Exploring the Plasticity of Cellular Fate Using Defined-Factor Reprogramming

Exploring the Plasticity of Cellular Fate

Using Defined-Factor Reprogramming

Abstract

Cellular fate, once established, is usually stable for the lifetime of the cell. However, the mechanisms that restrict the developmental potential of differentiated cells are in principle reversible, as demonstrated by the success of animal cloning from a somatic genome through somatic cell nuclear transfer (SCNT). An increased understanding of the molecular determinants of cell fate has also enabled the reprogramming of cell fate using defined transcription factors; recently, these efforts have culminated in the discovery of four genes that convert somatic cells into induced pluripotent stem cells (iPSCs), which resemble embryonic stem cells (ESCs) and can give rise to all the cell types in the body.

As a first step toward generating clinically useful iPSCs, we identified a small molecule, RepSox, that potently and simultaneously replaces two of the four exogenous reprogramming factors, *Sox2* and *cMyc*. This activity was mediated by the inhibition of the Transforming Growth Factor-β (Tgf-β) signaling pathway in incompletely reprogrammed intermediate cells. By isolating these stable intermediates, we showed that RepSox acts on them to rapidly upregulate the endogenous pluripotency factor, *Nanog*, allowing full reprogramming to pluripotency in the absence of *Sox2*.

We also explored lineage conversion as an alternative approach for producing a target cell type in a patient-specific manner, without first generating iPSCs. A

combination of pro-neural as well as motor neuron-selective factors could convert fibroblasts directly into spinal motor neurons, the cells that control all voluntary movement. The induced motor neurons (iMNs) displayed molecular and functional characteristics of *bona fide* motor neurons, actuating muscle contraction *in vitro* and even engrafting in the developing chick spinal cord when transplanted. Importantly, functional iMNs could be produced from fibroblasts of adult patients with the fatal motor neuron disease, amyotrophic lateral sclerosis (ALS).

Given the therapeutic value of generating patient-specific cell types on demand, defined-factor reprogramming is likely to serve as an important tool in regenerative medicine. It is hoped that the different approaches presented here can complement existing technologies to facilitate the study and treatment of intractable human disorders.

*This dissertation includes a supplementary file (**Movie 3.1. iMNs Induce Contraction of C2C12 Myotubes That Is Blocked by Curare Treatment**).*

Dedicated to my family:

Yoonho Son, Sohee Won, Miriam Y. Son
& Sookhee Lee

Acknowledgements

Many thanks are due some very special people who have made this journey possible.

I would first like to sincerely thank my advisor, Kevin Eggan, who has given me indispensable guidance and support throughout. Many a times I went into a meeting with Kevin, feeling inadequate or unsure about my work, only to come out feeling motivated and ready to face the challenge. Special thanks are also in order for my dissertation advisory committee, Florian Engert, Catherine Dulac and Alex Meissner, for their critical advice and encouragement over the years.

I am enormously indebted to Justin Ichida, an exceptional postdoctoral fellow in the Eggan lab who has mentored me patiently from day one. I have been privileged to work with Justin on all of the studies presented in this volume, and I cannot give him enough credit for pushing those works forward and for training me scientifically.

We had the good fortune of collaborating closely with excellent scientists in other laboratories. The chemical reprogramming studies in Chapter 2 were done in collaboration with Lee Rubin, Joel Blanchard and Kelvin Lam at Harvard Stem Cell Institute. Clifford Woolf and Brian Wainger at Children's Hospital, Boston, as well as Victor Rafuse and Jeremy Toma at Dalhousie University, Halifax, worked with us on the transdifferentiation studies in Chapter 3. I am deeply appreciative of their scientific insight and openness that led to the fruitfulness of these joint endeavors.

My colleagues and friends in the Eggan lab are some of the most intelligent, hard-working and fun people I have met. I have received help from everyone, in particular from Gabriella Boulting, Evangelos Kiskinis, Inna Tabansky, Julia Chung, Shila Mekhoubad, Sophie de Boer, Luis Williams, Steve Han, Jack Sandoe, Florian Merkle, Katie Koszka, Mariko Yamaki, Brandi Davis-Dusenbery, Seun Johnson-Akeju, Kaitlin Sandor, Ava Carter, Sean Singh, Nick Atwater and Diane Santos. It has been a pleasure working alongside such a terrific group of people.

My friends at Boston Korean SDA Church have helped to make the past five years in Boston, with five unforgiving winters and seemingly endless challenges, not only bearable but enjoyable. I am especially grateful to Daniel Park, Eunjin Choi, Juhyun Park, Greg Burnett, Sae Oh, Mary Park, Jason Sun, Carter Kyungmin Park, Samuel Koo and Jan Hyunkyung Ryu.

I would also like to acknowledge my classmates at the Department of Molecular and Cellular Biology at Harvard. All twelve of them have been very successful in their graduate careers, and I have no doubt that they will continue to excel.

There are people on the other side of the Atlantic that I must thank. I would not be here without the teachers and friends at my alma mater, Cambridge University, UK. Alan Winter, Margaret Stanley, Gavin Alexander and David Klenerman were my tutors at Christ's College, Cambridge. They graciously went beyond the call of duty and actively encouraged me to pursue my interest in science. Moreover, I was extremely lucky to meet friends who continually supported and challenged me, particularly Lulu Matakala-Chishinga, Amos Burke, Shreyas Mukund and Taeyang Kim – all remarkable individuals in every way.

I also owe much to these wonderful friends from Wokingham and Bracknell, UK, who are some of the kindest and most generous people one could hope to meet: Stella Heatley, Rachael Burrows, Mark Ainsworth, Maki Shimizu, Jena Hulbert, Steven Hulbert, Ashley Brooks, Eric Pilmoor and David Baraswayne; as well as John and Vaselinka Becejac, Charles and Janet Schlunt, Colin and Margaret Mitchell, Ron and Elaine Emerson, and Keith Berry.

During my most difficult times in graduate school, I was so fortunate to have the friendship of Guille Canas, Justin Ichida, April Hong, Minjeong Kim and Linda Oh. I was also privileged to share the brightest moments with them. They have each inspired me more than they know, and I am blessed to have these amazing people in my life.

Finally, I thank my family, who are the reason I did not give up and to whom I dedicate this work. I could not have asked for a more loving family, and am very excited to celebrate with them the end of this long chapter.

April 2012
Cambridge, MA

Table of Contents

List of Figures, Tables, and Movie Legend

Chapter 2

Figures

Tables

Chapter 3

Figures

Tables

Movie Legend

List of Abbreviations

4-OHT	4-Hydroxytamoxifen
5mC	5-Methylcytosine
α-BTX	α-Bungarotoxin
ACh	Acetylcholine
ALS	Amyotrophic lateral sclerosis
AP-MS	Affinity purification followed by mass spectrometry
ATP	Adenosine triphosphate
ATPase	Adenosine triphosphatase
Aza	5-Azacytidine
BAF	Brahma-related gene/Brahma-associated factor
BMP	Bone morphogenetic protein
BRG	Brahma-related gene
BRM	Brahma
cDNA	Complementary DNA
C/EBP	CCAAT/enhancer binding protein
ChAT	Choline acetyltransferase
ChIP	Chromatin-immunoprecipitated
CMP	Common myeloid progenitor
CNS	Central nervous system
CpG	Cytosine-phosphate-guanine
DA neuron	Dopaminergic neuron
DNMT	DNA methyltransferase
EGC	Embryonic germ cell
ES (or ESC)	Embryonic stem (cell)
FACS	Fluorescence-activated cell sorting
FGF	Fibroblast growth factor
FTD	Frontotemporal dementia
GABA	γ-Aminobutyric acid
GFP	Green fluorescent protein
GMP	Granulocyte/macrophage progenitor
GSK-3β	Glycogen synthase kinase 3β
H3K4me3	Histone H3 lysine 4 trimethylation
H3K27me3	Histone H3 lysine 27 trimethylation
HDAC	Histone deacetylase
HEF	Human embryonic fibroblast
HMTase	Histone methyltransferase
HSC	Hematopoietic stem cell
iβ-cell	Induced β-cell
iCM	Induced cardiomyocyte
ICM	Inner cell mass
iDA neuron	Induced dopaminergic neuron
IGF	Insulin-like growth factor
iHep	Induced hepatocyte

iMN	Induced motor neuron
iN (or iN cell)	Induced neuron (or induced neuronal cell)
iNPC	Induced neural progenitor cell
iNSC	Induced neural stem cell
iPS (or iPSC)	Induced pluripotent stem (cell)
IVF	*In vitro* fertilization
LIF	Leukemia inhibitory factor
MDB	Methyl-CpG binding domain
MeCP	Methyl-CpG binding protein
MEF	Mouse embryonic fibroblast
MEP	Megakaryocyte/erythrocyte progenitor
mES (or mESC)	Mouse embryonic stem (cell)
mRNA	Messenger RNA
NMJ	Neuromuscular junction
NPC	Neural progenitor cell
NSC	Neural stem cell
NuRD	Nucleosome remodeling deacetylase
PCR	Polymerase chain reaction
PD	Parkinson's disease
PRMT	Protein arginine methyltransferase
qPCR	Quantitative polymerase chain reaction
RA	Retinoic acid
RepSox	Replacement of *Sox2*
RFP	Red fluorescent protein
RT-PCR	Real-time polymerase chain reaction, or reverse transcription-polymerase chain reaction
RTT	Rett syndrome
SCNT	Somatic cell nuclear transfer
SHH	Sonic hedgehog
shRNA	Short-hairpin RNA
SMA	Spinal muscular atrophy
SMN	Survival of motor neuron
SOD1	Superoxide dismutase 1
SWI/SNF	Switch/sucrose nonfermentable
TDP-43	Transactivation-responsive DNA-binding protein 43
Tgf-β	Transforming growth factor-β
Tgfbr1	Transforming growth factor-β receptor 1
TH	Tyrosine hydroxylase
TSA	Trichostatin A
TTF	Tail-tip fibroblast
TTX	Tetrodotoxin
vChAT	Vesicular choline acetyltransferase
VPA	Valproic acid

Chapter 1

The Emergence of Cellular Reprogramming

Abstract

The pioneering amphibian cloning experiments demonstrate that differentiated cells retain the genetic information necessary for the development of an entire organism, but heritable epigenetic modifications and stable gene regulatory networks constrain cells to their committed lineages. However, master regulator transcription factors have been discovered that not only establish and maintain their associated cell fate in normal development and homeostasis, but also impose it on cells of unrelated lineages when overexpressed. Advances in nuclear reprogramming using such defined factors have revealed some fundamental principles of cell fate, and resulted in cellular conversions of high clinical importance.

Introduction

Our ability to survive and thrive depends on the function of a wide variety of cell types that are equipped for specific tasks. How this enormous cellular diversity is generated during development is a puzzle that still captivates developmental biologists. The genetic information for forming any cell type is contained in a single genome, such that a fertilized zygote gives rise to all the cells in the adult organism. As cells undergo successive stages of differentiation, they make a series of lineage commitment decisions that results in increased specialization and diversification. In nature, this is a unidirectional process that is rarely reversed. The progressive lineage restriction offers an obvious advantage: a cell stays committed to its past fate decisions to complete its developmental program and become a mature, functioning cell.

Conrad Waddington, in his book *The Strategy of the Genes* (Waddington, 1957), proposed an intuitive metaphor to describe this phenomenon, where a differentiating cell is likened to a ball rolling down a rugged mountain range. The path formed by peaks and valleys is highly branched, and each bifurcation point represents a developmental decision the cell must make. At the foot of the mountain, the balls comes to rest at various points of minimum potential energy that are separated from each other by significant heights; this is analogous to the inertia of terminally differentiated cell states.

Although this schema captures the tight link between cellular differentiation and the loss of developmental potential, it does not presuppose the nature of its valleys and hills – the molecular entities in or around the cell that instruct and restrict cell fate. Indeed, at the wake of the genetic revolution in the 1950s, the role of the nucleus during differentiation was still under debate (Briggs and King, 1952). One proposed mechanism

for ensuring that a cell does not aberrantly deviate from its lineage was to simply discard parts of the genome no longer necessary for the cell's identity; thus coupled to the selective destruction of genetic information, each lineage decision would be essentially irreversible.

In this opening chapter, I will outline how a dramatic refutation of this seemingly plausible model of differentiation laid the foundation for the field of regenerative biology and its promises for personalized medicine. Collectively, the body of work described below clearly demonstrates the possibility of manipulating cell fate, in particular by a remarkable process called nuclear reprogramming.

Defining Waddington's Landscape

The Constancy of the Genome

In order to answer this fundamental question on the mechanism of cellular differentiation, Briggs and King set up a heroic experimental system in the frog *Rana pipiens* (Briggs and King, 1952). Seeing that the egg cytoplasm normally supports the formation of an entire organism, they used it as a surrogate system in which to reveal the ability of a differentiated nucleus to direct development. In an approach called somatic cell nuclear transfer (SCNT), an enucleated egg cytoplasm was complemented by a donor nucleus whose potency they wished to test, and the embryo was allowed to develop. Strikingly, nuclei from the gastrula stage or earlier led to normal development of tadpoles in this assay, although success rate declined with increasing age of the donor (Briggs and King, 1957).

John Gurdon took this approach further in *Xenopus* to show that even more developmentally advanced nuclei retained this capacity (Gurdon 1960; Gurdon, 1962). He took intestinal epithelial cells of feeding tadpoles, of which greater than 99 per cent were differentiated cells readily identifiable by their morphology, as nuclear donors. Notably, if transplantation of a nucleus only led to the development of a partial blastula, cells from that arrested blastula were used as donors in a second round of transplantation into new recipient eggs. Combining the success rates of single and serial transfers, 7 per cent of donor nuclei were deemed competent to drive the formation of feeding tadpoles – a conservative estimate given the failures attributed to technical difficulties. Therefore, a strong statistical argument could be made for the idea that the information for generating complete cellular diversity is present in differentiated nuclei.

In more recent years, SCNT has been applied extensively to the cloning of mammals. Heralded by the birth of Dolly the sheep (Wilmut *et al*., 1997), whose genetic material derived from cultured mammary gland cells of a 6-year-old ewe, many other species including goats (Baguisi *et al.*, 1999), pigs (Polejaeva *et al.*, 2000), rabbits (Chesné *et al.*, 2002), and cats (Shin *et al.*, 2002) have been cloned. Although the existence of these cloned animals argues against the loss of genetic material as a widespread mechanism of differentiation, all of these cases lack a means to definitively trace the origin of the donor as a fully differentiated cell. A conclusive demonstration came when Hochedlinger and Jaenisch succeeded in cloning mice using lymphocytes as donors (Hochedlinger and Jaenisch, 2002). Each mouse clone contained an identical genetic rearrangement at the immunoglobulin or T cell receptor loci, definitively proving their derivation from a terminally differentiated, mature B cell or a T cell, respectively.

It is now thought that, barring a small number of cell types and somatic mutations, most somatic cells do not physically lose any of the genetic information first present in the zygote. One ramification of this conclusion is the theoretical possibility of reversing the developmental clock on cells – or, more broadly speaking, arbitrarily lifting or imposing cell fate restrictions independently of natural developmental processes.

Epigenetic Barriers

However, Waddington's hills are formidable barriers. As cells differentiate, the distinct states they adopt are highly resistant to perturbation, except in extreme, non-physiological situations such as during SCNT; spontaneous conversion of cell types is rarely observed in nature. An important mechanism for lineage restriction involves extra-genetic, or *epi*genetic, changes that are nonetheless abiding and heritable (Goldberg *et al*., 2007). Originally intended to loosely describe the interactions of genes that produce a stable phenotype (Waddington, 1942), the term 'epigenetics' now encompasses a number of highly conserved processes that help to establish and maintain stable cellular phenotypes without causing changes in the DNA sequence.

A well-characterized epigenetic mark is found at the level of the DNA molecule itself. Methylation of cytosine residues at CpG dinucleotides is associated with transcriptional repression, with stably silenced, heterochromatic regions being highly enriched in 5-methylcytosine (5mC) (Suzuki and Bird, 2008). Tissue-specific patterns of DNA methylation are established during development by *de novo* DNA methyltransferases (Dnmt3a and Dnmt3b), and are propagated by Dnmt1, the maintenance methyltransferase, to subsequent generations. Knocking out these enzymes

leads to early embryonic lethality (Li *et al.*, 1992; Okano *et al.*, 1999), and their knockdown is not tolerated in somatic cells types (Jackson-Grusby *et al.*, 2001), indicating the importance of maintaining global DNA methylation pattern in cellular differentiation and identity.

Histone octamers that help compact DNA into nucleosomes are targets of a great variety of covalent modifications that also contribute to cell identity (Wang *et al.*, 2004). The N- and C-terminal tails of histones, especially of histone H3, are extensively acetylated, methylated, phosphorylated and ubiquinated in a complex manner reflective of the transcriptional status at the region. In an informative study, Lander, Bernstein and colleagues used single-molecule based sequencing of chromatin-immunoprecipitated (ChIP) DNA to sample the genome-wide distribution of key histone marks at different stages of differentiation (Mikkelsen *et al.*, 2007). Interestingly, in pluripotent embryonic stem cells (ESCs), promoters of developmentally important genes possess 'bivalent' histone marks – that is, enriched for both activating (H3 lysine 4 trimethylation, H3K4me3) and repressive (H3K27me3) modifications. In more differentiated cells, these 'bivalent' marks often resolved to monovalent H3K27me3 in genes specifying unrelated lineages; genes important for the lineage of choice either remained bivalent or retained H3K4me3 only. These results are consistent with the involvement of these epigenetic modifications in globally orchestrating cell fate.

An important class of molecules, the ATP-dependent chromatin remodeling complexes, helps to impose lineage restriction by modulating the structure of the chromatin. In vitro studies indicate that all four families of vertebrate chromatin remodeling complexes increase the mobility of histone octamers along the DNA and

cause their repositioning (Côté *et al.*, 1994). However, their biological roles are far from uniform, ranging from maintaining euchromatin (Gaspar-Maia *et al*., 2009) to transcriptional activation and repression, as well as facilitating chromosomal segregation (Dirscherl *et al.*, 2004). Each multimeric complex is assembled combinatorially with more than one choice for many of the subunits; this molecular diversity translates to their functional non-redundancy and tissue- and stage-specific requirement. For instance, for the family of mammalian SWI/SNF complexes, BRG1 but not BRM is the necessary ATPase subunit in the pluripotent state (Ho *et al.*, 2009; Yan *et al.*, 2008); and the transition from neural progenitors to post-mitotic neurons requires the switching of another subunit, from BAF45A to BAF45B/C (Lessard *et al*., 2007).

As would be expected, there seem to be close molecular interactions among these epigenetic processes. One example is found in NuRD complexes of chromatin remodelers that allow cells to initiate differentiation by silencing pluripotency genes (Feng *et al.*, 2001; Zhang *et al.*, 1998). Interestingly, their subunits include methyl-CpG-binding domain (MBD) proteins that recognize methylated DNA, as well as histone deacetylases (HDACs) that remove activating histone marks. This provides a mechanistic link between DNA methylation and histone modifications associated with repressed transcription, as well as higher-order structural changes in the chromatin that further silence the region; satisfyingly, genome-wide patterns of DNA methylation and histone marks correlate well with each other (Meissner *et al.*, 2008). Although much more remains to be discovered in the way of the intersection of epigenetic processes – including non-coding RNAs whose roles as phenotypic regulators are only beginning to be appreciated (Bernstein and Allis,

2005) – the existence of molecular platforms for crosstalk highlights how distinct types of epigenetic modifications can give rise to a coherent phenotypic output.

If the epigenome is not merely a passive descriptor of the *status quo*, but actively endows stability to cell fate, then attempts to convert cellular identity would be resisted by the mechanisms described above; conversely, inhibition of epigenetic processes might make such a conversion more likely. Indeed, inhibition of HDACs by trichostatin A (TSA) was shown to increase the efficiency of bovine and murine SCNT (Enright *et al.*, 2003; Rybouchkin *et al*. 2006). This underscores the idea that modifications modulated by these enzymes guard against perturbations that can compromise a cell's identity; furthermore, it suggests that lowering epigenetic barriers, pharmacologically or genetically, can aid nuclear reprogramming.

Of course, for any reprogramming event to materialize, the epigenetic state of a differentiated cell must not be immutable, despite its stability even over cell divisions. Profound changes that occur during development, such as the rapid demethylation of sperm DNA following fertilization (Mayer *et al.*, 2000; Osward *et al*., 2000; Santos *et al.*, 2002), suggest that the epigenome can be highly responsive to signaling cues. The recent discovery of histone demethylases (Chang *et al.*, 2007; Shi *et al*., 2004) and pathways for demethylating DNA (Bhutani *et al*., 2010; Cortellino *et al*., 2011; Guo *et al*., 2011; He *et al.*, 2011) provides active mechanisms by which epigenetic modifications may be reversed. In fact, the possibility of SCNT is itself an argument for the dynamic nature of epigenetic regulation that does not preclude opportunities for a dramatic and global shift in cell state.

A Systems Approach and Gene Regulatory Networks

Do these epigenetic processes constitute the entirety of Waddington's barriers? In answering this question, let us consider the concept of cellular identity, or cell state. We may describe the state of a cell in terms of its suite of molecular features, such as the cohort of mRNAs and proteins it expresses. For example, the 20,000-25,000 putative protein-coding genes in the human genome can be considered independent variables whose expression levels collectively inform us of a cell's identity. Thus, the cell types we find in the body actually represent only a small subset of all theoretically possible cell states. In this framework, an intriguing question is this: why do certain states, or molecular configurations, manifest as stable cell types, while others are disallowed under physiological conditions? In other words, what are the constraints that render particular states accessible (minima in Waddington's landscape) or forbidden (maxima)?

The properties of epigenetic barriers – their stability and cell-type specific patterns – suggest that they are reliable indicators, and even stabilizers, of cell state. However, the question here is not one of mere correlation or maintenance, but of initial establishment. Orchestrating the genome-wide imposition of lineage-relevant restrictions would require recognition of specific genomic loci, especially given the erasure of epigenetic memory during early development (Mayer *et al.*, 2000; Osward *et al*., 2000; Santos *et al.*, 2002). The molecular machinery responsible for epigenetic regulation seems to lack intrinsic sequence specificity (Côté *et al*., 1998; Quinn *et al*., 1996), instead relying largely on their interaction partners to bind their correct targets (Armstrong *et al*., 1998; Cho *et al*., 1998; Fryer and Archer, 1998; Prochasson *et al*., 2003; Sullivan *et al*.,

2001; Wilson *et al*., 1996); therefore, they are unlikely to be the sole instructive force that defines the shape of Waddington's landscape.

Through the development of high-throughput experimental methods (Chen *et al*., 2008; Hinsby *et al.*, 2004; Kidder *et al*., 2008; Lo *et al*., 2006; Spooncer *et al.*, 2008), a systems perspective on cell state has begun to take shape (Kitano *et al*., 2002). Here, the cell is seen as an integrated system whose properties are a function of all of its component interactions. In essence, it is a description of cell state as a network consisting of nodes (cellular components) and their connections, known as edges (their interactions) (Ma'ayan *et al.*, 2008; Tyson *et al*., 2001).

There have been efforts to construct networks around a key molecular component, known for its central role in that particular cell type, that serves as an entry point. Wang and colleagues, for example, constructed a network of protein-protein interactions in mouse ESCs, starting from the pluripotency transcription factor Nanog: by an iterative use of affinity purification followed by mass spectrometry (AP-MS), they first identified proteins that bind Nanog, then in turn found their interaction partners (Wang *et al*., 2006). Transcriptional targets of ESC transcription factors such as Nanog (Chambers *et al*., 2003; Mitsui *et al*., 2003), as well as Oct4 (Nichols *et al*., 1998) and Sox2 (Avilion *et al*., 2003) have been mapped by several groups using chromatin immunoprecipitation (ChIP), revealing circuits rich in ESC-related genes (Boyer *et al*., 2005; Cole *et al*., 2008; Jiang *et al*., 2008; Kim *et al*., 2008; Liu *et al*., 2008; Loh *et al*., 2006; Sing *et al*., 2008) and reaffirming many of the findings from low-throughput studies. Other groups have pooled these datasets, each representing partial pictures, to derive more comprehensive networks representing the ES cell state (MacArthur *et al*., 2009; Muller *et al*., 2008).

Despite the current caveats, such as false positives and the question of directionality in an interaction (Ji *et al*., 2006; Mackay *et al*., 2007), the expectation is that these networks are not simply a birds-eye view of what traditional approaches offer, which are often linear chains of cause and effect concerning a small number of components. Mathematical analyses indicate that there are novel network properties that only emerge at the systems level (Hasty *et al*., 2001; Mogilner *et al*., 2006; Wilkinson *et al*., 2009) and which offer an explanation for how cellular fates are determined (Enver *et al*., 2009; Huang, 2009; MacArthur *et al*., 2009). Intuitively, the interrelationships within a network can render certain molecular configurations untenable, thus marking the high points on Waddington's landscape at which cells are never found; the low points, corresponding to allowed cell states, may be similarly defined. In this way, the landscape of cell fate is shaped by the sum of regulatory interactions that impose global constraints on the molecular configuration of the cell.

However, the real power of the systems approach is in predicting the evolution of cell state in time. A regulatory network that describes a cell state is a dynamic one, in which the individual components (such as gene expression levels) fluctuate in time. The labyrinth of cause-and-effect relationships that exist between components (such as transcriptional activation or repression) places restrictions on these local fluctuations and, by extension, on how the system as a whole may evolve over time (McAdams and Arkin, 1997; Thomas, 1998). This means that, if the state of the system at time t, or $F(t)$, is known, one can predict how it will have changed a short time after, i.e. $F(t + \Delta t)$, and so on – thereby delineating the temporal trajectory of the cell's internal state . If a sufficient understanding of the complete regulatory network, consisting of all the significant players

that contribute to cell state, could be achieved, it would be possible to rationalize the fate decisions that cells make during development and beyond.

The intrinsic inertia of physiological cell types is, by necessity, a time-dependent trait. At local minima in the cell fate landscape, not only are all individual interactions exactly balanced, but displacing the system by a small amount will generate forces that tend to restore it to the original point (Kauffman, 1969; Kauffman, 1993). These points of stability are termed 'attractors' in dynamic systems (Milnor, 1985; Strogatz, 2000), and their self-perpetuating character arises from features of the underlying circuitry, such as cross-antagonism and feed-forward loops (Laiosa *et al*., 2006; Lee *et al*., 2002; Shen-Orr *et al*., 2002; Swiers *et al*., 2006).

In fact, these features are frequently observed in simple circuits involving transcription factors of competing lineages. For example, Oct4 and Cdx2 specify two mutually exclusive populations in the early embryo: one that gives rise to the embryo and the other the placenta, respectively (Nichols *et al*., 1998; Niwa *et al*., 2000; Strumpf *et al*., 2005). Before making this lineage decision, cells of the pre-implantation embryo co-express both factors for a brief period of time. However, the precarious balance between the two factors quickly tips in favor of one when its expression, likely due to stochasticity, becomes dominant over the other. This is because each factor is capable of activating its own transcription in a feed-forward loop, as well as extinguishing the expression of the other (Niwa *et al*., 2005) – thereby effectively establishing and stabilizing its own expression at the expense of the other.

A strong attractor state, corresponding to a stable cell type, would require many such self-enforcing circuits interwoven into a prominent central network. Consistent with

this idea, transcriptional networks constructed for the ES cell state are characterized by a core group of key molecular components, typically transcription factors, that strengthen each other's expression and repress antagonistic factors (Boyer *et al*., 2005; Cole *et al*., 2008; Jiang *et al*., 2008; Kim *et al*., 2008; Liu *et al*., 2008; Loh *et al*., 2006; Sing *et al*., 2008). There is also direct experimental evidence which indicates that stable cell types possess properties of attractor states. Huang and colleagues showed that the differentiation of human HL60 promyelocytic progenitors into neutrophils using chemicals can take two distinct mechanistic routes (Huang *et al*., 2005), suggesting the latter cell type is an intrinsic attractor state discoverable through multiple trajectories. Moreover, the starting progenitor state itself may be an attractor state resistant to small perturbations: following insufficient chemical treatment, the cells that have begun to differentiate return to the original state (Huang *et al*., 2009).

In summary, Waddington's heights consist not only of epigenetic regulators but a host of other molecules whose collective interdependency constrains and guides the fate of a cell. Here is a deterministic picture of the cellular landscape, where the peaks and valleys are essentially encoded in the genome in the form of cellular components and their interactions; development, then, is the natural, stochastic unfolding of allowed cellular trajectories in this landscape that contains multiple, self-stabilizing attractor states.

Hopping Between Cell States

Master Regulators and Transdifferentiation

Having examined the nature of the barriers in Waddington's metaphor, some of the key implications are encouraging from the perspective of nuclear reprogramming. First, the heritable epigenomic stability is maintained dynamically and may lend itself to drastic changes. Second, intrinsically stable attractor states exist in the cell state landscape and can be discovered by multiple independent trajectories. We can further extend the latter claim: if a very strong perturbation forces cells out of their original attractor state and into the vicinity of another, they will gravitate toward the new attractor and eventually assume the identity of the new cell type accurately.

This idea reduces the challenge of cellular conversion significantly. Rather than precisely manipulating each molecular variable in the cell, we may only need to provide a sizable perturbation in the desired direction; the system would then be led toward a nearby self-propagating state by the inherent regulatory network interactions. Specifically, the right type of perturbation must be large enough to overcome the epigenetic barriers and disrupt the existing core gene regulatory network. It must also be directional; that is, it must bring the cell into close proximity of the intended target state.

The prime candidates for fulfilling these criteria are transcription factors, the class of genes that control the expression of other genes. Since gene expression is immediately related to phenotypic output, multiple pathways that control cell state converge at the level of transcription factor activity (Brivanlou and Darnell, 2002; Osborne *et al.*, 2001; Pawson, 1993). Naturally, they are found at the core of gene regulatory networks (Boyer *et al*., 2005; Cole *et al*., 2008; Jiang *et al*., 2008; Kim *et al*., 2008; Liu *et al*., 2008; Loh *et*

al., 2006; Sing *et al*., 2008) and are often necessary and sufficient for specifying lineages *in vivo* (Jessell, 2002; Lobe, 1992; Nichols *et al*., 1998; Niwa *et al*., 2000). At the molecular level, their intrinsic specificity for target DNA sequences allows them, as well as their interaction partners that include epigenetic regulators, to bind select regions throughout the genome and coordinate global patterns of gene expression. These properties makes transcription factors better suited to actuate arbitrary cell fate transitions than are other agents also associated with profound changes in cell state: including epigenetic regulators (that largely lack sequence specificity) and extrinsic cell signaling molecules (that require the presence of appropriate receptors and downstream effectors in the cell).

MyoD *and the Skeletal Muscle Lineage*

The first demonstration that a transcription factor can function outside its normal cellular context to mediate a cell fate conversion came from Harold Weintraub's group. Weintraub and colleague set out to identify the gene responsible for the conversion of mouse C3H10T1/2 embryonic fibroblasts into muscle precursors, or myoblasts (Davis *et al.*, 1987), induced by the DNMT inhibitor 5-azacytidine (Aza). This phenomenon of transdifferentiation, where a differentiated cell is converted to another type, occurred at a high efficiency (25-50%), and they reasoned that a small number of powerful muscle-inducing genes became de-repressed due to demethylation of DNA. To clone these genes, they screened a phage library using radiolabeled cDNA probes prepared from myogenic cell populations, and chose three candidates based on their specific expression in differentiating myoblasts and not in fibroblasts or in terminally differentiated myotubes.

Strikingly, introducing one of the cloned genes, *MyoD*, into C3H10T1/2 fibroblasts phenocopied the muscle-inducing effect of Aza treatment, demonstrating the ability of a single gene to induce a fate change in a stable cell type. *MyoD* is a basic helix-loop-helix transcription factor that has a skeletal muscle-specific expression pattern *in vivo*. Along with another helix-loop-helix factor, *Myf5*, it forms a core network that is required for muscle specification (Buckingham *et al.*, 2003). In this sense, *MyoD* is a prototypical 'master regulator' of the skeletal muscle state during development. However, Weintraub's discovery showed that *MyoD* can exert its full effect beyond its natural developmental setting. In the language of cellular landscapes, it serves the role of a directional perturbation that is strong enough to drive fibroblasts out of their attractor state and onto myoblast-bound trajectories that are inaccessible under physiological conditions.

An interesting comparison can be drawn with another method for inducing a myogenic fate in non-muscle cells, developed by Blau and colleagues, where human amniocytes are fused with differentiated mouse muscle cells (Blau *et al*., 1983). The resulting interspecies, non-dividing heterokaryons express human myosin heavy and light chains, as well as a human muscle-specific antigen, and downregulate fibroblast markers, indicating that the human fibroblast nuclei has adopted a muscle-specific gene expression program.

Unlike *MyoD*, which can induce a myogenic transcriptional program in fibroblasts but not hepatocytes (Choi *et al*., 1990; Weintraub *et al*., 1989), heterokaryon formation can activate muscle-specific gene expression in both cell types (Schafer *et al*., 1990). One possible explanation is that the the action of *MyoD*, such as its binding to

critical target promoters, requires additional factors that are not present in hepatocytes; in hepatocyte-muscle heterokaryons, the complete set of these factors would be derived from the muscle. Consistent with this hypothesis, fusing a *MyoD*-overexpressing hepatocyte (which is not reprogrammed) to an untransduced fibroblast (which does not express *MyoD* but can convert in response to *MyoD*) reprograms both cells, likely ascribed to a complementation effect (Schafer *et al*., 1990).

The differential response of different cell types to *MyoD* may also be related to their epigenetic status. It is conceivable that in certain cell types, genes whose activation is essential for myogenic conversion are stably repressed by one or more epigenetic mechanisms (Goldberg *et al*., 2007). Cells whose fates have diverged early from the myogenic lineage are more likely to have accumulated DNA methylation at muscle-specific promoters and lost activating H3K4me3 marks at key developmental genes that determine the myogenic fate; these loci may even have condensed into heterochromatin, effectively occluding the access of reprogramming factors and blocking fate conversion. Thus, the difficulty of conversion may correlate with the 'distance' between the two cell types in the cell state landscape; reprogramming to a more distant lineage may be less efficient, perhaps requiring a more complex strategy.

The Hematopoietic Lineages

The developmental lineages in the hematopoietic system have been thoroughly characterized over the years, from the long- and short-term hematopoietic stem cells (HSCs) through progenitor cells at various intermediate stages to the mature cell types (Orkin and Zon, 2008). At one of many binary decision points, a common myeloid

progenitor (CMP) commits to become either a megakaryocyte/erythroid progenitor (MEP) or a granulocyte/macrophage progenitor (GMP). Graf and colleagues showed that these closely related competing lineages can interconvert by ectopic expression of transcription factors: GATA-1, a central erythroid transcription factor, causes myeloblasts of the GMP lineage to convert to MEPs (Kulessa *et al*., 1995); conversely, PU.1, a myeloid transcription factor, turns MEPs into myeloblasts (Nerlov and Graf, 1998).

Significantly, GATA-1 and PU.1 not only regulate many downstream target genes, but each activates its own expression and inhibits the other's (Graf 2002; Orkin, 1990). These circuitry features – positive autoregulation and mutual repression – that help make up the core transcriptional network of a stable cell state seem to be the hallmarks of master regulator transcription factors (Arinobu *et al*., 2007; Boyer *et al*., 2005; Niwa *et al*., 2005; Zhang *et al.*, 1999). In the context of transdifferentiation, they allow each of these factors to effectively downregulate the existing transcriptional program, as well as establish the new state it is associated with.

Using yet another factor associated with the granulocyte/macrophage lineage, C/EBPα, the same group converted primary B lymphocytes into macrophages (Xie *et al*., 2004). In particular, terminally differentiated, antibody-producing B cells could give rise to phagocytosing macrophages whose origin was marked by immunoglobulin rearrangements. It is interesting that, while the less mature pro- and pre-B cells convert at 100% efficiency, mature B cells convert at one third the efficiency. This is reminiscent of Gurdon's *Xenopus* experiments in which the success of SCNT declined with the age of

the donor nuclei (Gurdon, 1960), likely due to the build-up of epigenetic mechanisms that are difficult to reverse.

The age-related decline in transdifferentiation efficiency may also be attributed to epigenetic regulation; but as *MyoD*- and heterokaryon-mediated conversion to muscle suggests, the repertoire of proteins expressed in the starting cell type may be another factor. Notably, PU.1 is expressed early in the lymphoid lineage as well as in myeloblasts, and can synergize with C/EBPα to enhance the reprogramming of CD19+ bone marrow cells to macrophages, likely by facilitating the conversion of a more recalcitrant cells in the population population (Xie *et al*., 2004). Therefore, within closely related lineages, less differentiated cells (such as pro- and pre-B cells) may share a greater portion of their protein repertoire with the target cell type, increasing the likelihood of having the components necessary for the action of the reprogramming gene; in contrast, the proteome of more advanced cells (such as mature B cells) may have further diverged and lack those enabling components, thus requiring additional factors for conversion.

In another study that highlighted the necessity of transcription factors in the acquisition and consolidation of cell fate, Busslinger and colleagues found that Pax5-/- pro-B cells lacking the B cell-specific factor could generate a host of hematopoietic cell types, such as macrophages, dendritic cells and granulocytes, given the right culture conditions (Nutt *et al*., 1999). Moreover, when transplanted into immunodeficient mice, Pax5-/- pro-B cells, as well as mature B cells in which Pax5 was deleted, gave rise to T cells (Cobaleda *et al*., 2007; Rolink *et al*., 1999). These results suggest that pro-B cells are unable to consolidate a mature B cell identity in the absence of the master regulator of

the B cell state. In effect, loss of Pax5 from the genome alters the cell state landscape into one where the B cell attractor state is deleted; therefore, a cell traveling down the valley of B cell lineage is eventually forced to choose from alternative fates that are most easily accessible to them.

Insights from Early Transdifferentiation Efforts

The first transdifferentiation studies, which also include attempts to convert between hepatic and pancreatic lineages (Ferber *et al*., 2000; Shen and Slack, 2000; Zaret, 2008), demonstrate these important concepts: the existence of master regulators of cell fate, typically in the form of transcription factors; and the plasticity of cell fate that can be revealed by those master regulators. In rationalizing the experimental observations, the analogy of Waddington's landscape appropriately describes many complex phenomena relating to cellular fate.

Broadly speaking, there seem to be two interrelated factors that contribute to the difficulty of reprogramming. First is the stability of the starting cell type (the depth of the original attractor state), which may have a significant epigenetic contribution. Secondly, the molecular differences between the starting and target cell types (the distance between the two attractor states) can facilitate or hinder the activity of reprogramming agents. Although the usefulness of the systems approach to cell fate remains to be fully explored, its bold prediction is that cellular reprogramming by defined factors is a generalizable phenomenon.

The Stem Cell Revolution

The Quest for the Pluripotent Cell

From a clinical as well as scientific standpoint, a critically desired cell state is one of maximum developmental potency, as in the totipotent zygote or in the pluripotent inner cell mass (ICM) of the pre-implantation embryo, poised to form any cell type of interest. For this reason, embryonic stem cells (ESCs) – the *in vitro* derivatives of the ICM that can be cultured and expanded indefinitely – are an ideal cell type to obtain in the laboratory. Following the derivation of ESCs from mouse blastocysts in 1981 (Evans and Kaufman, 1981; Martin, 1981), the first report of human ESC lines came in 1998, with functional demonstrations of their pluripotency *in vivo* (ability to form teratomas when injected into immunodeficient mice) and *in vitro* (ability to spontaneously generate cells of all three germ layers) (Thomson *et al*., 1998). In principle, human ESCs could be engineered into any cell type of interest to replace damaged or lost tissues, realizing one of the ultimate goals of regenerative medicine.

Since generating patient-specific ESCs is central to this process, there have been efforts to clone blastocysts from patients' somatic cells by SCNT (Cibelli *et al*., 2002; Egli *et al*., 2011a). But deriving an ESC line for every patient as an autologous source of replacement cells is challenging on multiple levels: including the scarcity of donated eggs that can be used as recipients in nuclear transfer (Egli *et al*., 2011b), as well as ethical concerns regarding the destruction of human embryos after the harvest of their ICM (Lanza *et al*., 2000). Moreover, in part due to technical difficulties, successful SCNT in human has yet to be reported, despite some well-guided efforts using oocytes and even *in vitro*-fertilized (IVF) zygotes at mitosis (Egli *et al*., 2011a). Nonetheless, these studies

hint at the existence of elusive yet powerful *trans*-acting factors that mediate this dramatic nuclear reprogramming. The fact that mitotic mouse zygotes, but not those at interphase, could mediate nuclear reprogramming (Egli *et al*., 2007; Egli and Eggan, 2010) suggests that these factors localize to the nucleus after fertilization and are inadvertently removed if enucleation is performed at interphase.

Fusion of pluripotent and differentiated cells has provided additional insight into the mechanism of nuclear reprogramming to pluripotency. Tada, Surani and colleagues demonstrated that fusing adult murine thymocytes to embryonic germ cells (EGCs) or ESCs produces tetraploid hybrids that are pluripotent by one of the most stringent criteria: contribution to chimeric embryos after injection into a blastocyst (Tada *et al*., 1997; Tada *et al*., 2001). The rapid timescale of molecular changes – activation of the *Oct4::*GFP transgene in thymocytes occurs within 48 hours of ESC-thymocyte fusion – suggests that ESCs contain reprogramming agents similar to those in the egg cytoplasm and the fertilized zygote.

Using a similar approach, Eggan and colleagues generated human somatic cell hybrids by fusing human BJ fibroblasts with HUES6, a human ESC line (Cowan *et al*., 2005). The resulting hybrids resembled ESCs in their growth characteristics, marker expression and global gene expression pattern, as well as developmental potential judged by their differentiation *in vitro* and *in vivo*. When an ESC reporter, *Rex1::*GFP, was introduced to BJ fibroblasts before fusion, the activation of this reporter was observed in hybrids, indicating that the differentiated nucleus had indeed undergone reprogramming. Since the ability to form pluripotent hybrids was not restricted to the particular pair of

cells, it is conceivable to reprogram human patient somatic cells in this way, albeit of limited therapeutic use due to the tetraploidy of the hybrids.

Induced Pluripotent Stem Cells (iPSCs)

Reports of factor-mediated transdifferentiation raised the possibility that reprogramming to an ESC state may also be accomplished by a small number of defined factors, in a manner that overcomes the limitations of SCNT and cell fusion approaches. But, though encouraging, these transitions had been between closely related lineages – from fibroblast to muscle (both of mesodermal origin) (Davis *et al*, 1987), or within the hematopoietic system (Kulessa *et al*., 1995; Nerlov and Graf, 1998; Xie *et al*., 2004), or between endodermal lineages (Shen and Slack, 2000) – that share significant developmental history. Reprogramming a somatic cell to a pluripotent state, on the other hand, entails a complete reversal of differentiation. It would be reasonable to suspect that the complex environment provided by the cytoplasm of an egg or an ESC is indispensible for resetting the nucleus to a primitive state, and cannot be reduced to the action of a few defined genes.

A groundbreaking discovery came in 2006, when Takahashi and Yamanaka achieved the improbable conversion of mouse fibroblasts into ESC-like cells by overexpressing just four genes: *Oct4*, *Sox2*, *Klf4*, and *cMyc* (Takahashi and Yamanaka, 2006). They reasoned that factors that maintain the ESC state could also establish the pluripotent state in the context of an already differentiated cell, and retrovirally overexpressed a combination of 24 candidate factors in mouse embryonic fibroblasts (MEFs) that report on the activation of the *Fbx15* promoter – a locus associated with, but

not required for, the ESC state. Upon discovering that rare ESC-like colonies appear in the culture, they narrowed down the reprogramming activity to the four transcription factors.

These first-generation induced pluripotent stem cells (iPSCs) exhibited many molecular and functional characteristics of ESCs, although they did not fulfill what are arguably the most stringent molecular and functional criteria: they showed incomplete demethylation of the promoters of core ESC factors such as *Oct4* and *Nanog*, and were unable to give rise to live chimeric pups (Takahashi and Yamanaka, 2006). Several groups soon showed that the use of essential ESC promoters to drive reporter expression allows the isolation of more completely reprogrammed murine iPSCs that contribute to a wide range of tissues in chimeric animals, including the germ line (Wernig *et al*., 2007; Okita *et al*., 2007). Most importantly, the technology translated to the human system: Yamanaka's and Thomson's groups were the first to independently report the generation of human iPSCs that closely resemble human ESCs (Takahashi *et al*., 2007; Yu *et al*., 2007).

A flood of studies have since improved the relevance iPSCs to regenerative medicine and deepened our understanding of the reprogramming process. Owing to the robustness of the approach, a variety of starting cell types could be used, including terminally differentiated cells (mature B lymphocytes (Hanna *et al*., 2008) and pancreatic β-cells (Stadtfeld *et al*., 2008a)) and those requiring fewer reprogramming factors but having different accessibility in patients (neural progenitor cells (Kim *et al*., 2009b) and keratinocytes (Aasen *et al*., 2008)).

In addition, safety concerns over the use of integrating viruses and the reprogramming genes, which possess varying oncogenic capacity (Hacein-Bey-Abina *et al*., 2003; Okita *et al*, 2007), were addressed in multiple ways. First, it was shown that iPSCs could be generated in the absence of the most potent proto-oncogene, *cMyc*, suspected of causing a high incidence of tumorigenesis in chimeras made with *cMyc*-harboring iPSCs (Nakagawa *et al.*, 2008; Wernig *et al*., 2008). Secondly, alternative methods for delivering the reprogramming factors were developed: these include the use of non-integrating viruses (Ferber *et al*., 2000; Stadtfeld *et al*., 2008b), transient plasmid transfection (Okita *et al*., 2008), protein transduction (Kim *et al*., 2009a; Zhou *et al*., 2009), and transposons (Woltjen *et al*., 2009). The use of small molecules has also been explored, both to replace one or more of the factors and to boost the efficiency of reprogramming (Huangfu *et al*., 2008a, b; Ichida *et al*., 2009). More recently, reprogramming factors delivered in the form of modified RNA were shown to not only generate transgene-free iPSCs, but do so at a high efficiency (Warren *et al*., 2010), raising the hope of large-scale production of transgene-free, patient-specific iPSCs.

Using Patient-Specific iPSCs

It has become clear that define-factor reprogramming to pluripotency is not an artifactual phenomenon, but one that could have a broad impact in the laboratory and in the clinic. Compared with SCNT or ESC fusion, it is the most scalable approach for producing patient-specific pluripotent stem cells that, due to their self-renewing nature, can serve as a limitless cellular source for downstream applications. With this major breakthrough at

hand, the necessary next step is to take advantage of their pluripotential and derive cell types that are clinically relevant.

The process of guiding the differentiation of ESCs or iPSCs, termed directed differentiation, is essentially a simulation of natural development in a dish through successive stages of cell fate specification (Gaspard and Vanderhaeghen, 2010; Murry and Keller, 2008; Peljto and Wichterle, 2011; Schwartz *et al*., 2008). In this case, the perturbation that propels them from one intermediate state to another can be provided by extracellular signaling molecules – the same agents that mediate the corresponding transitions *in vivo* – since the appropriate receptors and effectors should be present in the cells at each stage.

From an understanding of developmental signals, protocols have been developed for turning ESCs into a number of cell types representative of all three embryonic germ layers. For example, BMP4, which directs mesodermal specification in the embryo (Dosch *et al*., 1997) also induces a similar fate in ESCs (Murry and Keller, 2008; Zhang *et al.*, 2008). Further manipulations allow many mesodermal cell types – including blood cell types (Lengerke and Daley, 2010), cardiomyocytes (Dambrot *et al*., 2011), skeletal myoblasts (Barberi *et al*., 2007), and chondrocytes (Oldershaw *et al*., 2010; Toh *et al*., 2010) – to be derived.

