Human transcription factors responsive to initial reprogramming predominantly undergo legitimate reprogramming during fibroblast conversion to iPSCs

Ricardo R. Cevallos1, Yvonne J. K. Edwards1,2, John M. Parant3, Bradley K. Yoder2 & Kejin Hu1,*

The four transcription factors OCT4, SOX2, KLF4, and MYC (OSKM) together can convert human fibroblasts to induced pluripotent stem cells (iPSCs). It is, however, perplexing that they can do so only for a rare population of the starting cells with a long latency. Transcription factors (TFs) define identities of both the starting fibroblasts and the end product, iPSCs, and are also of paramount importance for the reprogramming process. It is critical to upregulate or activate the iPSC-enriched TFs while downregulate or silence the fibroblast-enriched TFs. This report explores the initial TF responses to OSKM as the molecular underpinnings for both the potency aspects and the limitation sides of the OSKM reprogramming. The authors first defined the TF reprogramome, i.e., the full complement of TFs to be reprogrammed. Most TFs were resistant to OSKM reprogramming at the initial stages, an observation consistent with the inefficiency and long latency of iPSC reprogramming. Surprisingly, the current analyses also revealed that most of the TFs (at least 83 genes) that did respond to OSKM induction underwent legitimate reprogramming. The initial legitimate transcriptional responses of TFs to OSKM reprogramming were also observed in the reprogramming fibroblasts from a different individual. Such early biased legitimate reprogramming of the responsive TFs aligns well with the robustness aspect of the otherwise inefficient and stochastic OSKM reprogramming.

1Department of Biochemistry and Molecular Genetics, School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA. 2Department of Cell Developmental and Integrative Biology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA. 3Department of Pharmacology and Toxicology, School of Medicine, University of Alabama at Birmingham, Birmingham, AL 35294, USA. *email: kejinhu@uab.edu
in the previous research, we recently developed the concept of reprogramome, which is the full complement of genes that should be reprogrammed. Based on the concept of reprogramome, we further developed another concept of reprogramming legitimacy, and these new concepts allow logical evaluations of an early transcriptional response of a gene to the reprogramming factors.

Each cell type has a defined and specific transcriptome. The transcriptome is governed by a defined set of transcription factors. iPSC generation, in essence, is a process of transcriptional reprogramming. To convert fibroblasts into iPSCs, it is essential to erase the fibroblast transcriptional program, and at the same time to establish the pluripotency transcriptional program. Among the transcriptional reprogramming of the entire reprogramome, the most critical process should be to erase the fibroblast transcription factor (TF) system, and at the same time to establish the pluripotent TF network. It is not clear what the full differences in the TF networks are between the starting fibroblasts and the endpoint iPSCs. Our previous analyses of reprogramming legitimacy were general in nature, and did not specifically examine transcriptional responses of the transcriptional factors to OSKM reprogramming.

Realizing the critical roles of TFs in defining cellular identities and cellular reprogramming, this research employed our new concepts of reprogramome and reprogramming legitimacy to evaluate specifically the early responses of human TFs to the OSKM reprogramming. The authors first defined the TF differences, that is, the set of transcription factors that should be reprogrammed, i.e., the TF reprogramome. Specifically, the repertoire of fibroblast-enriched and specific transcriptional factors, the TF downreprogramome, was defined. Similarly, the TF upreprogramome was defined. With the defined TF down- and up-reprogramomes, the reprogramming legitimacy was then evaluated. In agreement with the extremely low efficiency of reprogramming, we found that most members of the TF reprogramome were not responsive to OSKM reprogramming. Surprisingly, we also found that among the TFs that did respond to OSKM reprogramming, most of the transcriptional responses of TFs were legitimate in the context of reprogramming. The aberrant and unwanted reprogramming were also observed among the responsive TFs. The legitimate reprogramming of the majority of the responsive transcription factors to the early OSKM induction aligns well with the potency aspect of OSKM reprogramming, but at the same time, resistance of some transcription factors to OSKM induction along with some aberrant and unwanted TF reprogramming may underlie the low efficiency, long latency, and stochastic nature of iPSC reprogramming.

