Cell Reports

Resource

Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons

Highlights

- Pioneer factor Ascl1 opens closed chromatin at its target sites within 12 hr
- Rapid chromatin remodeling in reprogramming precedes neuronal maturation
- Ascl1 induces well-structured and stable nucleosome positioning by day 5
- Network model identifies Zfp238, Sox8, and Dlx3 as critical TFs for iN cell reprogramming

Authors

Orly L. Wapinski, Qian Yi Lee, Albert C. Chen, ..., Nadia Dahmane, Marius Wernig, Howard Y. Chang

Correspondence

wernig@stanford.edu (M.W.), howchang@stanford.edu (H.Y.C.)

In Brief

Wapinski et al. find that "on-target" pioneer factor Ascl1 induces rapid, genome-wide chromatin remodeling and nucleosome phasing that precedes other indicators of neuronal maturation. Building a network model incorporating chromatin accessibility and transcriptomic changes allows identification of Zfp238, Sox8, and Dlx3 as critical downstream factors for iN cell reprogramming.

Wapinski et al., 2017, Cell Reports *20*, 3236–3247
September 26, 2017
http://dx.doi.org/10.1016/j.celrep.2017.09.011

CellPress

OPEN ACCESS
CellPress

Cell Reports
Resource

Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons

Orly L. Wapinski,[1,10] Qian Yi Lee,[2,3,10] Albert C. Chen,[1] Rui Li,[1] M. Ryan Corces,[1] Cheen Euong Ang,[2,3] Barbara Treutlein,[4,5] Chaomei Xiang,[6,7] Valérie Baubet,[6,8] Fabian Patrik Suchy,[2] Venkat Sankar,[2,9] Sopheak Sim,[3] Stephen R. Quake,[3] Nadia Dahmane,[6,7] Marius Wernig,[2,*] and Howard Y. Chang[1,11,*]

[1]Center for Personal Dynamic Regulomes and Program in Epithelial Biology, Stanford University, Stanford, CA 94305, USA
[2]Institute for Stem Cell Biology and Regenerative Medicine, Department of Pathology, Stanford University, Stanford, CA 94305, USA
[3]Department of Bioengineering, Stanford University, Stanford, CA 94305, USA
[4]Max Planck Institute for Evolutionary Anthropology, 04103 Leipzig, Germany
[5]Department of Biosciences, Technical University Munich, 85354 Freising, Germany
[6]Department of Neurosurgery, University of Pennsylvania, Philadelphia, PA 19104, USA
[7]Department of Neurological Surgery, Weill Cornell Medicine, New York, NY 10065, USA
[8]Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
[9]The Harker School, San Jose, CA 95117, USA
[10]These authors contributed equally
[11]Lead Contact
*Correspondence: wernig@stanford.edu (M.W.), howchang@stanford.edu (H.Y.C.)
http://dx.doi.org/10.1016/j.celrep.2017.09.011

SUMMARY

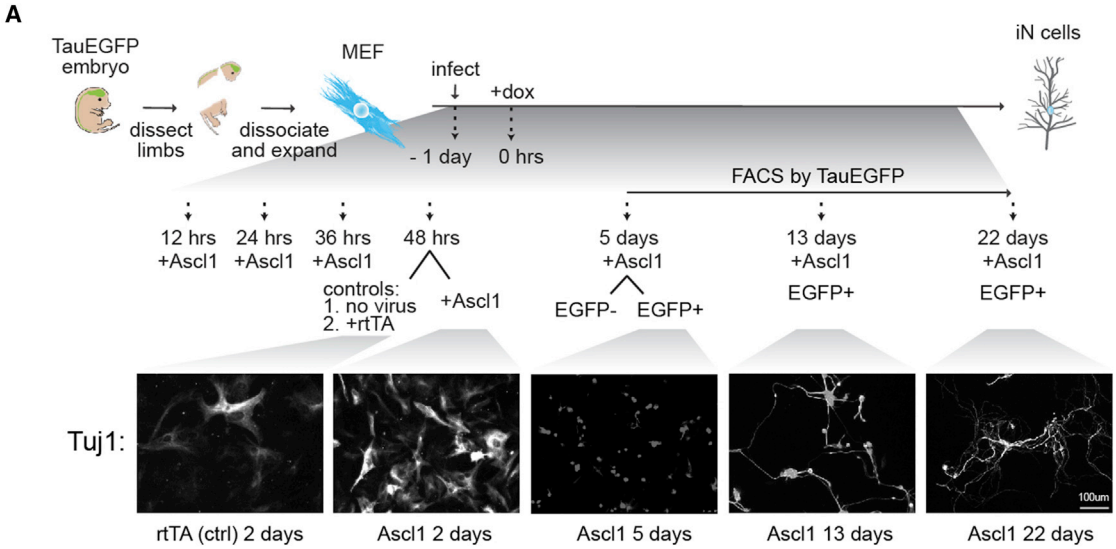
How transcription factors (TFs) reprogram one cell lineage to another remains unclear. Here, we define chromatin accessibility changes induced by the pro-neural TF Ascl1 throughout conversion of fibroblasts into induced neuronal (iN) cells. Thousands of genomic loci are affected as early as 12 hr after Ascl1 induction. Surprisingly, over 80% of the accessibility changes occur between days 2 and 5 of the 3-week reprogramming process. This chromatin switch coincides with robust activation of endogenous neuronal TFs and nucleosome phasing of neuronal promoters and enhancers. Subsequent morphological and functional maturation of iN cells is accomplished with relatively little chromatin reconfiguration. By integrating chromatin accessibility and transcriptome changes, we built a network model of dynamic TF regulation during iN cell reprogramming and identified Zfp238, Sox8, and Dlx3 as key TFs downstream of Ascl1. These results reveal a singular, coordinated epigenomic switch during direct reprogramming, in contrast to stepwise cell fate transitions in development.

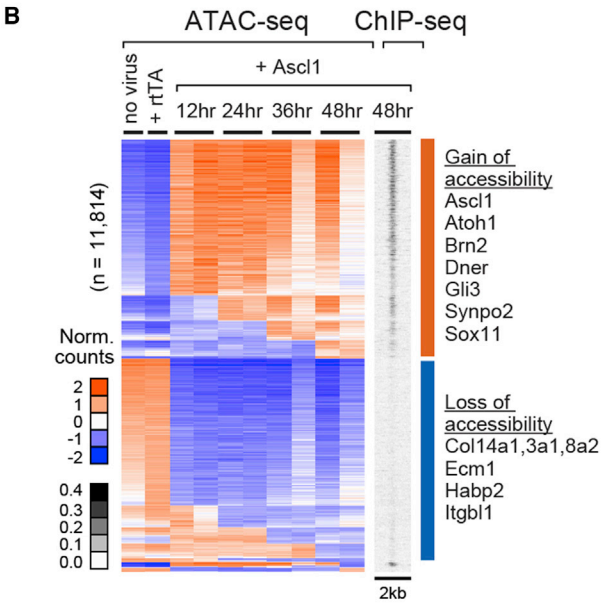
INTRODUCTION

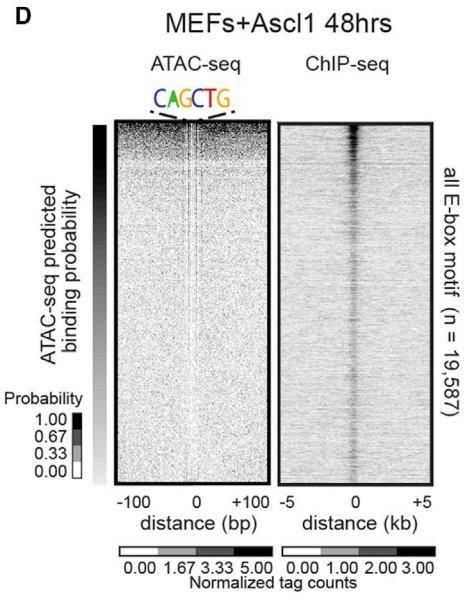
The ability of master regulators to redefine somatic cell fate has substantially added to our understanding of stem cell biology and development. Ectopic expression of select transcription factors (TFs) can reprogram terminally differentiated cell types into a pluripotent state (Okita et al., 2007; Park et al., 2008; Takahashi and Yamanaka, 2006; Takahashi et al., 2007; Wernig et al., 2007; Yu et al., 2007) or directly into somatic cells of unrelated lineages (Davis et al., 1987; Huang et al., 2011; Pang et al., 2011; Vierbuchen et al., 2010).

