REPROGRAMMING TO THE NERVOUS SYSTEM: A COMPUTATIONAL AND CANDIDATE GENE APPROACH

By

Bradly John Alicea

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Animal Science - Doctor of Philosophy

2013

ABSTRACT

REPROGRAMMING TO THE NERVOUS SYSTEM: A COMPUTATIONAL AND CANDIDATE GENE APPROACH

By

Bradly John Alicea

The creation of stem-like cells, neuronal cells, and skeletal muscle fibers from a generic somatic precursor phenotype has many potential applications. These uses range from cell therapy to disease modeling. The enabling methodology for these applications is known as direct cellular reprogramming. While the biological underpinnings of cellular reprogramming go back to the work of Gurdon and other developmental biologists, the direct approach is a rather recent development. Therefore, our understanding of the reprogramming process is largely based on isolated findings and interesting results. A true synthesis, particularly from a systems perspective, is lacking. In this dissertation, I will attempt to build toward an intellectual synthesis of direct reprogramming by critically examining four types of phenotypic conversion that result in production of nervous system components: induced pluripotency (iPS), induced neuronal (iN), induced skeletal muscle (iSM), and induced cardiomyocyte (iCM). Since potential applications range from tools for basic science to disease modeling and bionic technologies, the need for a common context is essential.

This intellectual synthesis will be defined through several research endeavors. The first investigation introduces a set of experiments in which multiple fibroblast cell lines are converted to two terminal phenotypes: iN and iSM. The efficiency and infectability of cells subjected to each reprogramming regimen are then compared both statistically and quantitatively. This set of experiments also resulted in the development

of novel analytical methods for measuring reprogramming efficiency and infectability. The second investigation features a critical review and statistical analysis of iPS reprogramming, specifically when compared to indirect reprogramming (SCNT-ES) and related stem-like cells. The third investigation is a review and theoretical synthesis which stakes out new directions in our understanding of the direct reprogramming process, including recent computational modeling endeavors and results from the iPS, iN and induced cardiomyocyte (iCM) experiments.

To further unify the outcomes of these studies, additional results related to Chapter 2 and directions for future research will be presented. The additional results will allow for further interpretation and insight into the role of diversity in direct reprogramming. These future directions include both experimental approaches (a technique called mechanism disruption) and computational approaches (preliminary results for an agent-based population-level approximation of direct reprogramming). The insights provided here will hopefully provide a framework for theoretical development and a guide for traditional biologists and systems biologists alike.

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

KEY TO SYMBOLS AND ABBREVIATIONS
(gene and protein names italicized according to convention)

α-ACT - alpha-actinin

Acsl1 - acyl-CoA synthetase long-chain family member 1

CA – cellular automata

c-Myc – regulator gene that codes for transcription factor activity

Col1α2 - collagen type 1a2 gene

DMEM - Dulbecco's Modified Eagle Medium

ES, ESC – embryonic stem cell

FEE - factor expression (early)

FEL - factor expression (late)

Fibr1 - fibrillin I

Fibu5 - fibulin V

FoxG1 – forkhead box G1

Fnectin - fibronectin gene

Gata4 – transcription factor that binds to GATA sequences

GEO – Gene Expression Omnibus

H3K4 – histone H3 at Lysine 4

H3K27 – histone H3 at Lysine 27

Hand2 – heart- and neural crest derivatives-expressed protein 2

HBK – Haken-Bunz-Kelso model of bistability

HE - heart (mouse fibroblasts)

iCM – induced cardiac muscle (cell)

iN – induced neuron (cell)

iPS – induced pluripotent stem (cell)

iSM – induced skeletal muscle (fiber).

Ink – *Ink* protein family of cell cycle regulators

Ker14 - keratin 14

KI - kidney (mouse fibroblasts)

Klf4 – kruppel-like factor 4

LI - liver (mouse fibroblasts).

LU - lung (mouse fibroblasts)

MAP Kinase (*Map2*) – mitogen-activated protein kinase

Mef2C – myocyte-specific enhancer factor 2C

Mesp1 – mesoderm posterior protein 1

MMLU – maloney murine leukemia virus

MOI – multiplicity of infection

MyoD1 - myogenic differentiation 1

Myf5 - myogenic factor 5

Myf6 - myogenic factor 6

Myt1L - myelin transcription factor 1-like

NANOG – homeobox protein known to regulate pluripotency

NeuroD - neurally-expressed basic helix loop helix transcription factor

Nkx2-5 – homeodomain-containing transcription factor.

NPC - neural progenitor cell

NR5A2 - Nuclear receptor subfamily 5, group A, member 2 (protein-coding gene)

Oct4 – octamer binding transcription factor 4

OKSM – *Oct4/Klf4/Sox2/c-Myc* retrovirus combination

ORF – open reading frame

p53 – tumor protein 53

PCA – principle componenet analysis

PECAM – endothelial factor

Pou3f2 - POU class 3 homeobox 2

qPCR – quantitative real-time PCR

Ras – Ras subfamily of GTPase proteins

RET - fibroblast cell line derived from Rett's Syndrome patient

SAF - fibroblast cell line derived from Schizophrenia patient

SCNT – somatic cell nuclear transfer, a technique for indirect reprogramming

SM - skeletal muscle (mouse fibroblasts)

Sox2 – sex-determining region Y-box 2

STM – Stochastic Threshold Model, a model for direct reprogramming

SYMO – sarcomyosin

Syn1 – synapsin 1

TA - tail tip (mouse fibroblasts)

Tbx5 – T-box gene, transcription factor

TE - testes (mouse fibroblasts)

TTX – tetrodotoxin

Tuj1 - Neuronal Class III β-Tubulin

YFP – yellow florescent protein

Zic1 - Zinc finger protein (Zic family member 1)

ZAP - *zic1/ascl1/pou3f* retrovirus combination

All cell lines abbreviations (human and mouse) are further defined in Figure 2.1.

CHAPTER 1: TOWARDS THE UNIFIED PRINCIPLES OF DIRECT CELLULAR REPROGRAMMING

1.0 Introduction

Cellular reprogramming via direct means, such as the virally-mediated induction of a generic somatic cell to a desired phenotype, is a potentially powerful but partially understood process. As a biological technique, cellular reprogramming has created opportunities for medical and biotechnological applications, particularly in the field of neuroscience. However, there are several aspects of the reprogramming process that are not well understood from a systems perspective. These aspects include reprogramming efficiency, input cell diversity, and the role of biological dynamics. While this dissertation will not provide a unified theory of reprogramming, each of these issues will be examined in detail. The second chapter will focus on reprogramming efficiency and input cell diversity. The third chapter will focus on comparing direct reprogramming with indirect reprogramming, in addition to understanding what these differences mean in the context of other stem-like cells. The fourth chapter will focus on reprogramming efficiency and biological dynamics. The final chapter will summarize these results and spell out directions for future research. By utilizing a variety of approaches, this dissertation will hopefully provide much needed context for understanding direct reprogramming at the systems-level.

1.1 Research Aims

The focus of this dissertation is on the basic biology of different types of induced cells, from pluripotent to neuronal. This dissertation will break ground in two ways. First, it will provide an informal comparison between different induced cell phenotypes

that may be used in the nervous system. These include induced pluripotent (iPS), induced neuron (iN), and induced skeletal muscle (iSM) cells. Second, it will provide a theoretical context for the large body of research on induced cell types that already exists. This will be accomplished in part by addressing three aims:

1) To examine the process of phenotypic respecification to two different destination phenotypes (iNC, iSMC), which provides novel information about the "architecture" of cellular reprogramming.

2) Through the use of novel comparisons and bioinformatics techniques, to examine the process of phenotypic respecification to the same phenotype using different methods (SCNT, direct reprogramming), which can place specific reprogramming regimens into a generalized context.

3) Re-interpret cellular reprogramming as a complex, dynamical system through the use of dynamic and population-level models that capture the underlying structure of large-scale phenotypic change.

These aims lead to three hypotheses that focus the work presented here around three general themes: the role of biological diversity in the reprogramming process, an evaluation of direct reprogramming with respect to indirect methods (e.g. SCNT), and the potential of dynamic and population-level models to elucidate the reprogramming process. The hypotheses are as follows:

H1) A comparison of reprogramming regimens that result in the production of induced neurons (iNCs) and induced skeletal muscle (iSM) fibers will allow us to better understand the role of biological diversity in the reprogramming process.

H2) Novel comparisons and bioinformatics techniques will allow us to uncover subtle yet important differences between cells reprogrammed to the same phenotype (pluripotency) using different methods (SCNT, direct reprogramming).

H3) Tested over time, dynamic and population-level models will allow for robust theories that characterize possibilities and limitations afforded by the underlying structure of the reprogramming process.

This dissertation consists of three papers, which are Chapters 2, 3, and 4, each of which provide a different perspective on direct reprogramming, and provide support for each stated aim. These chapters will deal with the following themes: the role of diversity in cellular reprogramming, comparing indirect and direct reprogramming, and new directions in direct reprogramming.

1.1.1 Role of diversity in cellular reprogramming. Chapter 2, entitled "Defining phenotypic respecification diversity using multiple cell lines and reprogramming regimens", includes a comparison between cell lines and reprogramming regimens (in this case, the creation of iN cells with the creation of iSM cells). A quantification method for the reprogramming efficiency and infectability variables used in the first paper will also be introduced. The findings in this paper support Aim #1, and make a contribution towards uncovering this fundamental architecture.

1.1.2 Comparing indirect and direct reprogramming. The second paper is entitled "Comparing indirect-derived ES cells and directly-derived iPS cells", and is a book chapter that both compares pluripotent cells created using indirect means (SCNT-ESCs) with iPS cells and places observed differences between these cell types in a statistical context. The comparison between indirect and direct reprogramming, along

with the meta-comparison between stem-like cell lines, contributes towards the second aim.

1.1.3 New directions in direct reprogramming. The third paper (an invited review) is entitled "New Directions in Cellular Reprogramming", and provides a one-of-a-kind review of systems biology and computational modeling approaches relevant to direct reprogramming. This paper provides a view of direct reprogramming as one of a complex, dynamical system. Such a viewpoint has the potential to provide insight into some of the mysteries surrounding the timing and overall kinetics of direct reprogramming, even in the absence of direct experimental observation.

1.2 Direct Reprogramming: diversity in a general mechanism

Direct cellular reprogramming has a brief but high-profile history (Jaenisch, 2012). Vierbuchen and Wernig (2012) discuss this history in detail, along with the variety of cell types that are possible using the direct reprogramming method. By using a relatively straightforward approach (delivery of a few key transcription factors and phenotype-specific culture medium), many different types of cell can be made from any given somatic cell type. To demonstrate this, a comparison of selected attributes amongst four major induced cell types (iPS, iN, iSM, and induced cardiomyocyte - iCM) can also be seen in Table 1.1. The direct reprogramming of readily cultured and maintained cell types such as fibroblasts to clinically important tissue types such as neurons, cardiac and skeletal muscle, hepatocytes, and pluripotent stem cells has tremendous promise in the understanding, diagnosis, and future treatment of many currently intractable human diseases (Patel and Yang, 2010; Kiskinis and Eggan, 2010). Despite the fact that many advances have been made into understanding *how* to

convert cells to alternate phenotypes (Caiazzo et al., 2011; Yoo et al., 2011), very little is known about which properties of input cell lines and underlying biochemical process contribute to the efficiency of this process. This is especially true of the parallels between different instances of reprogramming.

Table 1.1. Comparison of different attributes amongst different induced cell types.

	iPS	iN	iSM	iCM
Factors	*Oct4, Sox2, Klf4, c-Myc*[1]	*Ascl1*, Pou3f, *Zic1, Myt1L*	*MyoD, Myf5*, Myogenin, *Myf6*	*Gata4, Mef2C, Tbx5, Hand2*[2]
First Reported	2007	2011	1987[3]	2010
Reprogrammed *in vivo*	No	No	No	Yes
Self-renewing (immortalization)	Yes	No	No	No
Excitable (action potentials)	No	Yes	Partial	Yes
Exhibit functional integration	Yes	Partial	N/A	Yes

[1] Not essential. [2] Greatly enhances efficiency. [3] Done with a single factor (*MyoD*).

The initial condition of the input cells can introduce a great degree of variation in both the overall efficiency and other outcomes of the reprogramming process. Prior studies (Hanna et al., 2009) have highlighted how initial population size and the nature of input cellular material can affect the timing of key events and overall rate in the reprogramming process. Related to these outcomes is the inherent stochasticity of the direct reprogramming process, particularly with regard to fluctuations. High levels of variation can either occur independently or interact with large-scale phenotypic remodeling (Samoilov, Price, and Arkin, 2006). Of particular interest are the stochastic

effects of cell cycle and proliferation (Raj and van Oudenaarden, 2008) on the resulting infection and reprogramming efficiency of cells.

The research presented here will contribute to the existing literature by unifying work done on several induced cell types. While the factors used to directly reprogram somatic cells to a specific phenotype act to induce changes in different gene regulatory networks, the overall process of reprogramming is thought to be similar across contexts. We will now review the essentials of iPS, iN, and iSM phenotype creation, with an emphasis on the parallels of the different instances of direct reprogramming.

1.2.1 Short overview of iPS cells. The creation of pluripotent cells from somatic cells involves a resetting of the original cellular state. This has its roots in indirect reprogramming, starting with the experiments of John Gurdon (Gurdon and Byrne, 2003; Jaenisch, 2012). Attempts at achieving greater control over this process finally bore fruit in the mid-2000s, with the successes of the Thomson (Yu et al., 2007) and Yamanaka (Takahashi et al., 2007) groups. In these experiments, four to five transcription factors were delivered to somatic cells using virally-mediated techniques.

The transcription factors that enable iPS reprogramming act in a manner similar to master regulators (Khan et al., 2012) in the sense that a few key transcription factors may trigger a large number of gene expression, epigenetic, and other regulatory changes. This is also generally true for other types of direct reprogramming (e.g. iN, iSM, iCM). In fully reprogrammed cells (a fraction of the original population), continuous endogenous experession of these factors these lead to wholesale changes in cell phenotype. This has been understood in terms of regulatory network architectures that

6

also apply to the direct reprogramming of iN cells and iSMs (Artyomov, Meissner, and Chakraborty, 2010; MacArthur, Maayan, and Lemischka, 2009; Buganim et al., 2012).

1.2.2 Short overview of iN cells. The origins of creating neuronal cells from a somatic cell also involve resetting the original cellular state. Previous to direct techniques (see Zhang et al., 2001), neural precursors were required to create mature neural cells (e.g. neurons, astrocytes, and glia). However, using a combination of transcription factors that have been identified as diagnostic markers of neural precursor cells (Lujan et al., 2012) and genetic screens (Vierbuchen et al., 2010), it became possible by 2010 (Vierbuchen et al., 2010; Pang et al., 2011; Yang et al., 2011) to create inducible neural cells. As with iPS cells and iSMs, the delivery of transcription factors stimulates their endogenous upregulation (Soufi, Donahue, and Zaret, 2012). Sustained endogenous regulation of these key factors is the hallmark of a fully-reprogrammed cell. Unlike iPS cells, iNs do not form tightly-integrated colonies, although their true functional requirements are far from known at this point.

In the case of our experimentation (see Chapter 2), iN cells are mostly glutamitergic neurons with generic processes. More generally, the direct reprogramming approach allows us to create iN cells with many different neural attributes (see Vierbuchen and Wernig, 2012). This diversity of target phenotypes is made possible through the combinatorial use of key neuron-related transcription factors. These cell types include: dopaminergic neurons (using 6 factors), motor neurons (using 5 factors), Glutamitergic + GABAergic (using 3 factors), and neural progenitors (using 3 factors). Furthermore, the production and transplantation of dopaminergic iN cells under defined conditions is also possible (Kim et al., 2011; Caiazzo et al., 2011).

1.2.3 Short overview of iSM cells. The direct production of skeletal muscle is actually the first successful attempt at direct reprogramming. The work of Lassar (1987) demonstrated that *MyoD* delivered using cDNA molecules is sufficient to respecify somatic cells to skeletal muscle fibers. This experiment can be replicated using four transcription factors delivered via viral-mediation (see Chapter 2). While these fibers are not always contractile, they are phenotypically identifiable as skeletal muscle fibers. Other experiments circa 2010 (Ieda et al., 2010) and 2012 (Qian et al., 2012) show how iCM cells can also be produced using a 3-4 transcription factor combination. The effectiveness (and lack of effectiveness) of these factors have been demonstrated using genetic screens (Ieda et al., 2010) and *in vivo* validation (Qian et al., 2012; Song et al., 2012). This has potential for transplantation and other applications (Park et al., 2008a; Soldner and Jaenisch, 2012).

1.3 Applications to Nervous Systems: from transplantation to models

The direct reprogramming approach can be used to the better understand and perhaps even repair nervous systems. These applications can range from disease models (Kiskinis and Eggan, 2010) to transplantation (Chambers and Studer, 2011). Personalized, reprogrammed cell-based disease models (for concept, see Wieland and Fussenegger, 2012) can be created from abundant source of patient-specific somatic cells such as fibroblasts. The ability to derive cells with the same genetic background as a disease sufferer can allow us to observe the development and progression of disease phenotypes and regenerative processes both in context and *in vitro.* With the advent of next-generation sequencing techniques such as whole-genome RNA sequencing (RNA-seq, see Grabherr et al., 2011), exonuclease chromatin immunoprecipitation (ChIP-Exo,

see Rhee and Pugh, 2011), and whole-genome bisulfite sequencing (BS-seq, see Varley and Mitra, 2010), these cell populations can be understood as never before.

While cell therapy has a checkered history, there are also opportunities for using directly reprogrammed cells to restore function through the transplantation of cells with a respecified phenotype. While this has obvious medical applications, it may also help us understand how cells with a respecified phenotype integrate themselves into functional tissues and existing neuronal networks.

For proper application of directly reprogrammed cells to the nervous system, direct reprogramming must be better understood as a process. This can be done through the use of more precise quantitative measurements, large-scale and meta-analyses, and simulation-based approaches. We will now turn to the regulatory and biochemical processes of reprogramming, particularly as they relate to the mechanisms that trigger a respecified phenotype.

1.4 Mechanisms behind phenotypic conversion

This dissertation will focus on the potential differences between different types of input cell, different types of reprogrammed cell, and the mechanisms behind these differences. There are three themes in the literature that not only suggest differences in reprogramming capacity do exist between cellular populations, but also point to the existence of a complex mechanism. One is work from the 1970s and 1980s on the preferential conversion of 3T3 cells to an adipose cell fate (Green and Kehinde, 1976; Nixon and Green, 1983). In these experiments, the spontaneous conversion of 3T3 cells in culture was found to be highly variable and exhibited a preference for some cells over others. Later work also in 3T3 cells isolated the mechanism of adipose differentiation to

the differential expression of growth hormone receptors between isolated cell populations (Morikawa, 1986; Corin, Guller, Wu, and Sonenberg, 1990). In the case of reprogramming to the induced pluripotent stem (iPS) cell type using a retroviral approach, Polo et al. (2010) demonstrated that mouse cells from different sources (fibroblast, hematopoietic, and myogenic) will exhibit differential potentials for embryoid body formation and differentiation. This "differential potential" was largely found to be transient (e.g. disappeared with regard to time spent in the iPS state), but was based on indicators such as transcriptional and epigenetic similarity. More recent work suggests that successful phenotypic respecification is a rare event (Brambrink et al., 2008; Hanna et al., 2008) that involves the combined effects of cell line-specific behavior driven by stochastic processes.

1.4.1 Viral-mediated approaches. Although direct reprogramming can be done in a number of ways (RNA, etc), the most traditional (and reliable) technique is viral mediation (Bouard, Alazard-Dany, and Cosset, 2009). Figure 1.1 shows an example of the viral vector used in this method. Briefly, a series of transcription factors are included on a cassette which is then packaged on a retroviral vector. The *in vitro* cell population is exposed to a concentration of viral particles which infect the cells in culture. Upon exposure, individual cells may or may not take up the virus. The efficiency of this process is called the infectability of a cell population (Centlivre et al., 2011). In turn, cells that take up the viral particles and initially express high levels of the delivered transcription factors may or may not endogenously express the delivered factors. This fraction of the population (that exhibit endogenous expression and related phenotypic changes) contributes to the reprogramming efficiency of a cell population. Each regimen

involves the delivery of factors identified to be most critical in conversion to the desired phenotype. For example, conversion to iPS cells involves delivery of *Oct4*, *Sox2*, *Klf4*, and *c-Myc* (Stadtfeld and Hochedlinger, 2010). Although a four-factor recipe is shown in Figure 1.1 for each type of conversion, the actual number of factors used may be variable.
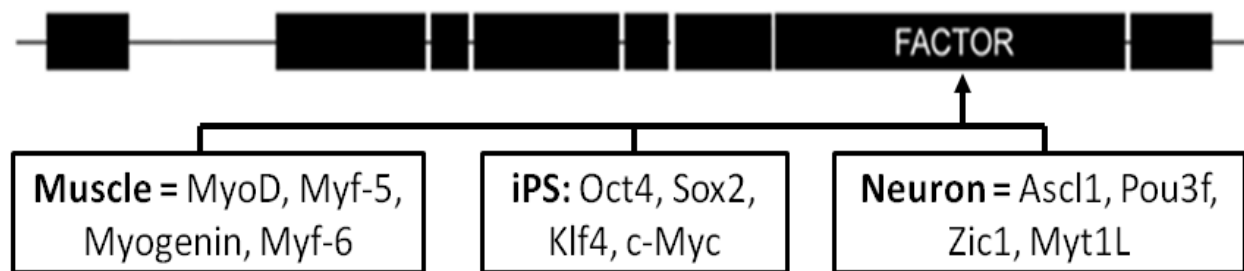
Figure 1.1. Generalized polycistronic vector with a cassette (labeled factor) that is interchangable with three four-factor regimens.

1.5 Direct cellular reprogramming as a stochastic, complex system.

To pursue systems-level computational modeling and simulation work, we must better understand the structure of direct cellular reprogramming. These structural underpinnings consist of an unpredictable process which is embedded within the complexity of the cell. Evidence of this can be seen in how slow kinetic (Lin et al., 2009) and stochastic mechanisms (Alfano, 1998) are used to characterize cellular reprogramming (Allen, 2003; Wilkinson, 2006) resemble a critical process (Bak, Tang, and Weisenfeld, 1988). Unlike protein folding, which exhibits fast kinetics and operates on a timescale of femto- to nano-seconds (Didiera et al., 2008; Karplus, 1999), reprogramming from one cellular state to another involves processes that occur at timescales ranging from that of protein folding to hours and even days (Brambrink et al., 2008; Stadtfeld et al., 2008). Because of this structure, cellular reprogramming can

be considered a course-grained phenomenon (Riniker, Allison, and van Gunsteren, 2012).

One consequence of this complexity at multiple scales is that large-scale phenotypic changes are initiated but not controlled by a small number of specific manipulations. These manipulations result in the initiation of hierarchical regulatory cascades (Erwin and Davidson, 2009), the propagation of which require changes above a given threshold and occur in bursts (Watts, 2002). This not only results in avalanche-like changes, but also implies that reprogramming to any one cell type (e.g. iPS) can occur through a large number of possible pathways (see Artyomov, Meissner, and Chakraborty, 2010). As an example, Loh and Lim (2011) have proposed that the delivered transcription factors (e.g. *Oct4*, *Sox2*, *Klf4*, *c-Myc*) often act in a combinatorial fashion to produce a mosaic of outcomes, from fully-reprogrammed cells to partially reprogrammed cellular phenotypes and apoptotic cells. This explains observations of low rates of conversion efficiency and a highly-variable response across replicates.

The known biochemical bases of direct cellular reprogramming demonstrate that that there are many potential questions regarding the complexity of direct cellular reprogramming that are only beginning to be defined. To better understand the role of complexity in the direct reprogramming process, we will now review three interrelated issues: the measurement of critical events at many levels of cellular regulation and the direct reprogramming process, the acquisition of cellular state as a critical point in the reprogramming process, and the treatment of critical changes in the cell due to cellular reprogramming as cellular switches.

1.5.1 Measurement of critical events and the direct reprogramming process. Chan et al. (2009), Mikkelsen et al. (2008), Liu et al. (2007), and Wadhwa, Kaul, and Mitsui (1999) have shown that the state of a reprogrammed cell observed at any one timepoint is the product of large number of undefined events that can be approximated by measurements of microRNA activity (Xu et al., 2009), immunohistochemistry (Chan et al., 2009), gene expression (Tang, 2008), and population dynamics (Mantzaris, 2006). In the literature on iPS direct reprogramming, these are characterized as discrete barriers, whether they involve changes in the expression of single genes (Hong et al., 2009), specific epigenetic changes (Maherali et al., 2007; Pasini et al., 2010), or the activation of an extended transcriptional network (Kim et al., 2008).

Yet not all mechanisms are created equal. Some candidate mechanisms that enable or enhance direct reprogramming have a global effect on phenotypic state, while the effect of others is limited to specific points in the reprogramming process. Maherali et al. (2007) and Theunissen et al. (2010) provide one example of a global mechanism by suggesting that Nanog expression alters the activation of *Oct4* and *Sox2* as well as the speed of reprogramming independently of proliferation rate. Evidence from single-cell analysis also suggests that the reprogramming process is marked by different phases (early vs. late) which can be identified by distinct regulatory regimes (Buganim et al., 2012).

1.5.2 Acquisition of cellular state as a critical point in reprogramming. One example of a critical point in the iPS reprogramming process is the acquisition of stemness. Parallel processes in iN and iSM reprogramming might be the extension of neuronal processes and cell body fusion, respectively. Properties of the mature iPS phenotype and other

stem-like cells are collectively referred to in the literature as stemness (Mikkers and Frisen, 2005; Orford and Scadden, 2008; Gokhale and Andrews, 2008; Spivakov and Fisher, 2007). Zipori (2004) argue that these properties define a cellular state, comparable to fibroblast-ness, neuron-ness, or muscle-ness. Stemness has traditionally been defined in terms of embryonic stem cells (ESCs) and 8-cell embryos. While there are many differences between different types of stem-like cell (see Chapter 3 for more information), stemness defines what properties make them distinct from somatic cells.

Given the great diversity among cells of a particular state (Sridharan et al., 2009; Fleming et al., 2008; Xie et al., 2010), it may appear difficult to find commonalities that define the boundaries of cellular state. In the case of stemness, there are indeed several overarching features of pluripotency. In Mikkers and Frisen (2005), stemness is defined as the stable suspension of cells in a specific developmental stage where cells do not leave the cell cycle. The stage is marked by expression and upregulation of *Oct4* and Nanog (Mikkers and Frisen, 2005). In Orford and Scadden (2008), stemness is defined as the continuation of the cell cycle caused by the downregulation of proteins of the Rb family and results in the global downregulation of cell specific genes. In Mikkers and Frisen (2005) and Gokhale and Andrews (2008), stemness is defined as by cells within a definable niche. It is interaction with the local environment that is thought to be critical for the maintenance of stemness (Mikkers and Frisen, 2005). The work of Ramalho-Santos (2004), Zipori (2004), Ivanova et al. (2006), and Azuara et al. (2006) provide us with other views on stemness and what unites stem-like cells as a single cell type.

These observations also apply to iPS cells, with the exception that the acquisition of stemness requires a reprogramming process where thresholds must be overcome on the way to acquiring these hallmarks. In this sense, a fully reprogrammed iPS cell is subject to several rounds of selection, each round having selected for cells with the capacity to reacquire stemness (for an evolutionary view on stem cell biology, see Mangell and Bonsall, 2008).

1.5.3 Critical changes in direct reprogramming as cellular switches. Another way to understand the occurrence of critical changes in the direct reprogramming process is to model the events that define a reprogramming cell as a series of switches that experience linear feedback. This approach best characterizes cellular differentiation and reprogramming as a series of parameters which can be tuned. Linear feedback is based on internal and external stimuli, and governed by improvements from a baseline rate. This was done in Mangell and Bonsall (2008) by applying a steady-state approach to this problem. As selected internal or external inputs are introduced, a linear feedback mechanism drives forward changes in reprogramming efficacy. In Park and Daley (2007), an example of a linear feedback mechanism (Fbx15) is given. In particular, Fbx15 is identified as something that defines stemness, but is not required in the transformation of a cell to pluripotency. Biological switches can be defined as genes that are modulated in a way that ultimately leads to large-scale phenotypic changes. The dynamics of a biological switch can range from all-on-none (Ferrell and Machleder, 1998) to high-dimensional (Cinquin and Demongeot, 2005), depending on the nature of the linear feedback. As the conductor of phenotypic changes, switches allow for the

movement of cellular phenotypes from one state to another (Huang, Eichler, Bar-Yam, and Ingber, 1998).

1.6 Conclusions

While relatively new, the science of direct cellular reprogramming has far-reaching applications. For specific applications to the nervous system, this area of research will herald advances in fields such as personalized medicine and tissue regeneration. In terms of basic science and theoretical synthesis, advances in the field of direct cellular reprogramming have broader relevance to our understanding of physiological adaptation/plasticity, cellular transformation (e.g. cancer), and developmental biology. In turn, systems biology and the complex system models can provide novel and needed perspectives to problems that are only beginning to be understood. Understanding cellular reprogramming as a dynamic, population-based process is essential to building a theory of direct reprogramming, and has applications to many other cellular and physiological processes.

The rest of this dissertation will feature case studies involving iN and iSM cells (Chapter 2) and iPS cells (Chapter 3). Chapter 4 will present literature review and theoretical synthesis that spans iPS, iN, iSM, and iCM cells. Chapter 5 will feature a summary of results from these chapters in addition to future research directions. This includes advances in experimental investigation and *in silico* modeling.

CHAPTER 2: DEFINING PHENOTYPIC RESPECIFICATION DIVERSITY USING MULTIPLE CELL LINES AND REPROGRAMMING REGIMENS

2.0 Abstract

To better understand the basis of variation in cellular reprogramming, experiments were performed that had two primary objectives: first, to determine the degree of difference, if any, in reprogramming efficiency among cells lines of similar type after accounting for technical variables, and second, to compare the efficiency of conversion of multiple similar cell lines to two separate reprogramming regimens – induced neurons and induced skeletal muscle. Using two reprogramming regimens it could be determined if converted cells likely are derived from a distinct subpopulation generally susceptible to reprogramming or are derived from cells with independent capacity for respecification to a given phenotype. Our results indicated that when technical components of the reprogramming regimen were accounted for, reprogramming efficiency was reproducible within a given primary fibroblast line but varied dramatically between lines. The disparity in reprogramming efficiency between lines was of sufficient magnitude to account for some discrepancies with results reported in the existing literature. We also found that the efficiency of conversion to one phenotype was not predictive of reprogramming to the alternate phenotype, suggesting that capacity for reprogramming does not arise from a specific subpopulation with a generally "weak grip" on cellular identity. Our findings suggest that parallel testing of multiple cell lines from several sources may be needed to accurately assess the efficiency of direct reprogramming procedures, and that testing a larger number of

fibroblast lines -- even lines with similar origins -- is likely the most direct means of improving reprogramming efficiency.

2.1 Introduction

Cellular reprogramming, when accomplished by either direct or indirect methods, is an inefficient process with many potential sources of variability. Several cellular characteristics, including cell type, species of origin, and age of the donor subject, are known influence reprogramming efficiency. There is an emerging awareness, however, thatholding these general cellular properties and technical aspects of the reprogramming regimen constant does little to hold results constant. This is observed in reprogramming experiments using either indirect means such as somatic cell nuclear transfer (SCNT – Campbell et al. 1996; Cibelli et al., 1998; Siripattarapravat et al., 2009; Wakayama et al., 1998), or direct methods such as plasmid transduction or infection with recombinant retroviruses expressing transcription factors (Davis, Weintraub, and Lassar, 1987; Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010; Ieda et al., 2010; Huang et al., 2011). One of the earliest studies of direct reprogramming, describing the conversion of mouse fibroblasts to skeletal muscle myotubes by transduction of the myogenic transcription factor *MyoD1*, found that reprogramming was not uniform across all cell lines. This was true even within a single cell type isolated from a single species (Davis, Weintraub, and Lassar, 1987). Comparison of five mouse fibroblast cell lines C3H10T1/2, NIH3T3, Swiss 3T3, Swiss 3T3 clone 2, and L Cells, transfected with a *MyoD* expression plasmid and selected to produce colonies of stably transduced cells yielded colonies of both the input and the conversion phenotype. The conversion efficiency varied dramatically from a maximum of 53% myoblastic colonies

in C3H10T1/2 cells to a minimum of 3% myoblastic colonies in L cells. Similarly, a report from Lattanzi *et al*. (1998) comparing the myogenic conversion of fibroblasts from different tissue sources infected with a high-titer (MOI 2,000) *MyoD* adenovirus vector found that murine dermis, muscle-, and bone marrow-derived fibroblasts converted at efficiencies of 59%, 43%, and 7% respectively, and human fibroblasts derived from the same tissues at respective efficiencies of 54%, 36%, and 6%. Together, these reports indicate that reprogramming variation may be observed regardless of whether vector delivery is relatively inefficient (plasmid transduction) or highly efficient (adenoviral infection). More recently, variation in the input cell population was postulated to account for reprogramming disparity in the conversion of fibroblasts to functional cardiomyocytes reported by several groups (Ieda et al., 2010; Yoshida and Yamanaka, 2012; Song et al., 2012; Chen et al., 2012).

Although variation in reprogramming efficiency may be frequently observed and reported, it remains unknown whether the observed variation arises from technical differences or from undefined differences intrinsic to the target cell lines used. If variation is still observed when technical elements are tightly controlled and factored into the calculation of reprogramming efficiency, it suggests that line-intrinsic characteristics play an important role in line-to-line variation.

Line-intrinsic variation in the number of cells amenable to respecification could arise through two hypothetical mechanisms. In the first, reprogrammed cells would have their origin in a subset of cells with general susceptibility to identity change. Like stem cells, these cells would be receptive to adoption of alternate fates, but might lack the active determinants associated with canonical stem cells. Alternatively, reprogrammed

cells could arise through chance reprogramming of cells to a specific alternate fate, but with no overall enhanced capacity for respecification.

Within cell lines of the same general type, are line-dependent differences in reprogramming capacity observed? If this is the case, then it is unclear whether reprogrammed cells arise from a subpopulation with a generally "weak grip" on identity, or from cells with independent capacity for respecification to a given phenotype If significant line-to-line differences are observed and cells with general lability of identity are the source, we hypothesized that we should observe a correlation in reprogramming efficiency using two separate conversion regimens applied to multiple target cell lines. Conversely, if cells with independent capacity for respecification are the source of converted cells, we would observe no significant correlation between conversion regimens. To test this hypothesis we used a rigorous measure of reprogramming efficiency that was robust across replicates within a specific reprogramming regimen. A parallel analysis was conducted consisting of 19 primary fibroblast cell lines converted to two alternate and disparate identities – induced neural cells (iN cells) and induced skeletal muscle cells (iSM cells). The results presented below provide insight into the nature of biological diversity as it is relevant to phenotypicconversion using direct cellular reprogramming methods.

2.2 Materials and Methods

2.2.1 Fibroblast lines. Human primary fibroblasts were obtained from skin-punch biopsies or gingival explants under MSU-approved IRB protocols (ADF, E2F, EAF and HSK) or were obtained from commercial sources including the ATCC (FET and HDNF) and The Coriell Institute (RET, SAF, and AUT). Additional information on these lines is

provided in Appendix B, Table 1. Mouse fibroblast lines were harvested from a single 5-month-old *nu/nu* mouse sacrificed by CO_2 overdose, the carcass disinfected with ethanol, and multiple tissues removed for dissection. 125mm^3 fragments of each target organ were removed, minced, and individual pieces placed in one well of a 6-well plate for outgrowth. Primary outgrowths were plated in fibroblast medium (DMEM, 10% FBS, 1X antibiotic/antimycotic (Invitrogen, Carlsbad, CA) and left undisturbed except for medium changes for two weeks and then passaged 1:1 to a new well. After an additional week of growth, the lines were passaged 1:2. This process was repeated to passage 5 when the wells were compared to identify those with typical fibroblast morphology, an absence of obviously non-fibroblastic cells, and similar growth characteristics. Thirteen mouse fibroblast lines were selected for two additional rounds of passage and expansion and then frozen as multiple aliquots for use in experiments.

2.2.2 Fibroblast characterization. Fibroblast RNAs from the mouse lines at passage 6-7 and the human lines at passages 8-10 were purified using Trizol (Invitrogen, Carlsbad, CA) or the RNeasy kit (Qiagen, Valencia, CA) and 2ug of purified RNA converted to cDNA using Superscript II (Invitrogen, Carlsbad, CA), following manufacturers guidelines. qPCR was performed on an ABI Prism 7000 analyzer using 1µl of cDNA and normalizing against nuclear lamin A or ARHGAP mRNAs as internal controls. Other genes used as internal controls (RPL27A, EED, and GR) gave similar results. Primers for qPCR analysis are shown in Appendix B, Table B.1.

Processing of human and mouse fibroblasts for immunocytochemical analysis to examine marker expression at the level of individual cells was performed as described below for iNC and iSMC analysis using primary antibodies as follows: anti-vimentin

1:1250 (Millipore AB5733), anti-fibronectin 1:750 (BD, 610077), anti-nestin 1:250 (Santa Cruz Biotechnologies, H-85), anti-*Sox2* 1:250 (Santa Cruz Biotechnologies, Y-17). Multiple images of each immunostained line were captured and approximately 1×10^{3} cells were examined at high magnification for the presence or absence of markers consistent with stem cell (Sox-2/nestin) or fibroblast identity (vimentin/fibronectin). NPCs used for comparison were fixed in parallel with fibroblasts and were generated as described below for iNC studies.

Relative infectivity for "Factor Expression Early" (FEE) was calculated by infecting 1×10^{5} actively growing cells with concentrated NITSC-NLS-YFP retrovirus at an MOI of approximately 0.5 and then counting the YFP-positive cells as a fraction of all cells at day 4 post-infection over three replicates. Since the precise cellular age of each line was unknown, to ensure that none of the lines analyzed was approaching cellular senescence that could influence reprogramming, each line was continuously passaged and counted for four rounds at the end of experiments to determine that all lines were still proliferative. One of the 13 mouse lines, LU3, failed by the third trial passage and was removed from the final analysis. Two human lines, AUT and HSK, produced iN cells and iSM cells and were partially characterized, but were not included in the final analysis because the matching frozen stocks were lost before the completion of some experiments due to freezer failure. Mouse lines TA6, KI2, and LU6 were not included in the experiments for the same reason.

2.2.3 iNC induction: plasmids and virus production. cDNAs encoding human *ASCL1*, *POU3F2*, and *ZIC1* were obtained from Open Biosystems (Huntsville, AL). *Myt1L* and *NeuroD1* ORFs were obtained from cDNA produced from human brain reference RNA

(Applied Biosystems, Carlsbad, CA). NITSC was produced by introducing the BstEII-ClaI fragment containing Neo-IRES-TTA-TetO from NIT (Genbank Acc# AF311318) into BstEII-ClaI cut pMSCVneo (Clontech, Mountain View, CA) and a polylinker for transgene expression, SfiI-MluI-PmeI-ClaI. Primers used for cloning of factors into NITSC are shown in Appendix B, Table B.2. The amplified YFP ORF with an MluI site inserted immediately before the stop codon was digested SfiI-PmeI and introduced into NITSC to produce the control vector NITSC-YFP. Remaining factor ORFs were PCR amplified with compatible MluI or Asc1 sites at the 5' end and ClaI at the 3' end for cloning into NITSC-YFP to produce the fusion protein constructs. NITSC recombinant MMLV particles were produced by three-way calcium phosphate transfection of HEK cells with gag-pol and VSV encoding plasmids to produce replication defective virus particles. Two days after transfection, viral supernatants were harvested, filtered, and introduced into fibroblast cultures using the carrier polybrene (8ug/ml) to improve infection efficiency essentially as described for lentiviral vectors in Suhr et al. (2009). Viral supernantants were frozen as aliquots and tested on MEFs to provide a rough titer and to establish the competence of viral preparations to produce iN cells prior to use with target cells. As indicated in the text, the YZIC/YASCL/YPOU3F (ZAP) combination appeared most potent on both mouse and human fibroblasts in preliminary experiments and was used for conversion unless otherwise noted. Equal volumes of each viral supernatant were used (e.g. for ZAP, typically 5ml of each viral supernatant for a total of 15ml infectious medium/10-cm plate).

2.2.4 iNC induction: infection and iNC conversion. For preliminary studies to determine the optimal timing and conditions for iNC conversion, approximately 1×10^6 growing

mouse or human fibroblasts in fibroblast medium at equal confluency were infected with viral medium and kept overnight to allow infection. The next day, virus-infected cultures were passaged by trypsin treatment to 6-well, 12-well, or 35mm tissue culture plates and allowed to remain 12-24 additional hours in fibroblast growth medium to attach and expand. After this time, the fibroblast medium was aspirated and replaced with iNC medium (DMEMF12 with N2 supplement and penicillin/streptomycin at 50u/ml) (Invitrogen, Carlsbad, CA). iNC medium was changed at 4-5 day intervals for the experimental duration, and the cells were kept in a 5% CO_2 environment at 37^oC. Cell culture plates coated with polyornithine/laminin (PORN/lam) or with no coating were used for pilot experiments interchangeably with little noticeable impact on iNC formation. To determine the optimal time for counting of iN cells, factor-infected MEFs and adult mouse and human fibroblast cultures were fixed and immunostained with the neural *TUJ1* antibody (see below) at 5-6 day intervals post-infection.

iNC conversion for determination of reprogramming efficiency across mouse and human cells lines was performed essentially as described above except that 1×10^5 target cells in one well of a 6-well plate were infected with 3 mls of the iNC (ZAP) viral cocktail and 24-hours later passaged to 3 wells of a 12-well plate. The passaged cells were allowed to rest an additional day and were then shifted to iNC medium with medium changes every 3-4 days until fixation and immunoprocessing on day 12 for mouse iN cells and day 24 for human iN cells. Experiments for quantification of reprogramming efficiency were performed in three separate replicates.

2.2.5 iNC induction: immunohistochemistry and imaging. Cells were fixed using 4% paraformaldehyde for 10 minutes followed by 3X PBS washes for 10 minutes each.

PBST (PBS with 0.3% Triton X-100) with 3% donkey serum (DS) was used for 30-60 minutes at room temperature to block, and was then replaced with PBST+1% DS with added primary antibody overnight at 4°C. Primary antibodies were used at the following dilutions: *TUJ1* - 1:3000 (Santa Cruz, Cat# sc-58888), MAP2ab -1:300 (Sigma, Cat#M1406), Synapsin 1 – 1:400 (Millipore, Cat#AB1543P), pan-neurofilament – 1:1000 (Covance, SMI311), Doublecortin – 1:400 (Santa Cruz, Cat#sc-8066), GAD – 1:250 (Santa Cruz, Cat#sc-7513), PSD95 – 1:250 (NeuromAb), GABAR3 – 1:250 (NeuromAb). After overnight incubation with primary antibody, wells were washed with PBST+1%DS 3X10 minutes and incubated with PBST+1% DS with the appropriate secondary antibody (Jackson ImmunoResearch, West Grove, PA) for 30-60 minutes. Wells were then washed with PBS 3X10 minutes to remove excess secondary, stained briefly with PBS+1ug/ml bis-benzamide to label nuclear DNA, and rinsed again. All plates and wells were stored at 4°C in the dark until imaging. Imaging was performed on a Nikon Eclipse TE2000 inverted stage fluorescence microscope.