**Results**

**Defining the set of transcriptional factors to be reprogrammed, the TF reprogramome.** Transcription factors (TF) are critical in defining any cell type. Human pluripotent stem cells (PSCs) should have a defined set of TFs, so do the starting cells for iPSC reprogramming, fibroblasts. In order to convert human fibroblasts to iPSCs, it is of paramount importance to reprogram the TFs to the expression levels of the pluripotent state from that of fibroblasts. It is not clear what the full TF differences are between the pluripotent cells and the starting somatic fibroblasts, i.e., the TF reprogramome. In order to find out the TF transcriptional differences between PSCs and fibroblasts, this study compared the expression of the entire set of human TFs based on RNA-seq data we recently published. Several groups attempted to define the repertoire of human TFs. The latest revised version by Lambert et al. was used in this study. The RNA-seq data were extracted for the Lambert set of 1639 human TFs, but the current report concerns with 1636 TFs only because ZNF788 is a pseudogene, and DUX1 and DUX3 were not annotated in the Ensembl database. Of the 1636 TFs, 315 were not expressed in both ESCs and fibroblasts, while 442 TFs were expressed in both cell types at similar levels. Two hundred and seventy-nine (279) TFs were enriched by at least 2-fold (q < 0.01) in fibroblasts compared to human ESCs, and they constitute the TF downreprogramome (Table S2, and Fig. 1A,D,E, and Fig. S1). There are 110 zinc finger, 18 HOX, 11 forkhead box, and 7 T-box TF genes in the downreprogramome. Within the TF downreprogramome, 93 TFs were expressed in fibroblasts only, constituting the TF erasome (Table S3, Fig. 1D). Of note, all of the 18 HOX TF genes are in the erasome, and there are 32 zinc finger TF genes in the erasome. Excluding the erasome, additional 71 TFs of the downreprogramome were highly enriched (by > 5-fold) (Fig. 1D). In the entire downreprogramome of 279 TFs, 217 were enriched by at least 3-fold.

Three hundred and ten (310) TFs are enriched in hESCs by at least 2-fold, and constitute the TF upreprogramome (Table S4, Fig. 1A–C, and Fig. S2). As expected, the established pluripotent TFs are in the upreprogramome including POU5F1, NANOG, SOX2, ZFP42, ZSCAN10, FOXD3, PRMD14, ZIC3, SALL4, and others. Interestingly, there are 198 zinc finger TF genes of different types (genes with the designations of ZNF, ZFP, ZIC, ZSCAN, GATA, KLF, SALL, ZBTB, and others) in the upreprogramome. Therefore, zinc finger TFs represent 63.9% of the TF upreprogramome. In contrast to the downreprogramome, there is not a single HOX gene in the upreprogramome. In fact, all the 39 HOX genes of human genome are silenced in hESCs. The upreprogramome contains 8 SOX, 4 POU, and 4 SALL genes but the downreprogramome includes no SALL and SOX genes and only one POU gene. Within the TF upreprogramome, 70 TFs were expressed in hESCs only, constituting the TF activatome (Table S5, Fig. 1B). Excluding the activatome, 80 additional TFs are highly enriched in hESCs (by > 5-fold) (Fig. 1B). There were 28 zinc finger TFs in the activatome. Two hundred and ninety of the human TFs could not be classified into the above categories because they fell into the marginal areas based on the selection criteria used here (Fig. 1A). Combining the TF downreprogramome and upreprogramome, the TF reprogramome includes 589 TFs that should be reprogrammed by at least 2-fold. Interestingly, the downreprogramome is characterized by HOX genes while the upreprogramome is featured by SOX, SALL, and POU TFs, and is dominated by zinc finger TFs of different types.

GO analyses of the upreprogramome and downreprogramome resulted in very different pictures. The majority of the pluripotent TF GO terms were generic while the fibroblast ones were very promiscuous. At the FDR < 0.01 level, the TF downreprogramome was overrepresented by 428 different GO terms (Table S6) while the TF
upreprogramome was overrepresented by 103 GO terms only (Table S7). Of these 103 pluripotent TF GO terms, 90 were shared with that of fibroblasts. Among the 13 unique GO terms for ESCs were “chromatin organization” (26 genes), “stem cell population maintenance” (15), “regulation of cell cycle arrest” (9), and “maintenance of cell number” (15) (Fig. S3). Among these GO terms unique to fibroblasts were “response to chemical” (99 genes), “response to stress” (79), “regulation of programmed cell death” (57), “limb development” (26), “muscle tissue development” (21), “brain development” (27), and “response to mechanical stimulus” (12) (Fig. S4). One interesting GO term in “reactome pathway” analyses of fibroblast-enriched genes was “transcriptional regulation of white adipocyte differentiation” (11 genes). In sum, the enriched GO terms for fibroblast TFs appear to be more tissue-specific and promiscuous, and those for pluripotent TFs are more generic and stem cell-related.