Similarly, to produce cells of the endodermal lineage, Activin A is first used to direct ESCs toward definitive endoderm. By subsequently taking these cells through developmental landmarks in a stepwise fashion, it is possible to generate hepatocytes, albeit with a fetal, rather than a mature, character (Basma *et al*., 2009; Hay *et al*., 2008; Touboul *et al*., 2010). Progress has also been made in directing the definitive endoderm-

fated cells through a pancreatic progenitor state (Borowiak and Melton, 2009; Kroon *et al*., 2008), then to a pancreatic endocrine cell fate (Chen *et al*., 2009) – and it is hoped that the final transition to insulin-producing β-cells, the cell type lost in type I diabetes, can be achieved.

On the other hand, ectodermal fate is considered the default differentiation path taken by pluripotent cells in the absence of mesoderm- and endoderm-inducing signals (Levine and Brivanlou, 2007). Accordingly, *in vitro* differentiation into neural cell types is efficiently initiated by blocking both Activin/Nodal/Tgf-β and BMP signaling to induce a neuroectodermal fate (Chambers *et al*., 2009). Dopaminergic neurons, which are lost in Parkinson's disease, can then be produced using Sonic Hedgehog (SHH) and FGF8 to generate neural progenitor cells (NPCs) with a mid-hindbrain character (Perrier *et al*., 2004). Motor neurons, whose loss in disease or injury leads to paralysis, possess a ventral spinal character which can be induced by the same signals, retinoic acid (RA) and SHH (Wichterle *et al*., 2002), that specify these spatial identities along the embryonic rostrocaudal and dorsoventral axes, respectively (Jessell, 2000).

Already, numerous groups have coupled these differentiation strategies with iPSC generation to produce patient-specific, disease-relevant cell types (Boulting *et al*., 2011; Brennand *et al*., 2011; Dimos *et al*., 2008; Ebert *et al*., 2009; Ku *et al*., 2010; Lee *et al*. 2009; Liu *et al*., 2010; Marchetto *et al*., 2010; Maehr *et al*., 2009; Park *et al*., 2008; Rashid *et al*., 2010; Seibler *et al*., 2011; Soldner *et al*., 2009; Zhang *et al*., 2010). Although transplantation is an attractive goal (Ilieva *et al*., 2009; Kroon *et al*., 2008; Lengerke and Daley, 2010), given the issues of safety, the immediate uses of these cells have been predominantly in the laboratory. Specifically, this approach provides

unprecedented opportunities for studying human disease *in vitro*, as well as platforms for testing and discovering drugs in the most relevant cellular context.

Initial attempts at setting up *in vitro* model systems of disease have been encouraging, in particular, for disorders of the nervous system (Han *et al*., 2011). This class of human disorder has been difficult to study, primarily due to the inaccessibility of the diseased tissue from live patients. There is limited insight to be gained from the study of post-mortem materials that do not reveal much in the way of disease initiation and progression, and animal models, which can only be established if the disease has a known genetic cause, have not led to potential therapies that translate well to the human system. Using iPSC-derived neurons offers the advantage of capturing patients' own genetic background in the cells that are born in the dish, so that even idiopathic disease processes might be recapitulated from their early stages.

Some of the first patient-specific iPSC-derived neurons were from fibroblasts of patients with spinal muscular atrophy (SMA) (Ebert *et al*., 2009), an early-onset disease of spinal motor neurons most often caused by the deletion of a ubiquitously expressed protein, Survival of Motor Neuron-1 (SMN1) (Monani, 2005). Interestingly, after six weeks of culture, a motor neuron-specific survival deficit was observed in the SMA-iPS-derived neuronal cultures relative to control (Ebert *et al*., 2009); moreover, tobramycin and valproic acid (Brichta *et al*., 2003; Sumner *et al*., 2003), known to increase the level of full-length SMN transcripts from the *SMN2* locus, could each ameliorate molecular and morphological phenotypes associated with the disease (Ebert *et al*., 2009).

Another neurodevelopmental disorder, Rett Syndrome (RTT), is the result of an X-linked mutation in methyl-CpG binding protein-2 (MeCP2) (Amir, 1999).

Interestingly, RTT-iPSC-derived neurons have a decreased number of glutamatergic synapses and abnormal electrophysiological properties relative to control. Significantly, the ability to produce these cells on demand enabled a screen of possible therapeutics discovered in the mouse model of RTT, and two of them, IGF1 and gentamicin (Tropea *et al*., 2009), were found to increase glutamatergic synapse numbers in the human context (Marchetto *et al*., 2010).

These and other studies, targeting neurological disorders (Brennand *et al*., 2011; Ku *et al*., 2010; Liu *et al*, 2010) as well as diseases of other systems, such as the liver (Rashid *et al*., 2010), provide a proof of concept for the utility of patient iPSC-derived cells as substrates for studying human disease. Of note, not all reported cases displayed important disease phenotypes as expected (Park *et al*, 2008; Seibler *et al*., 2011; Soldner *et al*., 2009; Zhang *et al*., 2010). This may be due to the inherent limitations of the approach, such as the contracted timescale of the experiments and the removal from the complex in vivo environment necessary for the unfolding of the pathology. Nonetheless, given the advantages, the use of iPSC-derived cells is likely to complement the existing methods for elucidating disease mechanisms and identifying therapies. Notably, human iPSC-derived cardiomyocytes can be used as a screening tool for assessing the cardiotoxicity of pre-clinical drugs (Braam *et al*., 2010), and iPSC-derived hepatocytes may similarly be used to evaluate their metabolic toxicity (Sullivan *et al*., 2010). Altogether, these early efforts demonstrate that using iPSCs is a promising approach for regenerative medicine.

A Second Look at Transdifferentiation

Conversion to Clinically Important Cell Types

A bottleneck in applying iPSC technology to a wide range of clinical settings may be the development of directed differentiation protocols for each new cell type of interest. This is no small feat, often requiring extensive knowledge of its development, as well as fine-tuning of the precise level of each signaling molecule (Gaspard and Vanderhaeghen, 2010; Murry and Keller, 2008; Peljto and Wichterle, 2011; Schwartz *et al*., 2008). In fact, many existing protocols suffer from poor efficiency and produce highly heterogeneous populations; for cell therapy applications in particular, this would necessitate not only the enrichment of the precise target cell type, but also a thorough removal of contaminating stem or progenitor cells that may become tumorigenic *in vivo*.

This clinical impetus, fueled by the success of defined-factor reprogramming in achieving what had appeared to be the most challenging conversion, has revived a strong interest in transdifferentiation. Arguably, it may be easier to identify the transcription factors important for a given cell fate, than it is to characterize the diffusive signaling molecules that induce it during development. Moreover, evidence indicates that a mere high-level expression of factors, rather than subtle tuning, may be adequate for causing drastic cell state transitions (Davis *et al*., 1987; Kulessa *et al*., 1995; Nerlov and Graf, 1998). Bypassing the iPSC stage and generating target cells directly from an easily accessible cell type, such as skin fibroblasts, could save a significant amount of time and resources, as well as mitigate concerns of residual proliferative cells in the population.

The series of transdifferentiation studies that shortly followed the discovery of iPSCs were of particularly high clinical relevance. In 2008, Melton and colleagues

reported the *in vivo* conversion of exocrine cells into induced β-cells (iβ-cells), which could ameliorate hypoglycemic phenotypes in a diabetic mouse model (Zhou *et al*., 2008). Then, in 2010, Wernig's laboratory succeeded in reprogramming fibroblasts into induced neuronal (iN) cells, albeit with unclear subtype specificity, capable of forming functional synapses *in vitro* (Vierbuchen *et al*., 2010). That same year, fibroblast-derived cardiomyocytes (iCMs) were also reported by Srivastava and colleagues (Ieda *et al*., 2010). Although demonstration of the *in vivo* functionality of the reprogrammed cells was lacking or limited, there are important parallels that can guide us in approaching the question: can any cell turn into any other?

Finding the Right Cocktail

Earlier efforts in transdifferentiation have tended to highlight the role of a single master regulator of a cell fate, such as the basic helix-loop-helix factor, *MyoD*, that specifies skeletal muscle cells (Davis *et al*., 1987). In contrast, the new lineage reprogramming studies each identified a combination of multiple defined factors as reprogramming agents, perhaps reflecting the increased difficulty of the transitions being attempted. The identity of the factors is informative and offers some logic to the selection of reprogramming genes.

The β-cell-inducing cocktail of *Pdx1*, *Ngn3* and *Mafa* (Zhou *et al*., 2008) collectively can govern all the stages of β-cell specification and maturation. The earliest marker of the pancreatic lineage in the foregut endoderm (Gu *et al*., 2002), *Pdx1* is a homeodomain transcription factor indispensible for pancreas formation (Jonsson *et al*., 1994), and is also required for the function of differentiated β-cells later on (Ahlgren *et*

al., 1998). *Ngn3* is a basic helix-loop-helix factor necessary for the specification of endocrine lineages (Desgraz *et al*., 2009; Gu *et al*., 2002), and may have an additional role in mature β-cell function (Wang *et al*., 2009); its ectopic expression is sufficient to induce an endocrine fate in a number of endodermal settings, including the chick endoderm (Grapin-Botton *et al*., 2001) and adult hepatic progenitor cells (Yechoor *et al*., 2009). *Mafa*, on the other hand, specifically marks β-cells and is important for their terminal differentiation; unlike *Pdx1* or *Ngn3*, it is able to induce insulin expression in early gut epithelium (Grapin-Botton *et al*., 2001) as well as in a pancreatic α-cell line (Zhang *et al*., 2005), and is known to directly bind the insulin promoter and activate transcription (Kataoka *et al*., 2002; Matsuoka *et al*., 2003; Olbrot *et al*., 2002).

Gata4, *Tbx5* and *Mef2c*, which reprogram fibroblasts into cardiomyocytes (Ieda *et al*., 2010), form a core transcriptional network that controls cardiac differentiation (Olson, 2006; Srivastava, 2006). In particular, *Gata4* is considered a 'pioneer factor', able to bind repressed promoters of cardiac-associated genes, thereby allowing other factors to also gain access (Cirillo *et al*., 2002; Smale, 2010). Interestingly, *Gata4* and *Tbx5*, along with a component of chromatin remodeling complexes, *Baf60c*, could cause the induction of cardiomyocytes at ectopic sites during early stages of development (Olson, 2006; Srivastava, 2006), and there is evidence of physical interaction of the gene products *in vitro* (Lickert *et al*., 2004).

The conversion of fibroblasts into iN cells – mediated by *Ascl1*, *Brn2* and *Myt1l* – is particularly impressive (Vierbuchen *et al*,. 2010), in that it involves the crossing of the boundaries between embryonic germ layers (mesoderm to endoderm) that have diverged at the beginning of embryonic development. *Ascl1*, also known as *Mash1*, seems to be a

mater regulator of the neuronal state, regulating the neuronal versus glial fate decision of multipotent neural progenitor cells (Lo *et al*., 1997; Nieto *et al*., 2001), as well as guiding the terminal differentiation of several neuronal types in various contexts (Casarosa *et al*., 1999; Fode *et al*., 2000; Parras *et al*., 2002). *Brn2* is expressed at multiple locations in the central nervous system, and often with *Ascl1* (Castro *et al*., 2006), suggesting their cooperation in neuronal differentiation; interestingly, *Brn1* and *Brn2* double knockout has a strong effect on the glutamatergic upper layer cortical neurons (McEvilly *et al*., 2002; Sugitni *et al*., 2002), and could explain the glutamatergic phenotype of the majority of the iN cells. *Myt1l* is expressed in all central and peripheral neurons as they exit mitosis during terminal differntiation (Cahoy *et al*., 2008; Weiner *et al*., 1997); although direct evidence of its cooperation with *Ascl1* is lacking, its close family member *Myt1* is known to act synergistically with other pro-neural basic helix-loop-helix factors in neural induction in *Xenopus* (Bellefroid *et al*., 1996; Wang *et al*., 2008).

We may thus attempt to rationalize the reprogramming cocktail of genes retrospectively, but the difficulty in identifying them *a priori* is that the outcome of novel physical or functional interactions between the factors in non-physiological settings cannot easily be predicted. In all of the above examples, as well as in iPS reprogramming, the working combination of factors was determined empirically by trial and error, starting from a larger pool of candidate factors and iteratively narrowing down the source of reprogramming activity. It remains to be seen whether an improved ability to simulate factor interactions could help automate this process with algorithms for *in silico* prediction.

Nonetheless, the selection of an initial test set of factors can be guided by some general principles. Prototypical master regulators of the target state – such as *Oct4*, *Pdx1*, *Ascl1* and *Tbx5* for the respective cell types they control – are prime candidates, as are factors that cooperate with them to form a core regulatory network. A target cell type-specific expression pattern and a strong knockout phenotype in the tissue of interest are criteria that can help narrow down the list of factors. Finally, pioneer factors, such as *Gata4*, that can access heterochromatic regions (Cirillo *et al*., 2002; Smale, 2010) may greatly facilitate the desired conversion, as well as those that can interact with chromatin remodeling complexes.

Tracing the Reprogramming Trajectories

The stability and fidelity of the reprogrammed state appear to be common features that mark these transdifferentiation paradigms as well as induced pluripotency. Consistent with the establishment of the endogenous transcriptional network, the reprogrammed cells persist long after the exogenous viral gene expression is extinguished. In addition, molecular hallmarks of the new cell state are accurately displayed: for example, iβ-cells express a wide range of β-cell markers, but not exocrine markers or hormones secreted by other endocrine cell types in the islet, indicating that the β-cell state has been established faithfully (Zhou *et al*., 2008).

However, in contrast to iPS reprogramming (Takahashi and Yamanaka, 2006), the in vivo conversion to β-cells is rapid, with the extra-islet insulin+ cells emerging at day 3 and reaching maximum insulin expression by day 10; it is also efficient, with one fifth of the infected cells converting after one month (Zhou *et al*., 2008). Comparable efficiencies

and speed of conversion are reported for the formation of iCMs and iN cells (Ieda *et al*., 2010; Vierbuchen *et al*., 2010), suggesting that this is a general feature of transdifferentiation. As demonstrated by lineage tracing, the iβ-cells originated from Cpa1-expressing differentiated exocrine cells, which also belong to the pancreatic lineage and share a significant portion of their developmental history with β-cells (Zhou *et al*., 2008). Hence, the trajectories between the two stable states, one might imagine, is much shorter and easier to navigate than in the case of iPSC generation. Similar argument might be made for the reprogramming of fibroblasts to cardiomyocytes – both mesodermal cell types – where, incidentally, the majority of iCMs do not arise from the c-kit+ stem-like cell population (Ieda *et al*., 2010).

Somewhat surprisingly, the fibroblast-to-neuron transition is also marked by an efficiency approaching 20%, although with the caveat that it may be a liberal estimate that uses Tuj1 expression to designate neurons (Vierbuchen *et al*., 2010). It may be that conversion between differentiated cell types is fundamentally easier than reversal to a primitive epigenetic state; even if the two somatic lineages have diverged early in development, they may be related to each other by a comparatively short distance in the cell state landscape. Accordingly, following the transdifferentiation-inducing perturbation, the cells would have a reasonably high chance of finding the target cell state and stably adopting it.

Since shared developmental history may be a facilitating factor in cellular conversion, and since many reprogramming factors are associated with an immature precursor state, it is conceivable that developmental paths are utilized during transdifferentiation. For instance, pancreatic exocrine cells might first be driven to a

progenitor state then differentiate into β-cells. However, neither Sox9 nor Hnf6, markers of the common progenitor between the two lineages, was detected during reprogramming, and the majority of iβ-cells had not gone through a BrdU-labeling proliferative stage characteristic of progenitors (Zhou *et al*., 2008). Similarly, lineage tracing during iCM generation demonstrated that many cells do not go through a *Mesp1*- or *Isl1*-expressing cardiac progenitor states (Ieda *et al*., 2010). Thus, these conversions seem to involve a direct imposition of the new cell fate, rather than dedifferentiation followed by partial recapitulation of development – a model also consistent with the rapid kinetics of reprogramming.

It may reasonably be argued that factor-mediated reprogramming is predominantly driven by transcriptional changes – where the non-physiologically high expression of the factors overrides the existing system by directly activating their downstream targets associated with the new state – and that epigenetic changes follow as a consequence. In this light, it is interesting that the enrichment levels of H3K4me3, H3K27me3 and DNA methylation at cardiac promoters in iCMs approach, but do not reach, those of neonatal cardiac cells (Ieda *et al*., 2010). The precise reason for this is unclear, but may be related to the ability of the reprogramming factors to interact with epigenetic remodeling machinery; certain factors, such as *MyoD*, that can recruit chromatin remodelers to some of its targets may be more potent in this regard (Albini *et al*., 2010; Berkes *et al*., 2004; Gerber *et al*., 1997). The lack of cell division that characterizes these transdifferentiation events may also contribute to incomplete epigenetic reprogramming, since cell division presents an opportunity for erasing old epigenetic marks and adding new ones (Egli *et al*., 2008). It is also tempting to speculate

that if developmental processes had been recapitulated, lineage conversion might have yielded better epigenetically reprogrammed cells.

It is interesting that certain factor combinations are specific to the starting cell type. *Pdx1*, *Ngn3* and *Mafa* were unable to generate iβ-cells from skeletal muscle *in vivo* or fibroblasts *in vitro* (Zhou *et al*., 2008), reinforcing the idea that greater developmental distance increases the difficulty of achieving a particular conversion. More intriguingly, different cocktails induce cardiomyocytes from the embryonic mesoderm (*Gata4*, *Tbx5* and *Baf60c*) and from fibroblasts (*Gata4*, *Tbx5* and *Mef2c*) (Olson, 2006; Srivastava, 2006; Vierbuchen *et al*., 2010). These observations suggest that the precise transcriptional and epigenetic state of the starting cell dictates the optimum trajectory toward the target state and the set of factors that can provide the appropriate perturbation.

Overall, the mechanism of transdifferentiation reveals important insights into the cell state landscape. If normal differentiation and iPS reprogramming can be considered transitions that are more or less 'vertical' but in opposite directions, then it seems that transdifferentiation takes 'horizontal' shortcuts. Put differently, in the context of the entire cell state landscape, the distance between two somatic cell types might be shorter than expected based on their phenotypic differences, resulting in a surprisingly efficient conversion.

Summary

It has been over half a century since the first amphibian cloning experiments enabled the crucial conceptual advance, that genetic information is conserved during development. Unlocking this latent potential in somatic cells has proved to be far from trivial, due to highly conserved epigenetic mechanisms and regulatory networks inherent in our genetic makeup. However, significant progress has been made in converting cellular fate with master regulator transcription factors, culminating in the generation of a number of clinically interesting cell types that include pluripotent stem cells and neuronal cells, starting from unrelated somatic cell types. It now seems plausible that conditions for any arbitrary cell fate transition may be found.

Though only in its infancy, regenerative medicine enabled by nuclear reprogramming technologies holds promises for many human conditions that are currently untreatable. In this regard, we note that there is limited functional data for the reported reprogrammed cell types, especially in the context of a live organism. Continued efforts to produce clinical-grade, patient-specific cells are needed to realize the potential therapeutic value of this approach.

Importantly, a consistent picture of the cell state landscape is beginning to emerge from reprogramming studies and system-wide analyses of cell fate. It will now be interesting to explore additional aspects of this space: for example, can we produce highly specialized cell types, such as a specific neuronal subtype, by transdifferentiation? Are there isolatable intermediate states during reprogramming? And can other agents, such as signaling molecules or chemicals, play important roles in this process? Our future efforts to navigate the cell state landscape may benefit from answers to these questions.

References

Aasen, T., Raya, A., Barrero, M.J., Garreta, E., Consiglio, A., Gonzalez, F., Vassena, R., Bilic, J., Pekarik, V., Tiscornia, G., *et al*. (2008). Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes. Nat Biotechnol *26*, 1276-1284.

Ahlgren, U., Jonsson, J., Jonsson, L., Simu, K., and Edlund, H. (1998). beta-cell-specific inactivation of the mouse Ipf1/Pdx1 gene results in loss of the beta-cell phenotype and maturity onset diabetes. Genes Dev *12*, 1763-1768.

Albini, S., and Puri, P.L. (2010). SWI/SNF complexes, chromatin remodeling and skeletal myogenesis: it's time to exchange! Exp Cell Res 316*,* 3073-3080.

Amir, R.E., Van den Veyver, I.B., Wan, M., Tran, C.Q., Francke, U., and Zoghbi, H.Y. (1999). Rett syndrome is caused by mutations in X-linked MECP2, encoding methyl-CpG-binding protein 2. Nat Genet *23*, 185-188.

Arinobu, Y., Mizuno, S., Chong, Y., Shigematsu, H., Iino, T., Iwasaki, H., Graf, T., Mayfield, R., Chan, S., Kastner, P., *et al*. (2007). Reciprocal activation of GATA-1 and PU.1 marks initial specification of hematopoietic stem cells into myeloerythroid and myelolymphoid lineages. Cell Stem Cell *1*, 416-427.

Armstrong, J.A., Bieker, J.J., and Emerson, B.M. (1998). A SWI/SNF-related chromatin remodeling complex, E-RC1, is required for tissue-specific transcriptional regulation by EKLF in vitro. Cell *95*, 93-104.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes Dev *17*, 126-140.

Baguisi, A., Behboodi, E., Melican, D.T., Pollock, J.S., Destrempes, M.M., Cammuso, C., Williams, J.L., Nims, S.D., Porter, C.A., Midura, P., *et al*. (1999). Production of goats by somatic cell nuclear transfer. Nat Biotechnol *17*, 456-461.

Barberi, T., Bradbury, M., Dincer, Z., Panagiotakos, G., Socci, N.D., and Studer, L. (2007). Derivation of engraftable skeletal myoblasts from human embryonic stem cells. Nat Med *13*, 642-648.

Basma, H., Soto-Gutierrez, A., Yannam, G.R., Liu, L., Ito, R., Yamamoto, T., Ellis, E., Carson, S.D., Sato, S., Chen, Y., *et al*. (2009). Differentiation and transplantation of human embryonic stem cell-derived hepatocytes. Gastroenterology *136*, 990-999.

Bellefroid, E.J., Bourguignon, C., Hollemann, T., Ma, Q., Anderson, D.J., Kintner, C., and Pieler, T. (1996). X-MyT1, a Xenopus C2HC-type zinc finger protein with a regulatory function in neuronal differentiation. Cell *87*, 1191-1202.

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.

Bernstein, E., and Allis, C.D. (2005). RNA meets chromatin. Genes Dev *19*, 1635-1655.

Bhutani, N., Brady, J.J., Damian, M., Sacco, A., Corbel, S.Y., and Blau, H.M. (2010). Reprogramming towards pluripotency requires AID-dependent DNA demethylation. Nature *463*, 1042-1047.

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

Borowiak, M., Maehr, R., Chen, S., Chen, A.E., Tang, W., Fox, J.L., Schreiber, S.L., and Melton, D.A. (2009). Small molecules efficiently direct endodermal differentiation of mouse and human embryonic stem cells. Cell Stem Cell *4*, 348-358.

Boulting, G.L., Kiskinis, E., Croft, G.F., Amoroso, M.W., Oakley, D.H., Wainger, B.J., Williams, D.J., Kahler, D.J., Yamaki, M., Davidow, L., *et al*. (2011). A functionally characterized test set of human induced pluripotent stem cells. Nat Biotechnol *29*, 279-286.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., *et al*. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Braam, S.R., Tertoolen, L., van de Stolpe, A., Meyer, T., Passier, R., and Mummery, C.L. (2010). Prediction of drug-induced cardiotoxicity using human embryonic stem cell-derived cardiomyocytes. Stem Cell Res *4*, 107-116.

Brennand, K.J., Simone, A., Jou, J., Gelboin-Burkhart, C., Tran, N., Sangar, S., Li, Y., Mu, Y., Chen, G., Yu, D., *et al*. (2011). Modelling schizophrenia using human induced pluripotent stem cells. Nature *473*, 221-225.

Brichta, L., Hofmann, Y., Hahnen, E., Siebzehnrubl, F.A., Raschke, H., Blumcke, I., Eyupoglu, I.Y., and Wirth, B. (2003). Valproic acid increases the SMN2 protein level: a well-known drug as a potential therapy for spinal muscular atrophy. Hum Mol Genet *12*, 2481-2489.

Briggs, R., and King, T.J. (1952). Transplantation of Living Nuclei From Blastula Cells into Enucleated Frogs' Eggs. Proc Natl Acad Sci U S A *38*, 455-463.

Brivanlou, A.H., and Darnell, J.E., Jr. (2002). Signal transduction and the control of gene expression. Science *295*, 813-818.

Buckingham, M., Bajard, L., Chang, T., Daubas, P., Hadchouel, J., Meilhac, S., Montarras, D., Rocancourt, D., and Relaix, F. (2003). The formation of skeletal muscle: from somite to limb. J Anat *202*, 59-68.

Cahoy, J.D., Emery, B., Kaushal, A., Foo, L.C., Zamanian, J.L., Christopherson, K.S., Xing, Y., Lubischer, J.L., Krieg, P.A., Krupenko, S.A., *et al*. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. J Neurosci *28*, 264-278.

Casarosa, S., Fode, C., and Guillemot, F. (1999). Mash1 regulates neurogenesis in the ventral telencephalon. Development *126*, 525-534.

Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J.A., Nguyen, L., Gossler, A., Gottgens, B., *et al*. (2006). Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. Dev Cell *11*, 831-844.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol *27*, 275-280.

Chang, B., Chen, Y., Zhao, Y., and Bruick, R.K. (2007). JMJD6 is a histone arginine demethylase. Science *318*, 444-447.

Chen, S., Borowiak, M., Fox, J.L., Maehr, R., Osafune, K., Davidow, L., Lam, K., Peng, L.F., Schreiber, S.L., Rubin, L.L., *et al*. (2009). A small molecule that directs differentiation of human ESCs into the pancreatic lineage. Nat Chem Biol *5*, 258-265.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., *et al*. (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Chesne, P., Adenot, P.G., Viglietta, C., Baratte, M., Boulanger, L., and Renard, J.P. (2002). Cloned rabbits produced by nuclear transfer from adult somatic cells. Nat Biotechnol *20*, 366-369.

Cho, H., Orphanides, G., Sun, X., Yang, X.J., Ogryzko, V., Lees, E., Nakatani, Y., and Reinberg, D. (1998). A human RNA polymerase II complex containing factors that modify chromatin structure. Mol Cell Biol *18*, 5355-5363.

Choi, J., Costa, M.L., Mermelstein, C.S., Chagas, C., Holtzer, S., and Holtzer, H. (1990). MyoD converts primary dermal fibroblasts, chondroblasts, smooth muscle, and retinal pigmented epithelial cells into striated mononucleated myoblasts and multinucleated myotubes. Proc Natl Acad Sci U S A *87*, 7988-7992.

Cibelli, J.B., Lanza, R.P., West, M.D., and Ezzell, C. (2002). The first human cloned embryo. Sci Am *286*, 44-51.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol Cell *9*, 279-289.

Cobaleda, C., Jochum, W., and Busslinger, M. (2007). Conversion of mature B cells into T cells by dedifferentiation to uncommitted progenitors. Nature *449*, 473-477.

Cole, M.F., Johnstone, S.E., Newman, J.J., Kagey, M.H., and Young, R.A. (2008). Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. Genes Dev *22*, 746-755.

Cortellino, S., Xu, J., Sannai, M., Moore, R., Caretti, E., Cigliano, A., Le Coz, M., Devarajan, K., Wessels, A., Soprano, D., *et al*. (2011). Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. Cell *146*, 67-79.

Cote, J., Peterson, C.L., and Workman, J.L. (1998). Perturbation of nucleosome core structure by the SWI/SNF complex persists after its detachment, enhancing subsequent transcription factor binding. Proc Natl Acad Sci U S A *95*, 4947-4952.

Cote, J., Quinn, J., Workman, J.L., and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. Science *265*, 53-60.

Cowan, C.A., Atienza, J., Melton, D.A., and Eggan, K. (2005). Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells. Science *309*, 1369-1373.

Dambrot, C., Passier, R., Atsma, D., and Mummery, C.L. (2011). Cardiomyocyte differentiation of pluripotent stem cells and their use as cardiac disease models. Biochem J *434*, 25-35.

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

Desgraz, R., and Herrera, P.L. (2009). Pancreatic neurogenin 3-expressing cells are unipotent islet precursors. Development *136*, 3567-3574.

Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R., *et al*. (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science *321*, 1218-1221.

Dirscherl, S.S., and Krebs, J.E. (2004). Functional diversity of ISWI complexes. Biochem Cell Biol *82*, 482-489.

Dosch, R., Gawantka, V., Delius, H., Blumenstock, C., and Niehrs, C. (1997). Bmp-4 acts as a morphogen in dorsoventral mesoderm patterning in Xenopus. Development *124*, 2325-2334.

Ebert, A.D., Yu, J., Rose, F.F., Jr., Mattis, V.B., Lorson, C.L., Thomson, J.A., and Svendsen, C.N. (2009). Induced pluripotent stem cells from a spinal muscular atrophy patient. Nature *457*, 277-280.

Egli, D., Birkhoff, G., and Eggan, K. (2008). Mediators of reprogramming: transcription factors and transitions through mitosis. Nat Rev Mol Cell Biol *9*, 505-516.

Egli, D., Chen, A.E., Saphier, G., Ichida, J., Fitzgerald, C., Go, K.J., Acevedo, N., Patel, J., Baetscher, M., Kearns, W.G., *et al*. (2011a). Reprogramming within hours following nuclear transfer into mouse but not human zygotes. Nat Commun *2*, 488.

Egli, D., Chen, A.E., Saphier, G., Powers, D., Alper, M., Katz, K., Berger, B., Goland, R., Leibel, R.L., Melton, D.A., *et al*. (2011b). Impracticality of egg donor recruitment in the absence of compensation. Cell Stem Cell *9*, 293-294.

Egli, D., and Eggan, K. (2010). Recipient cell nuclear factors are required for reprogramming by nuclear transfer. Development *137*, 1953-1963.

Egli, D., Rosains, J., Birkhoff, G., and Eggan, K. (2007). Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. Nature *447*, 679-685.

Enright, B.P., Kubota, C., Yang, X., and Tian, X.C. (2003). Epigenetic characteristics and development of embryos cloned from donor cells treated by trichostatin A or 5-aza-2'-deoxycytidine. Biol Reprod *69*, 896-901.

Enver, T., Pera, M., Peterson, C., and Andrews, P.W. (2009). Stem cell states, fates, and the rules of attraction. Cell Stem Cell *4*, 387-397.

Evans, M.J., and Kaufman, M.H. (1981). Establishment in culture of pluripotential cells from mouse embryos. Nature *292*, 154-156.

Feng, Q., and Zhang, Y. (2001). The MeCP1 complex represses transcription through preferential binding, remodeling, and deacetylating methylated nucleosomes. Genes Dev *15*, 827-832.

Ferber, S., Halkin, A., Cohen, H., Ber, I., Einav, Y., Goldberg, I., Barshack, I., Seijffers, R., Kopolovic, J., Kaiser, N., *et al*. (2000). Pancreatic and duodenal homeobox gene 1 induces expression of insulin genes in liver and ameliorates streptozotocin-induced hyperglycemia. Nat Med *6*, 568-572.

Fode, C., Ma, Q., Casarosa, S., Ang, S.L., Anderson, D.J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. Genes Dev *14*, 67-80.

Fryer, C.J., Nordeen, S.K., and Archer, T.K. (1998). Antiprogestins mediate differential effects on glucocorticoid receptor remodeling of chromatin structure. J Biol Chem *273*, 1175-1183.

Fusaki, N., Ban, H., Nishiyama, A., Saeki, K., and Hasegawa, M. (2009). Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. Proc Jpn Acad Ser B Phys Biol Sci *85*, 348-362.

Gaspar-Maia, A., Alajem, A., Polesso, F., Sridharan, R., Mason, M.J., Heidersbach, A., Ramalho-Santos, J., McManus, M.T., Plath, K., Meshorer, E., *et al*. (2009). Chd1 regulates open chromatin and pluripotency of embryonic stem cells. Nature *460*, 863-868.

Gaspard, N., and Vanderhaeghen, P. (2010). Mechanisms of neural specification from embryonic stem cells. Curr Opin Neurobiol *20*, 37-43.

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.

Goldberg, A.D., Allis, C.D., and Bernstein, E. (2007). Epigenetics: a landscape takes shape. Cell *128*, 635-638.

Graf, T. (2002). Differentiation plasticity of hematopoietic cells. Blood *99*, 3089-3101.

Grapin-Botton, A., Majithia, A.R., and Melton, D.A. (2001). Key events of pancreas formation are triggered in gut endoderm by ectopic expression of pancreatic regulatory genes. Genes Dev *15*, 444-454.

Gu, G., Dubauskaite, J., and Melton, D.A. (2002). Direct evidence for the pancreatic lineage: NGN3+ cells are islet progenitors and are distinct from duct progenitors. Development *129*, 2447-2457.

Guo, J.U., Su, Y., Zhong, C., Ming, G.L., and Song, H. (2011). Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell *145*, 423-434.

Gurdon, J.B. (1960). The developmental capacity of nuclei taken from differentiating endoderm cells of Xenopus laevis. J Embryol Exp Morphol *8*, 505-526.

Gurdon, J.B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. J Embryol Exp Morphol *10*, 622-640.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E., *et al*. (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415-419.

Han, S.S., Williams, L.A., and Eggan, K.C. (2011). Constructing and deconstructing stem cell models of neurological disease. Neuron *70*, 626-644.

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R., *et al*. (2008). Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell *133*, 250-264.

Harray F. (1994). Graph Theory (Westview: Boulder).

Hasty, J., McMillen, D., Isaacs, F., and Collins, J.J. (2001). Computational studies of gene regulatory networks: in numero molecular biology. Nat Rev Genet *2*, 268-279.

Hay, D.C., Fletcher, J., Payne, C., Terrace, J.D., Gallagher, R.C., Snoeys, J., Black, J.R., Wojtacha, D., Samuel, K., Hannoun, Z., *et al*. (2008). Highly efficient differentiation of

hESCs to functional hepatic endoderm requires ActivinA and Wnt3a signaling. Proc Natl Acad Sci U S A *105*, 12301-12306.

He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., *et al*. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science *333*, 1303-1307.

Hinsby, A.M., Olsen, J.V., and Mann, M. (2004). Tyrosine phosphoproteomics of fibroblast growth factor signaling: a role for insulin receptor substrate-4. J Biol Chem *279*, 46438-46447.

Ho, L., Jothi, R., Ronan, J.L., Cui, K., Zhao, K., and Crabtree, G.R. (2009). An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network. Proc Natl Acad Sci U S A *106*, 5187-5191.

Huang, A.C., Hu, L., Kauffman, S.A., Zhang, W., and Shmulevich, I. (2009a). Using cell fate attractors to uncover transcriptional regulation of HL60 neutrophil differentiation. BMC Syst Biol *3*, 20.

Huang, S. (2009b). Reprogramming cell fates: reconciling rarity with robustness. Bioessays *31*, 546-560.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys Rev Lett *94*, 128701.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008a). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol *26*, 795-797.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008b). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol *26*, 1269-1275.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., *et al*. (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell *5*, 491-503.

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.

Ilieva, H., Polymenidou, M., and Cleveland, D.W. (2009). Non-cell autonomous toxicity in neurodegenerative disorders: ALS and beyond. J Cell Biol *187*, 761-772.

Jackson-Grusby, L., Beard, C., Possemato, R., Tudor, M., Fambrough, D., Csankovszki, G., Dausman, J., Lee, P., Wilson, C., Lander, E., *et al*. (2001). Loss of genomic methylation causes p53-dependent apoptosis and epigenetic deregulation. Nat Genet *27*, 31-39.

Jessell, T.M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat Rev Genet *1*, 20-29.

Ji, H., Vokes, S.A., and Wong, W.H. (2006). A comparative analysis of genome-wide chromatin immunoprecipitation data for mammalian transcription factors. Nucleic Acids Res *34*, e146.

Jiang, J., Chan, Y.S., Loh, Y.H., Cai, J., Tong, G.Q., Lim, C.A., Robson, P., Zhong, S., and Ng, H.H. (2008). A core Klf circuitry regulates self-renewal of embryonic stem cells. Nat Cell Biol *10*, 353-360.

Jonsson, J., Carlsson, L., Edlund, T., and Edlund, H. (1994). Insulin-promoter-factor 1 is required for pancreas development in mice. Nature *371*, 606-609.

Kataoka, K., Han, S.I., Shioda, S., Hirai, M., Nishizawa, M., and Handa, H. (2002). MafA is a glucose-regulated and pancreatic beta-cell-specific transcriptional activator for the insulin gene. J Biol Chem *277*, 49903-49910.

Kauffman, S. (1969). Homeostasis and differentiation in random genetic control networks. Nature *224*, 177-178.

Kauffman, S.A. (1993). The origins of order: self-organization and selection in evolution (Oxford: Oxford Univ Press).

Kidder, B.L., Yang, J., and Palmer, S. (2008). Stat3 and c-Myc genome-wide promoter occupancy in embryonic stem cells. PLoS One *3*, e3932.

Kim, D., Kim, C.H., Moon, J.I., Chung, Y.G., Chang, M.Y., Han, B.S., Ko, S., Yang, E., Cha, K.Y., Lanza, R., *et al*. (2009a). Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins. Cell Stem Cell *4*, 472-476.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Kim, J.B., Sebastiano, V., Wu, G., Arauzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D., *et al*. (2009b). Oct4-induced pluripotency in adult neural stem cells. Cell *136*, 411-419.

King, T.J., and Briggs, R. (1955). Changes in the Nuclei of Differentiating Gastrula Cells, as Demonstrated by Nuclear Transplantation. Proc Natl Acad Sci U S A *41*, 321-325.

Kitano, H. (2002). Systems biology: a brief overview. Science *295*, 1662-1664.

Kroon, E., Martinson, L.A., Kadoya, K., Bang, A.G., Kelly, O.G., Eliazer, S., Young, H., Richardson, M., Smart, N.G., Cunningham, J., *et al*. (2008). Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. Nat Biotechnol *26*, 443-452.

Ku, S., Soragni, E., Campau, E., Thomas, E.A., Altun, G., Laurent, L.C., Loring, J.F., Napierala, M., and Gottesfeld, J.M. (2010). Friedreich's ataxia induced pluripotent stem cells model intergenerational GAATTC triplet repeat instability. Cell Stem Cell *7*, 631-637.

Kulessa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. Genes Dev *9*, 1250-1262.

Laiosa, C.V., Stadtfeld, M., and Graf, T. (2006). Determinants of lymphoid-myeloid lineage diversification. Annu Rev Immunol *24*, 705-738.

Lanza, R.P., Caplan, A.L., Silver, L.M., Cibelli, J.B., West, M.D., and Green, R.M. (2000). The ethical validity of using nuclear transfer in human transplantation. JAMA *284*, 3175-3179.

Lee, G., Papapetrou, E.P., Kim, H., Chambers, S.M., Tomishima, M.J., Fasano, C.A., Ganat, Y.M., Menon, J., Shimizu, F., Viale, A., *et al*. (2009). Modelling pathogenesis and treatment of familial dysautonomia using patient-specific iPSCs. Nature *461*, 402-406.

Lee, S.K., Jurata, L.W., Funahashi, J., Ruiz, E.C., and Pfaff, S.L. (2004). Analysis of embryonic motoneuron gene regulation: derepression of general activators function in concert with enhancer factors. Development *131*, 3295-3306.

Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., *et al*. (2002). Transcriptional regulatory networks in Saccharomyces cerevisiae. Science *298*, 799-804.

Lengerke, C., and Daley, G.Q. (2010). Autologous blood cell therapies from pluripotent stem cells. Blood Rev *24*, 27-37.

Lessard, J., Wu, J.I., Ranish, J.A., Wan, M., Winslow, M.M., Staahl, B.T., Wu, H., Aebersold, R., Graef, I.A., and Crabtree, G.R. (2007). An essential switch in subunit composition of a chromatin remodeling complex during neural development. Neuron *55*, 201-215.

Levine, A.J., and Brivanlou, A.H. (2007). Proposal of a model of mammalian neural induction. Dev Biol *308*, 247-256.

Li, E., Bestor, T.H., and Jaenisch, R. (1992). Targeted mutation of the DNA methyltransferase gene results in embryonic lethality. Cell *69*, 915-926.

Lickert, H., Takeuchi, J.K., Von Both, I., Walls, J.R., McAuliffe, F., Adamson, S.L., Henkelman, R.M., Wrana, J.L., Rossant, J., and Bruneau, B.G. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. Nature *432*, 107-112.

Liu, J., Verma, P.J., Evans-Galea, M.V., Delatycki, M.B., Michalska, A., Leung, J., Crombie, D., Sarsero, J.P., Williamson, R., Dottori, M., *et al*. (2010). Generation of

induced pluripotent stem cell lines from Friedreich ataxia patients. Stem Cell Rev *7*, 703-713.

Liu, X., Huang, J., Chen, T., Wang, Y., Xin, S., Li, J., Pei, G., and Kang, J. (2008). Yamanaka factors critically regulate the developmental signaling network in mouse embryonic stem cells. Cell Res *18*, 1177-1189.

Lo, L., Sommer, L., and Anderson, D.J. (1997). MASH1 maintains competence for BMP2-induced neuronal differentiation in post-migratory neural crest cells. Curr Biol *7*, 440-450.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J., *et al*. (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Ma'ayan, A. (2008). Network integration and graph analysis in mammalian molecular systems biology. IET Syst Biol *2*, 206-221.

Macarthur, B.D., Ma'ayan, A., and Lemischka, I.R. (2009). Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Biol *10*, 672-681.

Mackay, J.P., Sunde, M., Lowry, J.A., Crossley, M., and Matthews, J.M. (2007). Protein interactions: is seeing believing? Trends Biochem Sci *32*, 530-531.

Maehr, R., Chen, S., Snitow, M., Ludwig, T., Yagasaki, L., Goland, R., Leibel, R.L., and Melton, D.A. (2009). Generation of pluripotent stem cells from patients with type 1 diabetes. Proc Natl Acad Sci U S A *106*, 15768-15773.

Marchetto, M.C., Carromeu, C., Acab, A., Yu, D., Yeo, G.W., Mu, Y., Chen, G., Gage, F.H., and Muotri, A.R. (2010). A model for neural development and treatment of Rett syndrome using human induced pluripotent stem cells. Cell *143*, 527-539.

Martin, G.R. (1981). Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. Proc Natl Acad Sci U S A *78*, 7634-7638.

Matsuoka, T.A., Zhao, L., Artner, I., Jarrett, H.W., Friedman, D., Means, A., and Stein, R. (2003). Members of the large Maf transcription family regulate insulin gene transcription in islet beta cells. Mol Cell Biol *23*, 6049-6062.

Mayer, W., Niveleau, A., Walter, J., Fundele, R., and Haaf, T. (2000). Demethylation of the zygotic paternal genome. Nature *403*, 501-502.

McAdams, H.H., and Arkin, A. (1997). Stochastic mechanisms in gene expression. Proc Natl Acad Sci U S A *94*, 814-819.

McEvilly, R.J., de Diaz, M.O., Schonemann, M.D., Hooshmand, F., and Rosenfeld, M.G. (2002). Transcriptional regulation of cortical neuron migration by POU domain factors. Science *295*, 1528-1532.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al*. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al*. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Mikkola, I., Heavey, B., Horcher, M., and Busslinger, M. (2002). Reversion of B cell commitment upon loss of Pax5 expression. Science *297*, 110-113.

Milnor, J. (1985). On the concept of attractor. Commun Math Phys *99*, 177-195.

Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.

Mogilner, A., Wollman, R., and Marshall, W.F. (2006). Quantitative modeling in cell biology: what is it good for? Dev Cell *11*, 279-287.

Monani, U.R. (2005). Spinal muscular atrophy: a deficiency in a ubiquitous protein; a motor neuron-specific disease. Neuron *48*, 885-896.

Muller, F.J., Laurent, L.C., Kostka, D., Ulitsky, I., Williams, R., Lu, C., Park, I.H., Rao, M.S., Shamir, R., Schwartz, P.H., *et al*. (2008). Regulatory networks define phenotypic classes of human stem cell lines. Nature *455*, 401-405.

Murry, C.E., and Keller, G. (2008). Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development. Cell *132*, 661-680.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

Nerlov, C., and Graf, T. (1998). PU.1 induces myeloid lineage commitment in multipotent hematopoietic progenitors. Genes Dev *12*, 2403-2412.

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

Nieto, M., Schuurmans, C., Britz, O., and Guillemot, F. (2001). Neural bHLH genes control the neuronal versus glial fate decision in cortical progenitors. Neuron *29*, 401-413.

Niwa, H., Miyazaki, J., and Smith, A.G. (2000). Quantitative expression of Oct-3/4 defines differentiation, dedifferentiation or self-renewal of ES cells. Nat Genet *24*, 372-376.

Niwa, H., Toyooka, Y., Shimosato, D., Strumpf, D., Takahashi, K., Yagi, R., and Rossant, J. (2005). Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. Cell *123*, 917-929.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556-562.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247-257.

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

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. Science *322*, 949-953.

Olbrot, M., Rud, J., Moss, L.G., and Sharma, A. (2002). Identification of beta-cell-specific insulin gene transcription factor RIPE3b1 as mammalian MafA. Proc Natl Acad Sci U S A *99*, 6737-6742.

Oldershaw, R.A., and Hardingham, T.E. (2010). Notch signaling during chondrogenesis of human bone marrow stem cells. Bone *46*, 286-293.

Olson, E.N. (2006). Gene regulatory networks in the evolution and development of the heart. Science *313*, 1922-1927.

Orkin, S.H. (1990). Globin gene regulation and switching: circa 1990. Cell *63*, 665-672.

Orkin, S.H., and Zon, L.I. (2008). Hematopoiesis: an evolving paradigm for stem cell biology. Cell *132*, 631-644.

Osborne, C.K., Schiff, R., Fuqua, S.A., and Shou, J. (2001). Estrogen receptor: current understanding of its activation and modulation. Clin Cancer Res *7*, 4338s-4342s; discussion 4411s-4412s.

Oswald, J., Engemann, S., Lane, N., Mayer, W., Olek, A., Fundele, R., Dean, W., Reik, W., and Walter, J. (2000). Active demethylation of the paternal genome in the mouse zygote. Curr Biol *10*, 475-478.

Park, I.H., Arora, N., Huo, H., Maherali, N., Ahfeldt, T., Shimamura, A., Lensch, M.W., Cowan, C., Hochedlinger, K., and Daley, G.Q. (2008). Disease-specific induced pluripotent stem cells. Cell *134*, 877-886.

Parras, C.M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D.J., and Guillemot, F. (2002). Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. Genes Dev *16*, 324-338.

Pawson, T. (1993). Signal transduction--a conserved pathway from the membrane to the nucleus. Dev Genet *14*, 333-338.

Peljto, M., and Wichterle, H. (2011). Programming embryonic stem cells to neuronal subtypes. Curr Opin Neurobiol *21*, 43-51.

Perrier, A.L., Tabar, V., Barberi, T., Rubio, M.E., Bruses, J., Topf, N., Harrison, N.L., and Studer, L. (2004). Derivation of midbrain dopamine neurons from human embryonic stem cells. Proc Natl Acad Sci U S A *101*, 12543-12548.

Polejaeva, I.A., Chen, S.H., Vaught, T.D., Page, R.L., Mullins, J., Ball, S., Dai, Y., Boone, J., Walker, S., Ayares, D.L., *et al*. (2000). Cloned pigs produced by nuclear transfer from adult somatic cells. Nature *407*, 86-90.

Prochasson, P., Neely, K.E., Hassan, A.H., Li, B., and Workman, J.L. (2003). Targeting activity is required for SWI/SNF function in vivo and is accomplished through two partially redundant activator-interaction domains. Mol Cell *12*, 983-990.

Quinn, J., Fyrberg, A.M., Ganster, R.W., Schmidt, M.C., and Peterson, C.L. (1996). DNA-binding properties of the yeast SWI/SNF complex. Nature *379*, 844-847.

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.

Rolink, A.G., Nutt, S.L., Melchers, F., and Busslinger, M. (1999). Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. Nature *401*, 603-606.

Rybouchkin, A., Kato, Y., and Tsunoda, Y. (2006). Role of histone acetylation in reprogramming of somatic nuclei following nuclear transfer. Biol Reprod *74*, 1083-1089.

