Cell fate reprogramming through engineering of native transcription factors

Ralf Jauch

Genome Regulation Laboratory, CAS Key Laboratory of Regenerative Biology, Joint School of Life Sciences, Guangzhou Institutes of Biomedicine and Health, Chinese Academy of Sciences, Guangzhou 510530, China； Guangzhou Medical University, Guangzhou 511436, China.

School of Biomedical Sciences, Li Ka Shing Faculty of Medicine, The University of Hong Kong, Pokfulam, Hong Kong, China
e-mail: ralf@hku.hk

Abstract (120 words)

Cellular reprogramming using cocktails of transcription factors (TFs) affirms the epigenetic and developmental plasticity of mammalian cells. It demonstrates the ability of TFs to 'read' genetic information and to rewire regulatory networks in different cellular contexts. Silenced chromatin is not an impediment to the genome engagement by ectopically expressed TFs. Reprogramming TFs have been identified in diverse structural families that lack shared domains or sequence motifs. Interestingly, the reprogramming activity of non-redundant paralogous TFs can be switched with a few point mutations. These findings revealed that the sequence-function relationships influencing reprogramming are tied to subtle features directing genome wide binding. Therefore, endogenous reprogramming TFs are amenable to directed biomolecular engineering that opens up new avenues to optimize cell fate conversions.

Key words

Cellular reprogramming; Transcription factors; Protein engineering

Introduction

Reprogramming experiments have demonstrated that small sets of TFs are capable of activating genes in cells where they would otherwise be permanently silenced. As a consequence, non-natural cell state conversions occur even between cells from distant developmental domains. This discovery showed that the specifications of cell fates that follow tightly controlled hierarchical transitions *in vivo* are in fact reversible. Reprogramming assays provide a powerful tool to study the sequence-structure-function relationships in transcriptional control. Specifically, the directed transdifferentiation of cells between different lineages ('lineage reprogramming') and the reprogramming of somatic cells to induced pluripotent stem cells (iPSCs, 'pluripotency reprogramming') have provided a system to study how successive changes to the binding landscape of reprogramming TFs, associated changes in chromatin and gene expression programs direct cell fate transitions. TFs that direct cell fate conversions are typically identified through the sampling of candidate molecules that define the desired target cell; i.e. factors expressed in embryonic stem cells drive the induction of pluripotency [1]. However, lineage-specifiers unrelated to the pluripotency network could also be successfully used to direct cellular reprogramming [2,3]. This raises the question whether it is not the biological role in the target cell that confers reprogramming capacity. Rather, the presence of distinctive structural features allowing for the penetration of the 'hostile' epigenetic environment of the donor cells could endow a select group of factors to be able to master this feat. Indeed, it has been suggested that the pluripotency inducer Oct4, Sox2 and Klf4 (abbreviated as OSK) possess so-called 'pioneering' activity [4,5]. Pioneer TFs are defined as being able to access DNA in the context of nucleosome core particles [6]. As a consequence, chromatin is de-compacted and associated genes are activated. Because chromatin remodeling is a hallmark of cell fate conversions, it is believed that reprogramming TFs must also be the pioneer factors [6]. However, reprogramming TFs could be identified from a large range of unrelated classes of TFs without any detectable structural similarity. Some factors within a family of paralogous TFs were defined to possess pioneering activity whereas closely related factors do not [4,7]. Here I discuss insights from the time-resolved epigenetic profiling of pluripotency and neural lineage reprogramming as well as studies that biochemically dissect reprogramming TFs. Further decoding of the structural requirement directing chromatin binding and cell fate reprogramming will enable the re-design of more potent catalysts of artificial cell fate transitions.

Main text

Genomic time lapse of reprogramming factors in action

Detailed dissections of the cellular, genomic and molecular events during reprogramming uncovered mechanisms of how transcription factors bring about the radical changes associated with cell fate conversions [8-10]. Cataloging changes to gene expression programs has led to the definition of two transcriptional waves [11]. One is initiated immediately after the exogenous expression of reprogramming factors and is characterized by the mesenchymal-to-epithelial transition and the silencing of somatic genes. The second wave follows an extended transition period leading to the activation of the pluripotency gene expression network. A third 'stabilization' phase has been associated with the ultimate silencing of the ectopic transgenes [12].

Chromatin immunoprecipitation sequencing (ChIP-seq) has been used to profile genome engagement by reprogramming factors at fine-grained resolution [4,13-16]. The occupancy pattern is highly dynamic and includes both permanently bound enhancers as well as sites that

