Reprogramming cell fate with a genome-scale library of artificial transcription factors


*Cellular and Molecular Biology Training Program, University of Wisconsin–Madison, Madison, WI 53706; †Department of Biochemistry, University of Wisconsin–Madison, Madison, WI 53706; ‡Genetics Training Program, University of Wisconsin–Madison, Madison, WI 53706; §Department of Electrical and Computer Engineering, University of Wisconsin–Madison, Madison, WI 53706; ‡Morgridge Institute for Research, Madison, WI 53715; ‡Department of Medicine, University of Wisconsin School of Medicine and Public Health, Madison, WI 53792; ‡Wisconsin National Primate Research Center, University of Wisconsin–Madison, Madison, WI 53715; ‡Department of Pathology and Laboratory Medicine, University of Wisconsin–Madison, Madison, WI 53715; ‡Department of Cell and Regenerative Biology, University of Wisconsin School of Medicine and Public Health, Madison, WI 53706; ‡The Genome Center of Wisconsin, University of Wisconsin–Madison, Madison, WI 53706; and ‡Stem Cell Institute, University of Minnesota, Minneapolis, MN 55455

Edited by Michael R. Green, University of Massachusetts Medical School, Worcester, MA, and approved November 14, 2016 (received for review July 7, 2016)

Artificial transcription factors (ATFs) are precision-tailored molecules designed to bind DNA and regulate transcription in a preprogrammed manner. Libraries of ATFs enable the high-throughput screening of gene networks that trigger cell fate decisions or phenotypic changes. We developed a genome-scale library of ATFs that display an engineered interaction domain (ID) to enable cooperative assembly and synergistic gene expression at targeted sites. We used this ATF library to screen for key regulators of the pluripotency network and discovered three combinations of ATFs capable of inducing pluripotency without exogenous expression of Oct4 (POU domain, class 5, TF 1). Cognate site identification, global transcriptional profiling, and identification of ATF binding sites reveal that the ATFs do not directly target Oct4; instead, they target distinct nodes that converge to stimulate the endogenous pluripotency network. This forward genetic approach enables cell type conversions without a priori knowledge of potential key regulators and reveals unanticipated gene network dynamics that drive cell fate choices.

Significance

The ability to convert cells into desired cell types enables tissue engineering, disease modeling, and regenerative medicine; however, methods to generate desired cell types remain difficult, uncertain, and laborious. We developed a strategy to screen gene regulatory elements on a genome scale to discover paths that trigger cell fate changes. The proteins used in this study cooperatively bind DNA and activate genes in a synergistic manner. Subsequent identification of transcriptional networks does not depend on prior knowledge of specific regulators important in the biological system being tested. This powerful forward genetic approach enables direct cell state conversions as well as other challenging manipulations of cell fate.


Conflict of interest statement: A.Z.A. is the sole member of VistaMotif, LLC, and founder of the nonprofit WINStep Forward.

This article is a PNAS Direct Submission.

Data deposition: The RNA-seq and ChIP-seq data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo (accession no. GSE89221).

1 To whom correspondence should be addressed. Email: azansari@wisc.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1611142114/-/DCSupplemental.
the ATF can interact with other factors in the cell (10). Principles of cooperative assembly and synergistic activation were integrated in the design of our genome-scale ATF library (15, 16).

We used the following three criteria to choose among an array of DBD scaffolds: (i) ability to target multiple nodes of a gene network, (ii) regulatory potency of resulting ATFs, and (iii) efficiency of delivery into cells. Applying these criteria to the repurposed nuclease-inactivated CRISPR/Cas9 system, TAL-effectors, programmable small-molecule polyamides, and zinc fingers, we concluded that the zinc finger scaffold, engineered to enable cooperative and combinatorial assembly on DNA, would be the most effective DBD scaffold for the creation of an ATF library designed to trigger cell fate conversions (see Materials and Methods for details on choice of DBD).