A portion of the TF reprogramome are resistant to reprogramming. Yamanaka reprogramming is very inefficient, slow and stochastic. Previously, we reported that 953 genes are resistant to iPSC reprogramming at the initial stages10. We hypothesized that a portion of transcription factors in the reprogramome is among those genes irresponsive to OSKM reprogramming considering the low efficiency and long latency of iPSC reprogramming. To test this, the TF reprogramome was examined for the reprogramming statuses of its member genes upon OSKM reprogramming. Indeed, 108 out of the 279 fibroblast-enriched TFs were irresponsive at both time points examined (Figs. 2A, Fig. S5, and Table S8), while 188 out of the 310 PSC-enriched TFs demonstrated no significant transcriptional changes by OSKM reprogramming for 72 h (Figs. 2B, S6, and Table S9). Clustering analyses indicated that these irresponsible TFs remained similar to those in the starting naïve fibroblasts and in the fibroblasts transduced with GFP. In the upreprogramome, 124 zinc finger TFs were irresponsive to OSKM reprogramming including 4 ZSCAN zinc finger genes (ZSCAN2, ZSCAN10, ZSCAN16, and ZSCAN31). In the downreprogramome, 13 HOX genes were resistant to OSKM reprogramming.

As expected, 7 ESC-enriched genes with the “reactome pathway” GO term of “transcriptional regulation of pluripotent stem cells” were among the TF genes resistant to reprogramming. These include NANOG, LIN28A, ZSCAN10, FOXD3, PRDM14, ZIC3, and HIF3A. Interestingly, 66 ESC-enriched genes of the “reactome pathway” GO term of “generic transcription pathway” were also resistant to OSKM reprogramming.

Figure 1. Profiling reprogramming of human transcription factors for fibroblast-to-iPSC conversion. (A) Classification of differential expressions for the entire human transcription factor (TF) repertoire. (B) Number of TF genes in different degrees of enrichment in human embryonic stem cells (ESCs) as compared to fibroblasts. (C) A heat map showing enrichment of 310 human TF genes in hESCs. (D) Number of TF genes in different degrees of enrichment in human fibroblasts as compared to hESCs. (E) A heat map showing enrichment of 279 human TF genes in human fibroblasts compared with that in hESCs.
Transcriptional responses of transcription factors to OSKM induction. Next, we defined the set of TF genes significantly upregulated (TF upregulatome of OSKM) and downregulated by OSKM (TF downregulatome of OSKM). In the analyses of transcriptional impact of OSKM on TFs, we used two reference conditions: naïve fibroblasts and the fibroblasts transduced with GFP viruses. The use of GFP control can remove the transcriptional impact by viruses per se. We found that OSKM upregulated 53 TFs by at least 2-fold (q < 0.01) compared to both the naïve fibroblasts and the fibroblasts transduced with the GFP viruses (Fig. S7, and Table S10). As expected, the four reprogramming factors OCT4 (POU5F1), SOX2, KLF4, and MYC were among this list. Therefore, only 49 TFs were significantly upregulated by at least 2-fold at the initial stages. The fold upregulation ranged from 2- to around 17-fold and up to de novo activation of 12 TF genes. On the other hand, 70 TFs were downregulated by OSKM by at least 2-fold (q < 0.01) compared to the naïve fibroblasts and the fibroblasts transduced with GFP viruses (Fig. S8 and Table S11). The fold downregulation ranged from 2- to 14-fold, and up to silencing of 5 TF genes.

Most of the downregulation of TFs by OSKM is legitimate reprogramming. The legitimacy of a transcriptional change of a TF induced by OSKM should be evaluated by the relative expression levels of individual TFs in PSCs to that in fibroblasts (Fig. 3A). Upregulation of a gene is legitimate if its expression is higher in PSCs, while it is not when its expression is lower in PSCs. On the other hand, downregulation of a gene is legitimate if its expression is lower in PSCs while it is not when its expression is higher in PSCs. If the expression level of a gene is similar in both cell types both up- and down-regulations by the OSKM reprogramming factors are illegitimate, which constitute aberrant reprogramming.

