HYPOTHESIS

Direct lineage reprogramming via pioneer factors; a detour through developmental gene regulatory networks

Samantha A. Morris*

ABSTRACT

Although many approaches have been employed to generate defined fate *in vitro*, the resultant cells often appear developmentally immature or incompletely specified, limiting their utility. Growing evidence suggests that current methods of direct lineage conversion may rely on the transition through a developmental intermediate. Here, I hypothesize that complete conversion between cell fates is more probable and feasible via reversion to a developmentally immature state. I posit that this is due to the role of pioneer transcription factors in engaging silent, unmarked chromatin and activating hierarchical gene regulatory networks responsible for embryonic patterning. Understanding these developmental contexts will be essential for the precise engineering of cell identity.

KEY WORDS: Cell fate engineering, Direct lineage reprogramming, Gene regulatory networks, Pioneer transcription factors

Introduction

During embryonic development, cells progress along a route of increasing specialization and decreasing potential. It has long been thought that cell identity is irreversibly determined and rarely undergoes dramatic transformation, but advances over the past half-century, such as nuclear transfer (Gurdon et al., 1958), cell fusion (Blau et al., 1983) and factor-mediated reprogramming (Davis et al., 1987), have revealed the plasticity of cell identity. A major milestone in this field was marked by the *in vitro* reprogramming of cells to pluripotency via forced expression of the so called 'Yamanaka' transcription factors Oct4 (Pou5f1), Sox2, Klf4 and c-Myc (Takahashi and Yamanaka, 2006). The resulting reprogrammed cells can be directed to differentiate *in vitro* toward desired target populations by recapitulating the relevant embryonic programs (reviewed by Cohen and Melton, 2011), offering potential for regenerative therapy and permitting disease modeling, toxicology testing and drug discovery. Directing differentiation from a pluripotent state typically involves lengthy, multistep protocols and is therefore inefficient, often producing heterogeneous populations of developmentally immature cells (D'Amour et al., 2006; Oldershaw et al., 2010; Oshima et al., 2010; Rashid et al., 2010; Touboul et al., 2010).

An alternative strategy aims to generate target cell types by directly converting cell fate between fully differentiated somatic states. This approach, known as 'direct lineage conversion' or 'lineage reprogramming' (see Glossary, Box 1), was initially demonstrated *in vitro* by the conversion of cultured fibroblasts to

Departments of Developmental Biology and Genetics, Center of Regenerative Medicine, Washington University School of Medicine in St. Louis, 660 S. Euclid Avenue, Campus Box 8103, St. Louis, MO 63110, USA.

*Author for correspondence (s.morris@wustl.edu)

S.A.M., 0000-0001-8561-4340

contracting myocytes via forced expression of the transcription factor MyoD (Myod1) (Davis et al., 1987). Current approaches based on this strategy are designed with the intention of bypassing pluripotent or progenitor states; a 'shortcut' aimed at boosting the speed and efficiency of cell fate conversion. Numerous examples of direct lineage conversion have been reported using mouse and human cells, driven predominantly by transcription factor expression (Fig. 1) (reviewed by Morris and Daley, 2013; Vierbuchen and Wernig, 2011). Several studies have employed elegant genetic lineage tracing to demonstrate that conversion bypasses canonical progenitor states. For example, fibroblast to cardiomyocyte conversion does not require transition through an intermediate state expressing *Islet1* or *Mesp1*, two classic cardiac progenitor markers (Cahan et al., 2014; Ieda et al., 2010). Likewise, absence of progenitor marker expression and cell division during fibroblast to neural conversion has been taken as a measure of conversion 'directness' (Heinrich et al., 2010; Vierbuchen et al., 2010). In both these cases, however, it is apparent that the resultant cells are immature, suggesting that a developmental intermediary is accessed, albeit not the bona fide progenitor as we know it.

The immaturity of directly reprogrammed cells is something of a recurring theme in lineage reprogramming, and in some cases remnants of the previous cell fate persist (Cahan et al., 2014; Morris et al., 2014). For example, hepatic gene expression is not fully extinguished in neural cells derived from hepatocytes (Marro et al., 2011). Similarly, residual fibroblast gene expression is observed in macrophages generated from fibroblasts, where the resulting cells are unstable and de-differentiate upon removal of exogenous factors (Feng et al., 2008). With respect to target cell fate, direct conversion of fibroblasts to cardiomyocytes yields cells that do not fully recapitulate the profile of neonatal cardiomyocytes (Ieda et al., 2010). Similarly, conversion of fibroblasts to neurons can be achieved by expression of a single gene (Chanda et al., 2014), but requires additional factors to drive developmental maturation (Vierbuchen et al., 2010).

In some cases, it appears that cells undergoing lineage reprogramming do in fact transit through a bona fide progenitor state. Transient hematopoietic stem and progenitor cell signatures are observed during the conversion of B-cells to macrophages (Bussmann et al., 2009; Di Tullio et al., 2011; Morris et al., 2014), while the conversion of fibroblasts to induced hepatocytes yields cells bearing embryonic signatures (Huang et al., 2011; Sekiya and Suzuki, 2011). The transcription factor-driven hepatic conversion first reported by Sekiya and Suzuki (2011) in fact generates cells that have a broader endoderm potential than was initially anticipated (Morris et al., 2014). In addition to these studies, a growing number of reports demonstrate that progenitor intermediates can also be purposefully generated. Conversion from fibroblasts to hematopoietic progenitors (Pereira et al., 2013; Szabo et al., 2010), to neuronal stem/precursor cells (Lujan et al., 2012; Thier et al., 2012), to bipotent hepatic progenitors (Yu et al., 2013), to

angioblast-like progenitor cells (Kurian et al., 2013) and to osteoblasts (Yamamoto et al., 2015), and conversion from proximal tubule cells to nephron progenitors (Hendry et al., 2013) have all been achieved via transcription factor-mediated lineage reprogramming. Several recent methodologies have relied on a transient burst of Yamanaka factor expression, with the intention to direct cell fate to an alternative lineage before the pluripotent state is reached (Kurian et al., 2013; Thier et al., 2012; Zhu et al., 2014). Recently, however, two elegant studies have demonstrated that these cells do in fact passage through a pluripotent state (Bar-Nur et al., 2015; Maza et al., 2015). Altogether, studies such as these can reveal much regarding the specification of progenitor cell fate; however, the resulting cells themselves are generally unsuitable for immediate therapeutic application. Thus, immaturity in lineage reprogramming, whether accidental or by design, limits the application of reprogrammed cells, especially for disease modeling and drug toxicology testing *in vitro*, where it is often crucial to produce fully differentiated and functional cells.

A developmental logic for direct lineage conversion

Taken in its most stringent interpretation, direct lineage conversion aims to switch fate between two fully differentiated somatic states, circumventing intermediate progenitor states. In actuality, however, what we observe during lineage reprogramming is a spectrum of states, ranging from pluripotent stem cells to multipotent progenitors to lineage-committed yet phenotypically and functionally immature cells. In this Hypothesis article, I argue that there is minimal evidence to support direct conversion between terminally differentiated cell states. Far from a semantic argument, exploration of this concept is crucial in our efforts to precisely

engineer cell fate for disease modeling and regenerative therapy. I hypothesize that the majority of commonly employed lineage conversion factors engage developmental gene regulatory networks (GRNs; see Glossary, Box 1) to reinitiate molecular programs that correspond to different stages of cellular commitment and maturity, depending on the factors selected to drive conversion and the properties of the cells into which they are introduced. Based on current evidence, I posit that fate conversion via immature intermediates is more probable and feasible than direct conversion between two terminally differentiated states. This likelihood hinges on the capacity of pioneer factors (see Glossary, Box 1) to engage silent, unmarked chromatin, activating hierarchical GRNs at different levels of the developmental hierarchy. I argue that most current direct lineage conversions produce developmental intermediates, and that engineering strategies must accommodate maturation via the addition of further factors or exposure to the *in vivo* niche, or both.