2.2.6 Electrophysiology. For electrophysiological recordings, infected NPC-neurons (see below) or iN cells were cultured on PORN/lam coated 35mm plates at low density (2.5X10^5 cells/35mm plate) as described above. All recordings were made using the whole-cell configuration of the patch-clamp technique (Hammill et al., 1981). Patch glass pipet electrodes were double-pulled and heat-polished. The electrode was brought into contact with visually identified iNC targets to produce a high-resistance seal between electrode tip and cell membrane, and the whole-cell configuration was achieved by applying suction to the back of the electrode. For voltage clamp experiments, electrode capacitance was compensated prior to achieving the whole-cell

configuration, and membrane capacitance and series resistance were compensated after achieving this configuration. Membrane current and potential signals were amplified (List Electronic EPC-7), digitized (Digidata 14140A, Molecular Devices), and stored on a computer. Voltage steps and current injection pulses were generated and potential and current signals were analyzed using software written by Dr. John Dempster (Dept. Physiology, University of Strathclyde). In all voltage clamp recordings, the holding potential (Vh) was -80mV. Presence and properties of voltage-gated current was examined during positive voltage steps (30ms to 250ms depending on the experiment) to test potentials (Vtest) between -75mV and +50mV. To examine voltage-dependent, steady-state inactivation of voltage gated Na^+ channels, a double voltage step was used: a step to a conditioning potential (Vcon, -130mV - +40mV, 50ms) was applied immediately prior the step to the test potential (0mV). To examine whether cells had the capacity to generate action potentials, membrane potential was measured under current clamp.

The extracellular solution contained NaCl 135mM, KCl 5mM, glucose 10mM, $MgCl_2.6H_2O$ 1mM, $CacCl_2.2H_2O$ 2mM, and HEPES 20mM (pH 7.3). For recordings of isolated voltage-gated Na^+ current, the electrode solution contained CsCl 20mM, cesium methanesulfonate 130mM, $MgCl_2.6H_2O$ 2mM, glucose 10mM, EGTA 10mM, and HEPES 10mM (pH 7.3). For recordings of mixed voltage-gated Na^+ and K^+ current, and for recordings of membrane potential and action potentials, the electrode solution contained KCl 20mM, potassium methansulfonate 130mM, $MgCl_2.6H_2O$ 2mM, glucose 10mM, EGTA 0.01mM, and HEPES 10mM (pH 7.3).

2.2.7 Human NPC culture and neuron derivation. For the control human neurons in Alicea et al. 2013 (see Supplemental Figure 5), H9 human ES cells were differentiated to NPCs as described (Chang et al., 2009). H9-NPCs were propagated to passage 5 in iNC medium supplemented with 20ng/ml FGF-2. For differentiation, NPCs were plated on PORN/lam plates and FGF-2 was progressively withdrawn to a final concentration of 2ng/ml by day 20-24, when the cells were processed for immunostaining to confirm neuronal identity and subjected to electrophysiological analysis. Undifferentiated NPCs at 30-50% confluency were fixed and used as controls in the experiments shown in Alicea et al. 2013 (see Supplemental Figure 3).

2.2.8 iSMC conversion: plasmids and virus production. All factors for iSMC conversion were cloned from cDNAs produced from a piece the donor mouse skeletal muscle not used for fibroblast derivation in culture. The primers for cloning of the full-length *MyoD*, *MYF5*, *MYF6*, and myogenin ORFs by the same YFP-fusion strategy as the iNC factors are shown in Appendix B, Table B.3. Equal volumes (0.75 ml) of viral supernatant for each of the four myogenic factors was used to infect target fibroblasts, iSM cells were induced using iSMC medium ((DMEM, 0.1% FBS, 1X antibiotic/antimycotic (Invitrogen, Carlsbad, CA)), and all iSM cells were generated on uncoated plates, but otherwise virus production, infection, and conversion was performed as with iN cells.

2.2.9 iSMC induction: immunohistochemistry and imaging. All procedures were performed as with iN cells except for use of skeletal muscle-specific antibodies sarcomeric anti-α actinin used at 1:1000 (Sigma, A7811) and anti-sarcomeric myosin used at 1:500 (DSHB at University of Iowa).

2.2.10 Quantification of phenotypic conversion. Quantification of iNC/iSMC conversion

was done using Hoechst33342 staining for nuclei/DNA (blue), YFP fluorescence of the tagged proteins to indicate relative factor expression (green), and β-IIITubulin/*TUJ1* (iN cells) or sarcomeric α-actinin (iSM cells) immunostaining to indicate phenotypic conversion (red). Relative reprogramming efficiency was calculated either by dividing the red fluorescence value by the blue fluorescence or dividing the red/blue value by the green fluorescence value to include a factor-expression component in the calculation. For iSM cells, total sarcomeric α-actinin fluorescence was taken as the red value, whereas iNC conversion was measured as the number of red fluorescent cells that also had fibers of at least three soma lengths. For each well, five fields -- one in the center of each well and one at each compass point approx 1-cm from the well-edge -- were imaged at 100X-magnification for each separate channel and stored as a merged RGB image. For measurement of fluorescence, the RGB image was split into individual black and white channels and quantified using NIH ImageJ. For the graphs, the highest relative conversion value for iN cells or iSM cells for each group was set at 100 and the remaining values calculated as a fraction of that maximum. Calculation of correlation, significance, and ANOVA were performed using SigmaPlot 12 software.

2.3 Results

We sought to control external variables in the reprogramming process to promote better measurement of relative reprogramming efficiency. In keeping with this concept, all procedures and analyses were performed in parallel on all cell lines within a group and in at least three separate experiments. Fibroblasts were selected as the input cell type because they have been used extensively as the raw material for most reprogramming studies, are easily established and expanded for many passages

(without the need of cellular transformation), are adherent, and can be efficiently cryogenically preserved. Fibroblast lines were obtained from two species – human and mouse. Human fibroblasts were selected for study because they are of the most direct clinical relevance and mouse fibroblasts were selected because they have been shown capable of reprogramming to several output cell types and because we were able to establish multiple isogenic fibroblast lines from a single donor animal under highly controlled conditions. Using both groups of cells, we were able to compare fibroblast lines with similar morphological properties from multiple individuals with fibroblast lines of variable morphology and tissue-of-origin, but from a single subject. We chose to work primarily with non-embryonic cells to better relate our findings to applications in human or veterinary medicine, since older subjects are more likely to be the primary target of reprogramming-based therapy for cellular replacement or transplantation.

2.3.1 Establishment of input lines. Human fibroblast lines used for analysis are shown in Appendix B, Table B.3 (top panel). Lines for analysis were selected based on several general criteria. The lines needed to display robust growth in culture, be well-established (>passage 8 (P8)), display a homogenous morphology (Alicea et al. 2013; see Supplemental Figure 1, panel A), and be at least four passages from mitotic senescence. Most of the lines were from donor subjects with no known disease, but two were included from individuals with diagnosed neurological disorders – Rett syndrome and schizophrenia. Seven human fibroblast lines were subject to complete characterization and analysis with two additional lines only subjected to partial analysis because they did not satisfy all criteria for inclusion.

To examine the impact of genetic homogeneity on reprogramming efficiency, mouse fibroblast lines were generated from a single donor animal Appendix B, Table B.3 (bottom panel). To ensure that the individual mouse lines did not functionally approximate secondary lines or subclones that could occur if all lines were derived from a single explant, we harvested eight different mouse organs for establishment of fibroblast lines (Alicea et al. 2013; see Supplemental Figure 1, panel B) and then selected only those lines that displayed no evident non-fibroblastic cells and a consistently uniform fibroblastic morphology (Alicea et al. 2013; see Supplemental Figure 1, panel C). Lines also had to display a similar capacity for growth, passage, and survival of freeze/thaw that allowed them to be maintained side-by-side with other lines. In this way, approximately 50 lines were progressively winnowed to 12 lines that were the most representative of criteria such asgrowth characteristics and morphology. Unlike the human lines, the mouse lines were generated and maintained identically from their first day in culture – passaged, fed, frozen, thawed, and in every other respect cultured side-by-side. Ultimately, these lines, derived from seven different tissue sources (brain tissue did not give rise to robust fibroblast cultures), were frozen in multiple aliquots at P8 for use in experiments.

The phenotype of human and mouse fibroblast lines was further established by quantitative RT-PCR analysis to confirm an abundance of fibroblast-associated mRNAs and the absence of significant signal for indicators of other differentiated or progenitor cell types. All but two fibroblast lines displayed an abundance of the fibroblast-associated markers collagen type 1α2, vimentin, fibronectin, fibrillin I, and fibulin V. The two exceptions were mouse lines HE4 and KI3, which repeatedly tested negative for

mouse collagen type 1α2. HE4 also displayed weak vimentin immunopositivity (see below). However, both HE4 lines were positive for some fibroblast markers and were negative for indicators of stem cell types, so both lines were kept in the study. mRNAs for other common contaminating cell types (e.g. keratinocytes - keratin 14 and endothelial cells - PECAM) or for myogenic or neurogenic progenitor-type cells (e.g. *MyoD*, *Myf5* and *FoxG1*, *Sox2*, respectively) were not detected Alicea et al. 2013 (see Supplemental Figure 2).

To further characterize our lines at the level of individual cells each line was examined for the expression of fibroblast-associated markers or markers of stem cell types by fluorescent immunocytochemistry. *Sox2* and nestin were selected as stem cell markers because these markers label multiple classes of stem cell including pluripotent stem cells, neural stem cells, and muscle precursors such as mesenchymal stem cells and satellite cells, but are not expressed at significant levels in fibroblasts (Day, Shefer, and Yablonka-Reuveni, 2010; Day et al., 2007; Ellis et al., 2004; D'Amour and Gage, 2003; Vogel et al., 2003; Avilion et al., 2003; Zimmerman et al., 1994). The results revealed that essentially all cells appeared immunopositive for the fibroblast-associated markers vimentin and fibronectin and no cells appeared positive for the stem cell markers nestin and Sox-2 that were readily observed in control cultures of human neural stem cells (Alicea et al. 2013; see Supplemental Figure 3, panel A and B). Parallel analysis of selected mouse lines (Alicea et al. 2013; see Supplemental Figure 3, panels C and D) showed a result similar to the human lines, with the exception of HE4 that displayed weak vimentin staining. Additional RT-PCR and immuncytochemical analysis of lines such as ADF and AUT examining other markers of stem and progenitor

cell types including neural crest-derived cells and pluripotent cell types did not reveal any indication of other progenitor cell types within our fibroblast lines.

2.3.2 Conversion to induced neural cells. iN cells were selected as a conversion cell type to test comparative direct reprogramming efficiency for several reasons. The first was because several independent laboratories had confirmed the production of iN cells using similar methodologies (Vierbuchen et al., 2010; Pfistetrer, et al., 2011; Pang et al., 2011; Caiazzo et al., 2011). The second was that iN cells should display a distinct set of markers of neuronal phenotype and changes in cellular morphology that would allow them to be readily distinguished from input fibroblasts. Third, evidence-to-date suggested that, because they were produced directly from input cells without generation of an intermediate cell type or the requirement for cell division and formation of a progenitor colony, each iNC represented an individual reprogramming event.

For iNC conversion, human and mouse fibroblast cultures were infected with combinations of an MMLV-based retroviral vector encoding the neurogenic factors *ASCL1*, *POU3F2*, *ZIC1*, *MYT1L* or *NeuroD1* (Vierbuchen et al., 2010; Caiazzo et al., 2011) of human origin fused to YFP as shown schematically in Alicea et al. 2013 (Supplemental Figure 4). As predicted, all of the YFP-fusion proteins localized primarily to the nucleus, some presenting a mottled appearance and others appearing more diffuse (Figure 2.1, panel A). *ASCL1*-YFP generally presented as weak diffuse nuclear fluorescence with one or two bright punctuate bodies. Pilot studies indicated that *ASCL1* and *POU3F2* combined with either *ZIC1* or *MYT1L*/ND produced many *TUJ1*-positive iN cells in both human and mouse cultures using either fetal or adult fibroblasts, whereas cells infected with each neurogenic factor alone produced very few iN cells (Figure 2.1,

panel B). Mock-infected cells or cells infected with a YFP vector only and cultured under identical conditions produced no strongly immunoreactive *TUJ1*-positive cells with neuron-like morphology (Alicea et al. 2013; see Supplemental Figure 3, panel A). As described previously, the primary difference between the generation of mouse and human iN cells was the maturation rate (Vierbuchen et al., 2010; Pfistetrer, et al., 2011; Pang et al., 2011; Caiazzo et al., 2011). While mouse iN cells with relatively mature morphology were sometimes observed as early as day 4-5 and reached maximal differentiation by day 10-12 post-infection, human iNC maturation appeared more progressive, with cells at days 8-12 displaying a rounded cell soma and very short processes of 7-10μm, longer processes at day 18-20, and long processes with more complex branching and spine-like projections by day 24-30 (Alicea et al. 2013; see Supplemental Figure 5, frames B and C). Mature iN cells that displayed elongated processes (>3 soma lengths), were positive for multiple markers of mature neurons in addition to *TUJ1*, including MAP2a/b, synapsin 1, doublecortin, neurofilament 300kD, and others as shown in Figure 2.1, panels C and D. iN cells also displayed electrophysiological properties such as TTX-sensitive Na+ currents and action potentials that were essentially indistinguishable from human neurons generated under identical conditions from human ES cell-derived human neural progenitor cells (NPCs - Alicea et al. 2013; see Supplemental Figure 5, panel D).

To determine if, and to what degree, variation was observed in the reprogramming of primary fibroblast lines to neural cells, all lines were thawed, passaged, and plated at the same density for side-by-side infection and culture. On day 12 for mouse cells and day 24 for human cells, transduced and control cultures were

harvested in parallel for immunochemical staining and quantification of converted cells. iN cells with morphological characteristics and processes, essentially indistinguishable from neurons generated side-by-side from human ES-derived NPCs, were observed in all mouse and human fibroblast lines tested (Alicea et al. 2013; see Supplemental Figure 6, panels A and B), although in some lines they were rare or sometimes displayed shorter, less mature processes (e.g. SAF, LU6, TE5). Conversion efficiencies calculated as a simple percentage of output cells versus unconverted cells were in line with published reports of iNC conversion using methods similar to those in this report. Maximum conversion efficiency to iN cells was 0.72±0.08% for adult mouse fibroblasts (TA4), and 1.07±0.18% for human cells (RET)).

2.3.3 Induced neural cells: reprogramming efficiency. As shown in Figure 2.2, relative reprogramming efficiency could be calculated by dividing the red fluorescence value indicative of reprogrammed cells in each line by the Hoechst 33342 fluorescence value indicative of total cells present at the time of harvest (See Appendix A.1 for equation). Calculated in this way, our relative reprogramming measure reveals dramatic differences between cell lines of both mouse and human origin. The limitation of this method of determining conversion efficiency is that an unknown degree of the disparity may be caused by differences in factor delivery and production. Viral transduction and factor expression was included in the calculation of reprogramming efficiency in two ways. The first, referred to as "factor expression early" (FEE), was essentially a determination of relative infection efficiency. Each line used in the analysis was infected with a nuclear-localized YFP virus at an MOI of approximately 0.5 and then counted to determine the number of cells with yellow fluorescent nuclei as a fraction of all cells on

day 4 post-infection. The relative capacity of each line to take up and express virus at levels sufficient to be scored as positive relatively early after infection could then used as one means of including factor transduction efficiency into the calculation of reprogramming efficiency.

Factor expression late (FEL) was an alternate method of measuring factor expression that was determined by the level of YFP fluorescence produced by the transgenes at the time of harvest. The fluorescent signal at this stage was less uniform than in cells used in the calculation of FEE. By the time of harvest, the YFP signal in iNC cultures was often punctuate or faint and could not be quantified accurately as a percentage of positive cells, so FEL was measured as the overall intensity of YFP signal at the time of harvest instead of as a percentage of positive cells. Both FEE and FEL have their own individual merits, but both were included in our analysis to address the possibility that factor expression measured at different stages of the reprogramming process might influence the final determination of reprogramming efficiency.

2.3.4 Line-to-line variation and between-regimen correlations. As shown in Figure 2.2, after factoring factor expression into the calculation of reprogramming efficiency, clear line-to-line variation in the reprogramming capacity of cells of both mouse and human origin was still observed, irrespective of how transgene expression was factored into the calculation of efficiency. There was a strong and significant correlation between the iNC reprogramming efficiency calculated using both standards in the mouse cell lines (r=0.975, P < 0.0001), and a similar trend was observed in the human lines (r=0.664, P=0.104).
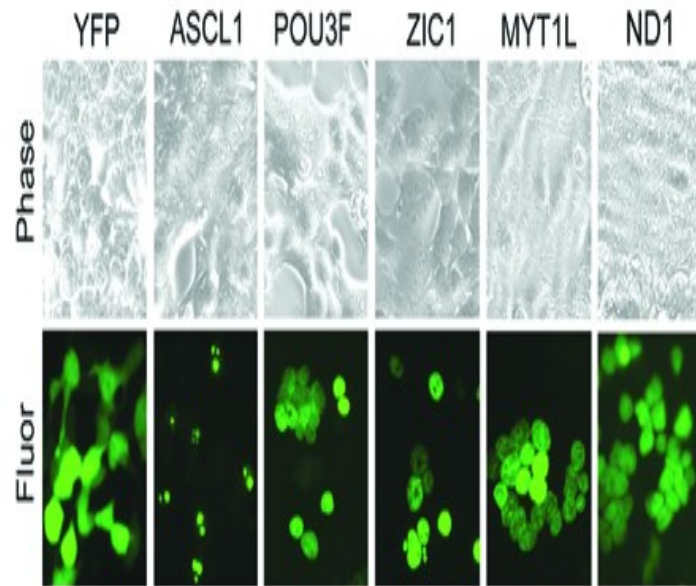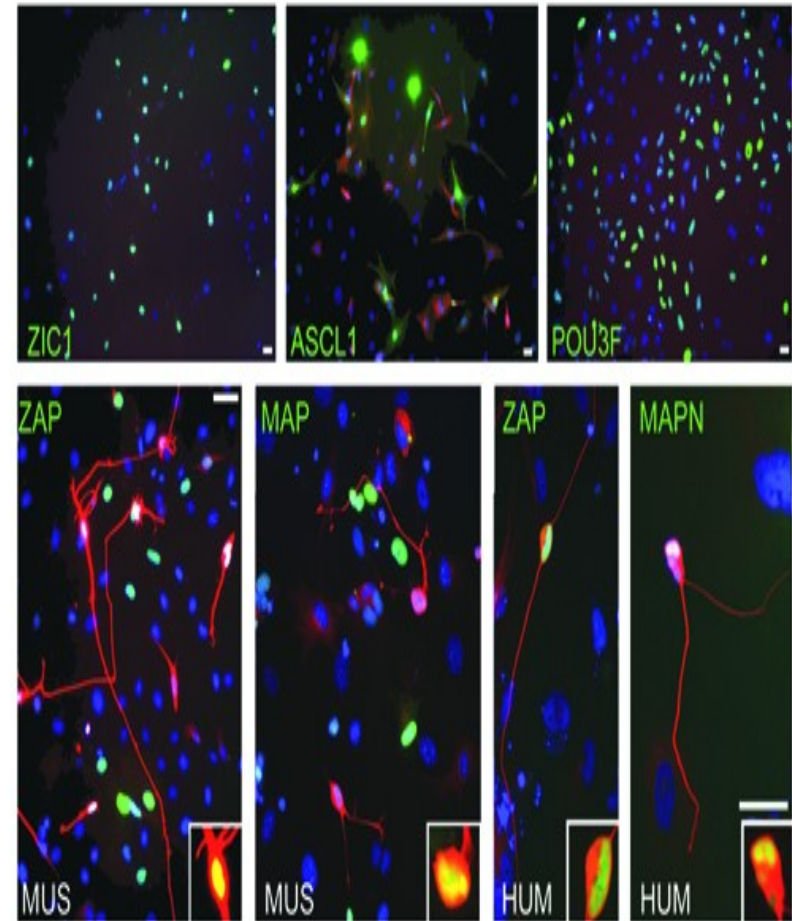
Figure 2.1. iNC conversion of mouse and human fibroblast lines. A) HEK cells transfected with YFP fusion protein for iNC conversion, as labeled. Phase- contrast images are in the upper panels and corresponding fluorescent images in the lower panels. B) Mouse embryonic fibroblasts infected with the individual factors (as labeled) and stained for β-III-tubulin/*TUJ1* in red (upper panels). Mouse (MEFs) or human fibroblasts (FET) infected with combinations of

Figure 2.1 (con't). *Zic1*/*Ascl1*/*Pou3f2* (ZAP) or *Myt1L*/*Ascl1*/*Pou3f2* (MAP) + *NeuroD1* (MAPN) (lower panels). Green fluorescence indicates expression of reprogramming factor(s). Blue color is bis-benzimide nuclear staining of DNA. Insets have the blue channel removed and the green channel intensified to show YFP-factor expression in the nucleus of all iN cells. C) Mouse iN cells produced from MEFs by day 10-15 post-infection immunopositive for multiple neural markers (red), including MAP2, pan-neurofilament (NF), doublecortin (DCX), or synapsin I (SYN). iN cells produced from adult mouse fibroblast lines (as labeled) with typical iNC morphology immunostained for β-III-tubulin/*TUJ1*(red). D) Human iN cells with typical morphology at day 24-30 post-infection and immunostained for multiple neuronal markers as in (C) in addition to PSD95, GABA receptor β 3 (GABAR-B3), and GAD1 (as labeled). GABAR-B3 and GAD1 iN cells were labeled using immunoperoxidase secondary antibody coupled with DAB staining. Scale bars are 10 μm unless otherwise labeled. For interpretation of the references to color in this and all other figures, the reader is referred to the electronic version of this dissertation.

Figure 2.1 (con't).
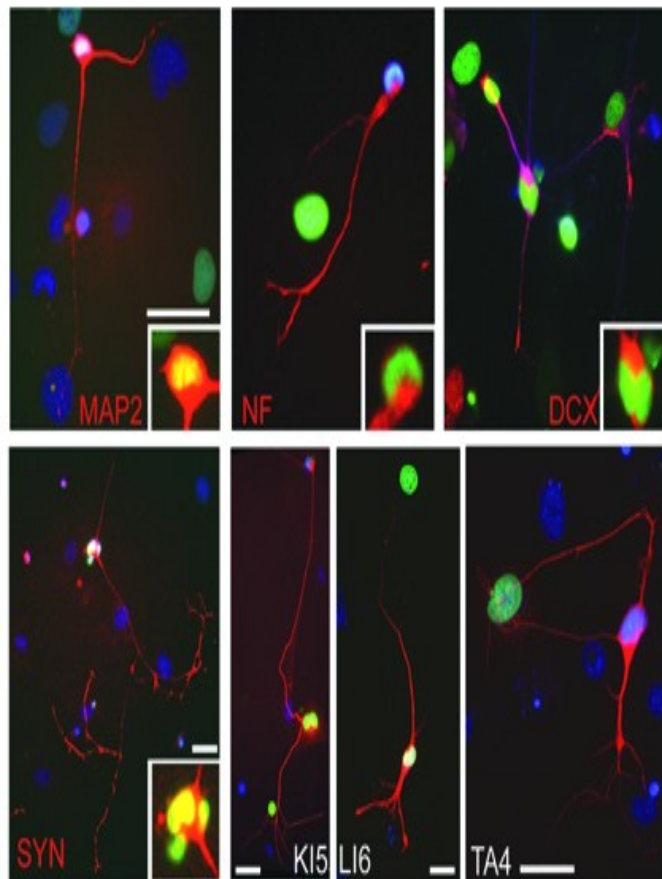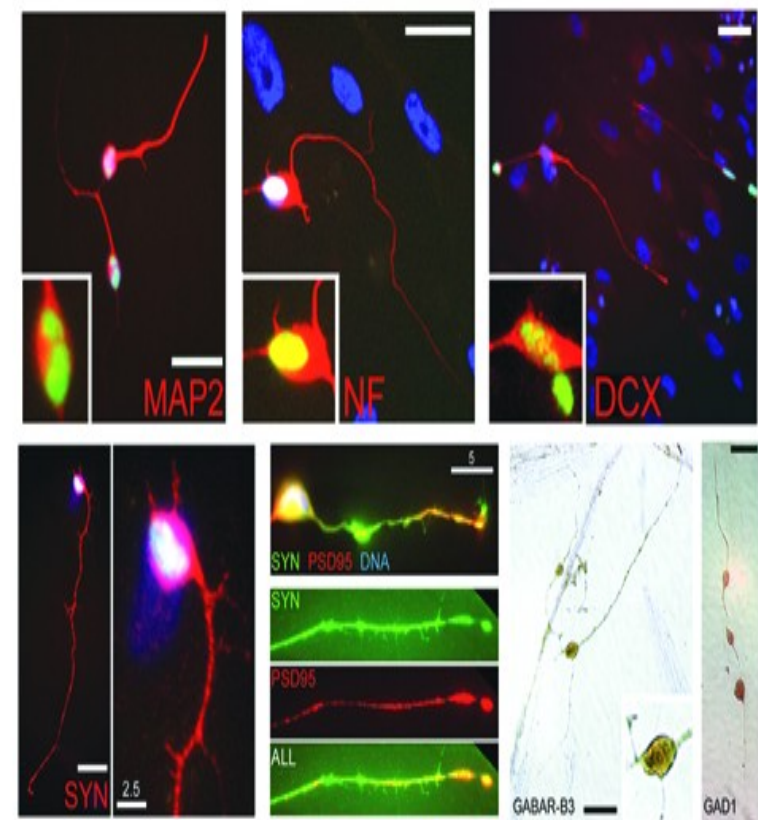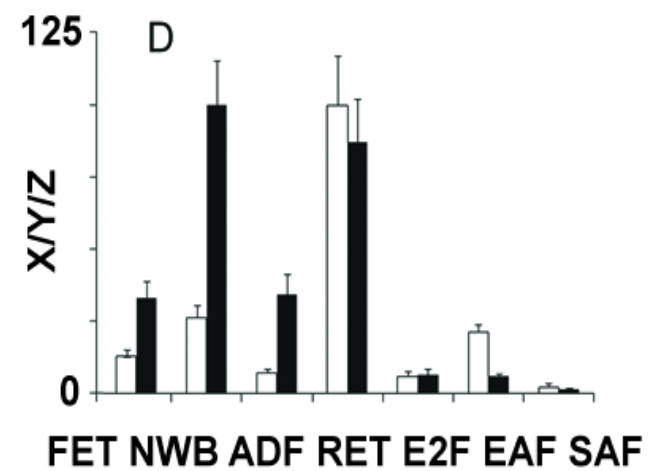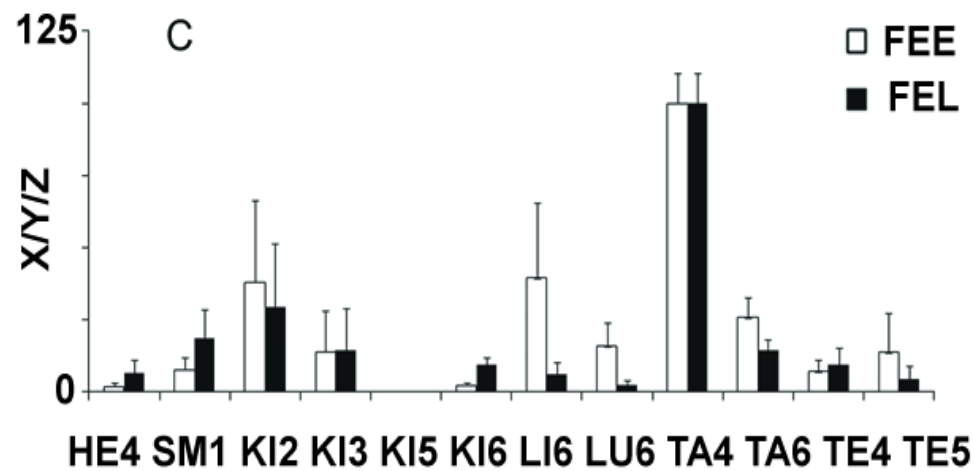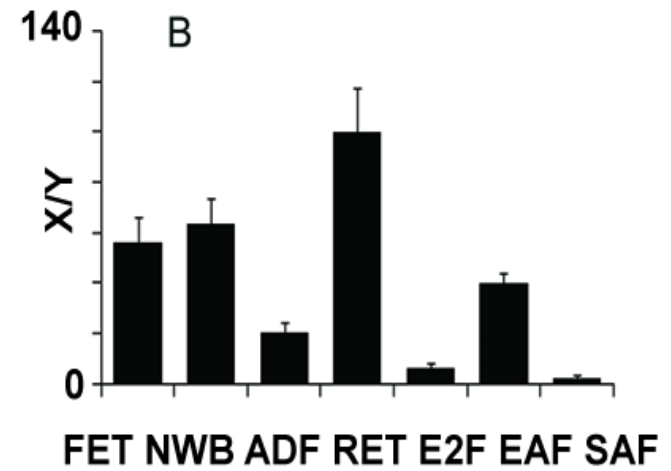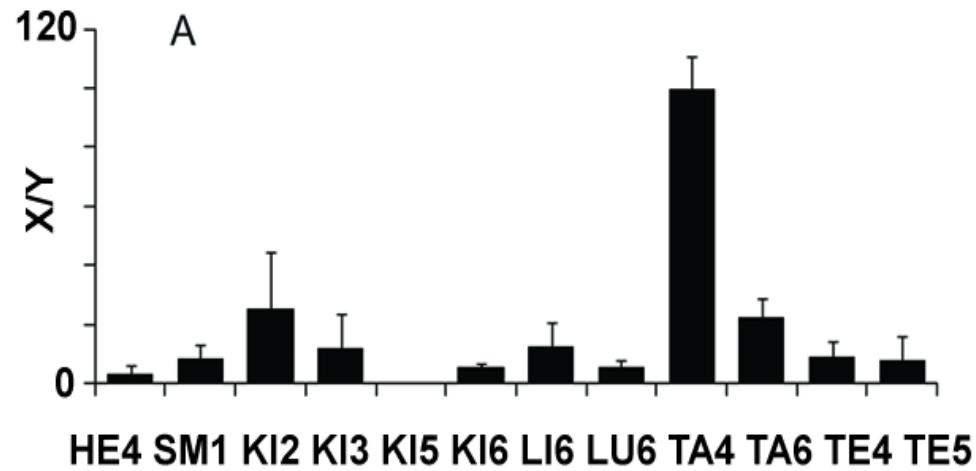
Figure 2.2. Data analysis for iNC conversion of mouse and human fibroblast lines. A) Relative conversion of mouse fibroblast lines to iN cells calculated as a function of cell number at the time of harvest. The highest efficiency of

Figure 2.2 (con't). conversion within each group was set to a value of 100. Bars indicate standard error of the mean (SEM). B) As in E, for the human fibroblast lines. C) As in E, but showing relative conversion of mouse fibroblast lines factoring in factor expression early (FEE) or late (FEL). D) As in G, for the human fibroblast lines. Y-axis label convention is as follows: X/Y = number of conversions/number of cells, X/Y/Z = number of conversions/number of cells/number of cells positive for factor.

More importantly, analysis of variance within human iNC cultures revealed that the differences in median reprogramming among lines using either FEE or FEL were significantly greater than would be expected by chance (FEE P=0.005, FEL P=0.006). Differences among iNC reprogramming efficiencies in the mouse lines were only significant to P= 0.150 (FEE) and P=0.108 (FEL), but appeared to be trending in the same direction. It was noted that although including factor expression into the calculation of reprogramming efficiency did not dramatically change the values for the best and worst converters, lines in the middle of the pack were more dramatically affected. Likewise, lines with intermediate conversion efficiencies also displayed more variation depending on the use of either FEE or FEL.

2.3.5 Conversion to induced skeletal muscle cells. A second reprogramming regimen was used for two reasons: to determine whether or not line-to-line variation observed with iNC reprogramming would also be observed using an alternate regimen, and to determine if the lines most amenable to iNC respecification were also those most reprogrammable to a second fate.

Reprogramming to iSM cells was performed using four myogenic factors known to promote skeletal muscle identity and expressed as YFP fusion proteins as with neurogenic factors (Figure 2.3, panel A). One factor was *MyoD1*, shown in multiple studies to induce the conversion of fibroblastic cells to myotube-like cells (Davis, Weintraub, and Lassar, 1987; Lattanzi et al., 1998). A second factor was the myogenic

transcription factor myogenin (Edmondson and Olson, 1989) that in our pilot studies also displayed some capacity to induce muscle marker expression and a shift to myotube-like morphology, though to a lesser extent than *MyoD* (Figure 2.3, panel B). We also included two additional factors -- *MYF6* (Miner and Wold, 1990) and *MYF5* (Braun et al., 1989) – that displayed minimal capacity to induce conversion on their own, but are known to support skeletal muscle maturation. Recombinant virus stocks for each of these four myogenic factors were combined in equal quantity and used as a multi-factor mix for iSMC conversion.

A shift of virus-transduced cultures from fibroblast growth medium to iSMC medium (DMEM+0.1% FBS) 3-4 days post-infection was found to best support conversion and maintenance of the iSMC phenotype. Uninfected fibroblasts (or cells infected with YFP only) showed no indication of conversion of cells to a myotube-like morphology under iSMC growth conditions. However, both mouse and human cells transduced with the iSMC virus cocktail showed abundant evidence of cell fusion and myotube formation by 6-8 days post-infection, and appeared to complete morphological maturation by 10-12 days of culture for mouse cells and approximately 18-20 days for human cultures (Figure 2.3, panels C and D). Immunocytochemical analysis of control and myogenic-factor infected fibroblasts confirmed that no cells immunopositive for markers of skeletal muscle myotubes were observed in control cultures (Figure 2.3, panel E), whereas factor-transduced cultures in some test lines displayed numerous elongated, tube-like cells positive for the muscle markers sarcomeric myosin and α-actinin (Figure 2.2, panel E). These data indicated that the viral reagents and culture regimen we used for iSMC reprogramming produced cells with multiple strong indicators

of the conversion cell phenotype that were essentially identical to induced skeletal muscle myotubes or myotubes produced from cultured myoblasts previously reported (Davis, Weintraub, and Lassar, 1987; Lattanzi et al., 1998; Lassar, Paterson, and Weintraub, 1986; Macpherson, Suhr, and Goldman, 2004).

2.3.6 Comparison of iSM cells to iN cells. The experimental design for the analysis of iSM cells was essentially identical to the process used above for iN cells. Maximum conversion efficiency calculated as a percentage of iSM cells to unconverted cells was similar to other published conversion rates of mouse and human fibroblasts to iSM cells (Davis, Weintraub, and Lassar, 1987; Vierbuchen et al., 2010; Lattanzi et al., 1998; Pang et al., 2011). Maximum iSMC conversion for adult mouse fibroblasts was 20.3±3% (KI6) and 60.4±6% for human fibroblasts (FET).

iSM cells, identified as elongated cells with strong α-actinin immunopositive fluorescence, were observed in ten of twelve mouse fibroblast lines tested. Lines such as KI6 and TA6 produced an abundance of iSM cells, while lines such as HE4, SM1, and LU6 produced only rare cells. Lines LI6 and TE4 produced no cells with the morphology and marker expression consistent with the output phenotype (Alicea et al. 2013; see Supplemental Figure 7, panel A). iSM cells were observed in all transduced human cultures (including the AUT and HSK lines not included in the full analysis), although immunopositive iSM cells were rare in the EAF and E2F lines derived from elderly donor subjects (Alicea et al. 2013; see Supplemental Figure 7, panel B).
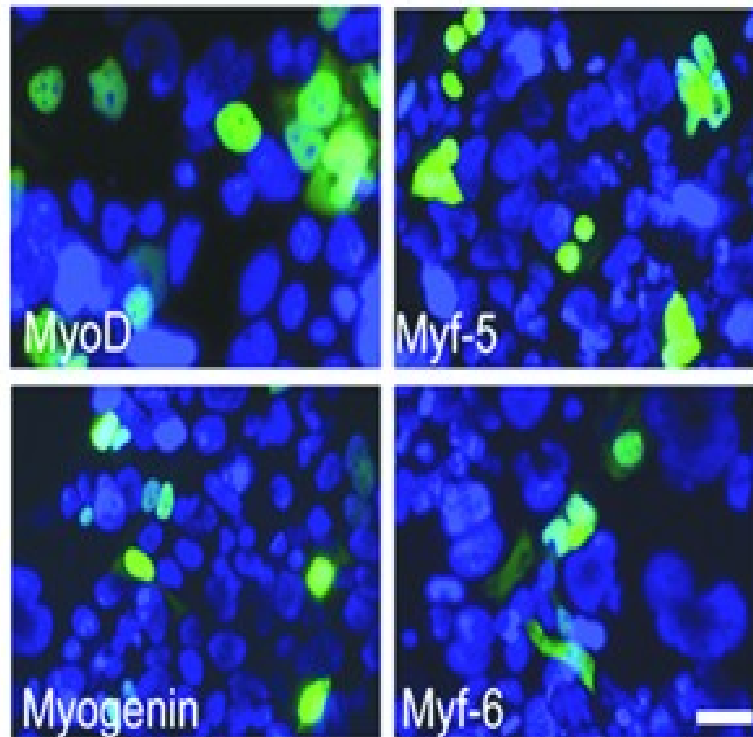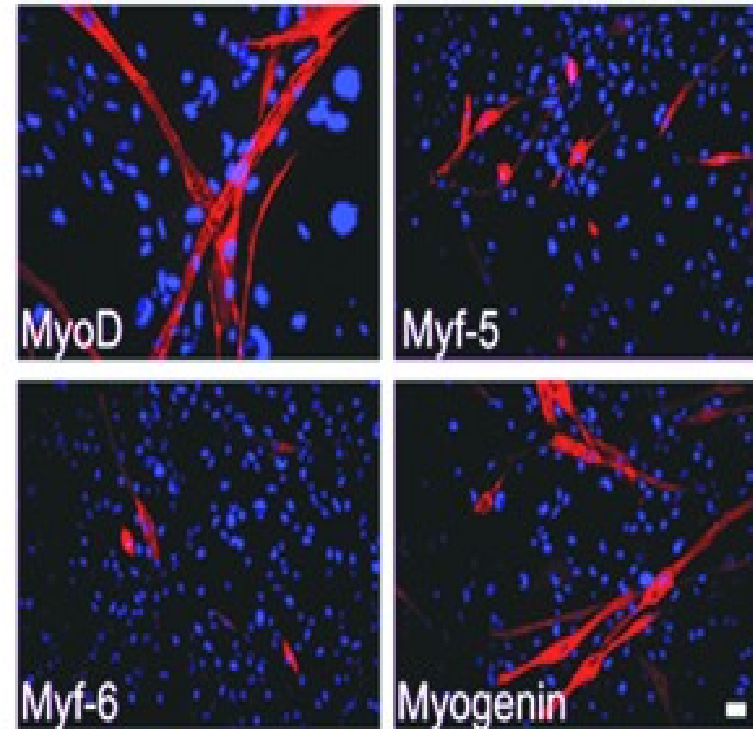
Figure 2.3. iSMC factors and induction of iSMC phenotype in mouse and human fibroblast lines. A) HEK cells transfected with YFP fusion protein vectors for iSMC conversion, as labeled. Green fluorescence is the nuclear-localized myogenic factors and blue is bis-benzimide staining. B) KI6 mouse fibroblasts infected with the four separate myogenic factor viruses (as labeled) and immunostained for sarcomeric α-actinin. C) Example of the morphological change observed in human NWB or mouse KI6 fibroblasts after infection and expression of iSMC factors (as labeled). D) Magnified image (400X) of phase contrast (upper) and fluorescent (lower) image of bis-benzimide stained KI6 iSMC myotubes. Arrows

Figure 2.3 (con't). indicate multiple nuclei within the fiber. E) Control or iSMC-factor infected mouse and human cells stained for skeletal muscle antigens sarcomeric α-actinin (α-ACT) or sarcomeric myosin (SMYO). Mouse cells were processed at Day 12 and human cells at day 24 post- infection. Scale bars are 10 μm.
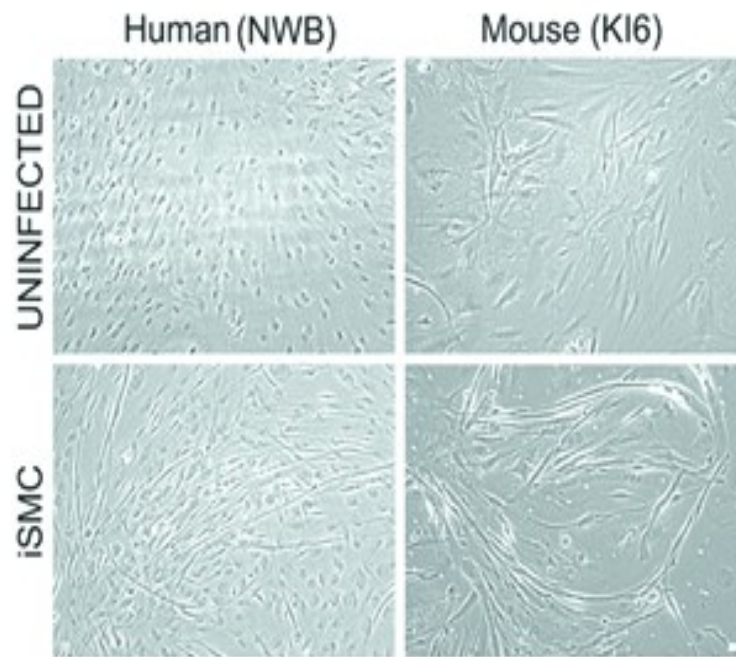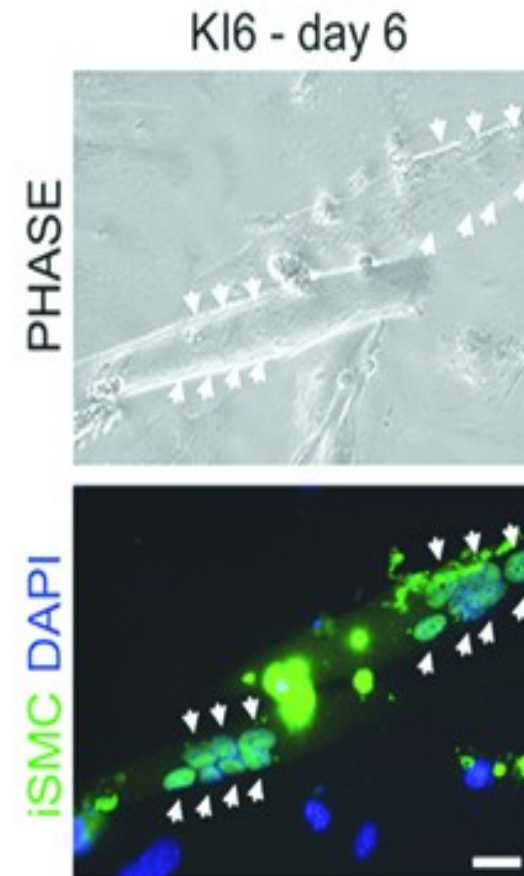
Figure 2.3 (con't).
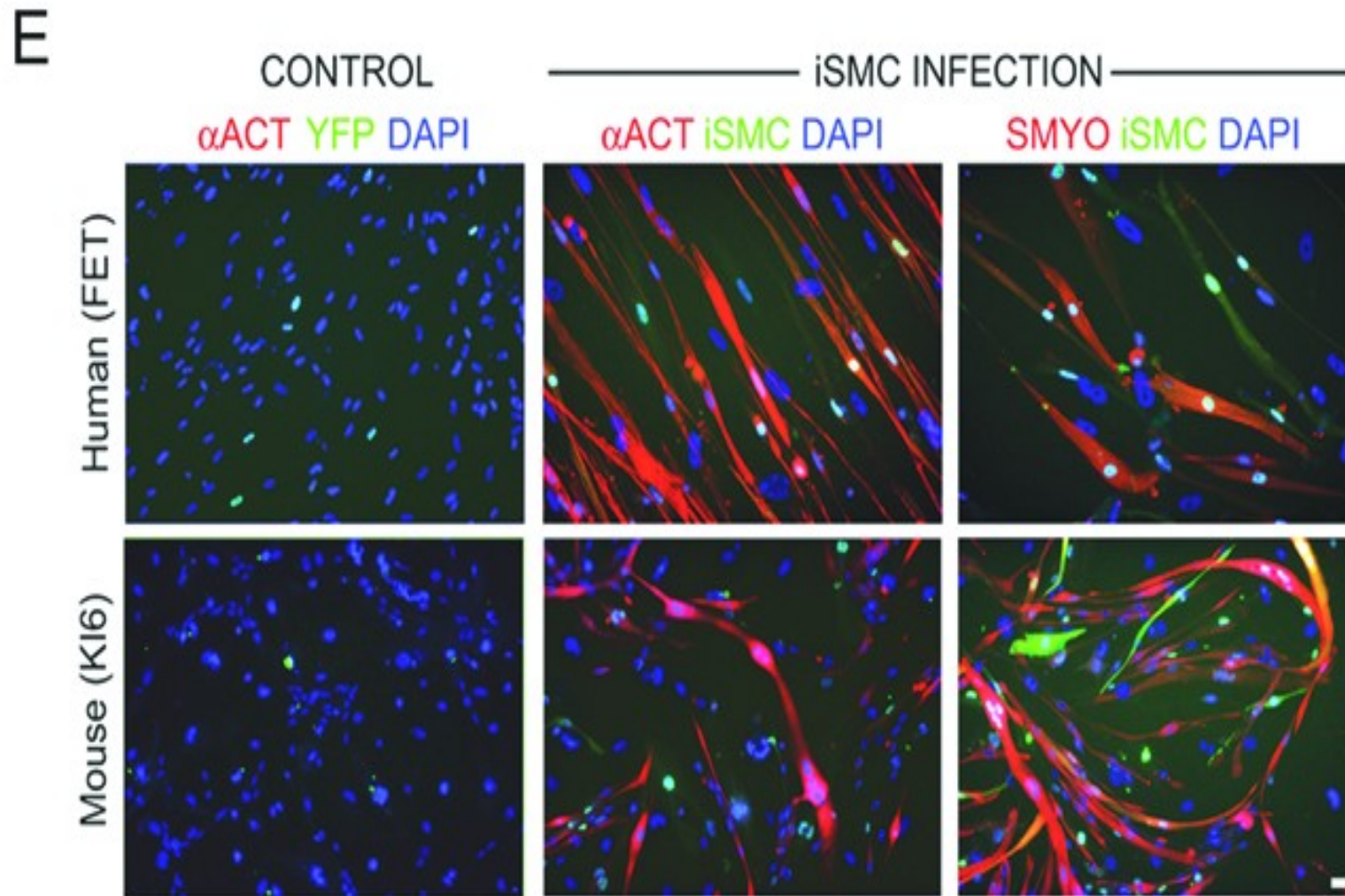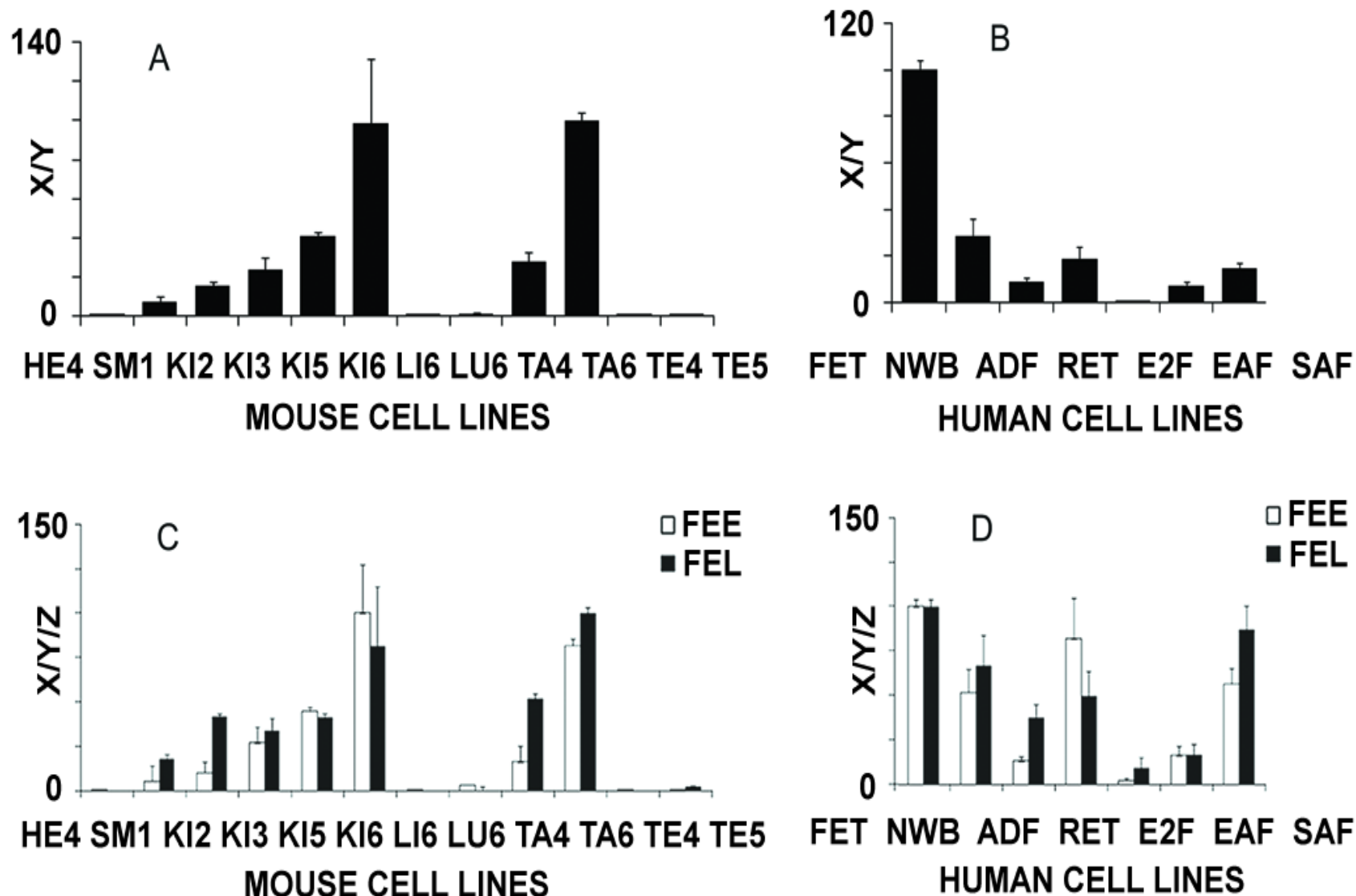
Figure 2.4. Data analysis for induction of iSMC factors and iSMC phenotype in mouse and human fibroblast lines. A-D) Relative conversion of mouse and human fibroblast lines to iSM cells as labeled (after Figure 2.2, A-D).