Using this logic, we evaluated the 70 TFs downregulated by OSKM for their reprogramming legitimacy. Surprisingly, only one TFs was enriched in ESCs by at least 2-fold and 49 were enriched in fibroblasts by at least 2-fold (Table S12). Because of this biased enrichment of the downregulated TFs for fibroblasts, the criteria were then loosened to a significance level of q < 0.05 for significant differences at any level. Fifty six out of the 70 genes downregulated by OSKM were expressed significantly higher in fibroblasts by at least 1.46-fold, indicating legitimate downreprogramming of these 56 TFs (Fig. 3B). A scrutiny of the 56 genes indicated that 21 were properly reprogrammed and became clustered with ESCs and away from fibroblasts and the fibroblasts transduced with GFP viruses (Figs. 3C, D, S9). Thirty-three of those 56 TFs were downregulated significantly towards the pluripotency levels although the downreprogramming is insufficient (Figs. 3C, E, S10), indicating a positive drive to the pluripotency states with some deficiency. In summary, 80% of the downregulated TFs by OSKM underwent legitimate downreprogramming.
Most of the upregulation of TFs by OSKM is legitimate reprogramming. Next, we evaluated the reprogramming legitimacy of the 49 upregulated TFs. Of note, only three of them were expressed higher in fibroblast by at least 2-fold. Because biased enrichment was observed again, the criteria were then loosened to the significance level of q < 0.05 at any difference level as done for the 70 TFs downregulated by OSKM. New criteria resulted in only 7 TFs that were enriched in fibroblasts by at least 1.44-fold (Fig. 4A). There was no difference in expression for nine of them, of which three were not expressed in both cell types. Impressively, 33 out of the 49 upregulated TFs were enriched in PSCs by at least 1.37-fold (25 TFs by at least 2-fold and 7 by 1.5- to 2-fold) (Fig. 4A), indicating that 67.5% of the upregulation is legitimate reprogramming. Detailed examination showed that 18 of the 33 PSC-enriched TFs were properly upreprogrammed and became clustered with the PSCs and away from the starting fibroblasts and the reprogramming GFP controls (Figs. 4B,C, and Fig. S11), while 11 of them were insufficiently upreprogrammed and remained clustered with the starting cells (Fig. 4B,D, and Fig. S12). However, 7 of the fibroblast-enriched genes upregulated by OSKM were wrongly upreprogrammed and became a separate independent group in clustering analyses (Fig. 4A,E). Of the 9 TFs with similar expression in both cell types, three can be considered as unwanted activation because they are not expressed in both cell types while four are unwanted upreprogramming (data not shown). In summary, legitimate reprogramming is predominant among the 49 upregulated TFs (67.5%).

Independent human fibroblasts from a different individual have a similar TF reprogramome. Fibroblasts are heterogeneous. Although the human fibroblast BJ used above has been widely employed in research of pluripotency reprogramming1-3,18, we investigated whether an independent fibroblast from a different individual has similar TF reprogramome. To this end, we used another model human fibroblast line also used in reprogramming study previously, CRL-2097 (denoted as CRL hereafter)19,20. We also used a different RNA-seq technology DNA nanoball sequencing (DNB-seq) because it is much cheaper than the Illumina.
mina technology we used for the above data. Our additional data indicated that the CRL fibroblast has a very similar TF upreprogramome to that of BJ (Supplementary Fig. S13, Supplementary Table S13). In short, for the new fibroblast CRL, 350 TFs were expressed significantly higher in ESCs, of which 263 TFs were shared with BJ (Fig. S13A,B). Out of the remaining 87 TFs not shared by BJ, 59 were still expressed significantly higher in ESCs than in the BJ when the sorting criteria were loosened (p < 0.05 with significant differences at any levels) (Fig. S13C). At the same time, out of the 47 ESC-enriched TFs in the BJ list but not in the CRL list, 28 were still enriched in ESCs compared to CRL when the selection criteria were loosened (Fig. S13D). Notably, only in one rare case we saw conflicting results between the two fibroblasts. DBP expression was significantly higher (2.1 ×) in ESCs compared to one fibroblast (CRL) but significantly lower (−1.6 ×) in ESCs compared to another fibroblast (BJ).

Similarly, the majority of the TF down-reprogramomes were shared by the two fibroblasts (Supplementary Fig. S14, and Supplementary Table S14).

The initial TF responses to Yamanaka factors in an independent human fibroblast are predominantly legitimate reprogramming. Next, we investigated whether the initial legitimate reprogramming of transcription factors can be observed in an independent fibroblast cell line. For this purpose, we sequenced RNA from the fibroblast CRL undergoing early reprogramming. As in BJ cells, OSKM were all overexpressed well in CRL cells in all samples (Supplementary Fig. S15). OSKM upregulated 219 TF at both 48 and 72 h post factor transduction (Fig. 5A and Supplementary Table S15). As seen with the BJ cells, these upregulated TFs are predominantly ESC-enriched (129 out of 219) (Fig. 5B). Classification of “insufficient up,” “proper up” and “over up” eliminate some genes with legitimate reprogramming because of the stringent sorting criteria applied. In fact, these 129 TFs can be considered as legitimate reprogramming (insufficient up, proper up and over up) (Fig. 3A) since overexpression of pluripotency factors might be beneficial to reprogramming as we have seen with the OCT4 and SOX2 reprogramming factors. We also examined the situation of these 129 TFs in BJ cells. None of these 129 TFs have significantly higher expression in BJ cells than in ESCs and the majority of them (116) are expressed significantly higher in ESCs than in BJ (Supplementary Fig. S15 and data not shown). None

