A Systematic Approach to Identify Candidate Transcription Factors that Control Cell Identity

Ana C. D'Alessio,[1,7] Zi Peng Fan,[1,2,7] Katherine J. Wert,[1] Petr Baranov,[3] Malkiel A. Cohen,[1] Janmeet S. Saini,[4,5] Evan Cohick,[1] Carol Charniga,[4] Daniel Dadon,[1,6] Nancy M. Hannett,[1] Michael J. Young,[3] Sally Temple,[4] Rudolf Jaenisch,[1,6] Tong Ihn Lee,[1,*] and Richard A. Young[1,6,*]

[1]Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, MA 02142, USA
[2]Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[3]Schepens Eye Research Institute, Massachusetts Eye and Ear, Harvard Medical School, Boston, MA 02114, USA
[4]Neural Stem Cell Institute, Rensselaer, NY 12144, USA
[5]Department of Biomedical Sciences, University at Albany, SUNY, Albany, NY 12201, USA
[6]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[7]Co-first author
*Correspondence: tlee@wi.mit.edu (T.I.L.), young@wi.mit.edu (R.A.Y.)
 http://dx.doi.org/10.1016/j.stemcr.2015.09.016

SUMMARY

Hundreds of transcription factors (TFs) are expressed in each cell type, but cell identity can be induced through the activity of just a small number of core TFs. Systematic identification of these core TFs for a wide variety of cell types is currently lacking and would establish a foundation for understanding the transcriptional control of cell identity in development, disease, and cell-based therapy. Here, we describe a computational approach that generates an atlas of candidate core TFs for a broad spectrum of human cells. The potential impact of the atlas was demonstrated via cellular reprogramming efforts where candidate core TFs proved capable of converting human fibroblasts to retinal pigment epithelial-like cells. These results suggest that candidate core TFs from the atlas will prove a useful starting point for studying transcriptional control of cell identity and reprogramming in many human cell types.

INTRODUCTION

Cell identity is controlled in large part by the action of transcription factors (TFs) that recognize and bind specific sequences in the genome and regulate gene expression. While approximately half of all TFs are expressed in any one cell type (Vaquerizas et al., 2009), a small number of core TFs are thought to be sufficient to establish control of the gene expression programs that define cell identity (Buganim et al., 2013; Graf and Enver, 2009; Morris and Daley, 2013; Sancho-Martinez et al., 2012; Vierbuchen and Wernig, 2012; Yamanaka, 2012). It would be valuable to identify these core TFs for all cell types; an atlas of candidate core regulators would complement the Encyclopedia of Regulatory DNA Elements (ENCODE) (Rivera and Ren, 2013; Stergachis et al., 2013), guide exploration of the principles of transcriptional regulatory networks, enable more systematic research into the mechanistic and global functions of these key regulators of cell identity, and facilitate advances in direct reprogramming for clinically relevant cell types (Henriques et al., 2013; Iwafuchi-Doi and Zaret, 2014; Soufi et al., 2012; Xie and Ren, 2013).

Core TFs that control individual cell identity have been identified previously, but systematic efforts to do so for most cell types have been relatively rare until recently. Early efforts focused on the experimental identification of genes that were differentially expressed in one cell type, compared to a small range of other cell types, and shown to have roles in controlling specific cell identities. Examples include *MyoD1*, which can convert fibroblasts to muscle cells upon overexpression in fibroblasts (Tapscott et al., 1988), and *Oct4*, whose loss results in the loss of the pluripotent cell population in the mammalian embryo (Nichols et al., 1998). More recently, cellular reprogramming experiments, where ectopic expression of TFs converts cells from one type to another, arose as a particularly stringent test of the ability of TFs to establish cell identity (Buganim et al., 2013; Graf and Enver, 2009; Morris and Daley, 2013; Sancho-Martinez et al., 2012; Vierbuchen and Wernig, 2012; Yamanaka, 2012). While powerful demonstrations of the role of TFs in control of cell identity, these experimental approaches are necessarily focused on specific cell types.

The development of genome-scale technologies has enabled more global attempts to predict candidate factors that control cell identity. Genome-wide gene expression and epigenome analysis across multiple cell types have been used to identify candidate core factors via computational methods (Cahan et al., 2014; Heinäniemi et al., 2013; Lang et al., 2014; Morris et al., 2014; Roost et al., 2015). While broad in scope, these studies assess their predictions using more easily scalable methods and typically do not assess whether predicted factors are sufficient to establish cell identity.

CrossMark

We describe here the identification of candidate core TFs across the largest collection of different human cell types to date. A computational approach was devised to systematically identify candidate core TFs for most known human cell types. Importantly, we demonstrate with ectopic expression experiments that these predictions can identify factors capable of converting cell identity, thus satisfying a stringent criterion that is not tested with most other approaches to identify key TFs. Expression of core factors identified for retinal pigment epithelial (RPE) cells was sufficient to reprogram human fibroblasts into RPE-like cells. These cells were functionally characterized for their similarity to RPE cells derived from healthy individuals and were shown to share many features—including morphology, gene expression, and the ability to perform canonical RPE processes—and to integrate into the host RPE layer in transplantation experiments. These results suggest that the atlas of candidate core TFs should be useful for reprogramming additional clinically important cell types and for systematically discovering the regulatory circuitries for these cells.

RESULTS

To identify candidate core TFs, we searched for TFs that fit two basic characteristics of known core TFs: the genes encoding known core TFs are typically expressed in a relatively cell-type-specific fashion and at relatively high levels (Figure 1A). The algorithm we developed quantifies both cell type specificity and the relative level of gene expression by using an entropy-based measure of Jensen-Shannon divergence (Cabili et al., 2011; Fuglede and Topsoe, 2004) to compare the expression of a TF in a cell type of interest (the query dataset) to the expression of that factor across a range of cell types that represent the diversity of human cell types and tissues (the background dataset) (Supplemental Information). The idealized case is when a TF is expressed at a relatively high level in the query dataset and not expressed in any other cell type in the background dataset. The algorithm generates a specificity score for each TF based on how well the actual expression data match with this idealized case and ranks each TF using a nonparametric rank-product approach to aggregate the results from multiple query datasets for a given cell type (Breitling et al., 2004).

This approach was used to score TFs for over 200 cell types/tissues collated from the Human Body Index collection of expression data, together with additional well-studied cell types (Figure 1B; Figure 2; Table S1; Supplemental Information). The complete atlas contains the scores for all TFs in all cell types, but for simplicity of additional analyses and manageability of experimental validation, we focused on the ten top-scoring TFs in each cell type as the primary candidate core TFs.

503 different TFs were considered candidate core TFs for one or more cell types or tissues. As expected given our methodology, the candidate core TFs were expressed at higher levels than non-core TFs (Figure 3A), and individual factors were generally considered candidate core TFs in limited numbers of cell or tissue types (Figure 3B). DNA-binding domain analysis indicates that the candidate core TFs have a different distribution of DNA binding domains compared to other TFs, with relatively increased frequencies of domains frequently associated with developmental regulators (homeobox and helix-loop-helix) and relatively decreased frequencies of other domains such as SCAN, KRAB, and C2H2 zinc finger (Figure 3C). The candidate core TFs are generally well conserved, as orthologs typically exist for the factors through vertebrate and metazoan species (Figure 3D). The genes encoding candidate core TFs are generally associated with super-enhancers, transcriptional regulatory elements that are associated with genes that play important roles in cell identity (Hnisz et al., 2013; Parker et al., 2013; Whyte et al., 2013). Gene set enrichment analysis (GSEA) shows that super-enhancers of each cell type are enriched among the highest scoring TFs in the atlas, and this enrichment occurs in a cell-type-specific manner (Figure 3E). These data are consistent with the highest scoring TFs in the atlas having roles in control of cell identity. More extensive characterization and discussion of the set of candidate core TFs is included in the Supplemental Information and Figure S1.

We examined whether factors from this atlas could induce a new cell identity as a stringent test of whether the atlas successfully identifies TFs that control cell identity. Ectopic expression of core TFs in fibroblasts can reprogram gene expression and produce cells with functional states similar to those that normally express those TFs (Buganim et al., 2013; Graf, 2011; Morris and Daley, 2013; Vierbuchen and Wernig, 2012; Yamanaka, 2012). Examination of the list of candidate core TFs predicted for embryonic stem cells shows good overlap with factors already used to reprogram murine or human fibroblasts to pluripotent stem cells (Table S2). Similar results are seen for several other cell types, including cardiomyocytes and hepatocytes (Table S2), and a comparison to a set of TFs that have been used for lineage reprogramming in human cells—summarized in Xu et al. (2015)—shows that roughly 70% of these lineage reprogramming factors are called candidate core TFs in the atlas (Table S3). To test factors from this atlas, RPE cells were chosen as the target cell type due to their growing relevance to cell therapy applications. Progressive degeneration of RPE cells is a major cause of age-related macular degeneration (AMD), and several clinical trials are currently assessing transplantation of
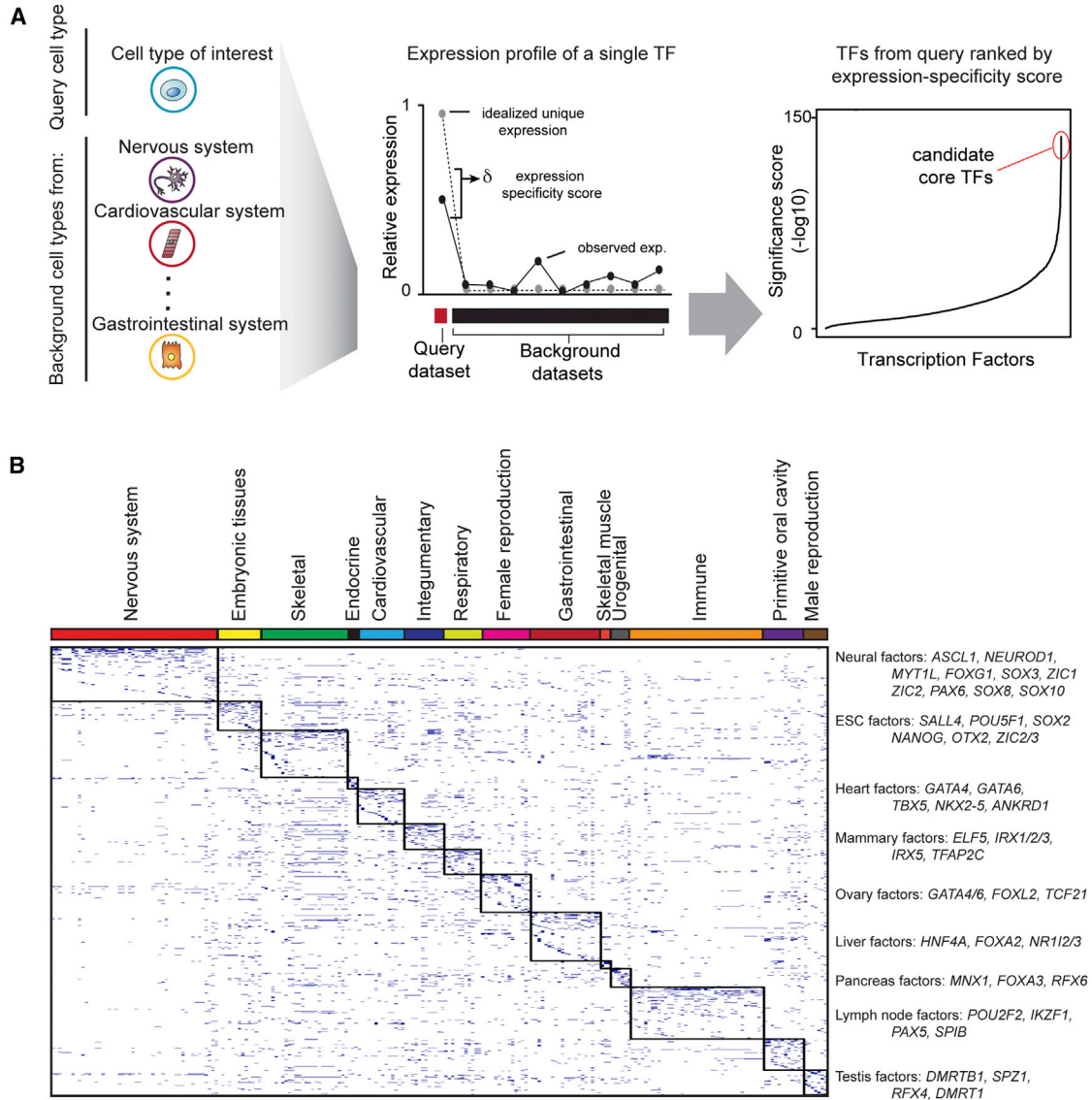
Figure 1. A General Approach to Identify Candidate Core TFs in Human Cells

(A) Computational approach used to identify candidate core TFs in human cells. Left panel: collection of gene expression profiles of a query cell type and representative cell types from the Human Body Index collection of expression data. Middle panel: expression profile of a single TF across a query dataset and a range of background datasets. The idealized case of expression level of a TF (gray circle, dashed line) is compared to the observed data to calculate the expression-specificity score of the TF. Right panel: plot depicting the distribution of significance scores of expression specificity for all TFs. Factors are arranged on the x axis in order of significance scores. Significance scores are indicated on the y axis. The highest scoring TFs are considered the best candidate core TFs and highlighted in the red circle.