As with iN cells, iSMC reprogramming efficiency calculates only the number of converted cells relative to all cells at the time of harvest (Figure 2.4, panels A and B). Alternatively, both infectability and reprogramming efficiency can be calculated using either FEE or FEL (Figure 2.4, panels C and D) in lieu of transgene expression. Either way, our measures revealed substantial differences among the cell lines of both species. There was a strong and significant positive correlation between the reprogramming efficiencies calculated using both FEE and FEL for mouse lines ($r=0.969$, $P = <0.001$) and human lines ($r=0.837$, $P=0.019$). More importantly, analysis of variance indicated differences in median iSMC reprogramming efficiencies significantly greater than would be expected by chance for both mouse ($P<0.001$) and human fibroblast cell lines ($P<0.05$) (Figure 2.4, panels C and D). These results confirm that although there are some differences in calculated reprogramming efficiencies depending on how or when factor transduction levels are measured, the overall results are similar and support the notion that there are substantial differences in the reprogramming capacity of individual lines, even of the same general type.

2.3.7 iNC vs iSMC reprogramming and factors involved in reprogramming efficiency. The reprogramming efficiencies for all cell lines and conversion regimens are combined in Figure 2.5 (panels A-D). Analysis of the correlation between iNC and iSMC conversion normalizing with either FEE or FEL revealed no significant correlation, indicating that efficiency of conversion to one phenotype is not predictive of reprogramming to another identity.

Since successful infection with MMLV-based vectors requires active cell division, and both reprogramming regimens produce a non-dividing, growth-arrested cell

phenotype, we sought to determine if growth-related cellular characteristics – infectivity and mitotic index – correlated with conversion efficiency. As shown in Figure 2.6 (panels A and B), comparing the reprogramming values without correction by FEE or FEL with relative growth and infectivity revealed no correlation with iNC conversion efficiencies. However, there was a significant positive correlation between infectivity/growth and conversion to iSM cells in both the human and mouse lines. Despite this correlation, there was still striking variability in the reprogramming efficiency of different lines, suggesting that multiple factors in addition to those regulating the cell cycle likely impact the reprogramming process.

2.4 Discussion

The initial impetus for this report arose from several observations in our laboratory. One observation was that cell lines of similar origin and properties often appeared to differ significantly in their capacity to reprogram to an alternate phenotype, and, while this difference was generally attributed to technical variables, this attribution was rarely, if ever, systematically investigated. Another was that reprogramming regimens described in the literature were sometimes not as efficient as reported, even when components such as viral vectors, media formulations, and culture conditions were very similar or even identical. Though it is recognized that properties of the input cell such as species of origin, cellular age, and cell phenotype can have a dramatic impact on reprogramming efficiency, our own observations and published studies where differences in reprogramming efficiency were observed even where the potential for technical variation was minimal (Davis, Weintraub, and Lassar, 1987; Lattanzi et al., 1998), led us question to what extent reprogramming efficiency could be expected to

vary among cell lines of the same type when methodological and technical issues were deliberately controlled.

The comparison of two conversion regimens over a large number of cell lines also permitted us to determine if converting cells resulted from a specific population of reprogramming-susceptible cells or if they likely derived from a "main" population of cells, each with their own independent capacity for adoption of a new identity. It was long thought likely that reprogrammed cells arose from cryptic populations of canonical stem cells of varying potency contaminating the main population, but direct reprogramming experiments from many sources to produce induced pluripotent stem cells (Takahashi and Yamanaka, 2006; Luo et al., 2011; Esteban et al., 2009; Ezashi et al., 2009; Liao et al., 2009; Takahashi et al., 2007; Yu et al., 2007), neurons (Vierbuchen et al., 2010; Pfisterer et al., 2011; Pang et al., 2011; Caiazzo et al., 2011; Ambasudhan et al., 2011; Yang et al., 2011; Son et al., 2011), hepatocytes (Huang et al., 2011; Sekiya and Suzuki, 2011), cardiomyocytes (Ieda et al., 2010; Chen et al., 2012; Efe et al., 2011) and other induced cell types have since largely disproved this proposition.

In our own study, it is unlikely that contaminating non-fibroblasts contribute significantly to the converted cell pool, considering the efficiency of reprogramming that we observed coupled with the probability of uptake of sufficient viral copies for conversion (particularly in the case of iN cells) and an absence of evidence of non-fibroblasts at the level of individual cells by immunocytochemical analysis. Therefore, if converted cells do arise through a specific "susceptible" population, it is likely that this population is a subpopulation of the fibroblasts themselves. Such cells could be a

purposeful biological variant, but they may also be "defective" cells that lack the determinants that allow them to maintain a fixed identity when challenged with a stronger competing program. These lacking determinants may be genetic in nature, arising from point-mutations that alter gene expression or function; epigenetic, arising from alterations in methylation or modification of histones that regulate gene silencing; or contextual, arising from some cellular characteristic that is determined by the cell's interaction with other cells (perhaps akin to something like side-population cells – Challen and Little, 2006).

2.4.1 General susceptibility explanation. Our results are at odds with the idea that converted cells arise from a fixed subpopulation of cells with *general* susceptibility to identity respecification. When subjected to two separate reprogramming regimens the efficiency of conversion to one phenotype did not predict reprogramming efficiency to a second phenotype. Although only two conversion regimens were tested, this result suggests that a mechanism such as a genome-wide failure of some aspect of the molecular machinery controlling cellular identity does not underlie phenotypic lability; but instead, that the process is likely to be stochastic, falling by chance on genes that permit the cell to respond only to specific transcription factors and to respecify only to some phenotypes.
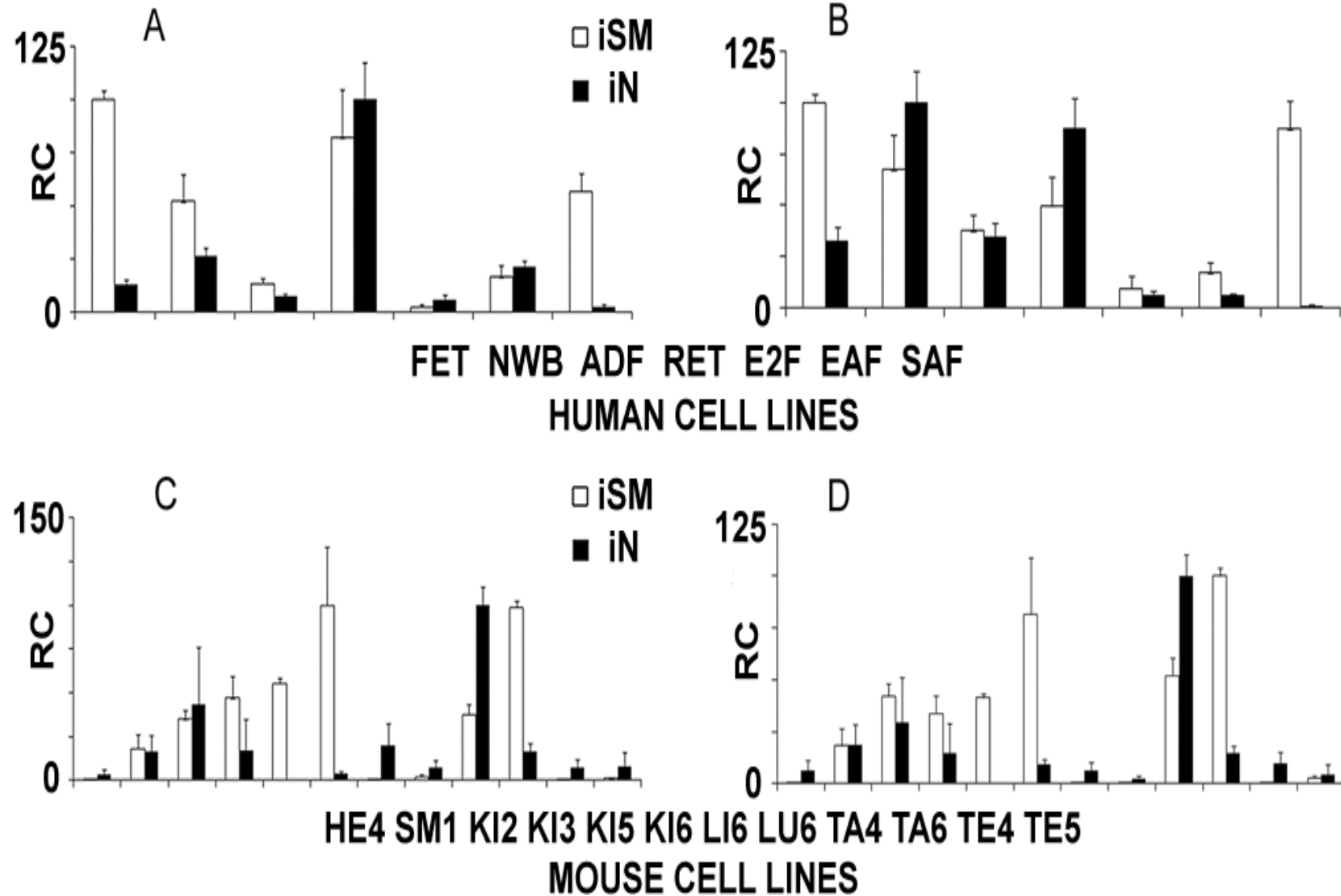
Figure 2.5. Comparative conversion of mouse and human fibroblasts to iN cells and iSM cells. A) Comparative conversion of human fibroblast lines to iN cells (black bars) and iSM cells (white bars) calculated as a function of FEE or B) as a function of FEL. C) As in (A) for mouse fibroblast lines. D) As in (B) for mouse fibroblast lines.
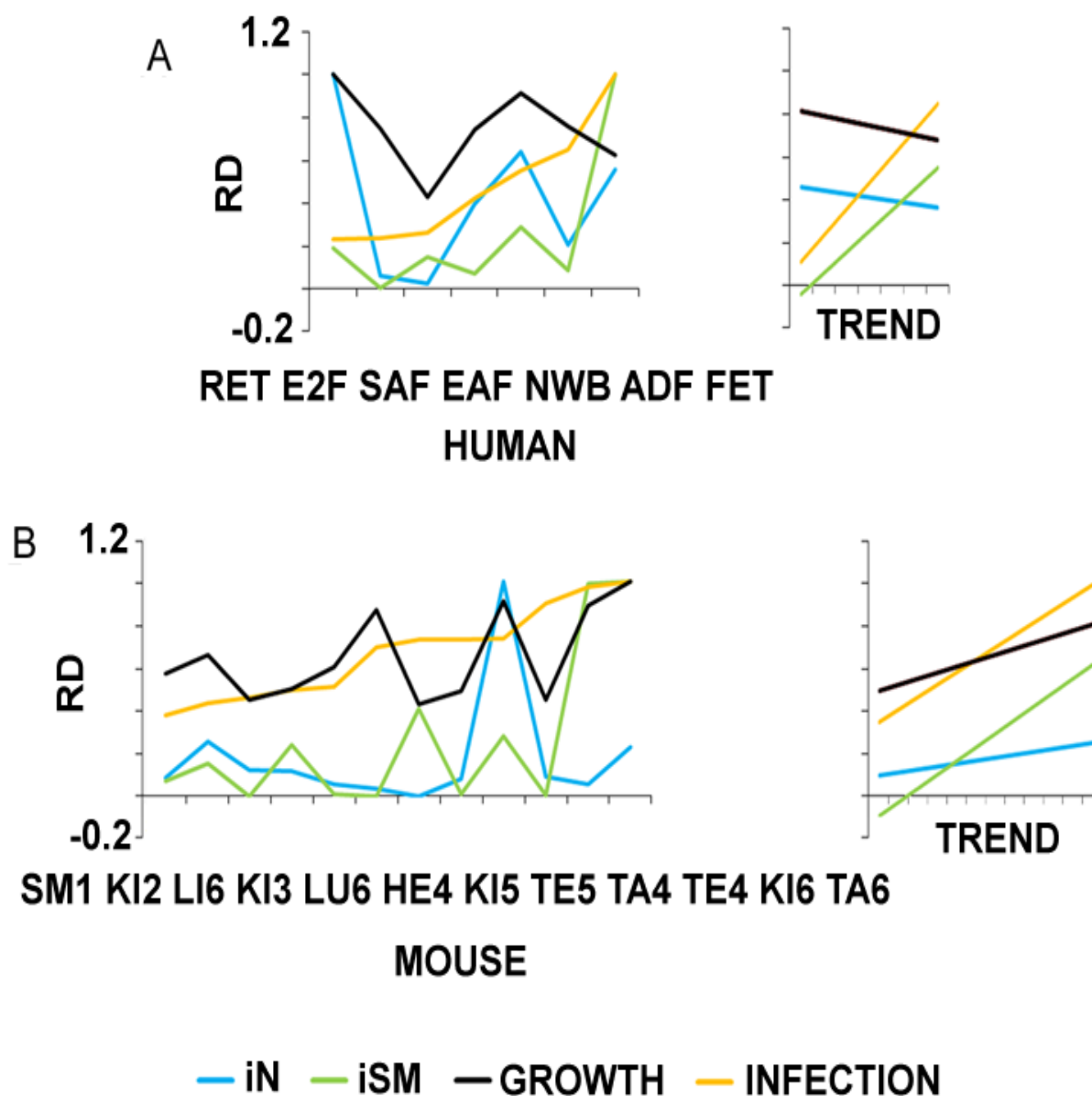
Figure 2.6. Data analysis for comparative conversion of mouse and human fibroblasts to iN cells and iSM cells A) Line graph of mean relative iNC conversion (iNC), iSMC conversion (iSMC), mitotic rate (GROWTH), and infection efficiency (INFECTION) for each of the human fibroblast lines sorted by increasing infection efficiency. At the right is a plot of the trend lines.

Our result also implies that the fraction of reprogramming-receptive cells in any given population may arise from any of a number of discrete genetic or epigenetic mutations (possibly even single-base changes) in critical determinants in or around the control regions of genes or classes of gene critical to the maintenance of a specific cell

phenotype. These mutations would also only be present in a fraction of the population – child cells of the cell harboring the mutation – and would be heritable and account for the apparent "stability" of reprogramming to a given phenotype within most lines. Whether these alterations are epigenetic, genetic, or some combination of both remains to be elucidated. If only tiny changes in the genome or epigenome are responsible for susceptibility to conversion, it may be difficult to isolate or even identify individual cells that are "good reprogrammers" *a priori*. Very sensitive methods that combine sequencing with quantification, such as whole genome sequencing, genome-wide bisulfite sequencing, or RNAseq may be able to identify even single-base changes that correlate with reprogramming efficiency and provide insight into gene targets and pathways that could be manipulated to improve phenotype-specific conversion.

2.4.2 Replicability and the tissue-of-origin explanation. Our results further indicated that, when technical components were held constant, the direct reprogramming capacity of independent primary fibroblast cell lines was reproducible within a single line from experiment-to-experiment, but varied dramatically from line-to-line irrespective of whether the lines were derived from different donors or from a single subject. It was notable that the disparity in reprogramming efficiency between different lines -- that in some instances differed by orders-of-magnitude -- was more than sufficient to account for discrepancies in reproducing published reprogramming regimens if the target cells were not of very similar type. Indeed, our results suggest that the production of primary lines of "very similar" type for use in reprogramming experiments by independent laboratories may, in itself, be a very challenging proposition. Fibroblasts cultured from different body regions are known to have distinct differences in gene expression for

transcription factors such as members of HOX gene clusters (Chang et al., 2002) despite sharing general characteristics and a common phenotype. It may be that propagation of tissue explants and repeated passage of resulting cells magnifies intrinsic differences such as those seen with HOX genes, even in lines derived from the same general tissue source. This would explain our results with cell lines like KI2, KI3, KI5, and KI6 that displayed dramatic differences in reprogrammability despite having originated from proximal tissue fragments from a single kidney harvested from a single mouse.

Along these same lines, we observed clear reprogramming differences in fibroblasts derived from different tissue sources, but cannot conclude that specific tissues repeatedly give rise to cells with superior or inferior reprogramming capacity until a larger number of samples can be assayed. In our hands, skeletal muscle-derived fibroblasts were poor converters to iSMC identity, indicating that in at least some cases, tissue of origin does not predict relative capacity to reprogram to that same tissue. A larger number of lines derived from more subjects will be needed to begin to see patterns in the reprogrammability of fibroblasts from different tissue sources.

2.4.3 Neurological disease confers a special phenotype explanation. Similarly, we observed that our two human cell lines isolated from patients with neurological disease (RET and SAF) displayed reprogramming efficiencies that were at opposite ends of the spectrum with regard to iNC conversion, but were very similar with regard to iSMC conversion. While it is tempting to speculate that iNC conversion was in some way influenced by the determinants of disease, the variability in efficiency that we observed between other human and mouse lines suggests that until a much higher number of

lines from multiple patients are compared directly, these differences as likely to arise by chance as from the determinants responsible for neural dysfunction. It will be interesting to determine in future experiments if "qualitative" measures of converted cells display as much variability as the number of cells converted. For example, if converted neurons, regardless of their total numbers, display an essentially equivalent level of maturation and electrophysiological response, it may be that lines from a lower number of donor subjects would be sufficient to produce meaningful results.

2.4.4 Final conclusions. Our results comparing the cell growth characteristics of individual lines and reprogramming efficiency suggest that there may be gross physical properties that either modulate or "predict" reprogramming efficiency. We expected and observed some correlation between either infection efficiency nd reprogramming or mitosis and reprogramming. However, a significant correlation was only observed with iSM cells and not with iN cells. With more cell lines and more detailed analysis of characteristics such as cell growth, the molecular determinants that account for this difference may be identifiable and may provide us with insight into the primary forces that govern phenotypic conversion.

Ultimately, our findings suggest that parallel testing of multiple cell lines from several sources may be needed to accurately determine the efficiency of direct reprogramming procedures, and by extension, that the most direct means of improving reprogramming efficiency for a given regimen may be to simply test more fibroblast lines, even if these lines are from a single source.

CHAPTER 3: COMPARING INDIRECT-DERIVED ES CELLS AND DIRECTLY-DERIVED IPS CELLS

3.0 Abstract

Comparing two types of reprogramming (somatic cell nuclear transfer and direct) to a single, pluripotent phenotype can reveal much about the underlying architecture of the reprogramming process. Unfortunately, existing comparisons are limited. To address this, two comparisons were made: a literature review that established what is known about such comparisons, and a quantitative analysis of secondary microarray data. Our analysis contributes to the existing literature by using the 8 cell embryo as the standard for comparing a range of ES and iPS cells. The goal of these comparisons was to establish differences between cell types of the same cellular state (pluripotency). Using a criterion independent of previous studies, it was found that iPS cells are transcriptionally closer to 8 cell embryos than are ES cells. However, when comparing both ES cells and 8 cell embryos with iPS, the same genes tend to be identified as different. Annotation using gene ontology terms revealed about half of these significantly different genes were ribosomal proteins. These results were confirmed in two ways. One of these was through a mutual information analysis that revealed elevated mutual information for all gene expression in iPS cells. The other was through an indirectly reconstructed gene network analysis, which revealed diistinct networks for genes that were different between the iPS-ES/iPS-8 cell and ES-8 cell comparisons. This information can be used to improve our choice of reprogramming approach (SCNT vs. direct) in future endeavors, as well as identifying groups of supportive (but not

essential) genes that might be used to augment efficiency in the reprogramming process.

3.1 Introduction

Cellular reprogramming for the derivation of pluripotent cells comes in two complementary forms. Indirect reprogramming involves a regulatory respecification of the nuclear genome by moving the nucleus a cell into another cell. The most commonly used technique involves transplanting the nucleus of a somatic cell into the cytosol of an enucleated oocyte (Hochedlinger and Plath, 2009), also called somatic cell nuclear transfer (SCNT). Direct reprogramming involves exogenous factor-driven (e.g. viral vectors, RNA, small molecules) respecification of phenotype (Hanna, Saha, and Jaenisch, 2010). This is done by delivering key factors (e.g. transcription factors) to a population of cells, and results in an induced-new phenotype, e.g. induced pluripotent stem cells (iPSC). Both techniques are effective at generating cells that resemble embryonic stem cells (ESC). The mechanisms underlying such dramatic transformation are poorly understood. In many ways, the science of direct and indirect cellular reprogramming is like poking a wild animal with a stick. Like the animal, a cell has a complex internal state (Sharov, 2012) that responds to external perturbations in a highly nonlinear fashion. In fact, predicting the response of a poked bear is far easier than predicting the response of a cell undergoing reprogramming, largely because the bear's behavior is better characterized. In the context of reprogramming practice, the analogy suggests that although we know how to kick-start the reprogramming process, we can do very little to control its outcome. However, our analogy also suggests that if we can better understand the molecular mechanisms (in this case, the distinguishing features of

gene expression) that underlie the stemness state, this information will allow us to better steer a reprogrammed cell in the desired direction.

3.2 Direct vs. Indirect Reprogramming

While both direct and indirect reprogramming can produce a similar cellular phenotype, there are a number of features which distinguish the two methods. Direct reprogramming has allowed for the creation of pluripotent cell types at a much higher throughput than with indirect methods. In this case, throughput refers to the number of cells that can be introduced to the reprogramming regimen in a single experiment rather than a measurement of efficiency. This is an important distinction, as estimates for efficiencies not only range widely, but are rather low as well. The most optimistic estimates of efficiency provided by Colman and Dressen (see Takahashi et al., 2007) are 3.4% for embryonic stem cells (ESCs) derived via indirect methods and 1-3% for pluripotent cells derived via direct methods. The focus of direct reprogramming (e.g. viral-mediated reprogramming (Sul, Kim, Lee, and Eberwine, 2012), RNA- (Li, Yang, Nakashima, and Rana, 2011) and miRNA- (Munsie et al., 2000) mediated reprogramming is at the level of population. By contrast, the focus of indirect reprogramming (e.g. SCNT (Hochedlinger and Jaenisch, 2006), cell fusion (Hansis, Barreto, Maltry, and Niehrs, 2004), and cell/protein extract (Hochedlinger et al., 2004) is at the scale of individual cells. In addition, direct reprogramming involves multiple, coordinated changes to the cell's genetic regulatory network. Ideally, the response to the direct reprogramming cocktail (e.g. introduction of *Oct4*, *Sox2*, *Klf4*, and *c-Myc*) involves a transcriptional regulatory cascade, or the successive activation transcription factors and downstream genes. This process uses many of the same genes and gene

regulatory networks that are upregulated or otherwise involved in regenerative processes and stem cells (Bolouri and Davidson, 2003), and should result in a new cellular phenotype. By contrast, SCNT acts by resetting the epigenetic landscape of the original cellular state without introducing large-scale genetic changes (Lathia and Rich, 2012).

While the final outcome for both procedures is similar (e.g. a pluripotent cell- broadly defined) at the onset of the reprogramming process, both techniques are fundamentally different, and therefore it is reasonable to expect both quantitative and qualitative differences in the ESCs produced using indirect methods (SCNT-ES) and direct methods (Figure 3.1).

Studies comparing iPSC and SCNT-ESC derived from the same original somatic cell line are scarce. However by performing statistical assessment of previously- collected microarray data and a thorough analysis of available data that compares SCNT-ESC and iPSCs against a common reference (ESC produced from fertilized embryos) we hope to convey the message that while SCNT-ES and iPSCs are mostly identical, there are however, key differences that have relevance when considering improving current approaches to the reprogramming process.

3.3 Known Differences between SCNT-ESCs and iPSCs.

Using the less sophisticated but most commonly cited methods for determining pluripotency, the signature of both SCNT-derived ESCs and directly reprogrammed iPSCs is one of a fully pluripotent cell, including the upregulation of *Oct4*, self-renewal, induction of teratoma formation, and the ability to be differentiated into cell types such as neurons and beating muscle (Hochedlinger and Jaenisch, 2006; Chin, Mason, Xie,

Volinia, and Singer, 2009; Guenther et al., 2010). Bock et al. (2011) took into account multiple iPSC and ESC lines and confirmed the result established in previous studies (Krause and Scadden, 2012), of a highly characteristic "pluripotency" program consisting of common gene expression patterns also called "stemness".

Even differences in the regulatory program may not be due to intrinsic differences in a hypothetical gene expression or epigenetic program. Differences observed between ESCs and iPSCs might be due to a microenvironmental niche (Byrne, Pedersen, and Clepper, 2007). Particularly, it is the existence of a functional hierarchy among signaling molecules, which contribute to either the initiation of large-scale phenotypic change or maintenance of the current state. When a cell is in the proper niche, a particular cellular state is reinforced. This requires interactions with neighboring cells as well as the extracellular milieu, which makes this contribution hard to account for (Yoon and Seger, 2006).

3.4 Different but Complementary Processes

The literature on direct comparisons between SCNT-ES and iPSCs suggests that SCNT-ESCs are closer to ESCs than are iPSCs (Stadtfeld et al., 2010; Byrne, 2011; Kim et al., 2010). Side- by-side comparisons of iPSCs and SCNT-ESC derived from the cell animal have shown specific differences. The source of that variation however is difficult to determine. One hypothesis suggests that the developmental context of a given cell is essential for determining how completely and successfully it reprograms. One example of this involves iPSC nuclei that have never been exposed to maternal reprogramming factors. In a study by Stadtfeld et al. (2010), transcriptional comparisons between genetically-identical mouse ESC and iPSCs revealed no significant differences

except for those of the Dlk1-Dio3 gene cluster, which are maternally-inherited and dysregulated during development to enable developmentally-related epigenetic silencing.
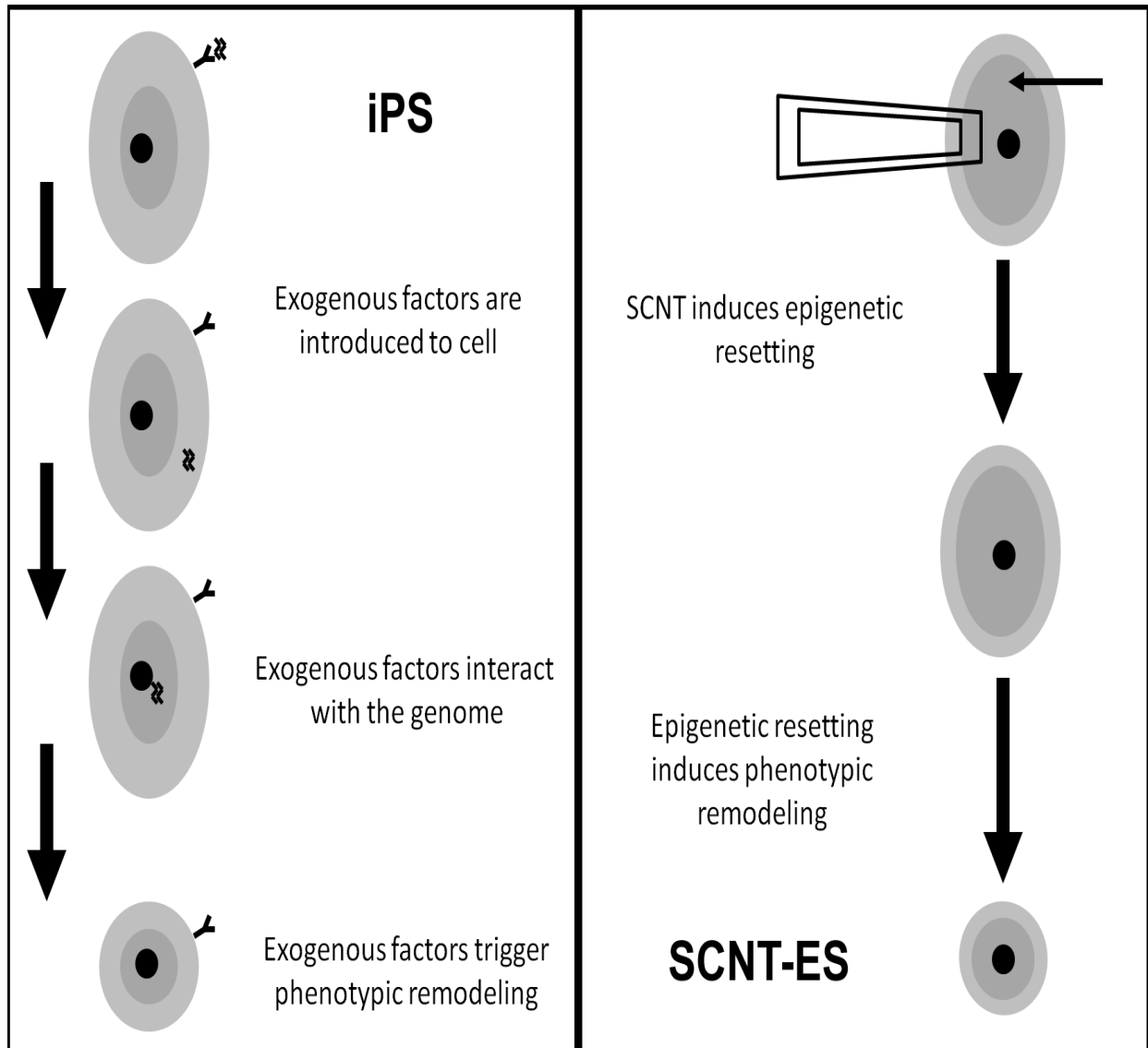
Figure 3.1. Overview of process for direct (via viral mediation, left) and indirect (via SCNT, right) reprogramming. LEFT: directly-reprogrammed cell morphology transforms along with the incorporation of reprogramming factors into genome and active expression of these genes. RIGHT: indirectly-reprogrammed cell morphology transformation that occurs in tandem with regulatory respecification of the cell's genome. Shaded boxes represent microenvironmental changes: for both processes, top box is supportive of fibroblasts, middle box is a transitional milieu, and bottom box is supportive of pluripotency.

3.4.1 Different but complementary processes: kinetics. There are also a number of overarching comparisons that can be made in the service of understanding these two types of reprogramming. This is based on a number of features common to both processes in different species and at different levels of genomic regulation, and may suggest that a re-interpretation of the stemness concept is needed.

The first example involves a comparison of the processes that distinguish direct and indirect reprogramming. This comparison is hypothetically-oriented (Takahashi et al., 2007), and focuses on the biochemistry of SCNT-ESCs. One of these processes involves the kinetics of genetic regulation in the service of phenotypic remodeling. The kinetics of the reprogramming process is fast in SCNT-ESC (Liu et al., 2008), and slow (Colman and Dreesen, 2009) and stochastic in iPSCs (e.g. involve multiple levels of genetic regulation - Takahashi et al., 2007). These differences may also be explained by the interactions between the nucleus and cytoplasm component during the process. In indirect reprogramming, this usually tight relationship is decoupled, whereas in direct reprogramming this is not the case (Takahashi et al., 2007). A study by Paull et al. (2013) suggests that even though the percentage of mitochondrial DNA transferred with the nucleus during indirect reprogramming is small (< 1%), it still enables subsequent development in an efficient manner.

There are also differences in mechanism with regard to chromatin remodeling complexes (Orkin and Hochedlinger, 2011), which are in an open state before pluripotency genes are activated in the oocyte but not in iPSCs (Awe and Byrne, 2013). According to the chromatin opening/fate transformative model of pluripotency (Chin et al., 2010), this should result in more complete reprogramming among SCNT-ESCs.

3.4.2 Different but complementary processes: cross-species. Another way to look at the differences between direct and indirect reprogramming involves the use of animal models representing two different species: rhesus monkey (*Macaca mulatta*) and mouse (*Mus musculus*). In a global transcriptome analysis of cell lines from rhesus monkey (Byrne, 2011), not only were SCNT-ESCs found to be closer to the ESCs standard than iPSCs, but SCNT-ESCs were deemed to be more completely reprogrammed. This same phenomenon was observed the mouse with regards to the epigenetic memory of the donor cell (e.g. SCNT-ESCs) were more completely reprogrammed (Kim et al., 2010). In the rhesus monkey study a cluster analysis of gene expression was used to determine "transcriptional closeness". In their analysis, the 500 significantly upregulated genes in ESCs lines are both central (*Oct3/4*) and peripheral (*NR5A2*) to the reprogramming process. In addition, a fair percentage of these genes are involved in biological processes such as cellular maintenance (16%) and metabolism (10%). A mouse study based on a genome-wide methylation assay clearly shows that there exists a difference between SCNT-ES and iPSCs in terms of epigenetic memory. Specifically, direct reprogramming does not involve a direct resetting of the cell's epigenetic markers and the potential exists for methylation marks indicative of the previous cellular state to survive the reprogramming process. According to Kim et al. (2010), differences due to epigenetic factors are substantial. This can be seen in the case of differentiation potential of neurally-derived iPSCs, which are more resistant to the conversion into blood whereas iPSCs derived from leukocytes are more prone to do so. It is worth mentioning that none of the original four transcription factors used by Yamanaka and collaborators has 'chromatin regulation' as primary function.

Using data from rhesus monkey, Byrne (2011) suggests that oocytes (e.g. SCNT-ESCs) contain factors that enable more complete reprogramming to the native ESC state. Yet we can also see that "supplemental factors" can also help iPSCs overcome the biological shortcomings of the direct reprogramming process. While the existence of an optimal set of enhancement conditions is debatable, these experimental outcomes raise the issue as to what contributes to this observed advantage. Perhaps we can uncover this reprogramming-favorable gene expression background using a quantitative analysis of secondary data.

3.5 Quantitative analysis of secondary data

The most common comparison at the gene expression level is done by looking at the transcriptome of the cells, and by comparing gene expression profiles. In one study, the genomic and epigenetic profiles of multiple iPSCs lines were shown to become both more like each other and to become more like ESCs over time (Chin et al., 2010). While both within- and between-species studies suggest that there are very few gene expression differences between SCNT-ESC and iPSCs, the question becomes whether or not these differences are due the methods themselves, developmental (SCNT-ES) vs. non-developmental (iPSC). It is also difficult to rule out the possibility that the differences found in the final product – a pluripotent cell – are due to natural variationthat defines distinct pluripotent cell lines. Furthermore, while the typical approach is to reveal significant differences that arise from essential changes to the transcriptome, a supportive environment must also exist in order for phenotypic remodeling to take place. In the SCNT-ESC/iPS comparison, we have two techniques

that result in the same cellular state (pluripotency). Yet while the resulting cells are of the same state, their supportive transcriptional milieu might be very different.

A related issue involves transcriptionally homogeneity of the pluripotent state itself. By looking at intra-state transcriptomic diversity, we might be able to uncover these supportive factors. However there is a need for a control cell type that has uniform cellular composition. We put forward the notion that the human 8-cell preimplantation embryo represents the ideal control. Its equivalent in the mouse, the four cell embryo, has demonstrated totipotency in multiple studies recorded in a strong body of literature that spans two decades (Kiessling et al., 2009; Telford, Watson, and Schultz, 1990). Our premise is that supportive genes can be revealed through an unconventional re-analysis of 15 microarrays that represent four sets of experiments and three cell types (iPS, ES, and 8-cell embryo). While not directly related to SCNT-ESC methodology (human SCNT-ESC are not available), these data can provide an objective perspective on what characteristics pluripotency confers on cells. Therefore, we used previously collected microarray data and two statistical models (mutual information and an objective distance metric) to better understand these differences.

Each comparison was made between multiple iPSCs, ESCs (Avery et al., 2008, and 8-cell embryos (Xie, Chen, He, Cao, and Zhong, 2007). In this case, fertilization-derived ESCs were used, assuming that their gene expression is mostly identical to an indirectly-reprogrammed (e.g. SCNT-ES) pluripotent cell. We used two sources of iPSCs: iPSC lines from our own laboratory (Cellular Reprogramming Laboratory, Michigan State University) and iPSC lines obtained from Gene Expression Omnibus

(GEO) (Masaki et al., 2007). The 8-cell embryo samples were included to demonstrate how both types of pluripotent cell compared to a totipotent gene expression signature.

3.5.1 Cell line information and methods. All analytical work is done using MATLAB, Excel, and R. Thirty-one (31) previously acquired human microarray studies that encompass fibroblast and pluripotent cell type diversity, and embryo time-series diversity. Pre-processing of microarray data is done using the AMP (automated microarray pipeline) tool located at http://compbio.dfci.harvard.edu/amp/. Microarray data are normalized using the RMA (robust multi-chip average) method. *n*-fold expression values are derived by dividing the individual values for each probe into the mean of the dataset. *z*-score transformations are also used to normalize each set of probes, and serve as an alternative to the n-fold expression criterion. Eighteen (18) previously acquired human microarray assays from five (5) experiments were used. These data encompassed fibroblast and pluripotent cell type diversity, and are shown in Table 3.1.

For ES cells (GEO database accession numbers: GSM194307, GSM194308, GSM194309) lines from Avery et al. (2008) are used. These samples are derived from human tissue, and used the Affymetrix Human Genome U133 Plus 2.0 platform. For one set of iPS cell lines (GEO database accession numbers: GSM245339, GSM245341, GSM245442, GSM257520, GSM257339, GSM257524) lines from Masaki et al. (2007) are used. These samples are derived from human tissue, and used the Affymetrix Human Genome U133 Plus 2.0 platform. For the set of 8 cell embryonic samples (GEO database accession numbers: GSM456652, GSM456653, GSM456654) lines from Xie et al. (2011) are used. These samples are derived from human tissue,

and used the Affymetrix Human Genome U133 Plus 2.0 platform. The iPS lines (104, 21010, 2555) each consisted of microarrays collected at the Cellular Reprogramming Laboratory, Michigan State University. These samples are derived from human tissue, and used the Affymetrix Human Genome U133 Plus 2.0 platform. For skin fibroblasts (GEO database accession numbers: GSM301264, GSM301265, GSM301266), the control condition from Duarte, Cooke, and Jones (2009) are used. These samples are derived from human tissue, and used the Affymetrix Human Genome U133 Plus 2.0 platform.

3.5.2 Parameteric description of microarray analyses. For the microarray analyses, each set of ith replicates for j^{th} cell types were averaged to yield a mean expression value. Each resulting vector was filtered for all common genes that also scored above the 50th percentile. Common genes were found by finding the intersection between n^{th} vectors (common identity based on probe ID). Pairwise comparisons were done by evaluating every k^{th} element between each vector m and n.