Direct lineage conversions are predominantly driven by pioneer transcription factors

An analysis of published lineage reprogramming methodologies reveals that fate conversions are chiefly driven by transcription factors associated with cell fate programming during embryonic development (Fig. 1). To successfully convert cell fate, developmentally silenced genes must be activated to specify target cell identity in host cells. Thus, not surprisingly, the most potent transcription factors endowed with reprogramming activity are pioneer factors. Pioneer factors are transcription factors that initiate transcriptional programs leading to cell fate change via their engagement with silent, unmarked chromatin (Iwafuchi-Doi and Zaret, 2014, 2016; Zaret and Carroll, 2011). These low-signal chromatin regions are characterized by their relative lack of histone modifications (Ho et al., 2014; Kharchenko et al., 2011), where transcription is instead likely to be repressed by the presence of linker histones (Iwafuchi-Doi and Zaret, 2016). Pioneer factor binding results in the local opening of this silent chromatin (Soufi et al., 2012; van Oevelen et al., 2015), although in the absence of any other transcription factors this is insufficient to induce changes in gene expression. Rather, pioneer factors recruit activators or repressors that by themselves are unable to engage with silent chromatin (Carroll et al., 2005; Gualdi et al., 1996; Sekiya and Zaret, 2007). In this way, pioneer factors act as master regulators of cell fate during normal development, directing competence for many different cell fates via their interaction with chromatin and subsequent cooperativity with lineage-specific transcription factors (Fig. 2).

Pioneer factors with the capability of inducing lineage conversions include three of the Yamanaka reprogramming factors, namely Oct4, Sox2 and Klf4 (Soufi et al., 2012, 2015), which also feature in several inter-germ layer lineage conversion strategies (Fig. 1). Gata4 is another well-characterized pioneer factor (Bossard and Zaret, 1998) responsible for driving both cardiac and hepatic lineage conversions. In addition, C/ebpα and PU.1 (Spi1) have been reported to act as pioneer factors (Heinz et al., 2010; van Oevelen et al., 2015), inducing conversions within the blood lineage. In the neural reprogramming field, Ascl1 acts as a pioneer factor and is employed in the majority of conversions to neural fate (Wapinski et al., 2013). Perhaps one of the best illustrations of how pioneer factors control cell fate decision-making is provided by the FoxA family of transcription factors (Foxa1, Foxa2 and Foxa3). In the early stages of mouse endoderm development, FoxA binds to an enhancer of the liver-specific gene albumin (*Alb*), which remains
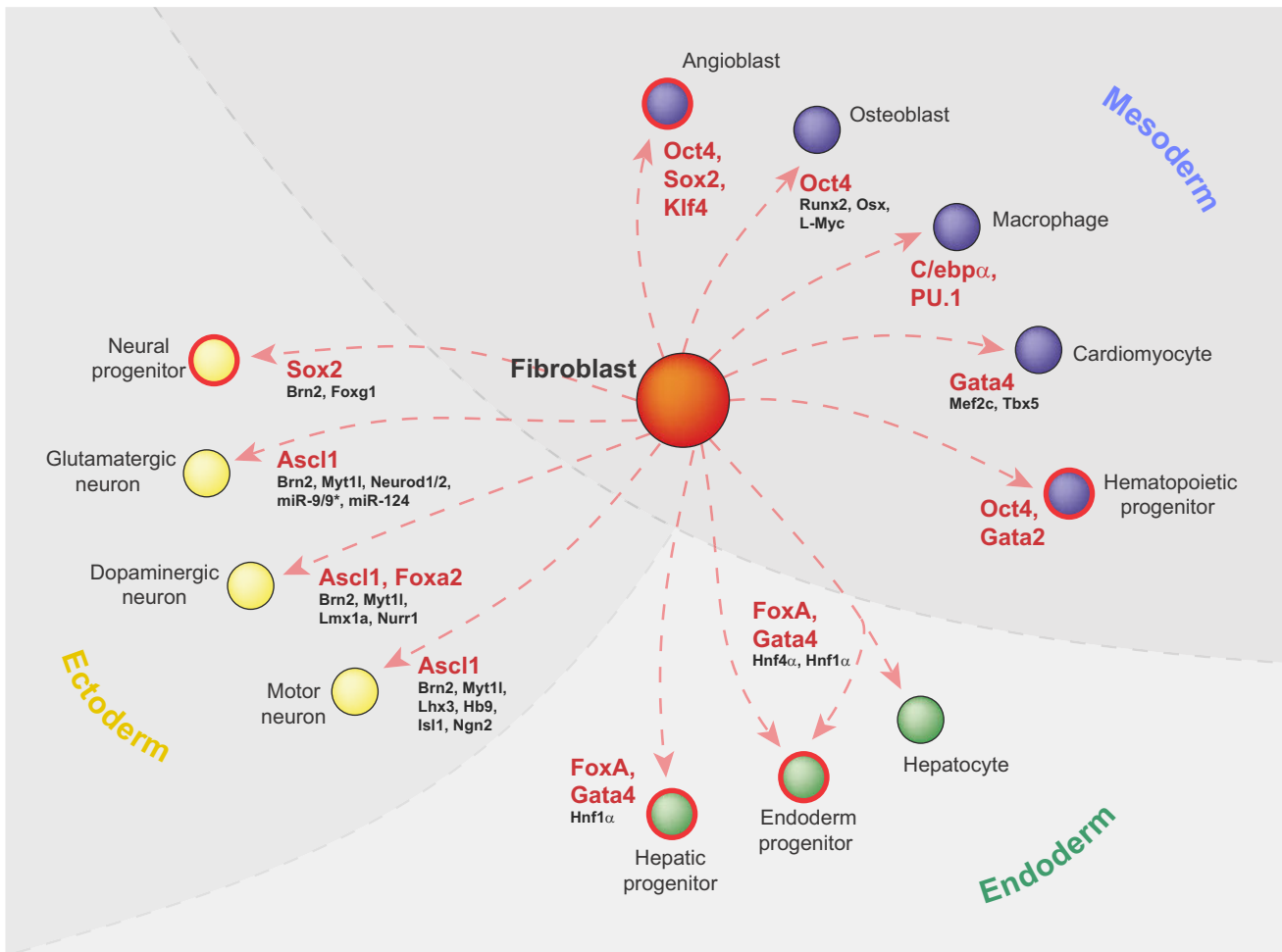
Fig. 1. Pioneer factors used in direct lineage conversions from fibroblasts. Pioneer transcription factors have been used in many instances of direct conversion of cell fate between differentiated states, both in mouse and human. These include the conversion of fibroblasts to cardiomyocytes (Ieda et al., 2010), hepatocytes (Huang et al., 2011; Sekiya and Suzuki, 2011; Zhu et al., 2014), glutamatergic neurons (Ambasudhan et al., 2011; Pang et al., 2011; Vierbuchen et al., 2010; Yoo et al., 2011), dopaminergic neurons (Caiazzo et al., 2011; Pfisterer et al., 2011), motor neurons (Son et al., 2011), hematopoietic progenitors (Pereira et al., 2013; Szabo et al., 2010), neural progenitors (Lujan et al., 2012; Thier et al., 2012), bipotent hepatic progenitors (Yu et al., 2013), angioblast-like progenitor cells (Kurian et al., 2013) and osteoblasts (Yamamoto et al., 2015). Other cell type conversions (not shown) include hepatocyte to neuron conversion (Marro et al., 2011), and cortical astrocytes to glutamatergic (Heinrich et al., 2010) and GABAergic (Berninger et al., 2007; Heinrich et al., 2010) neurons. This list is not exhaustive and continues to expand. Conversion from fibroblasts to progenitor states are circled in red. Pioneer factors involved in each conversion are shown in red text, cooperative factors are shown in black text.

silent. As liver differentiation progresses, FoxA subsequently recruits activators, such as Hnf4α, to activate *Alb* expression, which represents the initiation of the liver gene expression program (Gualdi et al., 1996). FoxA is also expressed more broadly throughout the endoderm, where its binding equips cells with hepatic potential that is never realized due to the inhibitory effect of the overlying mesoderm (Bossard et al., 2000; McLin et al., 2007). Clearly then, FoxA has the capacity to impart lineage competence in a context-specific manner (Wang et al., 2015). This central role of FoxA in development is reflected by its redeployment across disparate developmental programs (reviewed by Friedman and Kaestner, 2006) and its dominance in lineage conversion (Fig. 1). Although pioneer factors clearly endow cells with lineage competence, it remains the case that even when pioneer factors are co-expressed with their cognate activators the converting cells appear to transition via a developmental intermediate. Understanding the different trajectories that are initiated by pioneer factor expression will help us to better understand the molecular mechanisms that underpin direct lineage conversion.