Santos, F., Hendrich, B., Reik, W., and Dean, W. (2002). Dynamic reprogramming of DNA methylation in the early mouse embryo. Dev Biol *241*, 172-182.

Schwartz, P.H., Brick, D.J., Stover, A.E., Loring, J.F., and Muller, F.J. (2008). Differentiation of neural lineage cells from human pluripotent stem cells. Methods *45*, 142-158.

Seibler, P., Graziotto, J., Jeong, H., Simunovic, F., Klein, C., and Krainc, D. (2011). Mitochondrial Parkin recruitment is impaired in neurons derived from mutant PINK1 induced pluripotent stem cells. J Neurosci *31*, 5970-5976.

Shen, C.N., Slack, J.M., and Tosh, D. (2000). Molecular basis of transdifferentiation of pancreas to liver. Nat Cell Biol *2*, 879-887.

Shen-Orr, S.S., Milo, R., Mangan, S., and Alon, U. (2002). Network motifs in the transcriptional regulation network of Escherichia coli. Nat Genet *31*, 64-68.

Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstine, J.R., Cole, P.A., and Casero, R.A. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. Cell *119*, 941-953.

Shin, T., Kraemer, D., Pryor, J., Liu, L., Rugila, J., Howe, L., Buck, S., Murphy, K., Lyons, L., and Westhusin, M. (2002). A cat cloned by nuclear transplantation. Nature *415*, 859.

Singh, S.K., Kagalwala, M.N., Parker-Thornburg, J., Adams, H., and Majumder, S. (2008). REST maintains self-renewal and pluripotency of embryonic stem cells. Nature *453*, 223-227.

Smale, S.T. (2010). Pioneer factors in embryonic stem cells and differentiation. Curr Opin Genet Dev *20*, 519-526.

Smidt, M.P., and Burbach, J.P. (2007). How to make a mesodiencephalic dopaminergic neuron. Nat Rev Neurosci *8*, 21-32.

Soldner, F., Hockemeyer, D., Beard, C., Gao, Q., Bell, G.W., Cook, E.G., Hargus, G., Blak, A., Cooper, O., Mitalipova, M., *et al*. (2009). Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell *136*, 964-977.

Spooncer, E., Brouard, N., Nilsson, S.K., Williams, B., Liu, M.C., Unwin, R.D., Blinco, D., Jaworska, E., Simmons, P.J., and Whetton, A.D. (2008). Developmental fate determination and marker discovery in hematopoietic stem cell biology using proteomic fingerprinting. Mol Cell Proteomics *7*, 573-581.

Srivastava, D. (2006). Making or breaking the heart: from lineage determination to morphogenesis. Cell *126*, 1037-1048.

Stadtfeld, M., Brennand, K., and Hochedlinger, K. (2008a). Reprogramming of pancreatic beta cells into induced pluripotent stem cells. Curr Biol *18*, 890-894.

Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. (2008b). Induced pluripotent stem cells generated without viral integration. Science *322*, 945-949.

Strogatz, S.H. (2000). Nonlinear dynamics and chaos: with applications to physics, biology, chemistry and engineering (Westview: Boulder).

Strumpf, D., Mao, C.A., Yamanaka, Y., Ralston, A., Chawengsaksophak, K., Beck, F., and Rossant, J. (2005). Cdx2 is required for correct cell fate specification and differentiation of trophectoderm in the mouse blastocyst. Development *132*, 2093-2102.

Sugitani, Y., Nakai, S., Minowa, O., Nishi, M., Jishage, K., Kawano, H., Mori, K., Ogawa, M., and Noda, T. (2002). Brn-1 and Brn-2 share crucial roles in the production and positioning of mouse neocortical neurons. Genes Dev *16*, 1760-1765.

Sullivan, E.K., Weirich, C.S., Guyon, J.R., Sif, S., and Kingston, R.E. (2001). Transcriptional activation domains of human heat shock factor 1 recruit human SWI/SNF. Mol Cell Biol *21*, 5826-5837.

Sumner, C.J., Huynh, T.N., Markowitz, J.A., Perhac, J.S., Hill, B., Coovert, D.D., Schussler, K., Chen, X., Jarecki, J., Burghes, A.H., *et al*. (2003). Valproic acid increases SMN levels in spinal muscular atrophy patient cells. Ann Neurol *54*, 647-654.

Suzuki, M.M., and Bird, A. (2008). DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet *9*, 465-476.

Swiers, G., Patient, R., and Loose, M. (2006). Genetic regulatory networks programming hematopoietic stem cells and erythroid lineage specification. Dev Biol *294*, 525-540.

Tada, M., Tada, T., Lefebvre, L., Barton, S.C., and Surani, M.A. (1997). Embryonic germ cells induce epigenetic reprogramming of somatic nucleus in hybrid cells. EMBO J *16*, 6510-6520.

Tada, M., Takahama, Y., Abe, K., Nakatsuji, N., and Tada, T. (2001). Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells. Curr Biol *11*, 1553-1558.

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.

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.

Thomas, R. (1998). Laws for the dynamics of regulatory networks. Int J Dev Biol *42*, 479-485.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Toh, W.S., Lee, E.H., Richards, M., and Cao, T. (2010). In vitro derivation of chondrogenic cells from human embryonic stem cells. Methods Mol Biol *584*, 317-331.

Touboul, T., Hannan, N.R., 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.

Tropea, D., Giacometti, E., Wilson, N.R., Beard, C., McCurry, C., Fu, D.D., Flannery, R., Jaenisch, R., and Sur, M. (2009). Partial reversal of Rett Syndrome-like symptoms in MeCP2 mutant mice. Proc Natl Acad Sci U S A *106*, 2029-2034.

Tyson, J.J., Chen, K., and Novak, B. (2001). Network dynamics and cell physiology. Nat Rev Mol Cell Biol *2*, 908-916.

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

Waddington, C.H. (1942). Endeavour *1*, 18-20.

Waddington, C.H. (1957). The strategy of the genes: a discussion of some aspects of theoretical biology (London: Allen & Unwin).

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, S., Hecksher-Sorensen, J., Xu, Y., Zhao, A., Dor, Y., Rosenberg, L., Serup, P., and Gu, G. (2008). Myt1 and Ngn3 form a feed-forward expression loop to promote endocrine islet cell differentiation. Dev Biol *317*, 531-540.

Wang, S., Jensen, J.N., Seymour, P.A., Hsu, W., Dor, Y., Sander, M., Magnuson, M.A., Serup, P., and Gu, G. (2009). Sustained Neurog3 expression in hormone-expressing islet cells is required for endocrine maturation and function. Proc Natl Acad Sci U S A *106*, 9715-9720.

Wang, Y., Wysocka, J., Perlin, J.R., Leonelli, L., Allis, C.D., and Coonrod, S.A. (2004). Linking covalent histone modifications to epigenetics: the rigidity and plasticity of the marks. Cold Spring Harb Symp Quant Biol *69*, 161-169.

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al*. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618-630.

Weiner, J.A., and Chun, J. (1997). Png-1, a nervous system-specific zinc finger gene, identifies regions containing postmitotic neurons during mammalian embryonic development. J Comp Neurol *381*, 130-142.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. Proc Natl Acad Sci U S A *86*, 5434-5438.

Wernig, M., Meissner, A., Cassady, J.P., and Jaenisch, R. (2008). c-Myc is dispensable for direct reprogramming of mouse fibroblasts. Cell Stem Cell *2*, 10-12.

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.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Wilkinson, D.J. (2009). Stochastic modelling for quantitative description of heterogeneous biological systems. Nat Rev Genet *10*, 122-133.

Wilmut, I., Schnieke, A.E., McWhir, J., Kind, A.J., and Campbell, K.H. (1997). Viable offspring derived from fetal and adult mammalian cells. Nature *385*, 810-813.

Wilson, C.J., Chao, D.M., Imbalzano, A.N., Schnitzler, G.R., Kingston, R.E., and Young, R.A. (1996). RNA polymerase II holoenzyme contains SWI/SNF regulators involved in chromatin remodeling. Cell *84*, 235-244.

Woltjen, K., Michael, I.P., Mohseni, P., Desai, R., Mileikovsky, M., Hamalainen, R., Cowling, R., Wang, W., Liu, P., Gertsenstein, M., *et al*. (2009). piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells. Nature *458*, 766-770.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise reprogramming of B cells into macrophages. Cell *117*, 663-676.

Yan, Z., Wang, Z., Sharova, L., Sharov, A.A., Ling, C., Piao, Y., Aiba, K., Matoba, R., Wang, W., and Ko, M.S. (2008). BAF250B-associated SWI/SNF chromatin-remodeling complex is required to maintain undifferentiated mouse embryonic stem cells. Stem Cells *26*, 1155-1165.

Yechoor, V., Liu, V., Espiritu, C., Paul, A., Oka, K., Kojima, H., and Chan, L. (2009). Neurogenin3 is sufficient for transdetermination of hepatic progenitor cells into neo-islets in vivo but not transdifferentiation of hepatocytes. Dev Cell *16*, 358-373.

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.

Zaret, K.S. (2008). Genetic programming of liver and pancreas progenitors: lessons for stem-cell differentiation. Nat Rev Genet *9*, 329-340.

Zhang, C., Moriguchi, T., Kajihara, M., Esaki, R., Harada, A., Shimohata, H., Oishi, H., Hamada, M., Morito, N., Hasegawa, K., *et al*. (2005). MafA is a key regulator of glucose-stimulated insulin secretion. Mol Cell Biol *25*, 4969-4976.

Zhang, N., An, M.C., Montoro, D., and Ellerby, L.M. (2010). Characterization of Human Huntington's Disease Cell Model from Induced Pluripotent Stem Cells. PLoS Curr *2*, RRN1193.

Zhang, P., Behre, G., Pan, J., Iwama, A., Wara-Aswapati, N., Radomska, H.S., Auron, P.E., Tenen, D.G., and Sun, Z. (1999). Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. Proc Natl Acad Sci U S A *96*, 8705-8710.

Zhang, P., Li, J., Tan, Z., Wang, C., Liu, T., Chen, L., Yong, J., Jiang, W., Sun, X., Du, L., *et al*. (2008). Short-term BMP-4 treatment initiates mesoderm induction in human embryonic stem cells. Blood *111*, 1933-1941.

Zhang, Y., LeRoy, G., Seelig, H.P., Lane, W.S., and Reinberg, D. (1998). The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities. Cell *95*, 279-289.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., *et al*. (2009). Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell *4*, 381-384.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627-632.

Chapter 2

A Small Molecule Inhibitor of Tgf-β Signaling Replaces *Sox2* in Reprogramming by Inducing *Nanog*

Addendum

This chapter is adapted from the original article:
Justin K. Ichida*, Joel Blanchard*, Kelvin Lam*, Esther Y. Son*, Julia E. Chung, Dieter Egli, Kyle M. Loh, Ava C. Carter, Francesco P. Di Giorgio, Kathryn Koszka, Danwei Huangfu, Hidenori Akutsu, David R. Liu, Lee L. Rubin, and Kevin Eggan (2009). A Small-Molecule Inhibitor of Tgf-β Signaling Replaces *Sox2* in Reprogramming by Inducing *Nanog*. Cell Stem Cell 5: 491-503.

JKI, JB, KL, LLR and KE conceived the initial chemical screen. JKI, JB and KL performed the screen, discovered the hit molecules and optimized their concentrations. JKI, JB, KL, EYS, LLR and KE designed the subsequent experiments, and JKI, JB and EYS contributed equally to the functional and mechanistic characterization of RepSox. Specifically, JKI and JB generated the data in Figures 2.1 and 2.4; EYS helped JKI and JB with staining and functional pluripotency assays of RepSox-derived iPSCs; EYS and JKI, with the help of JEC, derived partially reprogrammed cell lines and treated them with chemicals; EYS performed cell cycle analysis on the partially reprogrammed cells and, with the help of JKI, performed mRNA microarrays, analyzed the expression data and identified *Nanog* as a downstream target of RepSox; JB performed RT-PCRs and Westerns using partially reprogrammed cell lines; JKI and EYS performed *Nanog* knock-down and overexpression experiments, as well as derivation of 'MONK' iPS lines and their characterization. DE, KML, ACC, FPDG and KK provided technical assistance and DH, HA and DRL provided helpful discussions.

Abstract

The combined activity of three transcription factors can reprogram adult cells into induced pluripotent stem (iPS) cells. However, the transgenic methods used to deliver reprogramming factors have raised concerns regarding the future utility of the resulting stem cells. These uncertainties could be overcome if each transgenic factor were replaced with a small molecule that either directly activated its expression from the somatic genome or in some way compensated for its activity. To this end, we have used high-content chemical screening to identify small molecules that can replace *Sox2* in reprogramming. We show that one of these molecules functions in reprogramming by inhibiting Tgf-β signaling in a stable and trapped intermediate cell type that forms during the process. We find that this inhibition promotes the completion of reprogramming through induction of the transcription factor *Nanog*.

Introduction

Retroviral transduction with three genes: *Sox2*, *Oct4*, and *Klf4*, can directly reprogram somatic cells to a pluripotent stem cell state (Okita et al., 2007; Takahashi et al., 2007b). Unfortunately, the resulting induced pluripotent stem (iPS) cells are suboptimal for applications in transplantation medicine and disease modeling because both the viral vectors used for gene transfer and the reprogramming factors they encode are oncogenic (Hacein-Bey-Abina et al., 2003; Nakagawa et al., 2008; Thrasher, 2007).

One potential solution is to identify small molecules that can efficiently reprogram cells, producing unmodified iPS cell lines better suited for downstream applications. Identification of such compounds would allow reprogramming that would not be impeded by the laborious nature of protein transduction or the safety concerns surrounding transgenic approaches (Kaji et al., 2009; Kim, 2009; Okita et al., 2008).

Several small molecules that catalyze reprogramming have already been described. Compounds that alter chromatin structure, including the DNA methyltransferase inhibitor 5-aza-cytidine (Aza) and the histone deacetylase (HDAC) inhibitor valproic acid (VPA), can increase reprogramming efficiency and even reduce the number of factors required for reprogramming (Huangfu et al., 2008a; Huangfu et al., 2008b; Mikkelsen et al., 2008; Shi et al., 2008b) . Treatment with these inhibitors presumably lowers the barrier to activation of endogenous pluripotency-associated genes. However, *Oct4* and *Sox2* not only activate genes required for pluripotency, they also function to repress genes promoting differentiation. It is therefore unlikely that this class of small molecules would be sufficient to completely replace the transgenic factors. As a

result, there remains a need to identify novel small molecules that can function in reprogramming.

Here, we report the discovery of compounds that can replace the central reprogramming factor *Sox2*. We demonstrate that one of these chemicals specifically acts by inhibiting Tgf-β signaling. Interestingly, this compound does not act by inducing *Sox2* expression in the target fibroblasts. Instead, we show that it enables reprogramming through the induction of *Nanog* transcription in a stable, partially reprogrammed cell type that accumulates in the absence of *Sox2*.

Results

A Screen for Chemical Mediators of Reprogramming

To identify small molecules that function in reprogramming, we transduced fibroblasts with viral vectors encoding *Oct4, Klf4,* and *cMyc* and then screened for compounds that allowed for reprogramming in the absence of *Sox2*. We favored this approach because it was unbiased with respect to the mechanism by which a given chemical could function and would not only deliver chemical compounds with translational utility but also provide novel insights into the mechanisms controlling reprogramming.

Activation of an *Oct4*::GFP reporter gene in colonies with an ES cell morphology has been shown to be a stringent assay for reprogramming (Meissner et al., 2007). In mouse embryonic stem (mES) cell culture medium supplemented with VPA, retroviral transduction of 7,500 *Oct4*::GFP transgenic mouse embryonic fibroblasts (MEFs) with *Oct4*, *Klf4*, *cMyc*, and *Sox2* (Boiani et al., 2004) routinely generated 100-200 GFP+ colonies (Figure 2.**1A**). In contrast, we observed no GFP+ colonies when *Sox2* was

omitted (Figure 2.1A). We used this robust difference to identify small molecules that can replace *Sox2*.

To facilitate the identification of cellular targets and signaling pathways affected by any compounds we discovered, we utilized a library of molecules with known pharmacological targets. We transduced *Oct4*::GFP MEFs with *Oct4, Klf4,* and *cMyc,* and then plated 2000 cells per well in 96-well format. To each well we added one of 200 distinct compounds for 7-11 days, while also treating with 2 mM VPA for the first 7 days (Figure 2.1B). It was our hope that this approach would allow us to identify both compounds that required chromatin remodeling to induce reprogramming (Huangfu et al., 2008a) and compounds that did not. After 16 days, we scored each well for the presence of GFP+ colonies with a mES-like morphology (Figure 2.1C) and identified 3 independent hit compounds (Table 2.1). Two of these compounds were distinct Transforming Growth Factor-β Receptor 1 (Tgfbr1) kinase inhibitors (E-616452 and E-616451 (Figure 2.1D-E) (Gellibert et al., 2004)), while the third was a Src-family kinase inhibitor (EI-275 (Figure 2.1F) (Hanke et al., 1996)).

Efficient Small Molecule Replacement of Sox2

Next, we optimized the effective concentration for each hit molecule (Figure 2.1G-I) and quantified the efficiency at which it synergized with VPA to replace *Sox2*. When 1500 MEFs were transduced with only *Oct4*, *Klf4*, and *cMyc* and then treated with VPA, we did not observe GFP+ colonies (Figure 2.1J). However, the addition of E-616452 (25 μM), E-616451 (3 μM), or EI-275 (3 μM), led to the formation of GFP+ colonies with an

Figure 2.1

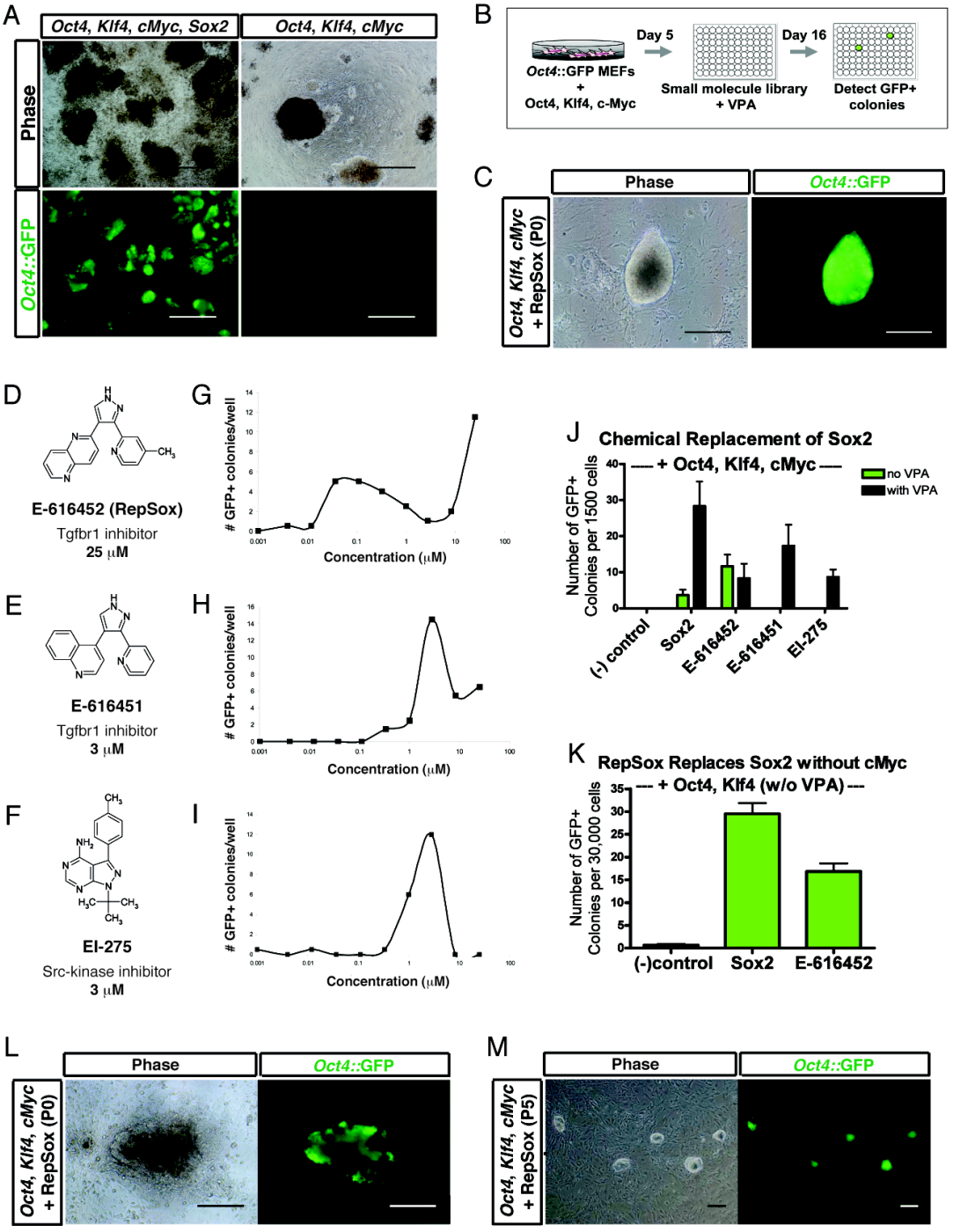
A
Oct4, Klf4, cMyc, Sox2
Oct4, Klf4, cMyc
Phase
Oct4::GFP
B
Day 5
Day 16
Oct4::GFP MEFs
+
Oct4, Klf4, c-Myc
Small molecule library
+ VPA
Detect GFP+
colonies
C
Phase
Oct4::GFP
Oct4, Klf4, cMyc
+ RepSox (P0)
D
E-616452 (RepSox)
Tgfbr1 inhibitor
25 μM
E
E-616451
Tgfbr1 inhibitor
3 μM
F
EI-275
Src-kinase inhibitor
3 μM
G
GFP+ colonies/well
Concentration (μM)
H
GFP+ colonies/well
Concentration (μM)
I
GFP+ colonies/well
Concentration (μM)
J
Chemical Replacement of Sox2
+ Oct4, Klf4, cMyc
no VPA
with VPA
Number of GFP+
Colonies per 1500 cells
(-) control
Sox2
E-616452
E-616451
EI-275
K
RepSox Replaces Sox2 without cMyc
+ Oct4, Klf4 (w/o VPA)
Number of GFP+
Colonies per 30,000 cells
(-)control
Sox2
E-616452
L
Phase
Oct4::GFP
Oct4, Klf4, cMyc
+ RepSox (P0)
M
Phase
Oct4::GFP
Oct4, Klf4, cMyc
+ RepSox (P5)

Figure 2.1. (Continued)
Identification of Small Molecules That Replace *Sox2.* (A) *Oct4*::GFP+ colonies form readily in *Oct4, Klf4, cMyc,* and *Sox2*-infected MEF cultures and do not form in *Oct4, Klf4,* and *cMyc*-infected MEF cultures. Scale bars represent 500 μm. (B) Overview of chemical screen for replacement of *Sox2.* (C) A P0 colony from *Oct4, Klf4,* and *cMyc*-infected MEFs + RepSox that displays a mES-like morphology and is *Oct4*::GFP+. Scale bars represent 200 μm. (D-F) Chemical structures and optimal reprogramming concentrations for each hit molecule. (D) E-616452 (RepSox). (E) E-616451. (F) EI-275 (F). (G-I) *Oct4*::GFP+ colony formation in *Oct4, Klf4, cMyc*-infected MEFs as a function of compound concentration. 2 mM VPA was used in all wells. (G) RepSox. (H) E-616451. (I) EI-275. (J) Quantification of small molecule replacement of *Sox2* in *Oct4, Klf4,* and *cMyc*-infected MEFs with and without VPA treatment. Colonies were counted at 30 days post-infection. (K) *Sox2* replacement by RepSox is not dependent on *cMyc* (no VPA treatment). (L-M) RepSox can replace *Sox2* in defined factor reprogramming of adult tail tip fibroblasts. (L) *Oct4*::GFP+ P0 colony derived from *Oct4, Klf4, cMyc*-infected tail tip fibroblasts treated with RepSox for 14 days. Scale bars represent 200 μm. (M) Passage 5 *Oct4*::GFP+ cell line derived from a P0 colony. Scale bars represent 500 μm.

	E-616452 (RepSox)	E-616451	EI-275
7-day treatment	11	8	0
11-day treatment	9	10	3

Table 2.1.
Preliminary Validation of Hit Molecules. Number of *Oct4::*GFP colonies detected for each chemical in the primary screen after transduction of *Oct4*, *Klf4*, and *cMyc* and VPA treatment.

ES cell morphology at a rate that was comparable to transduction with *Sox2* (Figure 2.1F).

Since the three compounds were identified in the presence of VPA, we next determined whether these molecules were dependent on this HDAC inhibitor for their reprogramming activities. We found that E-616451 and EI-275 could not induce the appearance of GFP+ colonies in the absence of VPA (Figure 2.1J), while E-616452 could do so and at a rate that was similar to a positive control transduced with the *Sox2* retrovirus (Figure 2.1J).

Although *cMyc* does increase the efficiency of reprogramming, it is not required for the generation of iPS cells (Nakagawa et al., 2008). Since the elimination of *cMyc* is an important step towards reducing the risk of tumor formation, we tested whether E-616452 could function in the absence of this oncogene. When added to MEFs transduced with only *Oct4* and *Klf4*, E-616452 induced the formation of GFP+ colonies with an efficiency similar to viral *Sox2* (Figure 2.1G).

Previous reports on small molecules that affect reprogramming have focused on MEFs or neural stem cells (NSCs). These cells may be reprogrammed more easily due to either their proliferative capacity or their expression of iPS factors (Huangfu et al., 2008a; Shi et al., 2008a; Shi et al., 2008b). However, it may be that chemical modulation of gene expression is cell-type specific and we therefore determined if the reprogramming compound we identified functioned in a more patient-relevant cell type. When we infected adult tail tip fibroblasts with *Oct4, Klf4,* and *cMyc* alone, we did not observe *Oct4*::GFP+ colonies. However, when we added E-616452, we readily observed reprogramming (Figure 2.1L). The resulting *Oct4*::GFP+ colonies could be expanded

into cell lines that maintained homogeneous *Oct4*::GFP expression and self-renewed similarly to mES and 4-factor control iPS lines (Figure 2.1M). Because it could efficiently replace transgenic *Sox2* in the absence of VPA and *cMyc,* as well as in both embryonic and adult fibroblasts, we chose to further characterize E-616452 and named it RepSox, for Replacement of *Sox2*.

RepSox-Reprogrammed Cells Are iPS Cells

Investigation of self-renewal capacity (Figure 2.2A), gene expression program, and pluripotency demonstrated that *Oct4::*GFP+ cells induced by the RepSox replacement of *Sox2* were *bona fide* iPS cells. PCR with primers specific to the *Oct4, Klf4, cMyc,* and *Sox2* transgenes confirmed that this cell line did not harbor transgenic *Sox2* (Figure 2.2B). Chromosomal analysis indicated it was karyotypically normal (Figure 2.2C).

The *Oct4*::GFP+ cells co-expressed alkaline phosphatase (Figure 2.2D) and the endogenous alleles of the *Nanog* and *Sox2* genes, suggesting pluripotency had been established (Figure 2.2E). The global transcriptional profile of cells reprogrammed with RepSox was similar to that of an iPS cell line produced with all four transgenes and as similar to those of mES cells (Pearson correlation coefficient = 0.95-0.97) as two distinct mES cell lines profiles were to each other (Pearson correlation coefficient = 0.96) (Figures 2.2F-H and Table 2). The profile differed significantly from that of the somatic MEFs (Figure 2.2F).

Cells produced with RepSox could readily form both embryoid bodies and teratomas that contained differentiated cell types of the three distinct embryonic germ layers (Figure 2.3B). In addition, we observed that these cells could respond to directed

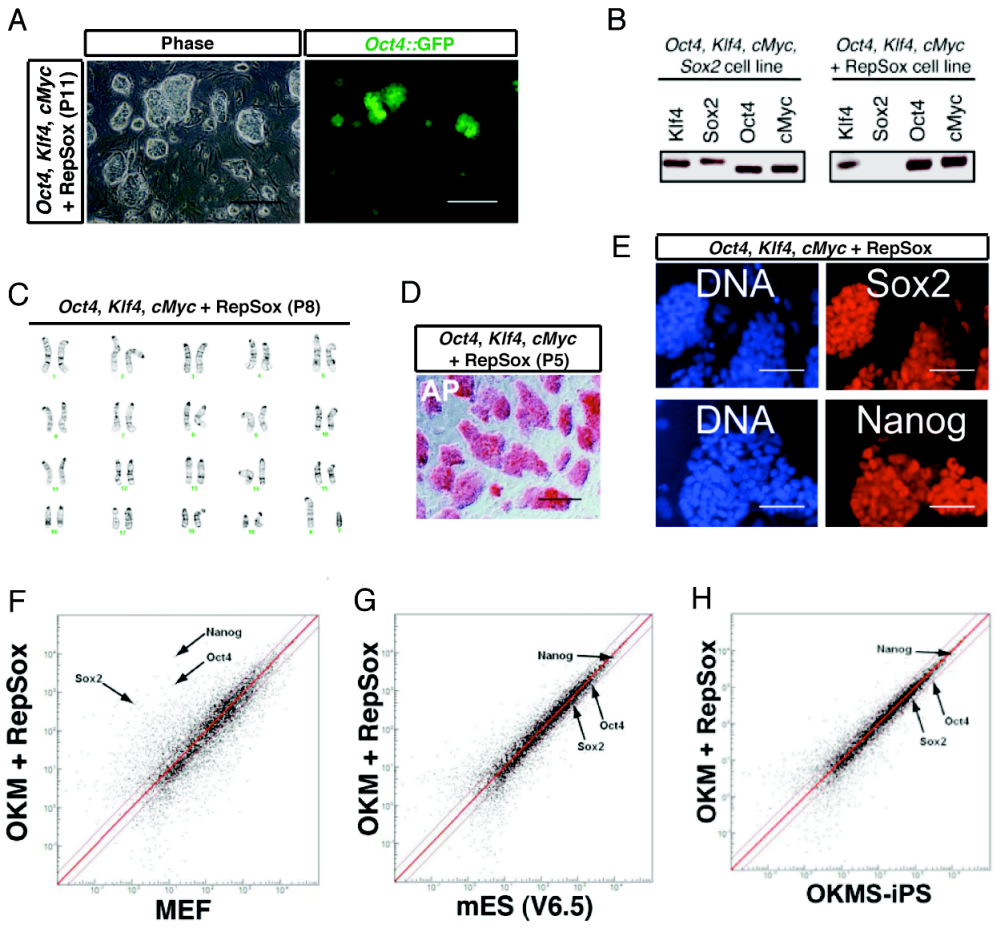

Figure 2.2.
Self-Renewal and Gene Expression of RepSox-Reprogrammed Cells. (A) An *Oct4*::GFP+ iPS line that was derived from a culture of RepSox-treated *Oct4, Klf4,* and *cMyc*-infected MEFs (OKM + RepSox Line 1) displays the characteristic mES-like morphology and self-renewal properties. Passage 11. Scale bars represent 500 µm. (B) PCR for viral transgenes using genomic DNA isolated from a control iPS cell line generated with *Oct4, Klf4, cMyc*, and *Sox2* and a RepSox-reprogrammed cell line generated with *Oct4, Klf4,* and *cMyc* + RepSox. (C) Normal karyotype of a passage 8 cell from *Oct4, Klf4,* and *cMyc* + RepSox line 1. 20 cells were counted and 5 cells were karyotyped by GTL banding. ll cells were karyotypically normal 40, XY. (D) *Oct4, Klf4,* and *cMyc* + RepSox Line 1 at Passage 5. Red color indicates alkaline phosphatase activity. Scale bar represents 500 µm. (E) Antibody staining of OKM + RepSox line 1 cells shows that they express markers of pluripotent stem cells Sox2 and Nanog. Scale bars represent 100 µm. (F-H) Microarray scatter plots showing correlation of global gene expression profiles. (F) OKM + RepSox Line 1 is transcriptionally very different from somatic MEFs, and (G) highly similar to the mES line V6.5, as well as (H) to an iPS line generated with *Oct4, Klf4, cMyc,* and *Sox2* (OKMS-iPS).

	mES1 (R1)	mES2 (V6.5)	OKMS-iPS	OKM + RepSox	MEF
mES1 (R1)	1.00	0.96	0.98	0.96	0.80
mES2 (V6.5)		1.00	0.99	0.97	0.81
OKMS-iPS			1.00	0.97	0.82
OKM+RepSox				1.00	0.79
MEF					1.00

Table 2.2.
Pearson Correlation Coefficients of Global Gene Expression. The following cell types were compared pairwise: two mES cell lines (R1 and V6.5); *Oct4, Klf4, cMyc,* and *Sox2* iPS line 1 (OKMS-iPS); *Oct4, Klf4,* and *cMyc* + RepSox iPS line 1 (OKM + RepSox); and *Oct4*::GFP MEFs (MEF).

differentiation signals *in vitro* and robustly differentiate into Hb9+/Tuj1+ motor neurons (Figure 2.3A).

In order to more definitively confirm the pluripotency of cells reprogrammed with RepSox, we tested their ability to contribute to chimeric embryos *in vivo*. We labeled cells with a lentiviral transgene encoding the red fluorescent Tomato-protein and injected them into blastocysts. Both embryos and adult mice with significant contribution from the iPS cells were obtained (Figures 2.3C-D). Although adult mice with high contribution from the iPS cells were observed, we found it difficult to assess the contribution of these cells to the germ-line, as the majority of animals developed tumors at or before the time of sexual maturity. However, we did observe that the reprogrammed cells could contribute *Oct4::*GFP+ cells to the genital ridges of embryonic chimeras, demonstrating contribution of these pluripotent cells to the germ-line (Figure 2.3E). Together, these results demonstrate that the RepSox-reprogrammed cells are indeed iPS cells.

RepSox Can Replace Sox2 *and* c-Myc *by Inhibiting Tgf-β Signaling*

Previous studies with RepSox suggest that it can act as an inhibitor of the Tgfbr1 kinase (Gellibert et al., 2004). Therefore, we investigated whether the mechanism by which RepSox functions to replace *Sox2* is through the inhibition of Tgf-β signaling. If Tgfbr1 is the functional target of RepSox, then a structurally unrelated inhibitor of Tgf-β signaling or depletion of Tgf-β ligands from the culture medium might also replace *Sox2*. The small molecule SB431542 (Figure 2.4A) is known to inhibit Tgfbr1 kinase and is structurally distinct from RepSox (Inman et al., 2002). When we treated fibroblasts

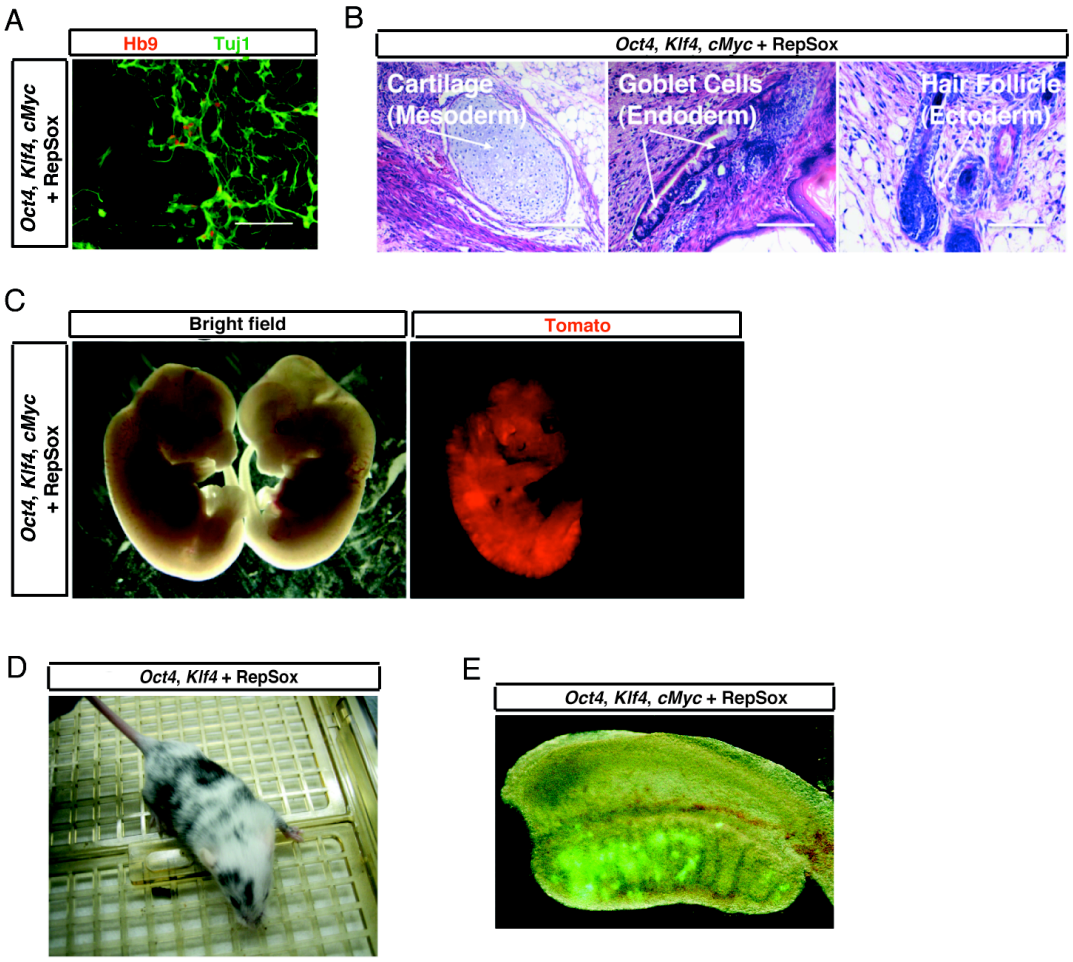

Figure 2.3.
RepSox-Reprogrammed Cells Are Pluripotent. (A) Motor neurons differentiated *in vitro* from OKM + RepSox line 1. Scale bar represents 200 µm. (B) Teratomas containing cells of allthe three germ layers formed by injection of OKM + RepSox line 1 cells into nude mice. (C) E12.5 chimeric mouse embryo (left, vs. non-chimeric littermate on the right) showing a high amount of contribution from OKM + RepSox line 1 cells constitutively expressing the dTomato red fluorescent protein. (D) 8 week-old chimeric mouse formed by injection of OK + RepSox line 1 cells (C57BL6 genetic background) into an ICR blastocyst. (E) *Oct4*::GFP+ cells derived from an OKM + RepSox cell line are present in the genital ridge of a male embryo at 13.5 d.p.c.

transduced with *Oct4*, *Klf4*, and *cMyc* with 25 μM SB431542, we observed ~10 GFP+ colonies per 7500 cells plated (Figure 2.4B). Likewise, when we transduced fibroblasts in the presence of either an antibody that neutralized a variety of Tgf-β ligands (R&D Systems, AB-100-NA) or an antibody specific to Tgf-β II (R&D Systems, AB-12-NA), *Oct4*::GFP+ colonies were generated (Figure 2.4B). In contrast, we observed no GFP+ colonies in transductions without these Tgf-β inhibitors. These results are consistent with the notion that at least part of the mechanism by which RepSox replaces *Sox2* in reprogramming is through the inhibition of Tgf-β signaling.

Our goal was to identify molecules that specifically replace *Sox2* instead of generally increasing reprogramming efficiency. If RepSox acts specifically to replace *Sox2*, then we would not expect it to stimulate reprogramming in the presence of transgenic *Sox2*. When RepSox- or Tgf-β antibody-treated MEFs were transduced with *Oct4*, *Klf4*, *cMyc* and *Sox2*, we observed less than a 2-fold increase in the number of GFP+ colonies over the untreated controls (Figures 2.4C-D). The magnitude by which RepSox stimulated reprogramming in this context was significantly less than the 10-fold increase that we observed following treatment with VPA, a compound thought to increase reprogramming efficiency (Figure 2.1J).

In order to further investigate the specificity of *Sox2* replacement by RepSox, we tested the ability of this molecule to individually replace *Oct4, Klf4,* and *cMyc* in reprogramming. We found that RepSox could not induce GFP+ colonies in the absence of either *Oct4* or *Klf4*, even in the presence of VPA (Figure 2.4E). In contrast, we found that RepSox did increase the number of *Oct4*::GFP+ colonies by 20-fold in the absence of *cMyc*, thereby fully replacing it in reprogramming (Figure 2.4F). In addition, the

Figure 2.4.

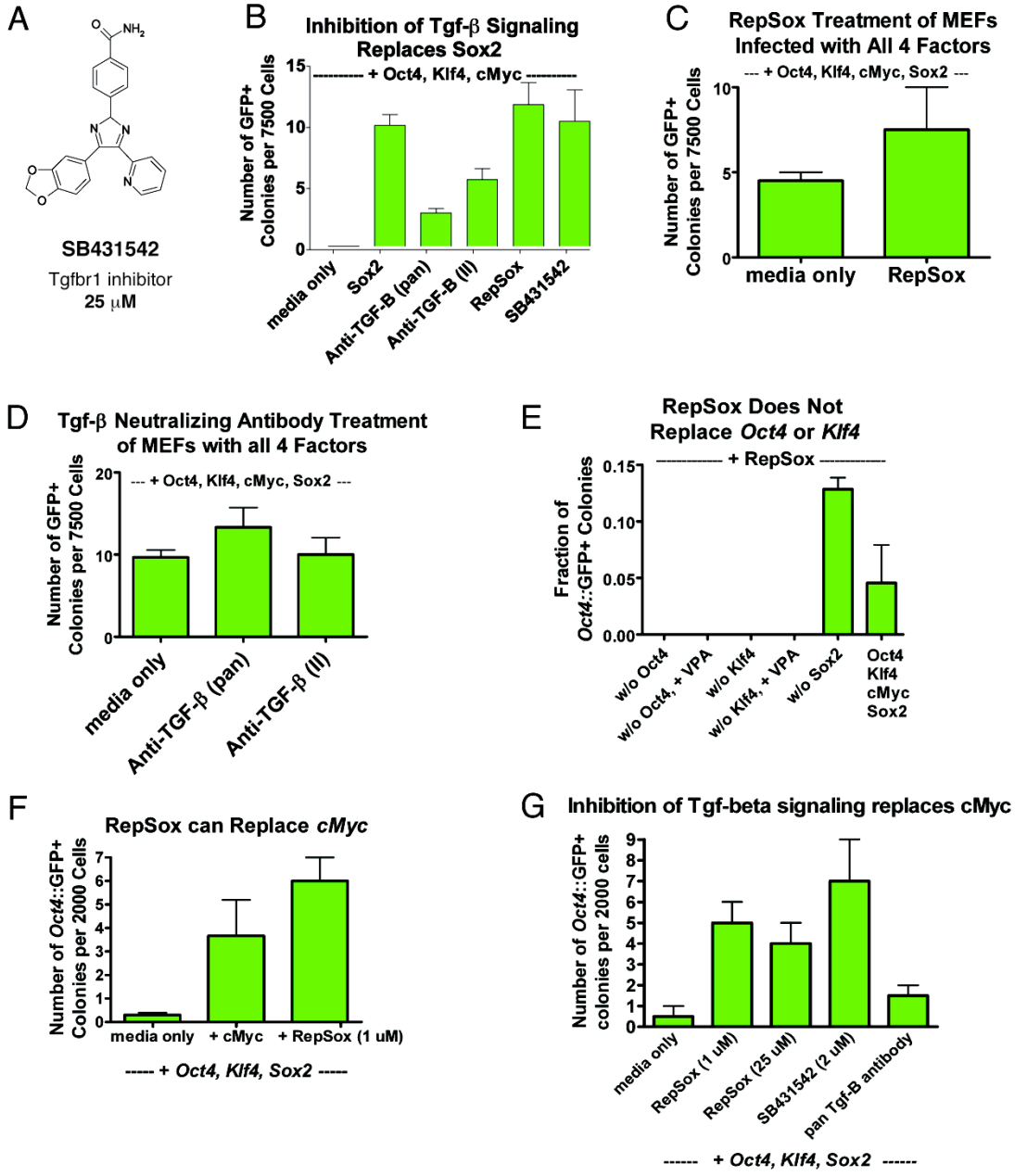
A
SB431542
Tgfbr1 inhibitor
25 μM
B
Inhibition of Tgf-β Signaling Replaces Sox2
--------- + Oct4, Klf4, cMyc ---------
Number of GFP+ Colonies per 7500 Cells
media only
Sox2
Anti-TGF-B (pan)
Anti-TGF-B (II)
RepSox
SB431542
C
RepSox Treatment of MEFs Infected with All 4 Factors
--- + Oct4, Klf4, cMyc, Sox2 ---
Number of GFP+ Colonies per 7500 Cells
media only
RepSox
D
Tgf-β Neutralizing Antibody Treatment of MEFs with all 4 Factors
--- + Oct4, Klf4, cMyc, Sox2 ---
Number of GFP+ Colonies per 7500 Cells
media only
Anti-TGF-β (pan)
Anti-TGF-β (II)
E
RepSox Does Not Replace Oct4 or Klf4
---------- + RepSox ----------
Fraction of Oct4::GFP+ Colonies
w/o Oct4
w/o Oct4, + VPA
w/o Klf4
w/o Klf4, + VPA
w/o Sox2
Oct4 Klf4 cMyc Sox2
F
RepSox can Replace cMyc
Number of Oct4::GFP+ Colonies per 2000 Cells
media only
+ cMyc
+ RepSox (1 uM)
----- + Oct4, Klf4, Sox2 -----
G
Inhibition of Tgf-beta signaling replaces cMyc
Number of Oct4::GFP+ colonies per 2000 cells
media only
RepSox (1 uM)
RepSox (25 uM)
SB431542 (2 uM)
pan Tgf-B antibody
------ + Oct4, Klf4, Sox2 ------

Figure 2.4. (Continued)
RepSox Specifically Replaces *Sox2* by Inhibiting Tgf-β Signaling. (A) Chemical structure of SB431542, an inhibitor of Tgfbr1 activity. , with the optimal concentration for *Sox2* replacement listed. (B) Inhibition of Tgf-β signaling by treatment of *Oct4, cMyc,* and *Sox2*-infected MEFs with SB431542 or Tgf-β neutralizing antibodies replaces *Sox2.* Colonies were counted at 30 days post-infection. (C) RepSox does not increase the efficiency of *Oct4*::GFP+ colony induction in *Oct4*, *Klf4, cMyc*, and *Sox2*-infected MEFs. Shown are the numbers of colonies per 7500 infected cells plated. Colonies were counted at 30 days post-infection. (D) Inhibition of Tgf-β signaling by Tgf-β neutralizing antibodies does not increase the efficiency of *Oct4*::GFP+ colony induction in *Oct4*, *Klf4, cMyc*, and *Sox2*-infected MEFs. Shown are the numbers of colonies per 7500 infected cells plated. Colonies were counted at 30 days post-infection. (E) RepSox does not replace transgenic *Oct4* or transgenic *Klf4* in reprogramming. We observed no *Oct4*::GFP+ colonies in RepSox-treated *Klf4, cMyc, Sox2*-infected MEFs or *Oct4*, *cMyc, Sox2*-infected MEFs out of 30,000 cells plated both with and without VPA treatment. We routinely observe 30-40 *Oct4*::GFP+ colonies when we plate the same number of *Oct4, Klf4, cMyc*-infected MEFs and treat with RepSox. Colonies were counted at 30 days post-infection.(F) RepSox can replace *cMyc* in reprogramming. Cells were transduced with *Oct4, Klf4,* and *cMyc* and treated with RepSox continuously starting at day 5 post-infection. Colonies were counted at 30 days post-infection. (G) Inhibition of Tgf-β signaling can replace *cMyc* in reprogramming. Cells were transduced with *Oct4, Klf4,* and *cMyc* and treated with inhibitors of Tgf-β signaling continuously starting at day 5 post-infection. Colonies were counted at 30 days post-infection.

structurally distinct Tgf-β inhibitor SB431542 and a Tgf-β-specific neutralizing antibody both increased reprogramming efficiency in the absence of *cMyc* (Figure 2.4G). From these experiments, we conclude that RepSox enables the replacement of the reprogramming activities provided by both transgenic *Sox2* and *cMyc*. In both cases, these complementing activities seem to be mediated through the inhibition of Tgf-β signaling.