are bound at restricted reprogramming stages (i.e. early, transient or late). The precise role of reprogramming TFs remains a subject of debate. In one view, OSK binding is highly sensitive to the epigenetic pre-disposition and the openness of their binding sites. It is suggested that DNA demethylation in many instances precedes Oct4 binding and that enhancers with primed H3K4me1 marks are preferentially targeted leading to the subsequent deposition of the activation mark H3K27ac [14]. Further, reprogramming TFs were reported to predominantly target pre-opened chromatin after their introduction into donor cells [14,15]. This leads to a multi-faceted interplay among the exogenous reprogramming factors (OSK) and the TFs that guard the somatic state [15]. A direct competition with somatic TFs and their active displacement from somatic enhancers was proposed to be a main consequence of OSK expression [15]. In this model chromatin changes precede OSK binding and accessibility is a main determinant for OSK occupancy [14,15].
In an alternative model, OSK effectively bind closed chromatin and actively direct changes to chromatin states. A key challenge of monitoring chromatin changes is the very low efficiency of reprogramming experiments. In pluripotency reprogramming, only a small fraction of donor fibroblasts successfully converts into iPSCs. As a consequence, bulk cell population at different reprogramming stages contains a mixture of refractory cells and cells that will successfully reprogram. Three recent studies tackled this problem by using a highly effective chemically defined medium [17], by using surface markers to separate cells that successfully reprogram from cells that fail [16] or by depleting components of the NuRD repressor complex leading to near deterministic reprogramming [13]. One of these studies inferred OSK binding by interrogating the presence of binding DNA motifs in open regions as determined by ATAC-seq (Assay for Transposase-Accessible Chromatin using sequencing) [17] whilst the two other studies integrated ATAC-seq and ChIP-seq for OS [16] or OSKM (OSK plus c-Myc) [13]. Collectively, results indicate that OSK utilize full consensus motifs to predominantly target closed chromatin in MEFs whilst only a minor fraction is recruited to active and pre-opened chromatin. The DNA methylation state has no major impact on Oct4 binding [16]. In this model, the compact chromatin of pluripotency enhancers containing nucleosome bound DNA is not a barrier for the initial genome engagement by reprogramming TFs (**Figure 1**). On the contrary, pre-opened somatic sites are devoid of OSK consensus motifs and only sparsely bound by OSK indicating indirect closing of these sites [13,16,17]. The promiscuous recruitment of reprogramming factors to closed chromatin in MEFs leading to the widespread opening could rationalize such a passive silencing mechanism. This way, new and otherwise inaccessible sites become available to somatic TFs and leading to their gradual depletion at their native enhancers and silencing of associated genes (**Figure 1**).

Are all cellular reprogramming systems driven by similar molecular events?

A growing list of TF cocktails could also be used to directly interconvert somatic lineages [18,19]. The directed transdifferentiation of fibroblasts to neurons with a Brn2, Ascl1 and Myt1l (BAM) cocktail showed that even cell types from different germ layers could be trans-differentiated with defined factors [20]. The basic helix-loop-helix (bHLH) factor Ascl1 is the most crucial component of this mix because it alone is able to accomplish neural lineage reprogramming albeit with reduced efficiency [21]. The potency of Ascl1 has been ascribed to its ability to pioneer the opening of chromatin. Shortly after its forcible expression, it targets E-boxes (short palindromic DNA elements preferred by all bHLH factors) regardless of their chromatin state [7,22]. This leads to chromatin opening within less than a day. The overall reconfiguration of the chromatin and the maturation of the neural gene expression program follow with a delay of several days or weeks [22]. The companion reprogramming factors Brn2 (a Pit-Oct-Unc (POU) family gene like Oct4) and Myt1l are unable to bind and open

closed chromatin. Nevertheless, at the onset of reprogramming, Myt1l still targets many locations that it would also bind in its 'native' neural environment [23]. Yet, Myt1l sites are in open chromatin and the Ascl1 driven pioneering does not markedly influence its binding pattern. Hence, the modus operandi of Ascl1 and Myt1l appears to be fundamentally different and mutually independent. Functionally, Myt1l actively represses non-neural genes (**Figure 1**). Thus, Myt1l resolves the heterogeneity and counteracting somatic programs initially induced by Ascl1 [24].

This poses the question as to whether pluripotency and lineage reprogramming follow fundamentally different trajectories or whether they are just different flavors of the same guiding principle (excellently reviewed in [25]). The two processes share a number of features including the progression through multiple transcriptional waves, phenotypically ambiguous intermediary states and extensive re-configuration of the chromatin. Moreover, roadblocks and facilitators of pluripotency reprogramming were also found to affect lineage reprogramming. For example, suppression of the chromatin assembly factor-1 (CAF-1) not only substantially improved pluripotency but also lineage reprogramming [26]. Likewise, constitutively active SMAD2/3 was suggested as a broad scope reprogramming facilitator regardless of the system [27]. Collectively, a picture emerges that reprogramming entails promiscuous chromatin opening followed by the action of lineage specifying 'safeguards' (i.e. Esrrb in pluripotency) or repressors (i.e. Myt1l in neural inductions) leading to the dissipation of a 'confused' cell state and faithful lineage commitment (**Figure 1**). The presence of chromatin opening pioneer factors appears to be required for reprogramming competent factor cocktails. Lineage specific TFs act as reprogramming enhancers but are optional because in favorable milieus the fate specification can be directed by endogenously activated factors (**Figure 1**).

The structural requirements for reprogramming capacity are unknown

A select group of factors is able to pioneer cell fate conversions. One could thus expect that the comparison of sequences or structural features would allow for the discrimination of factors that accomplish this feat. Yet, distinguishing structural domains or sequence motifs could so far not be identified. Rather, reprogramming factors were identified in most of the major mammalian TF families and include structurally highly diverse proteins without obvious evolutionary relationships such as bHLH, bZip (basic leucine zipper), POU, HMG (high mobility group), C_2H_2 zinc fingers, forkhead and GATA family factors [18,19]. However, within a family only few members can reprogram despite a high degree of sequence conservation and barely distinguishable consensus DNA sequences. Therefore, subtle properties of individual otherwise highly similar TFs specify whether they could direct the regulatory programs for cell fate conversion.

