CellPress

Review

Transcriptional Control of Somatic Cell Reprogramming

Yan Xu,[1,2,5] Meng Zhang,[1,2,3,5] Wenjuan Li,[1,2,3] Xihua Zhu,[1,2] Xichen Bao,[1,2] Baoming Qin,[2,4] Andrew P. Hutchins,[2,*] and Miguel A. Esteban[1,2,*]

Somatic cells and pluripotent cells display remarkable differences in most aspects of cell function. Accordingly, somatic cell reprogramming by exogenous factors requires comprehensive changes in gene transcription to induce a forced pluripotent state, which is encompassed by a simultaneous transformation of the epigenome. Nevertheless, how the reprogramming factors and other endogenous regulators coordinate to suppress the somatic cell gene program and activate the pluripotency gene network, and why the conversion is multi-phased and lengthy, remain enigmatic. We summarize the current knowledge of transcriptional regulation in somatic cell reprogramming, and highlight new perspectives that may help to reshape existing paradigms.

Resetting Gene Expression on the Road to Induced Pluripotency

The reprogramming of somatic cells to induced pluripotent stem cells (iPSCs) by exogenous factors, including the original cocktail devised by Takahashi and Yamanaka (OCT4, SOX2, KLF4, and c-MYC: OSKM), demonstrates the power of transcription factors to determine cell fate [1]. Nevertheless, our understanding of reprogramming mechanisms remains incomplete, raising concerns regarding potential applications [2]. Substantial progress has been made in defining the functional events necessary for reprogramming [3]: (i) a proliferation burst coupled to the bypass of apoptosis and cell senescence [4], (ii) loss of somatic cell characteristics coupled to a process of organelle remodeling and epithelialization (or mesenchymal-to-epithelial transition, MET) [5,6], and (iii) a metabolic shift that facilitates quicker – albeit less efficient – ATP production [7,8] (Box 1). These phenomena are linked to the progressive appearance of stem cell markers [e.g., alkaline phosphatase activity and surface antigens including SSEA-1 (for mouse), SSEA-3 and TRA-1-60 (for human reprogramming in primed conditions)], and finally the complete activation of the core pluripotency network [9–11]. However, while the reduction of somatic cell markers (e.g., *Thy1*, *S100a4*, and collagen family members in mouse fibroblasts) happens rapidly and in the majority of reprogramming cells, the acquisition of stem cell markers takes place in a significantly smaller fraction. Consequently, the final number of cells fully activating the pluripotency network is limited and most cells in the original population become trapped in an early intermediate cell state [11,12] or in a partially reprogrammed state termed the pre-iPSC state [13]. Gene expression studies of bulk populations with DNA microarrays have confirmed these findings in the mouse system, contributing to establishing the dogma that reprogramming is phased and stochastic, and hence inefficient [6,14,15]. In addition, analysis of specific cell intermediates in mouse reprogramming has defined two major waves of gene transcription separated by a period of relative quiescence [12] (Box 1), suggesting that the multi-phased nature of reprogramming is partly determined by the inability of OSKM to rearrange global gene transcription simultaneously. The first transcriptional wave of mouse reprogramming is induced in almost every cell in culture, and ensures the activation of a proliferative and cell remodeling response that is concomitant with the reduction of somatic gene expression. Conversely, the

Trends

Somatic cells confront massive barriers on the way to induced pluripotency, which impairs reprogramming efficiency and may induce abnormalities.

OSK act as pioneer transcription factors, interacting with distal regions in closed chromatin. This recruits coactivators and corepressors, inducing successive rounds of chromatin remodeling that make reprogramming permissive. Conversely, the role of c-MYC in reprogramming is complex and poorly understood.

Higher-order chromatin architecture is reorganized in reprogramming. This requires the reassembly of enhancer/promoter loops, interactions within and between topologically associated chromatin domains (TADs) and reorganization of lamina-associated domains (LADs).

Transcriptional pause release of pluripotency genes is rate-limiting for reprogramming. Gene body elongation is also regulated in reprogramming.

[1]Laboratory of RNA, Chromatin, and Human Disease, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[2]Key Laboratory of Regenerative Biology of the Chinese Academy of Sciences and Guangdong Provincial Key Laboratory of Stem Cells and Regenerative Medicine, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[3]University of Chinese Academy of Sciences, Beijing 100049, China

 http://dx.doi.org/10.1016/j.tcb.2015.12.003

second wave takes place in SSEA-1^{+} cells and is enriched for pluripotency regulators [12]. Analysis of gene expression in human reprogramming using immortalized secondary fibroblasts has also shown distinct waves of gene transcription, but there are differences in the number and nature of these waves compared with the mouse system [11] (Box 1). Understanding how the exogenous factors control gene activation/repression in the different transcriptional waves of reprogramming may clarify why the above-mentioned checkpoints need to exist at all, and consequently why the entire process takes so long (typically ~20 and ~30 days in mouse and human systems, respectively) compared to somatic cell nuclear transfer and cell fusion [16].

