KLF4 N-Terminal Variance Modulates Induced Reprogramming to Pluripotency

Shin-Il Kim,1,2 Fabian Oceguera-Yanez,1 Ryoko Hirohata,1 Sara Linker,1,3,6 Keisuke Okita,1 Yasuhiro Yamada,1,2 Takuya Yamamoto,1,2 Shinya Yamanaka,1,2,4 and Knut Woltjen1,2,5,*

1Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto 606-8507, Japan
2Institute for Integrated Cell-Material Sciences (WPI-iCeMS), Kyoto University, Kyoto 606-8507, Japan
3John P. Hussman Institute for Human Genomics, University of Miami, Miami, FL 33136, USA
4Hakubi Center for Advanced Research, Kyoto University, Kyoto 606-8501, Japan
5Gladstone Institute of Cardiovascular Disease, San Francisco, CA 94158, USA
6Present address: Salk Institute for Biological Studies, La Jolla, CA 92037, USA
*Correspondence: woltjen@cira.kyoto-u.ac.jp

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

SUMMARY

As the quintessential reprogramming model, OCT3/4, SOX2, KLF4, and c-MYC re-wire somatic cells to achieve induced pluripotency. Yet, subtle differences in methodology confound comparative studies of reprogramming mechanisms. Employing transposons, we systematically assessed cellular and molecular hallmarks of mouse somatic cell reprogramming by various polycistronic cassettes. Reprogramming responses varied in the extent of initiation and stabilization of transgene-independent pluripotency. Notably, the cassettes employed one of two KLF4 variants, differing only by nine N-terminal amino acids, which generated dissimilar protein stoichiometry. Extending the shorter variant by nine N-terminal amino acids or augmenting stoichiometry by KLF4 supplementation rescued both protein levels and phenotypic disparities, implicating a threshold in determining reprogramming outcomes. Strikingly, global gene expression patterns elicited by published polycistronic cassettes diverged according to each KLF4 variant. Our data expose a KLF4 reference cDNA variation that alters polycistronic factor stoichiometry, predicts reprogramming hallmarks, and guides comparison of compatible public data sets.

INTRODUCTION

Cellular identity can be guided by ectopic expression of master regulators (Graf, 2011). Deriving induced pluripotent stem cells (iPSCs) through the activities of OCT3/4, SOX2, KLF4, and c-MYC (Takahashi and Yamanaka, 2006) provides a potent model in which to study the role of transcription factor coordination in driving somatic cells toward pluripotency. Early mechanistic studies using mouse embryonic fibroblasts (MEFs) were conducted through de novo introduction of viral vectors, each expressing an individual (monocistronic) reprogramming factor (Brambrink et al., 2008; Stadtfeld et al., 2008a), where modulation of factor levels by viral titration led to altered reprogramming characteristics (Yamaguchi et al., 2011). Monocistronic reprogramming allows for variation in copy number and integration site, and as a result, stoichiometry is inconsistent on a cell-to-cell level. Therefore, this method was succeeded by the development of polycistronic factor cassettes that can produce multiple proteins from one single transcript (Kaji et al., 2009; Sommer et al., 2009). Although such fixed polycistronic stoichiometry revealed the importance of relative factor ratios in determining the quality of reprogramming (Carey et al., 2011), the principles that establish optimal stoichiometry remain undefined.

Studies of the mechanisms that underlie somatic cell reprogramming have revealed multi-step processes involving proliferation and cell-cell adhesion, along with molecular changes such as downregulation of lineage-specific genes and eventual upregulation of pluripotency markers (Plath and Lowry, 2011). Cell-surface markers were associated with reprogramming stages such as emergence of the embryonic stem cell (ESC) marker SSEA-1 (stage-specific embryonic antigen 1) (Polo et al., 2012; Stadtfeld et al., 2008a). Secondary reprogramming systems (Woltjen et al., 2009) helped define initiation, maturation, and stabilization as key stages in reprogramming toward pluripotency (David and Polo, 2014). Proliferation, colony formation, and a mesenchymal-to-epithelial transition (MET) define the initiation phase (Samavarchi-Tehrani et al., 2010), while stabilization is characterized by transgene independence and activation of pluripotency reporters such as Nanog and Dppa4 (Golipour et al., 2012). Thus, changes in global gene expression and epigenetics were associated with the progression of reprogramming through these stages (Theunissen and Jaenisch, 2014). However, discrepancies in reprogramming platforms influence reprogramming hallmarks, the severity of MET responses, lineage-specific gene repression and ectopic activation, the timing of cell-surface marker presentation, and the frequency of partial and complete reprogramming (Golipour et al., 2012; Mikkelsen et al., 2008; O’Malley...
Please cite this article in press as: Kim et al., KLF4 N-Terminal Variance Modulates Induced Reprogramming to Pluripotency, Stem Cell Reports (2015), http://dx.doi.org/10.1016/j.stemcr.2015.02.004
et al., 2013; Polo et al., 2012; Samavarchi-Tehrani et al., 2010; Wernig et al., 2008).

In order to clarify such method-dependent reprogramming hallmarks, we applied standard assays to compare polycistronic cassettes (constructed in-house or obtained from public resources) in a drug-inducible piggyBac (PB) transposon reprogramming system. The induced expression of various polycistronic cassettes in mouse fibroblasts evoked phenotypic and gene expression changes that we divided into two basic behavioral classes. An examination of individual factor sequences across the panel of polycistronic constructs revealed a re-occurring discrepancy in the Klf4 cDNA defining two coding region length variants. Increased length at the KLF4 N terminus was associated with higher protein levels and thus altered relative factor stoichiometry, ultimately impacting both the initiation and stabilization phases of iPSC derivation. Here, we report the consequences of KLF4 N-terminal variation in mono- and polycistronic reprogramming experiments, and apply these findings to recognize and reconcile differences in reprogramming characteristics implied hitherto.

RESULTS

A Transposon System for Collating Polycistronic Reprogramming Cassettes

Reprogramming studies in mouse have made use of unique polycistronic factor arrangements and delivery vectors. For uniform evaluation of factor-order effects, we employed a fundamental reprogramming scheme based on factor transposition in MEFs (Woltjen et al., 2009). The PB transposon vector (PB-TAC) employs doxycycline (dox)-responsive reprogramming cassette expression co-incidently with a mCherry reporter (Figure 1A). ROSA-rtTA; Nanog-GFP MEFs combine the m2-rtTA transactivator (Ohnishi et al., 2014) with a Nanog-GFP reporter (Okita et al., 2007). Thus, dox-responsive, PB-TAC-transgenic cells can be monitored throughout reprogramming initiation and maturation (day 2 [d2]–14) by mCherry fluorescence, while stabilization of pluripotency (d14–18) is indicated by activation of Nanog-GFP. Gain of factor independence through autonomous transgene silencing despite continued dox treatment is an established hallmark of the stabilization phase (Golipour et al., 2012), signaled here by a decrease in mCherry expression. For all polycistronic reprogramming cassettes tested, we routinely passaged populations on d8 and d18 using equal cell numbers without fractionation for extended culture and fluorescence-activated cell sorting (FACS) analysis. Dox-independent maintenance of iPSCs was verified by culture until d24, after which pluripotency was assayed by gene expression array and chimera contribution.