Sox2, Oct4 and Klf4 were reported to bind *in vitro* assembled nucleosome core particles with high affinity [5]. Yet, in a nucleosome context the curvature of the DNA changes and parts of cognate binding sequences are occluded by histone proteins. Thus, the binding configuration and the sequence specificity in a nucleosome context are expected to deviate from the available models for isolated TF:DNA complexes. Structures of the DNA binding domains bound to cognate DNA binding sites are available for most classes of TFs [28]. However, no structures of TFs bound to nucleosomes could so far be elucidated.

An elegant combination of biochemical and genomic techniques showed that a subset of otherwise intractable binding sites of Sox2 on nucleosomes is made accessible by PARP-1 [29]. The rotational phasing of Sox elements within nucleosomal DNA determines whether the binding is PARP-1 dependent or independent. This study indicates that pioneer activity is conditional and relies on the presence of co-factors as well as the position of binding elements

within the nucleosomal context.

Reprogramming activity can be switched with point mutations

Mutagenesis and domain mapping studies have begun to decode sequence-function relationships endowing individual TFs with reprogramming activities with a focus on Sox, POU and bHLH families TFs [30,31]. The comparison of reprogramming competent with incompetent members from paralogous families led to important insights. For example, the ectopic expression of MyoD, a bHLH factor, can reprogram fibroblasts to myoblasts [32]. The ubiquitous bHLH factor E12 is normally incapable of this feat but the exchange of three amino acids with MyoD bestowed E12 with myogenic activity [32].

In OSKM cocktails for iPSC generation, Sox2 can be replaced by its most closely related homologues Sox1, Sox3 and Sox15 but not by more distantly related Sox genes such as Sox17 [33-35]. Efforts to convert Sox17 into an iPSC inducer by swapping N or C-terminal domains with Sox2 failed [33]. However, interchanging glutamate 57 of helix 3 of the HMG box of Sox17 or Sox7 with the corresponding lysine 57 of Sox2 converts Sox17 into a high performance inducer of pluripotency in mouse or human cells (factors referred to as Sox17EK or Sox7EK) [33,34]. Mechanistically, this amino acid swap changes the DNA dependent heterodimerisation with Oct4 on alternative versions of composite *SoxOct* DNA elements [34,36-38]. Sox2, and the engineered Sox17EK protein, associate with Oct4 on the canonical *SoxOct* motif, which is found in the enhancers of many pluripotency genes. By contrast, Sox17/Oct4 dimers are preferentially formed on 'compressed' variants of the *SoxOct* element where one nucleotide separating Sox and Oct half-sites is eliminated. The compressed *SoxOct* element is found in enhancers of genes expressed during the differentiation of primitive endoderm [39].

Likewise, under standard reprogramming conditions, other POU family members such as Oct1 or Oct6 cannot replace Oct4 [35]. Mutational analysis of Oct4 showed a loss of reprogramming only for residues within the DNA binding domain but not for residues within the N or C termini of the protein [40]. Identifying functionally critical elements within its DBD is challenging for the bi-partite nature of the POU domain and the presence of a poorly conserved linker. Domain swaps and amino acid exchanges between Oct4 and Oct6 indicated that the linker connecting the two subdomains of the POU confers some unique function to Oct4 [41-43]. Oct6 could recently be converted into a pluripotency inducer by modifying several structural interfaces albeit the reprogramming efficiency of engineered Oct6 does not match wild type Oct4 [42]. One of the critical modifications reduces the preference of Oct6 to homodimerise on a palindromic DNA element termed MORE (more octamer recognition elements) leading to a preference for heterodimerization with Sox2 on *SoxOct* DNA.

Single molecule tracking studies provided detailed insights as to the enhanceosome assembly in pluripotent cells [44]. Protein complexes at Klf4 enhancer are formed in a hierarchical fashion initially nucleated by the Sox2/Oct4 complex. Subsequently, the TFs Esrrb and Stat3 and non-DNA binding factors such as p300 are recruited. Sox2 and Oct4 have a longer residence time on chromatin than other TFs presumably because of their direct dimerization on specific DNA sequences. Whether this dimerization is also required for their pioneer activity at the onset of reprogramming or only for enhanceosome assembly in pluripotent cells is currently unclear.

Collectively, reprogramming factors rely on specific molecular interfaces that mediate protein partnerships in the context of specific DNA sequences. These partnerships influence site selection in the context of chromatin. By contrast, transactivation domains and residues involved in the direct base readout of DNA sequences are unlikely candidates to set reprogramming factors functionally apart. Whilst the fusion proteins of reprogramming

factors with potent transactivation domains can accelerate and enhance reprogramming, they do not change reprogramming trajectories [45-50]. The uniqueness of reprogramming TFs is predominantly hardwired in their DNA binding domains. Sequence substitutions to these domains globally change the profile of chromatin engagement at the onset of reprogramming in an overall promiscuous but 'lineage biased' manner (**Figure 1**).

Outlook: Lessons for re-designing reprogramming factors

Recent genomic studies indicate that genomic and cellular events concomitant with pluripotency and lineage reprogramming are more similar than previously thought. This could explain why closely related factor combinations could be used to reprogram different lineages (such as the cocktails for iPSC generation (Oct4, Sox2, Klf4, c-Myc [1]) and transdifferentiation to neural progenitors (Brn4, Sox2, Klf4, c-Myc [12]). Nevertheless, reprogramming factors were identified from many TF families (reviewed in [18,19]). These factors lack shared molecular features other than structurally diverse domains mediating sequence-specific DNA recognition. A biological function in the target cell is not a necessary condition for reprogramming activity as demonstrated by the replacement of pluripotency reprogramming factors with lineage specifiers ('seesaw model') and the ability of the pluripotency factor Oct4 to also induce post-mitotic neurons [2,3,51]. This raises the question whether certain TF classes can be 'tuned' so they work for any donor/target cell combination. The common application of the OSKM cocktail in cell activation signalling directed (CASD) reprogramming approach has first outlined this possibility [52]. However, lineage-tracing experiments suggested that OSKM driven reprogramming transits through a pluripotent state before differentiating into a target lineage [53,54]. Such a pluripotent intermediate state could pose a safety concern for regenerative medicine application.