To demonstrate unbiased ability to change cell identity, we used our ATF library to screen for factors that induced pluripotency in mouse embryonic fibroblasts (MEFs) without exogenous delivery of Oct4. We created a library of $2.62 \times 10^6$ ATFs that encompass five times the number of factors as the entire sequence space of all 10-bp binding sites on double-stranded DNA (524,809 unique sequences). RNA-sequencing (RNA-seq) data, epigenetic landscapes, and comprehensive ATF binding profiles by cognate site identification (CSI) sequencing (CSI-seq) and ChIP-sequencing (ChIP-seq) were analyzed to determine the key nodes through which ATFs activate the pluripotency network.

Surprisingly, bioinformatic analysis reveals that the ATFs take fibroblasts through a different path to pluripotency than exogenously expressed Oct4. We demonstrate that this forward genetic approach enables the pursuit of elusive cell fate conversions in an unbiased manner.

Results

ATF Architecture. To determine the best architecture for a zinc finger ATF library, we tested the impact of each modular domain on the level of induction. The zinc finger backbone is derived from human EGR1/ZIF268 (early growth response 1), a well-studied scaffold for zinc finger ATFs (17–19). We fused VP64, a tetrameric repeat of the 11-aa activation region of VP16, a potent transactivation domain from the herpes simplex virus to the C terminus to the zinc fingers (Fig. 1A) (20). Although a variety of zinc finger-based libraries have been described before (21–27), a distinguishing and important feature of our ATF design is inclusion of a 15-aa peptide that serves as an ID, allowing dimerization of the ATF with another ATF through the hydrophobic surface of the first zinc finger of EGR1 (28). The inclusion of the ID in our ATF design adds a layer of control to the ATF library by allowing the ATFs to harness cooperative binding and synergistic activation (16).

Because we wanted to create an ATF library of high complexity while minimizing nonessential modules, it was necessary...
to determine the minimum number of zinc fingers required to have a transcriptional effect. Toward this end, we compared a two-zinc finger ATF with a three-zinc finger ATF in a luciferase assay with a palindromic cognate site for EGR1 upstream of the luciferase reporter. The two-zinc finger ATF could only activate the luciferase reporter 2-fold over background, whereas the three-zinc finger ATF was capable of activating 329-fold over background (Fig. 1B). Incorporation of the ID further increased activation of the three-zinc finger ATF almost by an order of magnitude (SI Appendix, Fig. S3A).

To determine how the ATFs impact genome-wide transcription, we performed RNA-seq in human cells expressing one of the four ATFs with different architectures. The ATFs either had the first two or all three zinc fingers of EGR1 as the DBD, with and without the ID. Compared with the mock-transfected control, the two-zinc finger ATFs had little impact on altering the transcriptional profile (Fig. 1C and D, and SI Appendix, Fig. S3B and C). On the other hand, the three-zinc finger ATF with the ID altered the expression of 104 transcripts (100 up-regulated, 4 down-regulated) (Fig. 1D and SI Appendix, Fig. S3C). Most of the genes were up-regulated compared with the mock-transfected control, and the repressed genes could be attributed to indirect effects of the ATFs. As the three-zinc finger ATF with the ID was capable of binding as a monomer or dimer, most of the subset of genes up-regulated by the three-zinc finger ATF without the ID could also be activated by the three-zinc finger ATF with the ID, and the ID increased the level of induction for a subset of genes (SI Appendix, Fig. S3B).