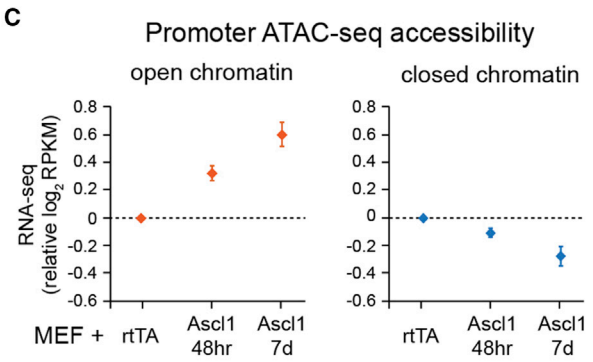
Given their ability to access their targets in a non-permissive state, it has been recently suggested that some of these TFs act as "pioneer factors" (Berkes et al., 2004; Gerber et al., 1997; Soufi et al., 2012, 2015; Wapinski et al., 2013). However, pioneering activity can differ markedly depending on the master regulator involved. In reprogramming fibroblasts to induced pluripotent stem cells (iPSCs), Oct4, Sox2, and Klf4 access different sites in fibroblasts than in iPSCs in a mostly cooperative manner (Chronis et al., 2017; Soufi et al., 2012) and DNA elements preferentially pioneered contain only parts of the canonical binding motifs (Soufi et al., 2015). Thus, secondary events must occur that eventually lead to proper TF binding as is reflected by sequential waves of gene expression programs that recapitulate steps in early embryonic development and patterning (Cacchiarelli et al., 2015).

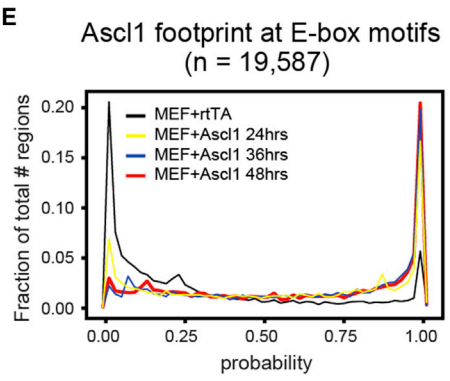
In contrast, during reprogramming of fibroblasts to induced neuronal (iN) cells using Ascl1, Brn2, and Myt1l, the pro-neural basic helix-loop-helix (bHLH) factor Ascl1 acts as an "on-target" pioneer factor, binding to its physiological targets even in closed chromatin regions, and actively recruits other TFs to some of its targets sites (Wapinski et al., 2013). This finding suggested that Ascl1 may be the most powerful of the three reprogramming factors and a strong activator of the neuronal program. Indeed, under optimized conditions, Ascl1 alone can reprogram fibroblasts to fully functional iN cells, albeit with lower efficiencies (Chanda et al., 2014). Although the on-target pioneering nature of Ascl1 is now well documented, very little is known about the subsequent dynamics of Ascl1 binding and the resulting alteration of the chromatin landscape over the course of reprogramming (Chanda et al., 2014; Wapinski et al., 2013; Yao et al., 2013). Recent single-cell RNA-sequencing (RNA-seq) experiments showed remarkably homogeneous initiation of transcriptome reprogramming followed by the emergence of several possible transcriptional programs (Treutlein et al.,











(legend on next page)

2016), highlighting the need to examine their possible origins at the chromatin level.

Here, we explored the chromatin dynamics of iN cell reprogramming induced by Ascl1 by measuring chromatin accessibility, a cardinal feature of active regulatory DNA. Eukaryotic genomes are extensively compacted by chromatin, except at active regulatory elements such as enhancers, promoters, and insulators. Prior methods of tracking chromatin accessibility were impractical as they often required tens of millions of cells, which was challenging to obtain due to low reprogramming efficiencies and cell death, particularly at later time points. The advent of assay of transposase accessibly chromatin with sequencing (ATAC-seq) provided a new and sensitive way to track open chromatin regions and predict TFs binding and nucleosome positions with as few as 500 cells (Buenrostro et al., 2013, 2015). Thus, ATAC-seq allowed us to better study epigenetic changes in a genome-wide fashion through the course of Ascl1-mediated reprogramming.

RESULTS

Ascl1 Induces Widespread Chromatin Remodeling in Fibroblasts within Hours

We used ATAC-seq to measure chromatin accessibility dynamics in mouse embryonic fibroblasts (MEFs) as they are reprogrammed into iN cells using Ascl1. Substantial transcriptional responses to Ascl1 in MEFs occur over the first 48 hr (Wapinski et al., 2013), preceding any overt morphological changes (Figure 1A). After 48 hr, the cultures become heterogeneous, and around day 5, the cells undergoing productive reprogramming start to induce the neuronal reporter gene TauEGFP. We therefore sampled cells for ATAC-seq at early time points in short intervals (12, 24, 36, and 48 hr) and isolated TauEGFP$^+$ cells at later stages of reprogramming (days 5, 13, and 22) by fluorescence-activated cell sorting (FACS) (Figure 1A).

TauEGFP$^+$ cells first appear with mostly rounded morphologies and occasional outgrowth of a short, non-tapering process. These morphologically distinct cells do not exhibit yet any functional neuronal properties. By day 13, a large number of cells exhibit clear neuronal morphologies with several dendritic branch points, express many pan-neuronal markers, and begin to express more mature synaptic markers such as synaptophysin. iN cells at day 22 have matured to fully functional neurons with the ability to fire action potentials and form synapses (Figure 1A) (Chanda et al., 2014). As controls, we used uninfected MEFs, as well as MEFs infected with the reverse tetracycline transactivator (rtTA) alone, and sampled these cells at 48 hr. Each ATAC-seq library met quality control metrics and was sequenced to obtain at least 20 million mapped reads (Qu et al., 2015). In total, we identified 233,650 sites of dynamic DNA access during the course of iN cell reprogramming.

Remarkably, thousands of genomic sites changed accessibility after a mere 12 hr of induction, swelling up to 11,814 sites that changed significantly compared to rtTA control by 48 hr (Figure 1B). We observed a high level of correlation between the two biological controls (Figure S1A), indicating that rtTA alone has little effect on chromatin accessibility. We used the StepMiner algorithm (Sahoo et al., 2007) to detect transitions in DNA accessibility patterns over time. Many of the sites gaining DNA accessibility (red) are either direct Ascl1 targets as defined by chromatin immunoprecipitation sequencing (ChIP-seq) at 48 hr after Ascl1 induction (Figure 1B; Figure S1B), or contained at least one CAGCTG E-box motif—which are enriched in Ascl1 ChIP-seq peaks (Figure S1C)—indicating the direct action of Ascl1 at its binding sites. In agreement with the notion that Ascl1 is primarily a transcriptional activator (Castro et al., 2011; Wapinski et al., 2013), the genes associated with sites gaining accessibility are on average induced (Figure 1C). Remarkably, however, quite a large number of sites are losing DNA accessibility in a similarly rapid manner and many of those sites do not contain E-box motifs, suggesting those changes are not directly regulated by Ascl1.

The overwhelming majority of chromatin sites that changed are located in intergenic or intronic regions (Figure S1D), consistent with the distribution of Ascl1 ChIP-seq peaks (Wapinski et al., 2013). Using genomic regions enrichment of annotations tool (GREAT) (McLean et al., 2010) to associate the enhancer and promoter regions to genes, we observed an enrichment of neuronal gene ontology (GO) terms, as well as a muscle GO term (Z disc) for opening sites. In contrast, extracellular matrix GO terms are enriched at closing sites. The observed changes at the chromatin level reflect the previously described

Figure 1. Rapid Chromatin Changes in Response to Ascl1 Induction during Early Stages of Reprogramming

(A) Top: Schematic of the study. Bottom: Representative Tuj1-immunostained images of cells at days 2, 5, 13, and 22 post-Ascl1 induction, as well as control cells only infected with rtTA. Tuj1 is upregulated within 48 hr of Ascl1 induction, but morphology of cells are still flat and fibroblast-like. By day 5, we observe rounded cell bodies that sometimes extend into short processes. By day 13, we see obvious neuronal morphology that continues to mature and gain complexity through to day 22.

(B) Left: Heatmap showing dynamic chromatin changes during first 48 hr of iN cell reprogramming. Sites gain (red) and lose (blue) DNA accessibility as early as 12 hr after induction of Ascl1, and these changes persist up to 48 hr. Opening chromatin regions are associated neuronal genes, whereas closing regions are associated with more general ECM terms. Right: Heatmap of normalized tag densities showing Ascl1 binding profile centered around the ATAC-seq peak on the corresponding row. Sites that gain DNA accessibility are more strongly bound by Ascl1 compared to sites that lose accessibility.