Understanding pioneer factor-driven direct lineage conversion in the context of developmental GRNs

It is clear that the lineage conversion strategies that utilize pioneer factor expression leverage fundamental developmental mechanisms. I propose that developmental GRN hierarchies provide a framework for understanding the relative probabilities for different cell fate conversion routes. GRNs are crucial determinants of cell identity that control transcriptional activity to drive lineage progression in an ordered manner. These networks describe a complex system of interactions among transcription factors binding to multiple cis-regulatory DNA sequences, resulting in the spatial and temporal regulation of gene expression (Davidson and Erwin, 2006). This meticulous choreography is inherently hierarchical, which helps to ensure correct developmental patterning (Peter and Davidson, 2011, 2016). During development, the embryonic body plan is initially laid down by upper-level GRNs, which are underpinned by evolutionarily conserved subcircuits referred to as 'kernels' (see Glossary, Box 1). This creates a primary spatial organization within
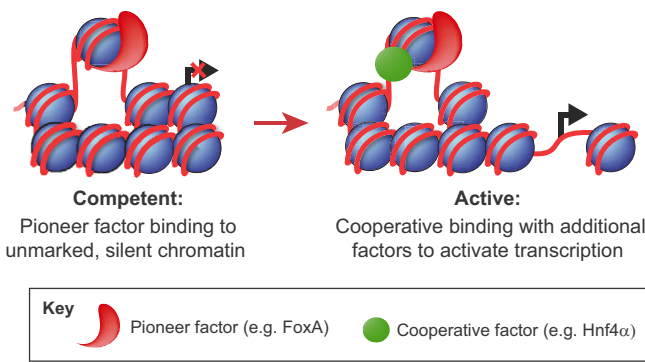
Competent:
Pioneer factor binding to unmarked, silent chromatin

Active:
Cooperative binding with additional factors to activate transcription

Key
Pioneer factor (e.g. FoxA)   Cooperative factor (e.g. Hnf4α)

Fig. 2. Mechanism of pioneer factor activity. Pioneer factor binding results in the local opening of silent, unmarked chromatin, which confers lineage competence, although in the absence of any other transcription factors this in itself is insufficient to induce changes in gene expression. Subsequent pioneer factor-mediated recruitment of activators or repressors, which by themselves are unable to engage with silent chromatin, initiates transcriptional programs leading to cell fate change.

the embryo, enabling the initiation of regional fate decisions via the implementation of successive GRNs (Fig. 3). Intermediate GRNs then specify cell identity in the appropriate regions in order to
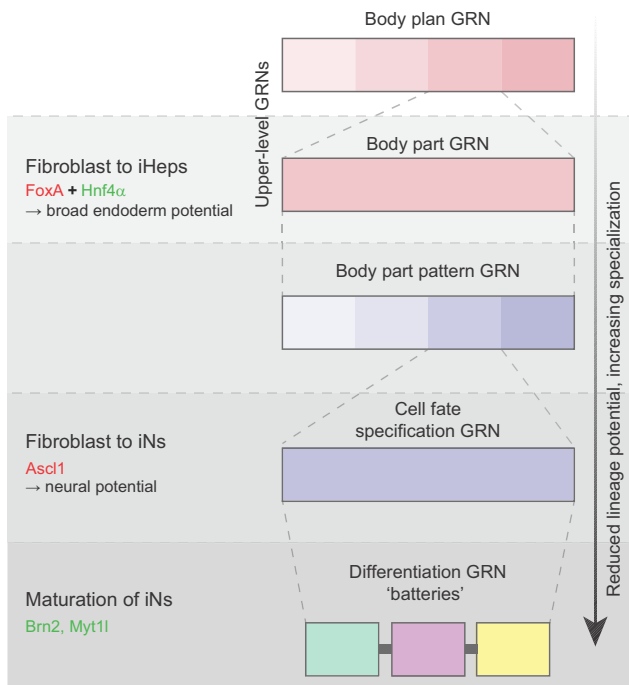

Fig. 3. Gene regulatory network (GRN) hierarchy and pioneer factor activity. The embryonic body plan is initially laid down by upper-level GRNs (see Peter and Davidson, 2011). This creates a primary spatial organization within the embryo, enabling the initiation of regional fate decisions via the implementation of successive upper-level GRNs. Intermediate GRNs specify cell identity in the appropriate regions in order to define tissue pattern, laying the foundations for cell fate specification GRNs. Finally, the hierarchy concludes with GRN 'batteries' that control the final stages of cell differentiation. In direct lineage conversion, FoxA factors engage with upper-level GRNs to generate induced hepatocytes (iHeps). Ascl1 engages with downstream cell fate specification GRNs to produce induced neurons (iNs). Installation of GRN batteries is achieved via expression of maturation factors (Brn2 and Myt1l) or a period of engraftment within the *in vivo* niche. Pioneer factors are shown in red text, cooperative factors in green text.

define tissue pattern, laying the foundations for cell fate specification GRNs (Davidson and Erwin, 2006). Within this regulatory hierarchy, pioneer factors play a core role. For example, the FoxA genes and *Gata4* are central to highly conserved endoderm and vertebrate heart kernels, respectively, and are crucial for the acquisition of developmental competence (Davidson and Erwin, 2006).

The developmental GRN hierarchy concludes with peripheral networks or 'batteries' (see Glossary, Box 1) that control cell differentiation (Fig. 3). The transcription factors that regulate these batteries can be described as 'terminal selector genes' (see Glossary, Box 1), which are responsible for directly regulating terminal differentiation genes, controlling specific cell identity as a result (Hobert, 2008). Within this hierarchical structure, complexity is gradually built up and relies upon the precise installation of networks in the preceding stage of development. Consistent developmental outcomes are ensured by this regulatory process, which is classically termed 'canalization' by Waddington (Waddington, 1942) and taken to mean that the number of available regulatory states decreases as cell specialization increases, concordant with a decrease in developmental potential. Canalization has several implications when considering mechanisms and strategies for direct lineage conversion. Mutations leading to reorganization at cis-regulatory nodes of upper-level GRNs might have many downstream changes that are detrimental to development of the organism. This idea is supported by the conservation of these GRNs over large evolutionary distances (Davidson and Erwin, 2006; Peter and Davidson, 2011). By contrast, changes further down the hierarchy, at the level of terminal selector genes, result in relatively minor perturbations to cellular phenotype (Hobert, 2008; Uchida et al., 2003). Within this hierarchical GRN context, we can consider direct lineage conversion driven by pioneer factors as a perturbation to upper-level circuits that would be catastrophic for normal development.