**Genome-Scale ATF Library Design.** Taking all these results into consideration, a scaffold that includes from N to C terminus: an ID, three zinc fingers derived from EGR1, NLS from EGR1, VP64 activation domain, and a 3x hemagglutinin (HA) tag was used to construct the ATF scaffold. The library of different DNA sequence-targeting ATFs was created by incorporating VNN codons, where V is A, C, or G, at the DNA recognition residues (~1, 2, 3, and 6). Use of VNN codons prevents incorporation of premature stop codons within the ORF and permits the incorporation of 16 different amino acids (Fig. 1E). The library was cloned into a second-generation lentiviral system to ensure efficient delivery to mammalian cells. The ATF is driven by the constitutively active EF-1α promoter, which resists silencing in mammalian cells compared with other constitutive promoters (Fig. 2A) (29). The sequence space for all 100 bp sequences on double-stranded DNA is 524,809 different sequence permutations. We created an ATF library with a complexity of 2.62 × 10^6, five times the targeted sequence space. Sanger sequencing of 100 clones confirmed success of our design with incorporation of all 16 aa at each recognition residue, suggesting diverse representation in the library (Fig. 1E and SI Appendix, Table S7).

**ATF Library Activates the Pluripotency Network.** We asked whether ATFs in the library could replace the key regulator of pluripotency, Oct4, in the mixture of TFs that triggers the pluripotency network, Oct4, Sox2, Klf4, and c-Myc (Oct4+SKM). To test a library, capable of sampling thousands of sites in the genome, it was necessary to have a robust readout of positive phenotypes (Fig. 2A). Toward this end, we used MEFs isolated from a transgenic mouse line that allows lineage tracing of endogenous Oct4 transcription (Fig. 2B) (30). In these cells, tamoxifen-inducible Cre recombinase (mER-Cre-mER) is expressed when the endogenous pluripotency associated gene, Oct4, is transcribed. In the presence of 4-hydroxytamoxifen, the recombinase removes Tomato from the ROSA locus, and transmembrane-bound GFP is expressed. Consequently, Tomato*GFP* MEFs become Tomato*GFP* cells after endogenous Oct4 is activated, and GFP expression is maintained in all their cell progeny (Fig. 2B).
Fig. 2. ATF library activates pluripotency network. (A) Genetic screen with an ATF library. (1) The ATF library was cloned into a second-generation lentiviral vector. The screen is performed in cells with a robust change in phenotype or a lineage-specific reporter. (2) Positive outcomes are isolated as single cells, such that combinations of ATFs, if any, can be captured (3). Integrated ATFs are identified from single cells (4). Identified ATFs are retested for validation. Once validated, downstream experiments can be performed to identify ATF target genes. (B) Testing the ATF library in mouse embryonic fibroblasts (MEFs) isolated from a transgenic mouse line that allows lineage tracing of endogenous Oct4 transcription. Upon induction of endogenous Oct4, tamoxifen-inducible Cre recombinase (mER-Cre-mER) is coexpressed. The recombinase removes Tomato from the ROSA locus, and transmembrane-bound GFP is expressed. Consequently, Tomato^–^GFP^+^ MEFs become Tomato^–^GFP^+^, and GFP expression is maintained in all their cell progeny. (C) Flow cytometry results at day 15 after introduction of TFs. Tomato^–^GFP^+^ MEFs transduced with the ATF library+SKM were isolated as single cells for further analysis. MEFs treated with Oct4+SKM (positive control) and untreated MEFs (negative control) were used for comparison. Double-positive (Tomato^+^GFP^+^; Q2) cells were also collected. Percentages are displayed under the quadrant number. (D) ATFs were identified from 11 single cells by two-step nested PCR of genomic DNA. Unique ATFs are depicted with a different color. One ATF had a frameshift mutation shortly after the ID, coding for a protein that does not have a zinc finger structure. A few ATFs, notably the light blue and orange ATFs, are expressed in most of the cells analyzed. Three ATFs are made up of two fingers (light blue, red, and pink). All cells except number 4 were collected as Tomato^–^GFP^+^ cells. Cell 4 was Tomato^+^GFP^+^ at day 15. (E) Three combinations of ATFs (C2, C3, and C4) successfully induced pluripotency with SKM. Micrographs of MEFs transduced with an Empty+SKM are Tomato^–^GFP^–^. iPS cells generated with ATFs+SKM are similar to those generated with Oct4+SKM and are Tomato^–^GFP^+^ . Two ATFs in each combination are the same (light blue and orange). (Scale bar, 100 μm.)
expression of pluripotency markers, OCT4, SOX2, and NANOG (Fig. 3A and SI Appendix, Fig. S5A). The capacity for ATF-derived iPS cells to differentiate into all three germ layers was assessed by formation of teratomas (Fig. 3B) and embryo bodies (EBs) (SI Appendix, Fig. S5 B–E). Immunocytochemistry of myosin light polypeptide 2 (mesoderm), forkhead box A2 (endoderm), and βIII tubulin (ectoderm) confirmed expression of germ layer markers at the protein level, and RT-qPCR of T, Nkx2.5, and Kdr (mesoderm); Tbr, Afp, and Foxa2 (endoderm); Nes, Nefl, and Sox17 (ectoderm) confirmed differentiation into all three germ layers at the transcriptional level. Beating cardiomyocytes were also observed from embryoid outgrowths from EBs derived from ATF-induced iPS cells (Movie S1).