(B) Representation of the collection of candidate core TFs for 233 tissue and cell types. Tissue and cell types are arranged on the x axis and ordered according to anatomical groups, represented by the colored bar at the top. Genes are arranged on the y axis. Blue dashes represent candidate core TFs in a cell type. Clusters of candidate core TFs in cell types representing an anatomical group are boxed. Representative genes are listed on the side.

RPE cells and stem-cell-derived RPE cells as a treatment for ocular disorders (Cyranoski, 2013, 2014).

For ectopic expression experiments, nine of the top-scoring RPE core TF candidates—*PAX6*, *LHX2*, *OTX2*, *SOX9*, *MITF*, *SIX3*, *ZNF92*, *GLIS3*, and *FOXD1*—were selected and cloned into doxycycline-inducible lentiviral expression vectors (Figure 4A). Human foreskin fibroblasts (HFF) were then transduced with a cocktail of all nine factors. Colonies showing a "cobblestone"-like morphology characteristic of RPE cells were evident after

Nervous system

globus pallidus
NFIA OLIG2 ST18 OLIG1 LHX4 NKX2-2 ZNF536 SOX8 NR2E1 SIX5
spinal cord
NKX2-2 RFX4 HOXB8 OLIG1 ST18 BHLHE41 SALL1 OLIG2 ASCL1 SOX2
corpus callosum
ST18 OLIG2 OLIG1 NKX2-2 C11orf9 ZNF536 SOX8 SOX10 NPAS3 SALL1
medulla
OLIG1 NKX2-2 ST18 OLIG2 POU4F1 SOX8 C11orf9 ZNF536 SALL1 BHLHE41
midbrain
OLIG1 ST18 OLIG2 C11orf9 NKX2-2 SOX8 ZNF536 SALL1 SOX10 ESRRG
vestibular nuclei superior
OLIG1 NKX2-2 OLIG2 NHLH2 PEG3 ST18 ARNT2 NPAS3 OTP SOX2
thalamus
SHOX2 OLIG1 OLIG2 OTX2 NKX2-2 NR2E1 ST18 SOX8 TCF7L2 C11orf9
hypothalamus
OLIG2 OLIG1 NR2E1 NKX2-2 PEG3 ST18 SOX8 C11orf9 ZNF536 ARNT2
ventral tegmental area
OLIG1 OLIG2 OTX2 NKX2-2 ST18 C11orf9 SOX8 POU3F2 PEG3 ASCL1
nodose nucleus
ST18 OLIG1 ST18 NKX2-2 OLIG2 C11orf9 ZNF536 SOX10 SOX8 SALL1 BHLHE41
substantia nigra
OLIG2 OLIG1 ST18 NKX2-2 C11orf9 POU3F2 SOX10 SOX8 ZNF536 NPAS3
subthalamic nucleus
OLIG1 ST18 NKX2-2 OLIG2 C11orf9 SOX8 SOX10 ZNF536 SALL1 BHLHE41
prefrontal cortex
NR2E1 LHX2 FOXG1 CSRNP3 MYT1L FEZF2 NEUROD6 RORB SATB2 LMO4
caudate
SIX3 NR2E1 MYT1L ONECUT2 ARNT2 DLX6 THRA EGR4 CSRNP3 PEG3
accumbens
SIX3 NR2E1 EGR4 MYT1L ARNT2 THRA FOXG1 PEG3 CSRNP3 ONECUT2
putamen
SIX3 EGR4 NR2E1 ONECUT2 MYT1L ARNT2 FOXG1 DACH1 DLX6 PEG3
frontal cortex
NEUROD6 RORB MYT1L LMO4 FOXG1 CSRNP3 FEZF2 TBR1 ARNT2 SATB1
cerebral cortex
NR2E1 LHX2 MYT1L RORB FOXG1 LMO4 FEZF2 ARNT2 OLIG1 OLIG2
occipital lobe
NR2E1 RORB MYT1L LHX2 FOXG1 OLIG1 CUX2 FEZF2 ARNT2 CSRNP3
temporal lobe
NR2E1 LHX2 FOXG1 MYT1L OLIG1 FEZF2 CSRNP3 ARNT2 DLX1 NEUROD6
frontal lobe
NR2E1 MYT1L LHX2 FOXG1 LMO4 NEUROD6 ARNT2 FEZF2 RORB CSRNP3
parietal lobe
NR2E1 FOXG1 LHX2 MYT1L LMO4 CSRNP3 ARNT2 FEZF2 RORB OLIG1
amygdala
NR2E1 LHX2 FOXG1 ARNT2 FEZF2 THRA MYT1L EGR4 OLIG1 OLIG2
hippocampus
NR2E1 LHX2 FOXG1 OLIG1 ARNT2 MYT1L NEUROD6 BHLHE22 OLIG2
gloubus pallidum (ext)
NR2E1 LHX8 SOX8 SIX3 ST18 SOX2 OLIG1 RFX4 ASCL1 NPAS3
gloubus pallidum (int)
SOX8 OLIG1 LHX8 ST18 OLIG2 NKX2-2 C11orf9 NR2E1 NPAS3 SIX3
substantia nigra pars compacta
EN1 TUB CSRNP3 SOX8 OLIG1 POU3F2 ST18 NPAS3 RFX4 MYT1L
substantia nigra reticulata
TUB EN1 SOX8 ST18 OLIG1 POU3F2 CSRNP3 SOX2 ZSCAN10 NKX2-2
thalamus lateral nuclei
SHOX2 OTX2 TCF7L2 LHX9 NHLH1 GZF1 CUX2 SCRT1 NR1I3 ETV1
thalamus subthalamic nucleus
OTX2 SHOX2 HES2 SP3 PEG3 MYT1L LHX9 ST18 FOXP2 SOX8
pyramidal neuron--superior frontal gyrus
CSRNP3 MYT1L LHX2 SATB1 LMO4 RORB THRB FOXG1 LZTS1 MEF2C
pyramidal neuron--singulate organ
SATB1 CSRNP3 MYT1L LHX2 RORB LMO4 THRB DLX1 FOXG1 PEG3
pyramidal neuron--midtemporal gyrus
LHX2 MYT1L MEF2C LMO4 THRB CSRNP3 DLX1 FOXG1 CUX2 RORB
pyramidal neuron--hippocampus
FEZF2 MYT1L LHX2 BHLHE22 CSRNP3 EGR4 FOXG1 LMO4 NEUROD6 E2F3
stellate neuron--entorhinal cortex
LHX2 CSRNP3 MYT1L EGR4 SATB1 LZTS1 ARNT2 FOXG1 LMO4 RFX3
pyramidal neuron--entorhinal cortex
LHX2 EGR4 CSRNP3 PKNOX2 LZTS1 RORB BCL11A LMO4 ARNT2 ZBTB33
cerebellum
ZIC1 NEUROD1 ZIC2 PAX6 EN2 ETV1 ZNF238 LHX1 MYT1 ZIC3
cerebellar vermis
NEUROD1 ZIC1 NFIA ZIC2 EN2 PAX6 ETV1 LHX1 ST18 NHLH2
cerebellar hemisphere
ZIC1 ZIC2 NEUROD1 PAX6 NFIA MYT1 EN2 ETV1 ZNF238 ZIC3
pons
RARA ESRRG ZIC1 POU4F2 ZIC2 AFF3 MYT1L PRDM2 ZNF536 ST18
fetal brain
SOX11 FOXG1 SATB2 DLX2 DLX1 LHX2 ARX NEUROD6 SOX1 MYCN
trigeminal ganglia
FOXD1 POU4F1 FOXD3 ISL2 ONECUT2 EGR2 SOX10 EBF2 BNC2 HEY1
dorsal root ganglia
POU4F1 FOXD3 ISL2 FOXD1 SHOX2 SOX10 GLI1 HEY1 EBF2 ONECUT2
Schwann cell
FOXD3 DLX2 RCAN1 RUNX1 GAS7 DLX1 FOXM1 RUNX2 MAF ARNTL
astrocyte
FOXD1 FOXL1 MYBL1 SIX3 RUNX2 E2F7 GLIS1 FOXC2 LHX2 FOXM1
definitive neural epithelium
OTX2 SP8 SOX11 HES5 SIX3 LHX2 SOX3 SOX1 FOXG1 SALL4
primary neural epithelium
OTX2 SIX3 SP8 SOX11 SALL4 LHX2 HES5 SOX2 PAX6 HESX1
neuronal progenitor
SOX2 ASCL1 SOX3 DLX2 NPAS3 SP8 HES6 SOX1 DLX1 ETV1
conjunctiva stem cell
PAX6 ZNF165 ELF3 MAFF ARNTL2 IRF1 TFAP2C GRHL1 POU2F3 FOXQ1
conjunctiva epithelium
ARNTL2 IRF6 MAFF BNC1 ATF1 EHF LRRFIP1 TFAM TFDP1 ELK3

Cardiovascular system

heart ventricle
NKX2-5 GATA4 ANKRD1 TBX5 GATA6 EBF2 HEY2 HOXB1 HAND1 IRX4
heart atrium
TBX5 NKX2-5 ANKRD1 GATA4 GATA6 STAT4 TCF21 MEOX1 EBF2 HSF1
heart
NKX2-5 ANKRD1 TBX5 KLF2 GATA4 IRX4 GATA6 HAND1 HEY2 HIC2
vena cava
ANKRD1 PURA NEUROG2 GATA5 ATF3 SCRT1 NFIX NFIL3 NR4A3 TCF15
saphenous vein
SHOX2 AEBP1 FOXL1 HOXC6 ARID5A MKX HOXC10 HOXC9 PRRX1 JUN
aorta
FOXL1 THRB FOXC2 FOXC1 OSR1 PRRX2 AEBP1 ATF3 GATA6 MECOM
coronary artery
GATA4 FOXS1 HEYL HES4 FOSL2 AEBP1 NR4A3 HEY2 TBX2 HIF3A
pericardium
NFIA BNC1 SIX5 ESR2 HIC2 NFIB SOX17 PRRX2 TSHZ2 RUNX1
coronary smooth muscle
E2F7 FOXL1 SMAD3 STAT1 FOXM1 FOXD1 FOXC2 PRRX1 TCF21 SNAI2
aortic smooth muscle
STAT1 PRRX1 SMAD3 E2F7 FOXD1 HIF1A TAF12 GLIS3 NPAS2 FOSL1
microvascular endothelium
FLI1 HHEX NFE2L3 ERG TBX1 NFATC1 EWSR1 FOXM1 GABPB1 E2F7
vein endothelium
FLI1 ERG HHEX BCL6B GABPB1 NFE2L3 SOX7 HOXA10 NFATC1 FOXC2
artery endothelium
HMGA1 MYBL2 SMAD3 FOXM1 FOSL1 TBX1 PTTG1 ELF4 TCF3
HUVEC
HMGA1 ALX1 FOXM1 FLI1 TFEC HHEX E2F7 FLI1 ELK3 ERG

