selected neuronal marker genes. (D) Heatmap of selected endocrine marker genes. (E) Flow cytometry for NCAM. (F) Antibody staining for neuronal markers. Scale bar, 300µm. Insert width, 100µm. (G) Summary of whole-cell patch clamp electrophysiology results, indicating fraction of patched cells with the colored action potential characteristic. (H) Example traces recorded following current injection seven days post doxycycline induction (DPI). Responses to each positive current step is shown, with the first step to elicit an action potential highlighted in red. (I) Example traces recorded following current injection fourteen DPI. (J) Example traces recorded without current injection twenty-one DPI. Error bars are standard errors of the mean. * P < 0.05. ** P < 0.01. *** P < 0.001.
2.4 Discussion

Here, we have presented a new method to systematically query virtually all individual human TFs to differentiate hiPSCs into any lineage. This was made possible by our creation of the human TFome library, the first comprehensive human TF ORF expression library containing 1,576 TFs. The entire collection is being made publicly available on several DNA repositories. We identified 243 TF hits that could individually induce differentiation in multiple hiPSC lines in only four days, even under pluripotency-maintaining conditions. We also used PiggyBac transposons as a method to rapidly engineer stable hiPSCs capable of highly efficient differentiation. In addition, we characterized two conversions: NKX3-2 differentiates hiPSCs into functional stromal fibroblast-like cells and NEUROG3 converts hiPSCs into functional neurons, but not endocrine cells. We envision this systematic approach and comprehensive resource will accelerate the identification of TFs to convert any cell type.

This study provides a global perspective of the TF-mediated stem cell differentiation landscape. Our finding that hundreds of TFs have the ability to induce stem cell differentiation indicates a widespread capacity for individual TFs to alter cell identity. This number is larger than those reported in previous studies. In one study, 8,000 mouse cDNA clones were over-expressed as sub-pools in mESCs, and 15 cDNA clones were found to activate a neuronal reporter under differentiation conditions (36). In another study, 700 human TF ORFs were expressed individually in mESCs and 24 TFs were reported to activate a neuronal reporter (37). There were no overlapping hits between these two neuron-specific studies. In a more recent study that was unbiased with respect to cell type, 137 TFs were tested in mESCs. Using gene expression microarrays, they found 21 TFs that induced conversion into a variety of cell types (41). Of our 243 hits, one was found in Falk et al., three was found in Theodorou et al. and four
was found in Yamamizu et al. Due to the large differences in screening conditions, including media conditions, TF delivery, duration of over-expression, species and cell lines, the poor overlap is not surprising.

There are a number of important parameters that would alter or improve our screen. First, the number of TFs we identified as differentiation-inducers may be an underestimate because we used pluripotency-maintaining conditions. Such conditions may have inhibited the effect of TFs that may otherwise have induced differentiation. One concern with performing a screen while withdrawing pluripotency-maintaining factors is higher spontaneous differentiation. This would make it more difficult to identify differentiation caused by TFs, and not by the lack of pluripotency-maintaining factors. Second, some TFs may require more than four days to induce differentiation. One issue with performing a longer screen is that stem cells proliferate more quickly than differentiating cells. This means cells that differentiated due to a potent TF would represent a smaller fraction of the population and be harder to detect as stem cells continue to proliferate. Third, the use of single copies of randomly integrated lentiviruses also meant the level of over-expression may have been strongly affected by the location of integration. The transgene may be silenced or may not achieve high expression compared to a safe harbor locus or having multiple copies of the same TF. Lastly, we were unable to determine the identity of 5 of the 11 TF-induced cell lines based on their transcriptomes. These cells may be in an early progenitor stage without a clear gene expression profile, in a cell state where we do not accurate transcriptomic data, or have lost pluripotency but also do not represent any cell type in vivo.

To more fully realize the power of the human TFome for cell conversion, further work and additional advances in technology will be necessary. First, the human TFome library should be viewed as the first of many versions. An expanded human TFome with comprehensive splice
isoforms would allow a systematic exploration of their effects. A synthetic human TFome, through modifications such as phosphomimetic mutations, addition of activator domains or removal of auto-inhibitory domains, may improve nuclear localization, protein stability, binding to protein partners and to DNA. Other genetic perturbations, such as RNAi knock-downs have revealed essential factors for pluripotency (84, 85) and could be used in combination with TFs. Epigenetic modifiers, microRNAs and transcriptional splicing factors may also be powerful tools in cell conversion.

Combinations of TFs will almost certainly be needed to convert between many cell types. Advances in DNA assembly and synthesis may enable the one-pot assembly of specific TFs together for targeted combinatorial screens (86). Natural partners and co-factors of TFs may be co-expressed to improve TF function. Scalable single cell-resolution readouts will also be needed to resolve the specific combinations received by a given cell. Recent in situ, droplet- and microwell-based single cell RNA-seq technologies may enable such single cell combinatorial studies (87, 88, 89, 90, 91). In fact, combinatorial genetic perturbations using CRISPR/Cas9 knockouts were performed with single cell RNA-seq as a readout (92, 93, 94). Furthermore, the temporal dynamics of TF expression are important (67), and improvements to orthogonal inducible promoters, either by small molecules, light or other means, and new technologies are needed to control this. Additionally, tuning the level of TF over-expression may also reveal subtle but important effects on cell conversion.

Lastly, a reference catalogue of every human cell type would greatly enable comparisons to in vitro differentiated cell types and may serve as blueprints to inform large-scale cell conversion studies. The Human Cell Atlas project aims to produce such a catalogue at a single cell resolution (95). Beyond adult tissues, it will also be important to obtain expression profiles
from embryonic stages of development to the aging human body. This would help pin-point not only cell identity but also the maturation status of *in vitro* derived cell. This is particularly important as many *in vitro* differentiated cell types remain embryonic in age. Effective computational methods to compare transcriptomes and inform additional genetic alterations to improve fidelity will also be needed (32). Through iterative refinement of transcriptomic and epigenetic states towards the target cell type, age and maturation status, we may truly achieve any desired cell type. As less-studied cell types are produced, especially in the context of supporting cells within tissues, improvements in culturing methods will also be needed. Efficient methods to deliver TFs into other cell types will need to be improved. This would facilitate the use of the human TFome library for large-scale systematic cell conversion studies in cell types other than hiPSCs. Overall, we anticipate this first human TFome library and screen will accelerate efforts to derive every human cell type *in vitro*.

2.5 Methods

2.5.1 Annotation of human transcription factors

A starting list of 1,591 TFs from Vaquerizas *et al.* (19) was used, based on 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 (HUGO) Gene Nomenclature Committee (HGNC) (96) were added: zinc fingers (including those containing C2H2), homeodomains (including LIM, POU, TALE, HOXL, NKL, 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 approved gene names using the
HGNC multi-symbol checker. The final set of TFs included in the human TFome contained 1,576 genes.

2.5.2 Construction of the human TFome

Gateway-compatible ORFs for the target set of 1,576 genes were obtained from the Center for Cancer Systems Biology CCSB-Broad Human ORFeome v8.1 collection (70), the Taipale lab (68), Arizona State University DNASU Plasmid Repository (69), and transOMIC (http://www.transomic.com). For the missing genes, Uniprot protein sequences were reverse-translated and codon-optimized for synthesis by Gen9 Inc. (Boston, USA). Where multiple isoforms exist, the one designated as “canonical” by Uniprot was selected for synthesis. If the selected isoform was longer than 4kb, the longest isoform less than 4kb was chosen due to synthesis constraints. Synthesized genes were cloned into pDONR221 using the same methods as above. Clonal, sequence-verified individual human TFome pDONR221 plasmids will be available on Addgene, DNASU and transOMIC.

2.5.3 Analysis of TF expression in tissues

GTEx (97) version 6 median tissue FPKMs were downloaded. For each TF, its median expression across tissues was computed. TFs that were synthesized were compared to TFs that were cloned from cDNAs.