Bypassing developmental states during direct lineage conversion: possible, but improbable?
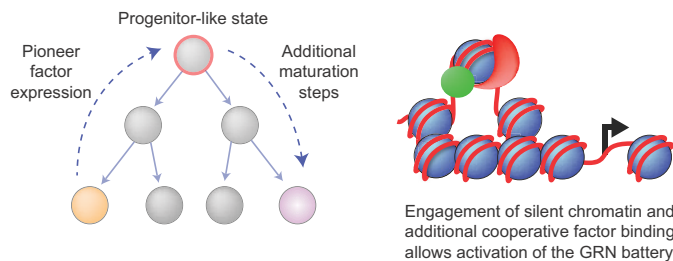
I hypothesize that direct lineage conversion relies on the engagement of developmental GRN hierarchies. The majority of cell fate conversions leverage pioneer factors to induce dramatic changes in cell identity toward varying degrees of immaturity. I propose that pioneer transcription factors engage with upper-level GRNs, from which point downstream GRN installation drives cells toward maturity. Within this structure, it is feasible that developmental GRNs might be accessed at various levels corresponding to different stages of development, accounting for the generation of cells ranging in lineage potential and self-renewal properties. For example, FoxA acts at several levels, from specifying competence for endodermal fate to driving the expression of the mature liver gene *Alb* (Gualdi et al., 1996; Wang et al., 2015). I hypothesize that pioneer factors activate upper-level GRNs via their normal developmental role of engaging silent chromatin, and that this is an unintended consequence of direct conversion design. Furthermore, I argue that pioneer factor-driven lineage conversion is more feasible as it allows for the initial endowment of lineage competence before expression of cognate activators can drive the conversion. Depending on where in the hierarchy pioneer factors engage GRNs, this process sometimes activates a progenitor-like program, whereas in other cases a more differentiated program is engaged, although full differentiation is not achieved. By contrast, expression of only downstream transcription factors that control terminal gene differentiation batteries would in theory provide a

direct and specific conversion, but these factors are unable to engage with silent chromatin to activate the desired batteries, making successful conversion improbable (Fig. 4). Does this challenge the idea that fate transitions can be truly direct, without activating some form of an embryonic program? To further explore this hypothesis it is helpful to survey evidence from established conversion strategies, where platforms have recently been developed to assess GRN establishment in reprogrammed cells (Cahan et al., 2014). Within this framework I will also consider the impact of host cell gene expression and chromatin landscape on pioneer factor behavior.

Pioneer factors induce broad fate: a case study of FoxA-driven induction of iHeps

Foxa1- and Hnf4α-driven conversion of fibroblasts to induced hepatocytes (iHeps) represents a clear example of a pioneer transcription factor-mediated conversion that unintentionally endows cells with an immature phenotype and a broad developmental potential. iHeps exhibit long-term self-renewal, accompanied by an immature gene expression profile (Morris et al., 2014; Sekiya and Suzuki, 2011). These properties have led to speculation that iHeps represent a developmental intermediate rather than fully differentiated hepatocytes (Willenbring, 2011). Analysis of GRN establishment in iHeps revealed coincident hepatic and intestinal signatures in the converted cells (Morris et al., 2014), and transplantation of iHeps demonstrated their potential to functionally engraft both into liver (Sekiya and Suzuki, 2011) and large intestine (Morris et al., 2014). These findings confirm that conversion to iHeps generates a progenitor-like state rather than committed hepatic fate as previously thought.

The factors employed in the production of iHeps feature prominently in the genesis of the endoderm lineage. Hnf4α is required for liver (Parviz et al., 2003) and intestinal (Garrison et al., 2006) development, as well as for adult liver function (Hayhurst et al., 2001). Similarly, Foxa1 is involved in both liver and intestinal development and their function (Bernardo and Keri, 2012; van der Sluis et al., 2008). These wide roles across the endoderm lineage are reflected by the broad potential of iHeps, suggesting that an upper-level endoderm GRN that is normally silent in fibroblasts is engaged by the pioneer factor Foxa1. This would establish a broad endoderm competence, working in concert with Hnf4α to induce specific endoderm lineages. Direct lineage conversion frequently relies on the overexpression of factors at levels much higher than normally seen during development, and in the absence of established developmental context. For example, to generate iHeps from fibroblasts, FoxA and Hnf4α are expressed at atypically high levels (Sekiya and Suzuki, 2011), which induces $Cdx2$ expression as an unintended consequence (Morris et al., 2014). Cdx2 specifies embryonic intestinal epithelium and is essential for intestinal gene expression (Gao and Kaestner, 2010). Therefore, in iHeps, Cdx2 presumably initiates a hindgut (intestine) GRN in parallel with FoxA- and Hnf4α-driven foregut (liver) specification.

iHeps fully differentiate following transplantation into either mouse liver or intestine (Morris et al., 2014; Sekiya and Suzuki, 2011), suggesting that installation of downstream GRNs proceeds normally in this context and is directed by the *in vivo* niche. Investigating whether iHeps do indeed transition through a developmental intermediate will require mapping of the mammalian GRNs from each developmental stage, akin to analyses performed in adult cell types and tissues (Cahan et al., 2014). This approach would explore a number of different possibilities. The first possibility is whether an upper-level GRN corresponding to early endoderm with both foregut and hindgut potential is activated. Alternatively, GRNs further downstream that equate to regionalized gut might be activated in parallel within the same cell. Given that foregut and hindgut programs repress each other in normal development (Gao et al., 2009), it is conceivable that iHeps actually consist of two populations of cells that possess discrete gut potential. Single-cell transcriptome analysis will help to dissect these possibilities, as will further investigation of the functional potential of these cells, i.e. do clonal iHep populations have the potential to engraft into distinct regions of the gut? Regardless of the precise mechanism, it is evident that FoxA and Hnf4α are able to engage silenced developmental GRNs to specify a broad endoderm fate; therefore, iHeps can be described as endoderm progenitor-like cells.

Beyond the endoderm, an unexpectedly broad fate specification also appears in the conversion of B-cells to macrophages by the pioneer factor C/ebpα (Bussmann et al., 2009; Di Tullio et al., 2011). In this case, the hematopoietic stem and progenitor GRNs transiently appear at the population level during the course of conversion (Morris et al., 2014). In addition, a recent report demonstrated that Ascl1, a neural pioneer factor, activates a neural progenitor-like state and myogenic program when overexpressed in fibroblasts (Treutlein et al., 2016). Although these engineered cell populations might not represent bona fide embryonic progenitors, it is nonetheless clear that they exist due to the engagement of developmental programs at some level. Interestingly, though, the potential of these cells is much broader than that of their closest *in vivo* correlates, suggesting that the pioneer factors are engaging with a much wider array of silent chromatin than expected.

A Pioneer factor-mediated conversion

Progenitor-like state

Pioneer factor expression

Additional maturation steps

Engagement of silent chromatin and additional cooperative factor binding allows activation of the GRN battery

B Terminal selector-mediated conversion

??

Terminal selector gene expression

Factors alone cannot engage silent chromatin; GRN battery not activated
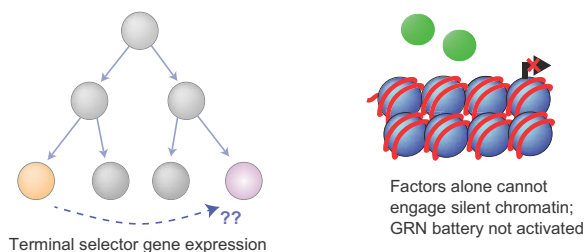
Fig. 4. Possible lineage reprogramming strategies. (A) Pioneer factor-mediated conversion of cell type A (orange) to cell type B (purple) depends on transition via a progenitor state (red circle), from which point cells must be matured. This approach is more feasible given the function of pioneer factors (red) to engage silent chromatin and recruit additional cooperative factors (green) to activate developmental GRNs. (B) Non-pioneer factor-mediated conversion aims to directly convert cells between fully differentiated states without reversion to a progenitor state. This could theoretically be achieved via the expression of terminal selector genes, which activate GRN batteries, although this approach is relatively infeasible given the low likelihood of these factors being able to engage silent chromatin in the host cell.

Understanding mechanisms of pioneer factor engagement with developmental GRNs to precisely engineer fate

In order to fully harness pioneer factor function for direct lineage reprogramming it is crucial that we understand how to constrain pioneer factors to their appropriate targets, and to understand how this induces the desired cell fate. Combinatorial binding of multiple transcription factors is a central feature of gene regulation and control of cell identity, and thus must be carefully orchestrated during development. This is reflected by binding of transcription factors to only a subset of their putative binding motifs in a cell type-specific manner (Mahony et al., 2014), and the fact that current conversions are clearly targeting a much broader set of GRNs than previously anticipated. Several possible mechanisms could explain this assignment of broad developmental potential via pioneer factor expression: (1) overexpressed pioneer factors promiscuously engage with silent chromatin; (2) the cellular context of the host cell with regard to transcriptional repressors and activators does not appropriately corral pioneer factor activity; and (3) pioneer factors are expressed within an incompatible chromatin landscape, leading to off-target effects.

