Transcription Factor-mediated Epigenetic Reprogramming

Published, JBC Papers in Press, September 5, 2012, DOI 10.1074/jbc.R111.319046

Camille Sindhu, Payman Samavarchi-Tehrani, and Alexander Meissner

From the Department of Stem Cell and Regenerative Biology, Harvard University and Harvard Stem Cell Institute, Cambridge, Massachusetts 02138 and the Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142

Input from various signaling pathways in conjunction with specific transcription factors (TFs), noncoding RNAs, and epigenetic modifiers governs the maintenance of cellular identity. Endogenous or exogenous TFs operate within certain boundaries, which are set, in part, by the cell type-specific epigenetic landscape. Ectopic expression of selected TFs can override the cellular identity and induce reprogramming to alternative fates. In this minireview, we summarize many of the classic examples and a large number of recent studies that have taken advantage of TF-mediated reprogramming to produce cell types of biomedical relevance.

For many years, it was unclear whether differentiation involves irreversible changes to the genome that would restrict a cell’s developmental potential. Early work by Briggs and King (1) and later Gurdon (2) addressed this subject using somatic cell nuclear transfer (SCNT) from various donor nuclei into an enucleated frog oocyte. Successful generation of viable organisms by SCNT demonstrated that the genome of a differentiated cell does retain all genetic information necessary for normal development. Later experiments were able to expand these seminal findings to mammals (3). SCNT using donor nuclei from genetically marked lymphoid cells (4) and olfactory receptor neurons (5) demonstrated further that even terminally differentiated and post-mitotic genomes could be reprogrammed. More recently, human somatic cell nuclei were also shown to be amenable to reprogramming by SCNT. However, in contrast to other species, this could be achieved only without prior removal of the oocyte nucleus (6).

Other cell types besides oocytes have been shown to possess factors capable of activating silenced loci in a somatic genome. For instance, introduction of a chicken erythrocyte nucleus into the cytoplasm of a HeLa cell results in chromatin decondensation and initiation of RNA synthesis from the previously inac-tive erythrocyte genome (7). Later studies in mice and humans demonstrated that fusion of somatic cells with embryonic stem cells (ESCs) or embryonic germ cells reactivates pluripotency-related genes from the somatic genome and creates pluripotent, albeit tetraploid (4N), cells (8–10). Although these have limited clinical value, their genesis has provided a useful experimental platform for studying cellular plasticity and reprogramming (11). Moreover, the fusion experiments offered evidence that ESCs, like oocytes, zygotes, and early blastomeres, contain factors that are sufficient to reprogram a somatic cell.

Initial evidence for the capacity of selected transcription factors (TFs) to direct cellular reprogramming came from the classic myoD experiments and subsequent lineage conversions in the hematopoietic system (reviewed in Ref. 12). However, a major advancement in the field occurred in 2006 with the landmark study by Takahashi and Yamanaka (13), who reported the generation of induced pluripotent stem cells (iPSCs) through ectopic expression of only four TFs.

Converting Cell States

In contrast to the limited plasticity that exists in vivo, there are many examples of ectopic TF-mediated reprogramming in vitro (Fig. 1). Early unrelated studies led to the discovery of the basic helix-loop-helix (bHLH) TF myoD, whose expression in mouse embryonic fibroblasts could induce their conversion to myoblasts (14). MyoD appears to be a master regulator of muscle-specific transcriptional programs, and its ectopic expression can induce activation of muscle-specific genes in a variety of non-muscle cells (15, 16). Interestingly, although MyoD could induce complete phenotypic conversion in mesodermal cell types, ectodermal and endodermal cells responded less effectively (16).