---

**Figure 4.** Most of the initial TF upregulation by OSKM is legitimate reprogramming. (A) Classification of the 49 TFs upregulated by OSKM into different expression statuses in PSCs and the reprogramming starting fibroblasts. (B) Further reprogramming characterization of the 33 ESC-enriched genes upregulated by OSKM. (C) A heat map showing proper upregulation of 18 ESC-enriched TFs. (D) A heat map showing insufficient upregulation for the 11 ESC-enriched TFs significantly upregulated by OSKM. (E) A heat map showing wrong upregulation by OSKM for the 7 fibroblast-enriched TFs.
of these 129 genes was significantly downregulated by OSKM in BJ cells, but the majority of them (104) were significantly upregulated by OSKM in BJ cells (Supplementary Fig. S16 and data not shown). These 129 TFs of legitimate reprogramming in CRL were also clustered with ESCs at 96 h of OSKM reprogramming of BJ cells, indicating their legitimate reprogramming in BJ cells as well (Supplementary Fig. S16). However, 65 TFs were upregulated by OSKM in CRL cells when they should not be, and 25 TF were upregulated when they should be downregulated (Supplementary Table S15). Nevertheless, the majority (59%) underwent legitimate upreprogramming and were largely conserved between the two fibroblast types.

A total of 118 TFs were downregulated by OSKM in the CRL cells at both 48 and 72 h (Supplementary Table S16). Out of the 118 genes, 89 were enriched by at least 2-fold in the fibroblast CRL, indicating legitimate downreprogramming (Fig. 5C,D). Importantly, none of these 89 TFs exhibited significantly higher expression in ESCs than in the other fibroblast BJ, and the majority of them (79 TFs) displayed significantly higher expression in BJ than in ESCs (Supplementary Fig. S17 and data not shown). Furthermore, none of the 89 genes was significantly upregulated and the majority (71 TFs) of those were significantly downregulated in BJ cells by OSKM at 96 h of reprogramming. Like in CRL, these 89 genes became clustered with ESCs in the reprogramming CRL cells at 96 h (Supplementary Fig. S17), indicating conserved legitimate downreprogramming between the two different fibroblasts.

Discussions
OSKM can convert a rare population of fibroblasts into the pluripotent state with an extended long latency. This is in sheer contrast to the oocyte reprogramming, which is authentic and fast. OSKM cannot activate the master transcriptional network of pluripotency directly. The key pluripotent factors, OCT4, SOX2, NANOG and others, are activated very late in the reprogramming process. Previous efforts have tried to identify molecular events underlying OSKM reprogramming of fibroblasts into pluripotency. However, those researches ignored the fact that 99% of the cells do not go in the direction towards pluripotency and represent the noise signals of the data. Those authors implicitly treated all the transcriptional responses to OSKM induction as positive reprogramming. To mitigate this limitation, the authors developed the concepts of reprogramome and reprogramming...
lending3,10. Using these concepts, a transcriptional response to the OSKM reprogramming can be evaluated as positive, negative (aberrant reprogramming), or unresponsive i.e., legitimate, illegitimate, and no responses, respectively. In previous reports, the authors evaluated the transcriptional responses of all human genes without specific examination of the transcription factors. In this report, the authors evaluated the reprogramming legitimacy of the transcriptional responses of the entire set of human transcription factors to OSKM reprogramming at the initial stages (48, 72, and 96 h). In agreement with the inefficiency, long latency, and stochastic nature of OSKM reprogramming, it was found here that the majority of human transcription factors (296 TFs) were unresponsive to OSKM induction. This report also identified some transcription factors that underwent aberrant reprogramming such as wrong and unwanted reprogramming. These data provide molecular interpretation for the inefficiency and stochastic nature of OSKM reprogramming.