(C) Average $\log_2$(fold-change) of genes with promoters (±5 kb from TSS) associated with open (red) or closed (blue) chromatin regions for each time point, relative to the MEF+rtTA controls (mean ± SEM, n = 2). Open chromatin is generally associated with promoters of upregulated genes, whereas closed chromatin is associated with downregulated genes.

(D) Heatmaps of normalized tag densities representing transposase accessibility and the occupancy profile of Ascl1 at 48 hr for all E-box motifs in the genome, sorted based on intensity of Ascl1 binding. Ascl1's footprint is clearly shown in regions with a stronger ChIP signal, with a correspondingly higher ATAC-seq predicted binding probability.

(E) Binding probability at E-box motifs inferred from ATAC-seq data at each of the indicated time points. The x axis is the probability of motif occupancy, and 1.00 (100%) indicates high probability of binding. The y axis is the fraction of all E-box motifs in the genome that have the indicated binding probability.

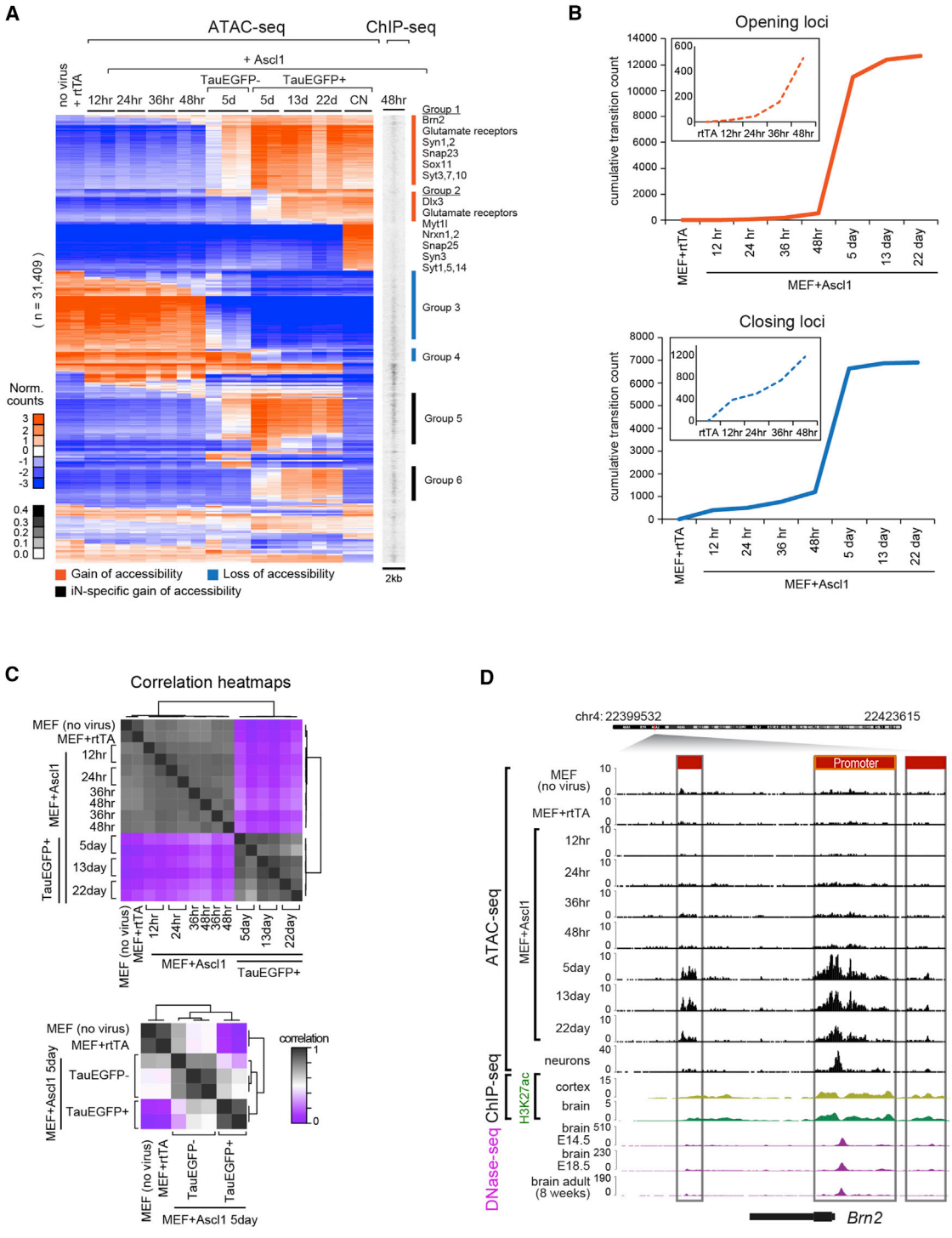


Figure 2. A Transition Point Occurs at Day 5 That Distinguishes between Early and Late Maturation Programs

(A) Left: Heatmap showing chromatin changes through the course of reprogramming (CN: E15.5 TauEGFP$^+$ cortical neurons). Majority of chromatin changes occur between 48 hr and day 5 after Ascl1 induction, and successful iN cell intermediates retain a stable chromatin conformation. Only subtle chromatin changes are observed at later time points. Groups 1 and 2 (red) are sites that increase in accessibility during reprogramming, whereas groups 3 and 4 (blue) decrease in accessibility. Groups 5 and 6 (black) represent sites that only become accessible in reprogramming cells but not in cortical neurons. Right: Heatmap of normalized tag densities representing Ascl1 binding profile at 48 hr centered on the corresponding ATAC-seq peaks. Peaks from groups 5 and 6 have a stronger Ascl1 ChIP-seq signal.

(B) Cumulative plots of transition counts for opening (red) and closing (blue) chromatin regions. Note the sharp transition point at day 5.

(legend continued on next page)

transcriptional response to Ascl1 (Figures S1E and S1F) (Treutlein et al., 2016).

We next calculated Ascl1's binding probability by computing the DNA accessibility, as defined by the number and pattern of ATAC-seq tag densities at all E-box motifs in the mouse genome, and compared the signal to Ascl1 genomic occupancy by ChIP-seq. We plotted heatmaps of tag densities representing transposase accessibility and Ascl1 ChIP-seq signal at 48 hr, which are centered around all E-box motifs in the genome and sorted based on the intensity of Ascl1 binding (Figure 1D). Ascl1's footprint is apparent within the ATAC-seq data in regions that have strong ChIP-seq signal. These sites have a high ATAC-seq binding probability as determined using CENTIPEDE (Pique-Regi et al., 2011). Because Ascl1's footprint strongly correlates with its genomic localization at 48 hr, we used the binding probability as a metric to define whether Ascl1 was bound at a specific site and at earlier time points for which we did not have ChIP-seq data. During the process, we observed a strong shift in Ascl1's binding probability from 0, for the vast majority of E-box motifs in the MEF+rtTA starting population, to 1, for more than 20% of the E-box motifs at 48 hr after Ascl1 induction (Figure 1E). Moreover, we implemented an enhancer cytometry assessment, which is a method to resolve the representative cellular populations on a mixed sample on the basis of chromatin accessibility at enhancers (Corces et al., 2016; Newman et al., 2015). We detected a homogeneous chromatin profile (Figure S1G), consistent with previous single-cell RNA-seq data (Treutlein et al., 2016).

These data indicate that reprogramming is a highly dynamic process, whereby chromatin remodeling begins as early as 12 hr after Ascl1 induction. Moreover, it indicates that the early increase in chromatin accessibility is largely and directly driven by Ascl1. Consistent with previously obtained transcriptomic data (Treutlein et al., 2016), sites that gain accessibility are enriched for both neuronal and muscle GO terms, suggesting a promiscuity of Ascl1 genomic targeting that can be observed at the chromatin level even in the early stages of reprogramming.

A Major Transition at Day 5 Distinguishes Early and Late Neuronal Programs

To further dissect the reprogramming mechanisms, we analyzed the ATAC-seq results at mature time points. Despite the large number of regions that changed in accessibility within 48 hr of Ascl1 induction, we observed a more striking number of sites that undergo chromatin changes between 48 hr and day 5 of reprogramming (Figure 2A). This chromatin "switch" (illustrated in Figure 2B) precedes phenotypic signs of neuronal maturation, such as membrane channel composition, axonal and dendritic specialization and transport, and synapse formation. After day 5, chromatin changes merely comprise less than 20% of all changes (Figures 2A and 2B). Moreover, pairwise Pearson correlation of ATAC-seq signals across all time points shows a sharp demarcation between 48 hr and day 5 (Figure 2C).