Immune system

spleen
IKZF3 NKX2-3 PAX5 TCF21 TLX1 RUNX3 AKNA EOMES MAFB IKZF1
lymph nodes
IKZF3 PAX5 RUNX3 POU2F2 IKZF1 SPIB SP140 HOXC6 AKNA HCLS1
tonsil
IKZF3 PAX5 SPIB POU2F2 MEF2B IKZF1 RUNX3 SP140 BATF HCLS1
activated B cell
IRF4 SPIB PAX5 IKZF3 RUNX3 NR4A3 REL NFKB2 IRF8 IKZF1
naive B cell
PAX5 IKZF3 IRF8 SPIB BACH2 IKZF1 IRF4 E2F5 SP140 RUNX3
germinal centre centrocyte
MEF2B MYBL1 PAX5 IKZF3 E2F5 MYBL2 BACH2 IRF8 POU2F2 SPIB
germinal center centroblast
MEF2B PAX5 MYBL1 MYBL2 IKZF3 E2F5 BACH2 SPIB ETS1 IRF8
activated CD4+ T cell
IKZF1 ETS1 STAT4 GATA3 GFI1 LEF1 IKZF3 SATB1 RUNX3 FLI1
naive CD4+ T cell
IKZF1 ETS1 RUNX3 IKZF3 LEF1 ZBTB25 MYBL1 CREM AKNA SATB1
activated CD8+ T cell
MYPN EOMES RUNX3 IKZF1 NFATC2 E2F6 IKZF3 MLLT10 SP140 HSFY1
regulatory T cell
IKZF1 ETS1 GATA3 LEF1 ZBTB25 IKZF2 RUNX3 RORA ZNF639 FLI1
effector memory CD4+ T cell
IKZF3 IKZF1 ETS1 MYBL1 GATA3 RUNX2 RUNX3 ZFP36L2 TCF7 FLI1
memory CD4+ T cell
IKZF3 ETS1 TCF7 IKZF1 ZFP36L2 LEF1 RUNX3 GATA3 MAF AKNA
natural killer cell
MYBL1 RUNX3 EOMES IKZF3 STAT4 IKZF1 ETS1 TBX21 CREM AKNA
hematopoietic stem cell
MYB TFEC RUNX1 IKZF1 NFE2 HOXA9 FLI1 RUNX2 BCL11A GFI1
myeloid progenitor
NFE2 TFEC MXD1 FLI1 IKZF1 MYNN HCLS1 CEBPD MAFB LRRFIP1
memory B cell
PAX5 SP140 IKZF3 IRF8 IKZF1 ZBTB32 SPIB BCL11A RUNX3 MYNN
thymus gland
TCF7 LEF1 SATB1 IKZF1 ETS1 MYB GFI1 IKZF2 NFATC3 TFDP2
neutrophil
MXD1 NFE2 SPI1 IRF1 HCLS1 IKZF1 TSC22D3 AKNA BACH1 BCL6
activated M2 macrophage
TFEC EGR2 MAFB ZNF267 NR1H3 IRF8 BACH1 ELF4 MSC HCLS1
M2 macrophage
TFEC EGR2 MAFB HCLS1 NR1H3 SPI1 IKZF1 LRRFIP1 IRF8 ZNF267
M1 macrophage
TFEC STAT1 EGR2 IRF1 ASCL2 SPI1 ZNF267 NR1H3 IRF8 BATF
activated myeloid dendritic
TFEC NR4A3 IRF4 MSC ETV3 BATF3 EGR2 NFKB1 ARNTL2 ZNF267
myeloid dendritic
TFEC IKZF2 SPI1 SNAI3 IKZF1 LRRFIP1 HCLS1 ELF4 ETV6 IRF8
monocyte from M1 macrophage
EGR2 SNAI3 SPI1 TFEC VDR BATF3 ETV6 HCLS1 ELF4 MAFF
alveolar macrophage
TFEC PPARG EGR2 RUNX1 BHLHE41 IRF8 BACH1 ZEB2 IRF5 HCLS1
alveolar macrophage (smoker)
TFEC EGR2 PPARG BHLHE41 NR1H3 SPI1 IRF8 RUNX1 BACH1 MAFB
monocyte
TFEC ZEB2 BACH1 MAFB SPI1 HCLS1 ELF4 NFKB1 EGR2 ZNF267
monocyte from dendritic differentiation
TFEC NFKB1 MAFG SNAPC1 BACH1 IRF8 RNF141 IRF1 IRF7 ETV3
monocyte (CD16+)
SPI1 IKZF1 FLI1 HCLS1 POU2F2 IRF5 EWSR1 ELF4 MAFB TCF7L2
erythroblast
NFE2 GFI1B KLF1 TAL1 E2F2 TFDP1 PTTG1 MYBL2 GATA1 E2F8
erythroid progenitor
GFI1B HES6 NFE2 KLF1 GATA1 TFDP1 MYB PTTG1 TAL1 E2F8
erythrocyte
GFI1B NFE2 HES6 TSC22D3 KLF1 ZNF23 ZNF165 CEBPG DDIT3 PTTG1
activated plasmacytoid dendritic
TFEC NR4A3 CREM NFKB1 RUNX3 SOX4 FOXP1 IKZF1 IRF7 STAT4
promyelocyte
CEBPE GFI1 MYB RNF141 NFE2 IKZF1 HCLS1 HLX SPI1 HMGB2
lymphatic vessel endothelium
TBX1 ESX1 FLI1 FOXM1 HHEX HMGA1 ERG E2F7 NFE2L3 HOXA1
naive CD8+ T cell
MYBL1 IKZF1 ETS1 RUNX3 IKZF3 EOMES LEF1 STAT4 FLI1 GATA3
natural killer T cell
MYNN RUNX3 IKZF1 MYBL1 ZNF33B SHOX MLLT10 STAT1 NFRKB MAF
plasma B cell
PAX5 MYNN TCF3 SP140 E2F6 VDR ZNF215 NFRKB IKZF3 POU2F2
platelet
ZEB2 TAL1 ZNF367 GFI1B MAX IKZF1 SMAD2 MYNN FLI1 MEIS1

Embryonic tissues

mesenchymal stem cell
FOXD1 ALX1 SNAI2 STAT1 HOXA9 RCAN1 RUNX1 MYBL1 HOXA1 ZNF281
mesenchymal stem cell (fetal)
FOXD1 SHOX LHX9 SMAD3 SIX1 RCAN1 AFF3 SNAI2 FOXM1 NR2F6
mesenchymal stem cell (diff)
DLX1 DLX2 FOXD1 E2F7 RCAN1 SIX4 BACH1 ZNF85 SIX1 ZNF281
germ line stem cell
LHX9 RUNX1 MSC PRRX1 ZNF281 SNAI2 GLIS3 E2F7 GATA6 MAFG
trophoblast
BNC1 E2F7 PITX2 RUNX2 NFE2L3 GABPB1 FOXF1 TAF12 HIF1A BACH1
amniocyte
MYBL2 FOXM1 PTTG1 NPAS2 E2F7 HOXA1 TFDP1 ILF2 FOXD1 ELF4
neuronal progenitor
E2F7 FOXL1 ANKRD1 DDIT3 HOXB6 HOXA1 SMAD3 FOXM1 HOXB7 SNAPC1
iPS
ZSCAN10 NANOG OTX2 POU5F1 ZFP42 ZIC3 SALL4 MYCN FOXH1 PRDM14
ESC
SALL4 OTX2 ZIC3 NANOG ZSCAN10 POU5F1 MYCN NR6A1 ZIC2 SOX2
embryonic body differentiation D2
OTX2 HESX1 ZFP42 SALL4 SIX3 ZIC3 ZSCAN10 NR6A1 SOX2 SOX11
NPC
OTX2 SALL4 SIX3 LHX2 SP8 SOX11 TCF3 PAX6 ZIC2 NR6A1
embryonic body differentiation D1
HAND1 ZFP42 PITX2 NR6A1 FOXM1 HOXA1 RFX6 SALL4 SP8 HOXB6
embryonic body differentiation D3
HAND1 SALL4 MSX2 ZFP42 ZSCAN10 HOXB6 FOXH1 NANOG NR6A1 POU5F1

Primitive oral cavity

pharynx epithelium
FOXA1 RUNX1 IKZF3 ELF3 FOXE1 NFE2L3 SPIB PAX5 KLF5 IKZF1
nasal epithelium
SIX3 FOXJ1 EPAS1 ELF3 NR2F6 DMRT3 FOXA1 STAT1 EHF PAX7
gingival epithelium (immortalized)
FOSL1 BNC1 IRF6 ZNFX1 STAT1 TP63 E2F7 HES6 IRX4 SOX15
pharyngeal mucosa
GRHL3 PITX1 TRIM29 GRHL1 HOPX MXD1 LHX8 EHF OVOL1 KLF5
oral mucosa
PITX1 TRIM29 GRHL3 MXD1 ARNTL2 PITX2 KLF5 HOPX LHX8 OVOL1
gingival epithelium (primary)
FOXE1 BNC1 PITX1 MAF FOXN1 TFAP2A PITX2 EHF GRHL1 GRHL3
oral epithelium
GRHL3 KLF8 ARNTL2 GRHL1 IKZF2 TP63 BNC1 MXD1 TFAP2C PITX2
tongue main corpus
TBX1 MYF6 PITX2 CSDA GRHL3 SIX1 MYOD1 DMRT2 MYF5 HOPX
tongue
TBX1 MYF6 CSDA PITX2 NEUROG2 SIX4 TEAD4 BARX2 PITX1 TBX15
tongue superior part papillae
EHF SIM2 PITX1 PAX1 FOXP2 MYOD1 PRDM1 MYF6 MEF2C GRHL3
salivary gland
ELF5 TFCP2L1 FOXC1 SIX1 ASCL2 ETV1 IRX2 EHF TLX1 SPDEF
odontoblast
LHX8 SP7 MSX1 RUNX2 DLX6 IRX5 DLX1 DLX3 DLX2 MSX2