The activity of reprogramming TFs can be switched and enhanced with a limited number of point mutations. Therefore, modifications to existing TF scaffolds could suffice to navigate the starting cell population into any desired direction. Taking advantage of this finding, we have recently established a screening method termed directed evolution of reprogramming factors by cell selection and sequencing (DERBY-seq) [55]. DERBY-seq enables the discovery of artificially evolved and enhanced transcription factors (eTFs) from libraries generated by randomizing critical molecular interfaces. eTFs are discovered by performing reprogramming experiments with pooled libraries, sorting reprogramming and non-reprogramming cells using phenotypic read-outs followed by genotyping through amplicon sequencing (**Figure 2**). If the feasibility of DERBY-seq could also be demonstrated in lineage reprogramming, an all-purpose solution for TF mediated reprogramming and transdifferentiation could be provided by a pooled library of molecular scaffolds capable of potently binding and activating the closed chromatin of the donor cells (**Figure 2**). Some members of these libraries will have the capacity to generate plastic intermediates biased towards the desired lineage (**Figure 2**). The maturation of the lineage would next be safeguarded by cues inherent to culture conditions and endogenously activated regulatory networks. Further insights into the sequence function relationships dictating chromatin engagement, target gene selection, and regulation would accelerate TF re-engineering by better-informed designs. The ultimate goal would be a library of tailored TF scaffolds from which cocktails could be selected governing the interconversion of any given donor/target cell combination.

Acknowledgements

R.J. is supported by the Ministry of Science and Technology of China (2016YFA0100700, 2017YFA0105103) by the National Natural Science Foundation of China (Grant No. 31471238, 31771454, 31611130038), a 100-talent award of the Chinese Academy of Sciences and by Science and Technology Planning Projects of Guangdong Province, China (2017B030314056 and 2016A050503038). I thank Rory Baldwin Johnson, Vikas Malik and Veeramohan Veerapandian for critical comments on the manuscript and help with preparing the illustrations.

Figure legends

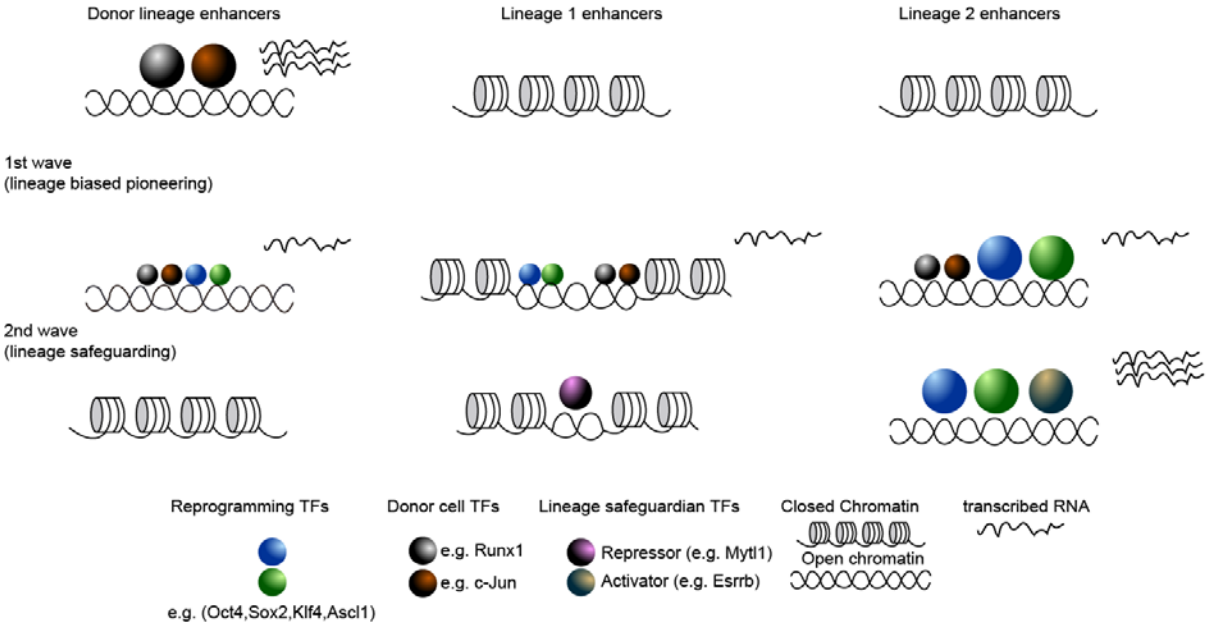

Figure 1: Scheme for TF driven reprogramming
The ability to bind closed and silenced chromatin appears to be a hallmark feature of both pluripotency and lineage reprogramming TFs. They only moderately bind donor cell enhancers and the somatic silencing occurs mainly by a passive re-distribution of donor cell TFs to newly opened locations. The initial genome engagement by reprogramming TFs is rather promiscuous but biased towards a certain lineage because of sequence specificities conferred by the DNA binding domains of the individual factors which can be profoundly influenced by TF dimerization. The maturation of a certain lineage depends on lineage safeguards (eg. Myl1 or Esrrb) that can either be provided as accessory components of the reprogramming cocktail or endogenously activated.