Valuable insights into TF-mediated reprogramming were gained through fate conversions within the hematopoietic lineage, where developmental hierarchies are well understood. Common lymphoid progenitors (CLPs) can be induced to give rise to myeloid lineage cells such as granulocytes and monocytes by ectopic expression of the IL-2 receptor, which in turn leads to expression of GM-CSF receptors and the associated responsiveness to myeloid commitment-inducing signals (17). Although both CLPs and pro-T-cells could undergo this form of lineage redirection, the further downstream and lineage-committed pre-T- and pro-B-cells could not, suggesting a possible link between differentiation state and amenability to cell fate conversion. In another example, ectopic expression of the zinc finger TF GATA-1, a key regulator of megakaryocyte-erythrocyte precursor lineage commitment, was found to elicit conversion of CLP and GM progenitors into megakaryocytes and erythrocytes (18, 19). GATA-1 activates megakaryocyte-erythrocyte precursor-specific genes while concomitantly down-regulating markers of the other lineages. Alternatively, when GATA-1 is expressed at low levels in GM progenitors, a fate decision is forced whereby eosinophils or mast cells are generated (20), suggesting a critical role for factor stoichiometry in fate determination.
The transcription factor PU.1 (Spi-1) belongs to the Ets family and is a key regulator of myeloid and B-lymphoid lineage specification, functioning through interaction with other TFs such as GATA-1/2 and C/EBPα. Interestingly, lineage choice between these two fates is determined in part by graded expression of PU.1. Through its interaction with GATA proteins, PU.1 transactivation of myeloid target genes is suppressed. Conversely, PU.1 inhibits the erythroid program by binding GATA-1 and inducing the formation of repressive chromatin structure at GATA-1 target loci. Ectopic expression of PU.1 in multipotent progenitor cell lines leads to suppression of GATA-1 and concomitant conversion into myeloid cells. The antagonism between GATA-1 and PU.1 dictates one of two fates and highlights the key role for TF interactions in the process.

Another classic example involves the basic leucine zipper transcription factor C/EBPα, which is required for the common myeloid progenitor-to-GM progenitor transition in vivo. Ectopic expression of C/EBPα in primary bone marrow cells, lymphocytes, or fibroblasts can elicit myelomonocytic cell-type characteristics, a classification that includes both macrophages and granulocyte precursors. The function of C/EBPα in this context is dependent on synergism with PU.1, which is required for deposition of H3K4me1 at enhancer elements of target genes, suggesting that these factors cooperate to define cell type-specific binding patterns at regulatory elements.

In addition to ectopic expression, loss of fate determinants can also induce cell fate changes. Differentiation of CLPs into the B-cell lineage depends on the TFs PU.1, E2A, and EBF1, which induce Pax5 to activate B-cell-specific genes while repressing genes associated with alternative lineages. In agreement with this model, Pax5−/− pro-B-cells fail to complete B-lymphopoiesis but are capable of differentiating into other hematopoietic cell types such as macrophages, dendritic cells, and granulocytes in response to specific signaling cascades. Additionally, Pax5 ablation in mature B-cells results in differentiative commitment to different hematopoietic cell types.

In summary, lessons from the hematopoietic system have provided strong evidence for the ability of TFs to redirect cell fate across related lineages derived from one germ layer or between specialized cell types of a particular lineage. Below, we will briefly review many of the more recent reprogramming studies.

Using a candidate gene approach, Melton and co-workers identified three bHLH TFs, Ngn3 (or NeuroD1), Pdx1, and MafA, whose forced expression can convert exocrine pancreas tissue into insulin-secreting endocrine β-cells in vivo. Expressing only the endocrine progenitor-defining TF Ngn3 in hepatic progenitor cells generated physiologically responsive pancreatic endocrine cells. In contrast, expression of Ngn3 in mature hepatocytes induced only insulin expression (but not transdif-
MINIREVIEW: Transcription Factor Reprogramming

ferentiation) in islet cells (41). An additional example of reprogramming in vivo comes from the inner ear, where the TFs Atoh1 and Prox1, among others, were found to regulate the development of sensory hair cells and supporting cells from a common progenitor (42, 43). Ectopic expression of the bHLH TF Atoh1 (also known as Math1) results in conversion of non-sensory cochlear cells into functional sensory hair cells (44, 45). Conversely, expression of Prox1 in sensory hair cells leads to the repression of Gf11 and Atoh1, factors required for sensory hair cell specification, resulting in cellular degeneration (43).

Recently, other groups have demonstrated direct reprogramming in vitro into additional endodermal cell types. The ectopic expression of GATA-4, Hnf1α, and Foxa3 in fibroblasts, along with inactivation of p19<sup>Arf</sup>, could give rise to induced hepatocyte-like cells (46). A separate study found that ectopic expression of Hnf4α and one of the three foxA genes in mouse embryonic or adult fibroblasts could also induce expression of multiple hepatocyte-specific features, converting the fibroblasts to hepatocyte-like cells (47).