Gastrointestinal system

hepatocyte
NR1H4 NR1I2 HNF4A NR0B2 FOXA2 CUX2 HHEX ZGPAT NR1I3 ATF5
liver
NR1I2 ATF5 HNF4A NR1H4 NR1I3 ONECUT2 PROX1 FOXA2 ZGPAT MLXIPL
colon epithelium
CDX1 HOXB13 ELF3 ISX HNF4A HNF4G VDR ZNF165 CDX2 SATB2
ileum epithelium
HNF4G NR1H4 CDX1 NR1I2 CREB3L3 ISX CDX2 NR5A2 HNF4A VDR
differentiated hepatocyte
HAND1 ZFP42 PITX2 MSX2 SALL4 BNC1 ALX1 SP8 SNAI2 GATA6
fetal liver
NFE2 GFI1B TAL1 KLF1 TFDP1 MYBL2 FOXM1 GATA1 E2F2 HNF4A
hepatic stellate cell
FOXD1 HOXA3 E2F7 GLIS3 GZF1 TAF13 ZNF281 BLZF1 SNAI2 TFDP1
colon fibroblast
TCF21 FOXF1 FOXD1 MAFG GLIS3 NPAS2 ISL2 RCAN1 MSC
esophagus epithelium
BNC1 TP63 SNAI2 HOXA1 FOXM1 E2F7 FOXD1 IRF6 SOX15 PTTG1
esophagus
PITX1 GRHL3 GRHL1 TRIM29 ARNTL2 MXD1 HOXA1 OVOL1 EHF PAX9
peritneum
WT1 PGR EMX2 GLIS3 ESR1 BNC2 POU5F1 SNAI2 MAF AR
small intestine
NKX2-3 HNF4G ISX CDX1 MNX1 VDR ASCL2 CREB3L3 ONECUT2NR1I2
colon
CDX1 CDX2 NKX2-3 MLXIP ATOH1 VDR ISX HOXB3 HAND1 FOXF1
colon cecum
NKX2-3 MECOM CDX1 TCF21 ZEB2 FOXF1 FOXF2 HNF4G HOXB6 SPIB
stomach pyloric
FOXL1 THRB FOXC2 FOXC1 OSR1 PRRX2 AEBP1 ATF3 GATA6 MECOM
stomach fundus
BARX1 FOXA2 FOXF1 MECOM ISL1 RFX6 HOXA5 FOXF2 NKX2-3 GATA6
stomach cardiac
FOXA2 BARX1 ISL1 RFX6 MECOM ISL2 GATA5 FOXF1 NKX2-3 IKZF3
stomach
BARX1 FOXQ1 FOXF1 FOXA3 FOXA2 GATA6 ISL1 PRDM1 MLXIP TLX2
small intestine duodenum
NR5A2 ONECUT2 RBPJL HES2 GSC2 HOXC13 MNX1 GATAD2A NR2F6 DEAF1
pancreas
NR5A2 RBPJL FOXA3 MNX1 CC2D1A MYCN RFX6 ONECUT1 ZSCAN2 ZBTB7A
small intestine jejunum
MYOD1 MYF6 CSDA MYOG SIX1 SIM1 PITX2 HSF2 MEF2C CDX4

Skeletal system

osteoblast (msc D21)
RUNX2 SNAI2 ZFHX4 FOXD1 ZEB1 HOXA1 PITX2 HIF1A TWIST1 HOXC10
white adipocyte (msc D21)
PPARG NR1H3 SNAI2 CEBPA ALX1 RUNX2 HOXA10 TAF12 GLIS3 STAT1
differentiated myofibroblast
GLIS3 MAFG SNAI2 MKX SIX2 SHOX TWIST1 FOXD1 HOXC6 STAT1
osteoblast
SNAI2 FOXD1 ATF1 STAT1 BACH1 SHOX RCAN1 RUNX1 E2F7 GLIS3
chondrocyte
SNAI2 FOXD1 MKX STAT1 SNAPC1 MYBL1 SIX1 GLIS3 BARX1 SIX2
chondrocyte (msc D21)
SNAI2 FOXD1 RUNX2 TWIST1 DLX5 TSC22D3 ZEB1 GLIS3 ZNF85 RUNX1
chondrocyte (msc D7)
SNAI2 FOXD1 RUNX2 TSC22D3 TWIST1 ELK3 GLIS3 ZFP36L2 ZEB1 ATF1
chondrocyte (msc D3)
SNAI2 GLIS3 FOXD1 TSC22D3 TWIST1 RUNX2 NFIL3 MAFF SNAPC1 HOXC10
chondrocyte (msc D1)
SNAI2 GLIS3 FOXD1 TSC22D3 RUNX2 NFIL3 SNAPC1 SNAI1 TWIST1 MAFF
differentiated fibroblast
FOXD1 FOXL1 DLX2 STAT1 MYBL1 E2F7 SIX1 SMAD3 AFF3 HMGA1
fetal fibroblast
TWIST2 SIM1 HOXA13 E2F7 TWIST1 FOXM1 BNC1 SHOX2 FOXD1 SNAI2
lipoma mesenchymal stem cell
RUNX2 FOXL1 FOXM1 SNAI2 PTTG1 HMGA1 ETV1 E2F7 TWIST2 MYBL1
adipocyte mesenchymal stem cell
E2F7 FOXM1 RUNX2 FOXL1 MYBL1 ZEB1 PTTG1 TWIST1 TWIST2 SNAI2
osteoblast (msc D7)
RUNX2 SNAI2 PITX2 FOXD1 ZNF85 HIF1A GZF1 TWIST1 STAT1 ZEB1
osteoblast (msc D3)
RUNX2 SNAI2 TBX1 PITX2 GZF1 FOXD1 HIF1A TWIST1 ZFY STAT1
osteoblast (msc D1)
RUNX2 FOXD1 SNAI2 PITX2 LHX6 HIF1A HOXA1 TRIM22 ZFY STAT1
adipose subcutaneous
DMRT3 EBF1 ZNF117 PPARG EBF3 DMRT2 HOXC6 TWIST2 HOXB7 PRRX1
adipose omental
PPARG EBF2 EBF1 TSHZ2 ZNF117 MAF RUNX1 MAFB CEBPA NFIB
synovial membrane
CREB5 EBF1 SHOX2 MEOX2 HOXC6 HOXC10 SOX5 EBF2 MAFB TSHZ2
joint tissue synovium
CREB5 EBF1 DLX4 SHOX2 MAFB ZFY TBX15 DLX3 NR1D1 EBF3
melanocyte
PAX3 ALX1 TFAP2A MITF E2F7 SNAI2 LZTS1 ZFY TCFL5 PKNOX2
foreskin epithelium
BNC1 ARNTL2 HOXA1 TP63 MBD1 DLX2 SOX15 IRF6 HOXA9 ZNF165
skin epithelial keratinocyte
ELF5 RORA BNC1 NFE2L3 OVOL1 GRHL1 MBD1 TRIM29 GATA3 GRHL3
skin epithelium immortalized
IRF6 BNC1 HOXA1 TP63 HOXC13 FOXM1 GABPB1 TP53 MYBL2 SMAD3
adipose tissue
PPARG CEBPA SHOX2 EBF3 DMRT3 HOXC10 EBF1 TWIST2 SOX18 ZSCAN4
bone marrow
NFE2 GFI1 GFI1B KLF1 MYB GATA1 SPI1 TAL1 IKZF1 HMGB2

Skeletal muscle

skeletal muscle
TAL2 SIM1 CSDA SIX1 MYF6 MYOD1 LMO1 PITX2 MEF2C TBX1
deltoid muscle
SIX4 MYF6 MEF2C HOXA10 ATF1 TP63 E2F8 SIX1 CNBP CEBPG
skeletal muscle superior quadricep
MYF6 CSDA MYOD1 SNAI3 SIX1 MEF2C HNF4G TAL2 ARX TEAD4

Urogenital system

urothelium
IRF6 SMAD3 BNC1 TP63 HES2 HOXA1 EHF STAT1 HMGA1 ZFP36L2
cumulus
NR5A2 E2F7 RUNX2 GATA6 HIF1A CBFB RUNX1 TSHZ2 PPARG GZF1
urethra
HOXD13 HOXA13 FOXF2 HOXD10 ISL2 ISL1 HOXD11 PRRX1 PRRX2 GLIS1
kidney medulla
SIM1 PAX2 PAX8 TFCP2L1 EMX1 ESRRB TFAP2B HNF1B HOXD3 HOXB6
kidney cortex
PAX2 TFCP2L1 EMX1 HNF4A PAX8 SIM1 HNF1B SLC2A4RG TCF21 DMRT2
kidney
EMX1 HNF4A PAX2 HNF1B SLC2A4RG ESRRB PAX8 TCF21 VDR DMRT2

(figure continued and legend on next page)

Figure 2. *Continued*

Figure 2. Candidate Core TFs for 233 Tissue and Cell Types

Tissue and cell types were grouped into categories corresponding to different anatomical systems in the human body. Within each category, tissue and cell types were ordered using hierarchical clustering. The distance matrix was calculated by first rank-ordering the specificity scores for all TFs in each tissue and cell type within a category and then finding the Kendall tau correlation coefficient for each pairwise comparison of tissue and cell types within the category. For each individual tissue or cell type, the ten top-scoring candidate core TFs are listed.

Respiratory system

trachea
FOXJ1 ELF3 EHF NKX3-1 FOXA1 SPDEF ELF5 XBP1 JUN TP73
bronchus
FOXJ1 NKX3-1 ELF3 EHF SPDEF FOSB FOXA1 RUNX1 ATF3 NR4A1
lung
NKX2-1 TCF21 MESP2 SMAD6 EPAS1 SPI1 FOXF1 TBX2 TBX5 FOSB
airway epithelium (immortalized)
BNC1 FOXM1 E2F7 MYBL2 PTTG1 IRX2 ZNF367 TFDP1 IRX1 FOXD1
airway epithelium (cultured)
EHF TP63 IRF6 BNC1 NR2F6 SMAD3 HOXA1 ARNTL2 HMGA1 FOXA1
small airway epithelium
FOXJ1 DMRT3 GLIS3 MYB FOXA1 SIX4 RFX3 ZNF396 SOX2 RFX2
large airway epithelium
DMRT3 FOXJ1 SIX4 GLIS3 FOXA1 SOX2 MYB RFX3 ELF3 ZNF396
trachea epithelium
DMRT3 FOXA1 FOXJ1 GLIS3 EHF ELF3 MYB RUNX1 SOX2 SIX4
lung fibroblast (primary)
FOXD1 FOXL1 STAT1 FOXF1 E2F7 HOXB6 TCF21 SMAD3 SNAI2 HOXB7
lung fibroblast (cultured)
FOXL1 TCF21 SMAD3 FOXD1 HOXB6 FOXF1 TBX5 MYBL1 PRRX2 ZEB1
pulmonary smooth muscle
FOXD1 ZEB1 TCF21 RORB ZNF277 SMAD3 RBPJ MECOM ALX1 PRRX2

Endocrine system

thyroid gland
PAX8 FOXE1 NKX2-1 HHEX PBX4 TFCP2L1 SALL4 GLIS3 HIRA TBX3
pituitary gland
POU1F1 SIX6 SIX3 TBX19 INSM1 ASCL1 LHX3 PEG3 MYT1L HES6
adrenal gland cortex
TBX3 HOXA5 NR0B1 GATA6 NR1H4 CREM RORA ETV5 HOXC6 PEG3

Integumentary system

skin
POU2F3 GATA3 RORA TRIM29 HOPX FOXN1 GRHL3 TP63 GLI4 TFAP2C
nipple
POU2F3 GATA3 IRX3 HR TP63 DLX3 RORA TFAP2A TFAP2C IRX5
ductal breast epithelium
RARA GATA3 ANKRD30A IRX2 IRX1 TFAP2C ELF5 EGR2 TFAP2B TRPS1
breast
ANKRD30A GATA3 TFAP2B IRX5 MYB FOXA1 TFAP2A IRX3 EGR2 TRPS1
mammary gland
ANKRD30A ELF5 IRX1 GATA3 IRX2 IRX5 LHX9 AFF3 IRX3 TFAP2C
skin fibroblast primary
TWIST2 MKX FOXD1 PRRX2 SHOX2 STAT1 SNAI2 RUNX1 PRRX1 TWIST1
skin fibroblast cultured
TWIST2 MKX FOXD1 PRRX2 SHOX2 STAT1 SNAI2 RUNX1 PRRX1 TWIST1
foreskin fibroblast
MKX SIM1 SHOX2 TBX5 SHOX TWIST1 TWIST2 RUNX1 HOXA1 MYBL1
foot fibroblast
STAT1 MKX E2F7 HOXA13 TWIST2 FOXL1 GLIS1 PRRX1 SNAI2 PRRX2
breast epithelium immortalized
BNC1 IRF6 HOXA1 TP63 IRX4 IRX2 SNAI2 MYC TFAP2A ZNF277
breast epithelium (cultured)
BNC1 IRF6 TP63 HOXA1 ETS1 HOXC13 HES2 DLX2 RUNX1 VDR
epithelial stem cell from mammosphere
RUNX1 SOX4 EGR2 BHLHE41 VDR IRX1 ZNF193 HOXA1 GLIS3 LZTS1