3.5.3 Discovery of potentially supportive genes. To demonstrate the actual difference between reprogrammed iPSC and ESCs, we used microarray data to establish an absolute distance in terms of fold-change between cell lines (see Figure 2). After normalizing for differences in platform types and technical variation, we analyzed 5,595 genes. The fold-change was then calculated for each sample, and the median of all replicates for a given gene, cell line combination was used for purposes of comparing absolute differences (e.g. distance). Only distances in the 50th percentile and above were considered for direct comparison. The goal of this analysis was to find the most different genes as characterized by our pairwise distance metric within the defined

intersection among the cellular states being compared (e.g. iPSC-ESC). Those lists were then annotated and subject to further analysis.

Table 3.1. Source data for secondary analyses.

Tissue Origin	Number of Microarrays	Organism
Skin	3	Human
ES cell culture	3	Human
iPS cell culture	6	Human
iPS cell culture (CRL lab)	3	Human
8 cell embryo	3	Human

Figure 3.2 shows us the result of this methodology. When comparing iPSCs and ESCs lines, 105 genes were selected as different using our criteria. A comparison of iPSCs and 8-cell embryo lines shows a similar result (106 genes). Comparisons between the 8-cell embryo and ESCs lines show a slightly larger set of genes with above threshold different expression (173 genes). Annotation of these genes using gene ontology (GO) showed that most of these genes are involved in either remodeling the cell or building structural components. This is to be expected in development (e.g. the ESC and 8-cell embryo samples), but should also be expected in conversion to an iPSC phenotype. Of particular note are the disproportionate representation of ribosomal proteins (roughly 1% of total genes assayed) in these lists, constituting 52% (iPSC-ES comparison), 51% (iPSC-8 cell embryo comparison), and 39% (ES-8 cell embryo comparison), respectively.
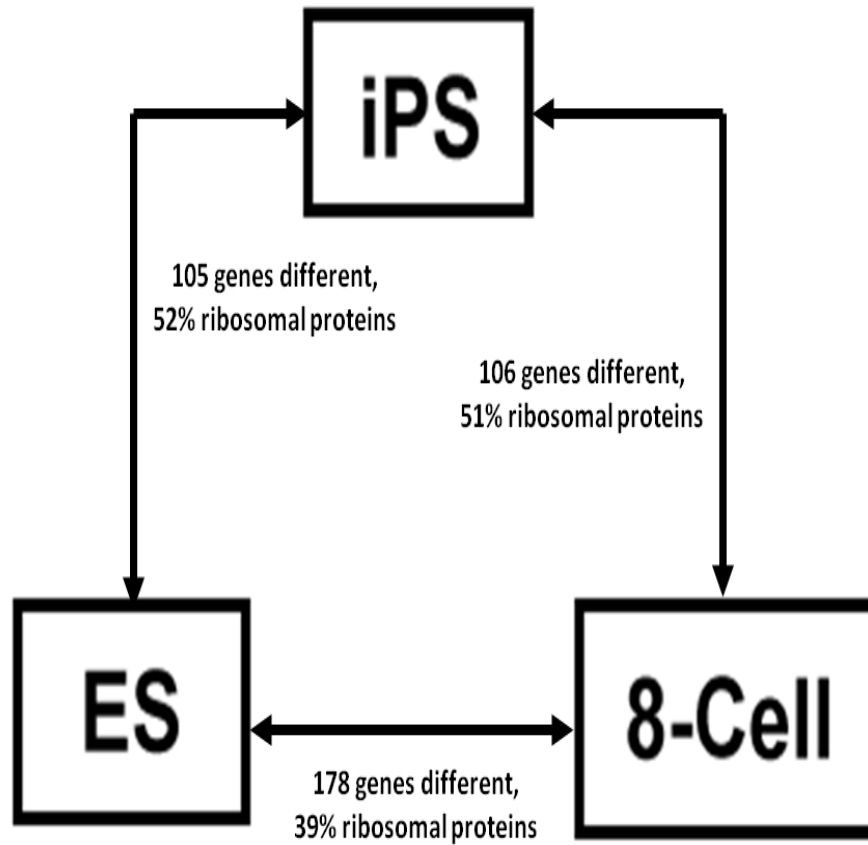
Figure 3.2. Gene expression differences between cell lines representing different types of pluripotent cells.

3.5.4 Further analysis using IPA Ingenuity. Further analysis using the IPA Knowledge Base (Ingenuity Systems Inc., Redwood City, CA) confirmed the GO annotations. Both comparisons with iPSCs yielded virtually the same genes, while the ES-8 cell comparison yielded a longer list with different constituent genes. Yet across all lists, the proportion of genes classified by Ingenuity as possessing a unique identity (64-65%) and proportion of genes whose expression is localized to the cytoplasm and nucleus (72-78% for cytoplasm, 9-15% for nucleus) is similar. For a list of the top 25 gene identities for each pairwise comparison, see Table 3.2. The IPA core analysis also provided information on indirect gene network reconstruction (Ideker, Oxier, Schwikowski, and Siegel, 2002) for each cell type comparison.

69

Genes associated with our iPSC-8 cell and iPSC-ESC comparisons yielded significant scores for the following gene network functions: protein synthesis, neurological/immunological diseases, and cancer. Genes associated with our ESC-8 cell comparison yielded significant scores for a different set of network functions: cellular development, cell death and survival, post-transcriptional modifications, and DNA replication/cellular assembly. This again shows that while functional differences exist between iPSCs and fertilization-derived ESC, these distinctions may not involve genes and pathways not typically associated with the core pluripotency process.

3.5.5 Mutual Information within and between pluripotent cell lines. To demonstrate that these iPSCs and ESCs lines share a common variance structures, and thus many of the same features important to defining the pluripotent phenotype, we measured the Mutual Information (MI) within and between cell types (see Figure 3.3). MI is an information-theoretic measurement of variance structure shared between two datasets (Steuer, Kurths, Daub, Weise, and Selbig, 2002). In the case of our pluripotent cell types, the greater the MI observed, the greater likelihood of similarity between cell lines.

Comparisons of MI content can be made within a cell type (among iPSCs replicates) and between cell types (a comparison of iPSCs and ESCs). Our limited analysis of 12,000 genes in Figure 3 demonstrates that there exists a greater amount of mutual information among replicates of the same cell lines (Figure 3, Part A) than there is between different cell lines (Figure 3, Part B). However, there are no large-scale differences between cell lines, nor are there large differences when compared to the within cell lines case. Focusing on the iPSC-ESC comparison, there is a greater amount of mutual information among ESCs lines than among iPSCs lines. As with result from

the literature, using SCNT-ESCs, this may be due to epigenetic memory among iPSC

samples depending on the tissue source of the original donor cell. While this is a

course-grained analysis, it does suggest that there is a core regulatory program shared

by all pluripotent cell types. This program may vary slightly due to transcriptional noise.

It also suggests a potential mechanism for initiating the changes required for phenotypic

respecification.

Table 3.2. Top 25 genes resulting from pairwise distance comparisons for iPSC-8 cell, ESC-8 cell, and iPSC-ESC. Genes with no annontation were not included in this table. RANK represents position of gene in rank order for that comparison.

ES-8 cell	RANK	iPS-8 cell	RANK	iPS-ES	RANK
NLRP4	1	CNOT6L	1	CNOT6L	1
KHDC1/RP11-257K9.7	2	DNAH5	3	DNAH5	2
PLAC1L	3	OR5H1	8	OR5H1	5
FOXR1	4	PER3	9	PER3	9
ZSCAN4	5	CFTR	13	CMTM6	13
ZNF280A	6	CMTM6	14	CFTR	14
FIGLA	7	GLS2	15	GLS2	15
UNC13C	8	SLC26A7	17	SLC26A7	17
MBD3L2	9	LOC283432	20	LOC283432	20
ZSCAN4	10	PDE6C	22	LOC441666	21
TRIM49	11	TPTE	23	HERC1	24
NLRP11	12	PDE8B	24	DNAJA4	25
LOC643224	13	LOC100127974	25	CCDC102B	26
PRAMEF12	14	TRPM1	26	LOC339260	27
BCAR4	15	TEX12	27	TLR4	29
RFPL1	16	SMAD4	31	GART	34
TRIM43	17	ATM	33	RXFP1	37
DPPA3/STELLAR	18	TRDN	35	NKTR	38
MAGEA2/A2B/A3/A6	19	CBLN4	36	PDE8B	39
TKTL1	20	OR2H1	37	TPTE	41
SYCP3	21	OR5J2	40	PSMA8	42
MAGEA3	22	C21orf109	43	SLC8A1	44
UNC13C	23	C14orf48	44	TLR4	46
HHLA2	24	PSMA8	45	LOC730092	47
NR2E1	25	RXFP1	46	TEX12	48
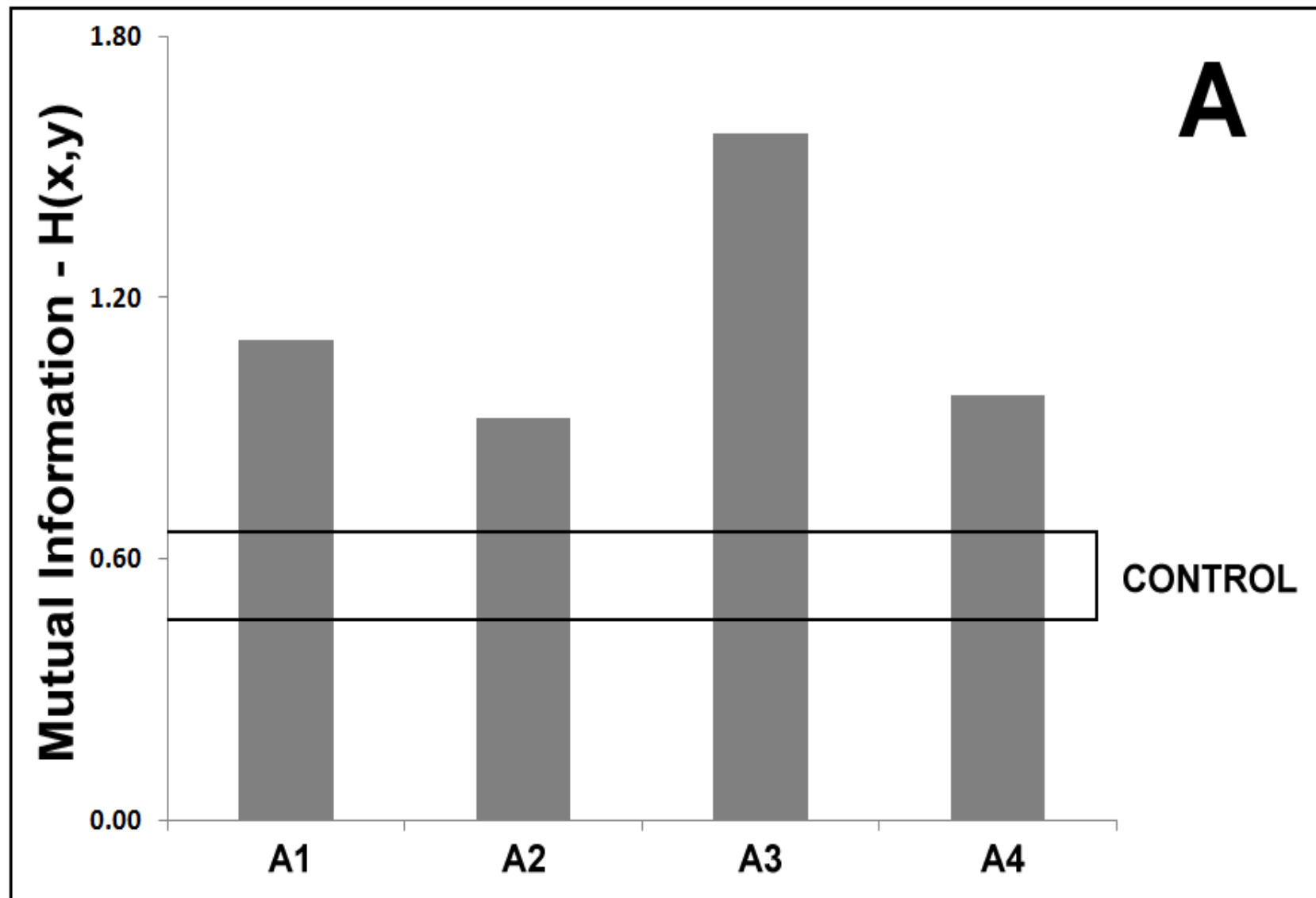
Figure 3.3. Mutual Information measurements within (A) and between (B) selected pluripotent cell lines. Outlined boxes (black, labeled control) represent the range of values for comparisons between comparisons of skin fibroblast microarray

Figure 3.3 (con't). data and all stem-like cell lines for the genes defined in this study. Data IDs are: A1 (iPS), A2 (iPS-CRL), A3 (ES), A4 (8cell), B1 (iPS, iPS-CRL), B2 (iPS, ES), B3 (iPS, 8cell), B4 (iPS-CRL, ES), B5 (iPS-CRL, 8cell), B6 (ES, 8cell).
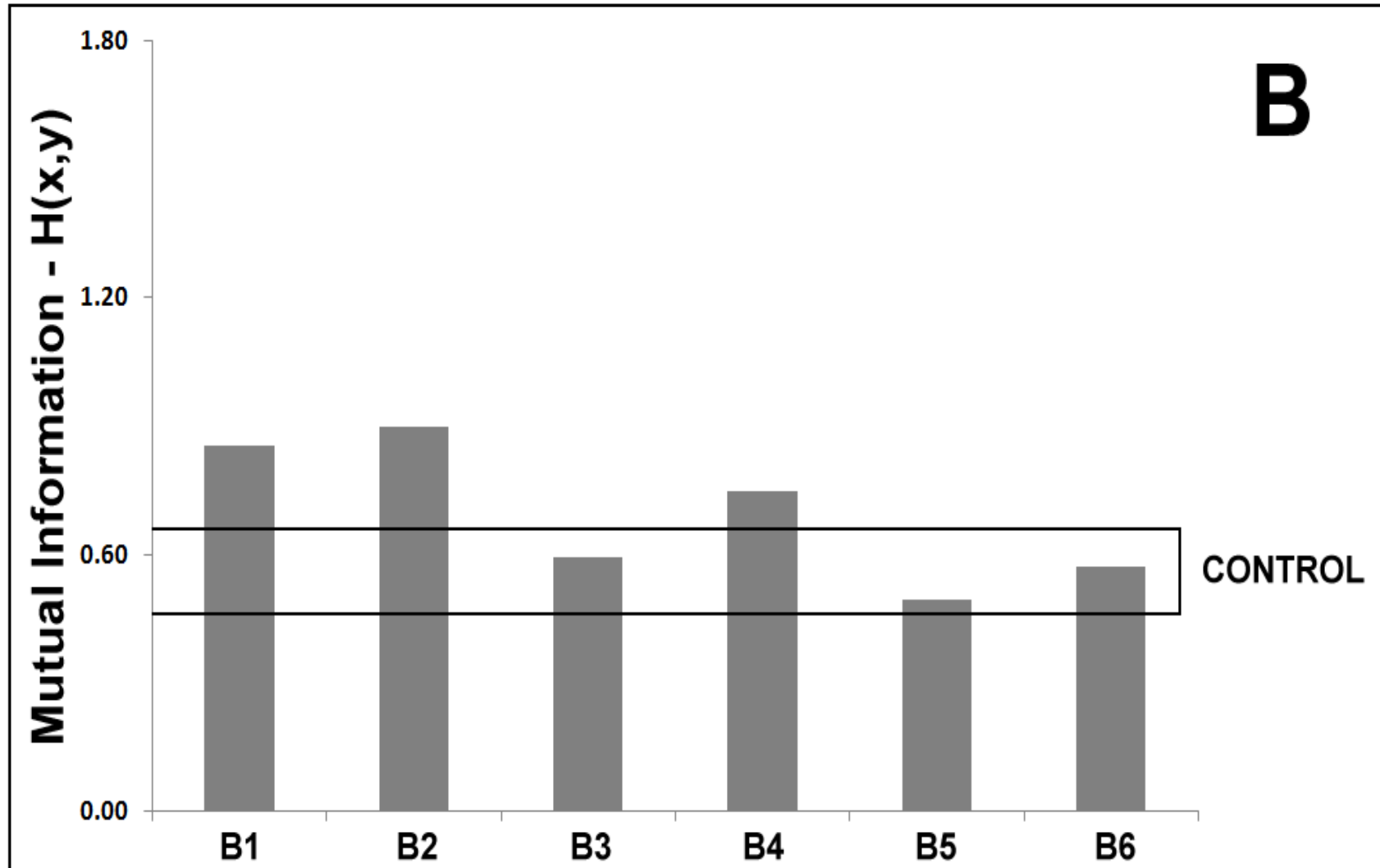
3.6 Conclusions

The first and most obvious difference between iPSCs and SCNT-ESC is at their origin. Depending on the method chosen, chromatin remodeling takes place either the moment the cell is fused to the enucleated oocyte (indirect reprogramming) or immediately after acquisition of a novel gene expression profile via viral vectors, mRNA, microRNA and/or small molecules. This novel pattern of gene expression is aided by changes in the extracellular environment trigger the process of phenotypic transformation that culminates in four distinctly different cellular stages: 1) no change; 2) cell death; 3) Pre-iPSCs and 4) iPSCs.

Despite the blind and oftentimes unpredictable nature of the reprogramming process, there are some insights that provide clues to future research directions. The first such difference involves phenotypic changes due to the reprogramming process. Cell reprogramming is said to exhibit slow kinetics (Stadtfeld, Maherali, Breault, and Hochedlinger, 2008). During this process, a reprogrammed cell must make multiple structural changes, many of which are indirectly tied to the changes in gene expression measured in most studies. Among these changes are translational mechanisms and the production of structural proteins. This is demonstrated by our quantitative result, and may explain why differences exist despite minor tweaks to the pluripotency program.

The second difference is related to feedback and other regulatory mechanisms associated with triggering and maintaining these phenotypic changes. The maintenance of the new stem-like state, perhaps better characterized as molecular reinforcement, is observed over time in cells reprogrammed to pluripotency. Under culture conditions that favor such cell type maintenance, molecular reinforcement might also serve as a

selection mechanism, favoring the survival of cells that are able to become more uniformly pluripotent over time.

In conclusion, the differences between directly-derived iPSCs and indirectly-derived ESCs may have little to do with the mode of reprogramming. When the mode of reprogramming is successful, there is almost no difference in the core regulatory program initiated by the cell. Perhaps more important are the consequences of reprogramming, and the variable kinetics this entails. From the utilitarian point of view, given this general lack of functional difference, Gurdon and Murdoch (2008) have suggested that nuclear transfer and direct reprogramming can be used to complement each other in a number of applications. However, to address the question of how pluripotency can be achieved by chromatin remodeling in a more predictable kinetics, SCNT can be a more reliable tool.

CHAPTER 4: NEW DIRECTIONS IN CELLULAR REPROGRAMMING

4.0 Abstract

While advances in cellular reprogramming techniques have been recognized by the Nobel committee (Nobel Prize Committee, 2012), the underlying systems biology of the reprogramming process is not well understood. In order to move from characterization to substantive theory, this review will focus on completing pieces of the reprogramming puzzle. These include modeling, population thinking, and natural variation in the reprogramming process. Initially, we will discuss the use of predictive and phenomenological models to characterize changes in cellular phenotype. While phenomenological models have dominated the literature, models that characterize the dynamics of the reprogramming process are also possible. Conceptual and computational models will be shown to provide context that experimental investigation cannot. Understanding the outcomes and constraints related to direct cellular reprogramming is to engage in population thinking. To this end, the concept of direct reprogramming as a population-level process will be presented in terms of potentially relevant models and the role of overall reprogramming efficiency in the process. Finally, we will discuss the role of variability in reprogramming, as there are several sources and effects that influence the efficiency and outcomes of the reprogramming endeavor. These include, but are not limited to: the combinatorial action of transcription factor activity, experimental replication, and effects of source variation on desired outcomes. This will be presented in the context of efforts to created induced cardiomyocyte (iCM) cells. We will conclude with a discussion of variability and reprogramming efficiency in a

broader context, and how this in turn might be used to shine light on the fundamental principles of direct cellular reprogramming.

4.1 Introduction

The conversion of one cell type to another via forced gene expression changes presents a biological paradox. On the one hand, the mechanisms of change are highly unpredictable. On the other hand, reprogramming is clearly not an epiphenomenon, as it can be done for multiple cell types that originate from different developmental germ layers. A cell has a complex internal state (Gianchandani et al., 2008) that responds to external perturbations in a highly nonlinear fashion. However, it is still the case that the conversion of fibroblasts to an array of cellular states is a largely blind process of trial-and-error. As such, failure in one lab might mean success in another, yet the reason behind the discrepancy never becomes clear. Although there have been attempts to improve the reprogramming process using drug treatments and small molecules (Feng et al., 2009), these attempts at process optimization obscure a larger set of issues. Through the development of better measurements, focusing on statistical structure of the reprogramming process, and understanding reprogramming as an emergent property, we can minimize the uncertainty when we challenge our cells with a reprogramming stimulus. The schematic in Figure 4.1 shows the "black box" nature of direct cellular reprogramming, and how convergent approaches might aid in a better understanding of the process.

Direct reprogramming is defined as the conversion of somatic cell types such as fibroblasts (Yamanaka and Blau, 2010) or pancreatic beta-cells (Stadtfeld, Brennard, and Hochedlinger, 2008) into another phenotype using a cocktail of transcription factors.

This cocktail is typically delivered to a population of cells in culture using a retroviral vector (e.g. infection), although alternate delivery strategies (e.g. using mRNAs or other small molecules) may be possible. While the number of transcription factors used is relatively small, the effects of direct reprogramming on cellular phenotype are widespread. This is because direct reprogramming takes advantage of the hierarchical structure and synergistic nature of genetic regulatory networks (Carter et al., 2002; MacArthur, Ma, and Lemischka, 2009) while retaining the genome of the host cell. Direct reprogramming has also been characterized as a slow kinetic process (taking days to weeks), with the generation of phenotypic changes being stochastic (and hence variable) with regard to time and across individual cells (Hockemeyer et al., 2008). This differs from indirect reprogramming, in which the chromatin state of the host cell is reset in order to mimic a desired cell type (Yamanaka and Blau, 2010). This triggers large-scale gene expression changes without directly manipulating hubs in genetic regulatory networks (Morgan et al., 2005), which is fundamentally different from what happens during direct reprogramming. While this difference leads to subtle changes in the resulting phenotypic, these features also suggest that there specific combinations of molecular mechanism states, microenvironmental conditions, and sets of induced transcription factors that provide maximal efficiency of conversion and a more controlled set of outcomes?

4.1.1 Direct reprogramming in context. To frame this question in context, we will now review what can be accomplished using direct reprogramming techniques. The first successful demonstration of direct reprogramming was in 1987, when Davis, Weintraub, and Lassar used a single transcription factor (*MyoD*) to create skeletal muscle fiber

(iSM) cells from fibroblasts. The next milestone in direct reprogramming research did not occur until the 2000s, when the five-factor approach of Jamie Thompson's group (Yu et al., 2007) and the four-factor approach of Shima Yamanaka's group (Takahashi et al., 2007) were used to convert fibroblasts into pluripotent stem (iPS) cells. Since this time, the creation of induced pluripotent (iPS) cells has become commonplace (Park et al., 2008; Patel and Yang, 2010). Even more recently, a series of papers have described the successful conversion of fibroblasts to a neuronal (iN) phenotype using a four-factor approach (Pang, 2011; Yang et al., 2011) and the conversion of fibroblasts to cardiomyocytes using three to four factors (Qian et al., 2012).

The direct reprogramming of readily cultured and maintained cell types such as fibroblasts to clinically important tissue types such as neurons, cardiac and skeletal muscle, hepatocytes, and pluripotent stem cells has tremendous promise in the understanding, diagnosis, and future treatment of many currently intractable human diseases (Egli, Birkhoff, and Egan, 2008; Kiskinis and Eggan, 2010). Specifically, these specialized cells can be created from a patient's own fibroblasts, providing a new window into diseases such as autism or diabetes (Park et al., 2008b; Ruder, Lu, and Collins, 2011). Direct reprogramming can also be used to enable innovative applications to regenerative medicine. In Song et al. (2012), six core myocardial transcription factors (*Gata4*, *Hand2*, *Mef2C*, *Mesp1*, *Nkx2-5*, and *Tbx5*) were used to convert mouse tail-tip fibroblasts into cardiomyocytes. These cells were then injected into mouse myocardium, and the over-expression of all factors was enforced. This led to the localized repair of cardiac scar tissue, and may lead to even more promising approaches in the future
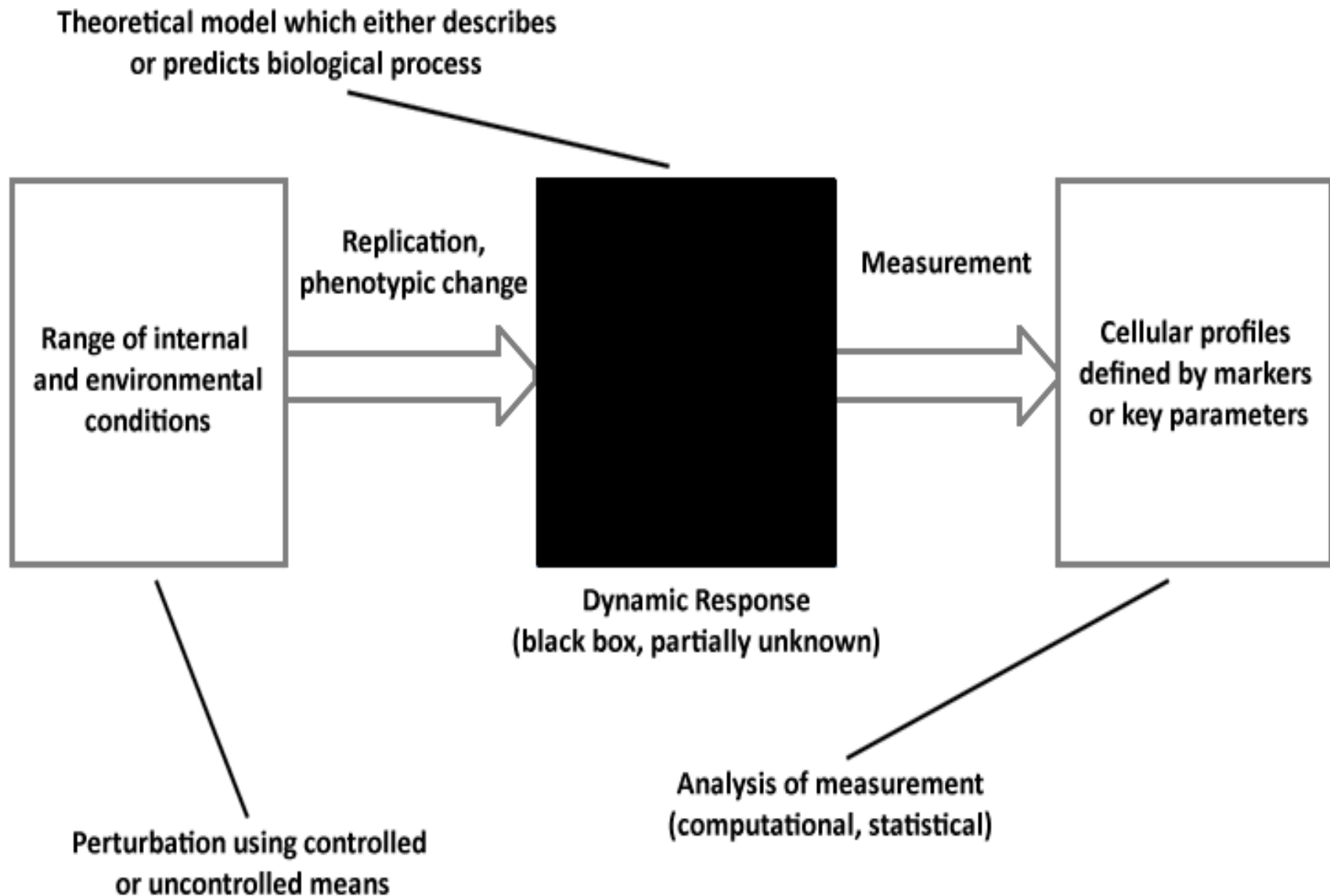
Figure 4.1. Example of how complementary cell biology experiments, modeling, and theory might be used to converge on the contents of the "black box" that defines direct cellular reprogramming.

4.1.2 Contributions of systems approach to cellular reprogramming literature. The goal of this review is to establish how modeling and systems biology approaches can advance the study of direct cellular reprogramming, and how they may ultimately improve upon the current state of the field. This includes a mix of review and introduction of allied theory and methods into the cellular reprogramming literature, and will span the existing literature on induced pluripotent stem cells (iPSCs), induced neuronal cells (iNCs), and induced muscle cells (skeletal iSMCs and cardiac iCMCs). In a recent review by Vierbuchen and Wernig (2012), the molecular mechanisms for a range of reprogramming regimens are reviewed and proposed as a series of key events in the reprogramming process. We will also do this in a way that highlights major challenges and issues across different regimens. However, unlike the Vierbuchen and Wernig review (2012), we will focus on both population-level processes and systems-level models. The first section will cover two distinct types of models relevant to the reprogramming process: phenomenological (descriptive and structural) models and predictive (probabilistic and functional) models. We will briefly touch on a number of models, including various dynamical models and a model that characterizes reprogramming as a slow kinetic process. The next section will introduce population thinking to the reprogramming community, and frame the issues of dynamics and efficiency in the context of cell populations. The final section will survey the existence and effects of variation in the reprogramming of cell populations, using the process of iCMC induction as an example. In the end, it should be apparent that theseexplanatory factors, which are novel to the field, provide an avenue to new research opportunities.
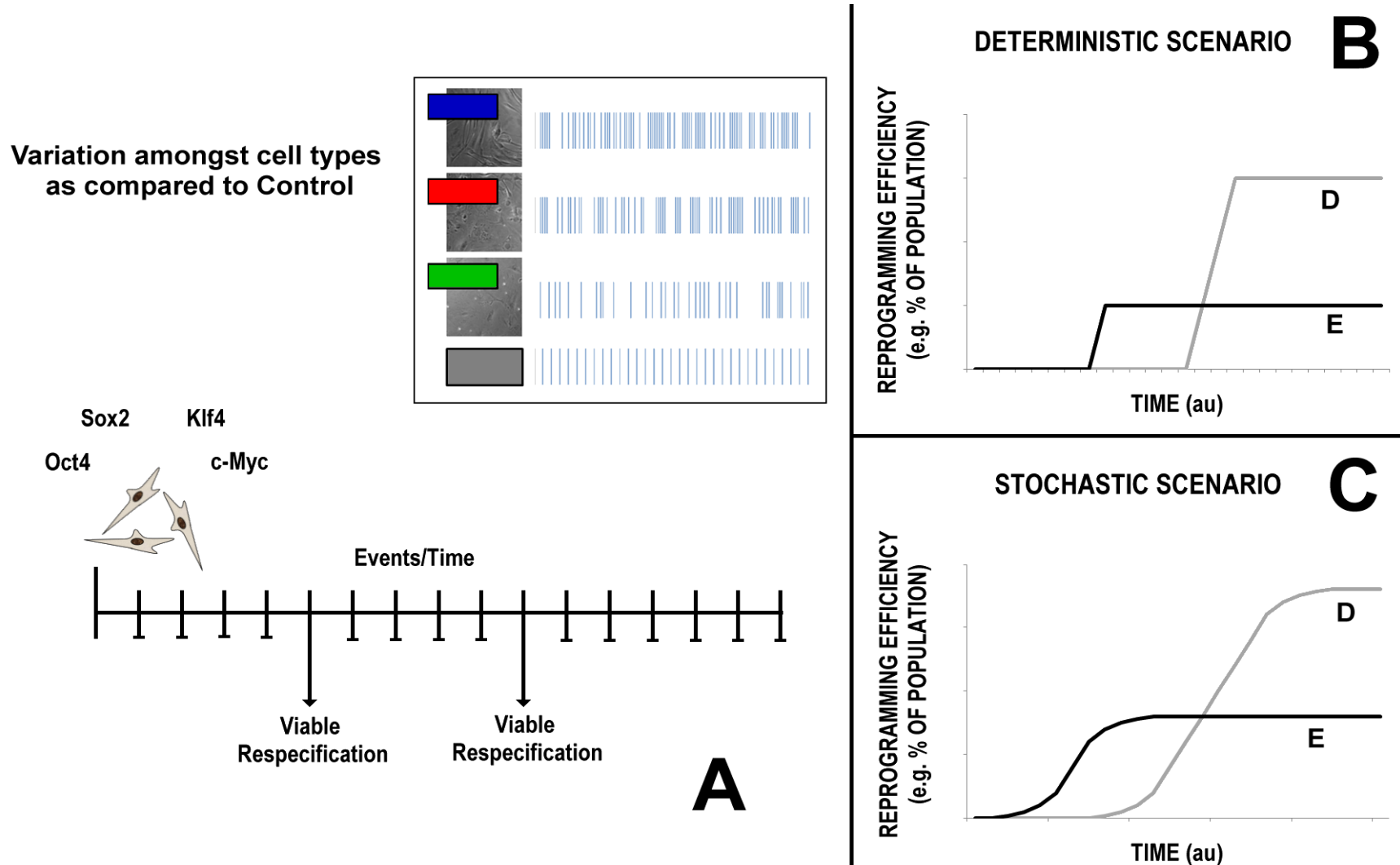
Figure 4.2. Current thinking about source of variation and the reprogramming process. A) Figure 1 taken in part from Sridharan and Plath (2008), and describes the rle of variation across biology and over time; B) and C) are the deterministic and stochastic scenarios, respectively, taken from Hanna et al. (2009), Figure 1. Legend for schematic functions in B) and C): D – democratic, E – elite.

4.2 Reprogramming in terms of models

To characterize the potential of cells to reprogram, there are two possible classes of model that can be used to simulate this biological system: predictive and phenomenological. While the relative advantages and disadvantages of each model type may depend on the input cell type, desired cell type, and even experimental context, models can be used to guide scientific practice and ultimately improve our control over the process.

4.2.1 Predictive Models. Direct reprogramming is neither highly controllable nor a highly efficient process. The diagrams shown in Figure 4.2 demonstrate current thinking about these issues. The top panel is based in part on the work of Sridharan and Plath (2008) that features a model focusing on reprogramming as an "ordered" process. In this model, variation in reprogramming outcomes is seen as a function of input cellular diversity and the relative rarity of reprogramming as an outcome when transcription factors are delivered to a cell population. Despite these sources of variation acting as perturbations, "order" in the reprogramming process can be maintained in many scenarios. The degree of orderliness in the reprogramming process is thought to have a direct influence on reprogramming efficiency. This can be compared to the model of Hanna et al. (2009), in which the outcome of direct reprogramming is constrained by both the degree of determinism (whether or not the process unfolds the same way every time) and democracy (whether or not all cells in the population have the ability to successfully reprogram) in the reprogramming process (see Figure 4.2, bottom panel).

One example of this disconnect can be found in the literature regarding general conception of stochastic processes relevant to phenotypic change. In general, existing

models for stem cell differentiation typically use rate constants and equations to characterize kinetics and cellular noise. In several instances (Tang, 2008; Wilkinson, 2008), these models focused on either a deterministic conceptualization of binding and degradation of proteins or specific pathways with discrete outputs. However, there also exist general models of chaotic noise-driven dynamics (Furusawa and Kaneko, 2009; Raj and van Oudenaarden, 2008) and population-level models (d'Inverno, Tiese, and Prophet, 2006) that are even more relevant to reprogramming. Given that our understanding of cellular reprogramming is coarse-grained with respect to more traditional models of cellular noise and kinetics, population-based models might provide alternative information (e.g. the role of natural diversity) about the reprogramming process. Viswanthan and Zandstra (2004) introduce a model of discrete dynamics for stem cell conversion *in vivo* that can be adapted to the reprogramming problem. In this model, simple rules govern the behavior of neighboring cells as they convert from stem cells to differentiated cells (Table 4.1). This results in the interaction of cellular colonies and populations over short distances and time intervals. What is interesting about this model of cellular conversion is that changes in cellular state can be triggered by a number of small-scale events. The Viswanthan and Zandstra model (2004) also makes the connection between cell cycle, cell immortality, and discrete dynamics. Collectively, these are important aspects of how reprogramming to a fully pluripotent state is an emergent phenomenon governed by a series of thresholds. Taken together, these features of real and simulated cell conversion suggest that a general theory is needed that focuses on the kinetics of the single cell, the cell culture environment, and local cellular populations.

4.2.2 Phenomenological Models. While predictive models help us select which events are most likely to occur, phenomenological models help us understand all possibilities that result from the dynamics of a biological process. One type of phenomenological approach is the phase space model (Ghosh et al., 2011). Phase spaces (or hyperspaces) are a concept originally from physics used to describe the state of *n*-dimensional systems as a surface. As phenomenological models of the reprogramming process, phase space representations are heuristic tools that help us understand reprogramming as a continuous, dynamic process.

The phase space approach has been used to model changes associated with transitions from stem cells to differentiated cells (MacArthur, Ma, and Lemischka, 2009; Cinquin and Demongeot, 2005), and bears a resemblance to the epigenetic landscapes of Waddington (Bhattacharya, Zhang, and Andersen, 2011). The advantage of the phase space approach, however, is that it treats reprogramming and the transition between cellular phenotypes as a continuous process. This allows us to trace possible paths from one cell type (e.g. fibroblast) to another (e.g. induced pluripotent cell) in the context of genotypic and/or phenotypic changes, in addition to inferring possible pathways members of a cell population might take towards an apoptotic or partially-reprogrammed state. Understanding the reprogramming process in the phase space phenomenological framework might also allow us to assess bistability (e.g. when a system is strategically poised between two stable states) in a manner not yet applied to direct cellular reprogramming.

In general, a phenomenological model is most useful in assessing the distribution of outcomes related to a single reprogramming event. The utility of phase space and

related models will be more explored at a deeper level during our discussion of population dynamics, as representing biological processes with regard to the demographics of a specific population is perhaps the greatest strength of a phenomenological approach. Related approaches such as graphs and trees might also be used to characterize the complex system of a reprogramming cell. This may assist in experimental design and ultimately enrich the interpretation of results. While these models may seem to be pure conjecture, they serve as an example of the design challenges for future experiments and therapeutic applications.

What both diagrams in Figure 4.2 show is a clear need to understand the sources of variation inherent in direct reprogramming. Whether this is possible at the level of principles or as a set of mechanisms is yet to be seen. However, direct reprogramming is a highly versatile and simple process that can be used for a variety of technical applications. While the biological mechanisms responsible for direct reprogramming may still be mostly unknown (Sridharan and Plath, 2009), there is also much potential promise for future applications.

4.2.3 Hybrid Models: Slow kinetics of reprogramming. Let us return to the concept of slow kinetics to better understand how predictive and phenomenological features of the reprogramming process might be combined in a single modeled. For this section, we will focus on conversion to iPS phenotypes, but the principles discussed here are applicable to all types of reprogramming regimen. Chan et al. (2009) demonstrates that for iPS cells there are three phases to the reprogramming process. The first is the global downregulation of somatic genes (Sridharan, 2009). The second phase involves a corresponding upregulation of pluripotent genes (Mikkelsen et al., 2008). The final

phase involves the acquisition of immortality by individual cells. As a hallmark of the pluripotent cellular state, acquiring immortality can be defined by the ability of a cell to escape senescence and upregulate genes and pathways such as *p53*, *Ink*, *Ras*, and MAP Kinase (Liu et al., 2007). Table 4.2 shows the role of various markers and their status during different stages in the process of reprogramming to an iPS phenotype.

Table 4.1. Example of a Computation Model for Cellular Differentiation: rules for cellular conversion in a small neighborhood of adjacent cells (Game of Life model). The ruleset[1] has two major components: a clock representing time, and a series of different states for each cell in an array representing part of a cell culturing substrate.

Rule class	First if condition	Operator	Second if condition	Then condition
Differentiated cells	internal clock = N_1			cell leaves its niche, reset clock to 0
	internal clock > N_1			increment internal clock by 1
Stem cells	counter at a stem cell location ≤ N_2	AND	all neighbors are also stem cells	do nothing and leave clock unchanged
	counter of a stem cell = N_2	NAND	all neighbors are stem cells	do nothing and leave clock unchanged
	counter > N_2			increment internal clock by 1
Empty spaces	counter at an empty space ≤ N_3	AND	there is a single stem cell neighbor	give birth to new stem cell and reset the clock to 0
	counter at an empty space > N_3	AND	there is a single stem cell neighbor	increment internal clock by 1
	no stem-cell neighbors			reset clock to 0

[1] The counter consists of three constants: N_1, N_2, and N_3. N_1 represents the time taken for a differentiated cell to leave a niche, N_2 represents the cycling phase of a cell, and N_3 represents the amount of time for empty spaces to be populated by a descendent cell.

It has also been shown that the state of a reprogrammed cell observed at any one timepoint is the product of large number of undefined events that can be

approximated by the measurement of epigenetic, immunohistochemical, and gene expression phenomena (Chan et al., 2009; Liu et al., 2007; Wadhwa, Kaul, and Mitsui, 1999; Karplus, 1999). In the literature, these are characterized as discrete barriers, whether they involve changes in the expression of single genes (Hong et al., 2009), or specific epigenetic changes (Maherali et al., 2007; Pasini et al., 2010). The morphology of induced pluripotent cell colonies can likewise be classified into three categories which may correspond to the proposed phases of reprogramming (Marson et al., 2008). In addition, there are mechanisms that are global rather than phasic. For example, in Maherali et al. (2007), it is suggested that the expression of Nanog alters reprogramming kinetics in a way that is independent of proliferation rate. The existence of both global and phasic responses to the reprogramming stimulus may ultimately suggest both bottom-up and top-down mechanisms that are essential for regulating the speed and extent of pluripotency acquisition.

One example of top-down mechanisms that regulate pluripotency acquisition involves epigenetic factors. For example, the current debate on stemness (Lander, 2009) can inform the phenomenological model by providing a model concerning the presence and timing of bivalent chromatin during the acquisition of pluripotency. In general, this epigenetic mechanism confers a dynamic stability to stem cells. Contemporary findings and thinking suggest that promoters for highly conserved genes across the genome are enriched for both active (e.g. H3K4) and repressive (e.g. H3K27) transcriptional states (Spivakov and Fisher, 2007). In the reprogramming process, this has also been observed by Sridharan et al. (2009) and Liu et al. (2007) as

a weak binding affinity for both H3K4 and H3K27 during the middle phase of iPS reprogramming.

Table 4.2. Various markers and mechanisms in the three-step kinetics of cellular reprogramming for cells converted to pluripotency (iPS phenotype).

Indicator	Stage I (< 6d)	Stage II (6-12d)	Stage II (> 12d)
Viral vector	Must be active	Can be inactive	Can be inactive
miR-145	Strongly Expressed	Expressed or weakly expressed	Weakly expressed or absent
HeMeK	Strong Binding	Weak Binding	Strong binding
HeMeK27	Lack of binding	Weak Binding	Strong binding
GDF3	Lack of expression	Weak expression	Strong expression
Nanog	Co-expresssion	Lack of expression	Co-expression
TRA-1-60	Weak expression	Expression	Expression
PRC complexes	PRC II associates with JARID	PRC II activates pluripotentcy-associated functions	Pluripotency maintained

Another key event in conversion to iPS phenotypes is the acquisition of immortality. This milestone seems to be a key event in the transition to a fully pluripotent phenotype. Properties of the fully pluripotent phenotype are collectively referred to in the literature as stemness (Mikkers and Frisen, 2005; Orford and Scadden, 2008; Gokhale and Andrews, 2008). Stemness has traditionally been defined in stem cells in terms of what makes stem cells functionally distinct from differentiated cells. Given the great diversity of stem cells, there are but a few overarching features of pluripotency. In Mikkers and Frisen (2005), stemness is defined as the stable suspension of cells in a specific developmental stage where cells do not leave the cell cycle. The stage is

marked by expression and upregulation of *Oct4* and Nanog (Mikkers and Frisen, 2005). In Orford and Scadden (2008), stemness is defined as the continuation of the cell cycle. This is caused by the downregulation of *Rb* family proteins, and results in the global downregulation of cell specific genes. In at least two studies (Mikkers and Firsen, 2005; Gokhale and Andrews, 2008), stemness is defined as by cells within a definable niche. Particularly, Mikkers and Frisen (2005) propose that interaction with the local environment is critical for the maintenance of stemness, which is a critical feature of the pluripotent cellular state. The acquisition of immortality is not a requirement of conversion to iN or iSM/iCM phenotypes. Whether the acquisition of coherent action potential generation serves as a comparable milestone for these reprogramming regimes is an interesting question that deserves further investigation.