Initially, we used an alkaline phosphatase-positive (AP+) colony formation assay on d10 to compare various polycistronic cassettes with two previously published versions: OSKM (Carey et al., 2009) and MKOS (Kaji et al., 2009). Across the range of vectors tested, OKMS, constructed through the combination of OSKM and pCX-OKS-2A (Okita et al., 2008), displayed a reprogramming phenotype.

Figure 1. KLF4 Isoforms Underlie Phenotypic Differences in Reprogramming

(A) Diagram of the PB reprogramming system and analysis steps. The PB-TAC transposon delivers inducible, reporter-linked (mCherry), polycistronic reprogramming cassettes (Poly) into ROSA-rtTA; Nanog-GFP MEFs (d-1). Cultures are induced with dox (d0, dox; filled arrows) and harvested on d8 for FACS and passage for late-stage analysis (d18). For AP staining, cultures are maintained without passage until d10. Full reprogramming and transgene independence is assessed by dox withdrawal from d18 to d24 (open arrows). Predicted expression periods for the mCherry and GFP reporters during iPSC derivation are shown with red and green graded bars. Blue polygons represent PB 3’ (left) and 5’ (right) inverted terminal repeats. tet0, dox-responsive promoter; IRES, internal ribosome entry signal; pA, polyadenylation signal.

(B) AP staining on d10 of OSKM and OKMS reprogramming cultures. Scale bar, 4,000 μm.

(C) Day 18 fluorescence microscopy of entire wells (composite 10 × 10 fields) or selected insets (3 × 3 subfields) for mCherry+ and Nanog-GFP+ (left). Scale bars, 4,000 μm (full well) and 1,000 μm (inset). FACS analysis of mCherry and Nanog-GFP expression in d18 Total and SSEA-1+ populations (right).

(D) Polycistronic cassette structure and sequence of the 2A-Klf4 N-terminal cleavage junctions. Cloned cDNA is compared with murine Klf4 GenBank mRNA sequences (U20344.1 and U70662.1), indicating the position of the predicted initiation codons and amino acid translations. The N-terminal 9aa of U20344.1 were introduced into OKMS (Klf4S) to produce OK MS (Klf4S).

(E and F) The OK MS construct was evaluated according to the assays outlined in (A)–(C). The results in (B), (C), (E), and (F) are representative of the results from at least three independent experiments (summarized in Figures S1A and S1B).

(G and H) Time-course analysis of mCherry cell expansion (G) and SSEA-1+ fractions (H) from OSKM, OKMS, and OK MS transfections cultured for the indicated number of days before (2, 4, 6, 8) and after (10, 14, and 18) passage. Means ± SE for three independent experiments.

(I and J) Time-course analysis of mCherry silencing (left) and Nanog-GFP activation (right) in SSEA-1+ cells on the indicated number of days after day 8 passage (10, 14, and 18). Means ± SE for three independent experiments.

See also Figure S1.
contradictory to OSKM and MKOS (Figure 1B and data not shown). Despite similar transfection efficiencies (~3%-5% mCherry+ cells at d2 for both OSKM and OKMS), OKMS-induced AP+ colonies were clearly more abundant by d10, indicating that OKMS initiated somatic reprogramming more robustly than OSKM (Figures 1B and S1A).

To evaluate iPSC stabilization, we assayed activation of the Nanog-GFP pluripotency reporter and silencing of factor-linked mCherry after d8 passage of OSKM and OKMS reprogramming cultures. Whole-well microscopic evaluation revealed that most OKMS-induced colonies were unable to activate Nanog-GFP by d18 (Figure 1C, left), whereas mCherry expression remained high. In contrast, OSKM cultures showed a majority of GFP+ colonies with silenced mCherry. FACS on d18 confirmed the proportions of reporter-positive cells (Figure 1C, right). Both the Total GFP+ and SSEA-1+ GFP+ fractions on d18 were consistently higher for OSKM reprogramming (mean 15.6% and 39.5%, respectively, versus 5.4% and 9.9% for OKMS; Figure S1B). Conversely, overrepresentation of Total mCherry+ and SSEA-1+ mCherry+ populations was characteristic of OKMS (mean 53.1% and 58.8%, respectively, versus 21.0% and 27.0% for OSKM; Figure S1B). GFP+ cells emerged only in the SSEA-1+ fraction (data not shown), and only in mCherrylow or mCherry− cells, such that double-positive GFP+ mCherry+ cells were not observed under either condition (Figure 1C, right). Reporter ratios remained mostly unchanged after d18 passage and an additional 6 days of culture in dox (d24; Figure S1C, left). Failure to activate Nanog-GFP was not associated exclusively with transgene interference, as dox withdrawal led to a complete loss of mCherry, but no increase in GFP for either the OSKM or OKMS populations (Figures S1C and S1D). For both vectors, dox-independent, Nanog-GFP+ iPSC clones contributed to chimera formation and clustered with ESCs in microarray transcriptome analyses, validating pluripotency (Figures S1E–S1G). These data suggest that OKSM more readily achieves stable reprogramming, whereas the majority of OKMS-induced cells remain in a transgene-dependent, partially reprogrammed state.

KLF4 Isoforms Underlie Phenotypic Differences in Polycistronic Cassette Reprogramming

We next sought to expose key differences underlying the distinct phenotypes observed for OSKM and OKMS, and initially examined the reprogramming factors at the DNA sequence level. Of particular interest, the 2A peptide cleavage site preceding Klf4 in OSKM demarcated an ATG initiation codon that included an additional nine N-terminal amino acids (9aa) compared with that for OKMS (Figure 1D). When we examined public databases, we found that these two distinct Klf4 isoforms correspond to open reading frames (ORFs) predicted from full-length cDNA cloning (GenBank accession numbers U70662.1 and U20344.1) (Garrett-Sinha et al., 1996; Shields et al., 1996), anticipating proteins of 474aa and 483aa (GenBank accession numbers AAC52939.1 and AAO4892.1); henceforth referred to as KLF4S and KLF4L. To examine the effects of KLF4S and KLF4L on reprogramming while ruling out differences in factor order and 2A cleavage peptides, we introduced the N-terminal 9aa noted in OSKM (MRQPPGESD) directly into OKSM, producing Ok9SM (where K9 is equivalent to KLF4L).