Female reproduction

vulva
TP63 IRX4 SOX7 GRHL1 POU2F3 TRIM29 GRHL3 TFAP2C FOXN1 KLF5
placenta
GCM1 TFAP2A GATA3 ZNF117 HOPX ZSCAN4 HES2 ZFAT AHR TWIST1
vagina
HOXD13 HOXA13 ISL1 ESR1 ISL2 TP63 AR EHF SNAI2 HOXD10
cervix
HOXA13 ESR1 HOXD13 OSR2 PGRAR EMX2 PBX1 SNAI2 ISL1
retrocervical infiltrate
RORB WT1 HOXA13 PGR HOXA11 OSR2 HOXA10 TSHZ2 NFATC2 TCF21
myometrium
WT1 EMX2 HOXD10 PGR HOXA11 HOXA10 BNC2 ESR1 OSR2 PBX1
endometrium
PGR DLX6 ESR1 MSX1 WT1 FOXL2 HOXA11 DLX5 MSX2 HOXA10
ovary
WT1 FOXL2 ARX TCF21 PEG3 GATA4 SNAI2 GATA6 HOXC6 BNC2
ovary epithelium
BNC1 WT1 LHX9 GATA6 ARX HOXC6HOXD3 HOXD8 LHX2 TCF21
urethra
HOXD13 HOXA13 FOXF2 HOXD10 ISL2 ISL1 HOXD11 PRRX1 PRRX2 GLIS1
fallopian tube
WT1 EMX2 FOXJ1 PGR ESR1 GLIS3 MEIS1 HOXA5 HOXD3 AR
uterus
EMX2 WT1 HOXA11 OSR2 HOXD10 HOXA10 PGR ESR1 HOXD9 BNC2
ovary epithelium (immortalized)
E2F7 FOXD1 FOXM1 STAT1 ETS1 FOXL1 HMGA1 SNAI2 FOSL1 PTTG1
oocyte (pooled)
OTX2 FOXR1 SOHLH2 FIGLA LHX8 SOX30 ZNF93 NR2E1 DMRTB1 HES6
oocyte
SOX30 ZNF93 TAF5 TAF4B SOHLH2 FOXR1 FIGLA OTX2 PTTG1 ZNF215

Male reproduction

prostate stromal cell
MKX HOXA13 FOXF1 MAFF FOXL1 SIX4 TCF21 CREM RCAN1 DDIT3
prostate gland
HOXB13 NKX3-1 FOXA1 HOXA13 SP8 CREB3L4 ELK4 SPDEF CUX2 MESP1
penis
SOX7 ISL1 POU2F3 HOPX ISL2 RORA TP63 GRHL1 SIM1 TRIM29
prostate epithelium
BNC1 E2F7 HOXA1 SNAI2 HOXA10 PTTG1 ZNF131 ZNF85 TFAP2A ZNF367
spermatogonial
RHOXF2 SOX30 DMRT1 CTCFL DMRTC2 TFDP3 BNC1 TAF5 MYT1L HSF2
testes
CREM DMRTB1 SPZ1 HSF5 RHOXF2 DMRT1 RFX4 ZBTB32 YBX2 DMRTC2
spermatozoal
SPZ1 AFF4 LHX2 HSFY1 HSF5 CREM ELK4 ZFY ZNF165 ZNF24

---

2 weeks of doxycycline induction (Figure 4A). Colonies were manually picked and further expanded into six independent RPE-like cell lines. Genotyping analysis showed that all six cell lines contained the *PAX6*, *OTX2*, *MITF*, *SIX3*, *GLIS3*, and *FOXD1* expression constructs (Figure 4B).

Two of the induced RPE (iRPE)-like cell lines, iRPE-1 and iRPE-2, were subjected to additional analysis. The iRPE cell lines exhibited characteristic expression of membrane-associated *TJP1* (*ZO-1*), together with a "cobblestone" sheet morphology involving individual cells connected by tight junctions (Figure 4C), and maintained an RPE-like morphology in the presence of doxycycline for over 6 months (12 passages). Additionally, immunostaining indicates that the iRPE cells showed co-expression of *CRALBP* and *RPE65* (Figure 4D), two well-known markers for RPE cells (Sparrow et al., 2010). Expression analysis shows that the candidate core TFs are expressed in both iRPE lines, and genes considered part of the RPE gene expression signature (Strunnikova et al., 2010) show substantial upregulation compared to fibroblasts (Table S4). Principal-component analysis (PCA) of genome-wide gene expression revealed that the two iRPE lines were as similar to primary RPE cells and induced pluripotent stem-cell-derived RPE cells as induced pluripotent stem cells are to embryonic stem cells (Figure 4E). Analysis of the genes differentially expressed between iRPE and fibroblasts shows that differentially expressed genes are enriched for genes considered part of the RPE gene expression signature (Figure 4F) (Strunnikova et al., 2010).

Ectopic expression of the RPE candidate core TFs results in cells that are functionally similar to RPE cells. RPE play crucial roles in the maintenance and function of retinal photoreceptors, including phagocytosis of shed outer segments of photoreceptors (Bok, 1993), transepithelial
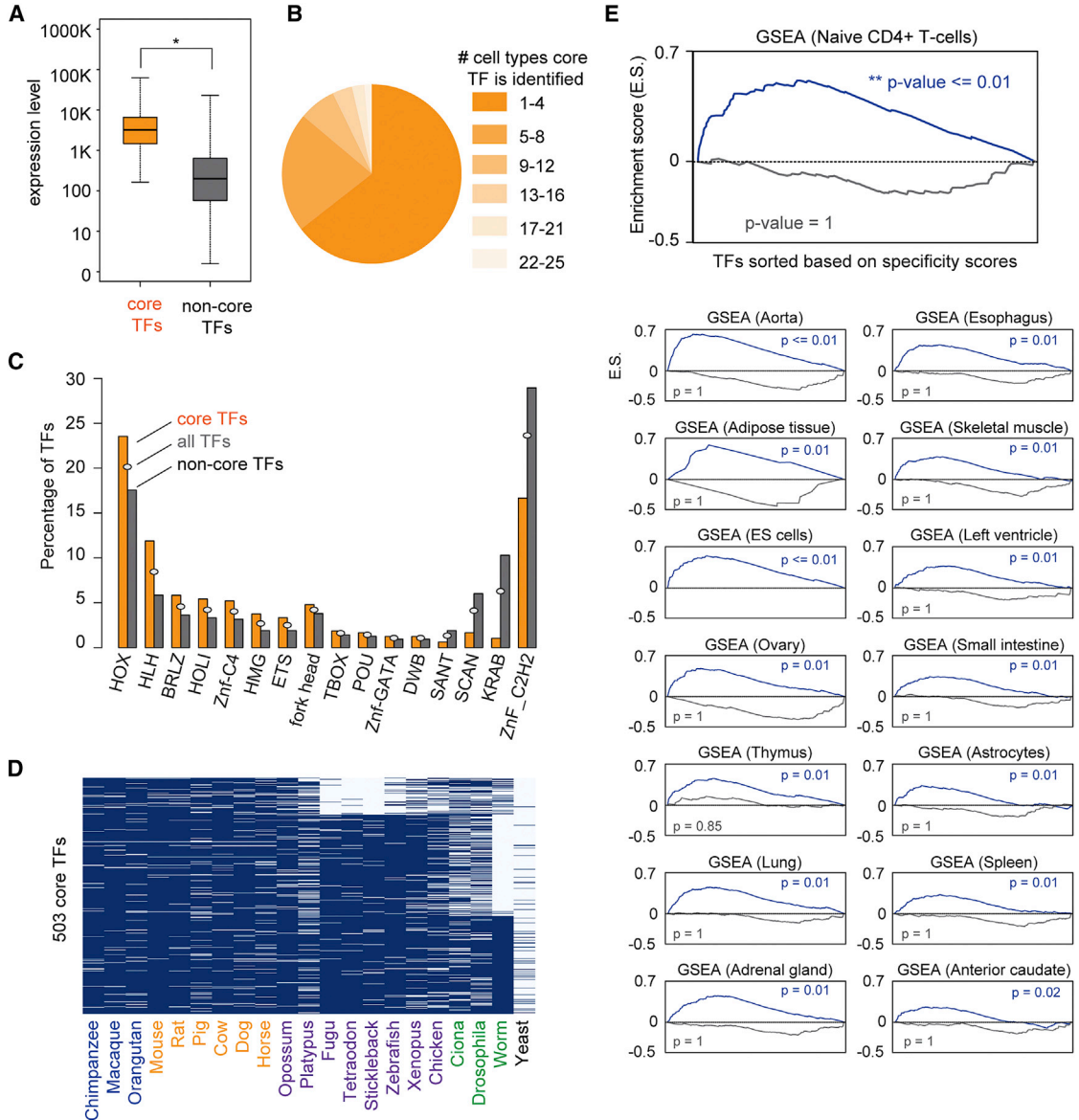
Figure 3. Characterization of Candidate Core TFs

(A) Box plots depicting the expression levels of candidate core TFs and non-core TFs. The significance of the difference between two groups was determined using a two-tailed Mann-Whitney test. For each plot, the top and bottom box edges mark the first and third quartiles, while the solid black line within the box marks the median. The top whisker line marks the largest data point that is within 1.5-fold of the interquartile range from the third quartile. The bottom whisker line marks the smallest data point that is within 1.5-fold of the inter-quartile range from the first quartile. Candidate core TFs are shown in gold. Non-core-TFs are shown in gray.

(B) Pie chart depicting the number of cell types in which a TF is considered as a candidate core TF.

(C) Bar chart representing the percentage of candidate core TFs and non-core TFs that are associated with different classes of DNA binding domains. The significance of the difference in distribution between candidate core TFs and non-core TFs across these categories is p < 0.003 and was determined using a chi-square test. The gray oval indicates the percentage of all TFs that are associated with the class of DNA binding domains as a point of comparison. Abbreviations for protein domains are: HOX, homeodomain; HLH, helix-loop-helix; BRLZ, basic region leucine zipper; HOLI, ligand binding domain of hormone receptor; ZnF_C4, c4 zinc finger in nuclear hormone receptors; HMG, high mobility group; ETS, erythroblast transformation specific; FH, forkhead; TBOX, T-box; POU, Pit-Oct-Unc; ZnF_GATA, zinc finger binding to DNA consensus sequence [AT]GATA[AG]; DWB, domain B in dwarfin family proteins; SANT, SWI3-ADA2-N-CoR-TFIIB DNA-binding domain; SCAN, SCAN domain; KRAB, Krueppel-associated box; ZnF_C2H2, zinc finger C2H2.