2.5.4 Pooled library cloning into lentiviral expression vector

To ensure high library representation, 19 parallel cloning and transformation reactions were performed. Each reaction involved a sub-pool of one 96-well plate containing TFs in
pDONR Gateway shuttling vectors. 75ng of pDONR-TF was cloned into 75ng of pLIX_403, the
doxycycline-inducible lentiviral vector (pLIX_403, Addgene plasmid # 41395), using 1µl LR
Clonase II (Invitrogen, 11789100) in TE pH 8.0 overnight at 25°C for 18 hours. 0.5µl proteinase
K was added and incubated at 37°C for 1 hour. 1µl of the reaction was transformed into one vial
of Stbl3 chemically-competent cells (Invitrogen, C737303), and spread LB agar containing
100µg/µl carbenicillin (Teknova, E0096) and incubated at 30°C overnight. Colonies were
counted to ensure library coverage of >200x. Colonies were scraped and resuspended in PBS.
Optical density was measured and plasmid was extracted using the endotoxin-free Midi Prep
plus kit (QIAGEN, 12943). DNA was quantified using the Quibit dsDNA broad range kit
(Invitrogen, Q32853). To determine coverage using next-generation sequencing, pDONR-TF and
pLIX_403-TF sub-pools were normalized by DNA amount and pooled into their own pDONR-
TF and pLIX_403-TF pools. They were then sheared using a Covaris S2 to 200bp and verified
for size using 1% E-Gel EX; 4µg was used for library prep with the NEBNext Ultra DNA
Library Prep Kit for Illumina (New England Biolabs, E7370L). To avoid over-amplification of
libraries, 1/5 of the libraries were amplified by quantitative PCR by adding SYBR Gold Nucleic
Acid Stain (Invitrogen, S11494) and running on a Lightcycler 480 (Roche). The number of
cycles required to reach mid-log amplification was determined, and the remaining library was
amplified using that number of cycles. Amplified libraries were run on a 1% E-Gel EX to verify
the size distribution and absence of adaptor dimers. Libraries were quantified using the KAPA
Real-Time Library Amplification kit (KK2702) and loaded onto an Illumina MiSeq v3 150-cycle
kit (MS-102-3001). Control pLIX_403 plasmids containing the resistance genes blasticidin,
hygromycin, and neomycin were generated by PCRing the ORF, BP cloning into pDONR221,
and then LR cloning into pLIX_403.
2.5.5 Data processing and analysis for library cloning.

To align reads to the TFome library, reference TF sequences were first indexed using the STAR aligner v2.5.2a (98) using the command STAR --runMode genomeGenerate with parameter --genomeSAindexNbases 9 to accommodate a reference “genome” with fewer bases, as recommended in the manual. Reads were aligned to the reference index using the STAR command, counted using bash scripts, and plotted in R.

2.5.6 Cell culture.

The PGP1 hiPSC line without genomically integrated Yamanaka factors was generated from fibroblasts (Coriell, GM23248) (99) using the CytoTune Sendai Reprogramming Kit (Life Tech, A16517) and will be deposited into a cell repository. They were adapted to feeder-free culture, verified for pluripotency by FACS, and karyotyped. CRTD5 hiPSCs were a gift from the Center for Regenerative Therapies Dresden ES/iPSC Facility and reprogrammed from BJ, a primary foreskin fibroblast line (ATCC, CRL-2522). ATCC-DYS0100 hiPSCs were purchased from ATCC (ACS-1019). Cell lines were verified by short tandem repeat (STR) profiling (Dana Farber Cancer Institute), regularly verified to be mycoplasma-free using PlasmoTest (InvivoGen, rep-pt1), and cultured between passages 8 and 40. hiPSCs were cultured in mTeSR1 (STEMCELL Technologies, 05850) without antibiotics on tissue-culture-treated plates coated with Matrigel (Corning, 354277). hiPSCs were passaged using TrypLE Express (Life Technologies, 12604013) and seeded with 10µM Y-27632 ROCK inhibitor (Millipore, 688001) for one day. Cells were frozen in mFreSR (STEMCELL Technologies, 5854) using a CoolCell LX (Biocision, BCS-405) overnight at □80°C, then in vapor-phase liquid nitrogen for long-term storage. □
2.5.7 Lentiviral production and transduction.

Lentiviral particles as one pool of the complete library were produced as previously described (100). Briefly, the pooled lentivirus was produced by transfecting pMD2G (Addgene plasmid 12259), psPAX2 (Addgene plasmid 12260), and pLIX_403-TFome using polyethyleneimine (Polyscience, 24765) into 293T cells. Media was exchanged after 24 hours and supernatants were harvested at 48 hours and 72 hours post-transfection. The supernatants were filtered (0.45-mM PES filter, Corning 431220), precipitated at 4°C overnight using a PEG solution (BioCat, K904-50-BV) and resuspended with PBS in 1/100 of the supernatant volume (100x concentrated). The particles were transferred in 50µl aliquots into 1.5-ml screw-cap tubes, snap-frozen on dry ice and stored at -80°C. Titrations were performed by qPCR as previously described (100) and titer as $1 \times 10^8$ IFU/ml.

Each hiPSC was independently transduced three times. 1.8 million hiPSCs were used for each transduction at MOI = 0.1 in mTeSR1 media. The mTeSR media was exchanged daily. 48 hours post transduction, 3µg/ml puromycin was added to the media to eliminate non-transduced PGP1 cells. Cells were maintained and propagated as mentioned before.

2.5.8 Immunofluorescence microscopy.

Cells were stained with antibodies as previously described (28). Cells were grown on 12mm coverslips (Warner Instruments, CS-12R15) coated with Matrigel, and fixed with 4% PFA (Electron Microscopy Sciences, 15714-S) in PBS for 20 minutes at room temperature, then washed three times with PBS and kept at 4°C. Fixed samples were incubated with block solution containing 10% normal donkey serum (Millipore, S30-100ml), 1% BSA, 0.5% Triton X-100 in 1PBS for 1 hour at room temperature, or overnight at 4°C. The stain buffer was similar to the
block buffer, except normal donkey serum was reduced to 3%. The stain buffer containing primary antibodies was incubated at room temperature for 1 hour, washed with stain buffer twice, incubated with secondary antibodies at room temperature for 1 hour, then washed once with stain buffer and twice with PBS. Coverslips were mounted on glass slides by incubating with Prolong Diamond Antifade with DAPI (Invitrogen, P36966) overnight at room temperature, and sealed with nail polish (Electron Microscopy Sciences, 72180). Slides were imaged on a Zeiss Observer.Z1 microscope equipped with a Plan-Apochromat 20×0.8 objective, a four-channel LED light source (Colibri), and an EM-CCD digital camera system (Hamamatsu).

2.5.9 Flow cytometry and fluorescence-activated cell sorting (FACS)

For flow cytometry analysis and FACS, cells were dissociated using TrypLE Express, washed, and resuspended in FACS buffer (PBS with 10% FBS). For surface antigens, live cells were stained with fluorophore-conjugated antibodies and the viability dye CellTrace Calcein Blue, AM (Life Technologies, C34853) at 1×10^7 cells/ml for 30 minutes on ice in the dark. This viability dye is particularly important when isolating TRA-1-60low 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), and run on a BD LSRFortessa. Compensation for spectral overlap was determined by staining AbC Total Antibody Compensation Beads (Life Technologies, A10497).
with single fluorophore-conjugated antibodies. All antibodies are listed in Table S4. Flow cytometry data was analyzed using FlowJo 10.2.

2.5.10 Morphological analysis by live-cell microscopy

Live cells were imaged on a Zeiss AxioObserver.Z1 microscope equipped with a Plan-Apochromat 20x/0.8 objective, a four-channel LED light source (Colibri), and an EM-CCD digital camera system (Hamamatsu).