RepSox Replace Sox2 *by Acting on Intermediates Formed during the Reprogramming Process*

The development of cocktails of small molecules that can effectively reprogram somatic cells may require a detailed knowledge of the mechanism and kinetics by which each compound acts. Therefore, we determined the optimal duration of time by which inhibition of Tgf-β signaling using RepSox can help induce reprogramming.

Initially, we pretreated MEFs with RepSox, applying the chemical for three days, and then removed it at the time of transduction with *Oct4, Klf4,* and *cMyc*. In these experiments, no *Oct4*::GFP+ colonies were formed (Figure 2.5A), suggesting that RepSox does not act on the initial somatic cells to replace *Sox2*. Consistent with this result, we did not detect a significant increase in the expression of endogenous *Sox2* or closely related *Sox* family members upon RepSox treatment (Figure 2.5B). In addition, RepSox treatment did not decrease the expression of the mesenchymal gene *Snai1* (Figure 2.5C), which is downregulated 5-40-fold by transduction of the 4 reprogramming factors (Mikkelsen et al., 2008). Thus, RepSox does not destabilize the pre-existing MEF transcriptional program.

Figure 2.5.

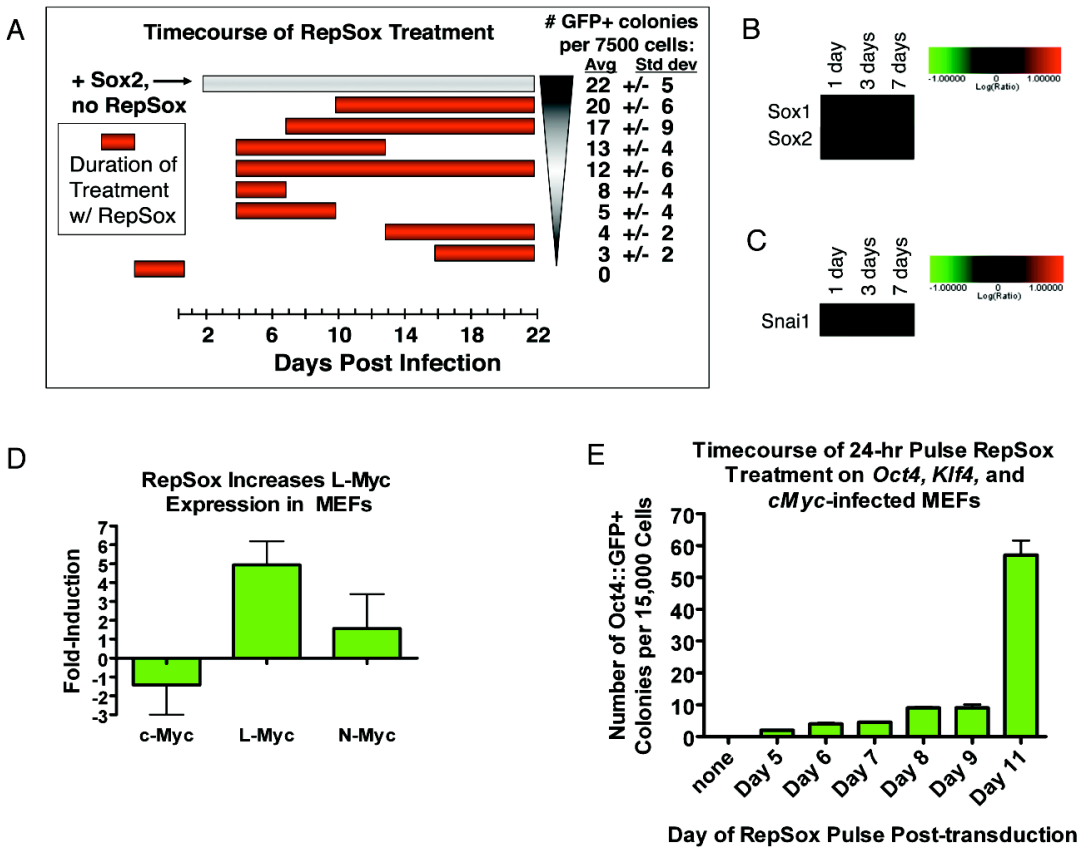
A
Timecourse of RepSox Treatment
GFP+ colonies per 7500 cells:
Avg Std dev
+ Sox2, ⟶ no RepSox
Duration of Treatment w/ RepSox
22 +/- 5
20 +/- 6
17 +/- 9
13 +/- 4
12 +/- 6
8 +/- 4
5 +/- 4
4 +/- 2
3 +/- 2
0
2 6 10 14 18 22
Days Post Infection
B
1 day
3 days
7 days
Sox1
Sox2
-1.00000 0 1.00000
Log(Ratio)
C
1 day
3 days
7 days
Snai1
-1.00000 0 1.00000
Log(Ratio)
D
RepSox Increases L-Myc Expression in MEFs
Fold-Induction
7 6 5 4 3 2 1 0 -1 -2 -3
c-Myc
L-Myc
N-Myc
E
Timecourse of 24-hr Pulse RepSox Treatment on *Oct4*, *Klf4*, and *cMyc*-infected MEFs
Number of Oct4::GFP+ Colonies per 15,000 Cells
70 60 50 40 30 20 10 0
none
Day 5
Day 6
Day 7
Day 8
Day 9
Day 11
Day of RepSox Pulse Post-transduction
F
Reprogrammed Colonies Appear at Day 14 Whether RepSox Treatment is Started at Day 7 or Day 10
Number of *Oct4*::GFP+ Colonies per 7500 Cells Plated
15 10 5 0
10 11 12 13 14 15 16
Day Post-transduction
Treatment started at day 7
Treatment started at day 10

Figure 2.5. (Continued)
A Short Pulse of RepSox Is Sufficient for *Sox2* Replacement and Most Effective at Later Time Points. (A) Time course of RepSox treatmentGraph showing the number of *Oct4*::GFP+ colonies induced by various timings of RepSox treatment of *Oct4, cMyc,* and *Sox2*-infected MEFs in mES medium. Colonies were counted at 24 days post-infection. (B-D) RepSox treatment in *Oct4, Klf4,* and *cMyc*-transduced MEFs does not induce the expression of *Sox*-family members or decrease the expression of fibroblast-specific genes, but it does increases *L-Myc* mRNA expression in MEFs. (B) *Sox*-family gene expression. Note that *Sox3* expression did not change significantly. Shown are changes relative to untreated controls. (C) Fibroblast-specific gene expression. (D) *L-Myc* expression analysis. Untransduced MEFs were treated with 25 μM RepSox for 7 days and mRNA expression was determined by microarray analysis. Fold-induction is relative to untreated control samples. (E) Timecourse of RepSox treatment showing the number of *Oct4*::GFP+ colonies induced by a 24-hr pulse of RepSox on *Oct4, cMyc,* and *Sox2*-infected MEFs in serum-free mES medium with knockout serum replacement (KSR mES). Colonies were counted at 24 days post-infection. Shown are average colony numbers +/- the standard deviation. (F) *Oct4*::GFP+ colonies appear at day 14 regardless of whether RepSox treatment is initiated at day 7 or day 10 post-transduction.

In contrast, we found that RepSox did increase by 5-fold the expression of *L-Myc*, a close homolog of *cMyc* that can functionally replace it in reprogramming (Nakagawa et al., 2008) (Figure 2.5D). Together these data suggest that although RepSox likely functions at the level of the initial somatic cell population to replace *cMyc*, it does not act on the starting MEF population to replace *Sox2*.

Because RepSox did not seem to act directly on the fibroblasts to replace *Sox2*, we investigated whether or not it functioned on intermediates that arose during reprogramming. To address this question, we varied both the duration and timing of RepSox treatment in order to determine when it was most effective. First, we transduced 7,500 MEFs with *Oct4, Klf4,* and *cMyc*, waited for 4 days, and then treated cultures with RepSox for either 3, 6, 9, or 18 additional days. Although a short 3-day treatment from days 4-7 induced a small number of *Oct4*::GFP+ colonies, the 9-day treatment from days 4-13 yielded the most *Oct4*::GFP+ colonies (Figure 2.5A).

Next, we varied the timing at which we initiated RepSox treatment, administering the compound beginning at day 4, 7, 10, 13, or 16 after transduction. We found that delaying the start of RepSox treatment increased its reprogramming potency, with optimal treatment beginning at 10 days post-transduction (Figure 2.5A). Together these results suggest that RepSox treatment is most effective between days 7-12 post-transduction.

To more precisely define the optimal treatment window, we determined the minimal duration of treatment required to induce reprogramming. We found that a treatment as short as only one day was sufficient to induce detectable reprogramming (Figure 2.5E). Delaying this short treatment yielded more reprogrammed colonies, with a

sharp increase at day 11 (Figure 2.5E). These results indicate that RepSox is most effective at replacing *Sox2* during days 10-11 after transduction and that therefore cultures of *Oct4*, *Klf4,* and *cMyc*-transduced MEFs give rise to intermediates capable of responding to RepSox treatment. These intermediates appear at day 4 post-transduction and peak at days 10-11.

Interestingly, when we tracked the timing of the initial appearance of reprogrammed colonies as a function of the timing of RepSox administration, we found that regardless of whether we began treatment at day 7 or day 10 post-transduction, *Oct4*::GFP+ colonies first appeared at day 14 (Figure 2.5F). This suggests that RepSox may not always be the rate-limiting step in this reprogramming process and that other, RepSox-independent events take place during the formation of the RepSox-responsive intermediates.

RepSox-Responsive Cell Lines

Our finding that a 24-hr pulse of RepSox can replace *Sox2* (Figure 2.5E) differs strikingly from the 5-10 day period of transgene expression normally required (Sridharan et al., 2009; Wernig et al., 2007) and suggests that RepSox could trigger a switch activating reprogramming. If RepSox acts to flip a switch in semi-stable intermediate cell types that accumulate in the absence of retroviral *Sox2* expression, we reasoned that it might also be possible to culture these responsive intermediates for prolonged periods of time. On the other hand, if RepSox acts during a critical window on very transient intermediates, this might not be possible. To distinguish between these models, we transduced *Oct4*::GFP MEFs with *Oct4, Klf4,* and *cMyc*, waited 10-14 days, and then

clonally expanded 10 iPS-like, GFP-negative colonies (Figure 2.6A). These cell lines continued to proliferate for at least 4 passages and often maintained an iPS-like morphology (Figure 2.6A) but never further activated expression of *Oct4*::GFP. However, when we treated these cell lines with a 48-hour pulse of RepSox, 5-10% of the colonies in 2 of the 10 lines became *Oct4*::GFP+ (Figure 2.6A-B). These results demonstrate that partially reprogrammed cells can accumulate in the absence of *Sox2* and that some, but not all, of these cells can be clonally expanded and cultured for prolonged periods while maintaining responsiveness to RepSox.

As we had shown that this particular reprogramming molecule seems to replace *Sox2* through the inhibition of Tgf-β signaling, we sought to determine whether RepSox treatment affected Tgf-β signal transduction pathways in these responsive cell lines. To this end, we determined the levels of phosphorylated Smad3 by western blot in cell line OKM 10 both with and without RepSox treatment. Without RepSox treatment, we detected relatively high levels of phosphorylated Smad3, suggesting that Tgf-β signaling was active (Figure 2.6C). In contrast, treatment with 25 μM RepSox almost completely eliminated Smad3 phosphorylation (Figure 2.6C), indicating that RepSox strongly inhibited Tgf-β signaling in these cells.

Because an increase in cell proliferation can also increase reprogramming efficiency (Hong et al., 2009) and possibly contribute to the replacement of transgenic *Sox2*, we measured the proliferation rate of partially reprogrammed OKM 10 cells both with and without RepSox. Treatment with RepSox decreased the proportion of cells in G2/M phase of the cell cycle (Figure 2.6D), indicating it does not increase the proliferation rate of these partially reprogrammed cells.

Figure 2.6.

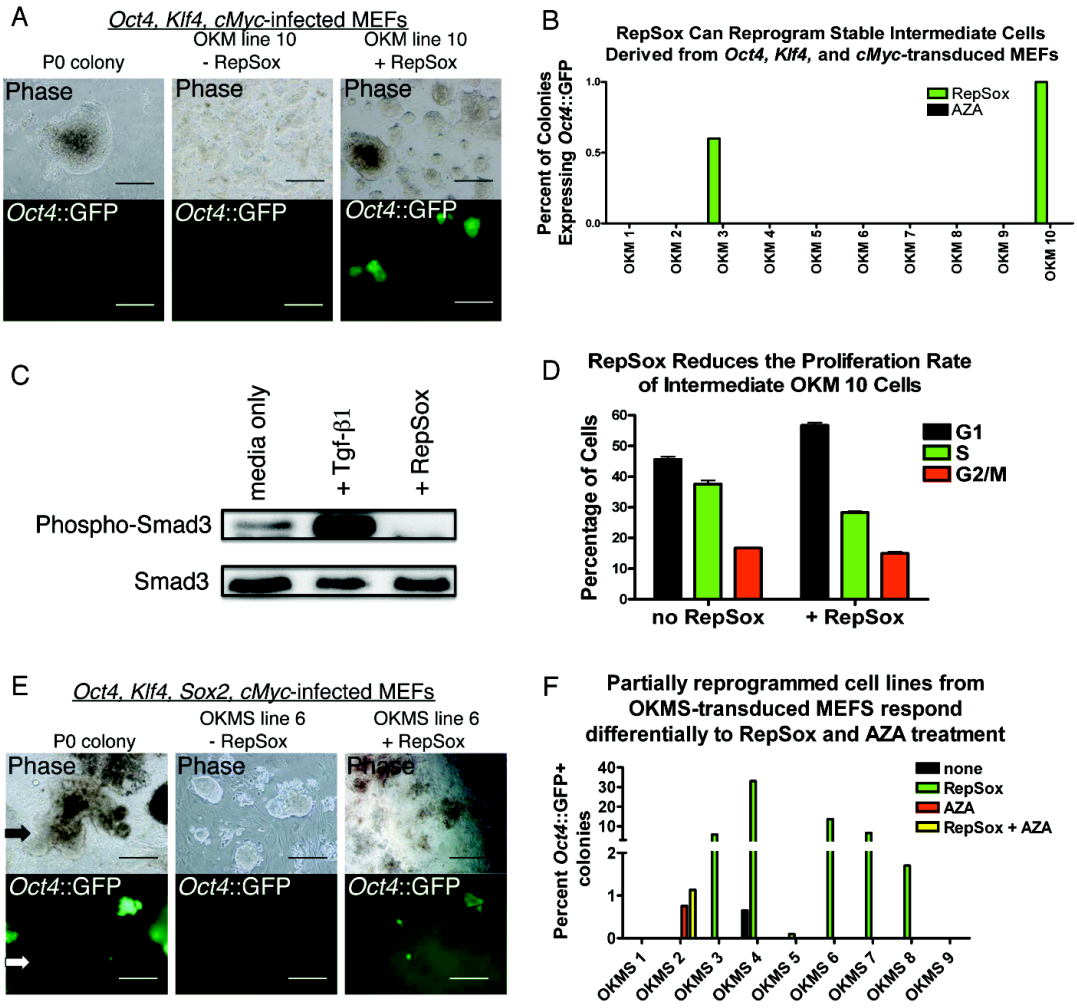
A
Oct4, Klf4, cMyc-infected MEFs
P0 colony
OKM line 10 - RepSox
OKM line 10 + RepSox
Phase
Oct4::GFP
B
RepSox Can Reprogram Stable Intermediate Cells Derived from Oct4, Klf4, and cMyc-transduced MEFs
Percent of Colonies Expressing Oct4::GFP
RepSox
AZA
OKM 1
OKM 2
OKM 3
OKM 4
OKM 5
OKM 6
OKM 7
OKM 8
OKM 9
OKM 10
C
media only
+ Tgf-β1
+ RepSox
Phospho-Smad3
Smad3
D
RepSox Reduces the Proliferation Rate of Intermediate OKM 10 Cells
Percentage of Cells
G1
S
G2/M
no RepSox
+ RepSox
E
Oct4, Klf4, Sox2, cMyc-infected MEFs
P0 colony
OKMS line 6 - RepSox
OKMS line 6 + RepSox
Phase
Oct4::GFP
F
Partially reprogrammed cell lines from OKMS-transduced MEFS respond differentially to RepSox and AZA treatment
Percent Oct4::GFP+ colonies
none
RepSox
AZA
RepSox + AZA
OKMS 1
OKMS 2
OKMS 3
OKMS 4
OKMS 5
OKMS 6
OKMS 7
OKMS 8
OKMS 9

Figure 2.6. (Continued)
Stable Intermediates Can Be Reprogrammed by RepSox. (A) Stable *Oct4*::GFP-negative cell lines derived from *Oct4*::GFP-negative colonies in *Oct4, Klf4,* and *cMyc*-infected MEF cultures can be reprogrammed by RepSox. *Oct4*::GFP-negative colonies were picked at day 14 post-infection, propagated, treated with 25 mM RepSox for 48 hours at passage 4, and scored for *Oct4*::GFP+ colonies 12 days after RepSox treatment. Scale bars in "OKM line 10 + RepSox" panels represent 500 μm; all other scale bars represent 200 μm. (B) Two of 10 stable, non-pluripotent intermediate cell lines derived from MEFs transduced with *Oct4, Klf4,* and *cMyc* can be reprogrammed with RepSox treatment but none can be reprogrammed with Aza treatment. (C) Western blot for phospho-Smad3 showing that RepSox inhibits Tgf-β signaling in line OKM 10 (OKM 10) cells. Lysates were generated from cells treated with 25 RepSox for 48 hours in mES media. (D) RepSox does not increase the proliferation of OKM 10 cells. (E) Stable *Oct4::*GFP-negative cell lines derived from *Oct4*::GFP negative colonies in *Oct4, Klf4, cMyc* and *Sox2*-infected MEF cultures can be reprogrammed by RepSox. Scale bars in "P4 line + RepSox" panels represent 500 μm; all other scale bars represent 200 μm. RepSox in mES media without feeders for 72 hours and subjected to cell cycle analysis by propidium iodide staining and flow cytometry. (F) Stable *Oct4::*GFP-negative cell lines derived from *Oct4*::GFP-negative colonies in *Oct4, Klf4, cMyc* and *Sox2*-infected MEF cultures can be reprogrammed by RepSox or by Aza, but lines responsive to RepSox are not responsive to Aza alone and lines responsive to Aza are not responsive to RepSox alone, indicating the presence of two different types of stable intermediates in the reprogramming cultures.

Cells That Respond to RepSox Treatment Are Distinct from Previously Described Intermediates

It has been shown that certain non-pluripotent, partially reprogrammed cell lines derived from MEFs transduced with *Oct4, Klf4, cMyc,* and *Sox2* can be fully reprogrammed with Aza or a combination of chemical inhibitors of Glycogen Synthase Kinase 3β (GSK-3β) and the Mek signaling pathway (2i conditions) (Mikkelsen et al., 2008; Silva et al., 2008). If the RepSox-responsive cell lines generated by overexpression of *Oct4, Klf4,* and *cMyc* were similar to these 4-factor cell lines, then they should also be reprogrammed by Aza or 2i. However, when we treated the 10 stable intermediate lines with either Aza or 2i for 48 hours, we found that none became reprogrammed (Figure 2.6B), indicating that the RepSox-responsive stable intermediates are distinct from partially reprogrammed cell lines described previously (Mikkelsen et al., 2008; Silva et al., 2009). Consistent with these results, *in vitro* assays of kinase activity revealed that RepSox does not inhibit the targets of the 2i cocktail (Table 2.3).

It occurred to us that some non-pluripotent cells derived from MEFs transduced with *Oct4, Klf4, cMyc*, and *Sox2* could potentially be held in a non-pluripotent state due to inappropriate levels of transgene expression and therefore might also be responsive to RepSox treatment. To test this hypothesis, we transduced *Oct4*::GFP MEFs with *Oct4, Klf4, cMyc*, and *Sox2*, then picked and clonally expanded 9 GFP-negative colonies at day 14 after transduction (Figure 2.6E). After treatment with RepSox, 5 of the 9 cell lines yielded reprogrammed colonies, with 2-33% of the colonies in each line becoming *Oct4*::GFP+ (Figures 2.6E-F). These results indicate that like the stable intermediate

	Average % inhibition	Standard error
Mek1	3	1
Mek2	-4	1
Erk1	7	0
Erk2	-2	2
GSK-3β	1	0

Table 2.3.
***In Vitro* Profiling of Kinase Inhibition Activities of RepSox.** RepSox does not inhibit the kinase targets of the 2i cocktail. Assays were performed in duplicate using the Z'-LYTE system (Invitrogen).

cells generated with only *Oct4, Klf4,* and *cMyc*, certain incompletely reprogrammed cells generated by *Oct4, Klf4, cMyc,* and *Sox2* transduction can also be reprogrammed by RepSox.

Next, in order to determine if these RepSox-responsive intermediate cell lines derived after *Oct4, Klf4, cMyc,* and *Sox2* transduction were similar to or distinct from previously described partially reprogrammed cell lines (Mikkelsen et al., 2008), we applied Aza to all 9 lines. After 48 hours of Aza treatment and 12 subsequent days in culture, none of the RepSox-responsive cell lines expressed *Oct4*::GFP (Figure 2.6F). However, one of the lines that had been refractory to RepSox treatment did express *Oct4*::GFP after Aza treatment, indicating that it had undergone complete reprogramming (Figure 2.6F). Together, these results show that there are a variety of intermediates that can form following retroviral transduction and that they vary in their responsiveness to reprogramming molecules.

RepSox Replaces Sox2 *by Inducing* Nanog *Expression*

The causal molecular events that drive reprogramming are difficult to detect because of the low efficiency at which somatic cells are successfully reprogrammed (Amabile and Meissner, 2009). However, when we administered RepSox to cell lines that had been partially reprogrammed by retroviral transduction, *Oct4*::GFP expression was induced in up to 33% of the resulting colonies (Figure 2.6F). We used this more efficient reprogramming system to identify the changes in gene expression induced by RepSox that enable it to bypass the requirement for transgenic *Sox2* expression.

We treated an *Oct4*::GFP-negative, partially reprogrammed cell line (OKMS 6) with RepSox and performed global gene expression analysis at 10, 24, and 48 hours following the initiation of treatment. To confirm that RepSox was inhibiting Tgf-β signaling in this intermediate cell line, we investigated expression changes in known Tgf-β-responsive genes after RepSox treatment. The *Inhibition of Differentiation* genes *Id1, Id2,* and *Id3* are repressed by Tgf-β signaling in mES cells (Ying et al., 2003). After treating the RepSox-responsive intermediate line OKM 10 with RepSox for 24 hours, we observed increased expression of *Id1, Id2,* and *Id3* (Figure 2.7A).

One way that RepSox could function to replace transgenic *Sox2* would be to induce the expression of endogenous *Sox2* or a *Sox*-family member, such as *Sox1* or *Sox3*, that can substitute for it in reprogramming (Nakagawa et al., 2008). However, we again did not observe a significant increase in the expression of *Sox1, Sox2, Sox3*, or any of the remaining *Sox*-family transcription factors within the first 48 hours of RepSox treatment (Figure 2.7B). Additionally, short-hairpin RNA (shRNA)-mediated depletion of *Sox1*, the most potent *Sox*-family member other than *Sox2* itself (Nakagawa et al., 2008), did not affect the rate of reprogramming in the presence of RepSox (Figure 2.7C). These results show that RepSox does not replace *Sox2* by directly activating endogenous *Sox2* or other closely related genes.

Next, we more broadly investigated changes in transcription factor expression following chemical treatment. We did not observe an increase in endogenous *Oct4* or *Klf4* expression at early time points following RepSox treatment. However, we found that the expression of the homeodomain factor *Nanog* was among the most increased following RepSox treatment. Relative to untreated controls, *Nanog* transcription

Figure 2.7.

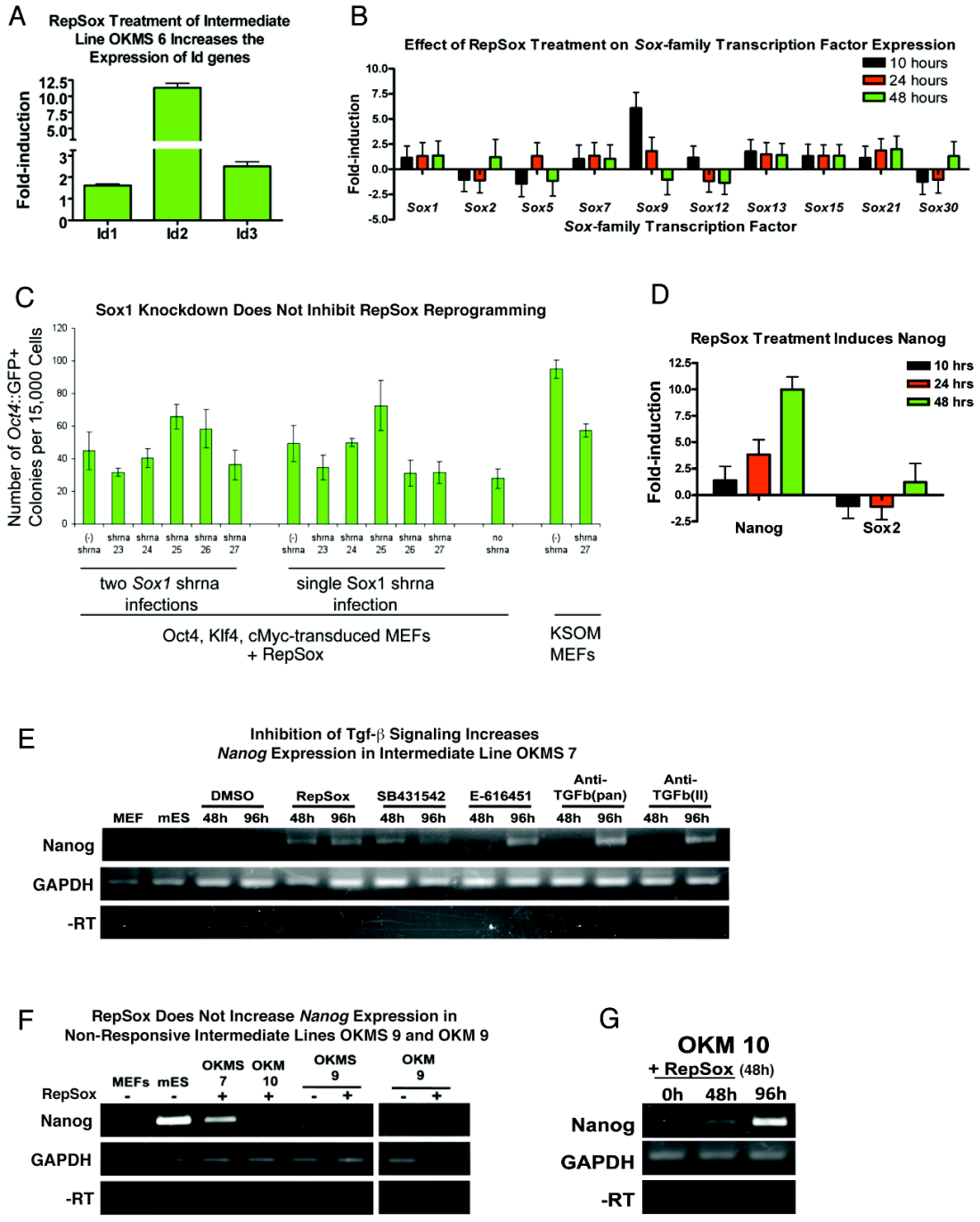
A
RepSox Treatment of Intermediate Line OKMS 6 Increases the Expression of Id genes
Fold-induction
Id1
Id2
Id3
B
Effect of RepSox Treatment on Sox-family Transcription Factor Expression
10 hours
24 hours
48 hours
Fold-induction
Sox1
Sox2
Sox5
Sox7
Sox9
Sox12
Sox13
Sox15
Sox21
Sox30
Sox-family Transcription Factor
C
Sox1 Knockdown Does Not Inhibit RepSox Reprogramming
Number of Oct4::GFP+ Colonies per 15,000 Cells
two Sox1 shrna infections
single Sox1 shrna infection
Oct4, Klf4, cMyc-transduced MEFs + RepSox
KSOM MEFs
D
RepSox Treatment Induces Nanog
10 hrs
24 hrs
48 hrs
Fold-induction
Nanog
Sox2
E
Inhibition of Tgf-β Signaling Increases Nanog Expression in Intermediate Line OKMS 7
DMSO
RepSox
SB431542
E-616451
Anti-TGFb(pan)
Anti-TGFb(II)
MEF
mES
48h
96h
Nanog
GAPDH
-RT
F
RepSox Does Not Increase Nanog Expression in Non-Responsive Intermediate Lines OKMS 9 and OKM 9
MEFs
mES
OKMS 7
OKM 10
OKMS 9
OKM 9
RepSox
Nanog
GAPDH
-RT
G
OKM 10
+ RepSox (48h)
0h
48h
96h
Nanog
GAPDH
-RT

Figure 2.7. (Continued)

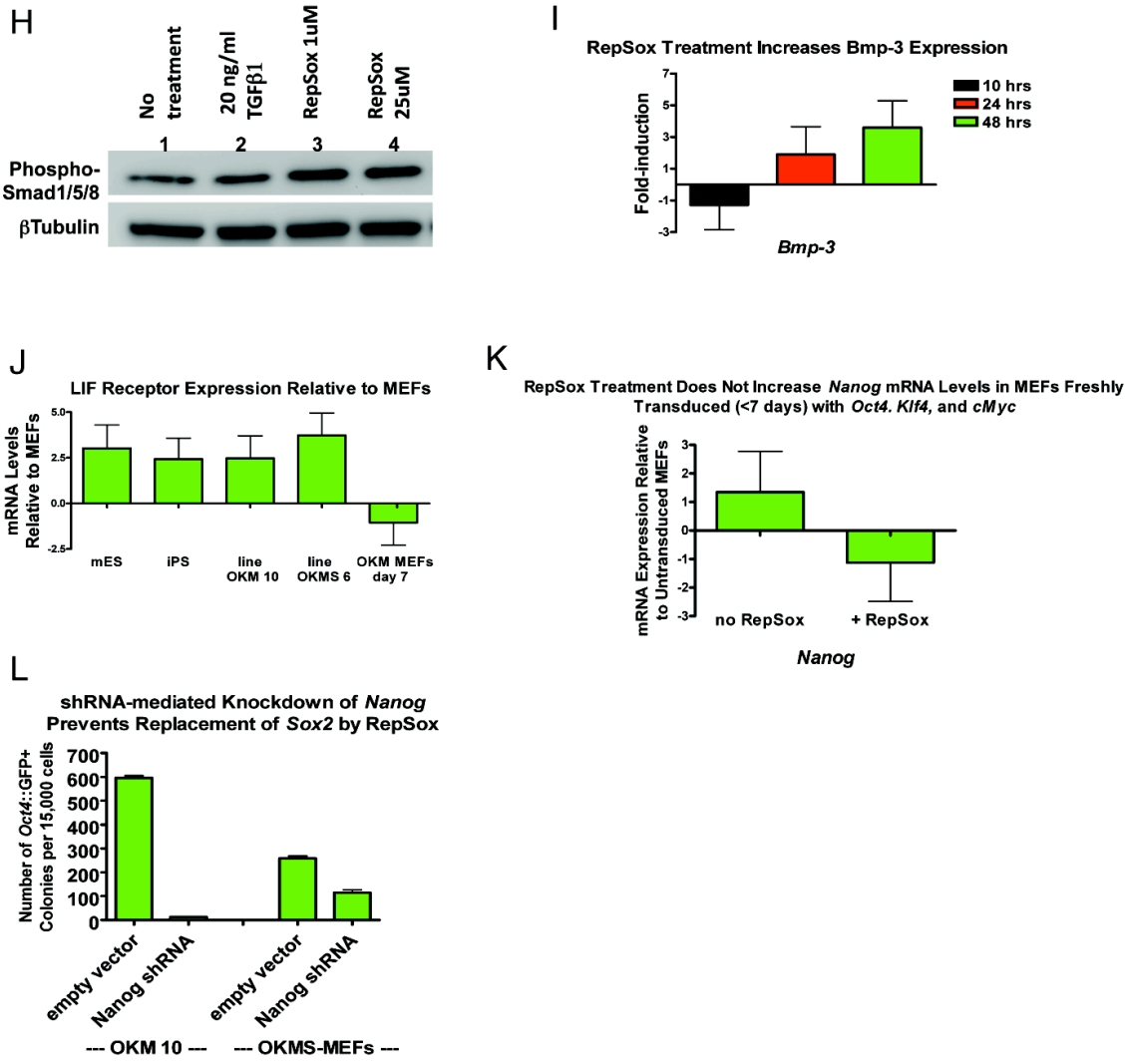

Figure 2.7. (Continued)
Induction of *Nanog* by RepSox Is Necessary for Its Reprogramming Activity. (A-B) RepSox-induced gene expression changes in in RepSox-responsive intermediate line OKMS 6 relative to untreated, time-matched controls. Cells were treated with 0 or 25 μM RepSox for 48 hours before RNA was harvested and analyzed by microarray. (A) *Id* gene expression increases. (B) *Sox*-family gene expression does not increase. Levels of *Sox3*, *4*, *6*, *8*, and *17* did not change significantly. (C) shRNA-mediated knockdown of *Sox1* does not inhibit reprogramming with RepSox. *Oct4, Klf4, cMyc*-transduced MEFs were transduced once or twice with lentiviral particles encoding 5 different *Sox1*-specific shRNA constructs or an empty vector control and subjected to RepSox treatment (25 μM) in KSR mES media. KSOM MEFs = *Klf4, Sox2, Oct4,* and *cMyc*-transduced MEFs. (D) RepSox treatment of RepSox-responsive line OKMS 6 strongly increases *Nanog* mRNA levels. Data wereas generated by microarray analysis and are relative to untreated controls. *Nanog* is induced faster and more significantly than *Sox2*, indicating it is upregulated before fully reprogrammed cells form. (E) RT-PCR analysis showing that Iinhibition of Tgf-β signaling increases *Nanog* expression in the RepSox-responsive intermediate line OKMS 7. (F) RepSox does not increase *Nanog* expression in non-RepSox-responsive intermediate lines OKMS 9 and OKM 9. Cells were treated with RepSox for 2 days in KSR mES media before RNA was harvested. Treatment was performed in KSR mES media. (G) A pulse of RepSox induces a persistent increase in *Nanog* expression in the RepSox-responsive intermediate line OKM 10. OKM 10 cells were treated with 25 μM RepSox for 48 hours and RNA samples were taken at 0, 48, and 96 hours (48 hours after removal of RepSox) and analyzed by RT-PCR for *Nanog* expression. (H-I) Bmp signaling increases in response to RepSox treatment. (H) Western blot for phospho-Smad1/5/8 shows an increase in the amount of the phosphorylated protein after a 48-hr RepSox treatment. (H) mRNA expression analysis shows that Bmp-3 levels increase upon RepSox treatment. Data are relative to untreated controls. (J-K) mRNA Expression analysis shows that non-pluripotent stable intermediate cell lines express the LIF receptor at the same level as mES cells, but freshly transduced MEFs do not, and MEFs do not upregulate *Nanog* significantly after RepSox treatment. (J) MEFs freshly infected with *Oct4, Klf4,* and *cMyc* (OKM MEFs day 7) express lower levels of the LIF receptor. (K) *Nanog* mRNA levels in MEFs freshly transduced with *Oct4, Klf4,* and *cMyc* (within 7 days) do not increase upon RepSox treatment. (L) shRNA-mediated knockdown of *Nanog* in OKM 10 cells inhibits replacement of *Sox2* by RepSox.

increased 4-fold within 24 hours and 10-fold after 48 hours of RepSox treatment (Figure 2.7D-E). In contrast, we did not observe a rapid increase in *Nanog* expression in 2 *Oct4*::GFP-negative intermediate cell lines that could not be fully reprogrammed using RepSox (Figure 2.7F). Therefore, we hypothesized that RepSox might replace *Sox2* by inducing *Nanog* expression.

Because we had determined that inhibition of Tgf-β signaling by several different small molecules and antibodies can replace *Sox2*, we reasoned that if the increase in *Nanog* expression was critical for *Sox2* replacement, the alternative inhibitors of Tgf-β signaling should also upregulate *Nanog*. To test this hypothesis, we treated the RepSox-responsive cell lines with RepSox, SB431542, or neutralizing antibodies and analyzed *Nanog* expression after 48 hours. In all cases, *Nanog* expression was strongly induced within 48-96 hours (Figure 2.7E).

If RepSox functions by increasing *Nanog* expression, then a short pulse of RepSox should induce a persistent increase in *Nanog* expression. To test this, we treated the RepSox-responsive intermediate cell line OKM 10 with RepSox for 48 hours, withdrew RepSox and then analyzed *Nanog* expression 48 hours later. A control time point taken just before RepSox withdrawal showed a significant increase in *Nanog* transcription (Figure 2.7G). 48 hours after RepSox removal (96 hours after the initiation of treatment), *Nanog* expression continued to increase (Figure 2.7G).

Previous reports have shown that chemical inhibition of Tgf-β signaling by SB431542 increases Bone Morphogenetic Protein (Bmp) signaling in embryonic stem cells (Xu et al., 2008). It has separately been shown that Bmp signaling in the presence of Stat3 induces *Nanog* expression in mES cells (Suzuki et al., 2006). The cross-talk

between the Tgf-β and Bmp signaling pathways may be the result of a common requirement for Smad 4, which mediates transcriptional events in the nucleus (Attisano and Wrana, 2002). Similarly, we observed an increase in the levels of phosphorylated Smad1 protein and *Bmp-3* mRNA in incompletely reprogrammed intermediates following RepSox treatment (Figure 2.7H-I). Furthermore, the stable, partially reprogrammed cells that responded to RepSox expressed the LIF receptor at levels equivalent to those found in mES cells (Figure 2.7J). Expression of this receptor suggests that its downstream signal transduction pathway could be active in these cells, resulting in the presence of activated Stat3, which is known to induce *Nanog* expression in conjunction with Bmp signaling.

If *Nanog* upregulation is the mechanism by which RepSox replaces *Sox2*, we would not expect *Nanog* to be upregulated in RepSox-treated MEFs, since RepSox does not act on the initial population of fibroblasts in this capacity. Indeed, within 7 days of transduction of MEFs with *Oct4, Klf4,* and *cMyc*, we did not observe an increase in *Nanog* expression upon RepSox treatment (Figure 2.7K). This may be explained in part by the observation that the LIF receptor, and thus activated Stat3, were not highly expressed in these cells (Figure 2.7J).

We decided to investigate whether *Nanog* upregulation is a necessary event during RepSox-mediated reprogramming. If RepSox replaces *Sox2* by increasing *Nanog* expression, then a forced reduction of *Nanog* expression should inhibit or even prevent reprogramming by RepSox. To test this hypothesis, we transduced the RepSox-responsive cell line with a lentivirus encoding an shRNA specific for *Nanog*. The *Nanog*-knockdown cells reprogrammed at a frequency that was 50-fold lower than cells

transduced with an empty control vector (Figure 2.7L). This effect was not due to a general decrease in reprogramming efficiency or differentiation of reprogrammed cells due to *Nanog* depletion because MEFs transduced with *Oct4, Klf4, cMyc, Sox2*, and the *Nanog* shRNA construct only suffered a 50% loss in reprogramming efficiency (Figure 2.7D). These results demonstrate that increased *Nanog* expression in this context was only necessary for the replacement of *Sox2* by RepSox.

If RepSox replaces *Sox2* by inducing *Nanog* expression, then retroviral transduction of RepSox-responsive intermediate cells (line OKM 10, Figures 2.6A-B) with *Nanog* should reprogram them. When we transduced line OKM 10 with *Sox2* as a control, .2% of the colonies expressed *Oct4*::GFP after 10 days, indicating that reprogramming could be induced in this cell line by *Sox2* (Figures 2.8A-B). When we transduced the same stable intermediate cell line with *Nanog*, it could also be reprogrammed, with .3% of the colonies expressing *Oct4*::GFP+ after 10 days (Figures 2.8A-B). In contrast, transductions with *Oct4* or *Klf4* resulted in only .04% and 0% reprogramming efficiencies (Figure 2.8B). These results suggest that *Nanog* can indeed functionally replace *Sox2* and induce reprogramming in these stable intermediates formed from *Oct4, Klf4,* and *cMyc*-transduced MEFs.

Because *Nanog* plays a key role in maintaining ES cells in an undifferentiated state (Chambers et al., 2003) and has been shown to enhance the efficiency of reprogramming (Silva et al., 2006; Silva et al., 2009; Yu et al., 2007), we decided to test whether *Nanog* could directly replace *Sox2* in reprogramming. If *Nanog* can indeed complement for the omission of *Sox2* in defined factor reprogramming, then MEFs transduced with *Oct4, Klf4, cMyc,* and *Nanog* might be as efficiently reprogrammed as

Figure 2.8.

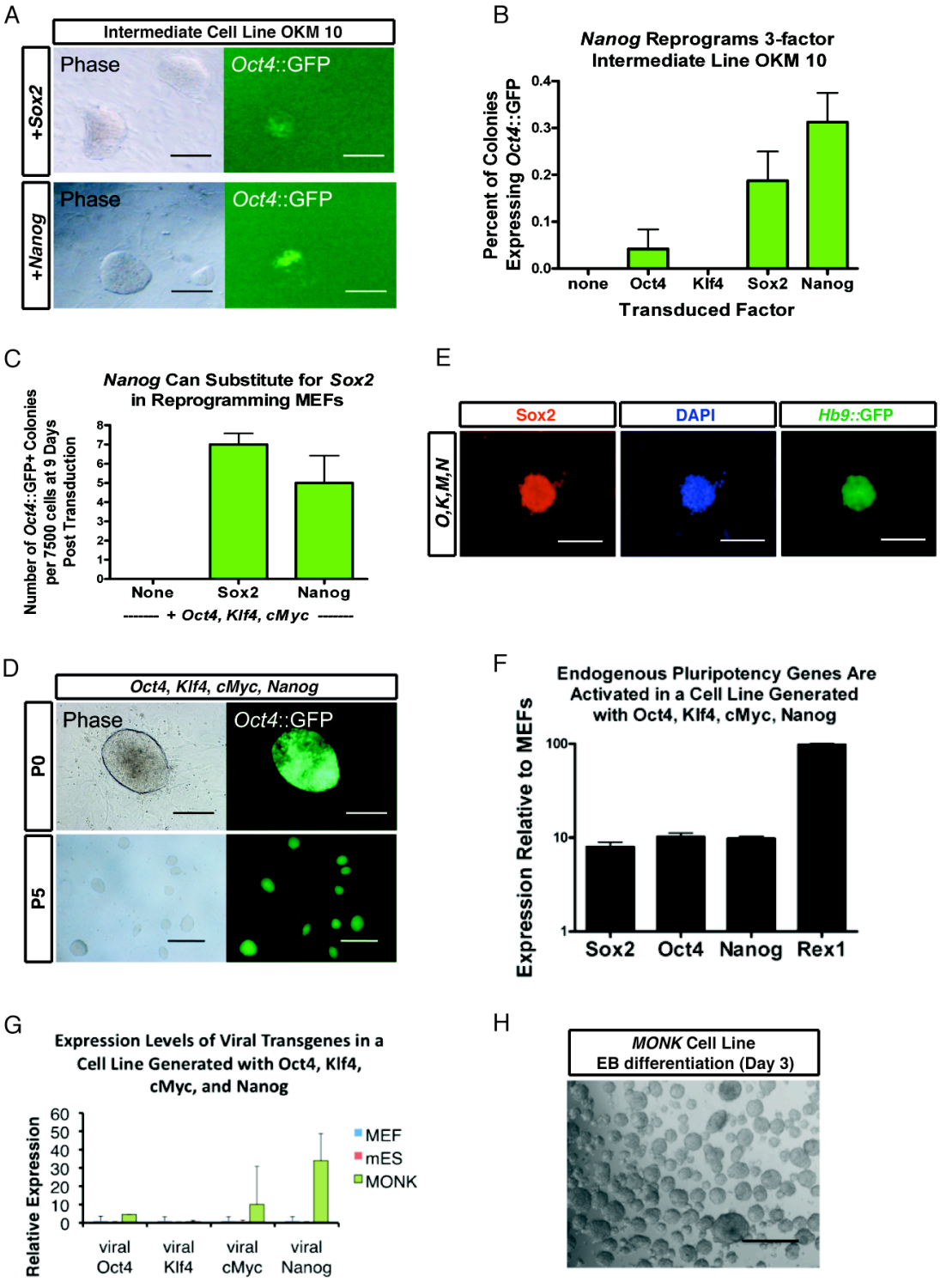
A
Intermediate Cell Line OKM 10
+Sox2
+Nanog
Phase
Oct4::GFP
B
Nanog Reprograms 3-factor
Intermediate Line OKM 10
Percent of Colonies Expressing Oct4::GFP
0.4
0.3
0.2
0.1
0.0
none
Oct4
Klf4
Sox2
Nanog
Transduced Factor
C
Nanog Can Substitute for Sox2
in Reprogramming MEFs
Number of Oct4::GFP+ Colonies per 7500 cells at 9 Days Post Transduction
8
7
6
5
4
3
2
1
0
None
Sox2
Nanog
------ + Oct4, Klf4, cMyc ------
E
Sox2
DAPI
Hb9::GFP
O,K,M,N
D
Oct4, Klf4, cMyc, Nanog
P0
P5
Phase
Oct4::GFP
F
Endogenous Pluripotency Genes Are
Activated in a Cell Line Generated
with Oct4, Klf4, cMyc, Nanog
Expression Relative to MEFs
100
10
1
Sox2
Oct4
Nanog
Rex1
G
Expression Levels of Viral Transgenes in a
Cell Line Generated with Oct4, Klf4,
cMyc, and Nanog
Relative Expression
60
50
40
30
20
10
0
MEF
mES
MONK
viral Oct4
viral Klf4
viral cMyc
viral Nanog
H
MONK Cell Line
EB differentiation (Day 3)

Figure 2.8. (Continued)
***Nanog* Replaces *Sox2* in iPS Reprogramming**. (A) Pictures of reprogrammed *Oct4*::GFP+ colonies induced by *Sox2* (top panels) or *Nanog* (bottom panels) transduction of line OKM 10. . Cells were grown in KSR mES media. Scale bars represent 200 μm. (B) *Nanog* transduction can reprogram line OKM 10 at a similar efficiency as *Sox2* transduction. Cells were grown in KSR mES media and *Oct4*::GFP+ colonies were counted at 9 days post transduction.(C) *Nanog* can substitute for *Sox2* in defined-factor reprogramming of somatic fibroblasts. (D-H) Cells were grown in KSR mES media. *Oct4*::GFP+ colonies were counted at 9 days post-transduction.MEF-derived, *Oct4, Klf4, cMyc* and *Nanog* (MONK)-induced *Oct4::*GFP+ cells exhibit characteristics of iPS cells. (D) A P0 colony (top panels), isolated and expanded (bottom panels, at Passage 5). Scale bars represent 100 μm. (E) Immunocytochemistry showing strong expression of *Sox2* from the endogenous allele. Scale bars represent 200 μm. (F) qPCR analysis shows activation of the endogenous *Sox2, Oct4, Nanog,* and *Rex1* in these cells. (G) qPCR analysis showing viral *Oct4, Klf4, and cMyc* are silenced but leaky expression from the *Nanog* transgene remains in MONK Line 1. (H) MONK Line 1 cells readily form embryoid bodies after 3 days in culture. Scale bar represents 500 μm.