(legend continued on next page)

transport of nutrients and ions between the neural retina and the blood vessels (Strauss, 2005), and secretion of growth factors and hormones (Ford et al., 2011). For assaying phagocytosis, mouse rod outer segments (ROS) were incubated with iRPE cells or HFF cells. ROS incorporation was measured using an antibody against rhodopsin, which specifically recognizes a component of ROS. Both iRPE cell lines stained positive for rhodopsin, indicating binding and incorporation of ROS into the iRPE cells by phagocytosis (Figure 5A; Figure S2). To measure ion transport barrier function, we analyzed transepithelial electrical resistance (TER), which detects functional tight junctions (Stevenson et al., 1986). iRPE cells demonstrated effective barrier function that was significantly higher than fibroblasts and was as effective as that observed for RPE cells (Figure 5B). To evaluate secretion of growth factors, iRPE cells were examined for production of vascular endothelial growth factor (VEGF), which is released preferentially to the basolateral side of RPE cells to prevent endothelial cell apoptosis in the blood vessels (Saint-Geniez et al., 2009). No VEGF release was detected when fibroblasts were assayed (Figure 5C). The iRPE lines exhibit polarized secretion of VEGF similar to that produced by RPE cells (Figure 5C). Subretinal transplantation experiments showed that iRPE cells survive in vivo when transplanted in albino rats, and some integrate to the host RPE layer as pigmented cells (Figure 5D). Taken together, these results provide the most extensive characterization of iRPE to date and indicate that cells generated with the factors from our atlas are similar to RPE cells in terms of morphology, gene expression, and functionality.

DISCUSSION

The atlas of candidate core TFs presented here (Supplemental Information) provides a powerful starting point for studies of transcriptional regulation of cell identity and eventual applications for therapeutic purposes. The atlas itself is easily expanded with additional genome-wide expression data, which are relatively easy to obtain compared to other data types, especially for cell types that may be available in limited quantities. The approach is easy to implement and can be adapted to next-generation sequencing data as sufficient numbers and variety of datasets become available and may, thus, be generally useful for a wide range of researchers. The approach presented here capitalizes on basic principles of the expression level of known core TFs: relatively high expression and relatively cell-type-specific expression. Additional principles commonly associated with core TFs, such as autoregulation, binding in regulatory regions, or motif enrichment in regulatory regions may be useful to integrate in future versions.

The iRPE cells here represent the results of a stringent test for whether our approach successfully identifies TFs that can control cell identity. The factors here differ from, but overlap with, a set of factors previously used for RPE reprogramming (Zhang et al., 2014). Significantly, known oncogenic TFs such as *MYC* and signaling molecules such as activin A or retinoid acid together with Sonic hedgehog (SHH) were components of previous factor cocktails but are not required here. The iRPE cells generated here were characterized for morphology and gene expression and found to be largely similar to RPE. Moreover, the iRPE cells generated here were shown to have functional similarity to RPE cells and, thus, represent a step forward in the characterization of iRPE cells and our understanding of their potential use. These cells require continued expression of the transgenes, as withdrawal of doxycycline causes the cells to revert back toward a fibroblast morphology. This dependency is similarly observed for many other transdifferentiated cells (Buganim et al., 2012; Huang et al., 2011; Lujan et al., 2012; Sheng et al., 2012; Vierbuchen et al., 2010) and indicates that establishment of a fully self-sustaining RPE identity will require additional study. We anticipate that analysis of additional factors from our ranked list, as well as analysis of additional transdifferentiated

(D) Heatmap depicting the presence (blue) or absence (white) of orthologous genes in a species for each candidate core TF. The candidate core TFs are arranged as rows, and species are shown as columns. Species labels are colored using the following scheme: blue (primate), orange (mammal), purple (vertebrates), green (metazoa), and black (eukaryote). In the image, rows are clustered according to k-means clustering (k = 3).

(E) GSEA enrichment plots depicting the relationship between super-enhancer associated genes and high expression-specificity scores. Top panel: GSEA plot for genes associated with super-enhancers in CD4+ naive T cells and expression-specificity score. Enrichment score is plotted on the y axis. The x axis represents genes ordered by specificity score. The relationship when ordered by the expression specificity scores from CD4+ naive T cells is shown in blue. The relationship when ordered by the expression specificity scores from a non-matching cell type (embryonic stem cells) is shown in gray for comparison. p values for each are shown. Subsequent panels show similar relationships in different cell types. For each panel, the cell type is indicated. Super-enhancer associated genes are from that cell type. Blue curves represent the relationship when ordered by expression-specificity scores for that cell type. Gray curves represent the relationship when ordered by expression-specificity scores for a non-matching cell type (embryonic stem cells). p values for each are shown. E.S., enrichment score.
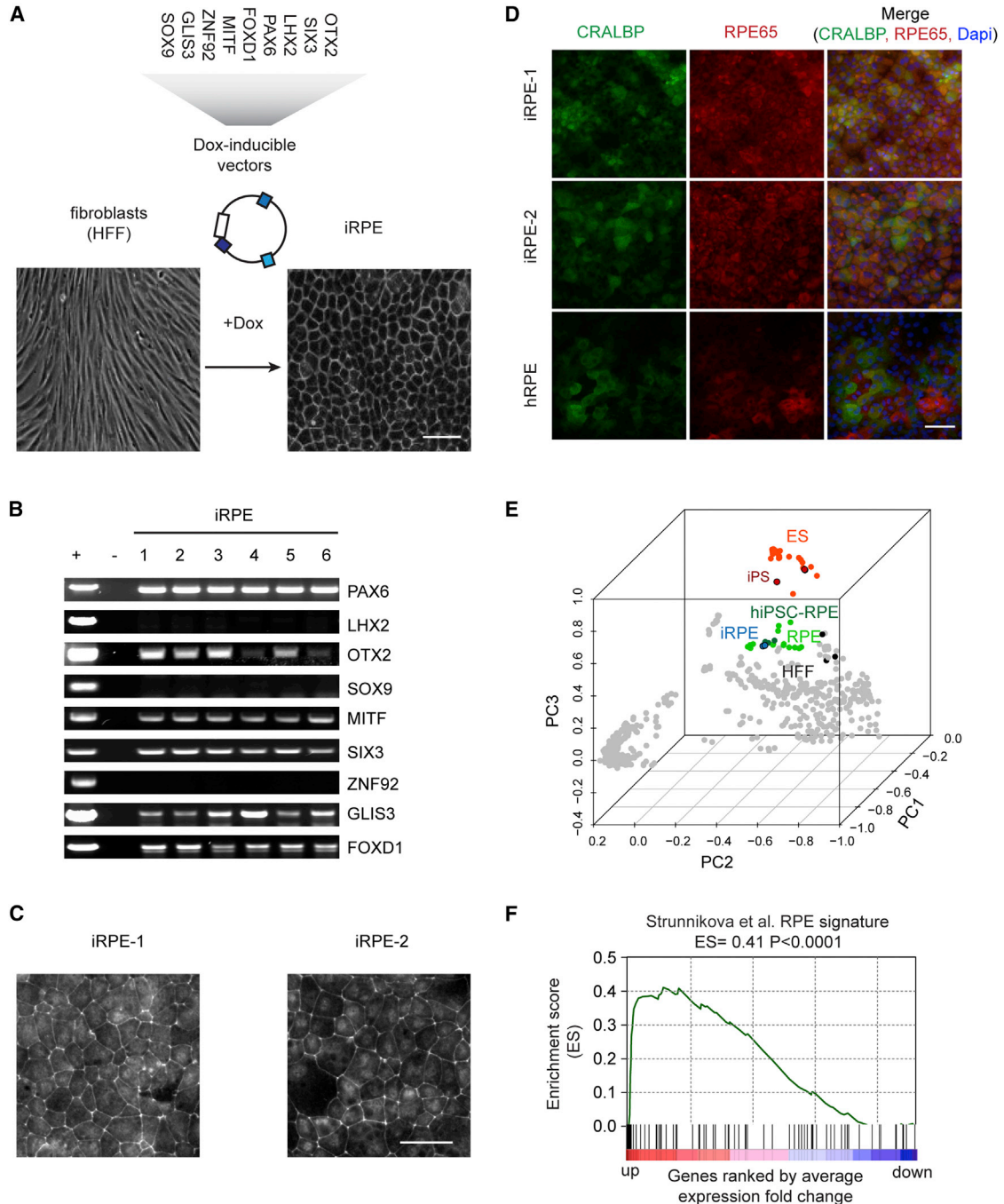
Figure 4. Ectopic Expression of RPE Candidate Core TFs Is Sufficient to Drive the Morphology and Gene Expression Program of Fibroblasts toward an RPE-like State

(A) Schematic outlining the ectopic expression of candidate core TFs in HFF. Lentiviral constructs were induced to express candidate core TFs with doxycycline (Dox). Scale bar, 50 μm.

(B) PCR and gel analysis of transgene integration for iRPE lines. Positive control (DNA of the constructs used to generate lentivirus) and negative control reactions are shown. Six different iRPE lines, labeled 1–6 are shown. Genes are indicated on the side.

(C) Immunostaining of iRPE-1 and iRPE-2 cells. Cells were immunostained with TJP1 (ZO-1). Scale bar 50 μm.

(D) Immunostaining imaging of RPE, iRPE-1, and iRPE-2 cells. Cells were immunostained for RPE cell markers CRALBP (green) and RPE65 (red) and with DAPI (blue). Scale bar, 50 μm.

(legend continued on next page)

and differentiated versions of RPE cells (Idelson et al., 2009; Kamao et al., 2014; Zhang et al., 2014), will prove useful in ultimately unraveling the complete transcriptional circuitry of RPE cells.

Multiple methods have been developed that can use high-throughput genomic data to identify factors critical for cell identity (Benayoun et al., 2014; Cahan et al., 2014; Davis and Eddy, 2013; Heinäniemi et al., 2013; Hwang et al., 2011; Lang et al., 2014; Morris et al., 2014; Roost et al., 2015; Zhou et al., 2011; Ziller et al., 2015). Many of these methods focus primarily on quantifying the differences between cell identities and less on the direct identification of factors controlling cell identity. Several of these approaches have experimentally verified that they are capable of identifying TFs important for cell identity, although none has demonstrated that the factors can establish cell identity to the extent shown here, possibly due to the extreme technical difficulty of these types of reprogramming experiments. Our expectation is that results for different methods of identifying candidate core TFs will eventually be compared and used in complementary fashions to gain insight on which TFs are critical for different cell types and which characteristics best define core TFs.

For the vast majority of human cell types, the core TFs and the transcriptional programs they control is poorly understood. Furthermore, much of disease-associated sequence variation occurs in transcriptional regulatory regions (Farh et al., 2015; Hnisz et al., 2013; Maurano et al., 2012), but the transcriptional mechanisms that lead to disease pathology are understood in only a few instances. The atlas of candidate core TFs described here should, therefore, facilitate future exploration of the functions of key regulators of cell identity, mapping of cellular regulatory circuitries, and investigation of disease-associated mechanisms.

EXPERIMENTAL PROCEDURES

Additional details are provided in the Supplemental Experimental Procedures.

Identification of Candidate Core TFs
Briefly, an entropy-based measure of Jensen-Shannon divergence (Cabili et al., 2011) was adopted to identify candidate core TFs based on the relative level and cell type specificity of expression of a given factor in one cell type compared to a background dataset of diverse human cell and tissue types. Expression datasets used are provided in Table S5. Additional details are provided in the Supplemental Experimental Procedures.

Cell Culture
Human RPE cells were maintained in epithelial cell medium supplemented with 2% fetal bovine serum, 1× epithelial cell growth supplement, and 1× penicillin-streptomycin solution. HFF were maintained in DMEM supplemented with 15% Tet System Approved fetal bovine serum, 2 mM L-glutamine, and 100 U/ml penicillin-streptomycin.

Construction of Lentiviral Vectors
The Lenti-X Tet-On Advanced Inducible Expression System was used. Plasmids containing the full coding sequence of *PAX6*, *OTX2*, *LHX2*, *MITF*, *SIX3*, *SOX9*, *GLIS3*, *FOXD1*, or *ZNF92* were obtained from Open Biosystems, Origene, or the Dana Farber/Harvard Cancer Center DNA Resource Core. Coding DNA sequences were amplified and cloned into the target vector via homologous recombination using the In-Fusion cloning system. Expression plasmids were transformed and maintained in STBL4 cells.