The first and most simplistic mechanism centers on the capacity of pioneer factors to engage with silent chromatin. Unnaturally high levels of transcription factor expression are typical during direct lineage conversion, and under these conditions pioneer factors might simultaneously engage targets associated with multiple fates in a promiscuous fashion. This broad specification of competence together with cognate activator expression would endow cells with greater lineage potential. To determine whether this hypothesis holds true, a good starting point would be to compare the binding sites of pioneer factors under conditions of low and high levels of exogenous expression by performing ChIP-Seq analysis. The second possible mechanism that could explain the broad developmental potential endowed by pioneer factor overexpression takes into account evidence demonstrating that repressors act to govern pioneer factor activity. Genomic location analysis showed that one third of FoxA-bound sites are near silent genes in the adult mouse liver (Watts et al., 2011). These silent sites are enriched for repressors such as Rfx1 and type II nuclear hormone receptors. Seven kilobases downstream of *Cdx2* is a 'shadow enhancer', where FoxA binding is detected and yet *Cdx2* remains silent. Here, Rfx1, which is expressed in the liver, restricts transcriptional activation via FoxA. I propose that in FoxA-driven conversion to iHeps, Rfx1 is expressed at low levels in fibroblasts, or perhaps high exogenous FoxA saturates Rfx1. As a consequence, FoxA binding at the shadow enhancer results in *Cdx2* expression and initiation of hindgut GRNs in parallel with foregut specification. Introduction of exogenous Rfx1 or modulating levels of exogenous factors in these particular engineering protocols could potentially direct the specification of a foregut as opposed to hindgut program. In support of this, lower levels of Foxa1 and Hnf4α expression generate cells resembling mature hepatocytes, as evidenced by their limited capacity for cell division (Morris et al., 2014). This suggests that low levels of these exogenous factors better reflect normal physiological conditions and do not saturate co-repressors such as Rfx1; thus, intestinal GRNs controlled by Cdx2 remain silent under these conditions.

In addition to repressors, the availability of transcriptional activators in starting cell types is likely to have a major influence on which GRNs are engaged by conversion factors. For example, during hepatic development, Hnf4α and Foxa2 enhancer occupancy and control of gene expression is differentiation dependent (Alder et al., 2014). Alder and colleagues demonstrated that Hippo

signaling status impacts nuclear Yap1 and Tead2 levels, which in turn induces enhancer switching by Hnf4α and Foxa2 as liver development progresses. This finding that transcription factor-enhancer interactions are not only tissue specific but also differentiation dependent adds an extra layer of complexity to cell fate engineering, suggesting that further considerations must be taken into account to ensure that the appropriate GRNs are accessed by conversion factors. One potential approach to refine engineering strategies would be to profile endogenous co-factor expression in starting cell populations in order to select the most appropriate cell type to drive conversion toward a specific target lineage. A comprehensive assessment of GRN activation in an array of starting cell types would help to build a more detailed picture of how the co-factor expression landscape directs cell fate to a more defined identity.

Finally, how might the host cell chromatin landscape impact pioneer factor behavior? Reprogramming to pluripotency provides some initial clues. In fibroblasts, genes required to establish pluripotency are initially 'locked' within H3K9me3-enriched heterochromatin. Reprogramming factors are barred from these regions at first, and the location of these regions is cell type specific. This barrier to reprogramming can be lowered by reducing H3K9me3 deposition (Soufi et al., 2012). Looking at this from another perspective, can a chromatin signature predict host cell-specific on/off target effects (see Glossary, Box 1) of FoxA engagement with the genome, and can this help assess which level of the GRN hierarchy will be accessed? The active histone modifications H3K4me2, H3K4me1 and H3K9ac are deposited at the time of, or following, FoxA binding (Sérandour et al., 2011; Taube et al., 2010). Moreover, reduction of H3K4me1 and H3K4me2 correlates with FoxA binding (Lupien et al., 2008), suggesting that there may indeed be a cell type-specific FoxA binding predictor. In support of this, a pattern emerges in reprogramming fibroblasts to neurons: a 'trivalent' chromatin signature of H3K4me1, H3K27ac and H3K9me3 correlates with pioneer factor binding of Ascl1 and is predictive of the reprogramming outcome (Wapinski et al., 2013). For example, Acsl1 targets in keratinocytes are not enriched for this trivalent signature and conversion to induced neurons (iNs) from these cells has so far been unsuccessful (Wapinski et al., 2013). That said, not all Ascl1 occupancy can be predicted by the trivalent signature, suggesting that the co-factor expression landscape also plays a crucial role in this conversion. Future experiments to cross-check these potential host cell-specific patterns with GRN activation in a given cell type would be very revealing.

Together, the current evidence suggests that the epigenetic landscape into which conversion factors are introduced influences their access to developmental GRNs. The case of Ascl1 is particularly interesting: this transcription factor has been described as an 'on-target' pioneer factor, meaning that its initial binding sites when expressed in fibroblasts correspond to its cognate binding sites in neural progenitor cells (Wapinski et al., 2013). Whether the on-target binding of Ascl1 during reprogramming generates cells closer to the target cell type of mature, functional neurons remains to be seen.

Even on-target pioneer factors have their off days

Neural conversions outperform other fate conversion protocols in terms of approximating mature target cell identity (Cahan et al., 2014). Most conversions to neural fate involve expression of the pioneer factor Ascl1 (Fig. 1). Ascl1 alone is sufficient to induce immature glutamatergic neurons, and Myt1l and Brn2 (Pou3f2) are

added to the cocktail primarily to enhance the neuronal maturation process (Chanda et al., 2014; Vierbuchen et al., 2010). This suggests that exogenous Ascl1 also accesses developmental GRNs, but the hierarchical level at which this occurs is not clear. In contrast to FoxA factors, Ascl1 is a more cell fate-specific factor in that it is restricted exclusively to neural fate specification: *Ascl1* is a pro-neural gene expressed in a subset of central and peripheral neural progenitors (Guillemot et al., 1993; Lo et al., 1991). Moreover, Ascl1 has been shown to act as an on-target pioneer factor (Wapinski et al., 2013), directly engaging with appropriate neural genes to drive conversion, regardless of the chromatin landscape. Following its binding to silent, low-signal chromatin, Ascl1 subsequently recruits other factors required for lineage conversion to those sites. This is in contrast to induction of pluripotency by the Yamanaka factors, where Oct4, Sox2 and Klf4 act as pioneer factors but are mislocalized in the initial stages of reprogramming (Soufi et al., 2012).

Do these findings suggest that fibroblast to neuron conversion is relatively direct? A recent single-cell RNA-sequencing analysis of Ascl1-mediated conversion reveals that iNs in fact transition through a 'fractional' neural progenitor cell state associated with the expression of several neural progenitor genes, although expression of the canonical neural progenitor genes *Sox2* and *Pax6* was not induced (Treutlein et al., 2016). This raises the possibility that atypical progenitor-like states are produced during lineage conversion via a unique intermediate transcriptional state, and also that as yet undiscovered equivalents to these states might exist *in vivo*. Moreover, during conversion in the absence of the Myt1l and Brn2 maturation factors, an alternative fate corresponding to a myogenic program is specified (Treutlein et al., 2016). This suggests that Ascl1, like FoxA, induces much broader potential than previously anticipated. This lends further support to my hypothesis, as presented here, that pioneer factor-driven direct conversion engages hierarchical developmental GRNs, and demonstrates the value of single-cell analysis in dissecting these mechanisms.