MEFs transposed with Ok9SM displayed a complete phenotypic switch. AP+ colony formation resembled that of OSKM (Figures 1E and S1A). Whole-well imaging on d18 strikingly revealed that OK9SM colonies were primarily GFP+ and there were few mCherry+ colonies (Figure 1F, left). The Total and SSEA-1+ fractions on d18 were predominantly GFP+ with extensive mCherry silencing, mirroring OSKM reprogramming behavior (Figures 1F, right, and S1B). Dox-independent, GFP+ Ok9SM clones were chimera competent and shared a gene expression profile with OSKM and OKMS iPSCs, and ESCs (Figures S1E–S1G).

Extended FACS analysis revealed the kinetics of mCherry+ cell expansion throughout the early phase (d2–8), equating Ok9SM with OSKM (Figure 1G). From our d8 FACS data, we found that a large proportion of OKSM mCherry+ cells were positive for SSEA-1. Therefore, we traced SSEA-1 dynamics from d2 to d8 (mCherry+ fraction; Figure 1H, left) and d10 to 18 post-passage (Total population; Figure 1H, right). Notably, only OKMS cultures were distinguished by rapid and robust SSEA-1 activation as early as d2 (29.4% versus 4.8% for OSKM, and 8.5% for Ok9SM), which continued through early reprogramming. For all constructs, SSEA-1+ cells remained mCherry+ through d8, at which point no Nanog-GFP activation could be detected. An examination of the increasing proportion of SSEA-1+ cells over d10–18 (Figure 1H, right) showed that activation of Nanog-GFP and coincident suppression of mCherry by Ok9SM were more consistent with OSKM, whereas for OKMS, both mCherry silencing and Nanog-GFP activation were less frequent (Figures 1I and 1J). Overall, these data indicate that elongation of Klf4S to Klf4L in Ok9SM subdues early reprogramming hallmarks but positively regulates stabilization of pluripotency.

Monocistronic Expression of Klf4S or Klf4L Does Not Differentially Influence Reprogramming

To determine whether the observed cassette-specific phenotypes were due to functional differences between KLF4S and KLF4L, we performed a series of monocistronic Klf4 expression experiments. We found that the transcriptome profiles of d6 mCherry+ cells arising from PB-TAC-Klf4S or -Klf4L MEF transfections were highly correlated (R = 0.999; Figure S2A). Gene enrichment compared with
mock-transfected MEFs revealed a common set of 476 genes enriched for Gene Ontology (GO) terms associated with extracellular matrix, skin development, and cell adhesion (Figure S2B). Activation of a KLF4-responsive Lefty1-luciferase reporter (Nakatake et al., 2006) was similar irrespective of the KLF4 variant (Figure S2C), indicating that the capacity for regulation of this ESC promoter was unchanged. Moreover, both KLF4s and KLF4l could similarly rescue leukemia inhibitory factor (LIF)-independent culture of 1A2 Nanog-GFP reporter ESCs (data not shown). Thus, independent expression of either Klf4 variant supported a preservation of functional capacity.

Considering the remarkable reprogramming phenotypes observed with polycistronic cassettes, we next addressed whether monocistronic expression of Klf4s or Klf4l in combination with polycistronic Oct3/4, Sox2, and c-Myc (OMS) had any detectable effect on iPSC derivation. OMS alone did not foster expansion of mCherry+ cells, which remained at a steady ~6% through d18, nor did it activate SSEA-1 or Nanog-GFP expression (data not shown). Interestingly, expression of either Klf4 variant in combination with OMS resulted in nearly indistinguishable reprogramming phenotypes using the assays outlined in Figure 1 (see also Figures 2A, S2D, and S2E). These data demonstrate that the phenotypes associated with KLF4s and KLF4l arise mainly in the context of polycistronic factor linkage.

KLF4 Levels Are Diminished in OKMS and Rescued by N-Terminal Elongation

In order to rationalize the phenotypic differences observed between polycistronic cassettes, we evaluated transgene expression at the transcriptional and translational levels. Quantification of polycistronic mRNA revealed similar expression levels for OSKM and OKMS (Figure S2F). Strikingly, western blotting on d2 revealed that dox-induced KLF4 was expressed at distinctly higher levels in OSKM compared with OKMS, whereas the levels of OCT3/4, SOX2, and c-MYC (OMS) were generally comparable (Figure 2B). High KLF4 stoichiometry was also observed in MKOS, which, like OSKM, encodes KLF4l (Figure S2G). Consistent with the PB transposon experiments, MEFs from gene-targeted Col1a1-OSKM (Ohnishi et al., 2014) and -OKMS ESC chimeras faithfully reproduced cassette-specific protein levels (data not shown). Moreover, OK4MS resulted in a marked increase in KLF4 levels compared with OKMS (Figure S2H). Surprisingly, N-terminal elongation of KLF4s in OKMS with a hemagglutinin (HA) epitope also rescued protein levels (OK4HAMS; Figure S2H) and even reprogramming phenotypes (Figure S2L), whereas N-terminal HA extension of KLF4l in OK4MS did not result in any additional phenotypic change. Concordant with monocistronic phenotypes, transfection of PB-TAC-Klf4s or -Klf4l in MEFs produced comparable protein levels on d2 (Figure S2J). These data indicate that the protein sequence at the KLF4 N terminus plays a role in specifically establishing polycistronic factor stoichiometry.

KLF4 Stoichiometry Regulates Reprogramming

To verify the reprogramming outcomes from adjusted factor stoichiometry, we supplemented OKMS with monocistronic factors in a PB-TAB transposon (Figure 2C). Supplementation of OKMS with Klf4s or Klf4l (Figure S2J) markedly reduced d10 AP+ colony numbers (Figures 2D and 2G), reminiscent of the reduced colony formation efficiencies observed when OSKM or OK4MS was employed. Sox2, but not Oct3/4 or c-Myc, also had a significant effect on reducing colony formation frequencies compared with OKMS alone. Thus, both KLF4 isoforms modulated OKMS factor stoichiometry to subdue reprogramming initiation.

In the late phase of reprogramming, we gauged stabilization via mCherry silencing and Nanog-GFP activation (Figures 2E, 2F, and 2H). Supplementation of OKMS with either Klf4s or Klf4l was capable of activating Nanog-GFP with co-incident repression of mCherry on d18. Interestingly, although Sox2 supplementation also showed a repressive effect on initiation, d18 mCherry* and GFP* proportions remained mostly unchanged, indicating that additional Sox2 fails to rescue the OKMS stabilization phase. Likewise, Oct3/4 and c-Myc failed to promote Nanog-GFP reporter activation or mCherry silencing. Thus, only Klf4 variants were capable of augmenting the OKMS phenotype, implying a sensitivity of reprogramming pathways to KLF4 expression thresholds.