Regulatory elements that become accessible at the day 5 chromatin transition in TauEGFP$^+$ iN cells (Figure 2A) are enriched for genes associated with neuronal processes, such as development of neurite networks (group 1), and synaptic maturation (group 2). They contain accessible peaks that most resemble TauEGFP$^+$ cortical neurons isolated from embryonic day 15.5 (E15.5) mice and are associated with various genes found in mature synapses, such as neurotransmitter receptors, neurexins, and synapsins (Figure 2A; Figure S2A). These sites are strongly associated with promoter regions compared to other groups (Figure S2B). Concurrently, sites that lose accessibility at day 5 are associated with broad biological functions and extracellular matrix GO terms (groups 3 and 4), both of which are associated with the starting fibroblast identity (Figure 2A; Figure S2A). Groups 5 and 6 consist of peaks uniquely accessible in iN cells, and not in cortical neurons. The vast majority of these regions are at intronic and intergenic regions, and lack any significant GO term enrichment (Figures S2A and S2B). These sites are strongly bound by Ascl1, which could contribute to aberrant gene induction or misassignment between ATAC-seq peak and nearest gene (Figure 2A).

The endogenous loci of *Brn2* (group 1) and *Myt1l* (group 2), both key factors that improve the efficiency of iN cell reprogramming (Chanda et al., 2014; Vierbuchen et al., 2010), also become open by day 5 (Figure 2D; Figure S2C). A closer examination of the *Brn2* locus reveals three putative regulatory regions (promoter and gene tail) that are accessible in primary neurons and gain accessibility through the course of MEF-to-iN cell reprogramming. These regulatory regions are marked by histone H3 lysine 27 acetylation (H3K27ac) and DNase-I hypersensitivity in the embryonic and adult mouse brain, both of which are associated with an active enhancer state (Figure 2D). This indicates that the initially repressed *Brn2* locus gains accessibility upon Ascl1 induction by day 5 and retains that accessibility through the rest of reprogramming.

Similarly, at day 5, we examined the chromatin landscape of TauEGFP$^-$ cells, which fail to reprogram due to Ascl1 downregulation and deviation from the reprogramming path (Treutlein et al., 2016). Correlation analysis shows that TauEGFP$^-$ cells lie between the starting population of MEFs and TauEGFP$^+$ intermediates, whereas day 5 TauEGFP$^+$ intermediates cluster closer to days 13 and 22 iN cells (Figure 2C). In particular, TauEGFP$^-$ cells fail to activate the regulatory regions of genes in group 2 and fail to repress genes in group 4 (Figure 2A).

In summary, our data suggest that day 5 post-Ascl1 induction represents the major transition point with respect to remodeling the fibroblast to a neuronal chromatin state.

Enhancer Remodeling and Nucleosome Phasing of Ascl1-Bound Sites Mark the Reprogramming Transition

Because Ascl1 induction triggered dynamic chromatin rearrangements, we next interrogated the global effects of Ascl1

(C) Top: Correlation plot of ATAC-seq peaks through the course of reprogramming show a sharp transition at day 5. Early time point samples cluster more closely to the controls, whereas day 5 TauEGFP$^+$ samples cluster more closely to day 13 and 22 cells. Bottom: A closer look at the day 5 transition point shows that TauEGFP$^-$ cells that fail to reprogram cluster between MEFs and TauEGFP$^+$ iN cell intermediates.
(D) Chromatin dynamics at *Brn2* locus. Genomic tracks showing ATAC-seq data during iN cell reprogramming and H3K27ac and DNase-I hypersensitivity in brain.

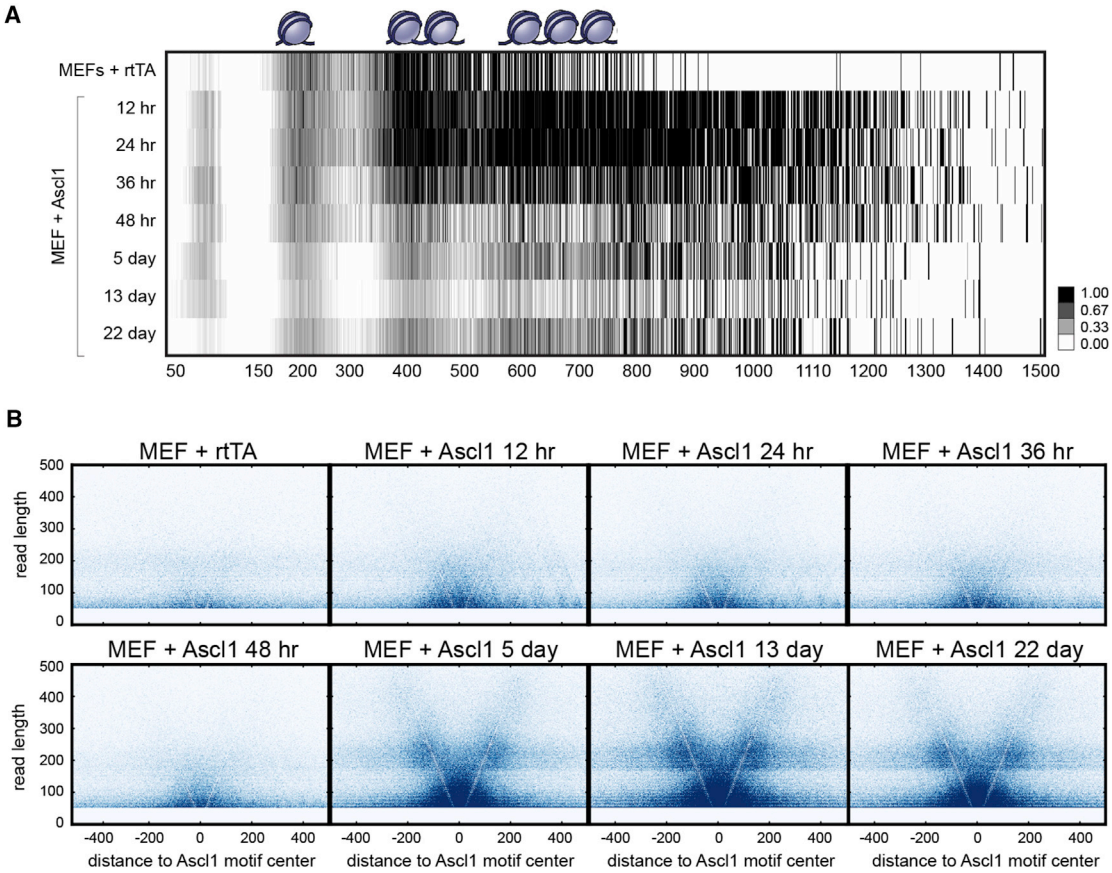


Figure 3. Enhancer Remodeling and Nucleosome Phasing at Ascl1 Sites during iN Cell Reprogramming

(A) Heatmap showing ATAC-seq read densities of reads within 100 bp of an E-box motif bound by Ascl1 at different insert sizes indicative of open chromatin and single nucleosomes, di-nucleosomes, or tri-nucleosomes. Note the striped pattern indicative of nucleosome phasing being established at day 5.

(B) V-plot of number of ATAC-seq reads as function of read length and distance from an E-box motif in the center Ascl1 binding sites. All data are centered on Ascl1 motifs across the genome. Note that the day 5 sample starts to exhibit the mature iN pattern: The short (<150-bp) fragments indicative of nucleosome-free DNA are at the Ascl1 motif center, and the accumulation of ~200-bp fragments (single nucleosome) starting on day 5 are positioned next to the Ascl1 sites, indicative of the +1 and −1 nucleosomes, followed by accumulation of ~400-bp fragments (second nucleosome) that are further away.

on the chromatin. A useful feature of ATAC-seq data are its ability to report nucleosome position and the degree of chromatin compaction (Buenrostro et al., 2013; Schep et al., 2015). The periodicity observed on recovered ATAC-seq fragments as multiples of approximately 150- to 200-bp lengths is reflective of the number of nucleosomes. Single nucleosomes generate an approximate 200-bp insert length, whereas compacted di- and tri-nucleosomes generate 400- and 600-bp fragments, respectively. Smaller than 150-bp reads account for nucleosomal free DNA.

We filtered the ATAC-seq reads that overlapped within 100 bp of an E-box bound by Ascl1 and binned the reads based on length. We then plotted the enrichment of differentially sized DNA fragments at each of the time points around Ascl1 binding sites (Figure 3A). This analysis showed a substantial genome-wide remodeling of the chromatin architecture near Ascl1 binding events during Ascl1-induced reprogramming.

Most striking was the gain of shorter fragments (less than 150 bp) that became prominent as early as 12 hr after Ascl1 induction relative to the starting MEF population, and remain enriched throughout the mature time points. These regions represent DNA that gains accessibility at Ascl1 target sites, demonstrating its pioneering activity (Figure 3A). Additionally, we noticed a global effect of nucleosome shuffling with no clear pattern occurring at the early time points. However, a clear transition occurs by day 5, when the nucleosome organization attains a stable configuration as noted by the nucleosomal periodicity maintained through late time points (Figure 3A).