2.5.11 TFome screen using FACS

Each independently TFome-transduced iPSC population was passaged into six wells of a six-well plate at 300,000 cells/well into mTeSR with 10µM ROCK inhibitor, 0.5µg/ml doxycycline, and 1µg/ml puromycin. The next day, the media was replenished with mTeSR1 with doxycycline, and from then on replaced daily. After four days of doxycycline induction, cells were dissociated using TrypLE Express, and counted using an automated cell counter Countess II (ThermoFisher Scientific, AMQAX1000). Typically, 3 million cells per 6-well plate and were then stained for PE anti-TRA-1-60 antibodies and Calcein AM live stain (Life Technologies, C1430) in mTeSR1 media at 10^7 cells/ml for 30 minutes in the dark, washed, and then filtered for debris using a single-cell strainer (Falcon, 352253). Cells were sorted on a BD FACSARia or Beckman Coulter MoFlo Astrios. Calcein-AM-positive cells were gated, then the bottom 20% and the top 20% of TRA-1-60 cells were designated as TRA-1-60-low and TRA-1-60-high, respectively. At least 50,000 cells were sorted per gate and condition. Cells were spun down and the cell pellet was frozen at −20°C.
2.5.12 Genomic extraction, PCR and library preparation for next generation sequencing.

Genomic DNA was extracted from sorted cells using the DNeasy Blood & Tissue Kit (QIAGEN 69506) on a QIAcube (QIAGEN). >5µg of genomic DNA was used for PCR. To prevent over-amplification, a calibration quantitative PCR was performed using KAPA HiFi HotStart ReadyMix spiked with SYBR Gold on a Roche LightCycler 480. Input gDNA volumes were adjusted to amplify at mid-log phase with the same number of cycles. Amplified PCR products were purified using QIAquick PCR purification kit (QIAGEN, 28106) on a QIAcube. Purified DNA was quantified using Qubit dsDNA broad range quantification kit (Invitrogen, Q32859). 1µg was sheared to 200bp on a Covaris sonicator E220 and used for library preparation using NEBNext Ultra DNA Library Prep, quantitative PCR for library amplification and quality control as described above. Samples were sequenced on an Illumina NextSeq 500 on high-output mode.

2.5.13 Analysis of sequencing data from TFome screening.

Sequencing reads of amplified TFs were aligned to reference sequences using STAR, and counted used bash scripts, as above. Samples were quality controlled by hierarchical clustering to assess technical reproducibility. Replicates that did not cluster within their sub-pool were considered as outliers and removed. Log₂(TRA-1-60-low/TRA-1-60-high) values were computed using DESeq2 (101). The top 25% of TFs from each hiPSC line were considered candidates, and overlaps where candidates were present in at least two of the three hiPSC lines or in all three were generated.
2.5.14 Nucleofection of PiggyBac and generation of stable cell lines.

PBAN, a Gateway-compatible, doxycycline-inducible, puromycin-selectable PiggyBac vector was constructed from PB-TRE-dCas9-VPR (Addgene #63800). Individual pDONR-TFs were cloned into PBAN using LR Clonase II. 500,000 to 800,000 hiPSCs were nucleofected with PBAN-TF and Super PiggyBac Transposase (SPB; System Biosciences, PB210PA-1) at a DNA ratio of 4:1 using Nucleofector P3 solution (Lonza, V4XP-3032). Nucleofected cells were transferred to a 6-well Matrigel-coated plate in mTeSR1 with ROCK inhibitor. When cells reached 80% confluence, 1µg/ml puromycin (Gibco, A1113803) was added. The next day, dead cells in suspension were washed away using PBS; if the remaining cells were sparse, ROCK inhibitor was added to prevent colony collapse.

2.5.15 Quantification of integrated TF copy number by digital droplet PCR.

Digital droplet PCR was performed on a Biorad QX100. 15ng of genomic DNA was used per reaction with ddPCR supermix (Biorad, 1863026), primers to amplify and HEX probe against RPP30 (Biorad, 10031243), primers and FAM probe against puromycin (forward primer: 5’- TGCAAGAACTCTTCCTCACG, reverse primer: 5’- CGATCTCGGCGAACACC, FAM probe: 5’ACATCGGCAAGGTGTGGGTCG), and AluI (New England Biolabs, R0137S).

2.5.16 RNA sequencing library preparation.

600µl TRIzol (Life Technologies, 15596-018) was added directly to cells, which were then incubated for 3 minutes and used for RNA extraction using Direct-zol RNA MiniPrep (Zymo Research, R2050). At least three replicates of control cells (without doxycycline) were processed in parallel in each set of library preps. RNA was quantified using Qubit RNA HS Kit
(Molecular Probes, Q32852) and RNA integrity was confirmed by the presence of intact 18S and 28S bands on a 1% E-Gel EX. 1μg RNA was used for Poly(A) isolation using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, E7490L) and the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, E7420L). To prevent library over-amplification, one-fifth of the PCR reaction was amplified by quantitative PCR using SYBR Gold Nucleic Acid Statin on a Roche Lightcycler 480. The remaining reaction was amplified using the number of cycles needed to reach mid-log amplification. Library size was visualized on a 1% E-Gel EX, and quantified using KAPA Library Quantification Kit as described before.

2.5.17 Analysis of RNA sequencing data.


- RNA-seq reads were aligned on four codes each with 12Gb memory using the command: STAR -quantMode GeneCounts.

- Gene counts per sample were merged into a master table and analyzed in R version 3.2.2. Differential expression analysis was performed using DESeq 2 (101), comparing each batch to its no-doxycycline control separately. Principal component analysis was performed in R. A local version of CellNet (32) was downloaded and FASTQ files were used. For the gene set enrichment analysis (GSEA) (102), genes were ranked by the sign of their log2 fold change.
compared to hiPSCs, divided by their adjusted P value, and then submitted to GenePattern (103) for GSEA on gene ontologies using the GSEAPreranked module with c5.bp.symbols gene set database with 1000 permutations, classic scoring scheme, and default minimum 15 gene and maximum 500 gene set size. FASTQ files will be available on NCBI GEO.

2.5.18 Collagen contraction assay.

Collagen contraction assays were performed according to Ngo et al. (80). NKX3-2 cells, grown in mTeSR1 either with or without doxycycline for 4 days, were dissociated and counted. 400,000 cells in 300 µl mTeSR1 were mixed with 150 µl collagen type I diluted to 3 mg/ml (BD, 354236) and 2 µl 1M NaOH, and set in a 12-well plate used as a mold. Collagen gels were left to solidify at room temperature for 20 minutes, then 500 µl mTeSR1 was slowly added to the gels. The gel was dissociated from the mold by running a P200 tip along the edge of the well. The plate was incubated overnight and images were captured using a Zeiss Axio Zoom.V16 Stereo Zoom Microscope with a color AxioCam MRm camera and a PlanNeoFluar Z 1x/0.25 objective. The area of the gel was quantified in Fiji.

2.5.19 Electrophysiology.

Neuronal function was assessed similar to Busskamp et al. (28). Rat astrocytes (Gibco, N7745100) were seeded onto Matrigel-coated coverslips and cultured in DMEM with GlutaMAX, 10% FBS, and N2 supplement (Gibco, 17502048). One day before seeding induced neurons, astrocytes were cultured in Neurobasal media with GlutaMax and B27 supplement. NEUROG3 hiPSCs were labeled by lentiviral transduction of constitutively expressed GFP (FUGW, Addgene 14883). 500,000 GFP-labelled NEUROG3 and ATOH1 hiPSCs were seeded in
a 6-well tissue culture plate and induced with doxycycline for three days on Matrigel-coated coverslips, then dissociated using TrypLE Express. Cells were counted and 1 million induced neurons were plated into pre-conditioned media on the astrocytes. Media was changed twice weekly.