MEFs transduced with *Oct4, Klf4, cMyc,* and *Sox2*. When we transduced MEFs with *Oct4, Klf4, cMyc*, and *Sox2* then scored cultures 9 days later, an average of 7 *Oct4*::GFP+ colonies appeared for every 7500 cells plated (Figure 2.8C). A control transduction with only *Oct4, Klf4,* and *cMyc* yielded no *Oct4*::GFP+ colonies (Figure 2.8C). Similar to the positive control transduction, MEFs transduced with *Oct4, Klf4, cMyc*, and *Nanog* gave rise to an average of 5 *Oct4*::GFP+ colonies for every 7500 cells plated (Figures 2.8C-D). These colonies could be picked and expanded and remained *Oct4*::GFP+ for at least 5 passages (Figure 2.8D). Immunocytochemistry indicated that these cells strongly activated *Sox2* expression from the endogenous allele (Figure 2.8E). Importantly, qPCR analysis demonstrated that they also transcribed endogenous *Oct4, Klf4, Nanog,* and *Rex1* (Figure 2.8F), indicating that a pluripotent gene expression program had been established. Furthermore, transgene-specific qPCR analysis showed that these cells had silenced the retroviral *Oct4*, *Klf4*, and *cMyc* transgenes, (Figure 2.8G). Additionally, *Oct4, Klf4, cMyc,* and *Nanog*-reprogrammed cells could readily form embryoid bodies *in vitro* (Figure 2.8H). However, we found that leaky expression of transgenic *Nanog*, which is a potent inhibitor of embryonic stem cell differentiation (Chambers et al., 2003; Chambers et al., 2007), reduced the amount of differentiation *in vitro* (Figure 2.8G). We anticipate that efficient differentiation of cells created with *Oct4, Klf4, cMyc,* and *Nanog* will eventually require the use of an excisable transgenic *Nanog* cassette to completely remove ectopic *Nanog* expression. Although definitive proof of the pluripotency of these cells will be required to conclude that *Nanog* expression is sufficient to replace *Sox2* in defined factor reprogramming, our results suggest this may be the case. Taken together,

our results demonstrate that RepSox inhibition of Tgf-β signaling bypasses the need for *Sox2* in defined-factor reprogramming through the induction of *Nanog*.

Discussion

We have used a phenotypic chemical screen to identify compounds that can replace the reprogramming transcription factor *Sox2* and have confirmed the mechanism by which the most potent compound acts: RepSox replaces *Sox2* by inhibiting the broadly expressed Tgf-β signaling pathway (Attisano and Wrana, 2002) in cultures containing stable intermediate cells that are trapped in a partially reprogrammed state. This inhibition in turn leads to sustained transcription of *Nanog*, through which reprogramming is achieved in the absence of *Sox2*. These results demonstrate the feasibility of replacing the central reprogramming transgenes with small molecules that modulate discrete cellular pathways or processes rather than by globally altering chromatin structure. Furthermore, they show that the mechanisms by which these molecules act in reprogramming can be distinct from those of the factor(s) that they replace.

Importantly, and unlike many other studies (Mikkelsen et al., 2008; Shi et al., 2008a; Shi et al., 2008b; Utikal et al., 2009), the approach that we report here for replacing *Sox2* did not rely on procurement of a highly specialized or rare cell type that already expresses *Sox2*. Furthermore, treatment with RepSox allowed the generation of iPS cells from both adult and embryonic fibroblasts with a frequency comparable to that of transduction with *Sox2*. Thus, reprogramming efficiency does not need to be compromised by small molecule replacement of transgenic factors.

We observed that instead of working on the initial fibroblast population to replace *Sox2*, RepSox acts on cellular intermediates formed by overexpression of *Oct4*, *Klf4,* and *cMyc*. Without RepSox treatment, these intermediates are trapped in an unproductive state. Unlike previously described partially reprogrammed cells (Mikkelsen et al., 2008; Silva et al., 2009), the RepSox-responsive intermediates could not be reprogrammed with Aza or 2i treatment, suggesting that they are distinct. In addition, we found that RepSox does not target any of the kinases inhibited by the 2i cocktail, indicating that it works through a different mechanism. Furthermore, 4-factor intermediates that reprogram with RepSox treatment are not responsive to Aza, indicating that they also are distinct.

These findings demonstrate that reprogramming can proceed in a step-wise fashion through different intermediates. Just as in a geographical setting where there are multiple routes to travel from point A to point B, there exist different intermediate states or "way stations" that somatic cells can transit through on the way to complete reprogramming. Interestingly, although our results indicate that defined-factor reprogramming with *Oct4, Klf4, cMyc*, and *Sox2* can occur in the absence of *Nanog*, its induction is required for chemical reprogramming of both our RepSox-responsive intermediates and the recently described 2i-responsive intermediates made from *Oct4, Klf4* and *cMyc* transduction of cells that express *Sox2* endogenously (Silva et al., 2009). This indicates that commonalities can exist in the reprogramming routes used by some sets of distinct intermediates.

Originally, we found it surprising that *Nanog* was not included in the initial set of defined reprogramming factors (Takahashi and Yamanaka, 2006) given its critical role in maintaining pluripotency in ES cells (Boyer et al., 2005; Chambers et al., 2003) and its

ability to stimulate reprogramming by cell-fusion (Silva et al., 2006). However, Takahashi and Yamanaka reported that a combination of 9 factors that included *Oct4, Klf4, cMyc*, and *Nanog*, but not *Sox2,* generated iPS colonies at a detectable rate (Takahashi and Yamanaka, 2006). This combination of factors included other genes that may have inadvertently lowered the rate of reprogramming, causing the combination of *Oct4, Klf4, cMyc*, and *Nanog* to be overlooked. Consistent with these data, work by Niwa and co-workers using inducible *Sox2*-null mES cells demonstrated that *Sox2* is dispensable for modulation of the Oct-Sox enhancers that regulate pluripotent-specific gene expression and instead mainly governs pluripotency in ES cells by regulating the expression of *Oct4* through other factors (Masui et al., 2007). Therefore, it is possible that *Nanog* may alleviate the requirement for *Sox2* in reprogramming by stimulating or maintaining *Oct4* expression. Indeed, *Nanog* is capable of maintaining *Oct4* expression in mES cells (Chambers et al., 2003). Thompson and co-workers also reported that *NANOG* expression enhanced the reprogramming of human fibroblasts, but that it was not able to replace *SOX2* in the presence of only *OCT4* and *LIN-28* (Yu et al., 2007). This may indicate that *Klf4* is required for *Nanog* to function optimally in reprogramming and suggests that either they or the genes they modulate interact during the reprogramming process.

It is well known that approximately 90% of genes with promoters bound by OCT4 and SOX2 in human ES cells are also bound by NANOG (Boyer et al., 2005). Our result suggests that either *Nanog* or *Sox2* may be sufficient to collaborate with *Oct4* to modulate these genes and drive reprogramming. Although *Nanog* is not required for pluripotency, it safeguards ES cells against neuroectodermal and, to a more limited

extent, mesodermal differentiation (Chambers et al., 2007; Vallier et al., 2009). Therefore, it is possible that *Nanog* functions in reprogramming by repressing differentiation signals, assisting in the transition to an undifferentiated state.

Interestingly, we found that RepSox is also able to functionally replace *cMyc* in reprogramming. Together, these observations highlight the fact that small molecules may functionally replace reprogramming transcription factors at either early or late stages of the process and that they can act by different mechanisms – by inducing the expression of the gene itself, or a closely related family member, or an unrelated gene that can functionally rescue the omission of the reprogramming transcription factor.

Our observation that a one-day treatment with RepSox can relieve the requirement for transgenic *Sox2* indicates that unlike reprogramming using transgenic *Oct4, Klf4,* and *Sox2*, where each transgene must be expressed for several days (Sridharan et al., 2009; Stadtfeld et al., 2008), small molecules can act as switches to induce stable changes in gene expression that promote the completion of reprogramming. This could be an important concept for achieving purely chemical reprogramming since our data show that chemicals such as RepSox can affect cellular processes differently depending on the timing of administration.

As we have shown here, there need not always be a discrete, one-to-one mapping between the functions of the reprogramming factors and their chemical replacements. Thus it may be that reiterative screening in the presence of *Sox2* replacement molecules will be required to identify compounds that can act in concert to replace *Oct4* and *Klf4*. However, it will be of significant interest to determine whether the novel reprogramming compounds we have identified can collaborate with those previously described (Marson

et al., 2008; Shi et al., 2008a; Silva et al., 2008) to replace the remaining reprogramming genes, opening a route to purely chemical reprogramming.

Materials and Methods

Derivation of MEFs, Cell Culture, and Retroviral Infection

MEFs were derived as previously described (Takahashi et al., 2007a). Retroviral infections were performed as previously described using the pMXs vector (Takahashi et al., 2007a). MEFs were infected with two to three pools of viral supernatant during a 72-hour period. The first day that viral supernatant was added was termed "day 1 post-infection." For quantification, *Oct4*::GFP+ colonies were counted at day 30 post-infection unless otherwise stated.

Derivation of Tail Tip Fibroblasts, Cell Culture, and Retroviral Infection

Adult tail tip fibroblasts were isolated from tails of 8-week old *Oct4*::GFP mice and cultured in DMEM supplemented with 40% fetal bovine serum and penicillin/streptomycin. For reprogramming experiments, P2 fibroblasts were infected by the same method as described for MEFs.

Small Molecule Screens

On day 4 post-infection, infected MEFs were trypsinized and re-seeded on irradiated feeders in 96-well plates at 2000 cells/well and cultured in mouse ES cell media (Knockout DMEM,15% Hyclone FBS, L-glutamine, penicillin/streptomycin, nonessential amino acids, β-mercaptoethanol, and 1000 U/ml LIF). The next day, compound stock solutions diluted in DMSO and VPA (Sigma) were added at a final concentration of 1 μM and 2 mM, respectively. VPA was removed after 1 week, and

compound was re-applied every other day with each media change. Plates were scored for GFP+ colonies after 11 days of compound treatment.

Quantification of Oct4::*GFP+ iPS Cells Generated with Small Molecule Hit Compounds, SB431542, and Tgf-β Antibodies*

Retroviral infection and compound or antibody treatment was performed as in the original chemical screen. To quantify the numbers of GFP+ colonies produced in different conditions, the number of colonies in each well was counted and at least 2 different wells were counted and averaged. Concentrations of compounds and antibodies were the following: VPA (Sigma)- 2 mM, RepSox (Calbiochem)- 25 μM or 1 μM as noted, E-616451 (Calbiochem)- 3 μM, EI-275 (Biomol)- 3 μM, SB431542 (Sigma)- 25 μM or 2 μM as noted, Tgf-βII-specific antibody (R&D Systems, AB-12-NA)- 10 μg/ml, pan-Tgf-β antibody (R&D Systems, AB-100-NA)- 10 μg/ml. Unless otherwise noted, all chemical treatments were continuous from initial administration at day 4-5 post-infection until GFP+ colonies were scored at day 30 post-transduction. Fresh chemical was added at each media change.

Lead Compound Titrations to Determine Optimal Dosage

Infections and VPA/compound addition was done as in the original chemical screen, and wells were scored for GFP+ colonies on day 25 after compound addition.

Generation of iPS Cells

GFP+ P0 colonies were picked manually and incubated in .25% trypsin (Gibco) for 20 minutes at room temperature before plating on a feeder layer in mES cell media. This process was repeated until passage 3, at which time colonies were trypsinized and passaged in bulk and maintained on feeders in mES cell media.

Karyotyping

Karyotype analysis was performed at Cell Line Genetics.

Antibody Staining for Sox2 and Nanog and Alkaline Phosphatase Staining

iPS cells were cultured on irradiated MEF feeders in chamber slides, fixed with 4% paraformaldehyde (PFA) and stained with primary antibodies against mSox2 (Santa Cruz, sc-17320), mNanog (CosmoBio, REC-RCAB0002PF), followed by staining with the appropriate secondary antibodies conjugated to Alexa Fluor 546 (Invitrogen). Nuclei were counterstained with Hoechst33342 (Sigma). iPS cells were assayed for alkaline phosphatase activity using the Vector Red alkaline phosphatase assay kit from Vector Laboratories.

***Spontaneous Differentiation of iPS Cells* in Vitro**

iPS cells were grown to 70–80% confluence in 10-cm plates (Falcon) in mES cell medium. To form embryoid bodies, cells were washed once with PBS to eliminate mES cell medium and then incubated with 1 ml of 0.25% trypsin (GIBCO) for 5–10 min at room temperature (21-25 °C). Cells were then resuspended in 10 ml of DM1 medium (DMEM-F12, GIBCO), 10% knockout serum (GIBCO), penicillin, streptomycin,

glutamine (GIBCO) and 2-mercaptoethanol (GIBCO), counted, and plated at a concentration of 200,000 cells per ml in Petri dishes (Falcon). Two days later, embryoid bodies were split from one dish into four Petri dishes containing DM1 medium and the medium was changed after 3–4 d. On day 10 the embryoid bodies were collected in a 15-ml Falcon tube, washed once with PBS and then fixed in PFA 4% at 4 degrees C for 1 hour. The EBs were then washed 4 times in PBS to remove the residual PFA and incubated overnight in a solution of 30% of sucrose. The next day, the cells were embedded in OCT and frozen at -80 degrees C. The block containing EBs were then sectioned with a cryostat into 10 μm sections. The sections were stained with primary antibodies against Alpha-fetoprotein (AFP)(Dakocytomation, A0008), Skeletal Myosin (MF20)(Developmental Studies Hybridoma Bank, MF20), or Beta-III-tubulin (TUJ1)(Sigma, T2200), and visualized by staining with a secondary antibody conjugated to Alexa Fluor 546 (Invitrogen).

Directed Differentiation of iPS Cells into Motor Neurons

iPS and mES (V6.5) cells were differentiated into motor neurons according to methods previously described for mouse ES cells differentiation [27]. The iPS and mES cells were grown to 70–80% confluence in 10-cm plates (Falcon) in mES cell medium. To form embryoid bodies, cells were washed once with PBS to eliminate mES cell medium and then incubated with 1 ml of 0.25% trypsin (GIBCO) for 5–10 min at room temperature (21-25 °C). Cells were then resuspended in 10 ml of DM1 medium (DMEM-F12, GIBCO), 10% knockout serum (GIBCO), penicillin, streptomycin, glutamine (GIBCO) and 2-mercaptoethanol (GIBCO), counted and plated at a concentration of 200,000 cells

per ml in Petri dishes (Falcon). Two days later, embryoid bodies were split from one dish into four Petri dishes containing DM1 medium supplemented with RAc (100 nM; stock: 1 mM in DMSO, Sigma) and Shh (300 nM, R&D Systems). Medium was changed after 3–4 d. On day 7, the embryoid bodies were dissociated into single-cell suspensions. The suspensions were pelleted in a 15-ml Falcon tube, washed once with PBS, and incubated in Earle's balanced salt solution with 20 units of papain and 1,000 units of DNase I (Worthington Biochemical) for 30–60 min at 37 °C. The mixture was then triturated with a 10-ml pipette and centrifuged for 5 min at 300 x *g*. The resulting cell pellet was washed with PBS and resuspended in F12 medium (F12 medium, GIBCO) with 5% horse serum (GIBCO), B-27 supplement (GIBCO), N2 supplement (GIBCO) with neurotrophic factors (GDNF and BDNF, 10 ng ml^{-1}, R&D Systems). The cells were counted and plated on poly-D-lysine/laminin culture slides (BD Biosciences) or on a layer of primary glial cells. 3-5 days later, the cultures were fixed with PFA and stained with primary antibodies against TUJ1 (Sigma, T2200) and HB9 (Developmental Studies Hybridoma Bank, 81.5C10), and visualized by staining with secondary antibodies conjugated to Alexa Fluor 488 and Alexa Fluor 546 (Invitrogen). For counting HB9+ cells, motor neurons were differentiated as above except in embryoid body culture without dissociation and plating. Embryoid bodies were sectioned as above and stained with the TUJ1 and HB9 antibodies along with the Alexa Fluor 488 and Alexa Fluor 546 secondary antibodies. Cultures were counterstained with Hoechst 33342 and HB9+ and total nuclei were counted. Numbers were derived from at least 3 different embryoid bodies per cell line.

Teratoma Production and Analysis

A confluent 10 cm dish of iPS cells was trypsinized, pelleted, resuspended in .2 mls of mES media, and injected subcutaneously into a CD1-Nude mouse. 3-4 weeks later, teratomas were harvested, fixed overnight with 4% paraformaldehyde, embedded in paraffin, sectioned, HE stained, and analyzed.

Production of Chimeric Mice

Female ICR mice were superovulated with PMS and hCG and mated to ICR stud males. 24-hours after hCG injection, zygotes were isolated from vaginally plugged females. After culture in KSOM media for 3 days, the resulting blastocysts were injected with ~5-10 iPS cells from a C57BL6 background pre-labeled with a lentivirus constitutively expressing the red fluorescent protein tdTomato and transferred into pseudopregnant females. Embryos were either harvested at day E13.5 or allowed to develop to term. Chimeric embryos were visualized on a Leica MZ16FA dissecting microscope using RFP and bright field channels. For 8-cell stage injections, zygotes were developed *in vitro* to the 8-cell stage, injected with iPS cells, further developed *in vitro* to the blastocyst stage, and visualized.

Genital Ridge Isolation and Visualization

The genital ridges of E13.5 embryos were mechanically isolated and visualized using a Leica dissection microscope.

Chemical Reprogramming of Stable Intermediate Cell Lines

Oct4::GFP-negative colonies in *Oct4, Klf4,* and *cMyc* or *Oct4, Klf4, cMyc*, and *Sox2*-infected MEF cultures were picked, plated on irradiated feeders, and single colonies were picked after 1 week. The resulting cell lines were passaged with trypsin and grown in mES media on feeders until passage 4, at which time they were treated with RepSox (25 μM), Aza (500 μM), or both for 48 hours. For 2i treatment, CHIR99021 (Stemgent) was used at 3 μM and PD0325901 (Stemgent) was used at 1 μM. *Oct4*::GFP+ colonies were scored 12 days after the beginning of chemical treatment. Treatments were performed in mES media containing FBS unless otherwise noted.

Whole-Genome Expression Analysis

For comparison to mES and iPS cells, cells reprogrammed with RepSox were grown to near confluence on an irradiated layer and RNA was isolated with Trizol (Invitrogen). In other experiments analyzing the effect of RepSox treatment, cells were harvested at less than 60% confluence. RNA was amplified and labeled with biotin using the Illumina Total Prep RNA Amplification Kit from Ambion, hybridized to Illumina Whole-Genome Expression BeadChips (MouseRef-8), and analyzed by an Illumina Beadstation 500. All lines were analyzed in biological duplicate or triplicate. Data were processed using Resolver software.

Western Blots

Lysates were generated from cells treated with 25 μM RepSox for 48 hours in mES media. Cells were harvested using a cell scraper into lysis buffer containing protease

inhibitors. For analysis of phospho-Smads, an anti-phospho-Smad2/3-specifc antibody (Epitomics) and an anti-phospho-Smad1/5/8-specific antibody (R&D) were used.

Cell Cyle Analysis

Cells were treated with 25 μM RepSox in mES media without feeders for 72 hours and subjected to cell cycle analysis by propidium iodide staining and flow cytometry. Cells were harvested with .25% trypsin and fixed with 70% ethanol overnight. Following at least one hour of incubation with propidium iodide staining solution (50 μg/ml propidium iodide in PBS, .1% BSA, .1% Rnase A) in the dark, samples were analyzed on a BD LRSII Flow Cytometer (BD Biosciences).

RT-PCR

For experiments measuring *Nanog* induction, cells were treated with 25 μM RepSox in KSR mES media. RNA was harvested with Trizol (Invitrogen) and treated with Turbo-free (Ambion) to remove DNA contamination. RNA was reverse transcribed using random hexamer primers and superscript III reverse transcriptase (Invitrogen). Primer sequences for endogenous genes were the following: *Nanog* (5'-CAGGTGTTTGAGGGTAGCTC and 5'- CGGTTCATCATGGTACAGTC), *Sox2* (5'-TAGAGCTAGACTCCGGGCGATGA and 5'- TTGCCTTAAACAAGACCACGAAA), *Oct4* (5'- TCTTTCCACCAGGCCCCCGGCTC and 5'-TGCGGGCGGACATGGGGAGATCC), *Rex1* (5'-ACGAGTGGCAGTTTCTTCTTGGGA and 5'-TATGACTCACTTCCAGGGGGCACT). The reverse primer (5'-

TTTCTACAAGAAAGCTGGGT) was used for all transgenes, plus the following forward primers: *Nanog* (5'- TTGGAATGCTGCTCCGCTCC), *Sox2* (5'- CTACAGCATGTCCTACTCGC), *Oct4* (5'- GCTATGGAAGCCCCCACTTC), and *Klf4* (5'- TGACTATGCAGGCTGTGGCA). qPCR was performed using these primers and SYBR green (Bio-Rad).

***shRNA-mediated Knockdown of* Nanog *and* Sox2**

OKM 10 cells or MEFs transduced 4 days earlier with *Oct4, Klf4, cMyc,* and *Sox2* (OKMS-MEFs) were transduced with shRNA constructs in the lentiviral vector pLKO.1 that were specific to murine *Nanog* (5'- CCGGCCTGAGCTATAAGCAGGTTAACTCGAGTTAACCTGCTTATAGCTCAGGTTTTTG) or *Sox2* (5'- CCGGCGAGATAAACATGGCAATCAACTCGAGTTGATTGCCATGTTTATCTCGTTTTTG) (Open Biosystems). Lentiviruses were packaged by co-transfection of pLKO.1-shRNA plasmids with VSVG envelope and delta 8.9 plasmids into 293T cells using Fugene 6. Starting two days after infection, the population was enriched for transduced cells by selection with 4 μg/ml puromycin for three days. For OKM 10 cells, RepSox treatment (25 μM) was initiated after puromycin selection. RepSox treatment was performed in KSR mES media for 9 days before GFP+ colonies were scored.

Reprogramming of Stable Intermediate Cell Lines by Viral Transduction

Oct4::GFP-negative cell lines were transduced using the same methodology and reagents as MEFs were in the original screen. Cells were infected with three rounds of viral

supernatant diluted 1:8 in MEF media in a 48-hour period on gelatin. Two days after the last viral supernatant was added, the cells were trypsinized and replated onto feeders. The media was changed to mES media containing knockout serum replacement (KSR) instead of FBS on the following day. *Oct4*::GFP+ colonies were counted at 9 days post-transduction.

Reprogramming of MEFs Using **Nanog**

MEFs were infected as described for the original screen, except that murine *Nanog* cDNA was cloned into the pMXs retroviral vector and used instead of pMXs-*Sox2*. Two days after the last viral supernatant was added, the cells were trypsinized and replated onto feeders. The media was changed to mES media containing knockout serum replacement (KSR) instead of FBS on the following day, and *Oct4*::GFP+ colonies were counted on day 9 post-transduction.

Acknowledgements

We thank E. Kiskinis, A. Arvanites, S. Bobrowicz, Weisenthal, R. Maehr, A. Kweudjeu, R. Gali, M. Yamaki, E. Massassa, R. Martinez, K. and Rosowski for technical assistance and S. Sullivan, K. Niakan, K. Rodolfa, S. Mekhoubad, I. Tabansky, C. Sasaki, D. Melton, and S. Chen for helpful discussions. This work was made possible by support provided by the Harvard Stem Cell Institute to L.R. and K.E. and by support from the NIH grant R01 HD046732-01A1 to K.E. E.S. and D.L. acknowledge support from the NIH (GM065400) and from the Howard Hughes Medical Institute. J.K.I. and F.P.D. are New York Stem Cell Foundation postdoctoral fellows. D.H. is a Helen Hay Whitney postdoctoral fellow. K.E. is a fellow of the John D. and Catherine T. MacArthur Foundation.

The authors are filing a patent based on the results reported in this paper.

References

Amabile, G., and Meissner, A. (2009). Induced pluripotent stem cells: current progress and potential for regenerative medicine. Trends Mol Med *15*, 59-68.

Attisano, L., and Wrana, J.L. (2002). Signal transduction by the TGF-beta superfamily. Science *296*, 1646-1647.

Boiani, M., Kehler, J., and Scholer, H.R. (2004). Activity of the germline-specific Oct4-GFP transgene in normal and clone mouse embryos. Methods Mol Biol *254*, 1-34.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Chambers, I., Colby, D., Robertson, M., Nichols, J., Lee, S., Tweedie, S., and Smith, A. (2003). Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. Cell *113*, 643-655.

Chambers, I., Silva, J., Colby, D., Nichols, J., Nijmeijer, B., Robertson, M., Vrana, J., Jones, K., Grotewold, L., and Smith, A. (2007). Nanog safeguards pluripotency and mediates germline development. Nature *450*, 1230-1234.

Gellibert, F., Woolven, J., Fouchet, M.H., Mathews, N., Goodland, H., Lovegrove, V., Laroze, A., Nguyen, V.L., Sautet, S., Wang, R.*, et al.* (2004). Identification of 1,5-naphthyridine derivatives as a novel series of potent and selective TGF-beta type I receptor inhibitors. J Med Chem *47*, 4494-4506.

Hacein-Bey-Abina, S., Von Kalle, C., Schmidt, M., McCormack, M.P., Wulffraat, N., Leboulch, P., Lim, A., Osborne, C.S., Pawliuk, R., Morillon, E.*, et al.* (2003). LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. Science *302*, 415-419.

Hanke, J.H., Gardner, J.P., Dow, R.L., Changelian, P.S., Brissette, W.H., Weringer, E.J., Pollok, B.A., and Connelly, P.A. (1996). Discovery of a novel, potent, and Src family-selective tyrosine kinase inhibitor. Study of Lck- and FynT-dependent T cell activation. J Biol Chem *271*, 695-701.

Hong, H., Takahashi, K., Ichisaka, T., Aoi, T., Kanagawa, O., Nakagawa, M., Okita, K., and Yamanaka, S. (2009). Suppression of induced pluripotent stem cell generation by the p53-p21 pathway. Nature *460*, 1132-1135.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008a). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol *26*, 795-797.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008b). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol.

Inman, G.J., Nicolas, F.J., Callahan, J.F., Harling, J.D., Gaster, L.M., Reith, A.D., Laping, N.J., and Hill, C.S. (2002). SB-431542 is a potent and specific inhibitor of transforming growth factor-beta superfamily type I activin receptor-like kinase (ALK) receptors ALK4, ALK5, and ALK7. Mol Pharmacol *62*, 65-74.

Kaji, K., Norrby, K., Paca, A., Mileikovsky, M., Mohseni, P., and Woltjen, K. (2009). Virus-free induction of pluripotency and subsequent excision of reprogramming factors. Nature.

Kim, D., Kim, C., Moon, J., Chung, Y., Chang, M., Han, B., Ko, S., Yang, E., Cha, K.Y., Lanza, R., KIm, K. (2009). Generation of Human Induced Pluripotent Stem Cells by Direct Delivery of Reprogramming Proteins. Cell Stem Cell *4*.

Marson, A., Foreman, R., Chevalier, B., Bilodeau, S., Kahn, M., Young, R.A., and Jaenisch, R. (2008). Wnt signaling promotes reprogramming of somatic cells to pluripotency. Cell Stem Cell *3*, 132-135.

Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A.*, et al.* (2007). Pluripotency governed by *Sox2* via regulation of *Oct3/4* expression in mouse embryonic stem cells. Nature Cell Biology *9*, 11.

Meissner, A., Wernig, M., and Jaenisch, R. (2007). Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nat Biotechnol *25*, 1177-1181.

Mikkelsen, T.S., Hanna, J., Zhang, X., Ku, M., Wernig, M., Schorderet, P., Bernstein, B.E., Jaenisch, R., Lander, E.S., and Meissner, A. (2008). Dissecting direct reprogramming through integrative genomic analysis. Nature *454*, 49-55.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2008). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol *26*, 101-106.

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

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of Mouse Induced Pluripotent Stem Cells Without Viral Vectors. Science.

Shi, Y., Desponts, C., Do, J.T., Hahm, H.S., Scholer, H.R., and Ding, S. (2008a). Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. Cell Stem Cell *3*, 568-574.

Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Scholer, H.R., and Ding, S. (2008b). A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell *2*, 525-528.

Silva, J., Barrandon, O., Nichols, J., Kawaguchi, J., Theunissen, T.W., and Smith, A. (2008). Promotion of reprogramming to ground state pluripotency by signal inhibition. PLoS Biol *6*, e253.

Silva, J., Chambers, I., Pollard, S., and Smith, A. (2006). Nanog promotes transfer of pluripotency after cell fusion. Nature *441*, 997-1001.

Silva, J., Nichols, J., Theunissen, T.W., Guo, G., van Oosten, A.L., Barrandon, O., Wray, J., Yamanaka, S., Chambers, I., and Smith, A. (2009). Nanog is the gateway to the pluripotent ground state. Cell *138*, 722-737.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364-377.

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. (2008). Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell *2*, 230-240.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F.H., Rodriguez-Esteban, C., and Izpisua Belmonte, J.C. (2006). Nanog binds to Smad1 and blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. Proc Natl Acad Sci U S A *103*, 10294-10299.

Takahashi, K., Okita, K., Nakagawa, M., and Yamanaka, S. (2007a). Induction of pluripotent stem cells from fibroblast cultures. Nat Protoc *2*, 3081-3089.

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

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.

Thrasher, A.J.a.G., H.B. (2007). Severe adverse event in clinical trial of gene therapy for X-SCID. http://wwwasgtorg/UserFiles/XSCIDstatementpdf.

Utikal, J., Maherali, N., Kulalert, W., and Hochedlinger, K. (2009). Sox2 is dispensable for the reprogramming of melanocytes and melanoma cells into induced pluripotent stem cells. J Cell Sci *122*, 3502-3510.

Vallier, L., Mendjan, S., Brown, S., Chng, Z., Teo, A., Smithers, L.E., Trotter, M.W., Cho, C.H., Martinez, A., Rugg-Gunn, P.*, et al.* (2009). Activin/Nodal signalling maintains pluripotency by controlling Nanog expression. Development *136*, 1339-1349.

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.

Xu, R.H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R.*, et al.* (2008). NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell *3*, 196-206.

Ying, Q.L., Nichols, J., Chambers, I., and Smith, A. (2003). BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. Cell *115*, 281-292.

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.

Chapter 3

Conversion of Mouse and Human Fibroblasts into Functional Spinal Motor Neurons

Addendum

A portion of this chapter is published as:
Esther Y. Son*, Justin K. Ichida*, Brian J. Wainger, Jeremy S. Toma, Victor F. Rafuase, Clifford J. Woolf, and Kevin Eggan (2011). Conversion of Mouse and Human Fibroblasts into Functional Spinal Motor Neurons. Cell Stem Cell 9: 205-218.

EYS, JKI and KE conceived the study and designed the experiments. EYS and JKI collaborated equally on all experiments and generated all iMNs (see below). BJW and CJW performed electrophysiological recordings. JST and VFR performed chick myotube cocultures and *in ovo* transplantation.

For mouse iMN reprogramming: EYS cloned the pMXs constructs; JKI harvested MEFs and EYS harvested glial cells; EYS and JKI performed the series of transductions for identifying the minimum set of iMN-inducing factors, as well as immunostaining; using cells generated in collaboration, EYS performed microarrays and analyzed the data with the help of JKI and BJW; JKI performed RT-PCRs; for C2C12 co-cultures, EYS prepared the cultures, and EYS and JKI quantified the contractions; for ALS disease modeling, EYS cultured the cells and JKI analyzed the results; EYS performed BrdU incorporation; JKI generated triple transgenic MEFs for the lineage tracing experiment, and EYS cultured the cells and analyzed the results.

For human iMN reprogramming: JKI and EYS derived and expanded the fibroblasts, and JKI derived new patient-specific iPS lines; JKI modified the fibroblasts with mCAT and performed the initial pMXs transductions; JKI and EYS cultured the retrovirally transduced human iMN cultures, and JKI performed vChAT staining; JKI cloned the FUW-tetO constructs and EYS cloned the pHAGE constructs; JKI, with the help of EYS, produced the lentivirus, except for pHAGE virus which was made at Harvard Gene Therapy Core; EYS, with the help of JKI, cultured the human iMNs; EYS performed and quantified TUJ1, MAP2 and HB9 stains as well as reporter expression.

Abstract

The mammalian nervous system is composed of a multitude of distinct neuronal subtypes, each with its own phenotype and differential sensitivity to degenerative disease. Although specific neuronal types can be isolated from rodent embryos or engineered from stem cells for translational studies, transcription factor mediated reprogramming might provide a more direct route to their generation. Here we report that the forced expression of select transcription factors is sufficient to convert mouse and human fibroblasts into induced motor neurons (iMNs). iMNs displayed a morphology, gene expression signature, electrophysiology, synaptic functionality, *in vivo* engraftment capacity and sensitivity to degenerative stimuli, similar to embryo-derived motor neurons. We show that the converting fibroblasts do not transit through a proliferative neural progenitor state, and thus form *bona fide* motor neurons *via* a route distinct from embryonic development. Importantly, we have generated functional human ALS patient-specific iMNs. Our findings demonstrate that fibroblasts can be converted directly into a specific differentiated and functional neural subtype, the spinal motor neuron, with potential applications in regenerative medicine.

Introduction

The mammalian central nervous system (CNS) is assembled from a diverse collection of neurons, each with its own unique properties. These discrete characteristics underlie the proper integration and function of each neuron within the circuitry of the brain and spinal cord. However, their individual qualities also render particular neurons either resistant or sensitive to particular degenerative stimuli. Thus, for each neurodegenerative disease, a stereotyped set of neuronal subtypes is destroyed, causing the hallmark presentation of that condition. Therefore, if we are to comprehend the mechanisms that underlie the development, function and degeneration of the CNS, we must first deeply understand the properties of individual neuronal subtypes.

Physiological and biochemical studies of individual neuronal types have been greatly facilitated by the ability to isolate distinct classes of neurons and interrogate them *in vitro*. Most studies have focused on neurons isolated from the developing rodent CNS. However, it is not routinely possible to isolate analogous populations of human neurons or to isolate and fully study differentiated central neurons. Pluripotent stem cells, such as embryonic stem cells (ESCs), may provide an inexhaustible reservoir of diverse neural subtypes, offering an attractive approach for *in vitro* studies (Wichterle et al., 2002). Although stem cells have shown great promise, to date, only a handful of neural subtypes have been produced in this way. Furthermore, in many cases the neuronal populations produced from stem cells have not been shown to possess refined subtype specific properties and may only superficially resemble their counterparts from the CNS (Peljto and Wichterle, 2011).

Experiments using the reprogramming of one set of differentiated cells directly into another suggest an alternative approach for the generation of precisely defined neural subtypes. Using distinct sets of transcription factors, it is possible to reprogram fibroblasts into pluripotent stem cells (Takahashi and Yamanaka, 2006), blood progenitors (Szabo et al., 2010), cardiomyocytes (Ieda et al., 2010) as well as functional, post-mitotic neurons (Caiazzo et al.; Pfisterer et al., 2011; Vierbuchen et al., 2010). We have therefore considered the idea that by using factors acting on cells intrinsically, rather than relying on morphogens that act extrinsically, it might be possible to more precisely specify the exact properties of a wide array of neuronal types. Most reprogramming studies have so far only produced induced neurons (iNs) with an unknown developmental ontogeny and a generic phenotype (Pang et al., 2011; Pfisterer et al., 2011; Vierbuchen et al., 2010). Recently, two studies have generated cells that resemble dopaminergic neurons based on the production of tyrosine hydroxylase (Caiazzo et al.; Pfisterer et al., 2011). However, it is unclear whether these cells are molecularly and functionally equivalent to embryo- or ESC-derived dopaminergic neurons. In particular, it has yet to be determined whether any type of neuron made by reprogramming can survive and properly integrate into the CNS. If neuronal reprogramming is to be successfully applied to the study of CNS function or degeneration, then it must be capable of producing specific neuronal types that possess the correct phenotypic properties both *in vitro* and *in vivo*.

To determine whether transcription factors can bestow a precise neural subtype identity, we sought factors that could reprogram fibroblasts into spinal motor neurons. Motor neurons control the contraction of muscle fibers actuating movement. Damage to motor neurons caused by either injury or disease can result in paralysis or death;

consequently, there is significant interest in understanding how motor neurons regenerate after nerve injury and why they are selective targets of degeneration in diseases such as spinal muscular atrophy (SMA) and amyotrophic lateral sclerosis (ALS). We therefore attempted induction of motor neurons both because of their significant translational utility and because the developmental origins and functional properties of this neural subtype are among the most well understood.

Here we show that when mouse fibroblasts express factors previously found to induce reprogramming toward a generic neuronal phenotype (Vierbuchen et al., 2010), they also respond to components of the transcription factor network that act in the embryo to confer a motor neuron identity on committed neural progenitors. Thus, we found that forced expression of these transcription factors converted mouse fibroblasts into induced motor neurons (iMNs). Importantly, we found that the resulting iMNs had a gene expression program, electrophysiological activity, synaptic functionality, *in vivo* engraftment capacity and sensitivity to disease stimuli that are all indicative of a motor neuron identity. We also show that the converting fibroblasts do not transition through a proliferative neural progenitor state before becoming motor neurons, indicating they are formed in a manner that is distinct from embryonic development. Finally, we demonstrate that this same approach can convert human fibroblasts into motor neurons; specifically, human ALS patient-specific iMNs have been generated and may serve as useful substrates for studying disease.

Results

Eleven Factors Convert Fibroblasts into Hb9::*GFP+ Cells with Neuronal Morphologies*

We hypothesized that transcription factors known to instruct motor neuron formation during development might also facilitate the conversion of other cell types into motor neurons. To test this idea, we used the literature to select eight candidate transcription factors that participate in varied stages of motor neuron specification (Jessell, 2000). In order to potentially aid the transition toward a neuronal phenotype, we supplemented the motor neuron specification factors with three factors that convert fibroblasts into induced neurons (iNs) of a generic character (*Ascl1*, *Brn2* and *Myt1l*) (Vierbuchen et al., 2010) (Figure 3.1A).

For reprogramming studies, we used mouse embryonic fibroblasts (MEFs) harvested from *Hb9::*GFP mouse embryos at day E12.5, allowing spinal motor neuron conversion to be monitored. Prior to use, cultures of MEFs were carefully screened for the absence of any contaminating GFP+ cells. First, we asked whether the action of the three iN factors alone could generate *Hb9::*GFP+ cells by transducing MEFs with retroviral vectors encoding *Ascl1*, *Brn2* and *Myt1l* (Figure 3.1A). Although cells with a neuronal morphology were observed, as previously reported (Vierbuchen et al., 2010), no *Hb9::*GFP+ cells emerged, even after 35 days (Figure 3.1B). This suggests that the iN factors alone do not generate motor neurons, consistent with the report that cholinergic neurons were not generated by these factors (Vierbuchen et al., 2010).

We next tested whether the eight motor neuron specification factors we selected could induce motor neurons in the absence of the three iN factors. Based on titering with a control virus encoding GFP, we determined that each factor was expressed in >95% of

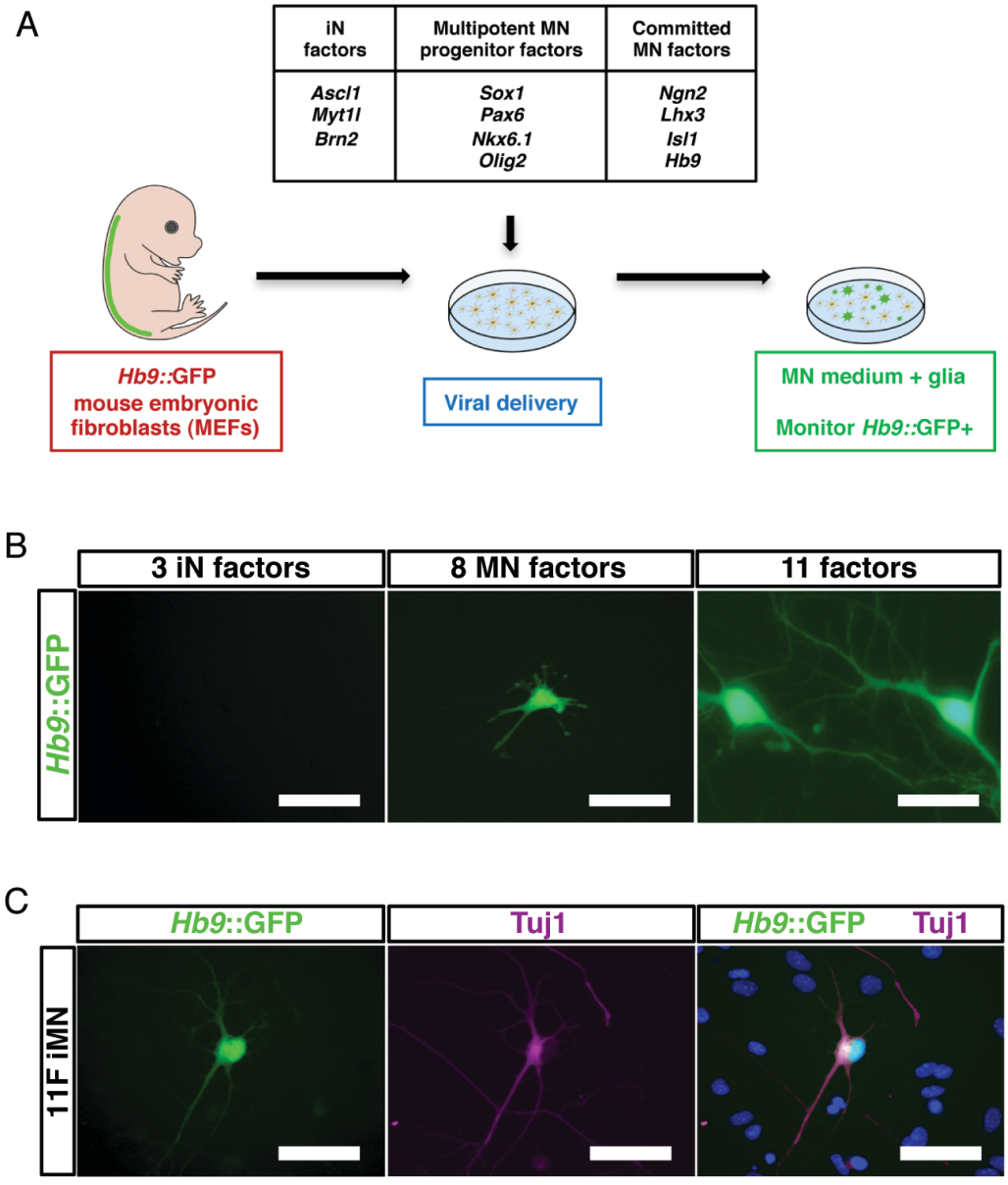

Figure 3.1.

Induction of *Hb9::*GFP+ Neurons from Mouse Fibroblast Cultures. (A) Experimental outline. 11 candidate transcription factors include eight developmental genes in addition to the three iN factors. (B) *Hb9::*GFP+ cells are generated from MEFs by transduction with 8 or 11 factors by day 35 post-transduction, but more efficiently by 11 factors. Scale bars represent 50 µm. (C) *Hb9::*GFP+ neurons express Tuj1 (purple). Scale bars represent 40 µm.

the fibroblasts. Encouragingly, a small number of *Hb9::*GFP+ cells were observed at 35 days post-transduction; however, they did not possess a normal neuronal morphology (Figure 3.1B). We therefore next asked whether the two sets of factors, iN factors and motor neuron specification factors, together could synergize to produce motor neurons. Indeed, when the aggregate set of 11 factors was transduced into fibroblasts, a significant number of *Hb9::*GFP+ cells emerged, which elaborated complex processes and all of which expressed a neuronal form of tubulin (n=50) (Figure 3.1C). We preliminarily designated these *Hb9::*GFP+ cells, induced motor neurons (iMNs).

iMNs Are Efficiently Induced by Seven Factors

To determine which of the 11 factors were necessary for generating iMNs, we omitted each gene one at a time (Figure 3.2A). Excluding either *Lhx3* or *Ascl1* eliminated iMN formation. However, reprogramming efficiency was either only slightly reduced or unchanged when each of the remaining factors were removed (Figure 3.2A). Interestingly, we observed that ectopic expression of *Hb9* was not required for iMN formation (Figure 3.2A), suggesting that, at least in that case, exogenous *Hb9* was not simply transactivating its own promoter. Similarly, we observed Isl1/2 expression by immunostaining in iMNs (80.6%, n=36), even when the *Isl1* retrovirus was omitted from the transduction (Figures 3.2A-C).

Although *Lhx3* and *Ascl1* seemed necessary for reprogramming, they were not sufficient to induce motor neuron formation (Figure 3.2D). However, when *Lhx3* was combined with the three iN factors (*Ascl1*, *Brn2* and *Myt1l*), we observed a modest number of *Hb9::GFP*+ cells (Figure 3.2E). Because these four factors could not

Figure 3.2.