KLF4 Stoichiometry Is Reflected in Gene Expression Analyses

Global gene expression has been used to define reprogramming paths. In order to link gene expression to initial reprogramming behaviors, we subjected d6 mCherry+ cells from PB-TAC-OSKM, -OKMS, -OK4MS, and -MKOS to microarray analysis (Figures 3A and 3B). Based on OSKM and OKMS GO term enrichment (data not shown) and results from previous reports, we established sample indexing criteria (Figure 3A, left). Krt14 and Krt17 (encoding structural proteins), as well as Sfi1 (a regulator of epithelial cell growth), are claimed to activate in reprogramming populations with preferred iPSC fate (O’Malley et al., 2013). Expression of Ocln and Cdhl indicates MET reprogramming behaviors (Samavarchi-Tehrani et al., 2010), while additional cell-surface changes are reflected by upregulation of Cadm1 and Futt (the fucosyltransferase responsible for producing SSEA-1). The developmental regulators Tbx21 and Bcl11a (transiently upregulated non-ESC markers; Polo et al., 2012) and the pre-iPSC marker Nrb1 (Mikkelsen et al., 2012) were monitored for changes in expression.
Please cite this article in press as: Kim et al., KLF4 N-Terminal Variance Modulates Induced Reprogramming to Pluripotency, Stem Cell Reports (2015), http://dx.doi.org/10.1016/j.stemcr.2015.02.004
show the extent of reprogramming. As early as d6, these ten marker genes were >2-fold differentially expressed between OKSM (green series) and OKMS (red series; Figure 3A, left), and <2-fold differentially expressed in a pairwise correlation of OSKM to OKSM (r = 0.993; Figure 3A, top right), certifying our indexing for further pairwise comparisons. Correlation and indexing of OSKM to OKMS verified the influence of KLF4 on early gene expression responses (Figure 3A, bottom right).

GO term analysis of commonly upregulated gene expression versus mock-transfected MEFs (PB-TAC-lacZ) was performed for cassettes containing Klf4L and Klf4S (Figure 3B and data not shown). For Klf4 isotypes (216 genes), only a few GO terms related to the biological processes “regulation of cell adhesion (laminins)” and “tissue morphogenesis,” as well as the molecular function of “phospholipase C activity,” were enriched with p values ≤ 1.0E-3 (data not shown). Nearly double the number of genes were activated by Klf4L constructs (512 genes; Figure 3B; Table S1). The common gene set for polycistronic Klf4L constructs significantly enriched GO terms for cellular components of the “plasma membrane” and “cell-cell junction” (consistent with an expected MET response), as well as distinct enrichment for “keratinocyte differentiation” and “epidermal cell differentiation” biological processes (p = 2.3E-11 and 5.4E-11; Figure 3B, right) with Benjamini scores = 3.6E-8 and 4.1E-8, respectively. These data are consistent with monocistronic Klf4 expression (Figure S2B).

KLF4 Isoforms Divide Reprogramming Cassettes into Two Major Classes

In order to more broadly understand the impact of KLF4-mediated reprogramming outcomes, we extended our analysis of publically available polycistronic constructs. DNA sequence surveys revealed that Klf4L and Klf4S reference cDNAs have been used interchangeably in the reprogramming field (Table 1). To confirm that public polycistronic vectors also display incongruous reprogramming behaviors, we selected MKOS (Kaji et al., 2009), STEMCCA (Sommer et al., 2009), EB-C5 (Chou et al., 2011), and WTSI (Yusa et al., 2009), and applied our battery of phenotyping tests (Figure S3). mCherry+ fractions and AP+ colony numbers varied among cassettes but were highest for OKMS and EB-C5 (Figures S3A and S3C). Notably, all four Klf4L cassettes (OKMS, STEMCCA, EB-C5, and WTSI) showed large proportions of mCherry+ SSEA-1+ cells on d8 (>40%, compared with <21% for Klf4L cassettes OKSM, OKSM, and MKOS; Figure S3B), a characteristic that was consistently associated with Klf4S (Figure 1H). In accordance with Klf4S construction and western blotting data (Figure S2G), MKOS AP+ colony numbers were low on d10, and became mostly GFP+ mCherry+ colonies on d18 (Figures S3D and S3E). On the other hand, the Klf4L cassettes STEMCCA, EB-C5, and WTSI all perpetuated mCherry expression after d8 passage (74.8%, 80.9%, and 79.6%, respectively), and activated Nanog-GFP in <3.0% of SSEA-1+ cells (Figure S3E). Thus, the phenotypes elicited by these additional public cassettes predictably correlated reprogramming hallmarks with Klf4 coding length.

Finally, gene indexing confirmed that MKOS (Klf4L) is distinct from STEMCCA (Klf4S; Figure 3C), substantiating the phenotypes reported in Figure S3. Moreover, unsupervised clustering of d6 mCherry+ cell gene expression profiles from all eight constructs (PB-TAC-OSKM, -OKMS, -OKSM, -OKNHA1MS, -MKOS, -STEMCCA, -EB-C5, and -WTSI) generated a dendogram (Figure 3D) that distinctly
bifurcated based on the Klf4 variant employed in each vector (Table 1). Thus, early gene expression patterns disclosed the nature of cloned Klf4. Intriguingly, despite unrelated N-terminal sequence identity (yet predicted by phenotype), OKN-HAMS activated 463/512 (~90%) of the core Klf4L genes (data not shown) and clustered with Klf4L vectors (Figure 3D). A standardized comparison of polycistronic reprogramming cassettes thus implicates KLF4 isoforms in differential modulation of the reprogramming transcriptome.

## DISCUSSION

Principally, we have uncovered a fundamental characteristic of cloned Klf4 that governs relative factor stoichiometry and reprogramming responses during mouse iPSC derivation with polycistronic constructs. At the root of our discovery lies the disparate coding prediction of full-length Klf4 cDNA (Garrett-Sinha et al., 1996; Shields et al., 1996), resulting in two proteins with distinct N termini defined herein as KLF4S and KLF4L. These two
Table 1. Survey of Publically Available Reprogramming Cassettes