Electrophysiological recordings were carried out at 20–25°C on an upright Olympus BX51WI microscope. Cells were bathed in artificial cerebral spinal fluid (ACSF) containing (in mM) 119 NaCl, 2.5 KCl, 4 CaCl2, 4 MgSO4, 1 NaH2PO4, 26.2 NaHCO3, and 11 glucose, 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 KMeSO3, 17.8 HEPES, 0.6 MgCl2, 1 EGTA, 4 Mg-ATP, and 0.3 Na-GTP. Traces were collected using a MultiClamp 700B amplifier (Molecular Devices), filtered with a 2kHz Bessel filter, digitized at 10kHz using a Digidata 1440A digitizer (Molecular Devices), stored using Clampex 10 (Molecular Devices), and analyzed off-line using customized procedures in Igor Pro (WaveMetrics). Cells were assessed for the presence of spontaneous action potentials in current-clamp mode. In the absence of spontaneous action potentials, cells were assessed by the injection of a set of current steps, ranging from −140 pA to 400 pA in 60 pA increments, with a duration of 0.6 s. Action potential parameters were quantified using the first action potential evoked at the lowest current injection that resulted in an action potential or the spontaneous action potentials. The threshold was defined as the voltage at which dV/dt of the action potential waveform reached 10% of its maximum value, relative to a dV/dt baseline taken 10 ms before the peak. Action potential amplitude was defined as the difference between the threshold value (in mV) and the maximum voltage at the peak of the action potential. The width was measured at half-maximum amplitude.
If cells were resting above 60 mV, a constant current injection was applied to bring the voltage down to between 60 mV and 70 mV.
Chapter 3: Vascularization of cerebral organoids using transcription factor-mediated orthogonal differentiation

3.1 Chapter summary

Organoids are powerful models to study human development and diseases in a physiologically relevant 3D environment. Although organoids are comprised of diverse lineage-specific cell types with spatial organization that resemble organs, many lack critical cell types or require long protocols for additional cell types to differentiate. Cerebral organoids lack vasculature due to the absence of endothelial cells, which leads to necrosis in the interior. Here, we devise an orthogonal differentiation strategy to vascularize cerebral organoids. By engineering different populations of cells and mixing them together, we can use the same media conditions to induce different cell fates. We demonstrate a splice isoform of ETV2 can potently differentiate human induced pluripotent stem cells (hiPSCs) into functional endothelial cells, even in neural-inducing media. We then show stable incorporation of vasculature within cerebral organoids for two weeks. Although perfusion, reduced necrosis and barrier properties remain to be demonstrated, these results demonstrate that orthogonal, transcription factor-mediated differentiation within an organoid can be a powerful method to incorporate additional cell types within developing tissues.

3.2 Introduction

Since the initial studies on intestinal crypt and optic-cup organoids (44, 45), many organoids with remarkable similarity to a myriad of human organs have been reported (42, 43). Cerebral organoids can model the human brain as a whole and has been used to model diseases...
such as microcephaly, Zika virus and autism spectrum disorder (46, 53, 54, 55, 56, 57, 58). The full repertoire of cell types within cerebral organoids was recently described by single cell RNA sequencing (62). Important cell types were notably absent, including brain microvascular endothelial cells, pericytes and microglia. Other cell types required months to emerge, such as oligodendrocytes and inhibitory neurons. Therefore, for cerebral organoids to be more accurate and efficient models of the brain, methods to incorporate these cell types are necessary.

Endothelial cells form the vasculature for nutrient and waste exchange. Necrosis at the interiors of cerebral organoids is likely due to the lack of vasculature (46, 62). During development, the vasculature develops as mesoderm progenitors differentiate into angioblasts to form the primary plexus (105, 106). Neuroectoderm cells secrete vascular endothelial growth factor (VEGF) to guide angioblasts to penetrate the neuroectoderm and to differentiate into leaky endothelial cells. Wnt signaling causes these endothelial cells to mature into brain microvascular endothelial cells with tight barrier properties typical of the blood-brain-barrier (107).

No published reports have vascularized cerebral organoids (43, 108). Mixing primary human umbilical vein endothelial cells (HUVECs) with hiPSC-derived liver progenitors can vascularized liver organoids (51), but when HUVECs are mixed with hiPSCs for cerebral organoid formation, they separate out of embryoid bodies (109). Directed differentiation of hiPSCs into both cerebral organoids and endothelial cells using current protocols is incompatible; directed differentiation to endothelial cells requires mesoderm induction, typically by Wnt activation using CHIR99021(110), whereas cerebral organoids require neural induction conditions.

Here, we present an orthogonal differentiation strategy to vascularize cerebral organoids. In this strategy, different populations of hiPSCs are engineered to respond to different cues such
that they adopt alternate fates in the same media conditions. We identified a potent splice isoform of ETV2 that differentiates hiPSCs into functional endothelial cells, even in neural-inducing conditions. We then mixed these cells with unmodified hiPSCs that can respond to neural-inducing cues in the media. This enabled the stable incorporation of vasculature within cerebral organoids for two weeks.

3.3 Results

3.3.1 An orthogonal differentiation strategy to incorporate cell types within organoids

We aimed to incorporate endothelial cells within cerebral organoids by using an orthogonal differentiation strategy. Two cell populations, each capable of differentiating into alternative fates, are mixed together prior to organoid generation (Figure 3.1A). One population consists of unmodified hiPSCs that differentiate into neural cells induced by external media conditions. The other population is engineered with a doxycycline-inducible endothelial TF that differentiates into endothelial cells (Figure 3.1B). Importantly, this strategy requires highly efficient endothelial conversion, in order to override external neural-inducing cues. Without triggering TF expression, the two populations are mixed together as hiPSCs for embryoid body formation and cerebral organoid generation. Once cerebral organoids are formed, doxycycline is added into the media to induce expression of the endothelial TF without affecting neural cells within the organoid. The endothelial cells that emerge form vasculature within the organoid. Perfusion can occur using endothelial-lined channels that connect to endothelial cells sprouting from the organoid. Therefore, the same media conditions can induce different fates in an orthogonal manner that does not affect each other.
Figure 3.1. Orthogonal differentiation strategy for vascularizing cerebral organoids.

(A) Two populations of cells - one that responds to external cues for neural differentiation (blue) and another population that has an inducible endothelial TF (orange) - are mixed together. Standard embryoid body formation and cerebral organoid generation proceeds as usual, without induction of the endothelial TF. After cerebral organoids are formed, doxycycline is added to vascularize the organoid. This occurs by inducing endothelial TF expression and converting the engineered cells into endothelial cells. Perfusion occurs using endothelial-lined channels connecting to endothelial cells sprouting from the organoid. (B) Genetic circuitry depicting same media conditions with external cues for neural differentiation and doxycycline for endothelial TF expression, but induce different fates in the two cell populations. The endothelial TF overrides external cues.
3.3.2 A splice isoform of ETV2 induces potent endothelial differentiation

To enable this orthogonal differentiation strategy, we first needed TFs that induce potent endothelial differentiation. ETV2 was previously described to produce endothelial cells from both hiPSCs and fibroblasts, with efficiencies of 42% and 3% respectively (29, 111). These reports optimized expression level and dynamics to achieve these efficiencies. We were concerned that non-ideal media conditions, such as neural-inducing conditions, may further reduce differentiation efficiency and affect endothelial function. Furthermore, constraints on TF expression may be incompatible with cerebral organoid development. From our TFome screen (Chapter 2), we noticed two splice isoforms of ETV2 were present in the library but did not have the resolution to determine potential differences in their differentiation efficiency. We hypothesized that splice isoforms could be optimized to enhance endothelial differentiation.