Employing a similar candidate gene approach as Yamanaka and Melton, Wernig and co-workers (48) demonstrated that expression of three factors, Ascl1, Brn2 (also known as Pou3f2), and Myt1l, in mouse embryonic and postnatal fibroblasts induced conversion to neural cells, termed induced neuronal (iN) cells. Although the exact identity of these cells remains unclear, iN cells can form functional synapses and are physiologically responsive (48). These three factors could also induce neuronal differentiation of human ESCs, although reprogramming of human fetal fibroblasts to functional iN cells required additional coexpression of NeuroD1 (49). To definitively demonstrate the neural conversion of a differentiated non-ectodermal cell type, Marro et al. (50) recently reprogrammed mouse hepatocytes into iN cells. The hepatic transcriptome was largely repressed in the iN cells, and they retained only a limited epigenetic signature of their starting state. These initial findings demonstrate the ability of ectopic TFs to reprogram somatic cells across germ layers. More recently, others have shown that human and mouse fibroblasts can be reprogrammed to specialized functional neural subtypes such as dopaminergic neurons (51, 52) and spinal motor neurons (53). Interestingly, fibroblasts can be induced to express neural lineage markers and exhibit neuronal morphology solely by forced expression of microRNAs miR-9/9* and miR-124, which repress the Baf53a phosphatase and SSEA1 (stage-specific embryonic antigen-1) could reprogram embryonic and adult fibroblasts directly into induced cardiomyocytes.

The diversity of cell types generated by TF-mediated reprogramming demonstrates the potential utility of this approach for therapeutic purposes. However, several key aspects of these converted cell types need further investigation. In particular, the stability and relative completion of functional, transcriptional, and epigenetic remodeling, as well as the in vivo equivalence of the generated cell types, remain unclear. A better understanding of potential risks posed by incomplete reprogramming or cellular memory is an important benchmark for the translational application of induced cell types.

Reprogramming Somatic Cells to Pluripotency

Perhaps the most striking example of factor-induced changes in cell state remains the reprogramming of somatic cells into pluripotent embryonic stem-like cells, termed iPSCs. The seminal work by Takahashi and Yamanaka (13) demonstrated that ectopic expression of Oct4, Sox2, Klf4, and c-Myc (OSKM) in embryonic and adult fibroblasts could convert these cells to iPSCs. Subsequently, numerous laboratories have improved the original protocol and generated iPSCs from a variety of species and cell types (58), highlighting the robustness and general conservation of the approach (Fig. 2).

Forced expression of OSKM initiates dramatic phenotypic (59) and molecular (60–63) changes in the targeted somatic cell. Transcriptional and epigenetic changes eventually lead to the reactivation of endogenous pluripotency genes and acquisition of molecular and, in some cases, functional pluripotency. Commonly used markers of reprogramming include alkaline phosphatase and SSEA1 (stage-specific embryonic antigen-1)
Although these markers appear earlier in the process, they are not indicative of complete reprogramming and are of relatively low stringency (65). Retroactive tracing of individual reprogramming fibroblasts has demonstrated that within the set of cells that undergo complete reprogramming, the initial response to reprogramming factor induction is characterized by a marked decrease in cell size and increased proliferation rate (59). The early response to the four TFs is marked by the transcriptional reversion from the somatic state, as evidenced by the down-regulation of the fibroblast-specific genes Snai1/2, Zeb1/2, and Cdh2 and the surface marker Thy1 (60, 63, 64). The loss of a mesenchymal phenotype coincides with the emerging expression of epithelium-associated markers such as E-cadherin and EpCAM, known as the mesenchymal-to-epithelial (MET) transition (63, 66). The importance of MET during reprogramming is highlighted by acceleration of reprogramming using bone morphogenetic protein ligand stimulation or expression of microRNAs from the miR-200 family (63), which have been implicated in MET (67). Notably, Klf4, the main driver of this transition, synergizes with Oct4 and Sox2 to direct reprogramming, and its absence can be replaced by stimulation with bone morphogenetic protein (68).

Establishment of the iPSC state in mice requires continuous expression of the OSKM transgenes for 8–10 days (64), leading to the eventual activation of endogenous pluripotency genes such as Oct4, Sox2, and Nanog after ~2 weeks. The final progression of reprogramming occurs only in a small percentage of SSEA1-positive cells (62). The mechanism whereby the endogenous stem cell network is transcriptionally reactivated in the final transition to pluripotency remains poorly understood.