Subsequently, we assayed the rearrangement of the chromatin landscape and nucleosome positioning at Ascl1-bound sites by mapping the distribution of differentially sized DNA fragments centered on Ascl1's motif (Figure 3B) (Henikoff et al., 2011). We found an enrichment of short fragments (less than 150 bp) as early as 12 hr after Ascl1 induction, suggesting a fast engagement of Ascl1 with the chromatin to make it more accessible. The enrichment of short reads at Ascl1 target sites increases throughout ensuing stages of reprogramming (Figure 3B; Figure S3A). Additionally, we observed an enrichment

of fragments of 200- and 400-bp approximate lengths starting at day 5 and continuing through day 22. Larger fragments are equidistantly located (±200 bp) from the Ascl1 target sites, likely representing the stabilization of new flanking nucleosomes around Ascl1 binding sites (Figures S3A and S3B).

Thus, we find that the early stages of reprogramming are characterized by rapid accessibility of Ascl1 to its cognate sites, resulting in local yet extensive rearrangement of the chromatin architecture. The most drastic epigenetic transition takes place at day 5, in which nucleosomes are phased into a distinct but stable configuration to promote the execution of a mature neuronal program that includes a larger number of regulatory regions. Nucleosome phasing has been reported to be a critical feature for enhancer activation following TF binding to DNA, as exemplified for the estrogen receptor cistrome (Carroll et al., 2006).

Network Analysis Reveals Hierarchical TF Regulation and Identifies Downstream Effectors of iN Cell Reprogramming

Given the extensive number of regulatory regions that become accessible at day 5, we interrogated the transcriptional network that confers mature neuronal programs. We created a comprehensive TF-motif database encompassing 321 TF motifs (ENCODE Project Consortium, 2012; Sandelin et al., 2004). For each motif site on the genome, a machine-learning algorithm (PIQ) was applied to assign a probability that the site is bound by its cognate TF at each time point using ATAC-seq data (Sherwood et al., 2014). Using this pipeline, we inferred TF-TF networks for iN cells at day 5, 13, and 22. To increase the stringency of the network, we integrated RNA-seq data for the corresponding time points to select target genes that are changing during the iN cell time course (Figure S4A). The end result is a network model that can be drawn as a hierarchical graph, in which Ascl1 is placed at the top (Figure 4A). Each node is a TF gene; each edge represents a relationship between two genes. The directionality of the edge indicates that the TF (protein) encoded by the upstream node occupies the regulatory region of the downstream node (gene). For example, Ascl1 → Zfp238 means that Ascl1 (protein) occupies the *Zfp238* locus, as inferred by the pattern of DNA access of the Ascl1 motif in the regulatory region within the *Zfp238* locus.

The network model of the three mature time points contained between 44 and 56 TFs and 273 to 391 edges in each time point-specific network. Moreover, to interrogate the network dynamics, we computed the gain and loss of each node's outedges between time points. A core of 200 edges was maintained across time points, suggesting that the vast majority of the TF-TF regulation is maintained from day 5 to 22. However, we identified 149 edges that are gained from day 5 to day 13, most of which are lost by day 22. Although largely static after day 5, the network analysis captured small transitory cell states. Of important note, many of the network members are transiently expressed and downregulated at later reprogramming stages (Figures S4B and S4C). Thus, the Ascl1-centered transcriptional network seems to remain similar during reprogramming, but its relevance may decrease at later stages (Treutlein et al., 2016).

When representing the network in hierarchical structure, Ascl1 is displayed at the top. At the second tier are TFs that are directly regulated by Ascl1 (dark blue and dark green). The third tier is composed of TFs that are not directly regulated by Ascl1 (light blue and light green). Blue nodes represent TFs that regulate other TFs in the network. For the most part, the core of the network is largely similar, with slight differences in binding and modest TF inclusions and exclusions occurring at the level of second and third tiers at days 13 and 22.

Previous studies have shown that *Zfp238* (or *Zbtb18*, *Rp58*) is directly bound by Ascl1 during reprogramming, and can partially replace Ascl1 by reprogramming MEFs when co-expressed with Myt1l (Wapinski et al., 2013). Reassuringly, we find that our network also predicts *Zfp238* to be a direct target of Ascl1, and that Zfp238 has a relatively high degree of connectivity (total number of edges leading to or out of the TF) within the network (Figure 4A; Figure S4D). Using PIQ to predict the probability of Ascl1 occupancy at later time points, we also observed on average a higher purity score in TauEGFP$^+$ cells (>0.9) compared to TauEGFP$^-$ cells (0.65–0.8) at day 5 at Ascl2 motifs. These predicted binding sites fall within actual Ascl1 binding sites at the *Zfp238* transcription start sites (TSSs) as shown by ChIP-seq at 48 hr, indicating that successful iN cell reprogramming is associated with high degree of Ascl1 occupancy at the *Zfp238* promoter at day 5 (Figure S4E). To assess the role of Zfp238 during reprogramming, we induced Ascl1 in MEFs from *Zfp238* knockout (KO) (Xiang et al., 2012) or wild-type (WT) littermate mouse embryos (Figure S4F). *Zfp238* KO MEFs produced significantly fewer Tuj1$^+$ and less Map2$^+$ iN cells (Figure 4B), and cell counts indicate that *Zfp238* KO does not induce cell death (Figure S4G). These results show that Zfp238 is an essential downstream effector of Ascl1-induced reprogramming.

Next, we queried the network to identify novel effectors of Ascl1's program and test the most salient features of the model. In particular, we asked whether the temporal expression of the TF or its connectivity weighed on its ability to control a significant component of the iN cell reprogramming transcriptional program. Using bulk (Wapinski et al., 2013) and single-cell (Treutlein et al., 2016) RNA-seq data over the course of iN cell reprogramming, we analyzed the expression pattern for all TFs present in the network. Because TF families share similar motifs, we looked at the expression pattern of all corresponding members, and selected candidates with similar transient early expression pattern as Zfp238 (Dlx3, Sox4, Sox8) (Figure 4C; Figures S4B and S4C). In addition, we selected candidates with high degrees of connectivity (Egr1, Plagl1, Zfp161) (Figure S4D).

The combined infection of all six candidate genes was not sufficient to reprogram MEFs. Overexpression of well-connected TFs Egr1, Plagl1, and Zfp161 in combination with Myt1l also did not form neurons. However, the combined expression of either Sox8 or Dlx3 with Myt1l resulted in the formation of Tuj1$^+$ cells with typical neuronal morphology. The reprogramming efficiency of Sox8, a direct target of Ascl1, in combination with Myt1l, was comparable to that of Zfp238 with Myt1l (Figures 4D and 4E; Figure S4H). Dlx3 is in a lower position in the network hierarchy and has a lower reprogramming efficiency than either Sox8 or Zfp238. This finding indicates that the temporal expression of the TF is a better indicator of reprogramming ability (Figure 4B), and nodes more closely connected to Ascl1 also have a better reprogramming ability.