To evaluate the differentiation efficiency of the ETV2 splice isoforms, we synthesized two missing splice isoforms for a total of four isoforms (Figure 3.2A) and examined their ability to differentiate hiPSCs into endothelial cells. We generated an inducible, stable hiPSC line for each ETV2 splice isoform, and induced differentiation for four days in pluripotency-maintaining conditions. We observed isoform 2, the second longest isoform, induced nearly complete conversion with 95±0.2% of cells staining for VE-Cadherin (Figure 3.2B). In contrast, isoform 1, the canonical isoform, produced only 48±1% VE-Cadherin-positive cells, in agreement with a previous report (29). Isoforms 3 and 4 lack a substantial number of residues near the N-terminus and had even lower efficiencies of 21±2% and 0±0% respectively. No major differences in the copy number of integrated ETV2 was observed in the four hiPSC lines, as all lines had greater than 10 copies per genome (Figure 3.2C). Furthermore, the minor differences in copy number did not correlate with the differentiation efficiency. To characterize the morphology of the
differentiated endothelial cells, we performed antibody staining for VE-Cadherin (Figure 3.2D). We observed strong staining on the cell membrane with a cobblestone morphology that is characteristic of endothelial cells. The stem cells that did not differentiate did not stain positive for VE-Cadherin. Together, these results indicate that ETV2 splice isoforms have a dramatic effect on the differentiation efficiency of hiPSCs into endothelial cells, and we used the novel splice isoform 2 to implement our orthogonal differentiation strategy.
Figure 3.2. A splice isoform of ETV2 induces potent endothelial differentiation.
(A) Splice isoforms of ETV2 annotated from Ensembl. Lines indicate introns, thin boxes indicate untranslated regions, and thick boxes indicate coding regions. (B) Flow cytometry on cells stained for VE-Cadherin after 4 days of doxycycline induction in pluripotency-maintaining conditions. (C) Quantification of genomic copy number integration of ETV2 in the four hiPSC lines. (D) Antibody staining for VE-Cadherin. Scale bars, 300µm. Error bars show standard error of the mean (s.e.m.). * P < 0.05. ** P < 0.01. *** P < 0.001. n.s. = not statistically significant.
3.3.3 Potent endothelial differentiation in neural-inducing conditions

With a potent differentiation protocol in pluripotency-maintaining conditions, we next tested whether this protocol translated to neural-inducing conditions. Efficient neural induction uses dual-SMAD inhibition with Noggin and SB431542 (13). The inhibition of TGF-beta signaling enhances hiPSC differentiation into neural cells. Due to the role of TGF-beta in both vascular development and also inhibition of endothelial proliferation (112, 113), it was unclear whether this protocol would translate to neural-inducing conditions.

To test whether we could achieve endothelial differentiation in neural-inducing conditions, we cultured doxycycline-inducible ETV2 isoform 2 hiPSCs in neural induction conditions, with or without doxycycline for four days. Strikingly, ETV2 hiPSCs cultured in the presence of doxycycline in neural conditions differentiated into endothelial cells, as stained by VE-Cadherin (Figure 3.3A). ETV2 cells cultured in the absence of doxycycline differentiated into SOX1-positive neural progenitors. This indicates that ETV2 successfully overrode neural-inducing cues. Unmodified hiPSCs differentiated into SOX1-positive neural progenitors regardless of doxycycline. This suggests that orthogonal differentiation could be possible, as the conditions to differentiate these two cell types did not affect one another.

To quantify the efficiency of neural and endothelial differentiation in neural-inducing conditions, we stained for SOX1 and VE-Cadherin and performed flow cytometry. ETV2 cells efficiently differentiated into VE-Cadherin-positive, SOX1-negative endothelial cells in the presence of doxycycline (Figure 3.3B). Nearly all unmodified hiPSCs regardless of the presence of doxycycline and ETV2 cells in the absence of doxycycline differentiated into SOX1-positive,
VE-Cadherin-negative neural progenitors. These results quantify the high orthogonality of differentiation into neural and endothelial fates in the same media conditions.
Figure 3.3. Highly efficient endothelial differentiation in neural-inducing conditions. (A) Antibody staining for the neural marker SOX1 and endothelial marker VE-Cadherin four days in neural induction media, with or without doxycycline. (B) Flow cytometry quantification of neural versus endothelial cells four days in neural induction media. Scale bar, 200µm. Insert, 100µm. Error bars are standard errors of the mean.
3.3.4 Endothelial cells remain stable after doxycycline withdrawal

Having shown the ability to induce endothelial differentiation in neural-inducing conditions, we next wanted to assess whether the endothelial cells were stable after withdrawing doxycycline. To test this, doxycycline was added for variable durations and then withdrawn. Doxycycline induction for only one day was sufficient to differentiate ETV2 cells into VE-Cadherin-positive endothelial cells without losing its phenotype (Figure 3.4). Similar results were observed after induction for two, three and the full four days. We did observe reduced cell numbers from shorter induction times, suggesting reduced endothelial proliferation. Nevertheless, this result show that continuous doxycycline addition is not required for endothelial differentiation or maintenance.
Figure 3.4. Endothelial cell phenotype is stable after doxycycline withdrawal. VE-Cadherin staining shows endothelial cells maintain a stable cell identity after doxycycline withdrawal. Scale bar, 100µm.
3.3.5 Transcriptomic characterization of differentiated endothelial cells

We next wanted to characterize the endothelial cells induced by this splice isoform. We first characterized them at a transcriptomic level. We performed RNA sequencing of 4-day induced endothelial cells and compared them to hiPSCs without induction. We used CellNet, a machine learning algorithm based on gene regulatory networks, to classify our differentiated cells. ETV2-induced endothelial cells were classified as endothelial and embryonic stem cell, indicating endothelial identity with residual stem cell signature (Figure 3.5A). We further used gene ontology enrichment analysis (103), which showed that vasculature development was the top gene ontology (Figure 3.5B). With this transcriptomic data, we then looked at the expression of a panel of endothelial markers. We observed expression of PECAM1 (platelet endothelial cell adhesion molecule), and VWF (von Willebrand factor), as well as receptors TIE1 and TEK and ESAM, and transcription factors in the development of endothelial cells, GATA2 and ERG (Figure 3.5C). Together, these indicate that ETV2 splice isoform 2 induces conversion into endothelial cells with high transcriptomic fidelity.
Figure 3.5. ETV2 isoform t has transcriptomic signature of endothelial cells.
(A) CellNet classification of ETV2 splice isoform 2 differentiated cells. (B) Vasculature development was the top gene ontology. Genes were ranked by their differential expression compared to hiPSCs, and shows highly ranked genes belonging to this gene ontology. (C) Heatmap of endothelial marker expression compared to hiPSCs.
3.3.6 Differentiated endothelial cells are functional

We next aimed to demonstrate that these induced endothelial cells were functional. Specifically, we aimed to determine if the cells could form tubes \textit{in vitro} that mimic capillaries. We used four-day differentiated endothelial cells for an angiogenesis assay, where cells are grown on a thick matrigel matrix to allow for tube formation. Overnight, we observed the formation of long tubes (Figure 3.6A). To show that these tubes formed lumens which could be used to exchange nutrients and waste, we performed transmission electron microscopy on cross-sections of the tubes and observed the presence of lumens (Figure 3.6B). The inner diameter was quantified to be 4.4±0.7μm (Figure 3.6C), which is typical of capillaries \textit{in vivo} (114). These lumen appeared to be formed by one single cell as shown by the presence of an intracellular cell junction visualized at higher magnification (Figure 3.6D). Together, these results demonstrate the endothelial cells are capable of functional lumen formation.
Figure 3.6. Differentiated endothelial cells are functional.

(A) Calcein AM staining of ETV2 isoform 2-induced endothelial cells in an angiogenesis assay on thick Matrigel. Arrows indicate tubes. Scale bars, 300µm. (B) Electron micrograph of ETV2 lumens (L). Scale bar, 1µm. (C) Quantification of lumen diameter. Error bars show standard error of the mean (s.e.m.) (D) Magnified electron micrograph of cell junctions sealing the lumen. Scale bar, 400nm.
3.3.7 Generation of stable vasculature within cerebral organoids

We then implemented our orthogonal differentiation strategy to vascularize cerebral organoids. We first dissociated unmodified hiPSCs and inducible ETV2 cells into a single cell suspension and mixed them at 80:20 ratio (Figure 3.7A). Instead of forming embryoid bodies using 96-well U-bottom plates (46), we used Aggrewell plates which reduce the variability of the size of embryoid bodies (115). Each Aggrewell contains 7,000 400µm-wide microwells of inverted square pyramids. Within six hours, embryoid bodies formed as spheroids with a characteristic internal cyst (Figure 3.7B). These were indistinguishable from embryoid bodies formed using only unmodified hiPSCs (data not shown). Over the course of three days, these embryoid bodies grew larger with no obvious abnormalities compared to embryoid bodies formed with only unmodified hiPSCs. This indicates that a mixed population of unmodified and inducible endothelial cells is capable of generating embryoid bodies.