Although c-Myc is routinely included in the reprogramming factor mixture, and expression of this gene enhances the efficiency of somatic cell reprogramming, it is not a core member of the ESC pluripotency network. The network controlled by c-Myc is largely separate from the core ESC pluripotency network and primarily includes targets involved in metabolic, cell cycle, and proliferative processes (69–71). In ESCs, c-Myc target gene promoters are highly enriched for the active histone modification H3K4me3 and depleted of the repressive modification H3K27me3; this is a striking pattern that is not exhibited to the same extent by targets of the core pluripotency factors (69). The core factors reside at targets enriched for both activating and repressive histone modifications, reflecting their role in maintaining ESC identity by activating ESC-specific genes while repressing expression of lineage-specific transcriptional regulators (72). Conversely, c-Myc targets are more expansive in function, encompassing genes not unique to the pluripotent state (61, 69). The recent discovery that c-Myc plays a role in regulating pause release of polymerase II transcriptional elongation is in agreement with many of the early phenotypic and molecular responses to reprogramming factor induction (73). Notably, the latter suggests that c-Myc may not be directly involved in reactivating the pluripotency network but rather act more broadly to drive robust transcription of paused genes during the early stages of reprogramming. Genome-wide localization studies of pluripotency factors, along with affinity-based identification of their interacting partners, have revealed an interconnection between TFs (74, 75), chromatin-remodeling, and modifying complexes (Table 1). Although it is well established that ESC chromatin is predominantly euchromatic (76, 77) and possess characteristic bivalent enrichment of both active and repressive histone modifications at developmental genes (78), functional links between specific epigenetic modifiers and members of the ESC core transcriptional network have only recently come to light. Wdr5, a member of the mammalian Trithorax complex and an effector of activating H3K4 methylation, is an interacting partner of Oct4 (79). This provides a framework for conceptualizing how transcriptional regulators of pluripotency may cooperate with chromatin-associated factors to regulate the balance between self-renewal and differentiation.

### Early Dynamics in TF-induced Reprogramming

Genome-wide chromatin mapping studies of selected epigenetic marks in populations of induced fibroblasts have provided many useful insights into early reprogramming dynamics (80). One of the earliest cellular responses to ectopic factor expression is de novo deposition or enhanced enrichment of euchromatin-associated H3K4me2 at a large number of promoters. Many of the target genes include developmental regulators and
MINIREVIEW: Transcription Factor Reprogramming

pluripotency genes (80). Importantly, these changes appear to precede transcriptional activation, which, in the case of the pluripotency-associated genes, will not occur until the late stages of reprogramming. Many of these dynamics can be observed in the absence of cell division. The fact that these regions are enriched for both targets of Oct4 and Sox2 suggests that one of the early steps in somatic cell reprogramming involves global coordinated epigenetic “priming.” The notable enrichment for H3K4 methylation at Oct4 target sites accords well with the aforementioned study by Ang et al. (79) connecting Oct4, Wdr5, and histone methylation of Oct4 targets. Importantly, the reestablishment and erasure of repressive marks such as H3K27me3 occur at a notably slower rate (80).

Interestingly, the majority of expression changes that are detectable during the first few cell divisions occur almost solely within regions of H3K4me3-enriched accessible chromatin environments (60, 80). This suggests that the action of the reprogramming factors is restricted, at least initially, to transcriptionally permissive genomic loci. Moreover, at these early stages, up-regulated (but not down-regulated) genes are strongly enriched for targets of c-Myc. Unlike Oct4 and Sox2, c-Myc is already endogenously expressed in somatic fibroblasts (13) and primarily regulates targets involved in cellular metabolism and proliferation. The observed transcriptional response to increased ectopic c-Myc is consistent with its proposed role in stabilizing transcriptional elongation, thereby enabling productive transcription at many shared (somatic and pluripotent) target genes (73). Consistent with these data and in accordance with a distinct binding pattern for c-Myc in ESCs (69), Plath and co-workers (61) showed that the most dramatic early changes in fibroblast gene levels were largely attributable to c-Myc expression.

Subsequent to the first few cell divisions, molecular events leading to up-regulation of endogenous pluripotency genes and transition to the iPSC state remain hard to define due to population heterogeneity. Treatment of reprogramming cells with chromatin-modifying agents can positively affect efficiency. For instance, the Dnmt inhibitor 5-azacytidine (5-AZA), which increases H3K4 methylation at Oct4 target sites accords well with the aforementioned study by Ang et al. (79) connecting Oct4, Wdr5, and histone methylation of Oct4 targets. Importantly, the reestablishment and erasure of repressive marks such as H3K27me3 occur at a notably slower rate (80).