When we specifically analyzed the reprogramming legitimacy of TFs in this report, a surprising discovery is that the majority of transcription factors, which did respond to OSKM induction, underwent legitimate reprogramming. This phenomenon was also observed in an independent human fibroblast cell line from a different individual. The population of transcription factors undergoing legitimate reprogramming is not small. 18 PSC-enriched TFs were properly upreprogrammed to the levels found in PSCs, while 21 somatic TFs were properly downreprogrammed to the levels of pluripotency. Additionally, 11 PSC-enriched TFs were significantly albeit insufficiently reprogrammed towards the pluripotent levels while 33 somatic TFs were significantly downregulated albeit insufficiently downreprogrammed towards the pluripotent levels. These observations may under-estimate the number of legitimate TF reprogramming since classification of legitimate reprogramming into proper, insufficient and over reprogramming usually eliminate some genes undergoing legitimate reprogramming as outlined in Fig. 3A. In fact, we observed many more legitimate TF reprogramming simply using the rationale in Fig. 3A (Fig. 5).

Transcription factors are critical in defining the transcriptional programs and identities of any cell type11–13. TFs are also critical in cellular reprogramming. In fact, all the four conventional pluripotency reprogramming factors are transcription factors21. Lineage-specific transcription factors can reprogram fibroblasts into the corresponding functional somatic cells22–24. Here, the legitimate reprogramming of a large set of transcription factors at the initial stages provides molecular underpinnings for the ability of OSKM to push some reprogramming fibroblasts to the pluripotent state. At the same time, the inability of OSKM to incite the required transcriptional changes of the transcription factors in the TF reprogramome at the early stages explains in part why OSKM is very inefficient, slow and stochastic.

Methods

Cell lines and cultures. The NIH-registered human embryonic stem cell (ESC) lines H1 (WiCell, Madison, WI) and H9 (WiCell, Madison, WI) were cultured in the chemically defined media as described before2,3,9,10. Briefly, hESCs were cultured on Matrigel-coated vessels with the E8 media25, and passaged using the EDTA-mediated dissociation when they reach 80% confluency.

Initial reprogramming. Our lentiviral reprogramming constructs were designed before and have been deposited in Addgene6 (pLVH-EF1a-GFP-P2A-OC14, Addgene #, 130692; pLVH-EF1a-GFP-P2A-OCX2, Addgene #, 130693; pLVH-EF1a-GFP-P2A-OC14, Addgene #, 130694; pLVH-EF1a-GFP-P2A-MYC, Addgene #, 130695). All these constructs have GFP co-expression for easy estimation of viral titers, as well as transfection and transduction efficiency. The reprogramming procedure was basically as reported before1,26 except for that reprogramming was stopped at the indicated time points when RNA was extracted. Human fibroblasts were transduced with the concentrated lentiviruses of the four Yamanaka reprogramming factors at the optimized...
Human transcription factors. Human transcript factors were based on the revised version of Lambert et al.17. DUX1 and DUX3 were missing from our RNA-seq data set because they are not annotated in Ensembl or UCSC database. The official symbol for ZNF645 used in the Lambert list is CBL12. The official symbol for T used in the Lambert list is TBXT. ZUFSP is ZUP1, and ZZZ3 is AC118549.1. ZNF788 (ZNF788P) was excluded from analyses since it is annotated in Ensembl and Genecards as a pseudogene. As a result, this research evaluated 1636 human TFs out of the 1639 TFs in the Lambert list.

Criteria for an active gene with similar expression in both cell types. A gene is considered active in both cell types with equal expression when the following conditions are met: (1) all replicates have a normalized read count greater than 50; (2) the q value should be greater than 0.01; (3) the fold difference between the two cell types should not be equal or greater than 2 regardless of the q value. In this group, all normalized individual read counts are > 50.

Defining the inactive gene set for both cell types. All normalized read counts are < 50 including all replicates for both cell types. The rationale of this value as the threshold of an active gene has been described before9.

Defining activatome (hPSC-specific gene set) anderasome (fibroblast-specific gene set). A gene will be a member of activatome or erasome when the following conditions are met: (1) the active cell type, e.g. hPSCs for activatome, should have a normalized read count of greater than 50 for all individual replicates while the silent cell type should have a normalized read count less than 50 for all individual replicates; (2) the fold differences should be greater than 2; (3) the q value should be less than 0.01.

Defining upreprogramome (hPSC-enriched gene set) and downreprogramome (fibroblast-enriched gene set). An enriched gene in any cell type should meet the following criteria: (1) the normalized read count should be greater than 50 for all replicates; (2) the enrichment should be at least 2-fold; (3) the q value should be < 0.01. The rationale for these criteria have been described before10.