As embryoid body formation does not incorporate all cells within the microwell, we wanted to ensure that inducible ETV2 cells were being incorporated within the embryoid bodies. This was important as previous attempts of mixing primary human umbilical vein endothelial cells (HUVECs) within embryoid bodies separated within hours of embryoid formation (109). To assess incorporation of inducible ETV2 cells, we genetically labelled unmodified hiPSCs and inducible ETV2 with lentiviruses encoding constitutively expressed GFP and mKate2 respectively to generate a stable line. Then the process of mixing these two populations to form embryoid bodies was repeated. We observed stable formation of embryoid bodies with contributions from both mKate2-labelled inducible ETV2 cells and GFP-labelled unmodified cells (Figure 3.7C). No obvious separation of mKate2-labelled cells was observed over the span of twenty-four hours. To obtain higher resolution confirmation that mKate2-labelled cells were
present within embryoid bodies, they were imaged by confocal microscopy. mKate2-labelled cells were clearly incorporated within embryoid bodies (Figure 3.7D). These results demonstrate that inducible ETV2 cells do not separate like HUVECs, and therefore are capable of stable incorporation within embryoid bodies.

To vascularize cerebral organoids, we harvested unlabeled four-day differentiated embryoid bodies and cultured them in suspension under neural induction media containing Noggin and SB431542. Doxycycline was added at this time for four days to induce endothelial conversion, and the cerebral organoids were fixed and stained eight days later (Figure 3.7A). VE-Cadherin-positive cells were observed within cerebral organoids cultured in the presence of doxycycline, but they were not observed in the absence of doxycycline (Figure 3.7E). This indicates successful production of endothelial cells within cerebral organoids. Anecdotally, vascularized cerebral organoids were typically larger in diameter than control cerebral organoids, but this difference was not quantified rigorously. Attempts were made to trace the vasculature network to determine if it interfaced with the media; however, such attempts were incomplete due to difficulty imaging the entire organoid. Nevertheless, these results demonstrate an orthogonal differentiation approach can successfully vascularize cerebral organoids.
Figure 3.7. Vascularization of cerebral organoids.

(A) Experimental scheme of mixing two populations to form vascularized organoids, with timeline and culturing conditions indicated below. (B) Brightfield time-lapse microscopy from single cell suspension in Aggrewell microwells to formation of embryoid bodies over seventy-two hours. Scale bar, 200µm. (C) Fluorescence time-course microscopy on GFP-labelled unmodified hiPSCs with mKate2-labelled inducible ETV2 cells forming embryoid bodies. Scale bar, 200µm. (D) Confocal microscopy of live embryoid bodies demonstrating incorporation of mKate2-labelled inducible ETV2 cells. Scale bar, 200µm. (E) Whole-mount antibody staining for SOX1 neural cells and VE-Cadherin endothelial cells by confocal microscopy of ETV2-incorporated cerebral organoids with or without doxycycline cultured for 14 days. Scale bars, 100µm.
3.4 Discussion

Here, we demonstrate an orthogonal differentiation approach to vascularize cerebral organoid. Two populations are mixed together in the same media condition, but each respond to different cues and undergo different fates: one population undergoes neural differentiation by external media conditions to form the cerebral organoid while another is engineered with inducible ETV2 that overrides external cues to become endothelial cells upon TF expression. This was enabled by the identification of a splice isoform of ETV2 that induced highly efficient endothelial differentiation in neural inducing conditions. Stable endothelial identity was maintained after doxycycline withdrawal. These cells were characterized by antibody staining, transcriptomic profiling, functional angiogenesis assays and electron microscopy that demonstrated lumen formation. Lastly, endothelial cells were stably incorporated within embryoid bodies and formed vasculature within cerebral organoids.

Our study demonstrates a new approach where endogenous and engineered cell types emerge simultaneously within organoids. We believe this has a number of advantages. First, it solves the challenge of cell separation when primary HUVECs migrated away from embryoid bodies during cerebral organoid generation. Second, it enables control over the timing of differentiation that is difficult with existing protocols, where missing cell types are added prior to 3D culture or after. Control over timing may be useful to mimic the emergence of cell types during specific time points of organ development. This could be achieved by timing TF expression via the addition of doxycycline or using conditional promoters. Third, the ability to induce alternative cell fates within the same media conditions using genetic components could enable sophisticated genetic circuits that carefully guide organoid development. This approach may be used to incorporate additional missing cell types, such as microglia and pericytes, as well
as speed up the differentiation of oligodendrocytes and inhibitory neurons. Fourth, this approach could be scaled as a mixed population of hiPSCs could be propagated and frozen. When vascularized cerebral organoids are needed, they can be grown to scale and vascularized by the simple addition of doxycycline. This alleviates the need to synchronize multiple cell types with different culturing conditions.

To more fully realize the potential applications of vascularized cerebral organoids, robust lumen formation with functional vasculature networks will need to be demonstrated. Electron microscopy of these organoids should show lumens and not only the presence of endothelial cells. Addition of labelled beads should travel within the organoid and be visible within the vascular networks. Furthermore, to reduce necrosis in the interior, current culturing methods based on diffusion, spinning bioreactors, or orbital shakers may not be sufficient; these organoids may need to be actively perfused using microfluidic channels. The best perfusion method would ideally remove all apoptotic signals as observed by antibody staining.

With this strategy, we can incorporate additional cell types to organoids and control their timing. Additional modes of control would be ideal: orthogonal inducible or conditional promoters would allow additional cell types to be differentiated separately, rather than all at once under doxycycline control. Genetic methods to precisely control the spatial location of cells would enable the creation of organoids with higher reproducibility or with high resemblance to a specific piece of tissue to be replicated.
3.5 Methods

3.5.1 ETV2 splice isoform analysis

ETV2 splice isoforms with a transcript support level of 1 were considered as high confidence isoforms and obtained from Ensembl (GRCh38). DNA-binding domain annotation was obtained from Uniprot.

3.5.2 Flow cytometry

Cells were dissociated using TrypLE express, washed, and resuspended in FACS buffer (PBS with 10% FBS). Cells were fixed with BD Cytofix fixation buffer (BD Biosciences, 554655) at 1 × 10^7 cells/ml for 20 minutes then washed with PBS twice and kept at 4°C or directly stained with FITC-conjugated mouse anti-human VE-Cadherin antibodies (BD, 560411) at 1 × 10^7 cells/ml for 30 minutes on ice in the dark. Stained cells were washed twice with FACS buffer, filtered into a strainer-capped tube (Falcon, 352235), and run on a BD LSRFortessa. Data was analyzed using FlowJo 10.2.

3.5.3 Quantification of integrated TF copy number by digital droplet PCR

Digital droplet PCR was performed on a Biorad QX100. 15ng of genomic DNA was used per reaction with ddPCR supermix (Biorad, 1863026), primers to amplify and HEX probe against RPP30 (Biorad, 10031243), primers and FAM probe against puromycin (forward primer: 5’- TGCAAGAACTCTTCTCCTACG, reverse primer: 5’- CGATCTCGGCGAACACC, FAM probe: 5’ACATCGGCAAGGTGGTGTTGGTCG), and AluI (New England Biolabs, R0137S).