From morphological and select gene marker analysis, we expanded our validation to global transcriptome analysis of the ATF-treated iPS cells. Comparison of genome-wide transcriptional profiles showed that the transcriptomes of ATF-induced iPS cells cluster tightly with mouse ES cells as well as iPS cells generated with Oct4<SKM (Fig. 3C and SI Appendix, Fig. S6A). RNA profiles of cells at early stages of reprogramming clustered with MEFs and the Empty+SKM control.

ATF-induced iPS cells show an up-regulation of pluripotency markers and a down-regulation of fibroblast markers (Fig. 3D). Using the 853 genes that make up the fibroblast gene regulatory network (GRN) and the 705 genes that make up the pluripotency GRN from CellNet (31), we compared the expression profiles of ATF-induced iPS cells to those of other pluripotent cells and MEFs (Fig. 3 E and F). Our genome-wide analysis indicates that the profiles of our ATF-induced iPS cells highly correlated with profiles of pluripotent cells generated using exogenous Oct4. It is important to note that, at early stages of reprogramming, ATF-treated cells have a remarkably different profile compared with Oct4<SKM-treated cells (Fig. 3F). These differences suggest other underlying regulators beyond what is characterized in the GRNs of CellNet guide cells to pluripotency. Once fully reprogrammed to the pluripotent state, global transcriptome profiles show ATF-induced iPS cells share more similarity among themselves than with Oct4<SKM or ES cells (Fig. 3C and SI Appendix, Fig. S6 B and C). ATF expression at early stages of reprogramming was readily detectable by RNA-seq; however, once converted to iPS cells, the lentiviral elements controlling the expression of ATFs are silenced, a further confirmation that the cells were fully reprogrammed. Once reprogrammed, the converted cells maintain the iPS cell state even in the absence of regulators that triggered the initial regulatory nodes that led to cell fate conversion.

**Signature Epigenetic Landscapes at ATF-Activated Pluripotency Genes.** The genome-wide chromatin modification landscapes in ATF-induced iPS cells were compared with iPS cells generated with Oct4<SKM. Specifically, ChIP-seq was performed on histone 3 lysine 27 acetylation (H3K27ac), the marker delineating active promoters and superenhancers that define cell identity (32, 33), and histone 3 lysine 9 trimethylation (H3K9me3), the marker that is strongly correlated to repressed regions of the genome that are bound by repressive histone marks by deepTools package shows strong H3K27ac peaks upstream of the gene and in the gene body for expressed genes, whereas broad H3K9me3 peaks marked genes that were not expressed (SI Appendix, Fig. S7 C and D) (37).