<table>
<thead>
<tr>
<th>Source</th>
<th>Addgene Collection</th>
<th>Depositing Scientist</th>
<th>Repository ID</th>
<th>Plasmid Name</th>
<th>Gene/Insert Name</th>
<th>Species</th>
<th>Type</th>
<th>Klf4\textsubscript{3}</th>
<th>Klf4\textsubscript{2}</th>
<th>AddGene</th>
<th>Depositor</th>
<th>Direct</th>
<th>PMID</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addgene</td>
<td>Klf4/KLF4</td>
<td>Miguel Ramalho-Santos</td>
<td>15950</td>
<td>pLOVE-Klf4</td>
<td>Kruppel-like factor 4</td>
<td>M. musculus</td>
<td>mono</td>
<td>n/d</td>
<td>n/d</td>
<td>n/c</td>
<td>–</td>
<td>18371358</td>
<td>Blclloch et al., 2007</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Shinya Yamanaka</td>
<td>17219</td>
<td>pMXs-hKLF4</td>
<td>KLF4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>–</td>
<td>○</td>
<td>○</td>
<td>18035408</td>
<td>Takahashi et al., 2006</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>George Daley</td>
<td>17227</td>
<td>pMIG-hKLF4</td>
<td>Kruppel-like factor 4 (gut)</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>18157115</td>
<td>Park et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Konrad Hochedlinger</td>
<td>19764</td>
<td>pLV-tetO-Klf4</td>
<td>Klf4</td>
<td>M. musculus</td>
<td>mono</td>
<td>●</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>18371448</td>
<td>Stadtfeld et al., 2008a</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Konrad Hochedlinger</td>
<td>19770</td>
<td>pAd-Klf4</td>
<td>Klf4</td>
<td>M. musculus</td>
<td>mono</td>
<td>●</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>18818365</td>
<td>Stadtfeld et al., 2008b</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Konrad Hochedlinger</td>
<td>19777</td>
<td>FU-tet-o-hKLF4</td>
<td>KLF4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>18786420</td>
<td>Maherali et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Juan Belmonte</td>
<td>20074</td>
<td>pMSCV-Flag-hKlf4</td>
<td>hKlf4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>18931654</td>
<td>Aasen et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Rudolf Jaenisch</td>
<td>20322</td>
<td>Tet0-FUW-Klf4</td>
<td>Klf4</td>
<td>M. musculus</td>
<td>mono</td>
<td>●</td>
<td>–</td>
<td>○</td>
<td>–</td>
<td>18371436</td>
<td>Brambrink et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell; Klf4/KLF4</td>
<td>Rudolf Jaenisch</td>
<td>20725</td>
<td>FUW-tetO-hKLF4</td>
<td>KLF4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>–</td>
<td>○</td>
<td>–</td>
<td>18786421</td>
<td>Hockemeyer et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>Klf4/KLF4</td>
<td>Rudolf Jaenisch</td>
<td>20727</td>
<td>FUW-tetO-lox-hKLF4</td>
<td>KLF4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>–</td>
<td>○</td>
<td>–</td>
<td>19269371</td>
<td>Soldner et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>Klf4/KLF4</td>
<td>Michel Sadelain</td>
<td>23243</td>
<td>pLM-mCherry-Klf4</td>
<td>mCherry_2A_KLF4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>●</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>19549847</td>
<td>Papapetrou et al., 2009</td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
Table 1.  