4.3 Population dynamics

The study of reprogramming kinetics, which has focused exclusively on conversion to iPS phenotypes, is a property of individual cells. While up to this point the study of cellular reprogramming has largely focused on the aggregate potential of certain cell types, the overall reprogramming efficiency and the relative survivability of a cell population are also important factors in determining how direct cellular reprogramming unfolds. While reprogramming efficiency and survivability can be assessed in a number of ways, it may also be that they contribute to highly informative cellular population dynamics. Compared to biochemical kinetics, population dynamics are a bit slower but may ultimately have almost as much influence on the reprogramming process.

4.3.1 Interaction models. It is imperative that any good model of reprogramming should involve a characterization of population dynamics. In the case of both cell culture and *in vivo* models of reprogramming, cell populations undergo a number of processes including demographic stochasticity, environmental selection, and effects related to initial population size (Nisbet and Gurney, 1982). If intrinsic properties of various cell populations are responsible for differences in reprogramming efficiency, they may be augmented by these processes. Of course, without extensive characterization, it is hard to establish a link between differences in reprogramming efficiency and a population's intrinsic properties.

Aside from understanding biological variation as a by-product of population process, we can also examine population dynamics using experimentally-tractable models. This will allow us to investigate the variation among cell populations in terms of their response to environmental challenges (Angka, Geddes, and Kablar, 2008). This can be accomplished by putting each cell population under growth and survival conditions independently. These populations are then counted to examine their demographics over several days. These results can be put together using mathematical modeling techniques, and will allow us to examine phenomena such as cell death and cell survivability (Palmer and Feldman, 2011), which are factors that indirectly influence differences in reprogramming efficiency.

To characterize growth and survival as an integrated process, we can use an additive population model of birth and death processes (Abrams, 2000). This results in a net growth rate measurement which characterizes the overall expansion of a given population in the first few days of infection. While this is not a direct measurement, it

does serve as a model that allows us to establish a population-specific baseline for cellular survivability. This may yield informativecorrelations with reprogramming efficiency and/or cell infection rate measured for these same populations, but in a different context. The application to cell reprogramming would involve coupling the rate of birth (e.g. cell division) with the rate of death (e.g. apoptosis, toxicity) in a single, idealized population. The output of such a model would demonstrate how population size fluctuates, particularly during the critical periods of the reprogramming process.

4.3.2 Dynamical models. Another way to capture the population dynamics of a reprogramming cell population is to use a dynamical model. Two examples of dynamical models are shown in Figure 4.3. The top of Figure 4.3 shows an example of a phase space representation discussed earlier (Bernew and Straubb, 1997). While meant to model all relevant dimensions of a complex system, phase space models are typically reduced to three-dimensional (*x, y, z*) surfaces for visualization purposes. As shown in the cartoon, the maxima and minima (movement along the *z* axis) of the phase space surface represents stable phenotypic states (e.g. neuron, fibroblast, iPS). Each cell in the population must move (in a metaphoric sense) towards these stable states (movement along the *x* and *y* axes) over time. The extent to which these cells cluster at or near these extreme values determines the relative rate of efficiency and extent of reprogramming.

A similar approach is the bistablity model, shown in cartoon form on the bottom of Figure 4.3 (Kelso, 2008). The bistability model is a two-dimensional model that approximates bistable phenomena. Unlike the phase space model, bistability models use a single dimension representing cellular state and another dimension representing

a potential for change. When the cell is in a stable state (center) it sits in the middle of two wells, and movement of a cell or population into one of these wells is characterized by its potential (y-axis). Originally applied to research ranging from bimanual performance (human cognition) to symmetry breaking (physics), bistability models might be well-suited to characterizing bistable phenomena in the reprogramming process such as histone binding and partial reprogramming.

4.4 Stochastic threshold model

It could be argued that direct cellular reprogramming is a critical process during which many cumulative changes trigger large-scale changes at random and uneven intervals (due to a stochastic process), much like what happens during an avalanche. This type of behavior has been observed in physical and neuronal systems alike (Bak, Tang, and Weisenfeld, 1987; Beggs and Plenz, 2003), and is a classic statistical signature of self-organization. The stochastic threshold model (STM) is useful for understanding these changes as they occur across a population of cells. Rather than independently monitoring the reprogramming process in every cell, we can use population-wide probability distributions to better understand the commonalities and rarities in reprogramming behavior among cells in their specific population.

There are four parameters (Table 4.3) that describe distributions and features in the stochastic threshold model (STM). The first of these is X_T, which describes a threshold value along the x-axis (e.g. a value for the proportion of cells in culture). The second of these is Y_T, which describes a threshold value along the y-axis (e.g. a value for the proportion of cells positive for one or more biochemical indicators).

Phase Space Model for Phenotypic Conversion
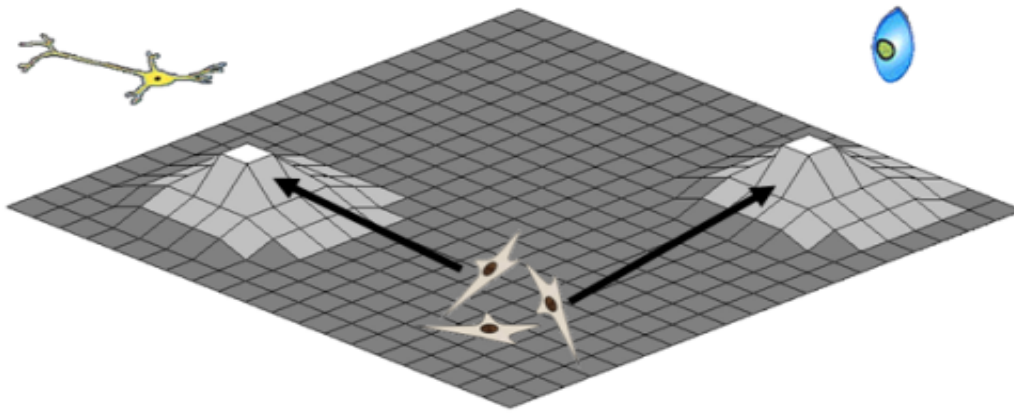
Bistability Model for Phenotypic Conversion
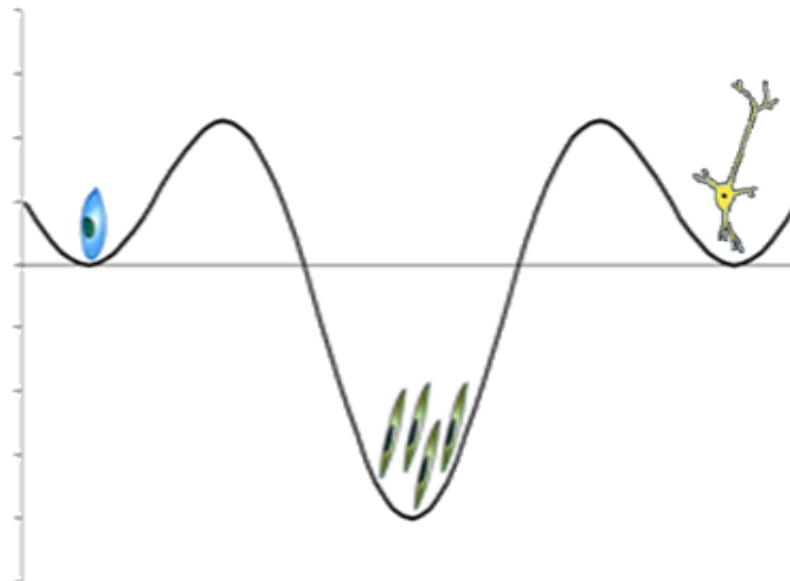
Figure 4.3. An example of a direct reprogramming hyperspace with stable states for iPS, iN, and fibroblast cells. LEFT: an exemplified *n*-dimensional hyperspace. RIGHT: an exemplified two-dimensional hyperspace demonstrating the potential bistability of cellular state (see Kelso, 2008 for HKB model example).

Taken together, these thresholds define a bounding box that enclose different fractions of our cell population representing fully reprogrammed cells, partially reprogrammed cells, and non-affected cells. These are defined by A, B, and C, respectively (see Figure 4.4). The other two parameters (*a* and *α*) are directly related to the statistical signature of the data. The distribution of reprogramming efficiency in a population should fit a power function, as shown in Figure 4.4. A power law signature direct reprogramming-related population dynamics is a non-uniform process characterized by a statistical signature in which only a few fully reprogrammed cells exist against a background population of non-converted cells (Reed and Hughes, 2002). Parameters *a* and *α* describe the slope and exponent (e.g. shape) of this curve, respectively (see Table 4.3).

Table 4.3. List of parameters that describe changes in the STM.

Parameter	Definition	Significance
X_T	Proportion of cells in culture of a certain state	Define transitions between fractions of the population
Y_T	Fraction of indicator-positive cells in population	Define transitions between fractions of the population
a	Shape of curve	Slope component of the power function model for the data
α	Shape of curve	Exponent of the power function model for the data

4.4.1 Scenarios under the STM. In an ideal situation, a small proportion of cells in the population should express the hallmarks of a reprogrammed cell as completely as possible. This includes phenotypic conversion in addition to the expression of all key biochemical indicators (category A). A larger proportion of cells (category B) feature partially-reprogrammed cells, which exhibit either cryptic phenotypes or only a subset of key biochemical indicators. A comparable or even larger proportion of cells define category C, and do not respond to the reprogramming stimulus, and do not express any of the hallmarks of a reprogrammed cell. Changes in the empirical distribution correspondingly contribute to changes in the X_T and Y_T parameter values.

For a given observation, the fewer the number of cells reprogrammed, the greater number of regulatory events must occur inside each cell before conversion is assured. While this threshold is variable for each cell in the population, the STM describes the mean behavior across all cells in a way that captures the non-uniform responses at the population level. While the STM can account for the state of a cellular population at different points in the reprogramming process, it does not explicitly account for cells that die during the course of reprogramming. Nor does this model directly account for the underlying genomic mechanisms. However, the STM is a good heuristic approximation for comparing how the outcomes of the reprogramming process are distributed between input cell lines and different target phenotypes.

4.5 Role of variability in cellular reprogramming

There are a number of issues that constrain the use of direct reprogramming as a biotechnological tool. These include both the phenotypic stability of a converted cell and the overall efficiency of viral transduction. Phenotypic stability can be assessed in many

ways, but the most common is the retention of a converted state after the removal of enhanced transcription factor activity. In the case of iPS cells, over-expression of the four factors is turned off after a period of time. The ability of fledgling iPS cell to retain a gene expression pattern typical of a pluripotent cell is an indicator of how complete the reprogramming process has transformed the cell (Duinsbergen et al., 2008). Since the reprogramming process is not uniform in every cell, many of cells initially infected by the retroviral vector either apoptose, differentiate, or remain in a partial-iPS state (Chan et al., 2009). A similar situation likely exists during conversion to any given cellular phenotype, as non-uniformity of process may result from combinatorial activity of the reprogramming factors on the cell's biochemistry. This can be problematic for applications such as cell therapy, as it can trigger inflammatory or carcinogenic responses. Considering variability from a systems perspective also enables us to understand both the potential causes and effects of variability in context. These include overall reprogramming efficiency, the combinatorial action of transcription factors, and the replication of experimental results.

4.5.1 Variability and efficiency. One outstanding problem in the reprogramming literature, and an issue that is characterized but not explained by the STM, is the high degree of variability in terms of efficiency. A comprehensive understanding of the mechanisms controlling reprogramming efficiency is compounded by two issues. The first is the way in which efficiency is measured. In most studies (Wernig et al., 2008; Stadtfeld et al., 2008), the efficiency is estimated from the number of colonies derived from an input cell population. More sophisticated methods (such as FACS sorting) can also be used to obtain the number of cells before infection and after mature colonies

appear. While these techniques serve their purpose as first-pass approximations, they do not account for variation in colony composition nor the population dynamics (e.g. growth and death rates) of cells from first infection to final count. This likely contributes to a significant amount of measurement noise within and between different cell types. The second component is appropriately characterizing the variability that *does* exist in the reprogramming process. One of the problems with current studies on reprogramming is that they do not provide estimates of variability between experiments, even replicates performed on clonal lines. What is needed is a series of estimators for expected variability within and between cell lines.

Due to the inherent complexity of the reprogramming process, there exists large variability in terms of reprogramming efficiency. Efficiency is generally defined as the fraction of input cells successfully converted to the desired phenotype. In a survey of the literature on efficiencies for converting somatic cells to iPS cells, Artyomov, Meissner, and Chakraborty (2010) report that the efficiency rate can vary from 0.001% to 29%. While the highest possible efficiency is most desirable, it is often the case that efficiency is highly variable across cell types and instances of reprogramming (e.g. independent infections). The underpinnings of this variation are largely unknown, and may pose a significant problem for the further development of this technology. Assessing efficiency is not simply a matter of better measurement and approximation, however. In the case of conversion to iPS cells, the resulting pluripotent cells are not created equal. Successfully converted cells can be further fractionated by their degree of conversion. So-called partial-iPS cells (Chan et al., 2009; Sridharan et al., 2009) are known to exist in iPS cell cultures, and demonstrate a high degree of phenotypic diversity. However,

partial-iPS cells are generally identified by their inability to express the reprogramming transcription factors independent of their transgene (Sridharan et al., 2009). Bearing a superficially pluripotent phenotype, however, they represent a transitional state which is neither truly pluripiotent nor unaffected by the reprogramming process. While it has not been demonstrated, there are likely such "partial" states in induced neuron (iN), muscle fiber (iSM), and cardiomyocyte (iCM) cultures as well. Such partially-converted states, if properly characterized, might also allow for statements about reprogramming process efficiency.

4.5.2 Combinatorial action of transcription factors and experimental replication. As a general phenomenon, cellular reprogramming is a repeatable process, even though overall efficiencies from experiment to experiment are highly variable. What might account for this paradox? By taking a systems-level view of cellular reprogramming, we can see that combinatorial transcription factor activity and experimental conditions between labs and context may play a bigger role than previous studies indicate. The optimal transcription factor combination for a desired cell type is generally selected using a series of screening experiments. For an example from cardiomyocytes, see Srivastava and Ieda (2012), although Chang et al. (2011) has worked out an algorithmic approach. Selective siRNA knockdown of downstream genes (Hong et al., 2009), selective inducement of delivered factors (Maherali et al., 2008; Markoulaki et al., 2009), and the addition of supplemental factors (Qian et al., 2012; Srivastava and Ieda, 2012), have also been used to improve efficiency. This is consistent with phenomenological model-based computational experiments looking at the reprogramming of fibroblasts to pluripotent cell fates have demonstrated that there are many potential pathways to the

successful induction (Artyomov, Meissner, and Chakraborty, 2010). It should also be

kept in mind that transcription factors can act synergistically or in combinatorial fashion

to result in a number of outcomes that are not easy to predict *a priori* (Loh and Lim,

2011).
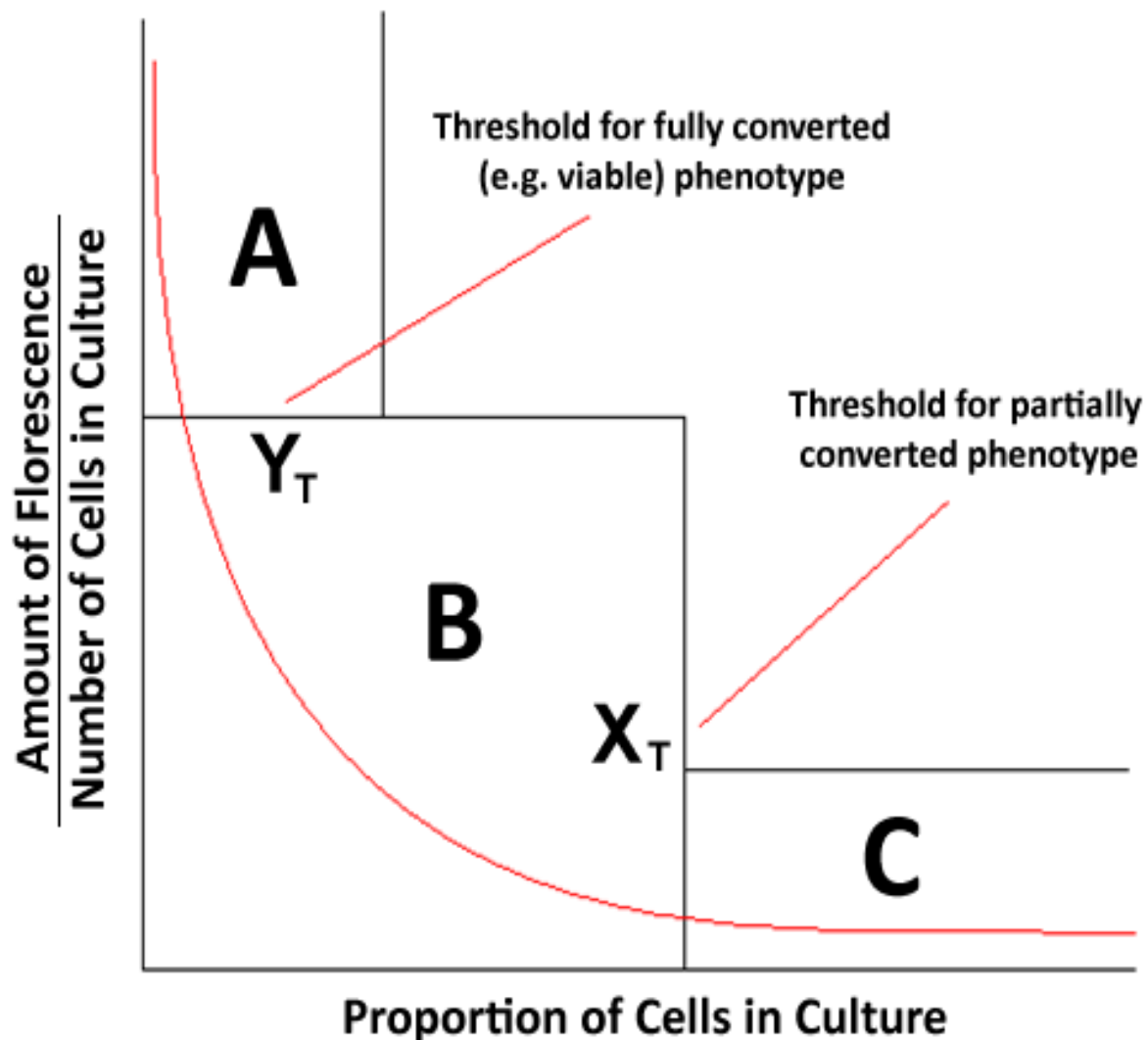
Figure 4.4. Schematic showing the details of the stochastic threshold model as it captures the dynamics of a reprogramming cell culture. Example shows hypothetical cell population being converted to iPS phenotype. Categories are components of a population's probability distribution, and are as follows: A) cells that sre fully converted during the reprogramming process, share a stable and hard-to-reverse phenotype, B) cells that are partially converted during the reprogramming process, sharing only some traits with category A, C) cells that are exposed to reprogramming stimulus but fail to convert to a new phenotype.

Another outstanding problem is the lack of experimental replication across experiments and laboratories. From a philosophy of science standpoint (Kuhn, 1962), the inability to replicate results is considered a violation of scientific method. There have been several recent papers citing the problems with reproducing well-known results in Psychology (Yong, 2012) and Medical Science (Ioannidis, 2005). In Yong (2012), interviews with a number of scientists revealed that most prestigious journals focus on positive results only. This may lead to articles that present parlor tricks, or one-time occurrences in which the experiment happened to work as expected (Yong, 2012). In light of this, perhaps an unfair standard (e.g. using the benchmark reproducible results as a good science standard) obfuscates our ability to truly understand the natural phenomenon (Buganim et al., 2012). In the case of cellular reprogramming in particular, perhaps replication should not be the primary goal. Instead, we should look at higher-order metrics (e.g. the variance structure over multiple replicates) rather than average outcomes. This might involve adapting principle component analysis (PCA) or a statistical learning technique in a problem-specific manner.

But why would this be preferable in terms of interpreting results? To understand this, recall that when we convert a cell from one phenotype to another, a few transcription factors trigger many changes in the cell that lead to a change in state. Therefore, it should not be assumed that this process always utilizes the same set of mechanisms. At least three papers (Artyomov, Meissner, and Chakraborty, 2010; Loh and Lim, 2011; Lee, 2011) have proposed that reprogramming is a combinatorial process triggered by initiating factors and having a mosaic effect on the cell's biochemical milieu. This suggests that there are multiple pathways for the successful

conversion of a phenotype, and that any endpoint measurement of this process will yield a multimodal statistical distribution. This outcome is not adequately characterized by either traditional clustering techniques or traditional parametric statistical analyses.

4.5.3 Effect of variability on desired outcomes. Since cellular reprogramming is in part a bioengineering technique, we must also consider cellular reprogramming as a series of design principles. To achieve this, we can leverage the potential cause and effects featured in previous sections. Design principles characterize the nature of a system for purposes of good engineering design (e.g. functionality, reliability, etc.). In another context, I have argued that the design principles for a "living" system must be fundamentally different than those for a mechanical or electrical system (Alicea, 2009). This is because living (e.g. biological) systems exhibit much more variation than non-living systems. Therefore, the design principles that govern direct cellular reprogramming should not focus on standardization and optimization, but rather on things such as customization and designing to take advantage of a living system's inherent variability. This might be especially helpful when using direct reprogramming to produce converted cells from disease- or patient-specific cell lines (Park et al., 2008a; Patel and Yang, 2010; Kiskinis and Eggan, 2010). One might imagine that cells could be classified not only by their tissue of origin and phenotypic characteristics, but also by their genotype or other biochemical profiles. Such approaches (e.g. personal pharmacogenomics) are already becoming commonplace in other areas of biological research, and should be a part of direct cellular reprogramming as well (Feala et al., 2010).

4.6 Variability example: enduced cardiomyocyte cells

We will now focus on specific methods and results from the cardiomyocyte literature. In this case, we can show that biological variation may be related to technical variation, but that overall significant variability exists even within a specific target phenotype. The first report of successful cardiomyocyte induction is from Ieda et al. (2010). This study is built around the observation that, unlike in the case of skeletal muscle fibers (*MyoD*) or pluripotency (*Oct4*), there is no "master regulator" transcription factor for cardiac muscle (Davis, Weintraub, and Lassar, 1987; Takahashi et al., 2007). To conduct their screening of appropriate factors, key developmental cardiac regulators were used in a reprogramming context. Fourteen (14) candidate genes were identified using a microarray of cardiomyocytes and fibroblasts, and were delivered individually and collectively into cultured cells. Ultimately, a cocktail containing *Gata4*, *Mef2c*, and *Tbx5* (GMT) was used to reprogram a fibroblast population. This resulted in successful reprogramming to a functional cardiomyocyte phenotype in a manner similar to other target phenotypes (e.g. iPS).

While the GMT factors provide the potential for reprogramming, it still may not be the best possible method. On the one hand, the reprogramming potential of GMT is not simply an artifact of the Ieda et al. (2010) methodology. Qian et al. (2012) found that *in vivo* delivery of GMT cocktail decreases infarct size and convert cells to a phenotype that fire action potentials, beat when stimulated, and exhibit cardiomyocyte-like gene expression. Moreover, the forced expression of four factors (*Gata4*, *Hand2*, *Mef2C* and *Tbx5* -- GMTH) in non-cardiomyocytes *in vivo* in mice results in functional cardiomyocyte-like cells and overall improves the rate of regeneration after injury (Song

et al., 2012). A cocktail containing all 14 factors resulted in 1.7% of cells providing a GFP+ signal (Ieda et al., 2010). Ultimately, the combination of *Gata4*, *Mef2c*, *Mesp1*, and *Tbx5* increased reporting activity of alpha-MHC to over 20%. The combination of *Gata4*, *Mef2c*, and *Tbx5* (GMT) were considered "sufficient", as adding *Tbx5* did not improve efficiency. As will be demonstrated in the next section, just because a given factor does not immediately add to the overall efficiency of the reprogramming process does not mean it is dispensable.

4.6.1 Cause and effect of variation in context. In a previous section, we proposed that the combinatorial action of transcription factors and the replication of results contribute to variability. This seems to hold true in the case of iCM cells, as there are a number of factors mentioned in the existing literature that could account for a lack of experimental replication. For example, the Ieda et al. (2010) study used the most common cell type in the heart (cardiac fibroblasts), but do not account for their diversity. A review by Srivastava and Ieda (2012) provides a good indicator of how important accounting for variation is to three studies using the GMT and GMHT cocktails. In Ieda et al. (2010), GMT factor expression was roughly 6- to 8-fold greater than in uninfected cardiomyocytes, while in Chen et al. (2012), GMT expression was 18- to 1000-fold greater than in uninfected fibroblasts. Perhaps the forced overexpression of transcription factors only plays a tangential role in successfully converting cells to the target phenotype, and perhaps there is variablity across contexts. For example, the approach of Ieda et al. (2010) yielded 15% of cells with cardiomyocyte-like expression patterns, but only 0.5% of cells became fully reprogrammed with the functional

attributes of cardiomyocytes. By contrast, the Chen et al. (2012) study reports difficulty in producing any cells that are fully reprogrammed.

From these induced cardiomycyte examples, we can clearly see that a host of mechanisms influence reprogramming efficiency. While some of these can be written off as technical variation, others are likely to be related to natural variation. One factor that might account for discrepancies across experiments is microenvironment. In Ieda et al. (2010) it is suggested that cardiac fibroblasts might reprogram more readily *in vivo*, where the presence of a cardiac microenvironment could reinforce the full conversion of otherwise partially-converted cells. But findings such as this consistently lead to the assumption that once we control for all known sources of variation, direct cellular reprogramming should become routine and highly precise (Srivastava and Ieda, 2012). We propose that this is not necessarily the case. We will now turn more explicitly to the potential of natural variation to influence the outcomes of direct cellular reprogramming.

4.7 Role of natural variability in cellular reprogramming

Since much of the modeling and population-level experimental work focuses on changes in state and overall efficiency, we must inquire as to the underlying mechanisms of direct cellular reprogramming. When our experiments involve a single cellular population reprogramming to a single target phenotype (e.g. iPS), the discussion is limited to sets of pathways and candidate biochemical markers. However, when we examine multiple cell populations reprogramming to multiple target phenotypes (e.g. iN and iSM), a set of mechanisms related to natural variation (e.g. genomic variants and innate differences between cell lines) must be considered.

4.7.1 Inherent preferences for reprogramming. Perhaps even more informative are experiments that demonstrate clear preferences of various cell lines for reprogramming to a single target phenotype (Alicea et al., 2013; see also Chapter 2). This preference has been demonstrated to be a property of fibroblasts populations with neither progenitor cell influence nor cells which are primed for gene expression reflecting the destination phenotype. This phenomenon can be shown in a contingency table (Table 4.4), in which all possible outcomes are featured for a set of populations exposed to both the muscle and neuronal reprogramming factors.

Table 4.4. All possible outcomes for experiments reprogramming cells to induced neuron (iN) and induced skeletal muscle fiber (iSM) phenotypes. (+) is equivalent to above-average reprogramming efficiency, while (-) is equivalent to below-average reprogramming efficiency.

	iN phenotype (+)	**iN phenotype (-)**
iSM phenotype (+)	Generalized plasticity	iSM Bias
iSM phenotype (-)	iN Bias	Active suppression/buffering

In this hypothetical set of experiments, cell lines that exhibit high reprogramming efficiency for both iSM and iN phenotypes are said to exhibit a generalized plasticity mechanism. This may involve changes in cell cycle timing or other changes to generalized cellular mechanisms (Egli, Birkhoff, and Eggan, 2008; Cox and Rizzinio, 2010). When a cellular population exhibits low reprogramming efficiency for both iSM and iN phenotypes (the diagonal cells in Table 4.4), this suggests active suppression of phenotypic change related to developmental buffering mechanisms (Chipev and Simon,

2002; Blasi et al., 2011). However, when a cell population exhibits a high reprogramming efficiency for one phenotype (e.g. iN) but not another (e.g. iSM), it is proposed that these cells exhibit reprogramming bias.

4.7.2 Potential driving mechanisms of reprogramming bias. The driving mechanism behind reprogramming bias is a preference towards changes specific to the genetic regulatory network of a given type. When a cell is exposed to the reprogramming factors, there is a reordering of the cell biochemistry that allows for major phenotypic changes to occur. Reprogramming bias is simply a functional directionality in these changes. For example, iSM bias will involve a preference for changes specific to the skeletal muscle regulatory network (Bismuth and Relaix, 2010). By contrast, iN bias will involve changes to specific types (e.g. glutamitergic) of neuronal regulatory network (Hobert, 2008). In terms of cellular populations, biased cells are non-progenitor cells that favor one specific phenotype (e.g. iSM) over another (e.g. iN). If enough individual cells in the population meet this criterion, the population can be said to exhibit strong bias. Previous studies suggest that determination of this bias may be highly centralized, as Nanog fluctuations may control this bias in iPS and other stem-like cells (Kalmar et al., 2009).

4.7.3 Potential effects of reprogramming bias. Reprogramming bias can also be understood in conjunction with the stochastic threshold model (STM) shown in Figure 4.4. At the population level, the thresholds and distribution function will differ for the same input cell population reprogrammed to different target phenotypes (Figure 4.5, left). In terms of a regulatory-specific process, the introduction of cell type-specific transcription factors will trigger an immediate response in a core set of genes, or a first-

order genetic regulatory network. This may correspond to genes most strongly associated with a specific phenotype. Changes in this network will trigger subsequent changes in a secondary set of genes, or a second-order genetic regulatory network. This may include genes that support but do not determine the target phenotype. If bias is demonstrated by the final phenotype, coordinated changes (e.g. changes that are significantly different from random occurrences) that enable the phenotype in question will characterize both networks.

Figure 4.5 also demonstrates that as a cell line exhibits bias (in this case, for an iSM phenotype), two features of the population emerge. One is that a greater number of cells both within and between replicates exhibit a higher reprogramming efficiency for the target phenotype under bias. Secondly, a greater number of cells both within and between replicates also exhibit partially reprogrammed phenotypes. While we are not isolating and identifying specific mechanisms for this bias, we are proposing that natural variation (rather than technical) may also contribute to the myriad results seen in the reprogramming literature, particularly when reprogramming to different target phenotypes.

4.8 Conclusions

To understand the variation and replication problems of the reprogramming process in context, Hanna et al. (2009) have proposed a model of latency and efficiency (Yamanaka, 2009). In terms of latency, all cells in a population can either reprogram at the same rate (deterministically) or at variable rates (stochastically). In terms of efficiency, either all cells in a population can reprogram given the right conditions, or only a few cells (e.g. an elite population) have this ability. As this model (shown in

108

Figure 4.2, right) features a continuum based on timing of the reprogramming event, it suggests that there is some set of mechanisms focused on the regulation and perturbation of cell cycle (Ding and Wang, 2011). Could a mechanism of this kind be variable across cell lines or even reprogramming contexts, and could this in turn account for the high degree of variation observed in efficiency across experiments and cell lines? A review by Kitada, Wakao, and Dezawa (2012) critically assesses the Hanna et al. model in the context of generating iPS cells from Muse and non-Muse cell lines. Their work suggests that the source of cells and other population-level properties determine whether or not reprogramming can be observed as being dominated by stochastic and elitist outcomes.

Another possibility is that the reprogramming ability of cells is affected by the physical environment. Cells in a population may become spatially restricted during the reprogramming process, by using a series of intracellular signals that resembles the process of pattern morphogenesis during development. In this case, all cells would have the same capacity to reprogram, but local microenvironmental restrictions would prevent most cells from becoming fully reprogrammed. Many of these candidate mechanisms involve a fair amount of speculation, so more models such as the one presented by Hanna et al. (2009) are needed in the field. Future developments in this direction may provide us with as-of-yet unknown dimensions to the reprogramming process.
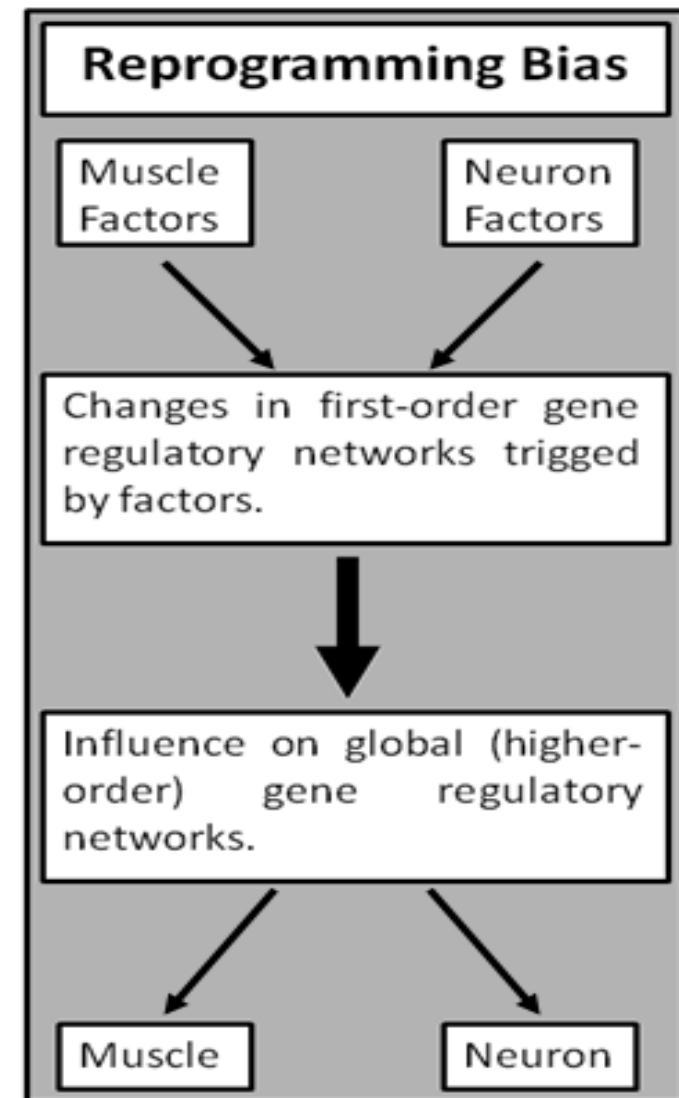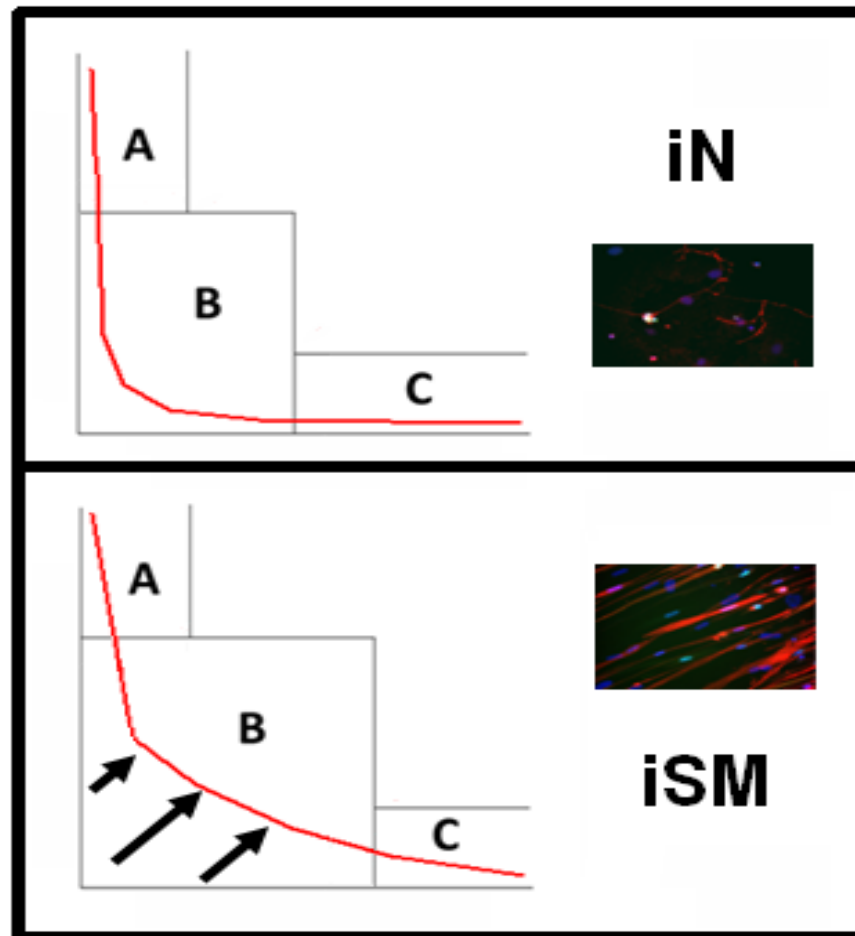
Figure 4.5. Reprogramming bias as a population-level effect. TOP: shape of function changes (shown by arrows) with bias for iSM as compared to iN, averaged across population (adapted from Figure 4.4). BOTTOM: proposed process during reprogramming that may potentially result in bias.

One theme that has been hinted at in the literature but not formally proposed is that cellular reprogramming is an emergent property of a cell's biochemistry. This refers to the coordination of changes associated with phenotypic conversion. These changes are either coordinated in time, or as a sequence of events (Subramanyam and Blelloch, 2009). In any case, the degree of change in phenotype is often greater than the sum of all changes responsible. Such a multiplicative response is suggested by Levin (2009) as playing a key role in the process of bodily regeneration. Yet this multiplicative response could also be a product of highly-complex regulatory mechanisms. Judson et al. (2009) found that adding a combination of miRNA molecules [91] to three iPS reprogramming factors (*Oct4*, *Sox2*, and *Klf4*) improved efficiency over both OKSM cocktail alone or the OKSM cocktail plus the same miRNA combination. The key outcome of this result involves the ability of *c-Myc* to negate the regulatory enhancement that miRNAs (particularly miR-145, see Appasani, 2008 for more information) provide to the *Oct4*, *Sox2*, *Klf4* pathway. This idea is followed up on in a paper by Shenoy and Blelloch (2012): in this case, it is suggested that the conversion to new cell fates may be driven by a double suppression mechanism. In this case, the delivered miRNAs suppress a repressor, which in turn activates a new cellular phenotype (Xu et al., 2009).

When we reprogram a population of cells, we are in reality perturbing a black box mechanism. Some ideas of what constitutes the direct cellular reprogramming black box are featured in Sridharan and Plath (2008). According to these authors, direct cellular reprogramming is described as an ordered yet inefficient process. Inefficiencies observed among fibroblast, liver, and stomach source cells as they are reprogrammed to an iPS state can be defined as differences in transcriptional regulation, epigenetic

regulation, and viral integration site location. Conversion of intermediate cell types to an iPS phenotype proceeds at low efficiency (Shenoy and Blelloch, 2012). So how are these inefficiencies to be overcome? Concrete solutions in form of a more traditional protocol (e.g. drug treatments of large effect or small molecule treatments) are somewhat useless: we are nowhere near being able to do this in any meaningful way, and hopefully this review has demonstrated why such a worldview will not produce what is intended. In the end, the direct reprogramming community must recognize that the black box mechanism is actually a problem best solved by accounting for multiple scales simulatenously. This involves an accounting of both complex phenotypic changes at the level of cellular phenotype and many individual changes at the level of genotype. Therefore, it is being proposed that future research in the area of cellular reprogramming include building models of the reprogramming process that account for diversity, using the appropriate statistical techniques (including customizing techniques used in other fields), and understanding the regulatory mechanisms that underlie the changes being initiated.

CHAPTER 5: CONCLUSIONS AND FUTURE DIRECTIONS

5.0 Introduction

In this dissertation, I have presented a broad overview of direct cellular reprogramming as applicable to the nervous system. This includes an analysis of the role of diversity in direct reprogramming (Chapter 2), insights into mechanism through comparison of direct and indirect reprogramming (Chapter 3), and how iN and iSM cells fit into the cellular reprogramming literature (Chapter 4). For the work presented in Chapter 2, an extended set of results will also be presented with details featured in Appendices C and D.

5.1 Review of Results

While each chapter provides its own insight into the reprogramming process, there are some overarching themes that provide directions for the further development of experimental investigations and approaches. Chapter 2 provides a purely experimental investigation using primary data. The comparison between iN and iSM reprogramming set up Chapter 3, which allows us to explore potential differences that emerge when alternate approaches to reprogramming (e.g. iPS and SCNT-ES) are used to reach the same phenotype (e.g. pluripotent cell). The coupling of computational analytical methods and secondary, high-throughput data set up Chapter 4, which frames all of these results and related insights into a common theoretical framework.

5.1.1 Role of diversity in cellular reprogramming. While Chapter 2 features a rigorous comparison of two reprogramming regimens, there are also a number of unexpected results and questions still unanswered. These unanswered questions revolve around a single issue: that a significant amount of variation exists in reprogramming efficiency

regardless of the target phenotype, even though there was weak correlative support for reprogramming performance between iN and iSM reprogramming.

Recall that the rigor of the Chapter 2 investigation relied in part on two novel measures of cellular reprogramming. The first of these was infectability (Appendix A.1.1), defined as the proportion of cells able to uptake the viral vector. The second of these was reprogramming efficiency (Appendix A.1.2), defined as the proportion of cells that express an immunocytochemical marker corresponding to the target phenotype of interest. Calculation of these two measures is conceptually related: infectability is the subset of all quantified cells that also express the YFP element, and reprogramming efficiency is the subset of all infected cells (as characterized by the infectability measure) that also expresses a marker gene of interest. Yet because the quantification of reprogrammed cells does not always involve a straightforward cell count (in the case of iSM quantification, positive area above a threshold value is used), reprogramming efficiency and infectability are not nested nor can they be characterized in the same metric space.

Despite the close relationship between the infectability and reprogramming efficiency measures, there appears to be a weak empirical correspondance between infectability and reprogramming efficiency. This correspondance is in fact highly cell line- and target phenotype-dependent (Table 5.1). When regression coefficiencts are calculated for all mouse cell lines for iSM and iN conversion, most cases provide no signal. Notably, however, a few lines (HE4, KI6, TA4, and TE4) exhibit one for conversion to the iSM phenotype. This generally supports the hypothesis that infectability of cell lines does not predict their reprogramming efficiency.

114

Table 5.1. Regression coefficients (R^2) that characterize the relationship between the reprogramming efficiency and infectability measures for all 13 cell types. R^2 values calculated for cell lines converted to both neuron (iN) and muscle (iSM). Regression coefficients based on immunocytochemistry data.

Cell type	R^2 (iSM)	R^2 (iN)
HE4	.74	.35
KI6	.77	.04
KI2	.05	.50
LI6	.62	.00
KI3	.08	.09
LU3	.29	.36
KI5	.03	.00
TA4	.91	.36
LU6	.02	.01
TE4	.99	.00
SM1	.66	.24
TE5	.42	.14
TA6	.41	.08

Another finding in Chapter 2 deserving of more study involves the existance of larger effects between cells from the same genetic background than those from different genetic backgrounds. While mouse cell line reprogramming exhibits little preference for tissue type of origin, there are nonetheless many-fold differences between the best performing lines and the average performers. This many-fold difference is less pronounced among thehuman cell lines. Furthermore, the gene expression assays in

Chapter 2 do not replicate previous reports that fibroblasts exhibit variation with respect to anatomical position (Chang et al., 2002; Rinn et al., 2006).