IPS cells. H3K9me3 peaks were annotated to genes with Homer to create Venn diagrams for genes. ChIP enrichment among treatments was determined to be significant at an FDR < 0.1 by DiffBind (n = 2).
The differentially marked histone modifications were determined by using DiffBind (38). Peaks with a false-discovery rate (FDR) < 0.1 were categorized into unique and overlapping sets for iPS cells generated with ATF combinations + SKM or Oct4 + SKM. Remarkably, despite differences in transcriptome profiles during early stages (Fig. 3F), pluripotent cells shared similar sets of peaks for H3K27ac regardless of whether they were generated with ATFs or with natural factors (Fig. 3G). Likewise, repressive H3K9me3 peaks were similar for ATF-induced iPS cells and Oct4-induced iPS cells, although there was greater overlap for Oct4+SKM and C3+SKM (Fig. 3F). Taken together, the active H3K27ac marks and the repressive H3K9me3 marks confirm that the ATFs+SKM induced significant remodeling of the chromatin structure in MEFs to a state that distinguishes them as pluripotent cells.

**DNA Sequence Specificity Landscapes of Pluripotency-Inducing ATFs.** DNA targets of ATFs were examined by CSI, a method that captures the comprehensive binding profile of a DNA-binding factor (39–41). CSI enables the discovery of sequence specificity across all possible randomized 25-bp sequence permutations. The ATF–DNA complexes are captured with an HA antibody, and bound DNA is PCR amplified for the next round of selection. Three rounds of selection are performed, and all three rounds are multiplexed and sequenced to obtain specificity and energy landscapes (SELs) and position weight matrices (PWMs). (B) SELs display the comprehensive binding preferences based on a chosen seed motif. The height of the peak is associated with affinity. Sequences that are 1–2 bp longer than the seed are arranged in concentric rings. Each ring outward from the 0-mismatch ring displays sequences to the corresponding number of mismatches. Within the mismatch rings, sequences are arranged by position of the mismatch, and then alphabetically. (C) SELs of four ATFs validated in this study. The first two ATFs (light blue and orange) appear in all three combinations. Each SEL displayed shows data after three rounds of enrichment. (D) Gene set enrichment analysis of genes with a CSI score of > 20 within ±1 kb of the TSS shows overrepresentation of genes from PluriNetWork for *Mus musculus*. The top-100-scoring 10-bp motifs for each ATF from C2 were used for this analysis. Binding sites were annotated to genes with Homer. Genes with a sum CSI score of >20 (equivalent to six or more ATF binding sites) were analyzed with Enrichr (WikiPathways). The combined score from Enrichr is calculated by multiplying the log of the P value by the z score (deviation from the expected rank). The adjusted P value corresponds to the Benjamini–Hochberg corrected P value.
**Fig. 5.** ATFs activate key regulators of the pluripotency network. (A) Workflow for determining ATF target genes for C2+SKM. Three pairwise comparisons were made: (1) C2+SKM iPS cells vs. Empty+SKM cells, (2) C2+SKM early cells vs. Empty+SKM cells, (3) C2+SKM iPS cells vs. Oct4+SKM iPS cells. Genes up-regulated greater than twofold \((p < 0.05)\) in the cells transduced with ATFs with ATF binding sites within a \(\pm 1\)-kb window of the TSS were determined to be potential targets. Binding sites were identified by using the top-five-scoring 10-bp motifs from CSI. These target genes were used to build the network in Fig. 6A with information from the literature and the STRING database. (B) Differentially expressed pluripotency genes with ATF binding sites within \(\pm 1\) kb of the transcriptional start site (TSS). ATF binding sites were derived from the top-five-scoring 10-bp motifs from CSI for C2+SKM. (C) ChIP-seq signal for HA tag on ATFs for five predicted targets in Fig. 5B. Additional ChIP-seq traces are in SI Appendix, Fig. S5A. Traces display total reads for C2+SKM and Empty+SKM cells at an intermediate stage before reprogramming to a pluripotent state.