Summary and perspectives

Throughout this article, I have presented evidence that direct lineage conversion largely produces immature cells. I hypothesize that this immaturity arises due to the use of pioneer factors that engage silent chromatin to activate developmental GRNs at different levels of the developmental hierarchy. Depending on the normal function of the pioneer factor during development, as well as the pre-existing transcriptional and epigenetic landscape of the host cell, the outcome of direct lineage conversion can vary widely. It is for this reason that we observe varying degrees of maturity in the cells that are generated. It is clear that current lineage conversion technologies do not produce fully differentiated cells – only the addition of maturation factors or a period of maturation in the *in vivo* niche can install peripheral differentiation batteries (Fig. 3). There are practical implications for these restrictions of lineage reprogramming. Perhaps most importantly, for disease modeling *in vitro* it is often crucial to produce fully differentiated and functioning cells. In this case, a detailed map of GRN hierarchies will be fundamental for the development of strategies to enhance cell maturation.

The currently favored path for direct lineage conversion employs pioneer factor expression. I hypothesize that this approach actually relies on the reinitiation of silenced developmental programs and the recruitment of co-expressed factors to drive conversion effectively. As a result, the converting cells transition through a developmental

intermediate and possess broader potential than anticipated. Such progenitor-like cells have been shown to respond to the *in vivo* niche and mature to become indistinguishable from target cell identity. I suggest that, at present, this represents the most feasible route for faithful cell fate engineering, given the ability of pioneer factors to engage silent chromatin and endow cells with lineage competence (Fig. 4A). One advantage of this approach lies in the potential to expand these immature cells, although a limitation is represented by possible difficulties in maturing these cells, a critical step for their utility in disease modeling and drug discovery *in vitro*. To address this, one possible strategy would be to recapitulate the developmental GRN hierarchy by expressing terminal selector transcription factors to install cell type-specific differentiation batteries. Indeed, this approach has proven successful in the conversion of human fibroblasts to iHeps, whereby the cells were first reprogrammed to a progenitor-like state, followed by maturation with a cocktail of factors including C/ebpα (Du et al., 2014). A series of maturation steps might assist in silencing the original host cell identity, as is observed upon maturation of engineered progenitors in the *in vivo* niche. This pioneer factor-mediated approach might represent the most feasible path to fully mature, functional cell identity.

One might argue that the use of pioneer factors in direct lineage conversion should be avoided due to their broad activation of progenitor-like states. As an alternative, only terminal selector transcription factors associated with terminal differentiation batteries would be expressed, potentially bypassing developmental GRN activation (Fig. 4B). Theoretically, this approach would drive a truly direct conversion between mature fates. In reality, however, this might not be feasible since these terminal selector transcription factors may not have the capacity to engage silent chromatin in host cells. For example, expression of terminal selector genes in *C. elegans* can only convert germ cells to neurons in the absence of LIN-53, a histone chaperone (Tursun et al., 2011), demonstrating the importance of chromatin context. In addition, simply activating target differentiation batteries in host cells might not silence the original cell identity.

Systems-level analyses will help disentangle the complex network biology that is required to manipulate cell identity. Several computational approaches have recently emerged and will be invaluable for these endeavors. The first, CellNet (Cahan et al., 2014), reconstructs GRNs using publicly available gene expression data for a range of cell types and tissues, and uses this approach to assess the similarity of engineered cells relative to their targets. Such GRN analysis of developmental intermediates will also help identify key transcription factors that regulate both upper-level GRNs and differentiation batteries. A similar computation approach is taken by KeyGenes, which employs a collection of human fetal transcriptional profiles collected from tissues and organs at different developmental stages to assess the cell identity of differentiating human tissues (Roost et al., 2015). To specifically identify pioneer factor activity, a computational method termed 'protein interaction quantitation' (PIQ) helps to identify pioneer factors via assessment of genome-wide DNase I hypersensitivity profiles (Sherwood et al., 2014). Using this approach, pioneer factor activity can be predicted via computational analysis of genomic DNA motifs in combination with knowledge of the 3D structure of the factor (Soufi et al., 2015). Finally, a recent platform called Mogrify (Rackham et al., 2016) combines gene expression data with regulatory network information from over 300 different human cell and tissue types to predict factors capable of inducing lineage conversion between human cell types. Integrating these data within a developmental context will

promote a better understanding of conversion mechanisms, leading the way to more precise engineering of cell identity.

Moving forward, it will be essential to address the reprogramming roles of additional regulators of cell fate, namely microRNAs (miRNAs), small molecules and epigenetic regulators. miRNAs in particular are key regulators of GRN architecture (reviewed by Bartel, 2009; Herranz and Cohen, 2010), and are becoming increasingly utilized in cell fate engineering strategies. For example, miR-9/9* and miR-124 expression in human fibroblasts induces neural fate, where conversion efficiency and maturation are promoted by introduction of ASCL1, MYT1L and NEUROD2 (Yoo et al., 2011). Similarly, miR-124 expression in concert with BRN2 and MYT1L reprograms human fibroblasts to neurons (Ambasudhan et al., 2011). Given the capacity of miRNAs to silence gene expression, it is probable that target cell type-specific miRNAs act to suppress host cell fate, thereby driving fate conversion. For example, miR-124 expression decreases the levels of non-neuronal gene transcripts when expressed in non-neural cells (Lim et al., 2005). This is a promising approach given that most converted cells retain remnants of their original cell identity (Cahan et al., 2014; Morris et al., 2014). In addition, miRNAs are likely to drive neuronal conversion by inducing changes in neuron-specific BAF chromatin remodeling complexes (Tang et al., 2013), fostering a neuronal ground state permissive for the activity of the terminal selector genes that control the batteries that drive conversion to specific neuronal subtypes (Victor et al., 2014). This parallels the induction of lineage competence via pioneer factor expression, followed by recruitment of cooperative factors to elicit changes in fate. Whether miRNA expression can support truly direct conversion remains an open question, and much more work is required to dissect the relationship between additional regulatory levels, such as miRNA dynamics and signaling, in lineage reprogramming. Ultimately, although much of the mechanistic detail regarding lineage reprogramming remains unknown, it is clear that this phenomenon invokes key developmental principles, and that pioneer factors play an important role in this. Harnessing the instructive nature of pioneer factors to activate broad lineage potential followed by refinement of cell fate with additional factors and/or niche signals might therefore represent the most efficient and logical strategy for producing fully differentiated cells.

References

Alder, O., Cullum, R., Lee, S., Kan, A. C., Wei, W., Yi, Y., Garside, V. C., Bilenky, M., Griffith, M., Morrissy, A. S. et al. (2014). Hippo signaling influences HNF4A and FOXA2 enhancer switching during hepatocyte differentiation. Cell Rep. 9, 261-271.

Ambasudhan, R., Talantova, M., Coleman, R., Yuan, X., Zhu, S., Lipton, S. A. and Ding, S. (2011). Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. Cell Stem Cell 9, 113-118.

Bar-Nur, O., Verheul, C., Sommer, A. G., Brumbaugh, J., Schwarz, B. A., Lipchina, I., Huebner, A. J., Mostoslavsky, G. and Hochedlinger, K. (2015). Lineage conversion induced by pluripotency factors involves transient passage through an iPSC stage. Nat. Biotechnol. 33, 761-768.

Bartel, D. P. (2009). MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233.

Bernardo, G. M. and Keri, R. A. (2012). FOXA1: a transcription factor with parallel functions in development and cancer. Biosci. Rep. 32, 113-130.

Berninger, B., Costa, M. R., Koch, U., Schroeder, T., Sutor, B., Grothe, B. and Götz, M. (2007). Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. J. Neurosci. 27, 8654-8664.

Blau, H. M., Chiu, C.-P. and Webster, C. (1983). Cytoplasmic activation of human nuclear genes in stable heterocaryons. Cell 32, 1171-1180.

Bossard, P. and Zaret, K. S. (1998). GATA transcription factors as potentiators of gut endoderm differentiation. Development 125, 4909-4917.