There are two possible explanations for this type of result that might be help guide future work. One possibility involves differences that are only detectable after reprogramming. This could involve context-dependent gene expression or regulatory changes induced by the reprogramming process. However, these types of changes (e.g. epigenetic modifications) would likely have a multiplicative effect on cells from different genetic backgrounds, which does not correspond with our findings. There may also be a role for genomic variants, although during respecification to the iPS phenotype the expression of many allele-specific variants remains consistent (Lee et al., 2009). An alternate hypothesis for large-scale differences in the same genetic background involves the dynamics of cell populations, specifically contributions from both stochastic effects and the differential survival capacity of some cell lines over others in times of reprogramming-related stress. In addition, there is some evidence that differences due to cellular age exist. This lack of correlation between reprogramming regimens and lack of definitive trends provide an opportunity for future work.

5.1.2 Comparing indirect and direct reprogramming. In this chapter, we demonstrated that different techniques (direct *vs.* indirect) yield subtle differences that have no significant functional distinction. This was done by coupling literature review with high-throughput analysis. There may be a lesson in these results for understanding the genetic regulatory networks that control changes in cellular state. Specifically, various trigger points for initiating the changes involved in reprogramming contribute to only superficial outcomes. More informative are the potential "supportive" mechanisms that

may underlie but not determine successful cellular reprogramming, and might only be uncovered by comparing cell lines within a particular cellular state.

Given the evolutionary conservation among mammals for the early stages of development, pluripotent cell phenotype, and the nature of our selected genes (mostly involved with the ribosome and cellular structure), we should expect a high degree of similarity when conducting this analysis on data from different species. However, we should also be cognizant of surprises. In the case of functional gene expression associated with the onset and progression of sepsis, Seok et al. (2013) found large-scale differences between mouse and human. Recall that we find almost identical sets of genes in our lists which involve iPS cell lines (see Table 3.1). Due to their association with a process related to establishing a new cellular identity (rather than generic developmental processes), the composition of these lists might very well change when replication is attempted in other species.

5.1.3 New directions in direct reprogramming. In this review, we synthesize many of the non-iPS direct reprogramming results (e.g. iN, iSM, and iCM) with more traditional iPS findings in a framework of conceptual models and quantification. The focus in this chapter was on the various modeling techniques that might shine light on how to better predict and control the reprogramming process. This involves both the application of existing models with great potential (e.g. predictive, phenomenological, and dynamical) and advocacy of using a population perspective to understand the reprogramming process. We also introduce the stochastic threshold model (STM) and the concept of reprogramming bias, which is related to understanding differences in efficiency within and between reprogramming regimens.

To better understand the correlates and potential causes of bias requires a two-pronged approach: a measurement of a gene's expression after reprogramming normalized by the same gene's expression before reprogramming, and indirect gene network reconstruction that allows us to evaluate expression within and between local neighborhoods. Recall that in Chapter 2, the evaluation of a cell line's potential for differential reprogramming capacity was done by measuring gene expression prior to reprogramming. By constrast, measuring reprogramming bias will require measurements taken both before and after the reprogramming factors have been delivered. Normalizing gene expression after reprogramming by the "prior to reprogramming" measurement will allow us to target only those genes that change their expression due to the effect of our reprogrmming stimulus and not other potential sources of variation. The other analytical component involves defining local neighborhoods in an indirectly-reconstructed network topology. Neighborhood discovery has been applied to identify and understand the genetic background of disease-causing genes by Nitsch et al. (2009). Explicitly identifying local neighborhoods allows us to dissect the role of local versus global effects on a genetic regulatory network. Using this technique, higher-order effects of gene expression variation and gene-gene intetactions such as pleiotropy can likewise be better understood (Tyler et al., 2009).

5.2 Advanced results for role of diversity in cellular reprogramming

In Chapter 2, we presented results for a set of experiments in which a variety of cells were reprogrammed to both iN and iSM phenotypes. In the human case, the cells were of a variety of ages from different genetic backgrounds. In the mouse case, the cells were from a variety of locations in the body, but of a single genetic background.

Despite these differences, variability was actually greater in the mouse experiments. To better understand these results (in mouse data only), we can characterize these data as a Poisson process. This involves a) fitting our data to an exponential function (e.g. a Poisson process – Figure 5.1, part A), and b) treating our raw data as cell counts over time, from which we can derive rates for each cell line (Figure 5.1, part B). In addition more extensive details regarding all measurements, including an automated means to make more specific measurements of reprogramming (e.g. "yellow" channel quantification – see Table A.1), are provided in Appendix A.

5.2.1 Direct reprogramming as stochastic Poisson process. From the concepts advanced in Chapters 2 and 4, we can say that direct reprogramming proceeds as a stochastic Poisson process. Poisson processes (see Consul, 1989) can be defined as events that are distributed across time and occur at a variable rate. The direct reprogramming process is also stochastic in that the exact timing of these events is never truly predictable. Ultimately, our aggregate and uncorrected counts of reprogrammed cells can be used to infer the distribution of these reprogramming events in a population over time. While this information is not precise, models such as this allow us to move towards cell-line and condition specific predictive models of reprogramming efficiency. In conjunction with time-series data, we may also be able to correlate bursts of reprogramming (e.g. time intervals during which many reprogramming events tend to occur) with the expression of key genes or activity of key pathways. Future work might clarify if these bursts are due to collective regularities related to the gene regulatory network function of specific cell lines or random effects related that emerge at the level of cell population.
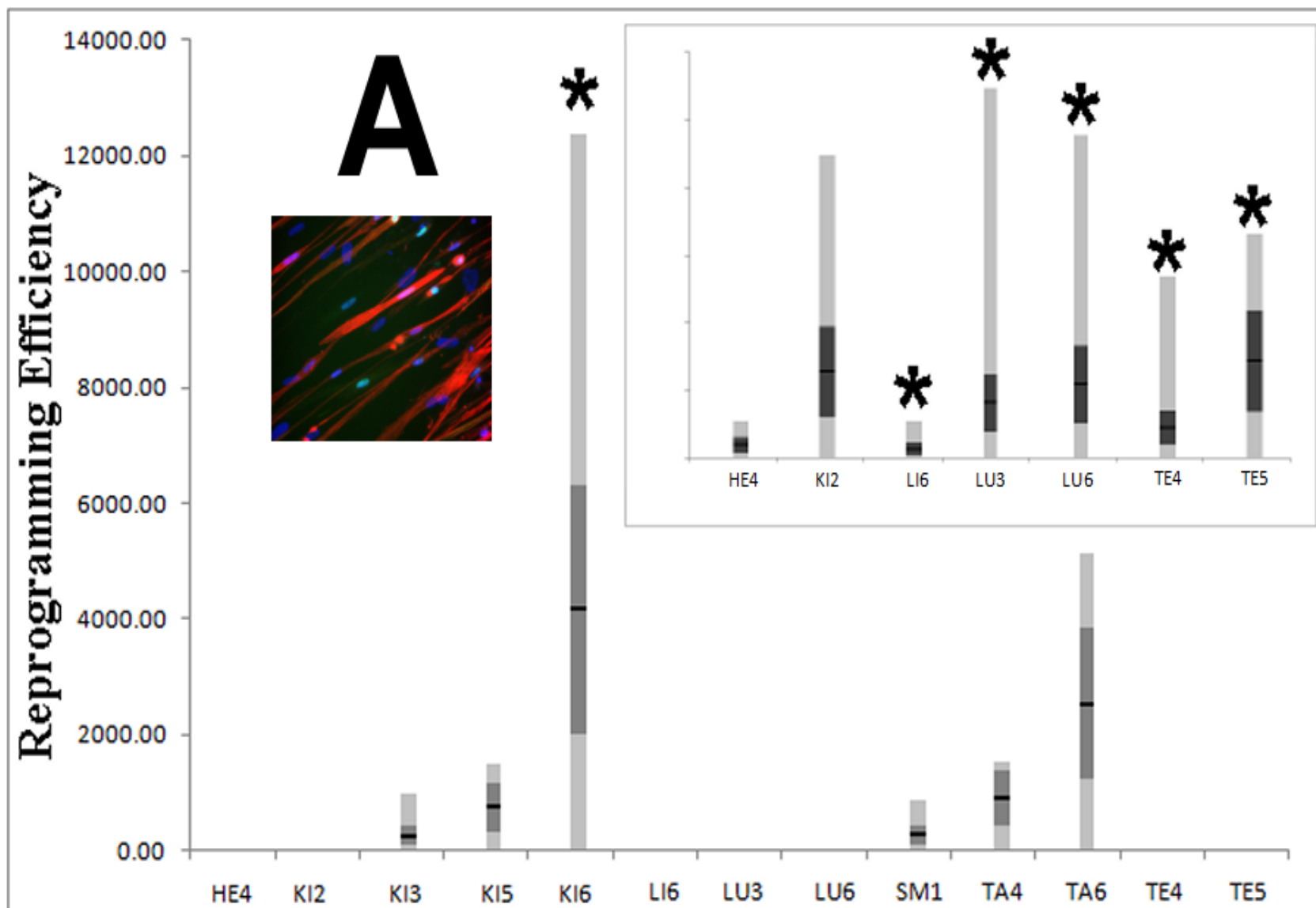
Figure 5.1. Rate-based information extracted from reprogrammed cell lines. Results of a Poisson exact test for iSM mouse cells (A) and iN mouse cells (B).

Figure 5.1 (con't).

121
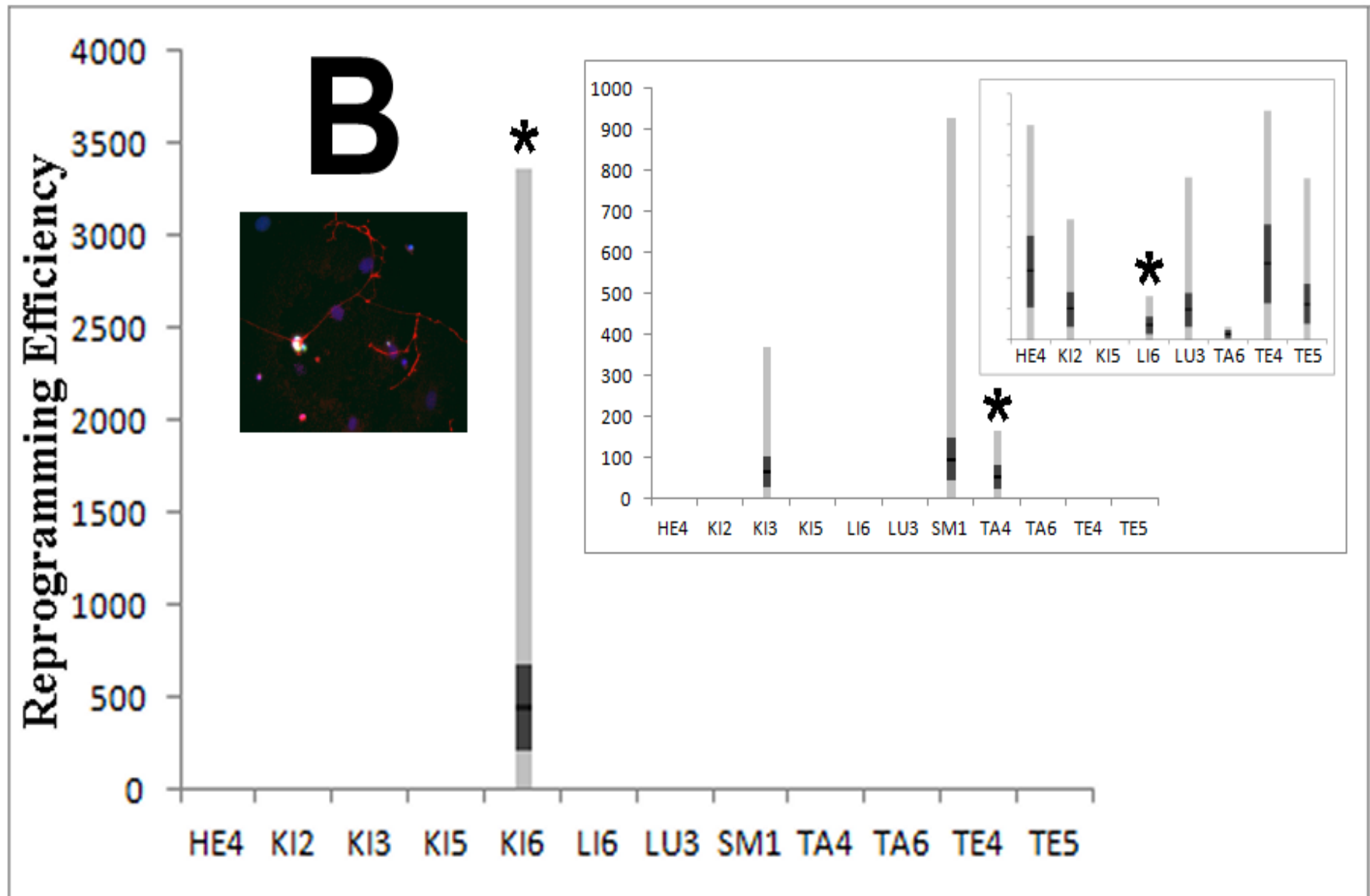
If these bursts are due to genetic regulatory network function, then reprogramming efficiency should be repeatable across experimental replicates. Figure 5.2 shows how this can be characterized using a histogram. Figure 5.2 demonstrates how many times a particular cell line end up in a given position, stated in the form of a frequency distribution. This can be thought of as a within and between cell line comparison that characterizes anomalies and trends among observed data. We will now test this using two examples. Two example of good performing cell lines are a mouse kidney cell line (Figure 5.2, Mouse) and a human IMR90 cell line (Figure 5.2, Human). The mouse kidney cells were the very best performing cell lines (mean performers) while also exhibiting high repeatability across replicates.

Theoretically, this suggests an innate mechanism is responsible, although our data suggest that population processes might also be responsible. This can be contrasted with the latter example, in which line #1 of human IMR90 cells are among the top performers (mean performance), but were highly variable across replicates. This suggests that the alternate hypothesis (emergent random effects) is more applicable here.

5.2.2 Potential measurements of reprogramming bias. We may also use our rank-order frequency distributions shown in Figure 5.2 to test for reprogramming bias as described in Chapter 4. This may be done by measuring the degree to which these distributions overlap for cells of the same line converted to muscle and neurons. While this resembles a joint distribution, the irregular nature of our distributions requires a specialized measure. This measure is taken by finding the difference in performance between each target phenotype for each histogram bin, with this being summed across

all bins. This results in an index that ranges from 0.0 (no overlap) to 1.0 (complete overlap). Bias can then be assessed by calculating 1 minus the overlap, and is consistent for distributions of any shape.

While this approach has the potential to enable us to uncover differential reprogramming performance between reprogramming regimens from replicates of reprogramming performance within single regimens, its robustness rests upon two assumptions. The first is that consistent performance across reprogramming regimens is expected to be directly proportional. A given cell line should have a similar rank when comparing the same cell lines destined for two different target phenotypes. The second is that bias occurs when reprogramming performance deviates from being directly proportional. In this way, we can quantify repeatability with respect to reprogramming performance. This quantification provides insights that can be applied both within and between reprogramming regimens.

5.2.3 Conceptual models for characterizing diversity in direct reprogramming. With few exceptions (Chang et al., 2002; Poss, 2010; Polo et al., 2010), the role of diversity in cellular function is generally treated as something to be minimized. In other cases, the massive complexity associated with genotype-phenotype interactions have been characterized using simple models, which are often incomplete in the face of an ever-expanding body of experimental results (Bhattacharya, Zhang, and Andersen, 2011). The results presented in chapters 2, 3, and 4 are an attempt to make sense of the diversity of cell types and phenotypic outcomes encountered in the study of cellular reprogramming. While in one way it makes sense to speak of reprogramming in terms of efficiency and phenotypic change, it also makes sense to interpret the results based

on what is happening at the level of cellular population processes. In Figure 5.3, two competing scenarios are introduced that presume the processes between different empirical outcomes.
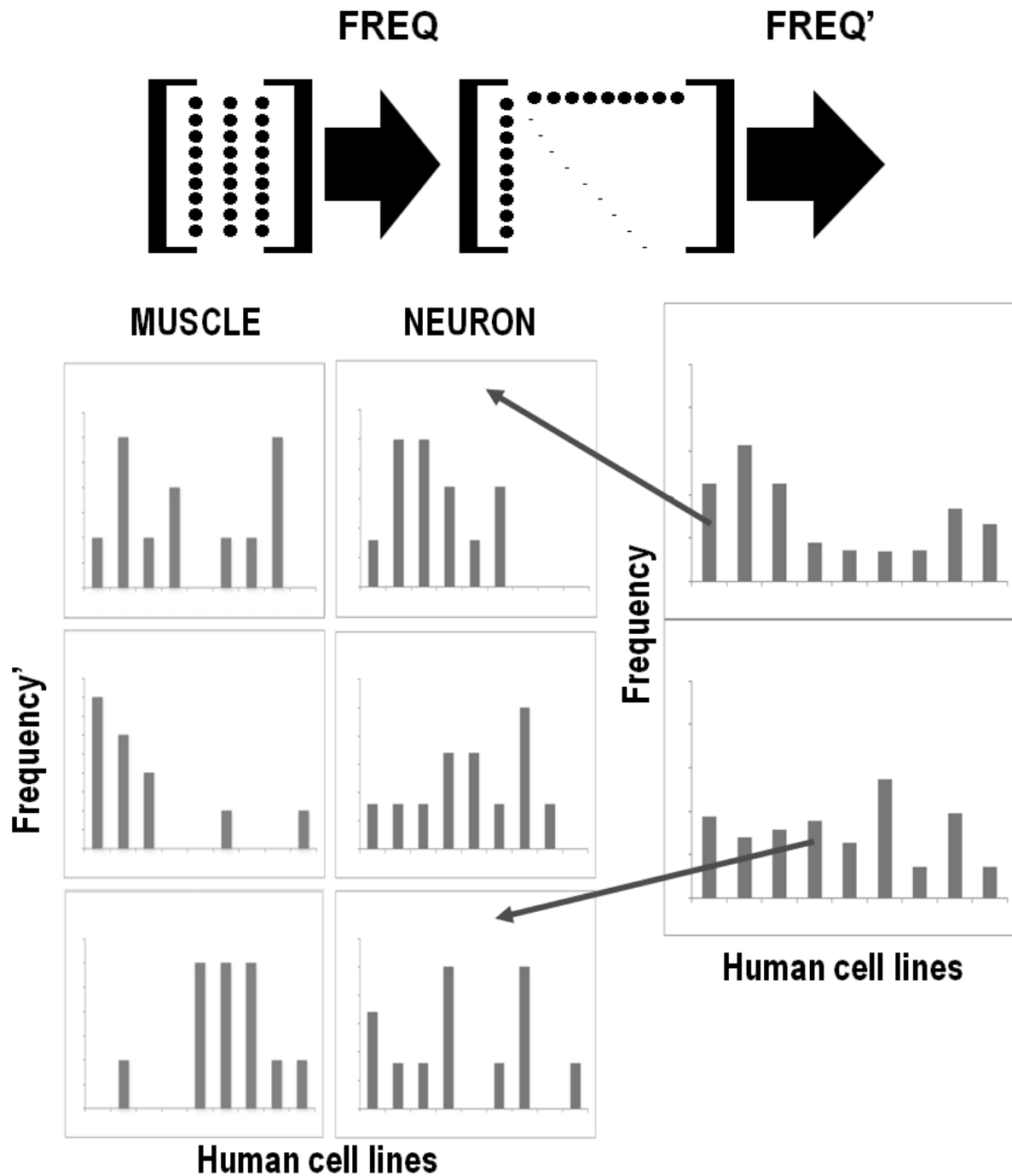
Figure 5.2. Rank-order frequency method for characterizing reprogramming efficiency performance and repeatability between experiments. Graphs with frequency' represents

Figure 5.2 (con't). the frequency of a given rank-order position across all replicates tested, while the graphs with the frequency y-axis represents frequency of each rank normalized by the inverse of each rank position and summed across all rank positions.
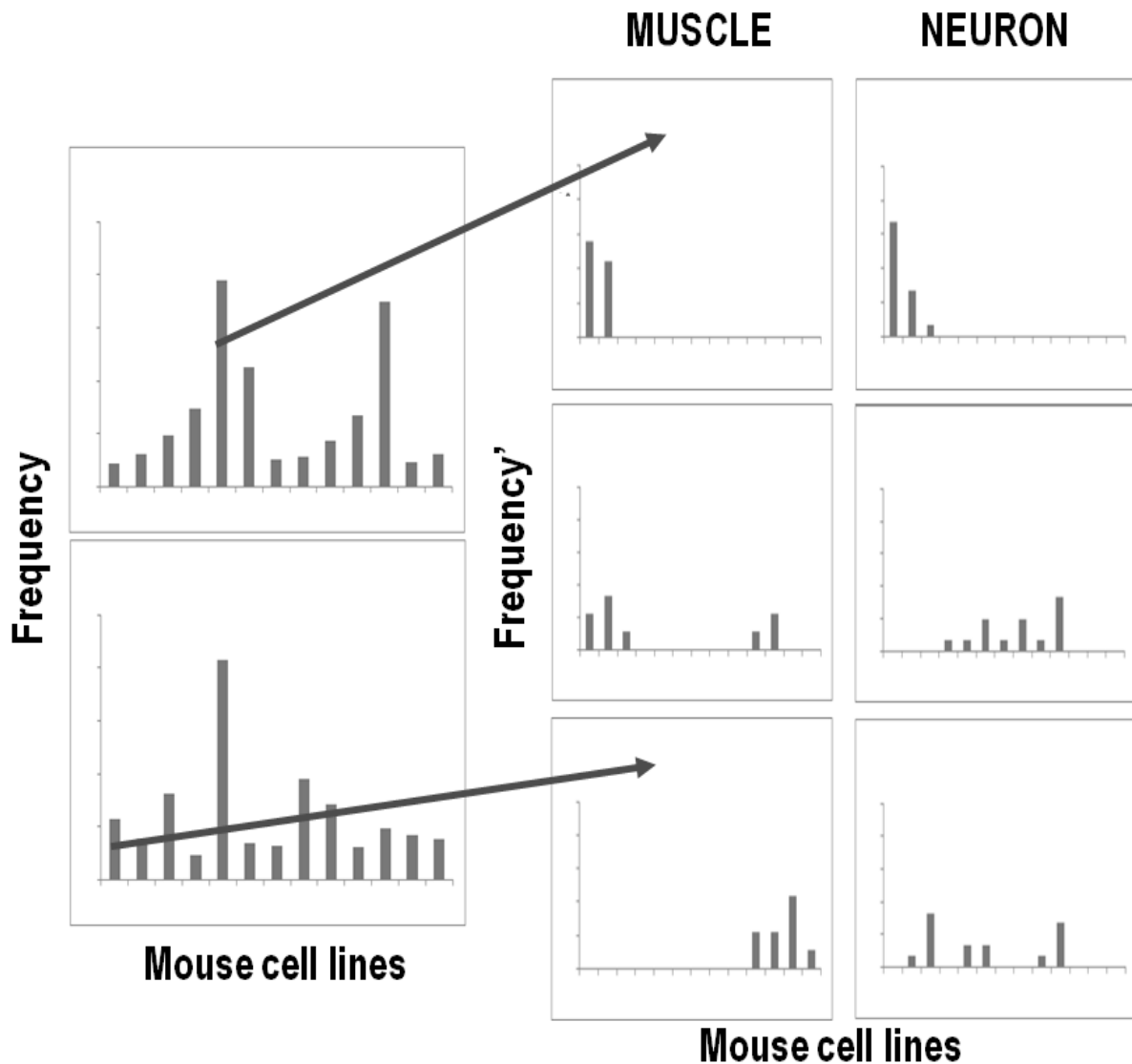
The **isolation and diffusion** scenario (Figure 5.3, top) explains what may cause consistently high reprogramming among a few clonal cell lines across experimental replicates. This process, analogous to genetic drift in population biology, relies on the existence of nominal advantages for certain cell types (such as kidney or liver cells). As cell lines of these types are isolated and cloned, some of the clonal lines exhibited an

amplified capacity for reprogramming. Whether these differences are due to microenvironment or innate factors is not clear. However, this scenario does suggest that high reprogramming efficiency has a real biological cause.

By contrast, the **1/f roulette** scenario (Figure 5.3, top) suggests that observed differences in reprogramming across cell lines are primarily a product of statistical sampling. This scenario may explain cases where high levels of reprogramming efficiency occur across a variety of cell lines across experimental replicates. This scenario also suggests that 1/f noise (or so-called pink noise) drives this process at multiple time-scales (Hausdorff and Peng, 1996). According to this scenario, all cell lines will exhibit a small number of replicates for which the level of reprogramming efficiency is high. These replicates are distributed randomly in time, so that over a series of observations, no single cell line will perform consistently with respect to reprogramming efficiency. This scenario does not suggest that high reprogramming efficiency has a real biological cause.

5.3 Future directions: *in vitro*

There are two potentially significant future directions for experimentally uncovering the systems-level complexity of direct cellular reprogramming. One direction involves assaying an actively reprogramming cell population at multiple timepoints. This experimental design would allow for transient features of the reprogramming process to be uncovered, particularly in the early phases of the reprogramming process. The other direction involves comparisons across a diversity of cell types using next-generation sequencing.

5.3.1 Time-series in direct reprogramming. A time-series experimental design that samples actively reprogramming cells at different time points may help us identify both critical points and transient events in the reprogramming process. This might involve the exopression of unique genes, or the assistive role of pathways and other factors not essential to the reprogramming process (e.g. *Oct4* or NANOG).
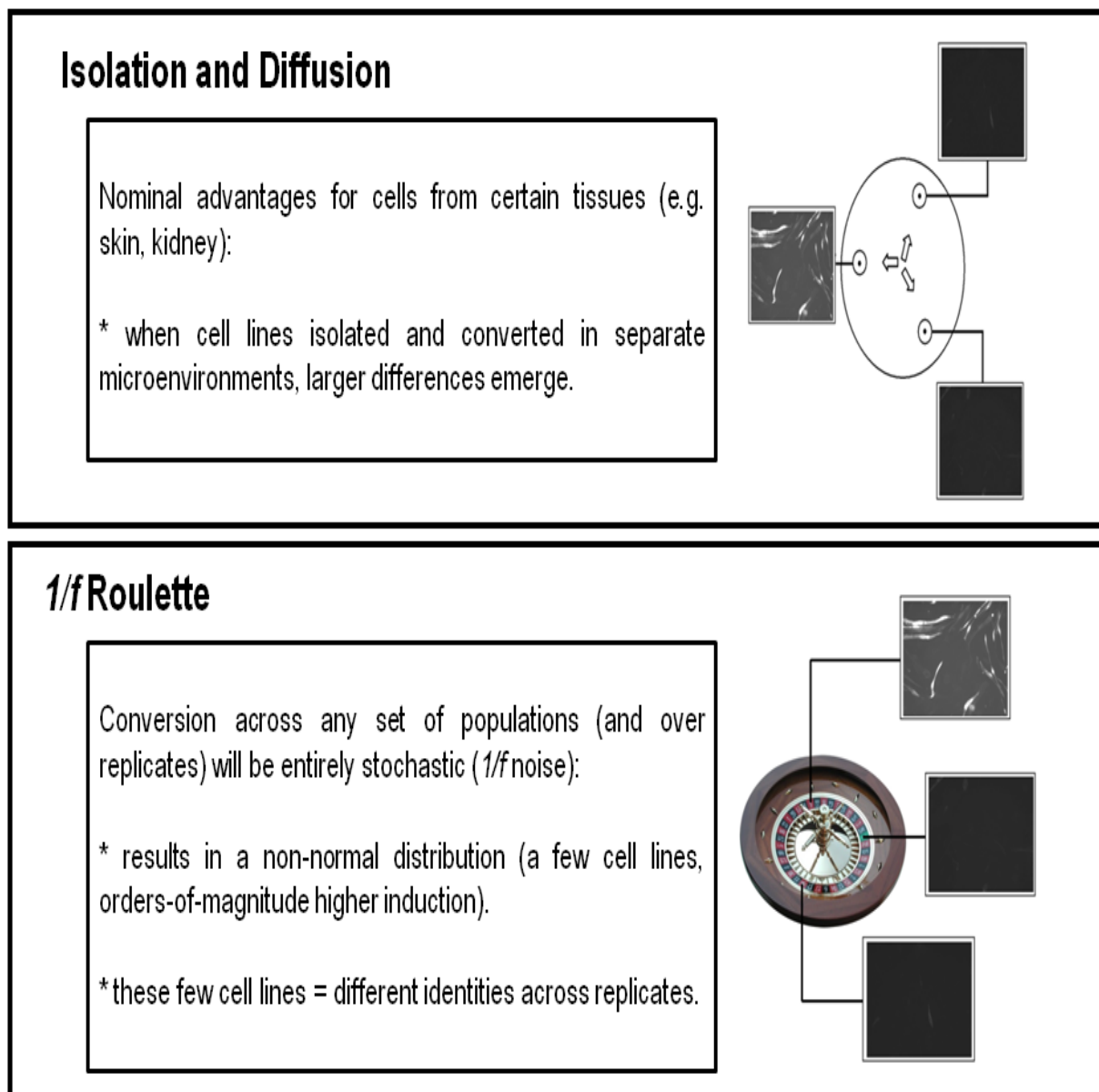
Figure 5.3. Hypothetical scenarios for preferential reprogramming capacity across cell line diversity. TOP: Isolation and diffusion scenario, BOTTOM: 1/f Roulette scenario.

It might even be possible to detect the early-onset signatures of critical events (e.g. sudden, large-scale changes to the cell) in the reprogramming process (for a non-reprogramming example, see Lade and Gross, 2012). This was demonstrated in a static context among pluripotent cells in Chapter 3, and this result might be improved upon using a time-series experimental design. Creating time-series datasets also allow us to compare and contrast the same events occurring dependently in time, such as general tendencies for gene expression during different phases of the reprogramming process.

Another application of the time-series design in the study of direct reprogramming is to use a reprogramming analogue. A reprogramming analogue involves treating uninfected cell lines with a compound that is known to have a disruptive effect (e.g. protein synthesis, transcription) to cell function. Taking measurements over a post-treatment timecourse allows us to characterize the role of RNA decay and aggregation mechanisms during this disruption. This ultimately fatal disruption may mimic the disruption to cellular processes and phenotype caused by direct reprogramming. To further evaluate this observation, I have developed an approach called *mechanism alteration*, which is introduced and demonstrated in Appendix C.

5.3.2 Direct reprogramming via next-generation sequencing. While the high-throughput gene expression of reprogrammed cells (e.g. iPS cells) has been characterized, the characterization of a multiple related cell lines is less well-known. This is especially true in terms of next-generation sequencing technologies. Next-generation sequencing combines whole-genome sequencing with a functional assay (e.g. gene expression levels) in an integrated fashion. Future directions might include combining RNA-seq

(gene expression) and BS-seq (methylation) to better understand gene expression in an epigenetic context at the systems level.

5.4 Future directions: *in silico*

In silico approaches are useful for simulating the possible behaviors of a reprogramming cell culture. I have done several presentations on the use of excitable cellular automata for purposes of modeling the infection dynamics, genomics, and phenotypic remodeling inherent in the reprogramming process. This represents a promising direction for future work, as a complex, agent-based model of direct reprogramming at the population level could resolve many mysteries of colony formation, cytotoxicity, and reprogramming efficiency.

5.4.1 Artificial life approximation of direct reprogramming. One highly suitable model for approximating reprogramming cell populations is the *cellular automaton*. Cellular automata are agent-based models of artificial life that allow for the exploration of spatially-explicit propagation behaviors (e.g. phenomena that traval from cell-to-cell) and emergent properties in these populations. For purposes of representing cellular reprogramming, cellular reprogramming can be combined with genetic algorithms to provide us with autonomous units that can be arrayed in parallel. More technical detail can be found in Appendix D.

5.5 Conclusions

In conclusion, what lessons have we learned from these experiments, analyses, and future directions that may be applied to repairing, modeling, or otherwise reconstructing the nervous system? Here we have explored three types of direct cellular reprogramming that produces cell types most relevant to the nervous system: iPS, iN,

and iSM/iCM cells. While conversion to each cell type requires a different set of transcription factors and operates at different efficiencies, there are also four lessons that apply to all cell types derived via direct reprogramming.

Two of these lessons are directly related to the potential transience of the direct reprogramming phenomenon. The first of these, perhaps not surprisingly, is that each type of reprogramming is unique. This uniqueness is manifest in the form of distinct regulatory mechanisms, and in some cases a mosaic response. Secondly, differences in efficiency demonstrate that even within cell types, there is no uniformity in the process. These results confirm the hypothesis that there is no single, dedicated pathway to a successfully reprogrammed cell, even within a specific cell type. One consequence of this non-uniformity is variability that is hard to control or compensate for across replicates or experimental settings. Yet non-uniformity of process does not imply that reprogramming is an epiphenomenon, as the induction of reprogramming is quite distinct from spontaneous phenotypic conversion.

The additional two lessons involve the contribution of multiple mechanisms at different scales to the direct reprogramming process. Thus, our third lesson is that population approaches require new conceptualizations of roles for individual genomes and population processes. In many cases, the outcomes of genetic regulation and the context of cellular phenotype are treated as disconnected processes. Perhaps this is not an acceptable strategy, particularly as the reprogramming efficiency of a population is implicitly linked to changes in gene expression. The final lesson of these studies involves a more fundamental understanding of within- and between-cell type differences in efficiency. This focus on biological diversity is a unique approach in cell and

molecular biology, and requires a movement towards systems-level and hybrid population/process models.

APPENDICES

APPENDIX A: TECHNIQUES FOR MEASURING REPROGRAMMING

A.1 MEASURES:

In the course of conducting the experiments in the Chapter 2, we developed several measures that provide a systematic, *in situ* assessment of reprogramming efficiency and infectability of reprogrammed cell lines. These include independent metrics of reprogramming efficiency (RE) and infectability (I), in addition to a yellow channel measurement that can differentiate between infected cells and cells which are both infected and reprogrammed. Image segmentation and pixel counts were done using Image J and MATLAB.

A.1.1 INFECTABILITY MEASURE.

Infectability (I) was developed to quantify the proportion of cells that have successfully taken up the virus and expressed the transgenic elements related to cellular reprogramming. Infectability can be defined by the following equation

$$\mathbf{I = \frac{YFP}{DAPI}} \qquad [1]$$

where YFP represents the green channel and DAPI represents the blue channel in a *r,g,b* color scheme.

A.1.2 REPROGRAMMING EFFICIENCY MEASURE.

Reprogramming efficiency (RE) was developed as a way to quantify all those cells or particles in an image that represent a successfully converted cell. Reprogramming efficiency can be defined by the following equation

$$RE = \frac{RED}{YFP : DAPI} \qquad [2]$$

where RED represents the red channel, YFP represents the green channel and DAPI represents the blue channel in a *r,g,b* color scheme.

A.1.3 TOWARDS GREATER SPECIFICITY IN MEASURING INFECTABILITY.

Because the reprogramming efficiency measure is more systematic than many others used in the literature it has the potential to directly differentiate between cells that are merely infected and cells that have been infected and express the genes most indicative of successful reprogramming. The yellow channel can be mathematically defined as the intersection of the red and green channels.

$$Y = R \cap G \qquad [3]$$

In the context of digitized images, this information can be provided by the red and blue color channels (in an *r, g ,b* color scheme). All color channels can be represented as an *m*-by-*n* matrix. To extract useful information, however, we must incorporate two additional pieces of information: conversion of each color channel into a binary matrix determined by a threshold intensity, and correction for background noise. Given these two constraints, the yellow channel (in matrix form) can be defined by the following equation.

$$y_{ij} = -b_{ij} + (r_{ij} - g_{ij}) \qquad [4]$$

$$\text{where } y_{ij} > t$$

where *t* is an arbitrary threshold value.

A.1.4 MATLAB FUNCTION FOR YELLOW CHANNEL AND AUTOMATED ANALYSIS

Table A.1. MATLAB code for yellow channel segmentation.

```
function [YM] = yellowchannel(t, 'import')
%%
% creates yellow channel (true positive green signal) for further analysis. t = threshold
value % between -1 and 254. 'import' is a rasterized graphic file (e.g. .tif, .jpg).
%%
x = image('import')
%%
% import images as x-by-y-by-z matrices.
%%
r = x(:,:,1);
g = x(:,:,2);
b = x(:,:,3);
rn = r/(r+g+b);
gn = g/(r+g+b);
bn = b/(r+g+b);
%%
% extract and normalize color channels. red = 255,0,0; green = 0,255,0; blue = 0,0,255.
%%
r2 = rn > t;
g2 = gn > t;
b2 = bn > t;
%%
% further normalize all channels relative to numeric threshold.
%%
y1 = r2 - g2;
b3 = b2 * -1;
y2 = b3 + y1;
%%
% generates a yellow-only channel without noise (low-intensity overlap of red and
green).
%%
[YM] = image(y2)
%%
% generate a binary color map (YM = yellow map). This output can be used to calculate
yellow channel statistics.
%%
```

APPENDIX B: SUPPLEMENTAL TABLES FOR CHAPTER 2

Table B.1. Primers using in cloning (primers used in the amplification of cDNAs or plasmid DNA for construction of retroviral vectors) analysis.

Sequences Used for Cloning		
PRIMER	**5'-3' SEQUENCE**	**SPECIES**
ASCL1 F	GAGAGAACGCGTGGCATGGAAAGCTCTGCC	Human
ASCL1 R	ACACACATCGATTCAGAACCAGTTGGTGAAGTCG	Human
MYF5 F	GAGAGAACGCGTATGGACGTGATGGATGGCTGCC	Mouse
MYF5 R	GTGTGTAATCGATTCATAGCACATGATAGATAAGCC	Mouse
MYF6 F	GAGAGAACGCGTATGATGATGGACCTTTTTGAAACTGG	Mouse
MYF6 R	GTGTGAATCGATTTACTTCTCCACCACTTCCTCCACGC	Mouse
MYOD F	GAGAGAACGCGTGGTATGGAGCTTCTATCGCCGCCAC	Mouse
MYOD R	ACACACATCGATTCAAAGCACCTGATAAATCGC	Mouse
MYOG F	GAGAGAACGCGTATGGAGCTGTATGAGAC	Mouse
MYOG R	GTGTGTAATCGATTCAGTTGGGCATGGTTTCATC	Mouse
MYT1L F	GAGAGAGGCGCGCCCGATGGAGGTGGACACCGAGG	Human
MYT1L R	CACACAATCGATTCAGACCTGAATTCCTCTCACAGCC	Human
NEUROD1 F	GAGAGAACGCGTGGTATGACCAAATCGTACAGCG	Human
NEUROD1 R	GTGTGTGTTTAAACCTAATCATGAAATATGGCATTGAGCTG	Human
POU3F2 F	GAGAGAGGCGCGCCCAATGGCGACCGCAGCGTCTAACC	Human
POU3F2 R	ACACACTATCGATTCAACGCGTCTGGACGGGCGTCTGCAC	Human
YFP F	CACAGGCCGCCTGGGCCATGGTGAGCAAGGGCG	Other
YFP R	AAACTTAACGCGTCTTGTACAGCTCGTCCATG	Other
ZIC1 F	GAGAGAGGCGCGCCCGGGAATGCTCCTGGACGCCGG	Human
ZIC1 R	ACACACATCGATTAAACGTACCATTCGTTAAAATTGGAAGAGAGCGCAC	Human

Table B.2. Primers using in Quantitative PCR (primers used in qRT-PCR quantification of gene expression) analysis.

PRIMER	5'-3' SEQUENCE	SPECIES

PRIMER	5'-3' SEQUENCE	SPECIES
PANL27 F	CCATCCAGACTGAGGAAGACCCGGAAAC	Human/Mouse
PANL27 R	GGGCAGAAGCTCTGGTTCCTC	Hu/Mo
ARHGAP1 F	TGCTGTGGGCCAAGGATGCG	Human
ARHGAP1 R	GGTCCGGGCTTGGGAACAGC	Human
COL1A2 F	CAGGGGCTCTGCGACACAAGG	Human
COL1A2 R	TCCGGCTGGGCCCTTTCTTAC	Human
EED F	GGAAGGAGCCAGGAAGCCGC	Human
EED R	ACTGTCGCAAATCGCGCCCA	Human
FIBR1 F	AGGAAACCAGAGCCAGTCGGG	Human
FIBR1 R	GGAATGCCGGCAAATGGGGACA	Human
FIBU5 F	GTGTGTGAACCAGCCCGGCA	Human
FIBU5 R	ACGTCTGCTGCAGGTTGCACG	Human
FNECTIN F	CGCCCTGGTGTCACAGAGGCTA	Human
FNECTIN R	TGGGGTGTGGAAGGGTTACCAG	Human
FOXG1 F	ACGGGGAGATCCCGTACGCC	Human
FOXG1 R	CCGCGAGCAGGTTGACGGAG	Human
KER14 F	GCAGCGGCCTGCTGAGATCAA	Human
KER14 R	CATTGGCATTGTCCACTGTGGCT	Human
LAMIN F	GTGCGCTCAGTGACTGTGGTTGA	Human
LAMINR	CGAGCGCAGGTTGTACTCAGCG	Human
MYF5 F	TCTCCCCATCCCTCTCGCTGC	Human
MYF5 R	CCACTCGCGGCACAAACTCGT	Human
MYOD F	CTCCAACTGCTCCGACGGCA	Human
MYOD R	TCGACACCGCCGCACTCTTC	Human
PECAM F	TCCACATCAGCCCCACCGGA	Human
PECAM R	TGGGCCACAATCGCCTTGTCC	Human
SOX2 F	GGGGGAAAGTAGTTTGCTGCCTC	Human
SOX2 R	CTGCCGCCGCCGATGATTGT	Human
VIMENTIN F	GAGCAGGATTTCTCTGCCTCTTCC	Human
VIMENTIN R	TCGTGATGCTGAGAAGTTTCGTTGA	Human
ARHGAP1 F	TTTGCCGAGCTTTGACAGGCG	Mouse
ARHGAP1 R	AATGGAGGCCAGCTTCAACTGG	Mouse
COL1A2 F	CAGGGGCTCTGCGACACAAGG	Human
COL1A2 R	TCCGGCTGGGCCCTTTCTTAC	Human
EED F	GGAAGGAGCCAGGAAGCCGC	Human
EED R	ACTGTCGCAAATCGCGCCCA	Human
FIBR1 F	AGGAAACCAGAGCCAGTCGGG	Human
FIBR1 R	GGAATGCCGGCAAATGGGGACA	Human
FIBU5 F	GTGTGTGAACCAGCCCGGCA	Human
FIBU5 R	ACGTCTGCTGCAGGTTGCACG	Human

Table B.2 (con't).