*Discussion*

Zinc finger, TAL effector, and CRISPR/Cas9 libraries have been tested for loss-of-function phenotypes, acquisition of resistance to a drug, or up-regulation of specific genes (21, 22, 27, 59–63); however, this study reports a gain-of-function screen with a genome-scale ATF library to reprogram fibroblasts to iPS cells, a feat that requires drastic transcriptional and epigenetic changes. Rather than engineering ATFs to target unique sites in the genome, we chose the zinc finger DBD primarily because of its ability to target a range of different DNA sites as well as the ability to interact with methylated and heterochromatic DNA.
Furthermore, we deliberately engineered in a synthetic protein–protein interaction module that endows the ATFs with a unique ability to sample binding sites as cooperative dimers and thereby stimulate target gene expression to greater extent due to transcriptional synergy (15). Based on design principles, we generated and screened a zinc finger ATF library of high complexity, not previously tested in mammalian cells. Furthermore, the capacity for the ATFs to cooperatively bind target genes provides the library with a unique feature to sample a larger set of binding sites cooperatively and activate to greater extent due to synergy. Conventional zinc finger libraries consisted of ATFs created by shuffling a limited number of zinc finger units, previously characterized to bind specific triplets of nucleotides (64). The library used in this study uses a much larger repertoire of residues, incorporating 16 of the 20 possible amino acids in the recognition residues, greatly expanding the target space of the ATFs. Our design, which is consistent with a survey of natural zinc fingers well suited for activating epigenetically silenced regions unlike TAL-fingers bind methylated DNA and heterochromatin (65, 68), making zinc fingers capable of binding sites of zinc fingers is an important advantage. A zinc finger ATF can have thousands of binding sites in the genome, much like natural TFs, and perturb the transcriptome on a genomic scale. By contrast, highly specific DBDs such as TAL-effectors or RNA-guided CRISPR/Cas systems that target single genomic loci are less likely to perturb multiple nodes in a network and alter the homeostatic state to induce a change in cell fate. ii) Regulatory potency. Among the DDs, zinc fingers have the unique ability to target both active and silenced regions of the genome, a feature important for cell fate conversions. Certain naturally occurring families of C2H2 zinc fingers bind methylated DNA and heterochromatin (65, 68), making zinc fingers well suited for activating epigenetically silenced regions unlike TAL-effectors, which are sensitive to DNA methylation (61, 69). Although CRISPR/Cas9 systems are set.

In addition to providing a means to perform a forward genetic screen, we demonstrate that our ATF library is a powerful resource for everyone who is interested in inducing cell fate conversion in the absence of a priori knowledge of natural TFs that might govern the desired cell type or phenotype. Furthermore, we describe a strategy toward identifying cell fate-defining transcriptional networks. By integrating expression data with in vitro binding site data, we were able to identify the nodes of the transcriptional network implicated in the induction of pluripotency. This technology enables the pursuit of elusive cell phenotypes or direct conversions, considered challenging to achieve by conventional methods. In summary, this study provides compelling support of our design principles and demonstrates that our ATF library can be used in a gain-of-function screen for complex cell fate conversions.

Materials and Methods

Design: Choice of DBD Scaffold. In designing an ATF library for cell fate conversions, we first focused on choosing a DNA-binding scaffold that would be most conducive for performing a forward genetic screen. As noted above, the following criteria were considered.

i) Ability to target multiple nodes. Although nuclease-inactivated CRISPR/Cas9 (66) and TAL-effectors (67) require at least at 10–20 bp to bind DNA, zinc fingers can be designed to target a wider range of sequences. The breadth of binding sites of zinc fingers is an important advantage. A zinc finger ATF can have thousands of binding sites in the genome, much like natural TFs, and perturb the transcriptome on a genomic scale. By contrast, highly specific DBDs such as TAL-effectors or RNA-guided CRISPR/Cas systems that target single genomic loci are less likely to perturb multiple nodes in a network and alter the homeostatic state to induce a change in cell fate.