Viral Preparation and Transduction of HFF
For ectopic expression experiments, a human fetal fibroblast line expressing rtTA Advanced was generated by viral-mediated integration of an rtTA-expressing construct (from pLVX-Tet-On-Advanced). Cells were grown in 1 mg/ml geneticin for 2 weeks to select for cells harboring the construct.

For virus preparation, replication-incompetent lentiviral particles were packaged in 293T cells in the presence of the envelope (pMD2) and packaging (psPAX) plasmids. Viral supernatants from cultures 36, 48, 60, and 72 hr post-transfection were filtered through a 0.45-μM filter. High-titer virus preparations for all nine TFs were then added to HFF in the presence of 5 μg/ml of polybrene (day 1). A second transduction with virus for all nine factors was performed the next day (day 2). After 2 days, transduced HFF were split and transferred to iRPE medium (discussed later) (day 3). The following day, iRPE medium was supplemented with 2 mg/ml doxycycline (day 4). Medium was replaced every 3 days, and fresh doxycycline was added with every medium replacement.

iRPE Growth Conditions
iRPE lines were plated on Matrigel Basement Membrane Matrix-coated plates. iRPE cells were grown on Minimum Essential Medium (MEM) Eagle Alpha Modification containing 5% Tet System Approved fetal bovine serum, 1× N1 Medium Supplement,

(E) PCA comparing the gene expression profiles of iRPE cells to gene expression profiles of other cell types. Principal components (PC1–PC3) are shown on the x, y, and z axes. The expression profiles of HFF (black), iRPE cells (blue), RPE cells (light green), induced pluripotent stem (iPS)-RPE cells (green), iPS cells (red), and ES cells (orange red), and 106 additional cell types (gray) are shown.

(F) GSEA enrichment score of a previously published RPE signature gene set (Strunnikova et al., 2010) compared with genes differentially expressed between iRPE and fibroblasts. Genes are ranked along the x axis based on differential expression in iRPE cells versus fibroblasts, with more expressed in iRPE (red) to more expressed in fibroblasts (blue). Black tick marks indicate a gene from the RPE signature set. Enrichment score is shown on the y axis.
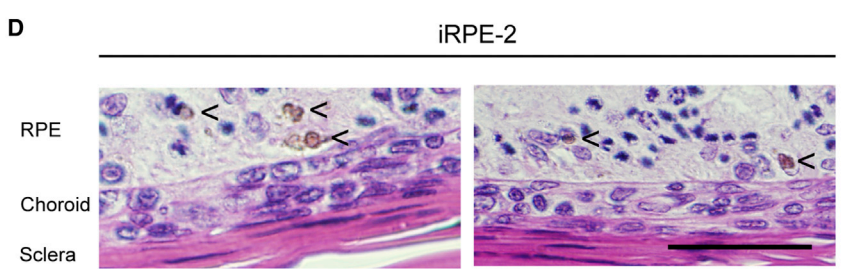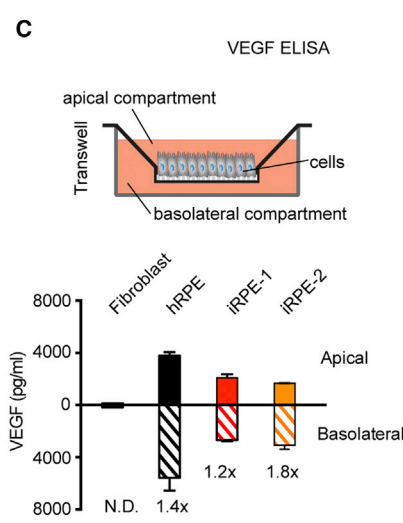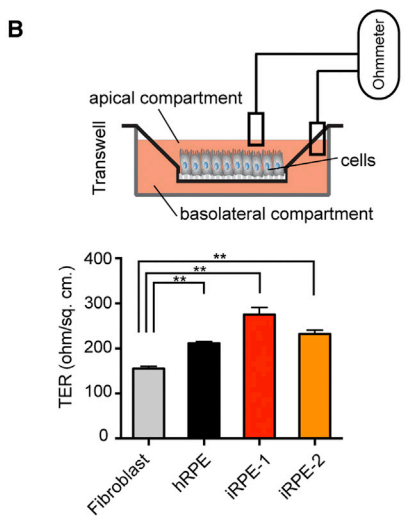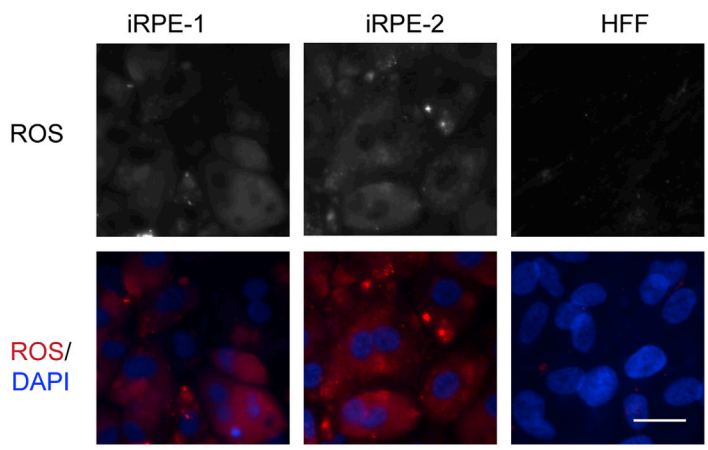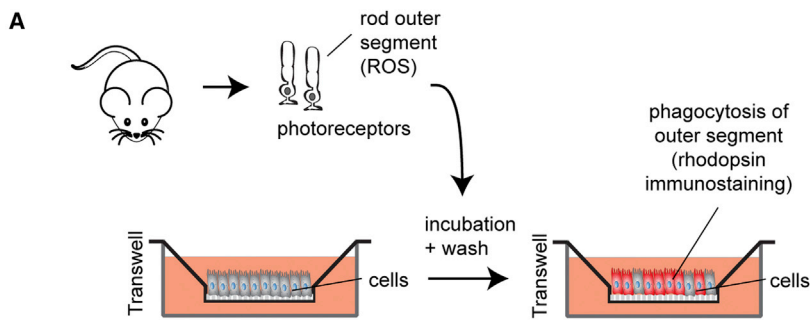
Figure 5. RPE-like Cells Have Functional Characteristics

(A) Schematic of the phagocytosis of ROS assay for iRPE function. Immunostaining for rhodopsin and DAPI are shown. The top row of images shows immunostaining for rhodopsin. The lower row of images shows the same fields with rhodopsin indicated in red and DAPI staining for DNA shown in blue. Scale bar, 25 μm.

(B) Schematic and results of TER assay for iRPE-1, iRPE-2, and hRPE cells (Salero et al., 2012). TER values for fibroblasts (gray), hRPE cells (black), iRPE-1 cells (red), and iRPE-2 cells (gold) are 155.2 ± 5 Ω/cm^2, 211.4 ± 4 Ω/cm^2, 275.6 ± 15 Ω/cm^2, and 232.2 ± 8 Ω/cm^2, respectively. TER was assayed in at least five biological replicates and is displayed as mean $\pm$ SD.

(C) Schematic and results for polarized release of VEGF assayed by ELISA. Values are shown for fibroblasts (nearly undetectable) and for hRPE (black), iRPE-1 (red), and iRPE-2 (gold), with the apical secretion values indicated with solid colors and the basolateral secretion values indicated with striped colors. The ratio of VEGF release (basolateral/apical) is shown below each bar. N.D., non detectable. ELISA was assayed in biological duplicates and is displayed as mean $\pm$ SD.

(D) Xenotransplant subretinal transplantations of wild-type albino Sprague-Dawley rats. H&E staining show pigmented donor cells iRPE-2 visible in the RPE layer. Single pigmented cells were identified in the host RPE layer in the doxycycline-treated group but not in the control iRPE group that did not receive doxycycline (data not shown). Pigmented cells are indicated with a "<" sign. Scale bar, 50 μm.

1% sodium pyruvate, 2 mM L-glutamine, 1× MEM Non-Essential Amino Acids, 1 mg/ml Geneticin, 100 U/ml penicillin-streptomycin and THT (20 mg/l hydrocortisone, 250 mg/l taurine, 0.013 mg/l triiodothyronine). Media was supplemented with 2 mg/ml doxycycline. Cells were incubated at 37°C with 5% CO_2 in a humidified incubator.

Genotyping

Cells were lysed, and genomic DNA was purified by treating samples with proteinase K, RNase A, and phenol-chloroform extraction. DNA was amplified using GoTaq Green Core Mix.

Immunostaining and Imaging

For immunostaining analysis, cells were grown in Corning transwell polyester membrane cell culture inserts for 8 weeks in iRPE medium supplemented with 2 mg/ml doxycycline. Medium was replaced every 3 days. Cells were then fixed and washed, and a 2-mm biopsy punch of the transwell membrane was transferred to a glass slide. Slides were incubated in blocking/permeabilizing solution, washed, treated with antibodies, and mounted. Slides were visualized under a fluorescence microscope (Zeiss Axio Observer D1).

Phagocytosis Assay

ROS were isolated following previously described protocols (Ryeom et al., 1996). Retinas were dissected immediately following sacrifice from 25 mice, ROS were isolated, and approximately 1.0×10^4 ROS were added to the supernatant of confluent cell cultures in transwells. All work was performed under protocols approved by the institutional animal care and use committee. The cells were then incubated for 2 hr at 37°C. Transwells were washed to remove all unbound ROS before fixation. Each transwell was fixed, immunostained for rhodopsin and DAPI, and visualized.

TER

iRPE and RPE cells (Salero et al., 2012) were grown in Corning transwell polyester membrane cell culture inserts for 8 weeks in iRPE medium supplemented with 2 mg/ml doxycycline. Medium was replaced every 3 days. Resistance was measured using the EVOM Epithelial Voltohmmeter.

VEGF-A Release

iRPE and RPE cells (Salero et al., 2012) were grown in Corning transwell polyester membrane cell culture inserts for 8 weeks in iRPE medium supplemented with 2 mg/ml doxycycline. RPE cells were collected from human cadaver donors for research purposes. The protocol was reviewed and found to be exempt category 4 by the Albany Medical College Institutional Review Board. Medium was replaced every 3 days with fresh doxycycline. Conditioned medium from apical and basal chambers of the same transwell insert was collected 24 hr following a complete medium change. VEGF-A protein secretion in conditioned medium was measured using a Human VEGF ELISA kit following the manufacturer's suggested protocol (Life Technologies).

Transplantation

To study the ability of iRPE to integrate into the native retina, we have performed subretinal transplantations into the wild-type rat retina. All animal studies were designed and performed according to the Association for Research in Vision and Ophthalmology guidelines for the use of animals in ophthalmic and visual research and approved by the institutional animal care and use committee. Three-week-old albino Sprague-Dawley rats (Taconic Biosciences) were used in these experiments. One day before the surgery, all animals were switched to cyclosporine-A-supplemented water (210 mg/l) and remained on immunosuppressive treatment until the end of the study. One group of iRPE-transplanted animals also received doxycycline in the water.

For the surgery, animals were anesthetized by intraperitoneal injection of ketamine/xylazine. Topical proparacaine (anesthetic) and tropicamide (mydriatic agent) drops were applied.