PRIMER	5'-3' SEQUENCE	SPECIES
Sequences Used for Quantitative PCR		
FNECTIN F	CGCCCTGGTGTCACAGAGGCTA	Human
FNECTIN R	TGGGGTGTGGAAGGGTTACCAG	Human
FOXG1 F	ACGGGGAGATCCCGTACGCC	Human
FOXG1 R	CCGCGAGCAGGTTGACGGAG	Human
KER14 F	GCAGCGGCCTGCTGAGATCAA	Human
KER14 R	CATTGGCATTGTCCACTGTGGCT	Human
LAMIN F	GTGCGCTCAGTGACTGTGGTTGA	Human
LAMINR	CGAGCGCAGGTTGTACTCAGCG	Human
MYF5 F	TCTCCCCATCCCTCTCGCTGC	Human
MYF5 R	CCACTCGCGGCACAAACTCGT	Human
MYOD F	CTCCAACTGCTCCGACGGCA	Human
MYOD R	TCGACACCGCCGCACTCTTC	Human
PECAM F	TCCACATCAGCCCCACCGGA	Human
PECAM R	TGGGCCACAATCGCCTTGTCC	Human
SOX2 F	GGGGGAAAGTAGTTTGCTGCCTC	Human
SOX2 R	CTGCCGCCGCCGATGATTGT	Human
VIMENTIN F	GAGCAGGATTTCTCTGCCTCTTCC	Human
VIMENTIN R	TCGTGATGCTGAGAAGTTTCGTTGA	Human
ARHGAP1 F	TTTGCCGAGCTTTGACAGGCG	Mouse
ARHGAP1 R	AATGGAGGCCAGCTTCAACTGG	Mouse
COL1A2 F	CTCATACAGCCGCGCCCAGG	Mouse
COL1A2 R	CGGTTGGCTAGCAGGCGCAT	Mouse
EED F	CGCCGGCGGGAACAGACATG	Mouse
EED R	TATTTGTGGGCGTGTCCGGGC	Mouse
FIBR1 F	AGGCCCCCTGCAGTTACGGT	Mouse
FIBR1 R	CCTCGGCCCATGCCCATTCC	Mouse
FIBU5 F	ACAACCCGATACCCTGGTGCCT	Mouse
FIBU5 R	CGAGGCCCTTTGATGGGGCG	Mouse
FNECTIN F	GAGCGACATGCTCTACAAAGTGCT	Mouse
FNECTIN R	CTGGGGGTGAGTCTGCGGTTG	Mouse
FOXG1 F	CGATCGCGGCTACCGGCTTC	Mouse
FOXG1 R	CACTCCCAGAGTCGCGCTCAC	Mouse
KER14 F	ACAGCCCCTACTTCAAGACCATCG	Mouse
KER14 R	CGCAGGCTCTGCTCCGTCTC	Mouse
LAMIN F	GCCTTCGCACCGCTCTCATC	Mouse
LAMINR	GCCGCTGCAGTGGGAACC	Mouse
MYF5 F	CCCCAACCTCAGCCACTGACC	Mouse
MYF5 R	GCCAGCAAATCCAGGCGGAGC	Mouse
MYOD F	GGAGATCCTGCGCAACGCCA	Mouse
MYOD R	GCAGCGGTCCAGGTGCGTAG	Mouse

Table B.2 (con't).

Sequences Used for Quantitative PCR		
PRIMER	**5'-3' SEQUENCE**	**SPECIES**
PECAM F	ACGAGAGCCACAGAGACGGTG	MOUSE
PECAM R	AGGGACGTGCACTGCCTTGAC	MOUSE
SOX2 F	GCTGCCTCTTTAAGACTAGGGCTG	MOUSE
SOX2 R	GCCGCCGCGATTGTTGTGAT	MOUSE
VIMENTIN F	GTCGAGGTGGAGCGGGACAAC	MOUSE
VIMENTIN R	CCGTTCAAGGTCAAGACGTGCCA	MOUSE

Table B.3. Properties of human and mouse fibroblast lines.

Human Fibroblast Lines					
STUDY ID	**LINE ID**	**DONOR**	**SEX**	**AGE**	**TISSUE**
FET	IMR90	Healthy	F	E16 wk	Lung
NWB	HDNF	Healthy	M	Newborn	Skin
ADF	MSU-HUMGM	Healthy	M	44	Gingiva
RET	GM17880	Rett syndrome	F	5	Skin
E2F	MSU-HUMAG10	Healthy	M	73	Skin
EAF	MSU-HUMAG07	Healthy	M	71	Skin
SAF	GM01792	Schizophrenia	M	26	Skin
AUT	GM07992	idic(15) autism	F	3	Skin
HSK	MSU-HUMSK	Healthy	M	41	Skin

Mouse Fibroblast Lines					
STUDY ID	**LINE ID**	**DONOR**	**SEX**	**AGE**	**TISSUE**
HE4	N/A	nude mouse	M	5 mo	Heart
SM1	N/A	nude mouse	M	5 mo	Skeletal Muscle
KI2	N/A	nude mouse	M	5 mo	Kidney
KI3	N/A	nude mouse	M	5 mo	Kidney
KI5	N/A	nude mouse	M	5 mo	Kidney
KI6	N/A	nude mouse	M	5 mo	Kidney
LI6	N/A	nude mouse	M	5 mo	Liver
LU6	N/A	nude mouse	M	5 mo	Lung
TA4	N/A	nude mouse	M	5 mo	Tail Skin
TA6	N/A	nude mouse	M	5 mo	Tail Skin
TE4	N/A	nude mouse	M	5 mo	Testis
TE5	N/A	nude mouse	M	5 mo	Testis
LU3	N/A	nude mouse	M	5 mo	Lung
MEF	N/A	mouse	F/M	E13 day	Embryo

LABELS: STUDY ID – Abbreviated line designation used this report.; LINE ID – Line common name.; DONOR – donor health status or strain; SEX – sex of donor (Male, Female, or of mixed sex); AGE – Age of donor (years, unless otherwise noted); TISSUE – fibroblast tissue of origin. *mRNA Expression:* results of quantitative PCR analysis of multiple markers. COL1A2 – fibroblast marker collagen type I alpha 2; FNECTIN - fibroblast marker fibronectin; FIBR1 – fibroblast marker fibrillin I, FIBU5 – fibroblast marker fibulin V; VIM – fibroblast marker vimentin; KER14 – keratinocyte marker keratin 14; PECAM – endothelial cell marker Platelet endothelial cell adhesion molecule/CD31; *FOXG1* – neural progenitor marker Forkhead box protein G1; *SOX2* – neural progenitor

marker SRY (sex determining region Y)-box 2, *MYOD* – myogenic progenitor marker *MyoD1*; *MYF5* – myogenic progenitor marker myogenic factor 5. *Immunocytochemistry:* summary of immunocytochemical analysis of fibroblast-associated (fibronectin, vimentin) and stem cell-associated markers (Sox-2, nestin) in fibroblast lines. N/A – not applicable; ND – not done; + = positive; +/- = weakly positive, - = negative.

APPENDIX C: MECHANISM DISRUPTION TECHNIQUE

C.1 INTRODUCTION

Mechanism alteration can be defined as disruption of a key cellular process that does not result in immediate cell death. The biochemical milieu of a cell can reveal much information about changes in its phenotype (for example, see Jarosz, Taipale, and Lindquist, 2010). Using several forms of mechanism alteration (cell cycle, protein synthesis, and ribosomal degradation), ways to better understand the dynamics of cellular information processing during a transformational process will be identified (Figure A.1). Mechanism alteration initiated by treatment with drug compounds will reveal adaptive responses in both the translatome (e.g mRNA associated with translation) and transcriptome (e.g. mRNA associated with transcription). These data can likewise be modeled to demonstrate regulatory features that lie between transcriptional and translation-associated RNA such as feedback and delays. The resulting data can also provide insight into how genes behave during transformation, and may lead us to a new view of how cells can convert from one phenotype to another. From these findings, the adaptive capacity of a group of genes or cellular population given environmental challenges may be inferred.

What dynamics can tell us about the control of cellular processes.

Due to the relative efficiency and accuracy of translation and transcription (Kirkwood, Rosenberger, and Galas, 1986), capturing the dynamics of RNA molecules associated with these processes can provide a window into changes associated with cell morphology, protein expression, and other outcomes (Rabani et al., 2011; Chechik and Koller, 2009; Barenco et al., 2009). The artificial manipulation of cellular RNA by exposing a cell population to specific compounds can yield insight into changes that

represent shifts in gene regulation due to aging or environmental stress. Simulated conditions of mass arrest of cell cycle, protein synthesis, and ribosomal degradation will be combined with a previously described method of assaying translational-associated RNA to build a model of RNA decay and regulation in adult human fibroblast populations. Changes over time due to the disruption of key cellular processes result in fluctuations that provide clues as to how cells process information. Information processing is a multivariate process that sets up adaptive responses, determines changes in mRNA levels, and ultimately affects the phenotypic state and long-term viability of a cell (Balazsi, van Oudenaarden, and Collins, 2011).

What decay can tell us about the control of cellular processes.

As complex systems and biological factories, cells produce, transform, and recycle materials using many different pathways (Garneau, Wilusz, and Wilusz, 2007). This results in biochemical fluctuations that can be observed through gene expression trends over time (Shalem et al., 2008). Therefore, while the manipulation of specific pathways can provide more explicit insights into cellular processes relative to specialized functions of a cell type, the manipulation of general mechanisms is superior in terms of understanding global functions associated with cellular transformation (e.g. cellular reprogramming).

C.2 INDUCTION OF MECHANISM DISRUPTION

Mass arrest of major features of a cellular phenotype, such as cell cycle, transcription, and protein synthesis, can be accomplished using either drug treatment or transgenic knockdowns. Drug treatments have been done on selected cell lines to assess the time course of mRNA decay after exposure to a drug stimulus. Drug

treatments can be uneven and its effects related to decay are unknown for many cell lines, but a generalized effect of gene expression can be observed. In previous studies this effect has proven to be a mosaic, as genes with different functions respond differentially to the drug stimulus (Perez-Ortin, 2008; Garneau, Wilusz, and Wilusz, 2007). Knockdowns of specific genes have also resulted in shutting down expression using a transgenic construct and measuring the remaining mRNA (Archer et al., 2008). While knockdown-decay studies are more focused on a specific gene and more completely knocks down a gene's expression, the focus in only on a single gene. While both methods can be leaky, there is a tradeoff between capturing a generalized decay response across many types of genes and capturing decay of a single gene.

Components of RNA Dynamics

To address the time-dependent and diversity-related properties of cell populations, measurements of mRNA half-life (e.g. decay, sequestration) will be induced as a way to investigate mechanism alteration. This forms the basis for a model with which to assess changes to a cell population in response to a stimulus. We propose that drug treatments can be used to interfere with normal cellular processes through mass arrest of RNA synthesis and other normal processes. It has been shown that drug treatments are a blunt instrument for examining RNA decay (Grunwald, Singer, and Czaplinski, 2008; Archer et al., 2008), and stands in contrast to transgenic approaches which shut off specific genes (Chen, Ezzeddine, and Shyu, 2008). Yet drug treatments allow for multiple genes to be examined in the context of a single stimulus, and can thus reveal features relevant to early reprogramming and cellular diversity.

C.3 UNCOVERING THE COMPLEXITY OF A DIRECT REPROGRAMMING ANALOGUE

Treatment using drug compounds produce dynamics that are much more complex than simple decay. There are several contemporary examples of how RNA dynamics are studied that place our results in context. Barenco et al. (2011) has examined RNA dynamics by examining the DNA-damage response in the MOLT4 pathway, which acts in a cell-line and context specific manner. Using this model, they are able to use a hidden variable dynamical model to partition out variance associated with decay and other signatures over time from time-series microarray data. By contrast, our model does not explicitly partition variance related to different cellular processes. However, in holding certain mechanisms constant, we can gain insight into adaptive responses on a longer time-scale than do Barenco et al. (2011) or Sharova et al. (2009). In this way, we can observe several potential nonlinear control mechanisms. What we do not know is whether these signatures are caused by previous gene expression patterns, a generalized adaptive response, or an indicator of pure decay.

Role of Degradation in Mechanism Alteration

Rabani et al. (2011) test the varying degradation hypothesis, which posits that changes in RNA level over time are strongly affected by changes in degradation rate. This can be characterized by either a single or continuous shift in RNA profiles, which tend to be gene specific. Likewise, we observe this gene-independent activity, which can be characterized either by a single shift (understood through nonlinear curve fitting) or a continuous shift (consistent with linear aggregation or decay response). These authors also use a technique called *4sU labeling* to separate newly transcribed RNA from the total pool (Rabani et al., 2011). In doing so, it appears that the processing of

mRNA at the site of transcription plays a role in shaping longer-term temporal functions. This is also consistent with the impulse model of Chechik and Koller (2009), who suggest that RNA dynamics are characterized by an abrupt early response coupled to a later transition towards a steady state. This regulatory output may be due to the pre-mRNA processing observed by Rabani et al. (2011) or ribosomal functions (Chechik and Koller, 2009). This is also consistent with the complex responses observed in our data, particularly across different functional classes of gene.

Role of Noise in Mechanism Disruption

In addition, we also propose that drug treatments are a way to hold constant the effects of intrinsic cellular noise (Isaacs et al., 2005; Ozbudak et al., 2002). Mitomycin C (see Sharova et al., 2009), Actinomycin D, and Saporin can be used to affect cell division, RNA synthesis, and ribosomal survival (integrity), respectively. Manipulating these pathways and then sampling transcriptome and translatome at 24 hours after initial treatment should allow us to approximate the nature of RNA turnover and aggregation which plays a key role in processes such as cellular transformation. Patterns of decay across three or four samples across time can be assessed using linear and nonlinear statistical curve-fitting techniques (Sharova et al., 2009; Larsson, Sander, and Marks, 2010), which when applied in tandem at a time-scale of days can reveal finer-scale fluctuations not due to intrinsic noise. We further propose that the effects of such a manipulation will be gene-specific. For example, cell-type specific genes should be affected differently than housekeeping genes. To explain the pulsatile nature of this response, we may turn to a simple control model (see D'Haeseleer et al., 1999; Bay et al., 2004; MacDonald, 1989) that explain aggregation, decay, and

feedback mechanisms that affect RNA measurements over time in response to a stimulus.

Towards a General Mathematical Model

To better understand exactly what the observed nonlinear responses mean in terms of regulatory mechanism, we may turn insights discrete dynamical equations (DDEs). DDEs have been used to better understand network-level control and stability in cellular systems (Orosz, Moehlis, and Murray, 2010; MacDonald, 1989). In this case, we propose that a DDE with conditions that represent second-order responses (see Figure A.2) will model delays, which in turn serve as signatures of delays in transcription and translation. For example, a delay in transcription will result in an observed aggregation of RNA in either the transcriptome or translatome. By contrast, a delay in transcription will lead to steep decay with later recovery of RNA in the transcriptome. This could also be true of a similar response in the translatome, but it cannot be distinguished from an increase in the speed of translation. It is of note that this response was not observed in the translatome for any of the assayed genes. The DDE model may serve as a gateway to future studies involving dynamic processes such as direct cellular reprogramming.
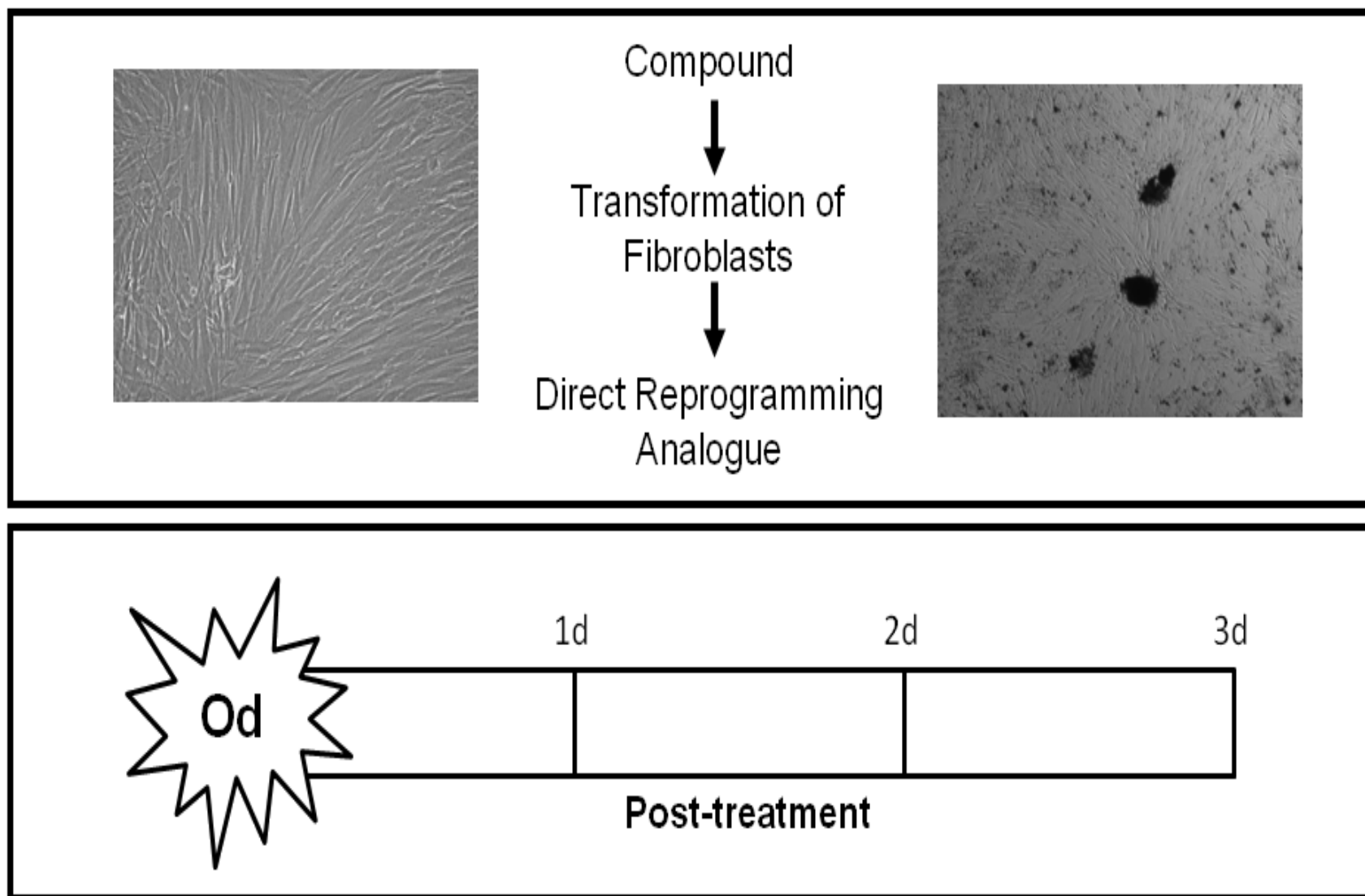
Figure A.1. The mechanism disruption approach. Images at top show in vitro cell culture before (left) and after (right) treatment with compound, which disrupts a major cellular mechanism. Measurements taken at time intervals specified on

Figure A.1 (con't). the scale at center. Graphs at bottom show generalized decay dynamics for several genes, which can be measured as associated with transcriptional and translational processes.
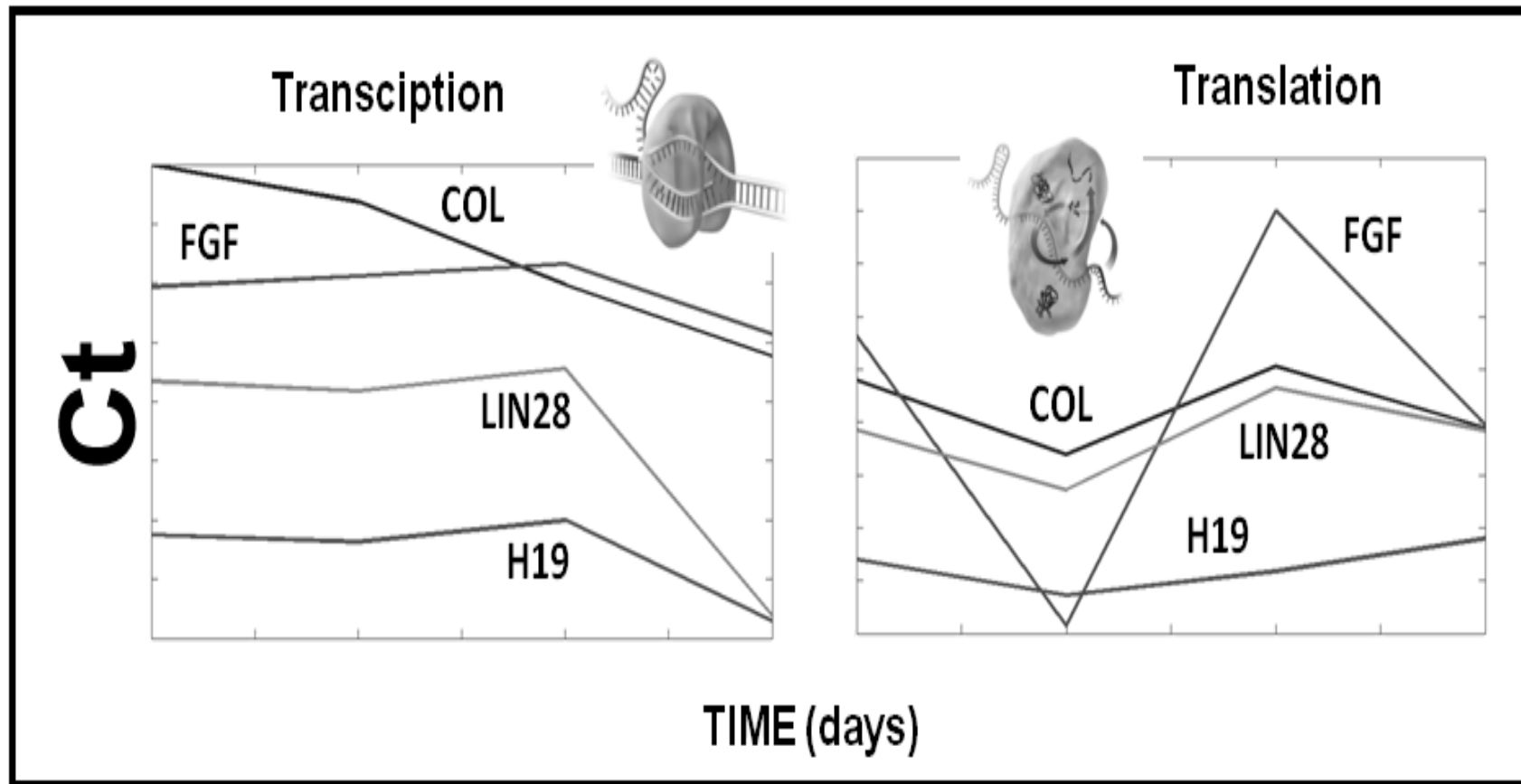
C.4 CONCLUSION

The systems modeling results reveal gene- and treatment-dependent fluctuations in course-grained control mechanisms with respect to time. We stress that these are course-grained because they are inferential in nature. While decay mechanisms are mostly responsible for controlling the rate of degradation, it has been found that at least one other pathway (miRNAs and siRNAs) also control and stabilize translation (Valencia-Sanchez et al., 2006; Fabian, Sonenberg, Filipowicz, 2010). This can occur either in tandem or independently, which makes the comparison of transcriptome and translatome a useful tool for understanding the nature and progression of mRNA dynamics.

First-order DDE model, general case

A

$$\frac{dy}{dt} = f[x(t - \tau)]$$

Aggregation, nonlinear decay are second-order (e.g. polynomial) effects

$$\tau = \begin{cases} -ax^2 + ax + b, & \text{aggregation} \\ ax^2 + ax - b, & \text{nonlinear decay} \\ -ax + b, & \text{decay} \end{cases}$$
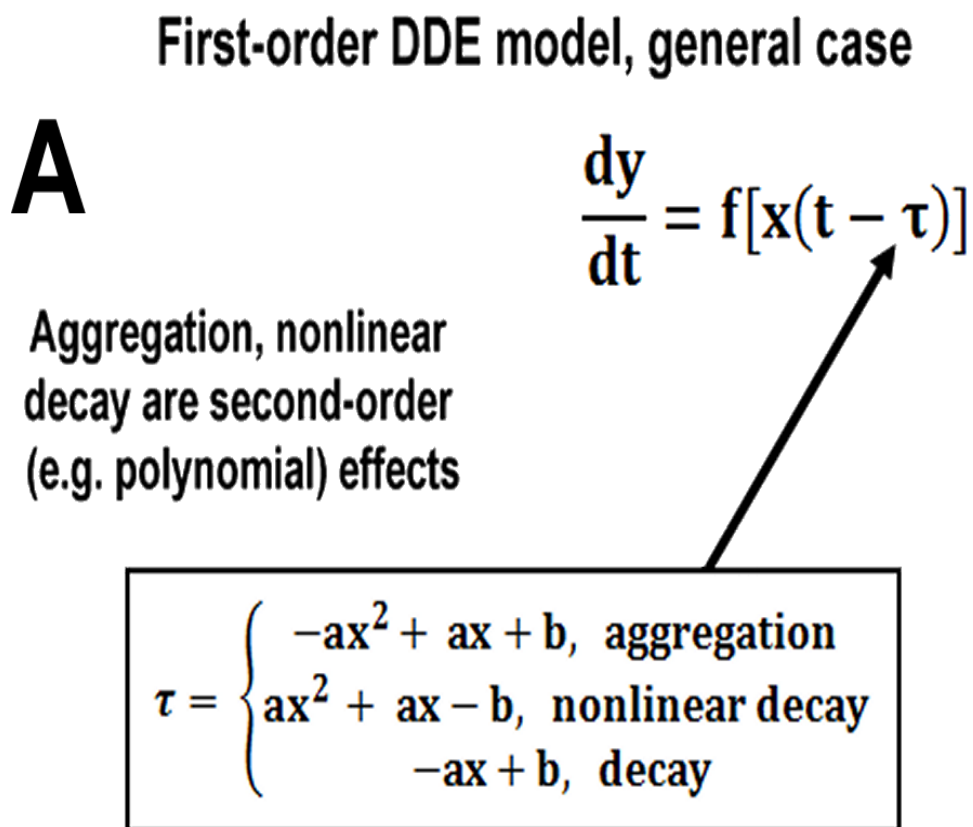
Figure A.2. Schematic demonstrating the concept of delays as it relates to transcriptional and translational processes. A) the delay model, expressed as a

Figure A.2 (con't). conditional discrete dynamical equation (DDE). B) pseudo-data demonstrating the dynamics of linear decay, aggregation, and nonlinear decay. C) interpreting the observed nonlinearities as a signature of delays in specific biological mechanism (in this case, transcription and translation).
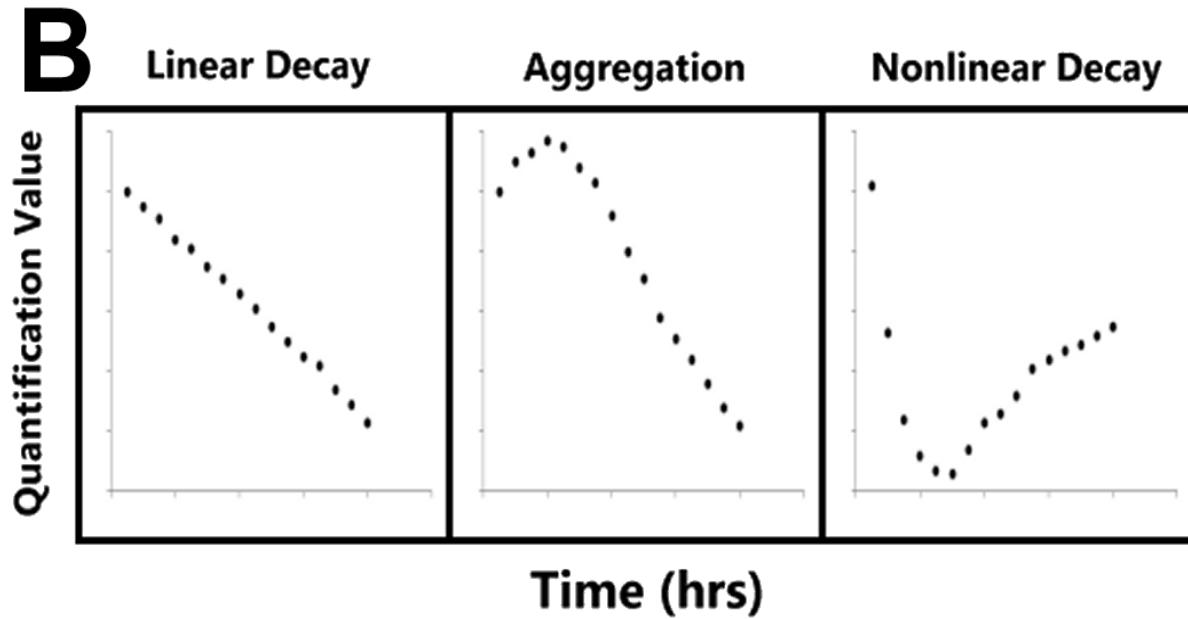
Figure A.2 (con't).
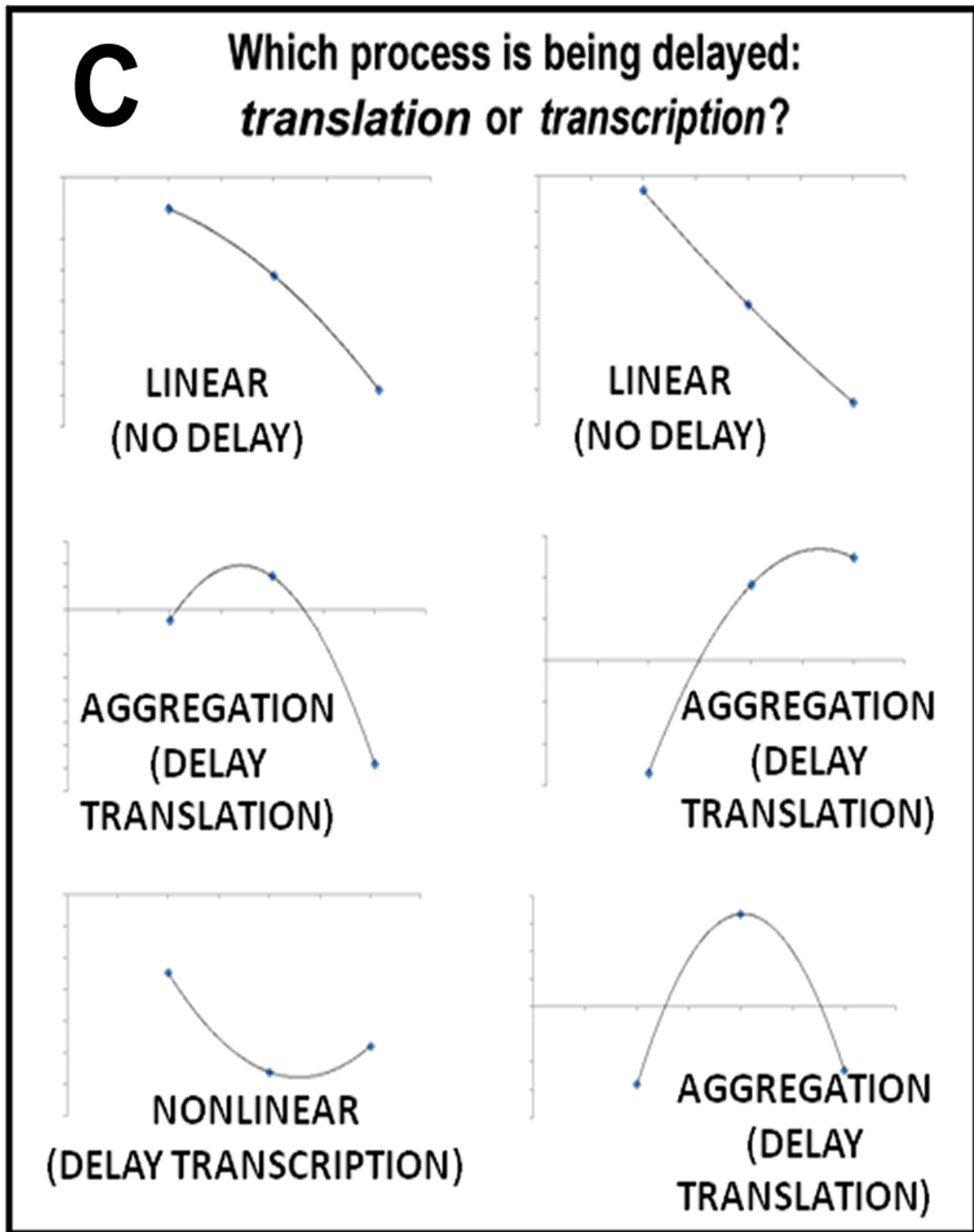
APPENDIX D: ADVANCES IN THE ARTIFICIAL LIFE OF DIRECT CELLULAR REPROGRAMMING

One future direction for direct cellular reprogramming research, particularly as it relates to the nervous system, is to use artificial life approaches to model "possible" scenarios. This has been summarized in the form of a cellular automata model presented as short presentations at the Dynamics Days conference in 2010 (Appendix D.1, "Dynamical Cellular Reprogramming Using Excitable Cellular Automata") and 2012 (Appendix D.2, "Dynamical Cellular Encodings for Exploring Cellular Reprogramming").

D.1 DYNAMICAL CELLULAR REPROGRAMMING USING EXCITABLE CELLULAR AUTOMATA

Cellular reprogramming is a process by which differentiated cells such as fibroblasts are converted into pluripotent cells that possess many characteristics of embryonic stem cells. This process occurs on the time scale of days to weeks, and involves phenotypic conversion given transgenic delivery of four transcription factors. Induced pluripotent cells, like naturally-occurring stem cells, can become any of the endoderm, mesoderm, or ectoderm-derived tissue types.

Reprogramming is spatially heterogeneous. One open problem in this area concerns how some cells are more reprogrammable than others. For example, reprogramming efficacy in response to a transgenic signal has been limited to 10% of cells in culture (Wernig et al., 2007), with various environmental manipulations and knockouts of key genes yielding up to 100% efficiency (Utikal et al., 2009). These findings suggest an underlying critical process within the cell reliant on population dynamics. Spatial heterogeneity is related to kinetics. The *in silico* approach presented here takes into account an intermediate scale of analysis, that of colony formation and

functional populations that might form *in vitro*. In order to approximate cellular reprogramming, we must consider a discrete model that approximates critically-organized biological kinetics, with cellular components that exhibit both transformational ability and considerable complexity in both space and time. Cellular automata (CA) models meet these requirements and have been used previously to model living systems (Ermentrout and Edelstein-Keshet, 1993; D'Inverno, 2006).

From Cell Culture to CA Model

An abstraction of reprogramming cells in culture to a series of heuristics (since many factors in reprogramming are empirically unexplored) requires two rules and two parametric constraints. Instances of each will be provided. These rules and constraints are attributes of a 2-D cellular automata model that treats the biological problem as a discrete, collective phenomenon that unfolds on a surface.

Parametric constraints vs. rules

This model utilizes a combination of rule-based and parameter-related constraint-based techniques. Simple rule-based models have previously been used to model the dynamics of stem cells *in vivo* (see Meissner, Wernig, and Jaenisch, 2002). However, processes in natural systems such as cellular reprogramming operate by satisfying a finite set of constraints (Wolfram, 2002):

Constraint #1: Infectability. How many cells of a specific cell type are infected under certain conditions?

Constraint #2: Autonomous factors. Are some cells in a population more likely to reprogram based on genetic or other internal factors?

Rule #1: Contact inhibition. If cells are packed together and unable to divide freely prior to infection, will the number of cells reprogrammed be reduced?

Rule #2: Substrate. What is the nature of the local cell population and surface on which cells are growing?

Constraint #1 (infectability), reprogramming *in vitro* as excitable media

A modified form of the Fitzhugh-Nagumo (Ilachinski, 2001; Izhikevich and Fitzhugh, 2009) model, used in this case to simulate discrete excitability, can be used to understand reprogramming of cells in parallel.

- cells in a culture are "infected" in waves (reprogramming factors are introduced to x,y locations).

- peaks of those waves = instances of successful infection. Stochastic process (Hanna et al., 2009), or subpopulation of cells spontaneously emerge above threshold (1/n cells at every timestep).

- interactions with neighboring cells, shifts in CA substrate = cells that remain alive, replicate to form colony or form colony with neighboring cells (cascading behavior).

Constraint #2 (autonomous factors), determining "stemness" using a secondary stimulus

One issue in induced pluripotent and stem cell biology research is "stemness" (Mikkers and Frisen, 2005), or what makes a cell pluripotent. There are multiple attributes of the internal cellular state which can be approximated using a dynamical representation of the molecular mechanisms behind pluripotency. In MacArthur et al. (2008), the internal state determines the pluripotency of the cell by using a secondary stimulus to induce transcriptional noise in key genes related to pluripotency. This is

thought to drive cell conversion, as chemicals like alkaline phosphotase can affect conversion rates in biological contexts.

The dynamics of a secondary stimulus is similar to those characterized by the Fitzhugh-Nagumo model, except that secondary stimuli act only on previously infected cells. Secondary stimuli act on compact genotypic representations in every automaton cell. This representation approximates key genes for pluripotency. Stimulating elements in the genotypic representation cause small- and large-magnitude outputs, mimicking the critical organization of genetic circuits in the biological reprogramming of cells (e.g. four factors activate many downstream genes).

In the CA model, introduction of the secondary stimulus operates in tandem with infectability to determine how likely an individual cell is to form a colony. The goal is to form structures typical of CA environments (e.g. gliders, see Ilachinski, 2001) to characterize the great diversity of colony morphologies found in actual cell cultures. Needed for further work are a better characterization of the underlying expression profile of pluripotency genes during reprogramming and a better working knowledge of differential response by cells in culture to drug inducement and paracrine signaling.

Rule #1 (contact inhibition), "game of life"-style rules

In both D'Inverno (2006) and Adimy (2006), a general set of rules regarding the maintenance and stability of pluripotency are derived. These rules are based on the following biological properties:

1) cell density as a function of cell division, death, and paracrine signaling.

2) phases of cell cycle as a function of stability (cells in S phase more likely to change state than those in G phase).

The state of these attributes in neighboring cells largely determines the state of a focal cell (position 0, 0) in a neighborhood. They also act in a feedforward manner with regard to infection and the secondary stimulus. For example, while the parametric constraints set up what the states are, the CA rules produce the emergent output. This is especially important for colony formation, as cells must act collectively to form viable colonies.

Rule #2 (substrate), higher-dimensional approximation

Unique to this CA approach is the use of "sliding" Von Neumann and Moore neighborhoods (Adimy, 2006) for purposes of simulating an adaptive and transformative surface. A "sliding" neighborhood involves the fusion and/or fission of cellular units that behave similar to a change in state and modify the local neighborhood accordingly. The "sliding" neighborhoods account for some *in vitro* idiosyncrasies in addition to a finite time window. Based on specific rates of infection and conversion in biological cells, it is hypothesized that cellular reprogramming is partially governed by collective environmental factors. In these examples, we have only considered a monolayer of cells. However, reprogramming cell cultures can form bilayers or even exhibit folding. Nevertheless, the excitable CA models presented here may provide "building blocks" (Schiff, 2007) for novel applications and reprogramming in novel conditions.

Details of Fitzhugh-Nagamo Model

A typical application of the Fitzhugh-Nagamo model is to characterize the electrical coupling of neurons. The model consists of coupled equations that have been adapted for potentials cellular behaviors that govern the respecification of phenotype

during cellular reprogramming. In this application, the potential is not electrical but the production of a changed phenotype.

$$V_n = \frac{V_m - V_m(d)}{d} - W_m + I$$

[5]

$$W_m = C_1(V_n + C_2) - C_1(W_n)$$

The symbolic variables are as follows: V is the cellular potential, W is the recovery variable, I is the magnitude of stimulus, C_1 and C_2 are constants related to the cellular potential, m and n are neighboring cells, and d is the dimensionality of cell culture (for most applications, either 2 or 3).

D.2 DYNAMICAL CELLULAR ENCODINGS FOR EXPLORING CELLULAR REPROGRAMMING

In cell biology, the concept of stemness is a heuristic used to characterize the properties that allow a stem-like cell to differentiate into multiple cell types. There have been attempts to define stemness using transcriptional, nuclear receptor (Chang and Stanford, 2008; Jeong and Mangelsdorf, 2009), and other types of data. However, the excitability inherent in this process (Kalmar et al., 2009) has not been addressed extensively at the population level. Recently, a process called cellular reprogramming (Meissner, Wernig, and Jaenisch, 2007; Ludwig et al., 2006) has allowed us to transform cells from a differentiated to a pluripotent state. The delivery of just four transcription factors to a population of differentiated cells results in a small subpopulation of these cells being transformed into pluripotent cell types and the formation of viable colonies (Meissner, Wernig, and Jaenisch, 2007).

Recent attempts to capture the underlying dynamics of cellular reprogramming have utilized cellular automata, stochastic dynamical (Hanna et al., 2009), and hybrid Cayley tree-dynamical models (Artyomov, Meissner, and Chakraborty, 2010). What is missing from this research is a model that can describe how self-organization contributes to the ability of a cell to transform both at the single-cell genomic level and the level of intercellular interactions within a cell population.

Dynamical Cellular Encodings

Dynamical cellular encodings may be a way to characterize the complex set of interactions and regulatory cascades that characterize differentiation, reprogramming, and the maintenance of stemness (Bilodeau and Savageau, 2006). The dynamical cellular encoding is an abstraction at the level of both a cell's genotype and a cell population. Incorporating elements of cellular automata (Dobrushin, Kryukov, and Toom, 1990), genetic algorithms (GAs), and epidemiological models, dynamical cellular encodings are proposed as a novel solution to the stochastic nature of reprogramming and a way to demonstrate how stemness is maintained despite environmental challenges. The hybrid algorithm and other details are shown in Figure A.1. Briefly, a population of automata (approximations of cells) is infected. A proportion of these automata become carriers of the virus. In these carriers, the viral element becomes a trigger for additional changes in the cell.

In this way, each cell's genotype is expressed in accordance with empirically observed slow kinetics (Hanna et al., 2009). This self-transformative process also allows for derivative effects such as both transient and longer-lasting responses. Each functional unit is analogous to a gene on a chromosome used in genetic algorithm

design (Vose, 1999). Each automata can have any number of functional units, each of which are switched on and off by the presence or absence of a viral element. The main features of each functional unit involve the combinatorial expression of gene products and the existence of a switch and integrator element downstream from the initial switch. The switch and integrator operates on the response profile given recombination of the functional unit using a series of simple calculus operations. There are also paracrine signaling effects that are transmitted from one automaton to another. These signals can be approximated using nearest-neighbor interaction rules on a two-dimensional topology. This reinforces the establishment and/or maintenance of pluripotency.

REFERENCES

REFERENCES

Abrams, P.A. The Evolution of Predator-Prey Interactions: Theory and Evidence. Annual Review of Ecology and Systematics, 31, 79-105 (2000).

Adimy, M. Stability of limit cycles in a pluripotent stem cell dynamics model. Chaos, Solitons and Fractals, 27, 1091–1107 (2006).

Alfano, F.D. A stochastic model of cellular transformation and its relevance to chemical carcinogenesis. Mathematical Biosciences, 149(1), 95-106 (1998).

Alicea, B., Murthy, S., Keaton, S.A., Cobbett, P., Cibelli, J.B., and Suhr, S.T. Defining phenotypic respecification diversity using multiple cell lines and reprogramming regimens. Stem Cells and Development, doi:10.1089/scd.2013.0040 (2013).

Alicea, B. Natural Variation and Neuromechanical Systems. Cogprints: http://cogprints.org/6698/ (2009).

Allen, L.J.S. An introduction stochastic processes with applications to biology. Pearson, London (2003). Chapter 1.

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

Angka, H.E., Geddes, A.J., and Kablar, B. Differential survival response of neurons to exogenous GDNF depends on the presence of skeletal muscle. Developmental Dynamics, 237(11), 3169-3178 (2008).

Appasani, K. MicroRNAs: from basic science to disease biology. Cambridge University Press, Cambridge, UK (2008).

Archer, S., Queiroz, R., Stewart, M., and Clayton, C. Trypanosomes as a model to investigate mRNA decay pathways. Methods in Enzymology, Volume 448, 359-377 (2008).

Artyomov, M.N., Meissner, A., and Chakraborty, A.K. A Model for Genetic and Epigenetic Regulatory Networks Identifies Rare Pathways for Transcription Factor Induced Pluripotency. PLoS Computational Biology, 6(5), e1000785 (2010).

Avery, K., Avery, S., Shepherd, J., Heath, P.R, and Moore, H. Sphingosine-1-phosphate mediates transcriptional regulation of key targets associated with survival, proliferation, and pluripotency in human embryonic stem cells. Stem Cells and Development, 17(6), 1195-1205 (2008).

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

Awe, J. and Byrne, J.A. Identifying candidate oocyte reprogramming factors using cross-species global transcriptional analysis. Cellular Reprogramming, 15(2), 126-133 (2013).

Azuara, V., Perry, P., Sauer, S., Spivakov, M., Jørgensen, H.F., John, R.M., Gouti, M., Casanova, M., Warnes, G., Merkenschlager, M., and Fisher, A.G. Chromatin signatures of pluripotent cell lines. Nature Cell Biology, 8, 532–538 (2006).