ii) Regulatory potency. Among the DDs, zinc fingers have the unique ability to target both active and silenced regions of the genome, a feature important for cell fate conversions. Certain naturally occurring families of C2H2 zinc fingers bind methylated DNA and heterochromatin (65, 68), making zinc fingers well suited for activating epigenetically silenced regions unlike TAL-effectors, which are sensitive to DNA methylation (61, 69). Although CRISPR/Cas9 systems can activate silenced genes, they show limited ability to increase the expression of genes that are already expressed at moderate levels (63, 70). Additionally, compared with zinc fingers and TAL-effectors that can up-regulate genes to biologically relevant levels (42, 71, 72), the magnitude of transcriptional change induced by nuclease-inactivated CRISPR/Cas9 systems with a single guide is not as robust (73). Recent modifications to the CRISPR/Cas9 system
system have improved their impact on the level of expression of target genes; however, these modifications come at the expense of increasing their size (70). Therefore, the rules governing polyclad cell permeability are still not well understood, making delivery problematic (78). In brief, the zinc finger DBD emerges as an optimal scaffold for design of complex ATF libraries that can trigger cell fate-defining gene networks.

Zinc Finger ATF Library. The scaffold of the ATF is comprised of N to C terminus: a 15-aa ID, the DNA binding domain of human EGR1, NLS from EGR1, VP64, and 3x HA tag. The ATF library was created by amplifying oligos with VNN codons at the −1, 2, 3, and 6 positions relative to the recognition helix of each zinc finger. The ATF library was cloned into the second-generation p53N vector by ligase-independent cloning as described in SI Appendix, SI Materials and Methods.

Cell Culture. Oct4: CreER<sup>Cre-mCherry</sup>, mTmG MEFs were grown in DMEM supplemented with 10% (vol/vol) FBS on plates coated with 0.1% gelatin. Mouse E14T ES cells and iP cells were grown in knockout DMEM supplemented with 15% (vol/vol) FBS, 1% nonessential amino acids, 2 mM l-glutamine, 1 x 10<sup>5</sup> units/mL leukemia inhibitory factor, 1 mM sodium pyruvate, and 100 μM β-mercaptoethanol. During reprogramming of Oct4-CreER<sup>Cre-mCherry</sup>, mTmG MEFs, 4-hydroxytamoxifen was added at 100 nM concentration. Cells were maintained in a humidified 37 °C incubator with 5% CO2. Additional details are in SI Appendix, SI Materials and Methods.

Luciferase Assay. The palindromic EGR1 binding site 5′-GGC-TGG-GGC-GGC-GGC-GGC-3′ was cloned upstream of the luciferase gene in the pGL3 basic vector (Promega). Luciferase assay (Promega; E4030) was performed according to the manufacturer’s guidelines. Additional details are in SI Appendix, SI Materials and Methods.

Retrospl production. Oct4, Sox2, Klf4, and c-Myc were packaged into retrovirus with Plat-E cells as described in ref. 79.

Lentivirus production. ATFs and the empty control were packaged into lentivirus with HEK293FT cells using a second-generation lentiviral system. Details are described in SI Appendix, SI Materials and Methods.

Identification of ATFs from Single Cells. Cells with a positive phenotype for Oct4 lineage tracing activation were isolated as single cells into a 96-well plate. Nest two-step PCR is described in SI Appendix, SI Materials and Methods.

EB Formation. For EB formation, pluripotent cells were seeded into ultralow-adhesion dishes at a concentration of 1 x 10<sup>5</sup> cells/mL in knockout DMEM supplemented with 15% (vol/vol) FBS, 1% nonessential amino acids, 2 mM l-glutamine, 1 mM sodium pyruvate, and 100 μM β-mercaptoethanol. Media was changed the day after seeding and every 2 d thereafter. EBs were collected on days 7, 11, and 14 for quantitative RT-PCR (RT-qPCR) and immunofluorescence. Cells were maintained in a humidified 37 °C incubator with 5% CO2.