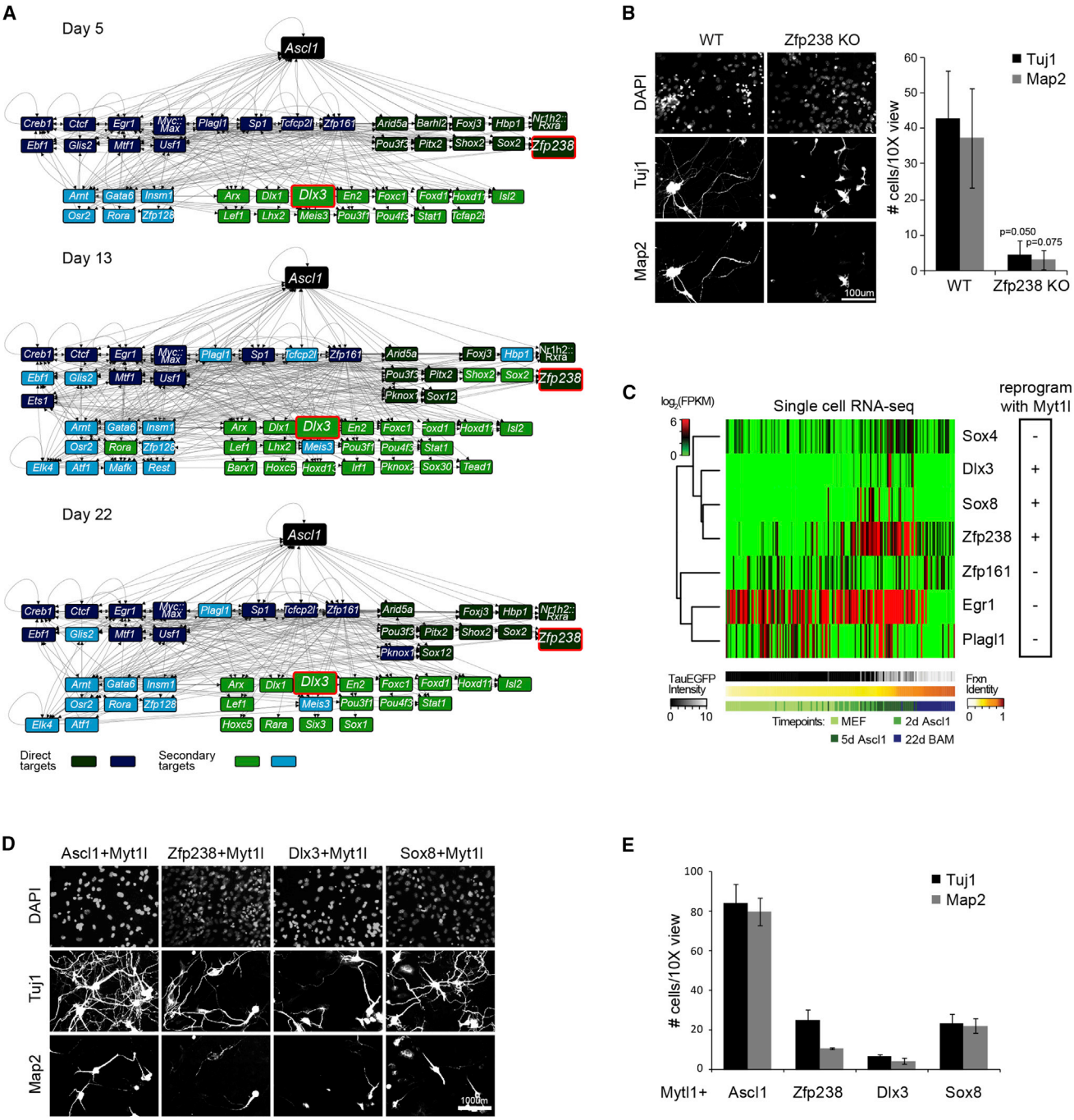


Figure 4. Network Analysis Validates Zfp238 as a Key Regulator and Identifies Downstream Effectors of Ascl1

(A) TF networks at the indicated time points of iN cell reprogramming. Each node is a TF, and each edge is a directed link indicating the first TF occupies the genomic locus of the second TF. TF nodes are arranged in the same order for all three time points. TFs directly bound by Ascl1 within ±5 kb of its TSS are in the second tier in dark blue or green, whereas TFs further downstream are in the third tier in light blue or green. Dark and light blue nodes indicate TFs that also bind to and regulate other TFs, whereas dark and light green nodes are TFs that have no outdegree. Key TFs described in the text are highlighted with a red border.

(B) *Zfp238* KO MEFs exhibit reduction in reprogramming capability, with a decrease in number of Tuj1$^+$ and Map2$^+$ neurons at 14 days after Ascl1 induction. Left: Tuj1 and Map2 immunostaining. Right: Counts of Tuj1$^+$ or Map2$^+$ neuronal cells (mean ± SEM, n = 3). Error bars indicate SE.

(C) Heatmap showing single-cell RNA-seq expression of downstream regulators during iN cell reprogramming, ordered by fractional neuronal identity. Cells at days 5 and 22 were sorted for TauEGFP. Dlx3, Sox4, and Sox8 have similar temporal expression patterns as Zfp238, being transiently upregulated in reprogramming cells around day 5. Zfp161, Plagl1, and Egr1 are highly connected nodes in the network (Figure S4B) but have more varied expression patterns.

(legend continued on next page)

DISCUSSION

In this study, we provide a view of the epigenomic landscape of somatic cell trans-differentiation, using direct reprogramming of fibroblasts to neurons. Surprisingly, we found a single major concerted chromatin switch in iN cell reprogramming. The chromatin switch can be considered "prescient" because the switch precedes and in fact anticipates the gene expression requirements for morphologic and functional maturation of iN cells.

The organizational logic of this concerted, rapid, and genome-wide chromatin remodeling "switch" that characterizes iN cell reprogramming seems to differ substantially from iPSC reprogramming based on recent chromatin mapping data of iPSC reprogramming (Cacchiarelli et al., 2015; Chronis et al., 2017), and also from physiologic development in either embryonic or somatic stem cells (Lara-Astiaso et al., 2014; Lopez-Pajares et al., 2015; Wang et al., 2015; Wu et al., 2016). Both iPSC reprogramming and normal development are thought to proceed through stepwise and gradual transitions, where prior steps bookmark enhancers for activation in subsequent stages (Lara-Astiaso et al., 2014). Recent studies of enforced ESC programming to motor neurons also revealed dynamic multi-step chromatin remodeling (Rhee et al., 2016; Velasco et al., 2017). It should be noted, however, that conventional Yamanaka-factor iPSC reprogramming is a slow and stochastic process (Buganim et al., 2012; Hanna et al., 2009; Smith et al., 2016). It will be interesting to assess whether a similar rapid chromatin switch occurs in recently developed high-efficiency iPSC reprogramming systems that involve the transient activation of Cebpα or timed, partial inhibition of Mbd3 (Rais et al., 2013; Di Stefano et al., 2014).

We propose that a TF needs to fulfill two main requirements to initiate such a concerted chromatin switch. First, it needs to be a pioneer factor capable of binding its cognate target sites in a "hostile" donor environment, rapidly reorganize the chromatin structure, and activate its target program. Second, it has to be a master regulator that can activate key endogenous TFs in the target program that can subsequently facilitate a major concerted cell fate transition to the target identity. Given that reprogramming of fibroblasts into cardiac and hepatic cells also involves lineage-specific pioneer factors, it will be interesting to explore whether a similarly rapid chromatin switch can be observed in these reprogramming scenarios (Huang et al., 2011; Ieda et al., 2010; Iwafuchi-Doi and Zaret, 2014; Sekiya and Suzuki, 2011).

In our iN cell system, Ascl1 is the pioneer factor and initially dominates the chromatin landscape, driving most of the increase in chromatin accessibility at the early time points and upregulating many Ascl1 target genes and neuronal genes. Reflecting our transcriptional findings, we detected some promiscuity in Ascl1-mediated chromatin remodeling at the early time points, whereby the regulatory regions of some myogenic genes gain accessibility within 12 hr of Ascl1 induction (Treutlein et al., 2016), suggesting that the determination of this alternative cell fate occurs on the chromatin level very early in reprogramming. Because both Ascl1 and Myod1 (a myogenic bHLH TF) bind a common E-box sequence (CAGCTG) (Castro et al., 2011; Fong et al., 2015), we speculate that the strong overexpression of Ascl1 led to the initiation of an off-target myogenic program that is not normally observed in a more tightly controlled endogenous environment. This is also consistent with our recent finding that Myt1l represses non-neuronal fates, including myogenic genes, suggesting potential synergy between Ascl1 and Myt1l (Mall et al., 2017).

By integrating ATAC-seq data at critical regulatory elements of the genome with gene expression, we built a network model that provides mechanistic insight into the hierarchy of events that occur during reprogramming. During this process, we identified critical TFs and determined their relative contribution to reprogramming. Remarkably, many of the genes in the Ascl1-induced reprogramming network are expressed in the ventral forebrain, just like Ascl1 itself. This suggests Ascl1-induced transcriptional cascades during reprogramming, which might be reminiscent of its physiological actions in vivo. Three candidates, Zfp238, Sox8, and Dlx3, were confirmed experimentally to play important functional roles. *Zfp238* is a target of Ascl1 and is implicated in neuronal differentiation (Baubet et al., 2012; Xiang et al., 2012). *Sox8* is also a direct target of Ascl1 and belongs to the SoxE subclass of the Sox family TFs. SoxE factors are implicated in many aspects of neuronal development, but the action of Sox8 specifically is not as well understood (Weider and Wegner, 2017). Dlx3 is reported to be expressed in the neural-related placodes during development (Beanan and Sargent, 2000), a site of additional proneural bHLH factor activity (Cau et al., 2000; Fode et al., 1998).

In conclusion, the epigenomic landscape during Ascl1-mediated direct reprogramming of fibroblasts into neurons shows a major concerted chromatin switch from the donor to target program that may be recapitulated in other direct reprogramming systems, but differs substantially from iPSC reprogramming and physiologic development. This poses intriguing implications in using direct reprogramming systems to model developmental mechanisms and also highlights properties required of pioneer factors that mediate concerted cell fate switches during development and for regenerative medicine.