Balazsi, G., van Oudenaarden, A., and Collins, J.J. Cellular decision making and biological noise: from microbes to mammals. Cell, 144(6), 910-925 (2011).

Bak, P., Tang, C., and Wiesenfeld, K. Self-organized criticality. Physics Review A, 38, 364–374 (1988).

Bak, P., Tang, C., and Wiesenfeld, K. Self-organized criticality: an explanation of the 1/f noise. Physics Review Letters, 59, 381-384 (1987).

Barenco, M., Brewer, D., Papouli, E., Tomescu, D., Callard, R., Stark, J., and Hubank, M. Dissection of a complex transcriptional response using genome-wide transcriptional modeling. Molecular Systems Biology, 5, 327 (2009).

Bay, S.D., Chrisman, L., Pohorille, A., and Shrager, J. Temporal aggregation bias and inference of causal regulatory networks. Journal of Computational Biology, 11(5), 971-985 (2004).

Beggs, J.M. and Plenz, D. Neuronal avalanches in neocortical circuits. Journal of Neuroscience, 23, 11167-11177 (2003).

Bernew, B.J. and Straubb, J.E. Novel methods of sampling phase space in the simulation of biological systems. Current Opinion in Structural Biology, 7(2), 181-189 (1997).

Bhattacharya, S., Zhang, Q., and Andersen, M.E. A deterministic map of Waddington's epigenetic landscape for cell fate specification. BMC Systems Biology, 5, 85 (2011).

Bilodeau, M. and Sauvageau, G. Uncovering stemness. Nature Cell Biology, 8(10), 1048-1049 (2006).

Bismuth, K. and Relaix, F. Genetic regulation of skeletal muscle development. Experimental Cell Research, 316(18), 3081-3086 (2010).

Blasi, A., Martino, C., Balducci, L., Saldarelli, M., Soleti, A., Navone, S.E., Canzi, L., Cristini, S., Invernici, G., Parati, E.A., and Alessandri, G. Dermal fibroblasts display similar phenotypic and differentiation capacity to fat-derived mesenchymal stem cells, but differ in anti-inflammatory and angiogenic potential. Vascular Cell, 3, 5 (2011).

Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., Gnirke, A., Eggan, K., and Meissner, A. Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell, 144, 439–452 (2011).

Bouard, D., Alazard-Dany, N., and Cosset, F-L. Viral vectors: from virology to transgene expression. British Journal of Pharmacology, 157, 153–165 (2009).

Bolouri, H. and Davidson, E.H. Transcriptional regulatory cascades indevelopment: Initial rates, not steady state, determine network kinetics. PNAS, 100(16), 9371-9376 (2003).

Brambrink, T., Foreman, R., Welstead, G.G., Lengner, C.J., Wernig, M., Suh, H., and Jaenisch, R. Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. Cell Stem Cell, 2, 151–159 (2008).

Braun, T., Buschhausen-Denker, G., Bober, E., Tannich, E., and Arnold, H.H. A novel human muscle factor related to but distinct from *MyoD1* induces myogenic conversion in 10T1/2 fibroblasts. EMBO Journal, 8, 701-709 (1989).

Buganim, Y., Faddah, D.A., Cheng, A.W., Itskovich, E., Markoulaki, S., Ganz, K., Klemm, S.L., van Oudenaarden, A., and Jaenisch, R. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell, 150(6), 1209-1222 (2012).

Byrne, J.A. Global transcriptional analysis of oocyte-based and factor-based nuclear reprogramming in the nonhuman primate. Cellular Reprogramming, 13(6), 473-481 (2011).

Byrne, J.A., Pedersen, D.A., and Clepper, L.L. Producing primate embryonicstem cells by somatic cell nuclear transfer. Nature, 450, 497–502 (2007).

Caiazzo, M., Dell'Anno, M.T., Dvoretskova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., Russo, G., Carninci, P., Pezzoli, G., Gainetdinov, R.R., Gustincich, S., Dityatev, A., and Broccoli, V. Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. Nature, 476, 224-227 (2011).

Campbell, K.H., McWhir, J., Ritchie, W.A. and Wilmut, I. Sheep cloned by nuclear transfer from a cultured cell line. Nature, 380, 64-66 (1996).

Carter, D., Chakalova, L., Osborne, C.S., Dai, Y-F., and Fraser, P. Long-range chromatin regulatory interactions *in vivo*. Nature Genetics, 32, 623-626 (2002).

Centlivre, M., Legrand, N., Steingrover, R., van der Sluis, R., Grijsen, M.L., Bakker, M., Jurriaans, S.,Berkhout, B., Paxton, W.A., Prins, J.M., and Pollakis, G. Altered dynamics and differential infection profiles of lymphoid and myeloid cell subsets during acute and chronic HIV-1 infection. Journal of Leukocyte Biology, 89, 785-795 (2011).

Challen, G.A. and Little, M.H. A side order of stem cells: the SP phenotype. Stem Cells, 24, 3-12 (2006).

Chambers, S.M. and Studer, L. Cell fate plug and play: direct reprogramming and induced pluripotency. Cell, 145, 827-830 (2011).

Chan, E.M., Ratanasirintrawoot, S., Park, I.H., Manos, P.D., Loh, Y.H., Huo, H., Miller, J.D., Hartung, O., Rho, J., Ince, T.A., Daley, G.Q., and Schlaeger, T.M. Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. Nature Biotechnology, 27(11), 1033-1037 (2009).

Chang, R., Shoemaker, R., and Wang, W. Systematic search for recipes to generate induced pluripotent stem cells. PLoS Computational Biology, 7(12), e1002300 (2011).

Chang, E.A., Beyhan, Z., Yoo, M.S., Siripattarapravat, K., Ko, T., Lookingland, K.J., Madhukar, B.V. and Cibelli, J.B. Increased cellular turnover in response to fluoxetine in neuronal precursors derived from human embryonic stem cells. International Journal of Developmental Biology, 54, 707-715 (2009).

Chang, W.Y. and Stanford, W.L. Translational Control: a new dimension in embryonic stem cell network analysis. Cell Stem Cell, 2, 410-412 (2008).

Chang, H.Y., Chi, J-T., Dudoit, S., Bondre, C., van de Rijn, M., Botstein, D., and Brown, P.O. Diversity, topographic differentiation, and positional memory in human fibroblasts. PNAS, 99(20), 12877-12882 (2002).

Chechik, G. and Koller, D. Timing of gene expression responses to environmental changes. Journal of Computational Biology, 16, 279–290 (2009).

Chen, J.X., Krane, M., Deutsch, M.A., Wang, L., Rav-Acha, M., Gregoire, S., Engels, M.C., Rajarajan, K., Karra, R., Abel, E.D., Wu, J.C., Milan, D., and Wu, S.M. Inefficient reprogramming of fibroblasts into cardiomyocytes using *Gata4*, *Mef2c*, and *Tbx5*. Circulation Research, 111, 50-55 (2012).

Chen, C-Y.A., Ezzeddine, N., and Shyu, A-B. Messenger RNA half-life measurements in Mammalian cells. Methods in Enzymology, 448, 335-357 (2008).

Chin, M.H., Pellegrini, M., Plath, K., and Lowry, W.E. Molecular analyses of human induced pluripotent stem cells and embryonic stem cells. Cell Stem Cell, 7, 263–269 (2010).

Chin, M.H., Mason, M.J., Xie, W., Volinia, S., and Singer, M. Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. Cell Stem Cell, 5, 111–123 (2009).

Chipev, C.C. and Simon, M. Phenotypic differences between dermal fibroblasts from different body sites determine their responses to tension and TGFβ1. BMC Dermatology, 2, 13 (2002).

Cibelli, J.B., Stice, S.L., Golueke, P.J., Kane, J.J., Jerry, J., Blackwell, C., Ponce de Leon, F.A., and Robl, J.M. Cloned transgenic calves produced from nonquiescent fetal fibroblasts. Science, 280, 1256-1258 (1998).

Cinquin, O. and Demongeot, J. High-dimensional switches and the modelling of cellular differentiation. Journal of Theoretical Biology, 233, 391-411 (2005).

Colman, A. and Dreesen, O. Induced pluripotent stem cells and thestability of the differentiated state. EMBO Reports, 10(7), 714-721 (2009).

Consul, P.C. Generalized Poisson Distributions: properties and applications. Marcel Dekker, New York (1989). Chapter 1.

Corin, R.E., Guller, S., Wu, K.Y., and Sonenberg, M. Growth hormone and adipose differentiation: growth hormone-induced antimitogenic state in 3T3-F442A preadipose cells. PNAS, 87(19), 7507-7511 (1990).

Cox, J.L. and Rizzino, A. Induced pluripotent stem cells: what lies beyond the paradigm shift. Experimental Biology and Medicine, 235, 148-158 (2010).

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

Day, K., Shefer, G., Shearer, A., and Yablonka-Reuveni, Z. The depletion of skeletal muscle satellite cells with age is concomitant with reduced capacity of single progenitors to produce reserve progeny. Developmental Biology, 340, 330-343 (2010).

Day, K., Shefer, G., Richardson, J.B., Enikolopov, G., and Yablonka-Reuveni, Z. Nestin-GFP reporter expression defines the quiescent state of skeletal muscle satellite cells. Developmental Biology, 304, 246-259 (2007).

D'Amour, K.A. and Gage, F.H. Genetic and functional differences between multipotent neural and pluripotent embryonic stem cells. PNAS USA, 100, S1, 11866-11872 (2003).

D'Haeseleer, P., Wen, X., Fuhrman, S., and Somogyi, R.   Linear modeling of mRNA expression levels during CNS development and injury. Pacific Symposium on Biocomputing, 41–52 (1999).

Didiera, P., Weiss, E., Sibler, A-P., Philibert, P., Martineau, P., Bigot, J-Y., and Guidoni, L. Femtosecond spectroscopy probes the folding quality of antibody fragments expressed as GFP fusions in the cytoplasm. Biochemical and Biophysical Research Communications, 366(4), 878-884 (2008).

Ding, S. and Wang, W.   Recipes and mechanisms of cellular reprogramming: a case study on budding yeast *Saccharomyces cerevisiae*. BMC Systems Biology, 5, 50 (2011).

D'Inverno, M., Theise, N., and Prophet, J.   Mathematical modeling of stem cells: a complexity primer for the stem-cell biologist. In Tissue Stem Cells, Potten, C.S. et al. eds. Taylor and Francis, New York (2006).

Dobrushin, R.L., Kryukov, V.I., and Toom, A.L.   Stochastic cellular systems: ergodicity, memory, and morphogenesis. Manchester University Press, Manchester, UK (1990).

Duinsbergen, D., Eriksson, M., t'Hoen, P.A., Frisen, J., and Mikkers, H.   Induced pluripotency with endogenous and inducible genes. Experimental Cell Research, 314(17), 3255-3263 (2008).

Edmondson, D.G. and Olson, E.N.   A gene with homology to the myc similarity region of *MyoD1* is expressed during myogenesis and is sufficient to activate the muscle differentiation program. Genes and Development, 3, 628-640 (1989).

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

Egli, D., Birkhoff, G., and Eggan, K.   Mediators of reprogramming: transcription factors and transitions through mitosis. Nature Reviews Molecular Cellular Biology, 9, 505–516 (2008).

Ellis, P., Fagan, B.M., Magness, S.T., Hutton, S., Taranova, O., Hayashi, S., McMahon, A., Rao, M., and Pevny, L.   *SOX2*, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. Developmental Neuroscience, 26, 148-165 (2004).

Ermentrout, G.B. and Edelstein-Keshet, L. Cellular automata approaches to biological modeling. Journal of Theoretical Biology, 160, 97-133 (1993).

Erwin, D.H. and Davidson, E.H.   The evolution of hierarchical gene regulatory networks. Nature Reviews Genetics, 10, 141-148 (2009).

Esteban, M.A., Xu, J., Yang, J., Peng, M., Qin, D., Li, W., Jiang, Z., Chen, J., Deng, K., Zhong, M., Cai, J., Lai, L., and Pei, D. Generation of induced pluripotent stem cell lines from Tibetan miniature pig. Journal of Biological Chemistry, 284, 17634-17640 (2009).

Ezashi, T., Telugu, B.P., Alexenko, A.P., Sachdev, S., Sinha, S., and Roberts, R.M. Derivation of induced pluripotent stem cells from pig somatic cells. PNAS USA, 106, 10993-10998 (2009).

Fabian, M.R., Sonenberg, N., and Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. Annual Review of Biochemistry, 79, 351–379 (2010).

Feala, J.D., Cortes, J., Duxbury, P.M., Piermarocchi, C., McCulloch, A.D., and Paternostro, G. Systems approaches and algorithms for discovery of combinatorial therapies. Systems Biology and Medicine, 2, 127 (2010).

Feng, B., Ng, J., Heng, J., and Ng, H. Molecules that promote or enhance reprogramming of somatic cells to induced pluripotent stem cells. Cell Stem Cell, 4, 301-312 (2009).

Ferrell, J. E. and Machleder, E.M. The biochemical basis of an all-or-none cell fate switch in *Xenopus* oocytes. Science 280, 895–898 (1998).

Fleming, J.M., Long, E.L., Ginsburg, E., Gerscovich, D., Meltzer, P.S., and Vonderhaar, B.K. Interlobular and intralobular mammary stroma: genotype may not reflect phenotype. BMC Cell Biology, 9, 46 (2008).

Furusawa, C. and Kaneko, K. (2009) Chaotic expression dynamics implies pluripotency: when theory and experiment meet. Biology Direct, 4, 17 (2009).

Garneau, N.L., Wilusz, J., and Wilusz, C.J. The highways and byways of mRNA decay. Nature Reviews Molecular Cell Biology, 8, 113-126 (2007).

Ghosh, S., Matsuoka, Y., Asai, Y., Hsin, K.Y., and Kitano, H. Software for systems biology: from tools to integrated platforms. Nature Reviews Genetics, 12, 821-832 (2011).

Gianchandani, E.P., Oberhardt, M.A., Burgard, A.P., Maranas, C.D., and Papin, J.A. Predicting biological system objectives de novo from internal state measurements. BMC Bioinformatics, 9, 43 (2008).

Gokhale, P.J. and Andrews, P.W. New insights into the control of stem cell pluripotency. Cell Stem Cell, 2(1), 4-5 (2008).

Grabherr, M.G., Haas, B.J., Moran Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., Chen, Z., Mauceli, E., Hacohen, N.,

Gnirke, A., Rhind, N., di Palma, F., Birren, B.W., Nusbaum, C., Lindblad-Toh, K., Friedman, N., and Regev, A. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology, 29(7), 644-652 (2011).

Green, H. and Kehinde, O. Spontaneous heritable changes leading to increased adipose conversion in 3T3 cells. Cell, 7, 105-113 (1976).

Grunwald, D., Singer, R.H., and Czaplinski, K. Cell biology of mRNA decay. Methods in Enzymology, 448, 553-577 (2008).

Guenther, M.G., Frampton, G.M., Soldner, F., Hockemeyer, D., Mitalipova, M., Jaenisch, R., and Young, R.A. Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells. Cell Stem Cell, 7, 249–257 (2010).

Gurdon, J. and Murdoch, A. Nuclear transfer and iPS may work best together. Cell Stem Cell, 2, 135-136 (2008).

Gurdon, J.B. and Byrne, J.A. The first half-century of nuclear transplantation. PNAS, 100, 8048-8052 (2003).

Hamill, O.P., Marty, A., Neher, E., Sakmann, B., and Sigworth, F.J. Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. Pflugers Archives, 391, 85-100 (1981).

Hanna, J.H., Saha, K., and Jaenisch, R. Pluripotency and Cellular Reprogramming: Facts, Hypotheses, Unresolved Issues. Cell, 143(4), 508-525 (2010).

Hanna, J., Saha, K., Pando, B., van Zon, J., Lengner, C.J., Creyghton, M.P., van Oudenaarden, A. and Jaenisch, R. Direct cell reprogramming is a stochastic process amenable to acceleration. Nature, 462(7273), 595-601 (2009).

Hanna, J., Markoulaki, S., Schorderet, P., Carey, B.W., Beard, C., Wernig, M., Creyghton, M.P., Steine, E.J., Cassady, J.P., Foreman, R., Lengner, C.J., Dausman, J.A., and Jaenisch, R. Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. Cell, 133, 250–264 (2008).

Hansis, C., Barreto, G., Maltry, N. and Niehrs, C. Nuclear reprogramming ofhuman somatic cells by *Xenopus* egg extract requires BRG1. Current Biology, 14, 1475–1480 (2004).

Hausdorff, J.M. and Peng, C-K. Multiscaled randomness: a possible source of 1/f noise in biology. Physical Review E, 54(2), 2154-2157 (1996).

Hobert, O. Regulatory logic of neuronal diversity: Terminal selector genes and selector motifs. PNAS USA, 105(51), 20067-20071 (2008).

Hochedlinger, K. and Plath, K. Epigenetic reprogramming and induced pluripotency. Development, 136, 509–523 (2009).

Hochedlinger, K. and Jaenisch, R. Nuclear reprogramming and pluripotency.Nature, 441, 1061-1067 (2006).

Hochedlinger, K., Blelloch, R., Brennan, C., Yamada, Y., Kim, M., Chin, L., and Jaenisch, R. Reprogramming of a melanoma genome by nuclear transplantation. Genes and Development, 18(15), 1875–1885 (2004).

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell, 3, 346-353 (2008).

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

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

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D. Cell fates as high-dimensional attractor states of a complex gene regulatory network. Physical Review Letters, 94, 128701 (2005).

Ideker, T., Oxier, O., Schwikowski, B., and Siegel, A.F. Discovering regulatory and signalling circuits in molecular interaction networks. Bioinformatics, 18(1), S233-S240 (2002).

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

Ilachinski, A. Cellular automata: a discrete universe. World Scientific, Singapore (2001).

Ioannidis, J.P.A. Why most published research findings are false. PLoS Medicine, 2(8), e124 (2005).

Isaacs, F.J., Blake, W.J., and Collins, J.J. Signal processing in single cells. Science, 307, 1886-1888 (2005).

Ivanova, N., Dobrin, R., Lu, R., Kotenko, I., Levorse, J., DeCoste, C., Schafer, X., Lun, Y., and Lemischka, I.R. Dissecting self-renewal in stem cells with RNA interference. Nature 442, 533–538 (2006).

Izhikevich, E.M. and Fitzhugh, R. Fitzhugh-Nagumo Model. Scholarpedia, 1(9), 1349 (2009).

Jaenisch, R. Nuclear cloning and direct reprogramming: the long and the short path to Stockholm. Cell Stem Cell, 11(6), 744-747 (2012).

Jarosz, D.F., Taipale, M., and Lindquist, S. (2010). Protein homeostasis and the phenotypic manifestation of genetic diversity: principles and mechanisms. Annual Review of Genetics, 44, 189-216.

Jeong, Y. and Mangelsdorf, D.J. Nuclear receptor regulation of stemness and stem cell differentiation. Experimental and Molecular Medicine, 41(8), 525-537 (2009).

Judson, R.L., Babiarz, J.E., Venere, M., and Blelloch, R. Embryonic stem cell-specific microRNAs promote induced pluripotency. Nature Biotechnology, 27(5), 459-461 (2009).

Kalmar, T., Lim, C., Hayward, P., Munoz-Descalzo, S., Nichols, J., Garcia-Ojalvo, J., and Martinez Arias, A. Regulated fluctuations in nanog expression mediate cell fate decisions in embryonic stem cells. PLoS Biology, 7(7), e1000149 (2009).

Karplus, M.   Protein dynamics: from femtoseconds to milliseconds. Journal of Molecular Structure: THEOCHEM, 463(1-2), 3-4 (1999).

Kelso, J.A.S. Haken-Kelso-Bunz Model. Scholarpedia, 3(10), 1612 (2008).

Khan, D.R., Dube, D., Gall, L., Peynot, N., Ruffini, S., Laffont, L., Le Bourhis, D., Degrelle, S., Jouneau, A., and Duranthon, V. Expression of pluripotency master regulators during two key developmental transitions: EGA and early lineage specification in the bovine embryo. PLoS One, 7(3), e34110 (2012).

Kiessling, A.A, Bletsa, R., Desmarais, B., Mara, C., Kallianidis, K., and Loutradis, D. Evidence That Human Blastomere Cleavage Is Under Unique Cell Cycle Control. Journal of Assisted Reproduction and Genetics, 26, 187–195 (2009).

Kirkwood, T.B.L., Rosenberger, R.F., and Galas, D.J.   Accuracy in Molecular Processes: its control and relevance to living systems. Chapman and Hall, New York (1986). Chapters 6 and 7.

Kitada, M., Wakao, S., and Dezawa, M. Muse cells and induced pluripotent stem cell: implication of the elite model. Cellular and Molecular Life Science, 69, 3739-3750 (2012).

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., and Jaenisch, R.   Functional integration of

dopaminergic neurons directly converted from mouse fibroblasts. Cell Stem Cell, 9, 413-419 (2011).

Kim, K., Doi, A., Wen, B., Ng, K., Zhao, R., Cahan, P., Kim, J., Aryee, M.J., Ji, H., Ehrlich, L.I.R., Yabuuchi, A., Takeuchi, A., Cunniff, K.C., Hongguang, H., Mckinney-Freeman, S., Naveiras, O., Yoon, T.J., Irizarry, R.A., Jung, N., Seita, J., Hanna, J., Murakami, P., Jaenisch, R., Weissleder, R., and Orkin, S.H. Epigenetic memory in induced pluripotent stem cells. Nature, 467, 285–290 (2010).

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. An extended transcriptional network for pluripotency of embryonic stem cells. Cell, 132(6), 1049-1061 (2008).

Kiskinis, E. and Eggan, K. Progress toward the clinical application of patient-specific pluripotent stem cells. Journal of Clinical Investigations, 120(1), 51–59 (2010).

Krause, D.S. and Scadden, D.T. Deconstructing the complexity of a microenvironmental niche. Cell, 149(1), 16-17 (2012).

Kuhn, T. The Structure of Scientific Revolutions. University of Chicago Press, Chicago (1962).

Lade, S.J. and Gross, T. Early warning signals for critical transitions: a generalized modeling approach. PLoS Computational Biology, 8(2), e1002360 (2012).

Lander, A.D. The 'stem cell' concept: is it holding us back? Journal of Biology, 8, 70 (2009).

Larsson, E., Sander, C., and Marks, D. mRNA turnover rate limits siRNA and microRNA efficacy. Molecular Systems Biology, 6, 433 (2010).

Lassar, A.B., Paterson, B.M., and Weintraub, H. Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. Cell, 47, 649-656 (1986).

Lathia, J.D. and Rich, J.N. Holding on to stemness. Nature Cell Biology, 14, 450–452 (2012).

Lattanzi, L., Salvatori, G., Coletta, M., Sonnino, C., Cusella De Angelis, M.G., Gioglio, L., Murry, C.E., Kelly, R., Ferrari, G., Molinaro, M., Crescenzi, M., Mavilio, F., and Cossu, G. High efficiency myogenic conversion of human fibroblasts by adenoviral vector-mediated *MyoD* gene transfer: an alternative strategy for *ex vivo* gene therapy of primary myopathies. Journal of Clinical Investigations, 101, 2119-2128 (1998).

Lee, C. Empirical information metrics for prediction power and experiment planning. Information, 2(1), 17-40 (2011).

Lee, J-H., Park, I-H, Gao, Y., Li, J-B., Li, Z., Daley, G.Q., Zhang, K., and Church, G.M. A robust approach to identifying tissue-specific gene expression regulatory variants using personalized human induced pluripotent stem cells. PLoS Genetics, 5(11), e1000718 (2009).

Levin, M.   Bioelectric Mechanisms in Regeneration: unique aspects and future perspectives. Seminars in Cell and Developmental Biology, 20, 543-556 (2009).

Li, Z., Yang, C-S., Nakashima, K., and Rana, T.M.   Small RNA-mediated regulation of iPS cell generation. EMBO Journal, 30(5), 823-934 (2011).

Liao, J., Cui, C., Chen, S., Ren, J., Chen, J., Gao, Y., Li, H., Jia, N., Cheng, L., Xiao, H., and Xiao, L.   Generation of induced pluripotent stem cell lines from adult rat cells. Cell Stem Cell, 4, 11-15 (2009).

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

Liu, H., Zhu, F., Yong, J., Zhang, P., Hou, P., Li, H., Jiang, W., Cai, J., Liu, M., Cui, K., Qu, X., Xiang, T., Lu, D., Chi, X., Gao, G., Ji, W., Ding, M., and Deng, H.   Generation of induced pluripotent stem cells from adult rhesus monkey fibroblasts. Cell Stem Cell, 3, 587–590 (2008).

Liu, N., Lu, M., Tian, X., and Han, Z.   Molecular mechanisms involved in self-renewal and pluripotency of embryonic stem cells. Journal of Cell Physiology, 211(2), 279-286 (2007).

Loh, K. and Lim, B.   A precarious balance: pluripotency factors as lineage specifiers. Stem Cells, 8(4), 363-369 (2011).

Ludwig, T.E., Levenstein, M.E., Jones, J.M., Berggren, W.T., Mitchen, E.R., Frane, J.L., Crandall, L.J., Daigh, C.A., Conard, K.R., Piekarczyk, M.S., Llanas, R.A., and Thomson, J.A.   Derivation of human embryonic stem cells in defined conditions. Nature Biotechnology, 24(2), 185-187 (2006).

Lujan, E., Chanda, S., Ahlenius, H., Sudhof, T.C., and Wernig, M.   Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. PNAS, 109(7), 2527–2532 (2012).

Luo, J., Suhr, S.T., Chang, E.A., Wang, K., Ross, P.J., Nelson, L.L., Venta, P.J., Knott, J.G., and Cibelli, J.B.   Generation of leukemia inhibitory factor and basic fibroblast growth factor-dependent induced pluripotent stem cells from canine adult somatic cells. Stem Cells and Development, 20, 1669-1678 (2011).

MacArthur, B.D., Ma'ayan, A., and Lemischka, I.R. Systems biology of stem cell fate and cellular reprogramming. Nature Reviews Molecular Cell Biology, 10, 672-681 (2009).

MacArthur, B.D., Please, C.P., and Oreffo, R.O.C. Stochasticity and the Molecular Mechanisms of Induced Pluripotency. PLoS One, 3(8), e3086 (2008).

MacDonald, N. Biological delay systems: linear stability theory. Cambridge University Press: Cambridge, UK (1989).

Macpherson, P.C., Suhr, S.T., and Goldman, D. Activity-dependent gene regulation in conditionally-immortalized muscle precursor cell lines. Journal of Cell Biochemistry, 91, 821-839 (2004).

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell, 3, 340-345 (2008).

Maherali, N., Sridharan, R., Xie, W., Utikal, J., Eminil, S., Arnold, K., Stadfeld, M., Yachechko, R., Tchieu, J., and Jaenisch, R. Directly reprogramming fibroblasts show global epigenetic remodeling and widespread tissue contribution. Cell Stem Cell, 1, 55-70 (2007).

Mangell, M. and Bonsall, M.B. Phenotypic evolutionary models in stem cell biology: replacement, quiescence, and variability. PLoS One, 3(2), e1591 (2008).

Mantzaris, N.V. Stochastic and deterministic simulations of heterogeneous cell population dynamics. Journal of Theoretical Biology, 241(3), 690-706 (2006).

Markoulaki, S., Hanna, J., Beard, C., Carey, B.W., Cheng, A.W., Lengner, C.J., Dausman, J.A., Fu, D., Gao, Q., Wu, S., Cassady, J.P., and Jaenisch, R. Transgenic mice with defined combinations of drug inducible reprogramming factors. Nature Biotechnology, 27(2), 169-171 (2009).

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., Calabrese, J.M., Dennis, L.M., Volkert, T.L., Gupta, S., Love, J., Hannett, N., Sharp, P.A., Bartel, D.P., Jaenisch, R., and Young, R.A. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. Cell, 134(3), 521-533 (2008).

Masaki, H., Ishikawa, T., Takahashi, S., Okumura, M., Sakai, N., Haga, M., Kominami, K., Migita, H., McDonald, F., Shimada, F., and Sakurada, K. Heterogeneity of pluripotent marker gene expression in colonies generated in human iPS cell induction culture. Stem Cell Research, 1(2), 105-115 (2007).

Meissner, A., Wernig, M., and Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. Nature Biotechnology, 25(10), 1177-1181 (2007).

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

Mikkers, H. and Frisen, J. Deconstructing Stemness. EMBO Journal, 24, 2715-2719 (2005).

Miner, J.H. and Wold, B. Herculin, a fourth member of the *MyoD* family of myogenic regulatory genes. PNAS USA, 87, 1089-1093 (1990).

Morikawa, M. Sensitivity of preadipose 3T3 cells to growth hormone. Journal of Cellular Physiology, 128(2), 293-298 (1986).

Morgan, H.D., Santos, F., Green, K., Dean, W., and Reik, W. Epigenetic reprogramming in mammals. Human Molecular Genetics, 14, R47–R58 (2005).

Munsie, M.J., Michalska, A.E., O'Brien, C.M., Trounson, A.O., Pera, M.F., and Mountford, P.S. Isolation of pluripotent embryonic stem cells from reprogrammed adult mouse somatic cell nuclei. Current Biology, 10, 989–992 (2000).

Nisbet, R.M. and Gurney, W.S.C. Modelling Fluctuating Populations. Wiley, New York (1982).

Nitsch, D., Tranchevent, L-C., Thienpont, B., Thorrez, L., Van Esch, H., Devriendt, K., and Moreau, Y. Network analysis of differential expression for the identification of disease-causing genes. PLoS One, 4(5), e5526 (2009).

Nixon, T. and Green, H. Properties of growth hormone receptors in relation to the adipose conversion of 3T3 cells. Journal of Cellular Physiology, 115(3), 291-296 (1983).

Nobel Prize Committee The Nobel Prize in Physiology or Medicine 2012. Nobelprize.org. http://www.nobelprize.org/nobel_prizes/medicine/laureates/2012/ (2012).

Orkin, S.H. and Hochedlinger, K. Chromatin Connections to Pluripotency and Cellular Reprogramming. Cell, 145, 835-850 (2011).

Orford, K.W. and Scadden, D.T. Deconstructing stem cell self-renewal: genetic insights into cell-cycle regulation. Nature Reviews Genetics, 9 115-128 (2008).

Orosz, G., Moehlis, J., and Murray, R.M. Controlling biological networks by time-delayed signals. Philosophical Transactions of the Royal Society A, 368, 439–454 (2010).

Ozbudak, E.M., Thattai, M., Kurtser, I., Grossman, A.D., and van Oudenaarden, A. Regulation of noise in the expression of a single gene. Nature Genetics, 31, 69-73 (2002).

Palmer, M.E. and Feldman, M.W. Survivability is more fundamental than evolvability. PLoS One, 7(6), e38025 (2011).

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Sudhof, T.C., and Wernig, M. Induction of human neuronal cells by defined transcription factors. Nature, 476(7359), 220-223 (2011).

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

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

Park, I-H. and Daley, G.Q. Debugging cellular reprogramming. Nature Cell Biology, 9(8), 871-873 (2007).

Pasini, D., Cloos, P.A., Walfridsson, J., Olsson, L., Bukowski, J.P., Johansen, J.V., Bak, M., Tommerup, N., Rappsilber, J., and Helin, K. JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. Nature, doi:10.1038/nature08788 (2010).

Patel, M. and Yang, S. Advances in reprogramming somatic cells to induced pluripotent stem cells. Stem Cell Reviews and Reports, 6(3): 367–380 (2010).

Paull, D., Emmanuele, V., Weiss, K.A., Treff, N., Stewart, L., Hua, H., Zimmer, M., Kahler, D.J., Goland, R.S., Noggle, S.A., Prosser, R., Hirano, M., Sauer, M.V., and Egli, D. Nuclear genome transfer in human oocytes eliminates mitochondrial DNA variants. Nature, 493(7434), 632-637 (2013).

Perez-Ortin, J.E. Genomics of mRNA turnover. Briefings in Functional Genomics and Proteomics, 6(4), 282-291 (2008).

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

Polo, J.M., Liu, S., Figueroa, M.E., Kulalert, W., Eminli, S., Tan, K-Y., Apostolou, E., Stadtfeld, M.,Li, Y., Shioda, T., Natesan, S., Wagers, A.J., Melnick, A., Evans, T., and Hochedlinger, K. Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nature Biotechnology, 28(8), 848-855 (2010).

Poss, K. Advances in understanding tissue regenerative capacity and mechanisms in animals. Nature Reviews Genetics, 11(10), 710-721 (2010).

Qian, L., Huang, Y., Spencer, C.I., Foley, A., Vedantham, V., Liu, L., Conway, S.J., Fu, J-D., and Srivastava, D. *In vivo* reprogramming of murine cardiac fibroblasts into induced cardiomyocytes. Nature, 485, 593–598 (2012).

Rabani, M., Levin, J.Z., Fan, L., Adiconis, X., Raychowdhury, R., Garber, M., Gnirke, A., Nusbaum, C., Hacohen, N., Friedman, N., Amit, I., and Regev, A. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. Nature Biotechnology, 29(5), 436-442 (2011).

Raj, A. and Van Oudenaarden, A. Nature, nurture, or chance: stochastic gene expression and its consequences. Cell, 135 (2), 216-226 (2008).

Ramalho-Santos, M. Stem cells as probabilistic self-producing entities. Bioessays, 26(9), 1013-1016 (2004).

Reed, W. and Hughes, B. From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature. Physical Review E, 66(6), 67103 (2002).

Rhee, H.S. and Pugh, B.F. Comprehensive Genome-wide Protein-DNA Interactions Detected at Single-Nucleotide Resolution. Cell, 147, 1408–1419 (2011).

Riniker, S., Allison, J.R., and van Gunsteren, W.F. On developing coarse-grained models for biomolecular simulation: a review. Physical Chemistry Chemical Physics, 14(36), 12423-12430 (2012).

Rinn, J.L., Bondre, C., Gladstone, H.B., Brown, P.O., and Chang, H.Y. Anatomic demarcation by positional variation in fibroblast gene expression programs. PLoS Genetics, 2(7), e119 (2006).

Ruder, W.C., Lu, T., and Collins, J.J. Synthetic biology moving into the clinic. Science, 333, 1248-1252 (2011).

Samoilov, M., Price, G., and Arkin, A. From fluctuations to phenotypes: The physiology of noise. Science STKE, 19(366), re17 (2006).

Schiff, J. Cellular Automata: a discrete view of the world. Wiley, New York. (2007). Chapter 5.

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

Seok, J., Warren, H.S., Cuenca, A.G., Mindrinos, M.N., Baker, H.V., Xu, W., Richards, D.R., McDonald-Smith, G.P., Gao, H., Hennessy, L., Finnerty, C.C., Lopez, C.M., Honari, S., Moore, E.E., Minei, J.P., Cuschieri, J., Bankey, P.E., Johnson, J.L., Sperry, J., Nathens, A.B., Billiar, T.R., West, M.A., Jeschke, M.G., Klein, M.B., Gamelli, R.L., Gibran, N.S., Brownstein, B.H., Miller-Graziano, C., Calvano, S.E., Mason, P.H., Cobb, J.P., Rahme, L.G., Lowry, S.F., Maier, R.V., Moldawer, L.L., Herndon, D.N., Davis, R.W., Xiao, W., and Tompkins, R.G.   Genomic responses in mouse models poorly mimic human inflammatory diseases, PNAS, 110(9), 3507-3512 (2013).

Shalem, O., Dahan, O., Levo, M., Martinez, M.R., Furman, I., Segal, E., and Pilpel, Y. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. Molecular Systems Biology, 4, 223 (2008).

Sharov, A.   Biosemiotics Short Course, Lecture 1. Embryo Physics Course, Second Life (2012).

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S. Database for mRNA half-life of 19,977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Research, 16(1), 45-58 (2009).

Shenoy, A. and Blelloch, R.   MicroRNA induced transdifferentiation. F1000 Biology Reports, 4, 3 (2012).

Siripattarapravat, K., Pinmee, B., Venta, P.J., Chang, C.C., and Cibelli, J.B.   Somatic cell nuclear transfer in zebrafish. Nature Methods, 6, 733-735 (2009).

Soldner, F. and Jaenisch, R.   iPSC disease modeling. Science, 338(6111), 1155-1156.

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

Song, K., Nam, Y-J., Luo, X., Qi, X., Tan, W., Huang, G.N., Acharya, A., Smith, C.L., Tallquist, M.D., Neilson, E.G., Hill, J.A., Bassel-Duby, R., and Olson, E.N.   Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature, 485, 599–604 (2012).

Soufi, A., Donahue, G., and Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. Cell, Volume 151, Issue 5, 994-1004 (2012).

Spivakov, M. and Fisher, A.G. Epigenetic signatures of stem-cell identity. Nature Reviews Genetics, 8, 263-271 (2007).

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

Sridharan, R. and Plath, K. Illuminating the black box of reprogramming. Cell Stem Cell, 2, 295–297 (2008).

Srivastava, D. and Ieda, M. Critical factors for cardiac reprogramming. Circulation Research, 111, 5-8 (2012).

Stadtfeld, M., Apostolou, E., Akutsu, H., Fukuda, A., Follett, P., Natesan, S., Kono, T., Shioda, T., and Hochedlinger, K. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. Nature, 465, 175-181 (2010).

Stadtfeld, M. and Hochedlinger, K. Induced pluripotency: history, mechanisms, and applications. Genes and Development, 24, 2239-2263 (2010).

Stadtfeld, M., Maherali, N., Breault, D.T., and Hochedlinger, K. Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse. Cell Stem Cell, 2(3), 230-240 (2008).

Stadtfeld, M., Brennand, K., and Hochedlinger, K. Reprogramming of pancreatic beta-cells into induced pluripotent stem cells. Current Biology, 18, 890–894 (2008).

Steuer, R., Kurths, J., Daub, C.O., Weise, J., and Selbig, J. The mutual information: detecting and evaluating dependencies between variables. Bioinformatics, 18(S2), S231-S240 (2002).

Subramanyam, D. and Blelloch, R. Watching reprogramming in real time. Nature Biotechnology, 27(11), 997-998 (2009).

Suhr, S.T., Chang, E.A., Rodriguez, R.M., Wang, K., Ross, P.J., Beyhan, Z., Murthy, S., and Cibelli, J.B. Telomere dynamics in human cells reprogrammed to pluripotency. PLoS One, 4, e8124 (2009).

Sul, J.Y., Kim, T.K., Lee, J.H., and Eberwine, J. Perspectives on cell reprogramming with RNA. Trends in Biotechnology, 30(5), 243-249 (2012).

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell, 131(5), 861-872 (2007).

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

Tang, M. The mean and noise of stochastic gene transcription. Journal of Theoretical Biology, 253, 271– 280 (2008).

Telford, N.A., Watson, A.J., and Schultz, G.A. Transition from maternal to embryonic control in early mammalian development: a comparison of several species. Molecular Reproduction and Development, 26, 90–100 (1990).

Theunissen, T.W., van Oosten, A.L., Castelo-Branco, G., Hall, J., Smith, A., and Silva, J.C.R. Nanog overcomes reprogramming barriers and induces pluripotency in minimal conditions. Current Biology, 21(1), 65-71 (2010).

Tyler, A.L., Asselbergs, F.W., Williams, S.M., and Moore, J.H. Shadows of complexity: what biological networks reveal about epistasis and pleiotropy. Bioessays, 31(2), 220–227 (2009).

Valencia-Sanchez, M.A., Liu, J., Hannon, G.J., and Parker, R. Immortalization eliminates a roadblock during cellular reprogramming into iPS cells. Nature, 460, 1145 (2009).

Valencia-Sanchez, M.A., Liu, J., Hannon, G.J., and Parker, R. Control of translation and mRNA degradation by miRNAs and siRNAs. Genes and Development, 20, 515–524 (2006).

Varley, K.E. and Mitra, R.D. Bisulfite Patch PCR enables multiplexed sequencing of promoter methylation across cancer samples. Genome Research, 1279-1287 (2010).

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

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

Viswanthan, S. and Zandstra, P.W. Towards predictive models of stem cell fate. Cytotechnology Reviews, 41(2/3), 1-31 (2004).

Vogel, W., Grunebach, F., Messam, C.A., Kanz, L., Brugger, W., and Buhring, H.J. Heterogeneity among human bone marrow-derived mesenchymal stem cells and neural progenitor cells. Haematologica, 88, 126-133 (2003).

Vose, M.D. The simple genetic algorithm: foundations and theory. MIT Press, Cambridge, MA (1999).

Wadhwa, R., Kaul, C., and Mitsui, Y. Cellular mortility and immortalization: a complex interplay of multiple gene functions. In A. Macieira-Coelho, Cell Immortalization. Progress in Molecular and Subcellular Biology, Volume 24. Springer-Verlag, Berlin (1999).

Wakayama, T., Perry, A.C., Zuccotti, M., Johnson, K.R., and Yanagimachi, R. Full-term development of mice from enucleated oocytes injected with cumulus cell nuclei. Nature, 394, 369-374 (1998).

Watts, D.J. A simple model of global cascades on random networks. PNAS, 99(9), 5766 –5771 (2002).

Wernig, M., Lengner, C.J., Hanna, J., Lodato, M.A., Steine, E., Foreman, R., Staerk, J., Markoulaki, S., and Jaenisch, R. A drug-inducible transgenic system for direct reprogramming of multiple somatic cell types. Nature Biotechnology, 26(8), 916-924 (2008).

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

Wieland, M. and Fussenegger, M. Reprogrammed cell delivery for personalized medicine. Advanced Drug Delivery Reviews, 64(13), 1477–1487 (2012).

Wilkinson, D.J. Stochastic modeling for systems biology. CRC Press, Boca Raton, FL (2006). Chapter 6.

Wolfram, S. A new kind of science. Wolfram Media, Champaign, IL (2002). Chapter 5.

Xie, D., Chen, C-C., He, X., Cao, X., and Zhong, S. Towards an evolutionary model of transcription networks. PLoS Computational Biology, 7(6), e1002064 (2007).

Xie, D., Chen, C.C., Ptaszek, L.M., Xiao, S., Cao, X., Fang, F., Ng, H.H., Lewin, H.A., Cowan, C., and Zhong, S. Rewirable gene regulatory networks in the preimplantation embryonic development of three mammalian species. Genome Research, 20, 804-815 (2010).

Xu, N., Papagiannakopoulos, T., Pan, G., Thomson, J.A., and Kosik, K.S. MicroRNA-145 regulates *OCT4*, *SOX2*, and *KLF4* and represses pluripotency in human embryonic stem cells. Cell, 137(4), 647-658 (2009).

Yamanaka, S. Elite and stochastic models for induced pluripotent stem cell generation. Nature, 460, 49–52 (2009).

Yang, N., Ng, Y.H., Pang, Z.P., Sudhof, T.C., and Wernig, M. Induced neuronal cells: how to make and define a neuron. Cell Stem Cell, 9(6), 517-525 (2011).

Yong, E.   Bad Copy. Nature, 485, 298 (2012).

Yoo, A.S.   MicroRNA-mediated conversion of human fibroblasts to neurons. Nature, 476, 228-231 (2011).

Yoon, S. and Seger, R.   The extracellular signal-regulated kinase: Multiple substrates regulate diverse cellular functions. Growth Factors, 24(1), 21-44 (2006).

Yoshida, Y. and Yamanaka, S.   Labor pains of new technology: direct cardiac reprogramming. Circulation Research, 111, 3-4 (2012).

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

Zhang, S-C., Wernig, M., Duncan, I.D., Brustle, O., and Thomson, J.A.   *In vitro* differentiation of transplantable neural precursors from human embryonic stem cells. Nature Biotechnology, 19, 1129-1133 (2001).

Zimmerman, L., Parr, B., Lendahl, U., Cunningham, M., McKay, R., Gavin, B., Mann, J., Vassileva, G., and McMahon, A.   Independent regulatory elements in the nestin gene direct transgene expression to neural stem cells or muscle precursors. Neuron, 12, 11-24 (1994).

Zipori, D.   The nature of stem cells: state rather than entity. Nature Reviews Genetics, 5, 873-878 (2004).