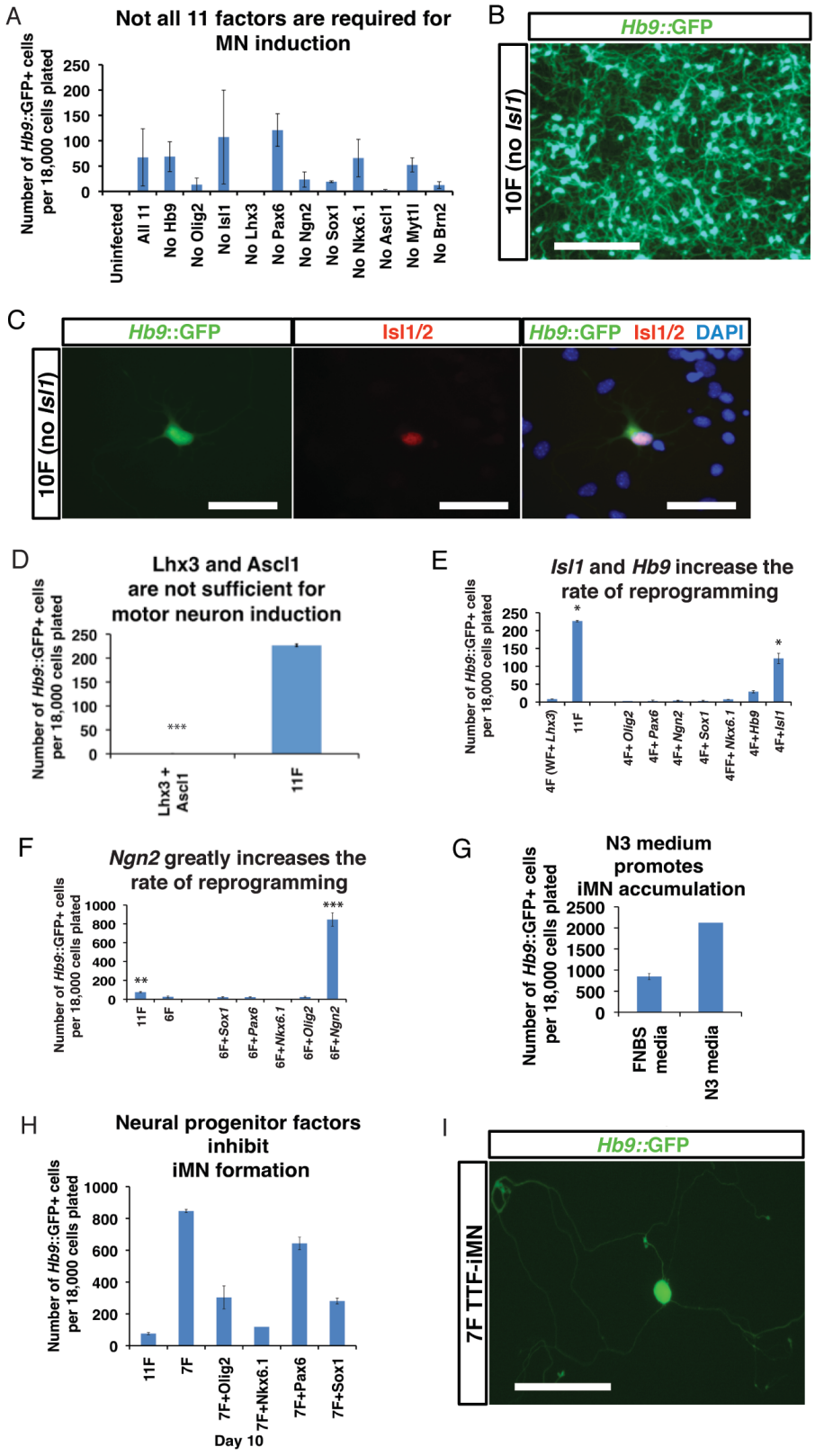
A
Not all 11 factors are required for MN induction
Number of Hb9::GFP+ cells per 18,000 cells plated
250
200
150
100
50
0
Uninfected
All 11
No Hb9
No Olig2
No Isl1
No Lhx3
No Pax6
No Ngn2
No Sox1
No Nkx6.1
No Ascl1
No Myt1l
No Brn2
B
Hb9::GFP
10F (no Isl1)
C
Hb9::GFP
Isl1/2
Hb9::GFP Isl1/2 DAPI
10F (no Isl1)
D
Lhx3 and Ascl1 are not sufficient for motor neuron induction
Number of Hb9::GFP+ cells per 18,000 cells plated
250
200
150
100
50
0
***
Lhx3 + Ascl1
11F
E
Isl1 and Hb9 increase the rate of reprogramming
Number of Hb9::GFP+ cells per 18,000 cells plated
250
200
150
100
50
0
*
*
4F (WF+ Lhx3)
11F
4F+ Olig2
4F+ Pax6
4F+ Ngn2
4F+ Sox1
4FF+ Nkx6.1
4F+Hb9
4F+Isl1
F
Ngn2 greatly increases the rate of reprogramming
Number of Hb9::GFP+ cells per 18,000 cells plated
1000
800
600
400
200
0
**
***
11F
6F
6F+Sox1
6F+Pax6
6F+Nkx6.1
6F+Olig2
6F+Ngn2
G
N3 medium promotes iMN accumulation
Number of Hb9::GFP+ cells per 18,000 cells plated
2500
2000
1500
1000
500
0
FNBS media
N3 media
H
Neural progenitor factors inhibit iMN formation
Number of Hb9::GFP+ cells per 18,000 cells plated
1000
800
600
400
200
0
11F
7F
7F+Olig2
7F+Nkx6.1
7F+Pax6
7F+Sox1
Day 10
I
Hb9::GFP
7F TTF-iMN

Figure 3.2. (Continued)
iMNs Are Efficiently Generated by 7 Factors. (A) Efficiency of reprogramming 35 days post-transduction when each factor is omitted from the 11-factor pool individually. Error bars indicate ±s.d. (B) *Isl1* is dispensable for generating iMNs. Scale bar represents 200 μm. (C) iMNs generated with 10 factors (without *Isl1*) express endogenous Islet (red). Scale bars represent 40 μm. (D) *Lhx3* and *Ascl1* are not sufficient to convert fibroblasts into motor neurons. Error bars indicate ±s.d. *** $p<0.001$ (Student's *t*-test, two-tailed). (E) Reprogramming efficiency is greater with *Hb9* or *Isl1* on top of 4 factors (*Lhx3*, *Ascl1*, *Brn2* and *Myt1l*) at day 21 post-transduction. Error bars indicate ±s.d. *$P<0.05$ (Student's *t*-test, two-tailed). (F) Addition of *Ngn2* to the 6-factor pool (*Hb9*, *Isl1*, *Lhx3*, *Ascl1*, *Brn2* and *Myt1l*) greatly enhances reprogramming efficiency as seen 10 days after transduction. Error bars indicate ±s.d. ***$P<0.001$; **$P<0.01$ (Student's *t*-test, two-tailed). (G) Efficiency of fibroblast-to-iMN reprogramming in two different media conditions. N3 medium promotes iMN accumulation. Error bars indicate ±s.d. (H) Adding each of the neural progenitor factors to 7 factors (*Ngn2* + 6 factors) inhibits iMN formation as seen 10 days after transduction. Error bars indicate ±s.d. ***$P<0.001$; **$P<0.01$ (Student's *t*-test, two-tailed).
(I) The 7 iMN factors convert adult tail tip fibroblasts into motor neurons. Scale bar represents 100 μm.

efficiently induce motor neuron formation, we next individually added each of the other factors back to this smaller set (Figure 3.2E). We found that either *Isl1* or *Hb9* were capable of increasing the efficiency of iMN induction, which was further enhanced when *Ngn2* was added to the other 6 factors (Figure 3.2F). Indeed, the efficiency of motor neuron induction with these 7 factors (*Ascl1*, *Brn2*, *Myt1l*, *Lhx3*, *Hb9, Isl1* and *Ngn2*) surpassed the activity of the 11 as a whole and, depending on the culture conditions used, reached between 5% and 10% of the number of MEFs transduced (Figures 3.2F-G). Adding any one of the remaining factors, which are all known to function in earlier stages of motor neuron specification (Lee et al., 2005), dramatically decreased the efficiency of reprogramming by the 7 factors (Figure 3.2H).

We reasoned that, although our apparently homogeneous MEF cultures lacked *Hb9::*GFP+ cells, they could be contaminated with rare embryonic neuronal progenitors that might be more responsive to reprogramming. To rule out the possibility that iMNs originated from such progenitors, we prepared fibroblasts from the tails of adult *Hb9::*GFP mice and transduced them with the optimal set of 7 factors. Again, GFP+ cells with neuronal morphologies emerged (Figure 3.2I), indicating that the ability to respond to the 7 iMN factors was not restricted to cells of an embryonic origin.

iMNs Possess a Motor Neuron Gene Expression Signature

To begin to assess whether iMNs had the known characteristics of cultured embryonic motor neurons, we carefully examined the phenotype of iMNs made with 10 factors (*Isl1* omitted). We found that iMNs were comparable in cell body size and projection length to both E13.5 embryo- and ESC-derived motor neuron controls (Figure 3.3A). To

Figure 3.3

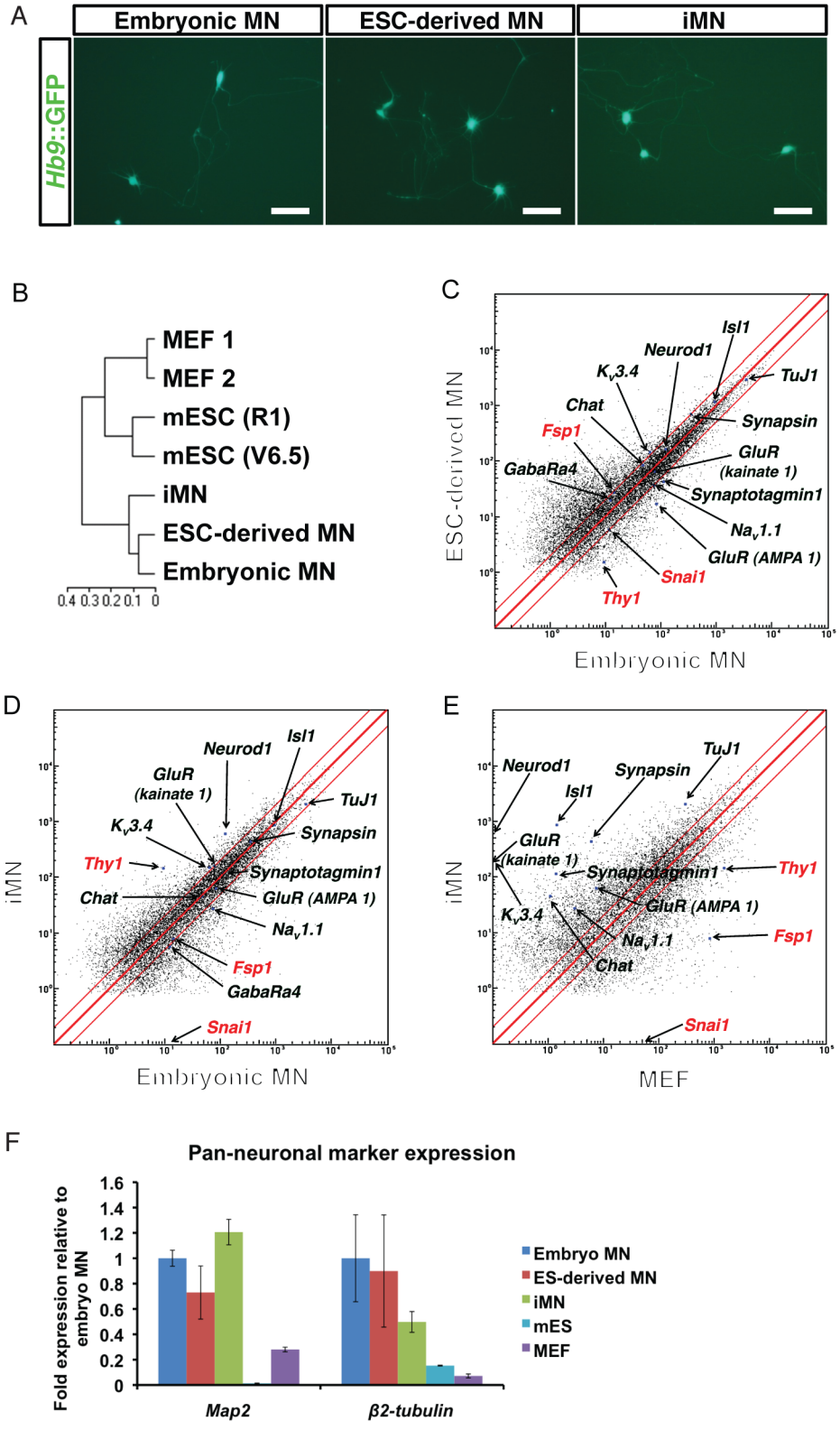
A
Embryonic MN
ESC-derived MN
iMN
Hb9::GFP
B
MEF 1
MEF 2
mESC (R1)
mESC (V6.5)
iMN
ESC-derived MN
Embryonic MN
0.4 0.3 0.2 0.1 0
C
ESC-derived MN
Embryonic MN
Isl1
Neurod1
Kv3.4
TuJ1
Chat
Fsp1
Synapsin
GluR (kainate 1)
GabaRa4
Synaptotagmin1
Nav1.1
GluR (AMPA 1)
Snai1
Thy1
D
iMN
Embryonic MN
Neurod1
Isl1
GluR (kainate 1)
TuJ1
Kv3.4
Synapsin
Thy1
Synaptotagmin1
Chat
GluR (AMPA 1)
Nav1.1
Fsp1
GabaRa4
Snai1
E
iMN
MEF
Neurod1
TuJ1
Synapsin
Isl1
GluR (kainate 1)
Synaptotagmin1
Thy1
Kv3.4
GluR (AMPA 1)
Nav1.1
Fsp1
Chat
Snai1
F
Pan-neuronal marker expression
Fold expression relative to embryo MN
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
Map2
β2-tubulin
Embryo MN
ES-derived MN
iMN
mES
MEF

Figure 3.3. (Continued)

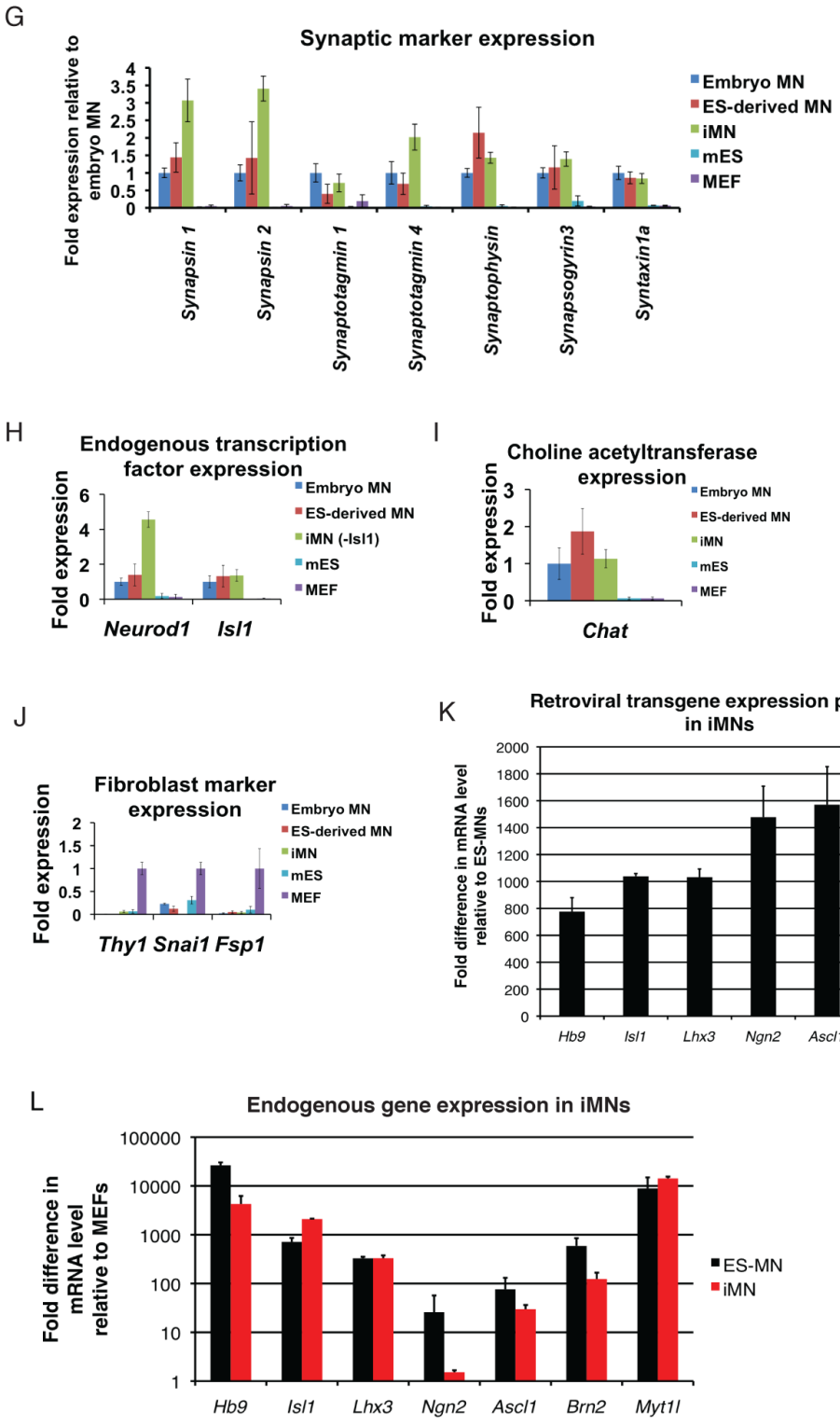

Figure 3.3. (Continued)
iMNs Possess Gene Expression Signatures of Motor Neurons. (A) iMN morphology is similar to that of embryonic and ESC-derived motor neurons. Scale bars represent 100 μm. (B-J) Global transcriptional analysis of FACS-purified *Hb9::*GFP+ motor neurons. (B) iMNs cluster with control motor neurons and away from MEFs. (C-E) Pairwise gene expression comparisons show that iMNs are highly similar to embryo-derived motor neurons and dissimilar from the starting MEFs; black labeling denotes genes expressed in motor neurons, red labeling denotes genes expressed in fibroblasts, and the red lines indicate the diagonal and 2-fold changes between the sample pairs. (F) iMNs express the pan-neuronal genes *Map2* and *b2-tubulin.* (G) iMNs express genes required for synapse formation. (H) iMNs endogenously express transcription factors expressed in motor neurons. (I) iMNs endogenously express choline acetyltransferase. (J) Fibroblast-specific genes are downregulated in iMNs. mRNA expression levels are shown relative to an embryonic motor neuron control (F-I) or relative to a MEF control (J). **(**K**)** qRT-PCR using primers for the viral transcripts of 7 iMN factors. iMNs have not silenced viral transgenes. Expression levels are shown relative to ESC-derived motor neurons. Error bars indicate ±s.d. (L) qRT-PCR data showing expression of endogenous transcripts of the 7 iMN factors relative to their levels in ES-MNs. Error bars indicate ±s.d.

determine how similar overall transcription in iMNs was to control motor neurons, we isolated the three motor neuron types by fluorescence-activated cell sorting (FACS) and performed transcriptional profiling (Figure 3.3B-E). For these analyses, RNAs isolated from MEFs and ESCs were used as negative controls. When we performed hierarchical clustering of the data, iMNs grouped closely to embryonic motor neurons, as did ESC-derived motor neurons (Figure 3.3B). In contrast, iMNs were very distinct from the initial MEF population. Thus, our results suggest that transduction of MEFs with these transcription factors results in a global shift towards a motor neuron transcriptional program.

When we examined the transcription of specific neuronal genes, we again found that iMNs and control motor neurons were very similar. Relative to either MEFs or ESCs, iMNs and both types of control motor neurons expressed elevated levels of β2-tubulins (*Tubb2a* and *Tubb2b*) and *Map2* (Figures 3.3C-F), as well as synaptic components such as synapsins (*Syn1* and *Syn2*), synaptophysin (*Syp*) and synaptotagmins (*Syt1*, *Syt4*, *Syt13* and *Syt 16*) (Figures 3.3C-E, G). iMNs also expressed known motor neuron transcription factors that were not provided exogenously (*NeuroD* and *Isl1*) (Figures 3.3C-E, H), as well as the gene encoding the enzyme cholineacetyltransferase (*ChAT*) (Figures 3.3C-E, I). In contrast, iMNs had downregulated the fibroblast program as exemplified by reduced transcription *of Snai1*, *Thy1* and *Fsp1* (Figure 3.3E and Figure 3.3J). Immunostaining confirmed that the iMNs expressed Map2 (100%, n=120) (Figure 3.4A), synapsin (Figure 3.4B), and vesicular ChAT (97.6%, n=124) (Figure 3.4C), indicating that they had indeed activated the enzymatic pathways for producing acetylcholine (ACh), the neurotransmitter released by motor neurons, and suggesting they should be

Figure 3.4.

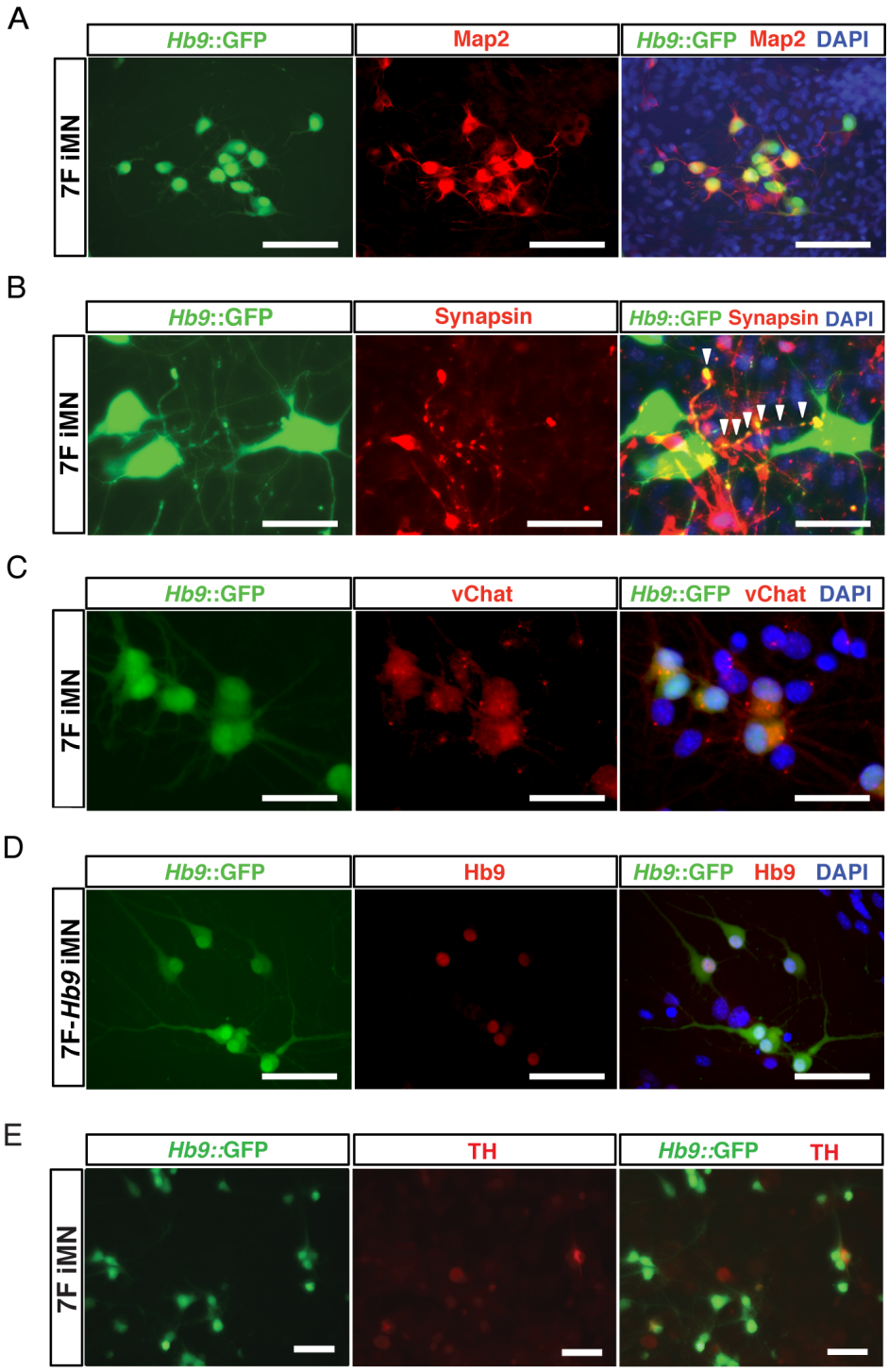
A
Hb9::GFP
Map2
Hb9::GFP Map2 DAPI
7F iMN
B
Hb9::GFP
Synapsin
Hb9::GFP Synapsin DAPI
7F iMN
C
Hb9::GFP
vChat
Hb9::GFP vChat DAPI
7F iMN
D
Hb9::GFP
Hb9
Hb9::GFP Hb9 DAPI
7F-Hb9 iMN
E
Hb9::GFP
TH
Hb9::GFP TH
7F iMN

Figure 3.4. (Continued)
iMNs Express Neuronal and Motor Neuron Proteins. (A) iMNs express the pan-neuronal marker Map2 (red). Scale bars represent 100 μm. (B) iMNs express synapsin (red). Scale bars represent 20 μm. (C) iMNs express vesicular cholineacetyltransferase (vChAT, red). Scale bars represent 40 μm. (D) iMNs generated without the motor neuron-selective transcription factor *Hb9* express the Hb9 protein (red). Scale bars represent 80 μm. (E) iMNs rarely express tyrosine hydroxylase (TH). A rare TH+ iMN with a low level of *Hb9::*GFP reporter expression is shown on the right. Scale bars represent 50 μm.

capable of forming functional synapses. In contrast, the vast majority of iMNs did not express tyrosine hydroxylase (3%, n=150) (Figure 3.4E), suggesting that they were not of a mixed neuronal character.

In order to determine if the iMNs truly adopted a new cellular identity through transdifferentiation, we performed qRT-PCR analysis to ask if they established an endogenous program of motor neuron gene expression (Table 3.1). As expected for a somatic cell type such as the motor neuron, the retroviral transgenes used for reprogramming were not silenced in the iMNs (Figure 3.4F), leaving it unclear as to whether the endogenous loci of these motor neuron genes had been activated. When we quantified the endogenous mRNA levels of the motor neuron-specific genes used for conversion, we found that all 7 transcription factors were expressed at levels similar to those in ESC-derived motor neurons (Figure 3.4G). Furthermore, immunostaining revealed that iMNs created without exogenous Hb9 still activated expression of this important transcription factor from the endogenous locus (87.9%, n=149) (Figure 3.4D). Together, these data indicate that the iMNs we produced had established a transcriptional program characteristic of motor neurons.

iMNs Possess Electrophysiological Characteristics of Motor Neurons

In order to determine if MEF- and tail tip fibroblast-derived iMNs possessed the electrophysiological properties of motor neurons, we performed whole-cell patch clamp recordings. The average resting membrane potential for iMNs was -49.5 mV (SEM 5.6, n=6), which was similar to that for control ESC-derived motor neurons (-50.5 mV, SEM 3.5, n=13). Depolarizing voltage steps in voltage clamp elicited fast inward currents

Gene	Forward Primer	Reverse Primer
Ascl1	CCAACTACTCCAACGACT	GGAGAGCCTGGCAGGTCC
Brn2	GCGCCGAGGATGTGTATG	AGGAAAGACTGTGGACC
Hb9	ACAACTTCCCGTACAGCAAT	CTTCCGCCCTGGAGGCAA
Isl1	GCGACATAGATCAGCCTGC	CATCTGAATGAATGTTCC
Lhx3	CCCCCACCCATGAGGGTGCT	GAGCCAGGGGAAGCAGAGGC
Myt1l	CGTGACTACTTTGACGGA	TCACCACTAGAGCAGCTGT
Ngn2	GCGTCATCCTCCAACTCC	AGAGGGAGACCCGCAGCT
Viral LTR	N/A	TTTGTACAAGAAAGCTGGGT

Table 3.1.
RT-PCR Primer Sequences.

followed by slow outward currents, consistent with the opening of voltage-activated sodium and potassium channels, respectively (Figures 3.5A-B, L). The inward current was blocked by addition of 500 nM tetrodotoxin (TTX), a potent antagonist of TTX-sensitive voltage-activated sodium channels (Figure 3.5C). A defining feature of a neuron is its ability to fire action potentials. In current clamp experiments with iMNs, depolarizing current steps produced single or multiple action potentials (90%, n=10), with overshoot, after-hyperpolarizations and a firing frequency similar to that reported for ESC-derived motor neurons and rat embryonic motor neurons (Alessandri-Haber et al., 1999) (Figures 3.5D-E, M-N).

We next tested whether iMNs express functional receptors for the excitatory and inhibitory neurotransmitters that normally act on motor neurons. As might be expected given the known receptor subunit transitions associated with development of immature neurons to a fully differentiated state, certain agonists yielded responses in some but not all neurons. Glycine and GABA are the major inhibitory neurotransmitters, and their ionotropic activity is mediated by opening chloride channels. Addition of 100 μM glycine (44.4%, n=9, Figure 3.5F) or GABA (72.7%, n=11, Figures 3.5G, O) elicited inward currents when cells were held at -80 mV. We also evaluated the response of iMNs to fast excitatory glutamatergic neurotransmitters and observed a strong response to the receptor agonist kainate (80%, n=15 cells, Figures 3.5H, P).

Consistent with our physiological analyses, and similar to control embryonic motor neurons and motor neuron populations described previously (Cui et al., 2006), the iMNs transcribed the genes encoding α and β subunits of voltage-gated sodium channels (Figures 3.3D-E and Figure 3.5I), as well as members of the Shaker-, Shaw-, and Eag-

Figure 3.5.

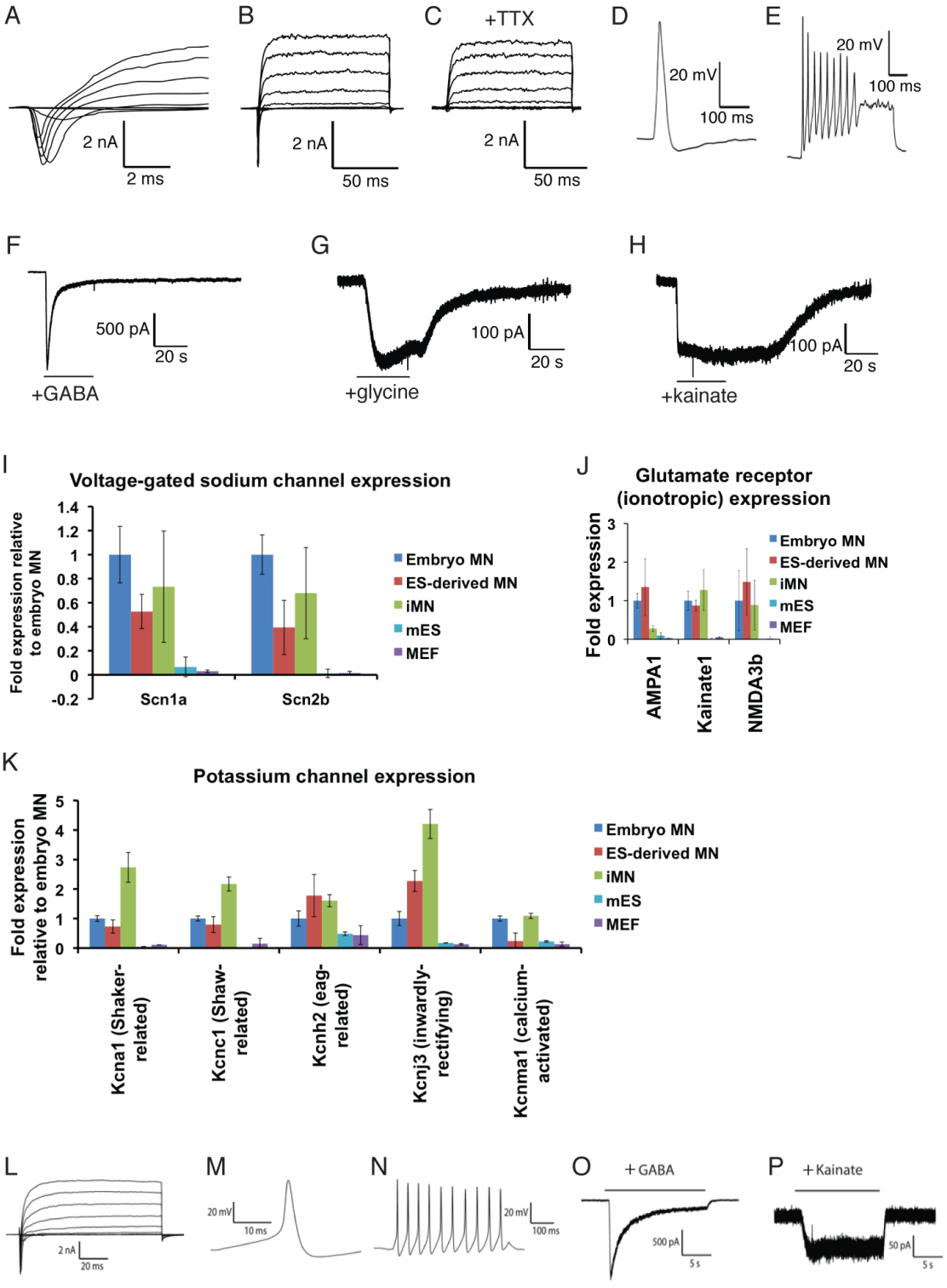
A
2 nA
2 ms
B
2 nA
50 ms
C
+TTX
2 nA
50 ms
D
20 mV
100 ms
E
20 mV
100 ms
F
500 pA
20 s
+GABA
G
100 pA
20 s
+glycine
H
100 pA
20 s
+kainate
I
Voltage-gated sodium channel expression
Fold expression relative to embryo MN
1.4
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
Scn1a
Scn2b
Embryo MN
ES-derived MN
iMN
mES
MEF
J
Glutamate receptor (ionotropic) expression
Fold expression
3
2
1
0
AMPA1
Kainate1
NMDA3b
Embryo MN
ES-derived MN
iMN
mES
MEF
K
Potassium channel expression
Fold expression relative to embryo MN
5
4
3
2
1
0
Kcna1 (Shaker-related)
Kcnc1 (Shaw-related)
Kcnh2 (eag-related)
Kcnj3 (inwardly-rectifying)
Kcnma1 (calcium-activated)
Embryo MN
ES-derived MN
iMN
mES
MEF
L
2 nA
20 ms
M
20 mV
10 ms
N
20 mV
100 ms
O
+GABA
500 pA
5 s
P
+Kainate
50 pA
5 s

Figure 3.5. (Continued)
Electrophysiological Activity of iMNs. (A-H) MEF-derived iMNs are electrophysiologically active. (A-B) MEF-derived iMNs express functional sodium and potassium channels (B) iMN sodium channel activity is appropriately blocked by tetrodotoxin (TTX). (D-E) iMNs fire a single and multiple action potentials upon depolarization. (F) 100 μM GABA induces inward currents in iMNs. (G) 100 μM glycine induces inward currents in iMNs. (H) 100 μM kainate induces inward currents in iMNs. (I-K) MEF-derived iMNs express genes required for ion channel function and neurotransmitter response. mRNA expression levels relative to an embryo-derived motor neuron control are shown for (I) sodium channel, (J) potassium channel, and (K) glutamate receptor genes. All motor neuron samples were FACS-purified by *Hb9*::GFP expression prior to mRNA extraction. (L-O) Tail tip fibroblast (TTF)-derived iMNs exhibit electrophysiology characteristic of motor neurons. (L) TTF-derived iMNs have functional sodium and potassium channels. (M-N) TTF-derived iMNs fire single and multiple action potentials. (O) 100 μM GABA induces inward currents in TTF-derived iMNs. (P) 100 μM kainate induces inward currents in TTF-derived iMNs.

related, inwardly rectifying, and calcium-activated families of potassium channels (Figures 3.3D-E and Figure 3.5K). In addition, iMNs transcribed genes encoding the receptor components required for responding to the neurotransmitter glutamate (Figures 3.3D-E and Figure 3.5J). Together, our physiological and gene expression analyses indicate that iMNs are excitable, generate action potentials and respond to both inhibitory and excitatory neurotransmitters in a manner characteristic of both ESC-derived and embryonic motor neurons.

iMNs Form Functional Synapses with Muscle

Our initial results indicated that iMNs have many of the phenotypic and electrophysiological properties of *bona fide* motor neurons. However, the defining functional characteristic of the spinal motor neuron is its ability to synapse with muscle and, through the release of acetylcholine (ACh), stimulate muscle contraction. To test whether iMNs could form functional neuromuscular junctions (NMJs), we co-cultured FACS-purified iMNs with myotubes derived from the C2C12 muscle cell line. We found that iMNs could establish themselves in these muscle cultures and sent projections along the length of the myotubes (Figure 3.6A).

Strikingly, we observed that several days following the addition of purified iMNs, C2C12 myotubes began to undergo regular and rhythmic contraction (Figure 3.6B). Regular contractions were not seen at this time point in myotubes that were cultured alone or with generic iNs (Table 3.2). To directly test whether the regular contractions of mytotubes were due to synaptic stimulation of ACh receptors, we quantified the frequency of myotube contraction and then added curare to the culture medium. As

Figure 3.6.

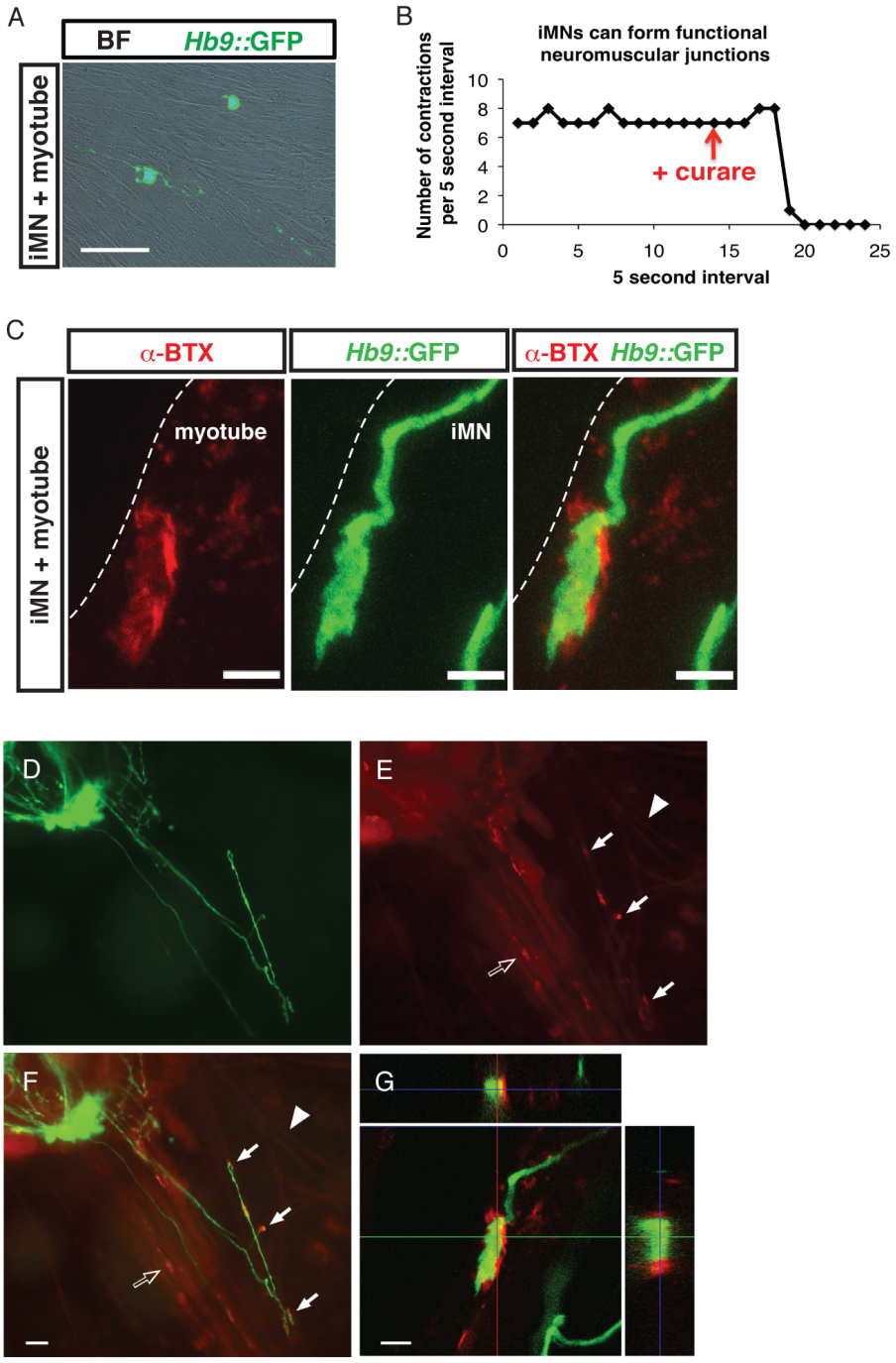
A
BF
Hb9::GFP
iMN + myotube
B
iMNs can form functional neuromuscular junctions
Number of contractions per 5 second interval
10
8
6
4
2
0
0
5
10
15
20
25
5 second interval
+ curare
C
α-BTX
Hb9::GFP
α-BTX Hb9::GFP
iMN + myotube
myotube
iMN
D
E
F
G

Figure 3.6. (Continued)
iMNs Form Functional Neuromuscular Junctions *in Vitro*. (A) iMNs co-cultured with a monolayer of C2C12 myotubes. Scale bar represents 200 μm. (B) iMN-induced contractions of C2C12 myotubes are blocked by 50 μM curare. The arrow indicates the timing of curare addition. (C-G) iMNs induce acetylcholine receptor clustering and form anatomical endplates on cultured myotubes. (C) iMNs cultured with chick myotubes form NMJs with characteristic α-bungarotoxin (α-BTX, red) staining. The dotted line outlines the boundaries of a myotube. Scale bar represents 5 μm. (D-G) A zoomed-out view of the iMN-muscle co-culture. (D) iMNs after 7 weeks of co-culture with chick myotubes. (E) Rhodamine-conjugated a-BTX staining showed ACh clustering occurred on the chick myotubes. (F) Merged image of (D) and (E) showed ACh receptors clustering preferentially occurred near the GFP+ axons (open arrowhead) and at the end of the neurites at putative endplates regions (arrows). ACh clusters were less pronounced on myotubes not associated with axons (arrowhead). (G) Confocal image depicting a GFP+ axon co-localized with acetylcholine receptors at a putative endplate in a 3-week co-culture. Imaging in both the *x–z* and *y–z* orthogonal planes confirms the close proximity of the receptors to the axon terminal. Scale bars represent 50 μm (F) and 5 μm (G).

	Total number of dishes	Dishes with twitching muscle	Dishes with >2 twitching areas
iMN on muscle	5	4	3
iN on muscle	2	0	N/A
Muscle only	10	0	N/A

Table 3.2.
iMNs Induce Contraction of C2C12 Myotubes. FACS-purified iMNs and iNs were plated on top of C2C12 myotubes. At day 10 after the start of co-culture, contraction was observed in myotubes cultured with iMNs. Those cultured alone or with iNs did not exhibit contractions at this time point.

Movie 3.1.
iMNs Induce Contraction of C2C12 Myotubes That Is Blocked by Curare Treatment. Video was filmed 10 days after iMNs were added to the myotube cultures. iMN-dependent contraction was observed for 5 minutes before the video was initiated. 17 seconds after video initiation, a final concentration of 50 μM curare was added to the culture to specifically block acetylcholine receptors on the myotubes. All muscle contraction stopped by 32 seconds after video initiation and did not resume for the remainder of the video, indicating that the contraction was dependent on the activity of the iMNs. Movie provided as a supplementary file.

curare selectively and competitively antagonizes nicotinic ACh receptors, its addition should only inhibit muscle contractions that result from stimulation of such receptors (Figure 3.6B and Movie 1). Shortly after the addition of curare, we observed a precipitous and sustained decline in the frequency of myotube contraction, indicating that the contractions were indeed dependent on the stimulation of ACh receptors.

In order to directly visualize NMJ formation in iMN cultures, we co-cultured iMNs with primary chick myotubes (Figure 3.6C-G). After one week of co-culture, we found that many *Hb9*::GFP+ iMNs survived even following withdrawal of neurotrophic support, suggesting that they had formed synapses with the muscle. Three weeks after co-culture had been initiated, staining with α-bungarotoxin (α-BTX) revealed ACh receptor clustering on the myofibers (Figures 3.6C-G). As occurs in ESC-derived motoneuron/chick myotube cocultures (Miles et al., 2004; Soundararajan et al., 2007), ACh receptors clustered preferentially near the iMN axons, although the clustering was not always clearly opposed to *Hb9*::GFP+ axons. This phenomenon is similar to what occurs during chick (Dahm and Landmesser, 1988) and mouse (Lupa and Hall, 1989) neuromuscular development where receptor clustering first appears near the innervating motor axons, but not always in direct contact. Imaging in the *x-z* and *y-z* orthogonal planes verified that ACh receptors clustered near iMN axons superimposed with the *Hb9*::GFP+ axons (Figure 3.6G). These results indicate that iMNs signal to the post-synaptic muscle fiber to induce appropriate receptor clustering which is necessary for neuromuscular transmission. Together, these data indicate that iMNs can make functional synaptic junctions with muscle.

iMNs Integrate into the Developing Chick Spinal Cord

Transplantation of motor neurons into the developing chick spinal cord provides a rigorous test of their ability to survive *in vivo*, migrate to appropriate engraftment sites in the ventral region of the spinal cord, and to properly respond to axon guidance cues to send their axonal projections out of the spinal cord through the ventral root (Peljto et al., 2011; Soundararajan et al., 2006; Wichterle et al., 2002). In order to test the ability of iMNs to survive and function *in vivo*, we transplanted FACS-purified iMNs or control ESC-derived motor neurons into the neural tube of stage 17 chick embryos at 12-16 days post-transduction (Figure 3.7A). Although the injection of the iMNs along the dorsal-ventral axis was not precisely controlled, we observed that *Hb9*::GFP+ iMNs engrafted in the ventral horn of the spinal cord in the location where endogenous motor neurons reside at stage 31 (Figure 3.7C). Like transplanted ESC-derived motor neurons (Soundararajan et al., 2006; Wichterle et al., 2002), the *Hb9*::GFP+ cells maintained Tuj1 expression and exhibited extensive dendritic arbors (Figure 3.7C). In addition, we asked whether iMNs project their axons out of the CNS. Endogenous and transplanted ESC-derived motor neurons send axonal projections out of the spinal cord through the ventral root towards musculature (Figure 3.7B) (Soundararajan et al., 2010; Wichterle et al., 2002). When *Hb9::*GFP ESCs are subjected to directed differentiation toward motor neurons, the resulting EBs contain both GFP+ motor neurons and distinct, non-motor neuronal subtypes that do not express GFP. In contrast to GFP+ motor neurons, GFP- non-motor neuron subtypes present within the same transplants extend extensive processes whose projections remain restricted to the developing spinal cord and do not exit through the ventral root (Soundararajan et al., 2010). Therefore, the chick transplantation assay can

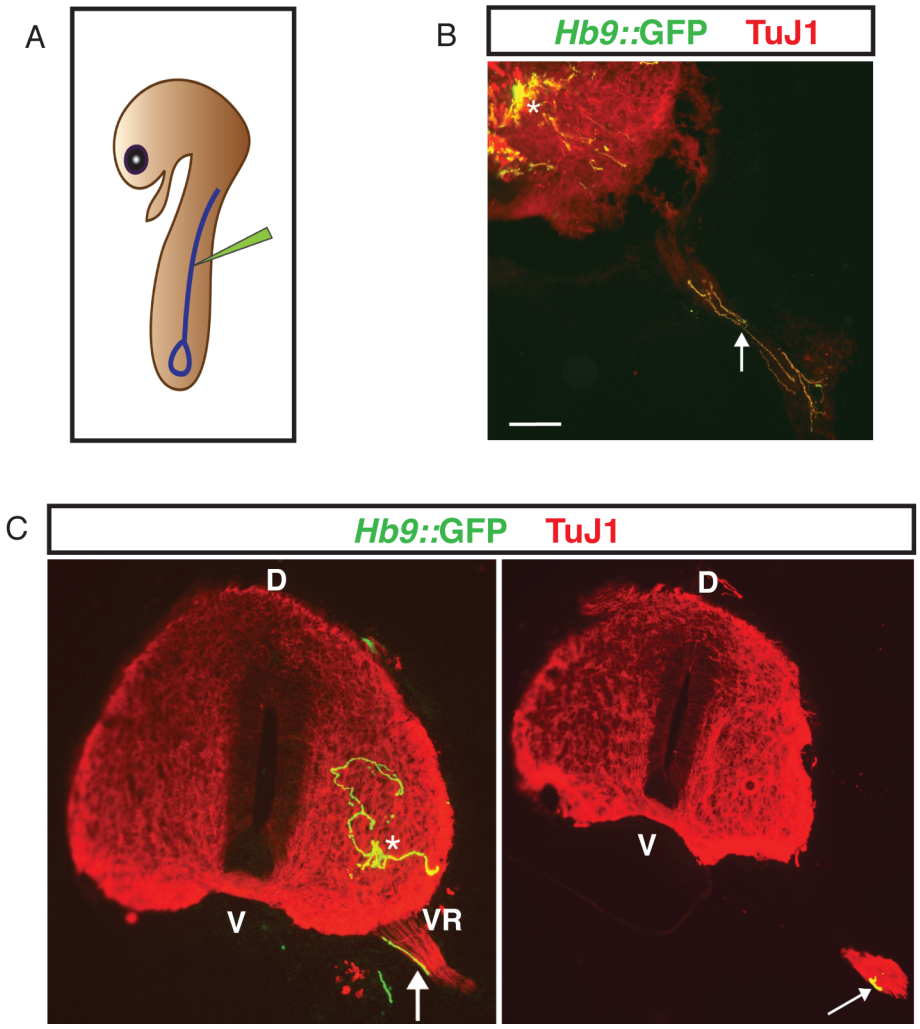

Figure 3.7.
***In Vivo* Engraftment Capacity of iMNs**. (A) Diagram showing the injection of iMNs into the neural tube of the stage 17 chick embryo. (B) Transverse section of chick neural tube injected with ESC-derived motor neurons (asterisks). The motor neurons engrafted into the ventral horn and extended axons out of the spinal cord through the ventral root (arrow) 5 days after transplantation into an E2.5 chick embryo neural tube. Scale bar represents 100 µm. (C) Transverse sections of the iMN-injected chick neural tube 5 day after transplantation. Arrows in both panels indicate the same axon of an iMN exiting the spinal cord through the ventral root. D: dorsal, V: ventral, VR: ventral root.

be used to measure motor neuron-specific axonal pathfinding. Indeed, after transplantation, we often observed *Hb9::*GFP+ iMNs in the ventral horn of the spinal cord, and in 80% (n=5) of these cases, we saw axons of *Hb9*::GFP+ iMNs projecting out of the spinal cord through the ventral root towards the musculature (Figure 3.7C). Thus, their *in vivo* engraftment capacity was similar to that observed for ESC-derived *Hb9::*GFP+ motor neurons (Figure 3.7B). Together, these data demonstrate that iMNs are able to engraft, migrate to appropriate sites of integration, and correctly respond to guidance cues *in vivo,* projecting their axons out of the CNS.

iMNs Are Sensitive to Disease Stimuli

ALS is an invariably fatal neurological condition whose hallmark is the selective and relentless degeneration of motor neurons. We reasoned that if iMNs fully phenocopied *bona fid*e motor neurons, they should also be sensitive to degenerative stimuli thought to contribute to ALS. To determine if this was the case, we co-cultured iMNs with glial cells from the *SOD1*G93A mouse model of ALS. We, and others, have shown that both embryonic and ESC-derived motor neurons are selectively sensitive to the toxic effect of mutant glia, while other neural cell types, such as spinal interneurons, are relatively unaffected (Di Giorgio et al., 2007) (Nagai et al., 2007). iMNs were co-cultured with either wild-type or mutant *SOD1*G93A glia and the number of *Hb9::*GFP+ iMNs quantified 10 days later. As we would expect if iMNs were indeed *bona fide* motor neurons, there was a sharp reduction in the number of iMNs co-cultured with mutant glia relative to those cultured with wild-type glia (Figures 3.8A-B), and the effect was similar

in magnitude to its reported effect on ESC-derived motor neurons (Di Giorgio et al., 2007) (Nagai et al., 2007).