EXPERIMENTAL PROCEDURES

Cell Derivation, Culture, and Viral Production

The following mouse protocols were approved by the IACUC committee in Stanford University; MEFs were isolated and co-infected with rtTA and doxycycline (dox)-inducible Ascl1 and allowed to reprogram as previously described (Vierbuchen et al., 2010). Cells were harvested at 12, 24, 36, and 48 hr, and days 5, 13, and 22 post-dox induction. Control MEFs were either not infected or infected with rtTA alone and harvested 48 hr after addition of dox. Neurons from E15.5 embryos were isolated and used as a positive neuronal control. See Supplemental Experimental Procedures for more details.

Right bar shows ability of each indicated TF to bypass Ascl1 and produce iN cells when co-expressed with Myt1l. Dlx3, Sox8, and Zfp238, when overexpressed in MEFs with Myt1l, gave rise to Tuj1$^+$ and Map2$^+$ cells with neuronal morphology.

(D) Tuj1 and Map2 immunostaining 14 days post-induction of Ascl1, Zfp238, Dlx3, or Sox8 with Myt1l. Zfp238, Dlx3, and Sox8 can all form iN cells when co-expressed with Myt1l in the absence of Ascl1.

(E) Counts of Tuj1$^+$ or Map2$^+$ neuronal cells 14 days post-induction of Ascl1, Zfp238, Dlx3, or Sox8 with Myt1l (mean ± SEM, n = 3).

ATAC-Seq

ATAC-seq was performed as described (Buenrostro et al., 2013). ATAC-seq libraries were sequenced on Illumina HiSeq2000 or NextSeq sequencers. Primary data processing was done as described (Buenrostro et al., 2013). Detailed information on computational analysis can be found in Supplemental Experimental Procedures. The accession number for the previously published ChIP- and RNA-seq data are GEO: GSE43916.

TF Network Analysis

TF-TF networks were constructed for the Ascl1 cells at days 5, 13, and 22. First, we used PIQ v1.2 (Sherwood et al., 2014) to assay mouse and human motif position weight matrices (PWMs) from JASPAR (Mathelier et al., 2014; Sandelin et al., 2004) and defined bound motifs to be sites with purity score greater than 0.9. Next, we filtered for TF-gene interactions that resulted in a change in the target gene's openness. Finally, we further prune the remaining edges by only keeping the edges whose target genes exhibited differential expression (adjusted p value $\leq$ 0.1) based on previously published RNA-seq raw counts (Wapinski et al., 2013). This gave a directed network between TFs, for TFs with open loci which bound near genes whose openness and RNA-seq expression changed relative to time 0. Then network was then organized and visualized using Cytoscape (Shannon et al., 2003). More detailed information can be found in Supplemental Experimental Procedures.

Single-Cell and Bulk RNA-Seq Analysis for Network TFs

Making the assumption that TFs from the same family bind similar motifs, the gene expression profile of the TFs from the networks of all three time points and all their family members was generated with gene expression (reads per kilobase of transcript per million mapped reads [RPKM]) as columns and either cells as rows for the single-cell RNA-seq dataset (Treutlein et al., 2016) or individual samples as rows for the bulk RNA-seq dataset (Wapinski et al., 2013). For the single-cell RNA-seq dataset, the list is filtered for TFs that are expressed with at least an RPKM of 1 in 10 or more cells. Hierarchical clustering is then performed on the remaining data frame using Pearson correlation as a distance metric. The rows are sorted according to fractional neuronal identity. The heatmap is then generated using ggplot2 in R, and visualized using TreeView. In Figure 4C, a similar heatmap is plotted with only the TFs we used to validate the networks. For the bulk RNA-seq dataset, the list is filtered for TFs that are expressed with at least one RPKM in two or more samples. The remaining data frame is used to plot a heatmap using heatmap.2 under the gplots package in R. Pearson correlation was used as the distance metric in clustering both columns and rows. The column values are scaled according to the heatmap.2 default.

Validation of Ascl1 Target Genes

Zfp238 KO and WT MEFs from littermates were obtained as previously described (Xiang et al., 2012) and co-infected with rtTA and dox-inducible Ascl1. Similarly, TauEGFP MEFs were co-infected with rtTA, dox-inducible Ascl1, and selected target genes following the reprogramming protocol (Vierbuchen et al., 2010). Day 14 after dox induction, cells were fixed, and immunostaining was performed with Map2 and Tuj1. For each replicate, the number of Tuj^+, $Map2^+$, and $DAPI^+$ cells were counted. Graphs show the average of n biological replicates, and the error bars indicate the SE. Student's t test was performed to determine statistical significance. More details of immunofluorescence and cell counting can be found in Supplemental Experimental Procedures.

Data and Software Availability

The accession number for the ATAC-seq data reported in this study is GEO: GSE101397.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and four figures and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2017.09.011.

AUTHOR CONTRIBUTIONS

Conceptualization, Methodology, Writing – Original Draft and Writing – Review & Editing, O.L.W., Q.Y.L., M.W., and H.Y.C.; Investigation, O.L.W., Q.Y.L., R.L., C.E.A, B.T., C.X., V.B., F.P.S., V.S., and S.S.; Software and Formal Analysis, O.L.W., Q.Y.L., A.C.C., and M.R.C.; Funding Acquisition, M.W. and H.Y.C.; Resources, S.R.Q., N.D., M.W., and H.Y.C.; Supervision, M.W. and H.Y.C.

ACKNOWLEDGMENTS

The authors would like to acknowledge P. Lovelace for support with FACS; Paul Giresi, Jason Buenrostro, and Yifei Men for discussions on bioinformatics pipelines; and other Chang and Wernig laboratory members for discussions and support. This work was supported by NIH Grant P50-HG007735 (to H.Y.C.), California Institute for Regenerative Medicine Grants RB5-07466 and RB4-05763 (to M.W. and H.Y.C.), a Howard Hughes Medical Institute Faculty Scholar Award (M.W.), and the New York Stem Cell Foundation-Robertson Award (M.W.). O.L.W. was supported by National Science Foundation. Q.Y.L. was supported by a National Science Scholarship from the Agency for Science, Technology, and Research. M.W. is a Tashia and John Morgridge Faculty Scholar in the Child Health Research Institute at Stanford.

Received: February 1, 2017
Revised: July 18, 2017
Accepted: September 3, 2017
Published: September 26, 2017

REFERENCES

Baubet, V., Xiang, C., Molczan, A., Roccograndi, L., Melamed, S., and Dahmane, N. (2012). Rp58 is essential for the growth and patterning of the cerebellum and for glutamatergic and GABAergic neuron development. Development *139*, 1903–1909.

Beanan, M.J., and Sargent, T.D. (2000). Regulation and function of Dlx3 in vertebrate development. Dev. Dyn. *218*, 545–553.

Berkes, C.A., Bergstrom, D.A., Penn, B.H., Seaver, K.J., Knoepfler, P.S., and Tapscott, S.J. (2004). Pbx marks genes for activation by MyoD indicating a role for a homeodomain protein in establishing myogenic potential. Mol. Cell *14*, 465–477.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Buenrostro, J.D., Wu, B., Litzenburger, U.M., Ruff, D., Gonzales, M.L., Snyder, M.P., Chang, H.Y., and Greenleaf, W.J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. Nature *523*, 486–490.

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. (2012). Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell *150*, 1209–1222.

Cacchiarelli, D., Trapnell, C., Ziller, M.J., Soumillon, M., Cesana, M., Karnik, R., Donaghey, J., Smith, Z.D., Ratanasirintrawoot, S., Zhang, X., et al. (2015). Integrative analyses of human reprogramming reveal dynamic nature of induced pluripotency. Cell *162*, 412–424.

Carroll, J.S., Meyer, C.A., Song, J., Li, W., Geistlinger, T.R., Eeckhoute, J., Brodsky, A.S., Keeton, E.K., Fertuck, K.C., Hall, G.F., et al. (2006). Genome-wide analysis of estrogen receptor binding sites. Nat. Genet. *38*, 1289–1297.

Castro, D.S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., Drechsel, D., Lebel-Potter, M., Garcia, L.G., Hunt, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. Genes Dev. *25*, 930–945.

Cau, E., Gradwohl, G., Casarosa, S., Kageyama, R., and Guillemot, F. (2000). Hes genes regulate sequential stages of neurogenesis in the olfactory epithelium. Development *127*, 2323–2332.