Bossard, P., Zaret, K. S., Ang, S.-L., Rossant, J., Ang, S.-L., Wierda, A., Wong, D., Stevens, K. A., Cascio, S., Rossant, J. et al. (2000). Repressive and restrictive mesodermal interactions with gut endoderm: possible relation to Meckel's Diverticulum. Development 127, 4915-4923.

Bussmann, L. H., Schubert, A., Vu Manh, T. P., De Andres, L., Desbordes, S. C., Parra, M., Zimmermann, T., Rapino, F., Rodriguez-Ubreva, J., Ballestar, E. et al. (2009). A robust and highly efficient immune cell reprogramming system. Cell Stem Cell 5, 554-566.

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

Caiazzo, M., Dell'Anno, M. T., Dvoretskova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T. D., Menegon, A., Roncaglia, P., Colciago, G. et al. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. Nature 476, 224-227.

Carroll, J. S., Liu, X. S., Brodsky, A. S., Li, W., Meyer, C. A., Szary, A. J., Eeckhoute, J., Shao, W., Hestermann, E. V., Geistlinger, T. R. et al. (2005). Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. Cell 122, 33-43.

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 Rep. 3, 282-296.

Cohen, D. E. and Melton, D. (2011). Turning straw into gold: directing cell fate for regenerative medicine. Nat. Rev. Genet. 12, 243-252.

D'Amour, K. A., Bang, A. G., Eliazer, S., Kelly, O. G., Agulnick, A. D., Smart, N. G., Moorman, M. A., Kroon, E., Carpenter, M. K. and Baetge, E. E. (2006). Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. Nat. Biotechnol. 24, 1392-1401.

Davidson, E. H. and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. Science 311, 796-800.

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 Tullio, A., Vu Manh, T. P., Schubert, A., Castellano, G., Månsson, R. and Graf, T. (2011). CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc. Natl. Acad. Sci. USA 108, 17016-17021.

Du, Y., Wang, J., Jia, J., Song, N., Xiang, C., Xu, J., Hou, Z., Su, X., Liu, B., Jiang, T. et al. (2014). Human hepatocytes with drug metabolic function induced from fibroblasts by lineage reprogramming. Cell Stem Cell 14, 394-403.

Feng, R., Desbordes, S. C., Xie, H., Tillo, E. S., Pixley, F., Stanley, E. R. and Graf, T. (2008). PU.1 and C/EBPalpha/beta convert fibroblasts into macrophage-like cells. Proc. Natl. Acad. Sci. USA 105, 6057-6062.

Friedman, J. R. and Kaestner, K. H. (2006). The FoxA family of transcription factors in development and metabolism. Cell. Mol. Life Sci. 63, 2317-2328.

Gao, N. and Kaestner, K. H. (2010). Cdx2 regulates endo-lysosomal function and epithelial cell polarity. Genes Dev. 24, 1295-1305.

Gao, N., White, P. and Kaestner, K. H. (2009). Establishment of intestinal identity and epithelial-mesenchymal signaling by Cdx2. Dev. Cell 16, 588-599.

Garrison, W. D., Battle, M. A., Yang, C., Kaestner, K. H., Sladek, F. M. and Duncan, S. A. (2006). Hepatocyte nuclear factor 4alpha is essential for embryonic development of the mouse colon. Gastroenterology 130, 1207-1220.

Gualdi, R., Bossard, P., Zheng, M., Hamada, Y., Coleman, J. R. and Zaret, K. S. (1996). Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. Genes Dev. 10, 1670-1682.

Guillemot, F., Lo, L.-C., Johnson, J. E., Auerbach, A., Anderson, D. J. and Joyner, A. L. (1993). Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. Cell 75, 463-476.

Gurdon, J. B., Elsdale, T. R. and Fischberg, M. (1958). Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei. Nature 182, 64-65.

Hayhurst, G. P., Lee, Y.-H., Lambert, G., Ward, J. M. and Gonzalez, F. J. (2001). Hepatocyte nuclear factor 4alpha (nuclear receptor 2A1) is essential for maintenance of hepatic gene expression and lipid homeostasis. Mol. Cell. Biol. 21, 1393-1403.

Heinrich, C., Blum, R., Gascón, S., Masserdotti, G., Tripathi, P., Sánchez, R., Tiedt, S., Schroeder, T., Götz, M. and Berninger, B. (2010). Directing astroglia from the cerebral cortex into subtype specific functional neurons. PLoS Biol. 8, e1000373.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y. C., Laslo, P., Cheng, J. X., Murre, C., Singh, H. and Glass, C. K. (2010). Simple combinations of lineage-

determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* **38**, 576-589.

Hendry, C. E., Vanslambrouck, J. M., Ineson, J., Suhaimi, N., Takasato, M., Rae, F. and Little, M. H. (2013). Direct transcriptional reprogramming of adult cells to embryonic nephron progenitors. *J. Am. Soc. Nephrol.* **24**, 1424-1434.

Herranz, H. and Cohen, S. M. (2010). MicroRNAs and gene regulatory networks: managing the impact of noise in biological systems. *Genes Dev.* **24**, 1339-1344.

Ho, J. W. K., Jung, Y. L., Liu, T., Alver, B. H., Lee, S., Ikegami, K., Sohn, K.-A., Minoda, A., Tolstorukov, M. Y., Appert, A. et al. (2014). Comparative analysis of metazoan chromatin organization. *Nature* **512**, 449-452.

Hobert, O. (2008). Regulatory logic of neuronal diversity: terminal selector genes and selector motifs. *Proc. Natl. Acad. Sci. USA* **105**, 20067-20071.

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.

Iwafuchi-Doi, M. and Zaret, K. S. (2016). Cell fate control by pioneer transcription factors. *Development* **143**, 1833-1837.

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. A., Gu, T. et al. (2011). Comprehensive analysis of the chromatin landscape in Drosophila melanogaster. *Nature* **471**, 480-485.

Kurian, L., Sancho-Martinez, I., Nivet, E., Aguirre, A., Moon, K., Pendaries, C., Volle-Challier, C., Bono, F., Herbert, J.-M., Pulecio, J. et al. (2013). Conversion of human fibroblasts to angioblast-like progenitor cells. *Nat. Methods* **10**, 77-83.

Lim, L. P., Lau, N. C., Garrett-Engele, P., Grimson, A., Schelter, J. M., Castle, J., Bartel, D. P., Linsley, P. S. and Johnson, J. M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. *Nature* **433**, 769-773.

Lo, L. C., Johnson, J. E., Wuenschell, C. W., Saito, T. and Anderson, D. J. (1991). Mammalian achaete-scute homolog 1 is transiently expressed by spatially restricted subsets of early neuroepithelial and neural crest cells. *Genes Dev.* **5**, 1524-1537.

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

Lupien, M., Eeckhoute, J., Meyer, C. A., Wang, Q., Zhang, Y., Li, W., Carroll, J. S., Liu, X. S. and Brown, M. (2008). FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958-970.

Mahony, S., Edwards, M. D., Mazzoni, E. O., Sherwood, R. I., Kakumanu, A., Morrison, C. A., Wichterle, H. and Gifford, D. K. (2014). An integrated model of multiple-condition ChIP-Seq data reveals predeterminants of Cdx2 binding. *PLoS Comput. Biol.* **10**, e1003501.

Marro, S., Pang, Z. P., Yang, N., Tsai, M.-C., Qu, K., Chang, H. Y., Südhof, T. C. and Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* **9**, 374-382.

Maza, I., Caspi, I., Zviran, A., Chomsky, E., Rais, Y., Viukov, S., Geula, S., Buenrostro, J. D., Weinberger, L., Krupalnik, V. et al. (2015). Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors. *Nat. Biotechnol.* **33**, 769-774.

McLin, V. A., Rankin, S. A. and Zorn, A. M. (2007). Repression of Wnt/beta-catenin signaling in the anterior endoderm is essential for liver and pancreas development. *Development* **134**, 2207-2217.