<table>
<thead>
<tr>
<th>Source</th>
<th>Addgene Collection</th>
<th>Depositing Scientist</th>
<th>Repository ID</th>
<th>Construct Data</th>
<th>Gene/Insert Name</th>
<th>Species</th>
<th>Type</th>
<th>Klf4s</th>
<th>Klf4l</th>
<th>AddGene</th>
<th>Depositor</th>
<th>Direct</th>
<th>PMID</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addgene</td>
<td>Klf4/KLF4</td>
<td>Rudolf Jaenisch</td>
<td>26274</td>
<td>pBRPyCAG-fmKlf4- DsRed-1p</td>
<td>Klf4</td>
<td>M. musculus</td>
<td>mono</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>20442331</td>
<td>Hanna et al., 2010</td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell;</td>
<td>Derrick Rossi</td>
<td>26815</td>
<td>pcDNA3.3_KLF4</td>
<td>5’UTR-KLF4-3’UTR</td>
<td>H. sapiens</td>
<td>mono</td>
<td></td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>–</td>
<td>20888316</td>
<td>Warren et al., 2010</td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell;</td>
<td>Steven Dowdy</td>
<td>36118</td>
<td>pCX4-KLF4</td>
<td>KLF4</td>
<td>H. sapiens</td>
<td>mono</td>
<td>n/d</td>
<td>n/d</td>
<td>n/c</td>
<td>–</td>
<td>–</td>
<td>22278060</td>
<td>Israel et al., 2012</td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell</td>
<td>Rudolf Jaenisch</td>
<td>20331</td>
<td>pFUW-K0</td>
<td>Klf4-F2A-Oct4</td>
<td>M. musculus</td>
<td>bi</td>
<td></td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>19109433</td>
<td>Carey et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell</td>
<td>Rudolf Jaenisch</td>
<td>20332</td>
<td>pFUW-KM</td>
<td>Klf4-F2A-cMyc</td>
<td>M. musculus</td>
<td>bi</td>
<td></td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>19109433</td>
<td>Carey et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell</td>
<td>James Thomson</td>
<td>21164</td>
<td>pSin4-CMV-K2M</td>
<td>Klf4, cMyc</td>
<td>H. sapiens</td>
<td>bi</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>19325077</td>
<td>Yu et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell</td>
<td>Shinya Yamanaka</td>
<td>27078</td>
<td>pCXLe-hSK</td>
<td>SDX2, KLF4</td>
<td>H. sapiens</td>
<td>bi</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>21460823</td>
<td>Okita et al., 2011</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell;</td>
<td>Shinya Yamanaka</td>
<td>41814</td>
<td>pCE-hSK</td>
<td>SDX2, KLF4</td>
<td>H. sapiens</td>
<td>bi</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>23193063</td>
<td>Okita et al., 2013</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>Klf4/KLF4</td>
<td>Shinya Yamanaka</td>
<td>19771</td>
<td>pCX-0KS-2A</td>
<td>Oct3/4-2A-Klf4-2A-Sox2</td>
<td>M. musculus</td>
<td>poly</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>○</td>
<td>18845712</td>
<td>Okita et al., 2008</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell;</td>
<td>Rudolf Jaenisch</td>
<td>20321</td>
<td>Tet0-FUW-OSKM</td>
<td>Oct3/4-Sox2-Klf4-cMyc</td>
<td>M. musculus</td>
<td>poly</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>○</td>
<td>19109433</td>
<td>Carey et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell;</td>
<td>Rudolf Jaenisch</td>
<td>20325</td>
<td>FUW-50KM</td>
<td>Sox2-P2A-Oct4-T2A-Klf4-E2A-cMyc</td>
<td>M. musculus</td>
<td>poly</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>○</td>
<td>19109433</td>
<td>Carey et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell;</td>
<td>Rudolf Jaenisch</td>
<td>20327</td>
<td>FUW-SOK</td>
<td>Sox2-F2A-Oct4-T2A-Klf4</td>
<td>M. musculus</td>
<td>poly</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>○</td>
<td>19109433</td>
<td>Carey et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell</td>
<td>Rudolf Jaenisch</td>
<td>20328</td>
<td>FUW-OSKM</td>
<td>Oct4-P2A-Sox2-T2A-Klf4-E2A-cMyc</td>
<td>M. musculus</td>
<td>poly</td>
<td></td>
<td>n/c</td>
<td>○</td>
<td>○</td>
<td>19109433</td>
<td>Carey et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene</td>
<td>stem cell</td>
<td>Keisuke Kaji</td>
<td>20865</td>
<td>pCAGMKOSIE</td>
<td>c-Myc-F2A-Klf4-T2A-Oct4-E2A-Sox2 (Alt name)</td>
<td>M. musculus</td>
<td>poly</td>
<td></td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>19252477</td>
<td>Kaji et al., 2009</td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Distribution</th>
<th>Addgene Collection</th>
<th>Depositing Scientist</th>
<th>Repository ID</th>
<th>Plasmid Name</th>
<th>Gene/Insert Name</th>
<th>Species</th>
<th>Type</th>
<th>AddGene</th>
<th>Depositor</th>
<th>Direct</th>
<th>PMID</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Addgene stem cell</td>
<td>Keisuke Kaji</td>
<td>20866</td>
<td>pCAG2LMK0Sim0</td>
<td>c-Myc-F2A-Klf4-T2A-Oct4-E2A-Sox2-ires-mOrange</td>
<td><em>M. musculus</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>19252477</td>
<td>Kaji et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene stem cell; Klf4/KLF4</td>
<td>James Thomson</td>
<td>20923</td>
<td>pEP4 E02S EM2K</td>
<td>Oct4 and Sox2; Myc and Klf4</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>○</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>19325077</td>
<td>Yu et al., 2009</td>
</tr>
<tr>
<td>Addgene stem cell</td>
<td>James Thomson</td>
<td>20924</td>
<td>pEP4 E02S CK2M EN2L</td>
<td>Oct4 and Sox2; Nanog and Lin28; Klf4 and Myc</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>○</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>19325077</td>
<td>Yu et al., 2009</td>
</tr>
<tr>
<td>Addgene stem cell</td>
<td>James Thomson</td>
<td>20925</td>
<td>pEP4 E02S EN2K</td>
<td>Oct4 and Sox2; Nanog and Klf4</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>19325077</td>
<td>Yu et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene stem cell; Klf4/KLF4</td>
<td>James Thomson</td>
<td>20927</td>
<td>pEP4 E02S ET2K</td>
<td>Oct4 and Sox2; SV40LT and Klf4</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>○</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>19325077</td>
<td>Yu et al., 2009</td>
</tr>
<tr>
<td>Addgene stem cell</td>
<td>Tim Townes</td>
<td>21627</td>
<td>pKP332 (Lenti-OSK1)</td>
<td>hOCT4 - 2A - hSOX2 - 2A - hKLF4</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>19415770</td>
<td>Chang et al., 2009</td>
<td></td>
</tr>
<tr>
<td>Addgene stem cell; Klf4/KLF4</td>
<td>Jose Cibelli</td>
<td>24603</td>
<td>OKSIM</td>
<td>Oct4 KLF4 Sox2 c-Myc</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>○</td>
<td>n/c</td>
<td>○</td>
<td>–</td>
<td>20030562</td>
<td>Ross et al., 2010</td>
</tr>
<tr>
<td>Addgene</td>
<td>Michel Sadelain</td>
<td>27512</td>
<td>pLM-fSV2A</td>
<td>OCT4_T2A_KLF4_ P2A_cMYC_ E2A_SOX2</td>
<td><em>H. sapiens</em></td>
<td>poly</td>
<td>–</td>
<td>○</td>
<td>–</td>
<td>21151124</td>
<td>Papapetrou et al., 2011</td>
<td></td>
</tr>
<tr>
<td>Addgene stem cell</td>
<td>Linzhao Cheng</td>
<td>28213</td>
<td>pEB-C5</td>
<td>Oct4, Sox2, Klf4, c-Myc, Lin28</td>
<td><em>M. musculus</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>21243013</td>
<td>Chou et al., 2011</td>
<td></td>
</tr>
<tr>
<td>Kotton Lab</td>
<td>Darrell Kotton, Gustavo Mostoslavsky</td>
<td>–</td>
<td>pHAGE-Tet-STEMCCA</td>
<td>Oct3/4-Klf4-ires- Sox2-cMyc</td>
<td><em>M. musculus</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>19096035</td>
<td>Sommer et al., 2009</td>
<td></td>
</tr>
<tr>
<td>WTSI</td>
<td>Alan Bradley</td>
<td>–</td>
<td>pPB-CAG.0SKM- puΔtk</td>
<td>0-T2A-S-T2A- K-F2A-M</td>
<td><em>M. musculus</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>19337237</td>
<td>Yusa et al., 2009</td>
<td></td>
</tr>
<tr>
<td>RIKEN BRC</td>
<td>Knut Woltjen</td>
<td>RDB13133</td>
<td>PB-TAC-OKMS</td>
<td>Oct3/4-Klf4- cMyc-Sox2</td>
<td><em>M. musculus</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>this study</td>
<td></td>
<td></td>
</tr>
<tr>
<td>RIKEN BRC</td>
<td>Knut Woltjen</td>
<td>RDB13132</td>
<td>PB-TAC-OSKM</td>
<td>Oct3/4-Sox2- Klf4-cMyc (from 20321)</td>
<td><em>M. musculus</em></td>
<td>poly</td>
<td>○</td>
<td>○</td>
<td>–</td>
<td>this study</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Source Distribution a</th>
<th>Addgene Collection b</th>
<th>Depositing Scientist</th>
<th>Repository ID</th>
<th>Plasmid Name</th>
<th>Gene/Insert Name</th>
<th>Species c</th>
<th>Type d</th>
<th>Klf4 e</th>
<th>Klf4 f</th>
<th>AddGene g</th>
<th>Depositor h</th>
<th>Direct i</th>
<th>PMID</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>RIKEN BRC</td>
<td>—</td>
<td>Knut Woltjen</td>
<td>RDB13135</td>
<td>PB-TAC-OK+9MS</td>
<td>Oct3/4-Klf4+Sox2</td>
<td>M. musculus</td>
<td>poly</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RIKEN BRC</td>
<td>—</td>
<td>Knut Woltjen</td>
<td>RDB13136</td>
<td>PB-TAC-STEMCCA</td>
<td>Oct3/4-Klf4-ires-Sox2-cMyc</td>
<td>M. musculus</td>
<td>poly</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RIKEN BRC</td>
<td>—</td>
<td>Knut Woltjen</td>
<td>RDB13134</td>
<td>PB-TAC-MKOS</td>
<td>c-Myc-F2A-Klf4-T2A-Oct4-E2A-Sox2</td>
<td>M. musculus</td>
<td>poly</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RIKEN BRC</td>
<td>—</td>
<td>Knut Woltjen</td>
<td>RDB13137</td>
<td>PB-TAC-EB-CS</td>
<td>Oct4, Sox2, Klf4, c-Myc, Lin28</td>
<td>M. musculus</td>
<td>poly</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>RIKEN BRC</td>
<td>—</td>
<td>Knut Woltjen</td>
<td>RDB13138</td>
<td>PB-TAC-WTSI</td>
<td>0-T2A-S-T2A-K-F2A-M</td>
<td>M. musculus</td>
<td>poly</td>
<td>—</td>
<td>—</td>
<td>0</td>
<td>0</td>
<td>—</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