3.5.4 Immunofluorescence microscopy
Monolayer cells were grown on 12mm coverslips (Warner Instruments, CS-12R15) coated with Matrigel, and fixed with 4% PFA (Electron Microscopy Sciences, 15714-S) in PBS for 20 minutes at room temperature, then washed three times with PBS and kept at 4°C. Fixed samples were incubated with block solution containing 10% normal donkey serum (Millipore, S30-100ml), 1% BSA, 0.5% Triton X-100 in 1×PBS for 1 hour at room temperature, or overnight at 4°C. The stain buffer was similar to the block buffer, except normal donkey serum was reduced to 3%. Samples were incubated with stain buffer containing rabbit anti-VE-Cadherin primary antibodies (Cell Signaling Technologies, 2500P) and/or goat anti-SOX1 (R&D Systems, AF3369) at room temperature for 1 hour, washed with stain buffer twice, and incubated with secondary antibodies at room temperature for 1 hour, then washed once with stain buffer and twice with PBS. Secondary antibodies include: donkey anti-rabbit secondary antibodies conjugated with Alexa Fluor 488 (Invitrogen, A21206) or Alexa Fluor 568 (Invitrogen, A10042), donkey anti-goat secondary antibodies conjugated with Alexa Fluor 488 (Invitrogen, A11055). Coverslips were mounted on glass slides by incubating with Prolong Diamond Antifade with DAPI (Invitrogen, P36966) overnight at room temperature, and sealed with nail polish (Electron Microscopy Sciences, 72180). Slides were imaged on a Zeiss Observer.Z1 microscope equipped with a Plan-Apochromat 20×0.8 objective, a four-channel LED light source (Colibri), and an EM-CCD digital camera system (Hamamatsu). Organoids were stained in a similar manner as above, except with over-night incubations, and imaged on a Zeiss LSM-780 confocal microscope.

3.5.5 RNA sequencing library preparation.
600µl TRIzol (Life Technologies, 15596-018) was added directly to cells, which were then incubated for 3 minutes and used for RNA extraction using Direct-zol RNA MiniPrep (Zymo Research, R2050). At least three replicates of control cells (without doxycycline) were processed in parallel in each set of library preps. RNA was quantified using Qubit RNA HS Kit (Molecular Probes, Q32852) and RNA integrity was confirmed by the presence of intact 18S and 28S bands on a 1% E-Gel EX. 1µg RNA was used for Poly(A) isolation using the NEBNext Poly(A) mRNA Magnetic Isolation Module (New England Biolabs, E7490L) and the NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs, E7420L). To prevent library over-amplification, one-fifth of the PCR reaction was amplified by quantitative PCR using SYBR Gold Nucleic Acid Statin on a Roche Lightcycler 480. The remaining reaction was amplified using the number of cycles needed to reach mid-log amplification. Library size was visualized on a 1% E-Gel EX, and quantified using KAPA Library Quantification Kit.

3.5.6 Analysis of RNA sequencing data.


RNA-seq reads were aligned on four cores each with 12Gb memory using the command: STAR -quantMode GeneCounts.

Gene counts per sample were merged into a master table and analyzed in R version 3.2.2. Differential expression analysis was performed using DESeq2 (101), comparing each batch to
its no-doxycycline control separately. Principal component analysis was performed in R. A local version of CellNet (32) was downloaded and FASTQ files were used. For the gene set enrichment analysis (GSEA) (102), genes were ranked by the sign of their log2 fold change compared to hiPSCs, divided by their adjusted P value, and then submitted to GenePattern (103) for GSEA on gene ontologies using the GSEAPreranked module with c5.bp.symbols gene set database with 1000 permutations, classic scoring scheme, and default minimum 15 gene and maximum 500 gene set size. FASTQ files will be available on NCBI GEO.

3.5.7 Angiogenesis assay.

For endothelial tubulogenesis assays, 70,000 ETV2-isoform hiPSCs were seeded into 48-well plates coated with 100µl Matrigel in mTeSR1 with doxycycline and ROCK inhibitor. The next day, the media was changed to the EGM-2 BulletKit (Lonza, CC-3162) with 500ng/ml doxycycline and replaced daily. On day 4, cells were stained live using 1µM Calcein Blue, AM (Life Technologies, C34853). Cells were imaged on a Zeiss Observer.Z1 microscope equipped with a Plan-Apochromat 20×0.8 objective, a four-channel LED light source (Colibri), and an EM-CCD digital camera system (Hamamatsu). 10µm-thick z-stacks that spanned the focal plane were captured and deconvolved using maximum z-projection in Fiji (104).

3.5.8 Transmission electron microscopy.

Cells were fixed overnight in 2.5% Glutaraldehyde, 1.25% Paraformaldehyde and 0.03% picric acid in 0.1M sodium cacodylate buffer (pH 7.4) at 4°C. They were then washed in 0.1M cacodylate buffer and post-fixed with 1% osmium tetroxide (OsO4)/1.5% potassium ferrocyanide (KFeCN6) for 1 hour, washed twice in water, once in maleate buffer and incubated in 1% uranyl
acetate in maleate buffer for 1 hour, followed by two washes in water and subsequent dehydration in the following grades of alcohol (10 min each; 50%, 70%, 90% and 2 × 10 min 100%). The samples were then placed in propyleneoxide for 1 hour and infiltrated overnight in a 1:1 mixture of propyleneoxide and TAAB Epon (Marivac Canada Inc. St Laurent, Canada). The following day, the samples were embedded in TAAB Epon and polymerized at 60 °C for 48 hours. Ultrathin sections (~80 nm) were cut on a Reichert Ultracut-S microtome, picked up onto copper grids stained with lead citrate and examined using a JEOL 1200EX transmission electron microscope. Images were recorded with an AMT 2 k CCD camera.

3.5.9 Cerebral organoid culture

Cerebral organoids were generated in a similar fashion as Lancaster et al. (46) with exceptions as follows. PGP1 hiPSCs were dissociated with TrypLE Express (Life Technologies, 12604013), resuspended in Aggrewell medium (STEMCELL Technologies, 05893) and counted using an automated cell counter (Countess II, AMQAX1000, ThermoFisher Scientific). The single cell suspension was transferred to Aggrewell400 plates (STEMCELL Technologies, 27945) for embryoid body formation. 600,000 cells were seeded into an Aggrewell containing Aggrewell medium with 10µM Y-27632 ROCK inhibitor (Millipore, 688001). The plate was spun down at 100 g for 3 minutes and placed in a tissue culture incubator overnight. The next day (day 1 of the protocol), embryoid body formation was verified by brightfield microscopy and the media was changed to neural induction media: DMEM/F12 + HEPES + GlutaMAX (Invitrogen, 11330-032) with N2 supplement (Gibco, A13707-01), non-essential amino acids (Gibo, 11140-050), Noggin (R&D Systems, 6057-NG-100) and SB431542 (Millipore, 616464). Half of the media was changed daily with neural induction media from days 1 to 3. On day 4,
embryoid bodies were harvested by pipetting with a wide-bore tip to dislodge them from the Aggrewells and filtered gently using a reversible cell strainer (STEMCELL Technologies, 27215). The strainer was reversed and washed with DMEM/F12 + HEPES + GlutaMAX into a 15mL tube. Embryoid bodies were transferred to ultra-low attachment 24-well tissue culture plates (Corning, 3473). On day eight, media was changed to neural differentiation media consisting of 1:1 of DMEM/F12 + HEPES + GlutaMAX (Invitrogen, 11330-032) and Neurobasal medium (Invitrogen, 12348-017) with non-essential amino acids (Gibo, 11140-050), N2 supplement (Gibco, A13707-01), B27 supplement without vitamin A (Gibco, 12587-010). Media was replaced every other day.

To produce vascularized cerebral organoids, PGP1 iPSCs and PGP1 inducible ETV2 iPSCs were dissociated, counted and mixed at a ratio of 80:20, then seeded into Aggrewells. 1µg/mL doxycycline was added into the media on day 4 and added daily for 4 days to induce ETV2 expression and endothelial trans-differentiation.