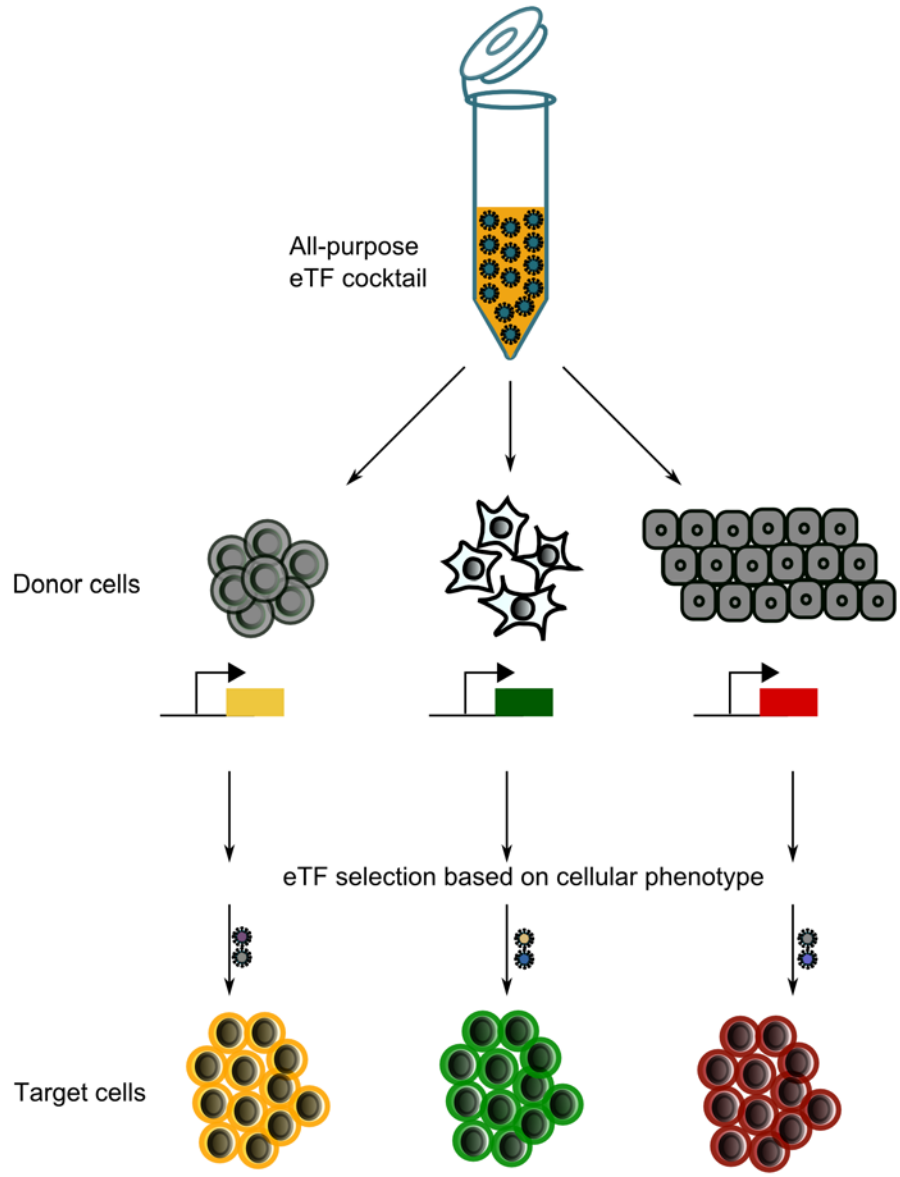

Figure 2: Towards an all-purpose reprogramming strategy based on the selection of artificially enhanced TFs from pooled libraries

Reprogramming TFs can be tailored by modifying their molecular interfaces and even point mutations can allow individual TFs to acquire new functions. Therefore, the structural scaffolds of endogenous reprogramming TFs could be modified to produce all-purpose libraries. These libraries can be screened for any given donor/target cell combinations to select for cocktails of artificially evolved and enhanced TFs (eTFs) that drive the desired cell fate conversion.

Annotated reference from last 2 years:

****Chronis .. Plath, 2017**
'In depth epigenetic characterization of four states during the mouse pluripotency reprogramming reporting that somatic TFs collaborate with OSKM on various set of binding sites including somatic enhancers and they only successively arrive at pluripotency enhancers.

****Knaup .. Polo, 2017**
'Interrogation of OS binding and epigenetic changes during mouse pluripotency reprogramming after sorting of reprogramming and non-reprogramming cells reveals the direct targeting of closed chromatin in fibroblasts bearing consensus SOX and POU motifs regardless of the underlying methylation state.'

****Zviran .. Hanna, 2017**
'Using a near-deterministic mouse iPSC generations system OSKM driven reprogramming was epigenetically dissected day-by-day indicating that OSK, but not M, predominantly bind enhancers that are closed in MEFs leading to the removal of repressive epigenetic marks followed by chromatin opening.'

****Li .. Pei 2017**
'Time resolved ATAC-seq study using a highly efficient chemically defined mouse iPSC generation system showing that OSK indirectly closes somatic enhancers via SAP30 and actively directs the opening of closed chromatin mediated by canonical DNA sequence motifs.'