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

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

Oldershaw, R. A., Baxter, M. A., Lowe, E. T., Bates, N., Grady, L. M., Soncin, F., Brison, D. R., Hardingham, T. E. and Kimber, S. J. (2010). Directed differentiation of human embryonic stem cells toward chondrocytes. *Nat. Biotechnol.* **28**, 1187-1194.

Oshima, K., Shin, K., Diensthuber, M., Peng, A. W., Ricci, A. J. and Heller, S. (2010). Mechanosensitive hair cell-like cells from embryonic and induced pluripotent stem cells. *Cell* **141**, 704-716.

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

Parviz, F., Matullo, C., Garrison, W. D., Savatski, L., Adamson, J. W., Ning, G., Kaestner, K. H., Rossi, J. M., Zaret, K. S. and Duncan, S. A. (2003). Hepatocyte nuclear factor 4alpha controls the development of a hepatic epithelium and liver morphogenesis. *Nat. Genet.* **34**, 292-296.

Pereira, C.-F., Chang, B., Qiu, J., Niu, X., Papatsenko, D., Hendry, C. E., Clark, N. R., Nomura-Kitabayashi, A., Kovacic, J. C., Ma'ayan, A. et al. (2013). Induction of a hemogenic program in mouse fibroblasts. *Cell Stem Cell* **13**, 205-218.

Peter, I. S. and Davidson, E. H. (2011). Evolution of gene regulatory networks controlling body plan development. *Cell* **144**, 970-985.

Peter, I. S. and Davidson, E. H. (2016). Implications of developmental gene regulatory networks inside and outside developmental biology. *Curr. Top. Dev. Biol.* **117**, 237-251.

Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., Björklund, A., Lindvall, O., Jakobsson, J. and Parmar, M. (2011). Direct conversion of human fibroblasts to dopaminergic neurons. *Proc. Natl. Acad. Sci. USA* **108**, 10343-10348.

Rackham, O. J., Firas, J., Fang, H., Oates, M. E., Holmes, M. L., Knaupp, A. S., FANTOM Consortium, Suzuki, H., Nefzger, C. M., Daub, C. O. et al. (2016). A predictive computational framework for direct reprogramming between human cell types. *Nat. Genet.* **48**, 331-335.

Rashid, S. T., Corbineau, S., Hannan, N., Marciniak, S. J., Miranda, E., Alexander, G., Huang-Doran, I., Griffin, J., Ahrlund-Richter, L., Skepper, J. et al. (2010). Modeling inherited metabolic disorders of the liver using human induced pluripotent stem cells. *J. Clin. Invest.* **120**, 3127-3136.

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

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

Sekiya, T. and Zaret, K. S. (2007). Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. *Mol. Cell* **28**, 291-303.

Sérandour, A. A., Avner, S., Percevault, F., Demay, F., Bizot, M., Lucchetti-Miganeh, C., Barloy-Hubler, F., Brown, M., Lupien, M., Métivier, R. et al. (2011). Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* **21**, 555-565.

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.

Son, E. Y., Ichida, J. K., Wainger, B. J., Toma, J. S., Rafuse, V. F., Woolf, C. J. and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* **9**, 205-218.

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

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.

Szabo, E., Rampalli, S., Risueño, R. M., Schnerch, A., Mitchell, R., Fiebig-Comyn, A., Levadoux-Martin, M. and Bhatia, M. (2010). Direct conversion of human fibroblasts to multilineage blood progenitors. *Nature* **468**, 521-526.

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.

Tang, J., Yoo, A. S. and Crabtree, G. R. (2013). Reprogramming human fibroblasts to neurons by recapitulating an essential microRNA-chromatin switch. *Curr. Opin. Genet. Dev.* **23**, 591-598.

Taube, J. H., Allton, K., Duncan, S. A., Shen, L. and Barton, M. C. (2010). Foxa1 functions as a pioneer transcription factor at transposable elements to activate Afp during differentiation of embryonic stem cells. *J. Biol. Chem.* **285**, 16135-16144.

Thier, M., Wörsdörfer, P., Lakes, Y. B., Gorris, R., Herms, S., Opitz, T., Seiferling, D., Quandel, T., Hoffmann, P., Nöthen, M. M. et al. (2012). Direct conversion of fibroblasts into stably expandable neural stem cells. *Cell Stem Cell* **10**, 473-479.

Touboul, T., Hannan, N. R. F., Corbineau, S., Martinez, A., Martinet, C., Branchereau, S., Mainot, S., Strick-Marchand, H., Pedersen, R., Di Santo, J. et al. (2010). Generation of functional hepatocytes from human embryonic stem cells under chemically defined conditions that recapitulate liver development. *Hepatology* **51**, 1754-1765.

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. et al. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**, 391-395.

Tursun, B., Patel, T., Kratsios, P. and Hobert, O. (2011). Direct conversion of C. elegans germ cells into specific neuron types. *Science* **331**, 304-308.

Uchida, O., Nakano, H., Koga, M. and Ohshima, Y. (2003). The C. elegans che-1 gene encodes a zinc finger transcription factor required for specification of the ASE chemosensory neurons. *Development* **130**, 1215-1224.

van der Sluis, M., Vincent, A., Bouma, J., Korteland-Van Male, A., van Goudoever, J. B., Renes, I. B. and Van Seuningen, I. (2008). Forkhead box transcription factors Foxa1 and Foxa2 are important regulators of Muc2 mucin expression in intestinal epithelial cells. *Biochem. Biophys. Res. Commun.* **369**, 1108-1113.

van Oevelen, C., Collombet, S., Vicent, G., Hoogenkamp, M., Lepoivre, C., Badeaux, A., Bussmann, L., Sardina, J. L., Thieffry, D., Beato, M. et al. (2015).

DEVELOPMENT

C/EBPα activates pre-existing and de novo macrophage enhancers during induced pre-B cell transdifferentiation and myelopoiesis. *Stem Cell Rep.* **5**, 232-247.

Victor, M. B., Richner, M., Hermanstyne, T. O., Ransdell, J. L., Sobieski, C., Deng, P.-Y., Klyachko, V. A., Nerbonne, J. M. and Yoo, A. S. (2014). Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. *Neuron* **84**, 311-323.

Vierbuchen, T. and Wernig, M. (2011). Direct lineage conversions: unnatural but useful? *Nat. Biotechnol.* **29**, 892-907.

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.

Waddington, C. H. (1942). Canalization of development and the inheritance of acquired characters. *Nature* **150**, 563-565.

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.

Watts, J. A., Zhang, C., Klein-Szanto, A. J., Kormish, J. D., Fu, J., Zhang, M. Q. and Zaret, K. S. (2011). Study of FoxA pioneer factor at silent genes reveals Rfx-repressed enhancer at Cdx2 and a potential indicator of esophageal adenocarcinoma development. *PLoS Genet.* **7**, e1002277.

Willenbring, H. (2011). A simple code for installing hepatocyte function. *Cell Stem Cell* **9**, 89-91.

Yamamoto, K., Kishida, T., Sato, Y., Nishioka, K., Ejima, A., Fujiwara, H., Kubo, T., Yamamoto, T., Kanamura, N. and Mazda, O. (2015). Direct conversion of human fibroblasts into functional osteoblasts by defined factors. *Proc. Natl. Acad. Sci. USA* **112**, 6152-6157.

Yoo, A. S., Sun, A. X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R. E., Tsien, R. W. and Crabtree, G. R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* **476**, 228-231.

Yu, B., He, Z.-Y., You, P., Han, Q.-W., Xiang, D., Chen, F., Wang, M.-J., Liu, C.-C., Lin, X.-W., Borjigin, U. et al. (2013). Reprogramming fibroblasts into bipotential hepatic stem cells by defined factors. *Cell Stem Cell* **13**, 328-340.

Zaret, K. S. and Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227-2241.

Zhu, S., Rezvani, M., Harbell, J., Mattis, A. N., Wolfe, A. R., Benet, L. Z., Willenbring, H. and Ding, S. (2014). Mouse liver repopulation with hepatocytes generated from human fibroblasts. *Nature* **508**, 93-97.

DEVELOPMENT