Defining the irresponsible TF genes to OSKM induction. For TF downreprogramme and upreprogramme, an irresponsible TF gene to OSKM induction should meet the following criteria: (1) the fold changes should be less than 2-fold (upregulation or downregulation), and any gene with a fold change of > 2-fold is removed from this list regardless of the significant levels; (2) any gene with a significant level of q < 0.01 is removed from the list regardless of the levels of fold changes. i.e., even though the fold change is 1.5-fold it will
be removed from the list of irresponsible TF genes if the change is significant; (3) exclude any gene for which the difference between OSKM reprogramming cells and the ESC become less than 2-fold regardless of the significance status.

For upreprogramme, the following additional criteria were applied. If the gene remain inactive after OSKM induction (normalized read count < 50), it is considered irresponsible even though the fold change is > 2 and is statistically significant.

**Data visualization.** Heat maps were prepared using the R package `pheatmap` (Version 1.0.12) in RStudio (Version 1.3.1073) on a desktop iMac (Version 10.15.6) as described. Boxplots were generated using the generic R function of `boxplot()` in RStudio as described recently. Both heat maps and box-plots were prepared using the log2-transformed read counts. Ladder plots were generated using the R package of `plotrix` (Version 3.7-8) in RStudio.

**Code availability**

Accession numbers

The RNA-seq data have been deposited in Gene Expression Omnibus (GEO) with the accession code of GSE148158, and GSE159410.