Initial Accessibility of OSKM to Target Sites

At the onset of reprogramming, exogenous OSKM face millions of potential binding sites among the myriad bases of the genome. However, much of this DNA, in particular those sites bound by pluripotency transcription factors in embryonic stem cells (ESCs), is packaged into nucleosomes and occluded by repressor complexes, non-permissive histone modifications, and higher-order chromatin structures [17]. To overcome these barriers, OSK (but not c-MYC) act as pioneer factors: transcription factors that bind to closed chromatin – but not to refractory heterochromatin – and then progressively endow competence for the activation of cell type-specific programs [18]. For this purpose, in the first 48 h of reprogramming OSK interact with distal genomic regions located in DNase I-resistant chromatin that lacks evident pre-existing histone marks [19] (Figure 1A, Key Figure). To access this non-permissive chromatin, OSK co-bind to degenerate DNA recognition motifs partially exposed on nucleosomes [20]. This suggests, paradoxically, that there is no obvious initial target preselection and it is instead the random engagement of the exogenous factors with chromatin that initiates the cascade of reprogramming. The high frequency of OSK co-bound DNA regions genome-wide, as opposed to regions bound by only one or two exogenous factors, also supports the idea that transcription factor cooperativity at target sites is important for reprogramming in the same

[4]Laboratory of Metabolism and Cell Fate, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China
[5]These authors contributed equally

*Correspondence: andrew@gibh.ac.cn (A.P. Hutchins) and miguel@gibh.ac.cn (M.A. Esteban).

Box 1. Key Cellular Events during Reprogramming

To achieve pluripotency, reprogramming cells must traverse through multiple roadblocks/checkpoints, including the apoptosis and cell senescence barrier, the MET, a metabolic switch, acquisition of early pluripotency genes, and finally the activation of the full pluripotency gene network (Figure I). In mouse, traversal across these roadblocks is mediated by two major transcriptional waves, named early and late waves [12]. Human reprogramming also shows distinct waves of gene transcription and, although many of the same mouse roadblocks are present, the order and timing are substantially altered [11] (Figure I).

Reprogramming is a stressful process that triggers the production of reactive oxygen species and a DNA damage response, leading to apoptosis via c-MYC-dependent activation of p53 and BAX [114], as well as OCT4-dependent CASPASE-3/8 activation [115], in the early phase of mouse/human OSKM reprogramming. The cell cycle is also crucial because quiescent cells are refractory to reprogramming [116], while rapidly cycling cells are capable of highly-efficient reprogramming [111]. Consequently, inhibition of the *Ink4/Arf/Cdkn2a* locus by ablation of JMJD3 (KDM6B) or over-expression of JHDM1A/B overcomes cell senescence and reprogramming efficiency is improved [24,94,117]. Importantly, the cell cycle of ESCs is different compared to somatic cells [118] and, by imposing ESC-specific cell cycle features on somatic cells (e.g., overexpressing cyclin D1), human iPSCs can be produced more efficiently [119].

Fibroblasts are mesenchymal and ESCs are epithelial-like, and as cells reprogram a near-uniform MET occurs [5,6]. OSKM are directly involved in the MET: OCT4, SOX2, and c-MYC inhibit TGF-β signaling and therefore repress SNAIL, the master transcription factor regulator of the epithelial-to-mesenchymal transition [120], while KLF4 directly activates epithelial genes including *Cdh1* [5]. Of interest, in human OSKM reprogramming the MET occurs much later than in mouse, and this may partly explain why reprogramming human cells is so lengthy [11].

Reprogramming involves a metabolic switch from a mitochondrial-based oxidative, to a glycolytic, metabolic program, a process seemingly independent of exogenous c-MYC [8]. This metabolic switch achieves the goal of quicker energy production but also aims to suppress reactive oxygen species generation by mitochondria because these are mostly detrimental for reprogramming [121]. Consequently, genes and chemical compounds that enhance glycolysis promote human reprogramming, such as the transcription factor hypoxia-inducible factor 1/2α (HIF1/2α) [122,123], although paradoxically in the late phase HIF2α is inhibitory because it induces the pro-apoptotic gene *TRAIL* [123].