WTSC, Wellcome Trust Sanger Institute; RIKEN BRC, RIKEN Bio Resource Center DNA Bank. Closed circles indicate the Klf4 variant, and open circles denote sequence validation steps for each construct. See also Figures 3 and S3.

aOnly vectors containing Klf4 are listed. Data from Addgene were current as of September 3, 2014. As Addgene is an independently curated and dynamically changing database, the authors do not warrant the accuracy or completeness of this resource beyond the date of data collection.


cCompared to mouse Klf4L, the human ORF is 87% and 90% similar at the nucleotide and amino acid levels, respectively, with 100% conservation across the first 35 amino acids. The effect of human Klf4S/L on reprogramming was not tested in this study.

d“Type” refers to the number of cistrons or ORFs in the construct.

eComparison of the ATG start site with U70662.1 (Klf4S) and U20344.1 (Klf4L) reference cDNAs. Note that for polycistronic vectors, this annotation does not take into account elongation by 2A peptides.

f“Form” refers to the number of cistrons or ORFs in the construct.

gValidation of Klf4 identity was from Addgene sequencing data. n/c, no coverage of the Klf4 start site, although available sequence confirmed the presence of Klf4.

hCassette sequences were verified empirically by the authors.
variants have seen unprejudiced use among vectors constructed for reprogramming (Table 1 and associated references). OKMS (Klf4S) presented low KLF4 levels, which were restored along with Klf4L-related phenotypes by elongation of the KLF4S N terminus by 9aa. Thus, OK+9MS excluded the influence of 2A peptide choice and factor order variation. Moreover, high KLF4-associated phenotypes were reproduced by supplementation of OKMS with either monocistronic Klf4S or Klf4L, limiting the effect to a polycistronic context. Relative stoichiometry was explored in a recent study, and it was found that chemically controlled degradation of KLF4 resulted in premature termination of reprogramming (Nishimura et al., 2014). Although precise factor titration and more direct biochemical assays may be required, our monocistronic reprogramming data argue against modification of an N-terminal functional domain, and rather support the idea that threshold levels of KLF4 are important for minimizing the occurrence of partial reprogramming.

Polycistronic cassettes encoding KLF4S tend to permit rapid expansion of transgenic cells with robust SSEA-1 activation. Yet, this majority population is ultimately impeded in the acquisition and stabilization of pluripotency, as indicated by transgene dependence and high-level expression of markers associated with pre-iPSCs (Mikkelsen et al., 2008; Theunissen et al., 2011). Polycistronic reprogramming with Klf4L or OKMS supplementation with Klf4L improved reporter activation and dox independence. Nanog-GFP reporter activation is consistent with a central role for KLF4 in the core ESC pluripotency network, as ectopic expression of Klf4 promotes epiblast stem cell conversion to ground-state pluripotency (Guo et al., 2009) and prevents loss of pluripotency in the absence of LIF (Niwa et al., 2009). Transgene silencing is considered a hallmark of complete reprogramming (Goli-pour et al., 2012). How the silencing phenomenon is related to the acquisition of pluripotency, and whether it is coordinated directly or indirectly through KLF4 are issues of considerable interest.

Individual expression of either Klf4S or Klf4L in MEFs led to a potent epidermal gene response (Figure S2B). Moreover, a similar response was noted as a common feature of d6 reprogramming populations induced with Klf4S cassettes (Figure 3B). A retrospective examination of early reprogramming experiments recalls reports of epithemis-related gene activation (Mikkelsen et al., 2008). Klf4 knockout mice die postnatally from failed epidermal stratification (Segre et al., 1999). Reciprocally, ectopic Klf4 expression during embryogenesis results in premature barrier formation (Jaubert et al., 2003). Therefore, it is pertinent that high KLF4 stoichiometry induces a cascade of epithelialization and expression of epithemis-associated genes such as Krt6a, Krt17, Sprt1a, Cnfn, and Tgm1 during early reprogramming (Figures 3A and 3B; Table S1). Interestingly, many of these genes are not expressed in ESCs, nor are they maintained in GFP+ iPSCs (data not shown), which brings into question their relevance in iPSC derivation. To address potential cellular heterogeneity, it will be important to link gene expression responses and reprogramming outcomes to specific subpopulations of cells. Thus, it remains to be determined whether the epidermal response elicited by excess KLF4 in the early phase directly contributes to high-fidelity reprogramming or to an alternate cell fate.

On the path toward induced pluripotency, investigators rely on gene expression patterns as milestones to gauge appropriate progression, and comparative analysis of publicly available gene expression profiles is a common approach. Clearly, our data show that inconsistencies in the relative levels of KLF4 to OCT3/4, SOX2, and c-MYC can fundamentally confound comparative analysis. Application of monocistronic Klf4S or Klf4L leads to nearly indistinguishable reprogramming phenotypes. Thus, typical clonal isolation of fully reprogrammed iPSCs with monocistronic vectors, which is subject to phenotypic selection for appropriate factor expression levels, should be mostly unaffected. Mechanistic studies, on the other hand, are prone to transcriptional noise arising from variable factor expression between cells and among transgenes. We expect the KLF4 threshold effects observed herein to be pronounced when the relative factor stoichiometry is fixed, for polycistronic transfection or even in secondary reprogramming systems with pre-integrated transgenes (Wernig et al., 2008; Woltjen et al., 2009). As such, we must be mindful that it is difficult to make direct comparisons across distinct reprogramming systems without first defining the inherent factor stoichiometry.

**EXPERIMENTAL PROCEDURES**

**Plasmid Construction**

The PB transponson-based PB-TAC expression vector (RDB13131) was generated by standard cloning (restriction digestion and ligation). PB-TAB vectors containing the four Yamanaka factors were generated by Gateway Cloning (Invitrogen) as described previously (Woltjen et al., 2009). Monocistronic and polycistronic reprogramming cassettes constructed in-house or obtained from public resources (Table 1) were cloned into pDONR221 using PCR and attB/P Clonase (Invitrogen) or pENTR2B using standard restriction endonuclease cloning or InFusion (Takara Clontech). Note that the cassette from pPB-CAG.OSKMBpuΔtk (Yusa et al., 2009) is in the order O-S-K-M, similarly to OSKM (Carey et al., 2009), and thus is referred to in the text as WTS1 to avoid confusion. Modifications such as N-terminal extension and HA tagging were produced using InFusion or the QS Site-Directed Mutagenesis Kit (New England Biolabs). Sequences of the main Gateway
Entry constructs (from *attL1* to *attL2*) were obtained using Sanger sequencing with Big Dye Terminator Ver.1 chemistry (ABI) and assembled as contigs using Sequencher (GeneCodes). Complete details regarding cloning procedures used, including InFusion amplification and sequencing primers, are available upon request.