Chanda, S., Ang, C.E., Davila, J., Pak, C., Mall, M., Lee, Q.Y., Ahlenius, H., Jung, S.W., Südhof, T.C., and Wernig, M. (2014). Generation of induced neuronal cells by the single reprogramming factor ASCL1. Stem Cell Reports *3*, 282–296.

Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., Ernst, J., and Plath, K. (2017). Cooperative binding of transcription factors orchestrates reprogramming. Cell *168*, 442–459.e20.

Corces, M.R., Buenrostro, J.D., Wu, B., Greenside, P.G., Chan, S.M., Koenig, J.L., Snyder, M.P., Pritchard, J.K., Kundaje, A., Greenleaf, W.J., et al. (2016). Lineage-specific and single-cell chromatin accessibility charts human hematopoiesis and leukemia evolution. Nat. Genet. *48*, 1193–1203.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell *51*, 987–1000.

Di Stefano, B., Sardina, J.L., van Oevelen, C., Collombet, S., Kallin, E.M., Vicent, G.P., Lu, J., Thieffry, D., Beato, M., and Graf, T. (2014). C/EBPα poises B cells for rapid reprogramming into induced pluripotent stem cells. Nature *506*, 235–239.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Fode, C., Gradwohl, G., Morin, X., Dierich, A., LeMeur, M., Goridis, C., and Guillemot, F. (1998). The bHLH protein NEUROGENIN 2 is a determination factor for epibranchial placode-derived sensory neurons. Neuron *20*, 483–494.

Fong, A.P., Yao, Z., Zhong, J.W., Johnson, N.M., Farr, G.H., 3rd, Maves, L., and Tapscott, S.J. (2015). Conversion of MyoD to a neurogenic factor: binding site specificity determines lineage. Cell Rep. *10*, 1937–1946.

Gerber, A.N., Klesert, T.R., Bergstrom, D.A., and Tapscott, S.J. (1997). Two domains of MyoD mediate transcriptional activation of genes in repressive chromatin: a mechanism for lineage determination in myogenesis. Genes Dev. *11*, 436–450.

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A., and Jaenisch, R. (2009). Direct cell reprogramming is a stochastic process amenable to acceleration. Nature *462*, 595–601.

Henikoff, J.G., Belsky, J.A., Krassovsky, K., MacAlpine, D.M., and Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. Proc. Natl. Acad. Sci. USA *108*, 18318–18323.

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.

Ieda, M., Fu, J.-D., Delgado-Olguin, P., Vedantham, V., Hayashi, Y., Bruneau, B.G., and Srivastava, D. (2010). Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell *142*, 375–386.

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

Lara-Astiaso, D., Weiner, A., Lorenzo-Vivas, E., Zaretsky, I., Jaitin, D.A., David, E., Keren-Shaul, H., Mildner, A., Winter, D., Jung, S., et al. (2014). Chromatin state dynamics during blood formation. Science *345*, 943–949.

Lopez-Pajares, V., Qu, K., Zhang, J., Webster, D.E., Barajas, B.C., Siprashvili, Z., Zarnegar, B.J., Boxer, L.D., Rios, E.J., Tao, S., et al. (2015). A LncRNA-MAF:MAFB transcription factor network regulates epidermal differentiation. Dev. Cell *32*, 693–706.

Mall, M., Kareta, M.S., Chanda, S., Ahlenius, H., Perotti, N., Zhou, B., Grieder, S.D., Ge, X., Drake, S., Euong Ang, C., et al. (2017). Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates. Nature *544*, 245–249.

Mathelier, A., Zhao, X., Zhang, A.W., Parcy, F., Worsley-Hunt, R., Arenillas, D.J., Buchman, S., Chen, C.Y., Chou, A., Ienasescu, H., et al. (2014). JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. Nucleic Acids Res. *42*, 142–147.

McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of *cis*-regulatory regions. Nat. Biotechnol. *28*, 495–501.

Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. Nat. Methods *12*, 453–457.

Okita, K., Ichisaka, T., and Yamanaka, S. (2007). Generation of germline-competent induced pluripotent stem cells. Nature *448*, 313–317.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Südhof, T.C., and Wernig, M. (2011). Induction of human neuronal cells by defined transcription factors. Nature *476*, 220–223.

Park, I.-H., Zhao, R., West, J.A., Yabuuchi, A., Huo, H., Ince, T.A., Lerou, P.H., Lensch, M.W., and Daley, G.Q. (2008). Reprogramming of human somatic cells to pluripotency with defined factors. Nature *451*, 141–146.

Pique-Regi, R., Degner, J.F., Pai, A.A., Gaffney, D.J., Gilad, Y., and Pritchard, J.K. (2011). Accurate inference of transcription factor binding from DNA sequence and chromatin accessibility data. Genome Res. *21*, 447–455.

Qu, K., Zaba, L.C., Giresi, P.G., Li, R., Longmire, M., Kim, Y.H., Greenleaf, W.J., and Chang, H.Y. (2015). Individuality and variation of personal regulomes in primary human T cells. Cell Syst. *1*, 51–61.

Rais, Y., Zviran, A., Geula, S., Gafni, O., Chomsky, E., Viukov, S., Mansour, A.A., Caspi, I., Krupalnik, V., Zerbib, M., et al. (2013). Deterministic direct reprogramming of somatic cells to pluripotency. Nature *502*, 65–70.

Rhee, H.S., Closser, M., Guo, Y., Bashkirova, E.V., Tan, G.C., Gifford, D.K., and Wichterle, H. (2016). Expression of terminal effector genes in mammalian neurons is maintained by a dynamic relay of transient enhancers. Neuron *92*, 1252–1265.

Sahoo, D., Dill, D.L., Tibshirani, R., and Plevritis, S.K. (2007). Extracting binary signals from microarray time-course data. Nucleic Acids Res. *35*, 3705–3712.

Sandelin, A., Alkema, W., Engström, P., Wasserman, W.W., and Lenhard, B. (2004). JASPAR: an open-access database for eukaryotic transcription factor binding profiles. Nucleic Acids Res. *32*, D91–D94.

Schep, A.N., Buenrostro, J.D., Denny, S.K., Schwartz, K., Sherlock, G., and Greenleaf, W.J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. Genome Res. *25*, 1757–1770.

Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature *475*, 390–393.

Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. *13*, 2498–2504.

Sherwood, R.I., Hashimoto, T., O'Donnell, C.W., Lewis, S., Barkal, A.A., van Hoff, J.P., Karun, V., Jaakkola, T., and Gifford, D.K. (2014). Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. *32*, 171–178.

Smith, Z.D., Sindhu, C., and Meissner, A. (2016). Molecular features of cellular reprogramming and development. Nat. Rev. Mol. Cell Biol. *17*, 139–154.

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.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell *161*, 555–568.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861–872.

Treutlein, B., Lee, Q.Y., Camp, J.G., Mall, M., Koh, W., Shariati, S.A.M., Sim, S., Neff, N.F., Skotheim, J.M., Wernig, M., and Quake, S.R. (2016). Dissecting

direct reprogramming from fibroblast to neuron using single-cell RNA-seq. Nature *534*, 391–395.

Velasco, S., Ibrahim, M.M., Kakumanu, A., Garipler, G., Aydin, B., Al-Sayegh, M.A., Hirsekorn, A., Abdul-Rahman, F., Satija, R., Ohler, U., et al. (2017). A multi-step transcriptional and chromatin state cascade underlies motor neuron programming from embryonic stem cells. Cell Stem Cell *20*, 205–217.e8.

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.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell *16*, 386–399.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635.

Weider, M., and Wegner, M. (2017). SoxE factors: transcriptional regulators of neural differentiation and nervous system development. Semin. Cell Dev. Biol. *63*, 35–42.

Wernig, M., Meissner, A., Foreman, R., Brambrink, T., Ku, M., Hochedlinger, K., Bernstein, B.E., and Jaenisch, R. (2007). In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. Nature *448*, 318–324.

Wu, J., Huang, B., Chen, H., Yin, Q., Liu, Y., Xiang, Y., Zhang, B., Liu, B., Wang, Q., Xia, W., et al. (2016). The landscape of accessible chromatin in mammalian preimplantation embryos. Nature *534*, 652–657.

Xiang, C., Baubet, V., Pal, S., Holderbaum, L., Tatard, V., Jiang, P., Davuluri, R.V., and Dahmane, N. (2012). RP58/ZNF238 directly modulates proneurogenic gene levels and is required for neuronal differentiation and brain expansion. Cell Death Differ. *19*, 692–702.

Yao, Z., Fong, A.P., Cao, Y., Ruzzo, W.L., Gentleman, R.C., and Tapscott, S.J. (2013). Comparison of endogenous and overexpressed MyoD shows enhanced binding of physiologically bound sites. Skelet. Muscle *3*, 8.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.