The subretinal injection was performed in one eye per animal using a 50-μm beveled glass needle connected to a 10-μl Hamilton syringe through polyethylene tubing. The success of the injection and lack of complications (hemorrhage, retinotomy, leakage of cells into the vitreous) was assessed by fundus examination. Antibiotic ointment was applied to the eye for recovery.

Experimental groups were as follows: iRPE with doxycycline treatment (n = 5), iRPE without doxycycline treatment (n = 5), hRPE (n = 5) as positive control, vehicle injection (n = 5), and non-injected eyes (n = 5) as negative controls—five groups total.

Two weeks after the injection animals were euthanized by CO_2 inhalation, eyes were enucleated and fixed in alcohol fixative (Excalibur Pathology), embedded in paraffin, and sectioned.

ACCESSION NUMBERS

The accession number for the raw and processed sequencing and microarray data reported in this paper is GEO: GSE64264 (http://www.ncbi.nlm.nih.gov/geo/).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and five tables and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.09.016.

AUTHOR CONTRIBUTIONS

A.C.D. contributed to the design of all experiments and performed reprogramming experiments. Z.P.F. developed the computational method used and provided all bioinformatics-based analyses. K.J.W. performed the phagocytosis assay. P.B. and M.J.Y. performed the in vivo transplantation experiments. M.A.C. performed iRPE colony selection. J.S.S. performed immunostaining and VEGF release experiments. E.C. provided maintenance of RPE and iRPE lines. C.C. performed the TER experiments. D.D. contributed to experimental design. N.M.H. generated the lentiviral constructs. R.J. and S.T. contributed to the conceptual development of the study. T.I.L. and R.A.Y. initially conceived the study and contributed to the design of experiments. A.C.D., Z.P.F., T.I.L., and R.A.Y. wrote the manuscript. M.A.C., K.J.W., R.J., and S.T. contributed critical comments.

REFERENCES

Benayoun, B.A., Pollina, E.A., Ucar, D., Mahmoudi, S., Karra, K., Wong, E.D., Devarajan, K., Daugherty, A.C., Kundaje, A.B., Mancini, E., et al. (2014). H3K4me3 breadth is linked to cell identity and transcriptional consistency. Cell *158*, 673–688.

Bok, D. (1993). The retinal pigment epithelium: a versatile partner in vision. J. Cell Sci. Suppl. *17*, 189–195.

Breitling, R., Armengaud, P., Amtmann, A., and Herzyk, P. (2004). Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. *573*, 83–92.

Buganim, Y., Itskovich, E., Hu, Y.C., Cheng, A.W., Ganz, K., Sarkar, S., Fu, D., Welstead, G.G., Page, D.C., and Jaenisch, R. (2012). Direct reprogramming of fibroblasts into embryonic Sertoli-like cells by defined factors. Cell Stem Cell *11*, 373–386.

Buganim, Y., Faddah, D.A., and Jaenisch, R. (2013). Mechanisms and models of somatic cell reprogramming. Nat. Rev. Genet. *14*, 427–439.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Cahan, P., Li, H., Morris, S.A., Lummertz da Rocha, E., Daley, G.Q., and Collins, J.J. (2014). CellNet: network biology applied to stem cell engineering. Cell *158*, 903–915.

Cyranoski, D. (2013). Stem cells cruise to clinic. Nature *494*, 413.

Cyranoski, D. (2014). Stem-cell method faces fresh questions. Nature *507*, 283.

Davis, F.P., and Eddy, S.R. (2013). Transcription factors that convert adult cell identity are differentially polycomb repressed. PLoS ONE *8*, e63407.

Farh, K.K., Marson, A., Zhu, J., Kleinewietfeld, M., Housley, W.J., Beik, S., Shoresh, N., Whitton, H., Ryan, R.J., Shishkin, A.A., et al. (2015). Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature *518*, 337–343.

Ford, K.M., Saint-Geniez, M., Walshe, T., Zahr, A., and D'Amore, P.A. (2011). Expression and role of VEGF in the adult retinal pigment epithelium. Invest. Ophthalmol. Vis. Sci. *52*, 9478–9487.

Fuglede, B., and Topsoe, F. (2004). Jensen-Shannon Divergence and Hilbert space embedding. Proceedings of the 2004 International Symposium on Information Theory (IEEE), p. 31.

Graf, T. (2011). Historical origins of transdifferentiation and reprogramming. Cell Stem Cell *9*, 504–516.

Graf, T., and Enver, T. (2009). Forcing cells to change lineages. Nature *462*, 587–594.

Heinäniemi, M., Nykter, M., Kramer, R., Wienecke-Baldacchino, A., Sinkkonen, L., Zhou, J.X., Kreisberg, R., Kauffman, S.A., Huang, S., and Shmulevich, I. (2013). Gene-pair expression signatures reveal lineage control. Nat. Methods *10*, 577–583.

Henriques, T., Gilchrist, D.A., Nechaev, S., Bern, M., Muse, G.W., Burkholder, A., Fargo, D.C., and Adelman, K. (2013). Stable pausing by RNA polymerase II provides an opportunity to target and integrate regulatory signals. Mol. Cell *52*, 517–528.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Huang, P., He, Z., Ji, S., Sun, H., Xiang, D., Liu, C., Hu, Y., Wang, X., and Hui, L. (2011). Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. Nature *475*, 386–389.

Hwang, P.I., Wu, H.B., Wang, C.D., Lin, B.L., Chen, C.T., Yuan, S., Wu, G., and Li, K.C. (2011). Tissue-specific gene expression templates for accurate molecular characterization of the normal physiological states of multiple human tissues with implication in development and cancer studies. BMC Genomics *12*, 439.

Idelson, M., Alper, R., Obolensky, A., Ben-Shushan, E., Hemo, I., Yachimovich-Cohen, N., Khaner, H., Smith, Y., Wiser, O., Gropp, M., et al. (2009). Directed differentiation of human embryonic stem cells into functional retinal pigment epithelium cells. Cell Stem Cell *5*, 396–408.

Iwafuchi-Doi, M., and Zaret, K.S. (2014). Pioneer transcription factors in cell reprogramming. Genes Dev. *28*, 2679–2692.

Kamao, H., Mandai, M., Okamoto, S., Sakai, N., Suga, A., Sugita, S., Kiryu, J., and Takahashi, M. (2014). Characterization of human induced pluripotent stem cell-derived retinal pigment epithelium cell sheets aiming for clinical application. Stem Cell Reports *2*, 205–218.

Lang, A.H., Li, H., Collins, J.J., and Mehta, P. (2014). Epigenetic landscapes explain partially reprogrammed cells and identify key reprogramming genes. PLoS Comput. Biol. *10*, e1003734.

Lujan, E., Chanda, S., Ahlenius, H., Südhof, T.C., and Wernig, M. (2012). Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. Proc. Natl. Acad. Sci. USA *109*, 2527–2532.

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. Science *337*, 1190–1195.

Morris, S.A., and Daley, G.Q. (2013). A blueprint for engineering cell fate: current technologies to reprogram cell identity. Cell Res. *23*, 33–48.

Morris, S.A., Cahan, P., Li, H., Zhao, A.M., San Roman, A.K., Shivdasani, R.A., Collins, J.J., and Daley, G.Q. (2014). Dissecting engineered cell types and enhancing cell fate conversion via CellNet. Cell 158, 889–902.

Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Schöler, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell 95, 379–391.

Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. Proc. Natl. Acad. Sci. USA 110, 17921–17926.

Rivera, C.M., and Ren, B. (2013). Mapping human epigenomes. Cell 155, 39–55.

Roost, M.S., van Iperen, L., Ariyurek, Y., Buermans, H.P., Arindrarto, W., Devalla, H.D., Passier, R., Mummery, C.L., Carlotti, F., de Koning, E.J., et al. (2015). KeyGenes, a tool to probe tissue differentiation using a human fetal transcriptional atlas. Stem Cell Reports 4, 1112–1124.

Ryeom, S.W., Sparrow, J.R., and Silverstein, R.L. (1996). CD36 participates in the phagocytosis of rod outer segments by retinal pigment epithelium. J. Cell Sci. 109, 387–395.

Saint-Geniez, M., Kurihara, T., Sekiyama, E., Maldonado, A.E., and D'Amore, P.A. (2009). An essential role for RPE-derived soluble VEGF in the maintenance of the choriocapillaris. Proc. Natl. Acad. Sci. USA 106, 18751–18756.

Salero, E., Blenkinsop, T.A., Corneo, B., Harris, A., Rabin, D., Stern, J.H., and Temple, S. (2012). Adult human RPE can be activated into a multipotent stem cell that produces mesenchymal derivatives. Cell Stem Cell 10, 88–95.

Sancho-Martinez, I., Baek, S.H., and Izpisua Belmonte, J.C. (2012). Lineage conversion methodologies meet the reprogramming toolbox. Nat. Cell Biol. 14, 892–899.

Sheng, C., Zheng, Q., Wu, J., Xu, Z., Wang, L., Li, W., Zhang, H., Zhao, X.Y., Liu, L., Wang, Z., et al. (2012). Direct reprogramming of Sertoli cells into multipotent neural stem cells by defined factors. Cell Res. 22, 208–218.

Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell 151, 994–1004.

Sparrow, J.R., Hicks, D., and Hamel, C.P. (2010). The retinal pigment epithelium in health and disease. Curr. Mol. Med. 10, 802–823.

Stergachis, A.B., Neph, S., Reynolds, A., Humbert, R., Miller, B., Paige, S.L., Vernot, B., Cheng, J.B., Thurman, R.E., Sandstrom, R., et al. (2013). Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. Cell 154, 888–903.

Stevenson, B.R., Siliciano, J.D., Mooseker, M.S., and Goodenough, D.A. (1986). Identification of ZO-1: a high molecular weight polypeptide associated with the tight junction (zonula occludens) in a variety of epithelia. J. Cell Biol. 103, 755–766.

Strauss, O. (2005). The retinal pigment epithelium in visual function. Physiol. Rev. 85, 845–881.

Strunnikova, N.V., Maminishkis, A., Barb, J.J., Wang, F., Zhi, C., Sergeev, Y., Chen, W., Edwards, A.O., Stambolian, D., Abecasis, G., et al. (2010). Transcriptome analysis and molecular signature of human retinal pigment epithelium. Hum. Mol. Genet. 19, 2468–2486.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science 242, 405–411.

Vaquerizas, J.M., Kummerfeld, S.K., Teichmann, S.A., and Luscombe, N.M. (2009). A census of human transcription factors: function, expression and evolution. Nat. Rev. Genet. 10, 252–263.

Vierbuchen, T., and Wernig, M. (2012). Molecular roadblocks for cellular reprogramming. Mol. Cell 47, 827–838.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. Nature 463, 1035–1041.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153, 307–319.

Xie, W., and Ren, B. (2013). Developmental biology. Enhancing pluripotency and lineage specification. Science 341, 245–247.

Xu, J., Du, Y., and Deng, H. (2015). Direct lineage reprogramming: strategies, mechanisms, and applications. Cell Stem Cell 16, 119–134.

Yamanaka, S. (2012). Induced pluripotent stem cells: past, present, and future. Cell Stem Cell 10, 678–684.

Zhang, K., Liu, G.H., Yi, F., Montserrat, N., Hishida, T., Esteban, C.R., and Izpisua Belmonte, J.C. (2014). Direct conversion of human fibroblasts into retinal pigment epithelium-like cells by defined factors. Protein Cell 5, 48–58.

Zhou, J.X., Brusch, L., and Huang, S. (2011). Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model. PLoS ONE 6, e14752.

Ziller, M.J., Edri, R., Yaffe, Y., Donaghey, J., Pop, R., Mallard, W., Issner, R., Gifford, C.A., Goren, A., Xing, J., et al. (2015). Dissecting neural differentiation regulatory networks through epigenetic footprinting. Nature 518, 355–359.