**MEF Isolation**

Nanog-GFP (Okita et al., 2007) transgenic reporter mice were maintained as homozygotes on a C57BL/6 background. For ROSA-rtTA; Nanog-GFP double transgenic MEFs, C57BL/6 homozygous ROSA26-rtTA females (Ohnishi et al., 2014) and inbred homozygous Nanog-GFP males were crossed. Mice bearing ColIa1-targeted OSKM and OKMS cassettes were generated as described previously (Ohnishi et al., 2014). MEFs were isolated from 12.5–13.5 days postcoitum transgenic embryos. Embryos were decapitated, eviscerated, and macerated, and then dissociated with 0.25% trypsin/EDTA at 37°C for 15 min and plated in DMEM containing 10% fetal bovine serum (FBS), penicillin-streptomycin, and L-glutamine. The presence of the desired transgenes and male gender were determined by PCR genotyping of residual embryonic tissues. Cells from male embryos were expanded 1:5 once and then frozen in liquid nitrogen. Cells for reprogramming were defrosted, counted, and plated for transfection without additional expansion. The genotyping primers are available upon request. Animal care and experiments using animal tissues were approved by the CiRA Animal Experiment Committee in accordance with Kyoto University guidelines.

**PB Reprogramming and iPSC and ESC Culture**

MEFs were seeded in DMEM containing 10% FBS, penicillin-streptomycin, and L-glutamine on gelatinized 0.1% 6-well plates at a density of 1.0 × 10^5 cells per 10 cm². After a 24 hr culture, Fugene HD (Promega) was used to transfect cells with 500 ng of transposons (PB-TAG-OSKM, -OKMS, -OKc⁶MS, -MKOS, -STEMCCA, -EB-C5, -WTsi, -OKmAHMS, -OKmAHMS, and -lacZ) plus 1,000 ng of pCyl43 PB transposase plasmid at a Fugene/DNA ratio of 8 μL:2 μg. For top-up experiments, cells were co-transfected with 500 ng of PB-TAG-OKMS along with 500 ng of PB-TAG-OK-c⁶MS and -lacZ) plus 1,000 ng of pCyl43 PB transposase plasmid at a Fugene/DNA ratio of 8 μL:2 μg. For FACS analysis, 3.0 × 10⁵ cells were stained on ice for 30 min with Alexa Fluor 647 conjugated SSEA-1 (480, sc-21702, Santa Cruz Biotechnology) antibody or the corresponding isotype-matched control, mouse IgM (1:50 dilution). Control staining with the appropriate isotype-matched control monoclonal antibodies (BD Biosciences) was included to establish thresholds for positive staining. Samples were analyzed using a BD LSRFortessa Cell Analyzer (BD Biosciences) with BD FACS Diva software (BD Biosciences). Flow cytometry data were analyzed and generated by FlowJo software (Tree Star). For cell sorting, the mCherry⁺ cell population was collected on a BD FACSAria II cell sorter (BD Biosciences).

**Microarray Analysis**

Total RNA was prepared from harvested cells using the RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions. cDNA synthesis and transcriptional amplification were performed using 200 ng of total RNA with the Whole Transcript (WT) Expression Kit (Ambion/Affymetrix). Fragmented and biotin-labeled cDNA targets were hybridized to GeneChip Mouse Gene 1.0 ST arrays (Affymetrix) according to the manufacturer's protocols. Hybridized arrays were scanned using an Affymetrix GeneChip Scanner. Probe signal intensities were normalized with the RMA algorithm in GeneSpring, Quality, correlation, and cluster analyses were performed using GeneSpring GX software v12.1 (Agilent).
Technology). GO term analysis was performed using the NIH DAVID Bioinformatics tool (http://david.abcc.ncifcrf.gov/). There is no straightforward way to derive gene numbers directly from probe sets, as the Affymetrix GeneChip Array platform makes use of redundant probes for a small subset of genes. Moreover, each expressed sequence tag (EST) is represented by a unique probe, despite the fact that multiple ESTs may be associated with a single gene transcript. For clarity, we refer to probe sets or entities as “genes” in the text.

Protein Analysis
For western analysis, we prepared total cell lysates by boiling 1 × 10^6 cells for 10 min in SDS sample buffer (25 mM Tris (pH 6.8), 2% β-mercaptoethanol, 3% SDS, 0.05% bromophenol blue, and 5% glycerol). Lysates were resolved on SDS-polyacrylamide gels, and proteins were detected by western blotting using ECL

Material Distribution
The PB-TAC reprogramming transposons are available through the RIKEN BRC DNA Bank (http://dna.brc.riken.jp) under accession numbers RDB13131–RDB13138. Gateway versions of the polycistronic reprogramming cassettes will be made available through AddGene (http://www.addgene.org). The PB transposase vector pCyL43 is available through the Wellcome Trust Sanger Center (http://www.sanger.ac.uk/form/Sanger_CloneRequests).

ACCESSION NUMBERS
Datasets for the d6 gene expression microarray analysis have been deposited in the Gene Expression Omnibus under accession number GSE65468.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures, three figures, and one table and can be found with this article online at http://dx.doi.org/10.1016/j.stemcr.2015.02.004.

AUTHOR CONTRIBUTIONS

ACKNOWLEDGMENTS
We thank Kazutoshi Takahashi, Masamitsu Sone, and Ren Shimmoto for their critical readings of the manuscript, and Katsunori Semi for bioinformatics advice. We also thank Chiho Sakurai, Michiko Nakamura, and Akiko Oishi for technical support; Toshikko Sato for microarray preparations; Akito Tanaka for blastocyst injections; and Kanako Asano and Kotaro Ohnishi for mouse husbandry. We appreciate the advice of Hiroaki Kii (Nikon, Japan) regarding image acquisition and analysis. We also thank Darrell N. Kotton and Gustavo Mostoslavsky (Boston University School of Medicine) for providing pHAGE-Tet-STEMCCA, Alan Bradley and Kosuke Yusa (Wellcome Trust Sanger Institute) for providing pPB-CAG.OSTMBpMutk, Hitoshi Niwa (RIKEN Center for Developmental Biology) for providing the pLefty1-luc reporter, and all Addgene contributors, especially Linzhao Cheng and Rudolf Jaenisch. This work was supported by the Cabinet Office, Government of Japan and the Japan Society for the Promotion of Science (JSPS) through the Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program), the Core Center for iPS Cell Research, the Research Center Network for Realization of Regenerative Medicine, and the Strategic International Collaborative Research Program of the Japan Science and Technology Agency (JST). K.W. is a Hakubi Center Special Project Researcher. S.Y. is a non-salaried scientific advisor of iPS Academia Japan. S.L. received funding from EPASI (JSPS/National Science Foundation, 2013). S.-I.K. is a JSPS Fellowship recipient (2011-2013) and JST Researcher.

Received: August 8, 2014
Revised: February 6, 2015
Accepted: February 6, 2015
Published: March 12, 2015

REFERENCES