Currently, it is unclear whether there are cell-autonomous mechanisms of motor neuron degeneration induced by mutant SOD1 that can lead to overt differences in motor neuron survival *in vitro*. To see whether iMNs could be used to answer this question, we asked if there is a survival difference between wild-type and *SOD1*G93A iMNs in culture with wild-type glia. We prepared MEFs from mouse embryos that overexpress the *SOD1*G93A transgene as well as harbor the *Hb9::*GFP reporter, and transdifferentiated them into *Hb9::*GFP+ iMNs alongside MEFs which only contain the *Hb9::*GFP reporter. We then FACS-purified *Hb9::*GFP+ iMNs of both genotypes in parallel and plated the same number of cells for each on wild-type glia. After 4 days in culture, we observed impaired survival of *SOD1*G93A iMNs relative to control iMNs (Figure 3.8C), suggestive of a cell-autonomous disease phenotype. Taken together, these results indicate that iMNs are useful for studying both cell autonomous and non-autonomous contributors to motor neuron degeneration in ALS.

Because there is significant interest in the identity of factors and pathways that modulate neuronal survival in the context of neurodegenerative diseases, we also tested whether iMNs were similar to motor neurons in their sensitivity to growth factor withdrawal. Indeed, when the neurotrophic factors GDNF, BDNF and CNTF were all withdrawn from the medium, iMNs were lost more rapidly (Figure 3.8D). Thus, iMNs share a neurotrophic support requirement similar to embryonic motor neurons, and we conclude that iMNs could serve as a suitable substrate for *in vitro* studies of motor neuron function, disease and injury.

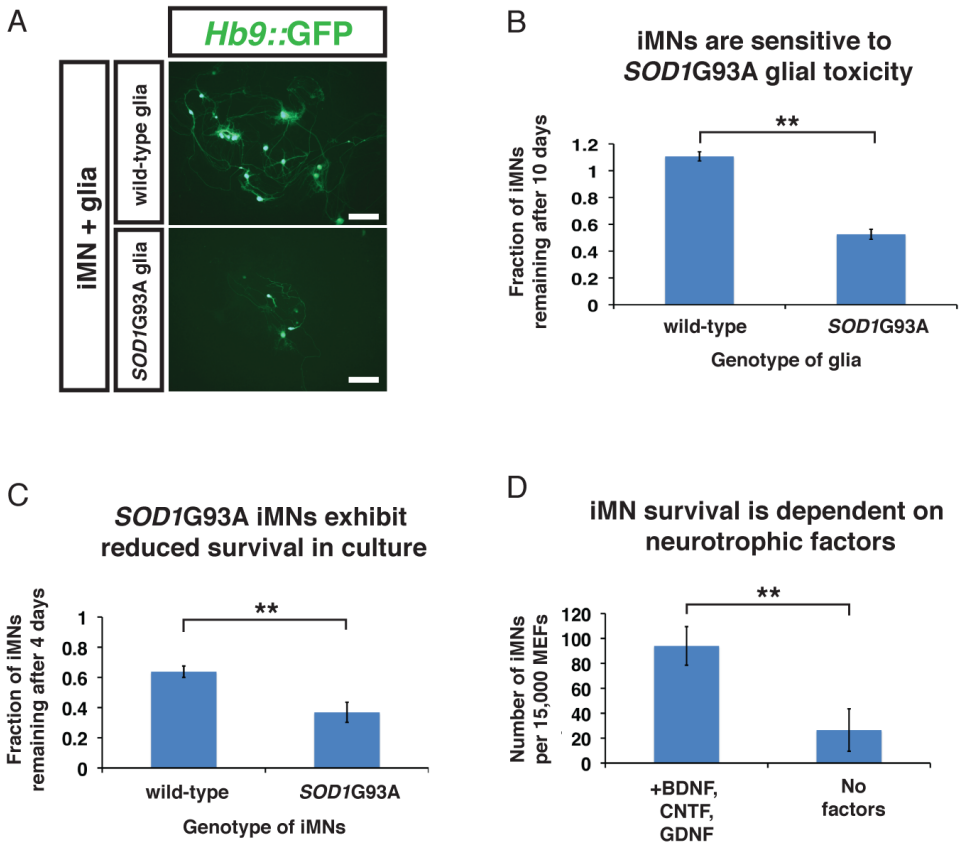

Figure 3.8.
iMNs Recapitulate ALS Disease Phenotypes *in Vitro*. (A) FACS-purified *Hb9::*GFP+ iMNs co-cultured with wild-type or the mutant *SOD1*G93A-overexpressing glia for 10 days. Scale bars reperesent 50 μm. (B) Quantification of (C). Error bars indicate ±s.d. **P <0.01 (Student's *t*-test, two-tailed). (C) *SOD1*G93A iMNs exhibit reduced survival in culture with wild-type glia. Error bars indicate ±s.d. **P <0.01 (Student's *t*-test, two-tailed). (D) Changes in iMN number after 9 days of culture in the presence or absence of neurotrophic factors (GDNF, BDNF and CNTF). Error bars indicate ±s.d. **P <0.01 (Student's *t*-test, two-tailed).

Fibroblasts Do Not Transit Through a Neural Progenitor State Before Becoming iMNs

The process by which the initial fibroblasts undergo conversion into another cell type in defined-factor reprogramming and transdifferentiation experiments remains poorly understood. In particular, it is currently unknown if the somatic cells reprogram through the same developmental intermediates that are found in the developing embryo, for example, by first de-differentiating and then re-differentiating through a neural progenitor state into a neuron, or if they instead convert more "directly". To address this question, we used a lineage tracing approach to ask if during the course of reprogramming, a gene commonly used to identify neuronal progenitors ever became expressed.

Motor neuron progenitor cells are highly proliferative in culture (Frederiksen and McKay, 1988; Jessell, 2000). To determine whether iMNs transited through a highly proliferative intermediate during the reprogramming process, we quantified the timing of cell division in the reprogramming cultures using 48-hour pulses of BrdU. Following transduction, we found that the cells incorporated decreasing amounts of BrdU at each subsequent time point and did not incorporate detectable levels of Brdu after 4 days post-transduction (Figure 3.9A). Consistent with a previous report (Vierbuchen et al., 2010), these results suggest that the transduced cells quickly become post-mitotic. Since 10% of the fibroblasts eventually become iMNs and because GFP+ iMNs do not begin to appear in culture until day 5 and the majority arise between 7 and 14 days in culture these results suggest that the iMNs are not being produced from highly proliferative neuronal progrenitors.

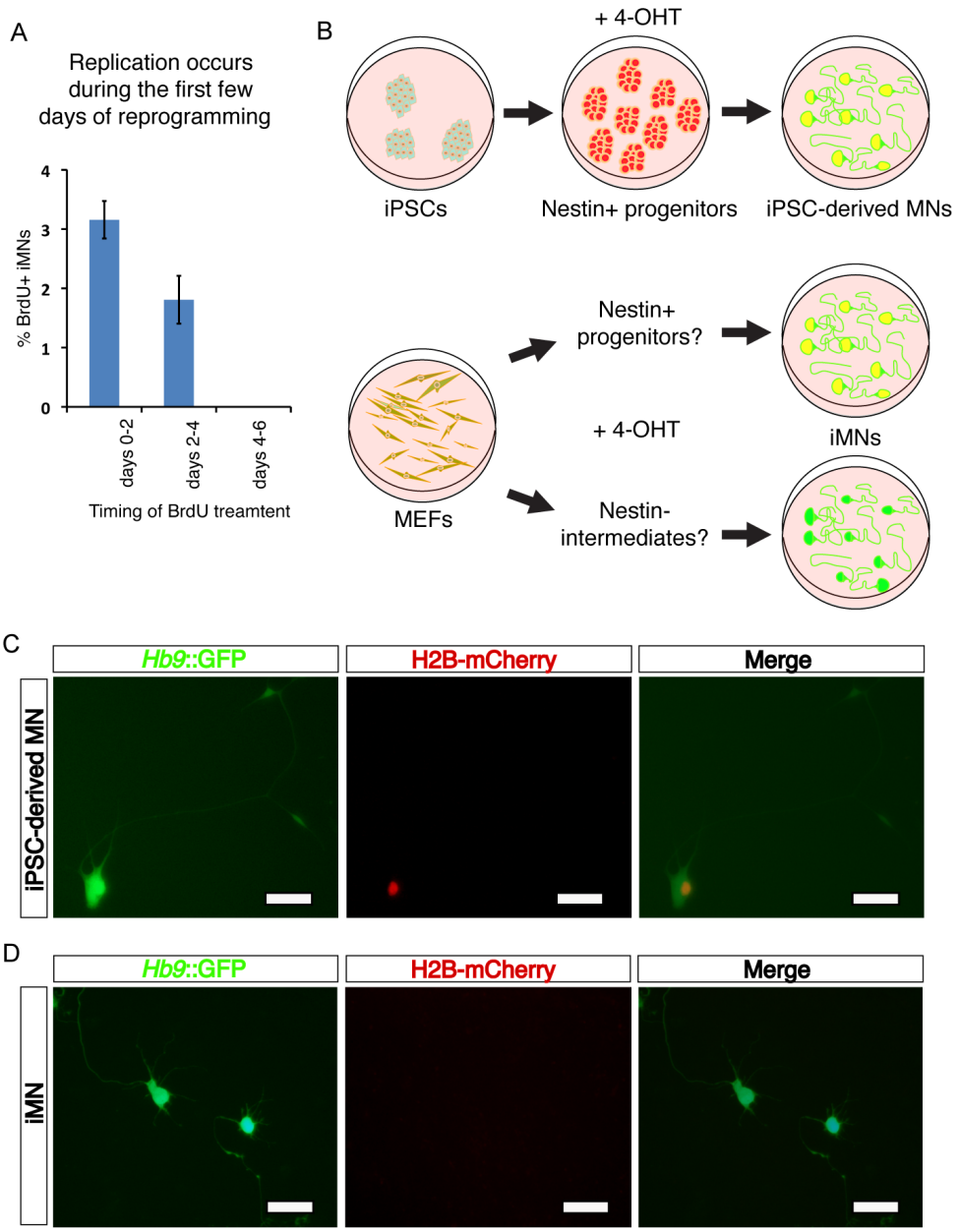

Figure 3.9.
Transdifferention Does Not Occur through a Nestin+ Neural Progenitor State. (A) Percentage of iMNs that have incorporated BrdU. (B) Outline of the lineage tracing experiment using *Nestin::*CreER; LOX-STOP-LOX-H2B-mCherry; *Hb9::*GFP iPSCs or MEFs. To detect Nestin+ intermediates, cultures were treated with 1-2 µM 4-OHT during directed diffentiation of iPSCs (positive control) or during transdifferentiation of fibroblasts by the 7 factors. (C) FACS-purified, mCherry+ *Hb9::*GFP+ motor neurons derived from the triple transgenic iPSCs in the presence of 1 µM 4-OHT. Expression of mCherry was observed in 3% of *Hb9::*GFP+ cells (n > 10,000) and indicates the activation of *Nestin::*CreER during directed differentiation. Scale bars represent 40 µm. (D) mCherry- *Hb9::*GFP+ iMNs generated from the triple transgenic MEFs by transdifferentiation in the presence of 2 µM 4-OHT. mCherry+ iMNs were never observed (n > 5,000), suggesting a Nestin+ state is not accessed during reprogramming. Scale bars represent 40 µm.

To more definitively test if the fibroblasts become motor neuron progenitors before differentiating into iMNs, we repeated the induction of a motor neuron identity using transgenic fibroblasts with a *Nestin*::CreER (Burns et al., 2007); lox-stop-lox-H2B-mCherry (Abe et al.); *Hb9*::GFP genotype (Figure 3.9B). Because Nestin is a well-known marker of neural progenitor cells in the mammalian CNS (Messam et al., 2002), we reasoned that if the fibroblasts transited through a progenitor state before becoming motor neurons, the resulting iMNs would activate expression of *Nestin::*CreER, recombine the reporter gene and thus express both mCherry and *Hb9::*GFP.

First, as a positive control for this experiment, we generated iPSCs from the fibroblasts, then used retinoic acid and sonic hedgehog (Wichterle et al., 2002) to differentiate the iPSCs into motor neurons. As this directed differentiation protocol mimics development, we expected the resulting motor neurons to originate from Nestin+ precursors. When we performed the differentiation without 4-hydroxytamoxifen (4-OHT), none of the resulting *Hb9*::GFP+ motor neurons expressed mCherry (Figure 3.9C). However, when we added 4-OHT to the differentiation, 3% of the motor neurons co-expressed mCherry ($n > 10,000$) (Figure 3.9C), verifying that the *Nestin*::CreER reporter successfully identified motor neurons that transited through a Nestin+ progenitor state. In contrast, when we treated the 7 factor-transduced MEF cultures with 4-OHT both before and during transdifferentiation, none of the resulting iMNs expressed mCherry ($n > 5,000$) (Figure 3.9D). These results confirm that fibroblasts do not become iMNs by transiting through a motor neuron progenitor cell state and further rule out the possibility that many of the iMNs are derived from contaminating neural progenitor cells in the MEF cultures.

Human iMNs Can Be Generated by Eight Transcription Factors

We next sought to determine whether the same, or a similar set of factors could be used to generate human iMNs from fibroblasts. To this end, human embryonic fibroblasts (HEFs) were derived from a human ESC line, HUES3, harboring the *Hb9::*GFP transgene (Di Giorgio et al., 2008). The HUES3-HEFs were first transduced with the mouse ecotropic receptor mCAT, then transduced with retroviruses containing the 7 iMN factors identified in the mouse system as well as *NEUROD1*, a pro-neural gene reported to enhance the conversion efficiency of human fibroblasts into iNs (Pang et al., 2011). 30 days after transduction, we observed *Hb9::*GFP+ cells with highly neuronal morphologies in the culture of 8 factor-transduced HUES3-HEFs (Figures 3.10A-B), whereas untransduced HUES3-HEFs never spontaneously expressed the transgene under the same conditions (Figure 3.10B). These putative human iMNs expressed vesicular ChAT (Figure 3.10C), indicating that they were indeed cholinergic in nature.

In order to assess the functionality of human iMNs made with 8 factors, we employed whole-cell patch clamp recording to look at their electrophysiological properties. Similar to their mouse counterparts, HUES3-HEF-iMNs expressed functional voltage-gated sodium and potassium channels (Figure 3.10D) and were able to fire action potentials (Figure 3.10E) when depolarized. Importantly, they responded appropriately to the addition of 100 μM kainate (Figure 3.10F) and 100 μM GABA (Figure 3.10G), demonstrating their ability to receive and respond to the major excitatory and inhibitory inputs, respectively, that govern spinal motor neuron activity. Therefore, functional iMNs can be generated from human fibroblasts by transdifferentiation.

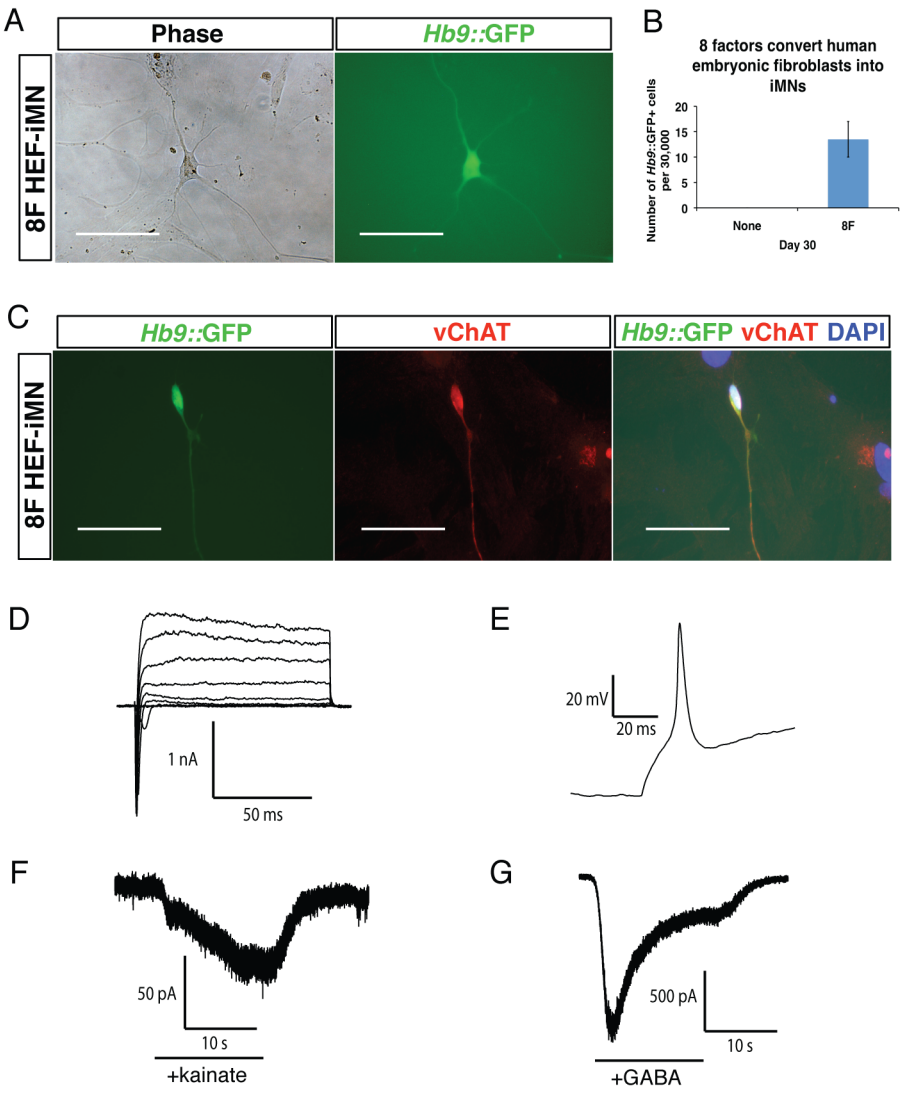

Figure 3.10.
Human iMNs Are Generated by Eight Transcription Factors. (A) An *Hb9::*GFP+ neuron generated from HUES3-HEFs by 8 transcription factors. Scale bars represent 80 µm. (B) Quantification of human iMN reprogramming efficiency at day 30 post-transduction. (C) Human iMNs express vChAT (red). Scale bars represent 80 µm. (D-G) Human iMNs are electrophysiologically active. (D) Human iMNs express functional sodium and potassium channels. (E) Human iMNs fire action potentials upon depolarization. (F) 100 µM kainate induces inward currents in human iMNs. (G) 100 µM GABA induces inward currents in human iMNs.

Functional ALS Patient-Specific iMNs Are Generated

Encouraged by these initial results using retroviruses, we set out to improve the human iMN reprogramming process. We noticed that a slightly greater number of *Hb9::*GFP+ cells were generated from HUES3-HEFs when *Hb9* was omitted from the cocktail (data not shown); therefore, all subsequent human iMN experiments were performed in the absence of the *Hb9* transgene.

Suspecting that the low transduction efficiency of the ecotropically packaged retrovirus was a major limiting factor, we cloned the 7 iMN genes (7F'; *NEUROD1* plus the 7 factors discovered in mouse, except *Hb9*) into a doxycycline (Dox)-inducible lentiviral vector (FUW-tetO). We transduced HUES3-HEFs with lentiviruses containing the 7F' cocktail and administered Dox. After 30 days, there were numerous *Hb9::*GFP+ cells in the culture, the majority of which had robust TUJ1 and HB9 expression (Figures 3.11A-B), indicating that the FUW-tetO lentivirus resulted in an increased efficiency of human iMN reprogramming.

While the transgenic *Hb9::*GFP reporter in HUES3 has worked well, establishing a similarly reliable reporter iPS line for each patient may not be practical. To this end, a small subset of the 7F'-transuced HUES3-HEFs had been subsequently infected with a lentiviral *Hb9::*RFP reporter, which contains a shorter *Hb9* promoter fragement (3.7 kb) (Marchetto *et al*., 2008) driving RFP expression. In this way, we could directly compare the performance of the lentiviral RFP reporter with the transgenic *Hb9::*GFP reporter that contains a well-validated, longer promoter fragment (9.2 kb) (Lee *et al*., 2004). After 5

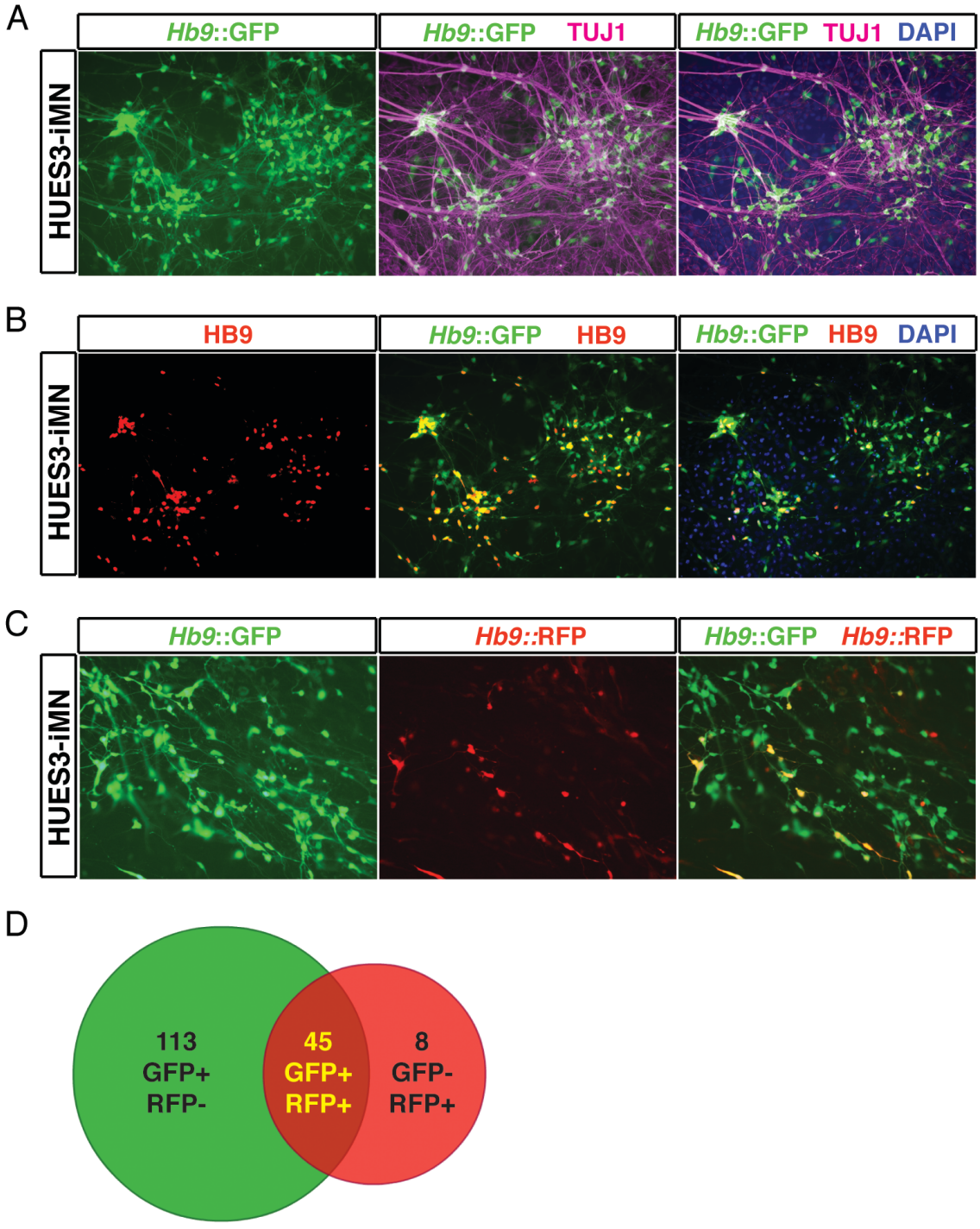

Figure 3.11.
Improved Efficiency of Human iMN Reprogramming. (A) HUES3-HEF-derived iMNs express TUJ1. Scale bars represent 200 µm. (B) HUES3-HEF-derived iMNs express HB9. Scale bars represent 200 µm. (C) Lentiviral *Hb9::*RFP reporter expression in HUES3-HEF-derived iMNs. Scale bars represent 200 µm. (D) Overlap of the transgenic *Hb9::*GFP reporter and the lentiviral *Hb9::*RFP reporter in HUES3-HEF-derived iMNs.

Subject	Genotype	Fibroblast types used
A11	Wild-type	Primary fibroblasts and iPSC-derived HEFs
A27	Wild-type	Primary fibroblasts
A36	TDP-43 Q343R	Primary fibroblasts and iPSC-derived HEFs
A39	SOD1 A4V	Primary fibroblasts and iPSC-deribed HEFs
A47	TDP-43 G298S	Primary fibroblasts
Rb9	SOD1 A4V	iPSC-derived HEFs

Table 3.3.
List of Human Fibroblasts from ALS Patients and Control Subjects.

Figure 3.12.

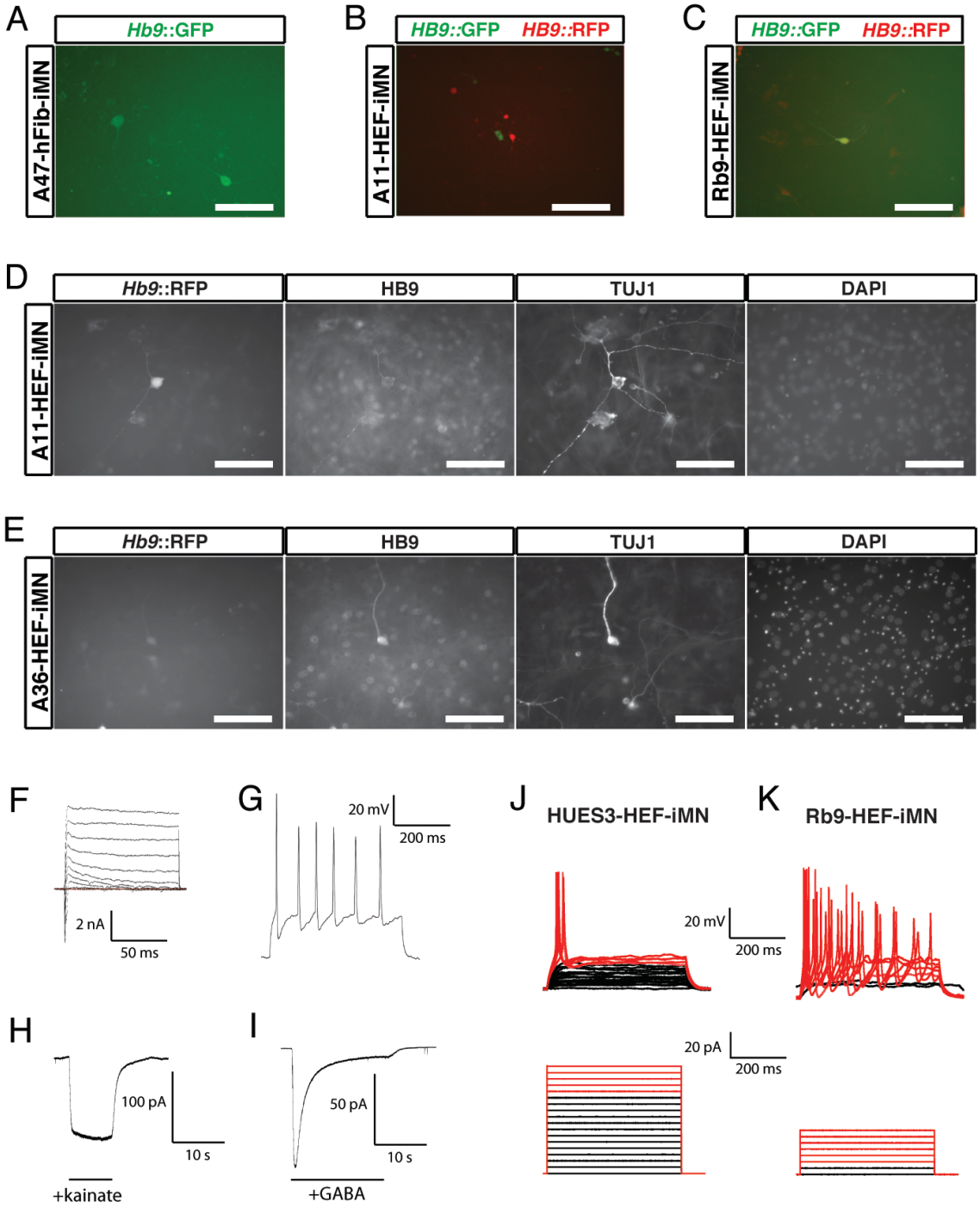
A
Hb9::GFP
A47-hFib-iMN
B
HB9::GFP
HB9::RFP
A11-HEF-iMN
C
HB9::GFP
HB9::RFP
Rb9-HEF-iMN
D
Hb9::RFP
HB9
TUJ1
DAPI
A11-HEF-iMN
E
Hb9::RFP
HB9
TUJ1
DAPI
A36-HEF-iMN
F
2 nA
50 ms
G
20 mV
200 ms
J
HUES3-HEF-iMN
20 mV
200 ms
K
Rb9-HEF-iMN
H
100 pA
10 s
+kainate
I
50 pA
10 s
+GABA
20 pA
200 ms

Figure 3. 12. (Continued)
Electrophysiologically Active ALS Patient iMNs are Produced. (A) A47-hFib-derived cells that express *Hb9::*GFP ePiggybac (ePB) reporter but not *Hb9*::RFP lentiviral reporter. Scale bars represent 100 μm. (B) A11a-HEF-derived cells that express either *Hb9::*GFP ePB reporter or *Hb9::*RFP lentiviral reporter. Scale bars represent 200 μm. (C) An RB9d-HEF-derived cell that expresses both *Hb9::*GFP ePB reporter and *Hb9::*RFP lentiviral reporter. Scale bars represent 200 μm. (D) *Hb9::*RFP+ TUJ1+ neuronal cells are generated from A11a-HEFs. Scale bars represent 100 μm. (E) *Hb9::*RFP+ TUJ1+ neuronal cells are generated from A36a-HEFs. Scale bars represent 100 μm. (F-I) A36-HEF-iMNs are elecrophysiologically active. (F) A36-HEF-iMNs express functional voltage-gated sodium and potassium channels. (G) A36-HEF-iMNs fire multiple action potentials upon depolarization. (H) A36-HEF-iMNs respond to 100 μM kainate. (I) A36-HEF-iMNs respond to 100 μM GABA. (J-K) ALS patient iMNs can show heightened excitability relative to wild-type controls. (J) HUES3-HEF-iMNs generate single action potentials in response to the current injections as shown. (K) Rb9-HEF-iMNs can have a lower threshold for firing and may be prone to generating trains of action potentials.

weeks of Dox treatment, there were many fewer RFP+ cells than GFP+ cells (53 versus 158, respectively) with neuronal morphologies, with some cells co-expressing both reporters (45 cells) (Figures 3.11C-D). Taking the expression of the transgenic *Hb9::*GFP reporter as a surrogate for a motor neuron state, this indicates that the lentiviral reporter vastly underestimates the reprogramming efficiency, since less than 30 % of GFP+ cells co-express RFP. On the other hand, 85 % of RFP+ cells also express GFP+, suggesting that the lentiviral reporter has a reasonable degree of fidelity and could be used to detect patient iMNs.

We went on to generate iMNs from a small panel of human fibroblasts from healthy adults as well as ALS patients with known disease-causing mutations in SOD1 or TAR DNA-binding protein 43 (TDP-43) (Neumann *et al*., 2006); adult primary fibroblasts as well as HEFs derived from established patient-specific iPS lines were used (Table 3.3). In addition to the *Hb9::*RFP lentiviral reporter, we also decided to test an *Hb9::*GFP ePiggybac transposon-based reporter containing the same promoter fragment as a possible alternative. The patient-derived fibroblasts were first modified with the ePiggybac reporter and selected for stably maintaining the construct. They were then transduced with 7F' and, the following day, with the *Hb9::*RFP lentiviral reporter.

After 40 days of Dox treatment, rare reporter-expressing cells emerged in all of the cell types; we saw a larger number of *Hb9::*RFP+ cells than *Hb9::*GFP+ cells, with only a limited degree of overlap, suggesting that the lentiviral reporter was more efficient than the ePiggybac reporter (Figure 3.12A-C). However, while the labeled cells expressed TUJ1, they did not convincingly express HB9 or possess complex neuronal morphologies (Figures 3.12D-E).

Despite the rarity of reprogrammed human iMNs, those marked by either reporter were electrophysiologically active, having functional voltage-gated sodium and potassium channels (Figures 3.12F) and firing action potentials (Figures 3.12G), as well as responding to neurotransmitters (Figures 3.12H-I). Surprisingly, many of the ALS patient-derived iMNs seemed much more excitable than what we commonly observe in non-ALS iMNs (Figures 3.12J-K): they not only generated trains of action potentials more readily, but did so with much smaller input current. A larger sample size would be needed to determine whether or not ALS iMNs behave in a significantly different manner in this assay.

Improved Efficiency of Patient iMN Generation

Although the FUW-tetO lentiviral system significantly increased the efficiency of HUES3-HEF reprogramming, there was a clear need to optimize patient-specific iMN reprogramming. Therefore, we decided to explore the use of another inducible lentiviral vector, pHAGE-tetO, that might permit greater transgene expression. We focused on fibroblasts derived from one patient, A39, who had an SOD*A4V* mutation that causes one of the most severe forms of ALS, with an average of three years from the time of disease onset until death.

Both primary fibroblasts (A39-hFibs) and iPS-derived HEFs (A39-HEFs) were transduced with 7F'-containing FUW-tetO or pHAGE-tetO lentiviruses, with a subset of cells also subsequently receiving the *Hb9::*RFP lentiviral reporter. The pHAGE virus seemed to infect the cells better, as judged by control transductions of Venus, and caused less toxicity (data not shown). Two weeks into Dox addition, *Hb9::*RFP+ cells began to

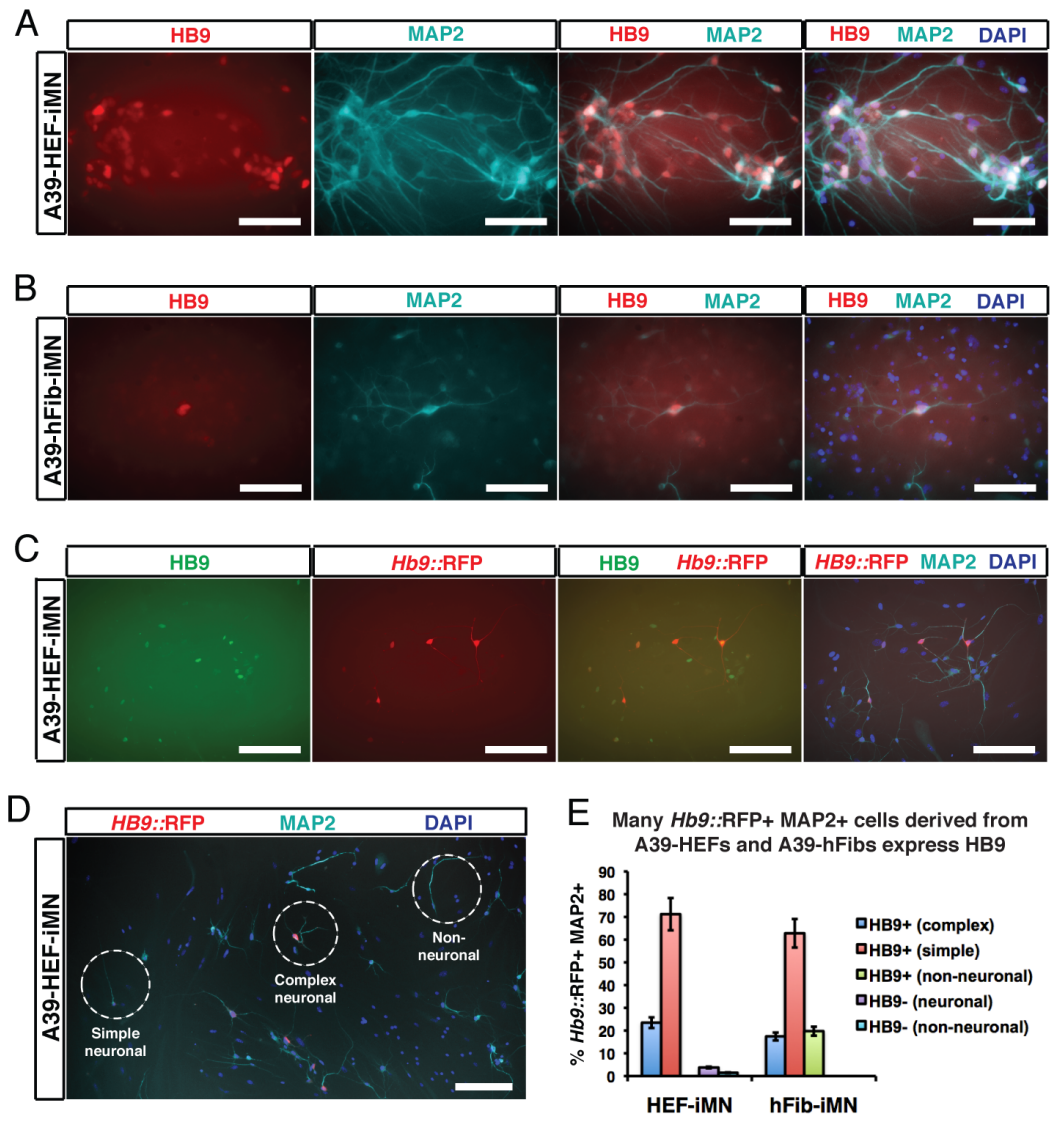

Figure 3.13.

Optimization of ALS Patient iMN Generation and Detection. (A) HB9+ MAP2+ neuronal cells are generated from A39-HEFs. Scale bars represent 100 µm. (B) HB9+ MAP2+ neuronal cells are generated from A39-hFibs. Scale bars represent 100 µm. (C) The lentiviral *Hb9::*RFP reporter labels HB9+ MAP2+ cells generate from A39-HEFs. Scale bars represent 200 µm. (D) Representative pictures of morphological categories for classifying human patient iMNs. Cells with a distinct soma and multiple and/or branched processes were classified as 'complex neuronal'; those with a distinct soma and a simple process, 'simple neuronal'; and those with a non-neuronal morphology, 'non-neuronal'. Scale bar represents 200 µm. (E) Quantification of HB9 expression in different categories of *Hb9::*RFP+ MAP2+ cells, generated from A39-HEFs or A39-hFibs.

emerge in cultures infected 7F'-pHAGE-tetO viruses. In contrast, FUW-tetO-infected cultures never generated significant numbers of *Hb9::*RFP-labeled cells at any timepoint (data not shown) and were excluded from further analysis.

After 4 weeks of Dox administration, we fixed and immunostained the pHAGE-tetO-infected cultures with antibodies for MAP2 and HB9 (Figures 3.13A-B). Surprisingly, the overall frequency of HB9+ MAP2+ cells was encouraging, amounting to 5.7 % of transduced A39-hFibs. Some of the HB9+ MAP2+ cells had complex neuronal morphologies (Figures 3.13A-B), suggesting that the quality of the putative iMNs had also improved. In the culture of infected A39-HEFs, clusters of Hb9+ MAP2+ cells were frequently observed (Figure 3.13A).

We noticed that the lentiviral *Hb9::*RFP reporter could label some of the HB9+ MAP2+ cells in the reprogramming cultures (Figures 3.13C-D), and asked how well it performed in these contexts. For this purpose, *Hb9::*RFP+ MAP2+ cells were scored for HB9 protein expression and also assigned to different morphological classes: complex neuronal, simple neuronal, and non-neuronal (Figure 3.13D). In the culture of 7F'-transduced A39-hFibs. 17 % of reporter-labeled MAP2+ cells stained positively for HB9 and had a complex morphology, while 63 % were HB9+ with a simple morphology (Figure 3.13E). Therefore, the fidelity of the reporter, estimated by HB9 expression in labeled, MAP2+ cells, is around 80 % – a slightly lower figure in the HUES3-HEF-iMNs. Conversely, this indicates that only 16 % of all HB9+ MAP2+ cells were labeled by the reporter in this case. (Figure 3.13E).

Taken together, these results clearly demonstrate that human iMNs from patient-derived fibroblasts can be generated using a suitable gene delivery method. The lentiviral

reporter for the motor neuron fate underreports on the number of HB9+ cells, but has a reasonable fidelity. Notably, a higher proportion of HEF-derived iMNs exhibited complex morphologies (24 % of RFP+ MAP2+ cells) compare with hFib-derived iMNs (17 % of RFP+ MAP2+ cells), suggesting that the HEFs are more amenable to the reprogramming process.

Discussion

We have shown that a small set of transcription factors can convert embryonic and adult fibroblasts into functional motor neurons. The iMNs expressed pan-neuronal and motor neuron-specific markers, as well as the receptors and channels that generate excitable membranes sensitive to transmitters, allowing them both to fire action potentials and receive synaptic input. These cholinergic iMNs also possessed the defining hallmark of motor neurons: the ability to synapse with muscle and to induce its contraction. Most importantly, iMNs are able to contribute to the developing CNS *in vivo*, migrating appropriately to the ventral horn and sending out axonal projections through the ventral root. We also demonstrated that the iMNs are sensitive to a degenerative ALS stimulus that selectively affects motor neurons. Thus we provide several lines of evidence that iMNs are functional motor neurons with consequent utility for the study of motor neuron physiology and disease susceptibility.

It is remarkable that the conversion to motor neurons occurs so efficiently given that the cells do not transit through a neural progenitor state. It was striking that under certain conditions, as many as one *Hb9::*GFP+ iMN was made from every 10 MEFs. This efficiency was substantially higher than iPSC reprogramming (Takahashi and

Yamanaka, 2006) and could be the result of a cooperative process in which establishment of a general neuronal program is augmented by specific patterning to a motor neuron identity. These results indicate that the massive changes in gene expression induced during defined-factor reprogramming can be executed efficiently even though they do not mimic embryonic development precisely. It will be of interest to determine whether this approach can serve as a general strategy for the production of many distinct neuronal subtypes.

Importantly, this method enables us to reprogram adult human fibroblasts into iMNs. We have significantly improved the efficiency of human iMN reprogramming, such that it would be feasible to use the iMNs for applications that require a relatively small number of cells. Using more powerful viral vectors, adding and removing genetic factors, or using small molecules might enhance the efficiency further. Developing a reporter system with better accuracy and efficiency of reporting would also be worthwhile, for example, by using a larger promoter fragment to drive the reporter gene expression.

It is intriguing that ALS patient fibroblast-derived iMNs seem prone to fire action potentials, perhaps more so than is typical for non-ALS iMNs. As an aberrant increase in neuronal activity is associated with muscle fasciculation in ALS, it is tempting to speculate that the readily excitable state observed in ALS patient iMNs is a possible disease-related phenotype that may be worth pursuing.

Going forward, direct reprogramming may be an ideal tool for rapidly surveying a cohort of ALS patient-derived motor neurons for disease-relevant phenotypes. By stratifying the patient population into phenotypic classes with clinical significance, a

more targeted approach to developing therapeutics may become possible. Applying it to the production of other neuronal subtypes could help tackle a variety of neurodegenerative diseases, and reveal more principles that govern the stability and plasticity of cell fate.

Materials and Methods

Molecular Cloning, Isolating Embryonic and Adult Fibroblasts, Viral Transduction, and Cell Culture

Complementary DNAs for the 11 candidate factors were each cloned into the pMXs retroviral expression vector, as well as FUW-tetO and pHAGE-tetO lentiviral vectors, using Gateway technology (Invitrogen). *Hb9::*GFP-transgenic mice (Jackson Laboratories) were mated with ICR mice (Taconic) and MEFs were harvested from *Hb9::*GFP E12.5 embryos under a dissection microscope (Leica). TTFs were isolated from *Hb9::*GFP-transgenic adult mice as previously described (Vierbuchen et al., 2010). The fibroblasts were passaged at least once before being used for experiments. HEFs were isolated from human ESCs or iPSCs by culturing them in DMEM + 20% fetal bovine serum without bFGF for at least three passages. Retroviral and lentiviral transductions were performed as described (Ichida et al., 2009). Glial cells isolated from P2 ICR mouse pups were added to infected fibroblasts two days after transduction. The next day, medium was switched either to mouse motor neuron medium containing F-12 (Invitrogen), 5% horse serum, N2 and B27 supplements, glutamax and penicillin/streptomycin, or to N3 medium (Vierbuchen et al., 2010). Both media were supplemented with GDNF, BDNF and CNTF, all at 10 ng/ml. Efficiency of mouse iMN generation was estimated by counting the number of *Hb9::*GFP+ cells with neuronal morphologies using a fluorescence microscope (Nikon).

Obtaining ESC-Derived and Embryonic Motor Neurons, FACS, Microarray Analysis, and qPCR

Motor neurons were derived from *Hb9::*GFP mouse ESCs and isolated by FACS using standard protocol (Di Giorgio et al., 2007). Embryonic motor neurons were harvested from *Hb9::*GFP E13.5 embryos. Briefly, whole spinal cords were washed in F-12 (Invitrogen) and incubated in 10 ml of 0.025% trypsin with DNase for 45 minutes with gentle agitation every 15 minutes. Media was added to the dissociated spinal cords and the cells were triturated, spun down at 1,000 rpm for 5 minute and resuspended in DMEM/F-12 with glutamax and penicillin/streptomycin. FACS was performed in the same way as with ESC-derived motor neurons. Total RNA isolation, RNA amplification and microarray analysis were performed as described previously (Ichida et al., 2009). qPCR was performed using iScript cDNA synthesis, SYBR green qPCR supermix (Bio-rad), and the primers in Table S1.

Immunocytochemistry

Antibody staining was performed as previously described (Ichida et al., 2009). The following primary antibodies were used: mouse anti-Hb9 (DSHB, 1:50), mouse anti-Islet (DSHB, 1:100); mouse anti-TuJ1 (Covance, 1:500); rabbit anti-vChAT (Sigma, 1:1000); rabbit anti-synapsin I (Millipore, 1:500); goat anti-Chx10 (Santa Cruz, 1:200); and rabbit anti-tyrosine hydroxylase (ThermoScientific, 1:300).