Received: 17 July 2020; Accepted: 19 October 2020

Published online: 12 November 2020

**References**

1. Takahashi, K. et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131, 861–872. https://doi.org/10.1016/j.cell.2007.11.019 (2007).
2. Shao, Z. et al. Reprogramming by de-bookmarking the somatic transcriptional program through targeting of BET bromodomains. Cell Rep. 16, 3138–3145. https://doi.org/10.1016/j.celrep.2016.08.069 (2016).
3. Shao, Z. et al. The acetyllysine reader BRD3R promotes human nuclear reprogramming and regulates mitosis. Nat. Commun. 7, 10869. https://doi.org/10.1038/ncomms10869 (2016).
4. Hu, K. All roads lead to induced pluripotent stem cells: the technologies of iPSC generation. Stem Cells Dev. 23, 1285–1300. https://doi.org/10.1089/scd.2013.0620 (2014).
5. Hu, K. Vectorology and factor delivery in induced pluripotent stem cell reprogramming. Stem Cells Dev. 23, 1301–1315. https://doi.org/10.1089/scd.2012.11.039 (2012).
6. Hu, K., Ianov, L. & Crossman, D. Profiling and quantification of pluripotency reprogramming reveal that WNT pathways and cell morphology have to be reprogrammed extensively. Heliyon 6, e04035. https://doi.org/10.1016/j.heliyon.2020.e04035 (2020).
7. Mikkelsen, T. S. et al. Dissecting direct reprogramming through integrative genomic analysis. Nature 454, 49–55. https://doi.org/10.1038/454070a (2008).
8. Polo, J. M. et al. A molecular roadmap of reprogramming somatic cells into iPS cells. Cell 151, 1617–1632. https://doi.org/10.1016/j.cell.2012.11.039 (2012).
9. Hu, K., Janov, L. & Crossman, D. Profiling and quantification of pluripotency reprogramming reveal that WNT pathways and cell morphology have to be reprogrammed extensively. Heliyon 6, e04035. https://doi.org/10.1016/j.heliyon.2020.e04035 (2020).
10. Hu, K. A PIANO (Proper, Insufficient, Aberrant, and NO Reprogramming) response to the Yamanaka factors in the initial stages of human iPSC reprogramming. Int. J. Mol. Sci. 21, 7093–7120. https://doi.org/10.3390/ijms21093229 (2020).
11. Gurdon, J. B. Cell fate determination by transcription factors. Curr. Top. Dev. Biol. 116, 445–454. https://doi.org/10.1016/B978-0-12-812206-4.00006-3 (2015).
12. Drouin, J. Minireview: Pioneer transcription factors in cell fate specification. Mol. Endocrinol. 28, 989–998. https://doi.org/10.1210/me.2014-1098 (2014).
13. Rosenbauer, F. & Tenen, D. G. Transcription factors in myeloid development: Balancing differentiation with transformation. Nat. Rev. Immunol. 7, 105–117. https://doi.org/10.1038/nri2024 (2007).
14. Fulton, D. L. et al. TFCat: The curated catalog of mouse and human transcription factors. Genome Biol. 10, R29. https://doi.org/10.1186/gb-2009-10-3-r29 (2009).
15. Wingender, E., Schoeps, T., Haubrock, M. & Donitz, I. TFCat: A classification of human transcription factors and their rodent orthologs. Nucleic Acids Res. 43, D97–102. https://doi.org/10.1093/nar/gku1064 (2015).
16. Zhang, H. M. et al. AnimalTFDB: A comprehensive animal transcription factor database. Nucleic Acids Res. 40, D144–149. https://doi.org/10.1093/nar/gkt965 (2012).
17. Lambert, S. A. et al. The human transcription factors. Cell 172, 650–665. https://doi.org/10.1016/j.cell.2018.01.029 (2018).
18. Kang, L. et al. The universal 3D3 antibody of human PODXL is pluripotent cytotoxic, and identifies a residual population after extended differentiation of pluripotent stem cells. Stem Cells Dev. 25, 356–368. https://doi.org/10.1089/scd.2015.0321 (2016).
19. Yu, J. et al. Human induced pluripotent stem cells free of vector and transgene sequences. Science 324, 797–801. https://doi.org/10.1126/science.1172482 (2009).
20. Yu, J. et al. Induced pluripotent stem cell lines derived from human somatic cells. Science 318, 1917–1920. https://doi.org/10.1126/science.1151526 (2007).
21. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676. https://doi.org/10.1016/j.cell.2006.07.024 (2006).
22. Davis, R. L., Weintraub, H. & Lassar, A. B. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell 51, 987–1000. https://doi.org/10.1016/0092-8674(87)90585-5 (1987).
23. Pang, Z. P. et al. Induction of human neuronal cells by defined transcription factors. Nature 476, 220–223. https://doi.org/10.1038/nature10120 (2011).
24. Ieda, M. et al. Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142, 375–386. https://doi.org/10.1016/j.cell.2010.07.002 (2010).
25. Chen, G. et al. Chemically defined conditions for human iPSC derivation and culture. Nat. Methods 8, 424–429. https://doi.org/10.1038/nmeth.1593 (2011).
26. Shao, Z., Cevallos, R. & Hu, K. Methods in molecular biology, Ch. 7. In Nuclear Reprogramming Vol. 2239 (ed. Hu, K.) (Springer-Nature, New York, 2020).
27. Andrews, S. FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc (2010).
28. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. https://doi.org/10.1093/bioinformatics/bts63 5 (2013).
29. Wang, L., Wang, S. & Li, W. RSeQC: Quality control of RNA-seq experiments. *Bioinformatics* **28**, 2184–2185. https://doi.org/10.1093/bioinformatics/bts356 (2012).
30. Ewels, P., Magnusson, M., Lundin, S. & Kaller, M. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048. https://doi.org/10.1093/bioinformatics/btw354 (2016).
31. Li, H. et al. The sequence alignment/map format and SAMtools. *Bioinformatics* **25**, 2078–2079. https://doi.org/10.1093/bioinformatics/btp352 (2009).
32. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): High-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192. https://doi.org/10.1093/bib/bbs017 (2013).
33. Anders, S., Pyl, P. T. & Huber, W. HTSeq—A Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169. https://doi.org/10.1093/bioinformatics/btu638 (2015).
34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550. https://doi.org/10.1186/s13059-014-0550-8 (2014).
35. Hu, K. Become competent in generating RNA-seq heat maps in one day for novices without prior R experience, Chapter 17. In *Nuclear Reprogramming - Methods in Molecular Biology*. https://doi.org/10.1007/978-1-0716-1084-8-17 (2020).
36. Hu, K. Become competent within one day in generating boxplots and violin plots for a novice without prior R experience. *Methods Protocols* **3**, 64. https://doi.org/10.3390/mps3040064 (2020).

**Acknowledgements**
The author appreciates the technical help in RNA-sequencing by Dr. Michael Crowley and the support in bioinformatics by Drs. David Crossman and Lara Ianov. This study is supported by the National Institutes of Health (1R01GM127411) and the American Heart Association (17GRNT33670780), and partly supported by the NIH Grant U54DK126087.

**Author contributions**
K.H. conceived the project, conducted the experiments, analyzed and interpreted the data, visualized the RNA-seq results, and wrote manuscript. R.C. conducted reprogramming and RNA preparation. Y.E. conducted the bioinformatics and manuscript writing. B.Y. and J.P. provided financial support and conducted manuscript preparation.

**Competing interests**
The authors declare no competing interests.

**Additional information**
Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-76705-y.

**Correspondence** and requests for materials should be addressed to K.H.

**Reprints and permissions information** is available at www.nature.com/reprints.

**Publisher’s note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020