Immunoﬂuorescence. EBs were plated on poly-L-lysine on day 14 to culture EB outgrowths. IPS cells were plated on glass slides coated with 0.1% gelatin for immunofluorescence. Antibody sources and dilutions are described in SI Appendix, SI Materials and Methods.

RT-qPCR. RNA was extracted from cells with RNeasy Mini Kit (Qiagen; 74104). RNA was converted into cDNA with SuperScript III First-Strand Synthesis System (Thermo Fisher; 18080051). qPCR was performed with Bullseye EvaGreen qPCR Mix with low ROX (Midwest Scientific; BEOPCR-LR). Primer sets are listed in SI Appendix, SI Materials and Methods.

ChIP. For ChIP, 5 x 10<sup>5</sup> cells were fixed in 1.5% (vol/vol) formaldehyde for 15 min. Harvested cells were ﬂash frozen, and then sonicated and lysed. Lysates were precleared and immunoprecipitated overnight with H3K27ac antibody (Abcam; ab17279), H3K9me3 antibody (Abcam; ab8980), or HA antibody (Abcam; ab9110) at 4 °C. Immunoprecipitated histone marks were puriﬁed with protein G magnetic beads (Life Technologies; 10004D) and a series of ﬁve washes. Cross-links of protein–DNA complexes were reversed by incubating at 65 °C for 6 h. Eluted DNA was treated with RNase A and Proteinase K. Additional details are in SI Appendix, SI Materials and Methods.

RNA-seq analysis. Reads were aligned with Bowtie2, version 2.2.5, to either the human genome hg19 (HEK293) or mouse genome mm10 (MEFs or cells derived from MEFs). Counts were quantiﬁed with Cufflinks, and differential expression was determined by Cuffdiff (80).

ChIP-seq analysis. Reads were annotated to the mouse genome mm10 with Bowtie2, version 2.2.5. Output sam ﬁles were converted to bam ﬁles, sorted, and indexed with Samtools 1.3. H3K27ac, H3K9me3, and HA peaks were called with MACS2 2.2.1. Differential peak signals were determined by DiffBind 1.16.2. ChIP peaks were visualized with Integrative Genomics Viewer (IGV). Additional details are in SI Appendix, SI Materials and Methods.

CSL. CSI was performed by incubating cell lysates containing zinc ﬁnger ATFs with randomized permutations of 25-bp sequences. The ATF–DNA complexes were immunoprecipitated with HA magnetic beads. Three rounds of enrichment were performed, and all three rounds of enrichment were sequenced. Experimental details and bioinformatic analysis are described in SI Appendix, SI Materials and Methods.

ACKNOWLEDGMENTS. We thank Sandra Tseng and Graham Erwin for help with ChIP-seq analysis. We thank José Rodriguez-Martinez for CSI analysis and Christina Shafer with RNA-seq analysis. We also acknowledge Mitchell Probosch with help with ﬂow cytometry, Jennifer Bolin for help with Illumina sequencing, and Bret Dufﬁn for the teratoma assay. We thank Laura Vanderplog for help with ﬁgure graphics. We are also grateful to Judith Kimble, Sushmita Roy, Garrett Lee, and Fang Wan for helpful discussions. This work was supported by the NIH Grant HL099773, W. M. Keck Medical Research Award, and Progenitor Cell Biology Consortium Jump-Start Award SU01HL099975-05 (Subaward 101330A). A.E. was supported by the Morgridge Biotechnology Wisconsin Distinguished Fellowship Award and the Stem Cell and Regenerative Medicine Training Award. A.S.K. was supported by Genomic Sciences Training Program Grant ST32HG002760. D.B. was supported by the National Science Foundation–Fanosca Science and Engineering Center grant.