Electrophysiology

Whole-cell voltage-clamp and current-clamp recordings were made using a Multiclamp 700B (Molecular Devices) at room temperature (21-23°C). Data were digitized with a Digidata 1440A A/D interface and recorded using pCLAMP 10 software (Molecular

Devices). Data were low-pass filtered at 2 kHz and sampled at 20 kHz (1kHz and 2 kHz, respectively, for transmitter application). Patch pipettes were pulled from borosilicate glass capillaries on a Sutter Instruments P-97 puller and had resistances of 2-4 MΩ. The pipette capacitance was reduced by wrapping the shank with Parafilm and compensated for using the amplifier circuitry. Series resistance was typically 5-10 MΩ, always less than 15 MΩ, and compensated by at least 80%. Leak currents were typically less than 200 pA with mean input resistance 675 MΩ and mean resting potential -49 mV. For study of voltage-gated conductances, linear leakage currents were digitally subtracted using a P/4 protocol and voltage was stepped from a holding potential of -80 mV to test potentials from -80 to 30 mV in 10 mV increments. Intracellular solutions were potassium-based solution and contained KCl, 150; $MgCl_2$, 2; HEPES, 10; pH 7.4 used for earlier experiments and KCl, 135; $MgCl_2$, 2; HEPES, 10; MgATP, 4; NaGTP, 0.3; Na_2PhosCr, 10; EGTA, 1; pH 7.4 used for later experiments with no obvious difference in sodium and potassium currents. The extracellular was sodium-based and contained NaCl, 135; KCl, 5; $CaCl_2$, 2; $MgCl_2$, 1; glucose, 10; HEPES, 10; pH 7.4). Based on the chloride Nernst potential of -2 mV, inward currents were expected following GABA and glycine treatment (Puia et al., 1990). Transmitters were not washed out, explaining the delayed current decay.

C2C12 Muscle Co-Culture

C2C12 myoblasts were expanded in DMEM with 20% fetal bovine serum and penicillin/streptomycin. When the culture reached 100% confluency, the serum content was reduced to 5% to induce differentiation. Flow-purified iMNs or iNs were added to

the myotubes after 7-14 days and the medium switched to either mouse motor neuron or N3 media. The co-cultures were monitored for myotube contractions under the microscope with 10x or 20x objectives. To stop contractions, a solution of tubocurarine hydrochloride was added to a final concentration of between 50 nM and 50 μM. Twitching myotubes were filmed using Nikon ACT-2U Imaging Software (Excel Technologies) and contraction frequencies determined.

iMN-Chick Myotube Co-Cultures and Immunocytochemistry

Myoblasts were isolated from the epaxial (longissimus) muscles of E10 White Leghorn chick embryos and plated in 24-well plates at a density of 100,000 cells/well. Cultures were maintained at 37^0C in F10 media (Gibco) supplemented with 0.44 mg/ml calcium chloride, 10% horse serum, 5% chicken serum and 2% penicillin:streptomycin. iMNs were added to the myotubes 5 days later in Neurobasal media (Gibco) supplemented with B27 (Gibco), 1% L-glutamine and 1% penicillin:streptomycin. Co-cultures were supplemented with 10ng/mL CNTF and GDNF every two days for the first week following the addition of the iMNs. Co-cultures were maintained for 3 weeks when they were prepared for immunocytochemistry. Antibody staining was performed as previously described (Soundararajan et al., 2006). A rabbit anti-GFP (Chemicon, 1:2000) primary antibody was used to visualize the iMNs and rhodamine-conjugated α-bungarotoxin (Invitrogen, 1:500) was used to visualize the AChRs. Images were acquired on a laser scanning-confocal microscope (Zeiss LSM 510). Orthogonal images were rendered and edited with LSM imaging software (Zeiss) and further contrast and brightness adjustments were performed on Photoshop version 7.0.

In Ovo *Transplantation of ESC-derived Motor Neurons and iMNs*

In ovo transplantations and immunohistochemistry were performed as previously described[12]. Briefly, E2.5 chick embryos were exposed; the vitelline membrane and amnion were cut to allow surgical access to the neural tube. An incision of 1-1.5 somites in length was made along the midline of the neural tube at the rostral extent of the developing hind limb bud (T7-L1) using a flame-sterilized tungsten needle (0.077 mm wire, World Precision Instruments). For control ESC-derived motor neuron transplantations, *Hb9::*GFP-transgenic mouse ESCs were differentiated into motor neurons as described previously (Soundararajan et al., 2006; Wichterle et al., 2002). A single embryoid body containing approximately 150-200 differentiated *Hb9::*GFP+ motor neurons was transplanted into the ventral lumen of the neural tube of E2.5 chick embryos as described previously (Soundararajan et al., 2006). For iMN transplantations, a sphere of iMNs mixed with non-transgenic, ESC-derived motor neurons containing approximately 200 cells was transplanted into the ventral lumen of the neural tube of E2.5 chick embryos. For all transplantations, the chick embryos were harvested five days later, fixed in 4% paraformaldehyde/PBS, cut on a cryostat and then processed for immunohistochemistry. The following primary antibodies were used: rabbit anti-GFP (Chemicon, 1:1000) and mouse anti-Tuj1 (Covance, 1:1000). Images were captured with a digital camera (C4742; Hamamatsu Photonics, Hamamatsu, Japan) in conjunction with digital imaging acquisition software (IPLab; Version 4.0; BD Biosciences, Rockville, MD, USA).

Glia-Neuron Co-Culture for Disease Modeling

*SOD1*G93A transgenic mice (Jackson Laboratories) were mated with ICR mice. Glial preps were derived from transgenic P2 pups and their littermates. 3 weeks later, confluent flasks of glial cells were passaged 1:2 onto 6-well plates and iMNs were plated on top. The co-cultures were kept in mouse motor neuron medium with neurotrophic factors and the media changed every other day for the duration of the experiment.

Nestin::*CreER Lineage Tracing*

MEFs were isolated from E13.5 embryos that were transgenic for *Nestin*::CreER, LOX-STOP-LOX H2B-mCherry, and *Hb9*::GFP. To generate iPSCs, the MEFs were transduced with retroviruses (pMXs vector) encoding *Oct4, Sox2,* and *Klf4*. Cells were cultured in mES media containing 13% Knockout Serum Replacement and colonies were picked, expanded, and verified by Nanog immunostaining. For the positive control, iPSCs were differentiated into motor neurons using retinoic acid and Sonic Hedgehog (Wichterle et al., 2002) in the presence or absence of 2 μM 4-OHT. iMNs were also created in the presence or absence of 2 μM 4-OHT.

Acknowledgements

We are grateful to A.C. Carter for help with lineage tracing experiments and E. Kiskinis, S. de Boer, G. Boulting, and J. Rivera-Feliciano for providing reagents and helpful discussions. We would also like to thank B. Tilton for assistance with FACS, K. Harrison for providing the mouse *Hb9* cDNA, N. Atwater for assistance with glia preparation, J. Sandoe and K. Sandor for help with molecular cloning, K. Koszka for help with mouse husbandry and M.Y. Son for help with the diagram in Figure 3.1. The authors declare no competing financial interests. Microarray data have been submitted to the GEO repository (Accession Number: GSE31160).

References

Abe, T., Kiyonari, H., Shioi, G., Inoue, K., Nakao, K., Aizawa, S., and Fujimori, T. Establishment of conditional reporter mouse lines at ROSA26 locus for live cell imaging. Genesis *49*, 579-590.

Alessandri-Haber, N., Paillart, C., Arsac, C., Gola, M., Couraud, F., and Crest, M. (1999). Specific distribution of sodium channels in axons of rat embryo spinal motoneurones. J Physiol *518 (Pt 1)*, 203-214.

Burns, K.A., Ayoub, A.E., Breunig, J.J., Adhami, F., Weng, W.L., Colbert, M.C., Rakic, P., and Kuan, C.Y. (2007). Nestin-CreER mice reveal DNA synthesis by nonapoptotic neurons following cerebral ischemia hypoxia. Cereb Cortex *17*, 2585-2592.

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.* Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. Nature.

Cui, D., Dougherty, K.J., Machacek, D.W., Sawchuk, M., Hochman, S., and Baro, D.J. (2006). Divergence between motoneurons: gene expression profiling provides a molecular characterization of functionally discrete somatic and autonomic motoneurons. Physiol Genomics *24*, 276-289.

Dahm, L.M., and Landmesser, L.T. (1988). The regulation of intramuscular nerve branching during normal development and following activity blockade. Dev Biol *130*, 621-644.

Di Giorgio, F.P., Boulting, G.L., Bobrowicz, S., and Eggan, K.C. (2008). Human embryonic stem cell-derived motor neurons are sensitive to the toxic effect of glial cells carrying an ALS-causing mutation. Cell Stem Cell *3*, 637-648.

Di Giorgio, F.P., Carrasco, M.A., Siao, M.C., Maniatis, T., and Eggan, K. (2007). Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. Nat Neurosci *10*, 608-614.

Frederiksen, K., and McKay, R.D. (1988). Proliferation and differentiation of rat neuroepithelial precursor cells in vivo. J Neurosci *8*, 1144-1151.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K.*, et al.* (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell *5*, 491-503.

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.

Jessell, T.M. (2000). Neuronal specification in the spinal cord: inductive signals and transcriptional codes. Nat Rev Genet *1*, 20-29.

Lee, S.K., Jurata, L.W., Funahashi, J., Ruiz, E.C., and Pfaff, S.L. (2004). Analysis of embryonic motoneuron gene retulation: derepression of general activators function in concert with enhancer factors. Dev *131*, 3295-3306.

Lee, S.K., Lee, B., Ruiz, E.C., and Pfaff, S.L. (2005). Olig2 and Ngn2 function in opposition to modulate gene expression in motor neuron progenitor cells. Genes Dev *19*, 282-294.

Lupa, M.T., and Hall, Z.W. (1989). Progressive restriction of synaptic vesicle protein to the nerve terminal during development of the neuromuscular junction. J Neurosci *9*, 3937-3945.

Messam, C.A., Hou, J., Berman, J.W., and Major, E.O. (2002). Analysis of the temporal expression of nestin in human fetal brain derived neuronal and glial progenitor cells. Brain Res Dev Brain Res *134*, 87-92.

Miles, G.B., Yohn, D.C., Wichterle, H., Jessell, T.M., Rafuse, V.F., and Brownstone, R.M. (2004). Functional properties of motoneurons derived from mouse embryonic stem cells. J Neurosci *24*, 7848-7858.

Nagai, M., Re, D.B., Nagata, T., Chalazonitis, A., Jessell, T.M., Wichterle, H., and Przedborski, S. (2007). Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. Nat Neurosci *10*, 615-622.

Neumann, M., Sampathu, D. M., Kwong, L. K., Truax, A. C., Micsenyi, M. C., Chou, T. T., Bruce, J., Schuck, T. *et al* (2006). Ubiquitinated TDP-43 in Frontotemporal Lobar Degeneration and Amyotrophic Lateral Sclerosis. Science *314* (5796): 130–3.

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

Peljto, M., Dasen, J.S., Mazzoni, E.O., Jessell, T.M., and Wichterle, H. (2011). Functional diversity of ESC-derived motor neuron subtypes revealed through intraspinal transplantation. Cell Stem Cell *7*, 355-366.

Peljto, M., and Wichterle, H. (2011). Programming embryonic stem cells to neuronal subtypes. Curr Opin Neurobiol *21*, 43-51.

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

Puia, G., Santi, M.R., Vicini, S., Pritchett, D.B., Purdy, R.H., Paul, S.M., Seeburg, P.H., and Costa, E. (1990). Neurosteroids act on recombinant human GABAA receptors. Neuron *4*, 759-765.

Soundararajan, P., Fawcett, J.P., and Rafuse, V.F. (2010). Guidance of postural motoneurons requires MAPK/ERK signaling downstream of fibroblast growth factor receptor 1. J Neurosci *30*, 6595-6606.

Soundararajan, P., Lindsey, B.W., Leopold, C., and Rafuse, V.F. (2007). Easy and rapid differentiation of embryonic stem cells into functional motoneurons using sonic hedgehog-producing cells. Stem Cells *25*, 1697-1706.

Soundararajan, P., Miles, G.B., Rubin, L.L., Brownstone, R.M., and Rafuse, V.F. (2006). Motoneurons derived from embryonic stem cells express transcription factors and develop phenotypes characteristic of medial motor column neurons. J Neurosci *26*, 3256-3268.

Szabo, E., Rampalli, S., Risueno, 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.

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

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Chapter 4

Promises of Reprogramming Technology

Abstract

Since the discovery of induced pluripotent stem cells (iPSCs), rapid progress has been made in manipulating cellular fate with defined factors. We and others have discovered small molecules capable of facilitating iPSC generation, adding to the molecular toolkit for inducing and dissecting the process of reprogramming. Transdifferentiation has produced terminally differentiated cells as well as more immature cells, and may provide multiple routes for producing a desired cell type. Improving the quality and efficiency of reprogramming, with methods for designing new conversions, could propel these technologies forward into therapeutic relevance. We propose that directly reprogrammed motor neurons may enable a rapid survey of a spectrum of ALS patients with the goal of classifying them into clinically distinct categories.

Insights from iPS Reprogramming Chemicals

Progress in Chemical Reprogramming

Our discovery of the *Sox2*-replacement molecule, RepSox, was one of the earliest demonstrations that small molecules could not only enhance iPS reprogramming efficiency, but also potently replace multiple reprogramming genes simultaneously, possibly through different mechanisms. Discovered from a functional complementation screen using a focused library of bioactive molecules, RepSox modulates a well-known signaling pathway, the Transforming Growth Factor-β (Tgf-β) pathway, in cells undergoing reprogramming. In particular, a subset of *Oct4*-, *Klf4*- and *cMyc*-infected, partially reprogrammed intermediate cells are highly responsive to RepSox: even with a pulse treatment, they rapidly upregulate *Nanog* and convert to fully reprogrammed iPSCs.

Of note, two other Tgf-β inhibitors could reportedly enhance four-factor iPSC generation in rat and human (Li *et al*., 2009; Lin *et al.,* 2009), indicating the involvement of Tgf-β signaling in related contexts. In these studies, inhibition of the MEK/ERK pathway, as well as activation of the Wnt pathway in the former case, were also found to facilitate the iPS conversion, and suggests that these pathways may also have relevance in reprogramming.

Interestingly, several inhibitors of epigenetic modulators have been reported to aid reprogramming; this indicates that, similar to SCNT, iPS reprogramming is hindered by the presence of epigenetic barriers. For example, valproic acid (VPA), an HDAC inhibitor (Huangfu *et al*., 2008a, b), increases both murine and human reprogramming efficiency and, at an extremely low rate (0.0001%), can replace *Klf4* and *Sox2* during

human iPSC generation. The DNMT inhibitor 5-azacytidine (Aza) can enhance late-stage murine reprogramming, even rescuing incompletely reprogrammed cells toward pluripotency (Mikkelsen *et al*., 2008). However, it is unclear whether global changes in chromatin structure or DNA methylation will have long-term adverse effects on the quality of iPSCs produced in this way.

Compounds of this functional class have also been found to cooperate with other molecules to replace one or more reprogramming factors: these include inhibitors of G9a histone methyltransferase (HMTase) (Shi *et al.*, 2008a), DNMTs (Shi *et al*., 2008b), HDACs (Zhu *et al*., 2010), and protein arginine methyltransferase (PRMT) (Yuan *et al*., 2011). Of note, the latter two are part of 6- and 2-chemical cocktails that reportedly enable reprogramming of human keratinocytes and MEFs, respectively, with *OCT4* alone; interestingly, both cocktails contain an inhibitor of Tgf-β signaling, A-83-10. However, these synergistic chemical combinations typically resulted in low efficiency reprogramming; more importantly, they have yet to be successfully replicated, suggesting that specific laboratory conditions might be required for their actions.

An outstanding goal in this area is to derived iPSCs with a purely chemical cocktail. Even though modified RNAs (Warren *et al*., 2010) or proteins (Zhou *et al.*, 2009) can generate transgene-free iPSCs, a chemical approach will be valuable for economic as well as logistical reasons. The success of identifying such a cocktail is likely to critically depend on finding a small molecule replacement for the central reprogramming factor, *Oct4*. Although an orphan nuclear receptor, *Nr5a2*, was shown to replace *Oct4* in murine reprogramming (Heng *et al*., 2010), no chemical has fulfilled this role in a reproducible manner. It remains to be seen whether a combination of approaches

– functional complementation screens, as well as luciferase screens for direct *Oct4* induction, for example – will achieve this milestone in the future.

Tgf-β Signaling and Reprogramming

The Tgf-β family of signaling molecules plays highly context-dependent roles in a plethora of processes ranging from development to cancer. Using specific neutralizing antibodies and other Tgf-β inhibitors, as well as kinase profiling, we showed that Tgf-β type I receptors, specifically ALK4/5 as well as ALK2, were the functional targets of RepSox during reprogramming. Thus, this well-known signaling pathway can also intersect with a highly artificial reprogramming system.

The precise molecular events that relate Tgf-β inhibition to a dramatic upregulation of *Nanog* in partially reprogrammed cells are unclear. One hypothesis is that RepSox tips the balance between Activin/Nodal and Bmp signaling – the two major arms of the Tgf-β superfamily that share a common requirement for Smad4 (Attisano and Wrana, 2002) – by specifically blocking the former. In mouse ESCs, Bmp signaling can activate *Nanog* in the presence of Stat3 (Suzuki *et al*., 2006). Another possibility is that RepSox treatment downregulates *Snail1/2* downtream of Tgf-β signaling, leading to de-repression of *E-cadherin* and its induction of of *Nanog*; these would be the reverse of the molecular changes induced by Tgf-β signaling during epithelial-mesenchymal transition (EMT) (Chou *et al.*, 2008; Willis *et al.*, 2008).

Given the importance of Tgf-β signaling in human ESC self-renewal (Xu *et al*., 2008), it is intriguing that its inhibition increases the efficiency of human iPS

reprogramming as well (Li *et al*., 2009; Lin *et al.,* 2009); it would be interesting to identify the contextual differences that would reconcile these observations.

Efforts to understand the full molecular mechanism of chemical-mediated reprogramming may reveal novel physical and functional interactions between signaling pathways and other regulators of the ESC state acting in *trans* or in *cis*. Such findings could help build a comprehensive algorithm for predicting reprogramming agents that include not only transcription factors but signaling pathway agonists and antagonists.

Partially Reprogrammed Cell States

The partially reprogrammed cell lines we isolated from iPS factor-transduced cultures responded in a repeatable and characteristic manner to reprogramming chemicals: a subset of them were in a RepSox-responsive state poised for *Nanog* activation and full reprogramming; others reprogrammed instead upon Aza treatment; still others did not respond to either chemical. These reproducible behavior patterns suggest that these cell lines are in distinct, non-overlapping states. Importantly, they are not representative of any naturally occurring cell type, yet maintain a stable phenotype.

Such non-productive reprogramming intermediates have been reported by other groups as well. Takahashi and Yamanaka found self-renewing, non-pluripotent colonies that had activated the *Fbx15::*β-geo cassette, when *Sox2* was omitted from the four-factor cocktail (Takahashi and Yamanaka, 2006). Meissner and colleagues also derived three partially reprogrammed lines from four-factor transduced MEFs and B cells (Mikkelsen *et al*., 2008); these lines were then extensively characterized with regard to their

differential propensities to reprogram, either spontaneously or in response to Aza with or without the knockdown of inappropriately expressed lineage-specifying genes.

Although sample size is limited, these observations in aggregate support an interesting prediction of the systems view of cell fate – the existence of numerous stable attractor states that are never accessed during development, due to the absence of natural trajectories that lead to them. In an intuitive picture, the cell state landscape is expected to be a highly complex terrain given the enormous number of inter-dependent variables required to describe cell state. During defined-factor reprogramming where a strong, non-physiological perturbation is provided, cells can gain access to unexplored parts of the landscape and encounter attractor states that would not otherwise be accessed. Hence, the trapped intermediate states would correspond to relatively stable, non-pluripotent attractor states from which a cell cannot escape; the precise configuration of each state would dictate what stimuli might rescue the cell to pluripotency.

These non-productive events during reprogramming are predicted by a computational model developed by Charkraborty and colleagues, where cell state is modeled as having a particular epigenetic signature associated with each of the developmentally important gene modules that they defined (Artyomov *et al.*, 2010). When reprogramming was simulated as an epigenetic change in a random module of the starting cell, followed by cell division cycles that permitted chromatin remodeling at each telophase, they found that reaching the pluripotent state was a rare, stochastically encountered outcome; dead-end loops, no-change, and even transdifferentiation were the other possibilities.

Interestingly, Jaenisch and colleagues showed that, when the cell-to-cell variability arising from differences in viral delivery is eliminated, the responses could be homogeneous over a long timescale (Hanna *et al*., 2009). They examined the reprogrammability of single 'secondary' pre-B cells from a clonal population, obtained from an iPSC-derived mouse with identical doxycycline-inducible lentiviral reprogramming transgenes in each cell. After 18 weeks of doxycycline administration, over 90% of the original cells had given rise to iPSC colonies. The efficiency and kinetics are consistent with a continuous, stochastic process that has one irreversible outcome among many reversible ones; given enough time, eventually the entire population will escape to the irreversible state.

Notably, major events that lead to non-productive intermediates seem to have been bypassed in this secondary system, where each cell has the same configuration of transgenes that permitted iPS reprogramming of the primary cells. In fact, some of our own partially reprogrammed lines can reprogram spontaneously, while others are highly resistant. Understanding the molecular basis for these different states of permissiveness could uncover roadblocks in reprogramming and delineate cell state trajectories. In a sense, these non-natural intermediates are offering a glimpse into the otherwise intractable regions of the cellular landscape.

Insights from Transdifferentiation

Progress in Transdifferentiation

The conversion of mouse and human fibroblasts into induced motor neurons (iMNs) is one of the first examples where a precisely defined neuronal type was produced by

transdifferentiation. In addition to extensive molecular and functional characterization *in vitro*, the engraftment capacity of mouse iMNs was demonstrated by their ability to integrate into the developing spinal cord of the chick embryo. We also demonstrated that human ALS patient-specific iMNs can be generated, as a complementary approach to using iPSC-derived motor neurons for downstream applications.

Notably, two groups have produced human induced neuronal (iN) cells from fibroblasts. Wernig and colleagues used *NEUROD1* on top of the three mouse iN factors (*Ascl1*, *Brn2* and *Myt1l*) to achieve 2-4% conversion of human fetal and postnatal fibroblasts into iN cells (Pang *et al*., 2011). On the other hand, Crabtree's laboratory achieved a similar result using micro-RNAs that control neuronal differentiation, miR-9/9* and miR-124, in conjunction with *NEUROD2* as well as *ASCL1* and *MYT1L* to enhance the efficiency (Yoo *et al.*, 2011).

From single-cell qPCR analyses and immunocytochemistry, it seems that both methods produce iN cells of predominantly cortical nature; however, the precise subtype identities represented in the iN cell populations remain unclear. Moreover, while iN cells made from postnatal sources exhibited many active membrane properties, including action potentials and postsynaptic currents (PSCs), only fetal fibroblast-derived iN cells showed response to GABA and L-glutamate, suggesting different levels of maturity.

To date, only one other specific type of neuron has been made directly from fibroblasts: the mesodiencephalic dopaminergic (DA) neurons that are lost in Parkinson's disease (PD) (Caiazzo *et al*., 2011; Pang *et al*., 2011; Pfisterer *et al*., 2011). Like motor neurons, their developmental pathway has been well characterized (Smidt and Burbach, 2007) given their clinical relevance, and key transcription factors that establish and

regulate the DA neuron fate are known. Three groups have each used a different cocktail chosen from DA neuron-specific factors and iN factors to generate putative induced dopaminergic (iDA) neurons, some more convincingly than others.

Parmar and colleagues supplemented the three iN factors with *Lmx1a* and *Foxa2* to convert over 5% of transduced human embryo-derived fibroblasts to tyrosine hydroxylase (TH)-expressing cells (Pfisterer *et al*., 2011). However, these cells only fired single action potentials upon depolarization, and no evidence of synapse formation was shown. Furthermore, there was limited molecular characterization and no functional demonstration pertaining to their specific identity.

A different combination of three factors, *Ascl1*, *Nurr1* and *Lmx1a*, was discovered by Broccoli and colleagues using *TH::*GFP reporter MEFs (Caiazzo *et al*., 2011). Mouse iDA neurons could be generated efficiently (15%) and exhibited many molecular and functional characteristics of DA neurons including release of dopamine. Significantly, human PD patient fibroblasts could also be converted to dopamine-releasing iDA neurons at 2% efficiency, although it is unclear whether they can produce multiple action potentials and respond to neurotransmitters.

Using a more extensive set of six factors – *Ascl1*, *Pitx3*, *Lmx1a*, *Nurr1*, *Foxa2* and *EN1* – Jaenisch and colleagues generated iDA neurons that most closely resemble endogenous DA neurons, albeit only in the mouse (Kim *et al*., 2011b). Remarkably, when transplanted into the brain of a mouse model of PD, they were able to engraft and extend TH+ projections into the deinnervated striatum by 4 weeks and increased the levels of dopamine in the area. At 8 weeks, iDA-transplanted PD mice had reduced amphetamine-

induced rotation scores, suggesting that these cells had a positive impact at the behavioral level.

The varying degree of functionality exhibited by the three flavors of putative iDA neurons may be a product of the precise factor combination, the starting cell type, or the peculiarities of different laboratories. We cannot rule out the possibility that some of the less functional iDA neurons represent incompletely reprogrammed states that are nonetheless closely related to the *bona fide* DA neuron state. It is intriguing to consider: if conditions could be strictly standardized, could the different factor cocktails induce equivalent states – that is, are there multiple ways of discovering the same DA neuron attractor state from the fibroblast state?

In answering this question, two recent reports of fibroblast-derived mouse induced hepatocytes (iHeps) of the endodermal lineage (Huang *et al.*, 2011; Sekiya and Suzuki, 2011) may provide additional insight. Starting from fibroblasts, each group used a distinct set of genetic conditions: *Gata4*, *Hnf1α*, and *Foxa3* with $p19^{Arf}$ inactivation; or, *Hnf4α* and one of *Foxa1*, *Foxa2* or *Foxa3*. The resulting iHeps were functional by a stringent criterion: they could engraft into the adult liver and partially rescue a genetic model of liver failure. Perhaps related to the incomplete rescue phenotype, they possessed global transcriptional programs similar, but not identical, to primary hepatocytes.

From the perspective of understanding reprogramming routes, a direct comparison would be useful between the two types of iHeps that seem to have many common functional properties, as well as a comparison of iDA neurons which may differ more greatly from each other. Similarly, we have noticed that *Hb9::*GFP+ iMNs can be produced with different factor permutations, which could provide an additional

experimental platform. Such studies, which could involve creating detailed genome- and epigenome-wide maps and carrying out quantifiable functional assays in parallel, may help to map out reprogramming trajectories and suggest ideas for improving the quality of reprogrammed cells.

Multiple Trajectories of Conversion

We have presented direct as well as circumstantial evidence against the appearance of progenitor-like intermediates during iMN reprogramming. The lack of BrdU incorporation or *Nestin* expression in the cells that give rise to iMNs argues that the Nestin+ proliferative neural progenitor state is not accessed. Although our results do not formally rule out other types of precursor intermediates, such as Olig2+ precursor cells, the fact that iMN generation is inhibited by progenitor factors makes a dedifferentiation-redifferentiation model unlikely.

As in iPS reprogramming, partially reprogrammed states may be stochastically encountered during the conversion, which may or may not lead to the fully reprogrammed states. The non-dividing nature of these cell types makes it currently difficult to isolate and characterize the possible unnatural intermediates, but the use of single cell-based methods may soon enable this type of analysis. However, findings from other lineage conversion paradigms, such as iβ-cell and iCM reprogramming, also suggest that the reprogramming factors seem to directly promote the final target cell state (Ieda *et al.*, 2010; Zhou *et al*., 2008). Although only speculative, bypassing a proliferative progenitor or stem-like state may be beneficial for preserving disease-causing genetic or epigenetic mechanisms during reprogramming.

Recently, two groups have made use of iPS reprogramming factors in transdifferentiation toward progenitor states. Bhatia and colleagues overexpressed a single gene, *OCT4*, in human dermal fibroblasts and cultured them with appropriate cytokines to generate cells expressing the pan-leukocyte marker CD45; these cells were capable of giving rise to the major non-lymphoid hematopoietic cell types (Szabo *et al*., 2010). This conversion does not involve the activation of pluripotency genes or embryonic hematopoietic genes, but occurs through the activation of a definitive hematopoietic program. The authors attributed the unexpected activity of *OCT4* in this context to its homology with *OCT1* and *OCT2*, which have roles in adult hematopoiesis.

However, two reports by Ding and colleagues suggest that a greater degree of stochasticity may be inherent in iPS factor-mediated transdifferentiation. They used the full cocktail of iPS factors – *Oct4*, *Sox2*, *Klf4*, and *cMyc* – with specific culture conditions to convert MEFs into contracting cardiomyocytes after 12 days (Efe *et al.*, 2011), or into bipotential induced neural progenitor cells (iNPCs) in a similar timeframe (Kim *et al.*, 2011a). The rapid kinetics and the culture media unsuited for iPSC generation, as well as molecular characterization of the reprogramming cell population, suggested that the pluripotent state is not reached during either conversion.

What might be the molecular basis for these surprising results? All iPS factors are known to bind many downstream targets (Sridharan *et al*., 2009), important among which are developmental regulators that are kept repressed in iPSCs and ESCs. During reprogramming, however, the pattern and outcome of iPS factor binding to *cis*-regulatory elements are likely to be imprecise, due to their high-level expression and the lack of other ESC-specific components. In turn, there can be stochastic activation of lineage-

specific genes, such as receptors and effectors for developmental morphogens, across the population. Therefore, providing a permissive environment for a somatic lineage but not iPSCs, as both laboratories have done, could capture and consolidate these transient states responsive to fate-inducing signals in the culture medium.

These observations can be directly related the aforementioned computational model from the Chakraborty group, which predicts that transdifferentiation into non-pluripotent states is one of the possible outcomes of perturbing a gene module at random by iPS factor overexpression (Artyomov *et al*., 2010). Given the extensive binding sites of these factors throughout the genome (Sridharan *et al*., 2009), enriched for developmentally important genes, a wide range of somatic states may be stochastically accessed in this way. This could help circumvent the need to generate individual viral constructs for new cell type-specific genes if the appropriate culture conditions could be found for accommodating the conversion to the desired cell type.

Since then, tripotent iNPCs and induced neural stem cells (iNSCs) have been produced by three different genetic cocktails. The former is a more canonical transdifferentiation approach using NPC-associated genes (Lujan *et al*., 2011); intriguigly, using *Sox2* and *FoxG1* generates bipotential iNPCs that are only neurogenic and astrogliogenic, whereas adding *Brn2* to the cocktail produces tripotential iNPCs that can also make oligodendrocytes, demonstrating that precise control over the identity of the target cell type is possible by this approach.

On the other hand, the iNSCs with self-renewing characteristics are produced using iPS factors with or without neural-specific genes (Han *et al*., 2012; Thier *et al*., 2012). It is interesting that *Oct4* expression is either limited to the first 5 days or omitted

altogether, while the remaining three iPS factors, which are expressed in NSCs, are used throughout in both studies. This suggests that, while stochasticity may play a part, the lineage-specific functions of the iPS factors are also being exploited in these processes.

Ultimately, the best trajectory for producing a desired cell type should be determined on a case-by-case basis. For making terminally differentiated neurons, it may sometimes be advantageous to first generate NPCs, using either of the above methods, or even iPSCs, followed by a partial or full recapitulation of development; these approaches that transit through proliferating cell states would be especially preferred if large quantities of cells are needed. In other cases, a direct transition to the final state that bypasses proliferative intermediate stages may be preferred, for example, for preserving unstable genetic elements in the starting cells or for producing a highly specific neuronal subtype. From these multiple independent approaches, it should be possible to develop an optimal protocol for generating the desired cell type for particular applications.

Future Challenges for Reprogramming

Improving the Efficiency

Several limitations must be addressed for transdifferentiation technologies to have a wide and meaningful impact in regenerative medicine. First is the low reprogramming efficiency of human cells relative to mouse cells, and of adult cells relative to immature cells: in our hands, mouse iMNs are routinely generated at 10-15% efficiency; in contrast, our best efficiency in generating patient-specific iMNs from primary adult human fibroblasts is around 5%. This seems to be a general pattern in induced

pluripotency as well as transdifferentiation, but is likely to be more severely limiting in the latter which produces a finite number of differentiated cells.

In this regard, optimizing the existing set of reagents could have immediate benefits. Systematically testing the factor combination and removing unnecessary or inhibitory genes has been shown to improve efficiency. Another integral component is the method of gene delivery: more potent viral vectors, or even modified RNAs, could be tested for possible improvement of efficiency. In addition, an accurate and efficient reporter system would be invaluable; for iMNs, using a longer *Hb9* promoter fragment to drive the reporter gene expression could help capture the maximum number of correctly reprogrammed cells.

To achieve an additional improvement in efficiency, it may be necessary to expand our reprogramming toolkit to include other classes of molecules. Micro-RNAs involved in determination and maturation of the target cell type are obvious candidates, similar to miR-9/9* and miR-124 in human iN cell generation (Yoo *et al*., 2011). Our experience in iPSC reprogramming also suggests that small molecules can be potent mediators of cell fate change; they can be discovered through unbiased screens (Ichida *et al*., 2009; Lyssiotis *et al*., 2009) or by educated guesses (Huangfu *et al*., 2008a). Similarly, morphogens and other signaling molecules could be added to this set.

In developing a complete protocol, combining the strengths of multiple approaches could prove beneficial. For example, Kaspar and colleagues combined defined-factor reprogramming with directed differentiation to generate motor neurons from human ESCs at 60-70% efficiency (Hester *et al*., 2011); in addition to the high-efficiency, this method may allow us to more precisely guide the fate of differentiating

cells into specific subtypes using additional factors. Of course, in the long-term, understanding the exact molecular nature of the difficulties in reprogramming adult human cells will be important for developing better strategies to overcome them.

Designing Novel Conversions

Another major challenge in lineage conversion is the initial identification of the optimum reprogramming condition. Although empirical determination by trial-and-error is likely to be always necessary, we may be able to assemble an efficient test set of candidate transcription factors using generalizable principles from existing studies. For transdifferentiation to a specific neuronal type, a good starting cocktail may be the iN factors supplemented by master regulators of the target state. Of note, most neuronal reprogramming paradigms to date seem to require *Ascl1* and, if applicable, factors central to the specific neuronal type. This suggests that a more complete cocktail may be built from a basal set of genes that are relatively easily predicted.

In this regard, insights from unbiased functional complementation screens for iPS reprogramming chemicals may be useful. If a screen had been performed using *Oct4*- and *Klf4*-infected MEFs with iPSC generation as the readout, without *a priori* knowledge of a gene that could fulfill the role of a positive control (*Sox2* in this case), it is likely that RepSox would still have been found. Similarly, kenpaullone might have been discovered as a chemical that could synergize with *Oct4* and *Sox2* to enable iPS reprogramming (Lyssiotis *et al*., 2009) even if *Klf4* had not been identified. Therefore, it may be possible to assemble a small set of target cell-specific factors, which do not accomplish the

desired cell fate change by themselves, to form the basis of a chemical screen to achieve full reprogramming.

Undoubtedly, developing algorithms for *in silico* prediction of reprogramming factors would improve our ability to navigate the cell state landscape. Global molecular profiling of known cell types provides reference maps from which high-quality regulatory networks might be constructed (Meissner *et al*., 2008; Mikkelsen *et al*., 2007). It remains to be seen whether these system-wide computational approaches can make predictions for new reprogramming paradigms that will pass experimental validation.

Using the Reprogrammed Cells

An immediate practical use of reprogrammed cells may be to model human disease in a dish. In particular, using transdifferentiated patient-specific neurons may be advantageous for the study of neurological disorders: first, it circumvents the generation of iPSCs; second, limited replication during reprogramming may preserve disease-associated epigenetic modifications or unstable genetic elements, such as hexanucleotide repeat expansions recently found to cause chromosome 9p21-linked ALS and frontotemporal dementia (FTD) (DeJesus-Hernandez *et al*., 2011; Renten *et al*., 2011). On the other hand, even if transdifferentiation efficiencies were vastly improved, the quantity of cells generated would be inferior to using directed differentiation of iPSCs.

Using iMNs for *in vitro* ALS modeling might be a useful case study, where we attempt to capitalize on the strengths of the transdifferentiation approach. In principle, a cohort of patient population could be screened for iMN phenotypes with relative ease, giving a preview of the disease characteristics associated with each genotype. Our small

sample of ALS patient iMNs which displayed high excitability suggests that lineage-reprogrammed cells may be a rapid means to reveal motor neuron phenotypes.

There are a number of ways in which patient iMNs may be evaluated in a standardized manner. First, morpholometric analyses of soma size, as well as the number and length of their neurites, could show gross morphological abnormalities arising in patient iMNs. Second, their electrophysiological properties – resting membrane potential, peak sodium and potassium currents, threshold voltage, and neurotransmitter response – could give important clues as to the nature of their functional defects. Third, the survival parameters of patient iMNs in the presence or absence of chemical and cellular inducers of stress, such as hydrogen peroxide or *SOD1*G93A glia (Di Giorgio *et al*., 2007; Nagai *et al*., 2007), could reveal pronounced disease-relevant phenotypes.

In addition, looking at the motor neuron-specific functionality of patient iMNs could be even more informative, and the two assays that we used in our study – muscle co-culture and transplantation – may be adapted for this purpose. Specifically, the number of neuromuscular junctions formed between an iMN and muscle can be quantified, as well as the amount of vesicle recycling at each synapse labeled by FM dyes (Gaffield and Betz, 2007). In transplantation assays, the degree of survival and the pattern of integration are parameters that could inform us of the subtype specificity of the iMNs as well as the ability to engraft (Soundararajan *et al*., 2006; Wichterle *et al*., 2002). However, in interpreting any results, careful attention should be paid as to whether an apparent phenotype is disease-related or is simply a result of sub-optimal reprogramming.

A main goal of this exercise would be to find criteria for classifying the patient population into phenotypic classes. One could imagine developing customized

therapeutics for each phenotypic class, then treating each newly diagnosed patient according to the behavior of their own iMNs. Also, the small-scale experiments performed with the iMNs could help identify the most important or representative genotypes to be further investigated by other methods using stem cells or animal models. If successful, this 'preview' approach would also be valuable for modeling other neurological spectrum disorders for which an appropriate neuronal cell type could be made by direct lineage conversion.

Concluding Remarks

Our work presented here contributes to a growing body of nuclear reprogramming studies, which collectively represents a significant step toward harnessing the latent plasticity of differentiated cells. The list of target cell types that can be generated by cellular conversion, and the human disease conditions that affect them, is growing rapidly. It is hoped that continued conceptual and technological advances in this area will eventually make an impact in diagnosing, treating and preventing human disorders.

References

Artyomov, M.N., Meissner, A., and Chakraborty, A.K. (2010). A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency. PLoS Comput Biol *6*, e1000785.

Attisano, L., and Wrana, J.L. (2002). Signal transduction by the TGF-beta superfamily. Science *296*, 1646-1647.

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.

Chou, Y.F., Chen, H.H., Eijpe, M., Yabuuchi, A., Chenoweth, J.G., Tesar, P., Lu, J., McKay, R.D., and Geijsen, N. (2008). The growth factor environment defines distinct pluripotent ground states in novel blastocyst-derived stem cells. Cell *135*, 449-461.

DeJesus-Hernandez, M., Mackenzie, I.R., Boeve, B.F., Boxer, A.L., Baker, M., Rutherford, N.J., Nicholson, A.M., Finch, N.A., Flynn, H., Adamson, J., *et al*. (2011). Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron *72*, 245-256.

Di Giorgio, F.P., Carrasco, M.A., Siao, M.C., Maniatis, T., and Eggan, K. (2007). Non-cell autonomous effect of glia on motor neurons in an embryonic stem cell-based ALS model. Nat Neurosci *10*, 608-614.

Efe, J.A., Hilcove, S., Kim, J., Zhou, H., Ouyang, K., Wang, G., Chen, J., and Ding, S. (2011). Conversion of mouse fibroblasts into cardiomyocytes using a direct reprogramming strategy. Nat Cell Biol *13*, 215-222.

Feng, B., Jiang, J., Kraus, P., Ng, J.H., Heng, J.C., Chan, Y.S., Yaw, L.P., Zhang, W., Loh, Y.H., Han, J., *et al*. (2009). Reprogramming of fibroblasts into induced pluripotent stem cells with orphan nuclear receptor Esrrb. Nat Cell Biol *11*, 197-203.

Gaffield, M.A., and Betz, W.J. (2006). Imaging synaptic vesicle exocytosis and endocytosis with FM dyes. Nat Protoc *1*, 2916-2921.

Han, D.W., Tapia, N., Hermann, A., Hemmer, K., Hoing, S., Arauzo-Bravo, M.J., Zaehres, H., Wu, G., Frank, S., Moritz, S., *et al*. (2012). Direct reprogramming of fibroblasts into neural stem cells by defined factors. Cell Stem Cell *10*, 465-472.

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.

Heng, J.C., Feng, B., Han, J., Jiang, J., Kraus, P., Ng, J.H., Orlov, Y.L., Huss, M., Yang, L., Lufkin, T., *et al*. (2010). The nuclear receptor Nr5a2 can replace Oct4 in the reprogramming of murine somatic cells to pluripotent cells. Cell Stem Cell *6*, 167-174.

Hester, M.E., Murtha, M.J., Song, S., Rao, M., Miranda, C.J., Meyer, K., Tian, J., Boulting, G., Schaffer, D.V., Zhu, M.X., *et al*. (2011). Rapid and efficient generation of functional motor neurons from human pluripotent stem cells using gene delivered transcription factor codes. Mol Ther *19*, 1905-1912.

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.

Huangfu, D., Maehr, R., Guo, W., Eijkelenboom, A., Snitow, M., Chen, A.E., and Melton, D.A. (2008a). Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol *26*, 795-797.

Huangfu, D., Osafune, K., Maehr, R., Guo, W., Eijkelenboom, A., Chen, S., Muhlestein, W., and Melton, D.A. (2008b). Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. Nat Biotechnol *26*, 1269-1275.

Ichida, J.K., Blanchard, J., Lam, K., Son, E.Y., Chung, J.E., Egli, D., Loh, K.M., Carter, A.C., Di Giorgio, F.P., Koszka, K., *et al*. (2009). A small-molecule inhibitor of tgf-Beta signaling replaces sox2 in reprogramming by inducing nanog. Cell Stem Cell *5*, 491-503.

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.

Kim, J., Efe, J.A., Zhu, S., Talantova, M., Yuan, X., Wang, S., Lipton, S.A., Zhang, K., and Ding, S. (2011a). Direct reprogramming of mouse fibroblasts to neural progenitors. Proc Natl Acad Sci U S A *108*, 7838-7843.

Kim, J., Su, S.C., Wang, H., Cheng, A.W., Cassady, J.P., Lodato, M.A., Lengner, C.J., Chung, C.Y., Dawlaty, M.M., Tsai, L.H., *et al*. (2011b). Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. Cell Stem Cell *9*, 413-419.

Li, W., Wei, W., Zhu, S., Zhu, J., Shi, Y., Lin, T., Hao, E., Hayek, A., Deng, H., and Ding, S. (2009). Generation of rat and human induced pluripotent stem cells by combining genetic reprogramming and chemical inhibitors. Cell Stem Cell *4*, 16-19.

Lin, T., Ambasudhan, R., Yuan, X., Li, W., Hilcove, S., Abujarour, R., Lin, X., Hahm, H.S., Hao, E., Hayek, A., *et al*. (2009). A chemical platform for improved induction of human iPSCs. Nat Methods *6*, 805-808.

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 U S A *109*, 2527-2532.

Lyssiotis, C.A., Foreman, R.K., Staerk, J., Garcia, M., Mathur, D., Markoulaki, S., Hanna, J., Lairson, L.L., Charette, B.D., Bouchez, L.C., *et al*. (2009). Reprogramming of

murine fibroblasts to induced pluripotent stem cells with chemical complementation of Klf4. Proc Natl Acad Sci U S A *106*, 8912-8917.

Meissner, A., Mikkelsen, T.S., Gu, H., Wernig, M., Hanna, J., Sivachenko, A., Zhang, X., Bernstein, B.E., Nusbaum, C., Jaffe, D.B., *et al*. (2008). Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature *454*, 766-770.

Mikkelsen, T.S., Ku, M., Jaffe, D.B., Issac, B., Lieberman, E., Giannoukos, G., Alvarez, P., Brockman, W., Kim, T.K., Koche, R.P., *et al*. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature *448*, 553-560.

Nagai, M., Re, D.B., Nagata, T., Chalazonitis, A., Jessell, T.M., Wichterle, H., and Przedborski, S. (2007). Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. Nat Neurosci *10*, 615-622.

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

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

Renton, A.E., Majounie, E., Waite, A., Simon-Sanchez, J., Rollinson, S., Gibbs, J.R., Schymick, J.C., Laaksovirta, H., van Swieten, J.C., Myllykangas, L., *et al*. (2011). A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron *72*, 257-268.

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

Shi, Y., Desponts, C., Do, J.T., Hahm, H.S., Scholer, H.R., and Ding, S. (2008a). Induction of pluripotent stem cells from mouse embryonic fibroblasts by Oct4 and Klf4 with small-molecule compounds. Cell Stem Cell *3*, 568-574.

Shi, Y., Do, J.T., Desponts, C., Hahm, H.S., Scholer, H.R., and Ding, S. (2008b). A combined chemical and genetic approach for the generation of induced pluripotent stem cells. Cell Stem Cell *2*, 525-528.

Soundararajan, P., Miles, G.B., Rubin, L.L., Brownstone, R.M., and Rafuse, V.F. (2006). Motoneurons derived from embryonic stem cells express transcription factors and develop phenotypes characteristic of medial motor column neurons. J Neurosci *26*, 3256-3268.

Sridharan, R., Tchieu, J., Mason, M.J., Yachechko, R., Kuoy, E., Horvath, S., Zhou, Q., and Plath, K. (2009). Role of the murine reprogramming factors in the induction of pluripotency. Cell *136*, 364-377.

Suzuki, A., Raya, A., Kawakami, Y., Morita, M., Matsui, T., Nakashima, K., Gage, F.H., Rodriguez-Esteban, C., and Izpisua Belmonte, J.C. (2006). Nanog binds to Smad1 and

blocks bone morphogenetic protein-induced differentiation of embryonic stem cells. Proc Natl Acad Sci U S A *103*, 10294-10299.

Szabo, E., Rampalli, S., Risueno, 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.

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

Warren, L., Manos, P.D., Ahfeldt, T., Loh, Y.H., Li, H., Lau, F., Ebina, W., Mandal, P.K., Smith, Z.D., Meissner, A., *et al*. (2010). Highly efficient reprogramming to pluripotency and directed differentiation of human cells with synthetic modified mRNA. Cell Stem Cell *7*, 618-630.

Wichterle, H., Lieberam, I., Porter, J.A., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385-397.

Willis, B.C., and Borok, Z. (2007). TGF-beta-induced EMT: mechanisms and implications for fibrotic lung disease. Am J Physiol Lung Cell Mol Physiol *293*, L525-534.

Xu, R.H., Sampsell-Barron, T.L., Gu, F., Root, S., Peck, R.M., Pan, G., Yu, J., Antosiewicz-Bourget, J., Tian, S., Stewart, R., *et al*. (2008). NANOG is a direct target of TGFbeta/activin-mediated SMAD signaling in human ESCs. Cell Stem Cell *3*, 196-206.

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.

Yuan, X., Wan, H., Zhao, X., Zhu, S., Zhou, Q., and Ding, S. (2011). Brief report: combined chemical treatment enables Oct4-induced reprogramming from mouse embryonic fibroblasts. Stem Cells *29*, 549-553.

Zhou, H., Li, W., Zhu, S., Joo, J.Y., Do, J.T., Xiong, W., Kim, J.B., Zhang, K., Scholer, H.R., and Ding, S. (2010). Conversion of mouse epiblast stem cells to an earlier pluripotency state by small molecules. J Biol Chem *285*, 29676-29680.

Zhou, H., Wu, S., Joo, J.Y., Zhu, S., Han, D.W., Lin, T., Trauger, S., Bien, G., Yao, S., Zhu, Y., *et al*. (2009). Generation of induced pluripotent stem cells using recombinant proteins. Cell Stem Cell *4*, 381-384.

Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J., and Melton, D.A. (2008). In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. Nature *455*, 627-632.