***Mall..Wernig,2017**
'Demonstration that Myt1lacts as a 'many-but-one' repressor inhibiting lineages except the neuronal one; counteracting the promiscuous pioneering activity of Ascl1.'

****Wapinski .. Chang , 2017**
'Time resolved ATAC-seq study during the transdifferentiation of murine fibroblasts to neurons showing that Ascl1 is able to bind E-box motifs in closed chromatin leading to rapid opening including sites not associated with neuronal genes.'

***Liu .. Kraus, 2017**
'Demonstration that the interaction with PARP1 is necessary for the binding of Sox2 to specific subsets of DNA binding sites in the context of inaccessible chromatin with high nucleosome occupancy'

****Xie .. Tjian 2017**
'Demonstration that Sox2/Oct4 binding nucleates an enhanceosome regulating Klf4 by facilitating the target search of Essrb and Stat3. Constructive assembly requires precisely grafted enhancer architecture. '

*** Jerabek .. Jauch, 2017**
'This study shows that Oct4 evolved a reduced preference to homodimerise on palindromic DNA sequences compared to somatic POU factors. This insight led to the conversion of Oct6 into a pluripotency inducer by protein engineering.'

***Ruetz .. Kaji 2017**
'Constitutively active Smad2/3 were shown to facilitate not only pluripotency but also lineage reprogramming thus suggesting a shared mechanism for cell fate conversions.'

***Chen..Gao, 2016**
This study profiles Oct4 binding and chromatin changes during mouse pluripotency reprogramming at high resolution and highlights a multitude of intricate binding patterns and a preference of Oct4 for primed and open enhancers.

***Veerapandian..Jauch, 2018**
'Demonstration that transcription factors are not optimally adapted to artificial reprogramming conditions but can be enhanced by directed evolution using the phenotype of reprogramming cells as readout.

****Tsunemoto..Baldwin, 2018**
This study shows that multiple transcription factor pairs, including the pluripotency gene Oct4, are able to induce the transdifferentiation of fibroblasts to post-mitotic neurons (iN). The resulting cells exhibit nuanced differences

in their expression signatures that depend on the donor TF pairs suggesting that neural subtypes can be produced by subtle modifications to the two-factor cocktail.

References

1. Takahashi K, Yamanaka S: **Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors**. *Cell* 2006, **126**:663-676.
2. Montserrat N, Nivet E, Sancho-Martinez I, Hishida T, Kumar S, Miquel L, Cortina C, Hishida Y, Xia Y, Esteban CR, et al.: **Reprogramming of human fibroblasts to pluripotency with lineage specifiers**. *Cell Stem Cell* 2013, **13**:341-350.
3. Shu J, Wu C, Wu Y, Li Z, Shao S, Zhao W, Tang X, Yang H, Shen L, Zuo X, et al.: **Induction of pluripotency in mouse somatic cells with lineage specifiers**. *Cell* 2013, **153**:963-975.
4. Soufi A, Donahue G, Zaret KS: **Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome**. *Cell* 2012, **151**:994-1004.
5. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS: **Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming**. *Cell* 2015, **161**:555-568.
6. Zaret KS, Mango SE: **Pioneer transcription factors, chromatin dynamics, and cell fate control**. *Curr Opin Genet Dev* 2016, **37**:76-81.
7. Wapinski OL, Vierbuchen T, Qu K, Lee QY, Chanda S, Fuentes DR, Giresi PG, Ng YH, Marro S, Neff NF, et al.: **Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons**. *Cell* 2013, **155**:621-635.
8. Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H, Jaenisch R: **Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells**. *Cell Stem Cell* 2008, **2**:151-159.
9. Stadtfeld M, Maherali N, Breault DT, Hochedlinger K: **Defining molecular cornerstones during fibroblast to iPS cell reprogramming in mouse**. *Cell Stem Cell* 2008, **2**:230-240.
10. O'Malley J, Skylaki S, Iwabuchi KA, Chantzoura E, Ruetz T, Johnsson A, Tomlinson SR, Linnarsson S, Kaji K: **High-resolution analysis with novel cell-surface markers identifies routes to iPS cells**. *Nature* 2013, **499**:88-91.
11. Polo JM, Anderssen E, Walsh RM, Schwarz BA, Nefzger CM, Lim SM, Borkent M, Apostolou E, Alaei S, Cloutier J, et al.: **A molecular roadmap of reprogramming somatic cells into iPS cells**. *Cell* 2012, **151**:1617-1632.
12. Golipour A, David L, Liu Y, Jayakumaran G, Hirsch CL, Trcka D, Wrana JL: **A late transition in somatic cell reprogramming requires regulators distinct from the pluripotency network**. *Cell Stem Cell* 2012, **11**:769-782.
13. Asaf Zviran NM, Yoach Rais, Hila Gingold, Shani Peles, Elad Chomsky, Sergey Viukov, Jason D. Buenrostro, Leehee Weinberger, Yair S. Manor, Vladislav Krupalnik, Mirie Zerbib, Hadas Hezroni, Diego Adhemar Jaitin, David Larastiaso, Shlomit Gilad, Sima Benjamin, Awni Mousa, Muneef Ayyash, Daoud Sheban, Jonathan Bayerl, Alejandro Aguilera Castrejon, Rada Massarwa, Itay Maza, Suhair Hanna, Ido Amit, Yonatan Stelzer, Igor Ulitsky, William J. Greenleaf, Yitzhak Pilpel, Noa Novershtern, Jacob H. Hanna: **High-Resolution Dissection of Conducive Reprogramming Trajectory to Ground State Pluripotency**. *bioRxiv* 2017.
14. Chen J, Chen X, Li M, Liu X, Gao Y, Kou X, Zhao Y, Zheng W, Zhang X, Huo Y, et al.: **Hierarchical Oct4 Binding in Concert with Primed Epigenetic Rearrangements during Somatic Cell Reprogramming**. *Cell Rep* 2016, **14**:1540-1554.
15. Chronis C, Fiziev P, Papp B, Butz S, Bonora G, Sabri S, Ernst J, Plath K: **Cooperative Binding of Transcription Factors Orchestrates Reprogramming**. *Cell* 2017, **168**:442-459 e420.
16. Knaupp AS, Buckberry S, Pflueger J, Lim SM, Ford E, Larcombe MR, Rossello FJ, de Mendoza A, Alaei S, Firas J, et al.: **Transient and Permanent Reconfiguration of Chromatin and Transcription Factor Occupancy Drive Reprogramming**. *Cell Stem Cell* 2017, **21**:834-845 e836.
17. Li D, Liu J, Yang X, Zhou C, Guo J, Wu C, Qin Y, Guo L, He J, Yu S, et al.: **Chromatin Accessibility Dynamics during iPSC Reprogramming**. *Cell Stem Cell* 2017, **21**:819-833 e816.
18. Graf T, Enver T: **Forcing cells to change lineages**. *Nature* 2009, **462**:587-594.
19. Guo C, Morris SA: **Engineering cell identity: establishing new gene regulatory and chromatin landscapes**. *Curr Opin Genet Dev* 2017, **46**:50-57.

20. Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC, Wernig M: **Direct conversion of fibroblasts to functional neurons by defined factors**. *Nature* 2010, **463**:1035-1041.
21. Chanda S, Ang CE, Davila J, Pak C, Mall M, Lee QY, Ahlenius H, Jung SW, Sudhof TC, Wernig M: **Generation of induced neuronal cells by the single reprogramming factor ASCL1**. *Stem Cell Reports* 2014, **3**:282-296.
22. Wapinski OL, Lee QY, Chen AC, Li R, Corces MR, Ang CE, Treutlein B, Xiang C, Baubet V, Suchy FP, et al.: **Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons**. *Cell Rep* 2017, **20**:3236-3247.
23. Mall M, Kareta MS, Chanda S, Ahlenius H, Perotti N, Zhou B, Grieder SD, Ge X, Drake S, Euong Ang C, et al.: **Myt1l safeguards neuronal identity by actively repressing many non-neuronal fates**. *Nature* 2017, **544**:245-249.
24. Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SA, Sim S, Neff NF, Skotheim JM, Wernig M, et al.: **Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq**. *Nature* 2016, **534**:391-395.
25. Firas J, Polo JM: **Towards understanding transcriptional networks in cellular reprogramming**. *Curr Opin Genet Dev* 2017, **46**:1-8.
26. Cheloufi S, Elling U, Hopfgartner B, Jung YL, Murn J, Ninova M, Hubmann M, Badeaux AI, Euong Ang C, Tenen D, et al.: **The histone chaperone CAF-1 safeguards somatic cell identity**. *Nature* 2015, **528**:218-224.
27. Ruetz T, Pfisterer U, Di Stefano B, Ashmore J, Beniazza M, Tian TV, Kaemena DF, Tosti L, Tan W, Manning JR, et al.: **Constitutively Active SMAD2/3 Are Broad-Scope Potentiators of Transcription-Factor-Mediated Cellular Reprogramming**. *Cell Stem Cell* 2017, **21**:791-805 e799.
28. Cramer P: **A tale of chromatin and transcription in 100 structures**. *Cell* 2014, **159**:985-994.
29. Liu Z, Kraus WL: **Catalytic-Independent Functions of PARP-1 Determine Sox2 Pioneer Activity at Intractable Genomic Loci**. *Mol Cell* 2017, **65**:589-603 e589.
30. Hou L, Srivastava Y, Jauch R: **Molecular basis for the genome engagement by Sox proteins**. *Semin Cell Dev Biol* 2016.
31. Malik V, Zimmer D, Jauch R: **Diversity among POU transcription factors in chromatin recognition and cell fate reprogramming**. *Cell Mol Life Sci* 2018.
32. Davis RL, Weintraub H: **Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12**. *Science* 1992, **256**:1027-1030.
33. Aksoy I, Jauch R, Eras V, Chng WB, Chen J, Divakar U, Ng CK, Kolatkar PR, Stanton LW: **Sox transcription factors require selective interactions with Oct4 and specific transactivation functions to mediate reprogramming**. *Stem Cells* 2013, **31**:2632-2646.
34. Jauch R, Aksoy I, Hutchins AP, Ng CK, Tian XF, Chen J, Palasingam P, Robson P, Stanton LW, Kolatkar PR: **Conversion of Sox17 into a pluripotency reprogramming factor by reengineering its association with Oct4 on DNA**. *Stem Cells* 2011, **29**:940-951.
35. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, Aoi T, Okita K, Mochiduki Y, Takizawa N, Yamanaka S: **Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts**. *Nat Biotechnol* 2008, **26**:101-106.
36. Ng CK, Li NX, Chee S, Prabhakar S, Kolatkar PR, Jauch R: **Deciphering the Sox-Oct partner code by quantitative cooperativity measurements**. *Nucleic Acids Res* 2012, **40**:4933-4941.
37. Merino F, Ng CK, Veerapandian V, Scholer HR, Jauch R, Cojocaru V: **Structural Basis for the SOX-Dependent Genomic Redistribution of OCT4 in Stem Cell Differentiation**. *Structure* 2014.
38. Chang YK, Srivastava Y, Hu C, Joyce A, Yang X, Zuo Z, Havranek JJ, Stormo GD, Jauch R: **Quantitative profiling of selective Sox/POU pairing on hundreds of sequences in parallel by Coop-seq**. *Nucleic Acids Res* 2017, **45**:832-845.
39. Aksoy I, Jauch R, Chen J, Dyla M, Divakar U, Bogu GK, Teo R, Leng Ng CK, Herath W, Lili S, et al.: **Oct4 switches partnering from Sox2 to Sox17 to reinterpret the enhancer code and specify endoderm**. *EMBO J* 2013, **32**:938-953.
40. Jin W, Wang L, Zhu F, Tan W, Lin W, Chen D, Sun Q, Xia Z: **Critical POU domain residues confer Oct4 uniqueness in somatic cell reprogramming**. *Sci Rep* 2016, **6**:20818.
41. Esch D, Vahokoski J, Groves MR, Pogenberg V, Cojocaru V, Vom Bruch H, Han D, Drexler HC, Arauzo-Bravo MJ, Ng CK, et al.: **A unique Oct4 interface is crucial for reprogramming to pluripotency**. *Nat Cell Biol* 2013, **15**:295-301.
42. Jerabek S, Ng CK, Wu G, Arauzo-Bravo MJ, Kim KP, Esch D, Malik V, Chen Y, Velychko S, MacCarthy CM, et al.: **Changing POU dimerization preferences converts Oct6 into a pluripotency inducer**. *EMBO Rep* 2017, **18**:319-333.