3.5.10 Genetic labeling of cells with fluorescent proteins.

Lentiviral plasmids encoding GFP were obtained from Addgene (#14883). An RFP version with membrane-bound mKate2 was obtained from Dr. Seth Shipman. Lentiviruses were produced as previously described (100). Briefly, lentiviruses was produced by transfecting pMD2G (Addgene plasmid 12259), psPAX2 (Addgene plasmid 12260), and the lentiviral vector using Lipofectamine 2000 (Invitrogen, 11668027) into HEK293T cells. Supernatants were harvested at 48 hours and 72 hours post-transfection, filtered (0.45-mM PES filter, Corning 431220) and precipitated at 4°C overnight using a PEG solution (BioCat, K904-50-BV). The viral pellets were resuspended with PBS in 1/100 of the supernatant volume (100x concentrated).
The particles were transferred in 50µl aliquots into 1.5-ml screw-cap tubes, snap-frozen on dry ice and stored at −80°C. Different volumes of concentrated virus was used to determine optimal transduction efficiency, as assessed by high percentage of cells labelled with GFP or mKate2.
Chapter 4: Discussion and future directions

On-demand production of any human cell or tissue is a major challenge for regenerative medicine. With the thousands of human cell types, how can we quickly find genetic recipes to convert any cell type into any other cell type? And if you can generate any cell type, how do you assemble them into a tissue? This dissertation has focused on developing methods to advance our technological capabilities on both fronts.

4.1 A strategy for sampling vast combinatorial space

In my first project, we developed a method to systematically query virtually all human TFs to convert hiPSCs into any differentiated cell type (Chapter 2). We created the human TFome expression library, the first comprehensive human TF library of 1,576 TFs that enabled such a systematic study. With this library, we screened for TFs that individually caused hiPSCs to differentiate and found 243 hits that could individually induce differentiation in multiple hiPSC lines. However, it is almost certain that many cell types will not be accessible with only one TF. Combinations of TFs are almost certainly required, and the ability to over-express any TF using the human TFome library will enable these studies.

I do not want to discount the challenges of such a combinatorial screen. New experimental methods are being developed by our laboratory and others as discussed in Chapter 2. Here, I want to illustrate the daunting task of exploring this vast combinatorial space and suggest a possible way forward. Starting from the collection of 1,576 TFs, there are 1.2 million pair-wise TF combinations, 640 million three-way TF combinations and over 250 billion four-way TF combinations. This is before factoring in the possibly hundreds of cells needed per combination to achieve reliable replicates. One study estimates the human heart to have
approximately 6 billion cells (117) and another estimates the human body to have 37 trillion cells (118). Therefore, a screen of these scales would literally require several organs’ worth of cells.

So are unbiased combinatorial screens impossible? Not necessarily. Taking inspiration from DNA microarrays for protein binding experiments, Martha Bulyk and colleagues were able to fit all possible DNA sequences of k-length onto a single microarray chip using de Bruijn sequences (119). In other words, one does not need to synthesize all possible DNA sequences of k-length; instead, if the oligonucleotides on a microarray chip are of length greater than k, there exists an algorithm to represent all k-length sequences within a smaller set of longer sequences.

A loose analogy could be considered here if DNA binding motifs were TFs and each oligonucleotide on the chip were a cell. It is important to stress that de Bruijn sequences may not be the algorithm to solve this combinatorial TF problem, but some computational approach likely exists.

The key idea is to “over-sample” TFs within each cell to explore all possible N-wise combinations using a number of cells that is smaller than the number of unique combinations. In other words, instead of transfecting the TFome library as a pool at an average integration of two copies per cell to explore all pairwise combination, one could transfect an average of ten TFs per cell to reduce the number of cells needed to sample the combinatorial space. To illustrate this more concretely, take an example of querying a set of six TFs for all three-way combinations. There are 20 unique combinations. In the naïve approach, one would integrate an average of three TFs per cell and expect on average that one in twenty cells will have a particular three-way combination. By contrast using the “oversampling” approach, we can modestly increase the number of integration to four TFs per cell and then determine what fraction of the population
harbors at least one of each of the three TFs from a particular three-way combination (Figure 4.1).
Figure 4.1. A strategy to explore vast combinatorial space.
A specific example of “oversampling” 4 TFs per cell to more effectively explore all three-way combinations. The fraction of cells that harbor at least one of each of the three TFs from the specific combination provides an estimate for its occurrence. The number of cells needed to have this occur is less in the “oversampling” approach than the naïve approach.
In brief, this probability follows a multi-nominal distribution, and shows that 8% of cells harbor this specific combination of three TFs. Thus we would expect one in approximately twelve cells to contain a particular three-way combination. This oversampling approach halves the number of cells needed compared to the naïve approach - and this is by minimal oversampling. A more convincing example is to query 500 TFs for all three-way combinations. There are over 20 million combinations; for a particular three-way combination, we expect one in 20 million cells to have exactly that combination. In the oversampling approach using 10 integrations per cell, we expect one in 100,000 cells to harbor a particular combination. This represents a 200-fold reduction in the number of cells needed for such a screen.

The goal of this section is to suggest a possible way to effectively sample a vast combinatorial space. The full mathematical details and experimental caveats are outside the scope of the present chapter. An obvious experimental caveat is the assumption that extraneous TFs do not impact the differentiation efficiency of a particular set of TFs that would otherwise induce cell conversion by themselves. How much of this interference exists will be important to determine empirically before using this approach. Additional calculations are needed to take into account the variation in the number of TFs integrated in a population, as it is unlikely that all cells receive exactly ten for instance. The number of cells needed to sample all combination with a given confidence interval is also important, as the calculations above only used average frequencies.

4.2 Beyond cell types: spatial control of 3D multicellular structures

In my second project, we developed orthogonal differentiation as a new approach to induce alternative fates in the same media conditions. We applied this to vascularize cerebral
organoids by mixing together one population of unmodified hiPSCs that differentiate into the neural lineage by responding to external cues, with another population engineered with inducible ETV2 that undergoes endothelial differentiation upon doxycycline induction. This was enabled by our identification of a splice isoform of ETV2 that induced endothelial differentiation even in neural-inducing conditions. These endothelial cells were stable, similar to endothelial cells by antibody staining and transcriptomic analysis, and function in forming lumens. This orthogonal differentiation approach is being applied to incorporate other missing or slow-to-differentiate cell types in cerebral organoids, and should be generally applicable to other organoids.

But will exogenously incorporated cells necessarily localize to their proper place within organoids or synthetic tissues? Will ETV2-induced endothelial cells form vasculature networks that are indistinguishable from those in human brains? Organoids have the remarkable ability to self-organize and classical experiments show that when tissues are dissociated and re-aggregated, they can self-organize with striking similarity to the original tissues (120, 121). We could hope that these cells will sort themselves and migrate to their proper locations. But it may be advantageous to have the ability to control where cells go.

Ideally, we would like to have complete genetic control over 3D multicellular structures in order to build “designer” organs that are superior to natural ones. We would not only need the cell types for that organ, but also control its shape and internal patterning. For instance, the structure of the avian lung is more efficient at gas exchange than mammalian lungs. This is due to the unidirectional flow of oxygen-rich air, in contrast to the bidirectional air flow of our lungs. Re-aggregation of human lungs or organoids are unlikely to form the structures of the avian one. The structure is encoded within its genome.
Spatial control of multicellular structures over a variety of scales will be needed (122, 123). For instance, synthetic short-range cell-cell interactions can be controlled in mammalian cells using SynNotch (124). Longer range patterning to form stripes has been achieved in bacteria (125), and similar strategies and thinking may work for mammalian cells. The ultimate goal would therefore be genetic control over self-assembly itself. We would want to program cells to spontaneously become any number of desired cell types at any desired location and pattern to build any organ - natural or synthetic. This would be a true feat for regenerative biology.
References


117 Adler, C. P. & Costabel, U. Cell number in human heart in atrophy, hypertrophy, and under the influence of cytostatics. Recent advances in studies on cardiac structure and metabolism 6, 343-355 (1975).