Some iPSCs Are Equivalent to ESCs

The functional and molecular equivalence of ESCs and iPSCs has been a key question since the initial report of iPSC derivation (13). Developmental potential can be determined with low stringency by in vitro differentiation and teratoma formation and with higher stringency in non-human species by chimera formation, germ-line contribution (84), or the production of entirely iPSC/ESC-derived animals by tetraploid complementation (85, 86). Multiple molecular screening strategies have been employed to compare ESC and iPSC lines quantitatively, including profiling transcriptional and epigenetic signatures on a genome scale. At the morphological, molecular, and functional levels, ESCs and iPSCs show a high degree of similarity (87), although the complete equivalence of these cell types remains controversial. Numerous studies have described molecular differences between iPSCs and ESCs (88–90). We have recently used three genomic assays to profile 20 human ESC and 12 iPSC lines and generated a reference map of DNA methylation, gene expression, and in vitro differentiation potential for the tested lines (91). Surprisingly, much of the variation observed within iPSC lines could be explained by the general variation between the different pluripotent lines tested, including the ESC lines. These findings accord with functional studies in the mouse, which have established that some (but not all) iPSC lines are both functionally and molecularly equivalent to ESCs (86, 89, 92).

Experimentally determined variations in gene expression levels and epigenetic profiles have yet to provide a molecular signature specific to iPSCs that can consistently discriminate between the various pluripotent populations (91, 93). Establishing the criteria for such a signature has proven challenging, given that variability exists among individual ESC and iPSC lines. To avoid confounding effects of differences in genetic background, a recent study derived and compared the molecular and functional characteristics of genetically matched mouse ESCs and iPSCs. The findings suggest that the expression patterns of ESCs and iPSCs are remarkably similar, with the exception of a few genes of the imprinted Dlk1-Dio3 locus, which appeared to be aberrantly silenced in the majority of iPSC clones examined (89). A more recent study has shown this disparity to be greatly affected by the stoichiometry of the TFs (OSKM), suggesting the inactivation of this locus may be linked to the specific strategy or conditions in which the cells are generated and is not necessarily inherent to reprogramming (94). An additional caveat to the generality of the observations summarized here is the relatively small number of ESC and iPSC lines that are used in many studies (Table 2). In humans, another concern is the limited number of ESC lines, such as H1


and H9, which have been used as references or controls in many studies. For instance, a recent study reported aberrant hypomethylation of promoter regions in iPSC lines compared with the H1 and H9 ESC lines (95). However, comparison with a panel of 32 pluripotent cell lines (91) demonstrated that hypomethylation of the reported regions is a general characteristic of both ESCs and iPSCs, with the exception of a few cell lines, including the H1 and H9 ESC lines used as controls in the original report. This suggests that rather than aberrant hypomethylation in the iPSCs, these loci exhibit possibly aberrant hypermethylation in the control ESC lines.

**Conclusion**

Coordinated activity of TFs and epigenetic modifiers, acting downstream or in concert with signaling molecules, establishes and maintains stable cellular identities by driving cell type-specific gene expression programs. Epigenetic modifications affect the ability of the transcriptional machinery to access the chromatin template, thereby influencing gene expression patterns that ultimately determine cellular identity. Although cellular differentiation appears to be largely unidirectional during development and fixed in adult organisms, elegant experiments over the recent decades have demonstrated the remarkable plasticity by which differentiated cells can be redirected to alternative fates. Cell state can be considered the phenotypic output of regulatory networks, which are in turn influenced by and reflected in cell-specific epigenetic landscapes. By manipulating transcriptional programs, it should be theoretically possible to reprogram a cell’s identity to any desired alternative state.

Many important questions remain open in the field of cell fate conversion experiments. It is still unknown how ectopic
MINIREVIEW: Transcription Factor Reprogramming

factors access their binding sites in a foreign chromatin context and how their function is influenced by the starting cell’s molecular ensemble. A mechanistic description for the relevance of epigenetic or transcriptional similarities between starting and intended cell types in cell fate reprogramming is also lacking. For example, the higher efficiency of conversion between closely related cell types may be partly attributable to higher transcriptional similarity and thus availability of essential cofactors between the starting and intended cell types. In addition to the transcriptional repertoire, however, a more closely related epigenetic landscape may also enable more efficient conversions. Ultimately, the knowledge gained from experimental manipulation of cell fates will have far-reaching implications not only for the basic understanding of the molecular mechanisms behind cellular identity but also for the eventual application to personalized regenerative medicine.

Acknowledgments—We thank members of the Meissner laboratory (in particular, Zachary D. Smith and Casey A. Gifford) for critical comments.

REFERENCES

Ronin/Hcf-1 binds to a hyperconserved enhancer element and regulates genes involved in the growth of embryonic stem cells. *Genes Dev.* **24**, 1479–1484

Tcf3 is an integral component of the core regulatory circuitry of embryonic stem cells. *Genes Dev.* **22**, 746–755


Chromatin and transcriptional signatures for Nodal signaling during endoderm formation in hESCs. *Dev. Biol.* **357**, 492–504