43. Kong X, Liu J, Li L, Yue L, Zhang L, Jiang H, Xie X, Luo C: **Functional interplay between the RK motif and linker segment dictates Oct4-DNA recognition**. *Nucleic Acids Res* 2015, **43**:4381-4392.
44. Xie L, Torigoe SE, Xiao J, Mai DH, Li L, Davis FP, Dong P, Marie-Nelly H, Grimm J, Lavis L, et al.: **A dynamic interplay of enhancer elements regulates Klf4 expression in naive pluripotency**. *Genes Dev* 2017, **31**:1795-1808.
45. Hammachi F, Morrison GM, Sharov AA, Livigni A, Narayan S, Papapetrou EP, O'Malley J, Kaji K, Ko MS, Ptashne M, et al.: **Transcriptional activation by Oct4 is sufficient for the maintenance and induction of pluripotency**. *Cell Rep* 2012, **1**:99-109.
46. Hirai H, Katoku-Kikyo N, Keirstead SA, Kikyo N: **Accelerated direct reprogramming of fibroblasts into cardiomyocyte-like cells with the MyoD transactivation domain**. *Cardiovasc Res* 2013, **100**:105-113.
47. Hirai H, Tani T, Katoku-Kikyo N, Kellner S, Karian P, Firpo M, Kikyo N: **Radical acceleration of nuclear reprogramming by chromatin remodeling with the transactivation domain of MyoD**. *Stem Cells* 2011, **29**:1349-1361.
48. Wang Y, Chen J, Hu JL, Wei XX, Qin D, Gao J, Zhang L, Jiang J, Li JS, Liu J, et al.: **Reprogramming of mouse and human somatic cells by high-performance engineered factors**. *EMBO Rep* 2011, **12**:373-378.
49. Wang L, Huang D, Huang C, Yin Y, Vali K, Zhang M, Tang Y: **Enhanced human somatic cell reprogramming efficiency by fusion of the MYC transactivation domain and OCT4**. *Stem Cell Res* 2017, **25**:88-97.
50. Zhu G, Li Y, Zhu F, Wang T, Jin W, Mu W, Lin W, Tan W, Li W, Street RC, et al.: **Coordination of engineered factors with TET1/2 promotes early-stage epigenetic modification during somatic cell reprogramming**. *Stem Cell Reports* 2014, **2**:253-261.
51. Tsunemoto R, Lee S, Szucs A, Chubukov P, Sokolova I, Blanchard JW, Eade KT, Bruggemann J, Wu C, Torkamani A, et al.: **Diverse reprogramming codes for neuronal identity**. *Nature* 2018.
52. Yu C, Liu K, Tang S, Ding S: **Chemical approaches to cell reprogramming**. *Curr Opin Genet Dev* 2014, **28**:50-56.
53. Bar-Nur O, Verheul C, Sommer AG, Brumbaugh J, Schwarz BA, Lipchina I, Huebner AJ, Mostoslavsky G, Hochedlinger K: **Lineage conversion induced by pluripotency factors involves transient passage through an iPSC stage**. *Nat Biotechnol* 2015, **33**:761-768.
54. Maza I, Caspi I, Zviran A, Chomsky E, Rais Y, Viukov S, Geula S, Buenrostro JD, Weinberger L, Krupalnik V, et al.: **Transient acquisition of pluripotency during somatic cell transdifferentiation with iPSC reprogramming factors**. *Nat Biotechnol* 2015, **33**:769-774.
55. Veerapandian V, Ackermann O, Srivastava Y, Malik V, Weng M, Yang X, Jauch R: **Directed eveoltuion of reprogramming factors by cell selection and sequencing**. *revised manuscript in preparation* 2018.